THE ROLE OF NON-REPLICATIVE DNA POLYMERASES IN MICROSATELLITE MUTAGENESIS

A Dissertation in
Genetics
by
Kimberly Dawn Jacob

© 2009 Kimberly Dawn Jacob

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

December 2009
The dissertation of Kimberly Dawn Jacob was reviewed and approved* by the following:

Kristin A Eckert  
Professor of Pathology  
Dissertation Advisor  
Chair of Committee

Sarah K. Bronson  
Associate Professor of Cellular and Molecular Physiology

Ralph Keil  
Associate Professor Biochemistry and Molecular Biology

Lisa Shantz  
Associate Professor of Cellular and Molecular Physiology

Maria Baker  
Assistant Professor of Medicine and Pediatrics  
Genetics Counselor/Medical Geneticist

Gavin P. Robertson  
Associate Professor of Pharmacology, Pathology, and Dermatology  
Director Intercollegiate Graduate Degree Program in Genetics

*Signatures are on file in the Graduate School
ABSTRACT

Microsatellites, or short tandem repeats (STR), are sequences of 1-6 base pairs per unit repeat and are found throughout the eukaryotic genomes, and to a lesser extent in prokaryotic genomes. Mutations in these sequences are sources of genetic variation and one proposed mechanism to explain these mutations is the polymerase slippage model. Although polymerase utilization of slipped DNA intermediates is a requisite step in this model, the cellular polymerase(s) responsible for microsatellite mutagenesis are unknown. In this study, we have examined the role of 2 non-replicative polymerases in microsatellite mutagenesis, *Escherichia coli* (*E.coli*) DNA Polymerase IV (Pol IV) and human DNA Polymerase β (Pol β). We investigated the effect of the *E.coli dinB* gene product (Pol IV) on mononucleotide and dinucleotide repeat stability, using a Herpes Simplex Virus thymidine kinase (HSV-tk) gene episomal reporter system for microsatellite mutations. For the control vector (HSV-tk gene only) we observed a statistically significant 3.5-fold lower median mutation frequency in Pol IV- than Pol IV+ cells (p<0.001, Wilcoxon Mann Whitney Test). For vectors containing an in-frame mononucleotide allele ([G/C]₁₀) or either of two dinucleotide alleles ([GT/CA]₁₀ and [TC/AG]₁₁) we observed no statistically significant difference in the overall HSV-tk mutation frequency observed between Pol IV+ and Pol IV- strains. To determine if a mutational bias exists for mutations made by Pol IV, mutational spectra were generated for each STR vector and strain. No statistically significant differences between strains were observed for either the proportion of mutational events at the STR or STR specificity among the three vectors. However, the specificity of mutational events at the STR alleles in each strain varied in a statistically significant manner as a consequence of
microsatellite sequence. Our results indicate that while Pol IV contributes to spontaneous mutations within the HSV-tk coding sequence, Pol IV does not play a significant role in spontaneous mutagenesis at [G/C]₁₀, [GT/CA]₁₀, or [TC/AG]₁₁ microsatellite alleles. Our data demonstrate that in a wild-type genetic background, the major factor influencing microsatellite mutagenesis is the allelic sequence composition.

Additionally, we examined mononucleotide mutagenesis in human cells. Our focus for this study was Pol β. Using an in vitro HSV-tk assay, we determined the signature mutation for Pol β on G₁₀ and C₁₀ templates. We observed a greater than 10-fold increase in polymerase error frequency for both mononucleotide templates, compared to GT₁₀ and CA₁₀ templates. Analyses of the mutational specificity found that greater than 71% of the mutational events occurred at the STR in the mononucleotide allele; of these, ~50% are unit-based insertion and 50% unit-based deletion events. These data indicate that the signature mutation for Pol β within the [G]₁₀ and [C]₁₀ mononucleotide alleles is approximately equal proportions of unit-based insertion and deletion events.

Previous research has shown that when over-expressed, Pol β increases spontaneous mutagenesis in mammalian cells for both microsatellite and non-microsatellite containing targets. A limited number of studies have examined mutagenesis in the presence of decreased levels of Pol β, and these have reported varied results. The effect of lowered Pol β levels on microsatellite mutagenesis has not been examined. To evaluate this, we attempted to knock down Pol β levels in non-tumorigenic human lymphoblastoid cells and determine the effect on spontaneous microsatellite mutation rates. We utilized an ex vivo shuttle vector system to examine the clonal mutation rates of the [G/C]₁₀ microsatellite sequence. Mutational spectra were generated from randomly selected
mutant clones to define the location of mutations. In wild-type cells, the mutation rate of the control vector was $2.4 \times 10^{-5}$. The $[G/C]_{10}$ repeat allele was less genetically stable with a median mutation rate of $5.5 \times 10^{-5}$, but the difference is not statistically significant.

Clonal cells containing the $[G/C]_{10}$ repeat allele were infected with a lentiviral vector containing a control shRNA sequence (to luciferase) or one of three different shRNAs to Pol β. Both mRNA and Pol β protein levels, as well as mutation rates were analyzed for each shRNA construct. Both qRT-PCR analyses of the mRNA levels and western blot analyses of protein levels of Pol β found no reduction for the expression of Pol β, indicating that the knockdown was unsuccessful. Comparison of the clones containing the control shRNA and those with each of the three Pol β shRNA constructs yielded no significant differences in the mutation rates. A comparison of the clones that had been infected with lentivirus and selected with puromycin to those that were uninfected revealed a statistically significant 3-fold decrease in overall mutation rate ($p<0.0001$, Wilcoxon Mann Whitney Test). Additionally, we observed altered mutagenesis of the $[G/C]_{10}$ allele. In the uninfected cells we observed 36% of the mutations at the STR; however in a pooled group of infected and selected cells 77% of the mutations were at the STR, which is a statistically significant increase ($p<0.0001$, Fishers Exact Test). A final role of infection and selection was a statistically significant altered specificity at the microsatellite motif shifting from a bias toward insertions in uninfected cells to more of a balance in the infected and selected cells ($p=0.04$, Fishers Exact Test).
# Table of Contents

List of Figures ........................................................................................................................................... x

List of Tables ................................................................................................................................................. xii

Acknowledgements ........................................................................................................................................ xv

Chapter 1: Introduction ............................................................................................................................... 1

1.1 DNA Polymerases ........................................................................................................................................ 3
  1.1.1 *E. coli* ................................................................................................................................................. 6
    1.1.1.1 Types of Polymerases, Role in Cell Mutagenesis, and Their Regulation of Function ....................... 6
    1.1.1.2 The role of *dinB*/Pol IV in spontaneous mutagenesis ................................................................. 10
    1.1.1.3 Mutations in Response to Altered Pol IV Levels ........................................................................... 11
  1.1.2 Mammalian ........................................................................................................................................... 13
    1.1.2.1 DNA Polymerase Families and Their Roles in Cells ................................................................. 13
    1.1.2.2 DNA Polymerase β ...................................................................................................................... 18
    1.1.2.3 Effect of Altered Pol β Levels on Mutagenesis ............................................................................ 21
    1.1.2.4 Pol β and Disease ......................................................................................................................... 24

1.2 Microsatellites ............................................................................................................................................ 25
  1.2.1 Microsatellites .................................................................................................................................... 25
  1.2.2 Overview/general Background .......................................................................................................... 26
  1.2.3 Models for Microsatellite Mutagenesis ............................................................................................. 31
  1.2.4 Reporter Systems for Determining Microsatellite Mutation Rate ..................................................... 34
  1.2.5 Role of Mismatch Repair in Microsatellite Instability ....................................................................... 39

1.3 Rationale of Thesis .................................................................................................................................... 45

Chapter 2: Materials and Methods ........................................................................................................... 48

2.1 *Escherichia coli* Strains .......................................................................................................................... 48

2.2 Human Cell Lines and Media ................................................................................................................. 48

2.3 Genotyping FCT21 and FTK44 *E. coli* strains ....................................................................................... 49

2.4 Construction of Microsatellite Containing Vectors .............................................................................. 49

2.5 Construction of Lentiviral Vectors ......................................................................................................... 50

2.6 Analysis of Mutation Frequency and Mutation Rate in *E. coli* ............................................................. 54

2.7 HSV-tk Mutational Specificity Analysis .................................................................................................. 56

2.8 Electroporation and Cloning of LCL721 Cells ....................................................................................... 57

2.9 Shuttle Vector DNA Isolation from LCL721 Cells .................................................................................. 59

2.10 Determination of Mutation Frequency and Mutation Rate in LCL721 Cells .......................................... 60

2.11 Packaging and Titering the Luciferase Lentiviral Vector ...................................................................... 60

2.12 Infection of LCL721 Cells with Lentiviral Vectors ............................................................................. 61

2.13 Protein Isolation and Concentration Determination ............................................................................... 62

2.14 Western Blots ......................................................................................................................................... 62

2.15 Isolation of Genomic DNA .................................................................................................................... 63
2.16 PCR of Infected Clones to Amplify the shRNA Insert or the Puromycin Resistance Gene ......................................................... 64
2.17 Isolation of RNA from Human Cells ........................................... 64
2.18 RT-PCR for Expression of Pol β mRNA ....................................... 66
2.19 qRT-PCR for Pol β mRNA Levels .................................................. 66
2.20 Transient RNA Transfection of Human Cells with shRNA to Pol β-1 ...... 68
2.21 Duplexing of shRNA Oligonucleotides ........................................ 69
2.22 HSV-tk \textit{in vitro} Assay ............................................................. 69

Chapter 3: Role of \textit{E. coli} DNA polymerase IV in Spontaneous Microsatellite Mutagenesis (Jacob and Eckert, \textit{Mutation Research} 619:93-103) ......................... 73

3.1 Rationale/Hypothesis .................................................................. 73
3.2 Genotyping the \textit{dinB}+ and \textit{dinB}- Strains .................................... 73
3.3 Results ......................................................................................... 74
  3.3.1 Effect of Mismatch Repair on Microsatellite Mutagenesis .......... 74
  3.3.2. Effect of Pol IV on Spontaneous Mutation Frequency .............. 76
  3.3.3. Mutational Specificity of the [G/C]_{10} Microsatellite Allele .......... 80
  3.3.4. Mutational Specificity at [GT/CA]_{10} and [TC/AG]_{11} Microsatellite Alleles ................................................................. 86
  3.3.5. Microsatellite Length and Sequence Specificity ...................... 86
3.4 Discussion ................................................................................... 89
  3.4.1. Pol IV and HSV-tk Mutagenesis ............................................ 90
  3.4.2. Pol IV and Microsatellite Mutagenesis .................................. 91

Chapter 4: Spontaneous Mutagenesis of the [G/C]_{10} Mononucleotide Microsatellite ................................................................. 95

4.1 Rationale/Hypothesis .................................................................. 95
4.2 Results ......................................................................................... 96
  4.2.1 Shuttle Vector Assay with 721 Cells and [G/C]_{10} Shuttle Vector ..... 96
    4.2.1.1 Mutation Rate Analysis of Human Cell Clones ....................... 97
    4.2.1.2 Mutational Spectra Analysis of Human Cell Clones ............... 97
  4.2.2 Pol β \textit{in vitro} reactions ......................................................... 101
    4.2.2.1 G_{10} Template ................................................................. 101
    4.2.2.2 C_{10} Template ............................................................... 102
4.3 Discussion ................................................................................... 105
  4.3.1 HSV-tk Assay in Human Cells ................................................ 105
  4.3.2 HSV-tk \textit{in vitro} Assay ........................................................... 108
  4.3.3 Comparison Between Systems ................................................. 111

Chapter 5: Development of a System to Genetically Manipulate Polymerase Levels in Human Cells ............................................................ 113

5.1 Rationale/Hypothesis .................................................................. 113
5.2 Results .................................................................................................................. 114
5.2.1 Evaluation of Transfection Methods to Introduce Genes of Interest .. 114
  5.2.1.1 Electroporation .................................................................................. 114
  5.2.1.2 Lipofection ..................................................................................... 114
5.2.2 Evaluation of Viral Transduction Methods ............................................. 115
  5.2.2.1 Infection of LCL 721 Cells and Detection of GFP Positive Cells .................................................................................................................. 115
5.2.3 Determination of Polymerase Levels in LCL 721 Cells ......................... 117
5.2.4 Derivation of Putative Knockdown Clones ............................................ 120
  5.2.4.1 Control Containing shRNA Vector .................................................. 120
  5.2.4.2 Pol β shRNA Containing Vectors .................................................... 120
5.2.5 Characterization of Knockdown and Control Clones .......................... 121
  5.2.5.1 RNA levels ..................................................................................... 121
  5.2.5.2 Protein Levels ................................................................................. 125
  5.2.5.3 Analysis of gDNA for shRNA Insert and the Puromycin Resistance Gene ......................................................................................................... 125
5.2.6 Determination of the Effectiveness of Pol β-1 Knockdown Sequence .......................................................... 128
5.2.7 Shuttle Vector Assay with Cells Containing the [G/C]_{10} Vector and Infected with the Control shRNA ................................................................. 130
  5.2.7.1 Mutation Rate Analysis of Control shRNA Containing Clones ................................................................................................................................. 130
  5.2.7.2 Mutational Spectra Analysis of Control shRNA Containing Clones ................................................................................................................................. 130
5.2.8 Shuttle Vector Assay with Pol β shRNA Infected Cells and
  [G/C]_{10} Vector ............................................................................................. 133
  5.2.8.1 Mutation Rate Analysis of Pol β-1, Pol β-2, and Pol β-3
  Clones Containing the [G/C]_{10} Vector ...................................................... 133
  5.2.8.2 Mutational Spectra Analysis of Pol β-1, Pol β-2, and Pol β-3
  clones Containing the [G/C]_{10} Vector ...................................................... 136
  5.2.8.3 Analysis of Uninfected versus Lentiviral Infected and
  Selected Cells ............................................................................................ 138
5.3 Discussion ............................................................................................................ 143
  5.3.1 Development and Validation of the Lentiviral Infection System ...... 143
  5.3.2 Mutagenesis of the Control shRNA Vector ............................................. 145
  5.3.3 Mutagenesis of Pol β shRNA Containing Clones .................................. 147
  5.3.4 Mutational Analysis of Uninfected versus Infected Clones............. 148
Chapter 6: General Discussion and Future Directions ........................................ 152

6.1 What Was Learned from these Studies? ...................................................... 152
6.2 Mutational Balance .................................................................................. 154
6.3 Future Experimentation ............................................................................. 159
   6.3.1 Studies Involving *dinB* and Pol κ ....................................................... 159
   6.3.2 Studies Involving Pol β ...................................................................... 160
   6.3.3 Additional Studies ............................................................................. 162
6.4 Overall Significance ................................................................................ 162

References ..................................................................................................... 164
List of Figures

Figure 1.1: Replicative and Non-Replicative DNA Polymerases: 5
Figure 1.2: The Base Excision Repair Process: 20
Figure 2.1: Sequences Used to Create the Microsatellite Containing Vectors: 51
Figure 2.2: Microsatellite Containing Vector: 52
Figure 2.3: Sequences and Locations of Pol β shRNA Sequences: 53
Figure 2.4: E. coli Mutational Methods: 55
Figure 2.5: Mutagenesis Analyses in Human Cells: 58
Figure 2.6: PCR Primers to Determine Integration of shRNA Cassette and Puromycin Resistance Gene: 65
Figure 2.7: qRT-PCR Sample Calculations: 70
Figure 2.8: HSV-tk in vitro assay: 72
Figure 3.1: Effect of MMR on Mutation Frequencies of Microsatellite Containing Vectors: 75
Figure 3.2: Effect of Pol IV on Mutation Frequencies of Microsatellite-Containing Vectors: 79
Figure 3.3: Range of Unit Changes at the [G/C]_{10}, [GT/CA]_{10}, and [TC/AG]_{11} alleles: 83
Figure 4.1: Pol β Mutational Spectra on G_{10} Template: 103
Figure 4.2: Pol β Mutational Spectra on C_{10} Template: 104
Figure 5.1: GFP Expression in 293 Cells: 116
Figure 5.2: FACS Analysis of LCL721 Cells following Lentiviral Infection with the shRNA Vector Containing a GFP Reporter: 118
Figure 5.3: Western Blot of Endogenous Protein Levels: 119
Figure 5.4: RT-PCR Primers for Analysis of Pol β mRNA Levels: 122
Figure 5.5: qRT-PCR Results for Relative Pol β Expression in Luciferase Control Containing Cells: 123
Figure 5.6: qRT-PCR Results for Relative Pol β Expression in Pol β-1 shRNA Containing Cells
........................................................................................................................................ 124

Figure 5.7: Sample Western Blot of Pol β knockdown clones........................................ 126

Figure 5.8: Ratio of Pol β:PCNA Calculated from Western Blot Analysis................. 127

Figure 5.9: qRT-PCR Results for Relative Pol β Expression with Transient Transfection of Pol β-1 siRNA ........................................................................................................ 129

Figure 5.10: Lentiviral Infection Lowers the Mutation Rate of the HSV-tk Gene .... 139

Figure 6.1: The Effect of MMR on Various Microsatellite Templates....................... 156
List of Tables

Table 1.1: *E. coli, S. cerevisiae*, and Human DNA Polymerases and Families ........ 14

Table 1.2 Number of Repeat Loci/mb ≥ 20 Base Pairs of DNA Sequence in Eukaryotes .................................................................................................................. 28

Table 1.3 Number of Microsatellite Repeat Tracts Found Throughout the Entire *E. coli* Genome ........................................................................................................ 29

Table 1.4 Genes Containing Mononucleotide Microsatellite Motifs that are Commonly Mutated in MMR Deficient Human Cancers ........................................ 32

Table 3.1: Summary of the Effect of *dinB* on Spontaneous Mutagenesis ......... 78

Table 3.2: Mutational Events Within the [G/C]_{10} Vector in *dinB*⁺ and *dinB*⁻ Strains ................................................................................................................................. 81

Table 3.3: Mutational Spectra of the HSV-tk Coding Region in *dinB*⁺ Cells (FCT21) ...................................................................................................................... 84

Table 3.4: Mutational Spectra of the Coding Region of the HSV-tk Gene in *dinB*⁻ Cells (FTK44) ........................................................................................................ 85

Table 3.5: Mutational Analyses of [GT/CA]_{10} and [TC/AG]_{11} Vectors .......... 87

Table 4.1: Mutation Frequencies and Rates for HSV-tk Control Clones .......... 98

Table 4.2: Mutation Frequency and Rate for 721 Clones Containing the [G/C]_{10} Microsatellite Containing Shuttle Vector ...................................................... 99

Table 4.3: Mutational Specificity for LCL721 Clones Containing the [G/C]_{10} Microsatellite Repeat ............................................................................................... 100

Table 4.4: Microsatellite Mutation Rates for Human Cell Clones .................. 106

Table 4.5: Pol β Microsatellite Error Frequencies from *in vitro* Assay .......... 109

Table 5.1: Mutation Frequencies and Mutation Rates for Luciferase Control Clones ......................................................................................................................... 131

Table 5.2 Mutational Specificity of Luciferase Control Clones Containing [G/C]_{10} Allele .................................................................................................................... 132

Table 5.3: Mutation Frequencies and Mutation Rates for Pol β-1 shRNA Containing Clones ........................................................................................................... 134

Table 5.4: Mutation Frequencies and Mutation Rates for Pol β-2 and Pol β-3 shRNA Containing Clones ....................................................................................... 135
Table 5.5: Mutational Specificity for the Pol β shRNA Containing Clones .......... 137

Table 5.6 Proportion and Number of Mutational Events in Uninfected and Lentiviral Infected Human Cell Clones ................................................................. 141

Table 5.7 Proportion and Number of Mutational Events Occurring at the STR .... 142
List of Abbreviations

AAF - N-2-acetylaminofluorene
BER – base excision repair
bp – base pair
Cm – chlorephimical
DNA – deoxyribonucleic acid
E. coli – Escherichia coli
EBV – Epstein-Barr Virus
FUdR – 5’-fluor-2’-deoxyuridine
gDNA – genomic DNA
GFP – green fluorescent protein
HSV-tk – Herpes Simplex Virus thymidine kinase gene
IR – ionizing radiation
MEF – mouse embryo fibroblasts
MMR – mismatch repair system
PCR – polymerase chain reaction
Pol - polymerase
RNA – ribonucleic acid
ROS – reactive oxygen species
S. cerevisiae – Saccharomyse cerevisiae
shRNA – short hairpin RNA
siRNA – small interfering RNA
STR – short tandem repeat (microsatellite)
UV – ultraviolet radiation
Acknowledgements

To my parents: Thank you for putting up with me through all of these years of school, the ups and downs, the changed majors (and changed boyfriends), and all of the uncertainty. Thank you for all of your love and support, because without it I would not be where I am today. Thank you for listening to me cry on the phone when I wanted to quit, and then giving me the option to move back home if that was what I wanted to do (even though all the while you knew I would never leave). Your trips out to see me, even if it was just a stop for lunch, were the things that really helped get me through the past 6 years. Thank you.

To My brother: Jim, as weird as this is for me to say, thank you for being there. I know for the first 16 years of your life all we did was fight, hit, punch, bite, and scream (well I did more screaming than you), and I never would have thought that we would be where we are today. Thank you for all of the advice, car services, and phone calls. I would have never thought that I would be able to spend 2 hours on the phone with my brother and not want to kill you. 😊 Having you there as support has meant so much to me.

To my extended family: Thank you for all of your support and motivation to get me through all of my years of schooling. Your constant concern for me and “when I was going to finish” really pushed me through. There is now a Doctor (The Weiss side) or a second Doctor (the Jacob side) in the family!

To Kristin: I don’t know how I could have done this without you. I can’t imagine a better mentor for me or a better lab to work in. You have become like a second mom to me through my years in the lab. I could come into your office, ready to cry, and you were always able to calm me down and help me figure out what to do. You never pushed me to be something or someone that I wasn’t. Thank you for that. Without your support I would have never made it to graduation, let alone matured into the scientist, and person, that I am today. I can’t imagine being anywhere else but here, and as I move on I am sad to leave. Are you sure you (and the lab) can’t come with me?

To the Eckert Lab: I really don’t think I could have asked for a better group of people to work with through my time in Hershey!

Suzanne: without your help I would be nowhere. You are the one who could answer any question that I came to you with and helped me build the confidence in the lab to be able to move onto what’s next knowing that I will be able to understand how to work in (or run) the lab. Also, thank you for being my friend. There are so many days that I don’t think I could have gotten through without being able to talk to you. Thanks for introducing me to the fun of tailgating at Penn State football games. I’ll for sure be back someday!

Sandeep: my fellow student (at least for the longest time, the only other student), can you believe we are finally at the end? Thank you for putting up with sitting next to me for all these years, with me always asking questions of you. Without your answers and input, I would have never thought to do some experiments that were the coolest in the end. I wish you the best of luck as you move onto your next step as well.
Kathy: without you the lab truly could not run. I know you don’t understand the science that we do, or the importance of your role in the lab, believe me when I say that you are integral to everything that goes on. The days that you are in lab are the most entertaining to me. Your crazy stories about things in your life made my day (I just remember the story about the patio falling off the house while you were on vacation!).

Noelle: although you haven’t been around as long as some others in the lab I truly feel that you stepped up and fit right in. I am going to miss our morning talk time, which although it is not scheduled, happens everyday. Without your help and motivation, the final push to finishing up my thesis projects would have been so much more difficult. Having you there to tell me that I can do it, even though I felt completely defeated, is something that I will always be immensely thankful to you for.

Beverly: (wow is it weird to write your full name) even though you are the “newest” addition to our lab, I feel that you have contributed so much to my thesis and to my success. You have brought a new devotion to science to the lab, which I think I definitely needed. Thank you for all of your help in coming up with new ways to look at things, and thinking up new ways to interpret all of my “negative” data. I wish you the best of luck as you continue to finish up your Ph.D. I know that you will go great places once you are finished with the program.

To My friends: THANK YOU!! For those of us who have made it through Grad School together, I’m not sure how we did it at times. If it wasn’t for the support and friendship of each and every single one of you, I don’t think I would have made it through. And even if I did, it wouldn’t have been very fun!

To Michelle: Thanks you for all of your support, even though for a while there you wanted me to quit and move to DC. I’m glad you stuck with me through everything, and we made it to the end. No promises on where I will end up, but I’ll try for DC!

To Jesse, Craig, and Katie: Thank you for all of the tearful talks about science, boys, and life. Thanks for feeding me at least once a week and letting me just come over and ignore the adults and play with the baby. Without you both (and Katie too) I think I would have fallen apart a LONG time ago. Good luck to both of you as you finish up your graduate degrees soon.

To Meredith: I love you babe! Thank you for all of the phone calls, emails, hugs, and random dinner trips. Any time I was upset, lonely, sad, or happy I knew I could count on you to be there. We had some great times together, and got through a lot (stupid boys and bed bugs), and will have lots more fun experiences to come!

To Ange and Matt: I think living with you (both) was awesome. You got me through some tough times, but we had a lot of fun together too. I hope to move to some awesome city so I can come visit (and help Jess nanny your kids!).

To the Core Group (Jean, Jess, John, Brian, Ange, Matt): You all made Grad School fun! The parties every Monday first year were fantastic. Without all of you I would have never been exposed to Boone’s Farm, or to Boone’s pong. The random trips to Moe’s and Christmas Tree Shoppe ☺ were great distractions. Brian, your cooking, although spicy, was delicious…and you even got me to start eating spicy and strange foods (Indian, Thai, sushi)! Thanks again to all of you.
To Sarah: Thank you for all of your help. I would have never figured out qRT-PCR without your help, and I don’t think I would have my sanity from writing my thesis without you to talk to everyday. You were my writing buddy that helped me get through the tough times and keep me focused to get things done.

To Callie: I knew that anytime I was miserable to things had gone wrong in lab (or life) that I could come to you, and you would always listen. Thank you for always being there and caring enough to tell me what I needed to hear, even if it’s not what I wanted to hear.

To Jenn: Thank you for all of the trips you made out here to Hershey to see me, when you had school of your own going on. I truly miss our trips to Moe’s, Nine West, and Chocolate World. I’m so glad that I ended up getting you in the divorce! 😊

To Nate: Although we had the most random trip throughout the time we have known each other, you have always been there when I needed you. You knew I would accomplish great things even when I doubted myself. You kept me motivated to finish my thesis, even if it was by nagging. 😊 Thank you for everything.
Chapter 1: Introduction

Spontaneous mutagenesis of DNA has a fundamental role in evolution as well as being implicated in aging, carcinogenesis and human genetic diseases [1-10]. These mutations are thought to arise as a result of a number of intracellular events, such as DNA synthesis errors, movement of transposable elements, and formation of DNA lesions [11-20]. For the purpose of this thesis, a mutation is defined as “An active process during DNA synthesis where a polymerase incorrectly synthesizes the template strand, altering the DNA sequence”.

Spontaneous DNA lesions can be created in many ways, i.e. by different environmental factors. One mechanism shown to induce spontaneous mutagenesis is the activity of reactive oxygen species (ROS). ROS are produced as byproducts of cellular metabolism in aerobic organisms as well as through exposure to ionizing radiation (IR). When ROS are generated they can interact with DNA to produce damage [21]. There is evidence that oxidative damage may lead to deamination, spontaneous alkylation, as well as loss or modification of DNA bases [19, 20, 22, 23]. Oxidative damage is thought to contribute to disease in humans, including cancer and neurodegenerative diseases, as well as accelerating the aging process [19, 24-26]. An environmental cause of induced mutagenesis is ultraviolet radiation (UV). UV is capable of producing pyrimidine dimers, namely cyclobutane pyrimidine dimers and pyrimidine (6,4) pyrimidone photoproducts [27-32] mutations in the DNA that, after replication, produce C/G to T/A transition mutations. It has been shown that both UVA and UVB produce DNA products that have mutagenic and carcinogenic properties [33, 34]. UVB is capable of producing melanoma in a transgenic mouse model [35]. In addition, “UV-like” damage can occur
metabolically, producing spontaneous mutations [12]. In this situation, mutations are made that appear to be like those produced by UV radiation and they are dealt with via the same pathways that handle UV radiation errors.

Other sources of spontaneous mutagenesis in the genome are transposable elements and recombination [11, 18]. Transposable elements are DNA sequences capable of inserting randomly throughout the genome, resulting in large changes in the DNA sequence where they integrate. These elements move rather infrequently; however, this rate can be increased by the presence of spontaneous DNA damage [36-38]. Recombination mechanisms are not well understood in mammalian cells, but there are three proposed mechanisms that may induce mutations leading to recombination events: double strand DNA breaks, and stalled or collapsed replication forks (which are two independent events). A stalled replication fork arises to allow the recruitment of additional DNA polymerases to replicate through or bypass a DNA lesion. Studies have shown that low levels of stalling of the replication fork result in activation of DNA repair mechanisms and temporary cell cycle arrest, to allow synthesis to continue [39, 40]. A collapsed replication fork occurs after attempts to rescue a stalled replication fork fail, where the entire replication fork and required proteins completely cease DNA replication. It has been shown that recombination events increase spontaneous mutagenesis at collapsed replication forks induced by single-stranded DNA breaks (recombination is required to reinitiate replication from these structures) [41, 42], while stalled replication forks inhibit recombination [43]. Errors caused by recombination have been shown to lead to a pre-disposition for cancer and early aging [44, 45].
Errors occurring during DNA replication or DNA repair are a third way that spontaneous mutations occur. Much of the endogenous DNA damage endured in cells results in changes to or loss of DNA bases or mutations that can be generated during replication. The accuracy of DNA synthesis on undamaged templates is due to accurate insertions of nucleotides by DNA polymerases, as well as the exonuclease function of a polymerase and the post-replicative mismatch repair (MMR) system [14, 46, 47]. Many studies have been performed in *E. coli* trying to examine the role of these pathways in spontaneous mutagenesis. Studies have shown that in bacteria deficient in recA, the enzyme required for recombination and DNA repair, still exhibit 50% of the mutations observed in wild-type strains [48]. This indicates that the polymerase errors observed in *E. coli* cells can account for half of the overall spontaneous mutagenesis observed. Similar studies performed on bacteria deficient in different DNA repair genes have shown that the other ~50% of spontaneous mutations are due to error-prone DNA repair [reviewed in [11]].

1.1 DNA Polymerases

Polymerases are enzymes that are responsible for several DNA metabolic pathways, including replication, repair, and recombination. Comparison of DNA polymerases shows very little variation of catalytic subunit structure among species at the structural level. Crystal structures have been solved for at least one member of each polymerase family (described below) and it has been shown that each of them have catalytic domains that resemble a hand, consisting of a palm, finger, and thumb (reviewed in [49]). Of the three functions of DNA polymerases, replication is the one most
commonly associated with the term polymerase. Figure 1.1 shows the known roles of polymerases in DNA replication.

