Stony Brook University

Structural Insights into Translesion Bypass Catalyzed by Mouse Polymerase *kappa*

A Dissertation Presented

by

Margaret Jane Luk-Paszyc

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Physiology and Biophysics

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Margaret Jane Luk-Paszyc

We, the dissertation committee for the above candidate for the
Doctor of Philosophy Degree,
Hereby recommend acceptance of this dissertation.

Caroline F. Kisker, Ph.D.
Co-Advisor
Professor, Rudolf Virchow Center for Experimental Biomedicine,
University of Würzburg, Germany
Formerly Associate Professor, Department of Pharmacological Sciences and
Department of Physiology and Biophysics, Stony Brook University

Carlos de los Santos, Ph.D.
Co-advisor
Associate Professor, Department of Pharmacological Sciences and
Department of Physiology and Biophysics, Stony Brook University

Roger A. Johnson, Ph.D.
Chairman
Professor, Department of Physiology and Biophysics, Stony Brook University

Raafat El-Maghrabi, Ph.D.
Research Associate Professor, Department of Physiology and Biophysics, Stony Brook University

Daniel F. Bogenhagen, M.D.
Professor, Department of Pharmacological Sciences, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Structural Insights into Translesion Bypass

Catalyzed by Mouse Polymerase *kappa*

by

Margaret Jane Luk-Paszyc

Doctor of Philosophy

in

Physiology and Biophysics

Stony Brook University

2007

DNA damage is toxic to the cell because a modified nucleotide's chemical structure may lead to replication errors, translational errors, or a stalled replication fork. Specialized Y-family DNA polymerases have the ability to insert nucleotides opposite these DNA lesions, thus rescuing the cell cycle, but at a mutagenic cost. While only one such polymerase exists in archaeal organisms, vertebrates have diversified this family to express four similar Y-family polymerases which evolved to bypass or accommodate different sets of DNA lesions. These four homologues are structurally very similar, but catalytically very distinct from one another. The focus of this dissertation is on the evolutionarily most ubiquitous of these specialized polymerases, polymerase κ (Pol κ) from mouse, which catalyzes error-prone and error-free translesion synthesis across specific DNA lesions. The studies described herein combine biochemical and structural

insights into how mouse polymerase κ is able to select and accommodate certain lesions, but is blocked by others.

This study evaluated the role of several unconserved residues that are located in the vicinity of the active site, and may be involved in protein-DNA interactions. A homology model of mouse Pol κ in complex with DNA illustrated that handling of the incoming nucleotide by the palm and fingers domains is strictly conserved, but all residues in contact with the DNA template upstream and downstream of the active site are virtually unconserved. I first focused on a residue in the Finger domain of mouse Pol κ, Met 134, since this bulky side chain is in a position that could sterically block linked-base lesions such as thymidine dimers. However, this change alone did not sufficiently improve Pol κ activity. The second part of my mutagenesis study uncovered two residues within the DNA major groove binding domain that govern error-prone extension, since mutating them to residues found in error-free Y-family polymerases reduces frameshift mechanisms. During the course of this study, I also uncovered a previously unreported primer 3′-OH slipping event, and this novel and error-prone activity is diminished when the same residues are mutated.

To gain further insight into translesion synthesis, I also solved a structure of mouse Pol κ in complex with DNA. While the apo- and complex structures of human Pol κ have been solved recently, neither structure explains how the polymerase binds to DNA and how a damaged base as large as a benzo[a]-pyrene diol epoxide-dG is delivered into the active site. A Y-family polymerase is structurally analogous to replicative DNA polymerases, but key features responsible for high fidelity replication, such as the 3′-5′ exonuclease domain or the mismatch sensing O-helix, are absent. Our structure portrays a

dimerized Pol κ catalytic core, in which one molecule has bound within the major groove of the DNA substrate. Due to the orientation of the DNA substrate, I believe that this is the structure of a quasi-stable polymerase loading intermediate, and I present a model which suggests how Pol κ initially binds to double-stranded DNA and extrapolate from other Pol κ structures how the damaged base is accommodated for translesion synthesis.

Dedication

I dedicate my doctoral dissertation to my parents, Jerzy and Marysia, and to my beloved Gabor, because their love, encouragement, and patience were the foundation of my success.

Table of Contents

Abstract of the Dissertation iii

List of Abbreviations x

Figure Coloring convention xiii

List of Figures xiv

List of Tables xvi

Acknowledgements xvii

Chapter 1: Main introduction 2

A. DNA Replication and the Necessity of Translesion Synthesis 2

B. Y-family diversity and lesion specificity 5

C. Regulation of Pol κ and Other TLS Polymerases 10

D. Overview of Y-family polymerase structure and function 12

Chapter 2: The Homology Model of Mouse Polymerase κ 17

I. Introduction 17

II. Methods 19

A. Generation of the mouse Pol κ homology model 19

B. Evaluation of resulting homology models 19

III. Results and Discussion 21

A. Evaluation of catalytically essential residues in the palm and finger domains 21

B. Met 134 in the presence of UV damaged DNA 23

C. The unconserved little finger domain 26

D. Little finger residues in complex with damaged DNA 30

Chapter 3: Mutational Analysis and Modulation of Lesion Bypass 33

I. Introduction 33

II. Methods 38

A. Full length mouse Pol κ 38

B. Site Directed Mutagenesis 38

C. Protein over-expression and purification 39

D. Synthesis and purification of oligonucleotide substrates 41

E. Primer extension assays 44

F. Data Analysis 46

III. Results and discussion of cisplatin insertion studies 47

A. Enzyme activity upon undamaged DNA 47

B. Bypass activity of Finger mutants in the presence of cisplatinated DNA 51

C. WT Pol κ activity in the presence of crosslinked DNA 52

D. Discussion 56

IV. Results of εdA extension studies 59

A. Fidelity of WT mouse Pol κ - qualitative comparison 59

B. A novel slippage mechanism by WT polymerase κ 62

C. A different fidelity profile with LF mutants 63

D. Improved error-free extension upon modified DNA with LF mutants 64

E. Mutants show little improvement in dGTP extension after an εdA:dG mispair 67

F. Mutants disrupt novel primer slippage activity 68

G. Frameshift deletion mechanism moderately slowed by mutants 68

H. Discussion 69

Chapter 4: The Binary Complex Structure of Mouse Polymerase κ-I and DNA 81

I. Introduction 81

II. Methods 85

A. Rationale of truncation constructs 85

B. Cloning and protein expression of Pol κ truncation constructs 85

C. Activity of crystallization constructs 89

D. Co-crystallization and data collection of the Pol κ-DNA complex 90

E. Structure determination and refinement of the Pol κ-I/DNA complex 94

III. Results and discussion 99

A. Activity of the Pol κ A and I constructs 99

B. Structure Determination of mouse Pol κ in complex with DNA 101

C. DNA binding mode within Y-family Pols 103

D. The Mouse Pol κ-I dimer 107

E. LF domain coordinates the DNA substrate 110

F. The DNA packs against a selective thumb motif 116

G. Broken "knuckles" 120

H. Final comments 125

Chapter 5: Concluding summary 128

A. Kinetics of lesion bypass 128

B. The binary complex structure of mouse Pol κ-I and DNA 132

Works Cited 134

Appendices 146

List of Abbreviations

2xYT: Luria-Bertani media supplemented with double Yeast Extract and Tryptone

4-OHEN: 4-hydroequilenin, an estrogen metabolite

8-oxo-dG: 7,8-dihydro-8-oxoguanosine

Å: Angstrom (10^{-10} m)

AAF: 2-acetylaminofluorene

Cα: alpha Carbon

AP: apurinic/apyrimidic site, or abasic sites

BPDE: benzo[a]pyrene diol epoxide

CBD: Chitin-binding domain, affinity domain fused to target protein in IMPACT system

Cisplatin: cis-diaminedichloroplatinum(II)

CPD: cyclobutane pyrimidine dimer

Da: Dalton (1/2 of a carbon-12 atom)

DLS: Dynamic Light Scattering

DNA: deoxyribonucleic acid

dATP: 2′-deoxyadenosine triphosphate

dCTP: 2′-deoxycytidine triphosphate

dGTP: 2′-deoxyguanosine triphosphate

dTTP: 2′-deoxythymidine triphosphate

ddNTP: 2′,3-deoxyX phosphate, where X is any nucleotide

D-PAGE: Denaturing (7M urea) polyacrylamide gel electrophoresis

DTT: 1,4-Dithiothreitol (Clelands Reagent)

E. coli: *Escherichia coli*,

εdA: 1,N^6-ethenodeoxyadenosine

EDTA: Ethylenediamine tetraacetic acid

GpG: 1,2 diguanosine cisplatin intrastrand crosslink

HhH: Helix-hairpin-helix motif, known as DNA-binding motifs

IMPACT-CN™: Intein Mediated Purification with an Affinity Chitin-binding Tag

IPTG: Isopropyl-β-D-1-thiogalactopyranoside

k_{cat}: kinetic rate constant, measures substrate turnover number/unit time (minutes)

K_m: substrate concentration at which the velocity of the reaction is half its maximum

LF: Little finger domain

OD_{600}: the spectroscopic absorbance measured at 600 nm, measures cell density

PCR: Polymerase chain reaction

PDB: Protein Data Bank

PEG: Polyethelene glycol

Pol: Polymerase, for clarity, all Pols in this document are DNA Polymerases.

P/T: hybridized primer/template DNA

RE: restriction endonuclease

Rms: root-mean-square

ROS: Reactive oxygen species

S. cerevisiae: *Saccharomyces cerevisiae*, budding (baker's) yeast

S. pombe: *Schizosaccharomyces pombe*, fission yeast

Sac: *Sulfolobus acidocaldarius* (formally known as Sso P1)

Sso: *Sulfolobus solfataricus* P2

SDS-PAGE: Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis

Tg: Thymine glycol

TLS: Translesion synthesis, in Chapter 3.

TLS: Translation/Libration/Screw, in Chapter 4.

Tris-HCl: Trishydroxymethylaminomethane buffer, titrated with hydrochloric acid

WT: wild type

Figure Coloring convention

All figures will be colored in the following manner unless otherwise noted:

Cartoon representations of polymerases	*Ball-and-stick representations*
Palm domain: red	Template DNA strand: cyan
Finger domain: blue	Primer DNA strand: yellow
Thumb Domain: green	Incoming nucleotide: white
Little Finger domain: violet	Nitrogens (when highlighted): blue
Tendril domain: orange	Oxygens (when highlighted): red
N-clasp domain: cyan	Phosphorus: orange
3′→5′ Exonuclease domain: white	

Unless otherwise indicated, all amino acids and secondary structure assignments named in this document are numbered as found in the mouse homologue of DNA Polymerase κ.

List of Figures

Figure 1.1 Conserved Motif/Domain Organization of Y-Family DNA Pols

Figure 1.2 Expression of mouse Pol κ in various mouse tissues

Figure 1.3 Cartoon depictions of replicative DNA polymerases

Figure 2.1 Comparison of nine of the ten catalytically essential residues in the palm and fingers domain between Dpo4 and mouse Pol κ

Figure 2.2 Cartoon representation of the unconserved position Ala 42/Met 134 in the presence of unmodified DNA

Figure 2.3 View of the unconserved position Ala 42/Met 134 in the presence of CPD-modified DNA

Figure 2.4 Organization of the Dpo4 little finger domain

Figure 2.5 Comparison of Dpo4 residues upon the LF domain of human Pol κ

Figure 3.1 Putative DNA adduction pathways

Figure 3.2 Diagram of LF domain mutants

Figure 3.3 Oligonucleotide substrates used in this study

Figure 3.4 Met 134 Gln mutation upon undamaged DNA

Figure 3.5 WT primer extension assays signifying the cisplatin bypass artifact

Figure 3.6 Representative autoradiograph of primer extension

Figure 3.7 Cartoon and stick representation of the linked-base lesion steric gate in human Pol κ

Figure 3.8 Qualitative assessment of WT and mutant extension fidelity after all εdA:dN base pairs

Figure 3.9 Diagram of the three insertion mechanisms after an εdA:dG mispair

Figure 3.10 Comparison of distances between LF domain residues and DNA in Dpo4 and Pol κ

Figure 3.11 Unconserved β-turn in thumb is sensitive to template destabilization

Figure 3.12 Surface representations of human Pol κ and Dpo4 in proximity to the DNA primer strand

Figure 4.1 Cartoon representation of apo human Pol κ

Figure 4.2 Schematic view of mouse Pol κ crystallization constructs including their molecular weight and pI

Figure 4.3 Crystals of the Pol κ-A and Pol κ-I ternary complex

Figure 4.4 Relative extension activities of various Pol κ constructs

Figure 4.5 Domain arrangement and dimerization in the mouse Pol κ-I structure

Figure 4.6 Comparison of mouse and human Pol κ complex structures

Figure 4.7 Association of the two Pol κ-I monomers

Figure 4.8 Superposition of Pol κ palm domain in various stages of DNA binding

Figure 4.9 Structural and enzymological analysis of the DNA in the model

Figure 4.10 Putative location of the Pol κ-I, molecule B LF′ domain and DNA

Figure 4.11 Cartoon and stick representation of the thumb:DNA interaction

Figure 4.12 Putative model for a Pol κ pre-loading complex

Figure 4.13 The "Broken" Finger domain

Figure 5.1 Localization of key protein-DNA interaction sites in Pol κ

List of Tables

Table 1.1 Table of known DNA polymerases

Table 3.1 Kinetic parameters from the cisplatin bypass study

Table 3.2 Kinetic parameters of post-εdA extension with WT and LF mutants in different sequence contexts

Table 4.1 Data collection and Refinement statistics

Acknowledgements

I would like to acknowledge our collaborators on the mouse DNA polymerase κ projects, Drs. Errol Friedberg and Dr. Caixia Guo at the University of Texas Southwestern Medical Center, and Dr. Paula Fischhaber at the University of California, Northridge for providing the mouse Pol κ plasmids and training me to manipulate damaged DNA oligonucleotides, Dr. Holly Miller at Rockefeller University (formerly of Stony Brook University) for invaluable discussions and years of friendship. I would like to thank Drs. Howard Robinson, Frank von Delft, and Annie Heroux for their technical assistance at beamlines X29 and X26C at the National Synchrotron Light Source, Upton, NY. I would also like to recognize Mr. An Lam and Mrs. Liqun Wang for their technical assistance. I would like to thank Dr. Irene Solomon and my mother and her co-workers for their endless supply of gathered cat whiskers.

I thank my graduate program administrator, Mrs. Melanie Bonnette, for rescuing my graduate career more than once. I also thank my committee members, Drs. Roger Johnson, Raafat El-Maghrabi, Carlos de los Santos, and Daniel Bogenhagen for their suggestions and guidance. I am grateful to the Kisker and Schindelin groups, Mr. Kelvin Luther, and Dr. Asim Okur, who became my family away from home. Finally, I whole-heartedly thank my advisors Dr. Caroline Kisker and Dr. Hermann Schindelin, whose guidance and support allowed me to achieve my dream.

.

MAIN INTRODUCTION

Chapter 1: Main introduction

A. DNA Replication and the Necessity of Translesion Synthesis

All organisms must replicate in order to sustain life. Unicellular organisms, such as yeast or bacteria, reproduce by division and therefore need to provide half of all organelles and replicate their genome or the daughter cells may not be properly equipped to survive (2). Multicellular organisms must also reproduce their cells not only to develop to maturity, but also to replace cells lost to metabolic fatigue or injury (2). For example, the epithelial cells that line the stomach must constantly be regenerated due to the caustic hydrochloric acid required for protein digestion and also enzyme activation, completely renewing themselves every 7 days (99).

Genomic integrity is also paramount to the viability of an organism. The cell must copy its entire genetic sequence quickly or populations will not be sufficiently sustained. Most importantly, it must be copied faithfully so that the following generations have the correct set of genetic instructions. Errors in DNA replication such as base misincorporation events may result in inadequate activity levels of essential proteins, which lead to disease, or worse, death. On the other hand, genetic errors that result in improved phenotypes will drive evolution. In either case, DNA replication should be as accurate as possible or risk changing the nature of the organism in an unpredictable way (25).

Every organism possesses DNA polymerases (Pols) that replicate the nucleic acid sequence in a template-directed manner and with high fidelity. DNA Pols can be divided into different families based on sequence homology (43, 97). Summarized in Table 1.1 are all of the known DNA Pols to date.

DNA Polymerase Superfamilies

Family	Prokaryotic	Eukaryotic	Archaea	Viral
A	Pol I	Pol γ, θ, ν		T3, T5, T7 pol
B	Pol II	Pol α, δ, ε, ζ	Pol B I-VI	Adenovirus, HSV, RB69, T4
C	Pol III			
D			Pol D	
X		Pol β, λ, μ		
RT		Telomerase		Reverse transcriptases
Y	Pol IV, V	Pol η, ι, κ Rev1	Dpo4, Dbh	

Table 1.1 Table of known DNA polymerases.
DNA polymerases are organized into one of 7 superfamilies based on conserved consensus sequences. Viral DNA polymerases are organism-specific, and therefore representative DNA polymerases are listed. Table updated from (43, 97).

