DNA repair fidelity and cancer: structural and kinetic insights from DNA polymerase Beta mutator variants

Chelsea Lynne Gridley

University at Albany, State University of New York, clgridley@gmail.com

Recommended Citation

DNA Repair Fidelity and Cancer:
Structural and Kinetic Insights from DNA Polymerase β Mutator Variants

By

Chelsea Lynne Gridley

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

School of Public Health
Department of Biomedical Sciences

May 2012
DNA polymerases are essential for genome replication and DNA repair in all living organisms. Precise DNA replication is critical for the preservation of genomic stability. Any insult, endogenous/exogenous, to cellular DNA requires properly functioning repair polymerases. In eukaryotes, DNA polymerase β, a small enzyme (39 kDa), plays an important role in DNA repair during the base excision repair pathway. Pol β catalyzes the incorporation of nucleotides in small stretches (1-6 nucleotides) of damaged double-stranded DNA. Should gap-filling synthesis by pol β be compromised, mutations in genomic DNA accumulate, which are frequently linked to human diseases, including cancers. For this reason, pol β mutants have been extensively studied, particularly in terms of nucleotide incorporation kinetics. Interestingly, the numerous kinetics studies of these mutator mutants have revealed a link between residues distant from the active site and an increased rate of nucleotide misincorporation. Our experiments focus on mutator mutants located in the 8 kDa-N-terminal region (K72E), the hydrophobic hinge, linking the fingers and the palm domain (I260Q), the fingers domain (Y265H, H285D, E295K), and a very flexible loop in the palm domain (P242R, D246V, E249K). We augment biochemical and kinetic studies with structural data in order to further understand the importance of these distant residues and ultimately, to understand the mechanism by which these mutator mutants misincorporate nucleotides at the molecular level.
Dedication

I dedicate this dissertation to the cheerleaders in my life: Mom and Dad; Bernie and Matt Gridley; Steve Smith & Family; Grandpa Duke; Dave & Joyce; Katrina & Kids; Emily & Rachael; The Vossler Families; Woka & WI; and God. I hope I made you proud! I love you!
Acknowledgements

With deepest gratitude, I thank my mentor, Dr. Joachim Jaeger, for his unwavering faith, support and guidance during the past four and a half years. I truly appreciate him for giving me the opportunity to work on challenging and exciting research projects. He encouraged my inquisitiveness, creativity, appetite for knowledge and understanding of all aspects of my research project and beyond.

I am also appreciative of Drs. Pata, Belfort, Banavali and Cunningham for serving in my thesis committee. They have provided guidance and encouragement, both professionally and personally during the course of the dissertation. The members of the Jaeger laboratory have been a pleasure to work with. Sneha Rangarajan, Virginia Muñiz, Kylia Goodner, Wes Sanders, Drew Barber and Timo Koenig have taught me a lot and made coming to work enjoyable. Thank you, Sneha, for making the all-nighters at the synchrotron entertaining and memorable. The members of the neighboring Pata laboratory were also very welcoming and helpful during my time here. They fostered a collaborative environment, which enabled successful troubleshooting in various areas of the research that went into this project.

X-ray diffraction at the Brookhaven National Laboratory was made possible by the teamwork of on-site technicians and the Structural Biology Core at the Wadsworth Center. I am grateful to the Sweasy laboratory of Yale, School of Medicine, for providing the pol β constructs. I appreciate the faculty members of the University at
Albany for offering courses that focused on critical thinking and practical applications of biology. I thank the Wadsworth Center and the Center for Medical Sciences staff for providing technical, mechanical and administrative support and assistance; their service over the years has been superb.

To all of my classmates and friends, thank you for always helping me to stay positive and motivated during this journey; best of luck to you!
Table of Contents

Dissertation Abstract........................................................................................................... ii
Dedication............................................................................................................................. iii
Acknowledgements............................................................................................................. iv
List of Figures and Tables................................................................................................. xii
List of Abbreviations.......................................................................................................... xv
Chapter 1. Introduction ........................................................................................................ 1
  1.1. Mutations in genomic DNA and human disease....................................................... 1
  1.2. DNA damage and repair ......................................................................................... 2
      1.2.1. The DNA-damage response .......................................................................... 4
  1.3. Base excision repair .............................................................................................. 7
  1.4. DNA polymerase β ............................................................................................... 13
      1.4.1. Mechanism of nucleotide transfer ................................................................. 14
          1.4.1.1. DNA pol β fidelity .................................................................................. 23
  1.5. DNA pol β and human cancer ................................................................................ 26
  1.6. Co-crystal structures of pol β ............................................................................... 28
  1.7. Variants of pol β and mutator activities ................................................................. 32
      1.7.1. The Tyr265His variant of pol β is a mutator polymerase ............................... 33
      1.7.2. Variants of pol β, in a flexible loop, exhibit mutator phenotypes ................. 34
      1.7.3. A pol β variant in the dNTP binding pocket is a mutator .............................. 37
      1.7.4. The apoenzyme structure of Met282Leu, is a pol β variant with decreased fidelity and nucleotide discrimination compared to wild-type ...................... 37

  2.1. Description of Participation ...................................................................................... 39
Chapter 3. Li, Y., Gridley, C.L., Jaeger, J., Sweasy, J.B. and Schlick, T.
Unfavorable electrostatic and steric interactions in DNA polymerase β
E295K mutant interfere with the enzyme’s pathway. Journal of the American
Chemical Society, Just Accepted Manuscript, 31 May 2012, DOI:
10.1021/ja300361r.............................................74

3.1. Description of participation.............................................74
3.2. Abstract.............................................................................75
3.3. Introduction......................................................................77
3.4. Methods............................................................................82
    3.4.1. Expression, purification and co-crystallization.........82
    3.4.2. Structure determination and refinement...............83
    3.4.3. System preparation.....................................................84
    3.4.4. Minimization and equilibration..............................85
    3.4.5. Transition path sampling.......................................86
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.6. Free energy barrier and rate constant calculations</td>
<td>90</td>
</tr>
<tr>
<td>3.5. Results</td>
<td>91</td>
</tr>
<tr>
<td>3.5.1. Crystallographic studies</td>
<td>91</td>
</tr>
<tr>
<td>3.5.2. Transition-state identification</td>
<td>92</td>
</tr>
<tr>
<td>3.5.3. Potential of mean force calculation</td>
<td>96</td>
</tr>
<tr>
<td>3.6. Discussion</td>
<td>104</td>
</tr>
<tr>
<td>3.7. Conclusion</td>
<td>109</td>
</tr>
<tr>
<td>3.8. Acknowledgments</td>
<td>110</td>
</tr>
<tr>
<td>3.9. Supplement Information</td>
<td>111</td>
</tr>
<tr>
<td>3.9.1. Shooting algorithm and test convergence of TPS</td>
<td>111</td>
</tr>
<tr>
<td>Chapter 4. Mutator phenotypes of pol β variants on dsDNA: Mg(^{2+}) versus Mn(^{2+})</td>
<td>123</td>
</tr>
<tr>
<td>4.1. Introduction</td>
<td>123</td>
</tr>
<tr>
<td>4.2. Materials and methods</td>
<td>128</td>
</tr>
<tr>
<td>4.2.1. In vitro primer extension assays</td>
<td>128</td>
</tr>
<tr>
<td>4.2.2. Misincorporation and mispair extension assays: Mg(^{2+}) versus Mn(^{2+})</td>
<td>128</td>
</tr>
<tr>
<td>4.2.3. Determining the misincorporation threshold: MnCl(_2) supplementing MgCl(_2)</td>
<td>129</td>
</tr>
<tr>
<td>4.3. Results</td>
<td>130</td>
</tr>
<tr>
<td>4.3.1. The mutator phenotype of pol β variant, I260Q, is enhanced in the presence of Mn(^{2+})</td>
<td>130</td>
</tr>
<tr>
<td>4.3.2. The H285D pol β variant’s mutator phenotype on the dsDNA substrate</td>
<td>137</td>
</tr>
<tr>
<td>4.3.3. E295K pol β is a variant linked to human gastric cancer with increased DNA binding and decreased activity compared to wild-type</td>
<td>138</td>
</tr>
<tr>
<td>4.3.3.1. E295K shows wild-type-levels of activity for primer extension on short dsDNA substrate</td>
<td>138</td>
</tr>
<tr>
<td>4.3.3.2. Crystallography and MD simulations of E295K show the effect of the complete charge reversal in the vicinity of position E295K</td>
<td>140</td>
</tr>
</tbody>
</table>
Chapter 5. Rangarajan, S., Gridley, C.L., Chave, K., Dalal, S., Sweasy, J.B. and Jaeger, J. The K72E DNA polymerase β mutant crystal structure explains loss of lyase function and reveal conformational changes in the polymerase domain..................................................................................................................146

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1. Description of participation</td>
<td>146</td>
</tr>
<tr>
<td>5.2. Abstract</td>
<td>148</td>
</tr>
<tr>
<td>5.3. Introduction</td>
<td>149</td>
</tr>
<tr>
<td>5.4. Materials and Methods</td>
<td>151</td>
</tr>
<tr>
<td>5.4.1. <em>Cloning, expression and purification of pol β</em></td>
<td>151</td>
</tr>
<tr>
<td>5.4.2. <em>Preparation of 5′-dRP substrate</em></td>
<td>153</td>
</tr>
<tr>
<td>5.4.3. <em>5′-dRP lyase assay</em></td>
<td>153</td>
</tr>
<tr>
<td>5.4.4. <em>In vitro primer extension assay</em></td>
<td>154</td>
</tr>
<tr>
<td>5.4.5. <em>In vitro misincorporation and mispair extension assays</em></td>
<td>155</td>
</tr>
<tr>
<td>5.4.6. <em>Single-turnover kinetics of misincorporation</em></td>
<td>156</td>
</tr>
<tr>
<td>5.4.7. <em>Crystallization of pol β apoenzyme</em></td>
<td>157</td>
</tr>
<tr>
<td>5.4.8. <em>Co-crystallization of pol β-dsDNA complexes</em></td>
<td>158</td>
</tr>
<tr>
<td>5.4.9. <em>Data collection and structure determination</em></td>
<td>158</td>
</tr>
<tr>
<td>5.4.10. <em>Molecular dynamics simulations of ternary complexes of K72E and wild-type pol β</em></td>
<td>159</td>
</tr>
<tr>
<td>5.5. Results</td>
<td>160</td>
</tr>
<tr>
<td>5.5.1. <em>K72E pol β lacks lyase activity</em></td>
<td>160</td>
</tr>
<tr>
<td>5.5.2. <em>K72E pol β polymerase activity</em></td>
<td>162</td>
</tr>
<tr>
<td>5.5.3. <em>K72E pol β misincorporates on dsDNA</em></td>
<td>165</td>
</tr>
<tr>
<td>5.5.4. <em>K72E pol β lacks discrimination at the level of ground state binding</em></td>
<td>166</td>
</tr>
<tr>
<td>5.5.5. <em>Local changes in the dRP lyase active site</em></td>
<td>170</td>
</tr>
<tr>
<td>5.5.6. <em>Crystallographic studies of the K72E lyase active in the presence of substrate</em></td>
<td>171</td>
</tr>
<tr>
<td>5.5.7. <em>Changes in the polymerase active site of K72E pol β</em></td>
<td>175</td>
</tr>
</tbody>
</table>
5.5.8. Molecular dynamics studies of lyase active site in a pol β-gapped DNA complex ........................................ 176

5.6. Discussion ................................................................................................................................. 181

5.6.1. Effects of the K72E pol β in lyase active site .................. 181
5.6.2. K72E is a mutator polymerase ..................................................... 182
5.6.3. Structural changes in the polymerase active site ...... 184

5.7. Acknowledgements .................................................................................................................. 184

6. Discussion ..................................................................................................................................... 187

6.1. Review of DNA polymerase β ................................................................................................. 187

6.2. The variant of I260Q: mutator phenotype and structure in the presence of 3’recessed DNA duplex .................................................................................................................. 189

6.3. The gastric-cancer associated pol β variant, E295K: inactive on short gapped DNA, yet active on short dsDNA .......................................................... 192

6.4. H285D is a highly active variant of pol β with a robust mutator phenotype on a variety of substrates ......................................................................................... 195

6.5. Concluding remarks and future directions ............................................................................. 196

References ........................................................................................................................................... 200

Appendices ........................................................................................................................................ 231


A.1.1. Description of participation .................................................................................................. 231

A.1.2. Summary ............................................................................................................................... 232

A.1.3. Introduction .......................................................................................................................... 233

A.1.4. Materials and Methods ......................................................................................................... 236

A.1.4.1. Bacterial strains and media ............................................................................................... 236

A.1.4.2. Chemicals and reagents .................................................................................................... 236

A.1.4.3. Construction of quad-variant of HIV-1 Nef .................................................................... 236

A.1.4.4. Expression and purification of HIV-1 Nef ....................................................................... 238

A.1.5. Results ..................................................................................................................................... 238
A.1.5.1. Production of HIV-1 Nef variant.................238
A.1.5.2. Expression and purification of protein......240

A.1.6. Discussion.................................................................243


A.2.1. Description of participation...............................246
A.2.2. Abstract.................................................................247
A.2.3. Introduction...............................................................248
A.2.4. Materials and methods.................................249
A.2.4.1. Materials.................................................................249
A.2.4.2. Site-directed mutagenesis...............................249
A.2.4.3. Expression and purification...........................250
A.2.4.4. Circular dichroism measurements...............251
A.2.4.5. Dynamic light scattering and sizing...........251
A.2.4.6. Analytical ultracentrifugation..........................251

A.2.5. Results and discussion........................................252
A.2.5.1. Purification and gel filtration.........................252
A.2.5.2. Circular dichroism.................................................253
A.2.5.3. Dynamic laser light scattering.....................256
A.2.5.4. Analytical ultracentrifugation.......................258

A.2.6. Conclusions...............................................................262
A.2.6. Acknowledgements................................................263

Appendix 3. Additional materials and methods..............264
A.3.1. Bacterial strains.......................................................264
A.3.2. DNA substrates.......................................................264
A.3.3. Optimizing crystallization conditions...............264
A.3.4. Crystallization of pol β - gapped DNA - ddNTP ternary co-crystals...........................................265
List of Figures and Tables

Chapter 1
Figure 1.1 DNA damage and the base excision repair pathway ........................................... 9
Figure 1.2 The DNA polymerase β gene and overall architecture: open and closed ................................................................. 11
Figure 1.3 Nucleotidyl transfer reaction and the two metal-ion mechanism for DNA polymerase β ................................................. 16
Figure 1.4 Single point mutations and polymerases with mutator phenotypes ... 35
Table 1.1 DNA damage response mechanisms and lesions acted upon .................. 6
Table 1.2 Key pol β residues and their roles in DNA polymerization ................ 19

Chapter 2
Figure 2.1 In vitro primer extension assays on short double stranded DNA used for co-crystallization ........................................ 54
Figure 2.2 The structure of apoenzyme I260Q pol β ...................................................... 60
Figure 2.3 I260Q ternary complex with dsDNA and ddTTP ........................................ 65
Table 2.1 Crystallographic Studies of DNA Polymerase β ........................................ 56

Chapter 3
Figure 3.1 Molecular snapshots near open (upper) and closed (lower) states for six transition state regions ........................................... 93
Figure 3.2 Normalized probability distribution of the order parameters for the transition states revealed (TS1 to TS6) ......................... 95
Figure 3.3 The free energy profile of the pol β E295K mutant system .................... 98
Figure 3.4 Representative active site conformation of E295K and wild-type pol β in the final closed form ........................................... 101
Figure 3.5 Electrostatic potential landscapes of residues around Arg258 ............. 102
Figure 3.6 “Sandwich-shaped” structure formed by Glu295/Lys295, Tyr296 and Tyr271 tends to cause steric hindrance for Arg258 in the E295K mutator complex ........................................... 103
Table 3.1 Transition states properties for the closing conformational profile of the E295K pol β mutant ........................................... 94
Table 3.2  Free energy barrier and rate \( k_{TST} \) estimated by transition-state theory.................................................................99

Table 3.3  The average critical distances for the chemical reaction following the final transitions of pol \( \beta \) DNA complexes in different systems.....106

Supplemental Figure 3.9.1.  Initial constrained trajectories for the conformational transitions obtained from TMD........116

Supplemental Figure 3.9.2  Correlation functions of order parameters for transition states in the closing pathway.................117

Supplemental Figure 3.9.3  Electron density map of the pol \( \beta \) E295K binary complex contoured at 1.25 RMSD above the mean.....118

Supplemental Figure 3.9.4  Superimposed free energy profiles of five pol \( \beta \) systems obtained from TPS........................................119

Supplemental Figure 3.9.5  Potential of mean force plots computed by BOLAS....120

Supplemental Figure 3.9.6  Summary of activity of pol \( \beta \) mutants..............................121
Locations of residues listed in Supplemental Table

Supplemental Figure 3.9.7  3.9.1 grouped by their effect on the enzyme’s activity when mutated........................................122

Supplemental Table 3.9.1  Kinetic data for wild-type pol \( \beta \) and mutants with correct base-pairing..................................................113

Supplemental Table 3.9.2  Sequence of transition states for the closing conformational profiles of five pol \( \beta \) systems identified by TPS.........................................................115

Chapter 4

Figure 4.1  Misincorporation and mispair extension activities of pol \( \beta \) variants...133

Figure 4.2  Misincorporation threshold determined with MnCl\(_2\): supplementing MgCl\(_2\): wild-type versus I260Q and H285D........................................135

Figure 4.3  Comparison of the binary and ternary complexes of E295K pol \( \beta \)....142

Table 4.1  DNA pol \( \beta \) mutator variant, E295K, affects the local environment and neighboring residues.........................................................144

Chapter 5

Figure 5.1  K72E has compromised dRP lyase activity compared to wild-type....161

Figure 5.2  \textit{In vitro} primer extension assays on the short dsDNA used for co-crystallization.................................................................163
Figure 5.3 Single-turnover experiments of dCTP incorporation opposite template G.................................................................170
Figure 5.4 Apoenzyme structures of K72E and wild-type pol β.........................175
Figure 5.5 Ternary complex of K72E.......................................................................180
Figure 5.6 Molecular dynamics analysis of the lyase active site in K72E pol β......182
Table 5.1 Crystallographic Studies.......................................................................177

Appendix 1
Figure A.1.1 Sequence and structure of HIV-1 Nef and quad-Nef variant.........238
Figure A.1.2 Cloning of HIV-1 Nef variants with site-directed mutagenesis by overlap extension.....................................................244
Figure A.1.3 Purification of HIV-1 Nef variants......................................................247

Appendix 2
Figure A.2.1 Far-UV circular dichroism of Nef protein........................................258
Figure A.2.2 Dynamic laser light scattering of mutant Nef....................................260
Figure A.2.3 Sedimentation velocity analysis of wild-type and mutant Nef........263
List of Abbreviations (In alphabetical order)

2-AP---------2′-deoxy-2-aminopurine

2xYT---------2xYT medium broth

3′-OH ------3′- hydroxyl

8-oxoG------7, 8-dihydro-8-oxoguanine

ABNR------adopted basis Newton-Raphson

Amp--------ampicillin

AP --------abasic site (apurinic/apyrimidinic)

atm --------atmosphere

AZT --------3′-azido-3′-deoxythymidine

BER---------base excision repair

CD --------circular dichroism spectroscopy

ddCTP ------2′,3′-dideoxycytidine-5′-triphosphate

DDR--------DNA damage repair

ddTTP ------2′,3′-dideoxythymidine-5′-triphosphate

DLS--------dynamic light scattering

dNTP ------2′-deoxynucleoside 5′-triphosphate

dP---------momentum perturbation magnitude

dRP---------5′-deoxy-ribose phosphate

DSB---------double stranded break

dsDNA -----double stranded DNA

E. coli------Escherichia coli
HhH--------helix-hairpin-helix

HIV-1 ------Human immunodeficiency virus type 1

HR--------homologous recombination

$k_{cat}$--------kinetic rate of catalysis

$K_d$--------dissociation rate

kDa--------kilodalton

LB--------luria broth

LTR--------long terminal repeat

MC--------Monte Carlo

MD--------molecular dynamics

NER--------nucleotide excision repair

NHEJ--------nonhomologous end joining

NMR--------nuclear magnetic resonance

nt--------nucleotide

pol β--------DNA polymerase β

PPi--------pyrophosphate

ROS--------reactive oxygen species

SD--------steepest descent

SDEL--------stochastic difference equation

SEC--------size exclusion chromatography

SIV--------Simian immunodeficiency virus

SSB--------single stranded break
TCEP-------tris (2-carboxyethyl) phosphine
TMD-------targeted molecular dynamics
TPS-------transition path sampling
UV--------ultraviolet
IPTG------isopropyl β-D-1-thiogalactopyranoside
MPD--------2-methyl-2,4-pentandiol
ASU--------asymmetric unit
RMSD-------root mean square deviation
AIDS-------acquired immune deficiency syndrome
CypA-------cyclophilin A
GFC--------gel filtration chromatography
DTT--------dithiothreitol
SOS--------sum of squares
Chapter 1. Introduction

1.1. Mutations in genomic DNA and human disease

Survival of species from one generation to the next relies upon the successful maintenance of genetic material. Cells are consistently threatened by exogenous and endogenous sources of damage that often cause lesions in DNA (Figure 1.1A). These DNA damaging agents must be intercepted prior to causing damage to DNA, or the cell must develop mechanisms to repair any damage to its DNA. Many human diseases have been linked to mutations in genomic DNA that result from DNA damage. Mutations in DNA arise through the natural aging process, lifestyle influences and environmental factors. Cancers in humans are the result of spontaneous and random mutations that cumulatively develop into wildly replicating, uncontrollable manifestations and genomic instability (Loeb 1974; Loeb 2000; Loeb 2003; Negrini 2010).

It is characteristic for cancer cells to display heterogeneous phenotypes and contain multiple abnormalities in genomic DNA. A hypothesis to explain the accumulation of numerous mutations in cancerous cells is the development of a mutator phenotype early in carcinogenesis (Loeb 2000; Loeb 2003; Venkatesan 2006).

The normal mutation rate in cells, one mistake out of every one billion base pairs (Drake 1998), is too low to account for the numerous mutations observed in tumors, suggesting that genomic mutations, which increase rates of mutation, create cancerous cells (Loeb
A logical hypothesis emerged: mutations in genes that control DNA replication and repair, cell-division and cell-cycle checkpoints (Paulovich 1997) or apoptosis could create the proposed mutator phenotype observed in tumors (reviewed (Preston 2010)).

In general, eukaryotic replicative DNA polymerases create one mistake for every 100,000 to 1,000,000 nucleotides incorporated (McCulloch 2008). Some human cancers have been found to contain variants of two eukaryotic replicative polymerases: pol δ (da Costa 1995; Venkatesan 2007; Daee 2010) and pol ε (Zhou 2008; Zhou 2009). Polymorphisms have also been identified in several genes encoding for DNA repair proteins (reviewed (Goode 2002)). Starcevic et al. review links between the eukaryotic repair polymerase, DNA polymerase β (pol β), and its variants and the occurrence of many human cancers (Starcevic 2004). The link between pol β and some cancers is most likely directly linked to the role of pol β in the maintenance of genomic DNA integrity.

1.2. DNA damage and repair

One well-known environmental source of DNA damage is the ultraviolet (UV) radiation given off by the sun. Solar UV radiation causes two types of DNA lesions: cyclobutane pyrimidine dimers and 6-4 photoproduc (Besaratinia 2011). These are bulky lesions that cause physical changes in the DNA structure and require repair via the nucleotide excision repair (NER) pathway (de Laat 1999). The chemical components of tobacco smoke also cause DNA damage in the lungs of both smokers and nonsmokers. Chemicals cause DNA damage in the form of small alterations and the production of
bulky adducts. Ionizing radiation, in the form of γ-rays and X-rays, causes double stranded DNA breaks (DSBs) (Mahaney 2010). DSBs in DNA can result in chromosomal rearrangements and require repair by nonhomologous end joining (NHEJ) (Leiber 2008) or homologous recombination (HR) repair (San Filippo 2008). While the environmental sources of damage may be controlled, to an extent, by lifestyle changes or increased personal protection around such sources, there is no escaping the chemical by-products of normal cellular metabolism.

Damage can alter the DNA and create mutations in several ways: change the DNA backbone with the addition of bulky adducts or fusion of doublets, cause breaks or gaps in the DNA, introduce mis-matched base pairs and cause missing bases. Our cells must counteract the constant intracellular exposure to damaging chemical agents to maintain genomic DNA integrity. Cellular processes including hydrolysis, oxidation and alkylation, create toxic byproducts. Uracil is an atypical base generated by the hydrolytic deamination of a cytosine base in exposed, single-stranded stretches of DNA (Barnes 2004).

Oxidation by reactive oxygen species (ROS), generated as a by-product during aerobic metabolism in the mitochondria, can cause damage to genomic DNA in a variety of ways. ROS generate a common base damage, 7, 8-dihydro-8-oxoguanine (8-oxoG), through the addition of an oxygen atom at position C8 of guanine (Nishimura 2002). Because oxidative base damage often results in the formation of 8-oxoG, up to 1,000
events per cell every day (Lindahl 1993), bacterial and mammalian cells have evolved specific enzymes, such as the OGG1 DNA glycosylase, that excises 8-oxoG from DNA, and the MutY bacterial DNA glycosylase - with an ortholog in mammals - prevents the incorporation of 8-oxoG into DNA (Barnes 2004). Thymine and uracil, when oxidized into thymine glycol and 5,6-dihydouracil, respectively, actually block the progression of their respective DNA and RNA polymerases (Batty 2000). Alternatively, oxidative damage can cause base substitutions and base modifications through nucleic acid damage (Tudek 2010). In situations where a glycosylase is recruited to the site of DNA damage to remove the damaged base, the gap created is filled and the nick sealed via the base excision repair (BER) pathway (see Chapter 1.3). Mutations in DNA can also be due to misincorporation by DNA polymerases and other enzymes specific to DNA replication and repair.

1.2.1. The DNA-damage response

Cells have evolved several mechanisms by which the effects of DNA damage are detected, reversed and, in some cases, prevented; these are grouped as DNA-damage response (DDR) mechanisms (Table 1.1). The various DDR mechanisms involve numerous proteins and enzymes, generally specific to the type of damage response mechanism elicited (Table 1.1) (Jackson 2009). DNA glycosylases function in the BER and single-stranded break (SSB) repair pathways. DNA polymerases that tend to be “error-prone” play an active role in trans-lesion bypass, a mechanism that requires a
polymerase capable of passing over a site of damaged DNA in order to prevent prolonged stalling during DNA replication. DSBs initiate NHEJ and HR repair pathways, both of which require a wide variety of proteins, protein kinases, polymerases and ligases. The responsibility of DNA repair rests on the shoulders of the various components of each DDR mechanism. A high fidelity polymerase cannot dependably cope with lesions and efficiently replicate across the trans-lesion, just as an error-prone polymerase cannot be relied upon to perform DNA replication. The tightly regulated coordination of DNA repair events, from DNA repair protein translation to recruitment of the necessary proteins, to facilitate repair, to the accurate performance of each protein in the steps of the required DNA repair mechanism, is intimately linked to the maintenance of genomic DNA integrity.
Table 1.1. DNA damage response mechanisms and lesions acted upon

<table>
<thead>
<tr>
<th>DDR mechanism</th>
<th>Prime lesions acted upon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct lesion reversal</td>
<td>O6 alkyl guanine</td>
</tr>
<tr>
<td>Mismatch repair</td>
<td>DNA mismatches and insertion/deletion loops arising from DNA replication</td>
</tr>
<tr>
<td>Base excision repair and SSB repair</td>
<td>Abnormal DNA bases, simple base adducts, SSBs generated as base-excision repair intermediates by oxidative damage or by abortive topoisomerase I activity</td>
</tr>
<tr>
<td>Nucleotide excision repair</td>
<td>Lesions that disrupt the DNA double helix, such as bulky base adducts and ultraviolet photo-products</td>
</tr>
<tr>
<td>Trans-lesion bypass mechanisms</td>
<td>Base damage blocking replication-fork progression</td>
</tr>
<tr>
<td>Non-homologous end joining</td>
<td>Radiation- or chemically-induced DSBs plus V(D)J and class-switching recombination intermediates</td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>DSBs, stalled replication forks, inter-strand DNA crosslinks and sites of meiotic recombination and abortive topoisomerase II action</td>
</tr>
<tr>
<td>Fanconi anaemia pathway</td>
<td>Inter-strand DNA crosslinks</td>
</tr>
</tbody>
</table>

*Adapted from (Jackson 2009)
1.3. Base excision repair

Repair polymerases are critical in preserving DNA integrity despite the damage accrued by ROS, radiation and unrepaired mistakes left behind the cellular replication machinery. The BER pathway specifically removes damaged nucleotides and apurinic/apyrimidinic (AP) sites that are generated from ROS (Wilson 1998). Human diseases, in particular, cancerous cells, result from, among other things, failure to repair damaged DNA (Loeb 1974). The removal of damaged bases creates a gap in double stranded DNA (dsDNA). This gap must be filled with template-matching nucleotides and then sealed, or ligated, to complete the repair. The multi-step BER process involves various enzymes, all of which must work accurately.

The BER pathway begins with the formation of an AP site at the lesion, likely through the action of a DNA glycosylase (Figure 1.1B). There are a variety of DNA glycosylases that recognize and remove different types of damaged bases. The mechanism can then proceed to one of two processes: short-patch or long-patch BER, depending on how many nucleotides need to be replaced, a single nucleotide or several nucleotides, respectively (reviewed in (Parikh 1997) and (Wilson 1998)). In either case, the AP site must be processed prior to DNA gap-filling synthesis. Figure 1.1B illustrates an AP endonuclease (APE) starting a cascade of events beginning with incision of the damaged DNA strand 5’ of the AP site. The resulting short gap in the DNA is flanked by a 3’-hydroxyl (3’-OH) group and a 5’-deoxy-ribose phosphate (dRP) group. The dRP lyase activity, one of DNA polymerase β’s catalytic functions, mediates the removal of the
dRP group in short-patch BER via β-elimination (Matsumoto 1995; Piersen 1996). This leaves the short gap, flanked by a 3’-OH group and a 5’-phosphate. With DNA pol β already participating in the BER process via dRP lyase activity, the polymerase is positioned to fill in the short gap, which is ultimately sealed with DNA ligase. Pol β prefers to synthesize DNA to fill short-gaps and its efficiency for correct versus incorrect nucleotide during incorporation decreases as the size of the gap increases (Brown 2011). Interestingly, when pol β was complexed in vitro with DNA and APE, the dRP lyase activity of pol β on a single-nucleotide gap increased by about 5-fold compared to the pol β-DNA complex in the absence of APE and the rate of gap-filling activity by pol β also increased slightly (Liu 2007). The consequence of mutations in genes encoding repair polymerases, particularly DNA pol β, often results in cancerous manifestations (Starcevic 2004; Albertella 2005; Sweasy 2006; Venkatesan 2006).
Figure 1.1 DNA damage and the base excision repair pathway

A

Figure 1.1 DNA damage and the base excision repair pathway

B

Diagram showing the base excision repair pathway:
- Damaged base
- DNA glycosylase
- AP endonuclease
- DNA ligase
- dRP
- DNA polymerase β
Figure 1.1. DNA damage and the base excision repair pathway

(A) Common agents that cause DNA damage are depicted above the double helix coil, in yellow/orange. The type of damage caused is indicated in the DNA itself, below the damaging agent. Below the DNA is the resulting product of the damage and the repair process or processes that are used to mend the damage. Abbreviations: cis-Pt and MMC, cisplatin and mitomycin C, respectively (DNA-crosslinking agents); (6-4)PP and CPD, 6-4 photoproduct and cyclobutane pyrimidine dimer, respectively; HR, homologous recombination; EJ, end joining. (Adapted from Hoeijmakers 2001). (B) The stages of BER are shown. The specific glycosylase (PDB ID: 1KO9) for repair is recruited first to flip the damaged base out of the DNA helix and into the active site of the enzyme where the base is cleaved from the sugar-phosphate backbone. Alternatively, an abasic site can result from spontaneous hydrolysis. BER next involves strand excision by APE-1 (PDB ID: 1DE8), an endonuclease that prepares the DNA strand for gap filling synthesis by pol β. Pol β cleaves the 5’-dRP group via lyase activity then performs nucleotidyl incorporation to fill in the gap (PDB ID: 3UXP). DNA ligase seals the nick to complete BER (PDB ID: 3RR5).
Figure 1.2. The DNA polymerase β gene and overall architecture: open and closed
Figure 1.2. DNA polymerase β and overall architecture: open and closed

(A) A linear depiction of the pol β protein, showing the division of subdomains by color, encompassing all 335 amino acids. The color scheme used throughout is red: 8 kDa N-terminal domain, residues 1-91; green: thumb subdomain, residues 92-151; magenta: palm subdomain, residues 152-262; blue: fingers subdomain, residues 263-335. The dRP lyase catalytic site is located within the 8 kDa domain and the conserved aspartate residues for nucleotidyl transfer are within the palm subdomain. (B) The apoenzyme structure of wild-type pol β is illustrated here with the same color scheme described above. It is conveniently oriented in the typical right-hand fashion such that the domains of the polymerase align with the right hand (PDB ID: 3UXN). (C) The closure of the fingers subdomain defines what is shown here, the ternary complex of pol β (substrate has been removed for clarity). Also shown is the movement of the 8 kDa domain to meet up with the fingers subdomain to enable optimal geometry in the active site (palm subdomain) for catalysis (PDB ID: 3UXP).
1.4. DNA polymerase β

DNA polymerase β is one member of the X-family of DNA polymerases; a family including polymerases λ, μ and σ (reviewed (Sweasy 2006)), the most thoroughly studied of the eukaryotic polymerases. Pol β shares the characteristic “right-hand polymerase” architecture (palm, thumb and fingers) with many other eukaryotic polymerases (Joyce 1995; Jaeger 1999; Steitz 1999) (Figure 1.2). This frequently used architecture allows us to generalize certain findings and information obtained from pol β studies to other polymerases in developing experimental hypotheses. Figure 1.2A shows how, structurally and functionally, pol β is divided into two domains based on dRP lyase and nucleotide transfer activity, in the N-terminal and the C-terminal domains, respectively. Pol β contains two helix-hairpin-helix (HhH) motifs (residues 55-79 and 92-118) that bind monovalent metals and interact with gapped DNA primer and downstream strands (Pelletier 1996b; Mullen 1997; Krahn 2004).

In vitro, pol β has been shown to be distributive in its catalysis of DNA (Singhal 1993; Prasad 1994) and with remarkably low fidelity (Roettger 2008) when compared to other eukaryotic DNA polymerases. Pol β is a small polymerase (39 kilo Dalton (kDa)) that lacks a proofreading, or endonuclease, domain (Figure 1.2A). The dRP lyase function of pol β has been enzymatically and structurally characterized, identifying several residues in the 8 kDa domain responsible for dRP lyase activity (E26, S30, H34, K35, Y39, K68, E71, K72 and K84) (Prasad 2005; S. Rangarajan (2012) Doctoral Thesis, SUNY Albany). Additionally, the polymerase has been structurally characterized via X-ray
crystallography as apo-polymerase, in complex with dsDNA and in complex with
dsDNA and an incoming 2’-deoxynucleoside 5’-triphosphate (dNTP) at near atomic
resolution (Figure 1.2B & C). These characteristics, along with its inherent lack of
proofreading activity makes pol β an ideal polymerase for structural, enzymatic and
fidelity studies.