DNA replication occurs in a semi-conservative fashion, where each strand of DNA will serve as the template, resulting in the production of two new double stranded DNA molecules. Replication is initiated at an origin of replication, where a DNA helicase facilitates the unwinding of the DNA double-helix. In both *E. coli* and *S. cerevisiae* origins of replication are very well defined, however the process differs, as *E. coli* have one origin of replication per circular chromosome [50] while *S. cerevisiae* and other eukaryotes have multiple origins per chromosome [51]. *E. coli* replication begins at OriC, which is characterized by three 13bp and four 9bp AT repeats which are specifically recognized by initiation proteins. Once these proteins are bound to and wound with the DNA helix, the AT sequences are used to begin the unwinding of the DNA [52]. In *S. cerevisiae* DNA replication begins at sequences called ARS (Autonomously Replicating Sequences) element, which contain an 11bp ARS consensus sequence recognized by the Origin Recognition Complex (ORC) and responsible for recruitment of replication proteins and initiation of DNA unwinding [53, 54]. In humans and other Eukaryotes the sequence of the replication origins is less well defined. Despite the lack of a consensus sequence for DNA replication, the steps required for replication to occur are the same across eukaryotes. A protein complex forms, known as the pre-replication complex (pre-RC) and is bound to the DNA. The interaction of the pre-RC with additional proteins loads the Mini Chromosome Maintenance (MCM) complex. Once the MCM complex has been loaded onto the DNA, this signals that the required factors and complexes are prepared, and DNA replication can be initiated [55].
Figure 1.1: Replicative and Non-Replicative DNA Polymerases: Shown are the DNA polymerases with defined roles in DNA replication and repair. A.) shows the replicative polymerases and B.) shows the non-replicative polymerases.
Once the strands are unwound, helicase produces a structure called the replication fork, where copying of the two DNA strands occurs simultaneously, however in two different mechanisms. On the 5’ to 3’ strand, referred to as the leading strand, an initial RNA primer is laid down by primase to initiate replication, and then the DNA polymerase begins synthesis and continues the length of the template. On the 3’ to 5’ strand, referred to as the lagging strand, discontinuous replication occurs as DNA polymerases are unable to synthesize in this direction. On this strand, primase lays down many RNA primers at a set interval along the template strand. A DNA polymerase then synthesizes the segments between the RNA primers, which are called Okazaki fragments, in the 5’ to 3’ direction from the RNA primer. The RNA primers are then removed, and a DNA polymerase fills in the gaps. DNA ligase then seals the gaps between the nucleotides, completing the replication process.

1.1.1 *E. coli*

1.1.1.1 Types of Polymerases, Role in Cell Mutagenesis, and Their Regulation of Function

Replicative DNA Polymerases in Mutagenesis:

Of the 5 known DNA polymerases in *E. coli*, two are shown to be involved in replication of undamaged templates under normal cellular conditions: Polymerase I (Pol I) and Polymerase III (Pol III). Pol I, the first DNA polymerase discovered [56], is a processive enzyme consisting of 3 subunits: a polymerase subunit, a 3’ to 5’ exonuclease subunit, and a 5’ to 3’ exonuclease subunit [56-59]. The function of Pol I is to remove the RNA primer with its 5’ to 3’ exonuclease and fill in gaps left between Okazaki
fragments [60]. DNA Pol III is also a processive enzyme that contains a polymerase domain and a 3’ to 5’ exonuclease subunit [61, 62]. This polymerase is responsible for the main replication of DNA in *E. coli*.

Studies have been performed looking at the phenotype of cells deficient in these polymerases. Okazaki et al. found that a Pol I strain is viable. However, further investigation showed that the joining of Okazaki fragments was slowed to ~10% of the levels in wild-type cells [63], lending to the idea that Pol I is required for replication on the lagging strand. Temperature sensitive mutants of Pol I have been created for the polymerase subunit, the 5’ to 3’ exonuclease subunit, and the 3’ to 5’ exonuclease subunit [64, 65]. When either exonuclease mutant was transferred to the restrictive temperature, discontinuous replication is halted, resulting in death of the cell [64]. This indicates that both exonuclease activities are essential for discontinuous replication of Okazaki fragments. Similarly, temperature-sensitive mutants of Pol III were examined and result in lethality when moved to the restrictive temperature [66], indicating that the *dnaE* gene (encoding Pol III) is essential for DNA replication and well as cell survival.

**Non-replicative DNA Polymerases in Mutagenesis:**

The SOS system in *E. coli* allows the cells to cope with unfavorable and undesirable conditions resulting in DNA damage. There is a cost for these cells, however, and the result is increased targeted and untargeted mutagenesis (review see [67]). Of the five *E. coli* polymerases, there are three that are involved in and regulated by the SOS repair system in *E. coli*: Polymerase II (Pol II), Polymerase IV (Pol IV), and Polymerase V (Pol V) [68, 69]. All three of these SOS-induced polymerases have been
shown to be involved in induced mutagenesis in bypassing DNA damage [70]. In the study by Napolitano et al., a plasmid was constructed that contained different DNA adducts, N-2-acetylaminofluorene (AAF) guanine or benzo(a)pyrene adducts. Replication of the plasmid in the presence and absence of each of the three SOS-induced polymerases was completed and the difference in mutagenesis was determined. Pol II is able to bypass AAF lesions with a consequence of minus two base pair (-2 bp) frameshifts, while Pol V is capable of error-free bypass. For the benzo(a)pyrene adducts, error-free or minus one base pair (-1 bp) frameshift events are observed with both Pol IV and Pol V [70]. These results showed that each of the three polymerases contributes to mutagenesis of the template, with a different resulting error sequence.

Pol II is encoded by the polB (also called dinA) gene product and has been shown to function in replication restart (reinitiation of DNA replication after a stalling or stopping of replication due to encountering a DNA lesion or another insult to the DNA) on templates that have UV-irradiation damage [68, 71, 72]. Pol IV is encoded by the dinB gene product [73] and thus far has been elusive as to its exact function. umuC and umuD genes encode Pol V by forming the heterotrimer umuD’2C [69, 74, 75]. This polymerase has been shown to be involved in mutagenesis of UV-damaged templates with a preference to insert G opposite a 3’-thymine of a TT [69] and requires umuD, recA, and SSB for activation [76].

The three polymerases in E. coli, which are regulated by the SOS system, have different degrees of regulation. Pols II and V are very tightly regulated, with Pol II having between 30-50 molecules present per cell and Pol V having less than 20 molecules present in non-SOS induced conditions [60]. Pol IV, however, is different in
regulation, having ~250 molecules present under non-SOS conditions [77]. Upon upregulation of the SOS system, a 10-fold increase in the levels of each polymerase is observed. Studies have detected that approximately 210-350 molecules of Pol II, 2500 molecule of Pol IV, and 200 Molecules of Pol V are present when the cells have upregulated the SOS response [60, 77]. The abundance of Pol IV compared to the other SOS polymerases may be explained by sequence differences in the promoter region of the dinB gene. In each SOS polymerase, the promoter contains a 20-nucleotide binding sequence for the lexA repressor. In the dinB gene, this domain differs from consensus by 7 nucleotides [78]. In the other polymerases of the SOS system, the sequence is not greater than 2 nucleotides different from the consensus [79]. It has been hypothesized that the greater variance from consensus in the dinB sequence may account for a less stringent regulation by lexA under non-SOS conditions [77]. Based on these results, it has been speculated that Pol IV may more frequently bind at a stalled replication fork than the other SOS polymerases, making it a “default” bypass polymerase in E. coli [77].

Other controls of dinB gene expression have been reported. Layton and Foster performed a study where rpoS, the stress-response sigma factor, was deleted from bacterial cells. When they looked at levels of Pol IV, they found the levels of Pol IV as well as the amount of transcript to be reduced. They found that in wild-type cells, the levels of dinB were increased in late stationary phase, while in the rpoS mutant cells the levels of Pol IV were not induced [80]. They concluded that induction of Pol IV is part of a cellular response to starvation, due to the induction in late stationary phase, and other stresses.
1.1.1.2 The role of dinB/Pol IV in spontaneous mutagenesis

The dinB gene is the first of 4 open reading frames in an operon [81], which is controlled by lexA and upregulated under SOS conditions [82]. Pol IV is a template-directed polymerase but lacks the 3’ to 5’ exonuclease (proofreading) activity and has a propensity to extend misaligned or bulged templates [73]. Unlike other SOS polymerases, Pol IV does not require recA, umuDC, uvrA, polA or polB for activity and works independent of mutS functions [83, 84]. Depending on the E. coli strain used, Pol IV can be expressed from the bacterial chromosome, from the F’ plasmid, or from both locations. Kim et al. have shown that expression from the chromosome only yields 250 molecules, from only the F’ plasmid yields 750 molecules, and expression from both the chromosome and the F’ plasmid yields 1000 molecules of the polymerase [77]. These data indicate that the location of expression of the polymerase could have an effect on mutagenesis, where the cells expressing the most copies of Pol IV have the largest effect on mutagenesis [77]. Conversely, it is possible that mutations could have a significant effect on Pol IV levels, raising them to deal with the damage or mutagenesis, thus also adding to the mutagenesis observed.

An in vitro study by Kobayashi et al. contributed a great deal of information regarding the mutagenesis produced by Pol IV. An analysis was performed to determine the efficiency of Pol IV to replicate through a 407nt sequence in the lacZα gene. Using this forward assay, the mutation rate was determined to be 3.0x10^{-2}. The mutational spectra showed a predominance of frameshift events (~80%), the large majority of which were -1 bp deletion events in mononucleotide runs. Also noted, was the propensity for the -1 bp frameshift events to occur adjacent to a 5’ template G (~60% of the -1 bp
frameshift events observed). Base substitution events accounted for ~17% of the mutations observed. Of these mutational events, a large majority (~70%) occurred at a template T site in the sequence [85].

It has been shown that UV-irradiation or other DNA damaging treatments are not needed for Pol IV to make mutations when expressed from a low-copy plasmid [84] and it has been suggested that Pol IV is an auxiliary polymerase involved in spontaneous mutation [86]. Wagner and Nohmi have also shown that mutations made by Pol IV are correctable by the mismatch repair system signifying that they may be true replication errors made on undamaged templates [83].

1.1.1.3 Mutations in Response to Altered Pol IV Levels

A number of studies have examined the effect of deleting or reducing the levels of Pol IV on spontaneous mutagenesis. One study deleted the dinB operon from the chromosome of E. coli and used a reversion assay, with the lacZ gene out of frame, to show a statistically significant two-fold reduction in both -1 bp frameshift and single nucleotide base substitution events [86]. McKenzie et al. conducted a study looking at the effect of deletion of parts of the dinB operon or a substitution mutation in the dinB gene, which affects the polymerase active site, on spontaneous mutagenesis. A two-fold decrease in the lacZ reversion frequency was observed with deletion of the dinB operon, but no difference was detected with the substitution mutant. They concluded that a substitution mutation does not have an effect on spontaneous mutagenesis and that the other genes in the operon may have an effect on growth-dependent mutation [87]. Kuban et al. also used the deletion of part of the dinB operon to examine the effect on
mutagenesis on a chromosomal or episomal lacZ target. They detected no effect of loss of dinB on mutagenesis of the chromosomal target; however the mutagenesis on the episomal target was decreased two-fold [88]. The conclusion of the study was that the location of the gene target plays a role in the mutagenesis caused by Pol IV. This may be due to the fact that the target gene was different between the episomal target and the chromosomal target. It may be possible that the episomal target has a higher propensity to be mutated than the chromosomal target.

Similarly, studies have been performed looking at the effect of over-expression of Pol IV on spontaneous mutagenesis. Kim et al. showed that expression of additional copies of the dinB gene from a multicopy vector resulted in an increase of base substitution events, with a higher occurrence of transversion substitution events than transition events. A propensity for -1 bp frameshift events was observed, and the specificity of these -1 bp frameshifts appears to favor homopolymeric runs of G or C base pairs that are 6 nucleotides or longer in length [84]. Wagner and Nohmi compared mutagenesis caused by a vector carrying an additional copy of Pol IV to a control vector and observed an ~20-fold increase in the mutation frequency. The study also showed that over-expression caused an increase in -1 bp frameshifts as well as single nucleotide base substitution, with a preference toward G:C bases, and that frameshift events occur more often than substitution events with a ratio of 2:1 [83]. An additional study by Kim et al. compared the mutagenesis caused by Pol IV when expressed from the chromosome, the F’ episome, or both. They concluded that expression of Pol IV from the F’ episome had a greater effect on mutagenesis than expression only from the chromosome [77], which was supported by Kuban et. Al in a later study [88]. An additional study by Kuban et al.
looked at the effect of over-expression of Pol IV, by addition of a medium-copy plasmid containing the dinB gene, on the leading vs. lagging strand. A previous study by this group determined that lagging strand mutagenesis was more accurate for base substitution events. They found that with over-expression of Pol IV, leading strand replication was less accurate [89]. Uchida et al. performed a study to try and achieve a particular phenotype with over-expression of Pol IV. By expressing Pol IV from an arabinose promoter, they found that over-expression resulted in immediate halting of the replication fork and resulted in death of the cells [90]. They proposed that Pol IV may act as a brake to halt the progression of the replication fork, serving as a checkpoint to promote genome stability.

1.1.2 Mammalian

1.1.2.1 DNA Polymerase Families and Their Roles in Cells

In the mammalian genome there are 15 known DNA polymerase genes, which can be classified into 4 DNA polymerase families based on their sequence homology [91, 92]. The polymerase families are shown in Table 1.1 and briefly described below.

A Family:

The A family of polymerases consist of both nuclear and mitochondrial DNA polymerases and can be found across many organisms and species [91, 93]. Polymerase γ (Gamma) is involved in replication and repair of mitochondrial DNA and contains a 3’ to 5’ exonuclease activity [94]. Mice with a knock-in of exonuclease-deficient Pol γ exhibit signs of premature aging and have elevated mutation rates [95, 96]. DNA polymerase ν (Nu) is encoded by the POLN gene and is primarily found in the testis, as
Table 1.1: *E. coli, S. cerevisiae*, and Human DNA Polymerases and Families

<table>
<thead>
<tr>
<th></th>
<th><em>E. coli</em></th>
<th><em>S. cerevisiae</em></th>
<th>Human</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol I</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>A</td>
</tr>
<tr>
<td>Pol II</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>B</td>
</tr>
<tr>
<td>Pol III</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
</tr>
<tr>
<td>Pol IV</td>
<td>-</td>
<td>Pol κ</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Pol V</td>
<td>-</td>
<td>Pol β</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol α</td>
<td>Pol α</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol δ</td>
<td>Pol δ</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol ε</td>
<td>Pol ε</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol γ</td>
<td>Pol γ</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Pol IV</td>
<td>Pol IV</td>
<td>Pol λ</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Rev1</td>
<td>Rev1</td>
<td>Rev1</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Pol η</td>
<td>Pol η</td>
<td>Pol η</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Pol ζ</td>
<td>Pol ζ</td>
<td>Pol ζ</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol ι</td>
<td>Pol ι</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol μ</td>
<td>Pol μ</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol θ</td>
<td>Pol θ</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol υ</td>
<td>Pol υ</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pol σ</td>
<td>Pol σ</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
well as in skeletal muscle and the heart [93]. The transcripts contain a nuclear localization signal (NLS), which when Pol ν is tagged with GPF, shows a complete localization to the nucleus [97]. The function of this polymerase has not been determined. Polymerase θ (Theta) has both polymerase and ATPase activities and is homologous to a Drosophila polymerase thought to function in cross-link repair [98-100]. Pol θ has been shown to bypass basic sites and thymine glycol lesions with efficiency [101].

**B Family:**

The B family of DNA polymerases includes the replicative DNA polymerases and representatives have been found among all species. Polymerase α (Alpha) synthesizes a short RNA primer that is loaded onto both the leading and lagging strand of DNA, and creates the iDNA that is then extended by other polymerases [102, 103]. Polymerase δ (Delta) has a 3’ to 5’ exonuclease and a putative NLS signal [104, 105]. Studies in yeast have postulated that Pol δ functions in DNA replication on the lagging strand [106-109] and revealed that deletion of Pol δ results in inviability [110]. Polymerase ε (Epsilon) is another member of this family and is often thought of as the partner of Pol δ. Similar to Pol δ, deletion of Pol ε in yeast results in lethality [110]. A recent report revealed that Pol ε may be the polymerase that replicates the leading strand during DNA synthesis [111]. The last member of the B family is Polymerase ζ (Zeta). Pol ζ is a heterodimer that consists of 2 subunits, Rev3 and Rev7 [112-115]. Studies in yeast have shown that Rev3 and Rev7 are responsible for the majority of spontaneous and DNA damage
induced mutations. Pol ζ is able to bypass a variety of DNA lesions and extend many different mispaired templates [reviewed in [116]].

**Y Family:**

The Y family of DNA polymerases consists of four enzymes that are all very similar to each other, but share little homology to any of the other polymerase families. The Y family polymerases are thought to have a specialized role in cells to restart stalled replication forks [reviewed in ref [117, 118]]. DNA Polymerase η (Eta) is encoded by the POLH gene and although detectable in S-phase cells, the levels of Pol η increase after exposure to UV-irritation [119, 120]. Mutations in Pol η result in xeroderma pigmentosum-variant (XP-V) due to a defect in the replication of thymine-thymine dimers caused by UV exposure. In *S. cerevisiae*, neither mutations in or over-expression of Pol η results in an drastic change in the rate of mutagenesis [121-124]. DNA Polymerase κ (Kappa), is encoded by the POLK (hdinB) gene, which is the homolog of *E. coli* Pol IV [125]. This particular polymerase is present in many species; however, is absent from *S. cerevisiae* and *D. melanogaster* [126]. Pol κ is relatively accurate compared to the other Y family polymerases, inserting an incorrect base pair once every $10^2$ to $10^3$ base pairs [85]. A unique characteristic of Pol κ is the tendency to extend mispaired primer-termini [127] with greater propensity than creating the same mispairs. Pol κ has been also shown to form nuclear foci with DNA damage [120, 128].

Polymerase ι (iota) is encoded by the POLI gene. Similarly to Pol κ, Pol ι is found in many species, but is not a component of the *S. cerevisiae* or the *C. elegans* genomes [129]. Pol ι functions in the formation of T·dGMP mispairs [130-132] as well as the
formation of A·dTMP Hoogsteen base pairs [133]. This polymerase is also capable of completing BER in vitro, with the same accuracy as Pol β [134, 135]. A study has shown in vivo that human cells with a deletion of Pol τ have increased sensitivity to oxidative damage, suggesting a biological role for the polymerase in repair of oxidative damage [136]. The REV1 polymerase is highly similar to the yeast REV1 polymerase [137, 138] and expression of REV1 mRNA has been detected in all human tissues examined [137]. REV1 is capable of inserting dCMP opposite different lesions and these insertion products can be extended by Pol κ or Pol ζ [132, 139, 140]. It is also thought that in addition to this function, REV1 acts as a scaffold for regulation of other proteins during replication, namely the other TLS polymerases. It has been shown that REV1 has binding domains for Pol κ, Pol τ, Pol η, and Pol ζ [141-144], although the purpose or function of this has not been determined.

**X Family:**

The X family of DNA polymerases is an ancient group of enzymes that consists of proteins with a function in DNA repair. This family is found in organisms from archaea bacteria through eukaryotes; however only one member of this family is found in *S. cerevisiae*. The first characterized member of the family was Polymerase β (Beta). Pol β is encoded by the POLB gene and is the major polymerase that functions in base excision repair (BER) [145-148]. Many different proteins interact with Pol β in BER. Polymerase λ (Lambda) is highly similar to Pol β [149, 150]. Pol λ in yeast is Pol4, and has been shown to be involved in non-homologous end joining (NHEJ) [151, 152]. The same function is proposed for human Pol λ [153, 154]. The similarities between Pol β
and Pol λ suggest that Pol λ may also function in BER. An additional member of this family is Polymerase μ (Mu). The exact role of this polymerase has not been determined, but it appears to copy DNA in a very inaccurate manner and has also been shown to incorporate rNTPs [155]. A role for Pol μ in NHEJ has been shown [156]. The function of Polymerase σ (Sigma) is still controversial. It appears that Pol σ has polymerase activity and functions in chromosome condensation and sister chromatin cohesion [157-159]. However other reports indicate that this polymerase may function on polyA tails, not as a DNA polymerase [160, 161].

1.1.2.2 DNA Polymerase β

DNA Polymerase β is a 39-kDa protein encoded by the POLB gene [162, 163], and is located on chromosome 8p11.2 [164]. The protein is composed of two domains, a 31-kDa domain containing the polymerase active site and an 8-kDa domain that contains the dRP lyase active site [reviewed in [165]]. The 8-kDa domain is essential for the polymerase to bind DNA [166, 167]. Pol β has been shown to be present as a low abundance protein in the cell under normal conditions. However, upon treatment with alkylating agents and exposure to oxidative stress mRNA and protein levels of Pol β were found to be increased [168].

Pol β has been shown to function in both short-patch and long-patch BER [145-148]. In short patch BER, Pol β removes the damaged 5’ dRP moiety from the DNA and fills in the single nucleotide gap left from the removal process. In long-patch BER, Pol β begins DNA synthesis by incorporating the first base in the gap during the gap-filling process [169].
The BER process is depicted in Figure 1.2 and will be described briefly [for a review see [170]]. In short-patch BER, where a single damaged base is repaired, a DNA glycosylase breaks the N-glycosidic bond, leaving an Apurinic/Apyrimidinic (AP) site. AP endonuclease (APE1) can then recognize the AP site and nick the DNA on the 5’ side of the incorrect base, leaving a free 3’OH. DNA Polymerase β can then use its lyase subunit to remove the DNA backbone and extend from the 3’OH, replacing the incorrect base. DNA ligase, with scaffold protein XRCC1, seals the nick in the DNA ending the repair process. Long-patch BER corrects DNA damage that is between 2 and 10 base pairs in length and requires many different proteins. In this process, the DNA glycosylase and AP endonuclease activities occur in the same fashion. DNA Pol β, or Pol δ/ε, coupled with PCNA replicates the DNA where the damaged base is located, causing a flap to be produced containing the removed DNA bases. FEN1 then removes the flap of DNA and DNA ligase seals the nick that was made, restoring the DNA.

Many different proteins interact with Pol β during BER and other processes. Interaction with XRCC1 recruits XRCC1 scaffold to the DNA as well as preventing strand displacement by Pol β [171-173]. Ligase I also helps inhibit strand displacement by Pol β [174-176]. APE1 has been shown to stimulate the dRP lyase activity as well as possibly functioning as an exonuclease [177-179]. Interaction of the dRP lyase domain with P300 causes acetylation, resulting in a decrease of the lyase activity [180]. Pol β interacts with the proteins TRF2, the RAD9/RAD/HUS1 complex, and TP53 helps to stabilize/promote the synthesis of the polymerase activity [181-183]. The WRN protein interacts with Pol β and functions to promote strand displacement, most likely during the BER process [175, 176].
Figure 1.2: The Base Excision Repair Process: For full explanation, see text
Pol β is required for survival in mice. The studies by Sugo et al. and Gu et al. have both shown that the PolB-/- embryos die immediately after birth [184, 185]. Further studies have shown that the actual window for death in the PolB-/- mice is between day E10.5 and 4 weeks after birth [185]. Examination of tissues from the homozygous null mice revealed a lack of inflation of airways in the lungs, as well as significant post-mitotic neuronal cell death [185].

1.1.2.3 Effect of Altered Pol β Levels on Mutagenesis

Many studies have been performed that examined the effect of altered DNA Pol β levels in mice, mouse embryo fibroblasts (MEFs), or human cells. These studies have looked at both over- and under-expression of Pol β levels. Similar studies have also been performed on cells treated with DNA damaging agents.

A study by Bergoglio et al. used CHO cells which ectopically expressed rat Pol β, and showed that deregulation of Pol β induces chromosome instability and tumorigenesis. They found that the cells over-expressing Pol β were more likely to be aneuploid, with both gains and losses of chromosomes, as well as having a higher rate of abnormal mitoses [186]. An in vitro study, which added Pol β to human cell extracts, concluded that involvement of Pol β in DNA replication has mutagenic consequences. When Pol β was added to the extracts, replication fidelity was lowered and a three-fold increase in mutation frequency was observed [187]. Chan et al. examined cancer patient cells that were found to be naturally over-expressing Pol β and observed a nine-fold increase in mutation frequency compared to cells with normal levels of Pol β. They also added Pol β to wild-type cells and detected a six-fold increase in mutation frequency [188]. Mice that
over-express Pol β in the lens epithelium had an onset of cortical cataracts within 4 days of birth [189]. A final study by Yamada et al. examined the role of over-expression of Pol β on the mutagenesis of mononucleotide microsatellite repeats. This study utilized a reversion system to show that over-expression of Pol β increased the mutation rate at both an A₁₇ and G₁₇ microsatellite sequence, compared to a vector containing no repeat. In addition they tested different levels of over-expression to determine the effect on mutagenesis. They observed an increase in mutation rate with less than a 50-fold over-expression of Pol β, leading to the conclusion that there was no correlation between the level of Pol β over-expression and mutation rate [190]. In summary, these studies show that an increase in Pol β expression leads to increased abnormalities to the cells as well as increased mutagenesis at microsatellite alleles.

A few studies have looked at the effect of DNA damaging agents on mutation rate in cells that over-express Pol β. Much like the studies without DNA damage, increases in mutation rate were detected. Using CHO cells over-expressing rat Pol β, Canitrot et al. detected increased mutation rates and decreased sensitivity to anticancer drugs upon treatment with cisplatin, mechlorethamine, and melphalan [191]. Frechet et al. also used CHO cells lines over-expressing Pol β and treated them with H₂O₂. They found that the over-expressing cells were more sensitive to H₂O₂ treatment, exhibited apoptosis, and had a 50-fold increase in mutation rate at the HPRT locus compared to wild-type cells [192]. An additional study by the same group treated the same cells with IR. They found that over-expressing cells were more sensitive to IR and had a two-fold increase in apoptosis compared to wild-type cells. Mutagenesis at the HPRT locus was increased
4.5-fold with a 5Gy (Gy is defined as the absorption of 1 Joule of energy by 1 kg of
matter) dose of IR [193].

A very limited number of studies have looked at the effect of knockdown of Pol β
in the absence of exogenous DNA damage. Those that have examined mutation rates
have largely varied outcomes. Sobol et al. found that MEFs containing a knockout of Pol
β resulted in a 1.5-fold increase in the mutation rate at the λcII gene [194]. A study using
embryonic Pol β +/+ and Pol β -/- mice found a two-fold decreased mutation frequency
in the brain of Pol β -/- mice, but no difference in the liver of the same mice [195]. A
study by Poltoratsky et al. examined a Burkett’s lymphoma cell line naturally deficient in
Pol β. They showed that cells deficient in Pol β were still capable of BER at similar
levels as wild-type cells, but had a 7.5-fold increase in base substitution frequency than
either wild-type cells or null cells with ectopic expression of Pol β [196]. One final study
examined the effect of a Pol β +/- mouse line expressing 50-60% reduced Pol β levels.
They showed that the heterozygous mice had reduced BER capability and increased
single strand breaks and chromosomal aberrations compared to wild-type mice [197].

Many different studies have looked at the effect of DNA damaging agents on Pol
β deficient cells. These reports show that the Pol β null mouse cell lines are normal in
cell growth and viability, but are sensitive to DNA alkylating agents and H₂O₂, resulting
in increased apoptosis and chromosome breakage [148, 198-201]. Sobol et al. also
showed that after treatment with methylmethane sulfonate (MMS), wild-type cells are
better able to repair damage caused by the treatment than are knockdown cells, as the
wild-type cells have Pol β and fully functional BER. The Pol β null cells showed no
detectable reduction in cell damage due to MMS treatment after the repair period,
explained by their lack of functional BER [194]. A study by Polosina et al. used RNAi to
Pol β to knock down levels in MEFs. The results showed that both the RNAi knockdown
cells and the Pol β deficient cell line have the same decreased levels of survival when
treated with MMS, but no effect with treatment by other DNA damaging agents,
validating the RNAi approach to creating a knockdown of Pol β [202]. Lastly, Cebelof et
al. published a study looking at the effect of a Pol β heterozygous mouse line. They
measured the basal mutation frequency in wild-type and +/- mice and detected no
difference. When the cells from the heterozygous mice were treated with dimethyl
sulfonate (DMS), an increase in mutation frequency was detected compared to wild-type;
no effect was observed on the mutation frequency with N-ethyl-N-nitrosourea (ENU) or
UV treatment [197]. In summary, the studies looking at knockdown of Pol β levels agree
that treatment with alkylating agents decrease the viability of the cells. The data looking
at the effect on mutagenesis vary, leaving a gap in knowledge to be filled.

1.1.2.4 Pol β and Disease

Mutations in the PolB gene have been observed in a number of different cancers
and recently it has been estimated that these mutations occur in ~30% of 189 tumors
studied [reviewed in [203]]. Pol β levels have been shown to vary greatly among
different cancers and different tissues. High levels of Pol β protein have been detected in
breast, colon, and prostate cancer tissues [204]. When compared to adjacent normal
tissues, increased levels of Pol β mRNA and protein have been observed in uterus,
prostate, ovarian, breast, lung and stomach tissues [205]. It was also shown that along
with the increased levels of Pol β found in these tissues, increased levels of Pol λ and Pol ι were also detected.

The mutations observed in Pol β cause non-synonymous amino acid changes throughout the polymerase, in both the 31-kDa and 8-kDa domains. Gastric cancer tissues and gastric cancer cell lines often harbor mutations in the POLB gene, one is Y265C [206, 207]. This variant of Pol β has been shown to have a mutator phenotype and can induce transformation in mouse cells [207-210]. Two additional Pol β variants, K289M found in colon cancer and I260M found in prostate cancer, are also capable of producing cellular transformation [211-213]. Altogether, these results indicate that altered activity of Pol β has a mutagenic outcome that can contribute to carcinogenesis.

1.2 Microsatellites

1.2.1 Microsatellites

Much is known about the complexity of the genome, yet the complete sequencing of the human genome has greatly expanded our knowledge of the nature and location of various DNA elements. Within the intronic, exonic, and intergenic regions are areas of repetitive DNA. These repeated sequences account for ~50% of the sequence of the human genome [214]. There are three types of repeated elements found in the genome: satellite DNA (long repeat length), minisatellite (intermediate repeat length), and microsatellite (short repeat unit length).

Satellite DNA is characterized by tandem repeats, of several million base pairs in length. The repeats show a large amount of variation in size, with repeat lengths up to several hundred nucleotides in length. In humans, satellite DNA is often observed in centromeric repeats and may perform functional roles [215]. Mini- and microsatellite
DNA are composed of shorter tandem repeat units. The length of each repeat unit is the main factor that distinguishes between the two elements, with minisatellites ranging from 6 to 60 bp per unit and microsatellites ranging from 1 to 6 bp per unit [216]. Although satellite DNA has the longest repeat unit length, microsatellite DNAs are the most abundant repeat motifs in the genome. The plethora of microsatellite motifs within the genome makes them an interesting component of DNA to study from a DNA replication standpoint. The fidelity of polymerase replication through these repetitive sequences is essential for genome maintenance and stability.

1.2.2 Overview/general Background

Microsatellite sequences are repeats of 1-6 nucleotides per unit repeat and can be found throughout the entire genome [216-219]. Approximately 3% of the human genome is comprised of microsatellite sequences and more than 1 million microsatellite loci are found in the genome [214, 220]. These sequences can be found throughout the genome including coding, intronic, and untranslated regions [219, 221, 222]. However, studies have shown that the distribution of microsatellites within the genome is nonrandom and that the length of the allele can be dependent on the location of repeat [223]. Of all of the microsatellite motifs, mononucleotide motifs, particularly A/T tracts, are the most prevalent in the human genome [220, 224].

