FUNCTION OF DAMAGE INDUCIBLE DNA POLYMERASE IV FROM Escherichia coli

A dissertation presented

by

Jason M. Walsh

to
The Department of Chemistry and Chemical Biology

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

in the field of
Chemistry

Northeastern University
Boston, Massachusetts
December, 2012
Function of Damage Inducible DNA Polymerase IV from
*Escherichia coli*

by

Jason M. Walsh

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Chemical Biology in the Graduate School of Science of Northeastern University, December 2012
Abstract

Y family polymerases are a family of DNA polymerases that carry out replication of damaged DNA in a process known as translesion synthesis (TLS) when normal polymerases cannot bypass the damaged base. TLS occurs in two distinct steps: the insertion of a nucleotide opposite the lesion, which can be correct or mutagenic, and the extension step, where 3-6 nucleotides (TLS patch) are added downstream from the lesion so that the normal replicative polymerase can then continue DNA replication. This extension step may be more essential to cell survival than the insertion step. This dissertation presents a comprehensive study to understand the basis of specificity of Y family DNA polymerases, analyzing both the amino acid residues of the DNA polymerases as well as using modified DNA constructs.

Using computational prediction methods THEMATICS and POOL developed at Northeastern, we have identified amino acid residues in DinB that are distant from the active site that are predicted to contribute to activity. Mutation of these amino acids have varying effects, from complete loss of activity to a reduction in the efficiency of incorporating dCTP opposite DinB’s cognate lesion $N^2$-furfuryl-dG. The inactive DinB variants were found to bind to damaged DNA as well as wild-type DinB does, indicating that the loss of activity is due to a defect in catalytic activity rather than substrate binding. Most striking is that those variants that were active were specifically defective for the essential extension step of TLS.

The DinB subfamily, which includes polymerase IV (DinB) from *E. coli*, Dpo4 from *S. solfataricus* and mammalian DNA polymerase kappa, are specialized to copy bases that are modified on the minor groove side of DNA such as the $N^2$ position on guanosine.
They are less efficient at replicating beyond DNA adducts that lie on the major groove side of DNA. DinB is unable to copy DNA containing the bulky cytosine analog tC (1,3-diaza-2-oxo-phenothiazine), even though the A-family DNA polymerase Klenow fragment is proficient at copying DNA that contains tC. However, DinB will utilize the triphosphate version of tC to incorporate it onto the growing primer strand and extend from it by inserting the next base. Other major groove DNA adducts such as $N^6$-furfuryl-dA and etheno-dA are not readily bypassed by the DinB subfamily of polymerases. All three polymerases remain faithful in addition of dTTP from $N^6$-furfuryl-dA. DinB extension is completely blocked by etheno-dA. Polymerase kappa and Dpo4 show weak extension on templates containing etheno-dA at the 60 minute time point, however, there is no observable dTTP addition opposite etheno-dA. We have identified a variant of DinB, R35A, that performs TLS on templates containing $N^6$-furfuryl-dA but not etheno-dA. This is an intriguing development in the study of Y family polymerases and their inherent function and specificity.
Acknowledgements

I would like to thank my advisor Dr. Penny J. Beuning. I am really appreciative of the opportunity that you gave me to succeed, and for all of your help through the years, fighting for me, and your continuing assistance throughout the process of writing my papers and all of the help you gave me in my eventual transition from graduate school to the real world. I could not have accomplished any of this without you.

I would like to thank my dissertation committee members: Dr. Mary Jo Ondrechen, Dr. David Budil, and Dr. Lee Makowski. Thank you all for taking time to read this dissertation, and offering me constructive questions, comments, and help to make this the best it could possibly be.

I would like to thank my collaborators: Ondrechen Research Group at Northeastern University; Wilhelmsson Group at Chalmers University, Sweden; Rozners Group at SUNY Binghamton.

I would like to thank the Department of Chemistry and Chemical Biology at Northeastern University.

I would like to thank all of the Beuning DNA Damage and Repair Laboratory members, past and present. Especially my post-doc Dr. Jana Sefcikova, and my partners in crime; Dr. Jaylene Ollivierre, Michelle Silva, Lisa Hawver, Phillip Nevin, Dr. Srinivas Somarowthu, Ramya Parasuram, Brenna Shurtleff, Nicholas DeLateur, Rajal Sharma, Paul Ippoliti. I would also like to thank the undergrads that have worked on these projects: Khadijah Balfour-Jeffrey, Kasia Wilk, Tara Dhingra, and Josh Araya.

Finally, I would like to thank my family. First and foremost my beautiful, patient, helpful and understanding wife Jacqueline… I never would have survived this thing if it wasn’t for your ability to kick me in the ass, or listen to me complain, pick me up from the city when I missed the good train, drive me in early, edit my writing, and do the household chores on the weekends when I was toiling away in the lab. You are my rock. Also Myles, I can’t wait to meet you, and hopefully one day you will be able to read this thing, and get a sense of how your old man spent his late twenties. Trigger, my best friend in the whole world, you opened my eyes, made me laugh and hung out with me whenever I felt down. My dad Raymond, mom Carolyn, brothers Jared and Jacob, Nana, aunts, uncles, cousins, my Braintree and WNEC homeboys, plus the Biology and Chemistry departments at Western New England College; you all helped mold a young me into what became a real life scientist.

I would also like to recognize my new family: Nicole, Noreen, Billy, Rosemary, Dan, and Pam.

Thank you all so very much, for all of your help.
Table of Contents

Abstract of Dissertation ................................................................. 2
Acknowledgements .............................................................................. 5
Table of Contents ................................................................................. 6
List of Figures ..................................................................................... 8
List of Tables ....................................................................................... 10
List of Abbreviations ........................................................................ 11
Chapter 1: Background ......................................................................... 17
  1.1 Introduction .................................................................................... 17
  1.2 Transcriptional and Post-Translational Regulation .......................... 22
  1.3 DNA Polymerase IV: DinB ................................................................. 24
  1.4 Specificity of DinB ........................................................................ 26
  1.5 DinB Variants ............................................................................... 30
  1.6 Cellular Interactions of DinB ............................................................. 32
  1.7 Non-natural Nucleotides as Probes for Polymerases ....................... 34
  1.8 Uses of Non-natural Nucleotides ..................................................... 35
  1.9 Synthetic Abasic Sites ................................................................. 36
  1.10 Minimal Requirements for Polymerization and Base Size Tolerance . 37
  1.11 Fluorescent Nucleotide Analogs .................................................... 38
  1.12 Major Groove Adducts ............................................................... 41
  1.13 Summary ..................................................................................... 43
  1.14 References ............................................................................... 44
Chapter 2: Effects of non-catalytic, distal amino acid residues on activity of E. coli DinB (DNA polymerase IV) .......................................................... 59
  2.1: Introduction .................................................................................. 59
  2.2 Materials and Methods ............................................................... 67
    2.2.1 Computational Methods .......................................................... 67
    2.2.2 Experimental Methods ........................................................... 68
  2.3 Results .......................................................................................... 69
    2.3.1 H6L ....................................................................................... 72
    2.3.2 D10E, D10N ........................................................................... 73
    2.3.3 Y106A, Y106F ....................................................................... 74
    2.3.4 K146A ................................................................................. 75
    2.3.5 K150A, K157A, K157I .............................................................. 76
  2.4 Discussion .................................................................................... 76
  2.5 References ............................................................................... 82
Chapter 3: Discrimination against the Cytosine Analog tC by *Escherichia coli* DNA Polymerase IV DinB

3.1 Introduction ............................................................................................................. 93
3.2 Materials and Methods ............................................................................................ 97
  3.2.1 Proteins and Nucleic Acids ............................................................................. 97
  3.2.2 Primer Extension Assays ................................................................................ 97
  3.2.3 DNA Binding Assays .................................................................................... 99
  3.2.4 Model of DinB bound to DNA containing tC ................................................ 100
3.3 Results ................................................................................................................... 100
  3.3.1 DinB discriminates against tC as the template base ...................................... 100
  3.3.2 DinB faithfully incorporates dGTP opposite tC ............................................ 103
  3.3.3 DinB requires primers at least three nucleotides beyond tC for efficient extension ................................................................. 105
  3.3.4 DinB adds dtCTP to a primer terminus.......................................................... 107
3.4 Discussion ............................................................................................................. 108

Chapter 4: Discrimination against major groove adducts by Y family polymerases of the DinB subfamily

4.1 Introduction ........................................................................................................... 121
4.2 Materials and Methods .......................................................................................... 124
4.3 Results ................................................................................................................... 126
  4.3.1 Cytosine analogs dP and pyrrolo-dC modestly inhibit DinB ......................... 126
  4.3.2 N⁶-furfuryl-deoxyadenosine inhibits TLS ...................................................... 128
  4.3.3 1,N⁶-ethenodeoxyadenosine (εdA) strongly inhibits TLS ............................. 132
  4.3.4 DinB R35A mutation allows bypass of N⁶-furfuryl-dA but not εdA ............. 133
4.4 Discussion ............................................................................................................. 135
4.5 References ............................................................................................................. 139

Chapter 5: Future Considerations ................................................................................... 143

5.2 References ............................................................................................................. 147
List of Figures

Figure 1.1 Comparison of the overall folds of (a) a replicative DNA polymerase, and (b) a Y-family DNA polymerase ................................................................. 18
Figure 1.2 The canonical Watson-Crick base pairs ................................................. 19
Figure 1.3 The domains and relative sizes of some Y family polymerases .............. 21
Figure 1.4 DNA damaging agents lead to the formation of lesions in DNA that disrupt replication and induce the SOS response .............................................. 23
Figure 1.5 Adducts of deoxyguanine bypassed by E. coli pol IV .............................. 28
Figure 1.6 Mechanisms for the generation of -1 frameshift mutations by E. coli pol IV .................................................................................................................. 29
Figure 1.7 Model of DinB with some important residues highlighted ..................... 31
Figure 1.8 2-Aminopurine (2AP) and its base pairs (a) 2AP base paired with thymine and (b) 2AP base paired with cytosine ................................................. 39
Figure 1.9 Fluorescent cytosine analogs tC and tC° form canonical base pairs with dG ................................................................................................................. 42
Figure 2.1 Structural comparison of (a) natural deoxyguanosine and (b) $N^2$-furfuryl-deoxyguanosine ($N^2$-f-dG) ................................................................. 60
Figure 2.2 Homology model of E. coli polymerase IV DinB ..................................... 71
Figure 2.3 Schematic of single nucleotide incorporation of each of the four deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP) ......................... 72
Figure 2.4 Schematic of primer extension assays with primer/template DNA containing $N^2$-f-dG correctly base paired with C ............................................. 73
Figure 2.5 (a) ConSurf model of DinB .................................................................. 78
Figure 3.1 Hydrogen bonding between (a) canonical guanine and cytosine and (b) guanine and the cytosine analog tC ................................................................. 96
Figure 3.2 Primer extension assay on DNA containing tC with (a) standing start primer and (b) running start primer .............................................................. 101
Figure 3.3 DinB cannot extend past tC in the DNA template. DinB primer extension on template containing C (a) and template containing tC (b) with primer 22 103
Figure 3.4 (a) DinB faithfully incorporates only dGTP opposite tC in the DNA template .............................................................................................................. 104
Figure 3.5 DinB extension on both tC-containing DNA template and C-containing DNA template with primers of varying lengths ............................................................... 106
Figure 3.6 DinB incorporates the tC analog and replicates at least one additional nucleotide ........................................................................................................... 107
Figure 3.7 Model of DinB bound to DNA ................................................................................ 111
Figure 4.1 (a) The canonical Watson-Crick base pairing of G:C and cytosine analogs pyrrolo-dC and dP (b) The canonical Watson-Crick base pairing of A:T and adenosine analogs N6-furfuryl-dA and etheno-dA........................................... 123
Figure 4.2 Primer extension as well as fidelity of incorporation of Y family polymerases E. coli DinB, human DNA polymerase κ and S. solfataricus Dpo4 on templates containing pyrrolo-dC................................................................. 127
Figure 4.3 (a) Primer extension of Y family polymerases E. coli DinB, human polymerase κ and S. solfataricus Dpo4 on templates containing N6-furfuryl-dA. (b) Fidelity of incorporation of Y family polymerases E. coli DinB, human polymerase κ and S. solfataricus Dpo4 on templates containing N6-furfuryl-dA. (c) Primer extension of Y family polymerases E. coli DinB, human polymerase κ and S. solfataricus Dpo4 on templates containing etheno-dA (d) Fidelity of incorporation of Y family polymerases E. coli DinB, human polymerase κ and S. solfataricus Dpo4 on templates containing etheno-dA. .......................... 128
Figure 4.4: Primer extension of Y family polymerases (a) E. coli DinB, (b) human polymerase κ, or (c) S. solfataricus Dpo4 on templates containing N6-furfuryl-dA and etheno-dA ......................................................................................................................... 131
Figure 4.5 Primer extension and fidelity of incorporation of wild-type DinB (25 nM) and DinB R35A (12.5 nM) with templates containing (a) N6-furfuryl-dA or (b) etheno-dA. (c) Position of residue R35 in DinB............................................................ 134
Figure 5.1 Diagram of the two phases of translesion synthesis (TLS): insertion opposite the lesion, and extension beyond it......................................................... 145
List of Tables

Table 2.1  Top 8% of POOL predicted residues.............................................66
Table 2.2  Kinetic Parameters for dCTP incorporation opposite $N^2$-f-dG by DinB
and Binding Affinity of DinB to DNA............................................................74
Table 3.1  Sequences of DNA primers and templates.......................................99
Table 3.2  Dissociation constants for DinB binding to DNA constructs.............102
Table 3.3  Steady state kinetic parameters for dGTP incorporation opposite tC
vs C...................................................................................................................105
Table 3.4  Steady state kinetic parameters for incorporation of dtCTP versus
dCTP opposite template G............................................................................108
Table 4.1  Steady state kinetic parameters for dTTP incorporation opposite $N^2$-f-
dA and A by DinB, DinB R35A, polymerase kappa and Dpo4............................129
Table 4.2  Steady state kinetic parameter comparison of DinB addition of dCTP
from $N^2$-f-dG and $N^6$-f-dA..........................................................................129
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>percent</td>
</tr>
<tr>
<td>*</td>
<td>position of $^{32}$P</td>
</tr>
<tr>
<td>′</td>
<td>prime</td>
</tr>
<tr>
<td>~</td>
<td>approximately</td>
</tr>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>β</td>
<td>beta</td>
</tr>
<tr>
<td>γ</td>
<td>gamma</td>
</tr>
<tr>
<td>η</td>
<td>eta</td>
</tr>
<tr>
<td>ι</td>
<td>iota</td>
</tr>
<tr>
<td>κ</td>
<td>kappa</td>
</tr>
<tr>
<td>ζ</td>
<td>zeta</td>
</tr>
<tr>
<td>γ-HOPdG</td>
<td>γ-hydroxypropano-deoxyguanosine</td>
</tr>
<tr>
<td>εdA</td>
<td>etheno-dA</td>
</tr>
<tr>
<td>2-AAF</td>
<td>N-2-Acetylaminofluorene</td>
</tr>
<tr>
<td>2-AF</td>
<td>2-aminofluorene</td>
</tr>
<tr>
<td>4NQO</td>
<td>4-nitroquinoline-1-oxide</td>
</tr>
<tr>
<td>8-oxo-G</td>
<td>8-oxo-guanine</td>
</tr>
<tr>
<td>$^{32}$P</td>
<td>Phosphrous-32</td>
</tr>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>A</td>
<td>Alanine</td>
</tr>
<tr>
<td>aaRSs</td>
<td>Aminoacyl-tRNA synthetases</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>AP</td>
<td>abasic</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic Acid</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BL21</td>
<td><em>E. coli</em> cells</td>
</tr>
</tbody>
</table>
BPDE   Benzo[a]pyrene diolepoxide
BSA   bovine serum albumin
C   carbon
C   cytosine
CEdG  $N^2$-(1-carboxylethyl)-2ʹ-deoxyguanosine
CPG   controlled pore glass
Cys  Cysteine
D   Aspartic acid
dA   deoxyadenine
dAPMP  deoxyaminopurine monophosphate
dATP  deoxyadenosine triphosphate
dC   deoxycytosine
dCTP  deoxycytosine triphosphate
dG   deoxyguanosine
dGTP  deoxyguanosine triphosphate
DMT   dimethoxytrityl
DNA   deoxyribonucleic acid
dNTP  deoxynucleotide triphosphate
dT   deoxythymine
dtCTP  deoxy-1,3-diaza-2-oxophenothiazine triphosphate
dTTP  deoxythymidine triphosphate
E   Glutamic Acid
E. coli  *Escherichia coli*
EDTA  ethylenediaminetetraacetic acid
Extn+  extension plus
F   Fluorine
F   Phenylalanine
FPLC  fast protein liquid chromatography
FRET  fluorescence/Förster resonance energy transfer
G   Glycine
G   guanine
Gln   Glutamine
Glu   Glutamic Acid
Gly   Glycine

g   g-force

GE   General Electric

H   Histidine
H   hydrogen

HEPES  4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV   human immunodeficiency virus
I   isoleucine
I2   molecular iodine
ID   identification number
Ile   isoleucine

IPTG   Isopropyl β-D-1-thiogalactopyranoside

K   Lysine

kc_{at}   catalytic constant; turnover number

K_{D}   dissociation constant

kDa   kilodalton
KF   Klenow fragment

K_{m}   Michaelis-Menten constant

KSI   ketosteroid isomerase

Leu   Leucine

L   liter
LB   Luria broth

L_{t}   DNA concentration

Lys   Lysine

M   molar

Me   metal

mer   unit length of nucleic acid

Met   methionine

mg   milligram
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg</td>
<td>magnesium</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mL</td>
<td>milliliter</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>N</td>
<td>Asparagine</td>
</tr>
<tr>
<td>N</td>
<td>nitrogen</td>
</tr>
<tr>
<td>N²-f-dG</td>
<td>N²-furfuryl-deoxyguanosine</td>
</tr>
<tr>
<td>N⁶-f-dA</td>
<td>N⁶-furfuryl-deoxyadenosine</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NFZ</td>
<td>nitrofurazone</td>
</tr>
<tr>
<td>NH</td>
<td>nitrile hydratase</td>
</tr>
<tr>
<td>NH₄</td>
<td>ammonium</td>
</tr>
<tr>
<td>nM</td>
<td>nanomolar</td>
</tr>
<tr>
<td>nm</td>
<td>nanometers</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>O⁶-BzG</td>
<td>O⁶-benzylguanine</td>
</tr>
<tr>
<td>O⁶-MeG</td>
<td>O⁶-methylguanine</td>
</tr>
<tr>
<td>OAc</td>
<td>acetate</td>
</tr>
<tr>
<td>O</td>
<td>oxygen</td>
</tr>
<tr>
<td>P</td>
<td>phosphorous</td>
</tr>
<tr>
<td>P</td>
<td>primer</td>
</tr>
<tr>
<td>P</td>
<td>Proline</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PGI</td>
<td>phosphoglucone isomerase</td>
</tr>
<tr>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenyl methyl sulfonyle fluoride</td>
</tr>
<tr>
<td>pol</td>
<td>polymerase</td>
</tr>
<tr>
<td>pols</td>
<td>polymerases</td>
</tr>
</tbody>
</table>
POOL Partial Order Optimum Likelihood
Pro Proline
Q Glutamine
R Arginine
Rh rhodium
RNA ribonucleotide
$R_t$ total protein concentration
S Serine
SDS-PAGE sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser Serine
SO$_4$ sulfate
SOS save our souls
SSB single stranded DNA binding protein
T Threonine
T thymine
T template; length of fully extended primer
t$_C$ 1,3-diaza-2-oxophenothiazine
TCA trichloroacetic acid
THEMATICS Theoretical Microscopic Anomalous Titration Curve Shapes
Thr Threonine
TLS translesion synthesis
tRNA transfer ribonucleic acid
Tyr Tyrosine
UV ultraviolet
V Valine
Val Valine
$V_{\text{max}}$ maximum velocity
W Tryptophan
wt wild-type
XPV Xeroderma pigmentosum variant
Y Tyrosine
This chapter is reproduced with permission from the following


Educational Use
Frontiers in Bioscience grants permission to all authors, readers and third parties of educational nature to reproduce and use published material and online resources as part of another publication or entity. This permission is granted free of charge provided that:
There is no charge, submission fee, royalty, honorarium, or any other monetary rewards for the use of the figure by the author, user, website, publisher, organizer or any other entity using the material.

The material is properly credited by including citing the source within the text or legend and including the full citation of the article in the reference section of educational material. When available, the DOI link should also be provided. If reproduced in CD format, the reference should be included in the same page that the material is included. If reproduced on a website, the reference should be linked to the article published in the Frontiers in Bioscience. Users who do not know the URL of the link can request it by providing the citation in an email to fbs@bioscience.org.

