Specificity of damage-bypass DNA polymerases \textit{E. coli} DinB and human pol \kappa

by Nicole M. Antczak

B.S. in Chemistry, St. John Fisher College

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
In partial fulfillment of the requirements
for the degree of Doctor of Philosophy

November 10, 2017

Dissertation directed by
Penny J. Beuning
Professor of Chemistry and Chemical Biology
Acknowledgements

I would like to thank my advisor, Dr. Penny Beuning. I am thankful for opportunity to do my PhD work in your lab and your guidance in writing papers and working towards my career goals. I appreciate all your help in sending different teaching career related events my way. Thank you for always believing in me and pushing me to be my best.

I would also like to thank my committee, Dr. Carla Mattos, Dr. Mary Jo Ondrechen, and Dr. Sam Lone for taking the time to read my dissertation and provide comments. Additionally to Carla and Mary for participating in my committee meetings every year and giving me additional feedback and support.

I would like to thank my collaborators: the Zhang group at Northeastern University and Rozners group at SUNY Binghamton for synthesizing our DNA, and Cisneros lab at University of North Texas for computational analysis of the SNPs.

I would like to thank the Northeastern University Department of Chemistry and Chemical Biology for the opportunity to pursue my PhD.

I would like to thank all my fellow Beuning laboratory members past and present for all their help and support. I would especially like to thank Dr. Jason Walsh and Dr. Lisa Hawver for all their training when I first joined the lab. I also want to thank my partner in crime, Caitlyn Mills, who is always there to lend an ear and discuss data. I would also like to thank all the undergraduates who have worked on these projects with me including Carl Palad, Morgan Packer, Jill Coghlan, John Lambert, and Emanuel Zlibut.

And finally, I would like to thank my family. My parents, Barb and Stan, and my sister Samantha for always being supportive of following my dreams and helping me move out to Boston. I would also like to thank my new husband Jeff for always being supportive, patient, and understanding. Thank you for always cooking dinner and having it ready for me when I get home late, waking up early to take care of Olive to let me sleep in, and driving me to the train or picking me up when I miss the bus or when the bus isn’t running. And thank you for being ready to move to Switzerland with me so I can continue to work towards my dreams of becoming a professor. I would also like to thank my professors at the St. John Fisher College Departments of Chemistry and Biology who helped mold me into a young scientist and open my eyes to graduate school and becoming a professor.
Abstract of Dissertation

DNA damage is a constant threat and can block replication. Y-family DNA polymerases can replicate damaged DNA through the process of translesion synthesis (TLS). Y-family DNA polymerases are conserved in all domains of life and bypass specific DNA damage in either an error-free or error-prone manner. This dissertation presents work to understand the damage specificity and nucleotide incorporation specificity of Y-family DNA polymerase orthologs E. coli DinB and human pol κ.

Using site-directed mutagenesis we have created alanine mutations in the active site loop 1, adjacent to the incoming nucleotide, of both DinB and pol κ. These mutations resulted in a range of activity and fidelity opposite undamaged DNA and the preferred lesion of both polymerases, $N^2$-furfuryl-dG. The variants that retained activity similar to the wild-type proteins had decreased fidelity, especially opposite $N^2$-furfuryl-dG while the variants with decreased activity did not show a decrease in fidelity.

Both E. coli DinB and human pol κ are able to bypass minor groove adducts on the $N^2$ position of deoxyguanine efficiently and correctly while adducts in the major groove, such as the $N^6$ position of deoxyadenine are blocking lesions. However, another DinB ortholog, archaeal Dpo4 can bypass major groove adducts. The active site loop of these three proteins differs in length and in order to determine if the active site loops are important for damage specificity we created loop swap chimeras of the three proteins. We found that pol κ is tolerant to increases in length of its active site loop but DinB is not tolerant to increases or decreases in the length of its active site loop.

Y-family DNA polymerases have been implicated in antibiotic resistance as well as cancer and chemotherapy resistance. We found that two pol κ mutations found in cancer tumors
have an increase in insertion and extension activity as well as decreases in fidelity opposite both minor groove and major groove adducts. It is believed that Y-family DNA polymerases could be targets to increase chemotherapeutic efficacy and decrease antibiotic resistance. Using a computational molecular modeling screen, a number of potential inhibitors of DinB were identified and we found that the compounds inhibit DinB at much lower concentrations than pol κ, and many do not inhibit pol κ even at high concentrations.

The work presented here reveals the important roles of the active site loop in DinB and pol κ in activity and fidelity. Point mutations in the DinB active site loop have a range of effects, whereas changing the length of the loop results in dramatic losses of activity. In contrast, human pol κ is far more tolerant to changes in its active site loop as well to potential inhibitors. This work suggests that DinB is overall more stringent in its lesion bypass activities than pol κ.
# Table of Contents

Copyright ......................................................................................................................... 2  
Acknowledgements ........................................................................................................ 3  
Abstract of Dissertation ............................................................................................... 4  
Table of Contents ........................................................................................................... 6  
List of Figures ................................................................................................................ 15  
List of Tables .................................................................................................................. 18  

Chapter 1: Introduction ................................................................................................. 19  

1.1 Replication, DNA damage, and translesion synthesis ........................................ 19  
1.2 SOS response, regulation, and cellular interactions of DinB ............................... 21  
1.3 DNA Polymerase IV: DinB .................................................................................... 24  
   1.3.1 DinB cellular interactions ............................................................................... 24  
   1.3.2 DinB DNA damage specificity ....................................................................... 25  
   1.3.3 DinB variants .................................................................................................. 27  
1.4 Eukaryotic DNA damage response and 9-1-1 complex .................................... 29  
1.5 Human DNA polymerase κ .................................................................................. 31  
   1.5.1 DNA damage specificity ............................................................................... 32  
   1.5.2 Pol κ structure and variants .......................................................................... 35  
1.6 Inhibitors of polymerases ....................................................................................... 38
Chapter 1: Introduction

1.7 Significance .................................................................................................................. 40

1.8 References .................................................................................................................... 42

Chapter 2: Loop 1 residues of DinB and pol κ are important for activity and fidelity .......... 55

2.1 Introduction .................................................................................................................. 55

2.2 Materials and Methods .............................................................................................. 57

2.3 Results ........................................................................................................................ 63

2.4 Discussion .................................................................................................................... 70

2.5 References ................................................................................................................... 76

Chapter 3: Pol κ is more tolerant than DinB to changes in the active site loop 1 ............... 79

3.1 Introduction .................................................................................................................. 79

3.2 Materials and Methods .............................................................................................. 82

3.3 Results ........................................................................................................................ 86

3.4 Discussion .................................................................................................................... 98

3.5 References ................................................................................................................... 105

Chapter 4: Human pol κ can tolerate a longer active site loop better than E. coli DinB ......... 109

4.1 Introduction .................................................................................................................. 109

4.2 Materials and Methods .............................................................................................. 110

4.3 Results ........................................................................................................................ 115

4.4 Discussion .................................................................................................................... 119
<table>
<thead>
<tr>
<th>Chapter 5: <em>In vitro</em> biochemical characterization of human pol κ SNPs</th>
<th>126</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>126</td>
</tr>
<tr>
<td>5.2 Materials and Methods</td>
<td>128</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>131</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>142</td>
</tr>
<tr>
<td>5.5 References</td>
<td>147</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6: Inhibitors of DinB are less effective against pol κ <em>in vitro</em></th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>150</td>
</tr>
<tr>
<td>6.2 Materials and Methods</td>
<td>152</td>
</tr>
<tr>
<td>6.3 Results and Discussion</td>
<td>153</td>
</tr>
<tr>
<td>6.4 Discussion</td>
<td>160</td>
</tr>
<tr>
<td>6.5 References</td>
<td>164</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7: Conclusions and future work</th>
<th>167</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 References</td>
<td>169</td>
</tr>
</tbody>
</table>

References .................................................................................. 170
## Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>β</td>
<td>beta</td>
</tr>
<tr>
<td>η</td>
<td>eta</td>
</tr>
<tr>
<td>γ</td>
<td>gamma</td>
</tr>
<tr>
<td>ι</td>
<td>iota</td>
</tr>
<tr>
<td>κ</td>
<td>kappa</td>
</tr>
<tr>
<td>%</td>
<td>percent</td>
</tr>
<tr>
<td>′</td>
<td>prime</td>
</tr>
<tr>
<td>ζ</td>
<td>zeta</td>
</tr>
<tr>
<td>γ-HOPdG</td>
<td>γ-hydroxypropano-deoxyguanosine</td>
</tr>
<tr>
<td>2-AAF</td>
<td>$N^2$-acetylaminofluorene</td>
</tr>
<tr>
<td>2-AF</td>
<td>2-aminofluorene</td>
</tr>
<tr>
<td>$^{32}$p</td>
<td>Phosphorous-32</td>
</tr>
<tr>
<td>4NQO</td>
<td>4-nitroquinoline-1-oxide</td>
</tr>
<tr>
<td>8-oxoG</td>
<td>8-oxo-guanine</td>
</tr>
<tr>
<td>A</td>
<td>Alanine</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic Acid</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BL21</td>
<td><em>E. coli</em> cells</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>BPDE</td>
<td>Benzo[a]pyrene diolepoxide</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>C</td>
<td>Cysteine</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>D</td>
<td>Aspartic Acid</td>
</tr>
<tr>
<td>dA</td>
<td>deoxyadenine</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenine triphosphate</td>
</tr>
<tr>
<td>dC</td>
<td>deoxycytosine</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytosine triphosphate</td>
</tr>
<tr>
<td>dG</td>
<td>deoxyguanosine</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>dT</td>
<td>deoxythymine</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymine triphosphate</td>
</tr>
<tr>
<td>E</td>
<td>Glutamic Acid</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherchia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>F</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>g</td>
<td>g-force</td>
</tr>
<tr>
<td>G</td>
<td>Glycine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>H</td>
<td>Histidine</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid</td>
</tr>
<tr>
<td>H</td>
<td>Hydrogen</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>I</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>K</td>
<td>Lysine</td>
</tr>
<tr>
<td>$k_{\text{cat}}$</td>
<td>catalytic constant, turnover number</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Michaelis-Menten constant</td>
</tr>
<tr>
<td>L</td>
<td>Leucine</td>
</tr>
<tr>
<td>L</td>
<td>liters</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>Leu</td>
<td>Leucine</td>
</tr>
<tr>
<td>LS</td>
<td>loop swap</td>
</tr>
<tr>
<td>$\mu$M</td>
<td>micromolar</td>
</tr>
<tr>
<td>M</td>
<td>Methionine</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>m</td>
<td>min</td>
</tr>
</tbody>
</table>
mer  unit length of nucleic acid
Met  Methionine
Mg   Magnesium
mg   milligram
min  minute
mL   milliliter
mM   millimolar
MW   molecular weight
N    Asparagine
$N^2$ffdG $N^2$-furfuryl-deoxyguanosine
$N^2$-AnthG $N^2$-CH$_2$(9-anthracenyl)dG
$N^6$ffdA $N^6$-furfuryl-deoxyadenosine
NaCl sodium chloride
NFZ nitrofurazone
nM   nanomolar
nt   nucleotide
$O^6$-MeG $O^6$-methylguanine
OAc  acetate
P    phosphorus
p    primer
P    Proline
PAGE polyacrylamide gel electrophoresis
PCNA Proliferating Cell Nuclear Antigen
<table>
<thead>
<tr>
<th><strong>PDB</strong></th>
<th>Protein Data Bank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phe</strong></td>
<td>Phenylalanine</td>
</tr>
<tr>
<td><strong>PMSF</strong></td>
<td>phenyl methyl sulfonyl fluoride</td>
</tr>
<tr>
<td><strong>Pol</strong></td>
<td>polymerase</td>
</tr>
<tr>
<td><strong>Pols</strong></td>
<td>polymerases</td>
</tr>
<tr>
<td><strong>POOL</strong></td>
<td>Partial Order Optimum Likelihood</td>
</tr>
<tr>
<td><strong>Pro</strong></td>
<td>Proline</td>
</tr>
<tr>
<td><strong>Q</strong></td>
<td>Glutamine</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>Arginine</td>
</tr>
<tr>
<td><strong>rNTP</strong></td>
<td>ribonucleotide triphosphate</td>
</tr>
<tr>
<td><strong>S</strong></td>
<td>Serine</td>
</tr>
<tr>
<td><strong>S. solfataricus</strong></td>
<td><em>Sulfolobus solfataricus</em></td>
</tr>
<tr>
<td><strong>SDS-PAGE</strong></td>
<td>sodium dodecyl sulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td><strong>Ser</strong></td>
<td>Serine</td>
</tr>
<tr>
<td><strong>SNP</strong></td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td><strong>SO$_4$</strong></td>
<td>sulfate</td>
</tr>
<tr>
<td><strong>SOS</strong></td>
<td>“Save Our Ship” international distress signal</td>
</tr>
<tr>
<td><strong>SSB</strong></td>
<td>single stranded DNA binding protein</td>
</tr>
<tr>
<td><strong>ssDNA</strong></td>
<td>single stranded DNA</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>Threonine</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>Thymine</td>
</tr>
<tr>
<td><strong>T-T</strong></td>
<td>Thymine-thymine dimer</td>
</tr>
<tr>
<td><strong>Thr</strong></td>
<td>Threonine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>TLS</td>
<td>translesion synthesis</td>
</tr>
<tr>
<td>$T_m$</td>
<td>melting temperature</td>
</tr>
<tr>
<td>Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Valine</td>
</tr>
<tr>
<td>Val</td>
<td>Valine</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>maximum velocity</td>
</tr>
<tr>
<td>W</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
<tr>
<td>Y</td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1. Y-family DNA polymerases have a unique little finger domain........................................... 21
Figure 1.2. SOS response in *E. coli* regulates the induction of Y-family polymerase Pol IV
(DinB) and Pol V (UmuD’C).................................................................................................................. 22
Figure 1.3. DNA adducts on the N\(^2\) position of deoxyguanine that DinB can bypass............. 26
Figure 1.4. Major groove adducts on dA, etheno-dA and N\(^6\)-furfuryl-dA, and dG, C\(^8\)-dG-2-
acetylamino-fluorene and C\(^8\)-dG-2-aminofluorene that block replication by DinB.. 27
Figure 1.5. Crystal structure of pol κ.................................................................................................... 36
Figure 2.1. Crystal structure of DinB (dark magenta).............................................................................. 57
Figure 2.2. Most DinB variants confer survival upon exposure to nitrofurazone. ......................... 64
Figure 2.3. DinB loop 1 residues have different activities on undamaged and N\(^2\)ffdG DNA
as compared to WT DinB.......................................................................................................................... 65
Figure 2.4. DinB loop 1 variants have increased misincorporation opposite N\(^2\)-furfuryl-dG
.................................................................................................................................................................. 66
Figure 2.5. Primer extension assays of wild-type pol κ and loop 1 variants: ................................. 68
Figure 2.6. Primer extension assays of wild-type pol κ and loop 1 variants................................. 69
Figure 2.7. Melting temperatures of WT DinB and pol κ and loop 1 variants of both proteins.................................................................................................................................................. 70
Figure 3.1. Alignment of pol κ (light blue, PDB 2W7O), DinB (green, PDB 4IR9), and Dpo4
(purple, PDB 3QZ7) based on alignment of the Cα atoms of the amino acids....................... 82
Figure 3.2. Both sets of DinBKappa loop swap chimeras have decreased activity with
undamaged and both damaged DNA templates..................................................................................... 88
Figure 3.3. DinBKappa loop swap chimeras have decreased misincorporation opposite
undamaged DNA and with both damaged DNA templates ................................................................. 89
Figure 3.4. DinBKappa loop swap chimeras have decreased misincorporation opposite
undamaged DNA and with both damaged DNA templates ................................................................. 90
Figure 3.5. KappaDinB loop swap chimeras have similar activity as WT pol κ on
undamaged and both damaged DNA templates..................................................................................... 91
Figure 3.6. KappaDinB loop swap chimeras have decreased misincorporation opposite
undamaged DNA and the damaged DNA N\(^2\)ffdG template.............................................................. 92
Figure 3.7. Pol κ V130I has similar activity as WT pol κ on undamaged and .................... 94
Figure 3.8. KappaDinB loop swaps (C) are more stable than the DinBKappa loop swaps
(A/B). .......................................................................................................................................................... 97
Figure 3.9. (A) KappaDinB LS6 homology model (yellow, z-score of -0.577 (good))
structurally aligned with WT pol κ........................................................................................................... 101
Figure 4.1 DinBDpo4 loop swaps show low activity on all DNA templates, (A) undamaged DNA, (B) $N^2$ffdG DNA, and (C) $N^6$ffdA compared to both WT proteins, DinB and Dpo4 ................................................................. 117

Figure 4.2 KappaDpo4 loop swap 1-5A (A-C) show very low to no activity on all DNA templates, (A/D) undamaged DNA, (B/E) $N^2$ffdG DNA, and (C/F) $N^6$ffdA ........... 118

Figure 4.3. DinBDpo4 LS and KappaDpo4 LS 1-5B have an increase in melting temperature in the presence of undamaged DNA and the correct nucleotide. ...... 119

Figure 4.4. The three introduced arginines in Dpo4DinB loop swap could decrease interactions with the little finger residues in Dpo4 due to electrostatic repulsion.. 121

Figure 4.5. The Dpo4Kappa loop swaps have a much smaller active site loop than WT Dpo4. ........................................................................................................... 122

Figure 5.1. Crystal structure of pol κ (PDB 4U6P (Jha et al. 2016)) with the residues of interest highlighted in dark blue (except Leu21 which is not in the crystal structure). In grey is the DNA and in red is the incoming nucleotide. Distances from the incoming nucleotide are in Table 5.2. ..................................................................... 129

Figure 5.2. WT Pol κ and SNPs insertion and extension past undamaged DNA or a lesion (A) undamaged dG DNA, (B) $N^2$ffdG DNA, (C) Undamaged dA DNA, (D) $N^6$ffdA DNA, and (E) etheno-dA DNA (using running start primer). ........................ 133

Figure 5.3. Pol κ WT and SNPs extension from a correct base pair ......................... 134

Figure 5.4. Pol κ WT and SNPs extension from a mismatched basepair ................. 135

Figure 5.5. Pol κ WT and SNPs misincorporation opposite a lesion show an increase in misincorporation ........................................................................................................... 137

Figure 5.6. Pol κ WT and SNPs misincorporation after a lesion show correct dA incorporation after lesions with the correct base pairing ............................... 138

Figure 5.7. Pol κ WT and SNP variants misincorporate after a lesion show correct dA incorporation from lesions with a mismatched base pair .............................. 139

Figure 5.8. Melting temperatures (°C) of WT pol κ and variants in the presence of undamaged DNA, $N^2$ffdG, $N^6$ffdA and etheno-dA DNA........................................ 141

Figure 6.1. Structures of the potential inhibitors found in the molecular modeling screen. .................................................................................................................. 151

Figure 6.2. (A) Pol κ is unaffected by phenyl phosphate while DinB (B) activity is affected even at low concentrations. ................................................................. 154

Figure 6.3. (A) Pol κ is unaffected by 2-hydroxy-5-nitrobenzoic acid (5-nitrosalicylic acid) while DinB activity is affected even at low concentrations. ......................... 155

Figure 6.4 Aurintricarboxylic acid affects both (B) DinB and (A) pol κ however a 10-fold higher concentration is required for pol κ inhibition ........................................ 156

Figure 6.5. (A) Pol κ is only affected by pamoic acid at very high concentrations while (B) DinB activity is affected at lower concentrations........................................... 157
Figure 6.6. (A) Pol κ is only affected by glycinamide HCl at very high concentrations while (B) DinB activity is affected at low concentrations. ................................................................. 158

Figure 6.7. (A) Pol κ is only affected by D-galacturonic acid at very high concentrations while (B) DinB activity is affected at low concentrations and only on damaged DNA template........................................................................................................ 159

Figure 6.8. (A) Pol κ is unaffected by glycolic acid while (B) DinB activity is affected at low concentrations but not at the higher concentrations .......................................................... 160
List of Tables

Table 2.1 DinB kinetics data on $N^2$fdG containing DNA ..................................................66

Table 3.1 DinBKappa loop swap A protein sequences .................................................................83

Table 3.2 DinBKappa loop swap B protein sequences .................................................................83

Table 3.3 Pol KappaDinB loop swap protein sequences .............................................................84

Table 4.1 DinBDpo4 loop swap protein sequences .....................................................................111

Table 4.2 Pol KappaDpo4 loop swap A protein sequences .........................................................111

Table 4.3 Pol KappaDpo4 loop swap B protein sequences .........................................................112

Table 4.4 Dpo4Kappa loop swap protein sequences .................................................................112

Table 4.5 Dpo4DinB loop swap protein sequences .....................................................................113

Table 5.1. POLK Single Nucleotide Polymorphisms studied .......................................................129

Table 5.2. Distance from SNP residues to incoming nucleotide ................................................130

Table 5.3. DNA primer and template sequences .......................................................................131
Chapter 1: Introduction

1.1 Replication, DNA damage, and translesion synthesis

Replication of DNA is required for cell survival and the replicative DNA polymerases that copy DNA are highly efficient and accurate, inserting the incorrect nucleotide less than one in every 100,000 additions (Ollivierre et al. 2011). Replicative DNA polymerases bind DNA and the incoming nucleotides by adopting a right handed fold, in which the thumb and fingers domains bind to the DNA and nucleotides (Federley et al. 2010, Patel et al. 2001). DNA polymerases catalyze the addition of the nucleotides via a nucleophilic attack of the 3’ hydroxyl group of the DNA primer on the α-phosphate of incoming nucleotides and release pyrophosphate. The active site residues, typically glutamic acid or aspartic acid, are located in the palm domain and coordinate the magnesium ions required to activate the nucleophilic attack (Adler et al. 1958, Bessman et al. 1958, Kornberg et al. 1958, Lehman et al. 1958).

DNA damage is ubiquitous, stemming from both endogenous and exogenous sources and can cause lesions on the DNA, which in turn leads to mutations and cell death. In order to bypass lesions, cells in all domains of life have specialized DNA polymerases that can replicate past lesions in a process called translesion synthesis (TLS) (Radman 1975). Many of the specialized DNA polymerases are in the family called the Y family of DNA polymerases (Ohmori et al. 2001). Y-family DNA polymerases are characterized by their low fidelity on undamaged DNA and their ability to bypass DNA lesions (Ohmori et al. 2001). The work presented in this thesis focuses on two of these Y-family polymerases, *E. coli* DinB and its eukaryotic ortholog human polymerase kappa (pol κ).
Y-family DNA polymerases are conserved in all domains of life. Other Y-family DNA polymerases include *E. coli* UmuD’2C (Pol V), *H. sapiens* Pol η, Pol ι, and Rev1, *S. solfataricus* Dpo4, and *S. cerevisiae* Pol η and Rev1. Y-family DNA polymerases have low homology to the members of other families of DNA polymerases but have high amino acid sequence similarity and identity within the family (Ohmori et al. 2001). Like members of other DNA polymerase families, they contain three domains, palm, fingers and thumb but they also contain a fourth domain unique to Y-family polymerases, the little finger or polymerase associated domain (PAD) (Friedberg et al. 2001, Ling et al. 2001, Trincao et al. 2001), as seen in Figure 1.1. Additionally, the fingers and thumb domains in Y-family polymerases are smaller than in other DNA polymerases which allows for larger, bulky lesions to fit in the active site (Chandani et al. 2010). Y-family DNA polymerases lack the 3’-5’ exonuclease activity possessed by other DNA polymerases, contributing to the elevated error frequencies of Y-family DNA polymerases (Zhang et al. 2000). Human pol κ has a unique N-terminal extension, the N-clasp, that is crucial for activity, and is found only in other eukaryotic pol κ homologs (Uljon et al. 2004).
Figure 1.1. Y-family DNA polymerases have a unique little finger domain. Shown above are the crystal structures DNA polymerase III α subunit from *Thermus aquaticus* (PDB 3E0D (Wing et al. 2008) and human DNA polymerase κ (PDB ID 4U6P (Jha et al. 2016)) highlighting the domains in different colors. The DNA primer/template is in gray the incoming nucleotide in red, the fingers domain in purple, the palm domain in green, the thumb domain in blue, the little finger domain in orange is unique to Y-family DNA polymerases, and the N-clasp, in magenta, is unique to eukaryotic polymerases in the pol κ family (Uljon et al. 2004).

1.2 SOS response, regulation, and cellular interactions of DinB

In *E. coli*, Y-family DNA polymerases and translesion synthesis are regulated by the SOS response to damaged DNA (Radman 1975). In 1967, Evelyn Witkin proposed that there was a repressor acting until DNA damage occurred, at which time the repressor would be inactivated and an operon would be induced until the DNA is repaired (Witkin 1967). It was later found that the repressor was the LexA protein. The SOS response is initiated when the replicative DNA polymerase encounters a lesion that it cannot bypass and causes single-stranded DNA (ssDNA) to develop, as shown in Figure 1.2. RecA binds to the single-stranded DNA to form a RecA/ssDNA nucleoprotein filament. LexA then binds to the RecA/ssDNA nucleoprotein filament which induces LexA to self-cleave. The cleavage allows for the expression of the SOS genes, of which there are at least 57 including *dinB* (Friedberg 2006).
Figure 1.2. SOS response in *E. coli* regulates the induction of Y-family polymerase Pol IV (DinB) and Pol V (UmuD’2C).