It has been shown that the distribution of microsatellites is fairly ubiquitous across each of the 22 autosomes as well as the X and Y chromosomes [220]. When looking specifically at mononucleotide repeats, once again each of the chromosomes had relatively equivalent numbers of repeats. When looking at the exonic density of
mononucleotide alleles, a varied result was observed. Chromosomes 7 and 16 had very high densities, while the Y chromosome had the lowest repeat allele frequency among all of the chromosomes [220]. Further analysis revealed that the density of these repeat alleles was slightly higher in the intronic regions than in intergenic regions across all chromosomes. When looking at the density of each type of mononucleotide sequence, it was observed that A/T sequences were much more prevalent than G/C sequences, consistently across all autosomes and the X and Y chromosomes [220]. The density of A/T mononucleotide microsatellites ranged from 3500 bp/MB of DNA, while the G/C motifs only reached a peak density of 100 bp/Mb. This results in an approximately 35-fold higher proportion of A/T mononucleotide microsatellites on each chromosome.

Although microsatellites are fairly well conserved across a broad taxonomic range, their genomic distribution differs radically. Characteristics such as microsatellite density per chromosome, average number of repeats in each microsatellite motif, as well as the different unit lengths of microsatellites varies greatly among different species [223, 225]. Across eukaryotes we see variation in microsatellites, with humans having the highest proportion of mono- and tetranucleotide microsatellite repeats less than 20 bp in length, while *Drosophila melanogaster* has the largest number of di- and trinucleotide repeats (Table 1.2) [223]. In contrast to this, prokaryotes, in particular *E. coli*, have a relatively low abundance of microsatellite alleles, the large majority of which are short in total length [226]. Table 1.3 shows the total number of microsatellite sequences of various lengths found across the *E. coli* genome and their distribution between coding and non-coding regions [227]. Compared to what is observed in eukaryotes, the
Table 1.2 Number of Repeat Loci/mb ≥ 20 Base Pairs of DNA Sequence in Eukaryotes

<table>
<thead>
<tr>
<th>Species</th>
<th>Mono</th>
<th>Di</th>
<th>Tri</th>
<th>Tetra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>183</td>
<td>93</td>
<td>32</td>
<td>136</td>
</tr>
<tr>
<td>D. Melanogaster</td>
<td>72</td>
<td>122</td>
<td>80</td>
<td>64</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>53</td>
<td>51</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>C. elegans</td>
<td>35</td>
<td>26</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>S. cerevisae</td>
<td>44</td>
<td>32</td>
<td>50</td>
<td>12</td>
</tr>
</tbody>
</table>

(data from [223])
Table 1.3  Number of Microsatellite Repeat Tracts Found Throughout the Entire *E. coli* Genome

<table>
<thead>
<tr>
<th># Repeat Tracts</th>
<th># Coding</th>
<th># Non-coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mono 3bp</td>
<td>163,354</td>
<td>127,407</td>
</tr>
<tr>
<td>4bp</td>
<td>42,901</td>
<td>31,777</td>
</tr>
<tr>
<td>5bp</td>
<td>13,837</td>
<td>9,830</td>
</tr>
<tr>
<td>6bp</td>
<td>4,123</td>
<td>2,696</td>
</tr>
<tr>
<td>7bp</td>
<td>1,000</td>
<td>544</td>
</tr>
<tr>
<td>8bp</td>
<td>217</td>
<td>92</td>
</tr>
<tr>
<td>9bp</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>10bp</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dinucleotide (&gt;6 bp)</td>
<td>7,575</td>
<td>5,779</td>
</tr>
<tr>
<td>Trinucleotide (&gt;9bp)</td>
<td>2,419</td>
<td>1,989</td>
</tr>
<tr>
<td>Tetranucleotide</td>
<td>52</td>
<td>41</td>
</tr>
</tbody>
</table>

(data from [227])
proportion of any microsatellite is dramatically lower in *E. coli*. The total density of microsatellites in human is estimated at 3.56 repeats/10 kb, while *E. coli* are estimated to have 0.01 repeats/10 kb [228].

Since uncovering the distribution of microsatellite alleles, they have been put to use for many different applications. These sequences are often used to answer questions regarding biological functions, such as those relating to personal identity, sex, paternity, epidemiology, and phylogenetics. Two examples of determining identity using microsatellites include the identification of a murder victim from skeletal remains [229] and the identification of a man who escaped to South America during WWII [230]. One of the most widely known cases of paternity testing using repeat alleles, was to determine if President Thomas Jefferson was the father of a child born to one of his slaves [231]. It was determined that he was. Due to their propensity to be polymorphic they are often used as genetic markers for population genetics, oncology, and forensics.

In the past decade, a relationship has been observed between the location and repeat length of microsatellites and human disease [232]. Polymorphisms at many mono-, di-, tri-, and tetranucleotide alleles have been suggested as being genetic risk factors for several diseases [233]. Two microsatellite alleles, a \([GT]_n\) and \([T]_n\), found in the CTFR gene affect the severity of the disease cystic fibrosis by altering gene splicing leading to altered gene expression [234, 235]. Polymorphisms in microsatellites associated with cancer genes (both tumor suppressor and oncogenes) have also been observed and may contribute to neoplastic progression (for a review see [236]). The length of a \([G]_n\) allele found in the EGFR gene is polymorphic, and has been shown to be inversely related to transcription of the gene [237]. This allele has also been shown to
have interracial length differences that are correlated with varying levels of EGFR in breast cancer [238]. Of great importance are a large number of genes containing mononucleotide microsatellite alleles, which have been found as commonly mutated in MMR deficient cancers [236] (a sample are listed in Table 1.4). As indicated by the list, the majority of the mononucleotide motifs ranges from 6-10 bp, indicating that shorter microsatellite repeats are targets for human cancers. This information provides great rationale for examining microsatellites that are shorter in length. Determining mutational mechanisms that may promote or alterations in the list of genes mutated in MMR deficient cancers as well as in other human diseases may help lead to new treatments and therapies.

1.2.3 Models for Microsatellite Mutagenesis

For an exonic sequence found in somatic cells, the estimated mutation rate is \( \sim 10^{-9} \) per locus per generation [6]. In contrast, microsatellite sequences have been shown to have much higher mutation rates ranging from \( 10^{-2} \) to \( 10^{-6} \) events per locus per generation [239-241]. There are a two proposed mechanisms to explain this instability in microsatellite sequences: unequal crossing over during replication and replication slippage.

Microsatellite mutations could be introduced through recombination events in the genome. It has been shown that recombination events lead to mutations in minisatellite DNA [242], but there is little data to show the same for microsatellite DNA. Microsatellite sequences are often associated with recombination hotspots, most likely due to the fact that the repetitive sequences are involved in recombination not because
Table 1.4 Genes Containing Mononucleotide Microsatellite Motifs that are Commonly Mutated in MMR Deficient Human Cancers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Repeat Motif</th>
<th>Cancer Type (% of Tumors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APAF-1</td>
<td>Proapoptotic Factor</td>
<td>A₈</td>
<td>C (8%), S (15%)</td>
</tr>
<tr>
<td>AXIN-2</td>
<td>Wnt Signaling</td>
<td>2-A₆, G₇, C₆</td>
<td>C (24%)</td>
</tr>
<tr>
<td>BAX</td>
<td>Proapoptotic Factor</td>
<td>G₈</td>
<td>C (45%), S (37%), E (33%), HNPCC (49%)</td>
</tr>
<tr>
<td>BCL-10</td>
<td>Proapoptotic Factor</td>
<td>A₈</td>
<td>C (13%), S (10%)</td>
</tr>
<tr>
<td>BLM</td>
<td>Response to DNA Damage</td>
<td>A₉</td>
<td>C (9%), S (25%)</td>
</tr>
<tr>
<td>Caspase-5</td>
<td>Proapoptotic Factor</td>
<td>A₁₀</td>
<td>C (48%), S (44%), E (28%)</td>
</tr>
<tr>
<td>CDX2</td>
<td>Transcription Factor</td>
<td>G₇</td>
<td>C (1%)</td>
</tr>
<tr>
<td>CHK1</td>
<td>Response to DNA Damage</td>
<td>A₉</td>
<td>C, (10%) E (10%)</td>
</tr>
<tr>
<td>IGFIIR</td>
<td>Growth Factor Receptor</td>
<td>G₈</td>
<td>C, (17%) S (21%), E (13%), HNPCC (7%)</td>
</tr>
<tr>
<td>MLH3</td>
<td>MMR</td>
<td>A₉</td>
<td>C (9%), HNPCC (15%)</td>
</tr>
<tr>
<td>MSH3</td>
<td>MMR</td>
<td>A₈</td>
<td>C (38%), S (39%), E (25%), HNPCC (51%)</td>
</tr>
<tr>
<td>MSH6</td>
<td>MMR</td>
<td>C₈</td>
<td>C (22%), S 38%), E (11%), HNPCC (24%)</td>
</tr>
<tr>
<td>PTEN</td>
<td>Cell Cycle</td>
<td>2-A₆</td>
<td>C (18%), E (9%)</td>
</tr>
<tr>
<td>RAD-50</td>
<td>Response to DNA Damage</td>
<td>A₉</td>
<td>C (28%), S (28%)</td>
</tr>
<tr>
<td>TGFβRII</td>
<td>Tumor Suppressor</td>
<td>A₁₀</td>
<td>C (82%)</td>
</tr>
</tbody>
</table>

C = Colon  
S = Stomach  
E = Endometrial  
(data from )
they are a consequence of it [243]. Another piece of evidence to indicate that recombination is not the likely cause of microsatellite mutagenesis looks at mutation rates between autosomes and the Y chromosome. Studies have shown that similar mutation rates were detected between autosomal and Y-linked markers [244, 245]. It has been shown that mutations arising during meiosis occur 6- to 20-times more frequently than those that occur during mitosis [246]. Given that fact that the Y chromosome does not participate in meiotic recombination which has higher levels of mutagenesis, and that the mutations rates are similar to that of autosomes, provides some evidence that meiotic recombination is not the main source of microsatellite mutagenesis. Furthermore, it has been shown that 63% of mutations observed on dinucleotide repeat tracts are multi-step events, with one occurrence requiring 5 steps to result in the mutation [247]. This provides additional evidence that recombination events are not the main mechanism to explain microsatellite mutations. A final piece of evidence to support the hypothesis that recombination is not the main for microsatellite errors is that correlation of recombination rate to microsatellite mutation rate or microsatellite density have found no significant correlation [247, 248].

The second mechanism proposed to explain microsatellite mutations is the replication-slippage model [249-251]. In this scenario, the DNA strands become dissociated at the replication fork, resulting in a region of non-paired DNA. In the case of a microsatellite repeat, when the strands reanneal they may misalign while synthesis by the DNA polymerase continues. The misalignment of DNA bases will create a loop out, which when not corrected by the postreplicative MMR system will alter the repeat tract length. With continued DNA synthesis and the absence of repair, a loop that occurs
on the template strand will result in the contraction of the repeat by the number of bases in the loop. If the loop occurs in the nascent strand, then the result would be an expansion of the number of bases in the loop.

There are two main ways that cells can deal with polymerase slippage to correct the misaligned frameshift intermediates: 3’ to 5’ exonuclease activity of the polymerase and MMR. For short STR runs, both of the mechanisms work well. However, in longer tracts, MMR becomes the favored mechanism [252, 253] because proofreading ability decreases as the length of the repeat increases [253-255].

1.2.4 Reporter Systems for Determining Microsatellite Mutation Rate

There are a few well-established experimental systems for examining microsatellite mutagenesis. Three eukaryotic systems in particular, one in S. cerevisiae and two in mammalian cell culture, are predominately used to determine the effect of DNA replication on the stability of microsatellite alleles. A predecessor to the eukaryotic systems was developed in E. coli. All of these systems will be described below.

The E. coli system for looking at microsatellite mutagenesis was developed and described by Levinson and Gutman. In this system, a [CA/GT]_{20} sequence was inserted in frame into the lacZ gene using BamHI restriction sites, resulting in a 48-bp insertion. Two control vectors were also created following the same procedure, one containing an 84-bp sequence with no repeat longer than 6 bp in length and a second, which was just the vector itself. By plating on selective media containing β-galactosidase (β-gal), mutation events that disrupt the functioning of the lacZ gene will result in white plaques. These mutation events may be in the artificial repeat or in the coding region of the lacZ gene. The percentage of mutations that were not in the repeat allele were determined by
plating with the control vectors [256]. Using this assay, the authors found that the frequency of mutation at the [CA/GT]_{20} allele was 50 to 150-times higher than the control vectors. By DNA sequencing they observed that 95% of the mutational events were single unit changes at the repeat allele and that -2 bp events outnumbered +2 bp events by a ratio of 5:1 [256].

The first eukaryotic system to be described was developed by the Petes lab and is used in *S. cerevisiae*. Two different assay schemes, one a forward assay and one a reversion assay are used. Both assays are related to those developed by Levinson and Gutman [257]. In the reversion system, a yeast LEU2 promoter is fused to the β-gal gene and a polylinker; the LEU2 promoter can function in both *S. cerevisiae* and *E. coli*. The microsatellite alleles are cloned out of frame into the BamHI site in the polylinker sequence. Mutants are scored by plating on bromo-chloro-indolyl-galactopyranoside (X-gal) plates. Colonies that do not contain a mutation, hence leaving the β-gal gene inactivated, will appear white on the plate. Those that contain a mutation that puts the β-gal gene back in frame will produce a blue colony [98]. The forward assay system has a fusion of the LEU2 promoter, a small portion of the β-gal gene, a small portion of HIS4, and the URA3 gene. In this plasmid, the microsatellite allele is inserted in frame, keeping the URA3 gene intact. If a plasmid were to have an alteration which left the URA3 gene out of frame, the cells would become Ura⁻, and can be selected on plates containing 5-fluoroorotic acid (5-FOA) [98].

Henderson and Petes examined the stability of [GT/CA]_{14} and [G/C]_{20}. With the reversion system they found that the mutation rate for the [GT/CA]_{14} allele was 1.5x10⁻⁴. The [G/C]_{20} allele was more unstable than the dinucleotide allele, with a mutation rate of
2.9x10^{-4}, and sequencing results showed that majority of the mutations observed were gains or losses of 1 or 2 bases. Using the forward system they tested the [GT/CA]_{16} allele and found that the mutation rate was 7.4x10^{-5}, a similar value to those achieved with the reversion assay, showing that both assays are capable of producing comparable results [98].

Many other studies have been performed using these assay systems all producing similar results: different microsatellite alleles have differing mutation rates. Wierdl et al. used the forward assay to examine a [GT/CA] tract of various lengths to determine how the mutation rate would change [258]. They observed that a 15-bp repeat had a mutation rate of 3.2x10^{-7}; however the 105-bp repeat had a mutation rate of 1.7x10^{-4}. This repeat had 500-fold higher mutation rate than the 15-bp repeat, showing that as the length of a repeat increase, so does the mutagenesis of the allele. This study also showed that as the repeats increase in length, the mutation events shift from a balance of insertion and deletion events, to a predominance of only deletion events [258]. One additional study looked at the effect of the presence of an interrupting sequence in a microsatellite allele. They found that the interrupted alleles, regardless of the interrupting sequence, helped to stabilize the microsatellite tract compared to those with no interruption. The mutation rate for the uninterrupted allele was 2.3x10^{-5}, while the interrupted alleles had an average mutation rate of 4.5x10^{-6} [259]. These authors suggest that this stabilizing effect is due to the requirement for a perfect realignment of the interrupting sequence during replication.

A similar system in S. cerevisiae, is used by the Jinks-Robertson lab. Two of their studies have contributed information about the four different mononucleotide
microsatellite alleles. Gragg et al. looked at each of the alleles in a wild-type strain. They found that the mutagenesis at the G and C alleles was nearly the same, and were ~100-fold higher than the A and T alleles [260]. Harfe and Jinks-Robertson looked at the four different alleles in different locations within the LYS2 gene each with different sequence contexts. They found that among the different locations the alleles behaved differently. In all four of the locations tested, the G/C allele was the most highly mutable. The same was also true when looking at the sequence context. Regardless of the context, the G/C allele was the most mutable [261].

Another reported system to study microsatellite mutagenesis is performed in mammalian cell culture by the Farber lab. This microsatellite reporter system is a reversion assay with 2 vector constructs, one in the +1 frame and one in the -1 frame. In both cases the plasmid contains a fusion gene of the neomycin resistance (Neo) gene and the HSV-tk gene, which is randomly integrated in the cells, rendering each cell with a different site of integration. The microsatellite alleles are inserted out of frame, either in the +1 or -1 orientation, into the 3’ end of the HSV-tk gene or the 5’ end of the Neo gene. The function of the Neo gene will be restored if a mutation occurs such that the reading frame of the gene is re-established [262, 263]. A major drawback with this assays system is the inability to detect all types of mutational events. Only events that restore the reading frame can be detected, leaving the effect of other classes of mutational events unaccounted for.

The first study looked at the effect of mutagenesis on a [CA/GT]_{17} alleles in a mouse cell line. They determined that the vector containing the repeat allele had a mutation rate 100-fold greater than compared to a 4 bp control. Sequencing analysis
showed that all mutants were \(-2\) bp deletions [262]. Lee et al. examined the difference in mutagenesis between a \([\text{CA}/\text{GT}]_{17}\) and a \([\text{GAAA}/\text{CTTT}]_{17}\) sequence in mouse cells. The results indicated that the mutation rate of the dinucleotide repeat was 5-times higher than the tetranucleotide repeat [264]. Also performed in mouse cells, Yamada et al. showed that longer alleles, \([\text{CA}/\text{GT}]_{17}\) or \([\text{CA}/\text{GT}]_{30}\), had 10-fold or 100-fold higher mutation rate than a \([\text{CA}/\text{GT}]_{8}\) allele respectively [263]. Mononucleotide alleles, \(\text{G}_{17}\) and \(\text{A}_{17}\), were examined and compared in mouse cell culture. Mutation rate analyses showed that the \([\text{G}/\text{C}]_{17}\) allele had a significant 17-fold higher mutation rate than the \([\text{A}/\text{T}]_{17}\) allele, and that the \([\text{A}/\text{T}]_{17}\) repeat had a similar mutation rate to that observed for the \([\text{CA}/\text{GT}]_{17}\) tract [265].

The human cell culture system used by the Eckert lab implements an oriP containing shuttle vector that is episomally maintained in Epstein-Barr Virus (EBV) transformed lymphoblastoid cells and contains origins of replication for both human and E. coli cells. The HSV-tk gene acts as the reporter gene for this forward assay and microsatellite alleles are inserted in frame into the 5’ region of the gene. Inactivating mutations in the microsatellite allele or the HSV-tk gene are detected by selective plating with 5-fluoro-2’-deoxyuridine (FUdR) [266]. Using this system, Hile et al. examined the mutagenesis of \([\text{TC}/\text{AG}]\) and \([\text{GT}/\text{CA}]\) alleles in a non-tumorigenic human lymphoblastoid cell line [266]. They determined that the mutation frequency for the control vector, containing only the HSV-tk gene, was \(1.8 \times 10^{-5}\). For insertion of a \([\text{TC}/\text{AG}]_{11}\) allele, the mutation frequency was \(22.5 \times 10^{-5}\), which resulted in an \(~18\) fold increase in mutation frequency was observed compared to a vector containing just the HSV-tk gene alone. As the length of the allele was raised to 17
and 20 units the mutation frequency was elevated 36- and 111-fold respectively. In contrast, the overall mutation frequency of the [GT/CA]$_{10}$ was elevated only ~9-fold and the mutation frequency at the [GT/CA]$_{10}$ microsatellite was 4.1x10$^{-5}$. When taken together, there was an ~2-fold increase in the microsatellite mutation rate for the [GT/CA]$_{10}$ allele, but a 12.5-fold increase for the [TC/AG]$_{11}$ allele, indicating that the later sequence is more mutable in this system [266].

Additionally, Eckert et al. used this same system to look at the microsatellite mutagenesis of [TTCC/AAGG]$_9$, [TTTC/AAAG]$_9$, or [TCTA/AGAT]$_9$ sequences. The mutation frequency for each of these repeats was significantly elevated compared to the control vector containing the HSV-tk gene alone (3.6x10$^{-6}$). However, the [TTCC/AAGG]$_9$ allele had a lower mutation frequency than did [TTTC/AAAG]$_9$, or [TCTA/AGAT]$_9$. This could be due to the difference in G+C content of the repeats, as the allele with the lower mutation frequency has a higher G+C content. Also included in this study was a [TC/AG]$_{11}$ allele. Analysis of this sequence showed a mutation frequency that was not significantly different from the [TTCC/AAGG]$_9$ allele, indicating that these two alleles have similar rates of stability [267]. The conclusion of this study was that biochemical factors, other than repeat unit length, contribute to microsatellite stability.

### 1.2.5 Role of Mismatch Repair in Microsatellite Instability

DNA replication is an error-prone process. As the new DNA strand is synthesized, it is possible for the DNA polymerases to mis-incorporate a nucleotide. When this occurs, mispaired base pairs must be corrected in order to reduce genetic
instability and promote survival. The DNA repair system capable of correcting these errors is the MMR system. The MMR proteins act to differentiate between the template and daughter strands, in order to make sure the newly inserted base is the one corrected, then excises bases including the incorrect base. Polymerases then replicate the newly excised DNA, incorporating the correct base for the DNA template. For a review of MMR see [268].

The best-characterized MMR system is that of MutHLS system in E. coli. During DNA replication, the newly synthesized strand remains unmethylated, allowing for the repair proteins to differentiate the parent and nacent DNA strands [269]. The MutS protein forms a homodimer (MutS$_2$) that can recognize and bind to the site of a mismatched base pair [270, 271]. MutH will bind to each of the hemimethylated sites on the daughter strand, but remains inactivated [272]. The last component, MutL, also forms a homodimer, which interacts with both MutS$_2$ and MutH, acting as a mediator between the two complexes and activating MutH [273-277]. Once the complexes are formed and in place, activated MutH nicks the newly synthesized strand at a GATC sequence near the hemimethylated site and MutL recruits DNA helicase II to unwind the DNA [278, 279]. The MutHLS complex moves along the DNA toward the site of the mispair, removing the bases as it goes with an exonuclease, which is selected based on the proximity of the incised DNA to the location of the mispaired base [280, 281]. Once the bases have been excised, DNA Polymerase III can come in and resynthesize the missing DNA bases [282].

Studies have been performed showing that deletion of each of these components results in inactive methyl-directed MMR and that cells lacking the DNA methylase have
increased mutation rates [283-285]. Lu et al. performed experiments using E. coli cell-free extracts, showing that strains deficient in mutS, mutL, mutH, and DNA helicase II are all inefficient at MMR [286]. A similar result was also shown in vivo for strains deficient in mutS, mutH, and mutL. Cells lacking these components of the MMR system were completely incapable of repairing mismatched bases via the MMR pathway [287].

The eukaryotic MMR system has been extensively studied. Two different homologues to the MutS complex exist, MutSα (Msh2/Msh6 protein complex) [288] and MutSβ (Msh2/Msh3 protein complex) [288]. MutSα functions in the repair of base substitutions and small loops (less than 10 bp) [289, 290] while MutSβ is involved in repair of both small and large loops (larger than 10 bp) [290, 291]. The MutL homologue has 3 different complexes: MutLα (MLH1/PMS2), MutLβ (MLH1/PMS1), and MutLγ (MLH1/MLH3). MutLα functions as the coordinator in human cells, while the functions of MutLβ and MutLγ are less understood [292-294]. There is no mutH homologue in eukaryotes. Replication in the eukaryotic system does not leave the DNA hemimethylated at adenine residues, thus the exact method of discrimination between template and daughter strands in eukaryotes remains to be determined. One early idea was that hemimethylation at cytosine residues could possibly be involved in strand discrimination [295-297], but this has since been shown not to be sufficient for strand discrimination [298, 299]. It is also though that strand discontinuities which occur as intermediates during DNA replication may provide signals which are similar to those present in the E. coli system [300]. Another idea is that PCNA, the replication clamp, interacts with proteins used for both replication and repair, might provide a physical link between DNA replication and repair where the termini at the replication fork would act as
the discriminating signal [301]. Yet another possibility is that signals from proteins that associate with individual DNA strands could be involved in a mechanism to establish parental and nascent strands [300].

Many studies have been performed looking at the effect of a defective MMR system on mutagenesis of microsatellite alleles. Using the system described previously, the Petes lab has contributed much information about the effect of MMR in *S. cerevisiae*. By deleting three different MMR genes in yeast strains, Strand *et al.* showed that the mutation rate was increased 100-700-fold, depending on the gene deleted [255]. Wierdl *et al.* examined the effect of the length of a microsatellite allele on mutagenesis in MMR− strains. They observed the same effect on mutation rate in MMR-deficient strains as they did in wild-type strains, with longer repeats being more unstable than short repeats [258]. This indicates that the events that occur are not due to a lack of MMR, but arise by another mechanism. Similarly, it was found that there was no effect on the mutagenesis of interrupted dinucleotide alleles with the loss of MMR [259]. This suggests that MMR is not required for the stabilization of interrupted microsatellite alleles. The effect of repeat size on microsatellite instability was examined using repeats of 1, 2, 4, 5, 8, and 20 bp per unit repeat. The results showed that in MMR− cells all of the repeat tract lengths were more unstable than the 20 bp tract, when compared to the mutagenesis in wild-type cells. This study also showed that in MMR− cells a [G/C]$_{18}$ motif is much more unstable than a [GT/CA]$_{16}$ repeat [302]. Hawk *et al.* looked at the effect of a [GT/CA]$_{16}$ repeat allele that was integrated at various location in the yeast genome. In wild-type cells they found a 16-fold variation among the sites examined. However, in MMR− cells they saw an increased mutation rate but the variation among the sites examined was gone [303].
They concluded that the site of integration of the alleles must effect the efficiency of MMR. One additional study looked at the mutation of PCNA (the clamp required for DNA synthesis and repair) and the effect on mutagenesis. The results showed a large increase in the mutation rate in the PCNA- S. cerevisiae strains, which was similar to but significantly greater than the effect seen in MMR- strains. They also noted that the most striking increase in mutation rate was observed in the mono- and dinucleotide alleles [304].

A study by Greene and Jinks-Robertson examined the mutagenesis of a 150-bp sequence found in the LYS2 gene. They observed that in a MMR- strain the mutation rate was increased between 200-250-fold. Also, it was noted that almost 100% of the mutational events were -1 bp frameshifts that occurred at a run of six As or four Cs [305]. Gragg et al. examined the effect of a MMR deficiency on each of the 4 mononucleotide alleles. Similar to wild-type cells, they found that the [G/C]_{10} and [C/G]_{10} alleles were ~100-fold more unstable than the [A/T]_{10} and [T/A]_{10} alleles. They also showed that in MMR- cells fewer mutations are observed for the [G/C]_{10} and [C/G]_{10} allele compared to those observed in wild-type cells. This would indicate that the MMR system is better able to repair mutations that arise in [A/T]_{10} and [T/A]_{10} tracts than those that occur in [G/C]_{10} and [C/G]_{10} tracts. The final finding was that MMR is better able to remove -1 bp frameshift events than +1 bp frameshift events in all four of the mononucleotide runs [260].

Studies have also looked at the effect of MMR in mammalian cells using the reversion assay system described above. In MMR- human cells from a cancer patient a [CA/GT]_{17} dinucleotide allele was 100-fold more unstable compared to wild-type human
cells, and all mutations in the MMR\textsuperscript{−} line were -1 bp frameshifts [306]. Yamada et al. examined the effect of microsatellite length on mutagenesis on H6 colorectal cancer human cells, which are MMR\textsuperscript{−}, using both the +1 and -1 reading frame vectors. They found that as the length of the allele increase, so does the mutation rate. The results also indicated that on short alleles both +1 and -1 events were observed, however as the length increase a bias toward +1 frameshifts was observed in MMR\textsuperscript{−} cells [263]. Boyer et al. used a vector in the -1 reading frame to show that in MMR\textsuperscript{−} human H6 cells, [G/C]\textsubscript{17} and [CA/GT]\textsubscript{17} alleles showed a similar increase in mutation rate compared to an [A/T]\textsubscript{17} allele [265]. This result together with the results from the wild-type cells suggests that a [G/C]\textsubscript{17} allele is more genetically unstable than a [A/T]\textsubscript{17} allele. The role of interruptions in microsatellite alleles was examined in MMR\textsuperscript{−} H6 human cells by Boyer et al. In this study, [G/C]\textsubscript{17} and [A/T]\textsubscript{17} alleles were compared to the same alleles interrupted to produce 2-8 bp tracts. They found that the interrupted alleles were more stable than the full length alleles, with the G tracts being more unstable than the A tracts [307].

A study by Tobi et al. looked at the effect of MMR on an interrupted [G/C]\textsubscript{29} microsatellite allele. They observed a lower mutation frequency for the plasmid only vector than for the [G/C]\textsubscript{29} allele vector, in both wild-type and MMR\textsuperscript{−} cells. A difference was observed in the types of mutations that occurred between the two cell lines. In the wild-type cells, only 18% of the mutational events were single base deletions, however in the MMR\textsuperscript{−} cells, 48% of the events were single base deletions [308]. A study performed by Tran et al. performed in yeast tested the effect of MMR on mononucleotide A alleles of various lengths. They observed that in MMR\textsuperscript{−} strains the mutation rate increased exponentially as the length increased from 4-bp to 14-bp. As the length approached 9-bp
and increased to 14-bp, a 1,700-21,000-fold increase in mutation rate was observed compared to the 4-bp length [309]. These results indicate that MMR can correct errors even as the length of an allele increases.

A study by Shah and Eckert used the system developed by the Eckert lab to examine the mutagenesis at [TTTC/AAAG]$_9$ and [TTCC/AAGG]$_9$ alleles in a lymphoblastoid cell line deficient in the MMR protein PMS2. They observed a 6-fold and 12-fold increase in mutation rate respectively in the PMS2 deficient cells compared to wild-type lymphoblastoid cells. In the absence of PMS2, a bias toward insertion events was observed which was equal in both alleles. However, in the presence of PMS2, a 29-fold decrease in mutation rate was observed for the [TTTC/AAAG]$_9$, but only a 6-fold reduction was observed for the [TTCC/AAGG]$_9$ allele [310]. They conclude that PMS2 is more protective toward tetranucleotide expansion events and that a bias is observed where the [TTCC/AAGG]$_9$ allele is stabilized better than the [TTTC/AAAG]$_9$ allele.

1.3 Rationale of Thesis

Given the knowledge that microsatellite alleles are found throughout the human genome and are postulated to play a role in many cellular functions, it seems imperative that the integrity of these sequences be maintained. Many studies have looked at the mutagenesis of dinucleotide and trinucleotide alleles as well as a few studies of tetranucleotide alleles. It is ironic however, that a very limited number of studies have been completed focusing on the mononucleotide alleles given that they are the most abundant in the genome.
In addition to studying the mutagenesis of mononucleotide alleles, it would be very beneficial to the mutagenesis field if the source of the errors could be determined. The best candidate for making errors in the repeat tracts would be DNA polymerases. Many studies have looked at the role of the replicative polymerases (Pols α, δ, and ε) in mutagenesis and tried to determine a role or lack thereof in microsatellite mutagenesis. Unfortunately, cells deficient in these polymerases are inviable, so a complete knockdown or making null cells is not possible. However, the non-replicative polymerases may play a role in microsatellite mutagenesis and are testable using knockdown or knockout systems.