1.4.1. Mechanism of nucleotide transfer

The 31 kDa domain of pol β (see Figure 1.2) catalyzes the gap-filling synthesis of DNA
(Figure 1.3A), which involves the addition of the correct (‘Crick’) nucleotide, as directed
by the template, opposite of the coordinating (‘Watson’) base pair.

\[ D = \frac{(k_{cat}/K_m)_{correct}}{(k_{cat}/K_m)_{incorrect}} \]

Where D represents the discrimination, or the degree of nucleotide selectivity of the
polymerase, determined by comparing the specificity constant for correct versus
incorrect nucleotide incorporation. The specificity constant, \( k_{cat}/K_m = k_{pol}/K_d \),
includes the rate of polymerization, \( k_{pol} \), the apparent dissociation constant, \( K_d \), and the
Michaelis constant, \( K_m \) and \( k_{cat} \) the kinetic rate of catalysis specific to the enzyme (Patel
Like many other polymerases and nucleases, pol β uses a two-metal-ion mechanism for nucleotide transfer, which was first described in crystallographic detail in the exonuclease (Klenow fragment) domain of *Escherichia coli* (*E. coli*) DNA pol I (Beese 1991) (reviewed in (Steitz 1994; Steitz 1999)). The palm subdomain (residues 152-262) is the site of metal ion coordination, Mg\(^{2+}\), in pol β and most other cases, where catalysis occurs between the primer-O\(^{3'}\) and the α-phosphate of the dNTP, supported by Mg\(^{2+}\). The two divalent metal ions, identified as A and B to distinguish their functional roles, are bound to carboxylates of three strictly conserved aspartates in pol β (Figure 1.3B). Metal A interacts with the 3’-OH, deprotonating the oxygen and stabilizing the negatively charged oxygen. In the case of pol β with Mg\(^{2+}\), metal A coordinates the dNTP as an α-monodentate and metal B coordinates the dNTP as an α,β,γ-tridentate (Sawaya 1997; Batra 2006). Metal A and neighboring amino acids stabilize and position the 3’O\(^{-}\) for an in-line nucleophilic attack on the α phosphate atom of the dNTP; this links the 3’O•nucleotide, forms a pentacovalent transition state and eventually and breaks the bond between α- and β-phosphodiester of the nucleotide. While both metals A and B are required for polymerase function, metal A is often referred to as the catalytic metal because it binds to the 3’O\(^{-}\) acting as a Lewis acid and coordinates with the polymerase in the active site only for the nucleotidyl transfer reaction. Metal B coordinates and stabilizes the incoming dNTP through charged interactions with the phosphate groups. At the conclusion of the nucleotidyl transfer reaction, metal A and metal B•pyrophosphate (PPi) dissociate from the polymerase to allow for translocation and the next round of catalysis to occur.
Figure 1.3. Nucleotidyl transfer reaction and the two metal-ion mechanism for DNA polymerase β.
Figure 1.3. Nucleotidyl transfer reaction and the two-metal-ion mechanism for DNA polymerase β

(A) The general nucleotidyl transfer reaction pathway with modifications specific to DNA pol β. The reaction steps are detailed in the text. The incoming dNTP is green, Mg$^{2+}$ ions are indicated by aqua spheres. The conformation states of the polymerase are represented by appropriate structures: apoenzyme/open, PDB ID: 3UXN; binary/partially closed, PDB ID: 3V7L; ternary/closed, PDB ID: 3UXP. (B) The two metal-ion mechanism as proposed by Steitz et al., labeled corresponding to pol β active site aspartates (Steitz 1994). See text for details.
Due to its lack of endonuclease function as seen in pol I type enzymes, pol β employs other mechanisms to avoid mistakes and increase fidelity. The fidelity of pol β refers to the ability of the polymerase to select and incorporate the correct nucleotide to faithfully complement the template strand from among a pool of nucleotide substrates. The binding of the correct nucleotide induces a conformational change in the fingers subdomain (residues 260-335) about α-helix M (Pelletier 1994; Pelletier 1996b). The movement of the fingers subdomain indicates that there is a step, prior to catalysis of nucleotide incorporation, during which pol β discriminates between nucleotide substrates. Several residues of pol β play a role in the DNA replication cycle (see Table 1.2 (Radhakrishnan 2006a)). Closing of the fingers subdomain causes several rearrangements of active site side chains: Asp192 rotates away from Arg258 to interact with the metals in the active site while Phe272 rotates down in between Arg258 and Asp192 (Sawaya 1997; Yang 2002); Arg258 then forms hydrogen bonds with Glu295 and Tyr296 (Sawaya 1997); the α-phosphate of the dNTP moves closer to the primer O3’ (Sawaya 1994; Batra 2006); the template base is repositioned to interact with the dNTP via closer interactions with the fingers subdomain the phosphodiester backbone of the DNA (Sawaya 1994); and a hydrogen bond is formed between the primer O2 and Tyr271 (Pelletier 1996b). Pol β is stabilized in the ternary complex (enzyme, DNA and dNTP) via hydrogen-bond donors: E295, Y296, Y271, N279 and R283 (Pelletier 1994; Sawaya 1997; Beard 1998; Opresko 1998; Shah 2001b; Li 2012).
<table>
<thead>
<tr>
<th>Residue</th>
<th>Conformational closing before chemistry</th>
<th>Role</th>
<th>Conformational opening after chemistry</th>
<th>Effect on fidelity¹</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser180</td>
<td>interacts with γ-P of dNTP</td>
<td></td>
<td>releases the H-bond with the primer</td>
<td>S180A, NE</td>
<td>(Kraynov 1997)³</td>
</tr>
<tr>
<td>Arg183</td>
<td>interacts with β-P of dNTP</td>
<td></td>
<td>releases the H-bond with the primer</td>
<td>R183A, NE</td>
<td>(Date 1990, Kraynov 1997)³</td>
</tr>
<tr>
<td>Gly189</td>
<td>interacts with γ-P of dNTP</td>
<td></td>
<td>releases the H-bond with the primer</td>
<td>G189A, ND</td>
<td>(Date 1991)³</td>
</tr>
<tr>
<td>Asp190</td>
<td>coordinates with the two metal ions</td>
<td>subtle rearrangements for metal-ion coordination</td>
<td>releases interactions with the two metal ions</td>
<td>D190S/E, ND</td>
<td>(Date 1991)³; (Radhakrishnan 2005)³</td>
</tr>
<tr>
<td>Asp192</td>
<td>flips away from Arg258 and binds the two metal ions</td>
<td>subtly rearranges for metal-ion coordination</td>
<td>side chain flips away from the active site</td>
<td>D192S/END</td>
<td>(Date 1991)³; (Radhakrishnan 2004, Radhakrishnan 2005, Yang, 2004)³</td>
</tr>
<tr>
<td>Asp256</td>
<td>interactions with catalytic metal and O3' of the primer</td>
<td>coordinates the catalytic metal, primer O3' and α-P</td>
<td>releases binding with the catalytic metal and the primer terminus</td>
<td>D256A, ND</td>
<td>(Menge 1995)³; (Radhakrishnan 2005)³</td>
</tr>
<tr>
<td>Arg258</td>
<td>rate-limiting motion, coordinated with fingers subdomain closing</td>
<td>interacts with Glu295/Tyr296, away from the active site</td>
<td>rate-limiting motion coordinated with fingers subdomain opening, interacts with Asp192</td>
<td>R258A, NE</td>
<td>(Menge 1995)³; (Radhakrishnan 2004, Yang 2002, Yang 2004)³</td>
</tr>
<tr>
<td>Residue</td>
<td>Conformational closing before chemistry</td>
<td>Role</td>
<td>Conformational opening after chemistry</td>
<td>Effect on fidelitya</td>
<td>References</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------</td>
<td>------</td>
<td>---------------------------------------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>lle260</td>
<td>present in the hydrophobic hinge of fingers subdomain; involved in fingers closing</td>
<td>interacts with Arg258, hindering the Arg258-Asp192 interaction</td>
<td>involved in fingers subdomain opening</td>
<td>I260Q, 22-fold↓</td>
<td>(Starcevic 2005a)†</td>
</tr>
<tr>
<td>Tyr271</td>
<td>interacts with primer terminus nucleotide base and the nascent base pair</td>
<td>interacts with nascent base pair</td>
<td></td>
<td>Y271A, NE; Y271F, 3.4-fold↓; Y271F/H, NE; Y271S, 3.5-fold↓</td>
<td>(Beard 1996, Kraynov 1997)†; (Radhakrishnan 2004, Yang 2004)°</td>
</tr>
<tr>
<td>Phe272</td>
<td>insulates the Arg238-Asp192 salt bridge; motion coordinated with fingers subdomain closing</td>
<td>blocks interactions between Arg238 and Asp192</td>
<td>side chain flips away from active site</td>
<td>F272L, 2.2-fold↓</td>
<td>(Li 1999); (Radhakrishnan 2005, Yang 2002, Yang 2004)°</td>
</tr>
<tr>
<td>Asp276</td>
<td>interacts with the incoming dNTP base and Arg40 of the 8 kDa domain</td>
<td>interacts with the incoming dNTP base</td>
<td>moves away from the incoming dNTP base</td>
<td>D276V, ND; D276E, 8.8-fold↓</td>
<td>(Skandalis 2001, Vande Berg 2001); (Yang 2004)°</td>
</tr>
<tr>
<td>Residue</td>
<td>Conformational closing before chemistry</td>
<td>Prechemistry</td>
<td>Conformational opening after chemistry</td>
<td>Effect on fidelity</td>
<td>References</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------</td>
<td>--------------</td>
<td>---------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Asn279</td>
<td>interacts with the template base; compensates for the loss of interactions involving Tyr271 in mispair and mutant systems</td>
<td>N279A, 8.7-fold↑; N279A, 2.1-fold↓; N279L, 2.0-fold↑; N279Q, 2.7-fold↓</td>
<td>(Beard 1996, Kraynov 1997)³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu295</td>
<td>forms a H-bond with Arg258</td>
<td>releases the H-bond with Arg258</td>
<td>E295A, 4.6-fold↓</td>
<td>(Kraynov 1997)³</td>
<td></td>
</tr>
<tr>
<td>Tyr296</td>
<td>forms a H-bond with Arg258</td>
<td>releases the H-bond with Arg258</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Effects are averaged over the number of observations from different groups and the different mispairs that were examined: ND, not determined; NE, no effect (a <2-fold difference); ↑, fidelity increase; ↓, fidelity decrease
³References reflecting experimental results
³References reflecting computational studies
Shaded cells reflect pol β residues with variants discussed in this dissertation.
(Adapted from (Radhakrishnan 2006))
Proposed nucleotide selectivity mechanisms include conformational changes of the polymerase in response to binding the correct versus incorrect nucleotide. Molecular dynamics (MD) simulations of matched and mismatched pol β-DNA complexes suggest that the closure of the fingers subdomain (residues 263-335), via positioning of α-helix N, is disrupted in the presence of a mispair in the active site (Arora 2005a). During misincorporation, the complex would not be fully stabilized, allowing for fast twitching (Santoso 2010a) and rapid dissociation of the incorrect nucleotide.

The distinct kinetic steps of the complex pathway of nucleotide transfer by pol β were first described for DNA pol I from *E. coli* (Bryant 1983; Polesky 1990). Johnson and co-workers used wild-type T7 DNA polymerase and an exonuclease-deficient T7 DNA polymerase to elucidate pre-steady-state kinetics of correct nucleotide incorporation (Patel 1991). The authors concluded, that for T7 DNA polymerase, the kinetic rate of catalysis (k_{cat}) is limited by the dissociation rate (K_d) of the DNA from the enzyme (Patel 1991). Figure 1.3A reviews these steps as applied to pol β: (i) pol β binds DNA template to form a binary complex, (ii) dNTP-Mg^{2+} binds, forming an open ternary complex, (iii) binding of the correct dNTP-Mg^{2+} induces a conformational change of the ternary complex, often referred to as the “closed” complex, (iv) binding of the catalytic Mg^{2+}, (v) nucleotidyl transfer reaction, (vi) catalytic Mg^{2+} dissociates, (vii) fingers subdomain reopens and PPI-Mg^{2+} are released (Patel 1991). Subsequent studies utilized the exonuclease-deficient T7 DNA polymerase to clarify the fidelity of replication; specifically, kinetics of misincorporation (Wong 1991). The characterization of
polymerase rate constants and fidelity of replication have been instrumental in unraveling similar features in the mechanism of action of DNA pol β.

1.4.1.1. DNA pol β fidelity

Fidelity of synthesis and repair are a major concern for all living cells. While mutations can sometimes allow an organism to evolve, they can also be detrimental to the organism’s survival. Because of this delicate balance and the role of DNA polymerases in genome synthesis and repair, they are often discussed in terms of their fidelity. The fidelity of a polymerase describes the frequency of correct versus incorrect nucleotide incorporation events. DNA polymerase fidelity is a mechanism controlled by a variety of kinetic, chemical and structural checkpoints inherent to each individual polymerase. A low-fidelity polymerase exhibits less dNTP discrimination and a greater propensity for incorporating a mis-matched base during DNA synthesis. The opposite is true for a high-fidelity polymerase. Alternatively, a low-fidelity polymerase can also be referred to as “error-prone” while a high-fidelity polymerase is more selective in nucleotide incorporation and has very low error rates.

DNA polymerase fidelity is determined by the kinetics of nucleotide binding and the chemistry of nucleotide incorporation (reviewed in (Johnson 2010)). The fidelity of dNTP incorporation, correct and incorrect, is most elegantly described as the difference in free-energy barriers between the ground state binding of dNTP (correct/incorrect) and
the highest-energy transition state of the reaction pathway (Fersht 1985). Nucleotide binding usually initiates a large conformational change in most DNA polymerases. One exception is in the error-prone Y-family of polymerases. Several Y-family polymerases have small fingers subdomains and few geometric constraints in the DNA and dNTP binding regions of the polymerase (Ling 2001; Silvian 2001; Zhou 2001). These differences, in part, explain the lesion-bypass ability of the Y-family polymerases. The conformation changes a polymerase undergoes during DNA synthesis can contribute to the overall fidelity of the polymerase. For example, pol I was found to adopt a stable conformation in between the open and closed states in response to binding a mismatched NTP via X-ray crystallography (Wu 2011). This intermediate structure was indicative of a conformational step during nucleotide selection for DNA pol I.

Conformational changes have been captured for pol β with X-ray crystallography and stopped-flow fluorescence studies. Pol β adopts a wide-open conformation in absence of substrate (Gridley 2012, see Chapter 2), partially closed binary forms when bound to DNA (Pelletier 1994) and a fully closed ternary form upon the binding of Mg$^{2+}$-dNTP (Pelletier 1994). These structures show how the polymerase responds, physically, to the presence of DNA and Mg$^{2+}$-dNTP. Initially, the large conformational change observed upon binding of the polymerase to the Mg$^{2+}$-dNTP substrate was predicted to be the rate-limiting step of the enzyme-catalyzed reaction.
The rate-limiting step of DNA polymerization can reveal a lot about the mechanism of fidelity. The idea that substrate binding and the accompanying conformational change is the rate-limiting step and, therefore, directly linked to polymerase fidelity, has been challenged by the recent work of Tsai and colleagues (reviewed in Showalter 2006). Using stopped-flow fluorescence, Tsai et al. followed the incorporation of dTTP opposite a 2'-deoxy-2-aminopurine (2-AP) template base. The results showed two phases of fluorescent changes, fast and slow; when the dTTP concentration was varied, it was determined that the fast phase corresponded to a conformation change prompted by dTTP binding (Zhong 1997). Subsequent analysis ruled out the slow phase and confirmed that the dNTP-induced conformational change is fast, therefore, not rate-limiting (Arndt 2001). Moreover, two rate-limiting steps have been identified by following the formation of proton via absorption change of pH: 3'-OH deprotonation and nucleotidyl transfer (Balbo 2011).

The fast and slow transition rates, fluorescently detected, were monitored in the presence of sucrose at 25°C. The rate of the fast phase decreased (Bakhtina 2005), an expected result per the hypothesis that the fast phase corresponds to a large conformational change and subdomain movements. Interestingly, the rate of the slow phase remained unaffected by sucrose or glycerol (Bakhtina 2005); this implies that the slow phase corresponds to the chemistry of the nucleotide incorporation reaction. The reverse rate of the conformation step was subsequently analyzed, using EDTA to remove Mg\(^{2+}\)-dNTP from a preformed ternary complex, and the conformational change
from closed to open was found to be significantly faster for mismatched dNTPs than matched dNTPs for wild-type pol β (Bakhtina 2009).

Taken together, these results indicate that the fidelity of pol β is limited, in part, by the chemistry step, which is directly linked to the geometric constraints imposed by the residues surrounding the active site, the DNA and the incoming dNTP. Furthermore, reverse rate studies determined that the mismatched ternary complex is unstable and is reversed more quickly than the matched ternary complexes, (Roettger 2008; Bakhtina 2009), a mechanism that would clearly contribute to pol β fidelity.

1.5. DNA pol β and human cancer

The main function of pol β is BER; a process of removing and repairing damaged nucleotides in DNA. Cancerous cells replicate very rapidly and therefore are countered by drugs targeting and preventing DNA synthesis. In cancer cells treated with chemotherapy or radiation, the responsibility of repairing damage to neighboring healthy cells falls on DNA repair polymerases, including pol β. The mutation rate during DNA synthesis in vitro by wild-type pol β is approximately $1 \times 10^{-4}$ (Kunkel 1985). Variants of pol β that result in decreased nucleotide discrimination and an increase in misincorporation further exacerbate the accumulation of DNA damage in eukaryotic cells. Damaged DNA, when transcribed and translated, can have one of two outcomes: 1, if DNA has been deleted or inserted, a frameshift mutation can occur and
result in abnormal protein products with an incorrect amino acid sequence, 2, the mutation may be silent and not alter the structure or function of the protein product, 3, the mutation may cause changes to the final protein product which can either be an advantage to the cell, and ultimately the organism, or it may be detrimental and possibly lead to cellular malfunction, diseases or apoptosis.

While not all DNA damage is harmful, and certainly not all damage can be prevented or corrected, it is generally accepted that an increase in the frequency of DNA damage events is coupled to an up regulation of repair proteins. This increase in demand for accurate DNA repair, together with an error-prone variant of pol β, results in an increased rate of mutation (reviewed in (Venkatesan 2006)). It is important to note that not all mutations of pol β are disadvantageous to DNA repair: a quad mutant of pol β was recently discovered to allow efficient bypass of abasic sites (Gieseking 2011). Gieseking et al. identified two substitutions, E232K and T233I that promote incorporation of dNTPs opposite of an abasic site; these variants show increased activity and decreased fidelity compared to wild-type (Gieseking 2011). One hypothesis is that the cell detects the heightened demand for repair polymerases and up regulates pol β expression; the overexpression of pol β could contribute to an increase in spontaneous mutations and genomic instability, particularly if the pol β in the cell is a variant with a mutator phenotype.
Canitrot et al. observed an increase in both the frequency and rate of spontaneous mutations in cells overexpressing pol β. Additionally, their results suggest that treatment with anti-cancer drugs can magnify the accumulation of mutator phenotypes in cells overexpressing pol β (Canitrot 1998). A pol β variant, with a mutator phenotype, could be the cause or the result of human cancers. The structure and function studies of pol β branched into human cancer research after a pivotal review identified variants of pol β in human tumors (Starcevic 2004). This research discovered that 30% of tumors expressed variants of pol β, and some were found to overexpress pol β (Starcevic 2004). Albertella et al. examined the overexpression of specialized polymerases, including DNA pol β, in human cancers (Albertella 2005). They found a correlation between increased expression at the mRNA level and pol β protein levels (Albertella 2005). In the tumors examined, pol β was overexpressed in one-third of them, at levels of 2-fold or more (Albertella 2005). While the majority of DNA polymerase cancer studies focus on wild-type pol β, the detailed study of mutator variants, at the molecular and cellular level, requires a thorough structural and functional approach.

1.6. Co-crystal structures of pol β

Wild-type human DNA pol β has been crystallographically characterized in great detail. An apoenzyme (Davies 1994), pol β•dATP binary complex (Pelletier 1994) and pol β•DNA template-primer (ddCMP-dideoxy 3’-terminated primer)•ddCTP ternary complex (Pelletier 1994) were the first structures of pol β. The ternary structure
identified by Pelletier et al. provided the first glimpse at the pre-transition state of pol β that included DNA, dNTP and metal ions. Subsequent structural characterizations of human pol β included binary and ternary complexes with gapped DNA substrate and a variety of dNTP analog substrates. The primary structures of the human and rat pol β differ by 15 amino acids and the overall secondary and tertiary structures of human and rat pol β remain unaffected. While initial activity studies determined $K_{m}^{dTTP}/k_{cat}$, 4.7 μM/51 min$^{-1}$ and 1.2 μM/42 min$^{-1}$ for rat and human pol β, respectively (Menge 1995), several examples in recent literature highlight the altered in vitro and in vivo activity of rat pol β due to only one amino acid change (see Chapter 1.7).

Pol β, as introduced earlier, is composed of two major subdomains, the N-terminal 8 kDa and the C-terminal 31 kDa, comprised of residues 1-87 and 88-335, respectively. These two domains, which can be readily cleaved with trypsin (Kumar 1990), perform two different enzymatic activities: 8 kDa, 5′-dRP lyase and 31 kDa, nucleotidyl transfer. The first crystal structures of rat pol β ternary complex with dsDNA and 2′,3′-dideoxycytidine-5′-triphosphate (ddCTP) show the large movements of the fingers subdomain relative to the palm subdomain (Pelletier 1994). The thumb and fingers subdomains are closed upon the DNA and active site. The negatively charged DNA is layered on top of a stretch of positively charged residues, predicted to stabilize the DNA. Most of the protein-DNA interactions are not sequence specific, an important feature of DNA polymerases. The detailed views of the active site geometry come from
the structures of pol β bound to a gapped substrate with incoming nucleotide (Sawaya 1997; Batra 2006; Batra 2008).

Sawaya et al. obtained high-resolution (2.2 Å) data from the ternary complex of human pol β - 2 nucleotide (nt) gap DNA - ddCTP with two Mg\(^{2+}\) ions in the active site (Sawaya 1997). This structure contributed significantly to our understanding of the metal-ion coordination of the triphosphate group in the active site as the ddCTP is incorporated and the missing 3’-OH allows the next dNTP to bind but blocks incorporation. Sawaya et al. concluded that the nucleotide-binding Mg\(^{2+}\) coordinates the α-phosphate oxygen of the ddCTP and both metals A and B coordinate the non-bridging α-phosphate oxygen (Sawaya 1997). Furthermore, a 1.65 Å ternary pol β complex with dideoxy-terminated primer shows an additional water molecule for coordination with the catalytic metal, while a 2.0 Å ternary complex with primer 3’OH and a dNTP analog show clear electron density for the 3’O of the primer terminus and two Mg\(^{2+}\) ions coordinated in the active site (Batra 2006). These structures confirm the earlier findings by Sawaya et al., particularly in terms of the metal-ion coordination of the dNTP phosphates. The gap-DNA substrate in these two structures assumes a kink in the active site of 90°. The DNA is stabilized through interactions with two HhH motifs in the fingers subdomain and 8 kDa domain (Pelletier 1996b) while the dNTP is stabilized, in part, by Arg149, Ser180, Arg183, Phe272, Tyr276 and Asp279 (Pelletier 1994; Sawaya 1997).
The three conserved aspartate residues (Asp190, Asp192 and Asp 256) in pol β function to coordinate the two metal ions in the active site during catalysis (Figure 1.3B), one of which exhibits a change in conformation related to the activity of the polymerase.

During catalysis, Asp192 participates in metal-ion coordination, however, in the absence of dNTP and the catalytic metal, with the fingers subdomain is in the open conformation, and the polymerase is in the inactive state, Asp192 rotates 120° about χ1 to interact with Arg258, via the formation of a weak salt bridge (Pelletier 1996b; Sawaya 1997). It is predicted that the transition from inactive to active conformations of Asp192 involves the disruption of the weak Arg258-Asp192 salt bridge as Asp192 interacts with the active site metals, while Phe272 is repositioned due to finger subdomain movement into the space between Arg258 and Asp192. As Asp192 participates in catalysis, Arg258 forms hydrogen bonds with Tyr296 and Glu295.

Located in the fingers subdomain of pol β is α-helix N (residues 275-288). This α-helix makes several critical contacts with the template, primer and incoming nucleotide in the closed ternary complex. Asp276 and Lys280 layer with the base of the incoming dNTP and the template nucleotide, respectively, along the minor groove edge of the DNA, while Tyr271, Asn279 and Arg283 form hydrogen bonds with the primer terminus, the incoming nucleotide and the template base, respectively, along the minor groove edge of the DNA (Pelletier 1994; Sawaya 1997). In the open conformation of pol β, the only interaction between the DNA and α-helix N is a hydrogen bond formed with Tyr271 and the template base (Sawaya 1997). The DNA template is positioned and stabilized after
closing of the fingers subdomain and during the transition state by several residues: Arg283, Asn294 and Glu295 (Kraynov 2000). The β-phosphate of the dNTP is stabilized in the active site via residues Arg183 and Ser180, particularly during the transition state (Kraynov 2000).

1.7. Variants of pol β and mutator activities

Mutator variants of polymerase β are identified as mutations in the pol β gene that lead to a variation of the protein; that has a more error-prone phenotype than wild-type pol β. Consequently, mutator variants of pol β exhibit lower levels of discrimination between matched and mismatched base pairs and, thus, are more likely to incorporate mismatched bases (Figure 1.4.A & B and see Table 3.9.1 for a comprehensive list).

Identification of these mutator mutants in vivo (Washington 1997) takes advantage of the ability of pol β to substitute for pol I in E. coli (Sweasy 1992). Additional pol β mutants with single amino acid substitutions located away from the active site were identified by in vivo complementation of polymerase-deficient E. coli (Kosa 1999a). At 37°C, the temperature sensitive pol I mutant is rendered inactive and pol β restores DNA replication and repair. In the presence of 3′-azido-3′-deoxythymidine (AZT) - a nucleoside analog drug that causes chain termination during DNA synthesis- and at 37°C, cells expressing drug-resistant mutants of pol β will survive. This results in the selection of drug-resistant variants of pol β with mutator properties that can compensate for pol I and are resistant to AZT (Kosa 1999a). The initial screens led to the

1.7.1. The Tyr265His variant of pol β is a mutator polymerase

The Y265C/H mutants were the first to be characterized following identification; with preliminary studies asking what role Y265 plays in the accurate synthesis of DNA by pol β. The Y265C mutant polymerase increased the frequency of frame shifts and base substitution errors \textit{in vitro} as compared to wild-type, presumably through primer/template misalignment and base misincorporation, respectively (Opresko 1998). Furthermore, this pol β variant is a mutator \textit{in vivo}, causing an increase in spontaneous mutations, creating small deletion mutations and increasing the frequency of base substitutions 7.8-fold over wild-type (Clairmont 1999). Another variant of Y265 to His generates spontaneous mutations with a frequency 40 times that of wild-type and has a 108-fold loss in dNTP discrimination at the level of \( k_{\text{pol}} \) for the misincorporation of dGTP opposite template A (Shah 2001b). Subsequently, Shah and colleagues interpreted the increased frequency of errors to be caused by differences in conformational changes upon nucleotide binding. Interestingly, the probability of Y265 to make direct contact with substrates is highly unlikely. Figure 1.4 shows the 265 residue is located the hydrophobic hinge region, which connects the fingers and thumb subdomains. Y265 variants with mutator phenotypes represent early examples of mutator pol β variants in which residues away from the active site have profound effects on the selectivity and
discrimination of pol β. While these conclusions and predictions based on the kinetic profiles of the Y265 mutants have never been characterized further with X-ray crystallography, another hinge mutant, Ile260Gln, has been shown to critically affect side chain orientations in the active site (See Chapter 2).

1.7.2. Variants of pol β, in a flexible loop, exhibit mutator phenotypes

Loop II (residues 240-253) of pol β is solvent-exposed and flexible, yet is important for primer strand alignment in the active site (Dalal 2004). Three residues in this loop were found to confer AZT-resistance to pol β, indicating the importance of loop II for activity and substrate specificity (Kosa 1999a). The pol β variant, Asp246Val, exhibits decreased fidelity compared to wild-type and lacks discrimination during dNTP binding (Figure 1.4) (Kosa 1999b; Dalal 2004). Fidelity studies of D246V indicate that the DNA is stabilized in different position and the base 5’ to the templating base actually becomes the template base in this pol β mutator (Dalal 2004). In addition, the neighboring residue, Glu249 is a proficient mispair extender when mutated to Lys (Kosa 1999b). The results of these studies suggest that residues located in the flexible loop II contribute the activity and fidelity of pol β, in part, by maintaining the proper positioning of the DNA within the active site.
Figure 1.4. Single point mutations and polymerases with mutator phenotypes
Figure 1.4. Single point mutations and polymerases with mutator phenotypes

Several single point mutations throughout the polymerase have been identified due to their mutator phenotypes in vivo and/or in vitro. The text describes several the observed changes in polymerase fidelity and/or efficiency pertaining to the single point mutations illustrated in sticks and spheres and labeled according to residue number. Pol β apoenzyme (A) and ternary (B) are colored according to subdomains as described in Figure 1.2. Note the location of each indicated amino acid and the active site cleft in the center of the palm subdomain. The closest residue to the active site is Phe272, and the most distant is Asp246, located on a flexible loop. Apoenzyme structure is PDB ID: 3UXN, ternary structure is PDB ID: 3UXP.
1.7.3. *A pol β variant in the dNTP binding pocket is a mutator*

One mutator variant of pol β, located in the dNTP binding pocket of pol β, F272L (Washington 1997), has been chemically and kinetically characterized (Li 1999). Li et al. found a 10-fold increase in the rate of spontaneous mutations in the F272L mutant as compared to wild-type (Li 1999), a hallmark of a mutator polymerase. The F272 residue is in the nucleotide binding pocket (Figure 1.4) and makes contacts with neighboring amino acids and the template base in the polymerase active site (Pelletier 1994). While the correlation between structure of the active site and decreased fidelity seems rather obvious, the subtler changes in fidelity exhibited by other pol β mutator mutants with changes in ‘non-active site’ residues, remains unresolved. The sensitivity and accuracy of high-resolution X-ray crystallography allows us to observe these subtleties in mutator mutants of pol β, not in a dynamic, but in a time averaged sense.

1.7.4. *The apoenzyme structure of Met282Leu, is a pol β variant with decreased fidelity and nucleotide discrimination compared to wild-type*

M282L pol β highlights the accuracy and sensitivity of X-ray crystallography. This mutator variant was identified through the genetic screen described by Washington and colleagues (Washington 1997). Pol β M282L has since been biochemically and structurally characterized (Shah 2001a). Pol β M282L exhibits decreased fidelity, 3-fold, and decreased nucleotide discrimination, 9- to 11-fold, relative to the wild-type enzyme (Kosa 1999a). Additionally, pol β M282L was crystallized in the apoenzyme form,
revealing shifts of less than 0.9 Å, restricted to residues in the fingers subdomain (R283, L287, I293, and E295) compared to wild-type (Shah 2001a). The disparities between wild-type pol β and M282L pol β in regard to fidelity and discrimination are more prominent in biochemical assays rather than the subtleties of the structural changes. Surprisingly, protein stability studies revealed the true reason for lack of fidelity in M282L: this mutant was found to be thermodynamically more stable than the parent enzyme (Shah 2001a).

Nucleic Acids Research, Submitted

2.1. Description of Participation
This manuscript is a direct reflection of my research. I designed the experiments, optimized the protocols and collected all of the data. J.B. Sweasy contributed the plasmids of the pol β variant. I performed all plasmid transformation and protein expression and purification techniques. I planned and performed all crystallization trials. I collected, cryoprotected and cryo-cooled the crystals with Sneha Rangarajan before performing diffraction experiments at the Brookhaven National Laboratory Synchrotron Radiation Source (Upton, NY). I optimized and collected crystal diffraction data. I also processed the data with the computer software packages listed in the Materials and Methods (Chapter 2.4.6). I performed molecular replacement and refinement strategies to determine the structure of the crystals described within. I then analyzed the electron density of the solution to interpret the structural findings.
Sneha Rangarajan was instrumental in the optimization of expression, purification and crystallization of the structures described within. The protocol for apoenzyme crystal growth was developed by Susan Firbank and provided a foundation from which to begin crystallization trials. The I260Q variant plasmid construct was created and provided by Shibani Dalal while Joann B. Sweasy and Joachim Jaeger funded the project.
2.2. Abstract

The I260Q variant of DNA polymerase β is an efficient mutator polymerase with decreased incorporation fidelity opposite A and C as templating bases. Previous modeling studies have suggested that I260Q harbors structural variations in its hinge region. Here, we present the high-resolution crystal structures of wild-type and I260Q rat polymerase β in the presence and absence of substrates. Both the I260Q apoenzyme structure and the closed ternary complex with double stranded DNA and ddTTP show ordered water molecules in the hydrophobic hinge, whereas this is not the case in the wild-type polymerase. Compared to wild-type polymerase β ternary complexes, there are subtle movements around residues 260, 272, 295 and 296 in the mutant. The rearrangements in this region, coupled with side chain movements in the immediate neighborhood of the dNTP binding pocket, namely residues 258 and 272, provide an explanation for the altered activity and fidelity profiles observed in the I260Q mutator polymerase.
2.3. Introduction

Faithful DNA replication and DNA repair are critical for the preservation of genomic integrity. The fidelity of a polymerase is defined as the ability to select the correct incoming deoxynucleoside triphosphate (dNTP) over an incorrect one to form a Watson-Crick base pair for incorporation into the new strand of DNA. Repair polymerases are critical in preserving DNA integrity. DNA polymerase β (pol β) is a key player in the base-excision repair (BER) pathway and is responsible for repairing small stretches (one to six nucleotides) of damage in double stranded DNA (dsDNA). Pol β, which is highly conserved across eukaryotic organisms (Uchiyama 2009), is the smallest and most thoroughly studied of the eukaryotic polymerases. The BER pathway removes damaged nucleotides and abasic sites that have been generated from reactive oxygen species and other forms of cellular damage (Maynard 2009). During BER, pol β recognizes and binds to 5’-deoxyribose phosphate (dRP) sites, left behind after the removal of damaged bases, where it removes the dRP moiety and fills in the gap with a NTP complement to the template base (Matsumoto 1995; Wilson 1998). Pol β is a distributive polymerase in vitro and exhibits a mutation frequency of $2.8 \times 10^{-4}$ to $6.6 \times 10^{-4}$ in a forward mutational assay on gapped DNA (Kunkel 1985), the preferred substrate for pol β (Pelletier 1994). Due to its small size and relative ease of handling, pol β has been a model repair polymerase to study the kinetics of nucleotide incorporation and the mechanism of DNA gap-filling repair processes: short-patch BER, long-patch BER and nucleotide excision repair. However, if the polymerase becomes altered in a way that leads to the generation of subsequent mutations during DNA gap-filling synthesis and causes the
accumulation of mutations in genomic DNA, then this can lead to abnormal manifestations, including human diseases such as cancer. In fact, several mutator mutants of pol β have been identified in a variety of human cancers (Starcevic 2004), indicative of a possible link between cancer and compromised repair polymerases.

A member of the X family of polymerases, pol β shares similar architectural features with other eukaryotic polymerases including pol λ and pol ε. Pol β shares the characteristic “right-hand polymerase” (palm, fingers and thumb) with many other eukaryotic polymerases (Steitz 1994; Jaeger 1999). Functionally, pol β is divided into two domains based on dRP lyase activity in the 8 kDa domain and nucleotidyl transfer activity in the 31 kDa domain. Pol β contains two sequence motifs that enhance the binding of the polymerase to the gapped DNA substrate: two helix-hairpin-helix motifs and carboxylate residues in the palm subdomain (Steitz 1994). The palm subdomain of pol β houses the strictly conserved aspartate amino acids in the pol X family (Asp190, Asp192 and Asp256), which bind two metal ions. Like all other polymerases and many nucleases, pol β uses a two-metal-ion mechanism for nucleotidyl transfer, which was first described in crystallographic detail in the exonuclease domain of Escherichia coli DNA polymerase I Klenow fragment (Beese 1991).

Many kinetic studies have characterized the numerous aspects of rat pol β activity: wild-type, mutator mutants, cancer-associated mutants, with double stranded or gapped DNA substrate, as well as incorporation rates for correct versus incorrect nucleotides.
Current literature highlights the altered in vitro and in vivo activity of pol β due to single amino acid changes: I260Q mis-incorporates nucleotides due to decreased dNTP discrimination during binding (Starcevic 2005a) and extends beyond the mispaired primer terminus more often than wild-type (Dalal 2008b); D246V exhibits decreased fidelity compared to wild-type and lacks discrimination during dNTP binding (Kosa 1999b; Dalal 2004); M282L shows mutagenic properties both in vitro and in vivo (Shah 2001b). Furthermore, three single amino acid mutant forms of pol β, E295K, I260M and P242R have been linked to gastric cancer (Iwanaga 1999; Lang 2007), and prostate cancer (Dalal 2005; An 2011). The vast majority of structural studies of pol β have been carried out on the wild-type polymerase, and thus far, structural information has only been obtained for two single-amino acid variants of pol β, M282L (Shah 2001a) and I260Q (Tang 2008), neither structure contains substrate. To better understand pol β structure/function relationships, high-resolution, complete structural characterizations of single point mutator mutants of DNA polymerase β in the presence of substrate are required.

The I260Q variant of pol β is an active polymerase with strong mutator properties compared to wild-type: it exhibits a 60-fold increase in reversion frequency in a Trp+ reversion assay in vivo and has an increase in mis-incorporation events in a qualitative gap-filling synthesis in vitro (Starcevic 2005b). I260Q exhibits 19-23-fold decreased fidelity opposite template base A and 16-30-fold decreased fidelity opposite template
base C, compared to wild-type pol $\beta$ (Starcevic 2005a). Interestingly, this residue is located in the hydrophobic hinge region at the boundary of the palm and fingers subdomains, an area known to influence the ability of pol $\beta$ to discriminate during nucleotide binding (Starcevic 2005a; Starcevic 2005b). The hydrophobic hinge (Ile174, Leu194, Thr196, Ile260, Tyr265 and Phe272) (Pelletier 1994; Starcevic 2005b) of pol $\beta$, and, in particular, the relative motion of the fingers subdomain about this hinge, critically affects polymerase fidelity (Starcevic 2005a). With residue 260 being 12-14 Å away from the active site, the mutator activity of the I260Q pol $\beta$ necessitates a thorough characterization of the structure of the mutant both local to the residue and globally to fully understand the mechanism of the mutator phenotype.