The cell, and therefore the genome, does not exist in a benign environment. The genome is constantly assaulted by endogenous and exogenous agents which can directly adduct to the nucleotides while genes are in the process of being transcribed or replicated (25). The bases that have been chemically modified (or lost) typically cannot fit (or be recognized) in the replicative Pol active site, which has evolved to interact tightly with the canonical bases and base pairs (31, 58, 59). Therefore, the damaged bases will block the DNA Pol from further processing, stalling the replication machinery (4) and resulting in chromosomal aberrations or apoptosis.

It has been estimated that an organism will undergo 100,000 damaging events per genome per cell per day (25). Fortunately, over 99% of the DNA lesions will be repaired by one of the many repair mechanisms available to each cell, but in times of overwhelming stress, some of these lesions will escape repair before DNA replication is initiated (25). As the number of stalled replication forks increase, the specialized Y-family DNA Pols are recruited to catalyze Translesion Synthesis (TLS) opposite and past these lesions, thereby providing an additional chance at repairing the damage (within the new daughter cell) before the lesion leads to a permanent error. TLS can also be mutagenic (24, 32, 132); since the active sites of Y-family DNA Pols need to be spacious enough to accommodate even bulky lesions such as benzo[a]pyrene diol epoxide adducts or linked base lesions such as thymidine dimers (26). The Y-family polymerases may incorporate the incorrect base opposite the lesion, leading to permanent nucleotide transversions in later replication cycles. However, it is a better option than cell death, and therefore the mechanism has been strictly conserved within every organism (97).

The importance of genomic integrity and TLS is exemplified by the disease Xeroderma Pigmentosum (XP). As a person is exposed to the sun, UV-induced cyclobutane pyrimidine dimers will form in the epidermal DNA, threatening replication if not repaired. Under normal circumstances, the XP nucleotide excision repair (NER) pathway, composed of at least 15 proteins, excises the lesion and several flanking nucleotides, synthesizes new DNA across the gap, and ligates the newly repaired strand. If any one of these enzymes is compromised, thymidine dimers are not repaired, and the epidermis erupts from severe sunburn and blisters (from apoptosis of the UV-damaged cells) within minutes of sun exposure (22, 23). Patients who suffer from XP have a sharply higher risk for developing secondary infections in the short-term and developing skin cancer from the constant injury to the tissue. During the investigations into the etiology of this disease, a small subset of patients was found to carry normal NER genes. Further genetic analysis showed that these patients instead carried a mutation in a novel protein, named XP-V (for Xeroderma Pigmentosum Variant), which later turned out to be the Y-family TLS DNA polymerase Pol η (74). Pol η is capable of catalyzing correct DNA synthesis across UV-induced pyrimidine dimers in order to rescue stalled replication forks, and its loss also results in the XP phenotype (55, 127). This clearly suggests that both the mechanisms of NER and TLS are required to maintain genomic integrity, and that just one efficient pathway is not sufficient to survive UV-induced DNA damage.

B. Y-family diversity and lesion specificity

Y-family DNA Pols are universally distributed from archaea to humans because

of their indispensable role in TLS. They exhibit generally low fidelity and low processivity, and lack 3′→5′ exonuclease activity so that damaged bases can be accepted and bypassed without constantly repairing the potential mismatches that they may create, thereby inhibiting progression (32, 117). They have been organized into 5 subfamilies based on sequence/motif similarity: UmuC/Pol V, Rev1, Rad30B/Pol ι, Rad30A/Pol η and DinB/Pol IV/Pol κ (93). As shown in Figure 1.1, all related proteins contain 5 conserved motifs that create most of the catalytic core (palm, thumb and fingers domain), but differ in sequence in the Little finger domains and in their regulatory C-terminal domains (30).

Each kingdom expresses a different subset of Y-family bypass DNA polymerases. For example, archaebacteria are limited to one such polymerase, Dpo4, a member of the DinB-subfamily that includes Pol κ. Eubacteria contain two Y-family DNA polymerases, DinB/Pol IV and Pol V, which work to promote bypass across different DNA lesions. Lower eukaryotes such as yeast contain several more; Rad30A (Pol η), Rad30B (Pol ι), and Rev1. Interestingly, S. cerevisiae (budding or baker's yeast) expresses a Pol κ polymerase, but S. pombe (fission yeast) does not, suggesting that some of the activities may differ from organism to organism, and some compensatory bypass may be catalyzed by the other Y-family DNA polymerases. Finally, higher eukaryotes express four Y-family polymerases: Pol η, Pol ι, Pol κ, and Rev1 (93, 96, 97). Over the course of evolution, this superfamily has diversified to bypass different DNA modifications by different accommodation mechanisms. Since multicellular organisms contain differentiated tissues that are exposed to different kinds and levels of cellular stress (ultraviolet damage, oxidative damage, alkylating damage, etc), it would be beneficial to

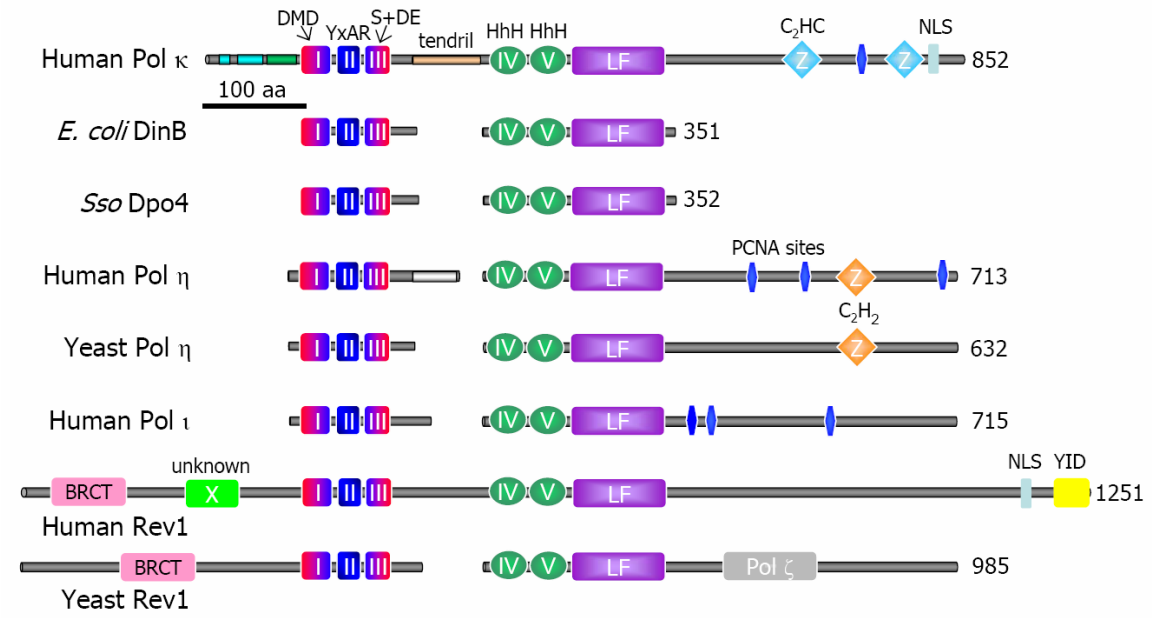


Figure 1.1 Conserved Motif/Domain Organization of Y-Family DNA Polymerases. Y-family DNA polymerases contain 5 conserved motifs that fold into the palm and fingers domains (Motifs I-III), and thumb domain (Helix-hairpin-Helix (HhH) motifs IV-V)(30). Pol κ contains unique sequences/structures not seen in the other Y-family DNA polymerases, such as the N-clasp (residues 1-100) and the tendril domain. BRCT, "BRca-1 C-Terminal" domain; LF, Little Finger domain; NLS, nuclear localization signal; YID, Y-family interaction domain; Pol ζ, interaction domain with Pol ζ. Image updated from Goodman, M.F, Ann. Rev. Biochem. (2002)(30).

control these specialized DNA polymerases selectively. This is in contrast to the unicellular organisms that need to express promiscuously bypass polymerases which accommodate all lesions.

The five Y-family DNA polymerase subfamilies differ in activity, regulation, and expression level. First, as mentioned, UmuC/Pol V proteins are bypass DNA polymerases limited to eubacteria. They consist of one UmuC plus 2 UmuD′ subunits combined with the Pol III holoenzyme, and the complex is able to bypass a wide range of lesions (87, 107). In eukaryotes, Rev1 also creates a complex along with Rev3 and Rev7 to create Pol ζ. Rev1 is considered to be a deoxycytidyl transferase, as it mainly inserts cytosines across most lesions by itself (72), but its ability to incorporate other bases increases when in it forms a complex with its partners. Rev1 can also associate with other Y-family DNA polymerases Pol κ and Pol η to promote extension past lesions in a concerted manner (34, 72, 112).

Pol ι, most likely a gene duplication of Pol η, typically catalyzes the primary incorporation event opposite the lesion, where it then works in concert with Pol κ and Pol η to extend past the lesion (48, 78, 112). However, Pol ι exhibits very low processivity and imposes Hoogsteen base pairing in order to insert nucleotides opposite lesions such as εdA but also upon undamaged DNA (80, 82-84). Bypass by Pol η is error-free with linked-base lesions such as UV photoproducts (47) and cisplatin/oxaliplatin intra-strand crosslinks (113). It is also capable of error-free insertion opposite the commonly miscoding lesion 7,8-dihydro-8-oxoguanosine (8-oxo-dG) (41), or across bulky and exocyclic adducts such as 4-hydroxyequilenin-dA (4-OHEN-dA) (126) and 1,N6-ethenodeoxyadenosine (εdA) (65). These activities are noteworthy considering Pol η's

elevated error rate on undamaged DNA in vitro (~1 in 200 bases (49)), suggesting that the active site has evolved to prefer the sometimes distorting effects of damaged bases. However, it is blocked by certain diastereomers of (+/-)trans-benzo[a]pyrene diol epoxide-dG (BPDE-dG) adducts (98), exhibits lower fidelity with 4-OHEN-dC (104) adducts, and poorly extends past N-acetyl-2-aminofluorene-dG adducts (2-AAF-dG) (73).

Pol κ can be described as the opposite of Pol η. It is the least error-prone TLS polymerase upon undamaged DNA in vitro (~1 in 500 bases), but more error-prone on lesions(91). It can successfully bypass only single-base adducts, including all diastereomers of the carcinogenic (+/-)trans-BPDE-dG adducts (98, 103, 128, 129) and (-)cis-BPDE-dG (103), thymine glycol (Tg)(20) with minor preference for the 5R-diastereomers, and many N2-dG adducts with varying success when considering stereochemistry and sequence context. It is partially blocked by the intercalating (+)cis-BDPE-dG (103), with only 16% of the substrate having a dCTP incorporated opposite the lesion, but is completely blocked by linked-base (40, 124) and intercalating BPDE-dA (98) lesions. It mostly incorporates mismatches or frameshift deletions when bypassing abasic (AP) sites (92), alkylation damage, or other oxidative damage such as 8-oxo-dG or εdA lesions.

Phylogenetically, Pol κ is the most ubiquitous of all Y-family DNA polymerases, including its homologues archaebacterial Dpo4 and bacterial DinB/Pol IV within its subfamily. However, the activities of these lower enzymes are more akin to Pol η in their ability to catalyze error-free bypass (9, 67-69). Recent molecular dynamics studies also suggest that Pol V is very similar in sequence to Pol η (63), making this eukaryotic subfamily more prolifically expressed than previously believed.

C. Regulation of Pol κ and Other TLS Polymerases

Eukaryotic Y-family DNA polymerases are damage-induced enzymes believed to be part of the Rad6/Rad18 epistasis group (6, 52). This pathway is activated during S-phase of the cell cycle in response to stalled replication foci, especially in tissues where exposure to endogenous and exogenous mutagens is more common (6, 7, 52-54). While the induction pathways of Pol η, Pol ι and Rev1 are generally ascribed to be Rad18/Proliferation Cell Nuclear Antigen (PCNA) activated, there is conflicting evidence whether the same is true for Pol κ. It was recently suggested that Pol κ repair of BPDE-damaged DNA is Rad18/PCNA mediated (6), however, Pol κ may only very transiently co-localize with PCNA or Rad18, or the other Y-family DNA polymerases *in vivo* (89). It has also been recently suggested that Pol κ enhances NER, as its presence is required for NER-mediated repair of UV-photoproducts, a surprise considering its inability to bypass linked-base lesions. However, this unexpected role in NER suggests a reason for the dissimilar expression patterns during the cell cycle, as it is believed to assist in the repair synthesis step normally attributed to Pol δ or ε (89, 90).

Within differentiated tissues, the results of northern-blot analysis and in situ hybridization, shown in Figure 1.2, revealed that mouse Pol κ is expressed in tissues that are exposed to higher levels of reactive oxygen species, such as the testes, adrenal cortex, ovaries, lung (in descending amounts), with lower levels observed in the brain, kidneys, heart, stomach, and skeletal muscle (114), in contrast to the constitutively expressed Pol η and Pol ι (77). The testes produce steroidal hormones and also exhibit an exponentially higher rate of DNA replication due to spermatogenesis (29). During normal hormonal biosynthesis, cytochrome P450 generates a large number of superoxide anions and free

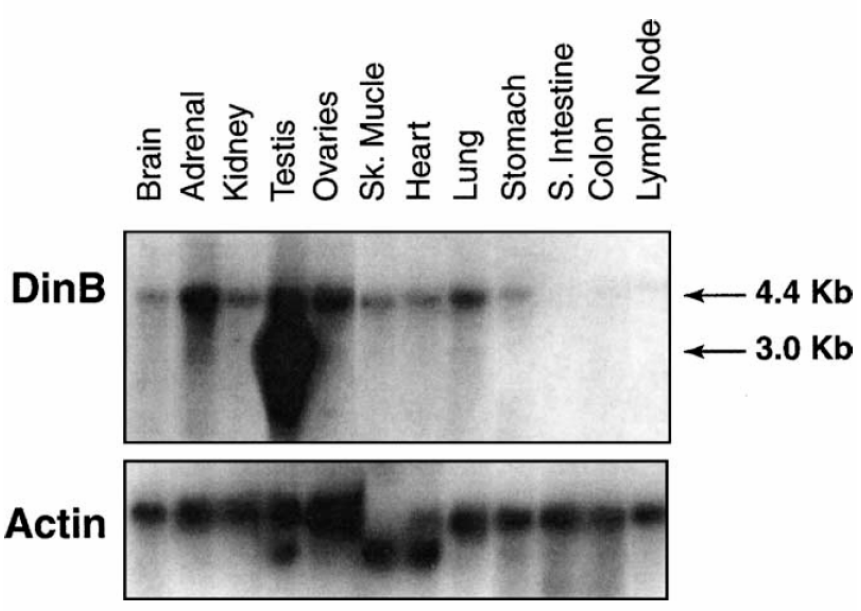


Figure 1.2 Expression of mouse Pol κ in various mouse tissues. Note that the tissues with the highest level of steroidogenesis (and subsequent oxidative damage) are the tissues with the highest Pol κ expression. The secondary band (~3.0 Kb) is one of three isoforms isolated from testes, in addition to full length Pol κ. Image reproduced from Velasco-Miguel *et al*., DNA Repair (2003)(114).

radicals that are typically counteracted by native antioxidant pathways (85, 86). However, in cases of extreme oxidative stress such as higher concentrations of Tg and 8-oxo-dG due to the presence of reactive oxygen species (ROS), the inflammatory response triggers lipid peroxidation resulting in electrophilic α, β-unsaturated crotonaldehyde and malondialdehyde and reactive enals (85, 86) such as 4-hydroxy-2-nonenal (HNE), for example, that attack deoxynucleotides and generate exocyclic etheno-adducts. While Pol κ has the ability to bypass a number of lesions, localization studies suggest that Pol κ prefers oxidatively-derived lesions in vivo (29). Wang et al. have demonstrated that human Pol κ is overexpressed in non-small cell lung cancer tumors (88), but this has been questioned in a recent survey of common cancers (1, 95).

D. Overview of Y-family polymerase structure and function

The Y-family is the latest addition to the DNA Pol superfamily with respect to its identification because of its lack of sequence similarity to classic DNA polymerases; polymerases within the Y-family, however, share considerable sequence homology with each other. Representative structures from 4 of the 5 subfamilies have been solved to date – *Sulfolobus solfataricus* (*Sso*) Dpo4 (66, 67), *Sulfolobus acidocaldarius* (*Sac*) Dbh (100, 133), and Human Pol κ (70) from the DinB/Pol IV subfamily, Yeast Pol η (109) and Rev1 (81), and Human Pol ι (83). As illustrated in Figure 1.3A-C, the structures of Y-family polymerases still adopt the overall structure of a replicative Pol, with a domain arrangement that bears the resemblance to a human right-hand grasping the DNA substrate.