UmuD$_2$ homodimer and RecA have been shown to interact with DinB and can modulate the mutagenic potential of the polymerase (Godoy et al. 2007). The heterodimer, UmuDD’, was
not shown to interact with DinB and RecA (Godoy et al. 2007). The co-overproduction of UmuD in a recA\(^+\) strain reduces -1 frameshifts caused by DinB overproduction; co-overproduction of non-cleavable UmuD (S60A) eliminates DinB frameshift mutagenesis while deletion of umuD resulted in an increase in -1 frameshifts (Godoy et al. 2007). UmuD\(_2\) inhibits a non-covalent step in the template slippage pathway of DinB that causes frameshifts (Foti et al. 2010).

Both Pol I and Pol II, independently, interfere with DinB-dependent mutagenesis in SOS-independent mechanisms (Hastings et al. 2010). The polymerase domain of Pol I, and not the 3’-5’ exonuclease domain, inhibits DinB mutagenesis. In the absence of either Pol I and Pol II, the mutation rate increases about 5-fold, and about 10-fold in the absence of both polymerases, implying that they independently block DinB from reaching the primer terminus (Hastings et al. 2010).

In addition to SOS regulation, stress response factor sigma, RpoS, regulates DinB expression (Layton et al. 2003). RpoS transcriptionally upregulates DinB by approximately two-fold (Frisch et al. 2010). The induction of DinB expression occurs in late stationary phase (Layton et al. 2003). DinB is also regulated by the heat shock chaperone GroE in a potentially indirect interaction (Layton et al. 2005).

*E. coli* DNA helicase Rep interacts with DinB both *in vitro* and *in vivo* (Sladewski et al. 2011). *In vitro*, it was shown that Rep stimulates DinB’s polymerase activity and the 6-clamp binding motif was involved while *in vivo* the interaction with Rep was shown to increase DinB’s mutagenic activity in stationary phase cells (Sladewski et al. 2011). It is also believed that NusA, a transcription elongation factor, interacts with DinB and recruits it to stalled RNA polymerase to fill a gap in the template strand (that is opposite a lesion) for transcription in the process of transcription-coupled TLS (Cohen et al. 2009, Cohen et al. 2010).
1.3 DNA Polymerase IV: DinB

DinB, DNA polymerase IV, is one of two Y-family DNA polymerases in *E. coli* (Ohmori et al. 1995, Wagner et al. 1999). The gene *dinB* (damage inducible) was found to be induced by DNA damaging agents (Kenyon et al. 1980, Ohmori et al. 1995) and regulated by the SOS response. Although DinB is ubiquitous, it is not necessary for life (Ohmori et al. 2001). DinB can accommodate bulky DNA adducts and misaligned primer templates and lacks 3’-5’ exonuclease proofreading (Wagner et al. 1999). DinB is the most abundant polymerase in the cells, with 250 molecules in normal cells and upregulated to 2500 molecules upon SOS induction (Kim et al. 2001). DinB has been implicated in antibiotic resistance (Cirz et al. 2005, Cirz et al. 2007a). It was found after exposure to the antibiotic ciprofloxacin that *dinB* was upregulated in *Staphylococcus aureus*, and that *E. coli* strains with *dinB* deleted were sensitive to ciprofloxacin (Cirz et al. 2005, Cirz et al. 2007a). It has also been shown that *dinB* is upregulated after exposure to beta-lactam antibiotics (Perez-Capilla et al. 2005, Petrosino et al. 2009).

1.3.1 DinB cellular interactions

The β processivity clamp targets DinB to the primer terminus and loaded β-clamp increases the processivity of the polymerase (Wagner et al. 2000). The DinB C-terminal peptide (\(^{346}\text{QLVLGL}^{351}\)) interacts with the β-clamp and the interaction is required for TLS and spontaneous mutagenesis (Lenne-Samuel et al. 2002). This involves interaction with a hydrophobic channel on the β-clamp surface (Bunting et al. 2003). The little finger domain of DinB makes additional protein-protein interactions with both monomers of the β-clamp, including both hydrophobic and ionic interactions (Bunting et al. 2003). DinB can bind to the β-clamp at the same time as the replicative polymerase Pol III to allow for rapid bypass of DNA
damage by DinB binding to a secondary site to promote dissociation of Pol III (Kath et al. 2014). A single mutation, Thr120Pro in DinB, which retains polymerase activity, has been shown to abrogate the biochemical interaction between the β-clamp and Pol III that is required for polymerase switching (Scotland et al. 2015). In an *E. coli* strain containing the *dnaN159* allele (encoding for a mutant β-clamp), high expression of DinB resulted in conditional lethality (Maul et al. 2005). This lethality due to DinB can be suppressed by an additional mutation in RecA (*recA730*), a chronically active RecA, which can impair DinB’s ability to access the replication fork and lead to lethality due to Pol II or Pol V induced mutations (Maul et al. 2005). It has also been shown that DinB interacts with the C-terminal tail of single stranded DNA binding protein (SSB) bound to DNA allowing DinB to travel along the DNA template and displace SSB (Furukohri et al. 2012). When in the presence of SSB with the C-terminal eight residues deleted, DinB is unable to displace SSB from DNA (Furukohri et al. 2012).

### 1.3.2 DinB DNA damage specificity

DinB bypasses specific DNA damage; some examples are shown in Figure 1.3. DinB has been shown preferentially to insert dCTP correctly opposite $N^2$-furfuryl-dG by 15-fold compared to undamaged dG (Jarosz et al. 2009, Jarosz et al. 2006, Kottur et al. 2015). Strains containing a *dinB* deletion show a sensitivity to nitrofurazone (NFZ) and 4-nitroquinoline-1-oxide (4NQO) (Jarosz et al. 2006). DinB can correctly bypass $N^2$-(1-carboxyethyl)-dG, which is an adduct formed from exposure to methylglyoxal which can be produced endogenously (Yuan et al. 2008a). It has been shown that DinB can weakly bypass $N^2,3$-ethenodG (a deoxyribose analog was used containing a fluorinated 2’ carbon) and $1,N^2$-ethenodG in an error prone manner (Chang et al. 2015). Other work has shown that DinB can bypass larger adducts on the $N^2$ position of deoxyguanine, like $N^2$-benzo[a]pyrene-dG, where it favors bypassing the R isomer.
over the S isomer when a dC is the next templating base but not when there is a dG as the next templating base (Seo et al. 2006, Shen et al. 2002). It has also been shown that DinB can bypass the $N^2$-benzo[a]pyrene-dG adduct in an error-free manner (Sholder et al. 2015). DinB is able to bypass DNA adducts due to acrolein, $N^2$-$\gamma$-hydroxypropano-dG crosslinks ($\gamma$-HOPdG), as well as DNA-peptide crosslinks mediated by acrolein with peptides of four or ten residues in an error-free manner (Minko et al. 2008b). DinB has been shown to bypass $N^2$-dG interstrand DNA crosslinks (ICL) faithfully, however the TLS activity is hindered (Kumari et al. 2008).

![Chemical structures](image)

**Figure 1.3. DNA adducts on the $N^2$ position of deoxyguanine that DinB can bypass.**

These $N^2$-dG adducts are minor groove adducts, whereas major groove adducts, like etheno-dA, $N^6$-furfuryl-dA, and the fluorescent base tC block DinB (Walsh et al. 2011a, Walsh et al. 2013). Etheno-dA is a much stronger blocking lesion than $N^6$-furfuryl-dA and DinB will not
replicate DNA containing etheno-dA in vitro, until it initiates synthesis at least three, preferably four bases, past the lesion (Walsh et al. 2013). DinB can correctly insert dG opposite the modified C analog 1,3-diaza-2-oxophenothiazine (tC) which protrudes into the major groove but cannot extend from the newly created terminus or from primer termini within three nucleotides past tC. DinB has a higher affinity to the tC-containing DNA template compared to a template containing unmodified C (Walsh et al. 2011a). DinB is also unable to bypass C8-dG adducts 2-acetylaminofluorene (AAF) and 2-aminofluorene (AF) (Suzuki et al. 2001).

1.3.3 DinB variants

Several residues of DinB have been studied either in vivo or in vitro or both. When the crystal structure of DinB was solved in 2013, Ser42, adjacent to the template base and incoming
nucleotide, was mutated to an alanine and was shown, both in vivo and in vitro, to have an increase in fidelity as compared to WT (Sharma et al. 2013). Ser42 likely interacts directly and through water-mediated polar interactions with the nascent base pair (Sharma et al. 2013). Near the site of the incoming nucleotide lies Phe13, which has been shown to be the steric gate of DinB, distinguishing between dNTPs and rNTPs (Jarosz et al. 2006, Nevin et al. 2015a). Phe13Val affects the overall bypass activity of DinB opposite $N^2$ffdG and BPDE-$N^2$-dG and allows for the incorporation of rNTPs (Jarosz et al. 2006) and DinB Phe13Ala also decreases activity on undamaged DNA and allows for the incorporation of rNTPs (Nevin et al. 2015a).

Near Phe13, Tyr79 has been shown to be important for extension from the $N^2$ffdG lesion (Jarosz et al. 2009). Mutations of Tyr79 (Tyr79Ala, Tyr79Val, and Tyr79Leu) confer a 10 to 50-fold increase in sensitivity of NFZ and 4NQO, suggesting the importance of Tyr79 in the bypass of $N^2$-dG adducts in vivo (Jarosz et al. 2006).

A number of distal residues predicted computationally by the POOL (Somarowthu et al. 2011, Tong et al. 2009) server to be important for catalysis were tested biochemically and found to have a range of impacts (Walsh et al. 2012). POOL (Partial Optimum Order Likelihood) is a machine learning method that uses electrostatic features as well as geometric surface properties and phylogenetic information to rank amino acids of a given protein according to their probable functional importance (Somarowthu et al. 2011, Tong et al. 2009). His6Leu and Asp10Asn are mutations of residues that are both located near the catalytic residues as well as steric gate residue Phe13, and both variants showed a decrease in dCTP incorporation opposite $N^2$ffdG and weak extension from the dC:$N^2$ffdG base pair (Walsh et al. 2012). Asp10Glu however, had no detectable activity for either insertion or extension. Tyr106 is considered to be a second shell residue, in a spatial location near the active site and near His6. The Tyr106Ala mutation led to a
large decrease in incorporation activity and no detectable extension activity while Tyr106Phe, which retains the aromatic ring, had a small decrease in activity and was the variant with the strongest extension activity (Walsh et al. 2012). Both the aromatic ring and hydroxyl group of Tyr106 are important for the activity of DinB. Lys146, a third shell residue, when mutated to Ala showed a drop in catalytic efficiency in the incorporation of dCTP opposite N2ffdG and did not have any detectable extension activity (Walsh et al. 2012). Lys150Ala had a decrease in activity while Lys157Ala had no detectable catalytic activity and Lys157Ile had weak insertion activity (Walsh et al. 2012). In vivo work has shown that DinB variants Asp8Ala, Asp8His, Arg49Ala, Arg49Phe, Asp103Ala, Asp103Asn, and Glu104Ala resulted in dramatically lower mutation frequencies (Wagner et al. 1999).

1.4 Eukaryotic DNA damage response and 9-1-1 complex

In response to DNA damage, proliferating cell nuclear antigen (PCNA) can be post-translationally modified by ubiquitin or SUMO to regulate the appropriate response, TLS by a Y-family DNA polymerase or DNA repair. The 9-1-1 complex, an alternative clamp, is another pathway that can function in the DNA damage response (Bergink et al. 2009, Helt et al. 2005, Parrilla-Castellar et al. 2004, Stelter et al. 2003, Ulrich 2009). The DNA damage response is mediated by proteins within the phosphatidylinositol 3-kinase-like protein kinase (PIKK) family (ATM, ATR, and DNA-PK), as well as proteins within the poly(ADP-ribose) polymerase family (PARP) (Ciccia et al. 2010). ATR is activated and recruited by replication protein A (RPA)-coated single stranded DNA (ssDNA) at stalled replication forks while ATM is recruited and activated by double strand breaks (DSBs) (Bartek et al. 2007, Cimprich et al. 2008, Jackson et al. 2009, Marechal et al. 2015, Shiloh 2003). ATM and ATR target the kinases CHK1 and CHK2,
which interact with ATM and ATR to inhibit cyclin-dependent kinase, arresting cell cycle progression (Bartek et al. 2007, Kastan et al. 2004, Riley et al. 2008). Additionally the ATR/ATM signaling induces DNA repair proteins both transcriptionally and posttranscriptionally, recruits repair proteins to the damage, and activates DNA repair proteins via the modulation of phosphorylation, acetylation, ubiquitylation, or SUMOylation (Huen et al. 2008).

PCNA is monoubiquitylated by Rad6-Rad18 at Lys164 in response to arrested replication forks, which has been shown to increase affinity for Y-family DNA polymerases (Lee et al. 2008, Lehmann et al. 2007, Waters et al. 2009, Yang et al. 2007). The overexpression of Rad18 has been shown to induce PCNA ubiquitination and the association between PCNA and pol κ; a form of PCNA that cannot be ubiquitylated does not interact with pol κ (Bi et al. 2006). This increase in affinity leads to error-prone TLS pathways. Polymerases interact with PCNA through a PCNA-interacting protein motif (PIP box), potentially competitively, but there are three PIP boxes as PCNA is a homotrimer so multiple simultaneous binding events could occur (Hishiki et al. 2009, Lehmann et al. 2007, Moldovan et al. 2007, Xu et al. 2001). Y-family DNA polymerases have different interactions with PCNA; both pol η and pol κ have canonical PIP boxes while pol ι has a noncanonical PIP box (Hishiki et al. 2009, Yang et al. 2007). PCNA can also be polyubiquitylated at Lys164; this polyubiquitylation promotes an error-free pathway of damage tolerance (Budzowska et al. 2009, Hoege et al. 2002, Stelter et al. 2003, Ulrich 2009). The SUMOylation of PCNA also occurs at Lys164 (Ulrich 2009). The SUMOylation of PCNA in humans and yeast prevents inappropriate homologous recombination (Andersen et al. 2008, Gali et al. 2012, Moldovan et al. 2012, Ulrich 2009). SUMOylated PCNA in humans also plays a role in preventing double strand breaks at stalled replication forks (Gali et al. 2012).
The 9-1-1 complex is also involved in the DNA damage response and is composed of Rad9-Hus1-Rad1 proteins (Parrilla-Castellar et al. 2004). The complex is recruited to stalled replication forks by TopBP1 and DNA polymerase α (Yan et al. 2009a, b) and is loaded on DNA by a clamp loader complex made up of Rad17-RFC (Ellison et al. 2003, Griffith et al. 2002, Majka et al. 2004, Parrilla-Castellar et al. 2004, Shiomi et al. 2002). Cells containing Rad9- and Rad9 knockdowns have been shown to be sensitive to genotoxic stresses such as ultraviolet light, ionizing radiation, and alkylation (Ellison et al. 2003, Roos-Matjus et al. 2003, Shin et al. 2012). In fission yeast (Schizosaccharomyces pombe) the clamp loader protein Rad17 is required for mutagenesis carried out by TLS pol κ and ζ (Kai et al. 2003). In S. pombe it has also been shown that checkpoint activation is necessary for the transcriptional induction of pol κ and both Hus1 and Rad1 of the 9-1-1 co-immunoprecipitate with pol κ (Kai et al. 2003). In S. cerevisae similar results were obtained for 9-1-1 complex interactions with pol ζ and were required for pol ζ mutagenesis (Sabbioneda et al. 2005). It is thought that the 9-1-1 complex both facilitates Chk1 activation and recruits TLS polymerases to promote bypass of DNA damage.

1.5 Human DNA polymerase κ

The eukaryotic ortholog of E. coli DinB is human pol κ, which was identified through homology searches of orthologs of E. coli DinB, and originally called pol theta (Johnson et al. 2000). It has been shown that altered levels of pol κ transcripts could be connected to cancer development, as in some colorectal cancers the transcripts are down regulated while in non-small cell lung cancers the transcripts are upregulated (Lemee et al. 2007, O. Wang et al. 2001). Single nucleotide polymorphisms of pol κ have been found in cancerous tumors and assayed in vitro (Kim et al. 2016, Song et al. 2014, Yadav et al. 2015) and will be discussed in a later chapter.
Aside from cancer, pol κ has been shown to contribute to the formation of covalently closed circular DNA in de novo Hepatitis B infections (Qi et al. 2016).

**1.5.1 DNA damage specificity**

E. coli ortholog DinB, pol κ is blocked by $N^2$-$\gamma$-hydroxypropano-dG ($\gamma$-HOPdG) as well as the $N^2$-hydroxynonenal-dG (HNE-dG), but can extend the primer when a cytosine is incorporated opposite these lesions by another polymerase (Washington et al. 2004, Wolfle et al. 2006). However, peptide crosslinks on the $N^2$-dG mediated by acrolein ($\gamma$-HOPdG) with crosslinked peptides of four and ten amino acids have been shown to be bypassed in an error-free manner by pol κ with the correct dCTP being incorporated (Minko et al. 2008b). Pol κ can also bypass $N^2$-$N^2$-dG interstrand crosslinks (ICLs) in an error-free manner (Minko et al. 2008a). Intrastrand DNA crosslinks between C$^8$-dG and N$^2$-dT are weakly bypassed by pol κ (Lee et al. 2016). It has been reported that pol κ cannot bypass abasic sites or can very weakly bypass them (Choi et al. 2010, Johnson et al. 2000). Unlike the previously mentioned $N^2$-$N^2$-dG interstrand DNA crosslinks (ICLs), ICLs induced by abasic sites block pol κ (Xu et al. 2015). It has been shown that pol κ cannot insert nucleotides opposite the 5’ thymine of a thymine-thymine dimer (formed from UV light exposure) but it can insert opposite the 3’ thymine and extend mismatched primer termini from that position (Johnson et al. 2000, Washington et al. 2002). Pol κ has been shown to extend from trans-4-hydroxy-2-nonenal-dG (HNE-dG) after insertion opposite the lesion by pol ι (Wolfle et al. 2006). Pol κ was shown to be more efficient at extending from a G:T mismatch with the 3’ T of a T-T dimer than a correct A:T from the 3’ T of a T-T dimer (Washington et al. 2002). Pol κ is three-fold more efficient when extending from a mismatch G:T of the 3’ T of a T-T dimer than a G:T mismatch on undamaged DNA (Washington et al. 2002). Pol κ has also been shown to be able to extend from undamaged mismatches, with the highest efficiencies with G-G, G-T, G-A, and T-C mismatches (Zhang et al. 2000).
Pol κ, like DinB, is blocked by many major groove adducts to different extents. Adducts on the $O^6$ position of deoxyguanine are increasingly blocking as the size increases, this includes $O^6$-methylguanine ($O^6$-MeG) (Choi et al. 2006b, Nevin et al. 2015b), $O^6$-benzylguanine ($O^6$-BnG), and $O^6$-4-oxo-3-(3-pyridyl)butyl-dG ($O^6$-PobG) (Choi et al. 2006b). Pol κ is slightly hindered by $O^6$-carboxymethyl-dG ($O^6$-CMG) as compared to undamaged dG but can only very weakly extend from a C:$O^6$-CMG base pair (Raz et al. 2016). DNA interstrand crosslinks on the $O^6$-dG position, like $O^6$-2'-deoxyguanosine-butylene-$O^6$-2'-deoxyguanosine, block pol κ (Xu et al. 2016). As compared to undamaged dA, adducts on $N^6$-dA block pol κ to differing extents. $N^6$-(2-Hydroxy-3-buten-1-yl)-2'-dA ($N^6$-HB-dA) (Kotapati et al. 2015), and $N^6$-(2-hydroxy-3-hydroxypropan-1,3-diyl)-2'-dA ($N^6$-$\gamma$HMHP-dA) (Kotapati et al. 2012) weakly hinder pol κ while S-[2-$N^6$-deoxyadenosinyl]ethyl]glutathione (Sedgeman et al. 2017), and $N^6$,$N^6$-(2,3-dihydroxybutan-1,4-diyl)-2'-dA ($N^6$,$N^6$-DHB-dA) block pol κ except at higher protein concentrations where there is weak extension (Kotapati et al. 2015). Other glutathione adducts, $S^-$-[4- ($N^6$- Deoxyadenosinyl)-2,3-dihydroxybutyl]glutathione and $N^6$-(2,3,4-trihydroxybutyl)deoxyadenosine were shown to strongly hinder, but not completely block, pol κ activity as compared to undamaged dA (Cho et al. 2013). It has also been shown that $N^6$-etheno-dA blocks pol κ activity (Levine et al. 2001, Walsh et al. 2013) while $N^6$-furfuryl-dA ($N^6$ffdA) blocks pol κ to a lesser extent (Walsh et al. 2013).

Adducts on dC and dT also hinder pol κ activity to different extents. It has been shown that major groove adducts 8-(hydroxymethyl)-3-$N^4$-ethenodC and 3-$N^4$-ethenodC both block pol κ activity (Singer et al. 2002). With a larger adduct on the $N^4$ position of dC, 4-hydroxyequilenin (4-OHEN-dC), pol κ is blocked to a lesser extent and can reach the end of the template strand, especially at higher protein concentrations (Suzuki et al. 2004b, Yasui et al. 2007). DNA-peptide
crosslinks on the major groove $C^5$ position of dC also hinder the activity of pol $\kappa$ (Wickramaratne et al. 2015). Psoralen-DNA interstrand crosslinks on dT also block the insertion activity of pol $\kappa$, however, pol $\kappa$ can extend from the ICL after misinsertion of dTTP by pol $\iota$ (Smith et al. 2012). Both $O^2$ (minor groove) and $O^4$ (major groove) of dT can be methylated and both products moderately block pol $\kappa$ activity (Andersen et al. 2012). Alkylated $O^2$-dT products beyond $O^2$-MedT, including $O^2$-ethyldT, $O^2$-$n$-propyldT, $O^2$-$i$-propyldT, $O^2$-$n$-butyldT, $O^2$-$i$-butyldT, and $O^2$-$s$-butyldT block pol $\kappa$ activity with the highest bypass activity with DNA containing $O^2$-ethyldT and $O^2$-$s$-butyldT at the higher protein concentration (Williams et al. 2016a). The same alkyl adducts were created on $O^4$-dT and it was found that pol $\kappa$ was blocked by all except $O^4$-MedT and $O^4$-ethyldT, where bypass of $O^4$-MedT was stronger than the ethyl adduct (Williams et al. 2016b). This has also been shown in vivo, as there was no appreciable change in the bypass of the lesions in a POLK deletion cell line (Wu et al. 2016). Pol $\kappa$ can bypass thymine glycols in an error free manner (Fischhaber et al. 2002, Yoon et al. 2010). Unlike DinB, which is stimulated by interactions with the E. coli processivity clamp, pol $\kappa$ is not stimulated by the human processivity clamp PCNA (Gerlach et al. 2001).

1.5.2 Pol $\kappa$ structure and variants

As mentioned previously, pol $\kappa$ and its eukaryotic homologs have an N-terminal extension of 75 amino acids, called the N-clasp, which is in the shape of an ‘L’ and increases the ability of the polymerase to bind the DNA as seen in Figure 1.3 (Lone et al. 2007, Uljon et al. 2004). It has been shown that the N-clasp is required for activity of the polymerase, especially for mismatch extension (Lone et al. 2007) and error-free bypass of the $N^2$-BPDE-dG adduct (Jia et al. 2008). The N-clasp has also been shown to be flexible during the bypass of $N^2$-AAF-dG, allowing for misincorporation of dTTP through wobble pairing (Lior-Hoffmann et al. 2014).
When comparing full length pol κ (1-870 amino acids) with truncated forms of 1-526, 19-526, and 68-526 residues, it was found that the truncation 68-526 lost nearly all of its activity (Lone et al. 2007).

Figure 1.5. Crystal structure of pol κ (PDB ID 4U6P (Jha et al. 2016)) highlighting the domains in different colors. The N-clasp domain, in magenta, wraps around the DNA and increasing the polymerases’ ability to bind DNA. The DNA primer/template is gray, the incoming nucleotide is red, the fingers domain is purple, the palm domain is green, the thumb domain is blue, and the little finger domain is orange.

