DNA Damage Specificity of \textit{Escherichia coli} DNA Polymerase DinB

A thesis presented

by

Paul Ippoliti

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements for the degree of Master of Science

in the field of Chemistry

Northeastern University

Boston, Massachusetts

August 2012
DNA Damage Specificity of *Escherichia coli* DNA Polymerase DinB

by

Paul J. Ippoliti

ABSTRACT OF THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Chemistry and Chemical Biology in the Graduate School of Science of Northeastern University, August, 2012
ABSTRACT

Exogenous and endogenous DNA damaging agents such as UV light cause lesions in DNA, which halt the progress of DNA replication. Microbes such as *E. coli* have a response mechanism known as the SOS response which is initiated upon DNA damage. The SOS response ultimately results in the expression of the Y-family DNA polymerases which catalyze translesion synthesis to copy damaged DNA (Walsh 2011). Y-family polymerases specifically bypass certain types of DNA lesions. Understanding DNA damage specificity of Y-family polymerases is necessary due to their mutagenic nature. They lack proofreading ability which removes incorrectly inserted nucleotides, and can therefore be error-prone (Jarosz 2007). Eventually, the insertion of the incorrect base can lead to mutations that confer a selective advantage in the cell, such as antibiotic resistance. In *E. coli*, there are two Y-family polymerases known as polymerase IV and polymerase V. Polymerase V, also known as UmuD’2C, is capable of bypassing lesions caused by UV light exposure namely thymine-thymine cyclobutane pyrimidine dimers and T-T (6-4) photoproducts (Tang 2000). Polymerase IV, otherwise known as DinB, is capable of bypassing N2-dG adducts (among others), but not T-T CPDs or (6-4) photoproducts (Tang 2000; Jarosz 2006). The goal of this study is to understand how DinB is able to bypass certain types of lesions, but not others. In this study, directed evolution methods are used to mutate the DinB gene and acquire variants that confer UV resistance. Characterization of these variants will help to understand which regions on the DinB protein allow for the specificity of lesion bypass. In addition, primer extension assays are used to study the lesion bypass efficiency of DinB variants in the Loop 1 region of the protein.
References


ACKNOWLEDGEMENTS

First of all, I would like to thank Dr. Penny Beuning for all of her guidance and support from the day I joined the lab as undergraduate and through my graduate work to attain my Master’s. I especially thank Penny for her availability and her willingness to help despite having an extremely busy schedule; you truly care about the success of your students, thank you.

Second, I would like to thank Dr. Mary Ondrechen and Dr. Lee Makowski for their support and for taking the time to review this thesis. I also thank Dr. Jana Sefkicova for all of her guidance and training in the lab and taking time away from your work to answer questions or to train me on a new assay. I would also like to thank Lisa Hawver, for training me when I first joined the lab as an undergraduate, and Jason Walsh for his assistance and training on several procedures I learned later on in my time here in the lab. I’d like to thank the other members of the DNA lab (in no particular order): Jaylene, Michelle, Philip, Nicholas, Susan, Caitlin, Ramya, Luke, Judith, Rachel, Jen, Caitlin G., Danielle, Ariana, and Grace. Finally, I would like to thank my family: Tony, Lynne, and Eric, for their love and support, and Samantha Sokup for all her love and dedication, and for putting up with me when I was stressed. Thank you
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... 3

ACKNOWLEDGEMENTS ....................................................................................................... 6

TABLE OF CONTENTS ........................................................................................................... 7

List of Figures .......................................................................................................................... 9

List of Tables .......................................................................................................................... 11

List of Abbreviations ............................................................................................................. 12

Chapter 1: Introduction ........................................................................................................... 16

1.1 Introduction ....................................................................................................................... 16

1.2 *E. coli* Pol IV ..................................................................................................................... 19

1.3 *E. coli* Pol V ....................................................................................................................... 25

1.4 DinB and UmuC orthologs in other bacteria ...................................................................... 29

1.5 References ........................................................................................................................ 32

Chapter 2: Understanding DinB Specificity through Directed Evolution ................................. 42

2.1 Introduction ....................................................................................................................... 42

2.2 Materials and Methods ..................................................................................................... 44

   a. Strains and Plasmids ......................................................................................................... 44

   b. Plasmid Preparation ......................................................................................................... 45

   c. Transformations .............................................................................................................. 46

   d. Hydroxylamine Mutagenesis .......................................................................................... 46

   e. Randomized Primers ....................................................................................................... 47

   f. UV Selection Assays ....................................................................................................... 48

   g. UV Survival Assays ........................................................................................................ 49

   h. Other Methods ............................................................................................................... 50

2.3 Results ............................................................................................................................... 51

   a. DinB UV Selection Assays .............................................................................................. 51

   b. DinB UV Survival Assays ............................................................................................... 53

2.4 Discussion .......................................................................................................................... 56
Chapter 3: Determining Activity and DinB Damage Specificity of the DinB Loop 1 Region

3.1 Introduction

3.2 Materials and Methods
   a. Expression and Purification of DinB Wild-type and DinB Variants
   b. Primer Extension Assays

3.3 Results
   a. Loop Swap Assays
   b. Primer Extension Results on Single Site-directed Mutations
   c. Fidelity of Incorporation

3.4 Discussion

3.5 References

Chapter 4: Understanding Bypass of DNA Containing Thymine-Thymine Cyclobutane Pyrimidine Dimers by E. coli DinB

4.1 Introduction

4.2 Materials and Methods
   a. Initial Attempts
   b. Final Attempts

4.3 Results

4.4 Discussion

4.5 References

Chapter 5: Future Consideration

5.1 Conclusions

5.2 Future Considerations
   a. Directed Evolution
   b. DinB Variants

5.4 References
List of Tables

Table 2.1: ..............................................................................................................Error! Bookmark not defined.

Table 4.1: ..............................................................................................................Error! Bookmark not defined.

Table 4.2: ..............................................................................................................Error! Bookmark not defined.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>percent</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>2-oxo-dA</td>
<td>1,2-dihydro-oxoadenine</td>
</tr>
<tr>
<td>4NQO</td>
<td>4-nitroquinoline</td>
</tr>
<tr>
<td>5-fodU</td>
<td>5-formyluracil</td>
</tr>
<tr>
<td>5-hmdU</td>
<td>5-hydroxymethyluracil</td>
</tr>
<tr>
<td>8-oxo-dG</td>
<td>7,8-dihydro-8-oxoguanine</td>
</tr>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>Å</td>
<td>angstrom</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>B[a]P</td>
<td>benzo[a]pyrene</td>
</tr>
<tr>
<td>BL21</td>
<td>competent <em>E. coli</em> cells</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>C^8-AAF</td>
<td>N-2-acetylaminofluorine</td>
</tr>
<tr>
<td>CEdG</td>
<td>N^2-(1-carboxyethyl)-2’-deoxyguanosine</td>
</tr>
<tr>
<td>CPD</td>
<td>cyclobutane pyrimidine dimer</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>dA</td>
<td>deoxyadenine</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenine triphosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------</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>EV</td>
<td>empty vector</td>
</tr>
<tr>
<td>FPLC</td>
<td>fast protein liquid chromatography</td>
</tr>
<tr>
<td>fs</td>
<td>femtosecond</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>Glu</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>ICL</td>
<td>inter-strand crosslinks</td>
</tr>
<tr>
<td>Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>IPTG</td>
<td>β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>J/m²</td>
<td>Joules per meter squared</td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>K_M</td>
<td>Michaelis-Menten constant</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
</tbody>
</table>
M molar
M-C mitomycin C
mg milligram
min minute
mL milliliter
mM millimolar
MMS methanesulfonate
NFZ nitrofurazone
ng nanograms
nm nanometer
nM nanomolar
PCR polymerase chain reaction
PDB Protein Database
Phe phenylalanine
Pol polymerase
ps picosecond
ROS reactive oxygen species
rpm revolutions per minute
SDS-PAGE sodium dodecyl sulfate-polyacrylamine gel electrophoresis
Ser serine
ssDNA Single-stranded DNA
T thymine
t-BHP tert-butylhydrogenperoxide
tC 1,3-diaza-2-oxophenothiazine
Thr threonine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLS</td>
<td>Translesion synthesis</td>
</tr>
<tr>
<td>T-T</td>
<td>thymine-thymine</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Val</td>
<td>valine</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>maximum velocity</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
<tr>
<td>xg</td>
<td>g-force</td>
</tr>
<tr>
<td>$\gamma$-HOP-dG</td>
<td>$\gamma$-hydroxypropano-deoxyguanosine</td>
</tr>
<tr>
<td>$\mu$g</td>
<td>microgram</td>
</tr>
<tr>
<td>$\mu$L</td>
<td>microliter</td>
</tr>
<tr>
<td>$\mu$M</td>
<td>micromolar</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Introduction

Various endogenous and exogenous agents, when they enter a cell, can cause damage to DNA, creating lesions and leading to mutations (Lindahl 1993). Replicative DNA polymerases, enzymes that catalyze DNA replication, are incapable of replicating past damaged DNA lesions (Lindahl 1993). Therefore, cells across all domains of life are equipped with specialized DNA polymerases that have the ability to replicate past damaged DNA lesions in a process known as translesion synthesis (TLS), which was first proposed over 30 years ago (Radman 1975). One such family of polymerases is known as the Y-family of DNA polymerases (Ohmori 2001). Despite this specialized function to bypass damaged DNA, Y-family polymerases are unable to replicate undamaged DNA with the same high fidelity as replicative polymerases, resulting in potential mutations (Ohmori 2001). The Y-family polymerases discussed in this chapter will focus on the two E. coli Y-family polymerases as well as Y-family polymerases found in other species of eubacteria.

The mechanism governing Y-family polymerases and the regulation of TLS in E. coli is known as the SOS response (Radman 1975; Friedberg 2006; Kim 2009). Under normal cellular conditions (i.e. non-stress conditions), a repressor protein inhibits the expression of the Y-family polymerase genes, an idea first proposed by Witkin in 1967 (Witkin 1967). This mechanism has since been clarified; namely, the repressor protein LexA was identified and shown to bind to operator sites repressing gene expression and to become inactivated upon damage of DNA (Kim 2009). The SOS response includes the activation of approximately 57 genes that are also involved in many other cellular processes (Beuning 2006a; Simmons 2008). When DNA damage is present and normal replication by replicative polymerases is inhibited, a region of single-
stranded DNA (ssDNA) is formed. RecA protein binds to and polymerizes on the newly formed ssDNA, forming a nucleoprotein filament. The LexA repressor protein then binds to the RecA/ssDNA nucleoprotein complex, which induces autocatalytic cleavage of LexA at the Ala84-Gly85 bond, thereby allowing the Y-family polymerase genes and others to be expressed (Horii 1981; Little 1994).

Y-family polymerases are found throughout all domains of life (Ohmori 2001), including *E. coli* and other species of bacteria which will be discussed in this chapter. However, Y-family polymerases also exist in other species such as the archael strain *Sulfolobus solfataricus* Dpo4 (Silvian 2001) and *S. cerevisiae* Pol η and Rev1 (Ohmori 2001) and *H. sapiens* Pol η, Pol ι, Pol κ, and Rev1 (Ohmori 2001; Goodman 2002). Y-family polymerases share common characteristic structural features such as the palm, finger, and thumb domains (Friedberg 2001). Also characteristic of Y-family polymerases is the presence of the little finger domain (Ling 2001). The overall size of the finger and thumb domains of Y-family polymerases is smaller than those of their replicative counterparts, which results in open, solvent-accessible active sites to allow for large, bulky lesions to enter the active site (Ohmori 2001; Chandani 2010). In addition, the Y-family polymerases lack the characteristic α helix, known as the ‘O-helix’ in *E. coli* pol I, which is used in replicative polymerases to improve their fidelity. The lack of this α helix presumably contributes to the ability of Y-family polymerases to accommodate damaged DNA lesions (Kaushik 1996; Ling 2001; Ogawa 2001; Beard 2003).
Figure 1.1: Shown above is a model of DinB from the homology model built in our lab (PDB file: ‘DinB_2012’) representing the four domains: thumb (blue), palm (red), finger (magenta), and little finger (green). The DNA is represented in yellow.

Prior to discussing DinB and UmuD’, it is necessary to briefly discuss *E. coli* UmuD. UmuD is the product of the *umuD* gene expressed along with *dinB* and *umuC* (Elledge 1983; Shinagawa 1983). UmuD2 is the predominating form of the protein for the first 20 to 30 minutes after induction by the SOS response (Opperman 1999). UmuD, in conjunction with UmuC, acts in a DNA damage checkpoint (Sutton 2001). When cells are grown at 30 °C and UmuD and UmuC are present at increased levels, they inhibit DNA replication in a role completely distinct from their role in TLS (Marsh 1985; Opperman 1999). The also slow the replication process after the cell has been exposed to UV light (Opperman 1999). When UmuD2 interacts with the RecA/ssDNA nucleoprotein filament, the filament facilitates UmuD autocatalytic cleavage, thereby removing the 24 N-terminal amino acids (Burckhardt 1988; Nohmi 1988; Shinagawa 1988). This cleavage is similar to the autocatalytic cleavage of LexA, also facilitated by the
RecA/ssDNA nucleoprotein filament (Paetzel 1999). However, the efficiency of cleavage is much greater for LexA than it is for UmuD\(_2\) (Burckhardt 1988). This cleavage process typically occurs about 20 to 40 minutes after the initiation of the SOS response (Opperman 1999; Friedberg 2006). The cleaved form of UmuD\(_2\) (UmuD’) then interacts with UmuC to form UmuD’:\(C\), otherwise known as Pol V, which is capable of performing TLS (Bruck 1996; McDonald 1998; Reuven 1999; Tang 1999; Friedberg 2006). Also, both UmuD and UmuC prevent RecA-dependent homologous recombination as a result of the interaction between UmuD’:\(C\) and the RecA/ssDNA nucleoprotein filament (Sommer 1993; Szpilewska 1995; Rehrauer 1996). Interestingly, full-length UmuD\(_2\) is involved in prevention of mutagenesis, whereas UmuD’ is involved in its facilitation since Pol V can be potentially mutagenic.