The goal of this project was to determine the role of non-replicative DNA polymerases on the mutagenesis of both [A/T] and [G/C] mononucleotide microsatellite alleles. For this study we chose to examine the role of *E. coli* DNA polymerase IV and human DNA polymerase β. Pol IV was chosen, as it was shown in the literature that overexpression of Pol IV increased the proportion of frameshift events, particularly in mononucleotide G/C sequences [84]. When Pol IV levels were reduced this effect was reversed. These data support the hypothesis that the mononucleotide motifs would show a reduction of mutation rate in the absence of Pol IV. Pol β was chosen to knockdown in human cells culture due to its function in BER. It has been shown that cells in culture are under oxidative stress [311] and that the BER pathway is the main mechanism to handle oxidative damage (for a review see [24]). We therefore hypothesized that knockdown of Pol β under these conditions would have an adverse effect on microsatellite mutagenesis.

To determine a role for these polymerases in microsatellite mutagenesis, cell lines were created that lack one of the non-replicative polymerases, either *E. coli* DNA
Polymerase IV or human Polymerase β. A vector containing the microsatellite allele was delivered to these strains and cell lines and the mutation frequency and rate as well as the type of mutagenesis was detected. The outcome of these experiments will inform us regarding the role of each of these polymerases in mutation at two different mononucleotide allele
Chapter 2: Materials and Methods

2.1 Escherichia coli Strains:

Strain FT334 is derived from E. coli K12 strain HB101 [312] with the genotype:  
\( tdk, upp, thi1, hsd20, supE44, lacY1, proA2, ara14, galK2, xyl5, mtl1, leuB6, rpsL20, \)  
\( recA13 \). Strain PP102 is mismatch repair deficient and is isogenic to strain FT334 with exception of the following alleles:  
\( recA306, srl::Tn10, \)  
\( mutL::Tn5[313] \).  

E. coli strains SMR 4562 and SMR 5830 were provided by Dr. Susan Rosenberg, Baylor College of Medicine. SMR 4562 is an independent isolate of FC40 [314]. SMR 5830 is isogenic to SMR4562, and is  
\( dinB10 \) [F’\( dinB10 \)] [315]. Spontaneous 5-flourouracil-resistant, 5-fluoro-2’deoxythymidine (F UdR)-resistant isolates of both strains were isolated by selective plating as previously described [316]. The usual mechanism for F UdR-resistance has been shown to be mutations at the  
\( tdk \) locus [317]. The  
\( dinB^+, tdk^- \) isolate obtained from SMR 4562 was designated FCT21; the  
\( dinB10 \) [F’\( dinB10 \)],  
\( tdk^- \) isolate obtained from SMR 5830 was designated FTK44.

2.2 Human Cell Lines and Media:

The 721 cell line is a lymphoblastoid cell line from a healthy female donor [318]. The cells were chosen as they are non-tumorigenic (by soft agar assay) and are EBV transformed, thus allowing for episomal replication. Culture media used is RPMI (Gibco) supplemented with 2.4% Hepes Buffer, 3% NaOCH$_3$, 10% fetal bovine serum (Hyclone, Fisher) and 50 μg/ml Gentamicin (Invitrogen). When cells are under selection the media is also supplemented with 150 μg/ml hygromycin B (Amersco) or 0.2 μg/ml puromycin (Clonetech) The 293TN cell line is a human kidney tumor cell line that
is transformed with SV40 large T-antigen and is neomycin resistant (Purchased from Systembio Science). Culture media used is DMEM (Gibco) supplemented with 2.4% Hepes Buffer, 3% NaOCH₃, 10% fetal bovine serum (Hyclone, Fisher) 50 µg/ml Gentamycin (Invitrogen), and 500 µg/ml Geneticin (G418) (Gibco).

2.3 Genotyping FCT21 and FTK44 E. coli strains:

A stab from a glycerol stock of each cell line was placed into a 10 ml overnight culture containing LB medium. The following day, this culture was diluted 1:100 into a 500 ml overnight culture containing LB. This culture was then used to isolate genomic DNA (gDNA) from each E. coli cell line following the company protocol for a Qiagen Tip-20 extraction. The gDNA was then subjected to PCR analysis using Pfu Ultra DNA Polymerase (Stratagene), and adding an additional 4mM MgSO₄ as well as DMSO to the reactions. The sequence of the primers used is as follows: ATGCCTTTTGTTTCATTCATGTGG (forward) and AGATAAGCCTCATCCAGTGG (reverse). The amplified products were then sequenced using the CEQ8000 (Beckman Coulter) following the company protocol.

2.4 Construction of Microsatellite Containing Vectors:

Vectors containing the [A/T]₁₁, [G/C]₁₀, and [C/G]₁₀ microsatellite sequences were constructed by oligonucleotide site-directed mutagenesis of plasmid pGTK₄ [protocol described in detail in [319]]. The pGTK₄ plasmid is identical to the pGTK₂ plasmid [320], but has a G→T base substitution at position 195 of the HSV-tk gene. The purpose of this base substitution is to eliminate a potential internal start site for translation.
of the HSV-TK protein. The construction of the plasmids was completed as follows. A DNA oligonucleotide (Integrated DNA Technologies (IDT)) was designed which contained the desired artificial microsatellite repeat, inserted in frame into the 5’ region of the HSV-tk gene. For the \([A/T]_{11}\) repeat an \(A_{10}\) sequence was inserted between bases 110 and 111 in the HSV-tk sequence. For both the \([G/C]_{10}\) and \([C/G]_{10}\) repeats the \(G_{10}\) or \(C_{10}\) sequence was inserted between bases 111 and 112. The sequence of each allele is shown in figure 2.1. Each oligonucleotide was hybridized to single stranded DNA (ssDNA), and the complementary strand was generated using T7 DNA polymerase (Invitrogen). The resulting double-stranded DNA (dsDNA) was digested with \(Bgl\)II and \(Bss\)HII (Invitrogen Life Technologies) and ligated to the pJY1 shuttle vector [266]. The ligated vectors were transformed by heat shock into \(E. coli\) strain FT334 (strain described in section 2.1). Transformants were isolated and assayed by \(Ava/Bgl\)II restriction enzyme digest and DNA sequence analysis as described. The transformants with the correct HSV-tk sequence and digestion pattern were then plated to determine both the HSV-tk\(^-\) and HSV-tk\(^+\) mutation frequency. HSV-tk\(^-\) mutants are selected in the presence of 40 mM FUdR, which is a pyrimidine analogue that will select for a nonfunctional HSV-tk gene. Trimethoprim, a folate analogue, is used to select for the HSV-tk\(^+\) phenotype [316]. After phenotypic plating, one clone from each of the three allele containing vectors constructed was selected as the vector that will be used for future studies. A diagram of the plasmid is shown in Figure 2.2.

2.5 Construction of Lentiviral Vectors:

Pol \(\beta\) shRNA containing vectors (Figure 2.3) were constructed and packaged virus was a generous gift from Dr. Robert Sobol (University of Pittsburgh). The
A.) [A/T]_{11} Allele:
5’-GCCATCAACACGCCTCTGCGTTACGGCTGCGCCTTTTTTTTTTTTCTCGAGGCCATAGCAACCGACGTACGGCGTTGCGC -3’

B.) [G/C]_{10} Allele:
5’-GGCCATCAACACGCCTCTGCGTTACGGCTGCGCGGGGGGGGGGGGTCTCGAGGCCATAGCAACCGACGTACGGCGTTGCGC -3’

C.) [C/G]_{10} Allele:
5’-GCCATCAACACGCCTCTGCGTTACGGCTGCGCCCCCCCCCCCCGTCTCGAGGCCATAGCAACCGACGTACGGCGTTGCGC -3’

**Figure 2.1: Sequences Used to Create the Microsatellite Containing Vectors:**
Oligonucleotides, with the sequences listed, were used *In vitro* to generate microsatellite containing vectors. The repeated sequence is indicated in red.
Figure 2.2: Microsatellite Containing Vector: This shuttle vector is designed for replication and selection in both human cells, oriP for replication and hph for antibiotic resistance, and *E. coli*, ColE1 for replication and cat for antibiotic resistance. The HSV-tk gene is the mutational target and inserted in frame into the gene are artificial microsatellite sequences (indicated by STR in red).
Figure 2.3: Sequences and Locations of Pol β shRNA Sequences: Three different Pol β shRNA sequences were designed from the Pol β mRNA. The shRNA in the 3’UTR is designated Pol β-1, the one to its left is Pol β-2, and the most 3’ shRNA is Pol β-3 [321].
luciferase shRNA vector (the control vector) was constructed using the pSIF-H1-Puro construction Kit from System Biosciences, following manufacturers protocol. Briefly, the luciferase insert provided by the company was ligated into the pSIF-H1 vector using T4 ligase (Invitrogen). The ligated vector was transformed into DH5α E. coli cells. T5rrrPlasmid DNA was isolated from selected colonies following the protocol from the Qiagen Mini-prep Kit (Qiagen). Sequencing reactions of the plasmid DNA were prepared according to Beckman Coulter (Fullerton, CA) protocol using the pSIF-H1 forward and reverse PCR amplification primers as sequencing primers. DNA sequencing was performed using the Beckman Coulter CEQ8000 following Company specifications.

2.6 Analysis of Mutation Frequency and Mutation Rate in E. coli:

A diagram of this procedure is illustrated in Figure 2.4A. Each E. coli strain was transformed with plasmids containing microsatellite repeats by electroporation. After a 60-minute expression period, an aliquot of approximately 20 transformed cells was used to inoculate a 3 mL LB culture containing 50 µg/ml chloramphenicol. A plasmid bearing population of cells was selected by overnight growth of the culture at 37°C for 18-24 hours. The mutation frequency of the HSV-tk gene was determined by selective plating on VBA (Vogel-Bonner minimal salts with the addition of: 0.3 mM of 19 amino acids (no asparagine); 30 mM glucose; 40 µg/mL thiamine; 40 µg/mL each cytidine, guanosine, adenosine; 500 µg/mL uridine) plates + 50 µg/ml chloramphenicol in the presence or absence of 40 µM FUdR as previously described [320]. Chloramphenicol (Cm) and FUdR were purchased from Sigma Chemical Co. (St. Louis, MO). The presence of FUdR selects for cells that have a HSV-tk deficient phenotype. This
Figure 2.4: *E. coli* Mutational Methods: For full description see text
phenotype can be achieved through expansions or deletions at the microsatellite which change the reading frame, as well as frameshifts, base substitutions, and rearrangements throughout the length of the HSV-tk gene and its promoter region [266, 316, 320]. For each microsatellite containing plasmid in each *E. coli* strain, independent overnight cultures from at least 10 to 20 electroporations were analyzed. The mutation frequency (MF) was calculated as the number of F UdR + Cm resistant colonies divided by the total number of Cm resistant colonies. The mutation rate was calculated as the number of mutations per cell generation. The HSV-tk mutation frequencies and mutation rates were analyzed statistically using the nonparametric Wilcoxon-Mann-Whitney Test.

2.7 HSV-tk Mutational Specificity Analysis:

The HSV-tk mutational target is carried on a multicopy ColE1 plasmid and the gene is under no selective pressure during growth of the bacterial culture. Therefore, each cell potentially contains more than one plasmid containing an inactivating HSV-tk mutation or a mixture of wild-type and mutant plasmids. In order to accurately assess the distribution of mutations between the microsatellite allele and the HSV-tk coding sequence within a given plasmid-bearing population, we isolated total plasmid DNA from a given culture. This population of DNA molecules was then sampled for individual HSV-tk plasmids to derive a mutational spectrum. In this manner, we are able to recover inactivating mutations that might otherwise be masked in a cell with a mixed plasmid population.

As shown in Figure 2.4B, plasmid DNA was isolated from 5 independent overnight cultures of each strain and vector combination. This DNA population was
introduced into *E. coli* strain FT334 by electroporation under conditions of 1 DNA molecule per cell, (data not shown) and immediately placed on ice. Each electroporation was then divided into 10-20 aliquots and each aliquot added to 1mL 1X VBA broth. After expression for 2 hours at 37°C the cultures were plated on selective medium and 1 FUdR-resistant mutant was picked from each plate [320]. This was done to ensure independence of the cultures and prevent the possibility of picking siblings for sequencing. The FUdR-resistant colonies were grown overnight in LB broth containing 50 µg/ml Cm and DNA was isolated from each the following day. Large insertions/deletions were detected by *Ava*I/*Bgl*II restriction enzyme digest [266]. DNA sequencing reactions were carried out according to Beckman Coulter (Fullerton, CA) protocol and were sequenced using a Beckman Coulter CEQ 8000 according to company approved procedure. Mutants from different overnight cultures arose independently, by definition. Among mutants isolated from a single culture, those that are of a different mutational event can also be considered independent. Because the types of independent events were similar across the populations, mutants from the five populations were combined to generate mutational spectra for each strain and plasmid. The variation in the types of mutational events observed within one microsatellite vector or cell type, or between different cell types were statistically analyzed using Fisher’s Exact test (two tailed) or Chi Square analysis with an α-value of 0.05.

2.8 Electroporation and Cloning of LCL721 Cells:

721 cells were electroporated and cloned as previously described [266], depicted in Figure 2.5, and described as follows. Approximately 1x10^7 721 cells were electroporated with 10 µg of the desired shuttle vector. The cells were allowed to recover
Figure 2.5: Mutagenesis Analyses in Human Cells: For full description see text
for a period of 3 days and were then put into high hygromycin B selection (300 µg/ml) to select the plasmid bearing population. After 5-7 days the cells were counted and cloned at a density of 1, 2, 5, or 10 viable cells per well by limiting dilution in a 96-well plate, and the hygromycin B selection was lowered to 100 µg/ml. The clones were allowed to grow and expand, adding media when needed. Once the cells expanded into P-100 dishes, the hygromycin B selection was raised to 150 µg/ml. The cells were expanded to ~3x10⁸ cells per clone for DNA isolation.

2.9 Shuttle Vector DNA Isolation from LCL721 Cells:

DNA was extracted from 3x10⁸ LCL721 cells per clone via an alkaline DNA extraction [322] [entire protocol described in [266]]. Briefly, the protocol involved incubation of the cell lysates with 100 µg/ml Proteinase K, incubation at 37°C for 2 hours, and phenol/chloroform extraction. The resulting DNA pellet consisting of EBV DNA, mitochondrial DNA, RNA, and shuttle vector DNA was digested in the presence of RNaseA with 150 units of XbaI, which will digest the mitochondrial and EBV DNA. After precipitation of the restriction enzyme digest, the pellet was resuspended in 10mM Tris-HCl+1mM EDTA (TE buffer) and passed through a Microcon-100 device. This will allow all of the small fragments and the RNA to pass through the filter, while retaining the shuttle vector DNA. The DNA is then eluted from the filter into 25-50 µl TE.
2.10 Determination of Mutation Frequency and Mutation Rate in LCL721 Cells:

The shuttle vector DNA isolated from individual human cells clones was electroporated into FT334 E. coli cells for mutation frequency and rate determination. Shuttle vector DNA (2 µl) was electroporated into 50 µl of E. coli cells, the cells were then placed in 2ml SOC, and grown on the shaker at 37°C for 2 hours. The mutation frequency was then determined by selective plating on VBA plates containing 50 mg/ml Cm or 50 µg/ml Cm+40 µM FUdR as described previously. The mutation frequency and rate calculations were performed as described above.

2.11 Packaging and Titering the Luciferase Lentiviral Vector:

The luciferase containing lentiviral vector was packaged using the pPACK1-Lentiviral Packaging Kit (System Biosciences) following company protocol. Briefly, 293TN cells were transfected with the luciferase shRNA lentiviral vector, the GFP control containing vector, and the pPACK Mix (which contains the vectors that encode the Gag, Pol, and Envelope genes needed for viral transduction). Lipofectamine 2000 (Invitrogen) was used as the transfection reagent. The media containing the pseudo-viral particles was collected every 24 hours for 96 hours, spun down, and filtered through 0.45µM syringe filters. The media was then placed in 1ml aliquots and stored at -80°C.

Titering was performed on LCL721 cells, the same cells to be infected with the pseudo-viral particles. The LCL721 cells were pelleted, resuspended in RPMI+10% FBS+8µg/ml Polybrene, and then seeded at 1x10^5 cells/ml in a 24-well dish. To each well 1 µl, 10 µl, or 100 µl of the media containing the pseudo-viral particles was added. The cells were then grown for 72 hours, collected, and analyzed by fluorescence
activated cell sorting to determine the number of GFP positive cells, as indication of the experimental and the GFP control. The percentage of GFP positive cells was used to determine the Multiplicity of Infection (MOI).

2.12 Infection of LCL721 Cells with Lentiviral Vectors:

LCL721 cells were infected with vectors that contained 1 of 3 shRNA sequences to the POLB gene or contained the luciferase shRNA sequence. To do this, 2.5x10⁵ total cells were pelleted and resuspended in 1ml (for Pol β shRNA vectors) or 1.5 ml (for the luciferase shRNA vector) RPMI+10% FBS+50 µg/ml gentamicin and placed in a 35 mm dish. To this either 1 ml (for the Pol β shRNA vectors) or 500 µl (for the luciferase shRNA vector) of packaged virus was added to the cells along with 5 µg/ml Polybrene. The amount of packaged virus added is determined to obtain an MOI of 1 for each of the virus stocks. The dish was then placed at 32°C overnight. The following day the cells were pelleted and placed in fresh medium, +10% FBS+50 µg/ml Gentamicin. The cells were then placed at 37°C for 4 additional days. On Days 6 and 12 the infection procedure was repeated and on Days 7 and 13 the cells were placed in new media and transferred to 37°C. After the 3 rounds of infection were completed, the cells were placed in medium that contained 150 µg/ml Hygromycin B and were allowed to recover until they reached a density of 5x10⁵ cells/ml. At this point, the cells were placed into puromycin selection at a concentration of 20 µg/ml. After 5-7 days in puromycin selection the cells were cloned at 1, 2, 5, and 10 cells per well and the clones were expanded and the shuttle vector DNA isolated as described previously.
2.13 Protein Isolation and Concentration Determination:

Protein isolation was performed with the NER-PER Kit (Roche) using a confluent plate of each cell line, at approximately $1 \times 10^6$ cells/ml. The basis of the kit is a step-wise isolation of protein from human cells, first isolating the cytoplasmic fraction entirely, then precipitating nuclei and isolating the protein from the nuclear compartment. The company protocol was followed for each sample isolated; however the nuclear fraction was recovered in half the volume recommended in order to increase the concentration of the protein samples. To determine the concentration of each protein sample the BioRad DC Protein Assay Kit was used. The proteins were diluted 1:10 for concentration determination. The samples were read in the spectrophotometer at a wavelength of 750nm. A BSA standard curve was run with each set of protein isolations and the standard curve produced was used to determine the concentration of each protein sample.

2.14 Western Blots:

Western blots were performed using the XCell SureLock system by Invitrogen, following company protocol. Briefly, protein samples were heated to 70°C for 10 minutes, then loaded into 12-well 10% Bis-tris pre-poured gels. MagicMark, a chemiluminescent standard ladder, was also loaded onto each blot, as well as a rainbow ladder used to monitor protein transfer. Antioxidant (500 µl) was added to the inner chamber of each running apparatus. The western gel was run for ~50 minutes at 200v. Upon completion of the run, the transfer apparatus was setup, using a 0.2 µ PDVF membrane. The transfer was completed at 30v for 1 hour and then PDVF membrane was dried. Before blotting of each membrane was completed, the membrane was charged in
methanol for 30 seconds, then washed in TBS+0.1% Tween-20 (TBST). The blot was then blocked in 5% dry milk in the cold room over night. The following day, the blots were washed with TBST the probed. The mouse-α-Pol β primary antibody (NeoMarkers) was used at a dilution of 1:100, the mouse-α-PCNA antibody (BD Transduction Laboratories) was used at a dilution of 1:2500 the mouse-α-Pol η (Santa Cruz) was used at a 1:400 dilution, the mouse-α-lamin A/C (Santa Cruz) was used at a 1:200 dilution, and the mouse-α-Pol δ (Santa Cruz) antibody was used at a 1:800 dilution. With all antibodies the blots were incubated for 1 hr at room temperature on a rocker. The goat-α-mouse secondary antibody (Jackson Labs) was used at a 1:2000 dilution and was incubated for 1 hour at room temperature. Upon completion of the antibody probing, the ECL-Plus chemilumincient reagent (GE Healthcare) was used to detect the proteins. The blots were then scanned on the Storm860 scanner and quantitated using ImageQuant software.

2.15 Isolation of Genomic DNA:

Genomic DNA was isolated from each of the Pol β-1 knockdown clones, as well as untreated LCL721 cells, and luciferase shRNA containing cells using the Qiagen Genomic DNA Isolation Protocol for the genomic tip-20 (Qiagen). For each isolation 5x10^6 total cells per clone were used.
2.16 PCR of Infected Clones to Amplify the shRNA Insert or the Puromycin Resistance Gene:

PCR primers were used to amplify either the shRNA insert or the puromycin resistance gene in Pol β-1 knockdown clones and luciferase shRNA clones to check for viral integration. To amplify the shRNA insert, the sequencing primers provided by SBI were used, shown in Figure 2.6A. These primers should yield an ~260 bp product if the shRNA insert is present. The PCR primers for the Puromycin resistance gene were designed using the PrimerQuest program from IDT, shown in Figure 2.6B. The primer set was subject to a BLAST search to ensure that it would not bind non-specifically, hence affecting the PCR results. Each of the PCR reactions was performed using the Taq Master Mix Kit (Qiagen) and a final primer concentration of 10µM. The PCR conditions were: 94°C for 3 min for initial denature. The 35 cycles of: 94°C for 45 seconds, 50°C for 30 seconds, 72°C for 1 minute. Then a final extension period of 72°C for 10 minutes. Upon completion of the PCR, a 0.8% agarose gel was run to examine the results of the PCR reaction.

2.17 Isolation of RNA from Human Cells:

RNA isolation was performed using Trizol (Invitrogen) following the manufacturer’s protocol. A total of 1x10⁷ cells were used to isolate protein from each clone. At the completion of the protocol the RNA pellet was resuspended in 200µl DNase/RNase free water and the concentration was determined by Nanodrop analysis.
A.) 5’-TGTTTTGGAGTTGGAATCTTAT-3’ (forward)  
5’-ATTTATGTATCTGTGGAGCCT-3’ (reverse)  

B.) 5’-ACCGAGCTGCAAGAAGTCTTCA-3’ (forward)  
5’-AGGAGGCTTCCATCTGCTT-3’ (reverse)  

Figure 2.6: PCR Primers to Determine Integration of shRNA Cassette and Puromycin Resistance Gene: Shown are the primer sequences for the amplification of A.) the shRNA cassette (provided from SBI) or B.) the puromycin resistance gene (designed sing PrimerQuest from ID)
2.18 RT-PCR for Expression of Pol β mRNA:

The primers for the shRNA integration were designed using Primer-BLAST (NBCI) and were ordered from IDT. Four primers were chosen using this program that would amplify products that spanned the entire length of the POLB gene. These primers and their amplicon locations are depicted in Figure 5.4. Each primer set was subjected to a BLAST search, to ensure that nonspecific binding would not occur. A trial RT-PCR reaction was performed on control RNA samples to ensure that each set of primer pairs amplified efficiently. The RT-PCR reactions were carried out using the SuperScript OneStep RT-PCR with Platinum Taq Kit (Invitrogen). For each reaction a final primer concentration of 0.2 µM and a dNTP (USB) concentration of 200µM were used. No additional magnesium was added. The RT-PCR protocol started with an initial 50°C for 30 minutes RT step, followed by 2 minutes at 94°C for initial denaturing. Then 40 cycles were completed as: 94°C for 15 seconds, 50°C for 30 seconds, 70°C for 30 seconds. A final elongation step was completed at 70°C for 10 minutes. The products of the RT-PCR reaction were run on a 0.8% agarose gel containing ethidium bromide to visualize the relative amount of Pol β mRNA present.

2.19 qRT-PCR for Pol β mRNA Levels:

Primers and probes for the TaqMan qRT-PR assay were designed using the TaqMan software. For Pol β, the accession number used for design was NM_002690.1. For TATA Binding Protein (TBP), the accession number used was NM_003194.3. Each mRNA sequence and splice junctions was input into the TaqMan program. Each probe was designed to span the location of a splice junction, to ensure that amplification of
gDNA does not occur. The probe designed to Pol β was labeled with 5’ FAM and 3’ BHQ1, and the probe to TBE was labeled with 5’ CalFlor Orange and 3’ BHQ1. The Pol β primers hybridized at position 113-131 bp (forward) and 194-214 bp (reverse) and the probe was located between base pairs 157-182. The TBP the primers hybridized at positions 1047-1067 bp (forward) and 1128-1147 bp (reverse) and the probe was located between base pairs 1088-1119. The primers were ordered from IDT and the fluorescent probe was ordered from BioSearch Technologies.

Preliminary experiments were completed to determine the initial starting amount of RNA needed for the reactions. To do this, a range of initial RNA was tested to determine the lowest amount that would yield accurate Ct values. The results indicated that the initial amount of RNA to use should be 200 ng.

For each qRT-PCR reaction, 200 ng of RNA was used with the Qiagen Quantitect RT Probe MasterMix Kit (Qiagen). For each sample, a 40 µl reaction was run following company protocol. Each qRT-PCR run was completed in triplicate with a standard, to allow for accurate concentration determination, a no RT control for each unknown sample, and a no RNA control for each dish. All reactions were set up in a cold block kept on ice. A master mix of all reagents (except the RT component) was added to each of the no RT control wells. After these wells had been aliquoted the RT component was added to the mix and the remainder of the wells were aliquoted. Each strip was mixed by inversion and then spun briefly. The strips were loaded into the Stratagene mx4000 qPCR machine and run on the following program: 30 minutes at 50°C, 15 minutes at 95°C, and then 40 cycles of: 94°C for 15 seconds, and 60°C for 1 minute.
For each run, a standard curve was determined using the standards loaded on the plate. Using this standard curve, the thresholds for each dye, and the Ct values, the initial amount of Pol β and TBP were determined. The average concentration of RNA was determined for the three wells of each unknown and the ratio of Pol β:TBP was determined. The standard deviation was determined for each of the probes individually, and overall standard deviation was determined as the square root of the square of each individual standard deviation. An example of each calculation is shown in Figure 2.7. The ratio of Pol β:TBP and the standard deviations were then graphically analyzed to determine the effect on knockdown of mRNA levels.

2.20 Transient RNA Transfection of Human Cells with shRNA to Pol β-1:

Transient transfection of LCL721 cells was carried out using Transit-TKO RNA transfection reagent (Mirus Bio LLC). For each transfection, 5 dishes were transfected, so that RNA samples could be collected every 24 hours for 5 days. To begin, 60µl of the Transit-TKO reagent was mixed with 7.5ml of complete RPMI and incubated for 20 minutes at room temperature. After the incubation, a final concentration of 25mM of duplexed RNA was added to the tube, mixed by gentle inversion, and let incubate for 20 minutes at room temperature. LCL721 cells were resuspended to a concentration of 5 x10^5 cells/ml in 1.5ml RPMI+10% FBS+50 µg/ml Gentamicin, and placed in a P100 dish. Transfection complex was added to the cell suspension dropwise, until all reagent was added. The cells were then incubated at 37°C, collecting 1 dish of cells every 2 hours for 5 days. RNA was extracted and the levels of Pol β were determined by qRT-PCR as described above.
2.21 Duplexing of shRNA Oligonucleotides:

RNA oligonucleotides with the sequence of the Pol β-1 siRNA were ordered from IDT. The oligos were then duplexed in a 1:1 ratio using the duplexing buffer provided by IDT. To duplex the oligos, each was resuspended to 1mM concentration in the duplexing buffer, and then mixed in equal parts in a microfuge tube. The tube was then heated to 94°C for 2 minutes and cooled to room temperature. The duplexed primers were then stored at -20°

2.22 HSV-tk in vitro Assay:

The development of this assay has been previously described in detail [320] and the assay is depicted in Figure 2.8. A DNA hybrid was made by hybridizing an oligonucleotide, IHTW02 or TK-MluR, to a single stranded DNA template, either pSSStu or pSAStu respectively, at a 1:2 molar ratio. 2pmol of the DNA hybrid was then combined with the reaction buffer, 50 mM Tris-HCl (pH 8.5), 50 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol, 1 mM dNTPs, 200 µg/mL BSA, and 10 pmol of Pol β/pmol of template DNA. Reactions were incubated at 37°C for 60 minutes and terminated with 15 mM EDTA. The reaction products were then washed with 9 volumes of TE and spun through a Microcon-30 device. The DNA products were then digested with EcoRV and MluI, and the small fragment (203 bp) was isolated by selective PEG precipitation. The small fragment was incubated overnight on ice, pelleted by centrifugation, and resuspended, and washed through a Microcon-30 filter. The final product was resuspended in ROH₂O and aliquots analyzed on a 0.8% agarose gel to determine the final concentration.
Figure 2.7: qRT-PCR Sample Calculations: The determined ng values (in triplicate) for each sample were normalized by dividing Pol β/TBP, and then averaged (called the triplicate average). These values were then divided by the triplicate average for the control sample (Average). For each gene examined individual standard deviation values were determined, cv1 for Pol β and cv2 for TBP. These were calculated at the standard deviation of the 3 samples/average of the 3 samples. The overall standard deviation (cv) was determined as the square root of the squares of individual standard deviations of Pol β and TBP (cv1 and cv2).

<table>
<thead>
<tr>
<th>Pol β</th>
<th>TBP</th>
<th>Normalized</th>
<th>Average</th>
<th>Relative levels</th>
<th>cv=Std Dev</th>
<th>cv1</th>
<th>cv2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.04E+02</td>
<td>1.12E+02</td>
<td>9.30E-01</td>
<td>8.67E-01</td>
<td>1.00E+00</td>
<td>1.79E-01</td>
<td>1.00E-01</td>
<td>1.48E-01</td>
</tr>
<tr>
<td>1.28E+02</td>
<td>1.47E+02</td>
<td>8.70E-01</td>
<td>8.02E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.18E+02</td>
<td>1.47E+02</td>
<td>1.08E+00</td>
<td>9.49E-01</td>
<td>1.09E+00</td>
<td>1.52E-01</td>
<td>1.23E-01</td>
<td>8.97E-02</td>
</tr>
<tr>
<td>1.39E+02</td>
<td>1.39E+02</td>
<td>1.00E+00</td>
<td>7.63E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.26E+02</td>
<td>1.65E+02</td>
<td>4.11E+00</td>
<td>3.99E+00</td>
<td>4.60E+00</td>
<td>1.17E-01</td>
<td>9.55E-02</td>
<td>6.82E-02</td>
</tr>
<tr>
<td>9.41E+02</td>
<td>2.39E+02</td>
<td>3.94E+00</td>
<td>3.92E+00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.64E+02</td>
<td>2.46E+02</td>
<td>3.92E+00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A gapped duplex (Gap) was created that consists of an intact inner strand of DNA and an outer strand that is missing a portion of the HSV-tk gene. This molecule was made by hybridizing a linear fragment to ssDNA. The linear fragment was obtained by digesting plasmid pStuI (sense strand) or pSAStuI (antisense strand) with StuI and MluI and the large fragment (~5.9kb) was recovered by PEG separation, centrifugation, and ethanol precipitation. The concentration was determined by ethidium bromide staining of electrophoresed samples in a 0.8% agarose gel. ssDNA was prepared by infection of *E. coli* strain DH5αIQ with R408 helper phage in the presence of carbenicillin and kanamycin. Construction of the Gap molecule was completed by hybridizing the large fragment with the ssDNA. To do this, the large fragment was incubated at 85°C for 12 minutes. During the final minute of the incubation the ssDNA was added. The tubes were iced for 5 minutes and then a final concentration of 2X SSC was added. The tubes were then incubated at 60°C for 30 minutes to allow the hybridization to occur. The product was run on a 0.8% agarose gel and purified using silica (Gene Clean). The concentration of the Gap molecule was determined by Nanodrop analysis.