Several residues of pol κ have been studied and shown to be important for activity. Asp198 and Glu199 are required for catalysis (Gerlach et al. 2001). Tyrosine 112 (Tyr112) has been shown to be the steric gate of pol κ, distinguishing between dNTPs and rNTPs, while Tyr112Ala and Tyr112Val both are able to insert rNTPs and extend the DNA primer (Nevin et al. 2015a, Niimi et al. 2009). Tyr112 is also important in mismatch extension as both Tyr112Ala and Tyr112Val were unable to extend from undamaged mismatched termini as well as BPDE-\(N^2\)-dG mismatches (Niimi et al. 2009). Met135 is important for catalytic activity and it is
believed to stabilize the template base (Vasquez-Del Carpio et al. 2009). Met135 makes van der Waals and stacking interactions with the template base, and it stabilizes the syn conformation of 8-oxoG via van der Waals and stacking interactions (Vasquez-Del Carpio et al. 2009). Leu508, in the little finger domain, is important in error free bypass of the 8-oxoG lesion and is not conserved among other Y-family pols like Dpo4 and pol η, which possess an arginine and lysine at this position, respectively. The positive charges in the other polymerases form an electrostatic interaction with a templating 8-oxoG, allowing for accurate bypass of the lesion (Irimia et al. 2009). Leu508Lys resulted in a decrease of dATP misincorporation opposite 8-oxoG, Leu508Ala resulted in an increase in dATP misincorporation, and finally Leu508Arg inserted dATP and dCTP equally but had much lower catalytic activity (Irimia et al. 2009). Leu508Lys likely stabilizes the anti conformation of the 8-oxoG and inhibits the dATP misincorporation while in WT pol κ uncharged Leu508 allows for the syn conformation of 8-oxoG and the dATP misincorporation (Irimia et al. 2009). Phe171 has been shown to be a molecular brake for pol κ in bypass of BPDE-\(N^2\)-dG as it slows down TLS through stacking interactions with the adduct (Sassa et al. 2011). Phe171Ala has an increase in activity opposite BPDE-\(N^2\)-dG but showed a decrease in correct dC incorporation as compared to incorporation opposite an undamaged dG. The catalytic efficiency of Phe171Ala opposite a thymine glycol adduct did not change compared to WT (Sassa et al. 2011). Nearby residues Arg175 and L197 were also mutated to alanine and showed decreases in activity on undamaged, BPDE-\(N^2\)-dG, and thymine glycol containing templates (Sassa et al. 2011). Multiple single nucleotide polymorphisms (SNPs) of pol κ have been biochemically characterized opposite a variety of lesions including \(N^2\)-CH\(_2\)(9-anthracenyl)dG, 8-oxoG, \(O^6\)-MeG, and abasic sites. The SNPs Arg246X, Arg298His, Thr473Ala, and Arg512Trp showed marked decreases in replication of undamaged and all
damaged templates tested (Kim et al. 2016). Glu29Lys, Phe192Cys, and Glu292Lys have similar activity to WT pol κ in their bypass efficiency (Kim et al. 2016). The SNP Thr44Met resulted in bypass impairment opposite \( N^2-\text{CH}_2(9\text{-anthracenyl})\text{dG} \) (\( N^2\text{-AnthG} \)), \( O^6\text{-MeG} \), and abasic sites but not 8-oxoG (Kim et al. 2016). Additional SNPs assayed only opposite undamaged and \( N^2\text{-AnthG} \) also had a range of activity. Arg219X was an inactive truncation, Glu419GLy and Tyr432Ser were greatly impaired in replication of undamaged and \( N^2\text{-AnthG} \) templates (Song et al. 2014). The SNPs Leu21Phe, Ile39Thr, and Asp189Gly all showed a decrease in bypass of \( N^2\text{-AnthG} \) but no decrease in activity was seen on undamaged DNA (Song et al. 2014). Arg219Ile retained activity similar to WT on undamaged and \( N^2\text{-AnthG} \) DNA while Ser423Arg showed a slight increase in polymerase activity on both undamaged and \( N^2\text{-AnthG} \) containing DNA (Song et al. 2014).

1.6 Inhibitors of polymerases

It has been thought that inhibiting Y-family DNA polymerases in \emph{E. coli} can combat antibiotic resistance while inhibiting Y-family DNA polymerases in humans can increase the efficacy of DNA-damaging chemotherapy (Cirz et al. 2005, Cirz et al. 2007b, Yamanaka et al. 2012). There have been several small molecules that have been found to inhibit human Y-family pols, these include \( \beta\text{-sitosteryl}(6'\text{-O-linoleoyl})\text{-glucoside} \) (from soybean extract), \( \text{penta-}O\text{-galloyl-beta-d-glucose} \), pamoic acid, aurintricarboxylic acid, ellagic acid, MK-866, candesartan cilexil, and manoalide (Dorjsuren et al. 2009, Horie et al. 2010, Ketkar et al. 2013, Mizushina et al. 2010, Yamanaka et al. 2012). The anti-cancer compound penta-\( O\text{-galloyl-beta-d-glucose} \) was shown to inhibit B- and Y-family pols most strongly but also inhibited X-family pol \( \beta \) (Mizushina et al. 2010). MK-866, \( \beta\text{-sitosteryl}(6'\text{-O-linoleoyl})\text{-glucoside} \), candesartan cilexil,
and manoalide have been shown to inhibit pol κ most strongly (Yamanaka et al. 2012), while pamoic acid, aurintricarboxylic acid, and ellagic acid were found to inhibit polys η and ι as well as DNA pol β (Dorjsuren et al. 2009). Additional compounds derived from fungi have also been shown to inhibit human Y-family DNA polys, these include 1-deoxyrubralactone (Naganuma et al. 2008), talaroflavone (Naganuma et al. 2008), penicilliols A and B (Kimura et al. 2009), 3-O-methylfunicone (Mizushina et al. 2009), and pinophilins A and B (Myobatake et al. 2012). Both 1-deoxyrubralactone and talaroflavone were able to inhibit both X and Y-family polys (pol β and pol κ were inhibited the most, and pol β was inhibited at IC₅₀ values nearly 10-fold lower). Penicilliols A and B were shown selectively to inhibit Y-family DNA polys η, ι, and κ compared to A, B, and X family polymerases. Pinophilin A and B were shown to inhibit A, B, and Y-family polys but not X-family polys. Pol κ was the Y-family pol most inhibited by 3-O-methylfunicone, found in Australian sea salt fungi. MK-866, candesarten ciletexil, and manoalide were found using a high throughput screening method analyzing over 15,000 compounds (Yamanaka et al. 2012).

There has been more work in the inhibition of other families of human polymerases, including the A, B, and X-families. RNA aptamers (30 nt) that were selected against pol β were shown to inhibit both X-family pol β and Y-family pol κ but did not show any inhibition with E. coli pol I Klenow fragment (Gening et al. 2006). Cholesterol hemisuccinate inhibits X-family polys β and λ at 2-6 µM while inhibition of A, B, and Y-family polys was at least 50-fold weaker and it did not inhibit the prokaryotic polymerases E. coli Pol I or Taq (Ishimaru et al. 2007). Two azaphilones, kasanosins A and B, derived from seaweed were found be inhibitors of polys β and λ (Kimura et al. 2008). Other specific inhibitors of pol λ have been isolated from a variety of fungi; these compounds include hymenoic acid (Nishida et al. 2008), and nodulisporol and
nodulisporone (Kamisuki et al. 2007). As mentioned previously, β-sitosteryl(6’-O-linoleoyl)-glucoside (from soybean extract) was shown to inhibit Y-family pols while two other compounds from soybean extracts, cerebroside and steroidal glycoside, were shown to inhibit pol λ selectively (Horie et al. 2010, Mizushina et al. 2014). Vitamins K_{1-3} and intermediates have been tested in polymerase inhibition. Vitamin K_{1} and K_{2} do not show inhibition to any family of polymerases, while Vitamin K_{3} inhibits A-family pol γ (Mizushina et al. 2011, Sasaki et al. 2008). Three intermediates between Vitamin K_{2} and K_{3} synthesized to assay for inhibition were able to inhibit A, B, X and Y-family polymerases to different extents (Mizushina et al. 2011) with intermediate 2 (MK-2) having the strongest inhibition (Mizushina et al. 2011).

Currently, inhibitors of E. coli and other prokaryotic DNA polymerases are not as commonly sought after. Work completed on some inhibitors of E. coli DinB and human pol κ will be discussed in Chapter 5.

1.7 Significance

The purpose of this work is to gain a better understanding of DNA damage specificity of E. coli DinB and human pol κ. Alanine scanning mutagenesis of the loop 1 residues of DinB and pol κ was performed to determine which residues in the loop adjacent to the active site are important for damage bypass and fidelity of incorporation (Chapter 2). In order to understand the contribution of the differences in loop size to major groove DNA adduct bypass, loop swaps of DinB, pol κ, and archaeal Dpo4 were created and assayed (Chapter 3, 4). Single nucleotide polymorphisms of pol κ found in cancer patients were biochemically characterized for their DNA damage specificity and fidelity of nucleotide incorporation (Chapter 5). Finally, potential
polymerase inhibitors were found to inhibit DinB but inhibit pol κ to a much less extent (Chapter 6).
1.8 References


, and Y-family DNA polymerases and human cancer cell proliferation." Journal of Natural Products 75(2): 135-141.


Bioorganic and Medicinal Chemistry 16(9): 5115-5122.


Chapter 2: Loop 1 residues of DinB and pol κ are important for activity and fidelity

Paul Ippoliti and Jason Walsh designed primers for DinB variants, created and purified variants and performed primer extension and misincorporation assays. Khadijah Balfour-Jeffrey performed nitrofurazone survival assays. Nicole M. Antczak designed primers for pol κ variants, created and purified variants and performed primer extension and misincorporation assays. N.M.A. also performed kinetics and thermal shift assays for DinB and pol κ variants.

2.1 Introduction

DNA damage is a constant threat to all life forms from both endogenous and exogenous sources. Radiation from UV light can cause thymine-thymine cyclobutane pyrimidine dimers and (6-4) photoproducts (Becherel et al. 1999, Fujii et al. 2004a, Fujii et al. 2004b, Tang et al. 2000). Cigarette smoke and diesel exhaust can induce the formation of $N^2$-benzo[a]pyrene-dG (Shen et al. 2002). And finally the $N^2$-furfuryl-dG and other $N^2$-dG adducts can be formed from exposure to nitrofurazone and 4-nitroquinoline oxide (Jarosz et al. 2006, Panigrahi et al. 1990, Whiteway et al. 1998). Many cancer drugs work by binding to DNA, like cisplatin, which forms interstrand and intrastrand crosslinks between adjacent guanines (Jamieson et al. 1999). These DNA adducts can block replication by replicative polymerases and in order for replication to continue cells employ DNA repair pathways or DNA polymerases called Y-family polymerases. Y-family DNA polymerases are found in all forms of life (Ohmori et al. 2001) and bypass damaged DNA by a process called translesion synthesis (TLS) (Radman 1975). *E. coli* DinB (DNA polymerase IV) and human pol κ are capable of bypassing adducts that are formed at the $N^2$ position of deoxyguanine (Irimia et al. 2009, Jarosz et al. 2006, Shen et al. 2002, Tang et al. 2000, Yuan et al. 2008a).
The DNA damage specificity of Y-family polymerases is not well understood. Even though Y-family polymerases have larger, more accessible active sites to accommodate bulky lesions, the active sites of the different Y-family polymerases are different sizes and shapes. The loop 1 residues directly adjacent to the nascent base pair in the active site are likely to play a role in replication by Y-family polymerases. The loop 1 region of DinB consists of residues 38 to 46 within the fingers domain, adjacent to the incoming nucleotide/template base, as seen in Figure 2.1. The loop 1 residues of pol κ include residues 135 to 141, as seen in Figure 2.1.

In DinB, it has been shown that loop 1 residue Ser42 interacts with the incoming nucleotide via polar interactions (both direct and water-mediated) and increases the fidelity of incorporation (Sharma et al. 2013). Sharma, et al., found that wild-type DinB misincorporates nucleotides with many combinations of template base and incoming nucleotide. DinB erroneously incorporates dCTP opposite template dA, misincorporates dCTP, dGTP, and dTTP opposite template dT, misincorporates dATP and dTTP opposite template dG, and misincorporates dATP, dGTP, and dTTP opposite template dC (Sharma et al. 2013). When the same assay was repeated with DinB Ser42Ala, there was no misincorporation opposite the templating bases dA and dG and much less dTTP misincorporation opposite dT and dC (Sharma et al. 2013). Previous work done in the Beuning laboratory with DinB shows that loop 1 residues are important for activity on undamaged DNA and 8-oxoG-containing DNA templates as well as correct nucleotide incorporation (Ippoliti 2012).

In pol κ, it has been shown that Met135 is important for the stabilization of the syn conformation of 8-oxoG by creating van der Waals and stacking interactions (Vasquez-Del Carpio et al. 2009). Creating a Met135Ala mutation caused a 36-fold drop in correct dCTP incorporation opposite an undamaged dG and a 56-fold decrease opposite 8-oxoG (Vasquez-Del Carpio et al. 2009).
In this work, we used alanine-scanning mutagenesis to probe the functions of the residues in the active site loop of DinB and human pol κ. We find that some residues are important for fidelity while other residues are important for replication activity.

Figure 2.1. Crystal structure of DinB (dark magenta) (PDB ID 4IRC (Sharma et al. 2013) aligned with the crystal structure of pol κ (dark blue) (PDB ID 4U6P (Jha et al. 2016). Highlighted in magenta and cyan are the loop 1 (inset) residues of DinB and pol κ, respectively. The DNA is in gray and the incoming nucleotide analog is in red.

The kinetics of the incorporation of dCTP (and dTTP where applicable) were determined on both DNA templates in this work. Additionally, site directed mutagenesis was performed on the loop 1 residues of human pol κ (residues 135-141) and the assays were repeated using pol κ and variants.

2.2 Materials and Methods

Protein Expression and Purification. Variants of both proteins were made via site-directed mutagenesis using a QuikChange kit (Agilent), using appropriately designed primers as well as the appropriate expression vector, pDFJ1 for DinB (performed by Jason Walsh and Paul Ippoliti)
(Beuning et al. 2006) and pBG101 for pol κ (Irimia et al. 2009). The DNA was sequenced to confirm the mutations (MGH DNA core, Macrogen, or Eton Bioscience).

For DinB, BL21 DE3 pLysS cells were transformed with the plasmids containing wild-type DinB or the loop 1 mutations. A 50-mL starter culture of the cells was grown overnight in Luria broth (LB) with ampicillin and chloramphenicol (100 µg/mL for ampicillin and 25 µg/mL of chloramphenicol). A 1-L culture of LB containing ampicillin and chloramphenicol (100 µg/mL for ampicillin and 25 µg/mL of chloramphenicol) was inoculated with the overnight culture and was allowed to grow for 2-3 h until an optical density (OD$_{600}$ nm) between 0.8 to 1.0 was reached. To induce expression of DinB, isopropyl β-D-thiogalactopyranoside (IPTG) was added to the culture at a final concentration of 1 mM and the cells were moved to 30 °C to shake for 3 h. The culture was then harvested at 6000 x g for 10 min at 4 °C. The pellet was then stored at -80 °C until purification.

The pellet was thawed on ice overnight at 4 °C in lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, and 2 mM β-mercaptoethanol) also containing a protease inhibitor tablet (Roche) and 10 µg/mL of phenylmethylsulfonyl fluoride (PMSF), a serine protease inhibitor. The cells were lysed by sonication and additional lysis was done with lysozyme and DNaseI. A freeze-thaw cycle was done and then the pellet was centrifuged at 12,000 x g at 4 °C for 1 h to remove the cell debris. The solubilized DinB was filtered and purified first with a FastFlow MonoS strong cation exchange column (GE Healthcare) using buffers Sa (50 mM HEPES, pH 7.5, 100 mM NaCl, and 2 mM β-mercaptoethanol) and Sb (50 mM HEPES, pH 7.5, 1 M NaCl, and 2 mM β-mercaptoethanol). Over five column volumes a gradient to 100% Sb was used to isolate the protein, and collected fractions were analyzed using SDS-PAGE. The fractions containing DinB were mixed 1:1 with PSa buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 1 M (NH$_4$)$_2$SO$_4$, and 2
mM β-mercaptoethanol) before loading on a phenyl sepharose column (GE Healthcare). Over five column volumes a gradient to 100% PSb (50 mM HEPES, pH 7.5, 100 mM NaCl, and 2 mM β-mercaptoethanol) was used to isolate the protein, and collected fractions were analyzed again using SDS-PAGE. The fractions containing DinB were collected and dialyzed twice with dialysis/storage buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, and 2 mM β-mercaptoethanol) for no less than 3 h with new buffer each time. Bradford assays were completed to determine the concentration of WT DinB and each variant using bovine serum albumin (BSA) as a standard. The protein was stored in the -80 °C.

For pol κ, Rosetta cells were transformed with the plasmids containing the catalytic core of pol κ (residues 19-526) fused to a His-GST tag, which we refer to as wild-type pol κ, or the corresponding plasmid with loop 1 mutations. A 50-mL culture of cells was grown overnight in LB containing 30 µg/mL of kanamycin. A 1-L culture of LB with 30 µg/mL of kanamycin was inoculated with the overnight culture and allowed to grow for 2-3 h until an OD$_{600\text{ nm}}$ of 0.6 to 0.8 was reached. To induce expression of pol κ, IPTG was added to the culture at a final concentration of 0.05 mM and the culture moved to room temperature to shake for an additional 3 h. The culture was then harvested at 6000 x g for 10 min at 4 °C. The pellet was then stored at -80 °C until purification.

For purification the pellet was thawed on ice overnight at 4 °C in lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, and 2 mM β-mercaptoethanol), ¼ protease inhibitor tablet (Roche), and 10 µg/mL of phenylmethylsulfonyl fluoride (PMSF). The cells were lysed by sonication and additional lysis was done with lysozyme and DNaseI. A freeze-thaw cycle was carried out and then the pellet was centrifuged at 17,100 x g at 4 °C for 1 h to remove the cell debris. The solubilized pol κ was filtered and purified first with a Histrap HP (GE Healthcare), a
nickel sepharose column to bind the histidine tag, using Histrap buffer A (50 mM HEPES, pH 7.5, 500 mM NaCl, and 2 mM β-mercaptoethanol) and Histrap buffer B (50 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM Imidazole pH 7.5, and 2 mM β-mercaptoethanol). To elute the protein from the affinity column 100% Histrap buffer B was used and the collected fractions were analyzed using SDS-PAGE. The fractions containing the protein were collected and diluted two-fold using GST binding buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, and 2 mM β-mercaptoethanol) before loading on a Glutathione Sepharose 4B (GST) column (GE Healthcare) using GST binding buffer and cleavage buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM ethylenediaminetetraacetic acid, and 1 mM dithiothreitol). The protein was loaded on the GST column, washed with 10 mL of GST binding buffer, then PreScission Protease was loaded on the column, and incubated overnight to remove the two tags by proteolysis. The untagged pol κ, with a four-amino acid scar (GlyProGlySer), was then eluted off the column using cleavage buffer. A Bradford assay was carried out to determine the concentration of the WT pol κ and variants using BSA as a standard and the protein was stored in the -80 °C.

**Nitrofurazone survival assays.** The strain AB1157ΔdinB containing the DinB single alanine loop 1 variants was exposed to increasing concentrations of nitrofurazone as described previously (Beuning et al. 2006, Hawver et al. 2011, Hawver et al. 2015, Shurtleff et al. 2009). Serial dilutions of overnight cultures were plated on LB agar plates containing 100 μg/mL ampicillin and 2 μg/mL nitrofurazone and incubated at 37 °C for 20-24 h. Error bars correspond to the standard error from at least three trials.
**Primer Extension Assay.** The primer extension assays were completed with 61-mer DNA template (undamaged or containing a single \(N^2\)-furfuryl-dG) and either running start or standing start DNA. The template DNA sequence was \(^5\)-

\[
\text{CGTTACTCAGATCGGCTGGCGAGACCTXGGCGTGCGTACTATCATATGC} \text{ } ^3\]

where \(X\) is either an undamaged G or \(N^2\)-furfuryl-dG. The running start sequence was \(^5\)-

\[
\text{GCATATGATAGTACAGCTGCAGCCGACGCX} \text{ } ^3\]

and the standing start sequence was \(^5\)-

\[
\text{GCATATGATAGTACAGCTGCAGCCGACGCX} \text{ } ^3\]. The DNA was synthesized by Operon, Eriks Rozners (SUNY Binghamton), or Ke Zhang (Northeastern University).

The DNA template containing a single \(N^2\)-furfuryl-dG was prepared as previously described (DeCorte et al. 1996, Jarosz et al. 2006). DNA was purified by denaturing polyacrylamide gel electrophoresis and the crush and soak method (Sambrook et al. 1989). DNA primers (running start or standing start) were end-labeled with \(^32\)P as previously described (Beuning et al. 2006, Sambrook et al. 1989).

For all assays, DNA template (unmodified or \(N^2\)ffdG) was combined with \(^32\)P-labeled primer (running start, or standing start) in a 1:1 ratio at 500 nM and annealed in annealing buffer (20 mM HEPES (pH 7.5) and 5 mM Mg(OAc)$_2$) by heating for 2 min at 95 °C, incubating at 50 °C for 60 min and cooling to 37 °C. Reactions were carried out in 1X reaction buffer (30 mM HEPES (pH 7.5), 20 mM NaCl, 7.5 mM MgSO$_4$, 2 mM β-mercaptoethanol, and 1% bovine serum albumin), 100 nM \(^32\)P labeled primer/template. For primer extension, to study replication past the lesion, running start primer and all four dNTPs at a final concentration of 500 µM were used to initiate the reactions containing 25 nM DinB or 10 nM pol κ. For misincorporation assays, the final concentrations in 1X reaction buffer used were 100 nM \(^32\)P labeled primer/template with standing start primer, 10 nM pol κ or 25 nM DinB, and 1 mM dATP,
dCTP, dGTP or dTTP in separate reactions. Kinetics assays for DinB were done at 100 nM $^{32}$P standing start primer/template, a range of protein concentrations from 1 to 10 nM depending on the overall activity of the protein and a range of 5 µM to 1000 µM dCTP (or dTTP in cases of misincorporation). Kinetics assays for pol κ were done at 100 nM standing start primer/template, a range of protein concentrations from 1 to 10 nM depending on the overall activity of the protein and a range of 0.1 µM to 100 µM dCTP (or dTTP in cases of misincorporation). All reactions for the assays were quenched with 85% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue, and 50 mM EDTA. Samples were analyzed on 12-16% polyacrylamide 8 M urea gels and imaged on Molecular Dynamics storage phosphor imaging screens with a Storm 860. Results were analyzed using ImageQuant software and for kinetics $V_{\text{max}}$ and $K_m$ were determined using GraphPad Prism to fit the data to the Michaelis-Menten equation. Gels are representative of at least two replicates. Kinetics values represent the average of at least three trials and error represents the standard deviation.

**Thermal shift assays.** Thermal shift assays were conducted using a Bio-Rad CFX as previously described (Nevin et al. 2015a, Nevin et al. 2015b) using T18 undamaged DNA. Reaction mixtures containing 5 µM each wild-type protein or chimera, 10 µM dd13 primer/T18 template DNA, 1 mM dNTP (dCTP or dTTP) in 1X reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO$_4$, and 2 mM β-mercaptoethanol). Reactions were incubated at room temperature for 20 min before addition of Sypro Orange (Invitrogen) to a final concentration of 20x. Values represent the average of at least three trials and error bars represent the standard deviation.
2.3 Results
Analysis of Loop 1 in DinB and pol κ. We identified active site loop 1 as directly adjacent to both nucleotides of the nascent base pair in crystal structures of *E. coli* DinB and human pol κ (Jha et al. 2016, Sharma et al. 2013) (Figure 2.1). This loop is reasonably well conserved among members of the DinB subfamily and distinct from other Y-family polymerases, although the loop length varies even among the DinB subfamily members (Gerlach et al. 1999, Lee et al. 2006). Given its proximity to the nascent base pair, we decided to test the role of this loop in polymerase activity by using alanine-scanning mutagenesis to obtain the Arg38Ala, Val40Ala, Ile41Ala, Ser42Ala, Thr43Ala, Asn45Ala, and Tyr46Ala mutations in DinB and Met135Ala, Leu136Ala, Ser137Ala, Thr138Ala, Ser139Ala, Asn140Ala, and Tyr141Ala mutations in pol κ. Upon alignment of the DinB and pol κ structures and loop sequences, as shown in Figure 2.1, it is clear that DinB Arg38 does not align with any of the pol κ loop 1 residues since the loop of DinB is longer than that of pol κ. Ser139 of pol κ aligns with DinB Ala44 so there is no corresponding mutation made in DinB.

Most DinB variants confer nitrofurazone resistance. DinB has been shown to confer resistance to nitrofurazone in *E. coli*. We first tested the ability of the DinB variants to confer resistance to nitrofurazone in an *E. coli* strain in which *dinB* has been deleted from the chromosome (Jarosz et al. 2006). As shown in Figure 2.2, cells containing Arg38Ala, Ile41Ala, and Asn45Ala have low survival upon exposure to 2 μg/mL of nitrofurazone. The remaining variants confer similar survival as wild-type DinB.
Figure 2.2. Most DinB variants confer survival upon exposure to nitrofurazone.