1.2  \textit{E. coli} Pol IV

DinB, initially identified as the product of the \textit{dinP} gene (Ohmori 1995), was discovered in 1980 as one of the damage inducible genes and was therefore named \textit{dinB} (Kenyon 1980). It was first discovered when DNA damaging agents were used to treat \textit{E. coli} cells and the resulting induced genes were identified (Kenyon 1980). The \textit{dinB} gene encodes one of the two \textit{E. coli} Y-family DNA polymerases (DinB, Pol IV) capable of bypassing lesions in DNA via translesion synthesis (Wagner 1999). DinB is the Y-family DNA polymerase that is considered to be conserved throughout all domains of life, although \textit{Saccharomyces cerevisiae} apparently lacks a DinB ortholog (Fuchs 2004). In non-SOS conditions, DinB is expressed at approximately 250 molecules per cell; however, the number of DinB molecules increases by approximately 10 fold under SOS induced conditions (Kim 2001). Of all five DNA polymerases in \textit{E. coli}, DinB is expressed at the highest level during SOS conditions (Kim 2001). Like other Y-family polymerases, DinB is a DNA-dependent DNA polymerase that is capable of copying damaged
primer-template structures and lacks 3’-5’ exonuclease proofreading abilities (Wagner 1999). The fidelity of the dinB gene product is lower than that of the replicative polymerase in E. coli, the Pol III holoenzyme (Kobayashi 2002). The presence of DinB during the TLS process can increase mutagenesis as a result of its relatively low fidelity and the lack of 3’-5’ exonuclease activity (Wagner 1999; Strauss 2000; Kim 2001; Kuban 2004; Wolff 2004; Satou 2005).

In addition to being expressed and activated by the SOS response, it has been proposed that DinB may also be induced in part by a process known as adaptive mutagenesis (McKenzie 2001; Tompkins 2003). In an experiment using Lac reporter strains of E. coli, there was a 7-fold decrease in mutations occurring in strains lacking the dinB gene than in the wild-type strain suggesting that DinB can induce mutations (Tompkins 2003). These mutations which result from the adaptive mutagenesis process can cause cells to have a selective advantage during stressful conditions (Foster 2007). The precise mechanism for adaptive mutagenesis and how it leads to high levels of DinB expression is not fully understood (McKenzie 2003; Slechta 2003; Hersh 2004). Overall, DinB expression can be considered a general stress response mechanism which can lead to high rates of mutagenesis and ultimately result in antibiotic resistance (McKenzie 2001; Cirz 2007; Smith 2007).

Currently, no crystal structure of DinB exists; however, homology models (Seo 2006; Jarosz 2009) have been constructed using the crystal structures of Dpo4 from Sulfolobus solfataricus (Ling 2001) and Dbh from Sulfolobus acidocaldarius (Silvian 2001) as templates. The steric gate residue, the amino acid residue of a DNA polymerase that prevents ribonucleotides from entering the active site (Astatke 1998), of DinB is Phe13 (Jarosz 2006). Changing the steric gate residue of DinB increases the frequency of ribonucleotide incorporation from less than $10^{-5}$ to $10^{-3}$ (Jarosz 2006).
DinB is known to bypass certain dG adducts. For example, DinB is effective in bypass of \( N^2 \)-dG adducts formed from benzo[a]pyrene (B[a]P), a bulky polycyclic carcinogen which is metabolically activated to form 7\( R,8S \)-dihydroxy-9\( S,10R \)-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene (Shen 2002; Seo 2006). In the presence of DinB, \( N^2 \)-B[a]P-dG adducts (Figure 1.2) are bypassed with high fidelity and efficiency with a misincorporation frequency of \( 10^{-2} \) to \( 10^{-4} \) (Shen 2002). In addition, DinB has also been shown to accurately and efficiently bypass \( N^2 \)-(1-carboxyethyl)-2’-deoxyguanosine (\( N^2 \)-CEdG) adducts (Yuan 2008). \( N^2 \)-CEdG minor groove adducts are formed endogenously from methylglyoxal, which is a byproduct of glycolysis (Thornalley 1990; Thornalley 1996), and are found to be detected at one lesion per \( 10^7 \) nucleotides in human melanoma cells (Yuan 2008). Furthermore, it has been established that an \textit{E. coli} strain containing a deletion of the \textit{dinB} gene is sensitive to both nitrofurazone (NFZ) and 4-nitroquinoline-1-oxide (4NQO) (Jarosz 2006). Both of these DNA-damaging agents form \( N^2 \)-dG adducts \textit{in vivo} (Panigrahi 1990; Whiteway 1998). DinB shows accuracy, and a 15-fold increased proficiency of dCTP insertion opposite \( N^2 \)-furfuryl-dG, a \( N^2 \)-dG structural analogue formed from NFZ, than opposite undamaged DNA (Jarosz 2006). DinB has also been shown to accurately bypass \( N^2-N^2 \)-dG inter-strand crosslinks (ICLs) which can halt DNA replication (Kumari 2008). Inter-strand crosslinks are typically repaired by cooperation between homologous recombination and nucleotide excision repair (Cole 1973), but recent data have shown that TLS may also play a role in repair of ICLs (Kumari 2008). \( N^2 \)-dG adducts formed from \( \alpha,\beta \)-unsaturated aldehydes such as acrolein, which forms \( \gamma \)-hydroxypropanodeoxyguanosine (\( \gamma \)-HOP-dG), can form DNA-peptide crosslinks (Minko 2008). DinB has been further shown to bypass these acrolein-mediated adducts as well as the inter-strand crosslinks and the peptide conjugates formed as a result of these adducts (Minko 2008; Minko 2009).
Moreover, DinB has also been shown to bypass lesions that are formed from alkylating agents such as methyl methanesulfonate (MMS) (Bjedov 2007). A cluster of residues referred to as the ‘aromatic triad,’ Phe12, Phe13, and Tyr79, are important for survival of *E. coli* cells in the presence of MMS (Benson 2011). Interestingly, it was found that strains that contain single-point mutations in the *dinB* gene in the ‘aromatic triad’ residues show fewer MMS-induced mutants than nitrofurazone-induced mutants which suggests that these residues not only play a significant role in TLS, but also are involved in modulating the accuracy of DinB in bypassing specific lesions (Benson 2011).

DinB is also involved in bypass of various reactive oxygen species (ROS) leading to A:T → G:C transitions (Hori 2010). DinB has been shown to preferably insert dATP opposite 5-formyluracil (5-fodU) and 5-hydroxymethyluracil (5-hmdU); both dCTP and dATP opposite 7,8-dihydro-8-oxoguanine (8-oxo-dG) with low efficiency; and both dCTP and dTTP opposite 1,2-dihydro-oxoadenine (2-oxodA) (Hori 2010). In addition, DinB was found to incorporate 8-hydroxy-dGTP opposite adenine and 2-hydroxy-dATP opposite both guanine and thymine *in vitro* (Yamada 2006).

It has been further shown that DinB is capable of adding dGTP opposite 1,3-diaza-2-oxophenothiazine (tC) consistently but is incapable of continuing TLS beyond the modified base.
This is intriguing since DinB more strongly binds to DNA primer/template constructs that contain the tC analog than it binds to undamaged DNA primer/templates (Walsh 2011). However, it was also found that DinB inserts tC opposite G in the DNA primer/template and is capable of extending the new primer terminus, suggesting that DinB shows asymmetric discrimination against the modified DNA template and the incoming nucleotide (Walsh 2011).

The error frequency of DinB on DNA that is undamaged is approximately $2.1 \times 10^{-4}$ for generating frameshift mutations and about $5.1 \times 10^{-5}$ for generating base substitution mutations (Kobayashi 2002). DinB is also known to bypass abasic sites which cause -1 frameshift mutations (Kobayashi 2002; Maor-Shoshani 2003). This is carried out by a ‘dNTP-stabilized’ misalignment mechanism in which dNTP is placed correctly opposite a template base downstream rather than placing an incorrect nucleotide opposite the next available template base (Kobayashi 2002). More recently however, another mechanism involving template slippage provides an additional explanation for the generation of -1 frameshift mutations (Foti 2010). The template slippage mechanism causes single base deletions on DNA containing homopolymeric runs (Foti 2010). When UmuD$_2$ is bound to DinB, a non-slipped conformation is preferred which prevents the generation of frameshift mutations (Foti 2010).

DinB interacts with several other proteins in the *E. coli* cell. It is known that four other proteins/subunits interact directly with DinB and affect its replication efficiency, UmuD$_2$, RecA, NusA, and the β-clamp subunit (Wagner 2000; Godoy 2007; Cohen 2009; Cohen 2010a; Cohen 2010b). The presence of the *umuD* gene product improves the TLS ability of DinB and also reduces the number of -1 frameshift mutations by DinB *in vitro* (Godoy 2007). It has been suggested that both RecA and UmuD$_2$ decrease the -1 frameshift mutation activity of DinB because they reduce the size of the active site (Godoy 2007). Deletion of the *umuD* gene led to
an increase in the frequency of -1 frameshift mutations; however, the resistance of DinB to nitrofurazone did not change suggesting that the ability of DinB to perform TLS and the mechanism for generating -1 frameshift mutations are separable functions (Godoy 2007). Both forms of UmuD, the full-length UmuD$_2$ and the cleaved form, UmuD’ are capable of preventing DinB’s involvement in adaptive mutagenesis as well (Godoy 2007). UmuD also seems to regulate DinB activity by stimulating DNA replication past correct base pairs while preventing nucleotide extension by DinB when bases have been paired incorrectly (Godoy 2007; Foti 2010).

The extreme C-terminus of DinB (347-351) binds to the β-clamp subunit of the Pol III holoenzyme which contributes substantially to both enzyme processivity in TLS and dNTP-binding affinity (Becherel 1999; Wagner 2000; Lenne-Samuel 2002; Bunting 2003). A second binding site has also been discovered between residues 303-305 of DinB and residues located near the dimer interface of the β-clamp subunit (Bunting 2003). It is believed that the coordinated binding of DinB to β-clamp subunit increases the processivity of the enzyme and helps to position it correctly at the replication fork (Bunting 2003; Heltzel 2009; Heltzel 2012).

NusA, which is primarily involved in three phases of transcription: elongation, termination, and anti-termination (Greenblatt 1981; Farnham 1982; Liu 1996), has also been shown to interact with DinB, playing a similar role as RecA, and helping to recruit DinB to gaps in the DNA template strand during transcription-coupled TLS, when RNA polymerase is stalled by a lesion in DNA (Cohen 2009; Cohen 2010a). NusA has also been discovered to be necessary for adaptive mutagenesis by DinB (Cohen 2010b). It was hypothesized that NusA interacts with DinB such that DinB is recruited to damaged sections where the RNA polymerase is stalled (Cohen 2009). The exact binding site of DinB is uncertain, yet it is believed to be located on the C-terminal domain of NusA (Cohen 2009) (Nakamura 1986). The NusA-DinB interaction is
predicted to bridge the gap between replication-coupled TLS and transcription-coupled TLS (Cohen 2010a). This work demonstrates a crucial connection between replication and transcription, especially in the presence of DNA damage.

1.3  *E. coli* Pol V

*E. coli* Pol V is the second of the two Y-family polymerases found in *E. coli*. Pol V consists of two different protein subunits, one UmuD’ 2 dimer and UmuC, to form UmuD’ 3:C (Kato 1977; Bagg 1981; Woodgate 1989; Bruck 1996). Pol V was discovered in the late 1970s when the umuC gene was found to produce a protein that allowed *E. coli* cells exposed to UV light to survive (Kato 1977). The ability of UmuC to bypass UV-induced DNA adducts via TLS was not discovered until well after it was characterized as being involved in SOS mutagenesis. Several models were proposed over the course of the next few decades (Walker 1984; Bridges 1985a; Bridges 1985b; Rajagopalan 1992) until the late 1990s when it was confirmed that UmuC was in fact a DNA polymerase (Bruck 1996; Tang 1998; Reuven 1999; Tang 1999). This came to fruition when it was found that UmuC exhibited low-fidelity lesion bypass on its own, but its fidelity and efficiency increased in the presence of RecA, SSB, and UmuD’ (Tang 1998; Reuven 1999). Interestingly, UmuC maintained its ability to function despite the absence of the pol III holoenzyme subunit (Tang 1998; Reuven 1999).

Similarly to the regulation of DinB, the protein constituents of Pol V, UmuC and UmuD, are both regulated by the SOS response and are located within the same operon (Elledge 1983; Shinagawa 1983; Schlacher 2007). Also similar to DinB, upon induction of the SOS response, the expression of the umu proteins increases 10-fold with UmuC increasing from approximately 15 to 200 molecules and UmuD increasing from approximately 180-2400 molecules (Woodgate 1991). Primary DNA repair processes such as nucleotide excision repair typically remove a
lesion once it has formed in DNA (Courcelle 2001; Courcelle 2005); however, in the event
where nucleotide excision repair is unable to take place, replication will recover upon induction
of Pol V (Courcelle 2009). It has been shown that when \textit{umuC} is not present, there is moderate
DNA synthesis recovery and when the \textit{recJ} gene is not present, there is poor DNA synthesis
recovery (Courcelle 2005). When both \textit{umuC} and \textit{recJ} are absent, there is no recovery of DNA
synthesis. Interestingly, in the absence of \textit{recJ}, \textit{umuC} is required for recovery of DNA synthesis
(Courcelle 2005; Courcelle 2006).