The purified small fragment from the Pol β reactions was hybridized into the constructed Gap vector. The small fragment was heated to 85°C for 5 minutes and then combined with the Gap vector and a final concentration of 0.5X SSC. The reaction was then heated to 45°C for 1 hour. An aliquot was removed and analyzed by overnight gel electrophoresis to ensure that complete hybridization had occurred. The resulting vectors were then electroporated into *E. coli* strain FT334 for mutation rate and specificity analysis as described previously.
Figure 2.8: HSV-tk in vitro assay: Purified DNA Pol β was added to a ssDNA template hybridized with a DNA primer. DNA synthesis was completed on the ssDNA and restriction enzyme digest was performed to isolate the newly replicated region containing the artificial microsatellite. This isolated DNA fragment was hybridized into a Gaped molecule, and introduced into *E. coli* for mutational analyses (Taken from [323])
Chapter 3: Role of *E. coli* DNA polymerase IV in Spontaneous Microsatellite Mutagenesis  (Jacob and Eckert, *Mutation Research* 619:93-103)

3.1 Rationale/Hypothesis

In this study, we examined the role of *E. coli* DNA Polymerase IV on spontaneous mutagenesis at microsatellite alleles. We inserted various mononucleotide and dinucleotide microsatellite repeat sequences in-frame into the 5’ end of the HSV-tk gene. Using a forward assay, we detect mutations that occur in the gene, both in the artificial repeat we have inserted as well as in the entire length of the coding region of the HSV-tk gene itself. We expect that *E. coli* strains that have reduced levels of Pol IV will have an altered mutational specificity at the microsatellite alleles, relative to cells that have normal levels of Pol IV.

3.2 Genotyping the dinB+ and dinB- Strains

Before starting any of the experiments to look at the effect of Pol IV on mutagenesis, we had to first determine that the genotypes of the strains were correct. gDNA was isolated from dinB+ (FTC21) and dinB- (FTK44) *E. coli* strains and subject to PCR analysis as described. The sequence for the dinB+ strain sequenced as wild-type; the dinB- strain had a point mutation at position 145 and 146 that changed the sequence from GC to TT, and resulted in an inactivation of the protein [73]; verifying the identity of our strains.
3.3 Results

3.3.1 Effect of Mismatch Repair on Microsatellite Mutagenesis

We previously reported that the quantitative effect of MMR deficiency on HSV-tk gene mutagenesis was relatively modest when the entire transformed population was sampled as a single unit [313]. In contrast, mutagenesis at a chromosomal locus (Rif^R) was elevated several orders of magnitude in the same MMR-deficient cultures. One potential reason for this differential effect is that a high number of pre-existing mutant plasmid DNAs are introduced into both MMR^+ and MMR^- populations during transformation. Alternatively, the number of cell generations between transformation and mutational sampling may have been insufficient to reveal the MMR effect in plasmid DNA. In the current study, we report the results of a modified mutagenesis protocol that was designed to increase the sensitivity of the plasmid-based assay for the detection of host cell genotype effects. The major change in our method is that we use only a small number of transformed cells to inoculate cultures, thereby minimizing the probability of pre-existing mutations while maximizing the number of cell generations.

To test the validity of this new method, we quantitated the cumulative effect of MMR loss. Plasmid vectors containing artificial microsatellite sequences were introduced into isogenic mutL^+ (FT334) and mutL^- (PP102) recA^- E. coli cells by electroporation. Approximately 20 transformed cells were then used to inoculate several independent cultures, the cultures were grown overnight, and plated under FUdR selection to determine the HSV-tk mutation frequency. For the control vector (containing no microsatellite) the HSV-tk mutation frequency was increased ~700-fold for the mutL-deficient strain compared to the mutL-proficient strain (Figure 3.1). This value is very
Figure 3.1: Effect of MMR on Mutation Frequencies of Microsatellite Containing Vectors: Vectors were introduced into mutL\textsuperscript{+} (FT334) and mutL\textsuperscript{-} (PP102) cells and the mutation frequencies were determined for the HSV-tk only as well as microsatellite containing vectors as described in Materials and Methods. The ratios of mutL\textsuperscript{-}/mutL\textsuperscript{+} mutation frequencies as determined for 10-20 cultures of each strain and vector combination are depicted in the graph. The median mutation frequencies for mutL\textsuperscript{+} strain are: 3.6x10\textsuperscript{-8}, control; 4.6x10\textsuperscript{-8}, [G/C]\textsubscript{10}; 22x10\textsuperscript{-8}, [GT/CA]\textsubscript{11}; 5.0x10\textsuperscript{-8}, [TC/AG]\textsubscript{11}. The median mutation frequencies for the mutL\textsuperscript{-} strain are: 2.5x10\textsuperscript{-5}, control; 85x10\textsuperscript{-5}, [G/C]\textsubscript{10}; 12x10\textsuperscript{-5}, [GT/CA]\textsubscript{11}; 19x10\textsuperscript{-5}, [TC/AG]\textsubscript{11}.
similar to the magnitude of the effect observed for the loss of MMR at the chromosomal RifR locus (900-fold) in the same strains [313]. This result demonstrates that the effect of MMR on mutagenesis at the plasmid-encoded HSV-tk gene target is more effectively quantitated by this experimental approach. In-frame insertion of the [G/C]₁₀ repeat allele into the 5’ region of the HSV-tk gene resulted in an ~18,000-fold increase in the mutation frequency in the MMR deficient cells, compared to MMR proficient cells. The difference in the median mutation frequencies observed for the two strains was statistically significant (Wilcoxon Mann Whitney Test, p=0.002). Insertion of the [GT/CA]₁₀ and [TC/AG]₁₁ alleles resulted in an ~500-fold and ~3,800-fold increases, respectively, in the mutation frequency of the mutL⁻ cells compared to mutL⁺ (p<0.001, p<0.001). In MMR-deficient cells, the addition of microsatellite repeats had an effect of the same or greater magnitude than the HSV-tk only control vector, confirming that these STR sequences truly are acted upon by the MMR proteins. A hierarchy was detected for the relative mutagenesis between the mutL⁻ and mutL⁺ strain, with the mononucleotide repeat [G/C]₁₀ having the greatest quantitative difference (Figure 3.1), followed by the two dinucleotide repeats, in the order [TC/AG]₁₁ then [GT/CA]₁₀. In addition, pairwise comparisons of the mutation frequencies for the three microsatellite containing vectors in the mutL⁻ strain were all statistically significant (p≤0.04). Overall, these results indicate that the effect of MMR at the STR alleles is sequence specific.

3.3.2. Effect of Pol IV on Spontaneous Mutation Frequency

To determine the effect of Pol IV on spontaneous mutagenesis, plasmid vectors were introduced into isogenic E. coli strains FCT21 (dinB⁺) and FTK44 (dinB⁻) and the
HSV-tk mutation frequencies determined. Growth curve analyses performed under conditions of the mutagenesis assay indicated that the cells were in stationary phase at the time of sampling for mutation frequency analyses (data not shown). For the control vector, the median HSV-tk mutation frequency was $6.8 \times 10^{-8}$ in the presence of $\text{dinB}$ and $2.0 \times 10^{-8}$ in the absence of $\text{dinB}$ (Table 3.1). This 3.5-fold higher mutation frequency observed in the presence of $\text{dinB}$ is statistically significant, indicating that Pol IV plays a role in spontaneous mutagenesis at the episomal HSV-tk gene (Wilcoxon Mann Whitney Test, $p < 0.001$). To determine the effect of $\text{dinB}$ on microsatellite mutagenesis, vectors containing different mononucleotide and dinucleotide repeats were introduced into $\text{dinB}^{+}$ and $\text{dinB}^{-} \ E. \ coli$ cells and the mutation frequency for each culture was determined (Figure 3.2; summarized in Table 3.1). In our vector nomenclature, the leading/lagging template strands are indicated as first the leading strand, then the lagging strand. For example, $[\text{G/C}]_{10}$ indicates that the 10 G’s are the leading strand template and the 10 C’s are the lagging strand template. For the mononucleotide $[\text{G/C}]_{10}$ vector, the median mutation frequency was $6.9 \times 10^{-7}$ in the presence of $\text{dinB}$ and $4.6 \times 10^{-7}$ in its absence. This difference in the median mutation frequency of the $[\text{G/C}]_{10}$ allele in the presence of Pol IV was not statistically significant ($p = 0.17$). To examine whether there is a replication strand bias for STR mutagenesis, we also measured mutation frequencies for a $[\text{C/G}]_{10}$ allele in each strain. The median mutation frequency was $7.1 \times 10^{-7}$ in the $\text{dinB}^{+}$ strain and $7.6 \times 10^{-7}$ in the $\text{dinB}^{-}$ strain. The result of statistical analysis indicated that there was no significant difference in the mutation frequency between the two cell types ($p = 0.84$). The $[\text{TC/AG}]_{11}$ dinucleotide repeat produced similar results to the mononucleotide allele. The median mutation frequency was $4.7 \times 10^{-7}$ in $\text{dinB}^{+}$ cells and
Table 3.1: Summary of the Effect of dinB on Spontaneous Mutagenesis

Median HSV-tk Mutation Frequency x10⁻⁸

<table>
<thead>
<tr>
<th>STR Allele</th>
<th>dinB⁺ (FCT21)ᵃ</th>
<th>dinB⁻ (FTK44)ᵇ</th>
<th>Ratioᵇ</th>
<th>p-valueᶜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>6.8</td>
<td>2.0</td>
<td>3.5</td>
<td>0.0012</td>
</tr>
<tr>
<td>[G/C]₁₀</td>
<td>69</td>
<td>46</td>
<td>1.5</td>
<td>0.17</td>
</tr>
<tr>
<td>[C/G]₁₀</td>
<td>71</td>
<td>76</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>[GT/CA]₁₀</td>
<td>23</td>
<td>34</td>
<td>0.68</td>
<td>0.46</td>
</tr>
<tr>
<td>[TC/AG]₁₁</td>
<td>47</td>
<td>33</td>
<td>1.4</td>
<td>0.20</td>
</tr>
</tbody>
</table>

ᵃ FCT21 is a derivative of SMR4562

ᵇ FTK44 is a derivative of SMR5830

ᶜ Wilcoxon Mann Whitney Test for 10-62 independent cultures of each strain and vector combination
Figure 3.2: Effect of Pol IV on Mutation Frequencies of Microsatellite-Containing Vectors: Mutation frequency data were analyzed by box plot [324] for each vector and cell type combination. The gray boxes represent the dinB+ (FCT21) values and the white speckled boxes represent dinB- (FTK44) values. Each box represents the 95% confidence interval for one allele and genetic background calculated from the number of independent cultures shown in parentheses. Horizontal bars represent median mutation frequencies. Vertical bars represent the range of values included in the statistic. Outliers exist for each individual vector and cell type combination, but were removed from the graph for simplicity purposes ([G/C]10 dinB+: 20, 26; [G/C]10 dinB+: 16; [C/G]10 dinB+: 19.4; [GT/CA]10 dinB+: 9.1, 9.4; [GT/CA]10 dinB+: 8.5; [TC/AG]11 dinB+: 13, 14, 16, 17, 22, 147; [TC/AG]11 dinB+: 18, 27, 46). Non-parametric statistical analysis showed no statistical difference between strains in median mutation frequencies for any of the repeat alleles tested.
3.3x10^{-7} in \textit{dinB}^- cells, a result that was again not significant (p=0.2). The second dinucleotide repeat, \textit{[GT/CA]}_10 produced a median mutation frequency of 2.3x10^{-7} in cells containing \textit{dinB} and 3.4x10^{-7} in cells that were deficient of \textit{dinB}. These median mutation frequencies produced an opposite effect of those observed with the control vector or other microsatellite alleles, in that the presence of Pol IV decreased the mutation frequency 1.5-fold at the HSV-tk target compared to cells that lack the polymerase. This opposing effect was not found to be statistically significant (p=0.46). Overall, these results indicate that the absence of Pol IV does not affect the mutation frequency observed for microsatellite containing vectors.

3.3.3. Mutational Specificity of the \textit{[G/C]}_10 Microsatellite Allele

The results of the mutation frequency analysis indicated that the presence of \textit{dinB} had no significant effect on mutagenesis of various vectors containing microsatellite alleles. This said, there is still a possibility that a mutational bias exists in the types of mutations that occur in cells that contain or lack the polymerase. To determine if a mutational bias does exist for mutations made by Pol IV, mutational spectra analyses were completed. As explained in the methods section, total plasmid DNA was isolated from five overnight cultures. Each population was sampled for mutations by re-introducing this DNA into \textit{tdk},\textit{recA}^- \textit{E. coli} strain FT334 under conditions of one molecule per cell. This HSV-tk mutational system detects mutation events that occur both at the microsatellite allele and also throughout the entire coding and promoter region of the HSV-tk gene [267]. The data in Table 3.2 summarizes the mutational events that occurred in plasmids recovered from both \textit{dinB}^+ and \textit{dinB}^- cells. In the \textit{dinB}^+ spectrum,
Table 3.2: Mutational Events Within the [G/C]₁₀ Vector in \( \text{dinB}^+ \) and \( \text{dinB}^- \) Strains

<table>
<thead>
<tr>
<th>Mutation Class</th>
<th>Number of Events (Proportion of Region)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{dinB}^+ ) (FCT21)</td>
</tr>
<tr>
<td>STR \ [G/C]₁₀</td>
<td></td>
</tr>
<tr>
<td>Median Mutation Frequency(^a)</td>
<td>(3.8 \times 10^{-7})</td>
</tr>
<tr>
<td>Number</td>
<td>61</td>
</tr>
<tr>
<td>Expansion</td>
<td>47 (.77)</td>
</tr>
<tr>
<td>Deletion</td>
<td>14 (.23)</td>
</tr>
<tr>
<td>HSV-tk Coding</td>
<td></td>
</tr>
<tr>
<td>Median Mutation Frequency(^a)</td>
<td>(2.4 \times 10^{-7})</td>
</tr>
<tr>
<td>Number</td>
<td>39</td>
</tr>
<tr>
<td>Base Substitutions</td>
<td>11 (.28)</td>
</tr>
<tr>
<td>Frameshifts</td>
<td>8 (.21)</td>
</tr>
<tr>
<td>Complex</td>
<td>2 (.05)</td>
</tr>
<tr>
<td>Large Insertions or Deletions(^b)</td>
<td>18 (.46)</td>
</tr>
<tr>
<td>Total Sequenced Events</td>
<td>100</td>
</tr>
</tbody>
</table>

\(^a\) Median mutation frequency for 5 independent cultures of each strain

\(^b\) Determined by \(AvaI/BglII\) restriction enzyme digest
61% of the mutational events occurred at the [G/C]₁₀ allele; similarly, 72% occurred at the [G/C]₁₀ allele in dinB⁻ cells, a difference that is not statistically significant (p=0.11, Fishers Exact Test). These populations correspond to a median mutation frequency at the [G/C]₁₀ allele of 3.8x10⁻⁷ for the five dinB⁺ cultures, compared to 1.7 x10⁻⁷ for the five dinB⁻ cultures. The mutational events at the microsatellite allele were expansions and deletions of 1 unit, with a bias toward expansions in both strains (77%, dinB⁺; 65%, dinB⁻) (Figure 3.3). No statistical difference was observed between the number of expansions and deletions at the microsatellite in the presence or absence of dinB (p=0.16). Overall, our results indicate that Pol IV does not play a significant role in spontaneous mutagenesis at the [G/C]₁₀ allele.

The remaining mutational events observed occurred at different locations through the coding region of the HSV-tk gene. As summarized in Table 3.2, the median HSV-tk coding region mutation frequency was 24x10⁻⁸ for the dinB⁺ populations and 6.4x10⁻⁸ for the dinB⁻ populations. This 3.75-fold difference is similar to what was observed for the control plasmid (Table 3.1). The coding region point mutations for each strain containing the [G/C]₁₀ vector are given in Tables 3.3 and 3.4. Overall, there was no large mutational hotspot observed in either cell type. The frameshift events observed were a fairly equal mix of insertions and deletions. In cells containing Pol IV, a similar percentage of frameshifts (21%) and base substitutions (28%) were observed. In the absence of Pol IV, a higher proportion of frameshifts (19%) over base substitutions (5% ) was observed. These results correlate with the results from the HSV-tk vector only control, showing that the HSV-tk gene functions as an internal control for the microsatellite-containing vector.
Figure 3.3: Range of Unit Changes at the $[G/C]_{10}$, $[GT/CA]_{10}$, and $[TC/AG]_{11}$ alleles: The x-axis represents the range of unit changes observed among the vectors. The y-axis represents the number of mutants containing that mutational event in each spectrum. $dinB^+$ is strain FCT21. $dinB^-$ is strain FTK44.
Table 3.3: Mutational Spectra of the HSV-tk Coding Region in \textit{dinB}^{+} Cells (FCT21)

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Location$^a$</th>
<th>Sequence$^b$</th>
<th>Mutation</th>
<th>Location</th>
<th>Sequence$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+G/C</td>
<td>-56</td>
<td>AGT\textbf{GGG}ACC</td>
<td>AT→GC</td>
<td>241</td>
<td>GGG\textbf{AAAA}</td>
</tr>
<tr>
<td></td>
<td>291</td>
<td>TAT\underline{CGTC}</td>
<td></td>
<td>671</td>
<td>TGCT\textbf{TGG}</td>
</tr>
<tr>
<td></td>
<td>571</td>
<td>TGCT\textbf{TACC}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-G/C</td>
<td>512 - 515</td>
<td>ATG\textbf{CCCGCC}</td>
<td>AT→CG</td>
<td>1042</td>
<td>GAT\textbf{TACG}</td>
</tr>
<tr>
<td></td>
<td>605 - 610</td>
<td>TGA\textbf{CCCCCGAGG}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+A/T</td>
<td>255 - 256</td>
<td>CGCA\textbf{ACTG}</td>
<td>GC→AT</td>
<td>256</td>
<td>ACG\textbf{CAAC}</td>
</tr>
<tr>
<td></td>
<td>268 - 269</td>
<td>TGCA\textbf{CTC}</td>
<td></td>
<td>1072</td>
<td>GAC\textbf{GCCC}</td>
</tr>
<tr>
<td></td>
<td>785</td>
<td>TGCA\textbf{AAAT}</td>
<td></td>
<td>656</td>
<td>CCG\textbf{GCAC}</td>
</tr>
</tbody>
</table>

$^a$ Nucleotide position of the mutated base in the HSV-tk gene. Numbering of the HSV-tk sequence begins at the G residue within the BglII recognition site [320].

$^b$ Wild-type sequence of the sense strand (5’ to 3’ direction). The underlined base(s) indicated the sequence context of the mutation. Insertions occur within or preceding the underlined base(s) and deletions or base substitutions occur at or within the underlined base(s).

$^c$ 3 mutations observed
Table 3.4: Mutational Spectra of the Coding Region of the HSV-tk Gene in dinB<sup>−</sup> Cells (FTK44)

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Location&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sequence&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Mutation</th>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>-G/C</td>
<td>486 – 493&lt;sup&gt;c&lt;/sup&gt;</td>
<td>ATCGGGGGGAGG</td>
<td>GC→CG</td>
<td>-16</td>
<td>TGCCGGG</td>
</tr>
<tr>
<td>+G/C</td>
<td>-66</td>
<td>CGGCACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-A/T</td>
<td>241 - 244</td>
<td>GGGAAAAACCA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Nucleotide position of the mutated base in the HSV-tk gene. Numbering of the HSV-tk sequence begins at the G residue within the BglII recognition site [320].

<sup>b</sup> Wild-type sequence of the sense strand (5' to 3' direction). The underlined base(s) indicated the sequence context of the mutation. Insertions occur within or preceding the underlined base(s) and deletions or base substitutions occur at or within the underlined base(s).

<sup>c</sup> 2 mutation observed
3.3.4. Mutational Specificity at [GT/CA]\textsubscript{10} and [TC/AG]\textsubscript{11} Microsatellite Alleles

Following the mutation frequency analysis of the two dinucleotide vectors, ten mutants from each of five populations per cell type were screened to determine the location of the mutational event, either at the microsatellite allele or in the HSV-tk coding region (Table 3.5). For the [GT/CA]\textsubscript{10} allele, 80% of the mutation events were at the microsatellite in dinB\textsuperscript{+} cells, compared to 71% in the dinB\textsuperscript{−} cells. Statistical analysis showed no significant difference in the proportion of mutations at the [GT/CA]\textsubscript{10} allele versus the HSV-tk coding region (p=0.45, Fisher’s Exact Test). The majority of mutational events for this vector were 1-unit deletions (meaning loss of a [GT/CA] unit) in both cell types (Figure 3.3), and the types of mutational events at the repeat allele were not statistically significant between strains (p=0.11). For the [TC/AG]\textsubscript{11} allele, a similar trend to that of the [GT/CA]\textsubscript{10} allele was observed. We detected 88% of the mutational events to be at the microsatellite in dinB\textsuperscript{+} cells and 98% in dinB\textsuperscript{−} cells (Table 3.5). Statistical analysis of the proportion of mutational events at the microsatellite allele versus the HSV-tk coding region was not statistically significant (p=0.06). For this microsatellite allele the majority of mutational events in both strains were alterations of 1 unit (Figure 3.3), similar to what was observed with the [GT/CA]\textsubscript{10} dinucleotide vector.

3.3.5. Microsatellite Length and Sequence Specificity

Given the variability in the location of mutational events between the different microsatellite vectors, further analyses of the data were performed comparing all three microsatellite repeat alleles. Chi Square statistical analyses were used to compare the three STR-containing vectors within a given strain to determine if the observed
Table 3.5: Mutational Analyses of [GT/CA]$_{10}$ and [TC/AG]$_{11}$ Vectors

<table>
<thead>
<tr>
<th>Mutation Class</th>
<th>Number of Events (Proportion of Region)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dinB$^+$ (FCT21)</td>
</tr>
<tr>
<td>[GT/CA]$_{10}$</td>
<td></td>
</tr>
<tr>
<td>Median Mutation Frequency</td>
<td>2.3 x10$^{-7}$</td>
</tr>
<tr>
<td>Events in Coding</td>
<td>9</td>
</tr>
<tr>
<td>Events at STR</td>
<td>35</td>
</tr>
<tr>
<td>Expansion at STR</td>
<td>6 (0.17)</td>
</tr>
<tr>
<td>Deletion at STR</td>
<td>29 (0.83)</td>
</tr>
<tr>
<td>[TC/AG]$_{11}$</td>
<td></td>
</tr>
<tr>
<td>Median Mutation Frequency</td>
<td>4.7 x10$^{-7}$</td>
</tr>
<tr>
<td>Events in Coding</td>
<td>6</td>
</tr>
<tr>
<td>Events at STR</td>
<td>42</td>
</tr>
<tr>
<td>Expansion at STR</td>
<td>14 (0.33)</td>
</tr>
<tr>
<td>Deletion at STR</td>
<td>28 (0.67)</td>
</tr>
</tbody>
</table>
mutational spectra were affected by either allele unit length or allele sequence mutants from each of five populations per cell type were screened to determine the composition. The first analysis compared the distribution of mutations within the microsatellite allele versus the HSV-tk coding region for the three STR containing vectors, in either the \( \text{dinB}^+ \) or \( \text{dinB}^- \) strain. The results were statistically significant irrespective of cell type \((p=0.0016, \text{dinB}^+; p=0.0007, \text{dinB}^-, \text{Chi Square})\). The next analysis examined the type of mutational events that occurred within each microsatellite region. For each strain, the mutational specificity among the alleles differed in a statistically significant manner \((p<0.001 \text{ both backgrounds, Chi Square})\). As mentioned previously, the \([G/C]_{10}\) mononucleotide allele showed a large proportion of expansion events occurring at the repeat region (Table 3.2 and Figure 3.3). This result is opposite to what was observed in the \([GT/CA]_{10}\) and \([TC/AG]_{11}\) dinucleotide alleles (Table 3.5).

The final test was to determine if there was a difference in the unit change of the mutational event at the repeat allele (gain or loss of 1-5 repeat units). The data in this case showed a hierarchy for the number of different unit changes observed, in the order of \([TC/AG]_{11}\) with the largest range, then \([GT/AC]_{10}\), and finally \([G/C]_{10}\) (Figure 3.3).

For the \([G/C]_{10}\) allele, we detected almost 100\% of the events to be insertions or deletions of only 1 unit. This however was not the case for the two dinucleotide alleles. Both the \([GT/CA]_{10}\) and \([TC/AG]_{11}\) vectors had events other than 1-unit changes, with the limits of the range extending as far as gain or loss of 5 units. The results of statistical analysis were again significant \((p<0.001, \text{Chi Square})\) regardless of the background of the cells. Overall, these results indicate that both the composition of the repeat allele and the unit length of the repeat play a role in microsatellite mutagenesis.
3.4. Discussion

Our initial experimental design was to examine the effect of Pol IV on mutagenesis of \([A/T]_{11}\) and \([G/C]_{10}\) mononucleotide microsatellites. A literature search performed before starting the experiments found that \textit{E. coli} performs transcriptional frameshifting in \([A/T]\) tracts longer than 9 base pairs in length [325]. Despite this information, we began this study with both mononucleotide templates to see if the same results were achieved in our hands. As we began to analyze independent mutants, the sequencing results revealed only 1 class of mutational events, deletions of 2 base pairs. This was unusual, as the characteristic mutation in a microsatellite allele is the gain or loss of 1 unit (which corresponds to one base pair for mononucleotides). Additional research in the lab showed that all classes of mutations could be detected with an \([A/T]_8\) allele within our system (data not shown). This indicated to us that the \textit{E. coli} frameshifting the 11 bp tract to 9bp, thus the result we are obtaining through sequencing did not represent the loss of 2 bp by polymerase errors. Therefore, we have decided not to move forward with the \([A/T]_{11}\) allele.

\textit{E. coli} polymerase IV is a distributive polymerase that lacks a 3’ to 5’ exonuclease (proofreading) activity and is prone to elongation of misaligned primer/template structures [73]. The over-expression of the polymerase causes -1 frameshifts in polypurine runs of G in undamaged DNA [84]. When looking at spontaneous mutagenesis, previous studies have shown a statistically significant decrease in the number of frameshift and base substitution events in the absence of \textit{dinB} [77, 86, 88] while over-expression of the \textit{dinB} gene results in a significant increase in the mutation frequency [83, 84]. Pol IV was chosen as a candidate to analyze for strand slippage.
within microsatellite alleles because a role for the polymerase in spontaneous mutagenesis was acknowledged, as was its role in mutagenesis at homopolymeric G sequences. In this study, we examined the effect of dinB on spontaneous mutagenesis at the HSV-tk gene and three different microsatellite alleles. The mutation frequency for the HSV-tk only control vector was 3.5-fold lower in the absence of dinB. Although we observed a significant decrease in mutation frequency with the loss of dinB at the HSV-tk coding sequence, we did not observe a significant effect with the loss of dinB on the mutation frequency of any of the three artificial microsatellite alleles examined, which is in contrast to previously published data.

### 3.4.1. Pol IV and HSV-tk Mutagenesis

The effect of Pol IV on spontaneous mutagenesis is complex, and dependent on at least two factors. First is the expression level of the Pol IV protein. Higher expression of Pol IV is detected from the F factor compared to expression from the chromosomal copy of the dinB gene [77] thus deletion of dinB from the chromosomal location has a negligible effect on mutagenesis [77]; whereas deletion from the episome produced a significant effect on mutagenesis [77, 88]. Pol IV protein levels also are dependent on the growth phase of the culture. Levels of Pol IV protein are transcriptionally induced in late stationary phase [80], and Pol IV does not contribute to spontaneous mutations during exponential growth when expressed at endogenous levels [161]. Second is the location of the mutational target. There are many reports in the literature in which both chromosomally and episomally located targets were used to look at a role for Pol IV in mutagenesis. Many of these studies were performed using a chromosomal target and
resulted in no difference in the mutation frequency in the presence of dinB [87, 88, 161],
while a few did observe a decrease in mutagenesis in the absence of dinB [83, 86]. Two
of these studies also examined an episomal mutational target and found that it was more
susceptible to mutagenesis than any of the chromosomal targets tested [88] in the
presence of dinB [83].

In this study we have used a dinB+ E. coli strain carrying both the F factor dinB,
as well as the chromosomal copy. The dinB− strain carries a nonfunctional dinB10 allele
in both a chromosomal and episomal location. Both cultures were sampled during early
stationary phase of growth. We have used the HSV-tk gene target, which is episomally
located, to show a 3.5-fold decrease in spontaneous mutation frequency in the absence of
Pol IV (Table 3.1). Mutational analyses at the coding region of the HSV-tk gene for the
[G/C]10 vector produced results that were very similar to the HSV-tk gene only vector.
An approximate 4-fold difference in the spontaneous mutation frequency at the HSV-tk
coding region was detected for dinB+ vs. dinB− cells (Table 3.2). Overall, our results are
consistent with a role for Pol IV in mutagenesis of an episomal target.

3.4.2. Pol IV and Microsatellite Mutagenesis

The mutational analyses of the [G/C]10 mononucleotide region showed no
statistically significant differences in mutation frequency of the dinB+ and dinB− strains
(Table 3.2 and Figure 3.2). This indicates that Pol IV does not play a role in the
spontaneous mutagenesis at the [G/C]10 repeat allele. Further analyses showed no strain
difference in the types of mutational events at the repeat allele. This result was surprising
to us, as Pol IV was shown to create -1 frameshifts in homopolymeric G6 runs [84]. The
difference between our study and the previous study [84] could be due to sequence context effects (lac gene vs. HSV-tk gene) or the length of the repetitive target (6 vs. 10 nucleotides).

Recently, Kuban et al. reported that dinB was found to make more errors on the lagging strand during synthesis than on the leading strand [89]. In our [G/C]_{10} vector, the mononucleotide G repeat is located on the leading strand of synthesis. To test whether replication strand bias may have been one of the reasons that we did not observe a significant difference in the mutation frequency at the mononucleotide allele, we constructed the opposite allele [C/G]_{10}, in which the G repeat is located on the lagging strand. The results of mutation frequency analyses of this vector showed no statistically significant difference with dinB status, just as was observed in the [G/C]_{10} vector (Table 3.1 and Figure 3.2).