If used online, the use should be for a timeline not longer than 1 month. The educational use includes, for example, the use of a figure, table or text in a presentation, another article, a book chapter, newsletter, thesis, dissertations, classroom material, academic course, academic conference material, training material or posting of an abstract on a website. If your use complies with the above guideline, you do not need to obtain permission from Frontiers in Bioscience for the use of material.

and


Journal of Nucleic Acids is an open access journal. Open access authors retain the copyrights of their papers, and all open access articles are distributed under the terms of the Creative Commons Attribution license, which permits unrestricted use, distribution and reproduction in any medium, provided that the original work is properly cited.
Chapter 1: Background

1.1 Introduction

Since the structure of DNA was determined (Franklin 1953; Watson 1953), biochemists have sought more detailed ways to study DNA and the proteins that interact with it (Kornberg 1957; Kleppe 1971), notably the DNA polymerases that carry out the job of copying the molecule of life. DNA polymerases generally adopt a right-hand fold, in which the thumb and fingers bind DNA and nucleotide (Figure 1.1) (Patel 2001; Federley 2010). DNA polymerases add nucleotides to the growing DNA strand via nucleophilic attack of the free 3’ hydroxyl group of the DNA primer on the alpha phosphate of the incoming deoxynucleotide with release of pyrophosphate.
Figure 1.1 Comparison of the overall folds of (a) a replicative DNA polymerase, *Bacillus stearothermophilus* DNA pol I (Johnson 2003) and (b) a Y-family DNA polymerase, *Sulfolobus solfataricus* Dpo4 (Ling 2001). The respective thumb domains are shown in green, palm domains are in red, and fingers domains are in blue. The little finger domain unique to the Y family DNA polymerases is in purple (Ling 2001). The “vestigial” exonuclease domain of *Bs* pol I has been omitted for clarity (Kiefer 1997) (c) Polymerase catalyzed DNA replication (phosphoryl transfer) reaction. Polymerization of DNA occurs at a free 3’ hydroxyl group of the deoxyribose. Polymerases use a divalent magnesium ion (Me\(^{2+}\)) to coordinate the negative charges of both the phosphate groups and the aspartic acid or glutamic acid in the active site of the polymerase (Berdis 2009).

DNA polymerase active site residues, which are usually glutamate or aspartate and are located in the palm domain, coordinate divalent magnesium ions that serve to activate the 3’OH nucleophile (Figure 1.1) (Adler 1958; Bessman 1958; Kornberg 1958; Lehman 1958; Burgers 1979). The catalytic cycle is generally accompanied by conformational changes in the fingers domain. In replicating DNA, DNA polymerases have to be able to
form all four base pair combinations specifically and efficiently in order to maintain the integrity of the genome (Figure 1.2); however, when replication errors occur, the mismatched bases can be removed by the exonucleolytic proofreading function of DNA polymerases (Berdis 2009). Replicative DNA polymerases possess an exonuclease domain that may be part of the same or a separate polypeptide that utilizes a metal-dependent mechanism to excise mismatched bases (Berdis 2009; Reha-Krantz 2010). The proofreading process involves translocation of the primer terminus from the polymerase active site to the exonuclease active site; after the phosphodiester bond is hydrolyzed to remove the mismatched base, the primer strand re-anneals to the template so that polymerization can continue (Lee 2010; Reha-Krantz 2010). Replication errors that escape proofreading can be repaired by the mismatch repair system (Friedberg 2006).

Based on sequence conservation, DNA polymerases are divided into A, B, C, D, X, and Y families. The A and B family DNA polymerases can be involved in replication or repair, whereas members of the C family are involved in DNA replication (Rothwell 2005). X family DNA polymerases are involved in repair and Y family DNA polymerases are specialized for copying damaged DNA (Rothwell 2005) in a process known as translesion synthesis (TLS). In general, replicative DNA polymerases cannot
copy damaged DNA; rather, a specialized TLS polymerase must be recruited to extend
primers a sufficient distance past distortions in DNA templates to allow replicative DNA
polymerases to recover synthesis (Fuchs 2004; Fujii 2004; Pages 2005; Jarosz 2009).
DNA replication past damage or unusual DNA structures requires the ability to both
insert a nucleotide opposite a modification in the template as well as to extend the newly-
genenerated primer beyond that position.

Translesion synthesis (TLS) was first postulated almost forty years ago (Radman
1975). It was observed that DNA damage induced the *E. coli* SOS response, which is
accompanied by mutagenesis of the DNA (Radman 1975). Originally DNA damage-
induced mutagenesis was thought to result from the modification of replicative DNA
polymerases, which allowed them to bypass DNA damage, albeit sometimes
mutagenically (Woodgate 1989). However it was later discovered that the UmuC/UmuD′
complex (UmuD′2C, pol V) and DinB (pol IV) are Y family DNA polymerases that have
the specialized ability to carry out potentially mutagenic TLS (Yang 2005).

Y family DNA polymerases (Ohmori 2001) are found throughout all domains of
life; they have five conserved sequence motifs but the overall size of the proteins can
vary considerably (Figure 1.3) (Ohmori 1995; Ohmori 2001; Boudsocq 2002; Friedberg
2006; Yang 2007; Pata 2010).
In addition to bacterial pol IV and pol V, the eukaryotic members of the family include Rev1, pol eta, pol kappa, and pol iota (Ohmori 2001). Y family polymerases possess a domain known as the ‘little finger’ domain found only in the Y family (Ling 2001). The Y family polymerases are characterized by small finger and thumb domains relative to replicative DNA polymerases, which result in an open, solvent-accessible active site in the palm domain of Y family members (Yang 2005; Chandani 2010). The active site of replicative polymerases contains an ‘O-helix’, the role of which is to act as a steric check on fidelity and allow only a correct base pair to be formed (Kaushik 1996; Ogawa 2001; Beard 2003). Y family polymerases lack the O-helix, contributing to their more open and flexible active sites and allowing them to accommodate lesions on the DNA template (Ling 2001; Jarosz 2007). The available crystal structures of Y family DNA polymerases (Pata 2010) tend to support the model of an open active site, as seen in the structure of *Sulfolobus solfataricus* Dpo4 in complex with DNA containing a thymine-thymine (T-T) cyclobutane pyrimidine dimer (Ling 2003). This structure demonstrates that both thymines are accommodated in the active site simultaneously (Ling 2003). Structures of other Y family DNA polymerases with or without DNA also generally show that these proteins have small finger domains and open, solvent-accessible active sites, suggesting a
structural basis both for their ability to accommodate DNA lesions and for their low fidelity when copying undamaged DNA (Ling 2001; Silvian 2001; Zhou 2001; Yang 2005; Chandani 2010; Pata 2010).

1.2 Transcriptional and Post-Translational Regulation

In *E. coli*, the expression of Y family polymerases along with other genes is induced via the SOS response to damaged DNA. This cellular response was named the SOS response by Miroslav Radman because there is a “danger signal which induces SOS repair” (Radman 1974). Evelyn Witkin suggested that there is a pathway in *E. coli* that is controlled by a repressor whose function is inactivated when DNA damage occurs and that again becomes activated as a repressor once the repair of DNA damage is complete (Witkin 1967). This repressor was discovered to be the LexA protein, the repressor of the SOS genes. The SOS response is initiated when a lesion in the DNA template prevents replicative polymerases from continuing with efficient replication, causing a region of single-stranded DNA (ssDNA) to develop (Figure 1.4).
Figure 1.4 DNA damaging agents lead to the formation of lesions in DNA that disrupt replication and induce the SOS response. Single-stranded DNA (ssDNA) develops and becomes coated with RecA creating a RecA/ssDNA nucleoprotein filament, which signals the initiation of the SOS response. At least 57 genes are regulated by the LexA repressor, which represses the SOS genes by binding to consensus sequences (“SOS boxes”) in the promoter regions. LexA cleaves itself upon its interaction with the RecA/ssDNA nucleoprotein filament (Friedberg 2006; Simmons 2008). The cleavage of LexA ablates its repressor function and allows for the expression of the *umuDC* and *dinB* genes, among others. UmuD₂ also undergoes a cleavage reaction facilitated by the RecA/ssDNA nucleoprotein filament to form UmuD'₂, the active form in SOS mutagenesis (Burckhardt 1988; Nohmi 1988; Shinagawa 1988). The dashed line in the UmuD₂ cartoon signifies that the arm is behind the globular domain, as the monomers are related to each other by a C₂ axis of symmetry. Both UmuD'₂C (pol V) and DinB (pol IV) perform translesion synthesis (TLS) to bypass DNA adducts. It should be noted that it is not yet known exactly how UmuC and UmuD'₂ interact and the cartoon merely indicates that they form a complex.

RecA is activated upon binding to ssDNA, forming a RecA/ssDNA nucleoprotein filament (Friedberg 2006). LexA then binds the RecA/ssDNA nucleoprotein filament, inducing LexA to cleave itself at its Ala⁸⁴-Gly⁸⁵ bond, approximately in the middle of the protein (Horii 1981). Once LexA is cleaved it no longer represses the SOS genes,
allowing at least 57 genes, including *umuC*, *umuD*, and *dinB*, to be expressed during the SOS response (Friedberg 2006; Simmons 2008).

The *umuD* gene products contribute an additional level of regulation of Y family DNA polymerases in *E. coli* (Friedberg 2006). Upon expression, UmuD₂ binds to the RecA/ssDNA nucleoprotein filament, stimulating the ability of UmuD to cleave itself at its Cys⁺²⁴-Gly⁺²⁵ bond and removing its N-terminal 24-amino acids to form UmuD'₂ by a mechanism similar to that of LexA (Burckhardt 1988; Nohmi 1988; Shinagawa 1988). The full-length UmuD₂ protein persists for approximately 20-40 min after expression is induced, after which time the cleaved form UmuD'₂ becomes the predominant form (Figure 1.4) (Opperman 1999). Full-length UmuD₂ and cleaved UmuD'₂ play distinct roles in the cellular response to DNA damage; UmuD₂ contributes to accurate DNA replication and repair while UmuD'₂ facilitates mutagenesis (Marsh 1985; Opperman 1999; Reuven 1999; Tang 1999; Sutton 2001; Godoy 2007; Ollivierre 2010). Therefore, this lag in the appearance of UmuD'₂ delays the use of a potentially mutagenic pathway, in part via direct interactions between the *umuD* gene products and Y family DNA polymerases.

### 1.3 DNA Polymerase IV: DinB

DinB is one of two Y family polymerases found in *E. coli* (Reuven 1999; Tang 1999; Sutton 2001). The *dinB* (*damage-inducible*) gene was identified as being induced upon treatment with DNA damaging agents (Kenyon 1980; Ohmori 1995). Subsequently, the *dinB* gene product was demonstrated to be a DNA-dependent DNA polymerase (Wagner 1999). DinB was also shown to possess the ability to accommodate misaligned
or bulged primer-template structures into its active site and to lack intrinsic 3′-5′
exonuclease proofreading activity (Wagner 1999).

DinB is expressed at a level of approximately 250 molecules per cell under non-
SOS induced conditions (Kim 2001). However, this number increases 10-fold after SOS
induction; therefore DinB is the most abundant DNA polymerase in *E. coli* during times
of cellular stress (Kim 2001). This level of upregulation of DinB leads to inhibition of Pol
III, decreasing the ability of Pol III to access DNA and ultimately leading to cell death
(Furukohri 2008; Uchida 2008).

A phenomenon known as adaptive mutagenesis involves *dinB*-dependent
increased mutability in starving, non-dividing cells (McKenzie 2001; Tompkins 2003). It
has been suggested that adaptive mutagenesis provides mutations that confer a selective
advantage in times of cellular stress (Foster 2007). DinB induction occurs late in
stationary phase and the higher levels may be maintained for at least several days with
maximum expression and mutagenesis occurring in cells that have active RNA
polymerase sigma factor (RpoS) (Layton 2003; Lombardo 2004). Adaptive mutagenesis
is a cellular starvation stress response system, which is partially distinct from the SOS
response. Notably, of the SOS genes, only DinB at elevated levels is required for stress-
induced mutagenesis (Galhardo 2009), although the exact molecular mechanism of
adaptive mutagenesis may not be entirely clear (McKenzie 2003; Slechta 2003; Hersh
2004). The GroE heat shock response chaperone system has also been shown to influence
DinB protein levels as well as adaptive mutagenesis, although no direct interaction
between GroE and DinB has been detected (Layton 2005).
Classically, expression of the \textit{dinB} gene is repressed by LexA and induced as part of the SOS response (Kenyon 1980; Simmons 2008). However, the \textit{dinB} gene can also be expressed under other conditions of cellular stress. For example, \textit{dinB} expression is induced by β-lactam antibiotic-mediated inhibition of the synthesis of bacterial cell walls independent of LexA (Perez-Capilla 2005). This suggests that transcription of the \textit{dinB} gene can be considered a general stress response mechanism. Increased mutagenesis by DinB, and possibly pol V, may be a contributing factor to antigenic variation or antibiotic resistance (McKenzie 2001; Cirz 2007; Smith 2007; Petrosino 2009).

1.4 Specificity of DinB

DinB displays a preference for certain damaged DNA substrates. For example, DinB possesses a 15-fold preference to insert C opposite $N^2$-furfuryl-dG and a 25-fold preference to extend from $N^2$-furfuryl-dG in comparison to undamaged dG in a DNA template (Figure 1.5) (Jarosz 2006; Jarosz 2009). While the endogenous source of $N^2$-furfuryl-dG has not yet been determined, by analogy to the formation of kinetin ($N^6$-furfuryl-dA), it may be a product of ribose oxidation (Barciszewski 1999). Strains in which \textit{dinB} has been deleted show a striking sensitivity to nitrofurazone and 4-nitroquinoline-1-oxide (4NQO), both of which are thought to form $N^2$-dG adducts, as well as possibly other adducts (Friedberg 2006; Jarosz 2006). DinB also efficiently and accurately bypasses $N^2$-(1-carboxyethyl)-2′-deoxyguanine ($N^2$-CEdG), which was detected in 1 in $10^7$ bases in melanoma cells and is formed as an adduct of methylglyoxal, a common byproduct of glycolysis (Figure 1.5) (Yuan 2008). The presence of DinB significantly increases bypass of the $N^2$-dG adduct of benzo[a]pyrene (B[a]P), a potent carcinogen consisting of a large, bulky polycyclic hydrocarbon (Figure 1.5) (Lenne-
Samuel 2000; Shen 2002; Fuchs 2004; Seo 2006). However, efficient bypass of some isomers of B[a]P also requires pol V, suggesting that even subtle changes in the conformation of adducts can alter how they are processed by DinB (Seo 2006). Acrolein is a potent toxin and has tumor initiating properties, but it is also an endogenous byproduct of fatty acid metabolism (Minko 2008). DinB inserts dCTP across from γ-hydroxypropano-deoxyguanosine (γ-HOPdG), the $N^2$-dG adduct of acrolein, as well as peptide cross-links to γ-HOPdG (Figure 1.5) (Minko 2008). In addition to bypass of DNA-peptide cross-links, DinB is also proficient in bypassing $N^2$-$N^2$-dG interstrand cross-links (Kumari 2008). DinB may have a functional duality as a bypass polymerase for certain metabolism-induced DNA lesions such as γ-HOPdG, $N^2$-CEdG, and $N^2$-furfuryl-dG, and as a general bypass polymerase capable of negotiating larger $N^2$-dG adducts such as benzo[a]pyrene.
Figure 1.5 Adducts of deoxyguanine bypassed by *E. coli* pol IV: \( N^2\)-dG-\( \gamma \)-hydroxypropano-dG (Minko 2008), \( N^2\)-(1-carboxyethyl)-2\'-dG (\( N^2\)-CE-dG) (Yuan 2008), \( N^2\)-furfuryl-dG (Jarosz 2006), \( N^2\)\,+\,[trans, anti]-benzo[a]pyryl-dG (Shen 2002; Fuchs 2004; Seo 2006). Unmodified dG is shown for comparison. The table shows the identity of the nucleotide inserted by pol IV opposite each lesion.

DinB exhibits varying efficiency of bypass of a number of lesions resulting from reactive oxygen species, including 8-oxoG, 2-oxoA, thymine glycol, 5-formyluracil, and 5-hydroxymethyluracil (Hori 2010), and is able to incorporate into DNA the oxidized nucleotides 2-oxo-dATP and 8-oxo-dGTP (Yamada 2006). DinB has been shown to bypass abasic (AP) sites *in vitro* generating either -1 or -2 frameshift mutations (Kobayashi 2002; Maor-Shoshani 2002). The presence of accessory proteins, specifically the beta processivity clamp and the gamma clamp loader, greatly increase the efficiency of bypass (Maor-Shoshani 2002) and decrease frameshift mutagenesis (Kobayashi 2002).
On undamaged DNA, DinB has a relatively high error frequency of $2.1 \times 10^{-4}$ for frameshift mutations and $5.1 \times 10^{-5}$ for base substitution mutations (Kobayashi 2002). The majority of the frameshift mutations are single base deletions (~81%), whereas a substantial number of frameshift mutations are two base deletions (~15%) (Kobayashi 2002). The mechanisms by which the -1 frameshift mutations occur in TLS by DinB is most likely a combination of Streisinger slippage (Streisinger 1966) and dNTP-stabilized misalignment in which a nucleotide downstream from the primer terminus is flipped out of the DNA helix and is skipped by DinB generating a -1 frameshift mutation (Figure 1.6) (Kobayashi 2002; Tippin 2004; Foti 2010).

![Figure 1.6](image)

**Figure 1.6** Mechanisms for the generation of -1 frameshift mutations by *E. coli* pol IV. Pol IV is represented by the crystal structure of Dpo4 (yellow) (Ling 2001), while UmuD₂ and UmuD′₂ are shown in blue. (a) dNTP-stabilized misalignment mechanism (Kobayashi 2002; Tippin 2004) (b). Template slippage mechanism I. Prior to SOS induction, full-length UmuD₂ regulates pol IV to enhance error-free synthesis (top). II. Upon SOS induction, cleaved UmuD′₂ does not bind pol IV and so pol IV is free to generate -1 frameshift mutations via the template slippage mechanism (bottom) (Streisinger 1966; Foti 2010).
The dNTP-mediated mechanism may be at odds however with the recently published mechanism of template slippage on homopolymeric runs by DinB (Foti 2010). This type of slippage, which generates -1 frameshift mutations, is inhibited by UmuD₂ (Figure 1.6) (Foti 2010). Base skipping occurs as the template strand opposite the 5′ terminus folds into the extrahelical space and the next nucleotide on the template strand becomes the base opposite the primer terminus (Kunkel 1988; Bebenek 1990). This mechanism was also observed in similar studies on the DinB ortholog *Sulfolobus acidocaldaricus* Dbh (DeLucia 2007).

### 1.5 DinB Variants

Currently, there is no high-resolution structure of DinB but homology models (Jarosz 2006; Seo 2009) have been constructed based on the crystal structures of other homologous proteins: *Sulfolobus solfataricus* Dpo4 (Ling 2001) and *Sulfolobus acidocaldaricus* Dbh (Silvian 2001). Therefore, interpretation of experiments in which DinB variants have been constructed relies on the use of homology models. Cells containing overexpressed wild-type DinB had have a mutation frequency of 6.8 x 10⁻³ (Wagner 1999), which is approximately 3600-fold greater than cells without overexpressed DinB (Wagner 1999). DinB mutations D8A, D8H, R49A, R49F, D103A, D103N (*dinB003*), and E104A led to between 850- and 3700-fold lower mutation frequencies (Figure 1.7) (Wagner 1999). D103 and E104 reside in the S[LI]DE box whose negatively charged residues in the active site coordinate the divalent magnesium ions needed for adding the incoming nucleotide to the DNA primer (Wagner 1999; Ling 2001). Along with D103 and E104, D8 is strictly conserved (Wagner 1999). R49 is predicted to lie in a loop region that is near the incoming nucleotide.
Figure 1.7 Model of DinB with some important residues highlighted (Jarosz 2006). F13 (brown) is involved in accommodation of the lesion in the active site (Jarosz 2006). D8 (green), D103 and E104 (orange), R49 (purple) each when mutated confer a lower mutation frequency than wild-type DinB (Wagner 1999). D103 and E104 are involved in coordinating the divalent magnesium ion necessary for catalysis (Wagner 1999). Y79 (black) is important in regulating the steric gate residue (Jarosz 2009). Residues 346QLVLGL351 (pink) comprise one of the interaction sites with the beta clamp (Becherel 2002), while 303VWP305 (red) comprise the site of interaction with the dimer interface of the beta clamp (Bunting 2003).