The ability of Pol V to perform TLS is dependent on the formation of the UmuD'\textsubscript{2}C
complex and the presence of RecA (Reuven 1999; Schlacher 2005). Full-length UmuD is
involved in preventing UmuC from engaging in TLS and therefore preventing mutagenesis by
UmuC. It has been found that by changing the Ser60 residue to alanine in full-length UmuD,
autocatalytic cleavage of UmuD\textsubscript{2} can be prevented (Nohmi 1988). Full-length UmuD is involved
in prevention of mutagenesis; thus it was found that UmuD\textsubscript{2} harboring this mutation results in a
large reduction of UV-induced mutagenesis (Battista 1990; Opperman 1999; Beuning 2006b).
Cells that contain UmuD\textsubscript{2}-Ser60Ala with UmuC experience sensitivity to UV light relative to
cells that contain wild-type UmuD and UmuC (Battista 1990; Beuning 2006b). Molecular
interactions between UmuC and UmuD have been difficult to determine; however, through
immunoprecipitation, glycerol gradient analysis, and yeast two-hybrid assays, the interaction of
UmuD' and UmuC has been confirmed (Woodgate 1989; Jonczyk 1996). The interaction of full-
length UmuD and UmuC has also been confirmed using affinity chromatography and velocity
sedimentation analysis (Woodgate 1989). The physical interaction between UmuD' and UmuC
consists of one UmuC protein bound to a dimeric UmuD' (Woodgate 1989). It was shown that
UmuD' binds to the 25 amino acid C-terminal end of UmuC (Woodgate 1989; Sutton 2001).
In general, Pol V is capable of bypassing lesions formed in DNA that DinB is incapable of bypassing. Pol V is capable of bypassing lesions caused from exposure to UV light such as thymine-thymine (T-T) \textit{cis-syn} cyclobutane pyrimidine dimers (CPD) and T-T (6-4) photoproducts (Figure 1.3) (Becherel 1999; Tang 2000; Fujii 2004a; Fujii 2004b). In addition to bypassing lesions from UV light, Pol V is also capable of bypassing abasic sites as well as $C^{8}$-dG lesions such as $N$-2-acetylaminofluorine ($C^{8}$-AAF) (Becherel 1999; Tang 2000; Maor-Shoshani 2003; Fujii 2004a; Fujii 2004b). Pol V bypasses lesions caused by UV light with greater fidelity when in the presence of the beta clamp subunit, the RecA/ssDNA nucleoprotein filament, and the single stranded binding protein (SSB) (Tang 2000). Similarly to DinB, Pol V sometimes performs TLS mutagenically. For example, Pol V inserts dGTP opposite the 3’T in T-T (6-4) photoproducts instead of dATP by a six-fold greater frequency (Tang 2000). However, Pol V is also known to bypass certain lesions with high accuracy. For instance, Pol V faithfully inserts dATP opposite T-T CPDs (Tang 2000) as well as dCTP opposite $C^{8}$-AAF (Fujii 2004b). Pol V has also been reported to bypass $N^{2}$-benzo[a]pyrene-dG adducts, $N^{6}$-benzo[a]pyrene-dA adducts and adducts resulting from oxidation (Shen 2002; Seo 2006; Neeley 2007). When replicating damaged DNA, Pol V bypasses with a much lower fidelity than the Pol III HE and even lower fidelity than DinB, with error frequencies of $10^{-3}$ to $10^{-4}$ (Tang 2000).
Figure 1.3: Shown are the DNA lesions caused by UV light that UmuD’\textsubscript{2}C is capable of bypassing.

In addition to facilitating the autocatalytic cleavage of the LexA repressor protein as well as facilitating the cleavage of full-length UmuD\textsubscript{2} to UmuD’\textsubscript{2}, RecA also plays a direct role in TLS (Schlacher 2006b; Patel 2010). RecA has been determined to both stimulate nucleotide insertion as well as extension (Pham 2002). In one model, it is argued that Pol V and two RecA molecules form a complex that essentially activates Pol V for TLS in the presence of ATP (Schlacher 2005; Sweasy 2005). The RecA/ssDNA nucleoprotein filament transfers RecA and an ATP from the 3’ end of the filament in \textit{trans} to Pol V which activates Pol V (Schlacher 2006a; Jiang 2009). It is also argued that the RecA/ssDNA nucleoprotein filament acts in \textit{cis} rather than \textit{trans} and facilitates TLS and the activation of Pol V (Fujii 2009).

In addition to RecA, it is also reported that the beta processivity clamp and even the gamma clamp loader may increase the processivity of Pol V by allowing the enzyme to remain bound to the DNA as well as provide additional stability (Pham 2001). It has been argued that the beta clamp will increase processivity approximately 3-fold or 5-fold in one study (Maor-Shoshani 2002) and about 100-fold in another study (Fujii 2004b).
1.4 DinB and UmuC orthologs in other bacteria

Although the Y-family polymerases of *E. coli* have been studied extensively, Y-family polymerases are present throughout all domains of life. The following will discuss recent developments in the study of Y-family polymerases in other species of eubacteria. The bacteria known to contain Y-family polymerases or potential Y-family polymerases are *Bacillus anthracis*, *Bacillus halodurans*, *Bacillus subtilis*, *Citrobacter freundii*, *Escherichia coli*, *Enterococcus faecalis*, *Lactococcus lactis*, *Mycoplasma genitalium*, *Morganella morganii*, *Mycoplasma pneumoniae*, *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, *Neisseria meningitides*, *Pseudomonas aeruginosa*, *Pasteurella multocida*, *Pseudomonas syringae*, *Streptomyces coelicolor*, *Salmonella enterica*, *Shigella flexneri*, *Serratia marcescens*, *Streptococcus pneumoniae*, *Salmonella typhi*, *Salmonella typhimurium*, *Ureaplasma urealyticum*, and *Vibrio cholera* (Ohmori 2001).

Y-family polymerases in the bacterium *Bacillus subtilis* have been found to be involved in mutagenesis (Sung 2002). The Y-family polymerases that are found in *B. subtilis* are UvrX, YqjH, and YqjW which are homologs of the *E. coli* Y-family polymerases DinB and UmuD’C (Ohmori 2001). Despite the homology between the Y-family polymerases in *B. subtilis* and *E. coli*, there are significant differences between the types of polymerases as well as the ways in which their expression is regulated through the SOS response (Winterling 1997; Barnes 2002). UvrX has been reported to be a protein from the prophage SPβ (Kunst 1997). UvrX has 25% sequence identity to the *E. coli* Y-family DNA polymerase DinB (Kunst 1997). The YqjH protein has 26% sequence identity to *E. coli* DinB and the YqjW protein has 26% sequence identity to *E. coli* UmuC. It has been suggested that these proteins are members of the Y-family of DNA polymerases (Permina 2002) and are part of the UmuC gram-positive family of Y-
family polymerases (Ohmori 2001). It was reported that inactivation of the \textit{yqjH} and \textit{yqjW} genes results in increased UV sensitivity and decreases the frequency of UV-induced mutagenesis (Sung 2003). Recently, it has been found that YqjH and YqjW are in fact Y-family DNA polymerases that protect sporulating cells of \textit{B. subtilis} (Rivas-Castillo 2010). Deletion of \textit{yqjH} and \textit{yqjW} genes decreased sporulation efficiency as well as increased sensitivity to chemical mutagens such as hydrogen peroxide, tert-butylhydrogenperoxide (t-BHP), mitomycin-C (M-C), and UV-C radiation (Rivas-Castillo 2010). It was concluded that YqjH and YqjW proteins are involved in TLS in sporulating \textit{B. subtilis} cells and cause spontaneous mutagenesis which is necessary for the sporulation process (Rivas-Castillo 2010).

\textit{Mycobacterium tuberculosis} contains Y-family DNA polymerases homologous to \textit{E. coli} DinB. There are two known Y-family DNA polymerases in \textit{M. tuberculosis}, both of which belong to the DinB superfamily of Y-family polymerases identified as DinB1 (or DinX) which is encoded by the gene \textit{Rv1537} and DinB2 (or DinP) which is encoded by the gene \textit{Rv3056} (Cole 1998). It was found that these proteins contain sequence similarity to their homologs in \textit{E. coli} (Wagner 1999) as well as in \textit{Pseudomonas aeruginosa} (Sanders 2006) leading to the presumption that DinB1 and DinB2 both have DNA polymerase activity. Unlike Y-family polymerases from \textit{E. coli} and most other eubacteria, DinB1 and DinB2 expression is not dependent on the RecA protein, the SOS response, or even the existence of damaged DNA (Brooks 2001; Rand 2003; Boshoff 2004; Kana 2010). In contrast to \textit{E. coli} Y-family polymerases, DinB1 and DinB2 are regulated by separate mechanisms whereby DinB1 is expressed in pulmonary tuberculosis (Rachman 2006) and DinB2 is expressed upon exposure to novobiocin (Boshoff 2004). This study has determined that these DinB homologs in \textit{M. tuberculosis} are therefore not upregulated upon introduction of DNA damage as in other
organisms; however, it was seen that the presence of DNA damage does result in the upregulation of C-family DNA polymerase DNAE2 (Boshoff 2003). DNAE2 is therefore predicted to play the primary role for adaptive mutagenesis in *M. tuberculosis* rather than the DinB homologs (Boshoff 2003).

The bacterium *Mycobacterium smegmatis* also contains sequence homologs to *E. coli* Y-family polymerases. It was found that the genome of *M. smegmatis* contains three DinB homologs encoded by the genes *msmeg*_1014, *msmeg*_3172, *msmeg*_6443 according to the KEGG PATHWAY Database (Kanehisa 2010). Interestingly, the key residues necessary for functional polymerase activity are conserved in the *msmeg*_1014 (also known as MsDpo4) (Sharma 2012). It has been shown that MsDpo4 is capable of performing template-dependent nucleotide insertion and is also capable of promoting mismatches on undamaged DNA templates (Sharma 2012). In addition, MsDpo4 has been shown to promote G:T and T:G mismatches more frequently than other DinB polymerases of other organisms showing that it has the ability to increase the frequency of mutations and is involved in adaptive mutagenesis (Sharma 2012).

Y-family DNA polymerase homologs are also present in species of the bacterial genus *Pseudomonas* (Ohmori 2001). *Pseudomonas aeruginosa* contains a homolog of *E. coli* polymerase IV known also as DinB (PaDinB) and is lacking in intrinsic proofreading capabilities (Sanders 2006). Further analysis of this polymerase unveiled that PaDinB promotes C to A transversions as well as induces -1 frameshift mutations (Sanders 2006). Strains lacking the *dinB* gene are sensitive to the DNA-damaging agents nitrofurazone and 4-nitroquinoline showing that PaDinB most likely plays a role in TLS similar to that of *E. coli* DinB (Sanders 2006). A DinB homolog was also found in *Pseudomonas putida* which was shown to be involved in 1-base pair
deletions in starving cells yet is also reported to be expressed in a RecA-independent process (Tegova 2004).

The purpose of this work is to understand better DNA damage specificity of DinB. We seek to determine the ability of DinB to bypass certain types of lesions while remaining inactive against others. Through methods of directed evolution (Chapter 2), we attempt to acquire a variant of DinB capable of bypassing lesions caused by UV light similarly to UmuD’2C. By mutating the DinB gene and screening these mutations by performing a UV selection assay, we examine potential variants that are able to bypass lesions caused by UV light. Through alanine screening of the loop 1 region of DinB (Chapter 3), we demonstrate the ability of certain variants to exhibit more or less TLS activity against damaged and undamaged DNA as compared to wild-type DinB. Finally, using molecular modeling tools in the YASARA suite of programs (Chapter 4), we explore the consequences of a thymine-thymine cyclobutane pyrimidine dimer, one of the lesions caused by UV light, entering the active site of DinB. The findings presented here contribute to our understanding of damage specificity as well as to our understanding of Y-family polymerases as a whole.

1.5 References


development of resistance to 5-nitrofuran derivatives in Escherichia coli."

DinR, the Bacillus subtilis SOS repressor." Journal of Bacteriology 179(5): 1698-1703.


analysis of the mutational spectrum in Escherichia coli rpoB to assess the role of


Escherichia coli: purification and interaction with UmuD and UmuD'." Proc Natl Acad

(2006). "Involvement of Y-family DNA polymerases in mutagenesis caused by oxidized

bypass of N-2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase in vitro
Chapter 2: Understanding DinB specificity through directed evolution

2.1 Introduction

Y-family DNA polymerases are a family of DNA polymerases capable of performing translesion synthesis (TLS). Various agents such as UV light or toxic chemicals can cause damage to DNA, forming lesions (Lindahl 1993). Ordinary replicating DNA polymerases such as DNA Pol III in *E. coli* do not have the ability to replicate past these lesions, which can inhibit replication (Radman 1975; Ohmori 2001). Therefore, biological systems have evolved to accommodate these lesions in various manners. One of these pathways involves employing Y-family polymerases to perform TLS (Radman 1975).

*E. coli* contains two of these polymerases called DinB (Pol IV) and UmuD’₂C (Pol V) (Ohmori 2001). One characteristic of these polymerases, as well as of Y-family polymerases in other organisms, is their specificity for certain types of DNA damage. For instance, DinB is specific for *N₂*-dG lesions, namely *N₂*-furfuryl-dG (Jarosz 2006), which is caused by exposure to nitrofurazone and 4-nitroquinoline (Panigrahi 1990; Whiteway 1998), *N₂*-benzo[a]pyrene-dG (Shen 2002; Seo 2006), and *N₂*-1-carboxyethyl-2’-dG (Yuan 2008). UmuD’₂C on the other hand is specific for lesions caused by UV light such as thymine-thymine (T-T) cyclobutane pyrimidine dimers and T-T (6-4) photoproducts (Becherel 1999; Tang 2000; Fujii 2004a; Fujii 2004b).

All Y-family polymerases have five conserved sequence motifs as well as four characteristic domains named according to the parts of a human right hand, namely: the thumb, palm, finger, and little finger domains (Friedberg 2001). Comparing the gene sequences of each of these Y-family polymerases in *E. coli*, all of the characteristic domains except for the little finger domain were found to be homologous (Ling 2001). Little to no homology was found
between the gene sequences of DinB and UmuC in the little finger region (Ling 2001). This finding led to the hypothesis that the little finger domain could be responsible or partially responsible for the specificity of these polymerases for certain types of DNA lesions.