DinB loop 1 residues have a range of activity on undamaged and $N^2$ffdG DNA. DinB variants harboring single-Ala mutations in loop 1 were assayed for primer extension activity. Figure 2.3 shows that Arg38Ala, Ile41Ala, Asn45Ala, and Tyr46Ala have little to no activity on both undamaged DNA and DNA containing the $N^2$ffdG lesion. However, both Arg38Ala and Tyr46Ala have slightly increased activity on the $N^2$ffdG damaged template compared to the undamaged template. The remaining variants, Val40Ala, Ser42Ala, and Thr43Ala, have activity similar to wild-type DinB on both DNA templates.
Figure 2.3. DinB loop 1 residues have different activities on undamaged and $N^2$ffdG DNA as compared to WT DinB. As shown previously (Jarosz et al. 2006), DinB and variants have an increase in activity opposite on DNA containing the $N^2$ffdG adduct (B) as compared to undamaged DNA (A) (Ippoliti 2012).

WT DinB and loop 1 variants have increased misincorporation opposite $N^2$ffdG.

Misincorporation assays were also performed with the wild-type DinB and loop 1 variants on both the undamaged and $N^2$ffdG containing DNA templates. On undamaged DNA, wild-type DinB and all the variants incorporate only correct dCTP, with the exception of Ile41Ala and Asn45Ala which are inactive (Figure 2.4). Wild-type DinB correctly inserts dCTP opposite the $N^2$ffdG lesion and to a lesser extent misinserts dTTP. The three variants Val40Ala, Ser42Ala, and Thr43Ala also misinsert dTTP. DinB Ser42Ala exhibited the most dTTP misincorporation while DinB Val40Ala and DinB Thr43Ala show less misincorporation than wild-type. The other active variants, DinB Arg38Ala and Tyr46Ala, do not exhibit dTTP misincorporation.
Figure 2.4. DinB loop 1 variants have increased misincorporation opposite $N^2$-furfuryl-dG. (A) Misincorporation assay of WT DinB and loop 1 variants on undamaged DNA. (B) Misincorporation assay of WT DinB and loop 1 variants on $N^2$-furfuryl-dG (Ippoliti 2012).

**DinB loop 1 variants kinetics assays.** Kinetics on Ile41Ala and Asn45Ala were not completed due to the very low activity of the proteins. As seen in Table 2.1, WT DinB, Val40Ala, and Ser42Ala have similar $K_m$ values for dCTP when the templating base is $N^2$-furfuryl-dG. The remaining active variants have higher $K_m$ and lower $k_{cat}$ values for dCTP than WT, Val40Ala, and Ser42Ala which corresponds to lower overall activity in the primer extension assays (Figure 2.3B, Table 2.1).

**Table 2.1.** DinB kinetics data on $N^2$ffdG containing DNA

<table>
<thead>
<tr>
<th>Protein</th>
<th>$V_{max}$ (μM min$^{-1}$)</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$</th>
<th>$k_{cat}/K_m$ (μM$^{-1}$ min$^{-1}$)</th>
<th>Fold decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N^2$ffdG:dC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT DinB</td>
<td>0.138 ± 0.037</td>
<td>52.6 ± 27.1</td>
<td>138</td>
<td>3.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Arg38Ala</td>
<td>0.054</td>
<td>167</td>
<td>5.42</td>
<td>0.033</td>
<td>95.2</td>
</tr>
<tr>
<td>Val40Ala</td>
<td>0.094</td>
<td>63.4</td>
<td>93.7</td>
<td>1.48</td>
<td>2.1</td>
</tr>
<tr>
<td>Ser42Ala</td>
<td>0.108 ± 0.018</td>
<td>55.0 ± 22.0</td>
<td>108</td>
<td>2.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Thr43Ala</td>
<td>0.181 ± 0.119</td>
<td>110 ± 1.41</td>
<td>36.2</td>
<td>0.328</td>
<td>9.42</td>
</tr>
<tr>
<td>Tyr46Ala</td>
<td>0.114 ± 0.035</td>
<td>156 ± 27.4</td>
<td>11.4</td>
<td>0.077</td>
<td>40.3</td>
</tr>
</tbody>
</table>

WT, Val40Ala, Ser42Ala 1nM; Thr43Ala5 nM; Arg38Ala, Tyr46Ala 10 nM
**Most pol κ loop 1 variants have primer extension activity similar to WT pol κ.** As seen with the DinB loop 1 variants, there is a range in activity with Met135Ala, Leu136Ala, Ser137Ala, Thr138Ala, and Ser139Ala having full extension ability on undamaged DNA after 60 min. Ser137Ala is the most similar to WT pol κ, as Ser42Ala is to DinB. Asn140Ala, and Tyr141Ala have much lower activity and no full extension (Figure 2.5). Unlike the DinB loop 1 variants, all of the pol κ variants have at least some primer extension activity.

On the DNA template containing N²ffdG, there is a similar trend. As shown in Figure 2.5B Ser137Ala, Thr138Ala, and Ser139Ala are again the most active and can extend to the end of the template harboring the N²ffdG lesion. Met135Ala and Leu136Ala have similar activity as that observed on undamaged DNA while Asn140Ala and Tyr141Ala have an increase in activity. Similar to the activity of pol κ variants on the undamaged DNA template, none of the pol κ variants are completely inactive on DNA containing N²ffdG.
Figure 2.5. Primer extension assays of wild-type pol κ and loop 1 variants: Met135Ala, Leu136Ala, Ser137Ala, Thr138Ala, Ser139Ala, Asn140Ala, and Tyr141Ala on (A) undamaged DNA and (B) $N^2$-furfuryl-dG DNA.

WT pol κ and loop 1 variants have increased misincorporation opposite $N^2$ffdG. The misincorporation assays were completed with standing start primer to determine which nucleotide can be inserted by each protein opposite undamaged dG or $N^2$ffdG. On the undamaged template it can be seen that WT pol κ correctly incorporates dCTP but also misincorporates dTTP opposite dG to an extent (Figure 2.6). As seen in Figure 2.6, Ser137Ala, Thr138Ala, and Ser139Ala correctly incorporate dCTP and like WT misincorporate dTTP. Ser137Ala has the highest amount of dTTP misincorporation and is also as active as or more active than WT in the primer extension assay (Figure 2.5). Met135Ala, Leu136Ala, Asn140Ala, and Tyr141Ala are more accurate than WT as there is no dTTP misincorporation seen after 30 minutes.

In the misincorporation assay opposite $N^2$ffdG, WT pol κ correctly incorporates dCTP and also misincorporates dTTP. Leu136Ala, Asn140Ala, and Tyr141Ala are more accurate than WT pol κ, as they only incorporate dCTP opposite $N^2$ffdG (Figure 2.6). Met135Ala, Ser137Ala, Thr138Ala, and Ser139Ala are less accurate opposite $N^2$ffdG as they show more dTTP misincorporation as compared to WT and opposite undamaged dG. Ser139Ala also misincorporates dGTP opposite $N^2$-ffdG.
Figure 2.6. Primer extension assays of wild-type pol κ and loop 1 variants: Met135Ala, Leu136Ala, Ser137Ala, Thr138Ala, Ser139Ala, Asn140Ala, and Tyr141Ala on (A) undamaged DNA and (B) $N^2$-furfuryl-dG DNA.

Some of the least active DinB and pol κ loop 1 variants have the lowest melting temperatures. As we have shown previously (Nevin et al. 2015a, Nevin et al. 2015b), there is an increase in melting temperature in the presence of DNA and the correct incoming nucleotide. For the DinB WT and loop 1 variants there is an increase in melting temperature in the presence of DNA and the correct nucleotide (Figure 2.7A). The DinB variants Ile41Ala and Asn45Ala have dramatically decreased $T_m$. Like other variants, the $T_m$ is increased in the presence of substrates, but still not to the level of that of wild-type.
Figure 2.7. Melting temperatures of WT DinB and pol κ and loop 1 variants of both proteins. The least active variants have the lowest melting temperatures. (A) DinB WT and loop 1 variants show an increase in melting temperature in the presence of DNA and the correct nucleotide. (B) WT pol κ and loop 1 variants.

2.4 Discussion

Wild-type pol κ, like its ortholog DinB, has high activity on both undamaged and $N^2$-furfuryl-dG templates (Figure 2.5) and previous work has shown that both DinB and pol κ are more active on $N^2$-furfuryl-dG DNA than undamaged DNA (Jarosz et al. 2006). Previous work has also shown that DinB bypasses $N^2$-dG adducts as well as undamaged dG with relative accuracy (Jarosz et al. 2006, Seo et al. 2006, Shen et al. 2002, Yuan et al. 2008b). The results presented here suggest that at relatively high concentrations of wild-type DinB and dNTPs, as well as prolonged incubation time, wild-type DinB does misincorporate nucleotides as has been seen previously (Jarosz et al. 2006, Sharma et al. 2013). Here we have shown that wild-type pol
κ misincorporates dTTP opposite both undamaged dG as well as $N^2$ffdG while DinB weakly misincorporates dTTP opposite $N^2$ffdG. The loop 1 variants for both DinB and pol κ do not show any damage specificity but are important in catalytic activity and correct nucleotide incorporation. Recent single nucleotide incorporation work has shown that pol κ misincorporates dATP, dGTP, and dTTP opposite an undamaged dG (Jha et al. 2016).

DinB variant Arg38Ala has weak activity on both undamaged and damaged DNA (Figure 2.3) and DinB Arg38Ala does not show any misincorporation (Figure 2.4). In Figure 2.1, it can be seen that Arg38 is located on the portion of the loop near the single-stranded DNA template that is entering the active site. The decrease in activity of Arg38Ala compared to wild-type DinB could affect the efficiency of DinB to replicate DNA, as the positive charge on Arg38 could help stabilize the negatively charged DNA as it is translocated during replication. This can also be seen in the in vivo nitrofurazone exposure, as cells containing the DinB Arg38Ala variant have low survival. This is also supported with the kinetics with $N^2$ffdG-containing DNA, which resulted in nearly a two order of magnitude drop in catalytic efficiency (Table 2.1) and represents the largest loss of activity of the variants characterized. Additional support for the large detrimental effect of the Arg38Ala mutation comes from the melting temperatures, as there is no increase in melting temperature in the presence of undamaged DNA as we have shown previously (Figure 2.7A) (Nevin et al. 2015a, Nevin et al. 2015b).

DinB Val40Ala (Figure 2.3) has strong activity on both undamaged and $N^2$-furfuryl-dG DNA, similar to wild-type DinB while pol κ Met135Ala (Figure 2.5) has weaker activity on both templates. Val40 and Met135 are located adjacent to the template base opposite the incoming nucleotide, shown in Figure 2.1, and therefore, we suspected these residues would contribute to DNA damage specificity based on their location near the damaged template base. Val40Ala and
Met135Ala are accurate opposite undamaged dG but Val40Ala also misincorporates dTTP opposite $N^2$-furfuryl-dG. Previous work has shown that Met135 is important for the stabilization of the template base and the Met135Ala mutation caused a 36-fold decrease in activity on undamaged DNA (Vasquez-Del Carpio et al. 2009). Kinetics of Val40Ala opposite $N^2$-ffdG result in similar $K_m$ and catalytic efficiency as WT DinB (Table 2.1). Kinetics for pol $\kappa$ Met135Ala are currently being pursued. Any stabilization of the template base done by these residues could be important for nucleotide incorporation when damaged DNA is present.

An alanine mutation to the residues neighboring DinB Val40 and pol $\kappa$ Met135, DinB Ile41Ala and pol $\kappa$ Leu136Ala, respectively, show little to no activity on either DNA template. In Figure 2.1, it is seen that Ile41 and Leu136 point away from the template nucleotide and incoming nucleotide base pair and towards the rest of the fingers domain. Leu136Ala has little activity on both undamaged and $N^2$-furfuryl-dG DNA and is more faithful than wild-type pol $\kappa$, incorporating only dCTP opposite both undamaged and $N^2$-furfuryl-dG. It is possible that these residues contribute to the stability and folding of the protein and removing the hydrophobic chain affects the structure, in turn affecting the activity. This is supported by the in vivo results in which cells harboring the Ile41Ala mutation have very low survival upon exposure to nitrofurazone (Figure 2.2), and by the observation that the $T_m$ of DinB Ile41Ala is below 20 °C, and more than 20 °C lower than that of WT DinB (Figure 2.7A).

DinB Ser42Ala and pol $\kappa$ Ser137Ala show high activity on both DNA templates and have an increase in misincorporation opposite $N^2$-ffdG relative to incorporation opposite dG, similar to both wild-type proteins. Since neither variant shows a preference for one type of DNA template, it can be concluded that the residues are not involved in $N^2$-ffdG damage specificity. As seen in Figure 2.1, both Ser42 and Ser137 are directly adjacent to the incoming nucleotide. It has
been shown that Ser42 interacts with the incoming nucleotide via direct and water mediated hydrogen bonds and when mutated to alanine there is a decrease in misincorporation (Kottur et al. 2015, Sharma et al. 2013). This differs from the results we obtained as Ser42Ala has an increase in dTTP misincorporation opposite $N^2$ffdG but not opposite undamaged DNA (Figure 2.3). It was also shown that cells containing the Ser42Ala had a decrease in survival upon exposure to 4 µg/mL nitrofurazone (Kottur et al. 2015), while we found that upon 2 µg/mL of nitrofurazone cells containing the same mutation had similar survival to cells with WT DinB (Figure 2.2). This difference could be due to the difference in nitrofurazone dosage, at a lower concentration the differences between Ser42Ala and WT might not be as clear. In addition, the work reported by Kottur uses a construct of DinB that is expressed with a His-tag that is subsequently removed by proteolysis, leaving a small scar, whereas our construct is not tagged.

Ser137Ala is much less accurate and has a much higher dTTP misincorporation on both templates as compared to wild-type pol κ. In pol κ, the loss of the hydroxyl group decreases the accuracy much more than in DinB as there is dTTP misincorporation opposite undamaged dG that is not seen with DinB Ser42Ala.

DinB Thr43Ala and pol κ Thr138Ala are less active than wild-type DinB and pol κ; the difference between the two variants lies in the misincorporation in which Thr43Ala does not misincorporate while Thr138Ala does opposite dG and $N^2$ffdG. However, Thr138Ala has a similar pattern of misincorporation as WT pol κ. As mentioned previously, Thr138 has been shown computationally to interact with the β-phosphate of the incoming dCTP during bypass of $N^2$-benzo[a]pyrene-dG adduct (Lior-Hoffmann et al. 2012). This provides an explanation for the decrease in activity as the incoming nucleotide would lose stability with the loss of the hydrogen
bonding or electrostatic power of the hydroxyl group. This is also supported by a nearly 10-fold decrease in catalytic efficiency opposite $N^2$ffdG by DinB Thr43Ala compared to WT DinB.

Unlike in DinB, there is a third residue with a hydroxyl in loop 1 of pol κ, Ser139, which aligns with Ala44 in DinB (Figure 2.1). Ser139Ala behaves similarly to Ser137Ala but its misincorporation opposite undamaged dG is similar to WT pol κ. In the crystal structure (Figure 2.1) Ser139 appears to be pointing away from the incoming nucleotide. Ser139 could be important structurally and removing the hydroxyl group could affect hydrogen bonding surrounding the loop.

DinB Asn45Ala shows no activity on both undamaged and $N^2$-furfuryl-dG DNA templates (Figure 2.3) while pol κ Asn140Ala has weak primer extension activity on both templates (Figure 2.5). DinB Asn45Ala, like Ile41Ala, may be folded improperly as the melting temperature of the protein is around 24 °C.

Pol κ Tyr141Ala is similar to Tyr46Ala of DinB, with low activity on both templates and no misincorporation seen on either template. Tyr141 and Tyr46 are in similar positions in both proteins, adjacent to the phosphates of the incoming nucleotide. The loss of the hydroxyl group could affect interactions with the incoming nucleotides and decrease the overall incorporation ability of the proteins. It has been shown that Tyr141 hydrogen bonds with the $\gamma$-phosphate of the incoming nucleotide during the transition state (Lior-Hoffmann et al. 2012); this supports the low activity of both pol κ Tyr141Ala and DinB Tyr46Ala. This is supported by the kinetics of DinB Tyr46Ala, as there is a 40-fold decrease in catalytic efficiency opposite $N^2$ffdG as compared to WT.

Overall the loop 1 variants for both DinB and pol κ are important for catalytic activity and the incorporation of the incoming nucleotide. The residues that are important for catalytic
activity are different from the residues that are important for correct nucleotide incorporation. In addition, some of the residues in the active site loop appear to play important roles in folding and stability.
2.5 References
DeCorte, B. L., D. Tsarouhtsis, S. Kuchimanchi, M. D. Cooper, P. Horton, C. M. Harris and T. M. Harris (1996). "Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N2 position of deoxyguanosine." Chemical Research in Toxicology 9(3): 630-637.


Chapter 3: Pol κ is more tolerant than DinB to changes in the active site loop 1

Reprinted with permission from
Human Y-Family DNA Polymerase κ Is More Tolerant to Changes in Its Active Site Loop than Its Ortholog Escherichia coli DinB
Nicole M. Antczak, Morgan R. Packer, Xueguang Lu, Ke Zhang, and Penny J. Beuning
Copyright 2017 American Chemical Society.

3.1 Introduction
Y-family DNA polymerases are conserved in all domains of life and are characterized by their ability to bypass damaged DNA in a process called translesion synthesis (TLS) (Friedberg 2006, Ohmori et al. 2001). As DNA damage is ubiquitous, it is important for cell survival to have robust mechanisms to tolerate DNA damage. DNA damage can be accommodated more readily by Y-family DNA polymerases than replicative DNA polymerases as the fingers domain in Y-family polymerases is smaller than that of replicative DNA polymerases, limiting contacts with the major groove side of the nascent base pair (Ling et al. 2001, Pata 2010, Wang et al. 2005, Yang et al. 2007).

Y-family DNA polymerases can bypass a wide variety of DNA damage from adducts due to reactive oxygen species to UV photoproducts, interstrand DNA crosslinks, and protein-DNA crosslinks, including those formed by DNA-damaging chemotherapy agents (Albertella et al. 2005, Bassett et al. 2004, Chen et al. 2006, Friedberg 2006, Kanemaru et al. 2017, Shao et al. 2014, Walsh et al. 2011b, Zhuo et al. 2017). The minor groove N2 position of deoxyguanine is particularly vulnerable to modifications and adducts of varying sizes (Basu et al. 2017, Nair et al. 2015, Yuan et al. 2008a). Adducts at the N2 position inhibit replicative DNA polymerases due to steric clashes. Although there is vast chemical diversity of DNA lesions, some Y-family DNA polymerases exhibit preferences for specific lesions. For example, Escherichia coli DinB (DNA
polymerase IV) has been shown to bypass proficiently deoxyguanine harboring minor groove adducts at the $N^2$ position, like $N^2$-furfuryl-dG ($N^2$-ffdG), which DinB bypasses 15-fold more efficiently than undamaged dG (Jarosz et al. 2006, Kottur et al. 2015, Kumari et al. 2008, Minko et al. 2008b, Seo et al. 2006, Shen et al. 2002). The DinB ortholog human pol κ, has also been shown to bypass $N^2$-dG adducts including bulky adducts like benzo[a]pyrene and $N^2$-ethyl-dG (Choi et al. 2006a, Choi et al. 2006d, Gowda et al. 2017, Hori et al. 2010, Minko et al. 2008a, Nair et al. 2015, Rechkoblit et al. 2002, Suzuki et al. 2002). *Sulfolobus solfataricus* Dpo4 is also considered a DinB ortholog and is able to bypass small $N^2$ adducts; however, as the adduct size increases, the efficiency and fidelity decrease (Zhang et al. 2009). Dpo4 can bypass the $N^2$-benzo[a]pyrene adduct in an error prone manner due to a structural gap between the little finger (or PAD) and the fingers domain (Bauer et al. 2007). Pol κ has been shown to bypass the $N^2$-benzo[a]pyrene-dG adduct in an error-free manner (Jha et al. 2016, Jha et al. 2017, Jia et al. 2008). Pol κ is more open on the minor groove side of DNA compared to other Y-family polymerases; the little finger and thumb domains shift the template strand to accommodate the large adduct and the pol κ linker between the catalytic core and little finger also moves to open the minor groove side. The N-clasp plays a critical role in stabilizing the conformation of the protein and the template strand of the DNA (Jha et al. 2016, Jia et al. 2008, Lone et al. 2007).

We and others have shown that DinB and pol κ are generally inhibited by major groove adducts (Basu et al. 2017, Choi et al. 2006b, Levine et al. 2001, Walsh et al. 2011a, Walsh et al. 2013, Xu et al. 2016). For example, both DinB and pol κ are inhibited by $O^6$-methyl-dG (Choi et al. 2006b, Nevin et al. 2015b). Human pol κ is also inhibited by a crosslink of glutathione to $N^6$-dA (Sedgeman et al. 2017). We previously showed that major groove modifications inhibit DinB activity, inhibit pol κ to a lesser extent, and only modestly affect Dpo4 activity (Walsh et al. 2016).
In particular, DinB and pol κ are both inhibited by the presence of the major groove modification N⁶-furfuryl-dA (N⁶ffdA) in the DNA template. In the case of both major groove lesions O⁶-methyl-dG and N⁶ffdA, DinB is inhibited more strongly than pol κ (Nevin et al. 2015b, Walsh et al. 2013). It was shown that the mutation R35A in DinB eliminated its discrimination against N⁶ffdA (Walsh et al. 2013). The active site loop that contains R35 is of different lengths in the different DinB-subfamily members (Figure 3.1). Pol κ has the smallest loop; the DinB loop is four amino acids larger than the pol κ loop and the Dpo4 loop is three amino acids longer than the DinB loop and seven longer than pol κ. Thus, it was hypothesized that the differences in this active site loop dictate the differences observed in bypass of major groove adducts. To test this hypothesis, a series of chimeric DNA polymerases were constructed that contain step-wise changes to the active site loops of DinB and pol κ. We found that changing the length and amino acid composition of the active site loop of DinB had a greater effect on its activity than analogous changes in pol κ and that pol κ was remarkably tolerant to changes in its active site loop.
Figure 3.1. Alignment of pol κ (light blue, PDB 2W7O), DinB (green, PDB 4IR9), and Dpo4 (purple, PDB 3QZ7) based on alignment of the Cα atoms of the amino acids. The active site loops are highlighted in the box and zoomed in. The primer/template is shown in gray and the incoming nucleotide is shown in red.

3.2 Materials and Methods

Proteins and DNA. Pol κ (residues 19-526) and DinB were expressed and purified as described previously (Beuning et al. 2006, Irimia et al. 2009). Pol κ and DinB chimeras were created by using QuikChange site-directed mutagenesis kits (Agilent). Mutations were confirmed by DNA sequencing (Macrogen, Cambridge, MA; or Eton Bioscience, Charlestown, MA) and the inferred amino acid sequences are shown in Tables 3.1-3.3. The chimeras were purified using the purification scheme of the respective parental protein.
**Table 3.1 DinBKappa loop swap A protein sequences**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT DinB</td>
<td>28PIAIGGSRERRGVISTANY45</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 1A</td>
<td>PIAIGGS___RRGVISTANY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 2A</td>
<td>PIAIGGS___GVISTANY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 3A</td>
<td>PIAVGSMGVISTANY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 4A</td>
<td>PIAVGSMSLSTANY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 5A</td>
<td>PIAVGSMSLSTSNY</td>
</tr>
<tr>
<td>WT Pol κ</td>
<td>127PIAVGSMSLSTSNY</td>
</tr>
</tbody>
</table>

**Table 3.2 DinBKappa loop swap B protein sequences**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT DinB</td>
<td>28PIAIGGRERRGVISTANY45</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 1B</td>
<td>PIAIGGRERRGV___NY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 2B</td>
<td>PIAVGRERRGVNY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 3B</td>
<td>PIAVGRERRGVNY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 4B</td>
<td>PIAVGSMSLGVNY</td>
</tr>
<tr>
<td>DinBKappa Loop Swap 5B</td>
<td>PIAVGSMSLSTNY</td>
</tr>
<tr>
<td>WT Pol κ</td>
<td>127PIAVGSMSLSTSNY</td>
</tr>
</tbody>
</table>
Table 3.3 Pol KappaDinB loop swap protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Pol κ</td>
<td>$^{127}$PIAVGSM$\underline{S}$MLSTNY$^{141}$</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 1</td>
<td>PIAVGSM$\underline{R}$ESMLSTANY</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 2</td>
<td>PIAVGSRERR$\underline{S}$MLSTANY</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 3</td>
<td>PIAVGS$\underline{M}$RR$\underline{G}$VLSTANY</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 4</td>
<td>PIAVGS$\underline{G}$RR$\underline{G}$VISTANY</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 5</td>
<td>PIAIGGS$\underline{R}$ERRGVISTANY</td>
</tr>
<tr>
<td>KappaDinB Loop Swap 6</td>
<td>$^{28}$PIAIGGS$\underline{R}$ERRGVISTANY$^{45}$</td>
</tr>
<tr>
<td>WT DinB</td>
<td>$^{28}$PIAIGGS$\underline{R}$ERRGVISTANY$^{45}$</td>
</tr>
</tbody>
</table>

The DNA template, 61-mer (5’-CGTTACTCAGATCGCTGCGAGAACCTGXCAGCTCCGAGCTGCAGCTGTACTATCAT-3’), Template18 (18-mer 5’-CCTGXCAGCTCCGAGCAACG-3’), where X is dG, $N^2$-furfuryl-dG, or $N^6$-furfuryl-dA, and primers, (running start, 30-mer 5’-GCATATGATAGTACAGCTGCAGCCGGACGCC3’; standing start 31-mer 5’-GCATATGATAGTACAGCTGCAGCC3’; dideoxy 13-mer (dd13) 5’-CGTTGCCGGACGCC3’) were obtained from Eurofins Operon or Integrated DNA Technologies. The 61-mer or 18-mer DNA templates containing a single $N^2$-furfuryl-dG (X) were prepared as previously described (DeCorte et al. 1996, Jarosz et al. 2006). The 61-mer and 18-mer DNA templates containing a single $N^6$-furfuryl-dA (X) were prepared as previously described from $O^6$-phenyl-deoxyinosine (Glen Research). DNA was purified by denaturing polyacrylamide gel electrophoresis and the crush and soak method (Sambrook et al. 1989). DNA primers (running start and standing start) were end labeled with $^{32}$P as previously described (Sambrook et al. 1989). The running start primer allows for one nucleotide to be inserted before the damaged or control base interest (X) and is used for primer extension assays, whereas with
standing start primer the first nucleotide added is opposite the damaged or control base of interest (X) and is used for the misincorporation assays.