The steric gate is a single residue in DNA polymerases that prevents the incorporation of ribonucleotides into DNA by sterically occluding nucleotides with a 2′ hydroxyl group (Astatke 1998). The steric gate residue of Y family DNA polymerases is most frequently tyrosine or phenylalanine (Jarosz 2006; DeLucia 2007; Shurtleff 2009). Mutation of the steric gate residue F13 in DinB has a marked effect on the ability of DinB to discriminate against ribonucleotides, increasing the frequency of misincorporation from $<10^{-5}$ to $10^{-3}$ (Jarosz 2006). It was hypothesized that the pocket in which F13 resides is involved in the accommodation of a lesion in the active site (Figure 1.7) (Jarosz 2006). The substitution F13V inhibits the ability of DinB to bypass N$_2$-furfuryl-dG, and also slightly enhances the ability of DinB to replicate undamaged DNA (Jarosz 2006).
Near the steric gate residue and conserved among all orthologous DinBs, Y79 is hypothesized to regulate the function of the steric gate residue (Jarosz 2009). Mutations at this position have a modest effect on primer extension on undamaged DNA but prevent DinB from extending more than a few nucleotides beyond a lesion and result in extreme cellular sensitivity to nitrofurazone (Jarosz 2009). Mutation of several other residues that are distal to the active site have been shown to specifically affect the extension step of TLS (Chapter 2) (Walsh 2012). As demonstrated in these studies, single mutations can have a large effect on both the DNA replication and TLS activities of DinB.

1.6 Cellular Interactions of DinB

DinB is regulated through protein-protein interactions with UmuD₂, RecA, and NusA, as well as with the beta clamp (Wagner 2000; Godoy 2007). Addition of RecA and UmuD₂ to a primer extension assay in which there are correctly paired bases at the primer terminus increases the polymerization proficiency of DinB [Godoy 2007]. It appears that the deleterious -1 frameshift mutator activity of DinB is a result of the elevated number of molecules of DinB present in a cell relative to the amount of UmuD₂ present [Godoy 2007]. Indeed, co-upregulation of both UmuD₂ and DinB suppresses the -1 frameshift mutation activity, while -1 frameshift mutations are elevated in the absence of umuD (Figure 1.4) [Godoy 2007]. Modeling studies suggest that RecA and UmuD₂ may suppress the -1 frameshift activity innate to DinB by decreasing the openness of the DinB active site [Godoy 2007]. This regulation may explain the dual nature of the polymerase activity of DinB, which accurately bypasses bulky N²-dG adducts but is also responsible for highly mutagenic -1 frameshift mutations [Brotcorne-Lannoye 1986; Kobayashi 2002; Jarosz 2006]. Deletion of umuD did not affect DinB-dependent
resistance to nitrofurazone, suggesting that the -1 frameshift mutator activity and TLS functions are to some extent distinct [Godoy 2007]. DinB residue F172 is highly conserved in DinB sequences from organisms that also harbor umuD and has been shown to mediate the interaction between DinB and UmuD (Figure 1.7) [Godoy 2007]. DinB also interacts physically with the beta processivity clamp. In the presence of the beta clamp, DinB is recruited to the primer terminus to form a stable complex with DNA, which substantially stimulates its processivity [Napolitano 2000]. The specific residues involved in this protein-protein interaction have been identified as $^{346}\text{QLVLGL}^{351}$ [Becherel 2002; Lenne-Samuel 2002], which is at the C-terminus of DinB (Figure 1.7). The co-crystal structure of the little finger domain of DinB with the beta clamp reveals a second interaction site between the two proteins at DinB residues $^{303}\text{VWP}^{305}$ and near the dimer interface of beta [Bunting 2003]. When the structure of full-length Dpo4 was superimposed on the structure of the DinB little finger, the DNA polymerase did not seem to be in the proper position to access the primer terminus [Bunting 2003]. This suggests that this conformation, while likely not catalytically relevant, may be one way in which the beta processivity clamp facilitates recruitment of DinB to replication forks [Bunting 2003; Wagner 2009]. DinB can remove pol III from the beta clamp when pol III is stalled at a primer terminus in vitro thus inhibiting the continuation of DNA synthesis by the pol III holoenzyme [Furukohri 2008].

Recently, a role for the NusA protein in stress-induced mutagenesis has been found that involves an interaction with DinB. NusA plays an important role in the elongation, termination, and anti-termination phases of transcription [Greenblat 1981; Farnham 1982; Liu 1996]. It was shown that DinB and NusA physically interact with one
another, so it was proposed that NusA recruits DinB to gaps that stall RNA polymerase during transcription [Cohen 2009]. Though the exact location of the interaction has yet to be identified, the C-terminal domain of NusA and surface residues near the \textit{nusA11}^{Ts} mutation are likely sites [Nakamura 1986; Cohen 2009]. Genetic interactions have been observed between \textit{nusA} and both \textit{dinB} and \textit{umuDC}, which may indicate a genetic link between TLS and transcription [Cohen 2009]. Moreover, NusA plays a role in transcription-coupled nucleotide excision repair as well as in recruiting DinB for transcription-coupled TLS of lesions that result in gaps in the DNA template that disrupt RNA polymerase [Cohen, Lewis 2010]. Additionally, NusA has been found to be required for DinB-dependent stress-induced, or adaptive, mutagenesis [Cohen, Walker 2010].

### 1.7 Non-natural Nucleotides as Probes for Polymerases

The genetic code is continuously expanding with new nucleobases designed to suit specific research needs. These synthetic nucleotides are used to study DNA polymerase dynamics and specificity and may even inhibit DNA polymerase activity. The availability of an increasing chemical diversity of nucleotides allows questions of utilization by different DNA polymerases to be addressed. Beyond the utility of non-natural nucleotides as probes of DNA polymerase specificity, such entities can also provide insight into the functions of DNA polymerases when encountering DNA that is damaged by natural agents. Thus, synthetic nucleotides provide insight into how polymerases deal with non-natural nucleotides as well as into the mutagenic potential of non-natural nucleotides.
1.8 Uses of Non-natural Nucleotides

Solid phase nucleic acid synthesis of DNA molecules facilitates the site-specific incorporation of a wide range of chemically modified bases and sugar-phosphate backbones, allowing the roles of specific atoms in DNA function and recognition to be probed. Synthetic non-natural nucleobases are useful for a variety of studies of DNA polymerase function, such as studies of DNA polymerase specificity, mutagenesis, and dynamics, as well as fluorescence resonance energy transfer (FRET) analysis of DNA polymerase interactions with DNA. The study of mutagenesis facilitated by DNA polymerases has attracted increasing interest because replication defects can lead to certain human diseases like the cancer-prone syndrome Xeroderma Pigmentosum-Variant (XPV) (Cordonnier 1999; Bresson 2002) and other diseases (Pages 2002; Loeb 2008; Takata 2011), as well as potentially contribute to antibiotic resistance (Napolitano 2000; Miller 2004). Moreover, specialized damage-bypass DNA polymerases are implicated in conferring cellular tolerance to cancer chemotherapy agents that act via DNA damage, thereby decreasing their effectiveness (Bassett 2004; Albertella 2005a; Albertella 2005b; Chen 2006; Ho 2011).

The four canonical bases vary in their chemical and geometric properties, but the C1’-C1’ distance of the standard Watson-Crick base pairs and the backbone C-O-P-O-C bonds remain constant regardless of the particular base pair (Kool 2002). Expansion of the nucleobase alphabet must take some of these structural considerations into account; usually non-natural bases need to have similar geometries as the natural bases, usually but not necessarily retain some level of hydrogen bonding capabilities, and usually have π
electron systems in order to retain the stability provided by base stacking. Hydrophobicity and base stacking interactions are also important for DNA structure (Kool 2002).

1.9 Synthetic Abasic Sites

Because Y family DNA polymerases are known to copy non-canonical DNA structures, their proficiency at copying synthetic abasic sites has been examined in some detail. The model Y family DNA polymerase *Sulfolobus solfataricus* Dpo4 copies synthetic abasic sites mainly by incorporating dA but also by generating small deletions (Fiala 2007). *E. coli* DinB (DNA pol IV) efficiently copies DNA containing a synthetic abasic site, but primarily by generating (-2) deletions (Maor-Shoshani 2003). Even though both *E. coli* Y family DNA polymerases can copy DNA containing abasic sites, Pol V is used to bypass abasic sites *in vivo*, probably because base substitutions are generally less harmful than frameshift mutations (Maor-Shoshani 2003). Human DNA pol eta copies abasic sites by incorporating predominantly dA but also dG (Masutani 2000; Kokoska 2003), whereas human pol kappa incorporates predominantly dA but also generates one nucleotide deletions (Zhang 2000; Sherrer 2011). Y-family member Rev1 from yeast incorporates dC opposite abasic sites, which have been suggested to be the cognate lesion of Rev1 (Otsuka 2002; Kim 2011; Pryor 2011). Due to the complicated responses of even the relatively forgiving Y family DNA polymerases to the synthetic model abasic site, it has been demonstrated that multiple DNA polymerases may be used to bypass DNA damage efficiently while minimizing mutations (Johnson 2000; Shachar 2009).
1.10 Minimal Requirements for Polymerization and Base Size Tolerance

Y family DNA polymerases are able to copy DNA containing non-canonical structures ranging from abasic sites to bulky DNA adducts (Washington 2003; Yang 2003; Jarosz 2007; Yang 2007; Waters 2009; Walsh 2011). Therefore, it was of interest to determine the minimal features of DNA required for replication. Short (three or 12) chains of methylene (CH$_2$) residues in the middle of canonical DNA templates were used to probe tolerance for minimal DNA backbones. *E. coli* pols I, II, and III were unable to replicate either DNA structure. On the other hand, both pols IV and V could replicate the three- or 12-methylene linker-containing DNA *in vitro*, although in an analogous situation to abasic sites, only pol V is observed to replicate these unusual structures *in vivo* (Maor-Shoshani 2003). Human DNA polymerases showed more subtle differences, in that pols eta, kappa, and iota could replicate a three-methylene linker by inserting nucleotides opposite the non-instructional segment, but only pols eta and kappa could fully bypass the modified gap (Adar 2006). Pols eta and iota could insert nucleotides opposite the 12-methylene linker, whereas pol kappa had little to no activity, and none of these three polymerases could completely bypass the 12-methylene linker (Adar 2006). Clearly, at least some Y family DNA polymerases are capable of replicating non-DNA segments.

In order to probe the size tolerance for bases in the active site, a series of dG analogs with increasingly large substituents at the $N^2$ position in the minor groove were constructed and used as the template base with a range of DNA polymerases. The $N^2$ modifications included methyl, ethyl, isobutyl, benzyl, CH$_2$-napthyl, CH$_2$-anthracenyl, and, in some cases, CH$_2$-benzo[a]pyrenyl derivatives (Choi 2004; Choi 2005; Choi
Bacteriophage T7 DNA polymerase (exonuclease) and HIV-1 reverse transcriptase are both able to bypass the $N^2$-methyl derivative efficiently, although significantly less efficiently than unmodified DNA, but are not able to bypass any of the larger adducts (Choi 2004). Moreover, even the methyl substituent caused a high frequency of misincorporation (Choi 2004). On the other hand, each of the human Y family DNA polymerases is more tolerant of the size-expanded bases (Choi 2005; Choi 2006a; Choi 2006c; Choi 2006d; Choi 2008). Rev1 is the most tolerant of $N^2$-dG-substitutions, followed by pol iota and pol kappa, whereas pol eta is the least tolerant, showing a decrease in activity of approximately two orders of magnitude between the CH$_2$-napthyl and CH$_2$-anthracenyl substituents (Choi 2008). A similar analysis of $O^6$-substituted bases showed that only Rev1 and pol iota could tolerate size-expanded substituents up to the benzyl substitution, but pol eta and pol kappa showed decreased activity even with an $O^6$-methyl substitution (Choi 2006b; Choi 2008). The use of a series of well-defined synthetic base modifications provides insights into the steric constraints of DNA polymerase active sites and allows detailed comparisons to be made between replicative and damage-bypass polymerases.

### 1.11 Fluorescent Nucleotide Analogs

The most common fluorescent base analog in use today is 2-aminopurine (2AP), which can form hydrogen bonds and base pair with either of the pyrimidines thymine or cytosine (Figure 1.8) (Wilhelmsson 2010). A recent crystal structure of DNA containing a 2AP:dC base pair in the active site of the Y567A variant of RB69 DNA polymerase suggests that the 2AP:dC pair may contain a bifurcated hydrogen bond between $N^2$-H of 2AP and N3 and O2 of dC (Reha-Krantz 2010).
In this example, the Y567A active site mutation in the nascent base-pair binding pocket is both less discriminating in the formation of mismatched base pairs and is better able to extend mismatched primer termini (Reha-Krantz 2010). The modified base 2AP is commercially available and has been used to study a number of DNA binding protein interactions including KF (Bloom 1993), EcoRI, DNA methyltransferase (Allan 1996), endonuclease (Lycksell 1987), and uracil DNA glycosolate (Stivers 1999). The fluorescence of this analog is sequence-context dependent, with the most pronounced effect occurring when the base is surrounded by other purines; much like other fluorescent nucleobases its fluorescence is quenched when it is within DNA (Ward 1969). KF has been shown to utilize 2AP and the fluorescence has been used to give insights into the dynamics of this protein as it synthesizes DNA (Bloom 1993; Hochstrasser 1994; Joyce 2008). For example in one FRET experiment with a labeled KF, the mechanism of the fingers closing conformational change was studied (Joyce 2008) and was found to be influenced by the added nucleotide. Specifically, mismatched nucleotides are detected before the polymerase “closes” on the DNA suggesting that the mismatched nucleotide itself may destabilize the “open” polymerase conformation (Joyce
2008). The role in the conformational change of the divalent cation (usually Mg$^{2+}$ or Ca$^{2+}$ but in this case an “exchange inert” Rh(III)) was also probed using 2AP (Lee 2009) and it was found that dNTP binding in the absence of the correct ion can induce the conformational shift (Lee 2009).

Fluorescence spectroscopy with 2AP can be used to study DNA polymerization on a millisecond time scale, and probe single events like nucleotide addition, base pairing interactions and subsequent excision via nuclease activity (Bloom 1993; Hochstrasser 1994). Insertion kinetics have been measured for the monophosphate version of 2AP (dAPMP vs dAMP) and dAPMP is found to be misincorporated at similar rates to the incorporation of the natural triphosphate dATP opposite dT by KF (Bloom 1993). This makes 2AP useful in studying polymerase activity as it is misincorporated about as frequently as dA is incorporated. However, this incorporation is influenced by the sequence surrounding the primer terminus, with double the rate of misincorporation of 2AP triphosphate if the nearest neighbor to the nascent base pair is dG, dC, or dA, as compared to dT (Bloom 1993).

Y family polymerases also have been studied using 2AP. Dbh adds dTTP correctly opposite 2AP in the template strand and binds various DNA substrates containing 2AP with $K_D$ values similar to those of natural DNA substrates (DeLucia 2007). Use of 2AP to monitor conformation changes during the base-skipping phenomenon, which can generate frameshift mutations as seen with Y family polymerases, provides evidence that the misincorporation pathway is distinct from the correct dNTP incorporation process (DeLucia 2007). Fluorescence from 2AP has been
also used to probe the proofreading mechanism by which bases are excised via nuclease activity of phage T4 polymerase (Subuddhi 2008).

The analog 2AP has been used together with the base analog pyrrolo-dC as a FRET pair as the excitation and emission wavelengths of these two nucleotide probes are compatible (Marti 2006), though this pair has not yet been utilized to study DNA polymerases. Pyrrolo-dC alone has been used to study DNA/RNA hybrids (Dash 2004), single-stranded DNA hairpins (Zhang 2009), and base pair flipping (Yang 2008). Two potential drawbacks of using 2AP are the sequence dependence of its fluorescence and that it can perturb the DNA structure or be mutagenic if it forms a wobble pair with dT (Wilhelmsson 2010). A 2AP:dT base pair destabilizes duplex DNA by ~8 °C relative to a dA:dT base pair (Petruska 1985).

1.12 Major Groove Adducts

The synthetic cytosine analog tC was developed first by Mateucci (Lin 1995), but then used as a probe of DNA polymerases by Wilhelmsson and co-workers (Wilhelmsson 2001; Sandin 2007; Stengel 2007). Although it has fluorescent properties, the extrahelical system that lies on the major groove side of DNA was the major driving factor in the research described in Chapter 3. The fluorescence quantum yield of this nucleotide analog, unlike 2AP, is not sensitive to the surrounding environment (Stengel 2007; Sandin 2009). This base also is incorporated into DNA, shows canonical base pairing with guanosine (Figure 1.9), and does not perturb the B-form structure of DNA. In fact, a dG:tC base pair stabilized DNA by 3 °C relative to a dG:dC base pair (Wilhelmsson 2010).
Different DNA polymerases have different efficiencies in utilizing tC in template DNA and in incorporating tC into the growing DNA primer strand. For example, KF utilizes template tC in preference to a template C, as KF apparently has a flexible enough active site to accommodate the extra cyclic ring system. Klenow also preferentially incorporates the tC nucleotide triphosphate in the growing DNA strand. *E. coli* Y-family DNA polymerase DinB (pol IV) (Ohmori 2001), also utilizes the tC nucleotide triphosphate more efficiently than dCTP, similar to Klenow (Walsh 2011). Primer extension by DinB is inhibited unless the primer terminus is at least 3-4 nucleotides beyond the tC analog, which suggests that the “TLS patch” of nucleotides required beyond non-cognate bases for DNA polymerases to resume efficient synthesis is shorter for a Y family DNA polymerase than for replicative polymerases. Moreover, the striking asymmetry of the DinB active site has also been observed in the case of B family DNA polymerases human polymerase alpha and herpes simplex virus I DNA polymerase when probed with non-natural nucleotide analogs (Lund 2011).
1.13 Summary

*E. coli* Y family DNA polymerases play important roles in conferring resistance to DNA damaging agents. The two Y family polymerases present in *E. coli* are proficient for bypassing distinct sets of DNA lesions, which suggests that they play roles complementary to each other in cells faced with DNA damage. The Y family polymerases also regulate DNA replication in response to DNA damage and other replication stress. Because of their role in mutagenesis, the Y family polymerases may be involved in antigenic variation or antibiotic resistance. Thus, understanding both the regulation and the inherent basis of specificity of Y family DNA polymerases is critical. Also, non-natural nucleotides continue to provide an important tool for the study of DNA and its interacting protein partners. In particular, DNA polymerases that are responsible for the systematic replication of DNA, whether accurate or mutagenic, are required to specifically recognize and efficiently base pair with a large number of non-canonical DNA structures. An increasingly expanding genetic alphabet of non-natural nucleobases provides the ability to obtain an unparalleled level of detail about how DNA polymerases discriminate among many different DNA structures.
1.14 References


substitution mutation requires a different set of genes in *Escherichia coli."
Lin, K. Y., R. J. Jones and M. Matteucci (1995). "Tricyclic 2'-Deoxyctydine Analogs -
Syntheses and Incorporation into Oligodeoxynucleotides Which Have Enhanced
Binding to Complementary Rna." Journal of the American Chemical Society
117(13): 3873-3874.
Ling, H., F. Boudsocq, R. Woodgate and W. Yang (2001). "Crystal structure of a Y-
family DNA polymerase in action: a mechanism for error-prone and lesion-bypass
regulator RpoS in adaptive mutation and amplification in *Escherichia coli."
Genetics 166(2): 669-680.
Lone, S., S. A. Townson, S. N. Uljon, R. E. Johnson, A. Brahma, D. T. Nair, S. Prakash,
Lyksell, P. O., A. Graslund, F. Claesens, L. W. McLaughlin, U. Larsson and R. Rigler
(1987). "Base pair opening dynamics of a 2-aminopurine substituted Eco RI
restriction sequence and its unsubstituted counterpart in oligonucleotides."
translesion replication across an abasic site by DNA polymerase IV of
Escherichia coli." DNA Repair 2: 1227-1238.
polymerase V of *Escherichia coli* by processivity proteins." Biochemistry 41(48):
14438-14446.
Marsh, L. and G. C. Walker (1985). "Cold sensitivity induced by overproduction of
translesion synthesis by human DNA polymerase eta." EMBO Journal 19(12):
3100-3109.
"SOS mutator DNA polymerase IV functions in adaptive mutation and not
and spontaneous mutation in *Escherichia coli.*" Journal of Bacteriology 185(13):
3972-3977.