Previously, it has been shown that by using directed evolution techniques, the substrate specificity of UmuD’2C could be altered to allow bypass of substrates normally bypasses by DinB. More specifically, mutations in UmuD’2C induced by hydroxylamine mutagenesis yielded a specific phenotype, in which *E. coli* strains lacking DinB were resistant to nitrofurazone (Tehrani unpublished). Hydroxylamine induces G:C to A:T mutations by causing a direct change in base pairing by favoring a tautomeric shift (Singer 1982). In this process, the amino group on cytosine is replaced with a hydroxyamino-group creating a base that acts like a uracil (Singer 1982). Adenine would be paired opposite uracil during a round of replication and followed by base excision repair which would replace uracil with thymine (Singer 1982).

Therefore, the main goal of this research is to acquire variants of DinB via directed evolution methods that will allow it to confer UV resistance. We are primarily using hydroxylamine mutagenesis among a few other methods to induce mutations in the *dinB* gene. Employing UV selection assays, we are screening for potential DinB variants that show resistance to damage caused by UV light. We will then perform further characterization of these variants to determine their proficiency at UV-induced lesion bypass and compare any mutations with native DinB in order to better understand its DNA substrate specificity.
2.2 Materials and Methods

a. Strains and Plasmids

The strains PB102 (ΔumuC ΔrecJ) and PB103 (ΔumuDC ΔrecJ) were used for all experiments.

Table 2.1: List of plasmids used in UV selection assays.\(^1\)

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Encodes</th>
<th>Antibiotic resistance gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGY9738 WT</td>
<td>(untreated) UmuD’₂C</td>
<td>Spectinomycin (Sommer 1993)</td>
</tr>
<tr>
<td>pGB2 EV</td>
<td>Empty vector</td>
<td>Spectinomycin</td>
</tr>
<tr>
<td>pWSK30 EV</td>
<td>Empty vector</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>pYG768 WT</td>
<td>(untreated) DinB wild-type</td>
<td>Ampicillin (Nohmi 1988)</td>
</tr>
<tr>
<td>pCC WT</td>
<td>DinB with UmuC C-terminus</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>pCCLF WT</td>
<td>DinB with UmuC C-terminus + UmuC little finger</td>
<td>Ampicillin</td>
</tr>
</tbody>
</table>

\(^1\)Throughout the text, the notation 1M and 1.5M indicates plasmids treated with 1 M or 1.5 M hydroxylamine, respectively.

In addition, two types of DinB chimeras were prepared by Dr. Beuning, one in which the C-terminus from the UmuC gene was spliced onto the end of the DinB gene, and another in which the little finger of DinB was removed and the C-terminus plus little finger of UmuC was spliced onto the end of the DinB gene (Figure 2.1). Figure 2.1 outlines the preparation of these chimeras.
Figure 2.1: The above diagram outlines the chimeras used that were created by Dr. Beuning. The top two diagrams show the DinB and UmuC genes as they normally are with the locations where restriction enzymes were used to cleave the necessary sections of DNA. The bottom two diagrams show the final constructs: DinB with the UmuC C-terminus and DinB with the UmuC C-terminus and little finger.

b. Plasmid Preparation

For the maxiprep procedure, a culture (5 mL Luria Broth (LB) supplemented with appropriate antibiotic, 60 µg/mL spectinomycin or 100 µg/mL ampicillin) was prepared from a single colony on a streaked plate (usually plated from frozen stocks on Luria Broth (LB) agar containing appropriate antibiotic). The culture was allowed to incubate for 8 h at 37 °C with shaking. After 8 h of incubation, overnight cultures were set up using 250 mL of LB, the appropriate amount of antibiotic and 500 µL of subculture. This was allowed to incubate overnight at 37 °C with shaking. The maxi-prep was performed using the QIAfilter™ Plasmid Maxi Kit (Qiagen). All procedures were followed according to the manufacturer’s instructions.

For the miniprep procedure, a single colony was used to inoculate 5-10 mL LB culture with
appropriate antibiotics for about 24 h at 37 °C while shaking at 200-250 rpm. The miniprep was carried out according to the instructions provided in the GeneJET™ Plasmid Miniprep Kit (Fermentas).

c. **Transformations**

Competent cells were prepared as described (Beuning 2006). The transformations were conducted as follows. A 50 µL aliquot of cells and 1.0 µL DNA (~30-35 ng) were used per transformation (all plasmid concentrations scaled to ~30-35 ng/µL). In micro-centrifuge tubes, the DNA and cells were mixed and set on ice for 10 min. They were then heat shocked by incubating at 37 °C for 5 min. Next, they were returned to ice for 10 min. An aliquot of LB (500 µL) without antibiotics was added to the mixture of DNA and cells. They were then allowed to recover at 37 °C for 1 h and then plated on selective media.

d. **Hydroxylamine Mutagenesis**

**Method 1 Mutagenesis:**

The 1 M hydroxylamine solution was prepared in the following way. Hydroxylamine hydrochloride (H$_3$NO-HCl, 0.35 g, Acros Organics), 450 µL 5 M NaOH, and 4.55 mL ice-cold sterile MilliQ water were combined to make a 1 M stock of hydroxylamine. The pH was maintained near 6.7 and the solution was stored at 4 °C. In a microcentrifuge tube, 480 µL 1 M hydroxylamine solution, 480 µL De-Ionized water, 240 µL 0.5 M sodium phosphate buffer (pH 6.5) prepared by Lisa Hawver, and 1.5 µg of plasmid (pYG768 WT or other) DNA were combined. The tube was allowed to incubate for 3 h in a 65 °C aluminum bead bath. The DNA was purified using the Fermentas GeneJet™ PCR Purification Kit as in Method 2 mutagenesis.
Method 2 Mutagenesis:

The hydroxylamine solution was prepared in the same way as in the Method 1 mutagenesis procedure described in the previous paragraph. The DNA was treated with hydroxylamine by incubating 10 µg (36 µL) DNA with 500 µL hydroxylamine solution at 37 °C overnight (~20-22 h) in a micro-centrifuge tube. The DNA was purified using the Fermentas GeneJet™ PCR Purification Kit. The purification procedure was followed according to the manufacturer’s instructions.

Various alterations were made in the hydroxylamine mutagenesis Method 2 protocol in order to maximize the probability of generating a greater frequency of mutations. We increased the concentration of hydroxylamine solution from 1 M to 1.5 M and 2 M as well as increased incubation time from 20 h to 25 h and 40 h. Other variations have included performing a second hydroxylamine mutagenesis on DNA that has already been mutated and acquiring several different stocks of hydroxylamine hydrochloride. Finally, the last variation of this procedure that we performed was to expose PB102 E. coli cells that have been transformed with pYG768 WT (DinB) to titrated amounts of hydroxylamine solution directly during the subculture step. These experiments were performed similarly to a survival assay except with the UV exposure and plating steps followed according to the UV selection assay protocol (see sections f and g).

e. Randomized Primers

Randomized primers were designed to induce mutations in the loop regions of DinB (Figure 2.2).
Figure 2.2: Randomized primers on the loop regions of the DinB protein designed to induce mutations at the highlighted regions. The numbers above the portions of highlighted sequence represent the amino acid position in the primary structure. N indicates a mixture of all four nucleotides; S indicates a mixture of C, G, and T nucleotides.

Using the standard QuikChange PCR protocol (Agilent), these primers were used to induce mutations in pYG768 WT (DinB) generating a PCR product with a library of mutations. DpnI digestion was employed to remove the parental plasmid encoding for DinB followed by transformation into XL1Blue cells and finally extracted once again using the Miniprep protocol (see section b). The resulting product was used in UV selection assays similarly to products from hydroxylamine mutagenesis (see section g for UV selection assay protocol).

f. UV Selection Assays

All experiments were performed in a similar manner with differences only in UV exposure levels and plasmids used in each experiment. All plasmids were either mini-prepped in the case of controls, or maxi-prepped followed by one of the mutagenesis methods. All of the plasmids were run on a 1.1% agarose gel prior to performing any experiments in order to normalize all of the plasmid concentrations. All gels were run using a Fisher Scientific FB300 electrophoresis system and a Fisherbiotech FB-SB-710 gel box. A Quick-Load 1Kb DNA ladder N0468S (NEB), 50 µg/mL was used as a control for all gels run. All gels were loaded with 10
µL of sample with 5 µL plasmid and 5 µL DNA dye and were run for approximately 70 min at 130 volts. DNA was visualized with ethidium bromide and gels were analyzed using ImageQuant after scanning on a Storm 860 imager. The recorded values for the volume of each band in the gel were referenced to the ladder control and a relative concentration was determined for each plasmid (in ng/µL). Once the plasmid concentrations were normalized to relatively equal concentration based on the gels, they were transformed into PB102 or PB103 cells. Once transformed, the cells were washed with 0.85% saline and exposed to UV light in suspension at the appropriate levels of exposure, usually 0, 25, 50, 75, and 100 J/m². The samples were then plated on agar plates with appropriate antibiotic and incubated at 37 °C overnight.

**g. UV Survival Assays**

UV survival assays were carried out in a similar way to the UV selection assays with a few notable changes. The assay begins with making 5 mL overnight cultures of appropriate controls (pYG768 WT, pGY9738 WT, and pWSK30 EV) as well as any selected hits from the UV selection assay to be validated. From the overnight cultures, 140 µL were transferred to 7 mL of fresh LB with appropriate antibiotics. These subcultures were grown for approximately 1 h to achieve an OD

\[ \text{OD}_{600\text{ nm}} \approx 0.2-0.4 \]

The samples were then normalized by OD and transferred to microcentrifuge tubes. The samples were washed with 0.85% saline and resuspended in 1 mL 0.85% saline. Next, the samples were transferred to small (50 mm) culture dishes and exposed to UV light at exposures of 0, 25, 50, 75, and 100 J/m². At each time point samples were removed and placed in 96 well plates. Then, 10-fold serial dilutions were performed using 180 µL saline and 20 µL of sample to achieve a final dilution of 1:100000. The samples (10 µL each) were then plated on large LB-agar culture plates using a multi-channel pipette. The samples were allowed to incubate at 37 °C overnight.
h. Other Methods

In addition, error-prone PCR was used as an alternative directed evolution method. We adapted an error-prone PCR method from Rasila et al., using the nucleotide analogues dPTP and 8-oxo-dGTP (Kim 2009; Rasila 2009). A 50 μL PCR reaction was prepared using 5 units of Taq polymerase (from Fisher Scientific) with 5 μL of 10x buffer B (500 mM KCl, 100 mM Tris-HCl), 3 μL of 25 mM MgCl₂, 2 μL of 5 mM dNTPs, 3 μL of 5 μM primers, 2.5 μL of 81 ng/μL pYG768 WT DNA plasmid, 2 μL of 2.5 μM each of the nucleotide analogues 2’-dPTP and 8-oxo-dGTP, and 26.5 μL of water. The PCR amplification method was carried out using a PTC-200 Peltier Thermal Cycler by MJ Research starting with step 1 at 95 °C for 5 min followed by 95 °C for 30 sec (step 2), 60 °C for 1 min (step 3), and 68 ° for 2 min (step 4). Steps 2 through 4 were repeated 34 times (step 5) and ending at 4 °C (step 6).

Furthermore, two mutator cell strains were used to induce mutations on the pYG768 WT plasmid once it was transformed into the cells. One of these strains is known as the XL1Red mutator strain which lacks three of the primary DNA repair pathways: mutS which corresponds to the mismatch repair pathway, mutT which corresponds to the oxo-dGTP pathway, and mutD which corresponds to the 3’-5’ exonuclease subunit of pol III (Rasila 2009). The other strain is referred to as the dut- ung- strain because it lacks dUTPase, which is an enzyme that degrades dUTP in the cell (Ladner 1997), and uracil N-glycosylase, which excises uracil bases (Pearl 2000).

The standard protocol used for using mutator cell strains is as follows. pYG768 WT DinB plasmid that has been mutated by hydroxylamine mutagenesis was transformed into the mutator cell strain (XL1Red or dut- ung-) and allowed to grow overnight on plates. The resulting
colonies were pooled in LB and allowed to grow in suspension overnight. The plasmids were then extracted the following day using the miniprep procedure. The resulting library was used as an experimental sample in UV selection assays.

2.3 Results

a. DinB UV Selection Assays

Several DinB UV selection assays were performed with plasmids that had been mutagenized with hydroxylamine hydrochloride. Many variations to the standard protocol were employed as outlined in section d of Materials and Methods. Only results from a few of the assays are described here.

Assays 1, 2, and 3 were identical in all parameters and the experimental plasmids used had been incubated with 1 M hydroxylamine hydrochloride which followed the general protocol outlined in section d of Materials and Methods. The only alteration in assay 3 was an additional UV exposure level of 65 J/m². In assay 4, DinB WT plasmids were exposed to hydroxylamine hydrochloride according to both method 1 and method 2 as outlined in section d of Materials and Methods. In addition, the DinB chimeras (described in section a of Materials and Methods) were employed here as well. DinB-CC and DinB-CCLF were exposed to hydroxylamine according to both method 1 and method 2 similarly to DinB WT. In assay 5, the same experiment was repeated as in assay 4 except with a newly purchased hydroxylamine hydrochloride (Acros Organics). In assay 6, the same parameters were used as in assays 4 and 5 with the same mutated plasmids prepared for assay 5; however, additional experimental plasmids were used (pYG768 WT, CC-WT, and CCLF-WT) that had been exposed to higher concentrations of hydroxylamine hydrochloride (1.5 M and 2 M).
In assay 8, a new UV selection method was employed in which the transformed cells were washed with 0.85% saline and exposed to UV light in suspension followed by plating. In the previous assays, the transformed cells were plated first, and then the plates were exposed to UV light. This was changed because it was possible that the cells were not getting an accurate dose of UV light when they were plated first. Washing with saline and exposing the cells while suspended in saline should allow for all the cells in suspension to get an equal and accurate dose of UV. Assay 8 utilized the same experimental plasmids used in the previous three assays. Assay 9 was performed in the exact same manner as assay 8. The following results are from DinB UV selection assay 9 (Figure 2.3), which of all the assays yielded the most positive results.