Both classes of DNA polymerases catalyze the same reaction: residues in the vicinity of the active site coordinate the nascent base pair so that the nucleophilic 3′-OH of the primer strand attacks the 5′α-phosphorus of the incoming nucleotide triphosphate in a Mg^{2+}-dependent manner. This reaction leads to the release of inorganic pyrophosphate and extends the primer strand by one nucleotide. The number of individual energetic steps are unknown because the enzyme must judge complementarity of the base pair, as well as negotiate any chemical adducts attached to the nascent template base. Motifs I-III contain the strictly conserved acidic catalytic triad, Asp 106, Asp 197, and Glu 198 in mouse Pol κ, which coordinates two Mg^{2+} ions that orient the triphosphate moiety for nucleophilic attack (28). The Finger domain (blue) facilitates nascent base pair stabilization, and the thumb domain (green) stabilizes the Pol-DNA interaction by fitting its helices along the minor groove of duplex DNA.

Despite its dissimilarity in primary sequence, the palm domain of Pol κ (residues 98-120, 190-213, 290-338; red) and other Y-family Pols share a high degree of structural similarity to all other DNA polymerases including A-family Pols *Bacillus* Pol I (50) and Klenow fragment (*E. coli* Pol I) (21), B-family Pols bacteriophage RB69 (115) and *Thermococcus gorgonarius* DNA Polymerase 1 (42), DNA Pol β (5) from the X-family, and HIV reverse transcriptase (45, 116), and even show similarity to the metal binding sites of adenylyl and guanylyl cyclases (3). It is composed of 3 α-helices which pack against 5 antiparallel β-strands and contains three strictly conserved acidic residues needed to bind two Mg^{2+} ions for polymerase activity. The remaining domains are structurally very different from the rest of the high fidelity DNA polymerases.

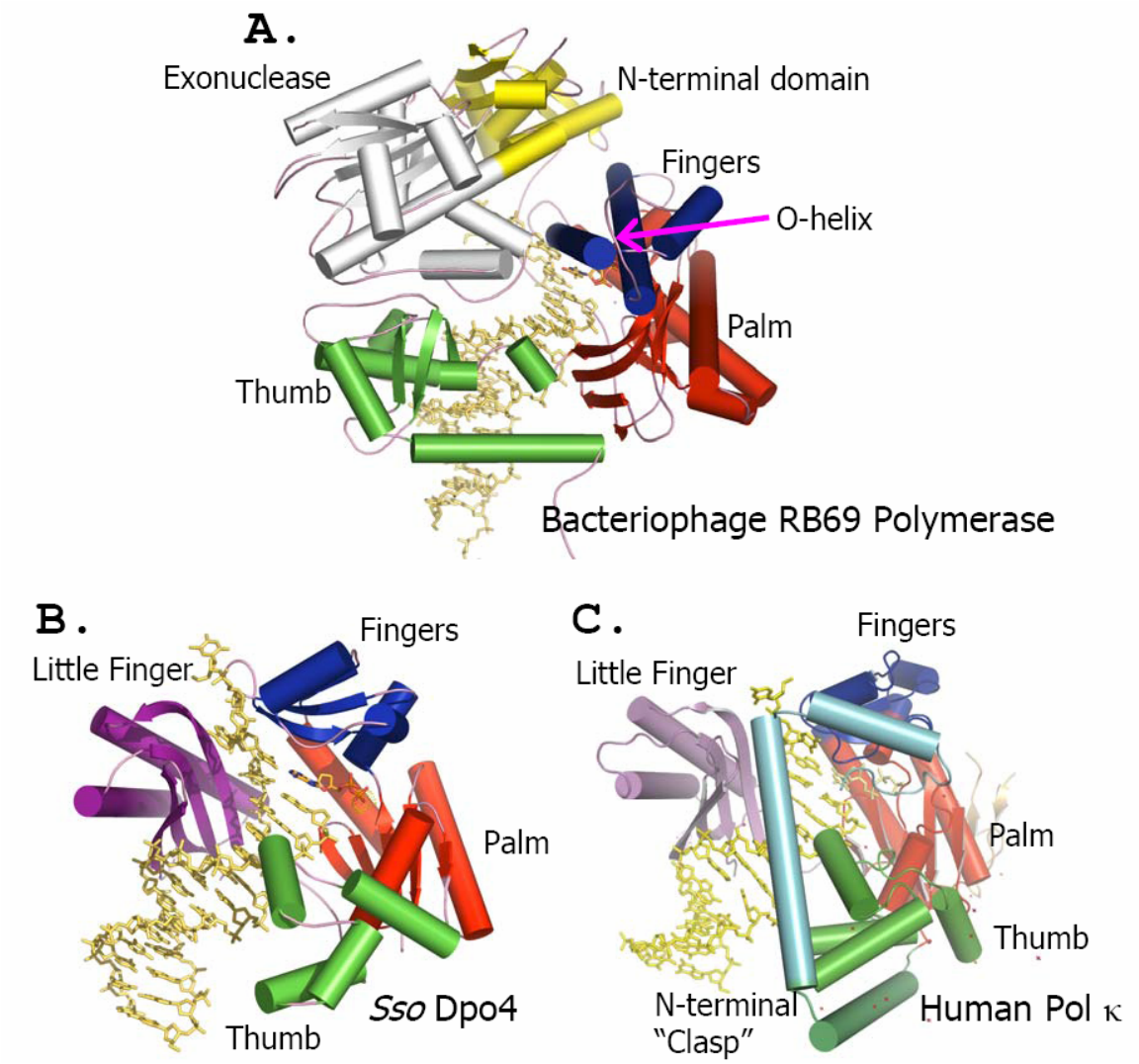


Figure 1.3 Cartoon depictions of replicative DNA polymerases. The palm domains of **(A)** RB69 (PDB entry 1IG9 (115)) and Y-family DNA polymerases **(B)** *S. solfataricus* Dpo4 (PDB entry 1JX4 (67)) and **(C)** human Pol κ (PDB entry 2OH2 (70)) have been superimposed and colored similarly to facilitate comparison of domain arrangement. Replicative DNA polymerases have error-sensing elements such as the 3′→5′ exonuclease domain (white) and the O-helix in the Fingers domain, whereas Y-family Pols do not need to possess such structures for translesion synthesis. Instead, they evolved a "little finger" domain (purple) to increase DNA template strand and major groove interaction. Images are recolored by PyMol using the convention described on page (xii).

Y-family DNA polymerases do not possess an exonuclease domain (Figure 1.3A, white), and the thumb and finger domains are significantly truncated compared with the high-fidelity DNA polymerases (67). These differences prevent the protein from forming extensive interactions with the DNA. The introduction of a novel exclusively Y–family domain, known as the "little finger" (LF) domain (Figure 1.3B, purple) amends this lack of DNA interaction surface. In the Dpo4 structure, Ling *et al*. point out that the LF provides contact with the major groove backbone, while the truncated thumb binds the minor groove (67). This arrangement of domains seems to guide the newly synthesized DNA as it threads out of the active site. The fingers domain is also smaller and lacks tertiary structure analogous to the Watson-Crick base pair checking "O-helix", usually present in high-fidelity polymerases (67).

Nucleotide insertion and O-helix base pair checking is thought to be facilitated by a conformational change of the finger domain towards the palm domain, resulting in a closed or "induced fit" state. It is unclear if the Y-family DNA polymerases adopt the same strategy; however, the structure presented in this dissertation suggests that mouse Pol κ may use an induced-fit mechanism to bind to a damaged template. In other Y-family DNA polymerases, such as *Sac* Dbh, a similar conformational shift does not seem necessary (125), but yeast Pol η may also undergo a conformational change upon substrate binding based on comparisons with T7 Pol and as suggested by pre-steady state kinetic analysis (109, 119). Due to these differences in domain architecture, the active site of the Y-family DNA polymerases are much more open and accommodating to base adducts and the distorting effects that these lesions have on the phosphate backbone of the double-stranded TLS product.

CHAPTER TWO

The Homology Model of Mouse Polymerase κ

Chapter 2: The Homology Model of Mouse Polymerase κ

I. Introduction

The Y-superfamily of DNA Pols has evolved into 5 subfamilies within higher eukaryotes, each targeting specific classes of lesions. While these four enzymes, Pol κ, Pol η, Pol ι, and Rev1, are structurally very similar, they are catalytically very different from one another. Pol κ is capable of bypassing BPDE-dG (7, 103, 128, 129) and Tg (20) adducts correctly, but misincorporates bases opposite 8-oxo-dG (46, 131), 2-AAF (92, 102) and εdA (65). It is also blocked by intercalating lesions such as BPDE-dA (98) and linked-base lesions such as UV-photoproducts and cisplatin intrastrand crosslinks (131). Pol η displays the opposite activity to Pol κ, successfully bypassing BPDE-dA (98), 8-oxo-dG (41), εdA (65), and UV-photoproducts (130), but being stalled on some stereoisomers of (+)BPDE-dG (98, 128). The other two DNA polymerases, Pol ι and Rev1, are far less processive and bypass lesions by primarily employing Hoogsteen (83, 84) or protein-template derived (81) mechanisms, respectively, instead of a mostly processive Watson-Crick base-pairing reaction.

In order to understand how the TLS Pols accommodate template strand lesions and promote (or inhibit) further extension of the growing primer, more information was required regarding the interaction between the polymerase and the DNA. In the absence of a crystal structure, a homology model was generated to evaluate putative protein-DNA contacts within the active site of mouse Pol κ. Since the sequence of Dpo4 is closest to Pol κ, but its activities are more akin to Pol η, I focused on unconserved residues in key positions that might explain some of Pol κ's divergent bypass activities. This chapter describes the results of the homology model that was generated, and discusses the

possible implications of key changes that were observed between Pol κ and Dpo4, two Y-family Pols from the DinB/Pol IV subfamily. The results of this model were used to design mutagenesis studies to evaluate the role of these positions in TLS.

This study also surveyed Pol κ's least understood domain, the Little Finger (LF) domain, unique among the Y-family DNA Pols. It is a ~100-residue domain composed of 4 β-strands (β11, β14, β12, β13 in Pol κ) interwoven in a modified jelly-roll motif, and two amphipathic α-helices (αQ and αR) positioned on one side of the β-sheet, which faces towards the solvent (110). The amino acid sequence of the LF domain is conserved within the subfamily, but varies substantially from subfamily to subfamily. The Dpo4-DNA complex structures demonstrated that the LF domain is positioned evenly in the major groove of the duplex DNA, with the single-stranded DNA template firmly held between the finger and little finger domains (67). Due to this observation, it is clear that both the LF domain and the finger domain play important roles in substrate binding and are most likely involved in lesion discrimination. Also, Boudsocq *et al.* discovered that if the LF domains of Dbh and Dpo4 were switched, each chimera inherited the lesion bypass properties from one another (10). The homology model was therefore generated to gain insight into the importance of individual amino acids with a focus on the finger and LF domains in Pol κ regarding their involvement in the above mentioned functions.

II. Methods

A. Generation of the mouse Pol κ homology model

The mouse Pol κ model was developed prior to the release of the original human Pol κ structure (110). All known Y-family DNA polymerase sequences (n=16) excluding Pol V were aligned in CLUSTALX (108) and manually edited in BIOEDIT (35) using the known Y-family Pol structures available at the time (Appendix A.1). Based on this alignment the sequence of mouse Pol κ (residues 100-514) was threaded manually onto the known structures of Dpo4 in complex with undamaged DNA (1JX4(67), Appendix A.2) and with a cyclobutane pyrimidine dimer (CPD) damaged DNA (1RYR(66)) using SWISS PDB VIEWER(33). The DNA substrates from both Dpo4 structures were then copied into the Pol κ homology model to determine protein-nucleic acid distances.

B. Evaluation of resulting homology models

Since SwissPDBViewer threads the sequence directly onto the carbon-α backbone, the root-mean-square-(rms) deviation value over all α carbons (Cα) is very low (< 0.2 Å). A manual walkthrough of the homology model using the modeling program O revealed several "bumped" residues in buried regions, meaning that several atoms were originally placed within the van der Waals distance of another nearby atom. This observation was not unexpected since Pol κ and Dpo4 only share 27% sequence identity. Since these residues were not located in critical interaction areas nor were they residues that could significantly destabilize a secondary structural element, they were repositioned to minimize the steric occlusions.

Once the apo structure of human Pol κ was released in 2004, I confirmed that our homology model was generally without error, except for β-strand β11 in the little finger domain which was misaligned in sequence by two residues (Thr 420 was erroneously replaced by Ser 422). The sequence of this region was highly variable between the two proteins, explaining the two residue misalignment. However, this did not cause a global shift for the remainder of the protein as the later strands were recognized and aligned correctly. Final secondary structure matching (SSM) superposition (61) of the palm and thumb domains from my homology model and the final mouse Pol κ model revealed an rms deviation of 1.68 Å over all main chain atoms, and rms deviation of 4.2 Å to the main chain atoms of the human ternary complex structure. The larger difference observed between the mouse model and the human Pol κ structure is dominated by differences in position of the LF domain and connecting linker, which are observed to be quite mobile between the apo and ternary complex structures.

III. Results and Discussion

A. Evaluation of catalytically essential residues in the palm and finger domains

Despite its dissimilarity in primary sequence, the palm domain of Pol κ and other Y-family Pols shares a high degree of structural similarity to all other DNA polymerases including A-family Pols *Bacillus* Pol I and Klenow fragment (*E.coli* Pol I), B-family Pols bacteriophage RB69 and *Thermococcus gorgonarius* DNA Polymerases, DNA Pol β from the X-family, and HIV reverse transcriptase. It also shares structural similarity to the metal binding sites of adenylyl and guanylyl cyclases.

As described by Yang in 2003, Dpo4 contains 10 catalytically essential residues required for polymerization activity(125). These residues, shown in Figure 2.1, are required for: (a) divalent cation coordination and subsequent polymerization, the catalytic triad Asp 7, Asp 105, and Asp 106 in the palm domain; (b) incoming nucleotide orientation, Tyr 48, Arg 51, in the fingers domain and Lys 159 in the palm domain; and (c) four residues in the "roof" identical between Dpo4 and mouse Pol κ, although other the residues in the roof region are not completely conserved. In Figure 2.1, the residues of interest of Pol κ overlap as expected of the active site in the fingers domain, Val 32, Ala 42, Ala 44 (not visible), and Gly 58. Out of these residues, the catalytic triad and the nucleotide orientation residues are perfectly aligned with those residues found in Dpo4 since it was used to create the model, therefore only the divergent Pol κ residues are visible in pink. Among the roof residues, most of the changes are fairly conservative. Ala 44 corresponds to Ala 136 in Pol κ, but is completely obscured in the image. Val 32 is replaced by Ser 131 in Pol κ, however, its position in the model is unknown since the alignment diverges to a

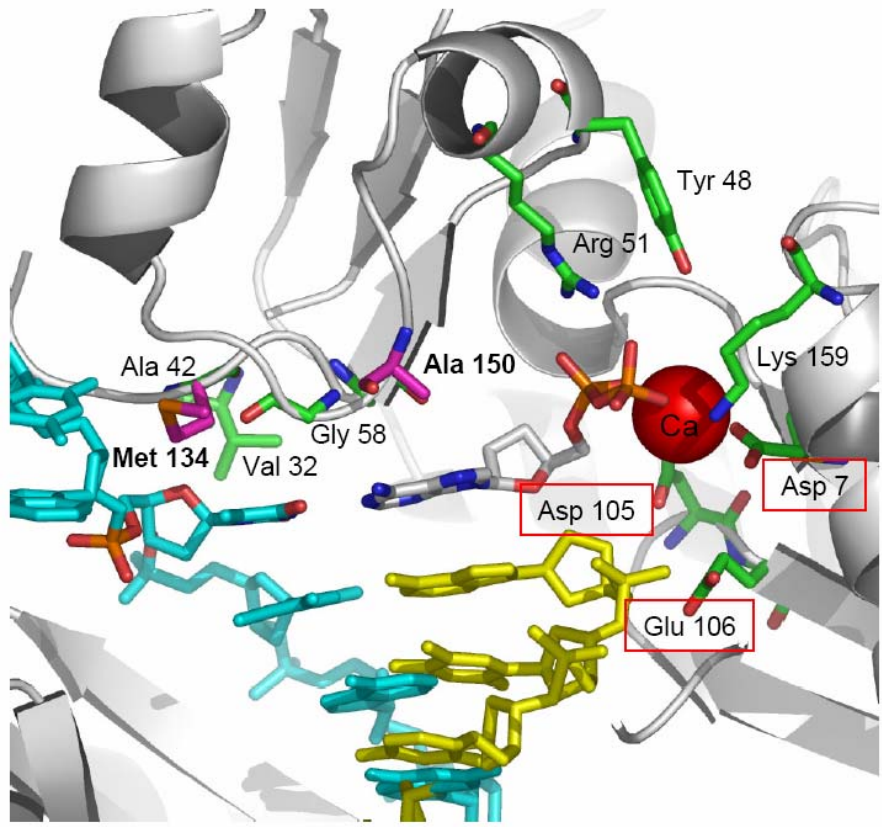


Figure 2.1 Comparison of nine of the ten catalytically essential residues in the palm and fingers domain between Dpo4 and mouse Pol κ. Dpo4 residues (green) overlap with the Pol κ (pink) residues perfectly, such that only the divergent Pol κ residues, Met 134 and Ala 150, are visible (bold). The red boxed labels indicate the catalytic triad, actively coordinating only one Ca^{2+} atom (red), since in this particular structure the γ-phosphate was already hydrolyzed. Remaining objects are colored as described on page (xii).

structurally unconserved loop. Capping the "roof" at the front of the active site base is Gly 58, which is replaced by Ala 150 in mouse Pol κ, creating a minor edge that covers the active site and may provide added guidance to the base-pairing reaction.