This chapter is reproduced with permission from the following

"Effects of non-catalytic, distal amino acid residues on activity of E. coli DinB (DNA Polymerase IV) " Jason M. Walsh, Ramya Parasuram, Pradyumna R. Rajput, Eriks Rozners, Mary Jo Ondrechen and Penny J. Beuning Environmental and Molecular Mutagenesis in press. http://dx.doi.org/10.1002/em.21730

Penny J. Beuning and Mary Jo Ondrechen conceived the study. Ramya Parasuram modeled the DNA, refined the homology model, carried out calculations for the residue predictions, and prepared figures. Pradyumma R. Rajput constructed the homology model. Eriks Rozners synthesized DNA constructs. Jason M. Walsh collected and analyzed the experimental data and prepared the manuscript draft, tables and figures.
This is a License Agreement between Jason M Walsh ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number
302051387153

License date
Nov 01, 2012

Licensed content publisher
John Wiley and Sons

Licensed content publication
Environmental and Molecular Mutagenesis

Book title

Licensed content author
Jason M. Walsh,Ramya Parasuram,Pradyumna R. Rajput,Eriks Rozners,Mary Jo Ondrechen,Penny J. Beuning

Licensed content date
Oct 4, 2012

Start page
n/a

End page
n/a

Type of use
Dissertation/Thesis

Requestor type
Author of this Wiley article

Format
Print and electronic

Portion
Full article

Will you be translating?
No

Order reference number

Total
0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively WILEY). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC’s Billing and Payment terms and conditions"), at the time that you opened your Rightslink account (these are available at any time at http://myaccount.copyright.com)

Terms and Conditions

1. The materials you have requested permission to reproduce (the "Materials") are protected by copyright.

2. You are hereby granted a personal, non-exclusive, non-sublicensable, non-transferable, worldwide, limited license to reproduce the Materials for the purpose specified in the licensing process. This license is for a one-time use only with a maximum distribution equal to the number that you identified in the licensing process. Any form of republication granted by this licence must be completed within two years of the date of the grant of this licence (although copies prepared before may be distributed thereafter). The Materials shall not be used in any other manner or for any other purpose. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Material. Any third party material is expressly excluded from this permission.

3. With respect to the Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Materials, or any of the rights granted to you hereunder to any other person.

4. The Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc or one of its related companies (WILEY) or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any
such right, license or interest with respect thereto.

5. NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

6. WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

7. You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

8. IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

9. Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

10. The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

11. This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

12. Any fee required for this permission shall be non-refundable after thirty (30) days from receipt.

13. These terms and conditions together with CCC’s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties’ successors, legal representatives, and authorized assigns.

14. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall prevail.

15. WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

16. This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

17. This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state’s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Wiley Open Access Terms and Conditions

All research articles published in Wiley Open Access journals are fully open access: immediately freely available to read, download and share. Articles are published under the terms of the Creative Commons Attribution Non Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. The license is subject to the Wiley Open Access terms and conditions:

Wiley Open Access articles are protected by copyright and are posted to repositories and websites in accordance with the terms of the Creative Commons Attribution Non Commercial License. At the time of deposit, Wiley Open Access articles include all changes
made during peer review, copyediting, and publishing. Repositories and websites that host the article are responsible for incorporating any publisher-supplied amendments or retractions issued subsequently.

Wiley Open Access articles are also available without charge on Wiley's publishing platform, Wiley Online Library or any successor sites.

Use by non-commercial users

For non-commercial and non-promotional purposes individual users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles, as well as adapt, translate, text- and data-mine the content subject to the following conditions:

The authors' moral rights are not compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be impugned).

Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.

If article content is copied, downloaded or otherwise reused for non-commercial research and education purposes, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.

Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

Use by commercial "for-profit" organisations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;
- Copying, downloading or posting by a site or service that incorporates advertising with such content;
- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack);
- Use of article content (other than normal quotations with appropriate citation) by for-profit organisations for promotional purposes;
- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;
- Use for the purposes of monetary reward by means of sale, resale, licence, loan, transfer or other form of commercial exploitation such as marketing products.

Print reprints of Wiley Open Access articles can be purchased from: corporatesales@wiley.com

Other Terms and Conditions:

BY CLICKING ON THE "I AGREE..." BOX, YOU ACKNOWLEDGE THAT YOU HAVE READ AND FULLY UNDERSTAND EACH OF THE SECTIONS OF AND PROVISIONS SET FORTH IN THIS AGREEMENT AND THAT YOU ARE IN AGREEMENT WITH AND ARE WILLING TO ACCEPT ALL OF YOUR OBLIGATIONS AS SET FORTH IN THIS AGREEMENT.

v1.7

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK50088515.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyight Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
Chapter 2: Effects of non-catalytic, distal amino acid residues on activity of E. coli DinB (DNA polymerase IV)

2.1: Introduction

DNA polymerases are generally highly efficient and accurate enzymes that use polymeric DNA as templates to synthesize new DNA. The DNA polymerase active site consists of negatively charged aspartate and glutamate residues that bridge the metal ions (2 Mg$^{2+}$) and position them into the correct orientation to activate and facilitate nucleophilic attack of the primer 3’-hydroxyl group onto the α-phosphate group of the incoming deoxynucleotide triphosphate (Berdis 2009). The metal ions also help stabilize the negative charge that accumulates on both this ternary complex transition state and on the pyrophosphate product of DNA synthesis (Berdis 2009). DNA is constantly subjected to insults from both environmental and endogenous sources (Friedberg 2006). The resulting damaged DNA generally cannot be copied efficiently or accurately by replicative DNA polymerases (Friedberg 2006). Thus, in addition to multiple DNA repair processes, organisms in all domains of life harbor specialized DNA polymerases that can copy damaged DNA in a process known as translesion synthesis (TLS) (Goodman 2002; Friedberg 2005; Prakash 2005; Friedberg 2006; Walsh 2011). Y family DNA polymerases are capable of carrying out TLS, which is an important mechanism of DNA damage tolerance, and also exhibit a relatively high error frequency when copying undamaged DNA. Because of their roles in DNA damage tolerance and in mutagenesis, Y family DNA polymerases are implicated in cancer (Cordonnier 1999; Bresson 2002; Pages 2002; Loeb 2008; Lange 2011), in resistance to DNA-damaging cancer chemotherapy agents (Bassett 2004; Albertella 2005; Chen 2006; Ho 2011), and are
proposed to play a role in antibiotic resistance (Napolitano 2000; Miller 2004; Cirz 2005; Cirz 2006; Cirz 2007). Notably, defects in Y family DNA pol eta can cause the cancer-prone disease Xeroderma Pigmentosum-Variant (XPV) (Kraemer 1985; Johnson 1999; Itoh 2000; Tanioka 2007; Inui 2008; Biertumpfel 2010).

DinB (DNA polymerase IV) is one of two Y family DNA polymerases in *Escherichia coli* (*E. coli*) (Friedberg 2002). DinB incorporates nucleotides across from deoxyguanine adducts at the $N^2$ position, namely carboxyethyl (Yuan 2008), benzo[a]pyrene (Shen 2002), and the furfuryl adduct (Jarosz 2006), with the latter being bypassed preferentially to that of natural dG in the template (Figure 2.1).

\[
\begin{align*}
\text{(a)} & \quad O \\
\text{(b)} & \quad O \\
\end{align*}
\]

Figure 2.1 Structural comparison of (a) natural deoxyguanosine and (b) $N^2$-furfuryl-deoxyguanosine ($N^2$-f-dG). DinB demonstrates a ~16 fold preference to incorporate dCTP opposite $N^2$-f-dG (Jarosz 2006) and thus $N^2$-f-dG is considered the cognate lesion of DinB.

The $N^2$-furfuryl-dG lesion is thought to arise from nitrofurazone (Jarosz 2006), which was used as an antibiotic in livestock until it was found to leave residue in edible tissues and was suspected of causing tumors in laboratory animals (Smith 1998). However, by analogy to the plant hormone kinetin ($N^6$-furfuryl-dA), the furfuryl moiety could also arise due to oxidation of ribose (Amasino 2005; Barciszewski 2007). Kinetin has also been found in DNA extracted from human cells (Barciszewski 2007) and in human urine.
(Barciszewski 2000). Incorporation of dCTP opposite $N^2$-furfuryl-dG is efficient and accurate for both DinB and its mammalian ortholog Pol κ (Jarosz 2006).

In addition to understanding the range of DNA damage that can be bypassed by DinB, it is of interest to determine how specific amino acids contribute to the specialized ability to bypass DNA damage. There are several important residues that have previously been identified in DinB (Walsh 2011), including the ‘steric gate’ residue (F13), mutation of which eliminates the enhanced ability of DinB to bypass its cognate lesion $N^2$-furfuryl-dG (Jarosz 2006). The steric gate residues found in many DNA polymerases (Astatke 1998; Bonnin 1999; Ogawa 2001; Yang 2002; DeLucia 2003; Shurtleff 2009; Sherrer 2011) including DinB (Jarosz 2006), prevents ribonucleotide incorporation (Patel 2000; Brown 2011). Residue Y79, which is near the steric gate residue and highly conserved in the DinB family, is thought to modulate the function of the steric gate residue (Jarosz 2009). Changing this residue has a profound effect on the ability of DinB to carry out the extension step of TLS, in which the primer is extended sufficiently beyond the lesion so that replicative DNA polymerases can resume replication (Jarosz 2009). This extension step of TLS may be as significant in overall bypass as the addition of a nucleotide opposite the lesion (Fujii 2004). The importance of positions other than the aromatic residues F13 and Y79 has been probed; specifically, mutations at position D8 (D8A, D8H), R49 (R49A, R49F), D103 (D103A, D103N), and E104 (E104A) confer a much lower mutation frequency (ranging from ~850 to ~3600 fold) than wild-type DinB (Wagner 1999). This is expected as D8, D103, and E104 are the highly conserved residues that coordinate the critical divalent cations (Wagner 1999; Ling 2001). In
addition, R49 lies in a loop region that is predicted to be in close proximity to the incoming nucleotide (Walsh 2011).

In order to systematically identify other residues that could contribute to polymerase activity, we applied computational methods to predict the functionally important residues in DinB. Theoretical Microscopic Anomalous Titration Curve Shapes (THEMATICS) uses perturbations in theoretical titration curves of ionizable residues to predict catalytic residues (Ondrechen 2001; Ko 2005; Wei 2007). Partial Order Optimum Likelihood (POOL) (Tong 2009; Somarowthu 2011b) is a monotonicity-constrained maximum likelihood machine learning methodology that incorporates the electrostatic features of THEMATICS as well as other input data, such as surface geometric properties and phylogenetic information, to rank all of the amino acid residues of DinB according to their probability of functional importance. Since there is no known experimentally determined crystal structure available for DinB, these methods were performed with a DinB model, which is based on homology to Dpo4 (Fiala 2007). POOL with THEMATICS and INTREPID (Sankararaman 2009) input features has been verified as a successful predictor of catalytic and binding residues (Tong 2009; Somarowthu 2011b) using the CSA-100 (Bartlett 2002; Porter 2004; Sankararaman 2009), a benchmark dataset of 100 non-homologous enzymes with experimental annotations of functionally important residues. Intriguingly, in addition to residues that immediately surround the substrate, other residues that are more distant from the substrate are often predicted to be important for activity. Such distal residues, although they do not come into direct contact with the substrate, may influence specificity or efficiency, possibly by their contact with those amino acid residues that do contact the substrate. These predictions of participation
in catalysis by distal residues has been verified experimentally for nitrile hydratase (NH) (Brodkin 2011), phosphoglucose isomerase (PGI) and ketosteroid isomerase (KSI) (Somarowthu 2011a), We typically define residues in contact with the substrate as first shell, those in contact with the first shell as second shell and those in contact with the second shell as third shell. Certain second-shell residues have been shown to have a profound impact on PGI function, while distal residues contribute little to the activity of KSI (Somarowthu 2011a), as predicted. Therefore, changing residues that are distally located from the active site can impact different enzymes differently. While the residues predicted by THEMATICS/POOL tend to be highly conserved residues, the use of these electrostatics-based methods for the prediction of active residues has distinctive advantages over simple selection of conserved residues. The chief advantage is selectivity (Wei 2007). Residues may be conserved for a variety of reasons and thus not all conserved residues are directly involved in catalysis. The specificity scores achieved by THEMATICS (Wei 2007) and POOL (Tong 2009; Somarowthu 2011b) are substantially higher that those obtained from simple sequence conservation-based methods. Indeed, as was demonstrated in earlier studies of NH (Brodkin 2011) and of KSI and PGI (Somarowthu 2011a), most of the residues predicted by THEMATICS prove to be important for catalysis, whereas negative control residues, i.e. conserved nearby residues that are not predicted by THEMATICS, are shown not to play a significant role in catalytic efficiency.

To our knowledge, systematic characterization of the contributions of remote residues to Y-family DNA polymerase activity has not yet been carried out, although remote residues in DNA pol η are associated with the cancer prone disorder Xeroderma
Pigmentosum-Variant (XPV) (Broughton 2002; Tanioka 2007; Biertumpfel 2010). Moreover, a systematic and thorough analysis of the roles of residues remote from the active site in DNA polymerase β has identified residues that are important for nucleotide selection and template alignment (Yamtich 2010). As one example, a triad of residues Arg333 Glu316 Arg182 in the fingers domain of pol β may disrupt packing of side chains in the hinge region and therefore alter the conformational change required for efficient DNA replication (Murphy 2011). Distal mutations in two separate domains in *Thermus aquaticus* DNA polymerase I can impact its ability to replicate DNA; S543N in the thumb region can hinder template-dependent pausing during replication (Ignatov 1999) and a number of mutations along the O helix (F667L, A661E, I665T) change the mutation spectrum (Suzuki 2000). Because of the large size of the DNA substrate and extensive contacts between DNA polymerases and DNA, it can be difficult to assign rigid definitions of shells. For the purposes of this work, we defined the first shell residues as the residues that are in contact with the incoming nucleotide substrate and the second shell residues as those that are in contact with the first shell residues.

THEMATICS (Ondrechen 2001; Ko 2005; Wei 2007) and POOL (Tong 2009; Somarowthu 2011b) were used to rank all amino acids of DinB in order of their probability of functional importance. Typically, the top 8-10% of the rank-ordered residues are used for functional residue prediction. Of the top 8% we chose to focus on the top ten residues. Among the top ten POOL-predicted residues are seven that are in direct contact with the substrate and three second- or third-shell residues (Table 2.1). The identified residues are H6, D103, E104, D8, K150, K157, Y106, R49, D10, K146 in POOL rank order (Table 2.1). As mutations at D8, R49, D103, and E104 have been
previously characterized (Wagner 1999), we constructed DinB variants that harbor the
single mutations H6L, K150A, K157A, K157I, Y106A, Y106F, D10E, D10N, and
K146A, and assessed the impact on catalytic efficiency of DinB with its cognate lesion
$N^2$-furfuryl-dG. We find a range of contributions of these remote residues to activity,
from modest effects to nearly complete loss of activity. Moreover, we find that these
DinB variants have greatly diminished ability to carry out the extension step of
translesion synthesis.
Table 2.1 Top 8% of POOL predicted residues Residues in bold, italic type are involved in metal coordination. Shaded residues are the top ten of the POOL-predicted residues.

<table>
<thead>
<tr>
<th>Rank</th>
<th>No.</th>
<th>Residue</th>
<th>Shell</th>
<th>Distance from dNTP (Å)</th>
<th>Closest first shell residue</th>
<th>Domain</th>
<th>Consurf Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>HIS</td>
<td>second</td>
<td>7.6</td>
<td>E104</td>
<td>Palm</td>
<td>-1.054</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>ASP</td>
<td>first</td>
<td>2.4</td>
<td>E104</td>
<td>Palm</td>
<td>-1.053</td>
</tr>
<tr>
<td>3</td>
<td>104</td>
<td>GLU</td>
<td>first</td>
<td>3.7</td>
<td>E104</td>
<td>Palm</td>
<td>-1.052</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>ASP</td>
<td>first</td>
<td>3.2</td>
<td>E104</td>
<td>Palm</td>
<td>-1.053</td>
</tr>
<tr>
<td>5</td>
<td>150</td>
<td>LYS</td>
<td>first</td>
<td>4.7</td>
<td>E104</td>
<td>Palm</td>
<td>-1.051</td>
</tr>
<tr>
<td>6</td>
<td>157</td>
<td>LYS</td>
<td>first</td>
<td>2.5</td>
<td>E104</td>
<td>Palm</td>
<td>-1.051</td>
</tr>
<tr>
<td>7</td>
<td>106</td>
<td>TYR</td>
<td>second</td>
<td>12.0</td>
<td>K150</td>
<td>Palm</td>
<td>-0.407</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>ARG</td>
<td>first</td>
<td>1.7</td>
<td>E104</td>
<td>Finger</td>
<td>-1.053</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>ASP</td>
<td>first</td>
<td>4.7</td>
<td>E104</td>
<td>Palm</td>
<td>-1.053</td>
</tr>
<tr>
<td>10</td>
<td>146</td>
<td>LYS</td>
<td>third</td>
<td>13.6</td>
<td>K150</td>
<td>Palm</td>
<td>-1.006</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>MET</td>
<td>first</td>
<td>3.3</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>98</td>
<td>GLU</td>
<td>second</td>
<td>12.9</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>158</td>
<td>PRO</td>
<td>second</td>
<td>5.8</td>
<td>K157</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>79</td>
<td>TYR</td>
<td>second</td>
<td>5.3</td>
<td>F13</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>54</td>
<td>ARG</td>
<td>first</td>
<td>2.4</td>
<td>E104</td>
<td>Finger</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>139</td>
<td>SER</td>
<td>second</td>
<td>6.8</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>66</td>
<td>CYS</td>
<td>third</td>
<td>14.8</td>
<td>E104</td>
<td>Finger</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>57</td>
<td>MET</td>
<td>second</td>
<td>5.8</td>
<td>E104</td>
<td>Finger</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>46</td>
<td>TYR</td>
<td>second</td>
<td>7.1</td>
<td>E104</td>
<td>Finger</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>153</td>
<td>SER</td>
<td>second</td>
<td>7.3</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>330</td>
<td>ARG</td>
<td>remote</td>
<td>14.3</td>
<td>E104</td>
<td>Little finger</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>11</td>
<td>CYS</td>
<td>first</td>
<td>2.6</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>101</td>
<td>SER</td>
<td>first</td>
<td>4.3</td>
<td>E104</td>
<td>Palm</td>
<td></td>
</tr>
</tbody>
</table>
2.2 Materials and Methods

2.2.1 Computational Methods

The DinB structure has not yet been determined, so a homology model was built with DNA and incoming nucleotide bound using the structure of the Dpo4 polymerase from *Sulfolobus solfataricus* with PDB ID 2imw (Fiala 2007) as the template using the homology feature in the YASARA/WHATif suite of programs (Krieger 2002). The model was refined with energy minimization using YAMBER3. The overall Quality Z-score of the final model was calculated to be -1.033 by YASARA. There are no outlier residues and 96.8% of residues are in the favored region of the Ramachandran plot (Lovell 2003). Because the DNA in the 2imw structure is a blunt-end DNA molecule, the DNA and the incoming nucleotide were modeled from 1jx4 (Ling 2001). The YASARA molecular modeling software was used to add missing atoms to the ligands bound in the final DinB model. The coordinates for the model are available on the Environmental & Molecular Mutagenesis journal website: doi: 10.1016/j.jmb.2011.03.069.

Ligands are removed prior to the analysis using THEMATICS. The electrostatic properties from THEMATICS (Ondrechen 2001), and the phylogenetic tree-based scores from INTREPID (Sankararaman 2009). (http://phylogenomics.berkeley.edu/intrepid/) were used as input features to POOL. POOL then generated an output file of all residues in the model, rank-ordered according to their probability of functional importance.