In assay 9, there were 13 colonies from the pYG768 1M plasmid that survived at 60 J/m$^2$ versus only 4 colonies from the pYG768 WT control plasmid surviving at 60 J/m$^2$. Likewise, there were 12 colonies from the CC 1M plasmid that survived at 60 J/m$^2$ versus 2 colonies from the control CC WT. There was a 3-fold increase in growth for the pYG768 1M versus the wild-type and a 6-fold increase in growth for the CC 1M versus the CC “wild-type”. At the 90 J/m$^2$ UV exposure level, both the pYG768 1M and CC 1M had 1 colony versus 0 colonies for the pYG768 WT and CC WT plasmids. It can therefore be seen that there is certainly better survival/growth of the plasmids that were treated with hydroxylamine hydrochloride at high UV exposure levels indicating the possibility that there are mutations in the DinB gene that confer resistance to UV.
Figure 2.3: DinB UV selection assay 9 yielded some positive results in which several colonies from the experimental plasmids pYG768 1M and CC 1M have survived at high exposures of UV light. Interestingly, the wild-type control plasmids for both the DinB WT and the DinB-CC WT (pYG768 WT and CC WT) did not have any surviving colonies at 90 J/m² whereas pYG768 1M and CC 1M did survive. These ‘hits’ were then further characterized using the UV survival assay (described below).

b. DinB UV Survival Assays

Several UV survival assays were performed with potential hits taken from UV selection assays that yielded positive results. All of the UV survival assays were performed according to the method described in section g of Materials and Methods. Any colonies from hydroxylamine treated plasmids that survived at high levels of UV exposure in the UV selection assays were further validated using this assay. Most of the assays that were performed did not yield significant survival of any of the hits. The following results are from the UV survival assays that were performed to further validate the hits taken from UV selection assay 9 described in the previous section.

From the several hits acquired in UV selection assay 9, four UV survival assays were performed to validate these hits. Data from three of these assays are shown here. In survival
assay 1 (Figure 2.4), two of the colonies, both taken from the culture plate that contained the pYG768 plasmids treated with 1 M hydroxylamine and exposed to 60 J/m² of UV, showed considerable survival. Hit P9160-4 showed almost as much survival at 100 J/m² as the UmuC wild-type control (pGY9738).

**Figure 2.4:** One of the three survival assays performed using the hits from UV selection assay 9 shows that two of the potential mutants (p9160-2 and p9160-4) survive roughly 10-fold more than the DinB WT (pYG768).

In survival assay 2 (Figure 2.5), three of the colonies, all taken from the culture plate that contained the CC plasmids treated with 1 M hydroxylamine and exposed to 60 J/m² of UV, showed approximately 10-fold more survival as compared to the DinB WT.
Figure 2.5: The second of three survival assays performed using the hits from UV selection assay 9 shows that three of the potential mutants (c9160-4, c9160-3, and c9160-2) survive roughly 10-fold more than the DinB WT (pYG768).

In survival assay 1, two of the colonies, both taken from the culture plate that contained the pYG768 plasmids treated with 1 M hydroxylamine and exposed to 60 J/m² of UV, showed remarkably good survival as compared to the DinB WT. In survival assay 3 (Figure 2.6), hit p9160-8 showed greater than 10-fold better survival as compared to the wild-type DinB at the 100 J/m² level of UV exposure. On the other hand, hit p9160-11 showed considerable survival of about 10-fold higher than wild-type at the 50 and 75 J/m² levels of UV exposure.
Figure 2.6: One of the three survival assays performed using the hits from UV selection assay 9 shows that 1 of the potential mutants (p9160-8) is surviving roughly 10-fold more efficiently than the DinB WT (pYG768).

Most of the survival assays shown were repeated again once the plasmids were isolated and re-transformed into fresh PB102 or PB103 cells. The results of these subsequent experiments showed no hits surviving better than the DinB wild-type. One theory is that the UV light, which is itself a mutagen, induced mutations elsewhere in the genome allowing these ‘hit’ colonies to survive high levels of UV exposure when in fact their survival had nothing to do with the *dinB* gene. Another theory is that these surviving colonies were merely false positives. The plasmid from hit p9160-2 was isolated and sequenced for mutations, but the results showed it was wild-type.

2.4 Discussion

Our results have shown that although this directed evolution method shows promising results, no variants were acquired that yielded mutations in the *dinB* gene that would allow an *E. coli* cell to confer UV resistance via TLS. Looking at the results of the UV selection assays and
survival assays presented in the previous section, it would appear that the experiments yielded several false positives. This lead to the idea that hydroxylamine mutagenesis was not a suitable method for directed evolution and so other methods needed to be considered. See chapter 4 for other directed evolution methods.

It was further considered that upon replacement of the amino group with a hydroxyamino-group on cytosine to cause it to ‘act’ like uracil, base excision repair would first remove uracil and restore the original cytosine, eliminating the possibility of mutation. We believed that the low frequency of mutations occurring could be a result of the cell repairing the mutations in the DinB gene caused by hydroxylamine hydrochloride. To account for this issue, the dut- ung- cell strain was employed. The dut- ung- strain lacks dUTPase, which is an enzyme that degrades dUTP in the cell (Ladner 1997), and uracil N-glycosylase, which excises uracil bases (Pearl 2000). It was thought that by first transforming the hydroxylamine treated plasmid into this deficient cell strain, it would fix (i.e. make permanent) the mutations in the dinB gene on the plasmid. The plasmids were then isolated by plasmid miniprep, followed by re-transformation into PB102 or PB103 cells and performance of a UV selection assay to select for DinB variants conferring UV resistance. Three such experiments were performed yielding no positive results, so this method was discontinued.

2.5 References


Chapter 3: Determining Activity and DNA Damage Specificity of the DinB Loop 1 Region

3.1 Introduction

All forms of life are exposed to various sources of chemical toxins and physical agents that can cause damage to DNA. A few of these types of agents are UV light, which induces thymine-thymine cyclobutane pyrimidine dimers and (6-4) photoproducts (Becherel 1999; Tang 2000; Fujii 2004a; Fujii 2004b), diesel exhaust and cigarette smoke, which induce a benzo[a]pyrene adduct at the $N^2$ position of deoxyguanosine (Shen 2002; Seo 2006), and nitrofurazone and 4-nitroquinoline oxide (Panigrahi 1990; Whiteway 1998), which induce adducts at the $N^2$ position of deoxyguanosine (Jarosz 2006). Most organisms employ the Y-family polymerases to bypass DNA adducts in a process known as translesion synthesis (TLS) (Radman 1975). In *E. coli*, there are two of these polymerases that are induced by the SOS response to DNA damage and are capable of bypassing specific types of DNA lesions (Ohmori 2001). For example UmuD’C (Pol V) is capable of bypassing adducts caused by UV light and DinB (Pol IV) is capable of bypassing deoxyguanosine adducts at the $N^2$ position (Becherel 1999; Tang 1999; Wagner 1999; Jarosz 2006).

DinB has been shown to be conserved throughout all domains of life and has the ability to copy damaged DNA templates, particularly $N^2$-dG adducts (Wagner 1999; Ohmori 2001; Jarosz 2006). DinB has also been shown to have higher fidelity when copying certain damaged templates than undamaged templates. For example, in comparison to undamaged DNA, DinB has a 15-fold preference for replicating past $N^2$-furfuryl-dG (Jarosz 2006). It is widely known that a single amino acid change in the peptide sequence of a protein can have a significant effect on that protein. This concept holds true for *E. coli* DinB as well, as it has been shown that by
changing the phenylalanine residue at position 13 to valine that the ability of DinB to perform TLS on $N^2$-furfuryl-dG was compromised (Jarosz 2006). Yet, the ability of DinB to bypass undamaged DNA is not affected significantly (Jarosz 2006). It was also shown that Phe13 is in fact the steric gate residue, which functions to prevent the incorporation of ribonucleotides (Jarosz 2006). Many other studies have been performed on DinB variants to determine mutation frequency, fidelity, or damage specificity. For instance, Wagner, et al. showed that cells harboring wild-type DinB have a mutation frequency of $68.5 \times 10^{10}$ with cells harboring variants including Asp8, Arg49, Asp103, and Glu104 exhibiting lower frequencies (Wagner 1999).

The purpose of this research is to identify amino acid residues located on the loop 1 region of DinB that affect its activity, particularly in terms of bypass efficiency and damage preference. The Ser42 and Val40 residues (both on the loop 1 region) were previously suggested to be important residues using the THEMATICS and POOL program developed by Ondrechen, et al (Ondrechen 2001). It has been proposed that the remainder of the residues on loop 1 may also be important for efficiency and specificity. Using alanine scanning mutagenesis and loop swapping the ability of DinB variants in this region to incorporate nucleotides opposite $N^2$-furfuryl-dG and on undamaged DNA templates was determined.

3.2 Materials and Methods

a. Expression and Purification of Wild-type DinB and Variants

The first step is to develop the desired variant via site directed mutagenesis. These variants were made by designing the appropriate primers and using a QuikChange kit (Agilent) as well as the appropriate DinB expression plasmid pDFJ1 by Jason Walsh (Beuning 2006). The mutations were then confirmed by DNA sequence analysis (MGH DNA Core or Macrogen).
Next, BL21 DE3 pLysS cells were transformed with the plasmids containing the DinB mutations for expression. For growth, the cells were incubated in LB supplemented with ampicillin (100 μg/mL) overnight. The overnight culture was then used to inoculate 1 L of LB supplemented with ampicillin (100 μg/mL) for approximately 3 h in order to reach an optical density (OD$_{600 \text{ nm}}$) between 0.8 and 1.0. In order to induce the cells to produce DinB, isopropyl β-D-thiogalactopyranoside (IPTG) was added to 1 mM concentration. The cells were then transferred to 30 °C to shake for an additional 3 h. The cells containing induced protein were then harvested by centrifugation at 6,000 xg for 10 min at 4 °C. Using SDS-PAGE analysis, the induction of DinB was monitored. The pellet was stored at -80 °C until purification (minimum of overnight).

For purification, the cells were thawed on ice overnight at 4 °C in Lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol) with a protease inhibitor tablet (Roche mini complete) and 10 μg/mL PMSF. Next, the cells were resuspended by sonication and then lysed with lysozyme and DNase. After a freeze-thaw cycle, the cell debris was removed by centrifugation at 12,000 xg at 4 °C for 1 h. Then, the DinB in the filtered resuspension was purified using fast protein liquid chromatography (FPLC) using two columns. First, the protein was purified using a FastFlow MonoS ion exchange column (GE Healthcare) with buffers Sa (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol) and Sb (50 mM HEPES, pH 7.5, 1 M NaCl, 2 mM β-mercaptoethanol) as a mobile phase. A gradient to 100% Sb was used over 5 column volumes in order to isolate the protein. The collected fractions were then analyzed using SDS-PAGE. The fractions containing the desired protein were combined and mixed 1:1 with PSa buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol, 1 M (NH$_4$)$_2$SO$_4$). This mixture was then purified using a phenyl sepharose column (GE Healthcare). This separation occurs with a gradient to 100% buffer PSb (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM β-
mercaptoethanol) over 5 column volumes. Finally, the protein was dialyzed twice against
dialysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol)
for at least 3 h each. Bradford assays were employed to determine the concentrations of each
protein using BSA as a standard. The protein was stored at -80 °C in 1 mL aliquots.

b. Primer Extension Assays

Primer extension assays were performed using a 61-mer DNA template as well as either
running start or standing start primers depending on the assay.

DNA Template:

\[5'\text{CGTTACTCAGATCAGGCCTGCGAAGACXTXGGCGTCCGGCTGCAGCTGTACTATCA}\]
\[\text{TATGC}^3'\]

Where X can be either dG (undamaged) or \(N^2\)-furfuryl-dG.

Running Start: \[5'\text{GCATATGATAGTACAGCTGCAGCCGGACGC}\]

Standing Start: \[5'\text{GCATATGATAGTACAGCTGCAGCCGGACGCC}\]

DNA was synthesized either by Prof. Eriks Rozners (SUNY-Binghamton) or Operon.
Post-synthesis modifications were carried out by Prof. Penny Beuning according to published
procedures (DeCorte 1996; Jarosz 2006).

In order to determine extension activity of DinB wild-type or one of the variants, running
start primer was employed which allowed incorporation of one nucleotide prior to the lesion. On
the other hand, when analysis of the kinetics of nucleotide incorporation opposite the lesion was
carried out, the standing start primer was employed. Primer extension assays to study extension
activity were used with all four dNTPs in order to determine replication past the lesion. Kinetics
assays were used with only dCTP (to be incorporated opposite dG or $N^2$-fururyl-dG) in order to determine the ability of DinB to incorporate a single nucleotide. The DNA primer and template were added in equal amounts to the reaction mixture which were annealed using an annealing buffer (20 mM HEPES pH 7.5, 5 mM Mg(OAc)$_2$) and heated to 95 °C for 2 minutes, then incubated at 50 °C for 1 hour, and then transferred to 37 °C until the assay was performed.

The assays were carried out in reaction buffer (30 mM HEPES pH 7.5, 20 mM NaCl, 7.5 mM MgSO$_4$, 2 mM β-mercaptoethanol, 1% bovine serum albumin and 4% glycerol) with 10 nM $^{32}$P-labeled primer/template. Reactions were initiated upon addition of dNTPs, and time points were taken up to 10 minutes (prior to addition of dNTPs, a zero point is taken). The reactions were quenched with 85% formamide, 50 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue. The samples were then analyzed on polyacrylamide gels consisting of 12-16% polyacrylamide and 8 M urea and then imaged on Molecular Dynamics storage phosphor imaging screens. The resulting data were analyzed using ImageQuant software (GE Healthcare). For the kinetics data, $V_{max}$ and $K_M$ were determined using GraphPad Prism analysis software.