While numerous kinetic studies of pol $\beta$ mutator mutants have revealed a link between residues distant from the active site (Kosa 1999b; Shah 2001b; Dalal 2004; Lin 2007), such as those in the hydrophobic hinge (Shah 2001b; Starcevic 2005a), and an increased rate of nucleotide mis-incorporation (Starcevic 2005b), the structural information explaining possible long-range steric or electrostatic constraints is lacking. Modeling suggested that the more bulky glutamine residue would occupy more space in both apoenzyme and co-crystal structures than the native isoleucine residue (Starcevic 2005a). Here, we present the first high-resolution crystal structures of rat DNA pol $\beta$ with a single alteration at position 260 from isoleucine to glutamine in the presence and absence of substrates. The structures show that subtle changes in the hydrophobic hinge and certain side-chains
alter the dNTP binding pocket at the primer binding site of pol β, and are possibly implicated in the altered activity and fidelity profiles of the I260Q mutator polymerase.

2.4. Materials and Methods

2.4.1. Bacterial strains

The bacterial strain Rosetta2 DE3 with genotype F′ompT hsdSb(rB mB) gal dcm (DE3) pRARE2 (CamR) was used for protein expression. The I260Q mutant was generated through Stratagene site-directed mutagenesis in a pET 28a(+) vector (Novagen) (Starcevic 2005b).

2.4.2. Expression and purification of proteins

The full-length construct of rat pol β was expressed in Rosetta2 DE3 cells as previously described for similar pol β constructs (Kosa 1999a). Cloning and expression was done as previously described (Werneburg 1996) with the exception of the use of pET28a plasmid. Briefly, the pET28a plasmid carrying one variant of pol β was transformed into E. coli strain Rosetta2 (DE3) (Stratagene) to over-express the various construct of pol β. The plasmid pET28a-pol β variants were generously provided by J.B. Sweasy (Yale University of Medicine, New Haven, CT, USA). The transformed cell products were grown in 2xYT media (components) with appropriate antibiotics (Kanamycin at 50 μg/mL for BL21(DE3); Kanamycin at 50 μg/mL and Chloramphenicol at 170 μg/mL for Rosetta2 DE3) at 37°C with vigorous agitation. Upon reaching doubling phase, or until
the OD$_{600}$ reached 06-0.8, 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration) was added to induce protein expression for ~16 hours at 20°C.

Centrifugation yielded cell pellets which were frozen at -80°C prior to purification. The purification of wild-type and I260Q pol β, is based on a published purification protocol of Klenow Fragment (Joyce 1983; Derbyshire 1991). Briefly, the cells were resuspended in 25 mM HEPES, pH 7.0, 100 mM NaCl, 5% glycerol with lysozyme and SigmaFAST Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma-Aldrich) and sonicated (5 sec pulse x 30 sec rest) for ten minutes while suspended in an ice-water bath, followed by ultracentrifugation. The protein mixture was filtered with 5 μm and 0.45 μm syringe filters before being loaded onto a HiTrap Heparin column and separated by a NaCl gradient [Buffer A: 25 mM HEPES, 100 mM NaCl, 5% glycerol, pH 7.0; Buffer B: 25 mM HEPES, 2 M NaCl, pH 7.0] using a liquid chromatography system (Aekta Prime; Amersham Biosciences). Fractions containing pol β were identified by electrophoresis on a 12.5% SDS PAGE gel, and were pooled, then diluted 1:5 with 25 mM HEPES, 25 mM NaCl, pH 7.0, and loaded onto a SP Sepharose column, and again separated by a NaCl gradient [Buffer C: 25 mM HEPES, 25 mM NaCl, pH 7.0; Buffer B: 25 mM HEPES, 2 M NaCl, pH 7.0]. Fractions containing pol β were identified again by electrophoresis on a 12.5% SDS PAGE, and were pooled, then concentrated to <5 mL before being loaded onto a gel filtration column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare Bio-Sciences Corp). Protein was purified and buffer exchanged over the gel filtration column with buffers specific to the crystallization of apoenzyme (20 mM (NH$_4$)$_2$SO$_4$, 100 mM HEPES, pH 7.0) or dsDNA-pol β complex (0.1 M MES, 10 mM
(NH₄)₂SO₄, 30 mM NaCl, pH 6.5). Following gel filtration of the polymerase, the protein was greater than 95% pure as determined by SDS PAGE and dynamic light scattering (data not shown). Concentration of the protein was calculated on the basis of an extinction coefficient of 23,380 M⁻¹cm⁻¹ and a molecular weight of 39 kDa.

2.4.3. In vitro primer extension assays

The substrates chosen in these primer extension assays are based on those used in our co-crystallization studies. The assay conditions are derived from previously published DNA pol β extension assays with slight modifications (Kosa 1999a). Briefly, the primer strand was labeled with a fluorescent probe at the 5'end: 5’-/56-FAM/ATG TGA G-3’, which when annealed to the template strand: 5’-TAC GTC GCG ACT GCT CAC AT-3’ created a 7-base pair duplex with a 13-nt 5’-overhang (Integrated DNA Technologies, Inc., Coralville, Iowa). Primer and template were annealed in 50 mM Tris pH 8.0 and 250 mM NaCl; annealing was verified with dynamic light scattering (data not shown). Enzyme to DNA ratios ranged from 50:1 to 10:1 where DNA remained at 45 nM and the enzyme concentration was changed to suit assay conditions. The assay was performed at 37°C in 20 mM Tris, pH 8.0, 60 mM NaCl, 10% glycerol, 1 mM dithiothreitol (final concentrations). The DNA and protein were mixed with reaction buffer, pre-incubated if indicated, and the reaction was initiated with the addition of dNTPs (50 μM final concentration) and MgCl₂ (10 mM final concentration). At various time points, the reaction was quenched with an equal volume of TBE-Urea Sample buffer, 89 mM Tris,
89 mM boric acid, 2 mM EDTA, pH 8.0, 12% Ficoll, 0.01% bromophenol blue, 0.02% xylene cyanole FF, 7 M urea, (Bio-Rad Laboratories, Inc.) and samples were heated to 95°C for five minutes before being separated on a 15% TBE-Urea Denaturing PAGE gel. Gels were then visualized on a Typhoon Imager (GE Healthcare Bio-Sciences, Corp.) for the detection of fluorescent bands indicating extended versus unextended primer.

2.4.4. Crystallization of the pol β apoenzyme

Pol β was typically concentrated to ~20 mg/mL prior to setting up crystallization trials. Protein to be crystallized in the apoenzyme form was buffer exchanged into 10 mM (NH₄)₂SO₄, 100 mM HEPES, pH 7.0 and crystallized by vapor diffusion in a sitting drop. Protein and well solution were mixed in equal volumes for crystallization. Initial protein concentration was 10 mg/mL using a well solution containing 9-12% PEG3350, 90-120 mM MES, pH 7.0, also initial concentrations. Crystals of the apoenzyme form were cryoprotected with 20-22% 2-methyl-2,4-pentandiol (MPD) prior to flash cooling in liquid nitrogen.

2.4.5. Crystallization of pol β ternary complexes

Protein to be co-crystallized with DNA and dTTP was buffer exchanged into 100 mM MES, pH 6.5, 30 mM NaCl and 10 mM (NH₄)₂SO₄. Pol β wild-type and mutant co-crystals contained primer (5'-ATG TGA G-3') and template (5'-CAA ACT CAC AT-3') (Integrated DNA Technologies, Inc., Coralville, Iowa) resuspended in water and
annealed 1:1 in 20 mM MgSO₄ (90°C for 2 minutes, 70°C for 2 minutes, 55°C for 1 minute, then cooled to 4°C by decreasing 0.5°C per minute) in a DNA Dyad Peltier Thermal Cycler (MJ Research, Inc.). Protein-DNA-ddTTP mixtures, at 230 μM – 290 μM – 2.5 mM, respectively, were aliquoted into 96-well MRC-2 Crystallization Plates (Hampton Research) for sitting drop screens and combined 1:1 with reservoir solution (1.5 μL:1.5 μL). Crystals were grown over a reservoir ranging from 120-200 mM NaCl, 6-14% PEG3350, 3% glycerol and 50 mM cacodylate, pH 6.5. Prior to flash cooling the crystals in liquid nitrogen, a cryo-protectant mixture of 15% glycerol, 8 mM ddTTP and well solution corresponding to the mother liquor was soaked into the drop. The excess ddTTP was added to maintain high occupancy for the active site ligands.

2.4.6. Data collection and structure refinement

X-ray intensity data were collected at beamlines X4C, X29 and X25 at the National Synchrotron Light Source Brookhaven National Laboratory, Upton, NY. Data integration and reduction were performed either using HKL2000 (Otwinowski 1997) or iMosflm.1.5.0 (Battye 2011). Structure determination was initiated using PHENIX.1.7.4 (Adams 2010). The structures of the P2₁ monoclinic apoenzyme pol β crystal forms were determined by molecular replacement using PHASER (McCoy 2007) with a composite model of apoenzyme pol β (X-ray structure of 31 kDa domain and NMR structure of 8 kDa domain) followed by rigid body refinement, full atomic refinement interspersed with manual rebuilding in COOT (Emsley 2010). The ternary complex, in an unrelated
monoclinic space group P2₁, was determined by molecular replacement using an
arrested rat pol β ternary complex (1HUO). Subsequent refinement and model building
were carried out again using PHENIX and COOT. Final structures were analyzed and
verified using COOT, LSQMAN (Kleywegt 1996), MolProbity (Davis 2007) and PyMol
(DeLano 2001).

2.5. Results

2.5.1. Pol β I260Q polymerase activity on the short co-crystallization oligonucleotides

The strong mutator activity of I260Q pol β has been characterized on a 1 (Starcevic
2005a) and 5 (Starcevic 2005b; Dalal 2008a) base-pair gapped DNA substrate. However,
as the mutator polymerase crystallized and diffracted well when complexed with
dsDNA, ddTTP and sodium, we determined the activity of pol β on these co-
crystallization oligonucleotides.

A simple primer-extension assay was designed based on the incorporation of a
fluorescent probe at the 5’-end of the short co-crystallization primer (5’-6-FAM-ATG
TGA G-3’). Wild-type pol β served as a positive control as its activity has been well
characterized for a wide range of substrates and conditions (Kosa 1999b; Shah 2001b)
and a polymerase/DNA mixture without dNTPs or MgCl₂ served as a negative control
(Figure 2.1A, lane 1). Figure 2.1A verifies that both wild-type and I260Q pol β are able
to extend the dsDNA substrate after 5 minutes in the presence (lanes 3-5 and 7-9) and absence (lanes 2 and 6) of pol β-DNA pre-incubation. Pre-incubation allows time for the polymerase to form a stable complex with the DNA substrate. I260Q pol β exhibits a 2.1-fold increase in primer extension compared to wild-type after five minutes, without pre-incubation, as determined by fluorescence intensity of product corresponding to 20 nucleotides in size (Figure 2.1A). This difference is negligible after thirty minutes of pre-incubation.

To determine the ability of I260Q pol β to mis-incorporate nucleotides on the same recessed DNA substrate, we omitted the first or second nucleotide required for faithful incorporation opposite template 3'-TA CAC TCG TCA GCG CTG CAT-5', dCTP and dATP, respectively. Wild-type pol β shows little to no activity in this mis-incorporation assay until after 30 minutes, where only minimal levels of extended primer are detected (Figure 2.1B: lanes 15 and 16). However, I260Q is able to mis-incorporate and extend mispaired dNTPs. While the absence of either dCTP (Figure 2.1B: lanes 3, 7 and 11) or dATP (Figure 2.1B: lanes 4, 8 and 12) results in a lower fidelity I260Q mutator polymerase, as expected, the amount of extended primer is greater for I260Q than wild-type. For primer extension, in the absence of dATP, the correct nucleotide, dCTP, is available for the first round of incorporation but to proceed, the polymerase must then add an incorrect nucleotide. The observed increase in extended primer with I260Q indicates that this mutator variant of pol β is more likely to allow a misincorporation event and create more extended primer product than wild-type in these conditions.
Overall, the dsDNA used for our co-crystallization trials is a suitable substrate for wild-type and I260Q pol β, and, furthermore, I260Q exhibits a mutator phenotype on this substrate whereas wild-type does not.
Figure 2.1. *In vitro* primer extension assays on short double stranded DNA used for co-crystallization.
(A) The activity of I260Q and wild-type pol β was tested by using a 5’FAM-labeled primer, otherwise of the same composition as that used in co-crystallization studies. The DNA:Enzyme ratio was 1:10. Unextended substrate at 7-nucleotides and fully extended product at 20-nucleotides are labeled. The polymerase/DNA complexes were incubated at 37°C prior to initiating primer extension via the addition of MgCl₂ and dNTPs for the length of time indicated below each lane. Unlike wild-type pol β, I260Q is more capable of extending the primer, even in the absence of pre-incubation. This is evident by the amount of unextended primer remaining in lanes 1 and 5 (equivalent to 7-nt) relative to the amount of fully extended primer (equivalent to 20-nt). See Materials and Methods for experimental details. (B) Activity of wild-type versus I260Q mutator pol β in the absence of dCTP or dATP, the first and second nucleotides to be incorporated during primer extension, respectively. While the mutator exhibits moderate extension in the absence of dCTP after 30 minutes, extension in the absence of dATP is observed after just 2 minutes. DNA:Enzyme ratio is 10:1 and all were pre-incubated for 5 minutes prior to the start of reaction.
### Table 2.1. Crystallographic Studies of DNA Polymerase β

<table>
<thead>
<tr>
<th>Rat Pol β Complex</th>
<th>WT apo</th>
<th>I260Q apo</th>
<th>I260Q-dsDNA-ddTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>77.86</td>
<td>77.9</td>
<td>100</td>
</tr>
<tr>
<td>b (Å)</td>
<td>67.83</td>
<td>67.35</td>
<td>56</td>
</tr>
<tr>
<td>c (Å)</td>
<td>83.29</td>
<td>82.37</td>
<td>93</td>
</tr>
<tr>
<td>β (°)</td>
<td>116.07</td>
<td>115.74</td>
<td>102</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>(2.58 - 2.5)(^a)</td>
<td>(2.14 - 2.1)</td>
<td>(2.8 - 2.75)</td>
</tr>
<tr>
<td>(R_{merge}) (%)(^b)</td>
<td>10 (56.8)</td>
<td>9.2 (51.6)</td>
<td>6.3 (52.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>89.6 (87.0)</td>
<td>93.77 (81)</td>
<td>90.1 (66)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>12.45 (1.8)</td>
<td>12.8 (2.0)</td>
<td>18.75 (1.6)</td>
</tr>
<tr>
<td><strong>Reflections (No.)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>64617</td>
<td>175067</td>
<td>196834</td>
</tr>
<tr>
<td>Unique</td>
<td>24378</td>
<td>36809</td>
<td>25065</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.271</td>
<td>1.119</td>
<td>1.479</td>
</tr>
<tr>
<td>(R_{work}) (%)</td>
<td>25.01</td>
<td>23.90</td>
<td>21.82</td>
</tr>
<tr>
<td>(R_{free}) (%)</td>
<td>31.66</td>
<td>29.52</td>
<td>29.20</td>
</tr>
<tr>
<td><strong>Residue Content</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid</td>
<td>650</td>
<td>650</td>
<td>656</td>
</tr>
<tr>
<td>Water</td>
<td>159</td>
<td>281</td>
<td>47</td>
</tr>
<tr>
<td>Cations</td>
<td>N/A</td>
<td>N/A</td>
<td>7 Na(^+)-ions</td>
</tr>
<tr>
<td>PDB ID</td>
<td>3UXN</td>
<td>3UXO</td>
<td>3UXP</td>
</tr>
</tbody>
</table>

\(^a\)Values in parenthesis refer to the highest-resolution shell.

\(^b\)\(R_{merge} = 100 \times \frac{\sum \sum I_{h,i} - I_h}{\sum \sum I_{h,i}}\) where \(I_h\) is the mean intensity of symmetry-related reflections, \(I_{h,i}\).
2.5.2. Apoenzyme structure of wild-type and I260Q DNA polymerase β

To identify structural differences between I260Q and wild-type in the absence of substrate, we crystallized each polymerase without DNA or dNTP. The wild-type polymerase and I260Q mutant crystallized in the same space group, P2₁, with comparable unit cell dimensions (Table 2.1). Data was collected at resolutions of 2.4 and 2.2 Å, for wild-type and I260Q, respectively. Each crystal form contains two molecules per asymmetric unit (ASU). The two molecules in the ASU for apo-wild-type and apo-I260Q superimpose with respective overall root mean square deviations (RMSD) of 1.058 Å and 1.109 Å (Figure 2.2A). For clarity, the wild-type or I260Q apoenzyme pol β molecule with the best corresponding electron density and lowest average temperature factors is shown.

Both apo polymerase structures are full-length and adopt a wide-open conformation of the polymerase and lyase domains (Figure 2.2B) not described before (Shah 2001a; Tang 2008). The distance between Asn24 in the 8 kDa lyase domain and Thr304 at the tip of the fingers subdomain is about 45 Å in both of our apo polymerases (I260Q shown in Figure 2.2A). In the hydrophobic hinge region of wild-type and I260Q apo-pol β, three residues are in close proximity to the side chain at position 260: Tyr296, Arg258 and Leu194 (Y296 and R258 shown in Figure 2.2B). The residues in this hinge must allow for the closure of the fingers subdomain over the palm during catalysis.
The I260Q mutation inserts a polar side chain into a hydrophobic region and causes a cascade of subtle changes, which contribute to the altered fidelity of this pol β variant. The distance between residue 260 and Tyr296 or Leu194 varies only slightly (± 0.4 Å) between wild-type and I260Q apo structures. Arg258 is known to form a weak salt bridge with Asp192 in the open polymerase conformation (Sawaya 1997). The positioning of Arg258, relative to the catalytic Asp192 residue varies between our wild-type and I260Q apo-polymerases: the Arg258 side chain is more proximal to Asp192 (2.7 Å) in the mutant than in the wild-type (3.5 Å). We observe the coordination of several well-defined water molecules in the I260Q apoenzyme structure (Figure 2.2C), which are not evident in the sufficiently detailed electron density maps of wild-type pol β (Figure 2.2D). The presence of water molecules (Figure 2.2C and Table 2.1) in this normally hydrophobic hinge region of I260Q is a direct result of the polar glutamine side chain and the corresponding changes around position 260. The glutamine at position 260 in the pol β variant is also more bulky and occupies more space within the hydrophobic hinge than the native isoleucine.

2.5.3. Structure of pol β I260Q ternary complex

The I260Q mutator pol β variant was crystallized as a ternary complex with a recessed primer:template (5’-ATG TGA G-3’ : 5’-CAA ACT CAC AT-3’), ddTTP and Na⁺ (Figure 2.3A). The structure of the pol β I260Q ternary complex was determined at 2.7 Å, belonging to the monoclinic space group P2₁ with two molecules in the ASU (Table 2.1).
The two molecules in the I260Q ternary complex ASU match closely and palm subdomains superimpose with 0.415 Å RMSD in Cα positions (overall RMSD = 0.861 Å).

An arrested ternary complex of wild-type rat pol β, PDB ID: 1HUO (Arndt 2001), with unit cell dimensions and crystal packing arrangement closely related to our ternary I260Q complex, was used as a trial structure for molecular replacement.
Figure 2.2. The structure of apoenzyme I260Q pol β
Figure 2.2. The structure of apoenzyme I260Q pol β

(A) The asymmetric unit of this new monoclinic crystal form (P2₁) contains two pol β molecules related by a translation of 36 Å approximately along a. Both molecules adopt a wide-open conformation not observed before. The distance between the N-terminal subdomain (Asn24 on helix A) and the tip of the fingers subdomain (residue Thr304) is about 45 Å. (B) Comparison of I260Q and wild-type apoenzyme pol β. The mutant is shown in subdomain colors (8K, red; thumb, green; palm, magenta; fingers, blue) and the wild-type structure is shown gray. Overall, the apoenzyme polymerase structures superimpose well. The amino acids near residue 260, namely Asp192, Leu194 (obstructed), Arg258, Phe272 and Tyr296, relative positions indicated with arrows, show only minor shifts. (C) Water molecules, illustrated as red spheres, enter the hydrophobic hinge region in I260Q apoenzyme pol β. The electron density maps of both I260Q molecules A (and B, not depicted) consistently show water molecules entering the hinge region between the side chains of Gln260, Tyr296 and Arg258. (D) Water molecules are not observed in the corresponding areas of the wild-type electron density maps.
Molecular replacement using 1HUO as a reference model for the I260Q ternary complex produced a single solution with an initial R-factor of 0.37 at 2.8 Å resolution. The 1HUO structure of wild-type pol β represents a coordinated complex consisting of rat pol β:dsDNA:[Cr(III)-dTMPPCP] (Arndt 2001). In that structure, the non-hydrolyzable analog of dNTP allowed the crystallization of the polymerase in the presence of a free hydroxyl group on the primer 3'-terminus, for the first time capturing the fully functional intermediate state. A related ternary structure of wild-type pol β, PDB ID: 2FMP, is a complex of human pol β:2-nt gapped DNA:[Mg²⁺-ddCTP] with Na⁺ coordinated as the catalytic metal (Batra 2006). While the equivalent atoms of the incoming nucleotides (ddTTP in I260Q-ternary and ddCTP in 2FMP) plus metal ions A and B are in superb agreement and superimpose with 0.398 Å RMSD, the overall structures superimpose poorly, with 2.622 - 3.053 Å RMSD, for the two molecules of I260Q-ternary pol β in the ASU. Although the crystallization conditions of the rat and human ternary pol β complexes included different metal ions (I260Q, Na⁺; 1HUO, Cr³⁺; 2FMP, Na⁺/Mg²⁺), there is generally excellent agreement in the residues lining the dNTP-binding pocket. Therefore, the size, charge and chemical nature of these cations do not affect the arrangement of active site residues and with the screening by surrounding residues will not affect the side chain rearrangements near the site of the mutation in the I260Q variant. However, the differences between rat and human pol β, 14 amino acid changes, and the DNA substrates, dsDNA in I260Q-ternary and 1HUO, yet 2-nt gapped DNA in 2FMP, prompted us to carry out a more detailed analysis of the site of mutation.
and the active site center between the rat pol β I260Q and nearly isomorphous wild-type (1HUO) ternary complexes.

Globally, the I260Q mutant is similar to 1HUO in terms of overall architecture and secondary structure elements (Figure 2.3B). For clarity purposes, the molecule with the best corresponding electron density and lowest average temperature factors will be used in our comparisons and discussions, for ternary pol β, I260Q and wild-type. The α-carbon atoms of 303 residues in I260Q and 1HUO can be superimposed with 1.071 Å RMSD. The α-carbon atoms of the N-terminal domains (residues 10-91) and those of the palm subdomains (residues 150-261) agree well with 0.56 Å and 0.45 Å RMSD, respectively. Subtle differences were observed in the fingers subdomain (residues 262-334: 0.80 Å RMSD) and palm residues adjacent to the site of mutation (as discussed below).

Since residue 260 is more than 16 Å away from the metal B in the active site, any effect on polymerase activity must be long-range and indirectly via neighboring residues. The amino acids in the vicinity of 260 interact with Ile and Gln differently in the wild-type versus the I260Q ternary complex of pol β. The amide side chain of Gln260 forms a hydrogen bond with the backbone of Glu295 (Figure 2.3C); this interaction does not occur in wild-type. The hydrogen bond between Gln260 and Glu295 locks-in the orientation of the glutamine side chain (2nd most common rotamer, χ1=-174°) increasing the stability of the closed conformation in the ternary complex of I260Q. Tyr296 and
Arg258, residues on either side of the hydrophobic hinge, highlight the increased constraints in the I260Q variant as compared to the flexible hydrophobic hinge in wild-type. In the I260Q-ternary complex, Tyr296 and Arg258 (the main conformation (70%) of the disordered Arg258) are 1.3 Å closer together than in wild-type (Figure 2.2C & D). Significant side chain movements are also observed in both wild-type and I260Q ternary complexes near I/Q260 involving the disordered Arg258 and Phe272 (Figure 2.3D). These two residues are immediately adjacent to the active site, namely residue Asp192, which directly affects polymerase activity.

The binding of dNTP-metal B causes the fingers subdomain to close upon the active site (Pelletier 1996b). This large conformational change is accompanied by more subtle side chain rearrangements creating the appropriate geometry in the active site. In wild-type (1HUO), Phe272 has rotated down and in between Arg258 and Asp192, keeping these residues 8.4 Å apart (Figure 2.3D). This movement allows Asp192 to participate exclusively in the critical ion-mediated interactions with the two other catalytic aspartates (Asp190 and Asp256). The mutator ternary complex of I260Q pol β shows the aromatic ring of Phe272 tilted (χ2=13°) towards Asp192 where in the wild-type, this aromatic ring (χ2=109°) is pointing away from Asp192 (Figure 2.3C & D). The side chain arrangement of Phe272 is important because of its backbone linkage to Tyr271, a residue that directly interacts with the primer terminus via hydrogen bonding with the minor groove edge of the primer DNA (Beard 2006).
Figure 2.3. I260Q ternary complex with dsDNA and ddTTP
Figure 2.3. I260Q ternary complex with dsDNA and ddTTP

(A) Omit electron density map showing dsDNA and ddTTP bound at the polymerase active site. The bias-free electron density ‘kick’ map was calculated in PHENIX by omitting the substrate atoms from the Fourier calculations and by adding small random motions to the remaining protein atoms. The density map is contoured at 2.5 Å RMSD above the mean density in the ASU. The protein is shown in subdomain colors (see Figure 2.2B for definitions) and the primer and template are shown in silver and gold, respectively. (B) Superposition of the ternary complexes of I260Q and wild-type (1HU0). Color-coding for I260Q as above, whilst the wild-type pol β backbone trace is shown in gray. Note, that the fingers subdomain of I260Q has moved closer to the DNA template as compared to the wild-type complex. (C) Close-up of the structural rearrangements near the site of mutation (I260Q). Note the close interaction of Asp192 and Arg258 in the apoenzyme structure (in lines). In the ternary complex (in sticks) Arg258 shows alternate side chain conformations. Asp192 has rotated into the active site to interact with the metal ions. One alternate conformation (70% occupancy) forms hydrogen bonds with Glu295 (red dashes). (D) Comparison of the relevant active site region in the I260Q and wild-type pol β ternary complexes. Color-coding for I260Q same as above, while the wild-type pol β backbone trace is shown in gray. The metal ions are shown as transparent spheres (Na+, red; Cr3+, grey). Metal ions B near the triphosphate moiety overlap completely.
Movements of Arg258 and Phe272 directly affect the conformation of Tyr271. In the wild-type ternary complex Tyr271 is 2.9 Å from N3 of the primer base, yet, in the I260Q ternary complex, this distance is increased by 1.8 Å. Arg258, no longer interacting with Asp192 in the ternary structure, is observed in two conformations in I260Q-ternary, both different than wild-type. The main conformation of Arg258 in the I260Q-ternary structure (at 70% occupancy) depicts the side-chain within 3.8 Å of Gln260, while the alternate conformation (at 30% occupancy) is now 6.3 Å away from Gln260 (Figure 2.3C). The conformation of Arg258 more distal to Gln260 is observed to interact with Asp192 in the active site (Figure 2.3C), an interaction that is supposed to be interrupted by the rotation of Phe272 in the closed, ternary structure (Figure 2.3D). Thus, the Arg258 switching motion represents a direct link between the presence (or absence) of a polar side chain at 260 and the active site of the polymerase via Asp192.

Finally, we compared the interactions between the ddTTP, Asn279 and nearby residues in the active site. In the human wild-type pol β, Asn279 has been shown to form a fairly strong hydrogen bond between the side chain amine group and O2 of the incoming ddCTP (Beard 1996). In the mutator pol β variant, I260Q, we observed a shift of the Asn279 side chain of about 0.4 Å, away from the ddTTP base when compared to the human or rat wild-type pol β. This small change leaves Asn279 about 3.3 Å away from the incoming nucleotide. The relaxed geometric constraints at the metal binding site (Asp192), primer base positioning (Tyr271) and dNTP-binding (Phe272 and Asp279) explains the mutator activity of I260Q pol β, detailed in our discussion.
2.6. Discussion

Fidelity of DNA replication and repair is dependent on the ability of the DNA polymerase to select and incorporate the correct dNTP to complement the template strand from among a pool of NTP substrates. Despite the lack of an endonuclease domain, or proof-reading activity, pol β only creates one error in every 5,000 incorporation events \textit{in vivo} (Kunkel 1985). As a repair polymerase, the DNA repair fidelity of pol β is crucial to maintain genomic integrity. A single point mutation, I260Q, alters activity and fidelity compared to wild-type, while being located in a region removed from the polymerase active site. I260Q pol β shows a marked decrease in selecting the correct incoming nucleotide (Starcevic 2005a), and, furthermore, an increased probability of incorporating the incorrect nucleotide (Starcevic 2005b) and extending beyond the mispair. Structural comparisons between the wild-type and I260Q polymerases, both in the absence of substrate and in their respective ternary complexes, has allowed us to gain structural insights and a better understanding of the altered phenotype of this mutator mutant.

Overall, the wild-type and I260Q apoenzyme structures differ only slightly in their overall conformations (Figure 2.2A), which is in good agreement with their similar secondary structure content as determined by circular dichroism spectroscopy (Starcevic 2005b). The two pol β polymerases also exhibit comparable levels of polymerase activity in an \textit{in vitro} gap filling assay with all four nucleotides (Starcevic 2005b) and in our \textit{in vitro} primer extension assay with all four nucleotides (Figure 2.1A). The catalytic
efficiency of mispair extension was determined for I260Q and wild-type on a gapped-DNA substrate in the presence of a mispaired primer terminus (A:dATP): I260Q bound the dNTP substrate \( (K_d = 0.6 \mu M) \) and catalyzed its incorporation \( (k_{pol}/K_d = 4.0 \times 10^4 M^{-1} s^{-1}) \) more efficiently than wild-type \( (K_d = 13.8 \mu M; k_{pol}/K_d = 2.75 \times 10^3 M^{-1} s^{-1}) \) (Dalal 2008b).

These findings indicate that the I260Q mutator polymerase is unable to discriminate correct versus incorrect dNTP, at the level of ground-state binding, in the presence of mispaired primer terminus, as compared to wild-type. Since well-diffracting crystals of I260Q pol \( \beta \) in complex with dsDNA/ddTTP were obtained, the mutator activity of this variant was tested on the same dsDNA. Not only is I260Q a strong mutator polymerase on the gapped-DNA substrates (Dalal 2008b), but it also exhibits a tendency to mis-incorporate and extend mispairs on our dsDNA substrate (Figure 2.1B).

Steady-state kinetic studies of pol \( \beta \) reveal a significant difference between wild-type and I260Q in terms of dNTP-induced conformational changes. With stopped-flow fluorescence, Roettger et al. (Roettger 2008) concluded that while wild-type and I260Q pol \( \beta \) both exhibit fast conformational changes in the presence of matched and mis-matched nucleotides, I260Q creates a more stable mis-match ternary complex than wild-type. Additionally, the wild-type polymerase is quick to release the mis-match \( (K_d, T:G = 489 \pm 26 \mu M; C:A = 227 \pm 22 \mu M) \) yet the I260Q mutant polymerase appears to stabilize the mis-match \( (K_d, T:G = 49 \pm 3 \mu M; C:A 45 \pm 3 \mu M) \) which may further explain the enhanced mis-incorporation properties of this polymerase (Roettger 2008). An early hypothesis suggested that the I260Q mutation might translate into altered binding of the
polymerase to DNA substrate and the accommodation of an incorrect nucleotide substrate in the active site of the mutant polymerase (Starcevic 2005b). Of particular interest though is the mechanism by which the mutation I260Q down-modulates the fidelity of the polymerase.

One of the critical conformational steps in catalysis involves the rotation of the pol β fingers subdomain toward the palm subdomain, induced by dNTP-metal ion binding in the active site (Pelletier 1996b). In pol β this movement is intimately linked to the hydrophobic hinge, of which, Ile260 is an integral residue (Starcevic 2005a). This step must be reversible to preserve the fidelity of the polymerase: an incorrect nucleotide, or mis-match, must be released to prevent mis-incorporation. Difficulty in closing of the fingers subdomain upon the palm subdomain may result in a phenotype of a more error-prone polymerase, now unable to efficiently and thoroughly “sample” the incoming nucleotide for proper Watson-Crick geometry according to the template base. Interestingly, both new crystal forms of pol β I260Q presented here show the coordination of water molecules in the “hydrophobic-hinge” region (Figure 2.2C). The appearance of water molecules is a result of the polar glutamine side chain, and the resulting subtle structural rearrangements of adjacent residues, as the hydrophobic hinge of the wild-type structure is free of ordered waters. For the I260Q mutation, some or all of the water molecules may need to be expelled before the fingers subdomain can close around the substrate, which may slow the process or make it less energetically favorable.
The consequences of the I260Q mutation become more evident when the enzyme binds DNA and ddTTP. We have shown that while both wild-type and I260Q apoenzyme structures are globally similar, there are small, yet structurally significant differences near residue 260 that could affect DNA binding. The key differences between Ile260 and Gln260 lie in their interactions with Arg258 and the carbonyl oxygen of Glu295. The interaction with Glu295 in helix M could further stabilize the closed conformation in the I260Q mutant. Furthermore, in wild-type and I260Q pol β apoenzyme, Arg258 interacts closely with Tyr271, Phe272 and Asp192. Upon binding of DNA/dNTP, Arg258 becomes repositioned closer to Gln260 prompting Asp192 to participate in metal coordination at the active site. Concurrently, Phe272 rotates to the space just above and between Arg258 and Asp192 as Tyr271 forms hydrogen bonds with the primer terminus.

In the apoenzyme structures of wild-type and I260Q, Ile/Gln260 is distant from Arg258, which is found in closer proximity of Phe272. In the co-crystal, however, Gln260 is in a position to predominantly interact with Arg258 conformer 1 (70% occupancy) and, consequently, Arg258 and Phe272 are now separated by more than 5 Å, leaving Tyr271 in a sub-optimal position for binding the primer terminus.

Furthermore, Phe272 now adopts a different side chain orientation ($\Delta \chi^2 \sim 70^\circ$) as compared to the wild-type ternary complex (Figure 2.3D). Ile260, due to its hydrophobic nature, is unable to create this strong “cascade” effect leaving Arg258 and Phe272 closer together while Tyr271 remains in a position to interact with the primer.
terminus for efficient nucleotide incorporation. It is the altered positioning of three crucial substrate-binding residues, Arg258, Tyr271 and Phe272, respectively, that plays a significant role in the lack of nucleotide discrimination observed in I260Q pol β.

2.7. Conclusion

To summarize, our structures, in their apoenzyme forms and in complex with primer/template, ddTTP and metal ions, provide a compelling structural foundation to argue that three subtle differences in the hinge region around residues Ile/Gln260, Arg258, Glu295 and Tyr296 coupled with side chain re-arrangements (mainly residues Arg258 and Phe272) immediately adjacent to the dNTP binding pocket are responsible for the altered fidelity profile of the I260Q mutator polymerase.