Shell definitions were established as follows: The distance between the incoming nucleotide and the predicted residues was calculated using YASARA. Any residue that is within a distance of 5 Å of the incoming nucleotide was considered to belong to the first shell, *i.e.* in direct contact with the incoming nucleotide. Any other residue that is within 5 Å of any of the first shell residues is considered to be in the second shell, and so on.
### 2.2.2 Experimental Methods

DinB was purified as described previously by Beuning, *et al.* (Beuning 2006) and stored in single-use aliquots at −80 °C. The DNA template containing a single $N^2$-furfuryl-dG adduct was prepared as described (DeCorte 1996; Jarosz 2006). DNA sequences used in primer extension assays are as follows: standing start primer, 31-mer 5′ GCATATGATAGTACAGCTGCAGCCGGACGCC 3′; MatchC primer, 32-mer, 5′ GCATATGATAGTACAGCTGCAGCCGGACGCCC 3′; and template containing $N^2$-furfuryl-dG (N), 61-mer, 5′ GGTTACTCAGATCAGGCCTGCGAAGACCTNGGCGT CCGGCTGCTGTACTATCATATGC 3′. DNA primer (standing start or MatchC) and the template containing $N^2$-furfuryl-dG were combined to a final ratio of 1:1 (100 nM) and annealed in annealing buffer [20 mM Hepes (pH 7.5) and 5 mM Mg(OAc)$_2$] by heating for 2 min at 95 °C, incubating at 50 °C for 60 min, and then cooling to 37 °C. The reactions were carried out with 100 nM $^{32}$P-end-labeled primer/template in a reaction buffer containing final concentrations of 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 (Beuning 2006). Experiments were carried out with dNTP concentrations ranging from 1 μM-5000 μM, unless otherwise noted, and DinB concentration varied based on the extent of activity of each variant. An aliquot for the zero point was removed prior to addition of dNTP, and reactions were initiated by the addition of dNTP (Beuning 2006). The final reaction volumes were 30 μL. Time points were typically taken from 0.5 min to 30 min and reactions were quenched with 85% formamide, 50 mM ethylenediaminetetraacetic acid, 0.025% xylene cyanol, and 0.025% bromophenol blue (Beuning 2006). To determine kinetic parameters, conditions were chosen such that less than 25% of the substrate was converted to product. Quenched reaction products were analyzed on
denaturing (8 M urea) 16% polyacrylamide gels, which were subsequently imaged on a Molecular Dynamics storage phosphor imaging screen with a Storm 860 imager. ImageQuant software (GE Healthcare) was used to analyze data. Kinetic parameters $V_{\text{max}}$ and $K_m$ were determined by assessing the percentage of primer extended at various time points in the experiment and were derived using GraphPad Prism® nonlinear regression analysis software (Segel 1993). DNA binding experiments were conducted by titrating 25 µM annealed DNA into a 60 µL solution of 1 µM DinB as previously described (Walsh 2011).

2.3 Results

Table 2.1 shows the top 8% of the POOL predicted residues from the homology model of DinB. Of these, 12 residues belong to the first shell, 11 to the second shell, three to the third shell, and two are in more distant shells. Among the POOL predicted residues, there are 20 in the Palm domain, six in the finger domain and two in the little finger domain.

When used as a standalone method without POOL and other input features, THEMATICS predicts two clusters of residues using a statistical cutoff of 0.99 (Wei 2007). The first cluster has eight residues H6, D8, E104, Y106, K150, K157, D103 and R49 of which H6 and Y106 are second-shell residues. These residues represent the eight top-ranked residues also predicted by POOL. THEMATICS also predicts a second cluster of two residues, R35 and D252, that is somewhat distant from the active site. These residues are not ranked high in the POOL predictions, but may play a role in interaction of the finger and little finger domains with DNA upon DNA binding and catalysis. In this
paper, we have investigated experimentally the top ten residues predicted by POOL, of which three are residues outside the first shell.

Some of the highly POOL-ranked residues are the catalytic residues D8, D103, and E104, and R49, which have previously been shown to be important for activity, and other residues that are near the incoming nucleotide (Wagner 1999). Therefore, POOL correctly predicts the importance of active site residues. These methods also identified a number of distal residues, those not directly involved in binding the substrate, that are likely to contribute to activity. In this work, we focused on the ten most highly ranked POOL-predicted residues (Table 2.1). The predicted residues studied here form a crescent-shaped cluster around, and including, the metal-coordinating residues D8, D103, and E104 (Figure 2.2).
Figure 2.2 Homology model of *E. coli* polymerase IV DinB. Highlighted residues include the identified distal residues (shown in magenta stick models) and their position in relation to known important residues (shown in space filling models), including the main active site residues D8 (orange), D103 and E104 (green) (Wagner 1999); the steric gate and its modulator, F13 (Jarosz 2006) and Y79 (Jarosz 2009), respectively (black); and the important residue R49 (light blue) (Wagner 1999). Images were rendered using the UCSF Chimera package (Pettersen 2004).

All of these residues are ionizable and may have either a steric or an electronic impact on the ability of DinB to catalyze dCTP addition across from *N*²-furfuryl-dG, its cognate lesion. For every variant constructed, primer extension activity with a mixture of nucleotides as well as fidelity of incorporation were determined. Each variant that was active incorporated only dCTP opposite template *N*²-furfuryl-dG (Figure 2.3). For those variants that were inactive, we also determined their ability to bind DNA. We find that
mutations at these amino acid positions have a variety of impacts on DinB function, as described in detail below.

![Diagram of single nucleotide incorporation](image)

**Figure 2.3** Schematic of single nucleotide incorporation of each of the four deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP) with a P indicating primer only, which are reactions in which no deoxynucleotide triphosphates were added. No misincorporation of dATP, dGTP, or dTTP opposite N²-f-dG was observed, while dCTP addition opposite N²-f-dG was observed for the variants that were active. The concentration of polymerase is noted above each set of reactions.

### 2.3.1 H6L

Histidine-6 is the highest POOL-ranked residue. Histidine-6 is a second shell residue and lies in the middle of the crescent near Y106 and K157 and is in close proximity to E104. Leucine was chosen as a reasonably close steric mimic of histidine. Compared to wild-type DinB, the H6L variant showed a small but statistically significant decrease (8.8-fold) in catalytic efficiency of dCTP incorporation opposite N²-furfuryl-dG, with a 5-fold decrease in $k_{cat}$ and a 1.7-fold increase in $K_M$. The H6L variant also showed
weak extension of the MatchC primer/$N^2$-furfuryl-dG-containing template DNA, but did not extend to the end of the template, in contrast to wild-type DinB (Figure 2.4).

![Diagram of primer extension assays](image)

Figure 2.4 Schematic of primer extension assays with primer/template DNA containing $N^2$-f-dG correctly base paired with C. Wild type DinB extended to the end of the template; only H6L and Y106F showed appreciable extension, but they did not complete extension to the end of the template. The other variants could not extend to the end of the template. DinB was present at 25 nM.

2.3.2 D10E, D10N

Aspartic Acid-10 is a first-shell residue that is 9th in the POOL rankings and is located near the catalytic metal ions, their coordinating residues D8, D103 and E104, and the important aromatic residues F13 (Jarosz 2009) that are known to have a role in DinB specificity. Specifically, D10 is near the steric gate residue F13, which is important for TLS (Jarosz 2006), and Y79, which modulates the activity of F13 and plays a role in TLS extension (Jarosz 2009). We constructed the mutation D10E to preserve the negative charge and D10N to preserve the length of the original amino acid. We did not observe any measurable activity with the D10E variant, although it could still bind DNA (Table 2.2), but the D10N variant was weakly active at insertion of dCTP opposite $N^2$-furfuryl-
dG. D10N showed a ~25 fold decrease in activity of insertion and was not able to fully extend to the end of the template, even when $N^2$-furfuryl-dG was correctly paired with the MatchC primer (Figure 2.4). The position of this residue is likely the main reason that such large effects were observed.

### Table 2.2 Kinetic Parameters for dCTP incorporation opposite $N^2$-f-dG by DinB and Binding Affinity of DinB to DNA

<table>
<thead>
<tr>
<th>DinB Construct</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}$ (min⁻¹)</th>
<th>$V_{max}$ (µM min⁻¹)</th>
<th>$k_{cat}/K_m$ (min⁻¹ µM⁻¹)</th>
<th>Fold Decrease</th>
<th>$K_{D,DNA}$ (µM)</th>
<th>Fold Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild Type</td>
<td>49 +/- 22</td>
<td>24</td>
<td>0.24 +/- 0.070</td>
<td>0.49</td>
<td>1</td>
<td>0.33 +/- 0.07</td>
<td>1</td>
</tr>
<tr>
<td>His6Leu</td>
<td>81 +/- 21</td>
<td>4.5</td>
<td>0.045 +/- 0.030</td>
<td>0.056</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys150Ala</td>
<td>27 +/- 5.9</td>
<td>0.92</td>
<td>0.023 +/- 0.0040</td>
<td>0.034</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys157Ala</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td>0.49 +/- 0.10</td>
<td>1.5</td>
</tr>
<tr>
<td>Lys157Ile</td>
<td>130 +/- 34</td>
<td>1.1</td>
<td>0.027 +/- 0.010</td>
<td>0.0083</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr106Ala</td>
<td>78 +/- 35</td>
<td>1.3</td>
<td>0.026 +/- 0.0030</td>
<td>0.017</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr106Phe</td>
<td>150 +/- 44</td>
<td>12</td>
<td>0.12 +/- 0.062</td>
<td>0.085</td>
<td>5.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp10Glu</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td>0.51 +/- 0.13</td>
<td>1.5</td>
</tr>
<tr>
<td>Asp10Asn</td>
<td>46 +/- 4.7</td>
<td>0.90</td>
<td>0.022 +/- 0.0050</td>
<td>0.020</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys146Ala</td>
<td>30 +/- 7.2</td>
<td>2.1</td>
<td>0.053 +/- 0.009</td>
<td>0.07</td>
<td>7.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ND, not detected

### 2.3.3 Y106A, Y106F

Tyrosine-106 is a second-shell residue that is near H6, K146, and K150, as well as the first-shell catalytic residue E104. Tyrosine has the ability to be involved in hydrogen bonding as well as pi stacking, so we first modified this position to alanine in order to eliminate both of those properties. The effects of this Y106A modification on overall catalytic efficiency were large, resulting in a decrease in $k_{cat}/K_m$ of 29 fold.
Interestingly, this overall decrease in catalytic efficiency resulted from only a 1.6-fold increase in $K_m$, but an 18-fold decrease in $k_{cat}$.

Modifying this position to phenylalanine (Y106F), which retains the aromatic pi system, gives catalytic efficiency closer to that of wild-type DinB, with a small but statistically significant 5.8-fold decrease. This mutation resulted in a ~3-fold increase in the value of $K_m$ and a 2-fold decrease in $k_{cat}$. This suggests that the pi system at this position contributes to catalytic efficiency of DinB and that the hydroxyl group is also important for activity.

2.3.4 K146A

Among the residues tested, lysine-146 is the furthest from either the DNA substrate or the known catalytic residues that ranked in the top ten POOL-predicted residues. Because it is behind second-shell residue Y106 with respect to the substrate, it is defined as a third-shell residue. Even though it is considered a third-shell residue and therefore only two other residues are between it and the substrate, K146 lies in a solvent exposed area on alpha helix F. DinB K146A gave a 1.6-fold reduction in $K_m$ and an 11.4-fold decrease in $k_{cat}$. However, we could not detect extension of the primer to the end of the template when DinB K146A was presented with a mixture of all four dNTPs, which may suggest an inability to extend from the new primer terminus. To test this possibility, K146A was also assayed for activity with a DNA substrate in which the $N^2$-furfuryl adduct was correctly paired with a C on the primer (MatchC). DinB K146A was not able to extend the primer when the terminus involved a base pair containing the template $N^2$-dG (Figure 2.4). Such extension could only be detected at very high concentrations (1.5 μM) of DinB K146A.
2.3.5 K150A, K157A, K157I

Based on our homology model, Lysine-150, and Lysine-157 are both very near the active site residues D8, D103, and E104, as well as the DNA and incoming nucleotide substrates. According to POOL, these two residues are ranked 5th and 6th, respectively. Changing either of these residues to alanine had significant effects on activity. DinB K150A was active, but with a 15-fold decrease in catalytic efficiency relative to that of wild-type DinB. Based on the homology model we hypothesize that this residue with its positively charged side chain may stabilize the backbone phosphate of the nucleotide at the end of the primer.

Lysine157 is closer than K150 to the catalytic residues and may be involved in binding to the phosphate of the incoming nucleotide. The variant K157A was not catalytically active; however, DinB K157A was found to bind DNA with similar affinity to that of wild-type DinB (Table 2.2), suggesting that the overall conformation of DinB is not greatly perturbed by this mutation. Introducing an isoleucine at this position restored weak activity for incorporation of dCTP opposite $N^2$-furfuryl-dG, suggesting that the chain length is a factor in the importance of this residue. Among the variants that exhibited detectable activity, K157I gave the largest change observed, with a ~59 fold reduction in catalytic efficiency. None of these K157 mutations were able to exhibit any further extension on DNA substrates from a MatchC primer (Figure 2.4).

2.4 Discussion

We used the THEMATICS and POOL programs, which predict functionally important residues, to uncover and probe residues that contribute to activity in DinB. In this work, we focused on the residues ranked in the top ten of the predicted residues.
These functional site predictors have been used to correctly predict important residues outside of the first shell in several different enzymes (Brodkin 2011; Somarowthu 2011a). In each of the previous cases, the substrates were small molecules. In this work, we tested two aspects of the prediction: (1) the performance with an enzyme such as DinB that makes extensive contacts with its large DNA substrate and (2) the performance of the prediction based on a homology model. The top ten POOL-predicted residues are all in or near the catalytic site and include known metal-coordinating residues (D8, D103, and E104) as well as R49, which is known to be critical for function (Wagner 1999). The predicted residues include other residues near the substrate, which are likely to interact directly with the incoming nucleotide, as well as second- and third-shell residues that are found to contribute to activity. Notably in this case, the largest effects observed upon mutation of the predicted residues were the catalytically inactive first-shell variants K157A and D10E; whereas the largest effect of a variant that could accomplish catalysis was the first-shell variant K157I, followed by a second-shell variant Y106A.

The K146 residue is a third-shell residue that is located near the well-conserved “FLAKIA” motif in Y family DNA polymerases that includes the highly-conserved residue K150 (Figure 2.5).
Figure 2.5 (A) ConSurf model (Glaser 2003) showing which amino acid residues are conserved in the DinB family. The degrees of conservation are color coded as blue being the least conserved to deep magenta being the most conserved. (B) The top ten POOL-predicted residues in DinB that are highly conserved, except for Y106, are highlighted. The substrates have been removed for clarity. Images were rendered using the UCSF Chimera package (Pettersen 2004).

DinB K146 appears to be involved in interactions with E98 and N235, the latter of which is part of the peptide linker that tethers the little finger domain to the polymerase domain in Y family DNA polymerases. Therefore, K146 could play a role in positioning the little finger domain. K146, as well as other variants assayed here, also appear to play an important role in the extension step of TLS, which can be as important in bypass as insertion opposite the lesion (Jarosz 2009), because primer extension to the full length of
the template demonstrated by the K146 variant was very poor. The specific effect of a third-shell residue on extension is unexpected.

Distal residues can have a variety of effects on the structure (Chalisser 2007; Chaptal 2007), substrate binding affinity (Heitman 2006; Carpten 2007; Chalisser 2007; Mirza 2007), stability (Hong 2007) and function (El Omari 2006; Dupre 2007; Klyuyeva 2008) of a wide variety of different enzymes (Lee 2011). The semantics of defining the ‘shell’ to which a specific amino acid residue belongs can be challenging, in particular in enzymes with large substrates, such as DNA polymerases. In DNA polymerases, the enzyme can make contacts with a larger proportion of the DNA substrate than an enzyme with a small molecule substrate. Aminoacyl-tRNA synthetases (aaRSs) are examples of enzymes that have very large nucleic acid substrates and whose activity is affected by mutations of distant residues (Lue 2007). The aaRSs must bind to and activate their cognate amino acids and tRNAs, and then transfer the activated aminoacyl adenylate to the tRNA (Arnez 1997; Lee 2011). There is a need for communication between distinct domains in aaRSs for efficient and accurate aminoacylation (Alexander 2001).

Interestingly, the mutation L570F in *E. coli* Leucyl-tRNA synthetase has a significant impact on the function of this enzyme, resulting in a ~30-fold decrease in activity from that of the wild type (Lue 2007). Crystal structure studies of *Thermus thermophilus* LeuRS show that the equivalent residue does not influence the interaction between the protein and the amino acid substrate (Cusack 2000). Rather, it is hypothesized that this residue comes into close proximity to highly conserved M576 which is located in a functionally important loop area in the protein that is responsible for the modulation of
the active site conformation that is essential for substrate recognition (Weimer 2009). Thus, the remote residue L570 contributes substantially to activity via its ability to mediate a conformational change. A more closely related protein to DinB is mammalian DNA pol η, a Y family DNA polymerase encoded by the XPV gene (Johnson 1999; Masutani 1999). Defective DNA pol η can cause the autosomal recessive disease XPV, which is characterized by UV-light-induced deterioration of a variety of tissues, as well as malignant skin carcinomas and melanomas (Kraemer 1985). Some of the pol η variants that are associated with XPV involve mutations that are distal to the polymerase active site triad including D13, D115 and E116 (Biertumpfel 2010), and are located relative to the catalytic triad in similar positions to the highly ranked DinB residues examined here. In particular, DNA pol η K220 is aligned with DinB K146.

Through the use of predictive computations we have demonstrated that residues that are removed from the active site in a DNA damage bypass polymerase have an impact on the catalytic efficiency of that enzyme. The ability of the TLS polymerase DinB to add dCTP opposite its cognate N2-furfuryl-dG lesion is altered when point mutations are introduced distal to the active site residues. In particular, we observed reductions in catalytic efficiency from 6-60 fold as well as instances of complete loss of activity. Using only computational methods to guide us, we identified residues that are important for TLS activity. These residues are postulated to impact the active site residues through steric, pi systems, or electrostatic interactions that are either lost or reduced in the mutations made. Identifying these distal residues and characterizing their contributions to activity may lead to a better understanding of the enzymes as well as to
controlled manipulation of enzyme activity and specificity (Somarowthu 2011a; Somarowthu 2011b).
2.5 References


of co-type nitrile hydratase from \textit{Pseudomonas putida}." \textit{Biochemistry} \textbf{50}(22): 4923-4935.


This chapter is reproduced with permission from the following

"Discrimination against the Cytosine Analog tC by *Escherichia coli* DNA Polymerase IV DinB" Jason M. Walsh, Imenne Bouamaied, Tom Brown, L. Marcus Wilhelmsson, and Penny J. Beuning Journal of Molecular Biology 409 89-100 (2011) dx.doi.org/10.1016/j.jmb.2011.03.069

Penny J. Beuning and Jason M. Walsh designed the study. Imenne Bouamaid, Tom Brown, and L. Marcus Wilhelmsson synthesized the tC nucleotide triphosphate and DNA constructs. L. Marcus Wilhelmsson modeled the tC DNA into the DinB homology model. Jason M. Walsh collected and analyzed the experimental data and prepared the manuscript, tables and figures.
This is a License Agreement between Jason M Walsh ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Elsevier Limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered Company Number</td>
<td>1982084</td>
</tr>
<tr>
<td>Customer name</td>
<td>Jason M Walsh</td>
</tr>
<tr>
<td>Customer address</td>
<td>23 Beech St</td>
</tr>
<tr>
<td>License number</td>
<td>3023210295538</td>
</tr>
<tr>
<td>License date</td>
<td>Nov 06, 2012</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Journal of Molecular Biology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Discrimination against the Cytosine Analog tC by Escherichia coli DNA Polymerase IV DinB</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Jason M. Walsh, Imenne Bouamaieed, Tom Brown, L. Marcus Wilhelmsson, Penny J. Beuning</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>3 June 2011</td>
</tr>
<tr>
<td>Licensed content volume number</td>
<td>409</td>
</tr>
<tr>
<td>Licensed content issue number</td>
<td>2</td>
</tr>
<tr>
<td>Number of pages</td>
<td>12</td>
</tr>
<tr>
<td>Start Page</td>
<td>89</td>
</tr>
<tr>
<td>End Page</td>
<td>100</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>full article</td>
</tr>
<tr>
<td>Format</td>
<td>both print and electronic</td>
</tr>
<tr>
<td>Are you the author of this Elsevier article?</td>
<td>Yes</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Order reference number</td>
<td>None</td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>Function of Damage Inducible DNA Polymerase IV from Escherichia coli</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2012</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>150</td>
</tr>
<tr>
<td>Elsevier VAT number</td>
<td>GB 494 6272 12</td>
</tr>
<tr>
<td>Permissions price</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>VAT/Local Sales Tax</td>
<td>0.0 USD / 0.0 GBP</td>
</tr>
</tbody>
</table>
**LIMITED LICENSE**

The following terms and conditions apply only to specific license types:

15. Translation: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. Website: The following terms and conditions apply to electronic reserve and author websites:

   Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if: This license was made in connection with a course, this permission is granted for 1 year only. You may obtain a license for future website posting, All content posted to the website must maintain the copyright information line on the bottom of each image. A hyper-text must be included to the Homepage of the journal from which you are licensing at [http://www.sciencedirect.com/science/journal/xxxxx](http://www.sciencedirect.com/science/journal/xxxxx) or the Elsevier homepage for books at [http://www.elsevier.com](http://www.elsevier.com), and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. Author website for journals with the following additional clauses:
All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier’s online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article’s Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. Author website for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com. All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. Website (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or for books to the Elsevier homepage at http://www.elsevier.com

20. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. Other Conditions:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500892011.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
Chapter 3: Discrimination against the Cytosine Analog tC by Escherichia coli DNA Polymerase IV DinB

3.1 Introduction

Replication of DNA and the faithful transmission of genetic information are of utmost importance to the survival of all organisms. DNA replication is a multistep process that utilizes several mechanisms for discrimination against the incorporation of incorrect bases, which increases the fidelity of the replication process (Kornberg 1992). Under normal circumstances this process is highly accurate, as DNA polymerization by replicative polymerases occurs with very high fidelity, making errors once in approximately $10^7$ nucleotides incorporated in *E. coli* (Kornberg 1992). When the contributions of proofreading and mismatch repair mechanisms are included, the overall accuracy increases to approximately 1 error in $10^{10}$ nucleotides (Kornberg 1992).