In Figure 2.1, an interesting change is seen further away from the active site, a Dpo4 Ala 42 → Pol κ Met 134 exchange. Ala 42 is located directly at a point through which all bases must pass in order to gain access to the active site (close-up, Figure 2.2A,B). The template base upstream from the active site (henceforth called the T_{+1} position) is rotated by 90°, such that only one nucleotide is presented into the active site (T_0 position) for unambiguous base pairing. Pol κ, on the other hand, has a bulkier methionine residue in this position. Although the longer side chain may cover more of the phosphate backbone (see Figure 2.2C,D), it should not impose any major restrictions to single bases since their movement into the active site does not seem to be impeded by any significant steric clashes. However, larger moieties or inflexible backbones, as seen with UV photoproducts, may be excluded suggesting that this residue may be one factor why Pol κ does not perform TLS on linked base lesions.

B. Met 134 in the presence of UV damaged DNA

The sequence of mouse Pol κ was also threaded onto the structure of Dpo4 in complex with CPD-damaged DNA to analyze possible effects of the modified DNA in proximity to the unconserved amino acids (close-up, Figure 2.3). Ala 42 in Dpo4 provides enough space for the CPD to enter the active site, however, even the most favorable of the putative Met 134 rotamers in mouse Pol κ positioned the side chain no further than 1.2 Å from the 5′T of the CPD. This suggests that having a methionine in this

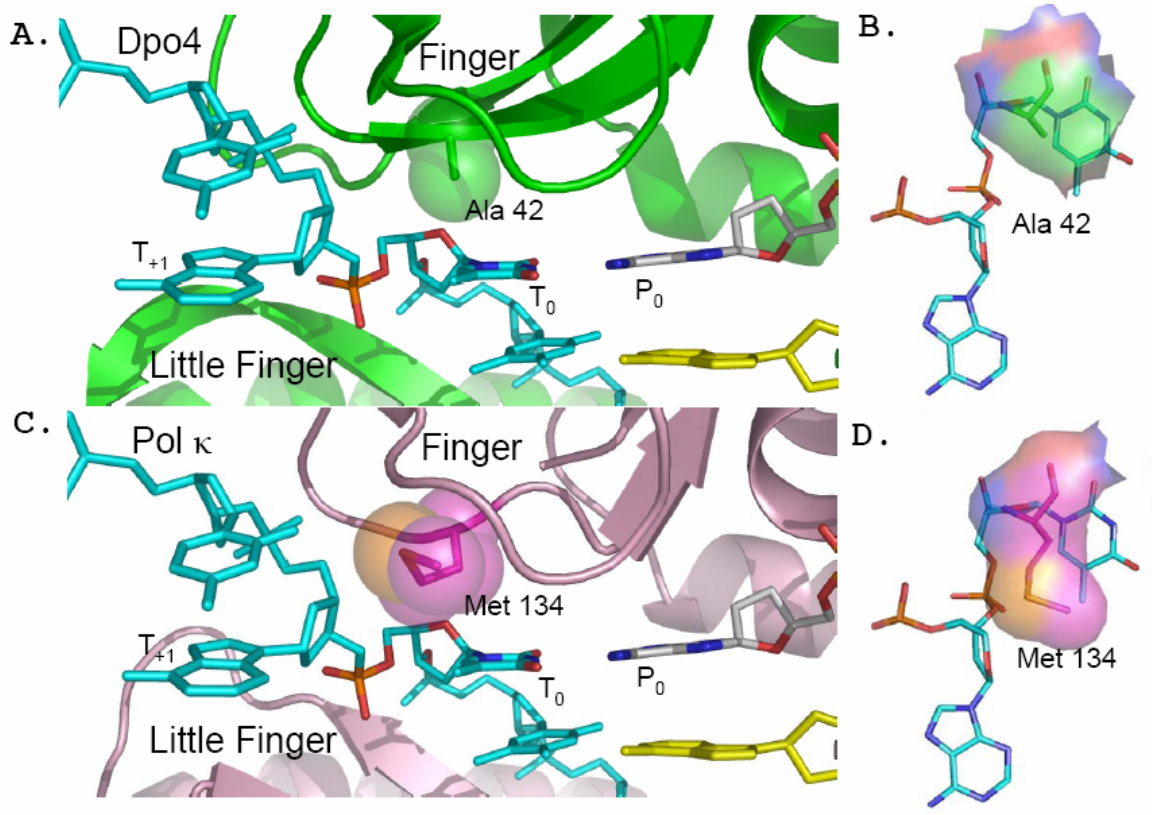


Figure 2.2 Cartoon representation of the unconserved position Ala 42/Met 134 in the presence of unmodified DNA. **(A)** In Dpo4 (67), Ala 42, depicted with its van der Waals radii, is directly positioned above the active site template base. An alanine in this position provides sufficient space for the movement of an individual base into the active site **(B)** Top view of Ala 42. **(C)** In Pol κ, Met 134, also depicted with its van der Waals spheres, occupies the same position as Ala 42 in Dpo4, promoting increased DNA substrate contact. **(D)** Top view of Met 134. Note that Met 134 also overlaps with the T_0 phosphate group, unlike the Ala in Dpo4. In all views, the Ala or Met are depicted in ball-and-stick representation and space-filling models to highlight their size. Bases in the active site are emphasized by being colored by atom as described on page (xii).

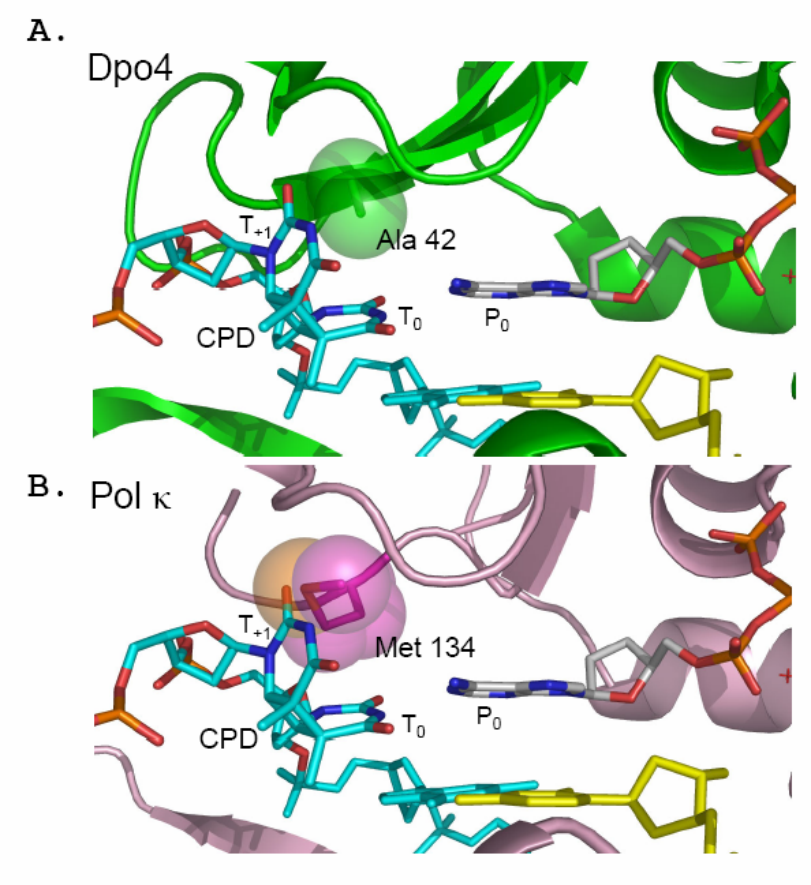


Figure 2.3 View of the unconserved position Ala 42/Met 134 in the presence of CPD-modified DNA. (A) Ala 42 in Dpo4 (67) provides enough room for the CPD to be accommodated, while **(B)** in Pol κ, the CPD is too large and rigid to proceed into the active site if Met 134 is present. Though the model depicts here that the 3′T is in the active site, it is unlikely that the CPD could be accommodated in this manner due to the inflexible backbone and the Met gate. Coloring and amino acid representation are the same as in the previous figure.

position is too large to accommodate the linked-base lesion. It has been reported that Pol κ cannot bypass CPDs or similar linked-base lesions, stalling one base before the 3′ end of the lesion(131).. Considering this steric hindrance imposed by the methionine, it is unlikely that the 3′T would enter the active site as shown in this model, since the phosphate backbone is highly distorted and inflexible, thus making it impossible for the 3′T to enter the active site by itself. It is therefore plausible to suggest that Met 134 is a single base steric gate, responsible for selecting only single bases to enter the active site. The increased contact of Met 134 with the DNA may also force distorted phosphates or damaged bases to adopt a conformation similar to undamaged DNA, directing proper presentation into the active site for base pairing.

C. The unconserved little finger domain

After the identification of putative residues in the fingers domain, the focus was placed on the alternate side of the T_{+1}/T_0 base interaction, the LF domain. Based on the Dpo4 structure, the sugar-phosphate backbones of the primer and template strands straddle the 4 β-strand LF β-sheet, mimicking 5^{th} and 6^{th} strands for this domain. As illustrated in Figure 2.4, the DNA phosphate backbones slide along the outermost β-strands β11 and β13, keeping the proper dimensions of the major groove. Interestingly, the phosphate groups are coordinated by the peptide bond amino groups of the β9 backbone (equivalent to Pol κ's outermost β11 strand) rather than through individual side chains. While this β-strand contains residues with short side chains facing the DNA substrate (with the exception of one Arg, which is positioned behind the phosphate

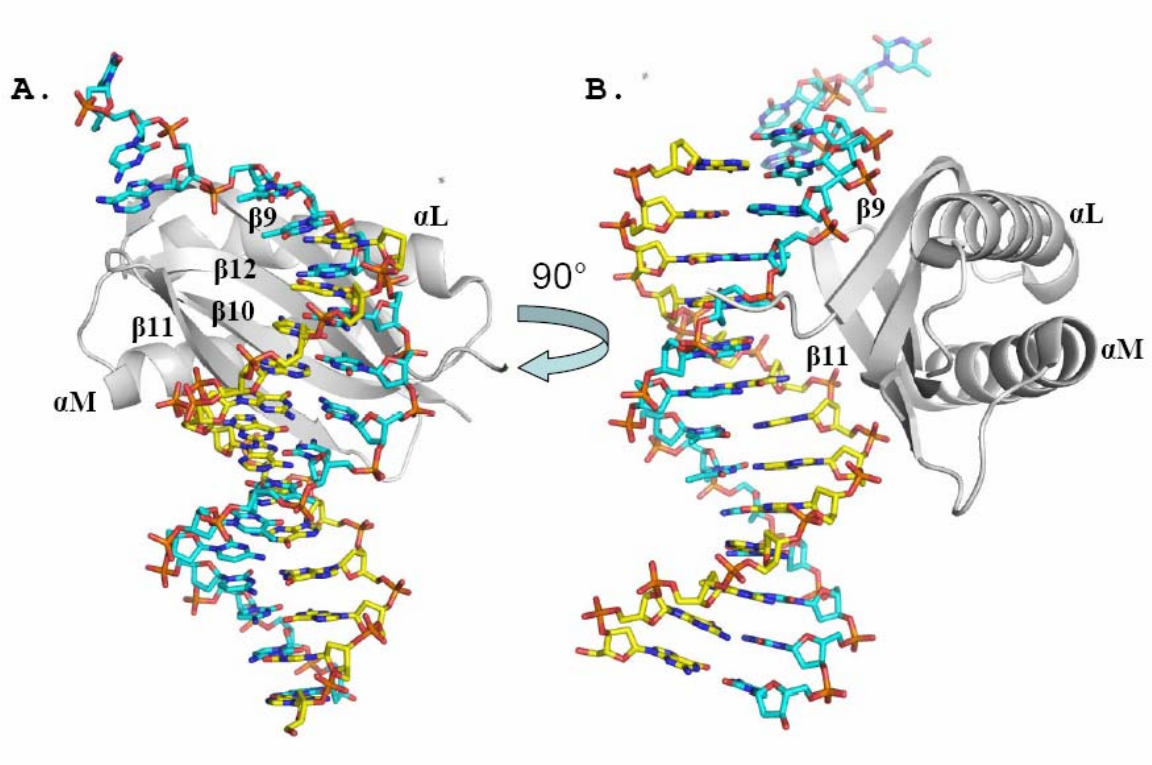


Figure 2.4 Organization of the Dpo4 little finger domain. Cartoon representation of *Sso* Dpo4 (PDB entry 1JX4(67)) with undamaged DNA substrate (sticks) bound to the outermost β-strands of the LF domain. **(A)** LF domain viewed looking upon the β-sheet. **(B)** Side view parallel to the β-sheet. Note how the phosphate groups of both the primer and template strands are coordinated as if they are a continuation of the LF β-sheet(67). DNA colors are as described on page (xii). Image recolored using PyMol.

backbone), the neighboring β14 strand is comprised of longer residues, Arg 331 and Arg 332, which fill out the concave surface of the β-sheet. This assortment of residues surrounding the T_{+1} and T_0 sites creates a rich interaction surface that supports the template base through the major groove, opposite from the previously discussed finger residues, stabilizing and guiding it correctly into the active site.

The Pol κ homology model reveals some interesting differences compared to the Dpo4 structure. While the sequence of the Pol κ β11 strand is slightly different, its short side chains still maintain the same function as in Dpo4. However, mouse Pol κ contains an Arg 332 (Dpo4) → Leu 507 substitution in strand β14. As shown in Figure 2.5, the bulky positive charge of Arg 332, which is located 2.8 Å away from the nascent template base phosphate major groove oxygen (O2), is now replaced by a shorter hydrophobic Leu with its terminal δ carbon-2 located 3.5 Å away from the same atom. This changes the local surface topology and the charge distribution, and therefore may affect the base stability in this critical interaction surface. The local positive charge is found instead on the third strand β12 at position Lys 460 in Pol κ, a position that is closer to the T_{-1} base according to our model, leaving only the conserved Arg 506 of mouse Pol κ (analogous to Arg 331 in Dpo4) to coordinate to the T_0 phosphate group. According to the various conformational rotamers, the guanidinium moiety of Dpo4's arginine and the terminal amino group of lysine in Pol κ can occupy the same space, but their intersection occurs away from the template phosphate groups, making it unlikely that Lys 460 can reach across to perform the same function as Arg 332. Considering the putative increase in flexibility for the template strand in this region due to the shifted positive charge, this change may at least partially explain the error-prone activities exhibited by Pol κ.