2.8. Acknowledgments

Access to the New York Structural Biology beamlines X4C at NSLS Brookhaven National Labs and on-site experimental and computational support by J. Schwanof and R. Abramowitz is gratefully acknowledged. Furthermore, we would like to thank A. Heroux, H. Robinson and other beam line staff at PXRR X25 and X29 at NSLS BNL for access and on-site support with data collection, reduction and back-up. Other support by the Macromolecular Crystallography and Biochemistry Instrumentation core facilities at the Wadsworth Center at NYS-DOH and support by Health Research Incorporated is
gratefully acknowledged. This work was supported in part by a grant 2R01CA08030 from the NCI (to JBS and JJ).
Chapter 3. Yunlang Li, Chelsea L. Gridley, Joachim Jaeger, Joann B. Sweasy and Tamar Schlick. Unfavorable electrostatic and steric interactions in DNA polymerase β E295K mutant interfere with the enzyme’s pathway. Journal of the American Chemical Society, Just Accepted Manuscript, 31 May 2012, DOI: 10.1021/ja300361r

3.1. Description of participation

In this manuscript, I designed and performed all of the crystallography experiments with E295K. I also analyzed the structural data and used that information to generate Supplemental Figure 3.8.3. The sections of the manuscript regarding methods, results and discussion pertaining to the crystal structure of E295K were prepared by myself along with Dr. Joachim Jaeger. Yunlang Li and Tamar Schlick designed and performed all other experiments described within and also wrote the manuscript. Joann B. Sweasy edited the manuscript for content and provided helpful suggestions.
3.2. Abstract

Mutations in DNA polymerase β have been associated with approximately 30% of human tumors. The E295K mutation of pol β has been linked to gastric carcinoma via interference with base excision repair. To interpret the different behavior of E295K compared to wild-type pol β in atomic and energetic detail, we resolve a binary crystal complex of E295K at 2.5 Å and apply transition path sampling to delineate the closing pathway of the E295K pol β mutant. Conformational changes are important components in the enzymatic pathway that lead to and ready the enzyme for the chemical reaction. Our analyses show that the closing pathway of E295K mutant differs from the wild-type pol β in terms of the individual transition states along the pathway, associated energies, and stability of the final closed form relative to the open form of the enzyme. In particular, the closed state of E295K is less stable than the open state by ~5 k_BT, in contrast to the wild-type pol β which has a more stable closed than open form by ~3 k_BT. This instability can be explained by the distorted active site in the closed state. In addition, the total energy barrier in the conformational closing pathway is 26 k_BT, much higher than that estimated for both correct (e.g., G:C) and incorrect (e.g., G:A) wild-type pol β systems (17 k_BT and 18 k_BT, respectively). In particular, the rotation of Arg258 is the rate-limiting step in the conformational pathway of E295K due to unfavorable electrostatic and steric interactions. The instability of closed relative to open state and the high-energy barrier in the conformational pathway may explain in part why the E295K mutant is observed to be inactive. Interestingly, however, following the closing of the fingers, but prior to key residue motions, the E295K mutant complex is
slightly more stable than the wild-type pol β. This suggests that the E295K mutant may associate with DNA more tightly but may not be able to continue the process of chemistry. Supporting experimental data come from the observation that the catalytic activity of wild-type pol β is hampered when E295K is present: this may arise from the competition between E295K and wild-type enzyme for the DNA. These combined results suggest that the low insertion efficiency of E295K mutant compared to wild-type pol β may be related to a less stable closed-form, destabilized by unfavorable electrostatic and steric interactions between Arg258 and other key residues. The active site is thus less competent for proceeding to the chemical reaction, which may also involve a higher reaction barrier than the wild-type, or may not be possible in this mutant. Our analysis also suggests further experiments for other mutants to test the above hypothesis and dissect the roles of steric and electrostatic factors on enzyme behavior.
3.3. Introduction

DNA polymerases are crucial to DNA replication and repair, which are of great importance to maintaining the genomic stability of living organisms. Various cancers, neurological diseases and premature aging have been related to malfunctions of DNA polymerases (Iwanaga 1999; Starcevic 2004; Lang 2007; Dalal 2008a; Hoeijmakers 2009; Jackson 2009; An 2011; Nicolay 2011). The eukaryotic DNA polymerase β (pol β), in the X-family of DNA polymerases, functions primarily in base excision repair (BER) (Seeberg 1995; Kunkel 2004), which involves binding to DNA and filling single-stranded gaps in DNA guided by the corresponding undamaged strands base with moderate accuracy (“fidelity”) (Friedberg 2003). In addition, pol β also has 5'-deoxyribose-5-phosphate lyase (dRP lyase) activity (Matsumoto 1995), which allows it to remove the sugar-phosphate overhang generated during BER. Since BER is considered to play a key role in cancer as well as aging (Maynard 2009), pol β is important in this context because of its prominent role in BER (Loeb 2008).

Pol β is shaped like a hand with fingers (residues 263-335), palm (152-262), thumb (92-151) subdomains, and an 8 kDa (1-91) domain (Steitz 1994). X-ray crystallography has provided exquisite views of two forms of the enzyme (Pelletier 1994; Sawaya 1994; Sawaya 1997): the closed (ternary, active) and open (binary, inactive) forms are related by a large subdomain motion of the fingers. Kinetic, structural, and computational studies (Werneburg 1996; Vande Berg 2001; Arora 2004, 2005b) have revealed that the catalytic pathway of pol β follows the general, three-step nucleotide insertion pathway.
for DNA polymerases: first, following DNA binding, pol β incorporates a 2'-deoxyribonucleoside 5'-triphosphate (dNTP) to form an open complex, which undergoes a conformational change to align active site residues and form a closed complex; second, the closed pol β complex catalyzes the nucleotidyl transfer reaction and forms the product complex which is also closed; third, the product complex undergoes a reverse conformational change back to the open form, allowing the release of pyrophosphate (PPi). During the conformational steps, besides the large fingers subdomain motion between open and closed form, side chain motions of several key residues such as Asp192, Arg258, Tyr271, Phe272, and Arg283 (Beard 1996; Kraynov 1997; Vande Berg 2001; Yang 2002; Radhakrishnan 2004b; Yang 2004b; Arora 2005b), as well as the movements of Mg²⁺ ions (Radhakrishnan 2004b; Yang 2004a; Arora 2005b), participate in transitioning the enzyme to/from the reaction competent state.

Depending on the polymerase in question, associated substrates, and other conditions, both conformational and chemical steps can be rate-limiting in general (Radhakrishnan 2006b). Therefore, an understanding of enzyme catalysis requires knowledge of many steps, conformational and chemical, with the latter being just one part of the big picture (Nagel 2010). The combined picture of the energy landscape including conformational and chemistry pathways helps link structural and energetic details with experimental methods and pursue enzyme design applications, for example, to explain the role of different residues, to design mutants with altered functions (Lin 2008; Schlick 2011) and to analyze heterogeneous enzyme mixtures where some enzyme become trapped in
conformational states prior to chemistry (Beard 2003; Tsai 2006; Johnson 2008; Kamerlin 2010; Karplus 2010; Kellinger 2010; Nashine 2010; Santoso 2010b; Ram Prasad 2011). The argument that conformational pathways that are not rate-limiting overall in the enzyme cycle are not relevant to fidelity (Ram Prasad 2011) is insufficient for analyzing complex kinetic observations and pursuing design applications.

Thirty percent of human tumors studied express DNA pol β mutations not present in normal tissue (Iwanaga 1999; Starcevic 2004; Dalal 2008a), including gastric carcinoma-associated pol β mutant E295K (glutamic acid to lysine) (Iwanaga 1999; Lang 2007) (See Supplementary Table 3.9.1 for comparison of kinetic data of pol β mutants). The E295K mutant has been shown to be inactive on a 45-nt DNA duplex containing a 1-nt gap in the middle, opposite template A (Lang 2007). It interferes with BER and induces sister chromatid exchanges (SCEs) and cellular transformation. However, the dRP lyase activity of the E295K mutant is retained, suggesting that it still can bind to the DNA. Interesting, not only does the E295K mutant itself lack activity on a single-gapped DNA substrate, but it also interferes with wild-type pol β in gap-filling synthesis during BER when both the mutant and wild-type pol β are present (Iwanaga 1999). Experimental evidence indicates that even half the concentration of E295K compared to wild-type pol β species reduces the amount of DNA synthesis products (Lang 2007). However, the reason for the inactivity of the E295K mutant and its mode of interference compared to the wild-type pol β remain unclear. It has been hypothesized that the E295K mutant lacks the ability to interact with Arg258 and sequester it away from Asp192 (Lang 2007).
To help unravel the factors that explain the inactivity of the E295K pol β mutant, we use enhanced configurational sampling with the CHARMM all-atom force field (including the cross-term energy correction map, or CMAP, specification for proteins) (MacKerell 1998; MacKerell 2000; Mackerell 2004) to study the closing pathway of E295K before chemistry and compare results to reference systems. Though far from perfect, modeling and simulation have gained accuracy and reliability to study biomolecular systems (Schlick 2011). Here, we apply transition path sampling (TPS) (Dellago 1998; Bolhuis 2000) simulations, an approach developed by Chandler and co-workers to traverse high barriers on the free energy surface and capture rare events that are not accessible to regular molecular dynamics (MD) simulations. TPS has been applied to study various systems from small molecules such as peptides and lipids (Bolhuis 2000; Bolhuis 2003; Hagan 2003; Evans 2004; Marti 2004a; Marti 2004b; Juraszek 2006) to large, complex systems (Radhakrishnan 2004b; Basner 2005; Radhakrishnan 2005; Quaytman 2007; Wang 2007; Saen-Oon 2008). In a previous work, we implemented TPS to study the closing and/or chemical pathways of a pol β complex binding dATP and dCTP, opposite dG (Radhakrishnan 2004b, 2005) as well as 8-oxoguanine (8-oxoG) (Wang 2007). The free energy profiles of those systems emerge as different both in terms of the transition states identified and the associated energy values; these differences, in turn, help interpret differing nucleotide insertion efficiency as revealed experimentally (Beard 1996; Shah 2001b; Vande Berg 2001).
Here, we show that the overall closing energy barrier of the E295K mutant is higher than any of the above four related systems (26 k_BT versus 15-18 k_BT) (Radhakrishnan 2004b, 2005; Wang 2007). The rate-limiting step in the conformational pathway before chemistry is the rotation of Arg258, which agrees with the previous hypothesis by Lang and colleagues (Lang 2007). Our analyses further show that the high-energy barrier caused by Arg258 is due to the drastic changes in the electrostatic environment around it, as well as the steric hindrance caused by the Lys295 with Arg258. The instability of the final closed form of the E295K mutant relative to the open form, together with the distorted active site observed in the closed form, also suggest the inactivity of the mutant. Interestingly, however, midway through the incorporation of the incoming nucleotide, following the fingers motion but prior to major key residue motions, the E295K mutant has lower energy than the wild-type complex; this suggests that the E295K mutant can bind DNA and lock the complex into a half-closed state. This is consistent with the experimental observation that the E295K mutant maintains the affinity for DNA but not the catalytic ability (Lang 2007). These marked differences of E295K from wild-type pol β also suggest further experiments on other mutants like E295Y and E295W, where neutral and bulky tyrosine and tryptophan replace Glu295.
3.4. Methods

3.4.1. Expression, purification and co-crystallization

The tagless construct of rat pol β was expressed in Rosetta2 DE3 cells as previously described for similar pol β constructs (Kosa 1999a). The purification of tagless pol β E295K protein is based on a published purification protocol of Klenow Fragment (Joyce 1983; Derbyshire 1991). Briefly, the cells were resuspended in 25 mM HEPES, pH 7.0, 100 mM NaCl, 5% glycerol with lysozyme and the SigmaFAST protease inhibitor cocktail. Cells were sonicated 5 x 30 seconds for ten minutes on ice, followed by centrifugation at 25,700 g. The protein mixture was syringe filtered through a 5 μm followed by a 0.45 μm filter then loaded onto a HiTrap Heparin column and separated by a NaCl gradient [Buffer A: 25 mM HEPES, 100 mM NaCl, 5% glycerol, pH 7.0; Buffer B: 25 mM HEPES, 2 M NaCl, pH 7.0]. Fractions containing pol β were pooled, diluted 1:5 and purified on a SP Sepharose column [Buffer C: 25 mM HEPES, 25 mM NaCl, pH 7.0; Buffer B: 25 mM HEPES, 2 M NaCl, pH 7.0]. The final purification step consisted of a gel filtration column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare Bio-Sciences Corp). Protein was purified and buffer exchanged over the gel filtration column with buffer specific to crystallization of the dsDNA-enzyme complex (100 mM MES, 10 mM (NH₄)₂SO₄, 30 mM NaCl, pH 6.5).

Binary complexes of E295K pol β were co-crystallized with DNA in 100 mM MES, pH 6.5, 30 mM NaCl and 10 mM (NH₄)₂SO₄. For E295K co-crystallization trials, the primer
(5′-ATG TGA G-3′) and template DNA (5′-CAA ACT CAC AT-3′) (Integrated DNA Technology) were resuspended in water and annealed 1:1 in 20 mM MgSO₄ (90°C for 2 minutes, 70°C for 2 minutes, 55°C for 1 minute, then cooled to 4°C by decreasing 0.5°C every cycle) in a DNA Dyad Peltier Thermal Cycler (MJ Research, Inc). Protein-DNA mixtures, adjusted to 231 μM and 289 μM, respectively, were dispensed into 96-well MRC-2 Crystallization Plates (Hampton Research) for sitting drop crystallization screens and combined 1:1 with reservoir solutions ranging from 120-200 mM NaCl, 6-14% PEG3350, 3% glycerol and 50 mM cacodylate, pH 6.5. Prior to flash-cooling in liquid nitrogen, a cryoprotectant mixture of 15% glycerol plus the mother liquor in the corresponding crystallization well was added to the harvested crystals. Concentration of the protein was calculated on the basis of an extinction coefficient of 23,380 M⁻¹cm⁻¹ and a molecular weight of 39 kDa.

3.4.2. Structure determination and refinement

X-ray intensity data were collected at 100 K on beam line X4C at the National Synchrotron Light Source in Brookhaven, NY. Data integration and reduction were performed using HKL2000 (Otwinowski 1997). The structure was determined and phases calculated by molecular replacement (AutoMR/PHASER in PHENIX.1.7.4) (Adams 2010) followed by rigid body refinement and full atomic refinement with PHENIX interspersed with manual rebuilding in COOT (Emsley 2010). The final
structure was analyzed and verified using COOT, MolProbity (Davis 2007) and PyMol (DeLano 2001).

3.4.3. System preparation

Our starting models are based on the wild-type pol β/DNA substrate complexes from the binary (PDB ID code 1BPX) and ternary (PDB ID: 1BPY) crystal structures (Sawaya 1997), respectively, as well as the binary crystal structure of the E295K mutant (PDB ID: 3V72). Since the binary structure does not contain an incoming nucleotide, a dCTP residue and two binding Mg\(^{2+}\) ions were added to pair with the dG template using the ternary crystal structure as reference. The Glu295 in the closed pol β complex was then modified to Lys using the CHARMM program (version c35b2) (Brooks 1983). Manual modifications for Lys295 and nearby residues were made to reduce possible collisions and provide more possible starting conformations for Lys295, as Lys295 is indicated by the crystal structure to move fast (see results below). Though we have prepared four different pairs of starting models (with different side chain orientations of several key residues such as Lys295, Arg283, and Phe272 etc.), after the equilibration and targeted MD (TMD) process described below, these structures converge well into one another. Therefore, the starting forms are likely reliable and do not affect our further analysis significantly. All missing atoms were added to the models according to the ternary crystal structure by CHARMM. The Na\(^+\) occupying the catalytic ion site in the crystal structure was modified to Mg\(^{2+}\) in the binary system.
The system was solvated with the explicit TIP3 water model in a water box via the VMD program (Humphrey 1996). The smallest image distance between the solute and the faces of the periodic cubic cell was set to 7 Å. The total number of water molecules is 12,828. To obtain a neutral system at an ionic strength of 150 mM, 47 Na\(^+\) and 27 Cl\(^-\) ions were added to the system. All of the Na\(^+\) and Cl\(^-\) ions were placed at least 8 Å away from both the protein and DNA atoms and from each other.

The initial model contained 44,918 atoms, 107 crystallographically resolved water molecules from the ternary complex, 12,721 bulk water molecules, two Mg\(^{2+}\) ions, incoming nucleotide dCTP and 47 Na\(^+\) and 27 Cl\(^-\) counter ions.

3.4.4. Minimization and equilibration

Initial energy minimizations and equilibration simulations were performed using CHARMM. The system was minimized with fixed positions for all heavy atoms of protein or nucleotides, using steepest descent (SD) for 10,000 steps followed by the adopted basis Newton-Raphson (ABNR) method for 20,000 steps. The equilibration process was started with a 100 ps simulation at 300 K using the single-time step Langevin dynamics, while keeping all the heavy atoms of protein or nucleotides fixed. The SHAKE algorithm was then employed to constrain the bonds involving hydrogen atoms. This was followed by unconstrained minimization consisting of 10,000 steps of
SD and 20,000 steps of ABNR. The system was then transferred to NAMD (Phillips 2005) and equilibrated for 1 ns at constant pressure and temperature. Pressure was maintained at 1 atmosphere (atm) using the Langevin piston method with a piston period of 100 fs, a damping time constant of 50 fs, and a piston temperature of 300 K. The temperature was maintained at 300 K using weakly coupled Langevin dynamics (Feller 1995) of nonhydrogen atoms with a damping coefficient of 10 ps$^{-1}$. Bonds to all hydrogen atoms were kept rigid using SHAKE, producing good stability with a time step of 2 fs. The system was simulated in periodic boundary conditions with full electrostatics computed using the particle mesh Ewald method (Darden 1993) with grid spacing on the order of $\leq 1$ Å. Short-range nonbonded terms were evaluated at every step using a 12 Å cutoff for van der Waals interactions and a smooth switching function. The final dimensions of the system were 77.5 Å x 77.3 Å x 77.4 Å.

Following minimization and equilibration, the overall root mean square deviation (RMSD) values ($\alpha$ atoms) of the open and closed form models relative to their initial structures are 1.14 Å and 1.26 Å, respectively.

3.4.5. Transition path sampling simulations

The TPS method relies on the idea of importance sampling using standard Monte Carlo (MC) procedures, which explore sequences of states constituting dynamical trajectories (Pratt 1986; Bolhuis 2002) through random walks. Starting from an initial trajectory
(generated here by TMD) that captures a barrier crossing, TPS samples the trajectory space using the Metropolis MC method by performing a random walk with the shooting algorithm (Bolhuis 1998); the random walk is biased to make sure that the most important regions of the trajectory space are adequately sampled (Bolhuis 2002). The frequency of a trajectory region being visited is determined by its probability so that, even when a random walk is initiated far from a representative transition pathway, the bias can drive the system to important regions of the transition space after sampling. Thus, despite the unphysical nature of the initial sampling trajectory obtained using TMD, the TPS protocol can lead the system to the most important regions and yield physically meaningful trajectories passing through saddle area. See supporting material for further details on TPS. The most challenging part of TPS is to describe the order parameters representing the transitions. Our prior work was valuable here (Radhakrishnan 2004b, 2005; Wang 2007), as described next.

To obtain the initial trajectories that connect the two states during the transition, we applied TMD simulations to connect our modeled open and closed forms of the pol β E295K mutant complexes. To choose appropriate order parameters for TPS simulations, we used the crystallographic data (Sawaya 1997), molecular dynamics (Arora 2004, 2005b) and prior TPS studies (Radhakrishnan 2004b, 2005; Wang 2007) on wild-type pol β as reference. Since these works have shown that key active-site residues (Asp192, Arg258, Tyr 271 and Phe272), the α-helix N on the fingers subdomain and the Mg²⁺ motion serve as measures of pol β’s closing pathway, we started testing values
associated with these residues and ions as well as the RMSD value of α-helix N atoms (residue 275 to 295). When piecing together the entire closing pathway, we found, in addition to above, the complete set of order parameters as listed in Table 3.1. We used the TMD code implemented in NAMD to generate the initial constrained trajectories. An energy restraint based on the RMS distance of the system relative to the final form was applied to force the open pol β complexes to close. The restraint energy can be expressed by:

$$ERMS = K[DRMS(X(t),X^{target}) - d_0]^2$$

In this equation, $K$ is a force constant, $DRMS$ is the relative RMS distance for a selected set of atoms between the instantaneous conformation $X(t)$ and the reference $X^{target}$, and $d_0$ is an offset constant (in Å). In our TMD simulations, the RMS distance is evaluated using the heavy atoms on the pol β mutant. A total force constant of 3,000 kcal mol$^{-1}$Å$^{-2}$ was applied to all heavy atoms of the complex. The offset parameter $d_0$ was set to decrease from 5.2 Å to 0 in 400 ps, as the mutant is driven from the open to the closed form. From the TMD trajectory, six transition regions were identified (Table 3.1 and Supplementary Figure 3.9.1).

From the TMD trajectory, we selected frames that bracket the transition regions associated with specific residues/ions and thumb conformational changes and performed unconstrained dynamics simulations. The atomic momenta of the frames
were perturbed and we integrated the equations of motion forward and backward over short trajectories of order 10 - 100 ps (see below) to generate new physical, unbiased trajectories that connected the open and closed states of pol β. Based on these unconstrained simulations we determined the adequate length of sampling trajectories for all the transition states. Specifically, for the mutant complex, the trajectories for Arg258 rotation were simulated for 20 ps, and those trajectories for other key residue and Mg$^{2+}$ motions were run for 10 ps. To capture the transition states of fingers closing in the two complexes, the sampling trajectories have to be propagated for 100 ps.

Using one of the newly generated physical trajectories as the starting trajectory, we perform path sampling for each individual conformational change with the shooting and shifting algorithm and a Monte Carlo protocol. The entire process was performed by using a PERL script that interfaces with NAMD to generate new trajectories. The velocity Verlet integrator in NAMD with a time step of 1 fs was used for generating the individual molecular dynamics trajectories in TPS. All other parameters are the same as those in the equilibration process. To obtain an acceptance rate of 30-45%, the momentum perturbation magnitude (dP) of each transition state are varied from 0.001 to 0.005. A total of 200 accepted trajectories were collected to map each transition state.

The convergence of the harvested sampling trajectories was verified by computing the autocorrelation function associated with order parameters to check for decorrelation of paths (see supporting material and Supplementary Figure 3.9.2). The new trajectories
are essentially decorrelated if the autocorrelation function shows a gradual transition between $<\chi_A^2>$ and $<\chi_A^2><\chi_B>$.

### 3.4.6. Free energy barrier and rate constant calculations

The free energy barriers for transition states were evaluated using the "BOLAS" protocol (Radhakrishnan 2004a), an efficient procedure for getting free energies with relatively low error bars using the TPS trajectory harvesting idea. Namely, we divided the reaction coordinate of each transition into 10 small overlapping windows and performed umbrella sampling to generate 200 trajectories on each window. We then combined the potential of mean force plots obtained from the sampling calculations on each window by adding/subtracting a constant to match the free energy values of the overlapping region. From the overall free energy plots, we calculated the free energy barriers for the conformational transitions. The error bar for the free energy calculations was determined by repeating umbrella sampling on one window of a transition for five times with the same initial trajectory but different starting pseudorandom numbers. The largest standard deviation (~3 k_BT) was used as the error bar, and is comparable to prior works (Radhakrishnan 2004b, 2005; Wang 2007).

The rate of the transition between adjoining metastable states is estimated using transition state theory as:

$$k_{TST}^{AB} = \frac{1}{\tau_{mol}^{AB}} e^{-\beta \Delta F_{\text{barrier}}^{AB}}$$
where the characteristic relaxation time $\tau_{\text{mol}}$ is estimated by the time taken from the gradual transition of the autocorrelation function $<\chi_i(0)\chi_i(0)>$.

3.5. Results

3.5.1. Crystallographic studies

The crystals of the pol $\beta$ E295K binary complex belong to space group $P2_12_12_1$ with one molecule per asymmetric unit (cell dimensions of $a=57.2$ Å, $b=73.8$ Å, $c=118.4$ Å and $\alpha, \beta, \gamma = 90.0^\circ$, $V_m = 2.66$ Å$^3$/Da). The crystal structure (PDB ID: 3V7L) was refined using data extending to 2.49 Å Bragg spacings and the final model shows excellent agreement with the density maps throughout ($R_{\text{free}} = 0.29$, RMSD bonds $= 0.009$ Å and angles $= 1.13^\circ$). A representative region of the density map near the site of mutation is shown in Supplementary Figure 3.9.3, with the map contoured at 1.25 Å RMSD above the mean electron density in the asymmetric unit. The DNA template strand and adjacent residues Arg258, Tyr271, Phe272, Arg283 and Tyr296 are very well defined. The residues adopt the same position seen in related binary complexes of wild-type pol $\beta$ (1BPY). However, the lack of electron density for Lys295 atoms C$\beta$ to N$\gamma$ is striking and indicative of fast moving, ill-defined side chain atoms.

The lack of definition in the electron density corresponding to the Lys295 side chain atoms is in excellent agreement with the flexibility of this residue observed during the
simulations. The temperature factors of the Lys295 side chain atoms are up to 1.5 times higher than those of the immediately adjacent tyrosine residues (Tyr271 and Tyr296).

3.5.2. Transition-state identification

Our analyses of detailed closing pathways before chemistry identify six transition states (Figure 3.1), compared to five and four transition states for matched (G:C) (Radhakrishnan 2004b) and mismatched (G:A) (Radhakrishnan 2005) wild-type systems, respectively. The sequence of changes along the pathway of the E295K mutant is illustrated in Figure 3.1: (1) flip of Asp192; (2) partial fingers closing; (3) flip of Phe272; (4) rotation of Arg258; (5) shift of Tyr271; (6) ion rearrangement characterized by the shift of nucleotide-binding Mg$^{2+}$ (Table 3.1). Each transition state is characterized by the probability distribution in Figure 3.2. The existence of two peaks Figure 3.2 (marked by 0 and 1) within one plot implies a conformational transition across the energy barrier. Therefore, a path-sampling trajectory of TS1 to TS6 captures the entire conformational pathway.
Figure 3.1. Molecular snapshots near open (upper) and closed (lower) states for six transition state regions

(1) Flip of Asp192.  (2) Partial fingers closing.  (3) Flip of Phe272.  (4) Rotation of Arg258.  
(5) Shift of Tyr271.  (6) Ion rearrangements.  Mg$^{2+}$, orange; key residues, purple; dCTP and C10, black; residue characterizing the transition state, red.
Table 3.1. Transition states properties for the closing conformational profile of the E295K pol β mutant

<table>
<thead>
<tr>
<th>TS</th>
<th>Event</th>
<th>χ-order parameter</th>
<th>χ&lt;sub&gt;max&lt;/sub&gt; State A</th>
<th>χ&lt;sub&gt;min&lt;/sub&gt; State B</th>
<th>τ&lt;sub&gt;mol&lt;/sub&gt; (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Asp192 Flip</td>
<td>Dihedral angle C&lt;sub&gt;γ&lt;/sub&gt;-C&lt;sub&gt;β&lt;/sub&gt;-C&lt;sub&gt;α&lt;/sub&gt;-D</td>
<td>150°</td>
<td>180°</td>
<td>6 ± 1</td>
</tr>
<tr>
<td>2</td>
<td>Partial fingers motion</td>
<td>RMSD of residues 265-295 with respect to closed form</td>
<td>3.2 Å</td>
<td>2.5 Å</td>
<td>60 ± 10</td>
</tr>
<tr>
<td>3</td>
<td>Phe272 Flip</td>
<td>Dihedral angle C&lt;sub&gt;α&lt;/sub&gt;-C&lt;sub&gt;β&lt;/sub&gt;-C&lt;sub&gt;γ&lt;/sub&gt;-C&lt;sub&gt;δ2&lt;/sub&gt;</td>
<td>110°</td>
<td>40°</td>
<td>3 ± 1</td>
</tr>
<tr>
<td>4</td>
<td>Arg258 rotation, fingers closing</td>
<td>Dihedral angle C&lt;sub&gt;γ&lt;/sub&gt;-C&lt;sub&gt;δ&lt;/sub&gt;-N&lt;sub&gt;ε&lt;/sub&gt;-C&lt;sub&gt;ζ&lt;/sub&gt;</td>
<td>100°</td>
<td>170°</td>
<td>12 ± 3</td>
</tr>
<tr>
<td>5</td>
<td>Shift of Tyr271</td>
<td>Distance of Tyr271: OH – Arg283: N&lt;sub&gt;ε&lt;/sub&gt;</td>
<td>6.9 Å</td>
<td>8.6 Å</td>
<td>8 ± 1</td>
</tr>
<tr>
<td>6</td>
<td>Ion motion</td>
<td>Distance of nucleotide-binding Mg&lt;sup&gt;2+&lt;/sup&gt; to O&lt;sub&gt;1α&lt;/sub&gt; of incoming nucleotide</td>
<td>4.4 Å</td>
<td>2.7 Å</td>
<td>7 ± 1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Time required to transverse the transition region as obtained from the correlation function of each transition path sampling (see Supplemental Figure 3.9.2)
Figure 3.2. Normalized probability distribution of the order parameters for the transition states revealed (TS1 to TS6). Labels 0 and 1 indicate the open and closed form conformations.
The first transition captured in our mutant system is the flip of Asp192, while in wild-type pol β systems, the flip of Asp192 almost always happens after the partial fingers closing, except in the 8-oxoG:C system, which lacks this transition state. Furthermore, following partial fingers closing, the next residue to rearrange in E295K is Phe272; this sequence occurs in the G:A mismatched system but not in the other wild-type systems, where the rotation of Arg258 occurs not after, but before, Phe272. The fingers subdomain remains half-closed (RMSD of residues 275-295 relative to the closed structure is ~2.5 Å) until the flip of Phe272 and rotation of Arg258 are completed (final RMSD is ~1.6 Å). The shift of Tyr271 was not characterized as a transition state in our previous TPS studies on wild-type pol β systems, though a similar shift of Tyr271 was captured using the stochastic difference equation (SDEL) method (Arora 2005b). Here, we observe that Tyr271 shifts after all other key residue motions are completed, which agrees with our previous observation that Tyr271 only moves after the transition state is stabilized by the fingers closing and the active site assembly (Arora 2005b). Following the motions of the fingers and key residues, subtle ion motions occur to seal rearrangements of the active site prior to the chemical reaction step.

3.5.3. Potential of mean force calculation

The free energy changes associated with each transition event are used to construct the overall reaction kinetics profiles in Figure 3.3 (also see Supplementary Figure 3.9.4). The barriers corresponding to TS5 (shift of Tyr271) and TS6 (ion rearrangements) are small
compared to TS1 to TS4; given the value $\approx 2 \pm 3 \text{ k} \theta \text{T}$, they may be negligible overall. The overall activation free energy for the conformational pathway in the mutant system is notably higher than the wild-type system: $26 \pm 3 \text{ k} \theta \text{T}$ versus $19 \pm 3 \text{ k} \theta \text{T}$. The much higher conformational energy barrier may hamper the system to reach the active closed form. In addition, the higher energy of the final closed form of the mutant compared to the starting open form of the mutant by $\approx 5 \text{ k} \theta \text{T}$ also contributes to the mutant’s hampered pathway. Based on these free energy values, we compute the rate constants for each transition in Table 3.2.

The high-energy barrier of the E295K mutant mainly arises from the large barrier of Arg258, which is $\approx 19 \text{ k} \theta \text{T}$, much higher than the 10 k$\theta$T barrier of Arg258 rotation in the wild-type system. In Figure 3.4 we compare the key residues around the active site in both the E295K mutant and wild-type system. Significantly, the hydrogen bond between Arg258 and Glu295 in the wild-type system, which may help stabilize the position of Arg258 in the closed form, is missing in the mutant system. Experiments show that this hydrogen bond is also crucial for promoting the binding of the catalytic ion in the final closed state (Kirby 2011). Besides the lack of this hydrogen bond, as shown in Figure 3.5, the electrostatic surface of the E295K mutant in both the open and closed forms is generally positive, while the surface of the wild-type system is more negative; thus the rotation of Arg258 in the mutant is unfavorable due to electrostatic repulsion.
Figure 3.3. The free energy profile of the pol β E295K mutant system

The superimposed dashed line is free energy profile of wild-type pol β system. The open and closed forms of the E295K mutant are drawn in green and red, respectively. Mg$^{2+}$, orange (except in the diagram of ion motions, where green and red are used); dCTP, black; Lys295, blue; key residues involved in transition states (Asp192, Arg258, Tyr271 and Phe272), purple.
Table 3.2. Free energy barrier and rate $k_{TST}$ estimated by transition-state theory

<table>
<thead>
<tr>
<th></th>
<th>Asp192 Flip</th>
<th>Fingers Closing</th>
<th>Phe272 Flip</th>
<th>Arg258 Rotation</th>
<th>Tyr271 Shift</th>
<th>Ion Motions</th>
<th>Total$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E295K Mutant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>5</td>
<td>10</td>
<td>4.5</td>
<td>19</td>
<td>1.5</td>
<td>0.5</td>
<td>26</td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>2</td>
<td>5</td>
<td>5.5</td>
<td>18.5</td>
<td>2.5</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>$k_{TST}^{A \rightarrow B}$ (s$^{-1}$)</td>
<td>$1.1 \times 10^9$</td>
<td>$7.6 \times 10^5$</td>
<td>$3.7 \times 10^9$</td>
<td>$4.7 \times 10^2$</td>
<td>$2.8 \times 10^{10}$</td>
<td>$8.6 \times 10^{10}$</td>
<td>–</td>
</tr>
<tr>
<td>$k_{TST}^{B \rightarrow A}$ (s$^{-1}$)</td>
<td>$2.2 \times 10^{10}$</td>
<td>$1.2 \times 10^8$</td>
<td>$1.4 \times 10^9$</td>
<td>$7.7 \times 10^2$</td>
<td>$1 \times 10^{10}$</td>
<td>$1.9 \times 10^{10}$</td>
<td>–</td>
</tr>
<tr>
<td><strong>Wild-type: matched system (G:C)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>4</td>
<td>14</td>
<td>4</td>
<td>10</td>
<td></td>
<td>7.5</td>
<td>17</td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>16.5</td>
<td>9</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>$k_{TST}^{A \rightarrow B}$ (s$^{-1}$)</td>
<td>$(5 \times 10^9)^c$</td>
<td>$(1.2 \times 10^9)^c$</td>
<td>$(5 \times 10^9)^c$</td>
<td>$(2.5 \times 10^7)^c$</td>
<td>2 $\times 10^9$</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$k_{TST}^{B \rightarrow A}$ (s$^{-1}$)</td>
<td>$(6 \times 10^8)^c$</td>
<td>$(1 \times 10^9)^c$</td>
<td>$(6 \times 10^8)^c$</td>
<td>$(1 \times 10^9)^c$</td>
<td>3 $\times 10^7$</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Wild-type: mismatched system (G:A)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>6</td>
<td>14.5</td>
<td>6</td>
<td>1</td>
<td></td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>$\beta \Delta F_{AB}^{\text{barrier}}$</td>
<td>6</td>
<td>2.5</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$k_{TST}^{A \rightarrow B}$ (s$^{-1}$)</td>
<td>$(6 \times 10^8)^c$</td>
<td>$(6 \times 10^9)^c$</td>
<td>$6 \times 10^8$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$k_{TST}^{B \rightarrow A}$ (s$^{-1}$)</td>
<td>$(6 \times 10^8)^c$</td>
<td>$(1 \times 10^9)^c$</td>
<td>$2 \times 10^7$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

$^a$The total energy barrier for conformational pathway before chemistry from open to closed state.

$^b$Equation is the free energy of the transition state region between basin A and B relative to basin A. Detailed free energy plots for each transition are shown in Supplemental Figure 3.9.5.

$^c$To make the energy values comparable, the orders of events in the table for systems other than E295K have been rearranged by values in parentheses. See Supplemental Table 3.9.2 for the sequence of transition states in each system.
We superimpose the final structures of both wild-type and mutant systems in Figure 3.6. The Lys295/Glu295 is buried between two tyrosines (Tyr296 and Tyr271), forming a sandwich-shaped structure. This “sandwich-shaped” structure is quite stable during our simulations, and it is also notable in the crystal structure of the E295K mutant (Supplementary Figure 3.9.3). Therefore, though Lys295 appears to be flexible during our simulation, the position, which Lys295 may adopt, is constrained in a limited space between those two tyrosines. Due to the longer length of the side chain of Lys295 compared to Glu295, the former is closer to Arg258 in the closed form and causes steric hindrance. Even with a hydrogen bond acceptor for Arg258, the hydrogen bond would be hindered, sterically, by Lys295. As a result, the rotation of Arg258 is difficult to complete in the E295K mutant system due to both the electrostatic repulsion and steric clash. This finding also agrees with the experimental observation that the E295A mutant can still complete its catalytic pathway though it has reduced activity (See Supplementary Table 3.9.1 and Supplementary Figure 3.9.6) (Kraynov 2000). Because alanine is neutral and less bulky than lysine, both electrostatic repulsion and steric clash in the E295A system would be reduced, and therefore the rotation of Arg258 would be more facile to accomplish compared to that in the E295K system.
Figure 3.4. Representative active site conformation of (A) E295K mutant and (B) wild-type pol β in the final closed form. Dashed lines indicate hydrogen bonds. Mg(A), catalytic ion; Mg(B), nucleotide-binding ion; C10, the base in the upstream primer DNA strand which will connect with dCTP during the chemical step.
Figure 3.5. Electrostatic potential landscapes of residues around Arg258 in (A) E295K mutant open form, (B) wild-type pol β open form, (C) E295K mutant closed form, and (D) wild-type pol β closed form.
Figure 3.6. “Sandwich-shaped” structure formed by Glu295/Lys295, Tyr296 and Tyr271 tends to cause steric hindrance for Arg258 in the E295K mutant complex. The residue Lys295 (red) is superimposed from the final closed form of the E295K mutant. All other residues are superimposed from the ternary crystal structure of wild-type pol β. Mg(A), catalytic ion; Mg(B), nucleotide-binding ion.
3.6. Discussion

We have described the conformational closing profiles of the pol β E295K mutant by TPS simulations. Important differences in the conformational profiles between the mutant and wild-type systems are observed concerning stability of the closed versus open state, the various transition states and their order, overall energy barrier, and electrostatic and steric factors.

First, the active site conformation in the final closed state of E295K is distorted (Table 3.3). As a result, the closed form of the mutant is relatively less stable than its open form by ~5 kBT. The active site of the mutant E295K deviates from that of the wild-type G:C with regard to O3’ - Pα, catalytic Mg^{2+} - O3’, and nucleotide-binding Mg^{2+} - O1α distances. Notably, the distance between the nucleotide-binding Mg^{2+} and O1α in the mutant system is significantly longer than that distance in all four wild-type systems (G:C, G:A, 8-oxoG:C, and 8-oxoG:A), representing a distorted active site. Following the transition, the crucial metal ion distances in the mutant system are far from the ideal values required for the nucleotidyl-transfer reaction; thus, the mutant system must undergo more significant rearrangements prior to the chemical reaction to assemble the metal ions in the proper positions, and these likely involve additional energy barriers (Arora 2004; Arora 2005a; Arora 2005b). In comparison to the G:A mismatched and 8-oxoG pol β closed state, E295K is overall less distorted than both G:A (in terms of the shift of Arg283, Mg^{2+} ions, and O3’ - Pα distance (Arora 2005a; Radhakrishnan 2005), also reflected in the lower energy of the final closed form compared to that of G:A.
mismatched system: ~5 k_BT versus ~9 k_BT) and 8-oxoG systems, which may have additional stabilizing features. However, the mutant must undergo larger barrier to reach that state (see below).