The analyses of the spontaneous mutation frequencies of the two dinucleotide repeat vectors also did not emulate that of the HSV-tk only control. In the [TC/AG]_{11} vector, we observed a similar trend to the control; however the observed decrease in mutation frequency was not significant. In the second dinucleotide vector, [GT/CA]_{10}, we observed a non-significant increase in the mutation frequency in the absence of dinB. The reason for this opposite effect on mutation frequency with the [GT/CA]_{11} vector may be sequence composition. Overall, these results indicate that Pol IV does not play a role in the spontaneous mutagenesis of the dinucleotide regions, as was observed for the mononucleotide repeat allele.

One possible mechanism to explain a role of Pol IV in coding region but not microsatellite region mutagenesis is MMR efficiency. We have shown in this study that
each of the microsatellite alleles is highly subjected to MMR. Errors made by Pol IV at the cII gene of lambda phage were shown to be correctable by MMR [83]. Therefore, it is possible that mutational intermediates generated by Pol IV at the microsatellite versus the coding region are differentially repaired during DNA replication. In this model, MMR more effectively repairs the polymerase errors made in the microsatellite alleles than those that are made in non-repetitive tracts (in our assay, the errors made in the HSV-tk coding region). This model is consistent with the trend we observed for lower microsatellite mutation frequencies in the absence of dinB that did not reach statistical significance (Tables 3.2 and 3.5). Our data also showed a larger number of insertion events in the mononucleotide [G/C]_{10} allele (Table 3.2). This result is consistent with both this proposed model and previous literature that showed MMR removed -1 bp frameshift intermediates more efficiently than +1 bp frameshift intermediates in a 10 bp mononucleotide G run [260].

Alternatively, a second model would be that the mutations in the microsatellite alleles arise primarily from mitotic recombination, implying that polymerase slippage is not playing a role in mutagenesis at the repeat tracts. In this model, the errors produced within the HSV-tk coding region would not arise due to polymerase events, but the errors at the microsatellite alleles are produced by recombination events. This model would be consistent with the specificity results we obtained for the range of changes at the microsatellite allele for the dinucleotide vectors (Figure 3.3). The primary mutational event at a repetitive allele is gain or loss of 1 unit repeat. For the dinucleotide vectors we observed a large range, up to gain or loss of 5 units. These larger unit changes could be the result of a recombination event.
A final model to explain our observations is that Pol IV is not utilized during DNA synthesis within the microsatellite region, but is engaged in DNA synthesis within the coding region. Further experimentation will be required to differentiate among the models. A description of experiments is discussed in Chapter 6.

There are many reports in the literature showing that the sequence composition and length of the repeat allele plays a large role in the mutagenesis of microsatellites [258, 260, 302, 326]. We compared various aspects of the mutational differences detected among the three microsatellite alleles. Chi Square analyses supported a significant effect of the allele sequence on the distribution, type, and range of mutational events observed. However, the background of the cells, proficient or deficient in \textit{dinB}, does not affect mutagenesis. Each repeat is unique in composition and may be handled differently by the gauntlet of polymerases present in the cell. In the case of this study, we conclude that \textit{dinB} does not affect the mutation frequency or specificity at the microsatellites examined. We also confirmed the differential mutagenesis of microsatellites that is dependent upon the composition of the repeat allele.
Chapter 4: Spontaneous Mutagenesis of the \([G/C]_{10}\) Mononucleotide Microsatellite

4.1 Rationale/Hypothesis

From analyses of the human genome sequence we have learned that among the 3 types of repetitive DNA sequences in the genome, microsatellite sequences are the most common. Further analyses of the different subunit length of microsatellite repeats, mono-, di-, tri-, tetra-, and pentanucleotide, have found that mononucleotide motifs are the most common repeat tract in the genome.

For both of the possible combinations of mononucleotide alleles, A/T and G/C, many studies have been performed in *E. coli* and *S. cerevisiae* looking at mutation frequencies and rates, as well as mutational specificity. In mammalian cells however, this is not the case, and limited number of studies have examined microsatellite mutagenesis, specifically in human cells. One group has examined microsatellite mutagenesis in mouse cells by using a reversion mutational system, which does not allow the detection of all possible mutational events, and thus may be misrepresenting the actual mutagenesis of repeat alleles. In contrast forward assays systems have been used in *E. coli* and *S. cerevisiae* that are capable of detecting all classes of mutational events. The desire for a forward assay system in mammalian cells led our lab to develop an OriP shuttle vector system that can be used for forward assays in human cells. A further advantage of this system is that it can also be used in *E. coli* cells. This system uses the HSV-tk gene as the mutational target and has the microsatellite alleles inserted into the 5’ region of the gene. It contains antibiotic resistance genes for selection in both human cells (hygromycin) and in *E. coli* (chloramphenicol).
The lab has also developed an *in vitro* forward assay system that utilizes the HSV-tk gene as a mutational target, and is described in detail in Chapter 2. The main purpose of this assay is to examine the role of individual DNA polymerases in mutagenesis, both at the microsatellite repeat allele and within an 88 bp coding region that is assayed in this system. By allowing an individual polymerase to replicate through the sequences, we are able to determine the error frequency and the mutational specificity induced by that particular polymerase.

Using a combination of these systems, the goal of this research was to determine the endogenous mutation rate and mutational specificity at the \([G/C]_{10}\) mononucleotide microsatellite, as well as to determine whether a particular polymerase might be involved in mutations at repeated sequences. We hypothesized that an increased overall mutation rate would occur at the repeat alleles, compared to a sequence with no repetitive nature.

### 4.2 Results

#### 4.2.1 Shuttle Vector Assay with 721 Cells and \([G/C]_{10}\) Shuttle Vector

Using this assay system described above, we set forth to determine the mutagenesis at mononucleotide repeats in human cells. Human lymphoblastoid cell line LCL721 was electroporated with the \([A/T]_{11}\) and \([G/C]_{10}\) shuttle vectors. The electroporated cells were then cultured, cloned, and expanded as described in Chapter 2. Our hypothesis was that a vector containing the \([G/C]_{10}\) mononucleotide allele would have an increased mutation rate, both overall and at the microsatellite allele, compared to a vector that contains no repeat allele.
4.2.1.1 Mutation Rate Analysis of Human Cell Clones

For the control vector the mutation rates ranged from 1.8 - 8.5 x10^{-5}, with a median mutation rate of 2.4x10^{-5}, shown in Table 4.1. For the vector containing the [G/C]_{10} repeat allele, the range of mutation rates was 4.1 x10^{-5} to 10.5 x10^{-5}, with a median mutation rate of 5.5 x10^{-5}, as shown in Table 4.2. Statistical analysis (Wilcoxon Mann Whitney non-parametric test) showed that the overall HSV-tk mutation rate for the control and [G/C]_{10} repeat vectors are not statistically different (p=0.15).

4.2.1.2 Mutational Spectra Analysis of Human Cell Clones

Given that the mutation rates between the control and [G/C]_{10} containing vectors were not statistically significant, we wanted to determine the specificity of mutagenesis at the microsatellite allele and the coding region of the HSV-tk gene. We sequenced 10-20 independent mutants from 5 different LCL721 clones containing the vector with the [G/C]_{10} allele, for a total of 72 mutants sequenced. For the [G/C]_{10} vector, the mutational specificity among the 5 clones examined showed a larger proportion of events occurring in the coding region of the HSV-tk gene than within the artificial microsatellite sequence. Table 4.3 shows the overall mutation rates for each of the 5 clones sequenced, as well as the estimated microsatellite and coding region mutation rates. The median mutation rate at the microsatellite allele was 2.9 x10^{-5}, while the coding region was 2.7 x10^{-5}. Of all of the mutational events observed, 36% occurred at the microsatellite repeat, compared to 64% within the coding region. Further analysis found that the specificity of mutations at the microsatellite was 81% insertion events and 19% deletion events.
### Table 4.1: Mutation Frequencies and Rates for HSV-tk Control Clones

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10^{-3})^a</th>
<th>HSV-tk Mutation Rate (x10^{-5})^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.17</td>
<td>0.7</td>
<td>2.7</td>
</tr>
<tr>
<td>1.20</td>
<td>2.4</td>
<td>8.5</td>
</tr>
<tr>
<td>1.21</td>
<td>2.0</td>
<td>7.0</td>
</tr>
<tr>
<td>1.25</td>
<td>0.5</td>
<td>1.8</td>
</tr>
<tr>
<td>1.27</td>
<td>0.5</td>
<td>1.8</td>
</tr>
<tr>
<td>1.29</td>
<td>0.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Median</td>
<td>0.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Range</td>
<td>0.5-2.4</td>
<td>1.8-8.5</td>
</tr>
</tbody>
</table>

^a Shuttle vector DNA isolated from individual LCL721 clones containing the [G/C]_{10} repeat was electroporated into *E coli* strain FT334. Bacteria were plated on selective media, Cm to select for *E coli* strain FT334. From these plate counts, the mutation frequency was determined for 10 independent LCL721 clones.

^b Mutation Rate was determined using the calculation: MF/N, where N is the number of population doublings the cells underwent between the time of cloning and DNA isolation.
Table 4.2: Mutation Frequency and Rate for 721 Clones Containing the [G/C]_{10} Microsatellite Containing Shuttle Vector

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10^{-3})&lt;sup&gt;a&lt;/sup&gt;</th>
<th>HSV-tk Mutation Rate (x10^{-5})&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.2</td>
<td>2.0</td>
<td>7.7</td>
</tr>
<tr>
<td>11.3</td>
<td>1.9</td>
<td>6.8</td>
</tr>
<tr>
<td>11.4</td>
<td>3.2</td>
<td>10.5</td>
</tr>
<tr>
<td>11.5</td>
<td>1.7</td>
<td>5.7</td>
</tr>
<tr>
<td>11.6</td>
<td>1.2</td>
<td>4.3</td>
</tr>
<tr>
<td>11.7</td>
<td>1.2</td>
<td>4.2</td>
</tr>
<tr>
<td>11.11</td>
<td>1.1</td>
<td>4.1</td>
</tr>
<tr>
<td>11.12</td>
<td>2.4</td>
<td>7.1</td>
</tr>
<tr>
<td>11.13</td>
<td>1.6</td>
<td>5.2</td>
</tr>
<tr>
<td>11.14</td>
<td>1.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Median</td>
<td>1.7</td>
<td>5.5</td>
</tr>
<tr>
<td>Range</td>
<td>1.1-3.2</td>
<td>4.1-10.5</td>
</tr>
</tbody>
</table>

<sup>a</sup> Shuttle vector DNA isolated from individual LCL721 clones containing the [G/C]_{10} repeat was electroporated into *E coli* strain FT334. Bacteria were plated on selective media, Cm to select for cells containing the plasmid and Cm+FudR to select for mutants. From these plate counts, the mutation frequency was determined for 10 independent LCL721 clones.

<sup>b</sup> Mutation Rate was determined using the calculation: MF/N, where N is the number of population doublings the cells underwent between the time of cloning and DNA isolation.
Table 4.3: Mutational Specificity for LCL721 Clones Containing the $[G/C]_{10}$ Microsatellite Repeat

<table>
<thead>
<tr>
<th>Clone</th>
<th>Overall MR (x10^{-5})&lt;sup&gt;a&lt;/sup&gt;</th>
<th>STR MR (x10^{-5})&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Coding MR (x10^{-5})&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Number of Mutants</th>
<th>Insertions&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Deletions&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.2</td>
<td>7.7 (17)</td>
<td>3.4 (7)</td>
<td>4.3</td>
<td>5 (0.71)</td>
<td>2 (0.29)</td>
<td></td>
</tr>
<tr>
<td>11.6</td>
<td>4.3 (17)</td>
<td>1.6 (5)</td>
<td>2.7</td>
<td>3 (0.60)</td>
<td>2 (0.40)</td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>4.2 (11)</td>
<td>1.4 (3)</td>
<td>2.8</td>
<td>3 (1.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>11.11</td>
<td>4.1 (17)</td>
<td>2.9 (5)</td>
<td>1.2</td>
<td>5 (0.80)</td>
<td>1 (0.20)</td>
<td></td>
</tr>
<tr>
<td>11.14</td>
<td>4.5 (10)</td>
<td>3.0 (6)</td>
<td>1.5</td>
<td>6 (1.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
</tbody>
</table>

Total Mutants 72 26 46 22 5
Summation - 36% 64% 81% 19%

MR: Mutation Rate
STR: $[G/C]_{10}$ microsatellite

For 5 of the LCL721 clones, the same shuttle vector DNA described in Table 4.2 was also electroporated into *E. coli* strain FT334 for mutational spectra analysis. Each electroporation was then split into 50µl aliquots, added to a tube containing 1ml VBA broth, and grown for 2 hours. Each tube was then plated on an individual Cm+FUdR plate and incubated overnight. This was done to ensure that each colony picked was unique, prohibiting the possibility that sibling cells would be examined. A single colony was selected from each plate, grown overnight, and plasmid DNA was isolated. The DNA was then amplified by PCR and sequenced on the CEQ8000 DNA sequencer (Beckman Coulter). Median values are indicated in red.

<sup>a</sup> Number of Mutants sequenced shown in parenthesis
<sup>b</sup> Proportion of mutants within $[G/C]_{10}$ microsatellite shown in parenthesis
4.2.2 Pol β in vitro reactions

With the completion of the experiments looking at the mutation rates and specificity in the LCL721 human cell clones it became apparent that there was an effect on the mutagenesis of the [G/C]_{10} microsatellite allele compared to a no repeat control. We next wanted to try and determine what polymerase might be responsible for introducing these mutations. We focused our attention on the non-replicative DNA polymerases (refer to Figure 1.1B), particularly Pol β, to determine its possible role in mononucleotide microsatellite mutagenesis. We chose to examine Pol β, as previous studies completed in our lab have indicated that Pol β resulted in differential mutagenesis across varied templates as well as showing a strand bias for microsatellite mutagenesis [323, 327]. Eckert and Hile also found that Pol β produced predominantly template directed frameshift mutations, a large majority of which were in short repetitive tracts of mononucleotide G and C tracts [328]. Additionally, Pol β has been found to be mutated in 30% of tumors examined (reviewed in [203]). Therefore, determining a baseline for mutagenesis caused by this polymerase could aid in understanding disease progression and/or treatment.

4.2.2.1 G_{10} Template

Using a DNA template containing the G_{10} sequence, the observed median Pol β error frequency was 45±3 x10^{-3}. DNA sequence analysis of mutants found that 71% of the mutation events occurred at the microsatellite repeat, corresponding to a Pol β error frequency at the G_{10} repeat allele of 38 x10^{-3}. Mutation specificity is shown in Figure 4.1. Breaking down the mutations, an almost even split of insertion and deletion events were detected (51% and 49%, respectively).
4.2.2.2 C\textsubscript{10} Template

The same assay was completed on a template containing a C\textsubscript{10} repeat sequence. In this case, the median Pol β error frequency was determined to be 27±5.5 \times 10^{-3}. As was observed for the complementary template, the large majority of the mutation events occurred at the repeat allele (88%), with the specificity shown in Figure 4.2. Based on this percentage, the estimated Pol β error frequency at the C\textsubscript{10} microsatellite repeat is 26 \times 10^{-3}. For the mutations that occurred at the repeat, 38% were insertions, while 62% were deletion events.

The range of Pol β error frequency values obtained for the G\textsubscript{10} and C\textsubscript{10} templates are not statistically different (p=0.1, WilcoxonMannWhitney Test), indicating that Pol β acts on the two repeat sequences in the same manner. Statistical analysis also supports this theory when looking at the percentage of mutations at the repeat allele and the distribution of the mutations between insertion and deletion events. For both templates, the percentage of mutations made by Pol β at the repeat allele is statistically higher than those made outside of the repeat (p=0.004 Fishers Exact Test). However the 71% of events observed at the STR for the G\textsubscript{10} template is not statistically different from the 88% observed on the C\textsubscript{10} template (p=0.65). Also, a comparison of the proportion of insertion and deletion events for each template indicated that there was not a statistically significant difference (p=0.35 Fishers Exact test).
Figure 4.1: Pol β Mutational Spectra on G₁₀ Template: Errors made by Pol β were determined by allowing the polymerase to replicate a ssDNA template followed by hybridization into a gapped heteroduplex molecule. This molecule was then electroporated into *E. Coli* strain FT334 and a single colony was selected per plate for DNA sequencing to determine the polymerase Mutational Specificity.
Figure 4.2: Pol β Mutational Spectra on C₁₀ Template: Errors made by Pol β were determined by allowing the polymerase to replicate a ssDNA template followed by hybridization into a gapped heteroduplex molecule. This molecule was then electroporated into E. Coli strain FT334 and a single colony was selected per plate for DNA sequencing to determine the polymerase Mutational Specificity.
4.3 Discussion

The initial experiments discussed in this chapter were conducted with both the [A/T]_{11} and [G/C]_{10} vectors. Unfortunately, as discussed in Chapter 3, we were unable to carry the experiments with the [A/T]_{11} repeat allele to completion. We did however continue the experiments with the [G/C]_{10} allele.

4.3.1 HSV-tk Assay in Human Cells

Using the shuttle vector assay we were able to determine the mutation rate for the [G/C]_{10} mononucleotide sequence. The median mutation rate for the microsatellite sequence was 29x10^{-6}. Table 4.4 provides a comparison of the mutation rate of all microsatellite sequences examined thus far in human cells from our lab [329]. The [G/C]_{10} microsatellite allele has a higher mutation rate than dinucleotide alleles of approximately the same repeat length, and some tetranucleotide sequences. We compared the microsatellite mutation rate for the [G/C]_{10} and [GT/CA]_{10}, two microsatellite alleles that are equal in number of repeated units. We observed a 15-fold increase for the mononucleotide allele, a value that was statistically significant (p<0.01, Wilcoxon Mann Whitney Test). The observed increase in the mutation rate is consistent with previous studies that examined mononucleotide sequences. Boyer et al. examined mouse cells containing vectors with different microsatellite alleles, including a [G/C]_{17} microsatellite sequence. They found that the mutation rate for the [G/C]_{17} repeat was significantly higher than both the [A/T]_{17} and [CA/GT]_{17} repeat alleles (values ranging from 10-100-fold increase for the [G/C]_{10} sequence) [265].
Table 4.4: Microsatellite Mutation Rates for Human Cell Clones

<table>
<thead>
<tr>
<th>Allele</th>
<th>STR Mutation Rate$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[G/C]$_{10}$</td>
<td>$29 \times 10^{-6}$</td>
</tr>
<tr>
<td>[GT/CA]$_{10}$</td>
<td>$\leq 1.9 \times 10^{-6}$</td>
</tr>
<tr>
<td>[TC/AG]$_{11}$</td>
<td>$3.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>TC/AG]$_{17}$</td>
<td>$9.8 \times 10^{-6}$</td>
</tr>
<tr>
<td>TC/AG]$_{20}$</td>
<td>$21 \times 10^{-6}$</td>
</tr>
<tr>
<td>[TTCC/AAGG]$_9$</td>
<td>$5.5 \times 10^{-6}$</td>
</tr>
<tr>
<td>[TTTC/AAAG]$_9$</td>
<td>$35 \times 10^{-6}$</td>
</tr>
<tr>
<td>[TCTA/AGAT]$_9$</td>
<td>$38 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

$^a$ Data are from thesis and reference [329]
The important implication of our results is in regard to total repeat length. Kelkar et al. performed the first genome-wide study to examine the mutability among orthologous human–chimpanzee microsatellites. In regard to mononucleotide alleles, they once again reported the abundance of A/T sequences compared to G/C sequences. However, they also reported that mononucleotide G/C sequences longer than 20 bp in length were not detected in their study. They also state that sequences longer than 17 bp are relatively rare in occurrence [330]. Given these results, the previous studies involving mononucleotide G/C alleles have examined allele lengths at the limit or longer than what are observed in the human genome. We have looked at a repeat that was only 10 bp in length, which falls well within lengths observed by Kelkar et al. Given that these results are more relevant to lengths observed in the genome, they may be a more accurate representation than those previously reported.

The sequence analysis we performed for the [G/C]_{10} allele resulted in 36% of the mutational events occurring at the microsatellite allele. Unfortunately it is difficult to compare our sequencing results with those in the studies previously discussed. Tobi et al. found many of the mutations occurred at locations that included parts of the repeat allele; however, they also encompassed sequence outside of the repeat in one mutational event [308]. The study by Boyer et al. was performed as a reversion assay, thus presenting a problem for direct comparison with our results, as only one class of mutational events can be observed for the repeat allele. Our study is a forward assay, which is able to detect all classes of mutational events. Therefore, our assay is more sensitive and can detect a larger array of mutational events, thus preventing direct comparison with previous results. We feel that our sequencing results for the [G/C]_{10} microsatellite allele provide
the first complete view of all possible mutagenesis events in human cells, as well as a more accurate representation to their proportions.

The sequence data that we have obtained for the \([G/C]_{10}\) microsatellite many also hold some clinical relevance. Duval et al. examined MMR deficient tumors and found that many of the genes commonly mutated contained mononucleotide repeats [236]. Of the genes examined containing mononucleotide microsatellites, the repeat tracts are no longer than 10 bp in length. This fact, taken in conjunction with the study by Kelkar et al. that found that shorter G/C sequences are found in the human genome, lends even more importance to our study. The mutagenesis we determined for the \([G/C]_{10}\) microsatellite allele could potentially lead to a diagnostic tool used for determining altered allele length in cancer or other diseases.

4.3.2 HSV-tk in vitro Assay

Using the HSV-tk assay designed in our lab, we examined the error frequency of Pol β synthesis using \(G_{10}\) and \(C_{10}\) containing templates. The results of our study indicated that Pol β makes more errors at the microsatellite allele than within the coding region of the HSV-tk gene. We also observed that Pol β makes the same types of errors, with nearly the same error frequency (38 vs. \(26 \times 10^{-3}\)) on both the \(G_{10}\) and \(C_{10}\) templates. The mutational specificity was approximately equal on both templates, indicating that Pol β is equally active on both templates and makes insertion and deletion events in similar proportions.

Overall, these results are significant when compared to other sequences examined in our lab. Table 4.5 lists all of the mononucleotide and dinucleotide sequences we have
### Table 4.5: Pol β Microsatellite Error Frequencies from *in vitro* Assay

<table>
<thead>
<tr>
<th>STR</th>
<th>Error Frequency (x10^-3)</th>
<th>Errors at STR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_{10}$</td>
<td>38.0 ± 3.0</td>
<td>71%</td>
</tr>
<tr>
<td>$C_{10}$</td>
<td>26.0 ± 5.5</td>
<td>88%</td>
</tr>
<tr>
<td>GT$_{10}^a$</td>
<td>1.1 ± 0.2</td>
<td>27%</td>
</tr>
<tr>
<td>CA$_{10}^a$</td>
<td>2.6 ± 1.0</td>
<td>42%</td>
</tr>
</tbody>
</table>

$^a$ N.S. and K.A.E., Unpublished observations
tested with Pol β as well as the overall error frequency and the error frequency at the microsatellite. (Dinucleotide data were kindly provided by Noelle Strubczewski) We found that the mononucleotide alleles are at least 10-fold more mutable than the dinucleotide sequences. For the dinucleotide alleles, a strand bias was observed where one allele was more mutable than its complementary sequence. A 2.4-fold difference was observed between the GT10 and CA10 sequences. We did not observe this trend with the mononucleotide alleles. The final difference when comparing the mononucleotide and dinucleotide allele, is that the mononucleotide alleles have a larger proportion of events occurring at the microsatellite allele than do the dinucleotide alleles. We observed greater than 71% of the mutational events occurred at the mononucleotide allele, while only 42% were at the STR for the dinucleotide alleles, a difference that is statistically significant (p<0.0001, Fishers Exact Test)

To explain the higher error frequency and mutational specificity at the mononucleotide allele compared to the dinucleotide alleles we have proposed the following hypotheses, which are not mutually exclusive. The first possibility involves the base stacking interactions that occur when a looped out base is encountered. In this theory, the interaction of the G/C base pairs stabilizes the repeat. Once a loop out is created, the thermodynamic properties of the bases stabilize the looped out bases. Without removal of these bases, the microsatellite allele is then altered in length; an expansion event for a loop in the primer and a deletion event for a loop out in the template strand. Another possibility is that the ability to form the loop is different. Due to the number of repeat units in the microsatellite, it is possible that a loop of one base is an easier event than a loop of two bases. If this is the case, then the higher mutation rate
in the mononucleotide could be explained by the higher frequency of creating the single base loop. The final possibility is that Pol β is better able to extend from a loop of one base then it can from a loop of two bases.

4.3.3 Comparison Between Systems

When comparing the data obtained from both the human cell and *in vitro* studies we observed similar increased mutation rates for the mononucleotide alleles compared to the dinucleotide alleles. In human cells, the mononucleotide allele was elevated 15-fold over the dinucleotide allele, and in the *in vitro* assay we observed a minimum 10-fold increase in the mononucleotide alleles. A similar difference between the mono- and dinucleotide alleles was observed when looking at the microsatellite mutational specificity. In human cells, we observed a lower proportion of mutational events at the microsatellite than we did in the *in vitro* assay for the mononucleotide alleles. Also, when breaking down the mutational events, insertion events outnumbered deletion events by a ratio of 4:1 in human cells; however, we observed nearly equal insertion and deletion events in the *in vitro* assay. The differences we observed between the human cell and *in vitro* assay may be explained by DNA repair. In human cells, MMR is active. Thus, if mutation events were created at different ratios by DNA polymerases, for example Pol β, they could be corrected, thus altering the mutational profile observed for any one polymerase.

To determine if MMR is altering the mutagenesis we have observed with any one polymerase in the *in vitro* assay we could use a MMR deficient cell line to test our templates of interest. We would examine the mutation rate in a cell that is MMR
deficient as well as in the same cell line complemented with a cDNA of the missing components of the MMR system. We would monitor for MMR protein levels by Western blot to ensure the complementation worked. This way, a comparison of mutation rates can be completed within the exact same cell background, therefore eliminating any cell line variation that could result in differential mutagenesis.
Chapter 5: Development of a System to Genetically Manipulate Polymerase Levels in Human Cells

5.1 Rationale/Hypothesis

The main goal for this project was to determine the effect of altered DNA polymerase levels on mutagenesis of our artificial microsatellite containing vectors. It has been shown that polymerases are altered in different cancers, leading us to inquire what effect these altered levels have on mutagenesis, particularly at microsatellite sequences. In order to achieve this goal, we first needed to develop a method to stably knockdown polymerase levels in our LCL721 human lymphoblastoid cell line. Once we were successful in knocking down polymerase levels, we could examine the effect of these altered levels on mutagenesis.

A study by Manavel and Eckert showed that human cells in culture are susceptible to oxidative stress induced microsatellite mutagenesis. In this study, non-tumorigenic human cells treated with H₂O₂ had a statistically significant increase in mutagenesis at the HSV-tk gene and a [TTCC/AAGG]₉ microsatellite allele, relative to untreated controls [311]. Additional information to support our experimental design lies in the fact that oxidative damage is repaired by the BER pathway, which utilizes Pol β to repair errors. Based on this knowledge, we chose to examine Pol β. If Pol β levels are reduced, then we would expect increased mutagenesis due to the fact that errors arising from oxidative stress cannot be repaired. In particular, guanine base pairs are especially susceptible to oxidative damage [331, 332]. Given the fact that cells in culture are under oxidative stress and that the BER pathway which usually deals with this damage will be
impaired, we hypothesize that reducing Pol β levels will result in an increase in mutagenesis at the mononucleotide [G/C]10 allele.

5.2 Results

5.2.1 Evaluation of Transfection Methods to Introduce Genes of Interest

5.2.1.1 Electroporation

To evaluate transfection efficiency of LCL721 cells by electroporation, we examined three different GFP containing vectors, each expressing GFP from a different promoter: EGFP-N1 (CMV Promoter), pHβAPr-EGFP (human beta actin promoter), and pL-UGIP (ubiquitin promoter). As a positive control, we also electroporated 293 cells with each of these vectors as these cells have successfully expressed the vectors previously (communication with Dr. John Harms). Cells were visually monitored for GFP expression for a period of seven days. For the LCL721 cells, no GFP was detected with any of the three GFP containing vectors during the entire seven days period. Despite the fact the two different electroporation experiments were completed for the 293 cells, in both cases the result was complete cell death within one day of electroporation. Based on these results, we concluded that electroporation was an inefficient method of choice for introducing genes into either the LCL721 or 293 cells.

5.2.1.2 Lipofection

We examined transfection efficiency by various lipofection methods, visualizing GFP expression. Three different reagents were evaluated to determine their efficiency for transient expression of GFP. Using Lioprectamine2000, we observed no GFP positive cells for the 721 cells after seven days. However, for the 293 cells, GFP positive cells were detected for two of the three vectors tested, as shown in Figure 5.1. From this
analysis we concluded that the EGFP-N1 and pHβAPr-EGFP were the most-likely candidates to give us GFP positive cells. The DMRIE-C reagent (Invitrogen) is a lipofection reagent designed for transfection of suspension cells. After transfection following company protocol (as described in Chapter 2) we were unable to visually detect any GFP positive cells for either the LCL721 cells or the 293 cells, using the EGFP-Ni and pHβAPr-EGFP vectors. FugeneHD transfection yielded GFP positive LCL721 cells. The best conditions were cells seeded at 1x10^6 cells/ml, with a 4:2 ratio of DNA to reagent, and either a 15 minute or 30 minute complex incubation time. Under these conditions our estimate is that 0.1-1% of the total cells were GFP positive. After a period of seven days we were no longer able to visually detect GFP expression in LCL721 cells.

5.2.2 Evaluation of Viral Transduction Methods

Although with the Fugene HD lipofection reagent we were able to obtain transient expression of GFP, we did not feel that this method would be feasible. Thus we evaluated whether lentiviral transduction would efficiently integrate our gene of interest into the LCL721 cells.

5.2.2.1 Infection of LCL 721 Cells and Detection of GFP Positive Cells

To test the lentiviral infection method, we received packaged and titered pseudo-viral particles with a vector containing a GFP reporter gene from our collaborator Dr. Robert Sobol (University of Pittsburgh) (for vector elements see Figure 2.2). The infection of LCL721 cells was performed as described in Chapter 2. After each infection,
Figure 5.1: GFP Expression in 293 Cells: Cells were transfected, using Lipofectamine2000, with different vectors containing the GFP gene under the control of different promoters. GFP positive cells were detected visually under 20X magnification. A) pHbAPr-EGFP  B) EGFP-N1  C) pL-UGIP  D) 293 cells bright field
the cells were examined visually for GFP expression. For each round of infection, the
unused portion of cells (or for the third infection, cells on the 6th day following infection)
were analyzed by FACS to determine the percentage of GFP positive cells. For each of
the three rounds of infection, we were able to detect some level of GFP positive cells. As
shown in Figure 5.2, after the first infection we were able to detect ~17% GFP positive
cells. After the second and third infection, this value increased to 56% and 50% respectively. We conclude that the lentiviral infection method was successful in
introducing a gene of interest into the LCL721 cells at levels that would be
experimentally feasible, and decided that three rounds of infection would ensure that the
largest proportion of cells were able to receive the vector containing our gene of interest.
We chose three rounds of infection to ensure that transmission of the full lentiviral
cassette would occur in the largest portion of our cells. It is possible that two rounds of
infection could leave the cells GFP positive, but not allow for full transmission of the
shRNA cassette. This would result in a transient expression of GFP by FACS analysis,
but the cells may not have received the shRNA sequence, thus Pol β would not be
knocked down.