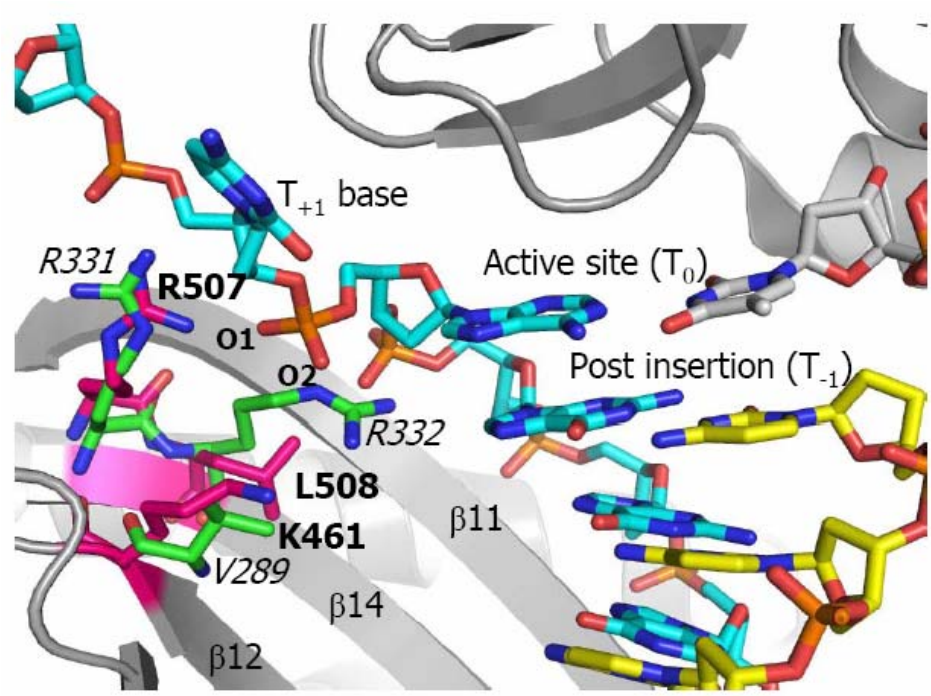


Figure 2.5 Comparison of human Pol κ and Dpo4 residues in the LF domain of human Pol κ. Several residues that interact with the T_0 and T_{-1} phosphates are different between the two proteins. Dpo4 (green, italicized) contains a positively charged residue Arg 332 that forms a salt bridge to the O2 atom of T_0 phosphate group while this interaction may not be possible in Pol κ (pink) where Arg 332 is replaced by Leu 507. The thumb domain has been removed for clarity.

D. Little finger residues in complex with damaged DNA

The importance of Arg 332 is emphasized in the structure of Dpo4 in complex with CPD-damaged DNA. The distorting effects of the thymine dimer impact both the finger domain interface as well as the LF interface. Due to the coupled nature of the CPD, the phosphate groups of the thymine dimer distort sharply as a result of the cyclobutane ring, causing an entirely constrained unit passing into the active site. Between the unmodified and UV-damaged DNA structures, the T_{-1} phosphate (phosphorus atom) moves over 3.5Å away from the finger domain and Ala 42, the T_0 phosphate (3′ CPD thymidine) is shifted 2.9 Å towards the finger domain, while the T_{+1} phosphate (5′ CPD thymidine) is most displaced, being 8.5 Å out of position and extends over the LF domain rather than projecting straight out between the finger and LF domains. This is in sharp contrast to the straight line of phosphates as depicted in Figure 2.4.

The LF finger residues respond to this distortion with their long side chains to accommodate the phosphates. Arg 332, which previously was oriented parallel to β-strand β12 (mouse Pol κ β14) in the Dpo4 structure containing undamaged DNA is displaced by the T_{-1} phosphate, and therefore points straight up to maintain its contact between the T_o and T_{-1} phosphate groups of the nascent base pair. Arg 331 does not show significant displacement other than switching the conformation of its guanidinium group towards the T_{+1} phosphate. Nonetheless, it still forms salt bridges with the major groove phosphate O2, thus aligning and handling the DNA substrate along the same side. An equivalent contact is not possible in Pol κ since Leu 507 is much shorter and cannot form a similar interaction. This observation implies that the minimum surface between the protein and the DNA is preserved, but there is no electrostatically-derived stability for

putative distorted phosphates as observed with the larger arginine in Dpo4. Lys 460 in the neighboring β-strand, may be too far away to impose any stabilizing influence on the T_0 base. Therefore, Pol κ's unique bypass activities upon damaged DNA could also be derived from the lack of contacts to the LF domain, since it is the only domain that interacts with the template DNA backbone and therefore accountable for its stability

CHAPTER THREE

Mutational Analysis and Modulation of Lesion Bypass

Chapter 3: Mutational Analysis and Modulation of Lesion Bypass

I. Introduction

Through evolution, the Y-family Pols have differentiated into 5 distinct subfamilies which accommodate and target different classes of DNA lesions (93). However, while incoming nucleotide binding and polymerization to the 3′ primer terminus have been conserved, other DNA contacts, especially in the LF domain, vary extensively between subfamilies.

In the previous chapter, amino acids potentially responsible for lesion recognition and accommodation in Pol κ were identified. As shown in Figures 2.2C and 2.5, these residues, Met 134 in the fingers domain, and Leu 507 and Lys 460 in the LF domain, create a surface in between the T_0 and T_{+1} template bases. In this strategic position, these residues may obstruct specific DNA lesions from entering the active site, either by sterically hindering the extra base moieties directly or the distorted phosphate backbone. The LF residues also coordinate the phosphate group of the template nucleotide in the active site, suggesting a continued stabilization of the base once it passes this gate. In order to assess the role that these three residues have in Pol κ's lesion selectivity, kinetic studies were conducted on point mutants of these three amino acids and on the wild type protein.

The first part of this chapter investigates the role of the fingers domain residue Met 134 with respect to translesion incorporation, which I established in my modeling study as a potential single base steric gate. Translesion insertion studies were performed in the presence of the diguanosine adduct, (1,2-GpG) cisplatin intrastrand crosslink (Figure 3.1A), an adduct that links two adjacent purines and is created by the anti-tumor

drug cis-diaminedichloroplatinum(II). Cisplatin therapy has proven to be a potent anti-tumor drug, due to its ability to adduct to DNA and stall replicative DNA Pols in rapidly reproducing tumor cells. However, in some patients, cisplatin treatment became less effective with continued administration. Later studies confirmed that a combination of upregulated TLS, nucleotide excision repair, and recombination repair mechanisms bypass/repair the crosslinks, thereby rescuing replication in these tumors (123). While it is the Y-family Pol η that bypasses cisplatin adducts without error *in vivo*, obtaining insight into the structural basis of linked-base adduct exclusion from the Pol κ active site is pertinent to understand how Pol κ selects DNA lesions for translesion synthesis. Mouse Pol κ mutants Met134Ala ("Ala-mutant") and Met134Gln ("Gln-mutant") were generated to alter the size of the presumed gate in order to evaluate the role of this position with respect to blocking linked-base lesions.

The second part of this chapter investigates post-lesion extension with respect to the LF domain. Primer extension assays were performed in the presence of $1,N^6$-ethenodeoxyadenosine (εdA, Figure 3.1B), an exocyclic adduct with numerous endogenous and exogenous sources that is readily extended by Pol κ from misaligned primers (40, 65, 118, 122), i.e. leading to an error-prone extension. In contrast, Pol η is able to bypass and extend past εdA in an error-free manner. These different activities suggest that some substrate destabilization is present in the active site of Pol κ, which favors extrusion of the template base and subsequent realignment of the primer terminus. The homology model identified two unconserved residues near the active site, Leu 507 and Lys 460, whose side chains do not form essential template DNA backbone contacts as compared to Dpo4, suggesting a potential source of the template strand destabilization.

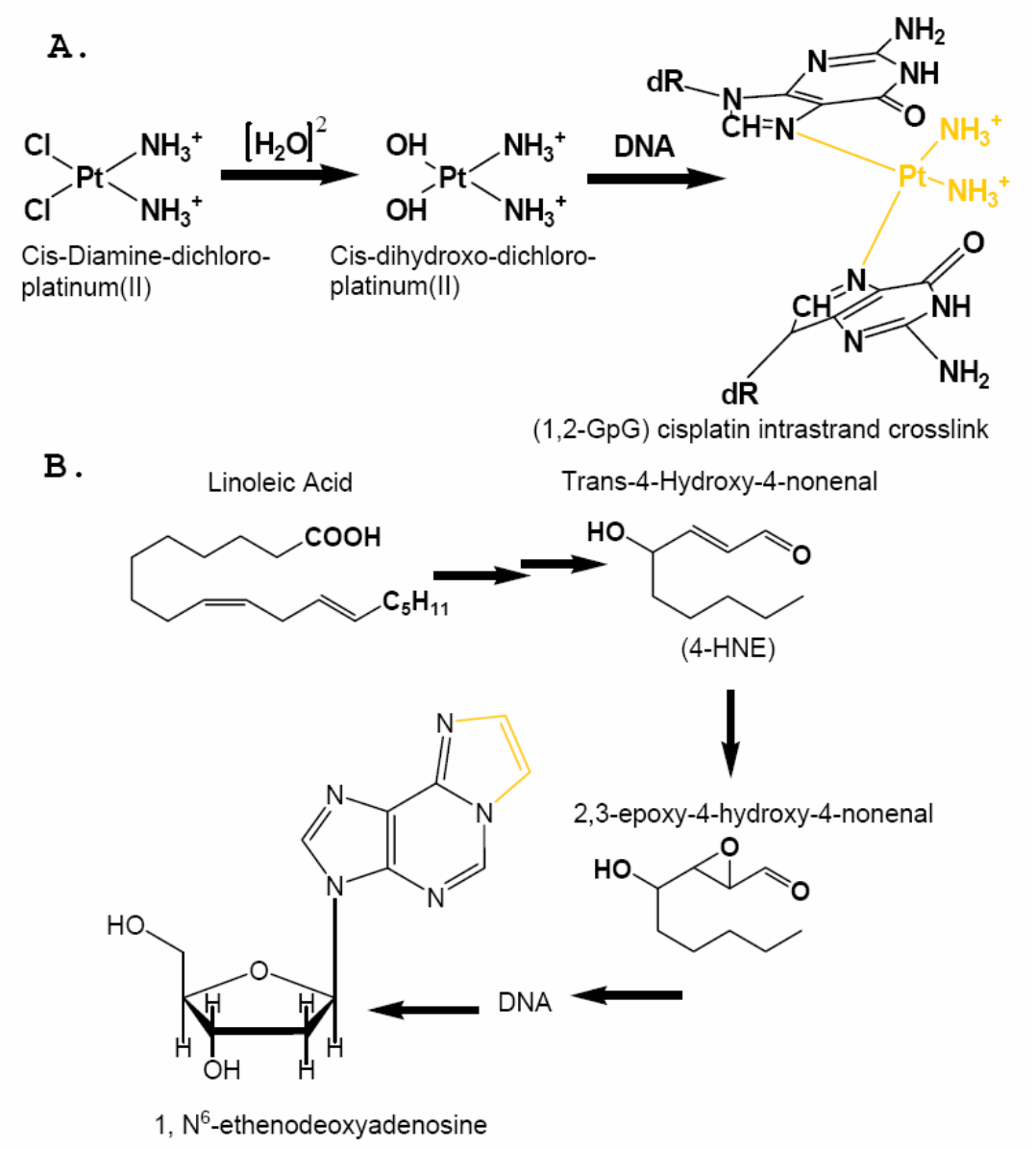


Figure 3.1. Putative DNA adduction pathways. (A) cisplatin intrastrand crosslinks (GpG); crosslinks occur in the major groove and generally do not perturb standard B-DNA geometry. **(B)** 1,N^6-ethenodeoxyadenosine (εdA) adducts; the exocyclic ring typically promotes syn/anti Hoogsteen base pairs, though hydrogen bonds can be created with all four nucleotides. The chemical adduct is highlighted in orange.

To study the impact of these LF residues with respect to Pol κ frameshift propensity, four mutants were generated (Figure 3.2) to evaluate the importance of the residues found on neighboring β-strands β12 and β14: an Open Mutant (mPol κ K460V/L507V) which create an extended concave uncharged pocket in the vicinity of the DNA template backbone, a Met-mutant (mPol κ K460V/L507M) which removes the Lys on β12 again and conservatively replaces the Leu with a longer but uncharged Met, a Closed Mutant (mPol κ K460F/L507M) that introduces not only the Met on β14, but a bulky Phe in place of the Lys on β12 resulting in a more constricted but hydrophobic surface, and lastly, a Dpo4-Revert Mutant (K460V/L507R), which mimics the pocket as found in *Sso* Dpo4. This can also be considered a close relative to yeast Pol η, which has a Lys in place of the Arg on strand β14. We thus explore the necessity of surface charge and topology in this potentially crucial region near the template DNA backbone.

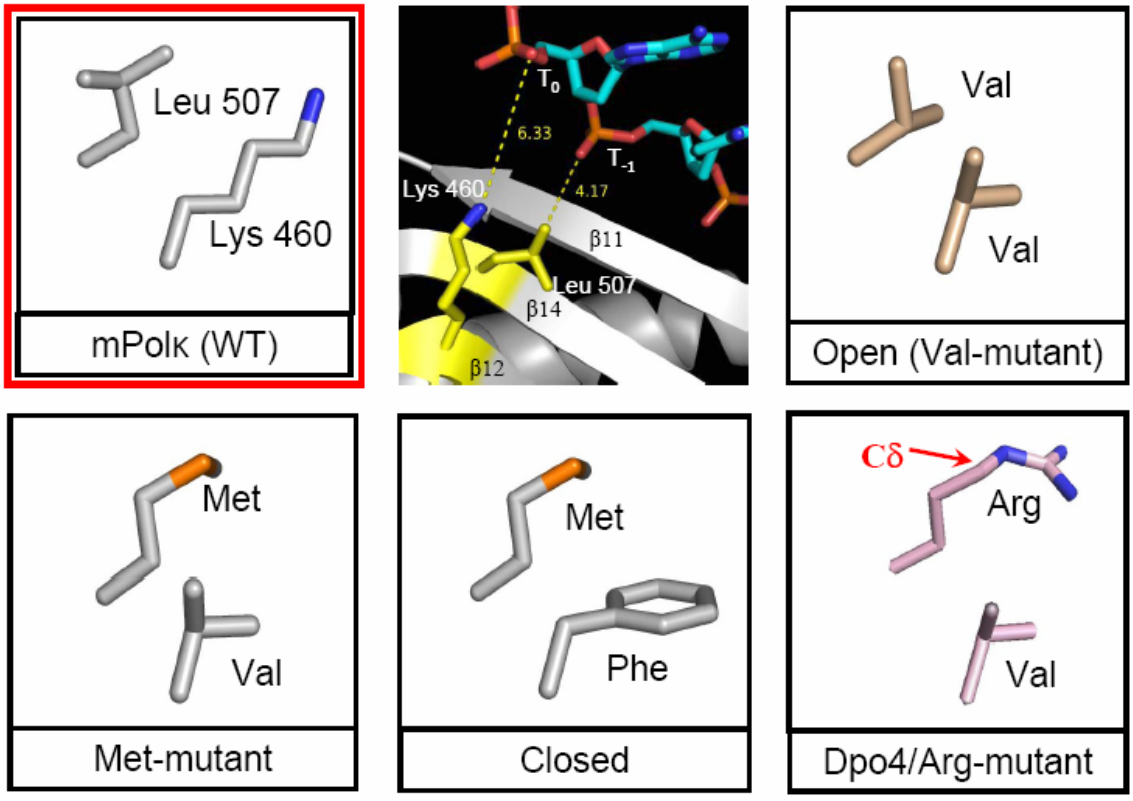


Figure 3.2 Diagram of LF domain mutants. Mutations in proximity to the T_0 and T_{-1} phosphates may alter bypass fidelity. Since Dpo4 and Pol κ have different residues in these positions and also have different activities upon various DNA lesions, several substitutions were made to evaluate the role of local charge and surface topology in these critical interaction surfaces. (**top center**) Mouse Pol κ LF domain contains Leu 507 and Lys 460, whose side chains are too far from the template strand sugar-phosphate backbone to make essential contacts to stabilize the strand. Distances are in Ångstroms. Image recolored from structure presented in Chapter 4. (**bottom right**) The arrow in the Dpo4/Arg-mutant indicates where the Arg 332 side chain bends to accommodate the CPD, as discussed in Chapter 2.

II. Methods

A. Full length mouse Pol κ

The full-length WT mouse Pol κ cDNA was received as a cloned construct in the pET-30a(+) S-tag/His-tag *E. coli* expression vector (Novagen, Madison, WI) from the lab of Dr. Errol Friedberg (University of Texas - Southwestern Medical Center, TX). This original full length construct did not produce crystals (discussed in Chapter 4), the purification tags were believed to be interfere with the formation of the crystal lattice. Therefore, the Pol κ cDNA was sub-cloned into the NdeI/SapI sites of the $pTXB_1$vector (IMPACT-CN protein purification system, New England Biolabs, Ipswich, MA) to increase yield and purity of the final samples. This bacterial expression vector fuses a C-terminal, 24 kDa high-affinity chitin binding domain bridged by a protein splicing sequence, or "intein", which cleaves itself off under reducing conditions without, the need for added proteases.

B. Site Directed Mutagenesis

Six mutants of mouse Pol κ were generated to evaluate the previously described residues of interest: two finger domain mutants, M134A (Ala-mutant) and M134Q (Gln-mutant), and four LF domain mutants, K460V/L507V (Open), K460V/L507M (Met-mutant), K460F/L507M (Closed), and K460V/L507R (Dpo4/Arg-mutant). These point mutations were introduced into the full-length cDNA ($pTXB_1$) using the QuikChange® Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA), and the primers in Appendix A as well as their antisense oligonucleotides. Plasmids containing the mutation were selected by DpnI digestion and used to transform BL21-CodonPlus-(DE3)-RIL *E. coli*

competent cells (Stratagene). Double mutants were created by subjecting the vector containing the first mutation to a second round of point mutagenesis, selection, and transformation.