Second, the flip of Asp192 occurs very early in the E295K mutant system. However, after the flip of Asp192, the free energy of the system increases by 3 k_BT, while in the wild-type G:C, G:A, and 8-oxoG:C systems the free energy decreases by 0-2 k_BT (in the 8-oxoG:A system the flip of Asp192 is not a step in the pathway). In the open form, Arg258 forms a salt-bridge with Asp192, which attracts Asp192 and stabilizes the position of Asp192 before it flips to Mg^{2+}. In the E295K mutant, the positively charged Lys residue replaces the negatively charged Glu. Therefore, the electrostatic environment along the side of Arg258 becomes more positive, attracting Asp192 more. As a result, Asp192 tends more to revert back (away from Mg^{2+}) in the mutant system, and thus the system following the Asp192 flip becomes less stable. In other words, unlike the wild-type systems, the flip of Asp192 in the E295K mutant does not make the remaining energy pathway “downhill”.
Table 3.3. The average critical distances for the chemical reaction following the final transitions of pol β DNA complexes in different systems.

<table>
<thead>
<tr>
<th>Critical distance</th>
<th>E295K (G:C)</th>
<th>Wild-type (G:C)</th>
<th>Wild-type (G:A)</th>
<th>Wild-type (8-oxoG:C)</th>
<th>Wild-type (8-oxoG:A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg(A)-O1α</td>
<td>1.8</td>
<td>1.8</td>
<td>3.8</td>
<td>6.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Mg(A)-O3’</td>
<td>5.6</td>
<td>4.2</td>
<td>5.9</td>
<td>2.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Mg(A)-OD1-Asp192</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Mg(A)-OD2-Asp256</td>
<td>1.7</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Mg(B)-O1α</td>
<td>2.7</td>
<td>1.9</td>
<td>1.8</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>O3’-Pα</td>
<td>4.8</td>
<td>4.1</td>
<td>5.3</td>
<td>6.7</td>
<td>5.5</td>
</tr>
</tbody>
</table>

*Mg(A) represents the catalytic Mg\(^{2+}\); Mg(B) represents the nucleotide-binding Mg\(^{2+}\); O1α represents O1α on the dNTP; O3’ represents O3’ on the CYT10 in the upstream primer DNA strand, which will be connected to the dNTP in the chemical step; Pα represents Pα on the dNTP.*
Third, the overall energy barrier for the partial closing of fingers – roughly the first half of the conformation pathway – is 13 kBT in the mutant system, slightly less than that of 14 kBT in the wild-type matched system and 14.5 kBT in the wild-type mismatched system. Furthermore, the partial-closed form before key residue motions besides Asp192 is also slightly more stable compared to those of the wild-type systems (energy level of 8 kBT and 9-12 kBT, respectively). Therefore, the E295K mutant may compete with the wild-type in the sense that it has better affinity to bind with DNA and transits to the partially-closed form. However, because of the high-energy barrier in the following steps, the partially-closed form of the E295K mutant is less likely to continue its conformational pathway and thus may become locked in the partially-closed form. This finding agrees with experimental observations: not only is the E295K mutant itself inactive on gapped DNA, but it also interferes with the ability of wild-type pol β to catalyze DNA synthesis on single nucleotide gaps (Lang 2007). E295K binds to single-nucleotide gapped DNA with an affinity that is similar to that of wild-type pol β, but could bind to the 3′OH of the gap and block access to wild-type pol β. Alternatively, the locking of E295K into a partially closed form may stabilize its association with DNA and block wild-type pol β from binding to DNA. Importantly, E295K has dRP lyase activity that is similar to that of wild-type pol β, suggesting that stabilization of the closed form of pol β is not necessary for this activity. Thus, the lyase reaction could involve a more open or completely open conformation of pol β and occur before closure or after the opening of the protein, following chemistry.
Last, but not least, the main difference in the conformation pathway of the mutant compared to wild-type systems lies in Arg258. We have shown that both the electrostatic effect and the steric hindrance affect the rotation of Arg258; yet, which effect is more significant still remains unclear. We suggest that mutations such as E295Y and E295W may help delineate this question. If steric hindrance is more important, mutations like tyrosine and tryptophan with neutral but bulky side chains should also inactivate the polymerase; if the electrostatic effect is more important, such mutations may deteriorate the catalytic ability of pol β but not inactivate it completely.

Kinetic data show that the overall catalytic process (both conformational and chemical changes) for the matched wild-type pol β system has an energy barrier of 26-28 k_BT (calculated from experimental data of the intrinsic rate constant of polymerization k_{pol}) (Beard 1996; Werneburg 1996; Ahn 1997; Ahn 1998; Kosa 1999b; Kraynov 2000; Shah 2001b; Vande Berg 2001; Dalal 2004; Dalal 2005; Murphy 2008; Yamtich 2010; Murphy 2011), which is similar to the conformational barrier for the mutant we obtain here, and chemistry is believed to be rate-limiting overall (Supplementary Figure 3.9.4). In the E295K mutant, the conformational pathway, especially the rotation of Arg258, already hampers significantly the enzyme’s ability to reach a chemical-reaction competent state. As observed in other DNA polymerases, an altered enzyme complex may also have a different rate-limiting step (chemical or conformational) from the wild-type system. For example, for Sulfolobus solfataricus P2 DNA polymerase IV, the rate-limiting step is chemical step for the matched (correct base) system, but the conformational step is rate-
limiting for the mismatched system (incorrect base) (Fiala 2004). For pol I from bacteria Bacillus stearothermophilus, the energy barrier for the chemical step is higher than that for the conformational changes (25.6 k_BT versus 24.3 k_BT) (Eger 1992) with a normal G:C base-pair; with a 8-oxoG:C base-pair, the energy barrier for the chemical step drops to 11.5 k_BT (Venkatramani 2008), which means that conformational changes may be rate-limiting for this 8-oxoG system. For the E295K pol β mutant, in vitro work indicates no incorporation of dNTP into gapped DNA even in excess amounts of substrate (J. Sweasy, unpublished work); thus, this mutant may not be able to go through the chemical step on this substrate. Additional studies are needed for a final determination.

3.7. Conclusion

Our binary E295K complex, together with computational investigation of the E295K pol β mutant using the transition path sampling simulations, reveals the conformational transition pathway before chemistry for this mutant. Compared to previous work on several wild-type systems (Radhakrishnan 2004b, 2005; Wang 2007), the conformational pathway of the mutant is significantly different regarding both the sequence of events and individual energy values. Six transition regions are identified, and these provide a combined energy barrier of 26 ± 3 k_BT. The final closed form of the mutant is less stable than the starting open form by ~5 k_BT, and is characterized by distorted active site conformation. The rate-limiting step in the conformational pathway is the rotation of Arg258. The rotation is hampered due to both steric and electrostatic effects.
Furthermore, the mutant’s half-closed form, which is after the partial fingers subdomain closing but before the key residue motions of Arg258, Phe272, and Tyr271, of the E295K mutant is slightly more stable than that of the wild-type pol β. This suggests the possibility that the mutant may bind with DNA and undergo fingers closing more easily than the wild-type, but due to the high energy barriers in the following steps it is difficult for the mutant to continue on its pathway. Therefore, when both the mutant and the wild-type pol β present, the E295K mutant competes with wild-type pol β, and the activity of the wild-type deteriorates.

3.8. Acknowledgements

We thank Dr. Ravi Radhakrishnan for providing the initial scripts for transition path sampling simulations. Research described in this article was supported in part by Philip Morris USA Inc. and Philip Morris International and by NSF award MCB-0316771, NIH award R01 ES012692, and the American Chemical Society’s Petroleum Research Fund award (PRF #39115-AC4) to T. Schlick, and NCI award 2R01CA08030 to J. Jaeger and J. Sweasy. Access to beam line NSLS-X4C at the New York Structural Biology Center and the core facilities at the Wadsworth Center is gratefully acknowledged. The computations were conducted using the resources of the CCNI supported by the New York State Foundation for Science, Technology and Innovation (NYSTAR), and the Dell computer cluster by New York University Information Technology Services (NYU ITS). Molecular images were generated using the VMD70 and PyMol68 programs.
3.9. Supplemental information

3.9.1. Shooting algorithm and test of convergence of TPS

The shooting algorithm (Bolhuis 1998; Bolhuis 2002; Dellago 2003) generates an ensemble of new trajectories by perturbing initial momenta of atoms in a randomly chosen time interval while ensuring conservation of Maxwellian distribution of velocities, total linear and angular momentum and detailed balance. The perturbation scheme employed in our work is also symmetric – the probability of generating a new set of momenta from the old set is the same as the reverse probability of generating the old set from the new set.

In particular, the ensemble of new trajectories \{x_\tau\} of length \( \tau \) are generated by a Metropolis algorithm according to a path action \( S(\chi) : S(\chi) = q(0) h_A(\chi_0) H_b(\chi_\tau) \), where \( q(0) \) is the probability of observing the configuration at \( t = 0 \) (\( q(0) \propto \exp(-\beta E(0)) \)) in the canonical ensemble. The newly generated trajectories are accepted or rejected based on selected statistical criteria that characterize the ensemble of trajectories (Bolhuis 1998; Bolhuis 2002; Dellago 2003; Radhakrishnan 2004a).

The ergodicity and convergence of each TPS run is confirmed by calculating the autocorrelation function of the order parameter \(<x_i(0)x_i(\tau)>\) associated with each transition state \( i \), where \(<\cdot>\) denotes the average over the ensemble of generated
trajectories. For each transition state, the autocorrelation function is plotted from time 0 where \( <\chi_i(0)\chi_i(0)> \approx \chi_A^2 \) to the time \( \tau \) where \( <\chi_i(0)\chi_i(\tau)> \approx \chi_A \chi_B \); this range is spanned during our sampling time \( \tau \) (see Supplemental Figure 3.9.2), indicating that the transition state regions between A and B are crossed during this interval. The time used for the gradual transition of the autocorrelation function \( <\chi_i(0)\chi_i(\tau)> \) from these plots can provide an estimate for the timescale of barrier crossing (\( \tau_{mol} \)) (Chandler 1978).

Thus, the length of the MD trajectories should be longer than the \( \tau_{mol} \) value to sufficiently cover the entire transition region.

The above procedure both conserves the equilibrium distribution of individual metastable states and ensures that the accepted MD trajectories connect the two metastable states for a particular transition. The shooting algorithm used in our work based on the Metropolis scheme also conserves microscopic reversibility. Hence, the ensemble of MD trajectories generated is guaranteed to converge to the correct ensemble defined by the path action and represents configurations that constitute the correct pathway for hopping between the metastable states.
Supplemental Table 3.9.1. Kinetic data for wild-type pol β and mutants with correct base-pairing. See also Supplemental Figures 3.9.6 and 3.9.7.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>$k_{pol}$ (s$^{-1}$)</th>
<th>$K_d$ (μM)</th>
<th>ΔG (k_BT)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>3 – 54</td>
<td>2.5 – 63</td>
<td>25.6 – 28.5</td>
</tr>
<tr>
<td>L22P (Dalal 2008a)</td>
<td>(No activity)</td>
<td>291 ± 45</td>
<td>N/A</td>
</tr>
<tr>
<td>R149A (Kraynov 2000)</td>
<td>11</td>
<td>12</td>
<td>29.6</td>
</tr>
<tr>
<td>I174S (Yamtich 2010)</td>
<td>6.7 ± 0.7</td>
<td>23 ± 6</td>
<td>27.7 ± 0.11</td>
</tr>
<tr>
<td>S180A$^b$ (Kraynov 2000)</td>
<td>1.0</td>
<td>70</td>
<td>27.1</td>
</tr>
<tr>
<td>R182E (Murphy 2011)</td>
<td>0.034 ± 0.002</td>
<td>131 ± 19</td>
<td>33.0 ± 0.06</td>
</tr>
<tr>
<td>R183A (Kraynov 2000)</td>
<td>2.6</td>
<td>5.9</td>
<td>28.6</td>
</tr>
<tr>
<td>S188A$^b$ (Kraynov 2000)</td>
<td>8.9</td>
<td>3.8</td>
<td>27.4</td>
</tr>
<tr>
<td>D246V (Dalal 2004)</td>
<td>31.8 ± 2.6</td>
<td>29.1 ± 6.2</td>
<td>26.1 ± 0.08</td>
</tr>
<tr>
<td>E249K (Kosa 1999b)</td>
<td>9.1 ± 0.5</td>
<td>25 ± 4</td>
<td>27.4 ± 0.06</td>
</tr>
<tr>
<td>I260M (Dalal 2005)</td>
<td>2.0 ± 0.1</td>
<td>6 ± 1</td>
<td>28.9 ± 0.05</td>
</tr>
<tr>
<td>I260Q$^b$ (Dalal 2008b)</td>
<td>10.8 ± 1.5</td>
<td>165 ± 40</td>
<td>27.2 ± 0.14</td>
</tr>
<tr>
<td>Y265F$^b$ (Shah 2003)</td>
<td>18.2 ± 0.9</td>
<td>63 ± 11</td>
<td>26.7 ± 0.05</td>
</tr>
<tr>
<td>Y265H$^b$ (Shah 2001b)</td>
<td>0.087 ± 0.003</td>
<td>1.2 ± 0.1</td>
<td>32.1 ± 0.03</td>
</tr>
<tr>
<td>Y265W$^b$ (Shah 2003)</td>
<td>19.4 ± 0.3</td>
<td>14 ± 1</td>
<td>26.6 ± 0.02</td>
</tr>
<tr>
<td>Y271A (Kraynov 1997)</td>
<td>0.58 ± 0.03</td>
<td>1.4 ± 0.2</td>
<td>30.1 ± 0.05</td>
</tr>
<tr>
<td>Y271F (Kraynov 1997)</td>
<td>3.30 ± 0.35</td>
<td>3.7 ± 1.2</td>
<td>28.4 ± 0.11</td>
</tr>
<tr>
<td>Y271S (Kraynov 1997)</td>
<td>1.0 ± 0.1</td>
<td>4.7 ± 0.5</td>
<td>29.6 ± 0.10</td>
</tr>
<tr>
<td>F272L$^b$ (Li 1999)</td>
<td>30 ± 1</td>
<td>77 ± 10</td>
<td>26.2 ± 0.03</td>
</tr>
<tr>
<td>D276R$^b$ (Liu 2001)</td>
<td>8.6 ± 0.87</td>
<td>170 ± 30</td>
<td>27.4 ± 0.10</td>
</tr>
<tr>
<td>D276V (Vande Berg 2001)</td>
<td>6.3 ± 0.9</td>
<td>0.6 ± 0.3</td>
<td>27.8 ± 0.14</td>
</tr>
<tr>
<td>N279A$^b$ (Kraynov 1997)</td>
<td>44 ± 10</td>
<td>1400 ± 600</td>
<td>25.8 ± 0.23</td>
</tr>
<tr>
<td>N279Q$^b$ (Kraynov 1997)</td>
<td>14 ± 2</td>
<td>610 ± 120</td>
<td>27.0 ± 0.14</td>
</tr>
<tr>
<td>K280A (Kraynov 2000)</td>
<td>12</td>
<td>6</td>
<td>27.1</td>
</tr>
<tr>
<td>K280G$^b$ (Beard 2002)</td>
<td>2.7 ± 0.1</td>
<td>289 ± 26</td>
<td>28.6 ± 0.04</td>
</tr>
<tr>
<td>M282L$^b$ (Shah 2001a)</td>
<td>39.8 ± 4.6</td>
<td>92 ± 28</td>
<td>25.9 ± 0.12</td>
</tr>
<tr>
<td>R283A (Ahn 1997)</td>
<td>0.83 ± 0.08</td>
<td>61 ± 17</td>
<td>29.8 ± 0.10</td>
</tr>
<tr>
<td>R283K$^b$ (Werneburg 1996)</td>
<td>0.05 ± 0.01</td>
<td>170 ± 60</td>
<td>32.6 ± 0.20</td>
</tr>
<tr>
<td>H285D (Murphy 2008)</td>
<td>2.5 ± 0.2</td>
<td>4.4 ± 0.9</td>
<td>28.7 ± 0.08</td>
</tr>
<tr>
<td>N294A (Kraynov 2000)</td>
<td>4.0</td>
<td>6.6</td>
<td>28.2</td>
</tr>
<tr>
<td>Mutant</td>
<td>$k_{pol}$ (s$^{-1}$)</td>
<td>$K_d$ (μM)</td>
<td>$\Delta G$ (k_BT)$^a$</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------</td>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>N294Q (Kraynov 2000)</td>
<td>2.6</td>
<td>1.6</td>
<td>28.6</td>
</tr>
<tr>
<td>E295A (Kraynov 2000)</td>
<td>2.0</td>
<td>10</td>
<td>28.9</td>
</tr>
<tr>
<td>E295K (Lang 2007)</td>
<td>(No activity)</td>
<td>28</td>
<td>N/A</td>
</tr>
<tr>
<td>E316R (Murphy 2011)</td>
<td>0.00185 ± 0.00006</td>
<td>20 ± 4</td>
<td>35.9 ± 0.03</td>
</tr>
<tr>
<td>R333E (Murphy 2011)</td>
<td>0.074 ± 0.003</td>
<td>70 ± 9</td>
<td>32.2 ± 0.04</td>
</tr>
</tbody>
</table>

$^a$k$_{pol}$ is the intrinsic rate constant of polymerization, $K_d$ is the equilibrium dissociation constant for the incoming nucleotide. $\Delta G$ is the overall energy barrier in pol $\beta$’s catalytic pathway and is calculated as $\Delta G = RT[\ln(k_BT/h)-\ln(k_{pol})]$ (Eger 1992), where R is the universal gas constant and h is the Planck constant.

$^b$S180A, S188A, I260Q, Y265H, F272L, N279A, N279Q and K280G are measured with a base-pairing of A:dTTP; Y265F, Y265W, D276R, M282L and R283K are measured with a base-pairing of T:dATP; all other mutants and the wild-type pol $\beta$ are measured with a base-pairing of G:dCTP. See Supplementary Figure 3.9.6 for a summary of pol $\beta$’s mutants, and Supplementary Figure 3.9.7 for the locations of the residues.
**Supplemental Table 3.9.2. Sequence of transition states for the closing conformational profiles of five pol β systems identified by TPS**

<table>
<thead>
<tr>
<th></th>
<th>E295K mutant (G:C)</th>
<th>Wild-type (G:A)</th>
<th>Wild-type (8-oxoG:C)</th>
<th>Wild-type (8-oxoG:A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1</td>
<td>Flip of Asp192</td>
<td>Fingers closing</td>
<td>Fingers closing</td>
<td>Fingers closing</td>
</tr>
<tr>
<td>TS2</td>
<td>Fingers closing</td>
<td>Flip of Asp192</td>
<td>Flip of Asp192 Arg258 half-rotation</td>
<td>Flip of Asp192</td>
</tr>
<tr>
<td>TS3</td>
<td>Flip of Phe272 Arg258 rotation</td>
<td>Flip of Phe272 Arg258 full-rotation</td>
<td>Arg258 full-rotation</td>
<td>Arg258 full-rotation</td>
</tr>
<tr>
<td>TS4</td>
<td>Arg258 rotation</td>
<td>Flip of Phe272 Arg258 rotation</td>
<td>Flip of Phe272</td>
<td>Flip of Phe272</td>
</tr>
<tr>
<td>TS5</td>
<td>Shift of Tyr271 Ion motions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS6</td>
<td>Ion motions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Figure 3.9.1. Initial constrained trajectories for the conformational transitions obtained from TMD
Supplemental Figure 3.9.2. Correlation functions of order parameters for transition states in the closing pathway
Supplemental Figure 3.9.3. Electron density map of the pol β E295K binary complex contoured at 1.25 RMSD above the mean

The final model is superimposed in stick representation. Note the lack of electron density for Lys295.
Supplemental Figure 3.9.4. Superimposed free energy profiles of five pol β systems obtained from TPS (Radhakrishnan 2004b, 2005; Wang 2007)

The energy barrier of the chemical step for wild-type G:C and G:A are also shown in dashed lines, which are derived from experimentally measured $k_{pol}$ values (Kraynov 1997; Ahn 1998; Kraynov 2000; Shah 2001b; Vande Berg 2001). See Supplementary Table 3.9.1 for the sequence of transition states in each system.
Supplemental Figure 3.9.5. Potential of mean force plots computed by BOLAS
Supplemental Figure 3.9.6. Summary of activity of pol β mutants

Green, active ($k_{pol} > 4 \text{ s}^{-1}$); blue, activity reduced ($k_{pol}$ between $4 \text{ s}^{-1}$ and $1 \text{ s}^{-1}$); red, lost activity ($k_{pol} < 1 \text{ s}^{-1}$ or cannot be measured). See also Supplemental Figure 3.9.7.
Supplemental Figure 3.9.7. Locations of residues listed in Supplemental Table 3.9.1 grouped by their effect on the enzyme’s activity when mutated

If on the same residue, different mutations change pol β’s activity differently, that residue is grouped according to the lowest activity of all the mutants on it. Green, blue and red as in Supplemental Figure 3.9.6. Black, dCTP and C10; orange, Mg²⁺.
Chapter 4. Mutator phenotypes of pol β variants on dsDNA: Mg$^{2+}$ versus Mn$^{2+}$

4.1. Introduction

To support faithful nucleotidyl incorporation during gap-filling synthesis, pol β must incorporate the correct nucleotide. In earlier sections, we discussed the importance of pol β in maintaining genomic integrity and highlighted the consequences of certain single amino acid mutations on the activity of the polymerase (Chapters 2 & 3). Several single amino acid variants show a mutator phenotype during gap-filling synthesis and during primer extension of double stranded DNA substrates. Thus far, we have introduced the various mutator phenotypes of a few pol β variants: Y265H, P242R, D246V, I260Q, Y265H, M282L and E295K. In this chapter, we explore the phenotypes of I260Q, E295K and an additional mutator variant, H285D.

The I260Q variant of pol β has a very well characterized mutator phenotype on a variety of DNA substrates. In a Trp+ reversion assay, in vivo, I260Q pol β generates reversions with a frequency 60-fold greater than wild-type (Starcevic 2005b). The mutation at 260 causes an increase in misincorporation events during gap filling synthesis in vitro (Starcevic 2005b). While the template base has a slight effect on the fidelity of I260Q pol β compared to wild-type, the mutant exhibits an overall decrease in fidelity on gapped DNA substrates. The residue is located in the hydrophobic hinge linking the fingers and
palm subdomains (Figure 1.4). In Chapter 2 we characterized two structures of I260Q: apoenzyme and in complex with dsDNA and ddTTP.

At 2.4 and 2.2 Å resolution, for wild-type and I260Q pol β, respectively, we compared the two polymerases in terms of overall conformation and also at the site of mutation. While the overall conformations were very similar, there were significant alterations in the vicinity of residue 260: Arg258 to Asp192 distance is decreased in the mutant compared to the wild-type, 2.7 and 3.5 Å, respectively, and several well-defined water molecules were coordinated in the hydrophobic hinge of the mutant, yet not in the wild-type polymerase (Gridley 2012, Chapter 2). For the I260Q variant, some or all of these water molecules may need to be expelled prior to closure of the fingers subdomain, which may create an energetically unfavorable situation.

The ternary complex of I260Q, at 2.7 Å, was compared to an isomorphous wild-type complex, PDB ID: 1HUO (Arndt 2001). Similar to the apoenzyme structures, the ternary complexes of wild-type and I260Q pol β were very similar and the significant differences were centered on the site of mutation. Significant differences include: the formation of a hydrogen bond between the amide side chain of Gln260 and the backbone of Glu295 – an interaction not observed in wild-type; altered positioning of the aromatic ring of Phe272 – tilted (χ²=13°) towards Asp192 in I260Q pol β and tilted (χ²=109°) away from Asp192 in wild-type pol β; and Tyr271 is shifted 1.3 Å away from N3 of the primer base in I260Q compared to wild-type (Gridley 2012, Chapter 2). The consequence of the
mutation at position 260 to Gln is sub-optimal positioning of Tyr271 for binding the primer terminus and an increase in the dNTP “pocket” to allow for mismatch incorporation and mispair extension.

In terms of a mutator phenotype for I260Q pol β we have shown that this variant can perform primer extension at levels comparable to wild-type on our dsDNA substrate, and additionally, this variant can both misincorporate and extend mispairs on the same substrate (Gridley 2012, Chapter 2). This is supported by misincorporation kinetics comparing wild-type versus I260Q: wild-type pol β is quick to release a mismatched nucleotide ($K_d$, T:G = 489 ± 26 μM; C:A = 227 ± 22 μM), while I260Q pol β appears to stabilize the same mismatch ($K_d$, T:G = 49 ± 3 μM; C:A = 45 ± 3 μM) (Roettger 2008). We expand on these characterizations of the mutator phenotype of I260Q in the present study.

The second mutator variant of pol β we examined in Chapter 3 was E295K. Due to its role in BER, pol β variants found in human cancers have been isolated and characterized for potential mutator properties. Iwanaga and colleagues collected tissue samples from gastric cancer patients, extracted DNA and RNA, isolated the coding sequence of pol β and ligated these PCR products into a plasmid vector for transfection into *E. coli* cells (Iwanaga 1999). The transformants carrying pol β mutants were then expressed in a HeLa cell line for characterization.
Out of 20 gastric cancer tissue samples, six contained missense point mutations in mRNA and genomic DNA, yet normal pol β was found in all tumor samples, indicating the cancer cells are heterozygous for pol β (Iwanaga 1999). When the 20 samples were tested for BER activity, all but one had activity levels comparable to wild-type; the E295K variant did not show any detectable levels of BER activity in this assay. Interestingly, in the cells expressing the E295K variant of pol β and wild-type, the presence of E295K inhibits the BER activity of normal pol β (Iwanaga 1999).

In the closed, ternary complex of pol β, the E295 residue is sandwiched between Y271 and Y296 and interacts with R258. In the absence of substrate, R258 interacts with a catalytic aspartate, D192 in the form of a weak salt bridge. Several events must occur to create the proper active site geometry and dNTP alignment for the in-line nucleophilic attack during catalysis to occur. As the fingers subdomain closes upon substrate binding, D192 is recruited for metal-ion binding and F272 rotates in between R258 and the active site, triggering the formation of a hydrogen bond between R258 and E295 (see Chapter 1.4 for details).

A mutation from Glu to Lys reverses the charge of this particular side chain, which would likely abolish the interaction between, now both positively charged residues 295 and 258 (in the closed conformation). Lang et al. performed a simple primer extension assay on a single nucleotide-gapped DNA duplex, comparing wild-type to E295K, and determined this variant of pol β is inactive. The inactivity of E295K could not be
attributed to DNA binding as it was found to bind to the gapped DNA duplex with similar affinity as wild-type (Lang 2007). In support of Iwanaga et al.’s results, Lang and colleagues found that the presence of E295K pol β interferes with wild-type’s ability to perform DNA gap-filling synthesis in the in vitro primer extension assay.

The H285D variant was identified after selectively pressuring pol β to develop resistance against the nucleoside analog drug 3’-azido-3’deoxythymidine (AZT) (Kosa 1999a). When incorporated into DNA during polymerization, the azido group in the 3’ position causes the polymerase to stall and frequently results in chain termination (Kosa 1999a). The mutation at position 285, a residue on α-helix N in the fingers subdomain, could affect the DNA-pol β contacts mediated through α-helix N, and in turn, alter polymerase fidelity. In terms of affinity for 5-base pair gapped DNA, H285D and wild-type have similar $K_{d(DNA)}$ values, 6.7 ± 0.8 and 9.2 ± 0.7 nM, respectively (Murphy 2008).

Furthermore, while subtly different in catalytic efficiency, the H285D variant can extend mispairs for all template bases in the absence of dCTP, more so than wild-type. This feature is attributed to the finding that H285D binds the incoming nucleotide more tightly than wild-type: dATP binding and catalytic efficiency values vary significantly between the two polymerases: wild-type $K_{d(dATP)}$, 183 ± 5 μM and $k_{pol}/K_{d}$, 1.63x10⁴ M⁻¹ s⁻¹; H285D $K_{d(dATP)}$, 4.4 ± 0.9 μM and $k_{pol}/K_{d}$, 5.7x10⁷ M⁻¹ s⁻¹ (Murphy 2008). Interestingly, H285D is not an active misincorporator mutant on a 5-base pair gap substrate, but can extend mispairs on this substrate.
4.2. Materials and methods

4.2.1. In vitro primer extension assays

To generate a baseline for primer extension capabilities, we performed the in vitro primer extension assay described in detail in Chapter 2 with slight modifications (see Materials and Methods in Chapter 2). The dsDNA substrate used: T: 5’-TAC GTC GCG ACT GCT CAC AT-3’ and P:5’-/56-FAM/ATG TGA G-3’, was annealed at 45 nM (Integrated DNA Technology). The enzyme:DNA ratio was increased to 15:1 (675 nM:45 nM). The enzyme and DNA were mixed with all four dNTPs (50 μM) and MgCl₂ (10 mM) in the reaction buffer (20 mM Tris, pH 8.0, 60 mM NaCl, 10% glycerol, 1 mM dithiothreitol, final concentrations) and incubated at 37°C. After ten minutes, the reactions were quenched with an equal volume of TBE-Urea Sample buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0, 12% Ficoll, 0.01% bromophenol blue, 0.02% xylene cyanole FF and 7 M urea – Bio-Rad Laboratories, Inc.) and samples were heated to 95°C for five minutes before being separated on a 15% TBE-Urea denaturing PAGE gel. Gels were then visualized on a Typhoon Imager (GE Healthcare Life Sciences) for the detection of fluorescent bands indicating extended versus unextended primer.

4.2.2. Misincorporation and mispair extension assays: Mg²⁺ versus Mn²⁺

To compare the mutator phenotypes of the three pol β variants, I260Q, H285D and E295K compared to wild-type, we modified the primer extension assay described above
(and in Chapter 2) to determine the misincorporation and mispair extension activity of each variant. Using the same 15:1 enzyme to DNA ratio, we omitted either the first or the second nucleotide for incorporation from the dNTP mixture used to initiate the reaction. Accordingly, either dCTP or dATP were omitted from the reaction. The enzyme and DNA were mixed with dNTPs (50 μM final concentration) and MgCl₂ (10 mM final concentration) or MnCl₂ (1 mM final concentration) and incubated at 37°C for either 10 or 30 minutes, after which the reactions were stopped with the TBE-Urea sample buffer detailed above and then viewed on the Typhoon Imager (GE Healthcare Life Sciences) for fluorescence detection.

4.2.3. Determining the misincorporation threshold: MnCl₂ supplementing MgCl₂

Following the same protocol as described in the previous two sections for primer extension, misincorporation and mispair extension, we performed preliminary experiments to determine the minimal threshold of Mn²⁺ concentration that allows for misincorporation events by wild-type pol β. The misincorporation event examined here is the second event during primer extension by omitting dATP from the dNTP mixture. The enzyme:DNA ratio tested was 10:1, 450 nM enzyme to 45 nM DNA. The enzyme and DNA were mixed with 10 mM MgCl₂ and 50 μM dNTPs (final concentrations) and the concentration of MnCl₂ was varied: 450 nM, 1 μM, 10 μM, 100 μM, 250 μM, 500 μM, 750 μM, 1 mM and 5 mM. The reactions were incubated at 37°C for 10 minutes, then quenched and viewed on the Typhoon Imager (GE Healthcare Life Sciences) for
fluorescence detection. The assay was then repeated with I260Q and H285D at similar concentrations of MnCl₂.

4.3. Results

4.3.1. The mutator phenotype of pol β variant, I260Q, is enhanced in the presence of Mn²⁺

In Chapter 2, we demonstrated the activity of I260Q on our dsDNA substrates used for crystallization. Here, we provide additional evidence of the mutator properties of I260Q compared to wild-type. In these experiments, we expanded the omission assay such that Mg²⁺ was replaced by or supplemented with Mn²⁺. Mn²⁺ is known to permit the coordination of mis-matched nucleotides in the active site (Tabor 1989; Werneburg 1996; Liu 2001). We hypothesized that even though wild-type pol β can utilize Mn²⁺ during gap-filling synthesis, at least in vitro, the wild-type polymerase will still be able to discriminate correct versus incorrect nucleotides, to some extent, while the mutator polymerases will be more likely to incorporate and extend beyond mispaired nucleotides in the presence of Mn²⁺.

The misincorporation assays were performed on the crystallization dsDNA and the first or second nucleotide for incorporation was omitted from the reaction (Figure 4.1A). Figure 4.1B shows the primer extension activity of three pol β variants, I260Q, H285D and E295K, compared to wild-type pol β. As expected, wild-type and I260Q pol β show
similar levels of primer extension with all four dNTPs and Mg\textsuperscript{2+} (Figure 4.1B). The misincorporation assays, with Mg\textsuperscript{2+} (Figure 4.1C) or Mn\textsuperscript{2+} (Figure 4.1D) were performed with the same three pol β variants, again using the wild-type enzyme as reference. Both I260Q and wild-type polymerases are unable to misincorporate the first nucleotide in the presence of Mg\textsuperscript{2+} after 10 minutes (Figure 4.1C). After 30 minutes, very minimal amounts of extended primer are observable in the case of I260Q but not for wild-type. However, substituting Mn\textsuperscript{2+} for Mg\textsuperscript{2+} facilitated the misincorporation by both polymerases opposite template G (-dCTP), and we observe more mispair extension for I260Q than wild-type at both time points, 10 and 30 minutes (Figure 4.1D).

When Mg\textsuperscript{2+} is available and the second incoming nucleotide is absent (dATP), wild-type and I260Q incorporate the first nucleotide well. Both polymerases appear to stall during misincorporation opposite T with Mg\textsuperscript{2+} (Figure 4.1C), yet Mn\textsuperscript{2+} allows near complete extension of the mispair for wild-type and I260Q pol β (Figure 4.1D). To determine if there is a threshold Mn\textsuperscript{2+} concentration at which wild-type exhibits decreased discrimination, we supplemented Mg\textsuperscript{2+} with increasing concentrations of Mn\textsuperscript{2+}. Figure 4.2A shows the misincorporation opposite template T (-dATP) by wild-type is detected between 10 and 100 μM MnCl\textsubscript{2}. A similar range of MnCl\textsubscript{2} concentrations, supplementing MgCl\textsubscript{2}, were used to compare the level of misincorporation by I260Q and H285D to wild-type (Figure 4.2B). The nonpermissive Mn\textsuperscript{2+} concentration for wild-type, 10 μM, did not hinder misincorporation and moderate extension by I260Q or H285D by
the same extent. I260Q pol β is the most robust misincorporator in all of our experiments on this dsDNA substrate.