5.2.3 Determination of Polymerase Levels in LCL 721 Cells

In order to determine if we were successful at altering polymerase levels in the
LCL721 cells, we first examined endogenous protein levels for Pol δ, Pol β, Pol η,
PCNA, and Lamin A/C (Figure 5.3). In the Jurkat cell lysate we were able to detect
Pol δ, Pol β, and PCNA, but not Pol η or Lamin A/C. For the MCF7 cells we were able
to detect Pol δ and PCNA in both fractions, and Lamin A/C in only the nuclear. We were
unable to detect either Pol η or Pol β in this cell line. In the LCL721 cell line we were
Figure: 5.2 FACS Analysis of LCL721 Cells following Lentiviral Infection with the shRNA Vector Containing a GFP Reporter
Figure 5.3: Western Blot of Endogenous Protein Levels: To determine endogenous protein levels, we used the NerPer Protein isolation kit (Peirce) as described in Chapter 2 to isolate protein from LCL721 cells as well as 293 cells. Western blots were run on 10% BisTris gels with nuclear extracts from various control samples as well as MagicMark, a chemiluminescent marker.
able to detect all of the proteins, with the exception of Lamin A in the cytoplasmic fraction. 293, the final cell line, expressed the Pol δ, Pol η, Pol β, and PCNA protein. Again, we were unable to detect the Lamin A/C protein in these cells.

5.2.4 Derivation of Putative Knockdown Clones

Human cells containing the [G/C]$_{10}$ mononucleotide repeat vector were used to determine the effect of altered Pol β levels in mutagenesis. A clone (11.7) (described in section 4.2.1) was chosen at random to complete these studies.

5.2.4.1 Control Containing shRNA Vector

A control lentiviral vector was created which contained an shRNA sequence to the luciferase gene, which is not expressed in our cells. The 11.7 clonal cells were infected three times with the control lentiviral vector as described in Chapter 2. After the three rounds of infection, the cells were placed into puromycin selection, followed by single cell cloning. The clones were then expanded to create 10 clonal cell lines that contained the control vector as well as the vector containing the [G/C]$_{10}$ repeat allele.

5.2.4.2 Pol β shRNA Containing Vectors

The pseudo-viral stocks for three different Pol β shRNA containing vectors were a generous gift to us from our collaborator D. Robert Sobol (University of Pittsburgh). Figure 2.3 depicts the location of our three shRNA sequences along the Pol β mRNA as well as their sequences. The shRNA containing vectors are named Pol β-1, Pol β-2, and Pol β-3, starting with the sequence in the 3’ end of the gene being Pol β-1 and the
moving toward the 5’ end of the mRNA. As was done for the control lentiviral vector, clone 11.7 cells were infected three times with one of the three Pol β shRNA containing vectors. After completion of the infections, the cells were placed in antibiotic selection, and then single cell cloned. The clones were expanded to result in 10-20 clonal cell lines that contained both the \([G/C]_{10}\) microsatellite vector as well as one of the shRNA sequences to Pol β.

5.2.5 Characterization of Knockdown and Control Clones

5.2.5.1 RNA levels

We analyzed the mRNA levels using RT-PCR. Four sets of RT-PCR primers were designed that spanned the entire length of the Pol β mRNA sequence (Figure 5.4). We observed that Pol β mRNA was present in each of the Pol β-1 and control clones examined. To determine whether mRNA levels varied quantitatively, we used qRT-PCR in a TaqMan assay. Pol β levels were compared to the levels of a non-specific housekeeping gene, TATA Binding Protein (TBP) in each clones. The RNA from each of the clones containing the Pol β-1 shRNA sequence, as well as the luciferase control shRNA clones, were analyzed. The control mRNA was isolated from LCL721 cells.

Figure 5.5 shows the results for the control shRNA containing vector and is a compilation of three individual experiments. As depicted, each of the clones has a varied level of Pol β, with all of the clones having 1.5 to 3.6-fold higher levels of Pol β expression than the LCL721 cells. Results for the Pol β-1 shRNA containing clones are shown in Figure 5.6. For each of the 22 clones analyzed in the three experiments, a range of mRNA levels was observed, with all clones having higher expression than LCL721
Figure 5.4: RT-PCR Primers for Analysis of Pol β mRNA Levels: Primers were designed using Primer-BLAST (NCBI) with accession number NM002690 for the Pol β gene sequence.

**PP12:** AGTACACCATCCGTCCCTTG (forward)
TTGCAAGCAAAAGCATGAAG (reverse)

**PP19:** GGCAGTTTCAGAAGAGGTGC (forward)
TGAAACCCTTTTCTAGGGCA (reverse)

**PP24:** AATCACCGACATGCTCACAG (forward)
CGTATCATCCTGCCGAATCT (reverse)

**PP25:** GAAAAGATTCGGCAGGATGA (forward)
CATCCATGTCACCACCTGGAC (reverse)
Figure 5.5: qRT-PCR Results for Relative Pol β Expression in Luciferase Control Containing Cells: The mRNA was analyzed as described in Chapter 2 and the ratio of clone:LCL721 was determined. The standard deviation was determined as the square root of the squares for both Pol β and TBP. Each bar represents the average of three qRT-PCR runs.
Figure 5.6: qRT-PCR Results for Relative Pol β Expression in Pol β-1 shRNA Containing Cells: The mRNA was analyzed as described in Chapter 2 and the ratio of clone:LCL721 was determined. The standard deviation was determined as the square root of the squares for both Pol β and TBP. Each bar represents the average of three qRT-PCR runs.
cells. However, analysis of each of the Pol β-1 clones shows that the error bars are overlapping and statistical analysis found no significant difference in Pol β expression among any of the clones (p=0.55). This indicates that the Pol β-1 shRNA containing clones do not have reduced levels of Pol β mRNA. All of the knockdown clones and the control containing clones have higher levels of Pol β than LCL721 cells, which may be a consequence of the lentiviral infection.

5.2.5.2 Protein Levels

In order to confirm the lack of effectiveness of shRNA knockdown, protein was isolated from each clone, as well as LCL721 cells containing no shRNA sequence. Western blots were completed to look at the levels of Pol β, relative to Proliferating Cell Nuclear Antigen (PCNA) as a nuclear control (Figure 5.7). For the Pol β knockdown clones, as well as the luciferase control clones, we detected a higher ratio of Pol β to PCNA than in the LCL721 cells (Figure 5.8). This is consistent with increased Pol β levels due to infection. We hypothesize that this effect is a result of the infection process, as the increase was detected in all of the lentiviral infected clones. Statistical comparison of each of the three Pol β shRNA containing constructs to the Luciferase control vector found no differences in the ratio of Pol β:PCNA expression (p=0.64, p=0.31, and p=0.6 respectively, Wilcoxon Mann Whitney Test).

5.2.5.3 Analysis of gDNA for shRNA Insert and the Puromycin Resistance Gene

gDNA was isolated from Pol β-1 knockdown clones, Luciferase control clones, and LCL721 cells and was subjected to PCR for both the shRNA insert and puromycin
Figure 5.7: Sample Western Blot of Pol β knockdown clones: To determine endogenous protein levels, we used the NerPer Protein isolation kit (Peirce) as described in Chapter 2 to isolate protein from clonal cells. Western blots were run on 10% BisTris gels with nuclear extracts from various samples as well as MagicMark, a chemiluminescent marker.
Figure 5.8: Ratio of Pol β:PCNA Calculated from Western Blot Analysis: Western Blots were run as previously described and probed with primary antibodies for Pol β and PCNA, and analyzed on a phosphoimager using ImageQuant software. The quantitated results were analyzed and the ratios graphed.
resistance gene as described in Chapter 2 and shown in Figure 2.6. Both PCR and sequencing of the gDNA for the shRNA insert resulted in no confirmation of the presence of this element. Experiments to determine if the puromycin resistance gene was present in our cells resulted in a band of the correct size by PCR analysis in luciferase control plasmid and Pol β knockdown clones but not in LCL721 cells. Sequencing of the purified PCR products found that in the luciferase control plasmid the product was in fact the puromycin resistance gene. However, in the Pol β knockdown clones the PCR product was not the puromycin resistance gene. This indicates to us that viral infection was not successful. In this case, the clones that were expanded must have had a mutation that rendered them puromycin resistant, thus allowing them to grow under the selection.

5.2.6 Determination of the Effectiveness of Pol β-1 Knockdown Sequence

We examined whether the Pol β-1 shRNA sequence was capable of knockdown in LCL721 cells, as our collaborator has obtained what appeared to be nearly 100% Pol β knockdown in MDA MB231 cells [321]. siRNA oligos of the same sequence were used to transiently transfect LCL721 cells as described in Chapter 2. RNA was isolated every 24 hours for a total of five days, and analyzed by qRT-PCR analysis. Figure 5.9 depicts the average results for the three reactions, in which all samples are normalized to LCL721. A decrease in Pol β levels was detected on Day 1, consistent with an effect of the siRNA on Pol β. However, this decrease disappears by Days 2 and 3.
Figure 5.9: qRT-PCR Results for Relative Pol β Expression with Transient Transfection of Pol β-1 siRNA: The mRNA was analyzed as described in Chapter 2 and the ratio of clone:LCL721 was determined. The standard deviation was determined as the square root of the squares for both Pol β and TBP. Each bar represents the average of three qRT-PCR runs.
5.2.7 Shuttle Vector Assay with Cells Containing the \([G/C]_{10}\) Vector and Infected with the Control shRNA

To determine the effect of lentiviral infection on mutagenesis of the \([G/C]_{10}\) repeat containing vector, we performed mutational analyses with shuttle vector DNA isolated from control luciferase shRNA containing clones.

5.2.7.1 Mutation Rate Analysis of Control shRNA Containing Clones

The first analysis performed on these clones was to determine the mutation rate of the \([G/C]_{10}\) containing vector. Mutation rates were determined for 10 individual clones each containing the repeat containing vector as well as the control shRNA vector. The range of mutation rates was $2.8 \times 10^{-6}$ to $7.0 \times 10^{-6}$, with a median value of $1.3 \times 10^{-5}$ (Table 5.1). A comparison of these clones to those that contain the \([G/C]_{10}\) repeat vector but were uninfected with the lentivirus (Table 4.2) was also completed. The decrease in mutation rate in the infected cells ($1.3 \times 10^{-5}$) compared to those that were uninfected ($5.5 \times 10^{-5}$) was statistically significant ($p<0.0001$, Wilcoxon Mann Whitney).

5.2.7.2 Mutational Spectra Analysis of Control shRNA Containing Clones

In order to determine the mutational specificity of the \([G/C]_{10}\) containing vector in cells containing the control shRNA vector, four clones were chosen for further analysis, for a total of 82 mutants examined (Table 5.2). We observed that 89% of the mutational events occurred at the \([G/C]_{10}\) microsatellite repeat, accounting for 73 of the mutants. This was different than the clones that were not infected with lentivirus (Table 4.3), where we observed 36% of the events to be at the microsatellite allele and is statistically significant ($p<0.001$, Fishers Exact Test). Also shown in Table 5.2, the range of STR
### Table 5.1: Mutation Frequencies and Mutation Rates for Luciferase Control Clones

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10⁴)</th>
<th>HSV-tk Mutation Rate (x10⁻⁵)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucif 1-1</td>
<td>6.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Lucif 2-1</td>
<td>2.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Lucif 2-3</td>
<td>2.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Lucif 2-4</td>
<td>7.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Lucif 2-5</td>
<td>3.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Lucif 2-6</td>
<td>3.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Lucif 5-1</td>
<td>1.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Lucif 5-2</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Median</td>
<td>2.8</td>
<td>1.3</td>
</tr>
<tr>
<td>Range</td>
<td>2.1-7.5</td>
<td>0.7-2.8</td>
</tr>
<tr>
<td>Clone</td>
<td>N</td>
<td>Overall MR (x10^{-6})</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Lucif 2-1</td>
<td>16</td>
<td>9.0</td>
</tr>
<tr>
<td>Lucif 2-3</td>
<td>18</td>
<td>9.0</td>
</tr>
<tr>
<td>Lucif 2-5</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>Lucif 5-1</td>
<td>24</td>
<td>7.0</td>
</tr>
<tr>
<td>Total Mutants</td>
<td>82</td>
<td>-</td>
</tr>
</tbody>
</table>

MR: Mutation Rate  
N: Number of Mutants Analyzed  
STR: \([G/C]_{10}\) microsatellite  
\(^a\) Proportion of insertions or deletions within the \([G/C]_{10}\) microsatellite
mutation rates was 6.4 to 9.2 x10⁶, with the median value being 7.7 x10⁶. This median value was ~4-fold lower than we observed for the uninfected cells. Of the events that occurred at the microsatellite, 56% were insertion events, while 44% were deletion events. Again, this was different than we observed for the uninfected clones, and was statistically significant (p=0.0002, Fishers Exact Test).

5.2.8 Shuttle Vector Assay with Pol β shRNA Infected Cells and [G/C]₁₀ Vector

5.2.8.1 Mutation Rate Analysis of Pol β-1, Pol β-2, and Pol β-3 Clones Containing the [G/C]₁₀ Vector

For each of the three Pol β shRNA containing constructs, 10-20 individual clones were analyzed to determine the mutation rate. Tables 5.3 and 5.4 show the mutation rate values for each of the clones containing one of the three shRNA constructs to Pol β. For the Pol β-1 clones the range of mutation rates was 0.51-6.3 x10⁻⁵, with a median value of 2.5 x10⁻⁵. Similarly, the Pol β-2 clones had a range of mutation rate of from 1.2-20 x10⁻⁵ and a median value of 3.1 x10⁻⁵. The last set of clones, containing the Pol β-3 shRNA construct, ranged in mutation rate from 0.49-6.3x10⁻⁵ and a median value of 1.9 x10⁻⁵. Statistical analyses found that there was no statistical difference in the mutation rates between any of the three Pol β shRNA constructs and the control shRNA-containing vector (p=0.64, p=0.31, and p=0.6 respectively, Wilcoxon Mann Whitney test). Also, a comparison of the mutation rates for each of the three Pol β shRNA constructs found no statistical difference among the clones in the mutation rates (p=0.44, ANOVA).
Table 5.3: Mutation Frequencies and Mutation Rates for Pol β-1 shRNA Containing Clones

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10^-4)</th>
<th>HSV-tk Mutation Rate (x10^-5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol β-1-1</td>
<td>20</td>
<td>4.3</td>
</tr>
<tr>
<td>Pol β-1-2</td>
<td>5.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Pol β-1-3</td>
<td>8.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Pol β-1-4</td>
<td>6.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Pol β-1-5</td>
<td>7.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Pol β-1-6</td>
<td>29</td>
<td>6.3</td>
</tr>
<tr>
<td>Pol β-1-7</td>
<td>20</td>
<td>4.5</td>
</tr>
<tr>
<td>Pol β-1-8</td>
<td>15</td>
<td>3.7</td>
</tr>
<tr>
<td>Pol β-1-9</td>
<td>12</td>
<td>2.9</td>
</tr>
<tr>
<td>Pol β-1-10</td>
<td>1.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Pol β-1-11</td>
<td>14</td>
<td>3.3</td>
</tr>
<tr>
<td>Pol β-1-12</td>
<td>2.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Pol β-1-13</td>
<td>14</td>
<td>3.5</td>
</tr>
<tr>
<td>Pol β-1-14</td>
<td>17</td>
<td>4.2</td>
</tr>
<tr>
<td>Pol β-1-15</td>
<td>6.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Pol β-1-17</td>
<td>8.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Median</td>
<td>10.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Range</td>
<td>1.8-29</td>
<td>0.4-6.3</td>
</tr>
</tbody>
</table>
Table 5.4: Mutation Frequencies and Mutation Rates for Pol β-2 and Pol β-3 shRNA Containing Clones

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10^{-4})</th>
<th>HSV-tk Mutation Rate (x10^{-5})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol β-2-6</td>
<td>16</td>
<td>4.2</td>
</tr>
<tr>
<td>Pol β-2-11</td>
<td>74</td>
<td>20</td>
</tr>
<tr>
<td>Pol β-2-12</td>
<td>17</td>
<td>4.3</td>
</tr>
<tr>
<td>Pol β-2-14</td>
<td>17</td>
<td>4.3</td>
</tr>
<tr>
<td>Pol β-2-15</td>
<td>6.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Pol β-2-16</td>
<td>11</td>
<td>2.8</td>
</tr>
<tr>
<td>Pol β-2-17</td>
<td>7.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Pol β-2-18</td>
<td>5.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Pol β-2-19</td>
<td>13</td>
<td>3.3</td>
</tr>
<tr>
<td>Pol β-2-20</td>
<td>4.8</td>
<td>1.2</td>
</tr>
<tr>
<td>Median</td>
<td>12</td>
<td>3.1</td>
</tr>
<tr>
<td>Range</td>
<td>4.8-74</td>
<td>1.2-20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>HSV-tk Mutation Frequency (x10^{-4})</th>
<th>HSV-tk Mutation Rate (x10^{-5})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol β-3-7</td>
<td>3.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Pol β-3-11</td>
<td>2.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Pol β-3-12</td>
<td>12</td>
<td>2.4</td>
</tr>
<tr>
<td>Pol β-3-13</td>
<td>11</td>
<td>2.2</td>
</tr>
<tr>
<td>Pol β-3-14</td>
<td>14</td>
<td>2.9</td>
</tr>
<tr>
<td>Pol β-3-15</td>
<td>7.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Pol β-3-16</td>
<td>30</td>
<td>6.3</td>
</tr>
<tr>
<td>Pol β-3-17</td>
<td>3.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Median</td>
<td>9.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Range</td>
<td>2.3-30</td>
<td>0.5-6.3</td>
</tr>
</tbody>
</table>
5.2.8.2 Mutational Spectra Analysis of Pol β-1, Pol β-2, and Pol β-3 clones Containing the [G/C]_{10} Vector

Of the clones that were examined for mutation rate, five were chosen for each of the three shRNA constructs for mutational spectra analyses as described in Chapter 2. Shown in Table 5.5 are the clones chosen for mutational spectra analysis as well as their STR mutation rates, and breakdown of STR mutational events. For the clones containing the Pol β-1 construct we observed 83% of the mutational events to be at the [G/C]_{10} repeat allele, and a median STR mutation rate of $2.3 \times 10^{-5}$. The mutation events at the STR break down to 56% insertions and 44% deletions. The Pol β-2 shRNA containing clones resulted in 63% of the mutational events at the microsatellite, with 70% insertions and 30% deletions and a median STR mutation rate of $1.1 \times 10^{-5}$. For the clones with the Pol β-3 shRNA we observed 78% of the mutational events to be at the STR. Of these mutations 63% were insertions and 37% were deletions, with a median STR mutation rate of $1.0 \times 10^{-5}$.

Statistical analyses were performed to determine if the location of the Pol β shRNA sequence has an effect on the mutagenesis of the [G/C]_{10} microsatellite sequence. Comparison of the STR mutation rates for clones containing each of the three Pol β shRNA sequences to the control containing shRNA clones found no statistical difference (p=0.13, ANOVA). Similarly, analysis of the STR mutation rates for each of the three Pol β shRNA containing clones found no statistical difference regardless of the location of the shRNA sequence (p=0.55, p=0.31, and p=0.69, Wilcoxon Mann Whitney; p=0.52, ANOVA). Analyses were also conducted to establish if the location of the Pol β shRNA sequence had an effect on the proportion of mutational events at the microsatellite allele.
Table 5.5: Mutational Specificity for the Pol β shRNA Containing Clones

<table>
<thead>
<tr>
<th>Clone</th>
<th>N</th>
<th>Proportion at STR</th>
<th>Overall MR (x10⁻⁵)</th>
<th>STR MR (x10⁻⁵)</th>
<th>Coding MR (x10⁻⁵)</th>
<th>Proportion Insertions ( ^a )</th>
<th>Proportion Deletions ( ^a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol β-1-1</td>
<td>10</td>
<td>40%</td>
<td>4.3</td>
<td>1.7</td>
<td>2.6</td>
<td>75%</td>
<td>25%</td>
</tr>
<tr>
<td>Pol β-1-4</td>
<td>15</td>
<td>93%</td>
<td>1.3</td>
<td>1.2</td>
<td>0.1</td>
<td>71%</td>
<td>29%</td>
</tr>
<tr>
<td>Pol β-1-9</td>
<td>20</td>
<td>80%</td>
<td>2.9</td>
<td>2.3</td>
<td>0.6</td>
<td>56%</td>
<td>44%</td>
</tr>
<tr>
<td>Pol β-1-13</td>
<td>20</td>
<td>100%</td>
<td>3.5</td>
<td>3.5</td>
<td>( \leq )</td>
<td>15%</td>
<td>85%</td>
</tr>
<tr>
<td>Pol β-1-17</td>
<td>16</td>
<td>81%</td>
<td>3.0</td>
<td>2.4</td>
<td>0.6</td>
<td>62%</td>
<td>38%</td>
</tr>
<tr>
<td>Pol β-2-6</td>
<td>22</td>
<td>91%</td>
<td>4.2</td>
<td>3.8</td>
<td>0.4</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Pol β-2-12</td>
<td>21</td>
<td>71%</td>
<td>4.3</td>
<td>3.1</td>
<td>1.2</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>Pol β-2-15</td>
<td>15</td>
<td>60%</td>
<td>1.8</td>
<td>1.1</td>
<td>0.7</td>
<td>67%</td>
<td>33%</td>
</tr>
<tr>
<td>Pol β-2-17</td>
<td>20</td>
<td>35%</td>
<td>2.9</td>
<td>1.0</td>
<td>1.9</td>
<td>57%</td>
<td>43%</td>
</tr>
<tr>
<td>Pol β-2-20</td>
<td>19</td>
<td>53%</td>
<td>1.2</td>
<td>0.6</td>
<td>0.6</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>Pol β-3-7</td>
<td>12</td>
<td>67%</td>
<td>0.9</td>
<td>0.6</td>
<td>0.3</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Pol β-3-14</td>
<td>13</td>
<td>69%</td>
<td>2.9</td>
<td>2.0</td>
<td>0.9</td>
<td>67%</td>
<td>33%</td>
</tr>
<tr>
<td>Pol β-3-15</td>
<td>16</td>
<td>69%</td>
<td>1.5</td>
<td>1.0</td>
<td>0.5</td>
<td>91%</td>
<td>9%</td>
</tr>
<tr>
<td>Pol β-3-16</td>
<td>21</td>
<td>86%</td>
<td>6.3</td>
<td>5.4</td>
<td>0.9</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td>Pol β-3-17</td>
<td>24</td>
<td>88%</td>
<td>0.7</td>
<td>0.6</td>
<td>0.1</td>
<td>79%</td>
<td>21%</td>
</tr>
</tbody>
</table>

N: Number of Mutants  
MR: Mutation Rate  
STR: [G/C]₁₀ Microsatellite  
\( ^a \) Proportion of insertions or deletions with the [G/C]₁₀ microsatellite
Comparison of clones from the three Pol β shRNA vectors to the control shRNA vector found that the proportion of events was not statistically different (p=0.41, ANOVA). The same comparison among the clones from the three Pol β shRNA constructs found no difference in the proportion of events (p=0.41, ANOVA), indicating each of the three Pol β shRNA sequences altered the location of mutations within the HSV-tk gene in the same manner. The final analysis was to determine if there was a detectable difference in the type of event that occurred at the repeat allele. When comparing the three Pol β shRNA constructs to the control shRNA no statistical difference was detected (p=0.06, Fishers Exact Test). Comparison between the three Pol β knockdown sequences also found no statistical difference (p=0.06, Fishers Exact Test), indicating that the location of the shRNA constructs used did not play a role in altering the specificity at the [G/C]_{10} microsatellite allele.

5.2.8.3 Analysis of Uninfected versus Lentiviral Infected and Selected Cells

Although as discussed above, no statistically significant difference was observed between any of the clones infected with a Pol β shRNA sequence compared to clones infected with the luciferase shRNA sequence, we wanted to determine if there was an alteration of mutation rate and mutational specificity upon lentiviral infection. To do this, the data obtained from the lentiviral infected clones were compared to the uninfected clones discussed in Chapter 4.

Shown in Figure 5.10, a comparison of the infected and uninfected clones resulted in a statistically significant decrease in mutation rate for each of the infected clones.
Figure 5.10: Lentiviral Infection Lowers the Mutation Rate of the HSV-tk Gene:
Mutation Rates were determined for each human cell clone as described in Chapter 2. The range of mutation rates were graphed as box plots, with the median value represented by the black horizontal bar within each box. Statistical analyses were performed on each set of infected clones compared to those that were uninfected, and a statistical decrease in mutation rate was observed in those clones that were lentivirally infected (p=0.007, ANOVA). One outlier was removed from the Pol β-2 plot by the software, with a mutation rate of 20x10^-5, compared to those that were uninfected (p<0.001 Lucif; p=0.001 β-1; p=0.009 β-2; p=0.001 β-3; all by Wilcoxon Mann Whitney Test).
Further statistical analysis of the uninfected to all of the infected clones also found a significant reduction in mutation rate for each of the infected clone populations (p=0.007, ANOVA). One final statistical analysis was performed using a pooled infected sample, containing all of the infected clones in one group. The median mutation rate value for this group was $2.0 \times 10^{-5}$, a 2.75-fold reduction in median mutation rate. Comparison of the pooled infected group to the uninfected clones also found a significant difference in mutation rates (p<0.0001, Wilcoxon Mann Whitney Test). These results indicate to us that lentiviral infection and/or selection have a suppressive activity on the cells, in which the mutation rate is lowered.

For the clones that were uninfected, 36% of the mutational events observed were at the repeat motif. We wanted to determine if clones that were infected and selected had an altered mutational specificity compare to those that were uninfected. To do this, we used the pool of the infected clones to compare the mutational specificity of both sets of clones. Shown in Table 5.6 are the proportion of events at the STR or coding region of the HSV-tk gene, as well as the number of events at each location for the uninfected and infected clones. By performing a Fisher’s exact test of the uninfected (27 events at STR, 45 events at coding) versus the pooled infected clones (268 events at STR, 78 events at coding), we observed a statistically significant shift of the mutational events from the coding region for the uninfected clones to the STR for the infected clones (p<0.0001), indicating that lentiviral infection and/or selection shifts the mutational specificity of the cells.

Additionally, we wanted to determine if the type of events that occurred at the STR were altered in the cells that were infected and selected. Shown in Table 5.7 are the
Table 5.6 Proportion and Number of Mutational Events in Uninfected and Lentiviral Infected Human Cell Clones

<table>
<thead>
<tr>
<th></th>
<th>Total Mutants</th>
<th>Prop At STR</th>
<th>Number at STR</th>
<th>Prop at Coding</th>
<th>Number at Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninfected</td>
<td>72</td>
<td>36%</td>
<td>27</td>
<td>64%</td>
<td>45</td>
</tr>
<tr>
<td>Lucif Infected</td>
<td>82</td>
<td>89%</td>
<td>73</td>
<td>11%</td>
<td>9</td>
</tr>
<tr>
<td>Pol β-1 Infected</td>
<td>81</td>
<td>83%</td>
<td>67</td>
<td>17%</td>
<td>14</td>
</tr>
<tr>
<td>Pol β-2 Infected</td>
<td>97</td>
<td>63%</td>
<td>61</td>
<td>37%</td>
<td>36</td>
</tr>
<tr>
<td>Pol β-3 Infected</td>
<td>86</td>
<td>78%</td>
<td>67</td>
<td>22%</td>
<td>19</td>
</tr>
<tr>
<td>Pooled Infected</td>
<td>346</td>
<td>77%</td>
<td>268</td>
<td>23%</td>
<td>78</td>
</tr>
</tbody>
</table>
Table 5.7 Proportion and Number of Mutational Events Occurring at the STR

<table>
<thead>
<tr>
<th></th>
<th>Total STR Mutants</th>
<th>Prop of Insertions</th>
<th>Number of Insertions</th>
<th>Number of Deletions</th>
<th>Number of Deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninfected</td>
<td>27</td>
<td>81%</td>
<td>22</td>
<td>19%</td>
<td>5</td>
</tr>
<tr>
<td>Lucif Infected</td>
<td>73</td>
<td>56%</td>
<td>41</td>
<td>44%</td>
<td>32</td>
</tr>
<tr>
<td>Pol β-1 Infected</td>
<td>67</td>
<td>56%</td>
<td>38</td>
<td>44%</td>
<td>29</td>
</tr>
<tr>
<td>Pol β-2 Infected</td>
<td>61</td>
<td>70%</td>
<td>43</td>
<td>30%</td>
<td>18</td>
</tr>
<tr>
<td>Pol β-3 Infected</td>
<td>67</td>
<td>63%</td>
<td>42</td>
<td>63%</td>
<td>25</td>
</tr>
<tr>
<td>Pooled Infected</td>
<td>268</td>
<td>61%</td>
<td>164</td>
<td>39%</td>
<td>104</td>
</tr>
</tbody>
</table>
breakdown of the proportion of insertion and deletion events as well as the number of each type of event for both uninfected and infected and selected clones. For the uninfected clones 81% of the mutational events were at the [G/C]_{10} microsatellite allele were insertion (22 events), and 19% were deletion events (5 events). For the pooled infected and selected clones 61% were insertion events (164 events) whereas 39% were deletion events (104 events). Statistical analysis of the proportion of insertion versus deletion events resulted in a statistically significant decrease in the number of insertion events at the microsatellite allele in the clones that were infected and selected (p=0.04, Fishers Exact Test). These results once again signify that lentiviral infection and/or selection alters microsatellite mutational specificity.

5.3 Discussion

5.3.1 Development and Validation of the Lentiviral Infection System

To introduce genes of interest into our LCL721 cells, we developed a lentiviral infection method. Initial experiments were completed using a lentiviral vector that contained the GFP gene. Using this vector, we were able to measure GFP expression both visually and by FACS analysis. As shown in Figure 5.2, after two to three rounds of infection ~50% of the cells were GFP positive. Given this information we decided that three rounds of infection would be optimal to ensure that we were able to transmit the full length of our lentiviral cassette.

We then used this lentiviral infection method to introduce our three Pol β shRNA constructs into the LCL721 cells. To determine if a reduction of Pol β mRNA was obtained, we used qRT-PCR to analyze RNA levels. As shown in Figures 5.5 and 5.6,
we did not observe a reduction in Pol β mRNA levels for any of the Pol β-1 clones. In fact, we observed that the ratio of Pol β:TBP was increased in the Pol β-1 knockdown clones compared to the control. Additionally, used western blot analyses to determine if a knockdown of Pol β protein levels was achieved. As shown in Figures 5.7 and 5.8, we did not observe a knockdown of protein levels. In fact, similar to the qRT-PCR data, it appeared that we increased the expression of Pol β with lentiviral infection of the cells, as shown by increased ratios of Pol β:PCNA for the three Pol β shRNA and the control shRNA infected clones.