C. Protein over-expression and purification

All full-length WT and mutant proteins were over-expressed and purified using the following method. Transformed cells harboring the respective $pTXB_1$ constructs were grown in 2 l of 2xYT broth containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C until an OD_{600} of 0.5-0.6 was reached. Protein expression was induced by the addition of IPTG to 1.0 mM, whereby the culture's temperature was lowered to 30°C and incubated for another 8 hours. Cells were harvested by centrifugation at 9300 x g for 20 minutes at 4°C and frozen for later purification. Typical wet pellet weights were 2-2.5 g from a 2 l culture.

One frozen pellet was resuspended in 20 ml chitin column buffer (CCB; 50 mM Na Phosphate, pH 8.0, 300 mM NaCl, and 1 mM EDTA, pH 8.0), supplemented with protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN) and 2 mg Deoxyribonuclease I. Cells were incubated for 45 minutes on ice to fragment the DNA, lysed three times by French pressure cell, and the cell debris was centrifuged at 20000 x g for 30 min at 4°C. The clarified extract was applied to 10 ml Chitin resin (New England Biolabs, Ipswich, MA) pre-equilibrated with CCB, and non-binding proteins were slowly washed away with >50 bed volumes (>500 ml) of CCB overnight. Cleavage of the intein-CBD tag was achieved by washing the chitin resin with 3 bed volumes CCB supplemented with 75 mM DTT and incubation at 4°C for 20 hours. The cleaved protein

was eluted with 40 ml CCB, and diluted 1:2 with 20 mM Tris-HCl, pH 8.0 to begin buffer exchange and to lower the salt concentration to 100 mM NaCl for anion exchange chromatography.

The sample was loaded onto a 25 ml Source Q anion exchange FPLC column (GE Biosciences, Piscataway, NJ) pre-equilibrated with degassed AE Buffer (25 mM Tris-HCl, pH 8.0 and 3 mM DTT), and eluted by a linear NaCl gradient (AE buffer + 1 M NaCl) over 10 column volumes. The fractions containing the target protein, which eluted at 150 mM NaCl over a total volume of 9-12 ml, were analyzed for purity by 8% SDS-PAGE and Coomassie staining. The purest fractions were pooled and dialyzed against Storage buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2.5 mM $MgCl_2$, and 20% PEG 35000). Samples were concentrated using a 1 kDa Slide-a-lyzer cassette in a PEG bath (Storage buffer + 20% PEG 35000) to <1 ml (Pierce Biotechnology, Rockford, IL).

The final protein concentration was determined spectrophotometrically and calculated by the Beer-Lambert Law; absorption of the protein sample was measured at 280 nm in a 1 cm pathlength quartz cuvette and divided by the calculated molar extinction coefficient (28020 M^{-1} cm^{-1}). Dynamic light scattering confirmed that samples contained one dominant species and the polydispersity of the hydrodynamic radius of Pol κ was less than 15%. This protocol yields 2-2.5 mg Pol κ from 1 l starting culture. Aliquots of the final sample, usually ~2 ml of 20 μM Pol κ, were flash frozen in liquid N_2 for later use.

D. Synthesis and purification of oligonucleotide substrates

For the cisplatin study, all oligonucleotides (oligos), shown in Figure 3.3A, were purchased from Invitrogen (Carlsbad, CA). Control template strands were synthesized as a single 36 base sequence while adducted templates were comprised of two shorter oligos: a 12-mer which was chemical modified, and a flanking 24-mer extension sequence which was ligated to the 3′ end of the short oligo to create the full 36 base template strand.

To create the cisplatinated template, 2 μmoles of 12-mer GG-oligo, (5′-CACTCGGTACCA-3′), 0.4 μmol 24-mer 3′-flanker (5′-GTCGATCCTGGGCTGCA-GGAATTC-3′), and 0.3 μmol GG-splint (5′-AGGATCGACTGGTACCGAGTGTTA-TTGACAGTC-3′) were purified by 20% denaturing polyacrylamide gel electrophoresis (D-PAGE). They were located on the gel by UV shadowing and excised, and desalted using a SEP-PAK Classic C18 cartridge (Waters Corp., Milford, MA), yielding 325 nmol, 87 nmol, and 35 nmol of each oligo respectively. The 3′-flanker was 5′-phosphorylated by T4 polynucleotide kinase (Invitrogen), since the 5′ phosphate is missing post-synthesis. The GG-oligo was treated prior to ligation with cisplatin using a well-established method by Gerlach et al. previously optimized for mono-adduct intrastrand crosslink formation (29). Ten milligrams of lyophilized cisplatin (cis-diaminedichloroplatinum(II), 263.15 g/mol) were soaked in 1 ml autoclaved Milli-Q deionized water for 16 hours at 25°C and wrapped in foil to protect the photosensitive chemical. This step converts the cisplatin to the adduction agent, cis-diaminedihydroxyplatinum(II), by replacing the chlorides with hydroxyl groups in a two-step aquation and deprotonation event. This intermediate then reacts with the N^7's of the two neighboring purines. Since cisplatin has a higher affinity

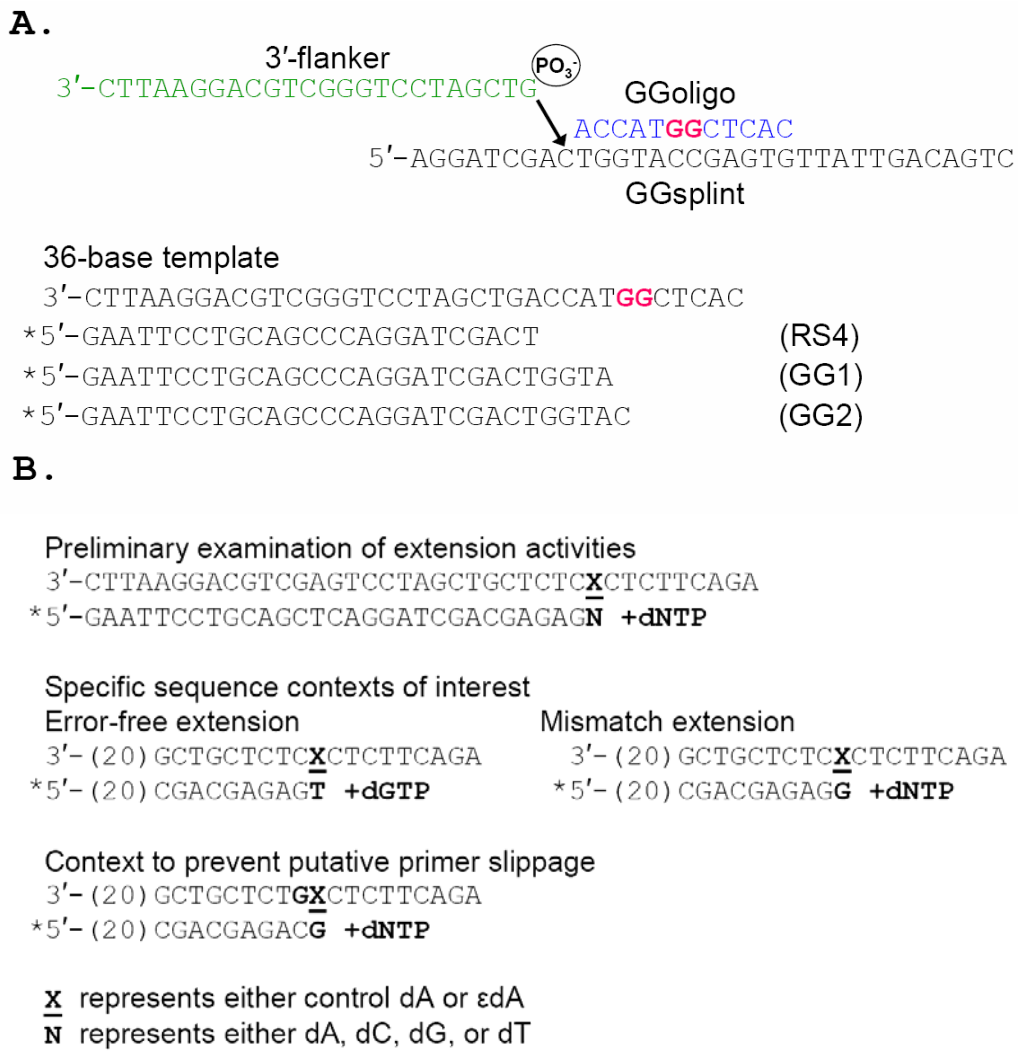


Figure 3.3 Oligonucleotides substrates used in this study (A) Cisplatin intrastrand crosslink bypass study; the di-guanine sequence is highlighted in red. Construction of the ligated template is illustrated, along with the final primer/template contexts. All standing start extension assays (GG1 and GG2 primers) were performed with dCTP as the next incoming nucleotide, while all four dNTPs were provided for running start assays (RS4). **(B)** Post-εdA extension study; various primer termini and incoming nucleotides were utilized to generate error-free and error-prone extension for analysis. Asterisks indicate the oligo that is 5′-radiolabeled with ^{32}P.

for guanosines than adenosines, the short GG-oligo sequence was designed with two adjacent guanines in the center of the sequence so that adduction will be specific (Figure 3.3A). One third of the preparation, ~110 nmol purified GG-oligo, was mixed with one molar equivalent of hydrated cisplatin and incubated for 6 hours at 25°C, wrapped in foil. The crude adduction product was purified by 20% D-PAGE. The successfully mono-adducted oligo migrated between the control 36-mer and a sample of a cisplatin di-adduct, which was created separately as a control by incubating with excess di-aqua-cisplatin. The mono-adduct band was carefully excised to minimize contamination from either species. The sample was again concentrated and desalted with a SEP-PAK C18 cartridge, lyophilized, and reconstituted in 20 mM Tris, pH 7.5 and 20 mM NaCl for storage. This method yields approximately 20 nmol adducted "GpG-oligo".

To create the 36-mer template, the entire GpG-oligo batch (~20 nmol), GG-splint, and 5′-phosphorylated 3′-flanker oligos were mixed in a 1:1.2:2 molar ratio, respectively. The mixture was precipitated by ethanol to decrease the volume of the reaction. It was then resuspended in 30 µl TE buffer and annealed by heating to 90°C to align the template sequences properly. The annealed sample, once cooled, was supplemented with 8 µl of 5x Ligation buffer (1X: 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM dATP, 1 mM DTT and 5% polyethylene glycol 8000) and 2 µl T4 DNA Ligase (Invitrogen) and incubated overnight (>12 hours) at 16°C to obtain the 36-mer modified template strand. The ligation reaction was stopped and ligated sample was separated from the GG-splint oligo by 20% D-PAGE. The final purified product was visualized by UV shadowing and was excised (since oligos were previously purified), and desalted using a SEP-PAK C18 cartridge. The final yield was 580 pmol and the sample was diluted to 1 µM solution.

For the εdA extension study, all oligos were either purchased from Operon (Huntsville, AL) or synthesized in the lab of Dr. Francis Johnson (Stony Brook University, NY) using an automated DNA synthesizer (Figure 3.3B). εdA-modified oligos were generated using εdA phosphoramidites (Glen Research, Sterling, VA) during the elongation process and then were purified by HPLC. All oligos were purified by 20% D-PAGE, detected by UV shadowing, excised, and desalted using a SEP-PAK Classic C18 cartridge. 10 pmol of gel-purified primer was 5′-endlabelled with [γ-^{32}P]ATP (GE Biosciences, Piscataway, NJ) by T4 polynucleotide kinase (New England Biolabs), according to the manufacturer's protocol. Primer-template substrates were prepared by annealing the radiolabeled primer to the template at 90°C in a 1:1.1 molar ratio in TE buffer supplemented with 50 mM NaCl.

E. Primer extension assays

Nucleotide incorporation opposite the cisplatin intrastrand crosslink was evaluated using the 36-mer template, shown in Figure 3.3A, containing either two unmodified dGs or a (1,2-GpG) cisplatin crosslink that was annealed to one of three 5′-radiolabeled primers: "RS4", to be used in 4-base running start extension studies with all four dNTPs, "GG1", for single nucleotide standing start dCTP incorporation experiments opposite the first 3′dG; or "GG2", for similar experiments beginning opposite the neighboring 5′dG. WT Pol κ and mutants (0.25-2.5 nM) were diluted with 50 mM Tris-HCl, pH 7.0, 5 mM $MgCl_2$, 10% glycerol, 1 mM DTT and 0.1 mg/ml BSA to the appropriate 10x concentration. Each sample consisted of 10 nM radiolabeled primer-template substrate, and the appropriate amount of enzyme in 50 mM Tris-HCl, pH 7.0, 10

mM NaCl, 2.5 mM $MgCl_2$, 1% glycerol, 1 mM DTT, and 0.1 mg/ml BSA. To start the reaction, 0.25-500 μM dCTP was added to the above reaction, which was then incubated at room temperature (26°±1°C) for 1-10 minutes as needed to observe <20% primer extension. Reactions were quenched with 20 μl Stop solution (95% formamide, 10 mM EDTA, pH 8.0, 0.001% xylene cyanol), heated to 90°C and subjected to electrophoresis on a 20% D-PAGE gel. Extended primers were visualized by autoradiography using a Storm Phosphorimaging system (Amersham Biosciences, Piscataway, NJ) and analyzed using the ImageQuant software. Control experiments using undamaged templates were performed in triplicate, but the experiments in the presence of cisplatin adducts were only completed in duplicate and halted prior to completion due to an artifact, which will be discussed below.

To study extension following the εdA lesion, a 36-mer template containing an unmodified dA or an εdA was annealed to a 5′-radiolabeled primer that either provided the correct base (herein named extnT) or a mismatched base (named extnG/A/C) opposite the εdA (Figure 3.3B). The reaction was performed in the presence of single nucleotides to observe the fidelity of the extension reaction within each sequence context. WT and mutant (0.1-5 nM) mouse Pol κ were diluted to 100 nM concentration. The primer extension assays were conducted with the same reaction buffer, with 10 nM radiolabeled primer-template substrate and 0.1-5 nM enzyme. To start the reaction, 0-2000 μM of each respective nucleotide were added to the reaction. The solution was then incubated at 37°±1°C for 1-20 minutes as needed to observe <20% primer extension. Reactions were quenched with 20 μl of Stop solution and heated to 90°C prior to electrophoresis on a 20% D-PAGE gel. Each experiment was performed in triplicate.

Separated reaction products were visualized as described earlier. In the cisplatin study, extension of the GG1 primer resulted in two dCTP incorporation events due to the adjacent dGTP in the template sequence; intensities of both bands were summed since both represented the occurrence of the primary incorporation event. In some εdA extension assays, higher concentrations of incoming nucleotides were required (>2 mM dNTP) resulting in magnesium chelation; experiments were modified accordingly to obtain a linear sub-saturating concentration curve to derive the specificity constant. The values from these measurements are therefore an underestimation of the maximal velocity rates, but do provide a range with which to compare kinetic parameters between variants.

F. Data Analysis

Steady-state kinetic parameters K_m and k_{cat} and their standard errors (SE) were determined from the primer extension/dNTP titration experiments with a minimum of 8 data points, and were calculated by the least-squares hyperbolic regression analysis software HYPER (18). In the Mg^{2+}-depleted experiments (because the nucleotide concentrations approached the concentration of divalent cation), an average slope from the linear portion of the curve was calculated from triplicate experiments. When [S] << K_m, the slope equals V_{max}/K_m; dividing this value by the enzyme concentration results in the specific activity, k_{cat}/K_m. These values are perhaps an underestimate for the true kinetic parameters, but provide some means of comparison. Slopes from linear regressions analysis and their standard errors were calculated using SIGMAPLOT 9.0 (Systat Software).

III. Results and discussion of cisplatin insertion studies

A. Enzyme activity upon undamaged DNA

Steady-state kinetic experiments were conducted on unmodified DNA with two standing start primer/templates to measure insertion opposite either the 3′-dG or 5′-dG (Figure 3.3A) and thereby determine any changes in activity as a result of the Met 134 substitutions. Results of these experiments are summarized in Table 3.1. WT mouse Pol κ inserts dCTPs on undamaged DNA in both contexts with a specificity constant (k_{cat}/K_m) of ~3-5 min^{-1} μM^{-1} and a calculated $K_{m(dCTP)}$ of 1-2 μM. These rates are comparable to those found in the published literature for normal dG:dCTP incorporation(46, 102), suggesting that the novel method used to prepare full length mouse Pol κ yielded fully active protein. The kinetic parameters for each dCTP insertion event are very similar, so only the first incorporation event will be discussed in detail. Dissimilarities with the second insertion event are addressed as needed.