Although wild-type pol β and I260Q show similar levels of misincorporation and mispair extension in the absence of the second incoming dNTP, the mutator activity of I260Q is verified by the misincorporation and extension opposite G, in the presence of either metal ion. Additionally, there is preliminary evidence of template switching DNA synthesis by I260Q when all four dNTPs are present with Mn²⁺ (Figure 4.1D, lane “0”). We observe a band of fluorescence with greater length than the predicted 20-nt fully-extended product. This could be due to the additional of one extra nucleotide or an indicator of the potential for some pol β variants to switch to an available 5’-template strand for further polymerization.
Figure 4.1. Misincorporation and mispair extension activities of pol β variants

A

3'-TAC ACT CGT CAG CGC TGC AT-5'
5'-/56-FAM/ATG TGA G -3'

dCTP
dATP

B

20nt-
7nt-

Control
WT
I260Q
H285D
E295K

C

Misincorporation/Extension with Mg$^{2+}$

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>WT</th>
<th>I260Q</th>
<th>H285D</th>
<th>E295K</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>20nt</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>7nt</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>10 minutes</td>
<td>Missing nucleotide</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
<td>C</td>
</tr>
<tr>
<td>30 minutes</td>
<td>Missing nucleotide</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
<td>C</td>
</tr>
</tbody>
</table>

D

Misincorporation/Extension with Mn$^{2+}$

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>I260Q</th>
<th>H285D</th>
<th>E295K</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>20nt</td>
<td>0</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7nt</td>
<td>0</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 minutes</td>
<td>Missing nucleotide</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>30 minutes</td>
<td>Missing nucleotide</td>
<td>0</td>
<td>C</td>
<td>A</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 4.1. Misincorporation and mispair extension activities of pol β variants

(A) Sequence of the primer:template DNA used for the assays depicted in B, C and D. First and second nucleotides for incorporation are indicated in boxes with arrows corresponding to the first template base, G, and the second, T. (B) Primer extension assay with 15:1 Enzyme:DNA (675 nM:45 nM), all four dNTPs (50 μM), Mg\(^{2+}\) (10 mM MgCl\(_2\)) at 37°C for 10 minutes. Fluorescent bands corresponding to non-extended primer (7-nt) and fully extended primer (20-nt) are indicated on the left. Pol β variant sample indicated below wells. Control sample is primer alone. Note all variants tested generate full-length product in this assay. (C) Misincorporation and mispair extension were tested with Mg\(^{2+}\) (10 mM MgCl\(_2\)). Either dCTP or dATP were omitted (dNTPs at 50 μM), forcing the polymerase to misincorporate either the first dNTP or the second to yield extended products. The missing nucleotide in each lane is indicated below the lane: 0 – all 4 dNTPs are included, this is the positive control for the sample; C – dCTP is omitted from the dNTP mixture; A – dATP is omitted from the dNTP mixture. The negative control, labeled “C” above the lanes, is the primer alone. The pol β variant tested is marked above the respective lanes. Top gel is 10 minutes, bottom is 30 minutes. Enzyme:DNA ratio and temperature are identical to B. (D) Misincorporation and mispair extension were tested with Mn\(^{2+}\) (1 mM MnCl\(_2\)). Assay conditions identical to C aside from Mn\(^{2+}\).
Figure 4.2. Misincorporation threshold determined with MnCl₂ supplementing MgCl₂: wild-type versus I260Q and H285D.
Figure 4.2. Misincorporation threshold determined with MnCl$_2$ supplementing MgCl$_2$: wild-type versus I260Q and H285D

(A) In the same primer extension assay as described above, the two metal ions were combined to determine the threshold at which wild-type begins to incorporate a mispaired nucleotide during the second incorporation event. N, negative control - no metal ions or dNTPs, P, positive control - 10 mM MgCl$_2$ and 50 μM all four dNTPs, remaining lanes contain 10 mM MgCl$_2$, MnCl$_2$ at the concentration indicated above lane, 45 nM enzyme, 450 nM dsDNA substrate and 50 μM dNTPs (dATP was omitted). Reactions were incubated at 37°C and quenched after 10 minutes. (B) Using the threshold determined for wild-type in A, 10-100 μM MnCl$_2$, we tested I260Q and H285D at similar concentrations of MnCl$_2$. All other parameters are the same.
4.3.2. The H285D pol β variant’s mutator phenotype on the dsDNA substrate

To verify the primer extension activity of H285D on the dsDNA substrate used for crystallography and in our kinetic studies, we repeated the previously described assays to compare H285D to wild-type (Figure 4.1B). In this assay, H285D is a very active polymerase and is able to complete primer extension by 10 minutes (Figure 4.1B), similarly to wild-type, I260Q and E295K. The H285D variant is the most active misincorporator and mispair extender variant we tested, even in the presence of Mg$^{2+}$. Figure 4.1C shows that in the absence of dATP, H285D is able to create fully-extended products after only 10 minutes, more so than the other three pol β variants tested in these assays. Additionally, a small amount of mispair extension is observed for H285D in the absence of dCTP after 30 minutes with Mg$^{2+}$ (Figure 4.1C). These findings are more pronounced when Mn$^{2+}$ is substituted for Mg$^{2+}$ in Figure 4.1D. After 30 minutes, H285D generates more mispair extended products in the absence of dCTP than any other variant. This misincorporation activity of H285D is also apparent in the presence of both metal ions (Figure 4.2B). The substrate and enzyme phenotype differences will be discussed further in Chapter 6.
4.3.3. E295K pol β is a variant linked to human gastric cancer with increased DNA binding and decreased polymerase activity compared to wild-type

4.3.3.1. E295K shows wild-type-levels of activity for primer extension on short dsDNA substrate

The activity of E295K pol β was also tested on the dsDNA crystallization oligonucleotides. We followed published protocols (Chapter 2 (Gridley 2012)), only changing the pol β:DNA ratio to 15:1 to mimic the assay performed by Lang et al. (Lang 2007). In this experiment, the dsDNA was created such that a fluorescent probe is at the 5’ end of the primer strand and there is a 7-base pair duplex with a 13-base pair overhang (Figure 4.1A). Incorporation or primer extension by the polymerase is monitored by quenching the sample with an EDTA-containing buffer, resolving the sample via electrophoresis on a 15% TBE-Urea PAGE, and visualizing on a Typhoon Scanner (GE Healthcare Sciences, U.S.A.). The fluorescent bands mark primer extension products.

As discussed below, we have shown that pol β E295K is active, to a degree similar to wild-type and I260Q, on this dsDNA substrate (Figure 4.1B). It is worth noting that the E295K purified polymerase loses activity in our assays over time. Initial attempts to use protein stored from 5 days to 1 month at 4°C yielded completely inactive E295K protein, where wild-type and I260Q protein retains activity after more than 30 days stored at 4°C. Due to the loss of activity, E295K protein was never frozen prior to crystallization
or activity assays and was always made fresh and used within three days of purification. After 30 minutes, the relative amount of extended product is similar for all pol β variants tested. We then tested the misincorporation potential of E295K in the presence of Mg²⁺ at 10 and 30 minutes. Figure 4.1C shows that in the absence of the first nucleotide (dCTP) the polymerase is unable to incorporate or extend a mispair. However, incorporation of mispaired nucleotides is detected after 10 minutes when the second nucleotide is omitted (dATP), however full-length product is not observed after 30 minutes (Figure 4.1C). Compared to wild-type and H285D, E295K creates less product in this misincorporation and mispair extension assay, indicating that it probably is a slower, less efficient enzyme.

The omission assay was repeated with Mn²⁺ instead of Mg²⁺, based on the hypothesis that the variant polymerases will be more likely to extend mispairs than wild-type in the presence of Mn²⁺. The E295K variant of pol β is a mispair extender with Mn²⁺ when the second nucleotide, dATP, is missing (Figure 4.1D). Interestingly, after 30 minutes of extension in the presence of Mn²⁺, E295K generates product containing more than 20 nucleotides (Figure 4.1D, E295K in the presence of all four dNTPs). This could be due to the polymerase adding one additional nucleotide in a template-independent fashion, or the higher molecular weight product could be a result of template switching (see Discussion, Chapter 6). Contrary to previous knowledge, we find E295K of rat pol β to be active on our dsDNA substrate, and to even possess a mutator phenotype in the
presence of both Mg$^{2+}$ and Mn$^{2+}$. The differing activity profiles of E295K might be partly attributed to the purity, freshness, age or oxidation state of the enzyme.

4.3.3.2. Crystallography and MD simulations of E295K show the effect of the complete charge reversal in the vicinity of position 295

We crystallized this mutator variant of pol β in the presence of substrate to observe changes in neighboring amino acids to position 295 and also to identify changes in DNA/dNTP alignment within the active site. Our crystal structure of a binary complex of E295K bound to dsDNA (PDB ID: 3V7L) reveals that the lysine295 side chain is poorly ordered (Supplemental Figure 3.9.3, Chapter 3 (Li 2012)). The fast movement of this positive residue could be due to the lack of $\pi$-$\pi$ stacking interactions of the carboxylate group with the adjacent aromatic side chains and/or a result of unfavorable interactions with the Arg258 side chain that has rotated away from Asp192.

Transition path sampling via MD revealed the conformational transition pathway for E295K, concluding that the overall closing energy barrier is higher in the mutant than in the wild-type pol β (Chapter 3 (Li 2012)). As predicted, the lysine at 295 introduces significant dynamic changes in the immediate environment and results in a high-energy barrier for the rotation of Arg258 away from Asp192 (Chapter 3 (Li 2012)). The binary complex structure of E295K with dsDNA is not fully closed because no dNTP or Mg$^{2+}$ ions are coordinated in the active site.
In addition to the transition path sampling by Tamar Schlick’s group, we independently performed molecular dynamics simulations on a closed ternary complex of E295K to supplement the binary structure of this pol β variant and enhance our understanding of the DNA binding properties and the dynamic behavior of residues 271, 295 and 296 in this region of the complex. Using a high-resolution structure of pol β:gapped DNA (PDB ID: 2FMS), we modeled in two most likely rotamers of lysine at position 295. We then put these models of E295K into a fully solvated, periodic boundary environment to monitor movement of neighboring side chains and the overall “flexing” of the enzyme complex over a 5 ns simulation.

As expected, we see Arg258 and Asp192 interacting at just 3.3 Å apart in the binary complex (Figure 4.3A). In our modeled and MD simulated structure of E295K in the ternary complex, the Asp192 side chain is flipped to interact with Mg^{2+} in the active site. Because of the positively charged lysine at 295, Arg258 no longer forms a hydrogen bond with 295 (Figure 4.3B). Additionally, we see a subtle, overall shift in the positions of several active site residues: E295K moves 2.5 -3 Å closer to the template DNA, the template bases at positions 0 and -1 move 3 and 5 Å, respectively, away from their previous neighbors, R258 and Y296 (Table 4.1). We propose that these changes specifically alter the geometry of the dNTP binding pocket enough to impede catalysis, while at the same time leaving DNA binding unaffected.
Figure 4.3. Comparison of the binary and ternary complexes of E295K pol β
Figure 4.3. Comparison of the binary and ternary complexes of E295K pol β

(A) The partially closed conformation of E295K (PDB ID: 3V72) shows poor electron density for the lysine at 295, “Lys295”, indicating fast side-chain movement. The salt bridge between Asp192 and Arg258 appears to be intact as these two side-chains are separated by only 3.3 Å. One nearby Na+ atom is shown as a purple sphere. The palm and fingers subdomains are colored magenta and blue, respectively. The DNA primer:template is not shown for clarity. (B) The fully equilibrated closed, ternary complex of E295K (starting PDB ID: 2FMS) after 5 ns of MD simulation. Primer strand of DNA is shown in orange, template is in gray. Notice the conformations of Arg258 and Asp192, pointing in opposite directions. The side chain at 258 repels K295, which juts closer to the DNA and dNTP. Subdomains are colored the same as in panel A. Significant changes in distance are listed in Table 4.1. See text for discussion.
Table 4.1. DNA pol β mutator variant, E295K, affects the local environment and neighboring residues: differences in distances

<table>
<thead>
<tr>
<th>Nearby Residue</th>
<th>NTP</th>
<th>Template $^d$</th>
<th>Primer $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Y N R Y</td>
<td>d</td>
<td>A6 G7 C8</td>
<td>C10 C10</td>
</tr>
<tr>
<td>2 2 2 2 2</td>
<td>T</td>
<td>N1 N1 N3</td>
<td></td>
</tr>
<tr>
<td>5 7 7 8 9</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 1 9 3 6</td>
<td>P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variant</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E295$^a$</td>
<td>4.2</td>
<td>5.1</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>4.1</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>4.1</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>9.1</td>
<td>7.5</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>K295, 1$^b$</td>
<td>-0.6</td>
<td>+4.4</td>
<td>+3.6</td>
</tr>
<tr>
<td></td>
<td>+0.8</td>
<td>-1.3</td>
<td>+4.2</td>
</tr>
<tr>
<td></td>
<td>+1.3</td>
<td>-1.6</td>
<td>+2.1</td>
</tr>
<tr>
<td></td>
<td>+1.3</td>
<td>-1.4</td>
<td>+1.8</td>
</tr>
<tr>
<td></td>
<td>+0.8</td>
<td>-0.8</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>+4.2</td>
<td>+2.0</td>
<td>+3.3</td>
</tr>
<tr>
<td></td>
<td>+2.1</td>
<td>+3.0</td>
<td>+4.7</td>
</tr>
<tr>
<td>K295, 2$^b$</td>
<td>+0.9</td>
<td>-1.3</td>
<td>-1.7</td>
</tr>
<tr>
<td></td>
<td>-1.3</td>
<td>+2.0</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>-2.0</td>
<td>-1.4</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>-1.6</td>
<td>+2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.4</td>
<td>+3.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K295 3/11$^c$</td>
<td>-2.7</td>
<td>-0.6</td>
<td>+0.9</td>
</tr>
<tr>
<td></td>
<td>+0.1</td>
<td>-0.8</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>-3.0</td>
<td>+2.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.0</td>
<td>+2.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.0</td>
<td>-2.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K295 6/11$^c$</td>
<td>-3.0</td>
<td>-1.1</td>
<td>+0.9</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>+0.5</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>-0.5</td>
<td>+1.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-5.2</td>
<td>+3.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.2</td>
<td>+3.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Native glutamic acid residue of PDB ID: 2FMS is located in the vicinity of the residues and substrate atoms listed across the top of the table. Measured distances, in Å, between E295 and those nearby are listed and used as a reference for the changes in these distances as a result of the lysine mutation.

$^b$K295, 1 and 2 represent two alternative side-chain rotamers of lysine, modeled into 2FMS using PyMol. Distances between these two rotamers and the residues listed are indicated in regards to the change from wild-type: +, refers to a decrease in the distance between the two respective residues; −, refers to an increase in the distance between the two respective residues.

$^c$K295 3/11 and 6/11 refer to two snapshots, 3 and 6, of the eleven total generate during the MD simulation described in the text. Again, the distance between the side chain at 295 and nearby residues are indicated in terms of the change in distances as compared to wild-type (E295): +, indicates a decrease in distance; −,
indicates an increase in distance. Note that the overall trend in the MD simulation is an increase in the distance between K295 and nearby protein residues, and a decrease in the distance between K295 and substrate atoms.

Template and primer strand atoms are indicated via base number and atom.
Chapter 5. Sneha Rangarajan, Chelsea L. Gridley, Karen Chave, Shibani Dalal, Joann B. Sweasy, and Joachim Jaeger. The K72E DNA polymerase β mutant crystal structures explain loss of lyase function and reveal conformational changes in the polymerase domain.

5.1. Description of Participation

My contribution to this manuscript began with streamlining the purification protocol for various pol β constructs: 6xHis-Tagged, shortened 6xHis-Tagged and native full-length. Before Sneha joined the lab, I worked out the high-level expression for pol β proteins using the E. coli strains, BL21(DE3) and subsequent Rosetta 2(DE3) cell lines. When Sneha joined the lab, I trained her in all aspects of site-directed mutagenesis, to create the shortened 6xHis-tagged pol β, protein expression and purification and creating crystallization screens.

We worked together to troubleshoot crystallization techniques and create the primer extension assays described throughout this thesis. We both contributed to all aspects of data collection, refinement, and structure determination. Data collection occurred at the Brookhaven National Laboratory several times each year where we worked around the clock to collect X-ray diffraction images for our various crystal forms.
This manuscript is a reflection of one of the pol β variants examined by Sneha. She wrote the majority of the text and I spent a significant portion of time editing and revising. Karen Chave was of great assistance in creating the pol β construct with the modified short histidine tag. The original plasmid constructs of pol β were provided by Shibani Dalal and Joann B. Sweasy.
5.2. Abstract

A Lys to Glu mutation at position 72 in pol β creates a charge reversal in the lyase active site preventing Schiff’s base formation and abrogating efficient 5’-deoxy-ribose phosphate removal. In this study, K72E was identified as a mutator polymerase on a 3’ recessed DNA substrate. The kinetic basis for the mutator phenotype is slightly tighter binding of the incoming nucleotide and an increased rate of misincorporation of nucleotides as compared to the wild-type enzyme. Crystal structures of K72E pol β detail how the mutant polymerase modifies the local side chain positions in the lyase domain leading to compromised deoxy-ribose phosphate lyase activity. The structures also elucidate how the mutant propagates its effect mainly through Tyr39 and Arg40 in the 8 kDa N-terminal domain to the polymerase active site/nucleotide-binding site, by disrupting interactions with Asp276 in the fingers C-terminal subdomain, which is critical in stabilizing the incoming nucleotide.
5.3. Introduction

The cellular genome incurs an extensive amount of damage from the surrounding environment. To repair such damage and restore genomic integrity, the cell has evolved several DNA repair mechanisms. Base excision repair (BER) is one of the crucial pathways responsible for recovery of single stranded breaks and removal of damaged bases (Barnes 2004). In this pathway, after recognition and excision of the damaged base, the DNA is cleaved at an abasic site (AP) by an AP endonuclease (APE) leaving behind a 3’ hydroxyl and 5’-deoxy-ribose phosphate (dRP) (Seeberg 1995). At this point, DNA polymerase β (pol β), a small DNA repair polymerase, removes the dRP moiety and then fills in the gap in the DNA (Matsumoto 1995). Many pol β mutations are known to be associated with human cancers (Starcevic 2004) and, given the role of pol β in DNA repair, it is important to understand the nature of these mutations that could lead to compromised DNA repair activity by pol β.

Pol β is a member of the X family of polymerases and shares the classic “right hand” structure with many eukaryotic polymerases (Steitz 1994; Jaeger 1999). It consists of two domains, the N-terminal, 8 kDa and the C-terminal, 31 kDa domains, separated by a protease sensitive linker region. While the nucleotidyl transfer reaction center rests in the 31 kDa, the 8 kDa has been shown to be responsible for the lyase activity. Although pol β is considered a low fidelity polymerase (Kunkel 1985), it inserts nucleotides in a template directed manner and efficiently selects the correct nucleotide from a pool of very similar nucleotide substrates. Some of the checkpoints include DNA binding,
deoxynucleoside triphosphate (dNTP) binding, extent of conformational changes upon correct nucleotide pairing and phosphodiester bond formation. However, it lacks proofreading capabilities, which makes it a suitable system to study incorporation infidelity. Additionally, it is a small enzyme relatively easy to characterize kinetically and structurally, making pol β a fitting candidate for understanding the mechanisms of repair polymerases.

A cluster of conserved carboxylates and other polar residues, located at the base of the polymerase cleft, define the polymerase active site in pol β. The carboxylates (Asp190, Asp192 and Asp256) are particularly important for the catalysis, a phosphoryl transfer reaction involving a nucleophilic attack by the 3’ hydroxyl of the primer terminus on the dNTP α-phosphate, with the release of the pyrophosphate. This type of a two metal-ion mechanism has been proposed for phosphoryl transfer reactions in the case of several polymerases including Klenow fragment (Beese 1991). The lyase activity, harbored in the N-terminal 8 kDa domain (residues 1-91), is mediated by the formation of Schiff’s base intermediate with the dRP moiety on the DNA, which is eventually removed in a β-elimination mechanism (Piersen 1996). Pierson et al. along with others have established Lys72 to be the key catalytic residue responsible for dRP excision from the 5’ incised AP sites, by forming a Schiff’s base with the dRP moiety (Piersen 1996; Matsumoto 1998). Substitution of Lys72 with Ala, Arg or Gln has been shown to largely abrogate the dRP excision activity to less than 10% as compared to that of the wild-type lyase activity, emphasizing the importance of lysine at position 72.
The present study focuses on mutating K72 to E72, which leads to a complete reversal of charge. We hypothesize that the mutation will not only affect the lyase activity by altering the local interactions and preventing the Schiff’s base formation as in the case of K72A (Prasad 1998), but also will have long range effects on the polymerase active site located about 20 Å away. While many studies of pol β variants have revealed a link between residues distant from the polymerase active site and nucleotide misincorporation, structural information explaining the long-range effect of the mutation on the polymerase active site, is still lacking (Kosa 1999b; Shah 2001a; Shah 2001b; Dalal 2004; Lin 2007).

This is the first time the K72E variant has been characterized in great detail through high-resolution X-ray crystallography and biochemical assays. While our biochemical analysis identifies that K72E misincorporates on the recessed double stranded DNA (dsDNA), the crystal structure sheds light on how subtle changes in the lyase domain may be able to cause changes in the side chain positions of residues important in stabilizing the incoming dNTP, presumably altering the fidelity of the polymerase.

5.4. Materials and Methods

5.4.1. Cloning, expression and purification of pol β

The K72E pol β variant was generated by site-directed mutagenesis from the wild-type construct in a pET28a(+) vector (Novagen) containing a modified amino terminal His6
The histidine tag was introduced such that the first four amino acids of the rat pol β gene (SRKR) were replaced with histidines. The protein was expressed in the bacterial strain, Rosetta2 DE3 ((with genotype F- ompT hsdSB(rB-, mB-) gal dcm (DE3) pRARE2 (Cam8)), and purification was based on a published protocol (Kosa 1999a). Briefly, the cells were resuspended in 500 mM NaCl, 100 mM Tris, pH 8.0, with SigmaFAST Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma-Aldrich) and sonicated (5 sec pulse x 30 sec rest) for ten minutes while suspended in an ice-water bath, followed by ultracentrifugation. The protein mixture was filtered with 5 μm and 0.45 μm syringe filters, applied to a His-Trap Nickel-NTA agarose column (GE Healthcare Bio-Sciences Corp., USA) and separated by an imidazole gradient [Buffer A: 20 mM Tris HCl, 100 mM NaCl, 20 mM imidazole, pH 8.0; Buffer B: 20 mM Tris HCl, 100 mM NaCl, 500 mM imidazole, pH 8.0] using a liquid chromatography system (Aekta Prime; Amersham Biosciences). Fractions containing pol β were identified by electrophoresis on a 12.5% SDS PAGE, were pooled and then concentrated to <5 mL before being loaded onto a gel filtration column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare Bio-Sciences Corp., USA). Protein was purified and buffer exchanged over the gel filtration column with buffers specific to the crystallization of apoenzyme (20 mM (NH4)2SO4, 100 mM HEPES, pH 7.0) or dsDNA-pol β complex (100 mM MES, 10 mM (NH4)2SO4, 30 mM NaCl, pH 6.5). Following gel filtration of the polymerase, the protein was greater than 95% pure as determined by SDS PAGE and dynamic light scattering (data not shown). Concentration of the protein was calculated on the basis of an extinction coefficient of 23,380 M⁻¹cm⁻¹ and a molecular weight of 39 kDa.
5.4.2. Preparation of 5′-dRP substrate

Oligonucleotides for this purpose were synthesized by WM Keck facility at Yale University. The sequence of the template was 5′-CTG CAG CTG ATG CGC UGT ACG GAT CCC CGG GTA C-3′. The 3′ end of the DNA template, which contained a single U at position 19 from the 3′ end, was radiolabeled using [−α\(^{32}\)P] ATP and terminal transferase and annealed with its complementary oligo. This DNA duplex was treated with Uracil DNA Glycosylase (UDG) at 2 U/1 pmol DNA in 50 mM HEPES, pH 7.5, at 37°C for 10 min, followed by treatment with human AP endonuclease (APE1), at 2 U/1 pmol DNA in Buffer R (10 mM MgCl\(_2\), 20 mM KCl and 2 mM dithiothreitol (DTT)) at 37°C for 10 min, which incises the phosphodiester backbone on the 5′-side of the AP site and leaves a 3′-OH and a 5′-dRP residue. Due to the labile nature of the AP site, the 5′-dRP-containing DNA substrate was prepared just before use.

5.4.3. 5′-dRP lyase assay

The assay was set up according to Haracska et al., with minor modifications (Haracska 2003). Typically, reaction mixtures (24 μl) contained the DNA polymerases (200–400 nM) and the 5′-dRP-containing DNA substrate (100 nM) in Buffer R (50 mM HEPES, pH 7.5, 10 mM MgCl\(_2\), 20 mM KCl and 2 mM DTT). Reactions were incubated at 37°C for 10 min. The reaction product was stabilized by the addition of 2 M sodium borohydride to a final concentration of 340 mM, followed by incubation on ice for 30 min. Stabilized (reduced) DNA products were ethanol-precipitated in the presence of 0.1 μg/ml of
tRNA, resuspended in water, and an equal volume of formamide dye was added. These products were resolved on a 20% polyacrylamide gel and visualized with a Storm 860 PhosphorImager (Molecular Dynamics, Inc.).

5.4.4. In vitro primer extension assay

Using the dsDNA substrates utilized in the co-crystal studies, we devised a fluorescent primer extension assay to verify in vitro polymerase activity of purified wild-type and K72E pol β with modified His6 tags. Briefly, the primer strand was labeled with a fluorescent probe at the 5’end: 5’-/56-FAM/ATG TGA G-3’, which when annealed to the template strand: 5’-TAC GTC GCG ACT G CAC AT-3’ created a 7-base pair duplex with a 13-nucleotide (nt) 5’-overhang (Integrated DNA Technologies, Inc., Coralville, Iowa). The oligonucleotides were annealed at 45 nM, primer:template 1:1.2, in 50 mM Tris, 250 mM NaCl, pH 8.0. Proper annealing was verified using dynamic light scattering (DLS) (data not shown). Enzyme to DNA ratios ranged from 50:1 to 10:1 where DNA remained at 45 nM and the enzyme concentration was changed to suit assay conditions. The assay was performed at 37°C in 20 mM Tris, pH 8.0, 60 mM NaCl, 10% glycerol, 1 mM DTT (final concentrations). The DNA and protein were mixed with reaction buffer, and the reaction was initiated with the addition of dNTPs (50 μM final concentration) and MgCl2 (10 mM final concentration). At various time points, aliquots were removed and quenched with an equal volume of TBE-Urea Sample buffer, 89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0, 12% Ficoll, 0.01% bromophenol blue, 0.02%
xylene cyanole FF, 7 M urea, (Bio-Rad Laboratories, Inc.) and samples were heated to 95°C for five minutes before being separated on a 15% TBE-Urea Denaturing PAGE gel. Gels were then visualized on a Typhoon Imager (GE Healthcare Bio-Sciences Corp., USA) for the detection of fluorescent bands indicating extended versus unextended primer. Tagless wild-type and variant pol β protein samples previously tested for in vitro activity served as positive and negative controls, respectively.

5.4.5. In vitro misincorporation and mispair extension assays

Using the aforementioned substrates, we performed additional primer extension assays: a “nucleotide omission primer extension” assay using three of the four dNTPs and a “one at a time primer extension” assay using only one of the four dNTPs. The missing substrates used in each reaction are indicated in the respective figures. The nucleotide omission primer extension assay contained 45 nM annealed primer-template DNA, 2.25 μM pol β and 500 μM dNTPs (final concentration). The one at a time primer extension assay contained the same concentrations of DNA and protein although the dNTPs concentration varied from 250 to 1000 μM. In both cases, aliquots were removed and quenched with an equal volume of 80% formamide containing 100 mM EDTA, with bromophenol blue and xylene cyanole dyes, at the end of 20 minutes. While the samples from the former assay were run on a 15% TBE-Urea gel as mentioned previously, the latter were run on a 17.5% polyacrylamide (19:1 dilution) 7.5 M urea, 1× Tris-borate-EDTA gel that was preheated and run at 50°C. The amounts of FAM fluorescence in the
unextended and extended primer bands were quantitated using a Typhoon 9400 scanner and ImageQuant software (GE Healthcare Bio-Sciences Corp., USA).

5.4.6. **Single-turnover kinetics for correct and incorrect nucleotide incorporation**

To elucidate the relative ability of K72E compared with the wild-type enzyme to incorporate correct and incorrect dNTPs into a primer-template substrate, we determined the equilibrium constant for dNTP binding, $K_d$, and the maximum rate of polymerization, $k_{pol}$, for correct [dCTP] and incorrect [dATP] incorporation for each enzyme. The reactions were conducted under single turnover conditions, on the same 3'-recessed substrate as mentioned above except that the substrate concentrations and reaction times were changed. For correct incorporations, the substrate concentrations were typically 50-1000 μM and the reaction times were 1-30 minutes. For incorrect incorporations, the substrate concentrations were 100-1500 μM and the reaction times were 1-60 minutes. The reactions resulted in the addition of one nucleotide onto the primer. The n (unextended) and n+1 (extended by one nucleotide) DNA products were resolved on a 17.5% polyacrylamide (19:1 dilution) 7.5 M urea, 1× Tris-borate-EDTA gel and the bands were quantified as mentioned previously.

The single turnover kinetic data were fit into a single exponential equation using Kaleidagraph software (Synergy Software):
\[ [product] = A(1 - e^{k_{obs}t}) \]

where A is the amplitude, \( k_{obs} \) is the observed rate constant and t is the time (Li 1999). A secondary kinetic plot was constructed by plotting the observed rate constant (\( k_{obs} \)) versus dNTP concentration, which was then fit into the hyperbolic equation:

\[ k_{obs} = \frac{k_{pol}[dATP]}{K_d + [dNTP]} \]

Where \( k_{pol} \) is the maximum rate of polymerization and \( K_d \) is the apparent equilibrium dissociation constant (Li 1999). Fidelity values were calculated using the following equation:

\[ Fidelity = \frac{(k_{pol}/K_d)_{correct} + (k_{pol}/K_d)_{incorrect}}{(k_{pol}/K_d)_{incorrect}} \]

5.4.7. Crystallization of pol β apoenzyme

Pol β was concentrated to 20 mg/ml prior to setting up crystallization trials. The apoenzyme crystals were grown in a sitting-drop 96-well MRC-2 Crystallization Plate™ (Hampton Research) by mixing equal volumes of protein, at 10 mg/mL, and reservoir solution consisting of 80 -110 mM MES and 5-14% PEG3350. The apo crystals were cryoprotected with 15-20% 2-methyl-2,4-pentandiol (MPD) prior to flash cooling in liquid nitrogen.
5.4.8. Co-cry stallization of pol β-DNA complexes

Prior to crystallization, the proteins were exchanged into a buffer containing 100 mM MES, pH 6.5, 30 mM NaCl and 10 mM (NH₄)₂SO₄ and then concentrated to 20 mg/ml. Co-crystals were grown using 5'-CAA ACT CAC AT-3' as template and 5'-ATG TGA G-3' as primer (mixed 1:1). The DNA was annealed in 20 mM NaCl, 100 mM Tris-HCl, pH 7.5 (2 min at 90°C, 2 min at 70°C, 1 min at 55°C and then cooled to 4°C by decreasing 1°C per cycle) in a DNA Dyad Peltier Thermal Cycler. Crystallization trials were then set up using sitting drop vapor diffusion method by mixing equal volumes of protein and well solution (1.5 μl plus 1.5 μl) in the 96-well MRC-2 Crystal Plates™ (Hampton Research). The reservoir solutions varied from 100-160 mM NaCl and 10-18% PEG3350. The co-crystals were cryoprotected with a solution that consisted of 15% glycerol and mother liquor, before flash cooling.

5.4.9. Data collection and structure determination

X-ray intensity data were collected at beamlines X4C and X25 at the National Synchrotron Light Source at Brookhaven National Lab (Upton, NY). Data integration and reduction were performed using HKL2000 (Otwinowski 1997) or iMosflm 1.5.0 (Battye 2011). Structure determination was initiated using PHENIX 1.7.4 (Adams 2010). The structure of the P2₁ monoclinic apo pol β crystal form was determined by molecular replacement (MR) using PHASER (McCoy 2007) with a composite model of apo pol β (X-ray structure of 31 kDa domain and NMR structure of the complete 8 kDa domain) followed by rigid body refinement, full atomic refinement interspersed with manual
rebuilding in COOT (Emsley 2010). Although the lattices and packing arrangement of the binary complexes of rat pol β are related to that of a human pol β complex (PDB ID: 1BPG), MR was necessary to determine the phases. Subsequent refinement and model building was carried out again using PHENIX and COOT. Final structures were analyzed and verified using COOT, Molprobity (Davis 2007) and PyMol (DeLano 2001).

5.4.10. Molecular dynamics simulations of ternary complexes of K72E and wild-type pol β

A model for a ternary complex of K72E pol β was generated using a high-resolution DNA co-crystal structure (PDB ID: 2FMS). Most amino acid changes between human and rat pol β are located at or near the surface and were modeled using the programs PyMol and VMD (Humphrey 1996). Initial close contacts were carefully analyzed in PyMol and removed by selecting frequently found side chain rotamers, that, as much as possible, resembled conformations found in the parent structure. The CHARMM27 force field was used to parameterize atomic interactions with short-range nonbonding terms calculated up to a 12 Å cutoff for electrostatic and Van der Waals interactions. Contributions from long-range interactions were considered using the particle-mesh-Ewald scheme implemented in NAMD2 version 2.8 (Phillips 2005). All hydrogen-bond lengths were held constant with the SHAKE-RATTLE-ROLL algorithm. The TIP3P model was employed for modeling the water molecules surrounding the protein. The ionic strength was adjusted to 0.2 mol/L for more appropriate screening of Coulomb forces. The PyMol optimized models were subjected to 5,000 steps of energy
minimization followed by 1 ps of a periodic-boundary equilibration molecular dynamics (MD) simulation at 310 K temperature with a time step of 1 fs. The equilibration phase was then followed by a 5 ns production MD simulation, also at 310 K, using a temperature bath with a coupling constant of 5 ps$^{-1}$. Root mean square deviation (RMSD) analyses were performed to evaluate the systems mobility and proper convergence. Trajectory analysis and molecular graphics images were generated using VMD and PyMol. NAMD2 was used for fully equilibrated periodic boundary MD simulations.

5.5. Results

5.5.1. K72E pol β lacks lyase activity

Lys72 has been shown to be the most important catalytic residue in mediating the lyase activity of pol β (Piersen 1996; Matsumoto 1998; Prasad 1998). Therefore, a mutation from Lys72 to Glu should abrogate the ability of the enzyme to form a Schiff’s base intermediate and thus, the mutant enzyme will no longer be able to support dRP lyase activity. In order to test this hypothesis, we performed the dRP lyase assay with two different protein concentrations (10 and 100 nM) (Figure 5.1 and Materials and Methods). As shown in Figure 5.1A, wild-type removes the dRP group very efficiently as suggested by the presence of 19 nt lower band while K72E pol β exhibits compromised lyase activity under the given reaction conditions. The data is quantified as shown in Figure 5.1B.
Figure 5.1. K72E has compromised dRP lyase activity compared to wild-type

The 5′-dRP-containing DNA substrate, LPSD (100 nM), was incubated in buffer R at 37°C for 10 min with increasing concentrations of pol β as indicated in the figure, followed by the addition of 340 mM NaBH₄ and stabilization of the reaction product as indicated in Materials and Methods section. The products were analyzed by a denaturing 20% polyacrylamide gel and visualized by autoradiography. (A) Image of the denaturing gel. (B) A graphical representation of the percentage of incised product versus protein concentration.
5.5.2. *K72E pol β polymerase activity*

Previous studies by Wilson and others indicated that mutations in the lyase active site could affect DNA binding and primer extension (Matsumoto 1998; Prasad 1998). With Lys72 being some $30 \ Å$ away from the polymerase active site, the effect of mutation at this remote site required further biochemical, kinetic and structural analysis. We began this characterization by determining whether the enzyme is active using a small 3’ recessed DNA substrate that readily binds to both wild-type and K72E mutant pol β. The same DNA oligomer also produced well-diffracting crystals with both the parent and the mutator enzyme as discussed later. A simple primer-extension assay, based on the incorporation of a fluorescent probe at the 5’-end of the short co-crystallization substrate, clearly indicates that both, the wild-type and K72E polymerases are active on this substrate (Figure 5.2A). After 30 seconds, the wild-type polymerase exhibits a three to four fold increase in primer extension compared to K72E, as determined by fluorescence intensity of product corresponding to 20 nucleotides in length (Figure 5.2A, quantified data not shown). However, both enzymes show full extension of the dsDNA substrate after 90 seconds as verified in Figure 5.2A. Wild-type pol β was used as the positive control since activity has already been determined on a variety of substrates (Kosa 1999b; Shah 2001b). The polymerase-DNA complex in the absence of dNTP and Mg$^{2+}$ served as the negative control (Figure 5.2A).
Figure 5.2. *In vitro* primer extension assays on the short dsDNA used for co-crystallization.

**A**

30 seconds

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>WT</td>
<td>C</td>
<td>K72E</td>
<td>C</td>
<td>WT</td>
<td>C</td>
<td>K72E</td>
</tr>
</tbody>
</table>

90 seconds

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>WT</td>
<td>C</td>
<td>K72E</td>
<td>C</td>
<td>WT</td>
<td>C</td>
<td>K72E</td>
</tr>
</tbody>
</table>

**B**

WT: 20 minutes

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

K72E: 20 minutes

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

**C**

WT: 30 minutes

<table>
<thead>
<tr>
<th>0.5mM</th>
<th>4mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>[dTTP]</td>
</tr>
</tbody>
</table>

K72E: 30 minutes

<table>
<thead>
<tr>
<th>0.5mM</th>
<th>4mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>[dTTP]</td>
</tr>
</tbody>
</table>
Figure 5.2. In vitro primer extension assays on short double stranded DNA used for co-crystallization

(A) The activity of K72E was tested by using a 5’FAM-labeled primer, otherwise of the same composition as that used in co-crystallization studies. The DNA:Enzyme ratio tested was 1:50 (45 nM:2.25 μM). Unextended substrate at 7-nts and fully extended product at 20-nts are labeled. In lanes marked ‘C’, the dNTPs and metals were omitted to serve as a negative control. (B) Activity of wild-type versus K72E pol β in the absence of dCTP or dATP, the first and second nucleotides to be incorporated during primer extension, respectively. The DNA:Enzyme ratio tested was 1:10 (45 nM:450 nM) in the presence of either MgCl₂ (unmarked) or MnCl₂ (marked with a star). The missing nucleotide in the reaction is listed below each lane: 4, all four dNTPs are missing for the negative control, 0, all four dNTPs are present for the positive control, C, dCTP is omitted, A, dATP is omitted. At the end of 20 minutes, the reaction was quenched and separated on a 15% TBE-Urea denaturing PAGE gel. In the presence of Mg²⁺, K72E extends in the absence of dATP while the wild-type does not (lanes 4 and 8). Mn²⁺ was seen to further accentuate the misincorporation ability in K72E as compared to the wild-type (lanes 5 and 9). (C) Incorporation of dTTP opposite G for the wild-type and K72E pol β. Protein-DNA complexes containing 2250 nM enzyme and 45 nM DNA substrate were mixed with 10 mM MgCl₂ and 0.5, 0.75, 1, 2, 3 or 4 nM dTTP. The reaction was terminated at the end of 30 and products were resolved by denaturing sequencing gel electrophoresis.
5.5.3. K72E pol β misincorporates nucleotides on dsDNA

In order to investigate the fidelity of DNA synthesis on a 3’ recessed substrate by the possible mutator K72E, we employed a nucleotide omission assay to determine the ability of K72E to misincorporate and extend from incorrect bases. With excess enzyme, the first or the second nucleotide, dCTP and dATP, respectively, required for correct incorporation were systematically omitted from the reaction mixtures. After 20 minutes, both enzymes show little to no activity in the absence of the first incoming nucleotide, dCTP (Figure 5.2B lanes 3 & 7), and extend the primer to the same extent, when all four nucleotides are present (Figure 5.2B lanes 2 & 6). As seen in Figure 5.2B, in the absence of the second incoming nucleotide dATP (and presence of Mg\(^{2+}\)), K72E misincorporates and extends the mispair (lane 8) whereas the wild-type does not (lane 4).