In addition to highly accurate replicative DNA polymerases, cells possess multiple DNA polymerases to fulfill other functions. One such polymerase family is the Y family, the members of which are specialized to copy damaged DNA, but they copy undamaged DNA with lower fidelity than replicative DNA polymerases (Sutton 2001; Friedberg 2006; Jarosz 2007; McCulloch 2008). Y family polymerases are conserved throughout all domains of life (Ohmori 2001; Yang 2005; Friedberg 2006). Although generally structurally similar to their replicative counterparts, there are several important differences between Y family and replicative polymerases. The overall folds of both replicative and Y family DNA polymerases resemble a right hand, including the thumb, palm, and finger domains (Ling 2001; Yang 2003; Yang 2005). Y family polymerases are characterized by small finger and thumb domains which result in an open and solvent-
exposed active site in the palm domain (Yang 2005). Specific to the Y family is a domain known as the ‘little finger’ which is thought to have an important function in both substrate selection and DNA binding (Ling 2001; Boudsocq 2004; Yang 2005). One major difference between replicative and Y family DNA polymerases in the active site is that the O-helix, which is present in replicative polymerases, is absent in the Y family members (Ling 2001). The O-helix is important for ensuring the fidelity of nucleotide incorporation into the primer DNA strand (Steitz 1987; Beese 1993; Steitz 1999). Y family polymerases are generally not as accurate as their replicative counterparts (Tang 2000; Yang 2003; Prakash 2005; Friedberg 2006; Jarosz 2007; Yang 2007; Washington 2010), but their function in induced mutagenesis may play an important role in evolution (Friedberg 2006).

The Y family polymerases are able to accommodate DNA lesions in their active sites and replicate damaged DNA in a process known as translesion synthesis (TLS) (Friedberg 2006), which may be accurate or error-prone. *Escherichia coli* DinB is a Y family DNA polymerase (Wagner 1999) that is known to bypass certain DNA adducts such as the dG adducts of benzo[a]pyrene (N²-B[a]P-dG) (Napolitano 2000; Seo 2006) and methylglyoxal, which forms N²-(1-carboxyethyl)-2′-deoxyguanine (N²-CEdG) (Yuan 2008). DinB possesses a 15-fold preference to bypass N²-furfuryl-dG compared to undamaged DNA (Jarosz 2006). Additionally, DinB accurately bypasses certain dG-dG and dG-peptide cross-links by inserting dC opposite the adduct-G in the template DNA and copies dG-dT intrastrand crosslinks accurately but stalls after incorporation of dA opposite the 3′-T in the crosslink (Bellon 2006; Kumari 2008; Minko 2008). On the other hand, DinB cannot bypass thymine-thymine cyclobutane pyrimidine dimers or
thymine-thymine (6-4) photoproducts, the principal DNA lesions resulting from exposure to UV light (Tang 2000). The $C^8$-dG adducts of 2-acetylaminofluorene and 2-aminofluorene also block the primer extension activity of DinB (Ogawa 2001). The error frequency of DinB in copying undamaged DNA is between $10^{-3}$ and $10^{-5}$, depending on the identity of the base pair (Kobayashi 2002), but in general DinB has the ability to copy DNA containing a range of $N^2$-dG adducts accurately by incorporating dC opposite the adduct-dG. These observations have led to the hypothesis that DinB is specific for minor groove adducts of dG.

We probed the specificity of DinB by determining its efficiency in utilizing the modified base 1,3-diaza-2-oxophenothiazine (tC), which is size-expanded in the major groove, as a template base or incoming nucleotide. The modified base tC is a fluorescent analog of C that forms a canonical Watson-Crick base pair with G, has increased base stacking interactions compared to C, and has been used to probe mechanisms of replication (Figure 3.1) (Marquez 1996; Wilhelmsson 2001; Engman 2004; Sandin 2005; Stengel 2007; Reha-Krantz 2009; Sandin 2009; Sinkeldam 2010; Wilhelmsson 2010). The tC nucleotide was originally developed as an analog of C with stronger binding to complementary G bases within RNA (Lin 1995), and was further derivatized with a guanidinium group, to form the so-called G-clamp that provides strong binding affinity toward G (Lin 1998; Wilds 2002; Wilds 2003).
Figure 3.1 Hydrogen bonding between (a) canonical guanine and cytosine and (b) guanine and the cytosine analog tC (Stengel 2007).

We find that DinB, like Klenow fragment of DNA pol I, can add the tC triphosphate across from template dG (Stengel 2007; Sandin 2009). We demonstrate that DinB inserts dGTP faithfully across from tC, but cannot extend from the newly-generated primer terminus. Moreover, we demonstrate that even with a nine-nucleotide running start DinB cannot complete TLS to the end of the template containing tC. This lack of activity is despite our finding that DinB binds DNA primer:template constructs that contain tC more strongly than it binds unmodified DNA primer:templates. In contrast, both Klenow fragment and human DNA polymerase α can insert dGTP opposite template tC and extend past the noncanonical base pair to the end of the template (Stengel 2007; Sandin 2009; Stengel 2009). This is an intriguing difference, as replicative polymerases are generally considered to be accurate and specific for canonical base pairs (Beese 1993), whereas Y family polymerases have more open and solvent-accessible active sites in order to accommodate bulky lesions in DNA (Yang 2003; Yang 2005; Friedberg 2006).
3.2 Materials and Methods

3.2.1 Proteins and Nucleic Acids

DinB was prepared as described in Beuning et al. (Beuning 2006). The concentrations obtained in DinB purifications are generally in the micromolar range and, therefore, concentrating the protein is not necessary if it is to be used in primer extension assays. DinB was stored in single-use aliquots at -80 °C. DNA containing tC was prepared using solid phase synthesis and the tC phosphoramidite was synthesized as described elsewhere (Sandin 2007). The nucleotide triphosphate was prepared as described in Sandin et al. (Sandin 2009). Unmodified DNA (Eurofins MWG Operon) was purified by denaturing polyacrylamide gel electrophoresis. DNA sequences are shown in Table 3.1. End-labeling with $^{32}$P was carried out as described (Maniatis 1982; Beuning 2006).

3.2.2 Primer Extension Assays

Primer extension assays were performed on a 61-mer DNA template, where the only variable is the position labeled X, which corresponds to G, C or tC (Table 3.1). Standing start primer (31-mer), in which the first nucleotide incorporated is opposite the analog, was generally used for determining the fidelity and kinetics of incorporation of a single nucleotide. The DNA primer and template were combined to a final ratio of 1:1 (500 nM) and were annealed in annealing buffer (20 mM HEPES, pH 7.5, 5 mM Mg(OAc)$_2$) by heating for 2 min at 95 °C, incubating at 50 °C for 50 min, and then cooling to 37 °C. The reactions were carried out with 100 nM $^{32}$P-labeled primer/template in a reaction buffer containing final concentrations of 30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM MgSO$_4$, 2 mM β-mercaptoethanol, 1% BSA and 4% glycerol. Experiments were carried out with 500 μM dNTPs unless otherwise noted and
10 nM DinB, except for the addition of dGTP from tC-containing template which was done at 25 nM DinB due to the low activity of DinB in processing this analog. An aliquot for the zero point was removed prior to addition of dNTP and reactions were initiated by the addition of dNTP (Beuning 2006). Final reaction volumes were 30 μL. Time points were typically taken from 0.5 min to 30 or 60 min and were quenched with 85% formamide, 50 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue (Beuning 2006). In experiments to determine kinetic parameters, conditions were chosen such that less than 25% of the substrate was converted to product. Products from the quenched reactions were analyzed on denaturing (8 M urea) 12-16% polyacrylamide gels, which were subsequently imaged on a Molecular Dynamics storage phosphor imaging screen with a Storm 860 imager. Data were analyzed using ImageQuant software (GE Healthcare) (Beuning 2006). This allows for the analysis of kinetic data and determination of $V_{max}$ and $K_M$ by assessing the percentage of primer extended at various time points in the experiment. Kinetic parameters were derived from GraphPad Prism® non-linear regression analysis software (Segel 1993).
Table 3.1 Sequences of DNA primers and templates

<table>
<thead>
<tr>
<th>DNA</th>
<th>Length</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer 22</td>
<td>22 mer</td>
<td>5′GCATATGATAGTGACGCTGCAG3′</td>
</tr>
<tr>
<td>Standing Start</td>
<td>31 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCC3′</td>
</tr>
<tr>
<td>MatchG</td>
<td>32 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCCG3′</td>
</tr>
<tr>
<td>Extn+1 Primer</td>
<td>33 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCCG3′</td>
</tr>
<tr>
<td>Extn+2 Primer</td>
<td>34 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCCG3′</td>
</tr>
<tr>
<td>Extn+3 Primer</td>
<td>35 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCCG3′</td>
</tr>
<tr>
<td>Extn+4 Primer</td>
<td>36 mer</td>
<td>5′GCATATGATAGTGACGCTGCAGCCGGAGCGCCG3′</td>
</tr>
<tr>
<td>Template DNA</td>
<td>61 mer</td>
<td>5′GGTTACTCAGATCCGCTGCAGCGACGCTGACGGAGCTG3′</td>
</tr>
<tr>
<td>(X = G, C, or tC)</td>
<td></td>
<td>GGCTGCAGCTGACTCATATCATATGC3′</td>
</tr>
</tbody>
</table>

3.2.3 DNA Binding Assays

Equilibrium dissociation binding constants for DNA binding to DinB were determined by using tryptophan fluorescence quenching. A 1-μM DinB solution was titrated with annealed 15 μM DNA primer:template mixture and the fluorescence intensity was measured after each addition with a Varian Cary Eclipse Fluorescence Spectrophotometer using λ_ex of 278 nm. Fluorescence emission was monitored from 290 nm to 500 nm (Favicchio 2009). The emission intensity maxima between 300-400 nm were obtained, corrected for the dilution factor, and the quenching (Q) was calculated by dividing the corrected intensity by the original fluorescence intensity of the 1-μM DinB solution alone (Favicchio 2009). Q was plotted against the final concentration of DNA and K_d was determined by fitting to the following equation:

\[ B = (L_t + K_d + R_t) - \left[ (-(L_t - K_d - R_t)^2 - 4L_tR_t)^{1/2} \right]/2 \]
where $B = \text{fraction bound}$, $L_t = \text{DNA concentration}$, $K_d = \text{dissociation constant}$, and $R_t = \text{total protein concentration}$ (Motulsky 1995; Swillens 1995).

### 3.2.4 Model of DinB bound to DNA containing tC

A homology model of DinB bound to DNA containing $N^2$-furfuryl-dG as the template base was used (Jarosz 2006). Using HyperChem (Hypercube, Inc.), an energy optimized tC base (Wilhelmsson 2003) was superimposed onto the thymine (residue T6, 3’ to the template base) that was replaced, i.e. the inner pyrimidine ring of the tC was superpositioned with the pyrimidine ring system of T. The T was deleted and tC was connected to the deoxyribose moiety. The adenine opposite tC (residue P13) was then replaced with a G. Finally, the furfuryl moiety on the guanine in the template (residue T5) was removed.

### 3.3 Results

#### 3.3.1 DinB discriminates against tC as the template base

In light of the studies in which DNA containing tC is efficiently replicated by the Klenow fragment and by human polymerase α (Stengel 2007; Stengel 2009), we wanted to determine the extent to which DinB could replicate a DNA template containing the tC analog. We expected that DinB, which is specialized to copy non-canonical DNA, would be proficient at copying tC as well. A comparison was made between the abilities of DinB and Klenow to replicate DNA containing tC. Klenow fragment can incorporate a nucleotide opposite tC, extend from the new primer terminus (Stengel 2007), and synthesize DNA to the end of the template, whereas DinB could not synthesize DNA past the site of tC (Figure 3.2).
Figure 3.2 Primer extension assay on DNA containing tC with (a) standing start primer and (b) running start primer, shown at the top of the Figure (* indicates $^{32}$P label), by 10 nM Klenow fragment (KF) or 10 nM DinB. Klenow fragment extended the primer to the end of the template, while DinB did not. DinB demonstrates limited ability to add a nucleotide opposite tC. Samples were analyzed by 12% polyacrylamide gel electrophoresis. Time points were 0, 1, 10, 45 min for each. “P” indicates the position of the primer and “T” indicates the position of the fully extended primer (the length of the template strand). Complete DNA sequences are given in Table 3.1.

However, DinB was able to add a nucleotide across from template tC, extending the standing start primer by one nucleotide, albeit weakly (Figures 3.2 and 3.3). At the longest time points, we observe modest extension past template tC by DinB (Figures 3.2 and 3.3). To probe the extent to which a change in binding affinity for the tC-containing DNA template could be responsible for the lack of catalytic activity of DinB on such a template, we determined the equilibrium dissociation binding constant, $K_d$, of DinB for DNA constructs containing either tC or C as the template base annealed to the standing start primer. We found that DinB has slightly greater affinity for the primer:template DNA containing tC than the primer:template containing unmodified C (Table 3.2), therefore the defect in catalytic activity of DinB with the substrate containing tC is not due to a lack of binding to DNA.
Table 3.2 Dissociation constants for DinB binding to DNA constructs

<table>
<thead>
<tr>
<th>DNA Primer / Template</th>
<th>$K_d$ (μM)</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standing Start / Template-C</td>
<td>0.38 ± 0.034</td>
<td>--</td>
</tr>
<tr>
<td>Standing Start / Template-tC</td>
<td>0.11 ± 0.051</td>
<td>3.5</td>
</tr>
<tr>
<td>MatchG / Template-C</td>
<td>0.21 ± 0.093</td>
<td>--</td>
</tr>
<tr>
<td>MatchG / Template-tC</td>
<td>0.053 ± 0.016</td>
<td>3.9</td>
</tr>
<tr>
<td>Extn+1 / Template-C</td>
<td>0.34 ± 0.099</td>
<td>--</td>
</tr>
<tr>
<td>Extn+1 / Template-tC</td>
<td>0.076 ± 0.022</td>
<td>4.5</td>
</tr>
</tbody>
</table>

We next assayed the ability of DinB to bypass tC with a large running start of nine nucleotides before encountering tC in the template. A primer extension assay was performed with a 22-mer primer to assess the ability of DinB to bypass the analog (Figure 3.3). DinB extends the primer to the end of the unmodified DNA template but is severely impeded when the template strand contains the tC analog (Figure 3.3). Similar to that observed with the standing start primer (Figure 3.2), DinB incorporates nucleotides up to and across from the tC analog, but cannot appreciably extend beyond tC (Figure 3.3).
Figure 3.3 DinB cannot extend past tC in the DNA template. DinB primer extension on template containing C (a) and template containing tC (b) with primer 22. The 31-nucleotide marker indicates the position adjacent to the tC analog; one nucleotide beyond that is addition opposite the analog. The 61-nt marker represents fully extended primer. Samples were analyzed by 14% polyacrylamide gel electrophoresis. Time points for each were 0, 0.5, 1, 2, 5, 10, 15, 20, 30, 45, 60 min.

3.3.2 DinB faithfully incorporates dGTP opposite tC

Our observation that DinB inserts a nucleotide opposite tC led us to wonder about the accuracy of this insertion. Single nucleotide incorporation assays were performed to assess the fidelity of insertion across from tC in the template (Figure 3.4a). DinB was not able to insert dATP, dCTP or dTTP across from the tC analog but faithfully adds dGTP across from tC (Figure 3.4a).
Figure 3.4 (a) DinB faithfully incorporates only dGTP opposite tC in the DNA template. The DNA construct used is diagrammed at the top (* indicates $^{32}$P label). Reactions were carried out with 500 μM dNTP and 10 nM DinB with standing start primer. Samples were analyzed by 16% polyacrylamide gel electrophoresis. Time points were 0 (primer control reaction) and 60 min. “P” indicates the position of the primer. (b) DNA construct used in (c) and (d) is shown (* indicates $^{32}$P label); X = C or tC. (c and d) DinB extends from MatchG primer annealed to template DNA containing C (c) but not to template DNA containing tC (d). Assays were carried out with 1 mM dATP with time points of 0, 0.5, 1, 5, 10, 30 min. “P” indicates the position of the primer.

We determined the efficiency of incorporation of dGTP opposite tC compared to incorporation of dGTP opposite unmodified C (Table 3.3) and found that DinB incorporates dGTP opposite template tC with a modest increase in $K_M$ but a more significant decrease in $k_{cat}$ (~11-fold). Overall, incorporation of dGTP is 12-fold less efficient opposite template tC than opposite template C (Table 3.3). Given that we do not detect misincorporation, we cannot determine an error frequency for DinB with the tC modification. From the limit of detection of the assay, we estimate the error frequency to be less than $\sim 6 \times 10^{-4}$. 
Table 3.3 Steady state kinetic parameters for dGTP incorporation opposite tC vs C

<table>
<thead>
<tr>
<th></th>
<th>$K_M$ (μM)</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$V_{max}$ (μM·min$^{-1}$)</th>
<th>$k_{cat}/K_M$ (min$^{-1}$ μM$^{-1}$)</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>dGTP*C</td>
<td>150 ± 24</td>
<td>17</td>
<td>0.17 ± 0.026</td>
<td>0.11</td>
<td>--</td>
</tr>
<tr>
<td>dGTP*tC</td>
<td>160 ± 31</td>
<td>1.5</td>
<td>0.038 ± 0.031</td>
<td>9.4 x 10$^{-3}$</td>
<td>12</td>
</tr>
</tbody>
</table>

We next determined whether DinB can extend from this newly-generated primer terminus opposite the tC modification. We used a synthetic primer constructed such that it contains a G (“MatchG,” Table 3.1) opposite tC in the template. Although DinB is highly proficient for extension from G opposite unmodified C, there is no extension apparent from the MatchG primer annealed to a tC-containing template (Figure 3.4b-d). Given that no extension of the primer opposite tC can be observed, we estimate the decrease in nucleotide insertion activity to be at least 1500-fold. We also determined the binding affinity of DinB for DNA constructs consisting of MatchG annealed to either a C-containing template or tC-containing template. DinB has a slightly higher affinity for the tC-containing DNA in this context (Table 3.2), indicating that the lack of activity is not due to a defect in binding.

3.3.3 DinB requires primers at least three nucleotides beyond tC for efficient extension

Since DinB could insert a nucleotide opposite the template tC analog but could not extend the newly-synthesized primer terminus, we determined the length of primer beyond the tC analog that is required to allow DinB to efficiently extend the primer. We compared the ability of DinB to extend beyond the tC analog with its ability to extend beyond unmodified C at the same position in the template (Figure 3.5).
When the primer terminus is one (Extn+1) or two (Extn+2) nucleotides beyond tC, primer extension by DinB is weak (Figure 3.5). DinB exhibits more robust activity with a primer terminus that is three (Extn+3) nucleotides beyond tC but still exhibits slightly weaker primer extension activity on the tC-containing DNA than on an unmodified DNA template (Figure 3.5). With a primer terminus four (Extn+4) nucleotides beyond the site of the tC analog, there is little difference in efficiency of synthesis on templates containing either tC or C (Figure 3.5). These observations suggest that in order to observe efficient extension of the primer its terminus must be 3-4 nucleotides beyond a modified base that is a poor substrate for DinB. We also determined the relative affinity of DinB for primer/template DNA with a primer (Extn+1) that terminates one nucleotide beyond the site of the tC modification, or unmodified C, in the template. DinB has modestly higher affinity for the tC-containing construct (Table 3.2), therefore the lack of DinB
activity on a template containing tC cannot be explained by a difference in binding to DNA.

### 3.3.4 DinB adds dtCTP to a primer terminus

Since it was previously demonstrated that the Klenow fragment preferentially incorporates the size-expanded tC analog into the growing DNA strand opposite G by approximately a factor of 10 (Sandin 2009), we next examined whether DinB could add the deoxytriphosphate version of the tC analog across from G in the template. Kinetic parameters for insertion of dtCTP or dCTP across from template G were determined (Figure 3.6a, Table 3.4). DinB exhibited a modest decrease in $K_M$ for dtCTP relative to dCTP and approximately a two-fold higher $k_{cat}$ for dtCTP than for dCTP. Overall, DinB is ~3.5-fold more efficient in incorporation of dtCTP than dCTP opposite template G. Thus, DinB shows a slight preference to add the dtCTP analog to templates containing G relative to the addition of the natural nucleotide dCTP.