When Met 134 was substituted with an Ala or a Gln, distinct changes in the kinetic parameters for correct incorporation were observed. When the roof residue Met was replaced by alanine as in Dpo4 (Ala 42), the k_{cat}/K_m was lowered by an order of magnitude to 0.64, dominated by a 10-fold increase in $K_{m(dCTP)}$. This suggests that while the rate of polymerization was not altered compared with the WT protein, the concentration of incoming dCTP required to initiate catalysis increased because of the shorter Ala residue in the roof of the active site. When Met 134 was substituted by glutamine, the specificity constant decreased significantly by two orders of magnitude compared with the WT protein. This change is also dominated by a significant increase in

A. GG1:control

enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dCTP}$ (μM)	k_{cat}/K_m	F_{ext}	F_{mut}
WT M134	6.3 ± 0.3	1.3 ± 0.2	**4.8 ± 0.8**	**1.000**	
M134A	7.0 ± 0.3	11 ± 1.6	**0.64 ± 0.1**		0.13
M134Q	2.8 ± 0.1	70 ± 8.6	**0.04 ± 0.005**		0.01

B. GG1:cisplatin

enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dCTP}$ (μM)	k_{cat}/K_m *1000	F_{ext}	F_{mut}
WT M134	0.13 ± 0.008	28 ± 5	**4.60 ± 0.87**	**9.6×10^{-4}**	
M134A	0.09 ± 0.100	23 ± 7	**3.90 ± 1.30**	**6.1×10^{-3}**	0.78
M134Q	0.02 ± 0.002	15 ± 3	**1.30 ± 0.31**	**3.3×10^{-2}**	0.26

C. GG2:control

enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dCTP}$ (μM)	k_{cat}/K_m	F_{ext}	F_{mut}
WT M134	4.7 ± 0.2	1.5 ± 0.2	**3.0 ± 0.4**	**1.000**	
M134A	5.4 ± 0.1	25 ± 1.7	**0.21 ± 0.01**		0.07
M134Q	2.3 ± 0.1	28 ± 3.7	**0.08 ± 0.01**		0.03

D. GG2:cisplatin

enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dCTP}$ (μM)	k_{cat}/K_m *1000	F_{ext}	F_{mut}
WT M134	0.12 ± 0.012	17 ± 5	**7.00 ± 2.30**	**2.3×10^{-3}**	
M134A	0.08 ± 0.007	26 ± 6	**3.10 ± 0.82**	**1.5×10^{-2}**	0.44
M134Q	0.02 ± 0.001	14 ± 3	**1.50 ± 0.33**	**1.8×10^{-2}**	0.22

Table 3.1 Kinetic parameters from the cisplatin bypass study. Values were calculated from hyperbolic non-linear regression analysis with standard errors. F_{ext} is the ratio of lesion/dG control specificity constants, while F_{mut} is the ratio of mutant/WT protein specificity constants. Representative gels of this extension reaction are provided in Figure 3.6.

$K_{m(dCTP)}$ of the reaction to 70 μM, suggesting that the slightly bulkier and polar side chain in the active site roof impedes normal catalysis. This likely occurs because the Gln, with its ε-amino or carbonyl ε-oxygen atoms, further constricts the steric gate, thus orienting the template base in an unproductive manner (Figure 3.4). The kinetic parameters for the second dCTP insertion (GG2 primer) are nearly identical, with the exception that the Gln-mutant exhibits a 3-fold lower K_m-dCTP for the second insertion, making it equal to the Ala-mutant.

Another way to analyze this mutant phenotype is to compare the specificity constants of the mutant and the wild type protein against the same substrate, which I name the "Frequency of Mutant Activity, or F_{mut}" (Table 3.1, column 6). This is defined as

$$F_{mut} = \frac{[k_{cat}/K_m]_{mut}}{[k_{cat}/K_m]_{WT}},$$

where F_{mut} can be described as the "percentage of WT activity retained." With this value, any observed changes in activity due to the point mutations can be taken into account. When Met 134 is substituted by an alanine, it exhibited 0.13 of the activity of WT Pol κ and is therefore 12% as active when correctly inserting dCTP upon undamaged DNA. Conservatively replacing the methionine with a glutamine decreases the overall activity of the enzyme 100-fold. Due to its proximity to the template strand, Met 134 likely plays a significant role in manipulating the hydrophobic base into position for base-pairing rather than a direct modulation of primer elongation. Any changes in surface topology or local electronic potential alter the probability for normal Watson-Crick base pairs to form.

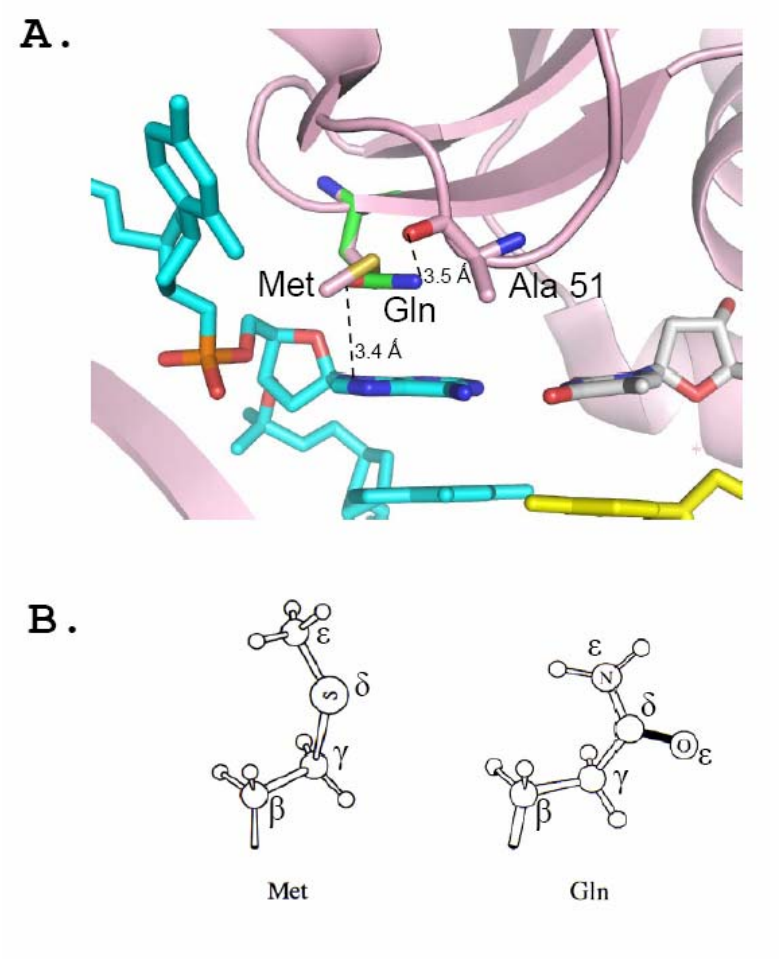


Figure 3.4 Effect of Pol κ Met 134 Gln mutation on extension of undamaged DNA **(A)** Cartoon and stick representation of mouse Pol κ finger domain residue Met 134, along with the most energetically favorable rotamer of Gln. The flexible methionine side chain has more degrees of freedom than the polar Gln, which is also slightly larger with an additional ε-amino group. Objects are colored as described on page (xii), and in this chapter, Pol κ is colored pink while Dpo4 is colored green. Distances are measured from the ε-moieties of Gln. **(B)** Ball-and-stick representation of Met and Gln side chains (α-carbons are absent) from Proteins, 2nd. ed, Creighton (1993), W.H. Freeman and Co, NY.

B. Bypass activity of Finger mutants in the presence of cisplatinated DNA

Because Pol κ was observed as being blocked by linked-base lesions, steady-state extension assays were performed with the steric gate point mutants to determine if they had any effect on accommodation and bypass of cisplatin crosslink. Cisplatinated templates were hybridized with standing start primers that promote insertion opposite either crosslinked dG (GG1 and GG2, respectively) in order to isolate the kinetics parameters of those events. The results (Table 3.1B) show that the K_m values remain constant for both the Ala- and Gln-mutants, but the turnover rates (k_{cat}) for the mutants compared to the WT have decreased 60-fold and 140-fold, respectively, due to the presence of the cisplatin intrastrand crosslink. To facilitate comparison of the enzymes' activities upon the various damaged and undamaged substrates, I also report the relative frequency of extension(31), which is defined as:

$$F_{ext} = \frac{[k_{cat}/K_m]_{damaged}}{[k_{cat}/K_m]_{undamaged}}.$$

For the first incorporation event opposite the 3′-dG of the crosslink, the calculated F_{ext} values show an notable trend. As expected upon the cisplatinated template, the WT enzyme's kinetic parameters show a sharp decrease in activity, reflected in its lower k_{cat}/K_m values and a F_{ext} value of 9.6 x 10^{-4}. However, in the case of the mutants, only the k_{cat} values decrease significantly while the K_m values remain similar to those observed opposite undamaged DNA. The F_{ext} values also demonstrate relative improvement in activity in comparison to WT Pol κ; WT-catalyzed cisplatin bypass was in the range of 10^{-4}, whereas the Ala- and Gln-mutants exhibited 10^{-3} and 10^{-2} F_{ext} values, respectively.

This suggests that these mutants are partially successful in alleviating the steric hindrance and permitting some cisplatin bypass.

A possible explanation for the improvement elicited by the substitutions is that they cause the template base to remain in the active site for a longer period of time, thus increasing the chance of a successful dG*:dC Watson-Crick base pair to form. Since a (1,2-GpG) crosslink prohibits the bases from adopting a *syn* (Hoogsteen) conformation, it is only a matter of time until the 3'dG is in a suitable position for base-pairing to the incoming nucleotide. Having an Ala in place allows base-pairing to occur more readily because of the additional space near the fingers domain.

However, this does not sufficiently explain how the bulkier Gln-mutant is more effectively able to accommodate the lesion compared with a methionine in that position. In addition, a puzzling observation was made: WT Pol κ catalyzed dCTP past a cisplatin adduct in a quantifiable way, contradicting all the known literature (29, 92, 131). For this reason, mutational analysis was halted in order to investigate the cisplatin crosslink bypass activity catalyzed by WT Pol κ.

C. WT Pol κ activity in the presence of crosslinked DNA

While evaluating WT Pol κ activity on the cisplatinated substrate, it was observed that 5-10% of the primers were extended when reactions were performed in the presence of all four dNTPs, with full extension of the primer being observed (Figure 3.5). Since reaction rates are calculated from experiments where only 20% of the primers are extended to isolate "single-hit" events, baseline activity resulting from some artifact would compromise data analysis. In previous studies extension of cisplatinated DNA was

observed as well, but at significantly higher enzyme concentrations (>10 nM Pol κ) and this was subsequently discounted (29, 92). In our experiments, cisplatin bypass was observed at a half maximal velocity with only 30 μM dCTP, which is well below the estimated dCTP concentration of 250 μM near DNA replication foci (75).

Standing start and running start experiments were performed to challenge this observation of bypass, and both contexts promoted 5-7% primer extension past the crosslink in the presence of 1-2.5 nM WT Pol κ (Figures 3.5 and 3.6). An notable pattern emerged from 4 dNTP extension on the cisplatinated substrate. An n+1 band accumulated opposite the 3′dG of the cisplatin adduct, indicating that the damaged base was entering the active site and stalling the polymerase to some extent. But products are also visualized beyond n+4 until the end of the template strand, indicating that once the adducted 3′dG was bypassed, primer extension continued to the end of the template, with no trace of n+2 and n+3 bands remaining post lesion. The observations suggest that a mono-adducted contaminant was likely present. However, subsequent analysis by mass-spectrometry of the adducted DNA clearly showed the presence of cisplatin (see Appendix C). Nevertheless, a mono-adducted DNA can not be ruled out.

The second notable observation was noted in the standing start experiment: full extension of the primer was seen even in the presence of 1 nM Pol κ, but the concentration of extended primer remained constant, suggesting that the artifact is due to a finite amount of contaminant, rather than a fractional amount of cisplatin crosslink bypass.

Several biophysical methods were used to analyze the nature of the artifact. First, the unmodified template, cisplatinated template, and freshly ligated cisplatin template

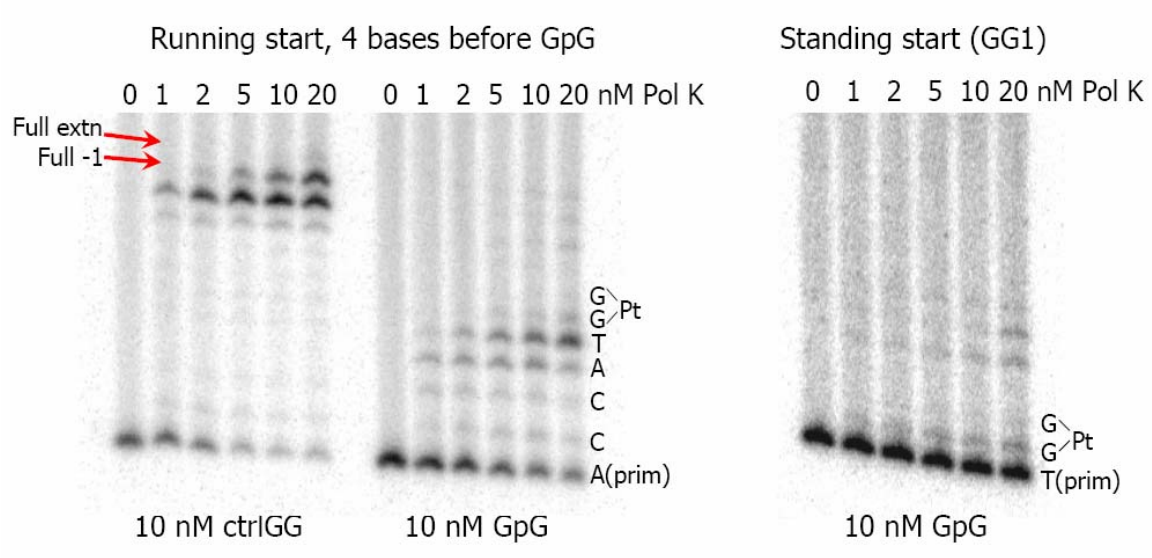


Figure 3.5 WT primer extension assays signifying the cisplatin bypass artifact. (Far left) Nearly full extension is observed on control DNA using WT Pol κ, which is known to stop one base before the end of the template. **(Center)** WT Pol κ upon cisplatinated DNA with a 4-base running start shows extension past the lesion at 2.5 nM enzyme. The template sequence is shown to the right, along with the location of the cisplatin intrastrand crosslink (Pt). **(Far right)** The artifact is more pronounced when the primer is terminated before the crosslink adduct. All experiments were performed in the presence of 50 μM of each dNTP for 10 minutes at 26°C. The background is due to insufficient formamide and EDTA in the quenching solution, but does not change the qualitative conclusion that cisplatinated template is being extended this experiment.

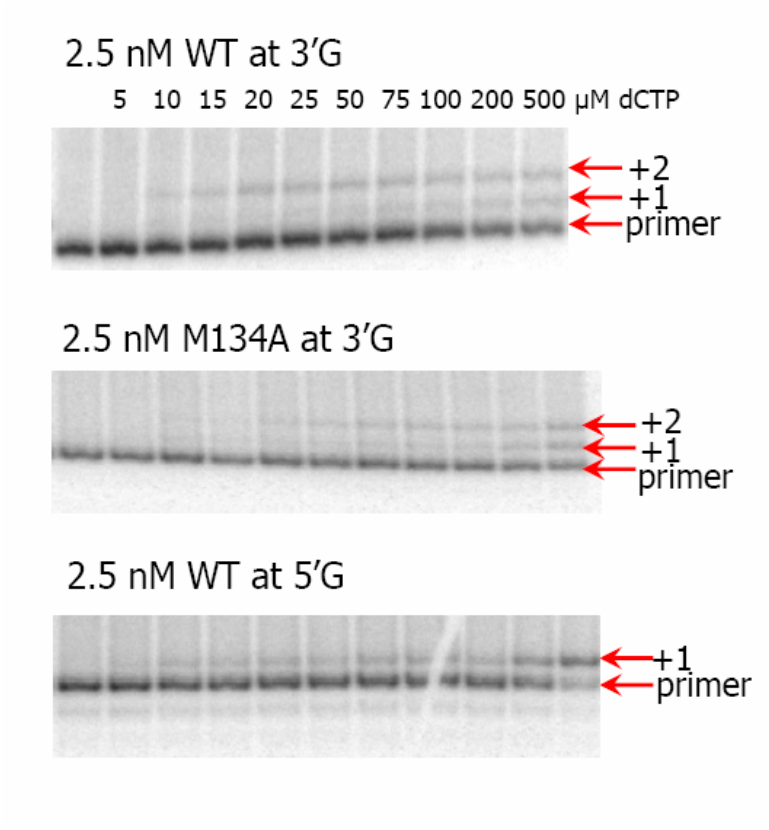


Figure 3.6 Representative autoradiograph of primer extension upon cisplatinated DNA. WT Pol κ and the M134A variant display measurable amounts of primer extension activity in the presence of a cisplatin intrastrand crosslink. Note the distinct n+1 and n+2 extension bands for the WT polymerase, indicating bypass upon cisplatinated DNA, a previously unreported event.