We repeated the primer extension assay where dATP was omitted, this time substituting MgCl\(_2\) for MnCl\(_2\) (Figure 5.2B, lanes 5 & 9, marked with a star). Mn\(^{2+}\) is known to alter polymerase fidelity (Liu 2001) and allow the coordination of mismatched nucleotides in the polymerase active site (Tabor 1989; Werneburg 1996). In the presence of Mn\(^{2+}\), the wild-type misincorporates in the absence of dATP, but extends the mispair poorly (Figure 5.2B, lane 5). K72E on the other hand, not only misincorporates in the absence of dATP, but also strongly extends the mispair (Figure 5.2B, lane 9). Thus, K72E misincorporates and extends the mispair and is termed a mutator on our 3’ recessed co-crystallization substrate.
To complement the assays where one dNTP was omitted from the reaction, we performed a “one-at-a-time” primer extension assay where only one nucleotide is given per reaction. The one at a time primer extension assay was performed on the same substrate as described above and confirms similar misincorporation characteristics of K72E. In this assay, incorporation of deoxycytosine (dC) yields the correct extension product. When dCTP is the sole nucleotide in the mixture, wild-type and K72E both correctly incorporate dCTP opposite template G (data not shown). However, when only dTTP is present, over a range of concentrations, along with Mg$^{2+}$, K72E misincorporates opposite G whereas the wild-type does not to the same extent (Figure 5.2C). In general, K72E appears to be able to misincorporate nucleotides on our 3’ recessed substrate and extend the resulting mispair much more efficiently than wild-type. In order to verify this nature of misincorporation, we carried out another single nucleotide incorporation assay with an increasing concentration of dTTP and in the presence of Mn$^{2+}$ (data not shown). The assay confirmed that K72E misincorporates T opposite G at a much higher frequency than the wild-type enzyme (data not shown).

5.5.5. K72E pol β lacks discrimination at the level of ground state binding

From the nucleotide omission and one at a time primer extension assays, it was clear that K72E exhibits misincorporation of nucleotides and extends the mispair (Figure 5.2B). In our preliminary assays, K72E pol β was found to misincorporate T opposite G on the 3’ recessed substrate (Figure 5.2C). We also determined the efficiency of K72E to
incorporate dCTP opposite G using the same substrate. The assay conditions were optimized by varying the polymerase excess such that about 90% of the DNA substrate was bound by the enzyme (data not shown). The percent of extended product was plotted over time for wild-type and K72E (Figure 5.3A, left panel, wild-type; right-panel, K72E). The observed rates of product formation by wild-type and K72E, \( k_{\text{obs}} \), were determined as described in “Materials and Methods” and were plotted against the various concentrations of dNTPs used in these experiments, as shown in Figure 5.3B. The measured equilibrium dissociation constant, \( K_d \), for incorporation of the correct nucleotide was determined to be 112 \( \mu \text{M} \) and 42 \( \mu \text{M} \) for the wild-type and K72E polymerases, respectively. Thus, the mutant exhibits a 2-fold reduction in the \( K_d \) as compared to the wild-type enzyme, thereby binding the incoming nucleotide much more tightly. Not much of a difference was observed at the \( k_{\text{pol}} \) level between the two enzymes.
Figure 5.3. Single turnover experiments of dCTP incorporation opposite template G
Figure 5.3. Single turnover experiments of dCTP incorporation opposite template G.

(A) Incorporation of dCTP opposite template G for the wild-type (left panel) and K72E (right panel) polymerases at 37°C. Protein-DNA complex containing 2250 nM enzyme and 45 nM DNA substrate was mixed with 10 mM MgCl$_2$ and 50, 250, 500, 750 or 1000 μM dCTP. The reactions were terminated by EDTA at different time points, and the product was resolved by denaturing sequencing gel electrophoresis. The data were fit to the single exponential equation to obtain $k_{obs}$. (B) This figure depicts the secondary kinetic plot of $k_{obs}$ versus the dCTP concentration for wild-type and K72E. The data were fit to a hyperbolic equation as described under “Materials and Methods”. The solid line represents the best fit of the data to the hyperbolic equation; diamonds, wild-type; squares, K72E.
5.5.5. Local changes in the dRP lyase active site

To gain insights into the structures of unbound polymerase and unoccupied dRP lyase active site, the K72E pol β variant was crystallized in the unliganded form. The unliganded crystal forms of both K72E and wild-type (Gridley 2012, Chapter 2) pol β belong to two different monoclinic space groups (P2₁), with pol β packing arrangements not described before. There is one molecule in the asymmetric unit (ASU) of K72E pol β, whereas there are two molecules, essentially related by a translation of x y z, in the ASU of the native polymerase. The structures of wild-type pol β and the K72E variant have been determined to 2.65 and 2.6 Å, respectively. The K72E apo polymerase and wild-type pol β structures superimpose with 1.598 Å overall RMSD respectively (Figure 5.4A). All molecules adopt a wide-open conformation of the fingers and thumb subdomains, with the N-terminal region separated by almost 50 Å from the bulk of the remaining polymerase core (residues 149-335) (Figure 5.4A). The distance between the dRP lyase- (Lys72) and the polymerase active sites (Asp190) is about 46 Å in the apo protein (Figure 5.4A). Comparison of the 8 kDa domains in the K72E and wild-type structures shows that the secondary structure is not grossly affected by changing Lys72 to Glu. However, due to the charge reversal from Lys to Glu, and the slight reduction in length of the Glu side chain, the mutation introduces significant local shifts in adjacent side chains (Figure 5.4B). Furthermore, the carboxylate group of Glu72 forms a double hydrogen bond with the side chain hydroxyl of Tyr39 (Figure 5.5B). Interestingly, in the wild-type protein, Lys72 points away from Tyr39. In fact, most arginine and lysine side chains around the lyase active site of the wild-type structure, Lys72 included, generally
point outward to the bulk solvent away from the 8 kDa domain. Hence, K72E pol β shows a collapsed, more compact side chain arrangement in the dRP lyase active site.

5.5.6. Crystallographic studies of the K72E lyase active site in the presence of substrate

In order to study the substrate-bound structures of K72E and wild-type polymerase β and examine possible effects of the mutation on substrate binding, the binary structures with a 3′-recessed duplex were determined at Bragg spacings between 2.27 and 2.3 Å (Figure 5.5A and Table 5.1). Both wild-type and K72E pol β crystallized in the same orthorhombic space group (P2_12_12_1). Interestingly, a lattice contact with a symmetry related DNA primer strand perfectly positions the two 5′ terminal bases of the primer strand (5′-ATG TGA G-3′) adjacent to and within the dRP lyase active site (Figure 5.4B). These crystal contacts provide detailed insights into the interactions and recognition of a ribose-phosphate moiety in the K72E mutant and wild-type lyase active site. Even in the presence of (symmetry-related) DNA, Glu72 maintains the double hydrogen bonds of 3.0 Å and 3.3 Å in length with Tyr39, as seen in the unliganded form of the mutant. Lys35 and Lys68 interact with the phosphate oxygens OP1 and OP2, locking in the DNA relative to the dRP lyase residues 39, 72 and 84, all of which are well defined in the density maps. The Nζ of Lys84 rotates inward by 160° around χ4, now pointing away from the substrate. In the case of wild-type pol β we notice that the surrounding polar or charged residues generally interact more favorably with the DNA primer.
Figure 5.4. Apoenzyme structures of K72E and wild-type pol β
Figure 5.4. Apoenzyme structures of K72E and wild-type pol β

(A) Overall structures of K72E and wild-type (PDB ID: 3UXN) pol β. The color-coding is by subdomains: 8 kDa in red; thumb in green; palm in magenta; fingers in blue. 

(B) Comparison of unoccupied dRP lyase active sites. In the mutator, Glu72 forms a strong interaction with the side chain OH1 of Tyr39. Furthermore, Lys68 Nz moves towards the Glu72 carboxylate group. The Nz nitrogens of Lys68, Lys72 and Lys84 point away from Tyr39 in wild-type pol β, poised to interact with the incoming substrate. In the K72E apo structure, the positions of these lysines point more toward the interior of the 8 kDa domain, away from the incoming substrate.
Table 5.1. Crystallographic Studies

<table>
<thead>
<tr>
<th>Complex: Rat Pol β</th>
<th>Wild-type:dsDNA</th>
<th>K72E apo</th>
<th>K72E:dsDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>25.0 - 2.5</td>
<td>19.53 - 2.66</td>
<td>30.319 - 2.27</td>
</tr>
<tr>
<td></td>
<td>(2.58 - 2.5)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(2.75 - 2.65)</td>
<td>(2.32 - 2.27)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10 (56.8)</td>
<td>14.2 (44.8)</td>
<td>6.2 (64.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>93.9 (95.3)</td>
<td>89.7 (84)</td>
<td>91.4 (72.7)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>12.45 (1.8)</td>
<td>19.5 (1.9)</td>
<td>38.1 (1.627)</td>
</tr>
<tr>
<td>Reflections (No.)</td>
<td>64617</td>
<td>33396</td>
<td>201278</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique</td>
<td>25251</td>
<td>11826</td>
<td>21805</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.271</td>
<td>1.596</td>
<td>1.273</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; (%)</td>
<td>25.01</td>
<td>25.58</td>
<td>22.53</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>31.66</td>
<td>33.48</td>
<td>27.70</td>
</tr>
<tr>
<td><strong>Residue Content</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid</td>
<td>650</td>
<td>325</td>
<td>325</td>
</tr>
<tr>
<td>Water</td>
<td>159</td>
<td>83</td>
<td>135</td>
</tr>
<tr>
<td>PDB ID</td>
<td>3V7J</td>
<td>3V7L</td>
<td>3V7K</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parenthesis refer to the highest-resolution shell.

<sup>b</sup>R<sub>merge</sub> = 100 x ΣhΣi[I<sub>hi</sub> - I<sub>h</sub>]/ΣhΣi[I<sub>hi</sub>], where I<sub>h</sub> is the mean intensity of symmetry-related reflections, I<sub>hi</sub>. 
5.5.7. Changes in the polymerase active site of K72E pol β

In order to understand how a mutation in the 8 kDa domain could possibly affect the polymerase active site, we examined the overall domain conformations relative to the palm domain, which harbors all critical polymerase active site residues (Steitz 1994). The comparisons show that the 8 kDa and fingers domains in the K72E variant have moved as compared to the wild-type conformation (Figure 5.4A). The corresponding rotation in the 8 kDa domain is about 3.5° with the hinge point located near residues Asp91 - Thr93. The fingers domain in K72E pol β opens up by 2° compared to the wild-type conformation with the hydrophobic hinge adjacent to the palm domain remaining stationary. Therefore, the overall conformation of the K72E mutant structure is slightly more open than wild-type, although both lattices and crystal packing are essentially identical (Figure 5.5B).

Whilst the backbone of the DNA template strand interacts with Arg40 in wild-type pol β, the corresponding Arg in K72E pol β is about 2 Å farther away from the DNA. Tyr39 and Arg40 in α-helix B (32 - 49) are directly adjacent, yet the side chains are pointing in opposite directions with the Tyr making critical interactions in the lyase active site and the Arg40 side chain facing the polymerase cleft. The conformation of the otherwise flexible Arg side chain is stabilized by Tyr36 and, to a lesser extent, by Lys27, both of which are buttressing against Arg40 thereby restricting torsional movements. Since Tyr39 is tightly interacting with the Glu72, Arg40 is moved farther away by about 2 Å in K72E pol β. The effect of this movement has critical implications on interactions
between Arg40 and Asp276, which is important in positioning the incoming nucleotide in the dNTP binding pocket, as discussed below.

5.5.8. Molecular dynamics studies of lyase active site in a pol β: gapped DNA complex

In addition to the crystallographic studies with the 3’-recessed duplex, we also carried out MD simulations of K72E pol β with gapped DNA. For reference, the wild-type complex was subjected to the same simulation protocol. In a detailed analysis of the MD trajectory we focused on the amino acids surrounding residue 72, namely Lys35, Tyr39, Lys68 and Lys84. The negatively charged carboxylate introduced in the largely basic dRP site of K72E pol β alters the side chain positions relative to the downstream duplex portion of the gapped DNA substrate (Figures 5.6A & B). By contrast with the wild-type pol β, these changes in the local network prevent the K72E pol β active site residues from tightly interacting with the downstream DNA. This is also evident from a scatter plot highlighting interactions between residues 39, 72, and 68 (Figure 5.6B).
Figure 5.5. Binary complex of K72E
Figure 5.5. Binary complex of K72E

(A) Comparison of K72E and wild-type pol β complexed with dsDNA superimposed with 1.598 Å RMSD in Cα positions. A lattice contact with a symmetry related DNA primer strand perfectly positions a ribo-phosphate moiety in the dRP lyase active site of both the K72E (red carbon atoms) and wild-type polymerase (gray carbon atoms). Lys35 and Lys68 interact with OP1 and OP2 locking in the DNA relative to other lyase residues 39, 72 and 84. (B) Superposition of the binary complexes of the wild-type and K72E polymerases. The mutant is shown in subdomain colors (see Figure 5.4A) and the wild-type is shown in gray. Note that the mutant structure is slightly more open than the wild-type structure.
Figure 5.6. Molecular dynamics analysis of the K7E lyase active site
Figure 5.6. Molecular dynamics analysis of the K72E lyase active site.

(A) MD simulations of K72E pol β complexed with a gapped substrate, shows that the Glu72 interacts with Tyr39 even in the presence of gapped DNA bound to the lyase active site. (B) Scatter plot analysis of (Lys or Glu)72/Nζ, Tyr39/OH1, Lys84/Nζ and Tyr39/OH1. The variation of distances shows that the wild-type active site retains more flexibility than the K72E mutant, which shows close interactions between the carboxylate oxygens and the OH1 of Tyr39. This interaction is confirmed crystallographically in K72E pol β (apo and dsDNA).
5.6. Discussion

Pol β lacks proofreading abilities, which makes it even more important for the polymerase to achieve utmost fidelity in order to preserve genomic integrity. Previous studies have indicated that a mutation remote to the polymerase domain could affect the fidelity of the polymerase (Kosa 1999b; Shah 2001b; Dalal 2004; Lin 2007). In this study, one such mutation, in the 8 kDa domain, K72E, about 30 Å away from the polymerase active site (Asp192), was investigated in detail, revealing the influence of the 8 kDa domain on the polymerase activity, not realized before. In this study, we identify K72E as a mutator polymerase and provide some of the kinetic and structural underpinning for the mutator phenotype.

5.6.1. Effects of the K72E pol β in lyase active site

Previous structural studies have shown that the lyase active site is dominated by a patch of basic residues that includes Lys 35, 68, 72 and 84 and Arg83 (Prasad 2005). Site-directed mutagenesis and kinetic studies of pol β have identified Lys72 to be the catalytic residue for the β elimination reaction. It is responsible for the dRP excision from the 5′-incised AP sites by forming a Schiff’s base with the dRP moiety (Piersen 1996; Prasad 2005). Comparison of the respective dRP lyase active site residues in K72E and wild-type pol β shows that the critical side chains interact much more tightly with the DNA in the wild-type enzyme, whereas in case of the mutant, they are pointing away from the substrate moiety (Figure 5.5A) towards the interior; for example, Glu72
interacts with Tyr39 much more strongly than that seen in the wild-type enzyme. A model put forward by Wilson and co-workers postulates a rotation of about 120° around the O5'-P bond to adequately place the dRP reactive groups close to relevant active site residues for effective catalysis (Prasad 2005). The 5' terminus of the symmetry related DNA in K72E crystal structure closely confirms this model. Consistent with experimental data for dRP lyase mutants at position 72, a water molecule near Tyr39 OH1 and Glu72 as modeled by Wilson et al. (Prasad 2005) is plausible but not resolved in the K72E electron density maps.

Taken together, the results of this study confirm our model that insertion of Glu72 modifies the local network of side chain interactions and aids in destabilizing the interactions with the downstream DNA. This altered local network around the 5' terminus of the DNA strand at the dRP lyase active site is due to long-range effects of a charge reversal at position 72 and the inability to form Schiff's base are likely the main contributing factors for the negligible lyase activity of the mutant as reflected in the dRP lyase assay (Figure 5.1A).

**5.6.2. K72E is a mutator polymerase**

After we found both the enzymes to be active on the crystallization substrate, we wanted to determine if K72E pol β was a mutator polymerase. For this purpose, we carried out two types of primer extension assays: nucleotide omission assay and one at a time primer extension assay. The nucleotide omission assay with three out of four
nucleotides (dATP missing) indicates that K72E exhibits an increase in extended DNA products when compared to the wild-type enzyme (Figure 2B). This suggests that K72E is more capable of either misincorporating and/or extending mispaired primer termini opposite template G as compared to wild-type pol β. Based on preliminary results, we carried out another one at a time primer extension assay, to examine if K72E misincorporates T opposite G or not. When we titrated the protein-DNA complex with an increasing concentration of dTTP, K72E exhibited an increased propensity to misincorporate T opposite G while the wild-type did not. At the end of 40 minutes, the wild-type pol β does not incorporate nucleotide even at 1 mM dTTP concentration whereas K72E incorporates T strongly, even at 0.5 mM dTTP concentration (Figure 2C). This confirms that K72E misincorporates nucleotides at a much higher frequency than the wild-type polymerase and that it prefers to misincorporate T opposite G. We also determined the catalytic efficiencies, $K_d$ and $k_{pol}$, for the correct nucleotide incorporation and found that K72E has a 2-fold reduction in $K_d$ as compared to the wild-type and therefore binds the incoming nucleotide much more tightly. Tighter binding may not be a preferable thing since it could lock the mutant in place on the DNA, leading to genomic instability. We did not observe any major differences in the $k_{pol}$ values between the two enzymes suggesting that the coordination of Mg$^{2+}$ ions and geometry of the active site residues are presumably unaffected.
5.6.3. Structural basis for the mutator phenotype

In a ternary complex of pol β with template/primer and dNTP, a closure of the polymerase brings the residues of the N-helix into very close proximity with the incoming dNTP and particularly residue Asp276. It has been reported that this residue may come within 3.5 Å of the C4 of the base of the incoming dNTP (Pelletier 1994; Pelletier 1996a; Sawaya 1997). According to the “induced fit” mechanism suggested for pol β, proper alignment of the substrates in the active site is only achieved when the correct dNTP is present. The correct fit is then the limiting step for the nucleotidyl transfer. Sawaya and others (Sawaya 1997), in their model, propose that during the ‘closing’ of the fingers subdomain, Asp276 is oriented towards, and makes contact with Arg40 of the 8 kDa domain, thus helping to determine the size of the active site. We propose that the mutator phenotype of K72E is due to an alteration in this interaction between Asp276 and Arg40, which stabilizes the incoming nucleotide and controls the size of the active site. As observed from the crystal structure of K72E, the Glu72 strongly interacts with Tyr39 by forming two hydrogen bonds and pulls it closer to itself. This in turn, pulls the adjacent Arg40 farther away by about 2 Å from Asp276, thereby increasing the size of the pocket enough to allow for misincorporations. The resultant active site is now big enough to incorporate G:T mispairs; however, not large enough to accommodate the bulkier G:G or G:A mispairs. This is very similar to the phenotype observed when Asp276 is mutated to Glu. The D276E mutant also misincorporates G:T mispairs much more efficiently as compared to the wild-type pol β (Skandalis 2001). The authors hypothesize that the size of the active site is increased as a result of
mutation to a bulkier residue (1.5 Å longer). We modeled the mutation D276E in wild-type pol β complexed with gapped DNA (PDB ID: 2FMS) and found that R40 exists in split conformations. E276 could have two rotamers, one of which increases the volume of the dNTP pocket while maintaining the bidentate salt bridge with R40, which concurs with the phenotype observed for the D276E mutant.

In a complex with gapped DNA (e.g. PDB ID: 2FMS), interactions between the fingers and thumb domains are mediated directly through the downstream duplex. However, in our co-crystal structure with the recessed duplex, and in a similar structure of wild-type pol β (PDB ID: 1HUO) where the downstream DNA is also not present, the direct interactions cannot be observed.

Thus, it is possible that rearrangements in the lyase active site or subtle repositioning of the 8 kDa domain could affect duplex binding in the polymerase cleft. What makes this study really interesting is the fact that the mutator phenotype is not the result of a direct mutation in R40, but is instead caused by a relay of rearrangements as a result of a the mutation at 72, which is situation farther away. This further supports the hypothesis that residues distant from the polymerase active site can influence fidelity.

In summary, while the biochemical analysis demonstrates that K72E variant of pol β is a mutator polymerase, the crystal structures provide insights into the possible mechanisms by which K72E, a mutation in the 8 kDa domain, could affect not just the
dRP lyase activity, but also have an influence on the polymerase activity, located farther away.

5.7. Acknowledgments

Access to the NYSBC beamlines X4C at NSLS Brookhaven National Labs and on-site support by J. Schwanof and R. Abramowitz is gratefully acknowledged. We would also like to thank A. Heroux, H. Robinson and other PXRR staff at X25 for support with data collection and back-up. This work was supported by NCI grant R01CA08030 (to JBS and JJ). Other support by the Biochemistry and Macromolecular Crystallography core facilities at Wadsworth Center/NYS-DOH, and support by Ordway Research Inc. is gratefully acknowledged.
Chapter 6. Discussion

6.1. Review of DNA polymerase β

The role of DNA repair enzymes, including pol β, is critical and indispensible for the maintenance of accurate genomic information. While some slight modifications in genetic information can enable evolution and adaptation of a species to enhance overall survival, some can alternatively cause problems and propagate disease.

Many diseases are caused by errors or unrepaired mutations in genomic DNA. Deletions and base substitutions in DNA, when translated, can yield mutant protein products. The BER pathway is one mechanism by which our cells thwart a subset of DNA damage, mainly counteracting ROS damage. Should one of the many enzymes involved in BER be mutated in both amino acid sequence and in activity, the efficiency of BER may be negatively impacted.

DNA pol β is an integral enzyme in BER, performing two activities in this pathway: dRP-lyase and nucleotidyl transferase in gap-filling synthesis. This dual-functioning enzyme is heavily relied upon for BER activity. The importance of pol β is evident in mice homozygous for pol β deletion, which do not survive beyond fetal development (Gu 1994). In humans, pol β mutator variants have been identified in several cancerous tumor cells (Canitrot 1998; Iwanaga 1999; Starcevic 2004; Albertella 2005), suggesting a relationship between pol β activity and cancer. To understand the relationship between
pol β and cancer, we must first look at the accuracy of nucleotidyl transfer and nucleotide discrimination, correct versus incorrect, for wild-type and the mutator variants of pol β.

The nucleotidyl transfer step is the underlying chemical reaction mediated by DNA polymerases. Pol β is one of the X family polymerases, along with λ, μ and terminal deoxyribonucleotidyl transferase. As the smallest of the polymerases, sharing homologous domains: fingers, palm and thumb, and conserved aspartate residues, for catalysis via the two metal-ion mechanism, pol β is a convenient model to study the nucleotidyl transfer reaction. In Chapter 1 we introduced pol β and highlighted some of the most important findings from the numerous kinetic, fidelity and structural studies that have been performed in this study. The consensus to date is that pol β prefers DNA with short gaps (1-6 nucleotides), on these short gaps, pol β is a processive polymerase - a distributive pattern is observed on longer gaps or on 3’ recessed dsDNA (no downstream duplex) - wild-type pol β is moderately error-prone and single amino acid variants of pol β can significantly impact the efficiency and/or fidelity of the polymerase (Kunkel 1985; Singhal 1993; Ahn 1998; Clairmont 1999).
6.2. The variant of I260Q: mutator phenotype and structure in the presence of 3’recessed DNA duplex

In Chapter 2, we presented structural evidence to accompany the efficiency and fidelity studies characterizing the pol β mutator variant, I260Q. The residue at position 260 belongs to a hydrophobic hinge linking the palm and fingers domains and is comprised of residues Ile174, Leu194, Thr196, Ile260, Pro261, Lys262, Tyr265 and Phe272. This cluster of residues has been referred to as a “hinge” because they collectively allow the rotation of the C-terminal domain about this axis to completely enclose the DNA, dNTP and metal-ions in the proper active site geometry. That allows for accurate and faithful nucleotidyl transfer to occur. The glutamine substitution at 260 introduces a polar, hydrophilic amino acid into this normally nonpolar, hydrophobic region.

The hydrophilic nature of the glutamine changes the hydrophobic makeup of this hinge indicated via coordination of water molecules in the vicinity of Q260 observed in our apoenzyme crystal structure (PDB ID: 3UXO, Gridley 2012). Water molecules are not evident in the matching, isomorphous wild-type apoenzyme crystal structure, providing clear evidence of subtle structural changes in side chains and bound solvent as a result of this single point mutation. Additionally, we observed that the glutamine residue occupies more space in the hydrophobic hinge than the native isoleucine residue (Gridley 2012). We discovered that the subtle changes in this hinge, caused by the Ile260Gln mutation, contribute to the overall phenotype of this pol β variant; beginning with the expulsion of some or all of the water molecules from this hydrophobic region.
prior to full subdomain closure, making the conformational change slower or energetically unfavorable.

The early investigations exploring the importance of residues in the hydrophobic hinge in polymerase fidelity and accuracy identified Ile260 to be important. Starcevic et al. mutated residue I260 to each of the other amino acids. They tested each variant for:

polymerase activity via complementation of *E. coli* pol I in bacterial cells; and a mutator phenotype via Trp$^+$ reversion screen, where a Trp$^+$ reversion occurs only through base substitution errors via dNTP misincorporation. While 6 of 20 amino acids at position 260 resulted in a loss of polymerase activity (D, N, E, R, K and Stop), only 2 of the 14 active polymerases exhibit a mutator phenotype (Gln and His) *in vivo* and *in vitro* (Starcevic 2005b).

The closed complex of I260Q pol $\beta$ bound to dsDNA, ddTTP and two metal-ions illustrates how the Gln260 mutation causes the altered positioning of critical substrate-binding residues, Arg258, Tyr271 and Phe272. Residues 260 and 258 interact more closely in the I260Q ternary complex than in the wild-type (Gridley 2012). As a result, the increased distance between 258 and 272 leaves Tyr271 in an unfavorable position for interacting properly with the primer terminus. In support of our structural results with this dsDNA substrate, we also found wild-type and I260Q polymerases capable of nucleotide incorporation in a primer extension assay with the same substrate (Gridley
The misincorporation and mispair extension profile of this variant was further explored in Chapter 4.

The I260Q polymerase actively misincorporates nucleotides with increased levels of mispair extension as compared to wild-type or the H285D variant. In the misincorporation assay where Mg$^{2+}$ was supplemented with Mn$^{2+}$, the I260Q variant creates more extended product in the presence of lower levels of Mn$^{2+}$ than the wild-type or H285D polymerases (Figure 4.2B). This suggests that I260Q has lower threshold for discrimination, particularly when the alternative metal ion, Mn$^{2+}$, is available. Coupled with our structural results, we find that the Gln at 260 alters the dNTP binding geometry of the active site, such that, the polymerase is less stringent in sampling for the correct nucleotide and is not affected, significantly, by the altered geometry of both an incorrect nucleotide at the active site, but also mispaired nucleotides in the upstream DNA.

An interesting and repeating result in our primer extension and omission assays was the detection of product greater than full-length. Particularly in the presence of Mn$^{2+}$, we observed products longer than 20 nucleotides, most evident when the assay contained all four dNTPs. This could simply be due to the addition of one nucleotide beyond the template in a template-independent fashion. Alternatively, this phenomenon can be attributed to a process called template switching. Template switching describes a process during DNA synthesis where the polymerase switches from the primary template to a secondary template (Odelberg 1995). The retrovirus, HIV-1, relies on
template switching during reverse transcription, increasing the frequency of genomic recombination (Onafuwa-Nuga 2009). Template switching is also referred to as copy-choice with respect to RNA recombination for many single stranded RNA viruses (reviewed in (Lai 1992)). Our hypothesis of template switching by wild-type pol β and the variants we studied is supported by evidence of template switching and the generation of recombinant DNA molecules by other DNA polymerases, namely, DNA pol I from both \textit{E. coli} (Masamune 1971) and \textit{Thermus aquaticus} (Odelberg 1995).

6.3. The gastric-cancer associated pol β variant, E295K: inactive on short gapped DNA, yet active on short dsDNA

E295K is an interesting mutator variant of pol β because of its links to gastric cancer (Iwanaga 1999), DNA binding properties (Iwanaga 1999; Lang 2007) and conflicting activity profiles established in the course of our study (Chapter 4). Additionally, while we determined the structure of the binary complex of E295K with dsDNA substrate, shown in Chapter 3, the structure is lacking strong electron density for the side chain at position 295 (Li 2012). In Chapter 3, we also presented evidence to support a hypothesis explaining the altered activity of this pol β variant.

E295K pol β is said to interfere with accurate BER by associating with DNA more tightly than wild-type, while at the same time, being unable to perform DNA synthesis to complete BER (Li 2012). This hypothesis is supported by the hampered BER activity of
wild-type pol β \textit{in vitro} due to the presence of the E295K variant (Iwanaga 1999). We provide evidence that the geometry of the active site of E295K is slightly changed in the ternary complex; determined by the increased distance between the dNTP-binding Mg\(^{2+}\) and O1\(\alpha\) compared to wild-type (Chapter 3, Li 2012).

The interactions between Arg258 again become important in the mutator phenotype of the E295K variant of pol β. In the wild-type closed, ternary complex of pol β, residues 295 and 258 interact in the form of a hydrogen bond. The positively charged lysine in the mutator polymerase changes the local environment in at least two ways: (i) abrogating the formation of a hydrogen bond with 258 in the ternary complex, and (ii) creating a more positively charged environment in the vicinity of Arg258. We predict that the loss of the 258-295 ion pair creates a ternary complex of E295K that is unable to translocate after the initial round of nucleotidyl transfer during gap DNA synthesis, and, furthermore, the rotation of Arg258 away from Asp192 must pass through a higher energy barrier than in wild-type pol β. This creates a scenario wherein subsequent rounds of incorporation on gapped DNA are hindered due to tight DNA binding and constraints in translocation. This also helps explain the activity of E295K we found on 3’ recessed DNA: pol β binds to dsDNA with less affinity than to gapped DNA, perhaps allowing the flexibility required to facilitate multiple rounds of nucleotidyl incorporation. As Asp192 rotates in response to dNTP and Mg\(^{2+}\)-ion binding, Arg258 rotates towards 260, 296 and 295, Phe272 and Tyr271 fill the space between 258 and 192.
This prevents 258 from influencing 192 and allows 192 to fully interact with Mg$^{2+}$ and participate in catalysis.

The lysine at 295 is loosely packed between Tyr296 and Tyr291, and the DNA template sits above these residues. In terms of side chain flexibility, K295 has a little room to move in the ternary complex, but for the most part is trapped in a depression in the DNA minor groove. As K295 is not likely to attract R258, the steric hindrance between these two residues makes the “flip” of R258 unfavorable in the mutant as compared to wild-type.

The binary complex of E295K depicts a fast-moving residue with ill-defined side chain atoms (Supplemental Figure 3.9.3). Following this observation, we performed molecular dynamics simulations of the fully closed, ternary complex of E295K (Chapter 4). The results indicated that the positively charged lysine is repelled away from Arg258 into the vicinity of the dNTP binding pocket. Our simulations show an overall increase in the distance between K295 and several nearby residues, namely, 258, 271, 279, 283 and 296, and an overall decrease in the distance between K295 and substrate residues: nitrogen atoms of template bases 6, 7 and 8, and the dNTP. Future studies of high-resolution X-ray structures of E295K ternary complexes will further address and help to verify the findings derived from the MD simulations.
In conclusion, the kinetic, structural and molecular dynamics studies of the E295K mutator variant of pol β reveal mechanisms to explain the altered phenotypes observed \textit{in vitro}. Since its discovery in gastric cancer tumor cells (Iwanaga 1999), E295K has been found to: bind to gapped DNA with wild-type-like affinity (Lang 2007); be inactive in BER (Iwanaga 1999); be unable to extend dsDNA or 5 base pair gapped DNA (Lang 2007); inhibit the BER activity of normal pol β (Iwanaga 1999); face an overall higher energy barrier to occupy the fingers-closed conformation of the ternary complex (Li 2012). The work presented in this dissertation adds the following information about E295K: this variant can extend dsDNA to full-length (Figure 4.1B) in the same assay conditions, aside from the dsDNA substrate, from Lang et al. (Lang 2007); can misincorporate dNTPs in the presence of Mg$^{2+}$ and can extend mispairs in the presence of Mn$^{2+}$ (Figure 4.1C & D); has a very flexible side chain with no stable conformation in the binary crystal complex (Supplemental Figure 3.9.3); and is predicted to avoid interacting with Arg258 via repulsion and move closer to the DNA and dNTP in the active site (Figure 4.3B).

\textbf{6.4. H285D is a highly active variant of pol β with a robust mutator phenotype on a variety of substrates}

The H285D variant of pol β was first identified by forcibly pressuring the polymerase to extend beyond the normally chain-terminating nucleoside analog, AZT (Kosa 1999a). The native histidine at 285 makes fairly close contacts with the tryptophan at 325 in the
ternary complex. These interactions may influence the DNA positioning through the fingers subdomain of the polymerase, translating this mutation into the altered phenotype we observe.

The phenotype of H285D is very similar to wild-type pol β. Both polymerases bind to 5-base pair gapped DNA with similar affinities (Murphy 2008) and both can actively extend dsDNA in a primer extension assay (Chapter 4.3.2, Figure 4.1B). Compared to the previously observed lack of mutator and mispair extension activity of H285D on 5-base pair gapped substrate, we detected high levels of fully-extended product in the absence of dATP, the cognate dNTP for the second round of incorporation on a short, 3’-recessed dsDNA substrate (Figure 4.1C & D). These results indicate that H285D pol β’s activity depends on the substrate. The protein-DNA interactions in the gapped complex differ somewhat from the dsDNA complex as expected, because the downstream DNA is stabilized through interactions with the 8 kDa domain and the matching template is stabilized, in part, by α-helix M.

6.5. Concluding remarks and future directions

Several milestones are presented in this dissertation to advance our understanding of DNA polymerases and fidelity via pol β. Pol β has been used as a model DNA polymerase for over 20 years and the contributions to the field include the two-metal-ion mechanism for nucleotidyl transfer, the overall conformational changes of polymerases
that enable efficiency and fidelity, contexts of conserved active site residues, the distinct roles of the two metal ions and the chemistry of catalysis. Here, we add critical information pertaining to the importance of residues distant from the known active site participants in the faithful incorporation and extension of nucleotides by template-directed DNA polymerases.

One amino acid difference, in a region of the polymerase remote of the active site, has the potential to not only influence nucleotide discrimination and primer extension, but also can permit the accumulation of mutations in DNA and propagate human diseases, including cancers. The mutator pol β variants introduced in this dissertation were characterized with structural, kinetic and computational tools to elucidate the mechanisms of their mutator phenotypes.

We presented the high-resolution crystal structure of wild-type pol β: apoenzyme and for the first time, I260Q pol β: apoenzyme and the closed-ternary complex with dsDNA-ddTTP, E295K: binary complex with dsDNA, K72E: apoenzyme and the binary complex with dsDNA. Additionally, we verified the activity of wild-type, I260Q, E295K, K72E and H285D on the dsDNA substrate used for crystallization of the binary and ternary complexes. Additionally, the mutator phenotypes of I260Q and H285D were compared to wild-type in the presence of Mg$^{2+}$ and Mn$^{2+}$. 
Overall, we identified a similar pattern: subtle changes in the vicinity of each amino acid variant leads to small, but significant changes in the interactions with neighboring residues, which, because of their location, is propagated to critical active site residues. I260Q modifies the hydrophobic environment of the palm-finger subdomain hinge. This impacts the conformational change of the fingers subdomain such that residues that interact with the DNA primer and dNTP are mispositioned. In the end, this results in suboptimal geometry at the active site, creating a mutator phenotype. In the future, long range molecular dynamics simulations, transition path sampling and QM/MM simulations should be performed in the context of defined, high-resolution pre- or post-insertion crystal structures to closely inspect the chemistry of catalysis with this variant to better understand how the altered active site geometry specifically contributes to the mutator phenotype. Additionally, the native substrate of pol β being gapped-DNA, it is likely that a high-resolution structure of I260Q-gapped DNA-dNTP will provide additional structural details to explain the effect of this mutation during BER.