![Figure 3.6](image)

**Figure 3.6** DinB incorporates the tC analog and replicates at least one additional nucleotide. DNA construct used is shown at the top. (a) Addition of dtCTP by DinB across from template G. (b) Addition of both dtCTP and dATP. (c) DinB does not add dATP alone. Reactions were carried out with standing start primer. Samples were analyzed by 16% polyacrylamide gel electrophoresis. Time points were 0, 1, 5, 10, 30, 45 min. “P” indicates the position of the primer.
Table 3.4 Steady state kinetic parameters for incorporation of dtCTP versus dCTP opposite template G

<table>
<thead>
<tr>
<th></th>
<th>(K_M) (μM)</th>
<th>(k_{cat}) (min(^{-1}))</th>
<th>(V_{max}) (μM·min(^{-1}))</th>
<th>(k_{cat}/K_M) (min(^{-1}) μM(^{-1}))</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>dCTP*G</td>
<td>190 ± 15</td>
<td>4.4</td>
<td>0.044 ± 0.0017</td>
<td>2.3 x 10(^{-2})</td>
<td>--</td>
</tr>
<tr>
<td>dtCTP*G</td>
<td>130 ± 74</td>
<td>11</td>
<td>0.11 ± 0.036</td>
<td>8.1 x 10(^{-2})</td>
<td>0.28</td>
</tr>
</tbody>
</table>

We next determined if DinB could extend a primer beyond the newly-added dtCTP. A mixture of dtCTP and dATP was added to a final concentration of 250 μM each, as the next nucleotide after G in the DNA template is T (Table 3.1). DinB was able to incorporate both tC and A (Figure 3.6b). No further nucleotide additions were observed, as the next template base is C and dGTP was omitted from these reactions. These observations suggest that DinB is accurate after incorporation of tC. This experiment was also conducted with only dATP available to DinB as a nucleotide substrate and in this case extension was not observed, indicating that the extension observed with the mixture of dtCTP and dATP is due to accurate incorporation of dtCTP and dATP (Figure 3.6c). In contrast, DinB only very weakly extends primers from one or two nucleotides beyond the position of tC in the template (Figure 3.5), suggesting a higher degree of stringency for modifications of the template than for modified incoming nucleotides.

3.4 Discussion

We have demonstrated that *E. coli* DinB, a Y-family DNA polymerase, will not tolerate the size-expanded base analog tC when it is present in the template strand. DinB synthesizes DNA up to and across from tC, accurately incorporating G, but cannot
efficiently bypass tC. It has also been observed that tC can tautomerize to a form that base pairs with A (Stengel 2009; Stengel 2010), but similar to human DNA primase and RNA polymerase (Stengel 2010; Urban 2010), we did not observe misincorporation by DinB. The tC analog is known to be bypassed by the A family polymerase Klenow fragment, as well as by human DNA polymerase α primase (Stengel 2007; Stengel 2009). Specifically, Klenow accurately inserts dGTP opposite template tC (Stengel 2007) and can then extend the primer to the end of the template (Stengel 2009). Remarkably, Klenow is also fairly proficient at bypassing 8-oxo-G and C\(^8\)-aminofluorene-dG (Shibutani 1993; Miller 1997). The overall fidelity of base pair formation on normal, undamaged DNA by Klenow has been determined to be 10\(^{-4}\), which is within the range of 10\(^{-3}\)-10\(^{-5}\) determined for DinB (Kuchta 1988; Kobayashi 2002).

On the other hand, DinB is slightly more efficient at inserting the deoxytriphosphate version of the tC analog relative to unmodified C into the primer strand (Table 3.4). DinB also extends from tC after it is incorporated (Figure 3.6), as was observed for Klenow (Stengel 2007; Tahmassebi 2009). Human DNA primase incorporates tCTP 2.5-fold better than CTP opposite template G and Klenow shows a ~10-fold preference for tCTP over CTP (Sandin 2009; Urban 2010). The increase in efficiency of incorporation of tCTP is generally ascribed to its increased hydrophobicity and increased stacking interactions compared to unmodified C (Sandin 2009; Urban 2010). These observations suggest that DinB exhibits stronger discrimination against modifications of the template and may have more relaxed specificity for the incoming nucleotide. Indeed, a similar trend has been observed with DinB and 8-oxo-G as DinB appears to be less efficient in replicating DNA templates containing 8-oxo-G than it is in
incorporating the nucleotide 8-oxo-GTP (Yamada 2006; Hori 2010). Other polymerases can also exhibit marked asymmetry between utilization of modified bases in the template compared to the incoming nucleotide or in fidelity of base pair formation (Einolf 1998; Urban 2009). This observation may reflect asymmetry in the polymerase active site (Einolf 1998) suggesting that the local environment of the template base compared to that of the incoming nucleotide is apparently quite different, which could lead to differences in efficiency of bond formation or translocation.

It is unclear why DinB specifically discriminates against tC as the template base or at the T-1 or T-2 positions in the template strand (Figure 3.5). Because tC is size-expanded in the major groove, our observations suggest that DinB displays discrimination against major groove adducts in the template strand. Whereas minor groove scanning is a well-known fidelity mechanism used by high-fidelity replicative DNA polymerases, it is not clear that major groove scanning could be generally used (Lutz 1996). Additionally, bases that are derivatized in the major groove are substrates for an array of A and B family DNA polymerases (Thum 2001; Kool 2002; Jager 2005). In structures of Y family polymerases bound to DNA, both the minor and major grooves are highly exposed to solvent at and near the template base (Broyde 2008; Pata 2010). The published homology model of DinB is consistent with this observation, as the incipient base pair and the adjacent two base pairs are predicted by the model to be nearly completely exposed to solvent in the major groove (Figure 3.7) (Jarosz 2006). We modeled a tC:G base pair just upstream from the nascent base pair in DNA bound to DinB, in which there is clearly no steric clash between the protein and the tC modified base (Figure 3.7). The Y family polymerase-specific little finger domain binds DNA in
the major groove, but few of these contacts are at the incipient base pair and so are unlikely to impart specificity at that position (Ling 2001; Yang 2003). The lack of sequence conservation of the little finger domain between different Y family polymerases may influence DNA binding and therefore could impart specificity, perhaps at DNA positions other than the template base (Ling 2001).

Figure 3.7 Model of DinB bound to DNA (Jarosz 2006) with the protein (yellow) shown as bonds (a) and surface (b). (c) and (d) Model of DinB bound to DNA containing a tC:G base pair with tC shown as bonds (c) and surface (d). The view is towards the major groove of DNA at the incipient base pair. The incipient base pair (top) and the next two base pairs are shown colored by identity of the atom and the rest of the DNA is pink. Figure was prepared with VMD (Humphrey 1996).

Crystal structures have been determined of Y family DNA polymerase Dpo4 with several different major groove adducts. In the structures of Dpo4 bound to DNA containing N⁶-benzo[a]-pyrene-dA, one structure shows the benzo[a]pyrene moiety stacked in the DNA resulting in a catalytically-incompetent conformation, while a second structure places benzo[a]pyrene in the major groove in a solvent-exposed conformation that disrupts the Watson-Crick base pairing at the template base (Ling 2004). Strikingly,
the overall conformation of Dpo4 in both structures is remarkably similar, showing only a small shift in the little finger domain (Ling 2004). Structures have also been solved of Dpo4 with DNA containing either \(O^6\)-methylguanine (\(O^6\)-MeG) or \(O^6\)-benzylguanine (\(O^6\)-BzG) paired with C in which the base pair shifts to form a ‘wobble’ pair (Eoff 2007a; Eoff 2007b). With either lesion, Dpo4 misincorporates T or A with fairly high frequency and base pairs between thymine and either lesion were observed in the structures (Eoff 2007a; Eoff 2007b). Dpo4 could extend efficiently from the C:\(O^6\)-MeG base pair, but less so from the T:\(O^6\)-MeG base pair; the trend was similar but with more inhibition in the case of extension from C:\(O^6\)-BzG and T:\(O^6\)-BzG base pairs, suggesting the degree of inhibition may correlate with the size of the adduct (Eoff 2007a; Eoff 2007b). In the experiments reported here, DinB was accurate in the addition of G opposite template tC and it seems that tC forms Watson-Crick base pairs with G. In addition, we could detect no extension from the G:tC base pair in which tC is the template base. In terms of substrate specificity, it is not entirely clear how appropriate Dpo4 is as a model for DinB (Chandani 2010), as Dpo4 efficiently bypasses adducts that are not substrates for DinB, for example, thymine-thymine cyclobutane pyridine dimers (Tang 2000; Boudsocq 2001).

The process of TLS requires both insertion of a nucleotide opposite a non-canonical base and extension of the newly-formed primer terminus. In some cases, insertion opposite a non-canonical, non-cognate base is relatively facile, while the extension step may be less efficient, as observed here. Indeed, replicative polymerases can sense distortions in the DNA caused by lesions 4-6 nucleotides from the incipient base pair (Carver 1994; Miller 1997; Johnson 2003; Fujii 2004). It has been shown that
the TLS polymerase pol V must synthesize 5-6 nucleotides beyond a lesion before the replicative polymerase DNA pol III can resume synthesis; a shorter “TLS patch” leads to proofreading by the pol III exonuclease (Fujii 2004). We find here that the activity of DinB is disrupted by the fluorescent analog tC up to 2-3 nucleotides after the 3’ end of the primer (Figure 3.5). Other investigators have also observed that DinB exhibits pausing within three nucleotides after certain DNA adducts (Kumari 2008; Minko 2008; Jarosz 2009). A DinB variant with a specific defect in copying DNA containing its cognate lesion \( \text{N}^2\text{-furfuryl-dG} \) also stalled three nucleotides after bypassing the lesion (Jarosz 2009). Our findings are in agreement with other work that suggests that DinB requires a slightly shorter buffer of three nucleotides beyond a non-cognate nucleotide than replicative polymerases in order to complete extension. Taken together, these observations suggest that Y family polymerases exhibit less stringency in the extension phase of TLS when confronted with non-cognate bases than replicative DNA polymerases, indicating that Y family polymerases may be able to recover synthesis within a shorter distance downstream from non-cognate DNA damage. However, the exact mechanism of polymerase switching between replicative and TLS polymerases is not fully understood.

In conclusion, Y family polymerases are known for their specialized ability to accommodate and bypass lesion-containing DNA. They also exhibit lowered fidelity on undamaged DNA and are therefore potentially mutagenic. We show that, unlike Klenow and DNA primase (Stengel 2007; Stengel 2009), Y family DNA polymerase DinB is unable to utilize templates containing the modified base tC, suggesting that DinB may specifically discriminate against major groove adducts. Moreover, DinB is remarkably
asymmetric with respect to tC, efficiently incorporating tCTP in the primer strand, while strongly discriminating against tC in the template. It remains to be determined the extent to which this is a general property of DinB.
3.5 References


Eoff, R. L., A. Irimia, M. Egli and F. P. Guengerich (2007b). "Sulfolobus solfataricus DNA polymerase Dpo4 is partially inhibited by "wobble" pairing between O-6-
methylguanine and cytosine, but accurate bypass is preferred." Journal of Biological Chemistry 282(2): 1456-1467.


Chapter 4: Discrimination against major groove adducts by Y family polymerases of the DinB subfamily

4.1 Introduction

The Y family DNA polymerases were first identified as a class of DNA polymerases in 2001 (Ohmori 2001) and are conserved throughout all domains of life (Yang 2005; Friedberg 2006; Pata 2010). They are characterized by their ability to copy damaged DNA in process known as translesion synthesis (TLS). The polymerase domains of Y family polymerases resemble a right hand, with domains identified as the thumb, palm, finger, and little finger (Yang 2005; Friedberg 2006). In contrast with high fidelity polymerases, the finger domains in Y family polymerases tend to be smaller, leading to virtually no major groove contacts at the nascent base pair (Pata 2010); thus, major groove adducts are not subjected to the same steric constraints as minor groove adducts, as they seem to protrude into a solvent-accessible area. The little finger, which is unique to the Y family, makes contacts on the major groove side of the DNA, although not at the nascent base pair. Large major groove DNA adducts often lie in the solvent-accessible area caused by the structural gap between the little finger and the fingers domains (Ling 2001; Shen 2002; Bauer 2007; Lone 2007)

Y family DNA polymerases bypass a wide variety of DNA damage (Suzuki 2001; Shen 2002; Jarosz 2006; Seo 2006; Bauer 2007; Godoy 2007; Yuan 2008; Jarosz 2009; Pence 2011), including interstrand DNA crosslinks (Kumari 2008), protein-DNA crosslinks (Minko 2008), other bulky adducts (Shen 2002; Choi 2006a; Bauer 2007), thymine-thymine dimers (Ling 2001), and damage induced by reactive oxygen species (Rechkoblit 2006; Zang 2006; Hori 2010). *E. coli* Y family DNA polymerase IV (DinB)
is proficient in its ability to bypass minor groove $N^2$ adducts of guanosine, in particular DinB bypasses $N^2$-furfuryl-dG ~15 fold more efficiently than it bypasses templates containing the natural nucleotide dG (Jarosz 2006). Human pol κ, a DinB ortholog that shares 35% sequence identity with DinB comparing their polymerase domains, is proficient in bypassing bulky adducts such as benzo[a]pyrene at the $N^2$ position (Zhang 2000; Rechkoblit 2002; Suzuki 2002; Choi 2006a; Choi 2006b). Substitution of palm domain residue F171 with Ala in human pol κ led to more efficient bypass of $N^2$-benzo[a]pyrene-dG (Sassa 2011). *S. solfataricus* Dpo4, also considered a DinB ortholog and whose polymerase domain shares 33% sequence identity with DinB, has a similar preference to bypass modified dG as it incorporates dCTP opposite 8-oxo-dG approximately ten times more efficiently than opposite unmodified dG (Zang 2006).

Dpo4 has been shown to bypass $N^2$-dG adducts $1,N^2$-ethenodeoxyguanosine, as well as 7-(2-oxoheptyl)-ethenodeoxyguanosine (Christov 2010). Minor groove $N^2$-dG lesions may block DNA synthesis by replicative polymerases, but are bypassed readily by these Y family DNA polymerases.

Adducts at the $N^6$ position of dA, or the $O^6$ position of dG, as well as extended extracyclical systems on cytosine analogs are considered to lie on the major groove side of DNA (Figure 4.1). One cytosine analog with a bulky major groove modification, 1-3,diaza-2-oxophenothiazine (tC), has been shown previously to inhibit primer extension by DinB (Walsh 2011). On the other hand, Dpo4 has been shown to bypass $O^6$-methylguanine, incorporating dCTP a majority of the time, but also misincorporating A and T (20% and 10% respectively) (Eoff 2007b). Human pol κ has also been shown to bypass $O^6$-methylguanine efficiently (Haracska 2002). Human DNA pols κ, η, ι and yeast
pol ζ have been previously tested for their ability to accomplish TLS on both stereoisomers of the bulky benzo[a]pyrene at the $N^6$ position of dA, and only pol η was capable of fully extending DNA beyond the (+)-trans-benzo[a]pyrene steroisomer (Rechkoblit 2002).

![Diagram of base pairing and analogs](image)

Figure 4.1 (a) The canonical Watson-Crick base pairing of G:C and cytosine analogs pyrrolo-dC and dP (b) The canonical Watson-Crick base pairing of A:T and adenosine analogs $N^6$-furfuryl-dA and etheno-dA.
In this report we investigate the ability of DinB, Dpo4 and pol κ to carry out TLS on two major groove-modified pyrimidines, dP and pyrrolo-dC, as well as two major groove adducts of adenosine, 1,N\textsuperscript{6}-ethenodeoxyadenosine (εdA) and N\textsuperscript{6}-furfuryl-deoxyadenosine (fA) (Figure 4.1). The εdA adduct is a result of exposure to the carcinogen vinyl chloride (Bartsch 1994), as well as endogenous aldehyde derived from lipid peroxidation. Human pol κ has been shown to perform very weak TLS on templates containing εA (Levine 2001), which we also observe; however, DinB is inactive on this adduct while Dpo4 is able to bypass it. N\textsuperscript{6}-furfuryl-deoxyadenosine was chosen as the other major groove purine adduct in this study because N\textsuperscript{2}-furfuryl-dG is considered to be a cognate lesion of DinB; it is possible that the furfuryl adducts of dG and dA could arise through similar mechanisms (Scopes 1976; Barciszewski 1999; Barciszewski 2000; Rechkoblit 2002; Jarosz 2006; Barciszewski 2007).

4.2 Materials and Methods

DinB was purified as described previously by Beuning, et al. (Beuning 2006), and stored in single-use aliquots at -80°C. Purification of Dpo4 (Wu 2011) and pol κ (Irimia 2009) were carried out as described. The DNA template containing a single N\textsuperscript{6}-furfuryl-deoxyadenosine was prepared as described from O\textsuperscript{6}-phenyl-dl (Glen Research) (Larson 1992). Pyrrolo-dC, ethenodA, and dP phosphoramidites and pyrrolo-dCTP were from Glen Research. Etheno-dATP was from Chemgenes. DNA sequences used in primer extension assays are as follows: standing start primer, 31-mer 5’-GCATATGATAGTACAGCTGCAGCCGGACGCC-3’; MatchT primer, 32-mer,5’-GCATATGATAGTACAGCTGCAGCCGGACGCCT-3’; MatchT+1 primer, 33-mer,5’-GCATATGATAGTACAGCTGCAGCCGGACGCCTA-3’; MatchT+2 primer, 34-
mer, 5′-GCATATGATAGTACAGCTGCAGCCGGACGCCTAG-3′; MatchT+3 primer, 35-mer, 5′-GCATATGAGTAGCTCGAGCCGGACGCCTAG-3′; MatchT+4 primer, 36-mer, 5′-GCATATGAATAGTACAGCTGCAGCCGGACGCCTAGG-3′; and template 61-mer, 5′-GGTTACTGCTGCTACAGCTGACCGACCTAGGT-3′; where N = A, C, N⁶-furfuryl-deoxyadenosine, or 1,N⁶-ethenodeoxyadenosine, pyrrolo-dC, or dP (Glen Research). DNA primer (standing start or MatchT) and the template containing the indicated adduct were combined to a final ratio of 1:1 (100 nM) and annealed in annealing buffer [20 mM Hepes (pH 7.5) and 5 mM Mg(OAc)₂] by heating for 2 min at 95 °C, incubating at 50 °C for 60 min, and then cooling to 37 °C. The reactions were carried out with 100 nM ³²P-end-labeled primer/template in a reaction buffer containing final concentrations of 30 mM Hepes (pH 7.5), 20 mM NaCl, 7.5 mM MgSO₄, 2 mM β-mercaptoethanol, 1% bovine serum albumin, and 4% glycerol (Beuning 2006). Experiments were carried out with dNTP concentrations ranging from 1 μM-5000 μM, unless otherwise noted, and DinB concentration was 25 nM unless noted. Dpo4 and polymerase κ concentrations were 25 nM for damaged templates and 1 nM for unmodified templates due to their higher activity in general as compared to DinB. An aliquot for the zero point was removed before reactions were initiated by the addition of dNTP (Beuning 2006). The final reaction volumes were 30 μL. Time points were typically taken from 0.5 min to 30 min and reactions were quenched with 85% formamide, 50 mM ethylenediaminetetraacetic acid, 0.025% xylene cyanol, and 0.025% bromophenol blue (Beuning 2006). To determine kinetic parameters, conditions were chosen such that less than 25% of the substrate was converted to product. Quenched reaction products were
analyzed on denaturing (8 M urea) 16% polyacrylamide gels, which were subsequently imaged on a Molecular Dynamics storage phosphor imaging screen with a Storm 860 imager. ImageQuant software (GE Healthcare) was used to analyze data. Kinetic parameters $V_{\text{max}}$ and $K_m$ were determined by assessing the percentage of primer extended at various time points in the experiment and were derived using GraphPad Prism® nonlinear regression analysis software (Segel 1993).

4.3 Results

4.3.1 Cytosine analogs dP and pyrrolo-dC modestly inhibit DinB

In order to extend our observation that the bulky modified pyrimidine tC blocks replication by *E. coli* DinB, we assayed the activity of DinB as well as that of human pol $\kappa$ and Dpo4 in bypass of pyrrolo-dC and dP. Pyrrolo-dC and dP (Figure 4.1) were chosen because they are slightly smaller than tC with only one additional ring in the major groove and because they are commercially available.
Figure 4.2 Primer extension as well as fidelity of incorporation of Y family polymerases *E. coli* DinB, human DNA polymerase κ and *S. solfataricus* Dpo4 on templates containing pyrrolo-dC. P indicates the position of the primer. Fidelity is not shown for dP because it was designed to base pair with dG or dA.