3.3 Results

a. Loop Swap Assays

The primary motivation behind the analysis of the loop regions is to determine whether they contribute to the specificity for damaged DNA. The DinB loop regions are located directly adjacent to the active site (Figure 3.1). It is known that DinB is incapable of bypassing lesions that are efficiently bypassed by pol V, specifically thymine-thymine cyclobutane pyrimidine dimers (T-T CPDs) and (6-4) photoproducts (Becherel 1999; Tang 2000; Fujii 2004a; Fujii 2004b). UmuC has similar loop regions located near the active site of the protein that have been
shown to be important for activity (Chandani 2010; Hawver 2011). Therefore, because of their location, it is possible that these loop regions play a significant role in damage specificity.

**Figure 3.4:** Shown is a homology model of DinB (Dinb_2012.pdb) developed in our laboratory highlighting the loop regions. Shown in the inset are the loop regions with loop 1 highlighted in magenta and loop 2 highlighted in cyan. DNA is shown in yellow.

We initially intended to assay these loop 1 and 2 variants of DinB against both $N^2$-furfuryl-dG as well as DNA containing a T-T CPD to determine if the variants affect damage specificity. To date, results only show bypass of undamaged DNA as well as $N^2$-furfuryl-dG. The two loop variants used were generated using site-directed mutagenesis described in Materials and Methods. The first is a loop 1 tri-variant with three site-directed mutations and the other is a loop 2 di-variant with two site-directed mutations (see Figure 3.2). These mutations were made to make the DinB loops more similar to those of UmuC.
Figure 3.5: Shown are the site-directed mutations in the Loop 1 tri-variant of DinB and the Loop 2 di-variant of DinB.

Both of these variants were assayed on undamaged DNA as well as DNA containing a $N^2$-furfuryl-dG adduct. The results can be seen in Figures 3.3 and 3.4.
Results show that both the L1 and L2 variants are active on both undamaged and damaged DNA templates. There is no considerable difference between the activities on either
DNA substrate, showing that there is no identifiable preference for DNA damage versus undamaged DNA.

b. **Primer extension results on single site-directed mutations**

We next focused on individual alanine mutations in the loop 1 region to analyze their activity on damaged and undamaged DNA to help better understand damage specificity. The activity of wild-type DinB and the DinB loop 1 variants were determined using the primer extension assays described in Materials and Methods. We analyzed the activity of WT DinB and the other variants on DNA containing the $N^2$-furfuryl-dG lesion and on the undamaged DNA templates. The single site-directed mutations are shown in Figure 3.5.

![Figure 3.5: Shown are the site-directed point mutations on the Loop 1 region of DinB that were analyzed using primer extension assays.](image-url)
Each of the residues at which the site-directed point mutations were made and their respective locations on the DinB protein can be seen in Figure 3.6. All of the mutations are alanine mutations except for Val40Cys which is a cysteine mutation. Since the Loop 1 tri-variant from section a contains Val40Cys as one of the mutations, we decided to analyze this variant individually as well. We plan to do the same for the other individual variants within loop swap variants. Arginine 38 is located in the region bordering where the single-stranded DNA template enters the active site. From the model, it is proposed that this residue may possibly play a significant role in guiding the DNA into the active site. Valine 40 is located closer to the active site directly adjacent to the template nucleotide opposite the incoming nucleotide and therefore may play a role in coordinating the template nucleotide for binding to the incoming nucleotide. Isoleucine 41 is facing the direction opposite the active site. It is predicted that this residue does not directly influence active site activity, but may support the overall structure of the loop region. Serine 42 and threonine 43 are both located directly above the incoming nucleotide and are predicted to play a significant role in stabilizing the incoming nucleotide for binding the template nucleotide. Asparagine 45 and tyrosine 46 are located along the region where the incoming nucleotide enters the active site. Tyrosine 46 is hypothesized to play a role in guiding the incoming nucleotide to the active site and asparagine 45 may assist in this role as well as help support the overall loop structure. Each of these DinB variants was assayed for primer extension on both undamaged DNA and DNA containing the $N^2$-furfuryl-dG lesion.
Figure 3.9: Shown are the residues where the site-specific mutations were made. Loop 1 and the corresponding residues are highlighted in magenta and loop 2 is highlighted in cyan. The DNA is highlighted in yellow and the remainder of the DinB protein is in gray.
Figure 3.10: Primer extension assays for the following variants of DinB including WT DinB: Arg38Ala, Val40Ala, Val40Cys, Ile41Ala, Ser42Ala, Thr43Ala, Asn45Ala, Tyr46Ala on $N^2$-furfuryl-dG lesion.

The DinB loop 1 variants on DNA containing the $N^2$-furfuryl-dG lesion show varying activity (Figure 3.7). The enzyme concentration for each of the variants is 25 nM with time points of 0, 20, and 60 min. The concentration of dNTPs is 1 mM. Wild-type DinB shows high activity as seen previously for the $N^2$-furfuryl-dG lesion (Jarosz 2006). The variants Ile41Ala and Asn45Ala show neither lesion bypass nor any nucleotide extension past the lesion. Variants Arg38Ala and Tyr46Ala show weak activity compared with activity on undamaged DNA whereas Val40Ala and Val40Cys show significantly greater activity. Lesion bypass is still observed with the Arg38Ala and Tyr46Ala variants. These variants do not show primer
extension to the end of the template, but insertion opposite the lesion is observed in the single nucleotide incorporation results (Figure 3.7). Both Ser42Ala and Thr43Ala show high activity comparable to wild-type with Ser42Ala having slightly higher activity.

Figure 3.11: Primer extension assays for the following variants of DinB including WT DinB: Arg38Ala, Val40Ala, Val40Cys, Ile41Ala, Ser42Ala, Thr43Ala, Asn45Ala, Tyr46Ala on undamaged DNA.

The DinB loop 1 variants show varying activity on undamaged DNA as well (Figure 3.8). The enzyme concentration for each of the variants is 25 nM with time points of 1, 20, and 60 min. The concentration of dNTPs is 1 mM. Wild-type DinB shows activity as seen previously (Jarosz 2006). The variants Ile41Ala and Asn45Ala show neither lesion bypass nor any extension
on undamaged DNA. Variants Arg38Ala and Tyr46Ala show weak activity on undamaged DNA. Lesion bypass is observed with these variants however. Val40Ala and Val40Cys show comparable activity to wild-type and their respective activites against $N^2$-furfuryl-dG DNA. Ser42Ala shows slightly higher activity on undamaged DNA compared to wild-type whereas Thr43Ala shows slightly lower activity.

c. Fidelity of incorporation

In addition to primer extension assays of loop 1 variants of DinB, each of these variants were screened for their fidelity of nucleotide incorporation. These assays were performed similarly to the primer extension assays except instead of adding a mixture of dNTPs; each dNTP was added in a separate reaction. This was performed for all of the loop 1 variants and the loop swap variants on both undamaged and damaged ($N^2$-furfuryl-dG) DNA.

Figure 3.12: Nucleotide incorporation results are shown for the DinB loop 1 variants on $N^2$-furfuryl-dG DNA. Each variant has a zero point which contains no nucleotide and 30 min time points for each nucleotide. The pattern shown for WT DinB on top is repeated for each variant.
The nucleotide incorporation results for the DinB variants on $N^2$-furfuryl-dG DNA show that some of the variants misincorporate dNTPs opposite the lesion (Figure 3.9). DinB has been shown to exhibit accurate nucleotide insertion of dCTP opposite $N^2$-furfuryl-dG and other $N^2$-dG adducts, with modest incorporation of dTTP (Jarosz 2006). Variants of DinB that do not follow this pattern are quite interesting. The most noticeable evidence of nucleotide misincorporation is exhibited by the Ser42Ala variant, which incorporates dTTP almost as well as dCTP opposite $N^2$-furfuryl-dG. Variants Val40Cys and Val40Ala exhibit this effect as well, however it is more similar to that of WT DinB. Also worth noting is that the Val40Ala and Val40Cys variants seem to extend dCTP one nucleotide beyond the lesion. The next nucleotide in the template sequence is dT, suggesting that these variants misincorporate dCTP opposite dT in addition to inserting both dCTP and dTTP opposite $N^2$-furfuryl-dG. It is also seen that the L1 tri-variant and L2 di-variant both seem to incorporate dATP opposite the $N^2$-furfuryl-dG lesion.

**Figure 3.13:** Shown are the nucleotide misincorporation results for the DinB loop 1 variants on undamaged DNA. Each variant has a zero point which contains no nucleotide and 30 min time points for each nucleotide. The pattern shown for WT DinB on top is repeated for each variant.
The nucleotide misincorporation results for the DinB variants on undamaged DNA show that none of the variants seem to misincorporate nucleotides except Val40Cys and Ser42Ala, which weakly misincorporate dATP (Figure 3.10). DinB has been shown to exhibit accurate nucleotide insertion opposite undamaged dG in addition to $N^2$-furfuryl-dG (Jarosz 2006). These variants do not seem to misincorporate nucleotides opposite undamaged dG as they did against $N^2$-furfuryl-dG. As was observed in nucleotide incorporation assays with DNA containing $N^2$-furfuryl-dG, Val40Ala and Val40Cys still appear to incorporate dCTP one nucleotide beyond the lesion, opposite template dT, showing that these variants misincorporate dCTP opposite dT.

3.4 Discussion

Primer extension results of loop 1 variants of DinB against undamaged dG template DNA and DNA containing $N^2$-furfuryl-dG show an array of interesting results. The loop 1 tri-variant and loop 2 di-variant show activity on both undamaged and damaged DNA. As mentioned in the previous section, our initial intention for making these variants was to mimic the loop regions of UmuC. By analyzing these variants on both DNA containing $N^2$-furfuryl-dG and DNA containing a T-T CPD, damage selectivity of the loop regions could be determined. To date, we have determined that these loop swap variants are active on $N^2$-furfuryl-dG. We then decided to focus on individual point mutations in the loop 1 region and analyze variants for activity to help understand DNA damage specificity.

As seen in the previous section, DinB variant Arg38Ala shows weak activity against both undamaged dG and $N^2$-furfuryl-dG DNA with only modestly increased activity on damaged DNA. In Figure 3.6, it can be seen that Arg38 is located in the region where the single-stranded DNA template is entering the DinB active site. Since DinB Arg38Ala exhibits a decrease in
activity when compared with wild-type DinB, it is possible that this residue affects the efficiency of DinB to replicate past any type of DNA, damaged or undamaged. It is possible that the positively charged Arg38 helps to stabilize the negatively charged template DNA strand as it enters the active site.

Both Val40Ala and Val40Cys show strong activity with undamaged DNA and \(N^2\)-furfuryl-dG showing that this residue does not contribute to discrimination against specific types of DNA. In Figure 3.6, it can be seen that Val40 is located directly adjacent to the template nucleotide opposite the incoming nucleotide. Therefore, we suspected that Val40 would contribute to DNA damage specificity based on its proximity to the damaged lesion; however, results show that this is not the case with the specific \(N^2\)-furfuryl-dG lesion studied here. Variant Ile41Ala shows no activity with either undamaged or damaged DNA. Therefore, it is difficult to determine what role Ile41 plays; however, it can certainly be seen that by changing this residue, it renders the protein inactive in replication.

Both Ser42Ala and Thr43Ala show considerably high activity on both undamaged and damaged DNA even when compared with wild-type DinB. Neither of these residues shows a preference for one type of DNA template leading to the conclusion that these residues are not involved in damage specificity. From Figure 3.6, it can be seen that both of these residues are located directly above the incoming nucleotide. Since both of these residues (serine and threonine) contain a hydroxyl group in their side chains, they most likely interact with the incoming nucleotide via hydrogen bonding. Eliminating these hydroxyl groups by changing the residues to alanine would eliminate these hydrogen bonds, possibly reducing nucleotide selectivity and promoting nucleotide misincorporation and/or higher activity. Asn45Ala shows very weak activity against either undamaged or damaged DNA (Figure 3.10).
Finally, Tyr46Ala shows similar results as Arg38Ala in that it exhibits low activity on both undamaged and damaged DNA templates. Tyr46 may play a similar role as Arg38 except that rather than being located in the region where the incoming DNA template strand enters the active site, it is located in the region where the incoming nucleotide enters the active site. Like the other residues analyzed here, Tyr46 most likely has no involvement in damage specificity, but may affect nucleotide incorporation.