**Primer Extension and misincorporation assays.** DNA template (where X is unmodified dG, N²ffdG, or N⁶ffdA) was combined with ³²P-labeled primer (running start for primer extension assays or standing start for misincorporation assays) in a 1:1 ratio (500 nM) and annealed with annealing buffer (20 mM HEPES, pH 7.5, and 5 mM Mg(OAc)₂) by heating for 2 min at 95 °C, incubating at 50 °C for 60 min and cooling to 37 °C. Reactions were carried out in 1X reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO₄, 2 mM β-mercaptoethanol, and 0.1 mg/mL bovine serum albumin) with 100 nM ³²P labeled primer/template, 10 nM protein, and 500 μM dNTPs. For misincorporation assays, the final concentrations in 1X reaction buffer were 100 nM ³²P labeled primer/template, 25 nM protein, and 1 mM dATP, dCTP, dGTP, or dTTP. Products were separated by denaturing 8 M urea 16% polyacrylamide gels and imaged on a Molecular Dynamics Storm 860 phosphorimager. Data were analyzed using ImageQuant TL1D Gel Analysis software (GE Healthcare) to determine percent nucleotide incorporation. Gels are representative of at least two replicates.

**Thermal Shift Assays.** Thermal shift assays were conducted using a Bio-Rad CFX as previously described (Nevin et al. 2015a, Nevin et al. 2015b) using T18 undamaged DNA or T18 with N²ffdG or N⁶ffdA modifications containing 3-5 μM each wild-type protein or chimera, 5-10 μM dd13 primer/T18 template DNA, 1 mM dNTP (dCTP or dTTP) in 1X reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO₄, and 2 mM β-mercaptoethanol). Reactions were incubated at room temperature for 20 min before addition of Sypro Orange (Invitrogen) to a final
concentration of 20x. Values represent the average of at least three trials and error bars represent the standard deviation.

**Homology models of the DinB and pol κ chimeras.** The FASTA sequences used for each protein were obtained from Uniprot, Q9UBT6 for pol κ (residues 19-526) and A7ZHZ2 for DinB. Residues 19-526 were used for pol κ as our laboratory uses the gene for the truncated protein and many crystal structures of pol κ are of this catalytic core. The loop regions of each protein were then swapped in the FASTA files to create the homology models. In the FASTA sequence of DinB, the sequence PIAIGGSRERRGVISTANYPARK was replaced with the corresponding pol κ loop sequence, PIAVGSMSMLSTSNYHARR, to create DinB with a pol κ loop (DinBKappa). This was repeated for pol κ by replacing the sequence specified above with the corresponding DinB loop sequence. The homology models were created using YASARA (Krieger et al. 2002) and were evaluated using a Ramachandran plot and Procheck via PDBsum. The Z score obtained from the homology modeling in YASARA is a measure of model quality calculated from the backbone dihedrals, side chain dihedrals, and packing interactions of the model, using the mean values and standard deviations of the corresponding scores for a set of high resolution crystal structures.

### 3.3 Results

**Design of polymerase chimeras.** In designing DinB and pol κ chimeras to test the roles of the active site loops, it was observed that the respective active site loops of DinB and pol κ have some conserved residues so those residues were not mutated; the changes (mutations, insertions, deletions) were made at most three at a time for each chimera to ascertain how the changes affect the protein activity and to allow a change of course if necessary. The active site loop of DinB is four amino acids longer than that of pol κ, so residues were first deleted from
DinB to make the loops the same size before changing the identities of the residues. Since we previously found that the R35A mutation in DinB allowed for bypass of the $N^6$ffdA adduct, we started by deleting R35 and E36 to make loop swap 1A (sequences, Table 3.1) followed by deleting R37 and R38 to make loop swap 2A and then changing the non-conserved residues by up to three at a time. Upon observing that LS2A-5A had much lower activity than LS1A (Figure 3.2), we created a different deletion site, deleting residues I41-A44 (LS1B-LS5B, sequences, Table 3.2). This approach created a mostly inactive set of chimeras (Figure 3.2). For the KappaDinB loop swap chimeras, the first step was to insert the missing four amino acids into pol $\kappa$ so the loop would be the same length as the DinB loop. First, R35 and E36 of DinB were inserted between pol $\kappa$ residues M32 and S33 (LS1, sequences Table 3.3), followed by R37 and R38 (LS2, Table 3.3). Like the DinB-Kappa loop swaps, the remaining changes were made changing up to three residues at a time.

**DinB-Kappa loop swaps have decreased activity and misincorporation on all DNA templates.** The first DinB-Kappa loop swap (sequences, Table 3.1), in which R35 and E36 are deleted, retains activity similar to that of WT DinB on all three DNA templates (Figure 3.2 A-C). The remaining DinB-Kappa loop swaps (LS2A-5A), which delete two additional residues and then change the identities of the residues in the middle of the loop, have decreases in activity with each change introduced, with LS5A having very low activity. Misincorporation assays revealed that all of the DinB-Kappa loop swap proteins retain DinB-like accuracy, incorporating only dCTP opposite undamaged dG unlike WT pol $\kappa$, which misincorporates dTTP (Figures 3.3 and 3.4). Opposite the $N^2$ffdG adduct, DinB-Kappa LS1A has misincorporation similar to WT DinB, with misincorporation of dTTP and a second dCTP observed, while the remaining variants
do not exhibit misincorporation. Incorporation opposite the $N^6$ffdA lesion is in general very low but the pattern of dTTP incorporation and dATP misincorporation is similar for all the proteins with LS4A and LS5A showing very low activity (Figures 3.3 and 3.4).

Figure 2. Both sets of DinBKappa loop swap chimeras have decreased activity with undamaged and both damaged DNA templates. Primer extension activity by WT DinB, DinBKappa LS1A-5A (A-C), DinBKappa LS1B-5B (D-F) and WT pol κ (all at 10 nM) on undamaged DNA (A/D), and templates containing minor groove adduct $N^2$ffdG (B/E), and major groove adduct $N^6$ffdA (C/F) at 0, 20, and 60 min time points.

DinBKappa loop swaps 1B-5B (sequences, Table 3.2) have no activity (at 10 nM) except for LS1B and LS5B on the template containing $N^2$ffdG (Figure 3.2 D-F). In misincorporation assays (Figure 3.3 D-F), even at a higher protein concentration (25 nM) there is incorporation only opposite the $N^2$ffdG adduct, with no activity observed on undamaged DNA or DNA containing $N^6$ffdA. There is weak incorporation only of dCTP by LS1B and LS2B, as well as
LS5B. Loop swaps 1B-5B start with the deletion site adjacent to the location of the nucleotide binding site, residues I41-A44, which is clearly detrimental to activity. Loop swap 5B contains one change, valine to serine to match the WT pol κ loop sequence (Table 3.1), which in the context of the shortened loop seems to confer a slight increase in activity.

Figure 3.3. DinBKappa loop swap chimeras have decreased misincorporation opposite undamaged DNA and with both damaged DNA templates. Representative misincorporation activity (of at least three replicates) by WT DinB, DinBKappa LS1A-5A (A-C), DinBKappa LS1B-5B (D-F), and WT pol κ (all at 25 nM) opposite undamaged dG (A/D), minor groove adduct N²fffdG (B/E), and major groove adduct N⁶fffdA (C/F) at 30 min time points. P indicates primer. The first reaction of each set is a control without added dNTPs.
Figure 3.4. DinBKappa loop swap chimeras have decreased misincorporation opposite undamaged DNA and with both damaged DNA templates. Percent nucleotide incorporation (representative data from at least three replicates) by WT DinB, DinBKappa LS1A-5A (A-C), and WT pol κ opposite undamaged dG (A), minor groove adduct $N^2$ffdG (B), and major groove adduct $N^6$ffdA (C). DinBKappa LS1B-5B had low activity on all templates and are omitted from this analysis.

KappaDinB loop swaps are more active than DinBKappa loop swaps. The pol κ active site loop is shorter than that of DinB by four residues (Figure 3.1), and therefore first the loop length was increased and then point mutations were made in the context of the longer loop. The
KappaDinB loop swaps generally retain robust activity on both undamaged DNA and DNA containing $N^2$ffdG, except LS6, which contains a change of a single amino acid from valine to isoleucine and is less active (Figure 3.5; sequences in Table 3.3). In general, the trend of relative activity levels of the KappaDinB loop swaps on DNA containing undamaged dG or $N^2$ffdG holds for DNA with $N^6$ffdA but with lower overall activity (Figure 3.5).

![Figure 3.5](image)

**Figure 3.5.** KappaDinB loop swap chimeras have similar activity as WT pol $\kappa$ on undamaged and both damaged DNA templates. Primer extension activity by WT pol $\kappa$, KappaDinB LS1-6, and WT DinB (all at 10 nM) on undamaged DNA (A), and templates containing minor groove adduct $N^2$ffdG (B), and major groove adduct $N^6$ffdA (C) at 0, 20, and 60 min time points.

Opposite undamaged dG, WT pol $\kappa$ inserts dCTP and weakly misinserts dTTP, whereas opposite $N^2$ffdG, pol $\kappa$ inserts dCTP and dTTP almost equally; in both cases DinB is more accurate, mainly inserting dCTP opposite dG or $N^2$ffdG. KappaDinB loop swaps show a decrease in dTTP misincorporation opposite undamaged dG and $N^2$ffdG (Figure 3.6). As the sequences of the KappaDinB loop swap proteins approach that of the DinB loop sequence, the pattern of nucleotide incorporation becomes more similar to that of DinB, with KappaDinB LS5 and LS6 having the lowest percentage of dTTP misincorporation opposite both dG and $N^2$ffdG (Figure 3.6). Misincorporation by the KappaDinB loop swaps opposite the $N^6$ffdA lesion is more similar to WT pol $\kappa$ than WT DinB as DinB misincorporates dATP to a greater extent, while WT
pol κ misincorporates dCTP more, although both proteins predominantly incorporate the correct dTTP (Figure 3.6).

Figure 3.6. KappaDinB loop swap chimeras have decreased misincorporation opposite undamaged DNA and the damaged DNA $N^2$fdfG template. Representative misincorporation activity (of at least three replicates) by WT pol κ, KappaDinB LS1-6, and WT DinB (all at 25 nM) opposite undamaged dG (A), minor groove adduct $N^2$fdfG (B), and major groove adduct $N^6$fdfA (C) at 30 min time points. P indicates primer. The first reaction of each set is a control without added dNTPs.

Given the decrease in misincorporation opposite $N^2$fdfG from KappaDinB LS5 to LS6, which involves only one mutation of Val to Ile, we decided to test the corresponding Val to Ile mutation in the context of WT pol κ. Thus, the mutation V130I was constructed in pol κ to assess the importance of the point mutation alone. V130 is found in the active site loop close to the
minor groove side of the incoming nucleotide. As seen in Figure 3.7 A-C, the primer extension activity of pol κ V130I is comparable to WT pol κ. V130I has similar misincorporation to WT pol κ, misincorporating dTTP opposite both the undamaged dG and the \( \text{N}^2\text{ffdG} \), unlike LS6, which exhibits less misincorporation than WT pol κ (Figure 3.7 D-E; Figure 3.6). Pol κ V130I also has similar nucleotide incorporation percentages to WT pol κ opposite \( \text{N}^6\text{ffdA} \) (Figure 3.7G). However, pol κ V130I shows higher primer extension activity with the DNA template containing \( \text{N}^6\text{ffdA} \); although V130 does not align precisely with DinB R35, mutation of which also relaxed discrimination against \( \text{N}^6\text{ffdA} \), these observations indicate the importance of this region of DinB and pol κ in modulating discrimination against major groove adducts (Walsh et al. 2013).
Figure 3.7. Pol κ V130I has similar activity as WT pol κ on undamaged and \(N^2\text{ffdG}\)
damaged DNA templates, but increased activity on DNA with \(N^6\text{ffdA}\). Primer extension
activity by pol κ V130I (at 10 nM) on undamaged DNA (A), and templates containing
minor groove adduct \(N^2\text{ffdG}\) (B), and major groove adduct \(N^6\text{ffdA}\) (C) at 0, 20, and 60 min
time points. Pol κ V130I has similar misincorporation activity as WT pol κ opposite
undamaged dG (D), minor groove adduct \(N^2\text{ffdG}\) (E), and major groove adduct \(N^6\text{ffdA}\) (F)
at 30 min time points. (G) Percent nucleotide incorporation for WT pol κ and V130I
(representative data from at least two replicates).

**KappaDinB loop swaps are stabilized more than DinBKappa loop swaps by preferred substrates.** We next assessed the overall stability of the protein chimeras, as we previously showed a specific stabilization of DinB and pol κ only in the presence of substrates with which the polymerases are active (Nevin et al. 2015a, Nevin et al. 2015b). We find that the melting temperatures of the KappaDinB loop swaps are higher than those of the DinBKappa loop swaps (Figure 3.8). Both WT DinB and pol κ, DinBKappa LS1A, and the KappaDinB loop swaps show an increase in melting temperature in the presence of undamaged DNA or \(N^2\text{ffdG}\)-containing...
DNA and an additional increase in melting temperature in the presence of these DNA constructs and the correct nucleotide. The additional increase in melting temperature in the presence of DNA due to the correct nucleotide is not observed with DinBKappa LS2A-5A. DinBKappa LS1B-5B do not show an appreciable increase in melting temperature with DNA containing dG or $N^2$ffdG or with correct incoming nucleotide dCTP. In the presence of undamaged DNA and the incorrect nucleotide dTTP there is generally a smaller but detectable increase in melting temperature for both WT DinB and pol κ, which both misincorporate dTTP opposite undamaged dG. DinBKappa LS1A and LS4A have the greatest increase in melting temperature in the presence of DNA of the DinB chimeras. The DinBKappa loop swaps that show little activity (Figure 3.2) also do not show an increase in melting temperature in the presence of the correct nucleotide or the incorrect nucleotide; however, WT DinB and DinBKappa LS1A have an increase in $T_m$ in the presence of the correct nucleotide dCTP and the incorrect nucleotide dTTP which both polymerases misincorporate (Figures 3.3, 3.8).

The KappaDinB loop swaps do not show as great an increase in melting temperature as WT DinB in the presence of undamaged DNA or $N^2$ffdG DNA but all have an increase in melting temperature in the presence of these DNAs and the correct nucleotide. Like WT DinB, WT pol κ has a modest increase in melting temperature in the presence of undamaged DNA and the incorrect nucleotide dTTP, which it also misincorporates, while the KappaDinB chimeras, which do not misincorporate dTTP opposite undamaged DNA, do not show an increase in melting temperature. V130I does not show an increase in melting temperature in the presence of undamaged DNA and the incorrect nucleotide dTTP even though it weakly misincorporates dTTP opposite undamaged dG.
In the presence of the major groove adduct, \( N^6 \text{ffdA} \), both the DinB chimeras and the pol κ chimeras have an increase in melting temperatures (Figure 3.8). However, in the few cases in which there is an increase in melting temperature in the presence of the major groove \( N^6 \text{ffdA} \) adduct and the correct or incorrect nucleotide, the increases are relatively modest, even for the proteins that show some bypass of the adduct.
Figure 3.8. KappaDinB loop swaps (C) are more stable than the DinBKappa loop swaps (A/B). The loop swaps have an increase in stability in the presence of the preferred DNA substrate and correct nucleotide that correlates with their activity. Values represent the average of at least three trials and error bars represent the standard deviation.
3.4 Discussion
In this work, we probed the active site loop region of *E. coli* DinB and human pol κ to better understand the importance of the loop in DNA damage specificity and nucleotide incorporation. The chimeras created had decreased misincorporation on all templates and overall the pol κ loop swap chimeras had more primer extension activity compared to the DinB loop swaps. There was no increase in TLS past the major groove $N^6$ffdA adduct for any loop swap chimera or increase in melting temperature in the presence of the $N^6$ffdA adduct and the correct or incorrect nucleotide. Intriguingly, pol κ V130I showed increased bypass of the non-preferred $N^6$ffdA lesion in the context of WT pol κ, although not in the context of the KappaDinB loop swap construct LS6.

As we have shown previously, there is a shift in melting temperature for both DinB and pol κ in the presence of undamaged DNA or other preferred DNA substrates and the correct incoming nucleotide but not with the incorrect incoming nucleotide (Nevin et al. 2015a, Nevin et al. 2015b). This stabilization indicates an active, correct ternary complex being formed, correlated with previous HDX-MS experiments that revealed protection in the fingers domain, located adjacent to the incoming nucleotide binding site (Nevin et al. 2015a, Nevin et al. 2015b). The loop swap proteins with low activity show stabilization in the presence of DNA but not in the presence of the correct nucleotide. This suggests that they can bind DNA but cannot form a productive ternary complex with the incoming nucleotide. The KappaDinB chimeras are all active and generally have 5-10 °C increases in melting temperature from the binary to the ternary complexes with undamaged and $N^2$ffdG DNA. Pol κ V130I also shows an increase with the correct nucleotide.

Previous HDX-MS studies of DinB and pol κ with minor groove adduct $N^2$ffdG and major groove adduct $O^6$-methylguanine showed similar levels of deuterium uptake as with an
undamaged DNA template, which was correlated with the formation of a binary complex. In the presence of $O^6$-methylguanine and the incorrect or correct nucleotide, there was no difference in deuterium uptake unlike ternary complexes formed with preferred DNA substrates. With the major groove adduct used in this work, $N^6$ffdA, we observed inhibition of polymerase activity as well as an increase in melting temperature with DNA but not with the correct or incorrect nucleotides. These changes in $T_m$, along with the previously reported HDX-MS results, indicate that DinB and pol κ can bind to major groove adducts forming a binary complex but cannot form an active ternary complex, an observation that correlates with the lack of activity in primer extension assays.

The single point mutation pol κ V130I resulted in overall similar activity and misincorporation relative to WT pol κ, unlike KappaDinB LS6, in which V130I is the only change within the context of the DinB loop and which had decreased activity and decreased misincorporation opposite dG and $N^2$ffdG. It is likely the combination of the changes in the active site loop result in the lower activity of LS6 relative to the effect of the V130I mutation alone (Figure 3.7, 3.9). In WT pol κ, V130 is near the minor groove of the DNA substrate and closer to the incoming nucleotide than to the template base; in the crystal structure V130 is pointing towards the rest of the fingers domain, away from the DNA (PDB 4U6P, Figure 3.9A) (Jha et al. 2016). The addition of the extra methyl group of the isoleucine could affect interactions with the neighboring residues within the fingers domain as the V130I has a slightly higher melting temperature than WT pol κ (42.2 °C and 38.6 °C, respectively).

Sholder et al., inserted the active site loop of DinB (residues 31-41) into Dpo4 with one additional mutation M76G and evaluated dCTP incorporation opposite the benzo[a]pyrene adduct. They found that dGTP misinsertion decreased approximately 10-fold; however, dCTP
incorporation increased 2-fold but was still ~5-fold lower than WT DinB (Sholder et al. 2015). It is believed that DinB has a larger minor groove opening than Dpo4 (G33/G74 versus V32/M76) to allow for correct Watson-Crick base pairs to form with minor groove adducts. The DinB loop contains V40 which is adjacent to the template base and plugs the major groove opening, preventing dGTP misinsertion, which was demonstrated by the A42V mutation in Dpo4 (Sholder et al. 2015). In our work, mutation of V40 in DinB to larger residues methionine and serine (Table 3.1 LS4A; Table 2 LS5B) in the context of different DinBKappa loop swaps resulted in less misincorporation compared to WT DinB. The larger residues substituted for DinB V40 could play similar roles in plugging the major groove as the changes made by Sholder et al.

The linker between the polymerase and little finger domains also plays a key role in polymerase activity, as swapping the linkers of Dpo4 and Dbh resulted in polymerase activity similar to that of the linker parental polymerase (Wilson et al. 2013). This was also true when assaying misincorporation activity of the chimeras; Dbh and chimeras containing the Dbh linker displayed the highest accuracy while Dpo4 and chimeras containing the Dpo4 linker misincorporated all three incorrect nucleotides opposite undamaged dG. Although these linkers and the active site loops studied here are in different positions relative to the active site, we found that nucleotide incorporation patterns of KappaDinB LS1-LS6 opposite undamaged dG and LS6 opposite $N_{2}^{2}$ffdG are similar to those of WT DinB. However, the DinB loop swaps containing the pol κ loops behave more similarly to WT DinB than WT pol κ.
Figure 3.9. (A) KappaDinB LS6 homology model (yellow, z-score of -0.577 (good)) structurally aligned with WT pol κ (light blue, PDB 4U6P) shows the additional methyl of the isoleucine (green) extending further into the fingers domain as compared to valine (medium blue). (B) DinBKappa LS5 homology model (yellow, z-score of -1.0 (satisfactory)) structurally aligned with WT DinB (green, PDB 4IR9) shows the little finger of the chimera tilted further away from the catalytic domain as the active site loop is smaller and contains uncharged residues serine and methionine. (C) Structural alignment of WT DinB (green, PDB 4IR9), WT pol κ (light blue, PDB 4U6P), and KappaDinB LS6 homology model (yellow) shows potential salt bridges between the inserted arginines in KappaDinB LS6 and the serine and glutamic acid (S423, E424) in pol κ that align with E251 and D252 of DinB.
The length of the active site loop is more significant in DinB than pol κ as seen in the primer extension assays that show low activity for both sets of DinB loop swaps while the KappaDinB loop swaps retain higher activity overall. Thus, pol κ is more tolerant than DinB to changes in the active site loop in terms of both activity and the pattern of nucleotide incorporation. Salt bridges appear to stabilize interactions between the active site loop residues R37 and R38 and little finger residues E251 and D252 in DinB and these charged arginine residues are replaced in the DinBKappa loop swaps with uncharged serines and methionines (Figure 3.9B). The serine at position 35 introduced in LS4A could interact with the phosphate backbone via hydrogen bonds and a methionine at position 36 is adjacent to the templating base. Mutation of this methionine to alanine in pol κ (M135A) results in a reduction in activity of pol κ as well as a decrease in correct dC incorporation opposite undamaged dG (Vasquez-Del Carpio et al. 2009). DinBKappa LS4A-5A both contain this methionine and have reduced activity but have decreased misincorporation compared to both WT proteins. Sholder et al., found that R37 and R38 are important for suppressing misincorporation opposite the large minor groove benzo[a]pyrene adduct. However, our loop swaps exhibit a decrease in misincorporation with the removal of the arginine residues. A homology model of DinBKappa LS5A shows that the active site loop is smaller than that of WT DinB and the little finger is tilted further away from the loop so it is likely there are fewer interactions between the little finger and loop region in the loop swaps (Figure 3.9B). This does not seem to affect the melting temperature of the protein alone, as the melting temperatures of most DinBKappa loop swap chimeras do not deviate substantially from WT DinB, except that both LS5A and LS5B have lower T_m values than WT DinB by ~5-7 °C. This could be due to poor folding of the protein without DNA present, as seen in the crystal structure of pol κ without DNA, in which the little finger is further from the remaining domains.
compared to structures with DNA bound (Uljon et al. 2004). There is currently no structure available of DinB without DNA.