With the conclusion that we did not obtain knockdown of Pol β, we next attempted to determine why the knockdown was unsuccessful. We first sought to determine if the viral integration was successful in our cells. We analyzed gDNA for both the shRNA cassette and the puromycin resistance gene. This examination found that neither the shRNA cassette nor the puromycin gene was present in any of the clones examined. This indicates that we selected for cells that did not receive the entire lentiviral product or did not received the viral infection at all. One potential reason for this is that a full length lentiviral integration did not occur, resulting in the shRNA cassette, which is at the very end of the integration, not being transmitted. A second possibility is that cells did receive the full lentiviral integration; however, a decrease in Pol β expression is lethal. If this was the case, any cells receiving the full integration would die, leaving those without the lentiviral cassette to survive in the puromycin selection, if they had acquired a mutation rendering them puromycin resistant.

To confirm that the shRNA sequence was capable of knockdown in our cells, we designed an siRNA to the Pol β-1 sequence and transiently transfected the LCL721 cells.
Although we did observe a small decrease in mRNA on Day 1, we did not detect an overall knockdown effect using this sequence (Figure 5.9). This result could indicate that the Pol β-1 sequence may not be the ideal sequence to achieve knockdown of Pol β in this cell line. The fact that we did see a slight knockdown on Days 1 and 2, and not an increase in Pol β levels, may indicate that the Pol β-1 sequence is capable of a very subtle decrease in Pol β levels in our cell line. The Pol β shRNA sequences were provided to us by our collaborator, Dr. Rob Sobol. The Sobol lab has achieved complete Pol β knockdown by Pol β-1 and intermediate knockdown by Pol β-2 and Pol β-3 in MDA-MB231 cells [321]. Based on this information, it is possible that the shRNA sequences provided to us are not optimal for the cell line that we are using, but are successful in other cell lines.

As mentioned, we observed that the levels of Pol β expression were increased at both the protein and mRNA levels in the cells that were infected by lentivirus. One possibility to explain this is that lentiviral infection is detected as DNA damage, thus activating the DNA repair pathways. The levels of Pol β could be increased to deal with this “damage”. Additionally is it possible that one of the components of the infection system can interact with the POLB gene or the regulators of the gene, in such a way that the gene is over-expressed, resulting in the increased levels of Pol β that we observed.

### 5.3.2 Mutagenesis of the Control shRNA Vector

Using the lentiviral system, we were able to infect our LCL721 cells containing the [G/C]_{10} shuttle vector with a control shRNA to be able to determine if the lentiviral
infection and puromycin selection process had an effect on mutagenesis. Comparison of mutation rates from uninfected and infected clones (refer to tables 4.2 and 5.1 and Figure 5.10) found that the median mutation rates were statistically different, $5.5 \times 10^{-5}$ for the uninfected and $1.3 \times 10^{-5}$ for the infected clones ($p<0.0001$, Wilcoxon Mann Whitney). This resulted in a 4.2-fold decrease in mutation rate for the cells infected with the control shRNA construct that was statistically significant. This result indicates that lentiviral infection and/or selection in puromycin has a suppressive effect on the mutation rate at the HSV-tk gene. Possible explanations for this will be discussed in section 5.3.4.

Analysis of the mutational specificity for the lentiviral-infected clones also showed differences from the uninfected clones (Tables 4.3 and 5.2). For the 4 infected clones analyzed, we detected 89% of the mutational events to be at the microsatellite, but only 36% in the uninfected clones ($p<0.001$, Fishers Exact Test). Similarly, when looking at the specificity at the repeat allele, we observed a difference, with the uninfected clones favoring insertion events 4:1 over deletions, while the infected clones were approximately equal in insertion and deletion events (Tables 4.3 and 5.2). Given these differences in the distribution of the mutational events as well as the specificity at the microsatellite, we conclude that lentiviral infection does have an effect on mutational specificity. By infecting the cells with lentivirus, a disruption of the endogenous cells conditions must occur, resulting in the changes in mutagenesis that we observed. Due to this fact, we conclude that the correct control for any future knockdown comparisons would be an infected, scrambled control, as these clones would be subjected to the same infection, cloning, and antibiotic selection procedure as the cells receiving the shRNA cassette.
5.3.3 Mutagenesis of Pol β shRNA Containing Clones

Using the same lentiviral methods previously discussed we initially set out to knockdown Pol β expression and determine the effect on mutagenesis of a vector containing the artificial [G/C]₁₀ repeat allele. Three different shRNA constructs were infected in cells (see Figure 2.3), the clones were expanded and then analyzed for Pol β levels, mutation rate, and mutational specificity. We found that the ranges of mutation rates, as well as the median mutation rates for the three different shRNA constructs were all similar as shown in Tables 5.3 and 5.4. This indicates that infection with all three shRNA constructs had the same effect on the overall mutation rate. Similarly, when looking at the mutational specificity, shown in Table 5.5, we observed the same trends among the three different infections. Although slight variations were observed in the proportion of events at the microsatellite as well as the distribution of microsatellite mutation events, no statistical significance was found. These results indicate to us that the three-shRNA infections have the same effect on mutagenesis.

Comparison of the clones containing the three shRNA constructs with the control shRNA containing clones showed an overlapping range of mutation rates and similar median mutation rates (Tables 5.1, 5.3, and 5.4). The lack of a statistically significant difference between the control and Pol β shRNA constructs indicates that the infections did not have an effect on the overall mutation rate of the HSV-tk gene. The mutational specificity between the control and Pol β shRNA infected clones was also similar, with 56%-89% of the mutational events occurring at the microsatellite between the four
different sets of clones. Again, this signifies to us, that no effect of mutagenesis was obtained by using any of the three Pol β shRNA constructs.

5.3.4 Mutational Analysis of Uninfected versus Infected Clones

As was discussed above, we did not obtain knockdown of Pol β levels in any of the clones we examined. In fact, what we observed was the opposite, an increase in Pol β levels. This finding, in conjunction with the fact that the normally low Pol β levels are transcriptionally upregulated upon treatment with alkylating agents, oxidative stress, or additional DNA damaging events [168] lead us to question if the upregulation we observed may be due to a DNA damage response. To determine if this effect was a global response to lentiviral infection and/or selection, determined the consequence of the lentiviral infection and selection process on mutagenesis of the HSV-tk gene containing the artificial [G/C]_{10} microsatellite sequence.

Comparison of the mutation rates at the HSV-tk gene between the uninfected and pooled lentivirally infected clones found a statistically significant 2.75-fold decrease in mutation rate (p<0.0001, Wilcoxon Mann Whitney). This result is consistent with and similar to result obtained by Shah and Eckert, using a similar lentiviral method. In this analysis, human lymphoblast cells were transduced with a cDNA construct using the same lentiviral system and protocol as was used in our study. A 2-fold decrease in mutation rate in the clones infected with both empty vector and a gene-specific complement lentivirus was observed, relative to uninfected cells [310]. Additionally, we observed that lentiviral infection and/or selection altered the mutational specificity in our cells, shifting the large proportion of mutation events to the STR region, and then the
events that occurred in this region were shifted away from insertion events to more of a balance between insertion and deletion events. A similar trend was also observed by Shah and Eckert. In addition to the observed a reduction in mutation rate, they too saw a shift of the mutational events at the microsatellite [310].

Mutational specificity differences between the uninfected and infected and selected clones showed a similar result as was found for the mutation rate, that lentiviral infection altered the specificity. Infection and selection of the lentiviral vector shifted the location of the majority of mutation events from the coding region of the HSV-tk gene to the STR repeat in a statistically significant manner (p<0.0001, Fisher’s Exact Test). Additionally, lentiviral infection and selection altered the specificity of the microsatellite mutation events. A shift from 81% insertions at the repeat motif in the uninfected clones to only 61% in the lentiviral infected and selected clones was observed. This alteration of the type of mutational event was statistically significant (p=0.04, Fisher’s Exact Test).

One possibility to explain the decrease in mutation rate and the shift in mutational specificity observed in the lentiviral infected and selected clones is an increased activity of DNA repair. Protein and mRNA analyses of our clones have shown that levels of Pol β, a polymerase that acts in a DNA repair pathway, is upregulated in infected cells compared to those which are uninfected. We hypothesize that when the cells are infected and/or selected by puromycin, the cells sense one or both of these actions as an insult to the DNA. In order to try and prevent the potential mutagenesis, the cell cycle is halted or DNA repair is activated to allow the cells to recover. Through the upregulated DNA repair activity or the slowing/stoppage of the cell cycle, the perceived “insult” caused by the infection and/or selection is prevented or repaired. However, as DNA repair has been
up regulated or activated, all mutagenesis is reduced, resulting in the 2.75-fold reduction in mutation rate observed here and the 2-fold reduction by Shah and Eckert [310].

Evidence to support the possibility of slowing or stopping the cell cycle was obtained but not previously discussed. With each round of lentiviral infection we found fewer and fewer cells were recovering between infections. The LCL721 cell line we use doubles roughly every 24 hours, so we would expect rapid cell growth after infection. We began the infection process with $2.5 \times 10^5$ total cells. Under the normal growth conditions for LCL721 cells we would expect a total of $3.2 \times 10^7$ total cells. The actual cell count $2.7 \times 10^5$ total cells after 1 round of infection. For the second round of infection $2.5 \times 10^5$ total cells were used again, and after 7 days we only recovered $1.25 \times 10^5$ total cells. After the third infection we only recovered $1 \times 10^5$ total cells. Once the cells had received the full three rounds of infection, 10-days to 2 weeks time was required to reach the concentration of cells needed to initiate the puromycin selection ($5 \times 10^5$ cells/ml).

This same phenomenon was observed by Shah and Eckert when working with cDNA complements LCL cells (unpublished data). The fact that the infected cells were not doubling on their normal schedule and that cell outgrowth was extremely lengthened indicates to us that the cell cycle may be slowed or halted completely to allow the cells to deal with the “insult” of lentiviral infection.

To try and determine if a global activation of DNA repair has occurred in our cells as a result of the lentiviral infection and/or selection in puromycin, I would like to perform both protein and mRNA analyses for other DNA repair genes, in various DNA repair pathways. Additionally, I would like to assay for markers of cell cycle regulation or inhibition, to determine if the cell cycle is, in fact, slowed or halted. To determine if
the cause of the increase in the levels of Pol β was caused by the lentiviral infection process or by the selection process in puromycin, I would like to perform protein and mRNA analyses on clones infected with lentivirus but containing a GFP marker for selection, not the puromycin resistance gene. If the results of these analyses indicate that Pol β is in fact increased in the clones, which contain GFP for selection, then we can conclude that the infection process is the culprit. However, if the induction of Pol β is not present in these clones, then we can conclude that the effect is due to the selection of the clones in puromycin.

Additionally, to determine if the increase in Pol β is due to a slowing or stopping of the cell cycle I would like to analyze cells within the initial infection period. As was mentioned, with each round of infection the cells were very slow growing, and took a 10-day or 2 week period to start growing after infection. I would like to examine the cells after each infection by flow cytometry to determine where they are in the cell cycle. These results would be compared to a wild-type cell flow cytometry analysis to determine if in fact the infected and selected cells are stalled in the cell cycle or if they have withdrawn into a quiescent stage.
Chapter 6: General Discussion and Future Directions

6.1 What Was Learned from these Studies?

The aim of my thesis research was to determine a potential role for non-replicative DNA polymerases in microsatellite mutagenesis. I utilized three systems to determine mutation rate and mutational specificity of the HSV-tk target gene that contained an artificial microsatellite repeat. Through these studies I have been able to examine the role of two non-replicative polymerases on microsatellite mutagenesis, as well as determine the endogenous mutation rate for the HSV-tk gene containing the [G/C]_{10} allele in human cells.

These studies indicate that *E. coli* DNA polymerase IV does not have a role in microsatellite mutagenesis of mononucleotide [G/C]_{10} or dinucleotide [GT/CA]_{10} and [TC/AG]_{11} alleles. However, in the absence of Pol IV, we detected a decreased mutation frequency of the HSV-tk coding region, indicating a mutagenic role of Pol IV. These results indicated the Pol IV does not act on mononucleotide [G/C] sequences, but does increase mutagenesis in non-repetitive regions. Mutations in the mononucleotide runs were biased toward deletion events, while mutations two dinucleotide repeats were biased toward insertion events. This different mutagenesis between mono- and dinucleotide repeats may be a consequence of differential MMR between the two types of microsatellites. MMR may be more effective at repairing insertion events in mononucleotide motifs, while it is more efficient at removing deletion events in dinucleotide runs.
Using our HSV-tk *in vitro* assay we examined the role of DNA Polymerase β in mononucleotide G$_{10}$ and C$_{10}$ sequence mutagenesis. We found increased microsatellite mutagenesis with both of the mononucleotide alleles compared to the dinucleotide alleles that have been tested in the lab. We observed a three- to four-fold increase in error rate for the mononucleotide alleles relative to the highest dinucleotide repeat sequence. In addition, we found that ~80-90% of the total mutational events occurred at the microsatellite allele, at approximately similar proportions. These results were in contrast to dinucleotide alleles of a similar number of repeat units, where less than 50% of the mutational events were at the repeat. These results provide evidence that Pol β acts differently between mononucleotide and dinucleotide repeat alleles, with higher mutagenesis within mononucleotide repeats.

We assayed the mutagenesis of the [G/C]$_{10}$ repeat allele in human lymphoblast cells using our HSV-tk containing shuttle vector. Overall, we observed a higher mutation rate for clones that contain the repeat motif within the shuttle vector than those clones that received the control vector. When looking at the specificity at the repeat, we found a 15-fold increase in mutagenesis at the [G/C]$_{10}$ allele over the [GT/CA]$_{10}$ dinucleotide microsatellite allele. Previously, no reports had used a forward assay to examine the mutagenesis of a mononucleotide G/C run of this length. The repeat length that we examined is also important, as genome studies have not found a G/C repeat longer than 20 bp [330]. Additionally, it has been shown that many target genes for MSI-H cancers containing mononucleotide repeats that are no longer than 10 bp in length [236]. Our results are therefore more directly relevant than previous studies examining longer mononucleotide repeat lengths.
Also using our human cell culture assay system, we examined clones that were naturally varying in levels of Pol β expression to determine the effect on mutagenesis. We compared three clones each with high, medium, and low expression of Pol β and found that the mutation rate at the [G/C]₁₀ allele was not different. This result signifies that the level of expression of Pol β does not have an effect on the mutagenesis of our mononucleotide allele in the absence of DNA damage.

Overall, we provided evidence for the role of two non-replicative DNA polymerases in the mutagenesis of the [G/C]₁₀ mononucleotide allele. In section 6.3, I discuss future research directions based on the experiments performed and the data obtained.

### 6.2 Mutational Balance

An equal balance between the number of mutations a polymerase can make and the number of errors corrected by MMR in the human genome would result in complete genome stability and we would see no mutational events. Obviously this is not the case, and there are two possibilities to explain the mutations present in the genome: either the MMR system does not accurately correct errors in microsatellite or the DNA polymerases make a larger number of errors than can be corrected.

I can conclude that MMR is active in our human cells, specifically on the [G/C]₁₀ template. In the absence of MMR there was a 38-fold increase in the mutation rate, indicating that MMR is active [310]. However, my data show that in the presence of MMR there is still a 25-fold increase in the mutation rate for the [G/C]₁₀ repeat. This results in a 13-fold difference, indicating that although MMR does correct some of the
errors, it is not able to correct all of the errors made by the DNA polymerases in human cells.

A comparison of each of the types of microsatellite repeats (mono-, di-, and tetranucleotide) examined by our lab in human cells might leave one slightly confused, as I have told you that the [G/C]$_{10}$ mononucleotide was as mutable as most of the tetranucleotide alleles previously examined (shown in Table 4.4). I feel that the explanation for this is a consequence of differential MMR. Shown in Figure 6.1 is a graphical representation of how effective MMR is on various microsatellite sequences [266, 310]. As you can see, the tetranucleotide alleles are not as mutable as the mononucleotide allele in the absence of MMR. This indicates that MMR has a greater effect on reducing mutagenesis of the mononucleotide allele than it does on the tetranucleotide alleles. With this said, the mononucleotide motif is actually subjected to a higher level of mutagenesis than what we truly observe, as MMR repair a large portion of mutations. The high level of mutagenesis observed for the tetranucleotide motifs appear to be the consequence of a lack of MMR activity, while the high level of mutagenesis observed for the mononucleotide repeat, as mentioned above, is caused by the polymerases making more mutations than the MMR system is able to correct.

Analysis of the data produced in *E. coli* also indicates that MMR has a great effect on reducing mutagenesis. Figure 3.1 shows the increase in mutagenesis on the mononucleotide and dinucleotide motifs in the absence of MMR. The same result, that MMR is highly active on the mononucleotide repeat, was observed in *E.coli* as was discussed in human cells. Additional, it has been shown that in the absence of MMR, an endogenous [G/C]$_7$ tract is highly mutable in *E. coli* [313]. Additional results of this
Figure 6.1: The Effect of MMR on Various Microsatellite Templates: Shuttle vector DNA containing the different repeat tracts was isolated from LCL721 cells and LCL1261 cells (PMS2 deficient) and then assayed for mutation rate. Shown in the graph is the ratio of the increase in mutation rate in the LCL1261 cells compared to the LCL721 cells [266, 310].
study found that the [G/C]$_7$ tract was the most destabilized in the absence of MMR, while the tetranucleotide repeats were the most stable. This data fits perfectly with what we have observed in human cells, giving greater credence to the idea that on the mononucleotide tracts polymerases make more errors than MMR can correct. In contrast, on other microsatellite repeats (di- and tetranucleotide) the polymerases make less errors, which are also corrected less efficiently, resulting in the hierarchy of mutagenesis observed.

The Pol β in vitro data I generated also supports the finding that MMR actively removes mutations, but does not correct all mutations for the mononucleotide repeat. This system lacks MMR, and for the mononucleotide allele we detected more than 71% of the mutational events to be at the repeat allele. This proportion is much greater than we observed in human cells. These data taken in sum indicate that polymerases generate a larger number of errors, as shown by the Pol β data, but they can be acted on by the MMR system to reduce the effect, as shown by the reduced proportion of STR events in human cells.

In addition, the models proposed in Chapter 4 to explain the difference in microsatellite mutagenesis between mononucleotide and dinucleotide repeats may help explain mutagenesis regardless of the fact that this system lacks MMR. I had proposed three potential models: the formation of the looped DNA, the stability of the looped DNA, and the utilization of the looped substrate by Pol β. These models also fit with data from the tetranucleotide alleles as well. A previous study performed in the lab looked at the effect of Pol β on mutagenesis at [GT]$_{10}$, [CA]$_{10}$, [TC]$_{11}$, [AG]$_{11}$, [TTCC]$_9$ and [AAGG]$_9$ alleles. The estimated error frequencies observed at the microsatellite
repeat for the tetranucleotides were on the same order of magnitude as the dinucleotide alleles examined (error frequencies ranged from 12- to 87 x10^{-4}) [323]. These error frequencies are all at least one order of magnitude lower than what was observed for the G_{10} and C_{10} mononucleotide alleles (380- and 260 x10^{-4}, respectively). Based on this order of magnitude difference, any of the three models proposed could explain the differential microsatellite error frequencies observed with Pol β. The difference could be in the formation of the loop, in that it is easier to create a single base loop than a two or four base loop. The stability of the looped out base could also be the cause. As discussed previously, a looped base in the mononucleotide G/C tracts may be bound more strongly, and thus protected, resulting in the mutagenesis observed for this template. Neither the dinucleotide or tetranucleotide templates have G/C bases long in nature. The final model could also explain the results, in that Pol β is better able to extend from a loop of a single base than of any other length, whether it be two, four, or any other number of bases.

Overall, the model that I am presenting for the balance of mutations is somewhat complicated, in that different microsatellite motifs are maintained differently. The mononucleotide repeat is very highly mutagenic and some errors are corrected by MMR, but not all. In contrast, the mutation rate observed for the di- and tetranucleotide sequences appears to be more the consequence of a lower level of polymerase errors, as MMR is not as active on these templates. In a general statement, mononucleotide repeats are the most mutable by polymerases in either the E. coli or human cell assay systems, but their mutagenesis is kept in check by the MMR system, while the polymerases are not as mutagenic on the other repeat motifs, as MMR does not correct errors in these tracts as efficiently.
6.3 Future Experimentation

6.3.1 Studies Involving dinB and Pol κ

To follow up on the *E. coli* studies performed, I would like to investigate two of the proposed models to explain the microsatellite mutagenesis results. The first model is that mutations are made within the microsatellite region by Pol IV, but are corrected by MMR before we are able to detect them. To test this model, I would construct *E. coli* strains that are dinB positive and negative, as well as mutL (a part of the MMR system) positive or negative. These strains would then be used in the same assays described to determine if a difference in mutation rate and specificity is measured in the absence of MMR. The second model is that recombination is the cause of the mutagenesis results that we obtained. To test this, similarly, *E. coli* strains would be constructed that were dinB positive and negative, as well as recA positive or negative. By performing the mutational studies we would be able to determine the contribution of recombination to mutagenesis of the HSV-tk gene as well as the microsatellite allele.

In addition to the studies in *E. coli* I would be interested in investigating Pol κ, the human homologue to dinB. Studies have been performed to gain general knowledge about Pol κ, but to date, no studies have determined the specific function of this polymerase. It has been shown that over-expression of Pol κ in mouse cells resulted in increased mutation frequencies [333], as well as increased HPRT mutation frequencies in human fibroblasts [128]. Also, it has been shown that non-small cell lung carcinomas have over-expression of Pol κ [318, 319], thus leading us to question of the role of Pol κ in mutagenesis. I would like to examine Pol κ using the assays we have available in our lab, to decipher a possible role for Pol κ in microsatellite mutagenesis. I would perform
the HSV-tk \textit{in vitro} assay, using purified human Pol $\kappa$, to determine the mutagenesis of Pol $\kappa$ on mononucleotide microsatellite templates. In addition to the \textit{in vitro} assays, I would use the lentiviral technology to examine the consequence of both over-expression and knockdown of Pol $\kappa$ on mutagenesis in human cells.

6.3.2 Studies Involving Pol $\beta$

With the results of the Pol $\beta$ experiments indicating that we did not obtain knockdown of the polymerase, I would like to further determine why the knockdown was ineffective. The study by Faumont \textit{et al.}, which found that EBV transformed cells have increased levels of Pol $\beta$ due to interactions of the LMP1 protein [334], lead to the idea of looking at additional lymphoblastoid cell lines for Pol $\beta$ analysis. We have been using the LCL721 cell line, as they are non-tumorigenic and MMR proficient. It is possible that these cells are not the ideal cell line for examining a knockdown of Pol $\beta$ levels. To test this, I would obtain an EBV transformed lymphoblastoid cell line that is deficient or has a knockdown of LMP1, or a lymphoblastoid cell line that is not transformed by EBV to determine if the levels of Pol $\beta$ are related to LMP1 expression. If in fact the Pol $\beta$ expression is lower in other cell lines, I would like to perform the transient siRNA knockdown experiment described in Section 5.2.6 to determine the plausibility of the lentiviral infection system working in a different lymphoblastoid cell line.

Once the optimal cell line for Pol $\beta$ knockdown is determined, I would perform the Pol $\beta$ shRNA studies to determine the effect on microsatellite mutagenesis. In addition to the studies on Pol $\beta$, I would also like to study Pol $\lambda$. Similar to Pol $\beta$, Pol $\lambda$ is a member of the X family of DNA polymerases and it has been shown that Pol $\lambda$ has
high similarity to Pol β [112, 113]. Additionally, it has been shown that Pol λ can substitute for Pol β in a reconstituted DNA repair assay, indicating that Pol λ might function in BER [149]. Based on this fact, I would like to determine the role of Pol λ on microsatellite mutagenesis, in both the HSV-tk In vitro assay as well as in human cells. Using lentiviral technology, I would like to create two new cell lines, one with a knockdown of Pol λ and a second that is deficient in both Pol β and Pol λ. By using these two cell lines, in conjunction with the Pol β knockdown data already obtained, I would be able to decipher the mutagenesis of both Pol β and Pol λ, and determine the consequence of absence of both X family polymerases.

An additional set of experiments that I would like to conduct involving Pol β would examine the role of different Pol β variants found in human cancers. As was mentioned earlier, 30% of tumors studied have mutations in Pol β [203], three of which (I260M, K289M, and Y265C) have been shown to cause cellular transformation and have been found in Prostate, Colon, and Gastric cancers respectively [210-212]. I would like to use these three variants of Pol β to determine their role in mutagenesis of the mononucleotide alleles G₁₀, C₁₀, A₈, and T₈. I believe that this study is of great significance, as many genes found to be mutated in a variety of MMR deficient cancers contain mononucleotide microsatellite alleles (see Table 1.4). By using the variants of Pol β found in cancers, we may be able to determine a role for these variants in mutagenesis of mononucleotide alleles already shown to be mutated in cancers.
6.3.3 Additional Studies

With the finding from Shah and Eckert, that MMR is active on the [G/C]$_{10}$ microsatellite allele [310], I would also be interested in examining the role of MMR on mutagenesis in a human cell context. To do this, I would use a cell line that is deficient in MMR, such as LCL1261 (deficient in PMS2). By using the same assays that are discussed in section 4.2.1, I would perform a full range analysis of mutation rate and mutational specificity in these cells. I would then be able to compare the results in the presence and absence of MMR to determine the effect on mutagenesis. I feel that this would be greatly beneficial, as the studies that have examined mononucleotide mutagenesis in MMR deficient cells have not been optimal due to the use of mismatched cell lines and only reversion assays [263, 265, 306, 307, 335].

6.4 Overall Significance

In overall relevance, why would one care about microsatellite mutations and the polymerases that may or may not cause them? One of the pieces of information that we have is that polymerase levels are altered in many cancers. Additionally, microsatellite sequences are often hotspots for mutation in cancers. If we can determine the signature mutation for each of the DNA polymerases, we would be better able to determine the consequences of polymerase mutations in cancer patients. The research in this thesis has shown that normal levels of polymerases can cause mutations in microsatellite sequences. We would therefore expect to see changes in the mutagenesis when polymerase levels are altered. We currently do not know the consequence of altered polymerase levels in
cancer patients. The knockdown studies here (and those proposed) were performed to mimic the potential situation in cancerous cells.

One of the hot topics of discussion for the future of cancer treatment is individualized therapy. When thinking about this, if a cancer patient has altered polymerase levels, it could be possible to give them a tissue specific treatment that would restore the polymerase to a more normal level. Hopefully, this treatment would help to reduce the mutagenesis caused by the polymerase at both microsatellite and non-microsatellite sequences. In order for this to occur, we must first understand the signature mutation of different polymerases on microsatellite alleles in the human genome. Once we can determine the signature we would be better able to detect the presence of such mutations and monitor the mutation rate during the cancer treatment.
References

58. Setlow, P. and A. Kornberg, Deoxyribonucleic Acid Polymerase: Two Distinct Enzymes in One Polypeptide. II. A PROTEOLYTIC FRAGMENT CONTAINING THE 5' rarr 3' EXONUCLEASE FUNCTION. RESTORATION OF INTACT
ENZYME FUNCTIONS FROM THE TWO PROTEOLYTIC FRAGMENTS. J.

59. Setlow, P., D. Brutlag, and A. Kornberg, Deoxyribonucleic Acid Polymerase:
Two Distinct Enzymes in One Polypeptide. I. A PROTEOLYTIC FRAGMENT
CONTAINING THE POLYMERASE AND 3'->5' EXONUCLEASE

and Company. 931.

61. Kornberg, T. and M.L. Gefter, Purification and DNA synthesis in cell-free

62. Kornberg, T. and M.L. Gefter, Deoxyribonucleic Acid Synthesis in Cell-free
Extracts. IV. PURIFICATION AND CATALYTIC PROPERTIES OF
p. 5369-5375.

63. Okazaki, R., M. Arisawa, and A. Sugino, Slow Joining of Newly Replicated DNA
Chains in DNA Polymerase I-Deficient Escherichia coli Mutants. Proceedings of
the National Academy of Sciences of the United States of America, 1971. 68(12):
p. 2954-2957.

64. Uyemura, D., D.C. Eichler, and I.R. Lehman, Biochemical characterization of
mutant forms of DNA polymerase I from Escherichia coli. II. The polAex1

65. Grindley, N.D.F. and W.S. Kelley, Effects of Different Alleles of Escherichia-
Coli-K12 Pola Gene on Replication of Non-Transferring Plasmids. Molecular &


67. Friedberg, E. and V. Gerlach, Novel DNA polymerases offer clues to the

68. Iwasaki, H., et al., The Escherichia coli polB gene, which encodes DNA
polymerase II, is regulated by the SOS system. Journal of Bacteriology, 1990.

69. Tang, M., et al., Roles of E. coli DNA polymerases IV and V in lesion-targeted

70. Napolitano, R., et al., All there SOS-inducible DNA polymerases (Pol II, Pol IV,
and Pol V) are involved in induced mutagenesis. The EMBO Journal, 2000.

71. Bonner, C., et al., DNA polymerase II is encoded by the DNA damage-inducible
7667.

72. Rangarajan, S., R. Woodgate, and M. Goodman, A phenotype for enigmatic DNA
polymerase II: A pivotal role for pol II in replication restart in UV-irradiated

73. Wagner, J., et al., The dinB gene encodes a novel E. coli DNA polymerase, DNA


# VITA

Kimberly Jacob  
69 University Manor East  
Hershey, PA 17033  
(717) 531-0003  ext 286069  
(412) 445-2921  
kdj125@psu.edu

| Graduate Education | 2003 - Present | Penn State Hershey College of Medicine  
Anticipated Graduation 12/2009  
Intercollegiate Graduate Degree Program in Genetics Doctoral Candidate  
Thesis Advisor: Kristin Eckert  
Thesis Project: **The Role Of Non-Replicative DNA Polymerases In Microsatellite Mutagenesis**

| Undergraduate Education | 1999 - 2003 | Allegheny College  
B.S. in Biology  
Minors in Religious Studies and Psychology  
Senior Comprehensive Project: Determination of the timing of expression and the expression pattern of the *microdiscs* gene in *Drosophila melanogaster* imaginal disc development.

| Publications |  
Jacob, K.D. and K.A. Eckert. The Role of DNA Polymerase β Variants Found in Human Cancer on Mutagenesis of Mononucleotide Microsatellite Alleles (Manuscript in Preparation)

Jacob, K.D. and K.A. Eckert. Lentiviral Infection Decreases Mutation Rate and Alters Mutational Specificity in Human Lymphoblast Cells (Manuscript in Preparation)


| Posters Presentations |  


Kimberly Jacob and Kristin Eckert. Effect of *E. coli* Polymerase IV on the Rate of Spontaneous Mutation at Microsatellite Alleles. Gordon Conference on Mutagenesis. Salve Regina University August 2006

Kimberly Jacob and Kristin Eckert. Effect of *E. coli* Polymerase IV on the Rate of Spontaneous Mutation at Microsatellite Alleles. 4th Annual Genetics Symposium. Penn State University April 2006

Kimberly Jacob and Kristin Eckert (2006). Effect of *E. coli* Polymerase IV on the Rate of Spontaneous Mutation at Microsatellite Alleles. Graduate Student Forum. Penn State College of Medicine March 2006

| Honors and Awards |  
Environmental Mutagen Society Student Travel Award (2008)  
Merit-based award in recognition of the excellence of research progress