The E295K mutator variant of pol β tells a similar story to that of I260Q as they are both residues neighboring the hydrophobic hinge of the polymerase. We found that the charge reversal of this residue prevents any interactions with R258, usually a hydrogen-bond partner in the ternary complex. Furthermore, the positively charged lysine becomes trapped in between neighboring tyrosines and below template strand sugar-base of the DNA. Interestingly, while other groups report that this variant is inactive on gapped DNA in terms of nucleotidyl transfer, we found that with freshly purified
protein and 3′-recessed DNA, this variant is as active as wild-type, I260Q and H285D. With the only experimental difference being the dsDNA substrate used for primer extension, it seems that this protein loses DNA synthesis activity much more rapidly than any of the other pol β variants we have tested thus far. Future studies of this variant are necessary because of the finding of the E295K variant in human gastric cancers. The series of experiments I would suggest include: monitor the decline of primer extension activity over time by repeating the assay each day, or every other day, until no extended primer is detected. The substrate should also be expanded to include variations of dsDNA and gapped DNA similar to previously published protocols. Also, structural changes in the polymerase may be insightful and the full ternary complex with dsDNA and/or gapped-DNA could provide this information. Ideally, the in vitro work should be supplemented with in vivo studies to determine the level of DNA synthesis inside the cell.
References


Mahaney, B. L., Meek, K. and Lees-Miller, S. P. (2010) Repair of ionizing radiation-


Appendices


A.1.1. Description of participation

This report describes a project designed by Dr. Joachim Jaeger. I performed all experiments initially with help from Ryan Wilson and developed alternative protocols myself. Dr. Janice Pata provided additional guidance in method design. This report was written by myself and Dr. Joachim Jaeger.
A.1.2. Summary

The HIV-1 retroviral accessory protein nef assists in facilitating viral replication and disease progression. Nef is post-translationally modified via myristoylation of the N-terminal glycine residues, which facilitates insertion into the lipid bilayer of the host cell. Nef contains three native cysteines, which may be involved in aspects of the alternating conformations observed in the native form. Here, we present an approach by which these conformations can be understood and observed spectroscopically. In developing a variant with site-specific mutations of single amino acids in four locations along the nef protein we created a platform for unique labeling with fluorescent dyes. We also explore the production of the variant proteins and, furthermore, an analysis of the variants via mass spectrometry.
A.1.3. Introduction

The virulence of human immunodeficiency virus (HIV) depends upon the activity of the nine genes encoded within its RNA genome: *gal, pol, env, tat, rev, nef, vif* and *vpu*, the last five yield accessory proteins. *Nef* is situated at the 3' end of the HIV-1 genome and partially overlaps the long terminal repeat (LTR). The *nef* gene encodes an accessory protein, 206 amino acids in length (Terwilliger 1986). This protein is post-translationally modified via phosphorylation of serine residues (Coates 1997) and myristoylation of the stringently conserved glycine residue positioned at the N-terminus (Allan 1985). The self-named *nef*-encoded protein, Nef, is required for efficient HIV-1 replication in primary CD4+ T cells, macrophages and some T cell lines (Miller 1994). Aside from regulating virulence and disease progression, one of the major functional roles of HIV-1 Nef is the down-regulation of CD4+ expression on the surface of T cells. After infection of T cells by HIV-1, Nef binds to the intracellular C-terminal domain of the CD4+/CD28 receptors, of which the N-terminal domains are displayed on the extracellular surface, and pulls them into the T cell, therefore, down-regulating the surface receptors. Nef has also been shown to interact with many other host proteins and, therefore, is an excellent target for drug development and therapeutic intervention. Attenuated strains of HIV-1 lacking functional *nef* gene have been reported to a non-progressive infection of the virus into acquired immune deficiency syndrome (AIDS) (Learmont 1999). The importance of Nef particularly, in terms of developing more highly specific drugs, is the focus of our studies here.
The Sydney Blood Bank studied a small group of people who became infected with the attenuated strain of HIV-1 following blood transfusions from a single source. Because nef has been partially deleted, along with an unusually modified LTR (Gurusinghe 1995), the individuals never present symptoms of AIDS even after nearly 20 years, considering 8 out 13 were determined to be infected with HIV (Gurusinghe 1995). The group did, however, show declined CD4 counts although the progression of this down-regulation was not comparable to typical HIV infections (Gurusinghe 1995).

Until recently, the mechanism by which Nef enhances the infectivity of HIV-1 was not well understood. Qi et al. determined that the host cellular protein, cyclophilin A (CypA), promotes HIV-1 infection; the expression of CypA, coupled to Nef, in HIV-1\(^{\text{nef}}\) (HIV-1 lacking functional Nef) infected cells led to increased infectivity, specifically during particle production (Qi 2008). CypA binds to the gag protein, and, when coupled with Nef, facilitates association of Nef and gag, crucial for enhanced infectivity (Qi 2008).

Using site-directed mutagenesis we have created a variant of HIV-1 Nef that contains only one cysteine residue at position 9 (Figure A.1.1A). The native cysteine residues were conservatively mutated to serines. A single cysteine at position 9 will enable labeling the N-terminus of the protein with a donor fluorophore. The strategic positioning of an acceptor near the core of the folded protein (Figure A.1.1B) will allow us to monitor conformational reorganization via fluorescence changes.
Figure A.1.1. Sequence and structure of HIV-1 Nef and quad-Nef variant

(A) Amino acid sequence of HIV-1 Nef; proposed mutations designed and created in this study are indicated in red (cysteine - to - serine) and blue (serine - to - cysteine). (B) Location of the cysteine residues on the composite model of HIV-1 Nef. The cysteines at positions 55, 142 and 206 are shown as green spheres. The N-terminal anchor domain (NMR model) is shown in blue and the core domain as determined by X-ray crystallography in red. The arrows indicate approximate dimensions within the molecule.
A.1.4. Materials and Methods

A.1.4.1. Bacterial strains and media

The HIV-1 nef gene (strain BH10) was cloned into the expression vector pET23a (Novagen) to generate a pETNef-6His plasmid product (Dennis 2005). The vector was propagated in DH5α cells (genotype, F- Δ lacZMΔ15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (r-, m+) phoA supE44 λ- thi-1 gyrA96 relA1) and expressed in BL21(DE3) cells (genotype, F-ompT hsdSB(rB-mB-) gal dcm (DE3)). Luria broth (LB) and 2xYT medium broth (2xYT) were prepared according to the manufacturer’s instructions (Difco). Ampicillin (Amp) was used at a concentration of 50 μg/mL.

A.1.4.2. Chemicals and reagents

All of the oligonucleotides used in this study were chemically synthesized by Integrated DNA Technologies, Inc. in Coralville, Iowa.

A.1.4.3. Construction of quad-variant of HIV-1 Nef

To create a variant of Nef devoid of cysteine, the site-directed mutagenesis by overlap extension protocol described by Sambrook & Russell was followed using the PicoMaxx High Fidelity PCR System polymerase mixture (Agilent Technologies) (www.genomics.agilent.com/files/manual/600420.pdf). The following F and R (forward and reverse, respectively) primers were used to change the three native cysteine
residues, C55, C142 and C206 to serine and then mutate the native serine at position 9 to a cysteine: C55SF, 5’ - ACA ATG CTG CTA GCG CCT GGC TAG A - 3’; C55SR, 5’ - TCT AGC CAG GCG CTA GCA GTA TTG T - 3’; C142SF, 5’ - ACT GAC CTT TGG ATG GAG CTA CAA GCT AGT AC - 3’; C142SR 5’ - GTA CTA GCT TGT AGC TCC ATC CAA AGG TCA GT - 3’; the primers used to generate S9C and C206S contained BamH1 and NdeI restriction sites as well: S9CF, 5’- GGA GAT ATA CAT ATG GGT GGC AAG TGG TCA AAA AGT TGC GTG GTT GGA TG -3’; C206SR, 5’- CTC GAA TTC GGA TCC TTA GTG ATG GTG TG GTG ATG GCT GTT CTT GAA -3’. The conditions for site-specific mutagenesis by overlap PCR were 20 cycles of: 95°C for 1 min, 50°C for 1 min and 72°C for 2 min, followed by one cycle of: 95°C for 1 min and 72°C for 10 min. Alternatively, QuikChange site-directed mutagenesis (Stratagene—An Agilent Technologies Company) was utilized with similar PCR conditions: 95°C for 30 s for one cycle followed by 18 cycles of 30 s at 95°C, 1 min at 55°C and 1 min at 68°C. PCR products were transformed into DH5α cells and plated onto agar, coated with Amp, then incubated overnight at 37°C prior to colony PCR amplification for sequence verification.

Transformation products were sampled and DNA was amplified with T7 forward and reverse primers using PicoMaxx enzyme (95°C for 2 min, 40 cycles of 1 min at 95°C, 1 min at 50°C and 1 min 20 s at 72°C, and then a final cycle of 1 min at 72°C). An additional control step involved a restriction enzyme digest of colony PCR amplification products with EcoRV: the PCR products were incubated with the restriction enzyme EcoRV in the preferred buffer (New England BioLabs, Inc.) for 60 min at 37°C.
**A.1.4.4. Expression and purification of HIV-1 Nef**

The protocol for expression and purification of non-myristoylated Nef, including our variants of Nef has been described (Dennis 2005). The Nef variants were transformed into competent BL21(DE3) cells, grown in 2xYT, in the presence of Amp, to OD$_{600} = 0.6$-0.8 and then induced overnight at 20°C with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were then harvested via centrifugation and resuspended in 20 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, 200 mM NaCl, 20 mM Imidazole and 6 μg/mL lysozyme, pH 8.0 and lysed by sonication. The protein mixture was purified via liquid chromatography (Aekta Prime; Amersham Biosciences) with a HiTrap Nickel-NTA column (GE Healthcare Bio-Sciences Corp.) [Buffer A: 20 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, 200 mM NaCl, 20 mM Imidazole, pH 8.0; Buffer B: 20 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, 200 mM NaCl, 500 mM Imidazole, pH 8.0]. Fractions containing Nef were identified by electrophoresis on a 12.5% SDS-PAGE gel, and were pooled, then dialyzed against 20 mM Tris, 50 mM NaCl, pH 8.0 prior to storage.

**A.1.5. Results**

**A.1.5.1. Production of HIV-1 Nef variant**

To ensure unique labeling with a single fluorescent dye via a disulfide bond, a Nef variant was created containing only one cysteine. Three native cysteines at positions 55, 142, and 206 were replaced with serines, while a native serine at position 9 was replaced with cysteine. We had already developed two individual Nef protein variants, either
with the C55S or C142S mutant incorporated. The C55S and C142S variants served as template double stranded DNA with which to apply QuikChange Site-Directed Mutagenesis and Site-Specific Mutagenesis Overlap Extension concurrently (see Materials and Methods for sequence of primers).

Products from the second round of PCR via Overlap Extension, after double digest by BamH1 and Nde1, were of predicted length, just over 206 amino acids, or 618 bases long (Figure A.1.2A). Following ligation then transformation, six colonies were collected from C142S transformation (no colonies formed from the C55S dsDNA template Overlap Extension). Upon PCR amplification of the colonies utilizing T7 promoter and terminator sequences, only two were confirmed to contain the recombined plasmid and nef variant (Figure A.1.2B, Lanes 2-7, recombinants: 2 and 4).

Products from QuikChange transformed colonies showed 100% success (Figure A.1.2B, Lanes 8-19). Six colonies were collected and amplified from each of the pair of trials, with C55S as dsDNA template, and C142S dsDNA template. The reactions performed via QuikChange were designed to begin with either C55S or C142S dsDNA template and insert C142S or C55S mutations, respectively.

Due to previous complications resulting in false-positives leading to sequences of empty pET23a plasmid, an additional checkpoint was designed for further validation. Our Nef variants contain a single restriction site for EcoRV while pET23a plasmid has none. A
small sample from the hopeful colony PCR results was digested with the EcoRV restriction enzyme yielding positive recombinants singly cut while leaving empty pET23a plasmid and a negative control, pKK, uncut (Figure A.1.2C).

A.1.5.2. Expression and purification of protein

The solution properties and amino acid composition of all Nef variants should be comparable to wild-type Nef since the residue changes are highly conservative. This should validate the use of the fluorescently labeled Nef variant in in vivo studies. The Nef variants, C142S and C55S/C142S/C206S/S9C were amplified by PCR and cloned into the pET23a bacterial plasmid. Both variants of Nef are non-myristoylated and were transformed into BL21(DE3) cells for high-level expression. C142S and the almost identical quad variant (C55/142/206S/S9C) bound to the nickel column with comparable affinity as indicated by their similar elution profiles (Figure A.1.3A & B). Levels of protein purity were also similar when visualized after electrophoresis separation on a 12.5% SDS PAGE (Figure A.1.3C & D).
Figure A.1.2. Cloning of HIV-1 Nef variants with site-directed mutagenesis by overlap extension

(A) Transformation products were PCR-amplified then treated to a double restriction enzyme digest with BamH1 and NdeI. Lane 1, 1 kB marker, Lane 2, PCR products from C55S dsDNA template, Lane 3, PCR products from C142S dsDNA template. Samples run on a 1% agarose gel prepared with 10 mg/mL ethidium bromide, samples loaded with bromethyl blue loading dye. (B) Site-directed mutagenesis overlap extension
products from C142S are shown in lanes 2-7. QuikChange C55S, lanes 8-13; QuikChange C142S, lanes 14-19. A 1 kB marker was run in lane 1. The PicoMaxx High-Fidelity polymerase was used with T7 primers, forward and reverse, for product amplification. The samples shown were amplified for 40 cycles and were run on a 1% agarose gel prepared with 10 mg/mL ethidium bromide. (C) PCR products were digested with the EcoRV restriction enzyme for one hour at 37°C then loaded onto a 1% agarose gel prepared with 10 mg/mL ethidium bromide. Lane 1, 1 kB marker; 2, Overlap extension colony PCR sample; 3 & 4, C55S template products; 5 & 6, C142S template products; 7, pET23a, untreated control; 8, pET23a, EcoRV digest product; 9, PKK, untreated control; 10, PKK, EcoRV digest product.
A.1.6. Discussion

The single base cloning approach and Nef mutations required a variation of primers, enzymes, ligase and competent cells. As expected, the growth and expression of the C55/142/206S/S9C variant as compared to the C142S variant is comparable (Figure A.1.3C &D). To ensure our non-myristoylated Nef variant would behave as the wild-type non-myristoylated Nef protein, most basic characteristics, expression condition and solution properties would be similar. As was the reasoning behind the cysteine to serine switch: both amino acids are essentially isosteric and polar with side chains displaying largely neutral behavior. The association of nef and the long terminal repeat, as discussed in the nef-defective HIV-1 in the Sydney trials, is apparent when looking at the genome and the structure of Nef as determined by NMR combined with X-ray crystallography (Figure A.1.1A & B).

In developing this variant gene, it is necessary to examine the plasmid into which the DNA is targeted to. It was upon realization of a single EcoRV restriction site only in the insertion, were we able to develop a method for stringently testing the positive results obtained from colony amplification PCR. It is also necessary to underscore pET23a’s single cuts created by the restriction enzymes NdeI and BamH1 (Figure A.1.2A & B). Due to pET23a Amp-resistance, the transformation product was allowed to develop on selective media. The necessity of EcoRV digestion came to light because both empty plasmid and recombinant Nef-harboring plasmid were allowed to grow on the agar, which illustrates the importance of be able to select only for the desired product.
Figure A.1.3. Purification of HIV-1 Nef variants

A

B
Figure A.1.3. Purification of HIV-1 Nef variants

(A) Chromatogram profile of the single amino acid variant of Nef, C142S. Protein eluted from a HiTrap Nickel-NTA Sepharose column via Imidazole gradient (see Materials and Methods). (B) Chromatogram profile of the Nef quad-variant (C55/142/206S, S9C). Imidazole gradient held at 11%, at the first sign of a peak in UV absorbance. (C) 12.5% SDS-PAGE electrophoresis separation of C142S Nickel-NTA Sepharose column fractions. (D) 12.5% SDS-PAGE electrophoresis separation of quad Nef Nickel-NTA Sepharose column fractions.

A.2.1. Description of participation

This studies summarized in this manuscript would not have been possible without the site-directed mutagenesis I performed as described in Appendix 1. The Nef project described in Appendix 1 was handed over to Dr. Muniz, who I trained to perform protein expression and purification of the Nef construct I created. My contribution to writing the manuscript was limited to revisions and sections involving the generation of the quad-mutant of HIV-1 Nef.
A.2.2. Abstract

HIV-1 Nef plays a critical role in viral pathogenicity and disease progression of AIDS as it interacts with many host factors, such as receptors, MHC-type molecules and kinases. The 3D structure of Nef contains a flexible N-terminal region and a compact core domain. The core domain harbors three cysteine residues that are not involved in disulfide bridges. Cysteine-to-serine replacements produces a stable mutant as studied by biophysical methods. Removal of thiols does not grossly alter the molecular properties of the protein. The introduction of new thiol groups allows derivatization for studying the Nef function in solution.
A.2.3. Introduction

Human and Simian Immunodeficiency viruses (SIV) contain a small gene (615 - 785 nucleotides in length) that encodes an auxiliary protein, which is expressed in abundance during early stages of viral replication (Allan 1985; Arhel 2009). Through the study of SIV Nef variants and construction of HIV/SIV chimaeras the protein has been shown to be a major determinant of pathogenicity and disease progression (Kestler 1991; Fackler 2002; Schindler 2006). Nef consists of two separate domains, a N-terminal “anchor” domain and a well-ordered C-terminal “core” domain (Freund 1994; Pandori 1996; Schorr 1996; Welker 1996; Arold 1997; Barnham 1997; Arold 2001). Nuclear magnetic resonance (NMR) studies revealed that the N-terminal anchor domain is largely unstructured and highly flexible (Grzesiek 1996; Lee 1996; Geyer 1999). Nef is post-translationally modified by phosphorylation and the addition of a myristate group at the N-terminus (Allan 1985; Guy 1987; Coates 1997). The N-myristate serves as a membrane anchor directing Nef to the plasma membrane where it can carry out a wide range of interactions and functions (Trono 1995; Welker 1998).

Multiple sequence alignments of the major representatives of Nef (HIV-1, HIV-2, SIV and others) indicate the positions of residues along the polypeptide chain are not conserved (see PFAM: PF00469). This suggests the thiols are presumably not involved in conserved intramolecular disulfide bridges but are located on or near the surface of the protein. Structural studies of HIV and SIV Nef support this finding (Grzesiek 1996; Lee 1996; Arold 1997; Kim 2010). Based on this notion, we have replaced all cysteine
residues in full-length Nef (BH10) with serines, which is most plausible in chemical properties and size (30.4 and $49.4 \times 10^{-24} \text{ cm}^3/\text{residue}$, for hydroxyl and thiol, respectively). Indeed, Hodge and colleagues report that in HIV-2 Nef single cysteine-to-glycine substitutions at positions 28, 32 or 55 drastically reduce covalent dimer formation and thermal stability of the HIV-2 variant \textit{in vitro} (Hodge 1995). We, however, find that in HIV-1 Nef (UniProtKB P03404), replacement of cysteines at positions 55, 142 and 206 by serines produces a stable, monomeric mutant protein as studied by gel filtration, circular dichroism, ultracentrifugation and dynamic light scattering.

\textbf{A.2.4. Materials & Methods}

\textbf{A.2.4.1. Materials}

The coding sequence for the HIV-1 \textit{nef} gene (strain BH10) was cloned into the prokaryotic expression vector pET23a (Novagen) as a NdeI-BamHI fragment to generate plasmid pETNef-6His (Dennis 2005). Chemicals, buffers, enzymes were obtained from Sigma Aldrich (St. Louis, USA), Stratagene (Santa Clara, CA, USA) and New England Biolabs (Ipswich, MA, USA). PCR primers for mutagenesis were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA).

\textbf{A.2.4.2. Site-directed mutagenesis}

See Appendix 1.3.
A.2.4.3. Expression and purification

Mutant Nef was cloned by overlap PCRs using a pET23a plasmid that harbored wild-type BH10 Nef (Dennis 2005). The site directed mutations of Ser9Cys, Cys55Ser, Cys142Ser and Cys206Ser were verified by sequencing. Both wild-type and mutant Nef were expressed by transformation of the appropriate plasmids into BL21 (DE3) cells in LB^Amp media. Expression was induced with 1 mM IPTG when A^600 reached 0.6 and cells harvested by centrifugation 3 hours after induction. Cell pellets were resuspended in 20 mM Na_2HPO_4/NaH_2PO_4, pH 8.0, 200 mM NaCl and 20 mM imidazole and lysed by sonication on ice. The lysate was centrifuged at 34,000 g and the supernatant filtered prior to loading on a pre-equilibrated His-Trap NiNTA column (GE Healthcare, USA). After extensive washing, the protein was eluted using an imidazole gradient (20 – 500 mM) using a liquid chromatography system (Aekta Prime; Amersham Biosciences). The fractions containing Nef were pooled and dialyzed against 50 mM HEPES, pH 8.0, 75 mM NaCl with protease inhibitors (Roche Ltd, CH). A second purification step involved a S70 gel filtration column (GE Healthcare, USA) in phosphate buffer (20 mM Na_2HPO_4/NaH_2PO_4, pH 8.0, 50 mM NaCl plus 1 mM tris(2-carboxyethyl)phosphine (TCEP)). Pure protein fractions were pooled and concentrated to 4 mg/mL (160 μM) using Vivaspin-20 concentrators (5000 MWCO, Sartorius) prior to storage at -20°C. Protein concentrations were determined spectroscopically at 280 nm and adjusted to 1.5 mg/mL (60 μM) on the basis of an extinction coefficient of 48,930 M^-1cm^-1 and a molecular weight of 24 kDa.
A.2.4.4. Circular dichroism measurements

Samples of wild-type and mutant Nef were dialyzed against 20 mM Na₂HPO₄/NaH₂PO₄, pH 7.0, 50 mM NaCl and 1 mM TCEP prior to analysis. CD spectra were recorded at 298 K from 260 to 190 nm at a scan rate of 10 nm/min on a Jasco J 710 CD spectrometer using 0.2 mm path length cuvette. A total of 10 scans were collected for all samples. Data were analyzed using the on-line version of the program suite DICHROWEB (Lobley 2002).

A.2.4.5. Dynamic light scattering and sizing

Samples of wild-type and mutant Nef at 3 mg/mL (120 μM) in 20 mM Tris pH 8.0, 50 mM NaCl with and without 1 mM TCEP were centrifuged at 25,700 g for 5 min in a microfuge and doubly syringe filtered (0.22 μm and 0.1 μm filter) prior to analysis. The samples were analyzed in batch mode in a Wyatt Nanostar™ instrument. Scattering intensities were accumulated using 15 sec acquisition times over temperatures ranging from 20 to 25°C. Sample polydispersity, the apparent hydrodynamic radius and the apparent molecular weight were determined using Wyatt DYNAMICS V.6. software (http://www.wyatt.com/solution/software/dynamics.html).

A.2.4.6. Analytical ultracentrifugation

Analytical ultracentrifugation of wild-type and mutant Nef were carried out with an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) using
absorbance radial scanning optics to measure solute concentration changes during the course of centrifugation. Samples of wild-type and mutant Nef, at 0.5 mg/ml (20 μM), respectively, were dialyzed against 25 mM HEPES, pH 7.5, 75 mM NaCl and 1 mM TECP. The sample and reference buffer volumes of 420 - 450 μL at a concentration of 2 mg/mL (80 μM) were run at 50,000 rpm in 12-mm Epon charcoal-filled double-sector centerpieces. Absorption measurements were made at 280 nm. The samples were at thermal equilibrium (20ºC) before starting the experiments. Scans were taken at a spacing of 0.003 cm in a continuous-scan mode. The data were analyzed by using the c(s) and c(M) methods implemented in SEDFIT (http://www.analyticalultracentrifugation.com). The experimentally calculated sedimentation coefficients from SEDFIT were converted to s20,w values within the same program.

**A.2.5. Results and Discussion**

**A.2.5.1. Purification and gel filtration**

Purification of Nef with Ni NTA affinity chromatography typically shows a single band and less than 90% purity as assayed by SDS-PAGE (data not shown). In addition, size exclusion chromatography (SEC) was employed as a purification step for further solution studies. Using 50 mM HEPES, pH 7.5, 75 mM NaCl and 1 mM TCEP buffer the gel filtration chromatography (GFC) elution profiles for wild-type and mutant Nef were of similar retention time and overall shape (data not shown). Despite the use of
reducing agents (TCEP or dithiothreitol (DTT)), the profiles for wild-type Nef were broader with the appearance of shoulders indicative of higher order oligomers (data not shown). Mutant Nef, however, eluted as a sharp narrow single peak suggesting that the protein is monomeric (data not shown). The calibration of the S70 SEC column confirms both proteins as a monomeric species of about 24 kDa in weight. Therefore, the presence of reducing agents is necessary to maintain Nef as a soluble, monomeric, and homogeneous species.

**A.2.5.2. Circular dichroism**

Circular Dichroism spectroscopy (CD) was used to assess the secondary structure content and the overall folding of wild-type and mutant Nef. CD was carried out with a range of Nef concentrations (0.5 - 1.0 mg/mL; 20 - 40 μM) to optimize the signal to noise ratio. The traces of the CD spectra of wild-type and mutant Nef are similar suggesting the thiol-group containing core domain (residues 55 - 206) retains the same secondary structure content (Figure A.2.1A). NMR studies of the N-terminal anchor domain suggest the structure of the first 57 amino acid residues is mostly random coil (Barnham 1997) and should contribute little to the CD signal in the range from 190 to 260 nm (Saxena 1971). Protein melting experiments based on the CD signal at 222 nm reveal the replacement of cysteine side chains with serines, results in a slightly less stable Nef mutant as compared to wild-type protein (Figure A.2.1B). The T_m experiments were
repeated three times resulting in a standard deviation of ± 0.1°C. The $T_m$ of mutant Nef was calculated to be 47.0°C as compared to 51.5°C for wild-type Nef (Groesch 2007).
Figure A.2.1. Far-UV circular dichroism of Nef proteins

(A) Analysis of secondary structure content of wild-type (gray line) and mutant (black line) Nef recorded at 298 K. The proteins were dialyzed against 20 mM Na₂HPO₄/NaH₂PO₄, pH 7.0, 50 mM NaCl and 1 mM TCEP and concentrations adjusted to 1.5 mg/mL (60 μM). The secondary structures of the two proteins are similar because there are no significant differences between the two spectra. (B) Using far-UV CD the thermal stability of wild-type (gray line) and mutant (black line) Nef were calculated to be 51.5 and 47.0°C, respectively.
A.2.5.3. Dynamic laser light scattering

Dynamic laser light scattering (DLS) was applied to both wild-type and mutant proteins to estimate the mass distribution and oligomeric state of the Nef samples. A high degree of polydispersity often indicates structural flexibility and indicates the presence of undefined oligomers or non-specific aggregates. The measurements were carried out between 1 mg/mL to 4 mg/mL (40 - 160 μM) of Nef to achieve an optimal and steady scattering signal in the DynaPro NanoStar™ (Wyatt Technology). Analysis of baseline error of the correlation function and sum of squares (SOS) error statistics of the individual readings, suggests that in the presence of 1 mM TCEP the mutant Nef (Figure A.1.2) is better behaved than wild-type. Monomodal data analysis of the scattering data (Wyatt DYNAMICS V.6. software) shows the apparent molecular weights of both samples to be 40 kDa indicating an extended, non-spherical but presumably monomeric species (data not shown). Occasionally, with older samples additional species at 156 kDa were detected, presumably indicating the presence of tetramers or higher molecular weight oligomers. These apparent molecular weights are an overestimate of the expected values as Nef has been shown to adopt a flat and non-spherical shape due to the disparate domain structures determined by SEC, X-ray diffraction and NMR spectroscopy. Taken together the DLS data suggest that, as compared to the wild-type protein, the mutant protein is predominantly monomeric, monodisperse and has a reduced tendency to form non-specific aggregates over time. This broadly confirms previous findings with unmodified as well as lipidated forms of wild-type HIV-1 Nef (Geyer 1999; Dennis 2005; Breuer 2006; Hoffmann 2007).
The light scattering measurements were carried out in 20 mM Tris, pH 8.0, 50 mM NaCl with and without 1 mM TCEP at 3 mg/mL (120 μM) protein concentration using a Wyatt Nanostar sizing instrument. (A) Light scattering intensity autocorrelation function for mutant Nef (dark line) showing a steep, simple decay indicative of a small, monodisperse sample. The fitted curve (gray line) corresponds to a spherical molecule of R = 3.8 nm diameter. (B) Error plot between autocorrelation function and fitted curve for the same time regime. (C) Monomodal data analysis using showing a predominantly monodisperse sample of R = 3.8 nm diameter. Note that in the presence of 1 mM TCEP the amount of aggregated mutant Nef protein is negligibly small (R > 90 nm).
A.2.5.4. Analytical ultracentrifugation

For globular, spherical proteins, DLS is a reasonably accurate method for obtaining diffusion coefficients for homogeneous solutes. However, due to the asymmetric and aspherical shape of Nef, sedimentation velocity experiments under a variety of conditions were performed to obtain more reliable estimates of the oligomeric state.

Briefly, we used freshly purified Nef samples in 20 mM HEPES, pH 7.5 and 75 mM NaCl with an excess of TCEP and repeated sedimentation velocity runs at zero, four and seven days after purification. Sedimentation coefficients for both wild-type and mutant Nef proteins were determined by using the c(s) method within the SEDFIT program (Schuck 2000).

Initially, with a fresh protein preparation at day zero, wild-type Nef was a predominantly single species with a sedimentation coefficient of s20,w at 2.2 S corresponding to a molecular weight (MW) of 24 kDa using the c(M) method for a single species within SEDFIT. Over time, additional higher order species corresponding to dimers, tetramers and, presumably, hexadecamers can be observed. This is in line with previous findings published by Dennis et al (Dennis 2005). A similar sedimentation velocity time course for mutant Nef (Day 0, Figure A.2.3A & B), however, shows the protein to remain a single species under the same buffer conditions and TCEP concentrations previously used for wild-type Nef analysis. The sedimentation coefficient s20,w for mutant Nef is 2.1 S indicating the mutant is a monomer (Figure
A.2.3C). By contrast to wild-type, there is little indication of a dimeric or other higher oligomeric species of mutant Nef.
Figure A.2.3. Sedimentation velocity analysis of wild-type and mutant Nef
Figure A.2.3. Sedimentation velocity analysis of wild-type and mutant Nef

(A) A representative analysis of the absorbance at 280 nm of mutant Nef (2 mg/mL (80 μM) in 25 mM HEPES, pH 7.5, 75 mM NaCl and 1 mM TECP) plotted against the radial position as the boundary migrates down the cell. The dots represent the experimental data and the solid lines correspond to the computer-generated best fit of the data to the Lamm equation using SEDFIT. (B) The panel below shows the residuals of the model fit against the experimental data of mutant Nef shown in panel A. (C) Experimentally calculated sedimentation coefficients using SEDFIT. Note that wild-type Nef produces a slightly wider profile presumably indicating lesser homogeneity as compared to the cysteine-to-serine Nef mutant.
A.2.6. Conclusions

We have produced and characterized a mutant form of Nef (BH10) where three cysteines located at positions 55, 142 and 206, have been replaced by serines. Comparison of various representative HIV-1 strains (type O and M) and a range of SIV strains shows that, overall the core domain of Nef is quite well conserved (PFAM00469); positions 55, 142 and 206, however, are not conserved in the multiple alignments. Using SEC, CD thermal melting, DLS and sedimentation velocity studies we were able to show that in vitro, mutant Nef behaves like the parent protein with respect to overall tertiary and quaternary structure. In terms of homogeneity and monodispersity as determined by SEC and DLS, the mutant exceeds the wild-type protein. The melting temperature of the mutant protein, which is lowered by 4.5°C compared to wild-type Nef, is the only less favorable property of the mutant we have found. Since there are no disulfide bonds in wild-type Nef, it is plausible that due to the reduction in volume from a thiol- to a hydroxyl group the side-chain packing around residue 142 could have affect protein stability slightly. A protein melting temperature of 47°C is well within the range typically found for stable mesophilic proteins (Kumar 2001). The other thiol groups in wild-type Nef at positions 55 and 206 are solvent exposed and are likely not contributing to the overall stability of Nef. The introduction of additional thiol groups at selected, crystallographically defined sites opens up possibility of studying the dynamic and oligomeric behavior of Nef in solution in greater detail by using fluorescence spectroscopy.
A.2.7. Acknowledgments

The use of equipment in the Wadsworth Center Biochemistry Core (D. LeMaster, Director) is gratefully acknowledged. We are indebted to all members of the J. Pata Lab for general advice and in particular Ryan Wilson for help with molecular cloning. This work is funded by NIH/NIAID (AI071879).
Appendix 3. Additional Materials and Methods

A.3.1. Bacterial strains

The bacterial strain DH5α (with genotype F- ϕ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (r-, m+) phoA supE44 λ thi1 gyrA96 relA1) was used to amplify the pol β-pET28a+ constructs prior to plasmid purification and subsequent transformation into bacterial expression cell lines. For protein expression, the pol β-pET28+ plasmid constructs were transformed into the E. coli strains: BL21(DE3) (with genotype F- ompT hsdSB(rB-, mB-) gal dcm (DE3)) or Rosetta2 DE3 (with genotype F- ompT hsdSB(rB-, mB-) gal dcm (DE3) pRARE2 (CamR)).

A.3.2. DNA substrates

All oligonucleotides utilized were purchased from Integrated DNA Technologies, Inc. Coralville, Iowa. Substrate for ternary crystal complexes, 2’,3’-Dideoxythymidine 5’-Triphosphate (ddTTP) and 2’,3’-Dideoxycytidine 5’-Triphosphate (ddCTP), were purchased from GE Healthcare. Ultra pure dATP, dCTP, dTTP and dGTP were obtained from Sigma-Aldrich (St. Louis, MO).

A.3.3. Optimizing crystallization conditions

A sparse matrix screen was designed to test a wide range of crystallization conditions with the least amount of protein-DNA consumption. Protein was diluted to various concentrations in 100 mM MES, 10 mM (NH₄)₂SO₄, 30 mM NaCl, pH 6.5. DNA was
annealed and mixed with the protein at a constant ratio of protein:DNA (varies for each sparse matrix screen). Using crystal bridges in sitting drop trays, 2 μL of protein-DNA mix was added to the bridge. While observing the mixture under a light microscope, 0.5 μL of sample well solution was added to the bridge. Total volume of well solution added should be slightly more than would be used in the crystallization conditions (for example, in a 2 μL:2 μL drop, add 2.5 μL total well solution). Based on the observed change to the protein-DNA mixture due to the addition of well solution, including precipitation in the form of clouding of the mixture, the next well solution tested would contain more or less precipitant (PEG3350 in this case). This allowed us to quickly determine what ranges of precipitant to use in our crystallization trials and could be easily repeated for each variant of pol β.

A.3.4. Crystallization of pol β – gapped DNA – ddNTP ternary co-crystals

The natural substrate for pol β is a one to six nucleotide gapped dsDNA duplex. The gap should be flanked with a 3’-OH and a 5’-PO4. We employed a strategy similar to a published protocol that yielded crystals of the human pol β complexed with a 1-nt gapped DNA duplex, with or without soaking in a non-hydrolyzable nucleotide analog (Batra 2008). The crystallization oligonucleotides, template (5’-CG CCG ACG GCG CAT CAG C-3’), 1-nt gap primer (5’-GCT GAT GCG C-3’), 2-nt gap primer (5’-GCT GAT GCG-3’) and downstream (5’/-5Phos/GTC GG-3’) complimented a previous protocol (Sawaya 1997). Template, primer and downstream oligonucleotides were annealed 1:1:1
in 50 mM NaCl, 10 mM Tris, pH 7.5 by heating to 95°C for 2 min, 5 min at 55°C, then cooled to 25°C at 0.5°C per minute (Sawaya 1997). Several crystallization trials were tested: protein concentrations varying from 6-16 mg/mL, DNA concentrations generally were kept at 10% more than protein concentration; well solution components ranged from 10-18% PEG 3350, 155-350 mM Na-Acetate, 50 mM HEPES, pH 7.5, sometimes substituting HEPES with 50 mM Imidazole. Streak seeding was also performed according to several published protocols yielding well-diffracting pol β-gapped DNA complexes (Sawaya 1997; Batra 2006; Batra 2008).