The salt bridges in DinB between R37 and E251, and between R38 and D252 could be replicated in the KappaDinB loop swaps. In the pol κ structure the residues in the little finger that align with DinB E251 and D252 are S423 and E424 (Figure 3.9C). In the KappaDinB loop swap variants it is possible that the inserted arginines (R37 and R38 in DinB) interact with E424 in pol κ. The additional contacts could stabilize the protein or increase interactions between the protein and the DNA and thus increase activity. Liu et al., observed a decrease in activity opposite the large minor groove benzo[a]pyrene adduct when replacing three residues in the mouse pol κ active site loop with arginines along with a fourth mutation F170W (S131R, S133R, F170W, and D171R; which correspond to S132, S134, F171, and D172 in human pol κ) to reduce the gap size between the catalytic core and little finger domain as well as create charge interaction pairs with residues in the little finger (S422E, K447E; which correspond to S423 and K448 in human pol κ) (Liu et al. 2014). These mutations also decrease dCTP incorporation opposite the benzo[a]pyrene dG adduct by 24-fold compared to WT pol κ. Replacement of S131 to S133 with Dpo4 residues V32 to G41 (including the mutation R36K), as well as G153I to mimic Dpo4, led to a smaller decrease in benzo[a]pyrene bypass and only a seven-fold decrease in dCTP incorporation opposite benzo[a]pyrene. Our results when adding four amino acids (including three arginines, LS1-2 sequences in Table 3.3) to replicate the DinB loop in pol κ show that the larger active site loop and the positively charged arginine residues allow for high activity and a decrease in misincorporation on undamaged DNA or DNA with the minor groove adduct N2ffdG, with no change on DNA containing the major groove adduct N6ffdA.
Overall, pol κ activity is less affected by changes in its active site loop than DinB. This could be due to changes in the interactions between the fingers domain and the little finger domain. In DinB, making the loop smaller removes salt bridges whereas in pol κ, adding multiple charged residues likely increases the interactions with the little finger, which may explain the generally increasing accuracy of the KappaDinB loop swap variants.
3.5 References
DeCorte, B. L., D. Tsarouhtsis, S. Kuchimanchi, M. D. Cooper, P. Horton, C. M. Harris and T. M. Harris (1996). "Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N² position of deoxyguanosine." Chemical Research in Toxicology 9(3): 630-637.


Chapter 4: Human pol κ can tolerate a longer active site loop better than *E. coli* DinB

4.1 Introduction

DNA damage is ubiquitous and the DNA polymerases that bypass DNA damage are conserved in all domains of life and have different specificities (Friedberg 2006, Ohmori et al. 2001). The bypass of DNA damage is carried out by Y-family DNA polymerases through the process of translesion synthesis (TLS) (Radman 1975). DNA adducts can arise from a variety of sources including UV radiation, reactive oxygen species, and chemotherapeutics that intercalate or covalently modify DNA. Different Y-family polymerases can be employed to bypass different types of DNA damage; for example, *E. coli* DinB (DNA polymerase IV) can bypass minor groove adducts on the \(N^2\) position of deoxyguanine (Jarosz et al. 2006, Kottur et al. 2015, Kumari et al. 2008, Minko et al. 2008b, Seo et al. 2006, Shen et al. 2002) while *E. coli* DNA polymerase V is able to bypass UV photoproducts (Tang et al. 2000). Human pol κ is an ortholog of DinB and both pol κ and DinB have been shown to bypass \(N^2\)-dG adducts proficiently including \(N^2\)-furfuryl-dG (N\(^2\)fdG), \(N^2\)-ethyl-dG and the bulky benzo[a]pyrene \(N^2\)-dG adduct (Choi et al. 2006a, Choi et al. 2006d, Gowda et al. 2017, Hori et al. 2010, Minko et al. 2008a, Nair et al. 2015, Rechkoblit et al. 2002, Suzuki et al. 2002). Dpo4 from *Sulfolobus solfataricus* is also considered to be a DinB ortholog. Dpo4 can bypass small \(N^2\)-dG adducts; however its efficiency and fidelity decreases as the adduct size increases (Zhang et al. 2009).

It has been shown that Dpo4 is less blocked by major groove adducts as compared to DinB and pol κ (Basu et al. 2017, Choi et al. 2006b, Levine et al. 2001, Walsh et al. 2011a, Walsh et al. 2013, Xu et al. 2016). DinB and pol κ are both inhibited by the major groove adducts \(O^6\)-methylguanine, \(N^6\)-furfuryl-dA, and etheno-dA, with DinB more strongly inhibited
than pol κ (Nevin et al. 2015b, Walsh et al. 2013). It has been shown that Arg35 in DinB is important in the discrimination against the major groove adduct $N^6$-furfuryl-dA, as an Arg35Ala mutation in DinB showed an increase in primer extension activity opposite $N^6$-furfuryl-dA (Walsh et al. 2013). We have also shown that the active site loop harboring Arg35 plays a role in activity and nucleotide specificity (Walsh et al. 2013). We previously created loop swap chimeric proteins of DinB and pol κ, which showed pol κ is tolerant to a larger active site loop and has a decrease in misincorporation when the active site loop is swapped for that of DinB. While DinB was not tolerant of a smaller active site loop, the nucleotide misincorporation pattern of the chimeras did not change relative to that of WT DinB (Antczak et al. 2017). Pol κ has the smallest active site loop, the DinB loop is four amino acids longer, and Dpo4 has the longest loop, three amino acids longer than DinB and seven longer than pol κ. We hypothesized that creating loop swaps with Dpo4 will allow for increased bypass of major groove adducts by DinB and pol κ.

In this work we created loop swap chimeras between the DinB, pol κ, and Dpo4 to assess the ability of the chimeras to bypass major groove adducts and assess the nucleotide specificities of incorporation. We found that increasing the DinB loop length affects the activity of the protein overall, increasing the pol κ loop length by seven residues decreases activity modestly. Dpo4 loop swaps are currently being created.

4.2 Materials and Methods
Proteins and DNA. Pol κ, DinB, and Dpo4 chimeras were created using QuikChange site-directed mutagenesis kits (Agilent) and mutations were confirmed by DNA sequencing (Macrogen, Cambridge, MA; or Eton Bioscience, Charlestown, MA). The resulting amino acid sequences are shown in Tables 4.1-4.5. Pol κ (residues 19-526), DinB, and Dpo4 were expressed
and purified as described previously (Beuning et al. 2006, Fiala et al. 2004, Irimia et al. 2009, Wu et al. 2011) and all chimeras were purified using the scheme of the parental protein.

Table 4.1 DinBDpo4 loop swap protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT DinB</td>
<td>$^{28}$PIAIGGSRERRGVISTANY$^{45}$</td>
</tr>
<tr>
<td>DinBDpo4 Loop Swap 1</td>
<td>PVVVIAIGGSRERRGVISTANY</td>
</tr>
<tr>
<td>DinBDpo4 Loop Swap 2</td>
<td>PVVVCVFSGSRSERRGVISTANY</td>
</tr>
<tr>
<td>DinBDpo4 Loop Swap 3</td>
<td>PVVVCVFSGRFERRGVISTANY</td>
</tr>
<tr>
<td>DinBDpo4 Loop Swap 4</td>
<td>PVVVCVFSGRFEDSGAVATANY</td>
</tr>
<tr>
<td>WT Dpo4</td>
<td>$^{27}$PVVVCVFSGRFEDSGAVATANY$^{48}$</td>
</tr>
</tbody>
</table>

Table 4.2 Pol KappaDpo4 loop swap A protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Pol κ</td>
<td>$^{12}$PIAVGSMMSLSTSNY$^{41}$</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 1A</td>
<td>PIAVCVFSGSMSALSTSNY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 2A</td>
<td>PIAVCVFSGRFESMSALSTSNY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 3A</td>
<td>PVVVCVFSGRFESMSALSTSNY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 4A</td>
<td>PVVVCVFSGRFEDSGALSTSNY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 5A</td>
<td>PVVVCVFSGRFEDSGAVATANY</td>
</tr>
<tr>
<td>WT Dpo4</td>
<td>$^{27}$PVVVCVFSGRFEDSGAVATANY$^{48}$</td>
</tr>
</tbody>
</table>
### Table 4.3 Pol KappaDpo4 loop swap B protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Pol κ</td>
<td>127PIAVGSMSMLSTSNY141</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 1B</td>
<td>PIAVCVFSGSMSMLSTANY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 2B</td>
<td>PIAVCVFSGRFESMSLSTANY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 3B</td>
<td>PVVVCVFSGRFESMSLSTANY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 4B</td>
<td>PVVVCVFSGRFEDSGMLSTANY</td>
</tr>
<tr>
<td>KappaDpo4 Loop Swap 5B</td>
<td>PVVVCVFSGRFEDSGAVATANY</td>
</tr>
<tr>
<td>WT Dpo4</td>
<td>27PVVVCVFSGRFEDSGAVATANY48</td>
</tr>
</tbody>
</table>

### Table 4.4 Dpo4Kappa loop swap protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Dpo4</td>
<td>27PVVVCVFSGRFEDSGAVATANY48</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 1</td>
<td>PVVVC__SGRFEDSGAVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 2</td>
<td>PVVVC__EDSGAVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 3</td>
<td>PVVVC__SGAVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 4</td>
<td>PVVVCGS_AVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 5</td>
<td>PIAVCSGSAVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 6</td>
<td>PIAVGSMSAVATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 7</td>
<td>PIAVGSMSLATANY</td>
</tr>
<tr>
<td>Dpo4Kappa Loop Swap 8</td>
<td>PIAVGSMSLSTSNY</td>
</tr>
<tr>
<td>WT Pol κ</td>
<td>127PIAVGSMSMLSTSNY141</td>
</tr>
</tbody>
</table>
Table 4.5 Dpo4DinB loop swap protein sequences

<table>
<thead>
<tr>
<th>Protein</th>
<th>Loop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Dpo4</td>
<td>$^{27}$PVVVCVFSGRFEDSGAVATANY$^{48}$</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 1</td>
<td>PVVVC__SGRFEDSGAVATANY</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 2</td>
<td>PIAVCGRFEDSGAVATANY</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 3</td>
<td>PIAIGGRFEDSGAVATANY</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 4</td>
<td>PIAIGGSRERDGAVATANY</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 5</td>
<td>PIAIGGSRERRGAVATANY</td>
</tr>
<tr>
<td>Dpo4DinB Loop Swap 6</td>
<td>PIAIGGSRERRGVISTANY</td>
</tr>
<tr>
<td>WT DinB</td>
<td>$^{28}$PIAIGGSRERRGVISTANY$^{45}$</td>
</tr>
</tbody>
</table>

The DNA template, 61-mer (5’-CGTTACTCAGATCAGGCTGCAAGACCTGXCGTGCCGCTGACGCTGTACTATCAT ATGC-3’), Template18 (18-mer 5’-CCTGXGCGTCCGCAACG-3’), where X is dG, $N^2$-furfuryl-dG, or $N^6$-furfuryl-dA, and primers, (running start, 30-mer 5’-GCATATGATAGTACAGCTGCAGCCGGACGC-3’; standing start 31-mer 5’-GCATATGATAGTACAGCTGCAGCCGGACGCC-3’; dideoxy 13-mer (dd13) 5’-CGTTGCCGGACGC-3’, containing ddC at the 3’-end) were obtained from Eurofins Operon or Integrated DNA Technologies. The 61-mer or 18-mer DNA templates containing a single $N^2$-furfuryl-dG (X) were prepared as previously described (DeCorte et al. 1996, Jarosz et al. 2006). The 61-mer and 18-mer DNA templates containing a single $N^6$-furfuryl-dA (X) were prepared as previously described from $O^6$-phenyl-deoxyinosine (Glen Research) (Larson et al. 1992). DNA was purified by denaturing polyacrylamide gel electrophoresis and the crush and soak method (Sambrook et al. 1989). DNA primers (running start and standing start) were end labeled with $^{32}$P as previously described (Beuning et al. 2006, Sambrook et al. 1989).
**Primer extension and misincorporation assay.** DNA template (where X is unmodified dG, $N^2$ffdG, or $N^6$ffdA) was combined with $^{32}$P-labeled primer (running start for primer extension assays or standing start for misincorporation assays) in a 1:1 ratio (500 nM) and annealed with annealing buffer (20 mM HEPES, pH 7.5, and 5 mM Mg(OAc)$_2$) by heating for 2 min at 95 °C, incubating at 50 °C for 60 min and cooling to 37 °C. Reactions were carried out in 1X reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO$_4$, 2 mM β-mercaptoethanol, and 0.1 mg/mL bovine serum albumin) with 100 nM $^{32}$P labeled primer/template, 10 nM protein, and 500 μM dNTPs. For misincorporation assays, the final concentrations in 1X reaction buffer were 100 nM $^{32}$P labeled primer/template, 25 nM protein, and 1 mM dATP, dCTP, dGTP, or dTTP. Products were separated by denaturing 8 M urea 16% polyacrylamide gels and imaged on a Molecular Dynamics Storm 860 phosphorimager. Data were analyzed using ImageQuant TL1D Gel Analysis software (GE Healthcare) to determine percent nucleotide incorporation. Gels are representative of at least two replicates.

**Thermal Shift Assays.** Thermal shift assays were conducted using a Bio-Rad CFX as previously described (Antczak et al. 2017, Nevin et al. 2015a, Nevin et al. 2015b) using T18 undamaged DNA or T18 with $N^2$ffdG or $N^6$ffdA modifications containing 3-5 μM each wild-type protein or chimera, 5-10 μM dd13 primer/T18 template DNA, 1 mM dNTP (dCTP or dTTP) in 1X reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO$_4$, and 2 mM β-mercaptoethanol). Reactions were incubated at room temperature for 20 min before addition of Sypro Orange (Invitrogen) to a final concentration of 20x. Values represent the average of at least three trials and error bars represent the standard deviation.
Homology models of the DinB, pol κ, and Dpo4 chimeras. The homology models were created as previously described (Antczak et al. 2017), the FASTA sequences used for each protein were obtained from Uniprot, Q9UBT6 for pol κ (residues 19-526), Q97W02 for Dpo4, and A7ZHZ2 for DinB. Residues 19-526 were used for pol κ as our laboratory uses the gene for the truncated protein and many crystal structures of pol κ are of this catalytic core. The loop regions of each protein were then swapped in the FASTA files to create the homology models. In the FASTA sequence of DinB, the sequence PIAIGGRERRGVRISVTANYPARK was replaced with the corresponding Dpo4 loop sequence, PVVVCVFSGRFEDGAVTANYEAR, to create DinB with a Dpo4 loop (DinB-Dpo4). This was repeated for pol κ by replacing the sequence, PIAVGSMMSMLSTSNYHARR, with the corresponding Dpo4 loop sequence. And finally, this was repeated twice with Dpo4 to create Dpo4-DinB and Dpo4-Kappa chimeras. The homology models were created using YASARA (Krieger et al. 2002) and were evaluated using a Ramachandran plot and Procheck via PDBsum. The Z score obtained from the homology modeling in YASARA is a measure of model quality calculated from the backbone dihedrals, side chain dihedrals, and packing interactions of the model, using the mean values and standard deviations of the corresponding scores for a set of high quality crystal structures.

4.3 Results

DinBDpo4 loop swaps have decreased activity overall. To test the roles of the active site loops in the activity and specificity of DinB, pol κ, and Dpo4, we constructed loop swap chimeras by making stepwise changes in the respective loops. The addition of three residues to the DinB loop to make it the same length as the Dpo4 loop (DinBDpo4 LS1, sequences Table 4.1) dramatically decreases the activity, whereas the activity stays the same for the remaining three loop swaps,
which each involve multiple point mutations (Figure 4.1). The DinBDpo4 loop swaps are more active on DNA containing the $N^2$ffdG lesion as compared to undamaged DNA, as seen with WT DinB (Jarosz et al. 2006). There is no increase in activity on the DNA containing $N^6$ffdA as hypothesized. The DinBDpo4 loop swaps retain low misincorporation opposite undamaged dG, only inserting dCTP, like WT DinB, and unlike Dpo4 which misincorporates dATP, dGTP, and dTTP. DinBDpo4 LS2 and LS3 have increased dTTP misincorporation opposite $N^2$ffdG while LS1 and LS4 have no dTTP misincorporation.
Figure 4.1 DinBDpo4 loop swaps show low activity on all DNA templates, (A) undamaged DNA, (B) $N^2$ffdG DNA, and (C) $N^6$ffdA compared to both WT proteins, DinB and Dpo4. The WT DinB and loop swaps have an increase in activity on the $N^2$ffdG containing template (B). The loop swaps retain incorporation patterns similar to that of WT DinB.

KappaDpo4 loop swaps have a decrease in activity and a decrease in misincorporation on all templates. KappaDpo4 LS 1-2A, 5A have very little activity and LS3-4A have no activity, even after 60 min (Figure 4.2 A-C). The KappaDpo4 LS A series begin with the insertion of four amino acids as well as the Met135Ala mutation; in order to determine if the mutation at position 135 was problematic, the same insertion was made starting with Ser139Ala as this mutation has similar activity to WT pol κ (Figure 2.5; Table 4.2-4.3). KappaDpo4 LS 1-5B have decreased
activity compared to WT pol κ (Figure 4.2 D-F) but they are more active than the previous set of LS 1-5A chimeras (Figure 4.2 A-C).

Figure 4.2 KappaDpo4 loop swap 1-5A (A-C) show very low to no activity on all DNA templates, (A/D) undamaged DNA, (B/E) N²ffdG DNA, and (C/F) N⁶ffdA compared to both WT proteins, pol κ and Dpo4. KappaDpo4 loop swaps 1-5B (D-F) show a decrease in activity but not as dramatic as the LS A chimeras.

Both DinBDpo4 and KappaDpo4 loop swaps have an increased melting temperature in the presence of the preferred DNA substrate and the correct nucleotide. As we have shown previously (Chapter 3; Nevin et al. 2015b), there is an increase in the melting temperature of
WT pol κ, WT DinB, and variants in the presence of preferred DNA and the correct nucleotide. The least active KappaDpo4 LS 3B melting temperature does not show the same increase as WT pol κ in the presence of DNA and the correct nucleotide (WT increases 10 °C as compared to protein alone, KappaDpo4 LS 3B increases 2 °C).

Figure 4.3. DinBDpo4 LS and KappaDpo4 LS 1-5B have an increase in melting temperature in the presence of undamaged DNA and the correct nucleotide. (A) DinBDpo4 LS and (B) KappaDpo4 LS B melting temperatures in the absence and presence of undamaged DNA and the correct or incorrect nucleotide.

4.4 Discussion

As shown in Chapter 3, pol κ is more tolerant than DinB to changes in the active site loop region. In this case both proteins had amino acids added to the active site loops (pol κ seven
amino acids, DinB three amino acids) to make them more similar to Dpo4 and even with seven additional residues pol κ retained more activity for most chimeras as compared to DinB. Dpo4DinB and Dpo4Kappa loop swaps are currently being created in order to express, purify, and assay the effects of the smaller DinB and pol κ loops on Dpo4.

DinBDpo4 loop swaps have a decrease in activity starting with LS 1, which is the insertion of three amino acids (Table 4.1, Figure 4.1). This is similar to previous results when residues were deleted to make the DinB loop into the pol κ loop (Chapter 3) (Antczak et al. 2017). The residues that were inserted were further from the DNA and incoming nucleotide in the DinBDpo4 chimera as compared to the residues deleted to create the pol κ loop within DinB and yet there is still a dramatic decrease in activity suggesting the active site loop length is important for DinB activity. The DinBDpo4 chimeras are still able to create binary and ternary complexes with undamaged DNA (Figure 4.3) as shown in the increase in melting temperature (Antczak et al. 2017, Nevin et al. 2015b).

KappaDpo4 loop swaps B retain more activity as compared to DinBDpo4 loop swaps. The second set of KappaDpo4 loop swaps, set B, has a small decrease in activity upon insertion of seven residues but can still reach the end of the DNA templates (Figure 4.2) when adding the residues to the Ser139Ala variant of pol κ (characterized in Chapter 2). KappaDpo4 loop swaps 1-2A had very low activity with the remaining LS 3-5A having no detectable activity; however this set of chimeras began with pol κ Met135Ala, which, as shown in Chapter 2, already has a decrease in activity as compared to WT pol κ. Like we have shown previously with the KappaDinB loop swaps (Chapter 3, (Antczak et al. 2017)) the KappaDpo4 loop swap B are not as strongly affected by the increasing active site loop length.
Based on the previous work in Chapter 3 (Antczak et al. 2017) and the work done thus far with DinB and pol κ it is possible that shortening the Dpo4 active site loop could have different results depending which loop was engineered in to it. In the case of the Dpo4DinB loop swaps, it could be postulated that with the introduction of Arg37 and Arg38 (sequences in Table 4.5) there would not be an increase in the interaction with the little finger similar to that seen with the KappaDinB loop swaps (Chapter 3, (Antczak et al. 2017)). The Dpo4 residues that align with DinB Glu251 and Asp252 are Lys252 and Arg253, which would likely cause repulsion with the newly introduced positive charged arginines in the active site loop (Figure 4.4). This repulsion could result in chimeras with low activity, as with the shorter active site loop the Asn254 would likely not be close enough to interact with any of the arginines and the positively charged residues would clash. The interactions could result in an unfolded or partially unfolded protein. It is possible that this would affect the melting temperature as well if the polymerase cannot fold completely.

Figure 4.4. The three introduced arginines in Dpo4DinB loop swap could decrease interactions with the little finger residues in Dpo4 due to electrostatic repulsion. Structural alignment of WT DinB (green, PDB 4IRC), WT Dpo4 (purple, PDB 3QZ7), and Dpo4DinB loop swap homology model (yellow) highlighting the residues in the little finger that could interact with the newly introduced arginines of the active site loop.
As for the Dpo4Kappa loop swaps, the chimeras may have activity similar to the DinBKappa loop swaps in which the first loop swap with a deletion of two amino acids will be still active but the remaining loop swaps with five to seven amino acids deleted will either have very low activity or be completely inactive. This could be due the active site loop and little finger residues unable to interact. It is possible that since pol κ has the N-clasp which increases contacts to the DNA, a smaller active site loop is sufficient. It has been shown that the N-clasp is important for pol κ catalytic activity (Jia et al. 2008, Lior-Hoffmann et al. 2014, Lone et al. 2007, Uljon et al. 2004) and could potentially be important for the activity of the Dpo4Kappa loop swaps and the DinBKappa loop swaps.

Overall increasing the active site loop size of both DinB and pol κ result in decreases in activity opposite all three DNA templates. The least active chimeras have negligible increases in melting temperatures in the presence of DNA and the correct incoming nucleotide. More work needs to be completed with misincorporation assays of the KappaDpo4 loop swaps, as well as assaying the Dpo4DinB and Dpo4Kappa loop swaps to make further conclusions.
4.5 References


DeCorte, B. L., D. Tsarouhtsis, S. Kuchimanchi, M. D. Cooper, P. Horton, C. M. Harris and T. M. Harris (1996). "Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N2 position of deoxyguanosine." Chemical Research in Toxicology 9(3): 630-637.


Chapter 5: *In vitro* biochemical characterization of human pol κ SNPs

5.1 Introduction

DNA damage is a constant threat, from both endogenous and exogenous sources. DNA adducts can lead to mutations and in turn to disease and cell death. Specialized DNA polymerases (pols) are required to replicate past DNA adducts in a damage tolerance process called translesion synthesis (Radman 1975). These specialized polymerases are Y-family DNA pols and are conserved in all domains of life. Y-family DNA pols are characterized by low fidelity on undamaged DNA and the ability to bypass DNA adducts (Ohmori et al. 2001). Y-family DNA pols bypass specific lesions and have different accuracies and efficiencies of bypass depending on the adduct. Pol κ is one of four Y-family polymerases in humans. Pol κ is able to bypass $N_2$-dG adducts such as $N_2$-furfuryl-dG and $N_2$-(1-carboxyethyl)-dG as well as larger, bulkier adducts like $N_2$-benzo[a]pyrene diol epoxide-dG (Choi et al. 2006a, Jarosz et al. 2006, Rechkoblit et al. 2002, Yuan et al. 2008a, Zhang et al. 2002) and has been shown to be blocked by major groove adducts (Choi et al. 2006b, Choi et al. 2008, Nevin et al. 2015b, Walsh et al. 2013). Pol κ has been shown to extend from mispaired primer termini on both undamaged DNA and DNA containing 8-oxo-7,8-dihydroguanine and $O^6$-methyl-dG as well as from a dG:dT mismatch at the 3’ end of a thymine-thymine dimer (Haracska et al. 2002, Lone et al. 2007, Niimi et al. 2009, Vasquez-Del Carpio et al. 2011, Washington et al. 2002, Wolfle et al. 2003).