In primer extension assays with Pyrrolo-dC and dP, DinB was more active than on DNA containing tC (Walsh 2011), although DinB and Dpo4 showed stalling 2-3 nucleotides after the site of the modification (Figure 4.2). Human pol κ was able to efficiently bypass both pyrrollo-dC and dP (Figure 4.2). No misinsertion was observed opposite the cytosine analogs (Figure 4.2) with the exception that both purines were added opposite dP (data not shown), which was not unexpected as dP was specifically designed to be a universal purine acceptor (Thoo 1992).
4.3.2 $N^6$-furfuryl-deoxyadenosine inhibits TLS

All three polymerases demonstrated weak bypass with templates containing fA with both running start (not shown) and standing start primers (Figure 4.3).

**Figure 4.3** (a) Primer extension of Y family polymerases *E. coli* DinB, human polymerase κ and *S. solfataricus* Dpo4 on templates containing $N^6$-furfuryl-dA. (b) Fidelity of incorporation of Y family polymerases *E. coli* DinB, human polymerase κ and *S. solfataricus* Dpo4 on templates containing $N^6$-furfuryl-dA. (c) Primer extension of Y family polymerases *E. coli* DinB, human polymerase κ and *S. solfataricus* Dpo4 on templates containing etheno-dA (d) Fidelity of incorporation of Y family polymerases *E. coli* DinB, human polymerase κ and *S. solfataricus* Dpo4 on templates containing etheno-dA.
Human pol κ and Dpo4 were modestly active, whereas DinB showed very weak activity. All three polymerases incorporate predominantly dT opposite $N^6$-furfuryl-dA, although slight misincorporation of dA and dC by pol κ and dA by Dpo4 is seen (Figure 4.3).

Steady-state kinetic parameters were determined, which show that there is less than a 10-fold difference among the three different DNA polymerases in efficiency of incorporation of dT opposite $N^6$-furfuryl-dA (Table 4.1).

<table>
<thead>
<tr>
<th>dTTP : fA</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$ (min⁻¹)</th>
<th>$V_{max}$ (μM*min⁻¹)</th>
<th>$k_{cat}/K_m$ (min⁻¹*μM⁻¹)</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol κ</td>
<td>12 ± 1.7</td>
<td>2.5</td>
<td>0.025 ± 0.002</td>
<td>8.6 x 10⁻²</td>
<td>--</td>
</tr>
<tr>
<td>DinB</td>
<td>39 ± 14</td>
<td>0.76</td>
<td>0.019 ± 0.00014</td>
<td>2.0 x 10⁻²</td>
<td>4.4</td>
</tr>
<tr>
<td>Dpo4</td>
<td>710 ± 210</td>
<td>9.7</td>
<td>0.24 ± 0.032</td>
<td>1.4 x 10⁻²</td>
<td>6.3</td>
</tr>
<tr>
<td>DinB R35A</td>
<td>27 ± 12</td>
<td>2.4</td>
<td>0.029 ± 0.006</td>
<td>8.9 x 10⁻²</td>
<td>0.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dTTP : A</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$ (min⁻¹)</th>
<th>$V_{max}$ (μM*min⁻¹)</th>
<th>$k_{cat}/K_m$ (min⁻¹*μM⁻¹)</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol κ</td>
<td>8.5 ± 6.9</td>
<td>47</td>
<td>0.047 ± 0.002</td>
<td>5.5</td>
<td>--</td>
</tr>
<tr>
<td>DinB</td>
<td>310 ± 140</td>
<td>31</td>
<td>0.15 ± 0.093</td>
<td>1.0 x 10⁻¹</td>
<td>54</td>
</tr>
<tr>
<td>Dpo4</td>
<td>5.9 ± 2.5</td>
<td>15</td>
<td>0.015 ± 0.003</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>DinB R35A</td>
<td>304 ± 89</td>
<td>29</td>
<td>0.029 ± 0.006</td>
<td>9.7 x 10⁻²</td>
<td>57</td>
</tr>
</tbody>
</table>

However, DinB is far less efficient in incorporation of dT opposite undamaged template dA than are pol κ or Dpo4 (Table 4.1). Notably, DinB is also less efficient (25 fold) in bypass of $N^6$-furfuryl-dA than $N^2$-furfuryl-dG (Table 4.2) (Jarosz 2006).

<table>
<thead>
<tr>
<th>DinB</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$ (min⁻¹)</th>
<th>$V_{max}$ (μM*min⁻¹)</th>
<th>$k_{cat}/K_m$ (min⁻¹*μM⁻¹)</th>
<th>Fold Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>dCTP*G</td>
<td>49. ± 22</td>
<td>24</td>
<td>0.24 ± 0.074</td>
<td>0.49</td>
<td>--</td>
</tr>
<tr>
<td>dTTP*G</td>
<td>39 ± 14</td>
<td>0.76</td>
<td>0.019 ± 0.00014</td>
<td>2.0 x 10⁻²</td>
<td>25</td>
</tr>
<tr>
<td>dCTP*A</td>
<td>310 ± 140</td>
<td>31</td>
<td>0.15 ± 0.093</td>
<td>1.0 x 10⁻¹</td>
<td>--</td>
</tr>
<tr>
<td>dTTP*G</td>
<td>147 ± 57</td>
<td>4.6</td>
<td>0.046 ± 0.016</td>
<td>3.2 x 10⁻²</td>
<td>3.7</td>
</tr>
</tbody>
</table>
Because of the poor primer extension on templates containing $N^\delta$-furfuryl-dA and the relatively modest decrease in catalytic efficiency of insertion opposite $N^\delta$-furfuryl-dA relative to undamaged dA, especially in the case of DinB, we suspected that the lack of primer extension could be due to a defect in extension from the new primer terminus after addition of the first nucleotide opposite $N^\delta$-furfuryl-dA. Therefore, we determined the activity of the DNA Pols with a primer that has dT properly base paired with $N^\delta$-furfuryl-dA at its terminus. When the adducts were correctly base paired with the primer MatchT, TLS by DinB is still hindered dramatically, while Dpo4 and pol κ are less disrupted (Figure 4.4).
Figure 4.4: Primer extension of Y family polymerases (a) *E. coli* DinB, (b) human polymerase κ, or (c) *S. solfataricus* Dpo4 on templates containing *N*<sup>6</sup>-furfuryl-dA and etheno-dA with primers that start from one nucleotide beyond the analog to four nucleotides beyond the analog, as illustrated at the top.
When the primer terminus is two or more nucleotides beyond the $N^6$-furfuryl-dA adduct there is only a slight difference in extension abilities between the three polymerases (Figure 4.4). DinB is still disrupted by fA when DinB begins synthesis at least one nucleotidde beyond the lesion, but initiating synthesis further beyond the lesion, with the MatchT+2 MatchT+3 primers, is more conducive to bypass and extension. Pol $\kappa$ and Dpo4 more efficiently added nucleotides and completed synthesis to the end of the template when the primer terminus was beyond the fA adduct. At MatchT+1, which is only one nucleotide beyond the adduct, pol $\kappa$ and Dpo4 efficiently synthesize DNA to the end of the template (Figure 4.4).

4.3.3 1,6-ethenodeoxyadenosine (εdA) strongly inhibits TLS

DinB, Pol $\kappa$, and Dpo4 are strongly or completely inhibited by the presence of εdA in the DNA template with both running (not shown) and standing start (Figure 4.3) primers. Pol $\kappa$ was only weakly active but was accurate, incorporating only dT (Figure 4.3). Dpo4 showed the most activity, albeit still quite weak, of the three pols, but incorporated dA opposite template εdA (Figure 4.3).

Extension from the damaged duplex DNA was then examined with primers that extend beyond the damaged site. Ethenoxyadenosine inhibits TLS by DinB even when the primer extends beyond the adduct. DinB showed no primer extension when MatchT+1, or MatchT+2 are annealed to a template containing εdA. With a DNA duplex of MatchT+3 and template containing εdA, DinB only completes primer extension very weakly. When assayed with the analogous DNA duplex with MatchT+4, DinB activity is restored sufficiently to extend the primer to the end of the template, although its activity is still weaker with εdA than with $N^6$-furfuryl-dA (Figure 4.4). The need for a primer that
is extended at least three nucleotides beyond εdA, which strongly inhibits DinB activity, is consistent with our previous observations with tC, which also strongly inhibits DinB (Walsh 2011).

This requirement for an extension of the primer several nucleotides beyond the damage that we observed with DinB is more moderate with its homologs Dpo4 and pol κ (Figure 4.4). With the MatchT+1 primer and εdA template, both pol κ and Dpo4 show slight disruption in primer extension, but both eventually extend to the end of the template. The disruption is slightly less with MatchT+2/εdA, while MatchT+3/εdA and MatchT+4/εdA are extended efficiently.

4.3.4 DinB R35A mutation allows bypass of N⁶-furfuryl-dA but not εdA

We attempted to identify residues of DinB that could play a role in modulating bypass of major groove adducts by examining a homology model of DinB. Fingers domain residue R35 is likely to be positioned to bind the template strand as it enters the active site, albeit not in the major groove. We hypothesized that the mutation R35A could remove specific contacts with the template and allow more flexibility at the template base to allow DinB to accommodate bulky major groove lesions. DinB R35A was able to bypass N⁶-furfuryl-dA more efficiently than wild-type DinB, but was not able to bypass εdA (Figure 4.5).
Figure 4.5 Primer extension and fidelity of incorporation of wild-type DinB (25 nM) and DinB R35A (12.5 nM) with templates containing (a) $N^6$-furfuryl-dA or (b) etheno-dA. (c) Position of residue R35 in DinB (Walsh 2012), shown in blue, compared to positions on Dpo4 (1JX4) (Ling 2001), shown in green and pol $\kappa$ (2OH2) (Lone 2007), shown in red complexed with the DNA from the DinB homology model (Walsh 2012). R35 lies in a loop region around the active site, and Dpo4 has a similar arginine residue at position 36. Polymerase $\kappa$ has a much smaller loop near this position, residues of which are highlighted from S132-M135. The inset shows a sequence alignment of the region of interest.
Both wild-type DinB and DinB R35A insert dTTP opposite $N^6$-furfuryl-dA as expected, with R35A being ~4.5 fold more efficient than wild-type DinB. There seems to be minor incorporation of A opposite $N^6$-furfuryl-dA by both wild-type DinB and DinB R35A as well, however this could be due to addition of dATP opposite the following nucleotide in the template, which is T, as DinB is known to form bulges in the template and insert a nucleotide opposite the next base in the sequence, generating deleterious -1 frameshifts (Foti 2010). Steady-state kinetics analysis provides an explanation for the improved activity of DinB R35A with $N^6$-furfuryl-dA, as wild-type DinB shows a ~5-fold decrease in catalytic efficiency relative to insertion of dTTP opposite dA, while DinB R35A shows nearly the same efficiency for insertion of dTTP opposite $N^6$-furfuryl-dA or dA (Table 4.1). Thus, the R35A mutation eliminates discrimination against $N^6$-furfuryl-dA relative to undamaged dA.

4.4 Discussion

DNA is constantly subject to damage from both internal and external sources. Y family DNA polymerases allow tolerance of DNA damage by replicating damaged DNA. The Y family polymerase DinB is specialized for bypass of minor groove adducted bases, such as at the $N^2$ position of G which is susceptible to damage from a variety of chemical compounds (Suzuki 2001; Shen 2002; Jarosz 2006; Seo 2006; Bauer 2007; Godoy 2007; Yuan 2008; Jarosz 2009; Pence 2011). Other Y-family DNA polymerases are capable of bypassing major groove adducts (Levine 2001; Rechkoblit 2002; Nair 2006; Eoff 2007a; Eoff 2007b; Jia 2008). We previously showed that the synthetic major groove modified base tC inhibits DinB, although DinB readily utilizes dtCTP as the incoming nucleotide (Walsh 2011). In this work, we used two synthetic nucleotide analogs, pyrrolo-dC and
dP, and two naturally-occurring modified bases, $N^6$-furfuryl-dA and etheno-dA, to probe discrimination of major groove modifications by the Y family polymerases *E. coli* DinB, human pol κ, and *Sulfolobus solfataricus* Dpo4. The furfuryl moiety is thought to form from the metabolism of ribose (Scopes 1976), as well as from treatment with nitrofurazone, a veterinary antibiotic and suspected carcinogen (Hiraku 2004). The damaged base εdA arises from both environmental sources such as vinyl chloride (Yang 2000) and endogenous sources such as the peroxidation of lipids (Chung 1996). Of the three polymerases studied, DinB showed the least proficiency in bypassing the major groove adducts. While base pair formation with the natural bases (dTTP with A vs. dCTP with G) occurred with similar catalytic efficiency by DinB, differing by only ~4 fold (Table 4.2), DinB incorporates the respective nucleotide triphosphate opposite fG or fA with a 25-fold preference for dCTP opposite fG relative to dTTP opposite fA. While DinB R35A had virtually identical activity to wild-type DinB in incorporation of dTTP opposite unmodified dA, the R35A variant was 4.5 fold more efficient at adding dTTP opposite fA (Table 4.1). Strikingly, DinB R35A has essentially the same catalytic efficiency on undamaged and damaged DNA containing $N^6$-furfuyl-dA (Table 4.1). Thus, this single amino acid change is sufficient to eliminate discrimination by DinB against the $N^6$-furfuyl-dA major groove modification. Upon generation of an alignment of the structures of Dpo4 (Ling 2001) and pol κ (Lone 2007) with the homology model of DinB (Walsh 2012) (Figure 4.5), it is apparent that Dpo4 R36 aligns well with DinB R35, yet wild-type Dpo4 was generally proficient in bypass of the major groove modified bases tested here. Moreover, pol κ has a very different loop structure in this region of the protein, which is much smaller than those of Dpo4 and DinB.
In general, human polymerase κ was the most efficient in bypass of the modified pyrimidines, and showed the highest catalytic efficiency for incorporation of dTTP opposite fA, although all three polymerases had catalytic efficiencies within 10 fold. Dpo4 was the most proficient in bypass of the two modified purines, even though it showed a large increase in $K_m$ when performing addition of dTTP opposite fA, but a corresponding $\sim$10-fold increase in $V_{\text{max}}$ compared to pol κ and DinB (Table 4.1). Both pol κ and Dpo4 are more efficient in insertion of dTTP opposite natural A compared to fA with pol κ being $\sim$64 fold more efficient and Dpo4 being $\sim$186 fold more efficient. However DinB is only five fold more efficient at insertion of dTTP opposite A compared to fA. Y family polymerases are specialized for copying damaged DNA and, in particular, DinB and pol κ are more efficient in replication of $N^2$-furfuryl-dG than unmodified dG. However, this work is evidence that the ability of Y-family polymerases to copy damaged DNA is specific to particular adducts, as certain major groove adducts strongly inhibit their activity.

These observations support the findings of Rechkoblit, et al. (Rechkoblit 2002), who found that human pol η was able to bypass the major groove adduct $N^6$-benzo[a]pyrene-dA, whereas pol κ could not carry out TLS when encountering this adduct (Rechkoblit 2002). Eukaryotic DNA polymerase ι is known to rotate the εdA to the $\text{syn}$ conformation to present a Hoogsteen face for hydrogen bonding, and can incorporate either dTTP (major product) or dCTP (minor product) (Nair 2006). Polymerase ι can also complete primer extension beyond both correct and incorrect primer termini with εdA base paired with a T or a C, respectively (Nair 2006). TLS in eukaryotes involves a division of labor, with one polymerase inserting a nucleotide
opposite the lesion and a second Y-family polymerase extending the newly-generated primer terminus (Haracska 2002; Prakash 2002). In addition, different Y-family polymerases appear to be specialized for bypass of distinct sets of lesions.
4.5 References


benzo[a]pyrene $N^2$-dG adduct are bypassed by different lesion-bypass DNA polymerases in *E. coli." DNA Repair (Amst) 5(4): 515-522.


Chapter 5: Future Considerations

This research has been focused on elucidating how *E. coli* Y-family DNA polymerase IV (DinB) functions. Y-family DNA polymerases are specialized DNA polymerases (Ohmori 2001) that are able to copy damaged DNA in a process known as translesion synthesis (TLS). DinB is especially proficient in bypassing adducts at the $N^2$ position of guanosine (Jarosz 2006), which lies on the minor groove side of DNA. Here we have shown that the DinB subfamily is inefficient at dealing with major groove adducts at the $N^6$ position on adenine, as well as certain cytosine analogs that contain extrahelical systems lying in the major groove (Walsh 2011). Y family DNA polymerases also exhibit relatively low fidelity when copying undamaged DNA. These DNA polymerases have roles in the prevention of cancer due to their ability to accurately copy DNA damage, as well as in generating antibiotic resistance because of their ability to generate point mutations when replicating undamaged DNA. I have made progress in many areas regarding this important family of proteins. One of my major findings is that DinB is unable to copy DNA containing bulky adducts that lie on the major groove side of DNA (Walsh 2011). This shifts the previously held paradigm of the Y family polymerases as promiscuous enzymes. My findings also reinforce the very specific nature of the DinB subfamily, as they not only prefer to bypass bulky minor groove adducts (Jarosz 2006), they are strongly blocked by bulky major groove adducts (Walsh 2011).

It would be of interest to expand the finding that DinB discriminates against bulky major groove adducts. A growing library of commercially-available DNA adducts that represent the products of environmental toxins or that are novel synthetic nucleotides will allow future researchers to probe discrimination against DNA damage by various DNA
polymerases. It could be investigated how efficiently different adducts are bypassed by both replicative polymerases as well as Y-family polymerases from a number of organisms that have not yet been thoroughly examined. This project could then be expanded by using site-directed mutagenesis to determine protein regions that are important for activity and specificity. Surprisingly, major groove adducts seem to lie in a solvent-exposed pocket in the DinB active site, even though bulky major groove adducts are blocks to replication by DinB. The work has shown that DinB displays a duality in its ability to utilize some synthetic nucleotides, showing varying activity depending on whether the adduct is located on the template strand or if it is the incoming nucleotide triphosphate being added to the growing primer strand (Walsh 2011). Investigation into the importance of the little finger could also be explored.

Chapter 2 describes work to experimentally characterize the roles that distal amino acid residues play in enzyme efficacy. Originally the theoretical work involved calculations to determine the rank of importance of amino acid residues on enzymes with small molecule substrates (including Ketosteroid isomerase and Phosphoglucose isomerase), work which then experimentally demonstrated the importance of distal residues to activity in the case of PGI but not KSI, as predicted (Somarowthu 2011a). A key question about the computational methods is whether they could be applied to a DNA polymerase, which has a much larger substrate (DNA) with which the DNA polymerase makes many contacts. We demonstrated that the amino acid residues that are highly ranked in the computational predictions of activity, which are distant from the active site, impact the activity of DinB (Walsh 2012). The next step in this project would be to examine the function of the remote residue mutations on DNA polymerase function in
**vivo.** This is particularly important because it was determined that remote residue mutations in DinB specifically affect the extension step, rather than the insertion step (Figure 5.1), of TLS, which has been shown to be as important if not more important than insertion (Jarosz 2009). Biological function could be characterized by assaying the survival of cells harboring the DinB variants in the presence of nitrofurazone, which is thought to generate the $N^2$-furfuryl adduct of deoxyguanosine. The results of the *in vivo* studies of the effect of distal residue mutations would then be compared to the biochemical data in order to determine the importance of the extension step of TLS to cellular survival in the face of DNA damage. Another direction of this project is to perform the POOL and THEMATICS calculations on other DNA polymerases, either those involved in TLS or normal replication, to determine how those remote residues impact activity.

![Diagram of the two phases of translesion synthesis (TLS): insertion opposite the lesion, and extension beyond it. The extension step may be as important as insertion in TLS (Jarosz 2009); distal residues of DinB have an impact on extension (Walsh 2012).](image)

Finally, protein-protein interactions are important for regulation of DinB and other Y family polymerases. The UmuD$_2$ protein regulates mutagenesis and has been shown to interact with DinB to prevent deleterious -1 frameshift activity of DinB (Godoy 2007). The nature and mechanism of this interaction has not been established. By
carrying out primer extension assays with DinB and a variety of UmuD2 constructs including a monomeric version (Ollivierre 2011) and a noncleaveable version (Nohmi 1988; Battista 1990) that are in hand, the interaction between DinB and UmuD2 can be probed at the molecular level. Different UmuD2 and DinB fragments could be purified to isolate the sites of interaction using fluorescence quenching assays. UmuD2 naturally contains one cysteine per monomer and lacks tryptophan, making it amenable to a variety of fluorescence-based assays. These experiments will allow a determination of the roles of protein interactions in DinB activity.
5.2 References