From the single nucleotide incorporation results, it can be seen that most of the variants as well as wild-type DinB exhibit misincorporation opposite $N^2$-furfuryl-dG DNA and show faithful incorporation of dCTP opposite undamaged dG. Previous results have shown that DinB bypasses $N^2$-dG adducts as well as undamaged dG with relative accuracy (Shen 2002; Jarosz 2006; Seo 2006; Yuan 2008). 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 in fact misincorporate nucleotides as has been seen previously (Jarosz 2006). In the conditions described in the previous section, DinB WT inserts dTTP opposite $N^2$-furfuryl-dG DNA, yet shows accurate nucleotide incorporation opposite undamaged dG. Ser42Ala also exhibits misincorporation of dTTP opposite $N^2$-furfuryl-dG DNA with a much greater effect. Removal of the hydroxyl group in serine seems to promote nucleotide misincorporation as well as higher activity. However, DinB Thr43Ala does not misincorporate nucleotides opposite $N^2$-furfuryl-dG DNA, which is in contrast to DinB Ser42Ala. Val40Ala and Val40Cys also exhibit extension beyond the lesion by one nucleotide as described in the results section. This further shows that Val40 is a significant residue and needs to be investigated in greater depth. Interestingly, wild-type DinB and the variants that misincorporate nucleotides opposite $N^2$-furfuryl-dG generally are accurate opposite undamaged dG.
3.5 References


Chapter 4: Understanding bypass of DNA containing thymine-thymine cyclobutane pyrimidine dimers by *E. coli* DinB

4.1 Introduction

The purpose of this project is to understand which residues in DinB may be involved in DNA damage specificity by analyzing the polymerase with different DNA lesions using the molecular modeling package Yasara (Vriend 1990). Specifically, DinB will be analyzed with two types of DNA, undamaged DNA and DNA containing the non-cognate lesion thymine-thymine cyclobutane pyrimidine dimer. A homology model of DinB was made using the crystal structure of Dpo4, a DinB ortholog from *Sulfolobus solfataricus* (PDB ID: 1XJ4) (Ling 2001) by Pradyumna Rajput using Yasara. Along with the DinB protein, the homology model also contains undamaged DNA in the active site of the protein. In addition to undamaged DNA, a thymine-thymine cyclobutane pyrimidine dimer was built and added to the existing DNA in the model. In order to investigate damage specificity, energy minimizations were performed for DinB on both the undamaged and T-T CPD dimer DNA templates. By analyzing the relative distances between various residues and the DNA, significant differences between both DNA templates and their effects on the DinB protein were observed. Based on these distances as well as the relative energy from the minimizations and the overall visual observations of the models, conclusions could be drawn about the specificity of DNA damage bypass by DNA polymerase DinB. It has been determined that the bulky three-dimensional nature of the thymine-thymine cyclobutane pyrimidine dimer causes distortion of the DinB active site as well as the opposite DNA strand. The dimer also causes steric hindrance as the template DNA strand is entering the active site of DinB.
4.2 Materials and methods

a. Initial attempts

First, the DinB homology model made by Pradyumna Rajput called ‘DinB_new’ was acquired. In making the homology model, the templates with PDB IDs: 2ASD-A, 1SRF-A, 2IMW-P, 2JEJ-A, 1JX4_A, 3RDI-A, 2UVV-A, 2AGQ-A, 2BQ1-A and 2UVU-A were selected as the final templates. A total of 50 models were made (five for each template) and then ranked according to their respective z-scores. The hybrid model was built based on the homology model made from the 2IMW-P template since it generated the best z-score. It has a reported sequence identity of 98% to the template sequence. When the newly created hybrid model was superimposed on the 2IMW-P template structure, the calculated RMSD was 0.818 with 324 residues aligned and an overall quality ranking of satisfactory. Using ERRAT, Pradyumna Rajput further analyzed the quality of the hybrid model. It produced an overall quality score of 98.193% where good high resolution structures generally produce values of around 95% or higher. Finally, using a Ramachandran plot, the quality of the hybrid model was further assessed. In the plot, 89.6% of the residues are reported in the favorable regions with no residues in the disallowed regions.

The initial plan for this project was to perform molecular dynamics simulations on the DinB homology model, which contains undamaged DNA in the active site. A DNA template with a thymine-thymine cyclobutane pyrimidine dimer in the active site of the DinB protein would be constructed and a similar molecular dynamics simulation would be performed. The data from these simulations would then be compared to provide insights into the effects of DinB binding to non-cognate damage.
The ‘DinB_new’ homology model was loaded into Yasara as a PDB file. Cell neutralization and a pKa prediction experiment were performed with in vivo conditions: pH = 7.0, NaCl concentration = 0.9%, water density = 1.0. The selected force field was YAMBER3 at Normal speed and a simulation cell of 10 Å ‘around all atoms’ was added. Once the cell neutralization and pKa prediction was completed, the energy minimization experiment was performed. Once this completed, a molecular dynamics simulation was begun with in vivo conditions: Temperature = 310 K, Water Density = 0.993 g/cc. The simulation parameters were set as follows: force field = YAMBER3 at Normal speed, Temperature = 310 K, Water Density = 0.993 g/cc, time step = 1 fs and recalculation every 10 steps, simulation snapshots every 5 ps. The simulation was run for about 2 ns but there were not enough data to draw any significant conclusions since the simulation time was too short.

Next, in order to construct the thymine-thymine cyclobutane pyrimidine dimer (T-T dimer) in the active site of the ‘DinB_new’ homology model, the existing DNA template in the current model was deleted and DNA with a T-T dimer was docked into the DinB protein. The file 1N4E.pdb, which is the crystal structure of a DNA decamer containing a cis-syn thymine dimer, from the PDB database, was downloaded and then loaded into the Yasara program with the DinB_new homology model. An attempt was made to dock the DNA into the DinB active site, but it did not work properly, as the Yasara program did not recognize the hydrogen bonds between the bases holding the two DNA strands together. Therefore, an alternative method was used to construct the T-T dimer in the active site of DinB by building the T-T dimer separately and then joining it to the existing DNA template strand in the homology model. The dimer was made by using the ‘build’ option in Yasara and selecting ‘dT’ under DNA nucleotide. Two dTs were added together, then the C5 atoms of the thymines were connected using the ‘add bond’
option and the same was done for the C6 atoms to complete the dimer. The whole dimer was then added to the end of the existing DNA after one of the nucleotides from the existing DNA was removed. Once this was complete, an energy minimization experiment was performed using the same parameters as were used for the DinB with undamaged DNA in the active site. Next, a molecular dynamics simulation was performed with the same parameters as the earlier simulation described above except with the following changes: Time step = 2 fs with a recalculation every 4 simulation steps, and a simulation snapshot taken every 10,000 steps. The simulation was allowed to run for about 10 ns. Once the simulation ended, it was observed that the DNA template strand containing the T-T dimer was incorrectly minimized (Figure 4.1). Ultimately, the molecular dynamics approach was discontinued and this project focused entirely on energy minimizations and analyzing the different types of DNA damage in the DinB active site using visual analyses.

**Figure 4.14:** The above image shows the incorrectly formed thymine-thymine dimer. As can be seen, the DNA backbone is arched over and the bottom nucleotide (from this perspective) should theoretically be on the top and vice versa. The DNA is shown in yellow and the DinB protein is shown in gray.
b. Final Methods

First, the DinB_new homology model was acquired and cell neutralization and pKa prediction, and energy minimization were performed exactly as described in the previous section. Using the minimized structure, the final reported energy was recorded and the model was analyzed by recording the distances between various atoms, mostly between the amino acid residues of the active site of the protein, particularly the residues on loops 1 and 2 and the atoms on the DNA, either the incoming nucleotide or the template strand. The exact distances and specific atoms that were measured are described in the Results section. The same experiments were performed for the DinB homology model containing the T-T dimer on the template DNA strand in the active site. The T-T dimer was made according to the method described in the previous section. This time, the dimer minimized properly (see Results section for details).

Next, two other models were analyzed. One of the models contained undamaged DNA entering the active site of the DinB protein. Using the DinB_new homology model as a starting point, three additional nucleotides were built: adenosine, thymine, and thymine, and they were added onto the end of the existing DNA template strand in the homology model. These three nucleotides were oriented such that they are entering the DinB active site. The same experiments used for the other two models were used for this model as well: cell neutralization and pKa prediction, and energy minimization. Finally, the final model made was similar to the model just described, except the two thymine nucleotides were joined to form a thymine-thymine cyclobutane pyrimidine dimer. They were joined at the C5 and C6 atoms as described for the other T-T dimer model. The same two experiments were performed for this model as well.
Finally, using these four models, the relative minimized energies were compared as well as the relative distances between atoms in the active site and atoms in the DNA. The overall models were compared visually and by analysis of the distances between the protein amino acid residues and the DNA in the active site pocket.

4.3 Results

The following section describes the results that were obtained from each portion of the experiment. First, the DinB homology model that contains undamaged DNA in the active site was analyzed. The overall conformation of the active site and the location of the loop structures and DNA were analyzed (Figure 4.2). In the image, the active site pocket where the template strand enters can be seen clearly as well as the incoming nucleotide which is subsequently joined to the template strand.

Figure 4.15: The above image shows the DinB polymerase with undamaged DNA in the active site. The DNA is shown in yellow and the DinB protein is shown in gray. Highlighted in magenta and cyan are the DinB loop structures known as Loop 1 and Loop 2, respectively. The
DNA strand on the right is the primer strand and the incoming nucleotide is in position to be joined to the primer.

Some of the key residues in the loop structure, particularly Thr43 and Ser42 on loop 1 and Ser55 on loop 2, are positioned directly adjacent to the incoming nucleotide and may play a significant role in guiding the nucleotide to the appropriate position for binding to the template DNA strand (Figure 4.3).

**Figure 4.16:** The above image shows the three of the key stabilizing residues on the loop structures. Ser42, Ser55, and Thr43 are all capable of forming hydrogen bonds with the incoming nucleotide so it is hypothesized that they play a significant role in accommodating the nucleotide in the active site. The DNA is shown in yellow, the DinB protein is shown in gray with loops 1 and 2 shown in magenta and cyan respectively.

Looking more closely at the model with template DNA containing a T-T dimer in the DinB active site (Figure 4.4), it can be seen that the presence of the dimer alters the conformation of the opposite DNA strand. The incoming nucleotide is positioned at an upward facing angle and the nucleotides directly opposite the dimer are slightly angled facing downward.
Also, it can be seen that the nucleotides are not lined up properly to allow the bases to easily form hydrogen bonds with the bases of the opposite strand as in normal DNA. It is also clear that the active site residues and overall conformation of the DinB active site are significantly altered.

Figure 4.17: The above image shows the T-T dimer on the DNA template strand located in the active site of DinB opposite the replicating strand, particularly the incoming nucleotide. As can be seen, the T-T dimer is bulkier than undamaged DNA and the nucleotides on the opposite strand are forced out of the way slightly. The incoming nucleotide is pushed upwards and a couple of the other nucleotides are pushed closer to the DNA backbone to avoid steric hindrance with the dimer.

Table 4.1 shows the dihedral angles of the T-T dimer backbone as well as the dihedral angles of two adjacent thymines that are not in the dimer conformation. It is reported that the dihedrals of the T-T dimer are smaller than those of the non-dimer, showing that the dimer is a more rigid and tighter structure compared to normal DNA. Table 4.1 also shows the lengths of the T-T dimer bonds. The distances between the respective C5 atoms of the two thymines and the C6 atoms are much longer compared to these distances when the C5 atoms are bonded to one
another as well as the C6 atoms in the T-T dimer. This also contributes to the overall rigidity and compactness of the T-T dimer structure. Looking again at Figure 4.4, the T-T dimer is clearly rigid and compact which forces the DNA backbone to be slightly distorted. Overall, the dimer structure is quite bulky and rigid and causes geometrical or spatial problems in the DinB active site as compared to normal DNA which is more flexible.

**Table 4.2**: Dihedrals and Bond Distances in the active site pocket of DinB

<table>
<thead>
<tr>
<th>Non-dimer Dihedrals</th>
<th>Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4, C5, O5, P of DT 2</td>
<td>147.292</td>
</tr>
<tr>
<td>P (DT 2), O3, C3, C2, of DT1</td>
<td>92.565</td>
</tr>
<tr>
<td>Dimer Dihedrals</td>
<td>Angle (°)</td>
</tr>
<tr>
<td>C4, C5, O5, P of DT 2</td>
<td>136.815</td>
</tr>
<tr>
<td>P (DT 2), O3, C3, C2, of DT1</td>
<td>62.958</td>
</tr>
<tr>
<td>Non-dimer bond distances</td>
<td>Distance (Angstroms)</td>
</tr>
<tr>
<td>C6 DT1 and C6 DT 2</td>
<td>6.128</td>
</tr>
<tr>
<td>C5 DT1 and C5 DT2</td>
<td>5.56</td>
</tr>
<tr>
<td>Dimer bond distances</td>
<td>Distance (Angstroms)</td>
</tr>
<tr>
<td>C6 DT1 and C6 DT 2</td>
<td>1.591</td>
</tr>
<tr>
<td>C5 DT1 and C5 DT2</td>
<td>1.613</td>
</tr>
</tbody>
</table>

Figure 4.5 shows the distances between various atoms from residues in the active site region of the DinB protein and atoms from the incoming nucleotide or other residues in the second shell of the DinB protein. The purpose of this plot is to examine the change in the distances between these atoms in the model with DNA containing a T-T dimer and compare them to the model with undamaged DNA in order to determine which areas of the active site are most affected spatially by the presence of the T-T dimer. As can be seen, other than a few significant differences in the distances between residues and the incoming nucleotide on the right side of the graph, the major differences are seen in the region where the template DNA strand is located in the active site. It is clear that the presence of a T-T dimer in the template DNA strand causes significant spatial alterations in the DinB active site. The distances between residue V40
and I41 and between I31 and the incoming nucleotide base are significantly longer when the T-T dimer is present compared to when the DNA is undamaged.

**Figure 4.5:** The relative distances between the residues of DinB in the active site area of the protein and the atoms on the incoming nucleotide.

Figure 4.6 shows the model of the DinB protein with the undamaged template strand extended, showing how it enters the active site. This image shows the entire extended template strand as well as the DinB active site with the incoming nucleotide positioned appropriately to bind to the growing replicating strand and the opposite nucleotide on the template strand. The extended portion of the template DNA contains an adenine and two thymine bases.
Figure 4.6: The above image shows the DinB protein with undamaged DNA in the active site as well as the extended single-stranded template entering the active site. The extended portion consists of thymine, thymine, and adenine going from left to right in the image. Loops 1 and 2 are labeled in magenta and cyan respectively, the remainder of the DinB protein is labeled in gray, and the DNA is labeled in yellow.

Looking more closely at the entering template strand, it can be seen that the extended portion of the strand fits nicely into the entrance pocket as can be seen in Figure 4.6 and 4.7. A few potentially significant residues are highlighted (Figure 4.7) and are expected to play a vital role in the coordination of the template strand into the active site of the DinB protein: Arg35, Arg38, and Phe295. Figure 4.8 presents a closer look at the two adjacent thymine nucleotides on the extended portion of the template strand. Ideally, this is how normal thymine nucleotides are oriented in undamaged DNA.
**Figure 4.7:** The above image shows the same model as in Figure 4.6, except from a different angle. This angle shows the single-stranded portion of the DNA template entering the active site. Three critical residues are shown, Arg38 which is part of the loop 1 structure, Arg35, and Phe295 which are both hypothesized to play a critical role in guiding the template strand into the active site. Loops 1 and 2 are labeled in magenta and cyan respectively and the other key residues are labeled in red. The remainder of the DinB protein is in gray and the DNA is in yellow.