Single nucleotide polymorphisms (SNPs) are differences in base pair positions in genomic DNA between individuals (Brookes 1999). The different base in the DNA could be translated to a mutated protein or cause a truncated protein to be translated. Mutated or truncated proteins could cause disease or cell death. There is no human disease yet linked to a mutation in
pol κ (Kim et al. 2016). However, Dai, et al., found two noncoding POLK SNPs that were associated with breast cancer risk (Dai et al. 2016). Shao, et al., found several noncoding POLK SNPs linked to platinum chemotherapy toxicity and response as well as progression-free survival in non-small cell lung cancer patients (Shao et al. 2014). Several POLK SNPs in the dbSNP have been analyzed with in silico prediction tools to determine those which would affect pol κ activity and biochemically characterized (Kim et al. 2016, Song et al. 2014). Variants Leu21Phe, Ile39Thr, Asp189Gly, Arg219Ile, Glu419Gly, Ser423Arg, and Tyr432Ser (in a construct containing pol κ residues 1-526) were biochemically characterized using undamaged DNA and DNA containing the minor groove adduct $^{N2}$-CH$_2$(9-anthracenyl)G (Song et al. 2014). On undamaged DNA, WT pol κ, and the Leu21Phe, Ile39Thr, Asp189Gly, Arg219Ile, and Ser423Arg variants were all able to extend to the end of the template, while Glu419Gly and Tyr432Ser variants had much less activity than WT pol κ. WT pol κ is able to reach the end of a template containing $^{N2}$-CH$_2$(9-anthracenyl)G as are Arg219Ile and Ser423Arg variants but not as well as with the undamaged template. Leu21Phe, Ile39Thr, and Asp189Gly variants were much less active (Song et al. 2014). In the second study, Glu29Lys, Thr44Met, Phe192Cys, Glu292Lys, Arg298His, Ala471Val, Thr473Ala, and Arg512Trp variants were characterized on DNA templates containing $^{N2}$-CH$_2$(9-anthracenyl)G, 8-oxo-7,8-dihydroguanine, $^{O6}$-methyl-G, or an abasic site (Kim et al. 2016). The SNP variants Glu29Lys, Thr44Met, Phe192Cys, Glu292Lys, and Ala471Val had similar activity as WT pol κ on undamaged DNA while Arg298His, Thr473Ala, and Arg512Trp variants had much lower activity that WT. The same trend was seen with $^{N2}$-CH$_2$(9-anthracenyl)G, 8-oxoG, $^{O6}$-methyl-G, or DNA containing an abasic site, except Thr44Met had a decrease in activity opposite every lesion except the 8-oxoG (Kim et al. 2016).
In this work *POLK* SNPs Leu21Phe, Ile39Thr, Pro169Thr, Phe192Cys, Arg219Ile, Glu292Lys, Arg298His, Ser423Arg, and Tyr432Ser were biochemically characterized using undamaged DNA, or DNA containing $N^2$-furfuryl-dG, $N^6$-furfuryl-dA, and etheno-dA for both insertion and extension activity. Leu21Phe, Ile39Thr, Arg219Ile, Ser423Arg, and Tyr432Ser were found in cancer tumors and determined to be statistically significant. The remaining SNPs were found in dbSNP (and the other five are also found in this database). It was found that the SNPs could be grouped into three different categories, those with activity similar to WT (Leu21Phe, Ile39Thr, Pro169Thr, Arg219Ile and Glu292Lys), those that are more active than WT (Phe192Cys, Ser423Arg), and those that are less active than WT (Arg298His, Tyr432Ser).

### 5.2 Materials and Methods

#### Proteins and DNA.

Pol κ was expressed and purified as described previously (Irimia et al. 2009). Pol κ SNP variants, listed in Table 5.1, were created by using Quikchange site-directed mutagenesis kits (Agilent). Mutations were confirmed by DNA sequencing (Macrogen, Cambridge, MA; or Eton Bioscience, Charlestown, MA). The DNA template containing a single $N^2$-furfuryl-dG was prepared as previously described (DeCorte et al. 1996, Jarosz et al. 2006). The DNA template containing a single $N^6$-furfuryl-dA was prepared as previously described from $O^6$-phenyl-deoxyinosine (Glen Research)(Larson et al. 1992). The unmodified DNA and etheno-dA templates and primers (running start, standing start, MatchC, MatchG, and MatchT) were from Eurofins Operon. The DNA sequences for templates and primers are in Table 5.2. DNA was purified by denaturing polyacrylamide gel electrophoresis and the crush and soak method (Sambrook et al. 1989). DNA primers (running start, standing start and MatchC/G/T) were end labeled with $^{32}$P as previously described (Beuning et al. 2006, Sambrook et al. 1989).
Table 5.1. *POLK* Single Nucleotide Polymorphisms studied

<table>
<thead>
<tr>
<th>ID</th>
<th>Amino Acid Mutation</th>
<th>Domain</th>
<th>Tumor Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3104729</td>
<td>Leu21Phe</td>
<td>N-clasp</td>
<td>Prostate</td>
</tr>
<tr>
<td>rs3094258</td>
<td>Ile39Thr</td>
<td>N-clasp</td>
<td>Prostate, Melanoma</td>
</tr>
<tr>
<td>rs148385845</td>
<td>Pro169Thr</td>
<td>Fingers</td>
<td>Lung*</td>
</tr>
<tr>
<td>rs150515841</td>
<td>Phe192Cys</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>rs3104717</td>
<td>Arg219Ile</td>
<td>Palm</td>
<td>Prostate</td>
</tr>
<tr>
<td>rs142203892</td>
<td>Glu292Lys</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>rs151251843</td>
<td>Arg298His</td>
<td>Palm</td>
<td>Large intestine*</td>
</tr>
<tr>
<td>rs35257416</td>
<td>Ser423Arg</td>
<td>Palm</td>
<td>Melanoma</td>
</tr>
<tr>
<td>rs77612491</td>
<td>Tyr432Ser</td>
<td>Palm</td>
<td>Melanoma</td>
</tr>
</tbody>
</table>

*1000 genomes

Figure 5.1. Crystal structure of pol κ (PDB 4U6P (Jha et al. 2016)) with the residues of interest highlighted in dark blue (except Leu21 which is not in the crystal structure). In grey is the DNA and in red is the incoming nucleotide. Distances from the incoming nucleotide are in Table 5.2.
Table 5.2. Distance from SNP residues to incoming nucleotide

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Domain</th>
<th>Distance from incoming nucleotide (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile39</td>
<td>N-clasp</td>
<td>22.3</td>
</tr>
<tr>
<td>Pro169</td>
<td>Fingers</td>
<td>14.2</td>
</tr>
<tr>
<td>Phe192</td>
<td>Palm</td>
<td>19.7</td>
</tr>
<tr>
<td>Arg219</td>
<td>Palm</td>
<td>31.1</td>
</tr>
<tr>
<td>Glu292</td>
<td>Palm</td>
<td>25.6</td>
</tr>
<tr>
<td>Arg298</td>
<td>Palm</td>
<td>18.2</td>
</tr>
<tr>
<td>Ser423</td>
<td>Little finger</td>
<td>27.2</td>
</tr>
<tr>
<td>Tyr432</td>
<td>Little finger</td>
<td>35.6</td>
</tr>
</tbody>
</table>

*Distance measured from Cα to αPO₄. Leu21 not measured as it is not in the crystal structure (PDB 4U6P).

**Primer Extension Assays.** DNA template (unmodified, N²ffdG, or N⁶ffdA) was combined with ³²P-labeled primer (running start, standing start, MatchC, MatchG or MatchT) in a 1:1 ratio (500 nM) and annealed with annealing buffer (20 mM HEPES (pH 7.5) and 5 mM Mg(OAc)₂) by heating for 2 minutes at 95 °C, incubating at 50 °C for 60 minutes and cooling to 37 °C. Reactions were carried out in 1X reaction buffer (30 mM HEPES (pH 7.5), 20 mM NaCl, 7.5 mM MgSO₄, 2 mM β-mercaptoethanol, and 1% bovine serum albumin), 100 nM ³²P labeled primer/template and 5 nM protein (25 nM protein was used for etheno-dA containing templates) and 500 μM dNTPs. For misincorporation assays, the final concentrations in 1X reaction buffer used were 100 nM ³²P labeled primer/template and 5 nM protein and 1 mM dATP, dCTP, dGTP or dTTP. For etheno-dA-containing templates, 25 nM protein (wild-type pol κ and variants) was used. Gels are representative of at least three replicates.
**Table 5.3.** DNA primer and template sequences

<table>
<thead>
<tr>
<th>DNA</th>
<th>Length</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running</td>
<td>30 mer</td>
<td>5'-GCATATGATAGTACAGCTGCAGCCGGACGC-3'</td>
</tr>
<tr>
<td>Start</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing</td>
<td>31 mer</td>
<td>5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3'</td>
</tr>
<tr>
<td>Start</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MatchC</td>
<td>32 mer</td>
<td>5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3'</td>
</tr>
<tr>
<td>MatchT</td>
<td>32 mer</td>
<td>5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3'</td>
</tr>
<tr>
<td>MatchG</td>
<td>32 mer</td>
<td>5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3'</td>
</tr>
<tr>
<td>P13</td>
<td>13 mer</td>
<td>5'-GCTTGGCCGGACG-3'</td>
</tr>
<tr>
<td>18 mer*</td>
<td></td>
<td>5'-CCTAXGCGTCCGGCAAG-3</td>
</tr>
<tr>
<td>61 mer*</td>
<td></td>
<td>5'-CGTTACTCAGATCAGGCTGCAAGTACXGGGCTCCGGCTGACGTACTATCATATGC</td>
</tr>
</tbody>
</table>

* X = A, G, N²ffdG, N⁶ffdA, Etheno-dA

**Thermal Shift Assays.** Thermal shift assays were conducted using Primer13 and undamaged as well as N²ffdG, N⁶ffdA, and etheno-dA Template18 DNA (sequences in Table 5.2). DNA was annealed in the same manner at stated for the primer extension assays except at a final concentration of 75 μM. Protein, at a final concentration of 5 μM, was combined in the presence or absence of DNA (at a final concentration of 10 μM), and dNTP (dCTP or dTTP at 1 mM) in 15 μL of 1X assay buffer (30 mM HEPES (pH 7.5), 20 mM NaCl, 7.5 mM MgSO₄, 2 mM β-mercaptoethanol). The reactions were incubated for 20 min at room temperature before Sypro Orange (20X, Life Technologies) was added. The fluorescence was detected as the temperature was increased from 10 °C to 90 °C in a Bio-Rad CFX real-time PCR instrument. Values represent the average of at least three trials and error bars represent the standard deviation.

**5.3 Results**

*Some SNPs are as active as or more active than WT pol κ.* Using running start primer, we assayed the insertion and extension ability of WT pol κ and SNPs on undamaged (dG and dA), N²ffdG, N⁶ffdA, and etheno-dA DNA templates. We have previously shown that on the N⁶ffdA
template WT pol κ is not as active as it is on an undamaged template (Antczak et al. 2017, Walsh et al. 2013). SNPs encoding pol κ variants Leu21Phe, Ile39Thr, Pro169Thr, Phe192Cys, Glu292Lys, and Ser423Arg have similar activity as WT on the undamaged template (Figure 5.2). The remaining variants, Arg219Ile, Arg298His, and Tyr432Ser, have the least activity. A similar trend is observed for both furfuryl damaged templates, $N^2$ffdG and $N^6$ffdA. With both major groove adducts, $N^6$ffdA and etheno-dA, the pol κ and its variants are able to insert one nucleotide before the lesion very well but cannot fully extend the template, unlike the activity with minor groove $N^2$ffdG or undamaged templates (Figure 5.2). On etheno-dA containing DNA we see there is a large amount of insertion before the lesion but very little full extension of the templates except with WT, Leu21Phe, Ile39Thr, Pro169Thr, Phe192Cys, Glu292Lys, Ser423Arg, and Tyr432Ser which are able to fully extend the template to different extents. The remaining SNPs, Arg219Ile and Arg298His, are able to insert before the lesion but are unable to insert opposite the lesion and extend to the end of the template.
Figure 5.2. WT Pol κ and SNPs insertion and extension past undamaged DNA or a lesion (A) undamaged dG DNA, (B) $N^2\text{-}furfuryl\text{-}dG$ DNA, (C) Undamaged dA DNA, (D) $N^6\text{-}furfuryl\text{-}dA$ DNA, and (E) etheno-dA DNA (using running start primer).

SNPs can extend from damaged template with the correct base pairing We assayed the extension ability of WT pol κ and SNPs on damaged templates with a primer that creates the correct Watson-Crick base pair opposite the damage (dC opposite $N^2\text{-}furfuryl\text{-}dG$ and dT opposite $N^6\text{-}furfuryl\text{-}dA$ and etheno-dA). There is an increase in extension from base pairs containing a damaged template base compared to insertion and extension using running start primer (Figure 5.3). Pol κ Phe192Cys and Ser423Arg show an increase in extension ability from all three damaged base pairs as compared to WT pol κ. Etheno-dA remains a blocking lesion for pol κ as an extender since there is still much less full extension from the correct etheno-dA:dT base pair. However, Leu21Phe shows a large increase in extension from the etheno-dA:dT mismatched base pair relative to pol κ and the other variants.
Figure 5.3. Pol κ WT and SNPs extension from a correct base pair with undamaged DNA or a lesion (A) undamaged dG DNA (MatchC primer), (B) N²ffdG DNA (MatchC primer), (C) undamaged dA DNA (MatchT primer), (D) N⁶ffdA DNA (MatchT primer), and (E) etheno-dA DNA (MatchT primer).

SNPs can extend from mismatched termini from damaged templates. We assayed the ability of the WT pol κ and the variants to extend primers containing mismatches at the damaged template base (N²-ffdG:T, N⁶ffdA:G, or etheno-dA:G and undamaged controls) (Figure 5.4). It has been shown previously that WT pol κ can extend from G:T primer termini mismatches (Haracska et al. 2002, Vasquez-Del Carpio et al. 2011, Washington et al. 2002, Wolfle et al. 2003). There is slightly more extension from the N²-furfuryl:dT mismatch than the undamaged dG:dT mismatch (Figure 5.4). Pol κ Leu21Phe, Phe192Cys, and Ser423Arg show more extension than WT from the N²-furfuryl:dT mismatch. The two proteins that have weakest primer extension activity with running start primer, Arg298His and Tyr432Ser, have no visible extension from the dG:dT mismatch on either template, undamaged dG or N²-furfuryl-dG. As
with undamaged dG:dT, there is little extension activity from the A:G mismatch. Pol κ and variants do not show as much extension activity on DNA templates containing the $N^6$-furfuryl-dA major groove adduct as compared to the increase in activity with the minor groove $N^2$-furfuryl adduct. There is a much clearer increase in extension activity with an etheno-dA:dG mismatch at the primer terminus. Pol κ Ser423Arg have more extension activity in general compared to WT pol κ and compared to the ability of WT pol κ to extend an undamaged A:G mismatched primer terminus.

Figure 5.4. Pol κ WT and SNPs extension from a mismatched basepair with undamaged DNA or a lesion (A) undamaged DNA (MatchT primer), (B) $N^2$ffdG DNA (MatchT primer), (C) TemplateA DNA (MatchG primer), (D) $N^6$ffdA DNA (MatchG primer), and (E) etheno-dA DNA (MatchG primer).

SNPs have similar nucleotide incorporation to WT pol κ when inserting opposite a lesion

On an undamaged template G, WT pol κ incorporates dC, and the only variant to have dT misincorporation is Ser423Arg (Figure 5.5). However, when inserting opposite $N^2$ffdG, WT and
most of the pol κ SNP variants correctly incorporate dC and also misincorporate dT, with the exception of Ile39Thr, Arg298His, and Tyr432Ser, which only incorporate dC. With an undamaged A template base, WT correctly inserts dT as well as misinserts dC, as do the majority of the SNP variants except Arg298His, and Tyr432Ser which only insert dT. In contrast, with the major groove adducts N6ffdA and etheno-dA, WT pol κ and the variants correctly incorporate dT opposite only N6ffdA. With etheno-dA is the template base, higher concentrations (25 nM) of WT pol κ and several SNPs, Leu21Phe, Ile39Thr, Phe192Cys, Arg219Ile, Glu292Lys, and Ser423Arg, incorporate dT and misincorporate dG while the remaining variants, Pro169Thr, Arg298His, Tyr432Ser, only incorporate dT and we see very little incorporation after 30 minutes by both Arg298His and Tyr432Ser (25 nM protein).
Figure 5.5. Pol κ WT and SNPs misincorporation opposite a lesion show an increase in misincorporation opposite the $N^2$ffdG lesion and the etheno-dA lesion. (A) Undamaged dG DNA, (B) $N^2$-furfuryl-dG DNA, (C) TemplateA DNA, (D) $N^6$ffdA DNA, and (E) etheno-dA DNA (using standing start primer).

**Insertion after a lesion from the correct base pair** The DNA template used here harbors a T as the template base after the lesion. WT and all the SNP variants insert dA correctly after the undamaged dG:dC base pair. WT and the SNP variants also insert only the correct dA after a base pair containing the minor groove lesion $N^2$ffdG (Figure 5.6). When extending from the correct undamaged dA:dT base pair we see that all proteins, except the least active two Arg298His and Tyr432Ser, incorporate A but also misincorporate what appears to be three guanines. The second and third dG residues would be correctly incorporated opposite the two dC residues in the template following T (Table 5.2). A second dA is also misincorporated by Ser423Arg (Figure 5.6). Surprisingly, extension from the major groove adduct $N^6$ffdA:dT base
pair results in only correct incorporation by all proteins (Figure 5.6). Extension from the major
groove adduct etheno-dA:dT basepair yielded low levels of correct dA incorporation by WT,
Arg219Ile, Glu292Lys, and Ser423Arg, while the remaining variants showed no nucleotide
incorporation after 30 min, which is consistent with the very low primer extension activity
(Figure 5.2).

Figure 5.6. Pol κ WT and SNPs misincorporation after a lesion show correct dA
incorporation after lesions with the correct base pairing. (A) Undamaged dG DNA, (B)
N²ffdG DNA (using MatchC primer for undamaged dG and N²ffdG), (C) TemplateA DNA,
(D) N⁶ffdA DNA, and (E) etheno-dA DNA (using MatchT primer for TemplateA, N⁶ffdA,
and ethenod-dA).

Insertion after a lesion from a mismatch. In general, all of the proteins studied here exhibit
low activity extending from a dG:dT mismatch compared to an etheno-dA:dG mismatch (Figure
5.4). With undamaged dG:dT mismatch there is only correct dT incorporation seen by WT,
Leu21Phe, Pro169Thr, and Ser423Arg. A similar pattern is seen from the $N^2$ffdG:dT mismatch as only dT incorporation is seen by WT, Leu21Phe, Ile39Thr, Pro169Thr, Phe192Cys, Arg219Ile, Glu292Lys, and Ser423Arg. There is no incorporation of any nucleotide by either Arg298His or Tyr432Ser even after 30 minutes, this correlates to the very low primer extension activity by both variants in the extension from the dG:dT and $N^2$ffdG:dT mismatches.

Extension from the undamaged dA:dG mismatch also shows a similar pattern, very low incorporation levels, with only dT incorporation seen by every variant except Arg298His and Tyr432Ser. The insertion after the $N^6$ffdA:dG mismatch does not show any incorporation by any variant, this matches with the primer extension as the levels are much lower than with the extension from the dA:dG mismatch and the etheno-dA:dG mismatches. The insertion after the etheno-dA:dG mismatch has the greatest amount of incorporation seen, although this is done at 5-fold protein concentration as compared to the other templates (5 nM vs 25 nM). There is correct dT incorporation by all variants, including the two variants with the weakest primer extension activity. Also seen is multiple dG misincorporation, albeit weak, by WT, Leu21Phe, Pro169Thr, Phe192Cys, Arg219Ile, Glu292Lys, and Ser423Arg.
Figure 5.7. Pol κ WT and SNP variants misincorporate after a lesion show correct dA incorporation from lesions with a mismatched base pair. (A) Undamaged dG DNA, (B) $N^2$-ffdG DNA (using MatchT primer for undamaged dG and $N^2$ffdG), (C) TemplateA DNA, (D) $N^6$ffdA DNA, and (E) etheno-dA DNA (using MatchG primer for TemplateA, $N^6$ffdA, and ethenod-dA).

**Thermal shift assays of protein stability.** We next assessed the overall stability of pol κ and variants alone and with their substrates using a thermal shift assay. Many of the pol κ SNP variants have similar melting temperatures as WT pol κ (40.7 ± 1.2 °C), except Tyr432Ser (37.6 ± 2.1 °C) and Arg298His (32.8 ± 0.6 °C), which are lower (Figure 5.8). Arg298His and Tyr432Ser are also the least active of the all the variants. As shown previously with WT pol κ and other variants, there is a negligible increase in the melting temperature in the presence of any DNA template (Antczak et al. 2017, Nevin et al. 2015a, Nevin et al. 2015b); however, with the addition of DNA and the correct nucleotide there is an increase in melting temperature in the
range of 7 to 12 °C for all the proteins. That large stabilization is not seen with the incorrect nucleotide or the non-preferred DNA lesion (major groove lesions) as we have shown previously (Antczak et al. 2017, Nevin et al. 2015b).

Figure 5.8. Melting temperatures (°C) of WT pol κ and variants in the presence of undamaged DNA, $N^2$ffdG, $N^6$ffdA and etheno-dA DNA. The largest increase in melting temperature is in the presence of the preferred DNA substrates (undamaged and $N^2$ffdG) and the correct nucleotide (dCTP). (A) SNPs in the presence of undamaged DNA and $N^2$ffdG DNA in the presence of correct (dC) and incorrect nucleotide (dT). (B) SNPs in the presence $N^6$ffdA and etheno-dA DNA and in the presence of the correct or incorrect nucleotide.
5.4 Discussion
Pol κ has been shown to bypass minor groove adducts efficiently and is blocked by major groove adducts (Choi et al. 2006a, Choi et al. 2006b, Jarosz et al. 2006, Rechkoblit et al. 2002, Walsh et al. 2013, Yuan et al. 2008a, Zhang et al. 2002). It has also been shown to be an extender from mismatched primer termini (Haracska et al. 2002, Lone et al. 2007, Niimi et al. 2009, Vasquez-Del Carpio et al. 2011, Washington et al. 2002, Wolfle et al. 2003). In this work, we investigated the in vitro activity of nine variants encoded by single nucleotide polymorphisms of human pol κ. The nine variants can be separated into three categories; enhanced insertion and extension activity compared to WT, similar activity to WT, and diminished insertion and extension activity compared to WT. We found that the most active variants, Phe192Cys and Ser423Arg, efficiently insert opposite minor groove and major groove lesions as well as extend from correct and mismatched damaged base pairs. Pol κ Arg298His and Tyr432Ser are the least active variants and most blocked by the major groove adducts. Arg298His and Tyr432Ser variants also cannot extend from mismatched damaged bases but can extend from the minor groove lesion with the correct base pair. The variants Leu21Phe, Ile39Thr, and Pro169Thr have similar insertion and extension activity as seen with WT. Arg219Ile and Glu292Lys have decreased insertion activity but extension from the lesions with both correct and mismatched base pairs is similar to that of WT.

Structurally these variants are spread across the polymerase domains of pol κ except the thumb domain (Table 5.1). Pol κ Phe192Cys is one of two variants that have increased activity compared to WT pol κ, especially in extension from mismatches. In the structure of pol κ (Figure 5.1), Phe192 is in the palm domain adjacent to the linker between the polymerase domain and little finger or PAD. Phe192 appears to interact with the adjacent alpha helix, and thus changing this phenylalanine to a cysteine would decrease the bulk from the phenylalanine and potentially
allow for more flexibility and movement leading to an increase in activity. Phe192Cys has a melting temperature of 41.8 °C; in the presence of undamaged or $N^2$ffdG-containing DNA there is no change in $T_m$ and in the presence of DNA and the correct nucleotide there is a modest increase with undamaged DNA of approximately 5 °C, while with $N^2$ffdG there is an increase of approximately 7 °C. Pol κ Ser423Arg is also overall more active than WT pol κ and particularly when extending from major groove adducts, whether paired with correct or incorrect bases at the primer termini. Ser423 is in the little finger or PAD and is solvent accessible; the hydroxyl group could hydrogen bond with the backbone of the DNA template. Replacing this hydrophilic residue with a larger charged arginine residue could allow for more stabilization of the template strand. Pol κ Ser423Arg has a melting temperature (41.7 °C) similar to WT pol κ and, like Pol κ Phe192Cys, there is no increase in melting temperature in the presence of undamaged or $N^2$ffdG-containing DNA (Figure 5.8). Both variants have similar incorporation and misincorporation as WT except opposite an undamaged dG, in which case pol κ Phe192Cys misincorporates dT and pol κ Tyr432Ser misincorporates dG. Previous work done with Phe192Cys and Ser423Arg variants (1-526 amino acids) resulted in activity similar to WT on undamaged DNA and $N^2$-AnthG-containing DNA (Kim et al. 2016, Song et al. 2014).

The Leu21Phe and Ile39Thr mutations are both in the N-clasp and both variants retain activity similar to WT pol κ with both the minor groove $N^2$ffdG adduct and major groove adducts $N^6$ffdA and etheno-dA. Pol κ Leu21Phe also shows a large increase in extension from an etheno-dA:dT base pair as compared to WT pol κ. It has been shown previously that both Leu21Phe and Ile39Thr variants were markedly less active in bypass of the bulkier $N^2$-AnthG, especially in the extension step (Song et al. 2014). The bulkier adduct in the basepair at the primer terminus could cause more perturbation in the N-clasp compared to the smaller furfuryl adduct, as it has been
proposed that the N-clasp can tolerate distortions at the primer terminus (Jia et al. 2008, Lone et al. 2007). Additionally, our construct, unlike the construct used by Song et al., does not contain the first 18 residues of pol κ, which have been shown to increase DNA binding as well as positively alter the N-clasp’s ability to bind DNA and extend from mismatches (Lone et al. 2007). This difference could account for the different activities by the same variants in different constructs. Both the Leu21Phe and Ile39Thr variants have a similar melting temperature compared to WT pol κ and when in the presence of any of the DNA molecules studied here. When the correct nucleotide and undamaged DNA is present the melting temperatures of the Leu21Phe and Ile39Thr variants increase by approximately 5 °C, unlike WT in which there is an increase of approximately 10 °C as shown previously (Antczak et al. 2017, Nevin et al. 2015a, Nevin et al. 2015b). In the presence of 2ffdG and dCTP, the melting temperatures of the Leu21Phe and Ile39Thr variants increase by approximately 10 °C, similar to that of WT pol κ. However, the Leu21Phe and Ile39Thr variants do now show an increase in melting temperature in the presence of major groove adducts and incoming nucleotides, which correlated with their low activity on DNA containing major groove lesions.

Pol κ Arg219Ile has decreased insertion and extension activity compared to WT while the Glu292Lys variant activity is similar to WT. In the structure of WT pol κ, Arg219 appears to be engaged in hydrogen bonding with a water molecule that could also interact with Trp214. Changing the arginine to a hydrophobic isoleucine eliminates the interaction with the water but likely introduces an interaction between the isoleucine and the phenyl ring of Trp214. The carboxyl group of Glu292 could interact with the backbone of Phe286 in the neighboring beta strand. Lys292 has a longer carbon sidechain, and may be projecting more into the solvent than glutamic acid. Both 219 and 292 are near a region of pol κ that cannot be crystallized, which
means it is likely disordered and it is unclear whether that region of approximately 60 amino acids interacts with the rest of the palm domain. This could potentially explain why the Arg219Ile mutation has a greater effect on the activity of pol κ than Glu292Lys; there could be different interactions with the residues not seen in the crystal structure that are important in catalysis and especially the translocation step. However, the melting temperatures of pol κ Arg219Ile are similar to WT in the presence of all templates and the correct/incorrect nucleotides, indicating the Arg219Ile mutation likely does not have a large effect on protein folding or DNA or nucleotide binding.

The Arg298His and Tyr432Ser variants have much lower activity than WT pol κ on all primer template combinations. Arg298 has two possible hydrogen bonds for stability with Gln332 and Asn330. When arginine is replaced by histidine there is a loss of interaction with Asn330. In a model with His298, the residues are too far apart to hydrogen bond and thus the mutation may affect the folding of the protein, consistent with the thermal shift assay, where the melting temperature of the Arg298His variant is only 32.8 ± 0.6 °C, almost 10 °C lower than WT. When DNA (undamaged or $N^2$ffdG), is present there is a small increase (3-5 °C) in melting temperature but when the correct dNTP is present, there is a 9-10 °C increase (to 45.2 ± 0.3 °C), similar to WT. The lack of activity could be due to the low melting temperature of the variant, since in the presence of DNA the melting temperatures are only around 37 °C and so it is likely that there is a large fraction of unfolded protein in our assays. In the crystal structures, Tyr432 is highly solvent accessible and could interact with the neighboring helix in the little finger. Like Arg298His, the melting temperature of the Tyr432Ser variant is low, 37.6 ± 2.1°C, and there is no or a relatively small increase in the presence of DNA or in the presence of undamaged DNA and the correct dCTP, but there is a 10 °C increase in $T_m$ with $N^2$ffdG-containing DNA and
dCTP. The change from the larger phenol ring of tyrosine to the much smaller serine could allow for more flexibility in the little finger (PAD) and could be why when protein is alone the melting temperature is lower than other variants. The low activity of these two proteins could be correlated with the low melting temperatures although it appears that both variants bind DNA and the correct nucleotide based on the corresponding thermal shifts.

Overall, most of the SNPs, like WT, have a greater increase in melting temperature in the presence of $N^2$ffdG and dCTP compared to undamaged DNA and dCTP although many have similar *in vitro* activity as WT. There is no clear increase in activity for the SNPs on undamaged DNA versus DNA containing $N^2$ffdG. The SNPs are all slightly blocked by major groove adducts like WT but can extend better from the $N^6$ffdA:dT basepair than from the ethenodA:dT or mismatched dA:dG primer termini.
5.5 References


Chapter 6: Inhibitors of DinB are less effective against pol κ in vitro

6.1 Introduction

Cells are exposed to mutagens constantly and have multiple pathways to deal with damage by either repair or tolerance. Translesion synthesis (TLS) is one such pathway to bypass unrepaired adducts on DNA. Y-family DNA polymerases, including *E. coli* DinB and human pol κ, are specialized to perform TLS. Both DinB and pol κ have been shown to bypass proficiently minor groove \( N^2\)-dG adducts and DNA-protein crosslinks (Jarosz et al. 2006, Minko et al. 2008b), which can be caused by polycyclic aromatic hydrocarbons and acrolein, respectively.

Y-family DNA polymerases are attractive drug targets because bacterial Y-family polymerases contribute to antibiotic resistance (Cirz et al. 2005, Cirz et al. 2007a, Yamanaka et al. 2012) and human Y-family polymerases decrease the effectiveness of chemotherapy drugs whose mechanism of action is DNA damage (Guo et al. 2009, Loeb et al. 2008). It has been shown that the antibiotic ciprofloxacin induces the SOS response in *Staphylococcus aureus* which leads to an increase in the levels of the Y-family DNA polymerases DinB and UmuC (Cirz et al. 2005, Cirz et al. 2007a). Pol κ mRNA has been found to be elevated in non-small-cell lung cancers (O. Wang et al. 2001). It was found that overexpression of pol κ increased the number of point mutations in mice (Ogi et al. 2002). Pol κ has also been shown to be involved in the tolerance of mitomycin C, a chemotherapeutic agent that causes DNA-DNA interstrand crosslinks (Minko et al. 2008a). Many chemotherapeutic drugs covalently bind to DNA creating interstrand crosslinks that interfere with DNA replication and transcription but can be bypassed by Y-family DNA polymerases employing TLS, making these drugs less effective. One way to increase the efficacy of chemotherapy drugs is to inhibit Y-family polymerases (Guo et al. 2009).
There have been two compounds that have been found to selectively inhibit human Y-family DNA polymerases, 3-O-methylfunicone (Mizushina et al. 2009) and 1-deoxyrubralactone (Naganuma et al. 2008). Pol κ was the most strongly inhibited polymerase of the human Y-family DNA polymerases by both 3-O-methylfunicone and 1-deoxyrubralactone. There was no effect on polymerases from other organisms, including plant and prokaryotic polymerases (Mizushina et al. 2009, Naganuma et al. 2008). A high-throughput method to discover inhibitors of pol κ using fluorescent oligonucleotides has been developed and a number of hit compounds (Candesartan cilexetil, manoalide, MK-886) were identified and characterized further (Yamanaka et al. 2012).

![Chemical structures](image)

*Figure 6.1. Structures of the potential inhibitors found in the molecular modeling screen.*
Previous work done in the lab identified potential inhibitors of DinB using a computational binding screen of a library of compounds. The compounds identified were glycolic acid, pamoic acid, phenyl phosphate, 2-hydroxy-5-nitrobenzoic acid, aurintricarboxylic acid, d-galacturonic acid, glycinamide, 3-nitrosalicylic acid, and ellagic acid (Figure 6.1). In this work, the compounds were tested in vitro against both DinB and pol κ using primer extension assays with varying amounts of compound. We found that most of the compounds inhibited DinB to a greater extent than pol κ.

6.2 Materials and Methods

Proteins. DinB and pol κ were expressed and purified as described previously (Irimia et al. 2009, Jarosz et al. 2006). The DNA template containing a single $N^2$-furfuryl-dG was prepared as previously described (DeCorte et al. 1996, Jarosz et al. 2006). DNA was purified by denaturing polyacrylamide gel electrophoresis and the crush and soak method (Sambrook et al. 1989). DNA running start or standing start primer was end-labeled with $^{32}$P as previously described (Beuning et al. 2006, Sambrook et al. 1989).

Primer Extension Assays. The primer extension assays were completed with 61-mer DNA template (undamaged or containing a single $N^2$-furfuryl-dG) and either running start or standing start DNA. The template DNA sequence was $^5'$-CGTTACTCAGATCAGCTGCAGCCGGACGC-3' where X is either an undamaged G or $N^2$-furfuryl-dG. The running start sequence was $^5'$-GCATATGATAGTACAGCTGCAGCCGGAGC-3' and the standing start sequence was $^5'$-
GCATATGATAGTACAGCTGCAGCCGGACGCC-3'. The DNA was synthesized by Operon, Eriks Rozners (SUNY Binghamton), or Ke Zhang (Northeastern University).

DNA template (unmodified or N²ffdG) was combined with ³²P-labeled primer (running start) in a 1:1 ratio (500 nM) and annealed with annealing buffer (20 mM HEPES (pH 7.5) and 5 mM Mg(OAc)₂) by heating for 2 min at 95 °C, incubating at 50 °C for 60 min and cooling to 37 °C. Reactions were carried out in 1X reaction buffer (30 mM HEPES (pH 7.5), 20 mM NaCl, 7.5 mM MgSO₄, 2 mM β-mercaptoethanol, and 1% bovine serum albumin), 100 nM ³²P labeled primer/template and 10 nM protein (1 nM protein was used for some pol κ assays, as noted) and initiated with 500 µM dNTPs. For assays with compounds, a solvent control was used if the compound was dissolved in anything other than water and then a range of 0.1 to 100 µM compound of interest was added to each reaction. Pamoic acid was dissolved in HEPES as it is not soluble in water, aurintricarboxylic acid was dissolved in ethanol, and 2-hydroxy-5-nitrobenzoic acid was dissolved in methanol.

6.3 Results and Discussion

Overall, most compounds selectively inhibited DinB over pol κ. Some inhibitor assays with pol κ were done at 10-fold lower enzyme concentration (1 nM, pamoic acid, D-galacturonic acid, and glycinamide HCl) as pol κ is more active than DinB and there was no effect seen at 10 nM pol κ. The remaining inhibitor compounds (glycolic acid, phenyl phosphate, and 2-hydroxy-5-nitrobenzoic acid) were completed at 10 nM of both DinB and pol κ.

Phenyl phosphate inhibits DinB at concentrations as low as 0.1 µM while there is no effect on pol κ activity. Pol κ activity at 10 nM protein is not affected by phenyl phosphate while DinB activity is decreased at the lowest concentration tested (Figure 6.3).
Figure 6.2. (A) Pol κ is unaffected by phenyl phosphate while DinB (B) activity is affected even at low concentrations. The pol κ and DinB concentrations tested were 10 nM in the presence of phenyl phosphate.

Pol κ is unaffected by 2-hydroxy-5-nitrobenzoic acid (5-nitrosalicylic acid) while DinB activity is affected even at low concentrations. DinB is activity is inhibited with both undamaged and $N^2$-ffdG DNA, at the lowest concentration, 0.1 µM of 2-hydroxy-5-nitrobenzoic acid tested (Figure 6.3). On undamaged DNA, there is a small increase in activity in the presence of 100 µM phenyl phosphate. On the other hand, pol κ is not inhibited by 2-hydroxy-5-nitrobenzoic acid.
Figure 6.3. (A) Pol κ is unaffected by 2-hydroxy-5-nitrobenzoic acid (5-nitro salicylic acid) while DinB activity is affected even at low concentrations. The pol κ and DinB concentrations tested were 10 nM in the presence of 2-hydroxy-5-nitrobenzoic acid. As the compound was dissolved in methanol, the control reaction contains methanol.

Both pol κ and DinB activity are inhibited by aurintricarboxylic acid, though DinB is inhibited at a 10-fold lower concentration. DinB activity is inhibited in the presence of 0.1 µM of aurintricarboxylic acid while pol κ activity is not affected at concentrations below 1 µM of the compound. Both proteins have no detectable activity at 10 µM or 100 µM aurintricarboxylic acid (Figure 6.4).
Figure 6.4 Aurintricarboxylic acid affects both (B) DinB and (A) pol κ however a 10-fold higher concentration is required for pol κ inhibition. The pol κ and DinB concentrations tested were 10 nM in the presence of aurintricarboxylic acid. As the compound was dissolved in ethanol, the control reaction contains ethanol.

Pol κ is only inhibited by pamoic acid at the highest concentration (100 µM) while DinB activity is inhibited at concentrations as low as 1 µM. In the presence of pamoic acid, the pol κ concentration tested was 1 nM, while DinB was tested at 10 nM. As shown in Figure 6.5, Pol κ activity is inhibited with 100-fold higher pamoic acid concentrations.
Figure 6.5. (A) Pol κ is only affected by pamoic acid at very high concentrations while (B) DinB activity is affected at lower concentrations. The pol κ and DinB concentrations tested were 1 nM and 10 nM, respectively, in the presence of pamoic acid. As the compound was dissolved in HEPES, the control reaction contains HEPES.

Glycinamide hydrochloride more strongly inhibits DinB activity (at 10 nM) than pol κ, even though the pol κ concentration was 1 nM (Figure 6.6). There is a modest decrease in activity for DinB on both DNA templates. The decrease occurs in the presence of as low as 0.1 µM glycinamide hydrochloride and the activity remains the same even at the much higher concentrations tested. Although pol κ concentration is 10-fold lower than DinB, glycinamide hydrochloride does not inhibit pol κ as strongly at concentrations below 100 µM.
Figure 6.6. (A) Pol κ is only affected by glycinamide HCl at very high concentrations while (B) DinB activity is affected at low concentrations. The pol κ and DinB concentrations tested were 1 nM and 10 nM, respectively, in the presence of glycinamide hydrochloride.

As shown in Figure 6.7, d-galacturonic acid also shows specificity between the two polymerases, as DinB is inhibited by d-galacturonic acid at concentrations as low as 0.1 µM compound. There is a more apparent inhibition of DinB activity on the N²ffdG template as compared to the undamaged DNA template.
Pol κ is only affected by D-galacturonic acid at very high concentrations while DinB activity is affected at low concentrations and only on damaged DNA template. The pol κ and DinB concentrations tested were 1 nM and 10 nM, respectively, in the presence of D-galacturonic acid.

Glycolic Acid decreased DinB activity at low concentrations while at higher concentrations did not affect the activity of DinB. Assays with glycolic acid present an anomaly, as assays with low concentration affected DinB activity much more than the assays with much higher concentrations of glycolic acid (Figure 6.8). Pol κ did not appear to be affected by glycolic acid (10 nM protein).
Figure 6.8. (A) Pol κ is unaffected by glycolic acid while (B) DinB activity is affected at low concentrations but not at the higher concentrations. The pol κ and DinB concentrations tested were 10 nM in the presence of glycolic acid.

6.4 Discussion

Overall the compounds tested were shown to inhibit *E. coli* DinB preferentially over human pol κ, with the exception of aurintricarboxylic acid. Only in the presence of glycinamide hydrochloride or D-galacturonic acid was there a difference in inhibition of DinB on the undamaged and $N^2$-ffdG templates.

Glycolic acid is used in chemical skin peels for rehabilitating photodamaged skin and is naturally occurring in some foods (Monheit 2004, Murad et al. 1995). Glycolic acid has also been used with polylactic acid for controlled protein and drug delivery systems (Cohen et al. 1991, Makadia et al. 2011). It has been shown that polyglycolic acid delivery systems do not induce an inflammatory response as compared to non-biodegradable nanospheres (Dailey et al. 2006, Sundback et al. 2005). Since glycolic acid is used as a drug delivery system it would be a
poor inhibitor choice in general as those systems should not affect cellular or protein function in humans. It does inhibit DinB at low concentrations and could be used against bacterial cells but in vivo work would need to be done to determine any effect on cellular growth.

Phenyl phosphate has been shown to be an inhibitor of alkaline phosphatase as well as 5’ nucleotidase with increasing concentrations of phenyl phosphate (above 80 mM) showing greater inhibition (Belfield et al. 1970). The enzyme 5’ nucleotidase hydrolyzes AMP to adenosine and phosphate (Belfield et al. 1970). Phenyl phosphate does not inhibit pol κ but does slightly inhibit DinB at concentrations as low as 0.1 µM. It is possible that at concentrations above 80 mM, there would be more inhibition since both polymerases bind nucleotides.

Aurintricarboxylic acid has been shown to be an inhibitor of protein synthesis as well as several RNA polymerases (Grollman et al. 1968, Stewart et al. 1971) and was predicted to inhibit any nucleotide binding enzyme (Blumenthal et al. 1973). The mechanism of action was revealed to be competition between aurintricarboxylic acid and nucleotides for binding to the active site (Gonzalez et al. 1980). It was found to be a non-specific enzyme inhibitor, inhibiting a wide variety of enzymes including DNase I, glucose 6-phosphate dehydrogenase, lysozyme, trypsin, and chymotrypsin but did not inhibit seryl-tRNA synthetase (Bina-Stein et al. 1976). Aurintricarboxylic acid was shown to have nanomolar IC50 against pols β, ι, and η (Dorjsuren et al. 2009). Our results support that it is an inhibitor of protein-nucleic acid interactions as both DinB and pol κ are inhibited by the compound.

Salicylic acid derivatives (like 2-hydroxy-5-nitrobenzoic acid) have been shown to inhibit nuclear factor kappa B (NFκB) activity and could be useful in treatment of diseases, with NFκB playing a major role in the progress of colitis (Kim et al. 2012). It has been found that 5-nitrosalicylic acid could be used as prodrug of 5-aminosalicylic acid, which is used in the
treatment of inflammatory bowel disease (IBD) (Saphier et al. 2012). The cause of IBD is still unclear but it is thought that it is an immune response to gut bacteria. Although 5-nitrosalicylic acid could be a prodrug of 5-aminosalicylic acid, we have shown it inhibits DinB a bacterial protein. This could make it a potentially non-ideal drug candidate as it is nonspecific and will inhibit not only NFκB but also bacterial DinB.

Pamoic acid has been shown to inhibit pol 1 more strongly (IC₅₀ 4.9 µM) as compared to, pols β and η (IC₅₀ 60 µM and 79 µM, respectively) (Dorjsuren et al. 2009). We have shown here that pamoic acid inhibits DinB starting at concentrations between 0.1 µM and 1 µM while there is no inhibition of pol κ until the concentration of compound reaches 10 µM -100 µM.

Glycinamide hydrochloride has been tested in yeast strains to screen for anti-cancer activity. It was found to be inactive in the following deletion strains, Δrad50 (2005), Δmec2-1 (2005), Δsgs1mgt1 (2005), Δcln2rad14 (2005), Δbub3 (2005), and Δmlh1rad18 (2005). Our data supports this as there is no inhibition of human pol κ and modest inhibition of DinB so glycinamide HCl may not affect the DNA damage response or interact with other polymerases.

D-galacturonic acid is naturally abundant as it is a constituent of pectin and it is believed pectin could be used for the biotechnological conversion of sugars to fuels (Richard et al. 2009). E. coli can convert D-galacturonic acid to D-glyceraldehyde-3-phosphate and E. coli has been engineered to convert D-galacturonic acid to ethanol (Doran et al. 2000). In the presence of D-galacturonic acid DinB activity is only inhibited with damaged DNA.

Many of the compounds we tested did not inhibit pol κ to the same extent as DinB and this could be due to the fact that the molecular modeling screen was done only with DinB. In addition, the pol κ N-clasp could block some of the compounds from entering the active site, especially larger molecules. Additional misincorporation assays and melting temperature assays
will need to be completed. Also primer extension assays with polymerases from other families will need to be completed as well to check the specificity for DinB.
6.5 References

(2005). NCI Yeast Anticancer Drug Screen. Data for the Bub3 strain, Cln2Rad14 strain, mec2-1 strain, Mih1 Rad19 strain, Rad50 strain, Sgs1 Mgt1 strain. PubChem BioAssay Database, National Cancer Institute Division of Cancer Treatment and Diagnosis.


DeCorte, B. L., D. Tsarouhtsis, S. Kuchimanchi, M. D. Cooper, P. Horton, C. M. Harris and T. M. Harris (1996). "Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N2 position of deoxyguanosine." Chemical Research in Toxicology 9(3): 630-637.


Chapter 7: Conclusions and future work

The focus of this work has been on understanding the DNA damage and nucleotide specificity of Y-family DNA polymerases *E. coli* DinB and human pol κ. DinB and pol κ have been shown preferentially to bypass minor groove adducts on the $N^2$ position of deoxyguanine (Jarosz et al. 2006) and are blocked by major groove adducts (Basu et al. 2017, Choi et al. 2006b, Levine et al. 2001, Nevin et al. 2015b, Walsh et al. 2011a, Walsh et al. 2013, Xu et al. 2016) while DinB archaeal ortholog Dpo4 can bypass major groove adducts. In order to understand this difference in damage specificity, active site loop chimeras of DinB, pol κ, and Dpo4 were created and it was found that DinB had low tolerance to changes in the active site loop while pol κ was much more tolerant (Chapters 3, 4). Y-family DNA polymerases are known to misincorporate nucleotides opposite both undamaged and damaged DNA. In this work, we found that there are multiple residues in both DinB and pol κ that are important for fidelity of nucleotide incorporation opposite undamaged dG and the $N^2$fdG lesion (Chapter 2).

Human Y-family DNA pols are known to play a role in cancer and chemotherapy resistance and bacterial Y-family DNA pols are important in the development of antibiotic resistance. A large number of single nucleotide polymorphisms of pol κ have been identified in cancers and in this work, we biochemically characterized several SNPs to understand their potential roles (Chapter 5). Since the Y-family DNA pols are implicated in chemotherapy resistance and antibiotic resistance it is believed that inhibitors could be developed to increase the efficacy of chemotherapy and antibiotic courses. In Chapter 6, we have shown that there is a potential to inhibit *E. coli* DinB selectively over human pol κ. Continuation of this work could be expanded to include other polymerases from the other DNA pol families, like replicative DNA pols and other Y-family pols from both bacteria and eukaryotes. The compounds could also be
used in *in vivo* assays and in conjunction with antibiotics. Expanding beyond the biochemistry, additional derivatives could be made and tested.

In Chapters 3 and 4, we have shown that the active site loop length of DinB is important for DinB activity (both shortening and lengthening) while pol κ is tolerant of longer loops (Antczak et al. 2017). We found that the active site loop region is not part of the damage specificity as DinBDpo4 loop swaps and KappaDpo4 loop swaps do not show an increase in bypass of major groove adducts as compared to WT DinB or WT pol κ. It has also been shown that the linker between the polymerase and little finger domains plays a key role in polymerase activity, as swapping the linkers of Dpo4 and Dbh resulted in polymerase activity similar to that of the linker parental polymerase (Wilson et al. 2013). A next step beyond testing the Dpo4 loop swaps would be to create linker swaps and little finger swaps between the three proteins and assay activity against major groove adducts. Another interesting experiment would be to add the pol κ N-clasp at the N-terminus of DinB or Dpo4 to assess the effects of the N-clasp as it has been shown to be important for pol κ activity (Jia et al. 2008, Lior-Hoffmann et al. 2014, Lone et al. 2007, Uljon et al. 2004).
7.1 References
References

(2005). NCI Yeast Anticancer Drug Screen. Data for the Bub3 strain, Cln2Rad14 strain, mec2-1 strain, Mih1 Rad19 strain, Rad50 strain, Sgs1 Mgt1 strain. PubChem BioAssay Database, National Cancer Institute Division of Cancer Treatment and Diagnosis.


DeCorte, B. L., D. Tsarouhtsis, S. Kuchimanchi, M. D. Cooper, P. Horton, C. M. Harris and T. M. Harris (1996). "Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N2 position of deoxyguanosine." Chemical Research in Toxicology 9(3): 630-637.


Ippoliti, P. (2012). *DNA Damage Specificity of* *Escherichia coli* *DNA Polymerase DinB*. Master of Science, Northeastern University.


Washington, M. T., R. E. Johnson, L. Prakash and S. Prakash (2002). "Human DINB1-encoded DNA polymerase kappa is a promiscuous extender of mispaired primer termini."


185