**Figure 4.8:** The above image shows the two thymine residues in the non-dimer conformation on the undamaged template strand. The DNA is labeled in yellow and the DinB protein is labeled in gray. The red color refers to Arg35 and magenta refers to the N-terminus of loop 1.
In the model with the extended template strand containing a T-T dimer (Figure 4.9), it can be seen that the rigid conformation of the dimer as explained earlier alters the overall conformation of the entering template strand.

**Figure 4.9:** The above image shows the thymine-thymine dimer on the template DNA strand as it enters the active site of the DinB protein. Loops 1 and 2 are labeled in magenta and cyan respectively. The remainder of the DinB protein is in gray and the DNA is in yellow.

Figure 4.10 shows this alteration even further since it shows the pocket where the template strand enters the protein. It can be seen that the entering DNA with the T-T dimer is severely forced out of position when compared with the undamaged DNA strand in Figure 4.7. The DNA is forced away from the Phe295 residue and is instead much closer to the two Arg residues, with part of Arg38 going through the center of the dimer which is highly unfavorable.
Figure 4.10: The above image shows the template DNA strand containing the thymine-thymine dimer as it is entering the active site of the DinB protein similar to Figure 4.7 except from a different angle. Loops 1 and 2 are labeled in magenta and cyan respectively and the other key residues are labeled in red. The remainder of the DinB protein is in gray and the DNA is in yellow. Other critical residues are labeled in red. As can be seen, the DNA here compared with the DNA in Figure 4.5 does not fit as nicely into the pocket. The DNA backbone is kinked and the dimer is rigid causing the template strand to have an awkward orientation as it is entering the active site.

Table 4.2 compares the dihedral angles of the dimer when it is located directly in the active site and when it is on the extended template strand entering the active site pocket of DinB. It also compares the dimer bond distances similar to Table 4.1 between the dimer in the active site and when the dimer is part of the extended template strand. As can be seen, the dimer is more compact when it is in the active site pocket compared to when it is entering the active site on the extended template strand since the dihedral angles and bond distances are smaller in the pocket versus out of the pocket.
Table 4.3: Dihedrals and distances outside of the active site of DinB where the template strand enters DinB

<table>
<thead>
<tr>
<th>Dimer Dihedrals (out of pocket)</th>
<th>Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4, C5, O5, P of DT 2</td>
<td>170.079</td>
</tr>
<tr>
<td>P (DT 2), O3, C3, C2, of DT1</td>
<td>96.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimer bond distances (out of pocket)</th>
<th>Distance (Angstroms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6 DT1 and C6 DT 2</td>
<td>2.703</td>
</tr>
<tr>
<td>C5 DT1 and C5 DT2</td>
<td>2.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimer Dihedrals</th>
<th>Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4, C5, O5, P of DT 2</td>
<td>136.815</td>
</tr>
<tr>
<td>P (DT 2), O3, C3, C2 of DT1</td>
<td>62.958</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimer bond distances</th>
<th>Distance (Angstroms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6 DT1 and C6 DT 2</td>
<td>1.591</td>
</tr>
<tr>
<td>C5 DT1 and C5 DT2</td>
<td>1.613</td>
</tr>
</tbody>
</table>

Figure 4.11 shows the distances between the residues of the DinB active site and the template DNA. The left side of the graph shows the portion of the template DNA that is farthest from the active site and the right side of the graph shows the portion of the template DNA that is closest to the active site. As can be seen, the presence of the T-T dimer greatly distorts the conformation of the template DNA strand and its position relative to the pocket entering the active site as well as the active site pocket itself. Almost all of the distances are longer when the dimer is present showing that there is a significant change in the configuration of the entering template strand relative to the DinB protein.
Figure 4.11: The above plot shows the distances between the key residues near the entering template strand and the atoms of the template DNA.

Finally, Figure 4.12 shows the distances between residues of the DinB active site and the incoming nucleotide where the template DNA contains a T-T dimer in the active site. It compares this configuration where the loop residues are the native DinB residues and where the loop residues are changed to match the residues of UmuC, the polymerase in *E. coli* capable of bypassing T-T dimers in DNA. As can be seen in the graph, there are no significant changes in the distances of the residues to the incoming nucleotide when the loop residues are either native or changed to UmuC.
4.4 Discussion

Taking a closer look at the DinB active site with undamaged DNA template, it can be seen that it can accommodate undamaged DNA quite well without any steric strain or hindrances to the active site (see Figure 4.2). Highlighted in Figure 4.3 are some of the residues that are expected to play key roles in guiding the incoming nucleotide to the opposite nucleotide in the template strand, particularly Ser42, Thr43, and Ser55. Each of these residues, which are located on the loop structures directly above the incoming nucleotide, is capable of forming hydrogen
bonds. It is possible, therefore, that these residues form hydrogen bonds with atoms of the incoming nucleotide and help to coordinate the nucleotide so that it can bind to the replicating strand and the opposite nucleotide on the template strand. Another key residue that is known to play a role in accommodating nucleotides in the active site of DinB is Phe13, which is not shown in the results section (Jarosz 2006). It is located behind the incoming nucleotide and is considered a second shell residue (Walsh 2011).

Looking more closely at the thymine-thymine dimer in the active site of DinB (Figure 4.4), it can be seen that the dimer itself is much more bulky and rigid than undamaged DNA and is therefore geometrically unfavorable in the DinB active site. Although not seen in the figure, it was observed that the loop residues are greatly distorted when the T-T dimer is present in the active site. Loop 1 forms a beta sheet conformation but this conformation is altered when the T-T dimer is present in the active site. In addition, the incoming nucleotide is forced up and out of the way when the T-T dimer is present and other nucleotides on the opposite replicating strand are bent inward and forced closer to the DNA backbone. This leads to the conclusion that the T-T dimer, when present in the DinB active site, is geometrically unfavorable and forces the active site residues and the nucleotides on the opposite DNA strand to be pushed out of the way. This also leads to the conclusion that the DinB active site pocket is clearly too small to accommodate such a large bulky lesion as a T-T dimer. Even looking at the types of lesions that DinB typically bypasses, it can be seen that the small size of the DinB active site is not necessarily a problem for accommodating those lesions. For example, most dG adducts such as $N^2$-furfuryl-dG are geometrically planar, as are ordinary nucleotides. The DinB active site, based on visual observation, seems to accommodate planar adducts appropriately since it is relatively small in size. Therefore, it is no surprise that the DinB polymerase, with its small active site, cannot
accommodate such a large, rigid, more three dimensional structure as a T-T dimer. This also leads to the assumption that UmuC, the polymerase capable of bypassing T-T dimers, most likely has a larger active site that allows for such a bulky lesion to enter. However, without a crystal structure of UmuC, it is not possible to determine this exactly. Even by changing the loop residues to match the loop structures to those of UmuC, there is very little difference observed that could explain how UmuC might accommodate the bypass of a T-T dimer. Perhaps it confers a different degree of flexibility but it is still difficult to determine why UmuC accommodates this lesion.

Looking more closely at the models with the extended DNA template strand as it is entering the DinB active site, it can be seen that the undamaged extended template fits nicely into the entrance pocket of DinB (see Figure 4.7). Some of the highlighted residues, particularly Phe295, seem to play a critical role in coordinating the template strand in the active site. Some of the other key residues such as Arg35 and Arg38 are most likely playing a significant role in coordinating the template strand since these residues are positively charged at in vivo pH conditions (approximately 7.5). Since these arginine residues are positively charged, they could be significantly involved in coordinating the negatively charged DNA backbone. Taking a look at the model with the extended DNA template strand that contains a T-T dimer (Figure 4.10), it can be seen that the presence of the T-T dimer greatly distorts the conformation of the DNA as well as its position as it is entering the DinB active site. As explained earlier, the bulkiness of the T-T dimer causes the residues of the DinB active site to be forced away causing problems in the conformation of the protein itself. Similarly, the large, bulky, and rigid T-T dimer probably has the same effect on the DNA entrance channel of DinB. Additionally, the entire template strand has shifted much closer to the region with the two arginine residues that were mentioned above,
Arg35 and Arg38. As explained, the arginine residues are positively charged and the DNA backbone is negatively charged. Since the T-T dimer forms a rigid, more constrained conformation than normal DNA, it brings the two thymine nucleotides closer together through the C5/C6 bonds (see Table 4.1). This in turn further exposes the negatively charged DNA backbone which then interacts more strongly with the positively charged arginine residues. It can be concluded therefore that the positively charged arginine residues are most likely interacting with the negatively charged DNA backbone more when the DNA is in the dimer conformation causing the entire DNA template strand to be shifted closer to those residues as seen in Figure 4.10. The overall rigidity of the template strand with the T-T dimer also removes any flexibility that is normally seen with undamaged DNA, making the template strand and the T-T dimer even more geometrically unfavorable.

In conclusion, it can be seen that the T-T dimer is highly unfavorable both in the active site of the DinB polymerase as well as in the template entrance pocket. This can therefore explain why DinB is incapable of bypassing thymine-thymine cyclobutane pyrimidine dimer lesions in DNA. In addition, it can be seen that certain residues, particularly, Arg35, Arg38, and Phe295 most likely play a key role in accommodating the DNA template into the DinB active site and therefore may play a role in damage specificity. It can also be seen that other residues, particularly those on the loop structures such as Ser42, Thr43, and Ser55 play a key role in accommodating the incoming nucleotide to be bound to the template strand, but most likely have little role in determining damage specificity.
4.5 References


Chapter 5: Conclusions and Future Consideration

5.1 Conclusions

The original objective for this work was to identify which residues or regions of *E. coli* Y-family polymerase DinB contribute to DNA damage specificity of the polymerase. To answer this, we approached the problem from two different angles. First, we attempted directed evolution methods in order to transform the function of DinB to perform the tasks usually associated with the second of the two *E. coli* Y-family polymerases, UmuD’2C. Through hydroxylamine mutagenesis of the DinB gene followed by selection with UV light exposure, we attempted to isolate a variant of DinB that would be capable of bypassing thymine-thymine cyclobutane pyrimidine dimers, one of the DNA adducts formed from UV exposure. It was further believed that the little finger of UmuC and/or its C-terminal region may give it the unique ability to bypass adducts caused by UV light, hence explaining why DinB is unable to do so naturally. We therefore created chimeras of DinB containing UmuC’s little finger and C-terminal region and performed directed evolution methods on these as well. Despite having no conclusive results to date, we have obtained some positive hits which will continue to be characterized.

In addition, the second avenue by which we attempted to identify residues of DinB that contribute to DNA damage specificity was to study the loop 1 region of the protein. Based on observations of the DinB model, we know that the loop 1 region of DinB is located directly adjacent to the active site where the incoming nucleotide meets the single-stranded template. Using primer extension assays of DinB variants of the loop 1 residues; in which each residue was converted to alanine in a single point mutation, we were able to observe how the residue affects the activity of DinB in performing TLS. We found that none of the residues contribute to
specificity for the $N^2$-furfuryl-dG lesion, as the variants generally exhibited similar activity on damaged and undamaged DNA, but that some of them do contribute to nucleotide selection. Variants Ser42Ala, Val40Ala, Val40Cys misincorporate dTTP opposite $N^2$-furfuryl-dG showing that these residues (Val40 and Ser42) located in the loop 1 region of DinB are involved in nucleotide selection as DinB bypasses $N^2$-furfuryl-dG and possibly other adducts. Residues Arg38, Ile41, Asn45, and Tyr46 were also identified as being extremely important for DinB activity.

Finally, we have performed modeling work, using homology model, structure building, and energy minimization modules in the YASARA suite of programs, on DinB to understand its inability to bypass thymine-thymine cyclobutane pyrimidine dimers. We modeled the dimer into the single-stranded DNA template in both the active site of DinB as well as in the region where the template strand enters the active site. We found that the active site of DinB and the opposite DNA strand are sterically hindered by the presence of the dimer when the dimer is located in the active site. We also found that the dimer causes a distortion in the DNA strand that creates an unfavorable conformation when the dimer enters the active site. The amino acid residues Arg35, Arg38, and Phe295 seem to be involved in guiding the template DNA strand into the active site of DinB and could affect DNA damage specificity. We conclude therefore that the thymine-thymine dimer is geometrically unfavorable in the DinB active site which could provide an explanation as to why DinB is unable to bypass adducts of UV light in vivo.
5.2 Future Considerations

a. Directed Evolution

We are continuing to characterize the hits that have been obtained using the directed evolution methods. We plan to continue screening for mutants and assaying potential hits in order to acquire variants of DinB capable of bypassing adducts of UV light.

In addition, error-prone PCR was used as an alternative directed evolution method. We adapted an error-prone PCR method from Rasila et al., using the nucleotide analogues dPTP and 8-oxo-dGTP (Rasila 2009) (Kim 2009). We also used mutant strains as a possible alternative for directed evolution, particularly the XL1Red strains and the dut-ung- strain. Finally, we attempted to use randomized primers as well in order to screen through several genetic orientations to acquire a successful variant capable of bypassing lesions caused by UV light. These methods are explained in further detail in sections e and h of Chapter 2. They will continue to be refined in order to identify variants of DinB with the ability to bypass UV damage.

b. DinB variants

The next steps in the analysis of the DinB loop 1 variants is to perform kinetics assays in order to determine the precise efficiency of each variant and analyze how each residue affects the overall activity of the polymerase. These experiments are currently being performed for each of the active loop 1 variants. Further analysis will also be conducted concerning the preferred ability of DinB Ser42Ala, Val40Ala, and Val40Cys variants to misincorporate dTTP opposite $N^2$-furfuryl-dG.
Furthermore, site-directed mutations have been made for Arg35Ala and Phe295Ala to determine if these residues contribute to DNA damage specificity by DinB. Primer extension assays will be performed with these variants using both undamaged dG template as well as $N^2$-furfuryl-dG DNA and perhaps other DNA adducts. Finally, we plan to obtain DNA template containing a site-specific thymine-thymine cyclobutane pyrimidine dimer in order to perform primer extension assays with all DinB variants to help better understand DNA damage specificity of not only the loop 1 region, but other important regions of DinB as well.

5.4 References