(used to evaluate the possibility of contamination, thereby causing further reactions over a prolonged time), were 5′-radiolabeled with ^{32}P and analyzed by 20% D-PAGE. The results showed that both cisplatinated samples migrated in exactly the same manner with no additional contaminating bands. Then, samples of the constructed template, along with freshly ligated cisplatinated oligos, were analyzed by MALDI-TOF mass spectrometry, but no obvious contaminants of lesser molecular weight were observed within a 2% confidence level, confirming that non-adducted template was not present. Cisplatin intrastrand crosslinks are very stable, which leaves mono-adducts to be formed from di-aqua (or di-chloro) cisplatin that did not coordinate to the second dG. This would reveal itself as templates with the additional mass of a chloride (35.4 amu), a hydroxyl group (17.0 amu), or a water molecule (18.0 amu), and should be readily visible by mass spectrometry. While there are some minor higher molecular weight signals (Appendix C, bottom), it is difficult to tell if this is evidence for mono-adduct or merely noise from the analysis.

D. Discussion

From these results, it is unclear why cisplatin bypass was observed in our experiments. I cannot rule out the possibility that the crosslinked substrate used in the experiments may contain a small amount of cisplatin mono-adduct on the 3′dG that either never formed a covalent bond or subsequently dissociated from the 5′dG in our template. This suggestion derives from primer accumulation opposite the 3′dG position (Figure 3.5, right). In this case, the mono-adduct would be recognized as a single adduct modified at the guanosine N^7 that could be easily accommodated by the active site of Pol κ (70) (and

presumably the Ala- and Gln-mutant's active site), as there is plenty of room. Since our samples had been obtained from our collaborator's laboratory, creating new sample was put on hold to pursue an εdA adduct first, for which it was clear that no contaminants were present.

From the homology model, it appears that nearly every residue that was in contact with the DNA template was not conserved. I explored the active site residue Met 134, which is substituted by an alanine in Dpo4 (Ala 42). It is located directly in the region where the upstream T_{+1} base faces away from the active site and must swing >90° when processing along the DNA binding channel. However, eukaryotic Pol κ has a methionine in this position. I, and others (110), presumed that this bulky side chain excludes thymine dimers since this Met 134 would sterically prohibit the distorted phosphate backbone from entering the active site. However, mutation of mouse Pol κ's Met to an Ala did not conclusively improve bypass of linked base lesions such as an (1,2 GpG) intrastrand cisplatin crosslink. Only after the structure of fully active human Pol κ in a ternary complex was solved (70), it became clear that the novel N-terminus, which adopts the fold of a three-helix "N-clasp", forms hydrophobic interactions to the fingers domain and creating a single base "steric" gate between Met 134 and an aromatic Phe 49 (Figure 3.7). This arrangement promotes bottlenecking and base-stacking as the single bases slide through this 5Å wide gate. Replacing Met 134 with an Ala may have more effectively accommodated the cisplatinated template, but the unaltered presence of Phe 48 (Phe 49 in human Pol κ) did not open the gate wide enough as evidenced by the unchanged k_{cat} value. Therefore, the single methionine to alanine mutation could not improve bypass of linked base lesions.

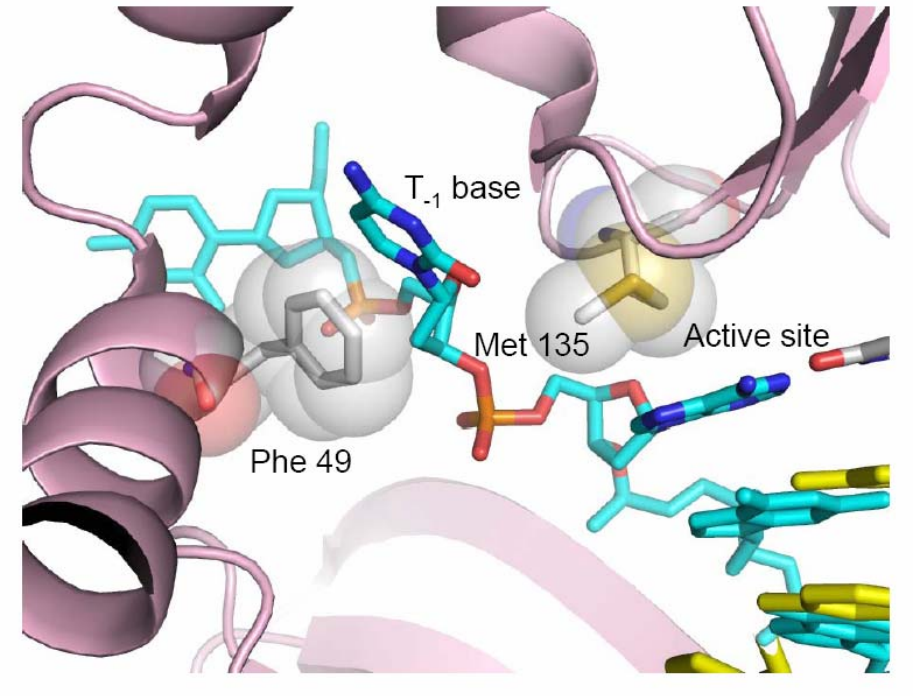


Figure 3.7 Cartoon and stick representation of the linked-base lesion steric gate in human Pol κ. Met 135 (Met 134 in human) and Phe 49, both depicted with their van der Waals spheres, create a steric gate only 7Å wide through which only certain lesions can fit. While the Ala 134 mutant would provide more space due to its shorter side chain, the stacking influence of Phe 49 is still unfavorable to the constraints of a cisplatin intrastrand crosslink. Image of PDB entry 2OH2 (70) recolored using PyMol; as described on page (xii).

IV. Results of εdA extension studies

A. Fidelity of WT mouse Pol κ - qualitative comparison

Since Pol κ had been observed to generate frameshift deletions when bypassing certain lesions in a sequence-dependent manner, we examined extension activity after every primer-terminus opposite an εdA with WT Pol κ and the LF mutants. Reactions contained 10 nM primer-template, excess incoming nucleotide, and an appropriate amount of each enzyme (10 – 100 nM) to maximize observations of minor mismatch events or other rare events <20% extension of its minor products. An interesting pattern emerged from this qualitative investigation, as seen in Figure 3.8A-D. First, correct extension post-εdA (with a dGTP) was observed with all enzymes after every εdA:dN; the weakest dGTP extension is observed after an εdA:dC mispair. Second, extension following an εdA:dG mispair results in several extension products, dATP>>>dTTP>dGTP, with the correct extension being the least preferred. As explained in Figure 3.9, the predominance of dATP, is a result from Pol κ's preference to extrude the lesion and to extend from misaligned primers, resulting in (-1) frameshift deletion. This activity is sequence-dependent and has been reported several times (65, 118).

The novel observation from this study is a dC:dTTP mismatch following the εdA:dG mismatch. According to our template context as shown in Figure 3.9 (Bottom), either WT Pol κ is misincorporating this dTTP opposite the 5′C, or a primer misalignment activity is observed for the first time. When the primer terminus dG, or its adjacent 3′dG, becomes extruded, εdA could re-enter the active site for an additional incorporation cycle opposite the dTTP (Figure 3.9A). Considering Pol κ's propensity for frameshift errors even on undamaged DNA sequences, this activity could easily be

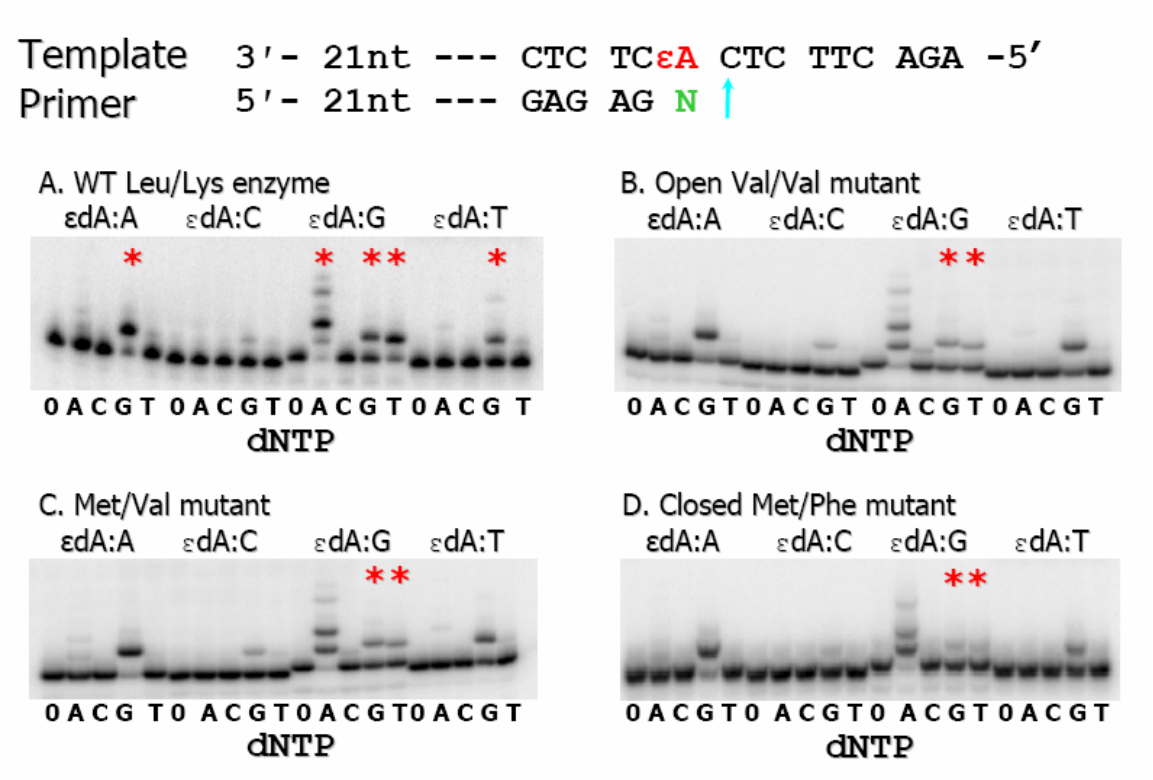


Figure 3.8 Qualitative assessment of WT and mutant extension fidelity after all εdA:dN base pairs. (Top), Primer/template sequence context used for this study, where the light blue arrow indicates the template base in the active site; **(Bottom)**, single nucleotide extension assays reveal which nucleotides are inserted by the 4 enzymes, **(A)** WT, **(B)** Open-mutant, **(C)** Met-mutant and **(D)** closed mutant, when presented with different εdA:dN base pairs. The mutated residues are indicated, (Leu 507/Lys460), based on β-strand order in proximity to the template strand. Red asterisks indicate insertion activity worthy of comparison. After an εdA:dG base pair, the mutants appear to extend less often in the presence of dATP and dTTP than WT Pol κ, suggesting a slight improvement in extension fidelity.

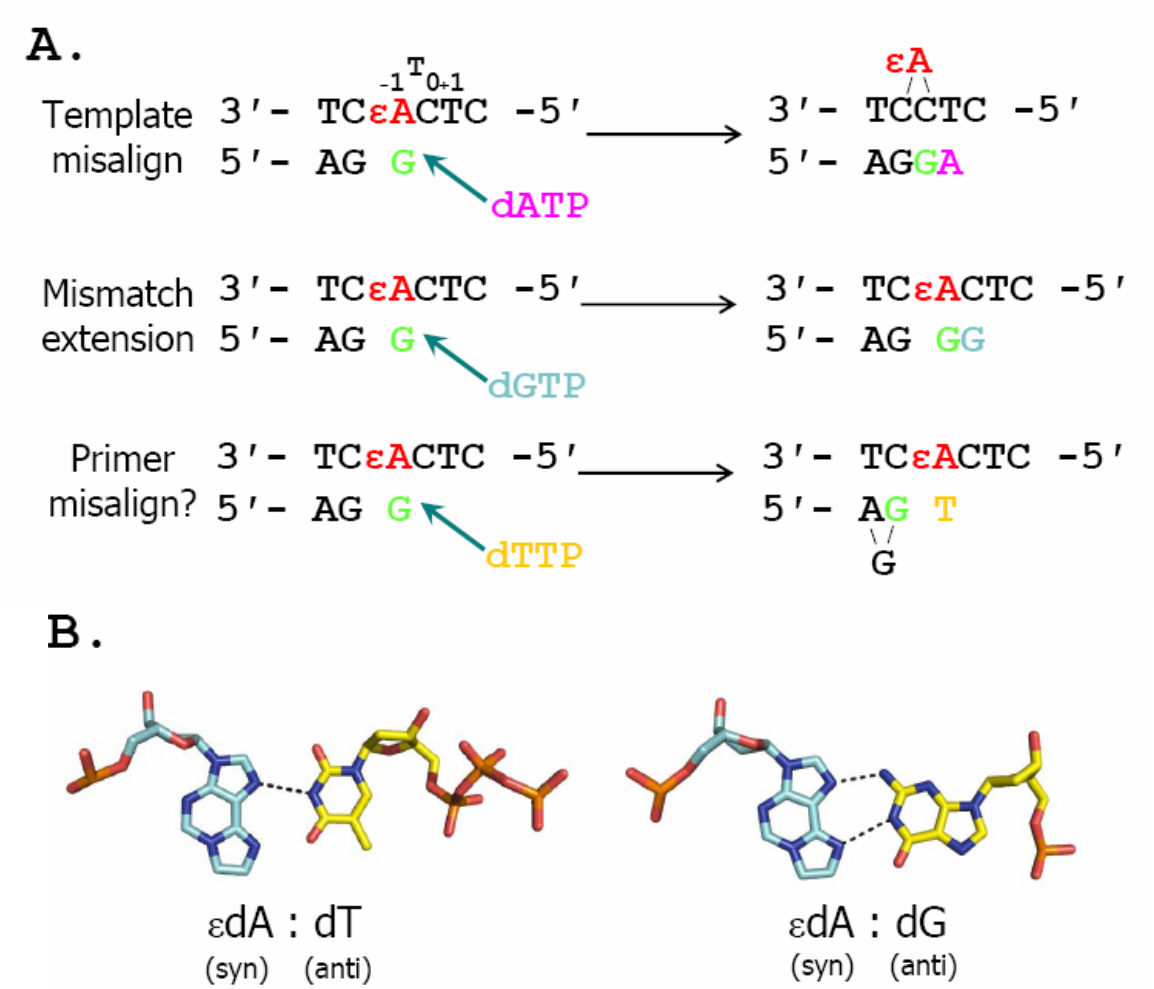


Figure 3.9 Diagram of the three potential insertion mechanisms for Pol κ after an εdA:dG mispair. (A), (top) A (-1) frameshift results when the damaged base is extruded and the primer terminus is stabilized by the neighboring 5′dC, promoting incorporation opposite the T_{+2} base; (Center) no bases are extruded in normal extension; (Bottom) the primer strand slips due to instability, promoting +1 frameshift deletions. **(B)**, Hoogsteen base pairs are formed with εdA:dG or εdA:dT. Polarities of hydrogen bonds in εdA:dG pair are in same direction, making them slightly less stable (64). Base pairs are recolored from Nair *et al* (82) and Leonard *et al* (64), respectively.

envisioned and detailed steady-state primer extension assays are required to dissect this novel activity.

B. A novel slippage mechanism by WT polymerase κ

As discussed above, further studies were required to explain the unexpected dTTP insertion activity exhibited by mouse Pol κ. Since the dG-terminated primer could misalign either forward or backwards due to the presence of dCs on either side of the damage, we changed the template sequence to 3′-T**G**XCT-5′ (sequence written in reverse to facilitate comparison to Figure 3.9) to remove the terminal dG's putative primer misalignment partner. If Pol κ is in fact misaligning its primer template, this substitution should affect its kinetics parameter behavior. While this sequence is designed to block primer misalignment activity, template misalignment is also prevented by the lack of dA binding partners 5′ to the lesion in this sequence.

Due to this sequence alteration, (a) the K_m-dTTP increases significantly to 600 μM, (Table 3.2G), and (b) the k_{cat} increased ~5-8 fold in comparison to the other sequence context. In fact, it is >10-fold higher than WT Pol κ-catalyzed error-free bypass of εdA. The k_{cat} of the reaction depends on WT Pol κ to stabilize the template base correctly for nucleotide incorporation, and therefore these results are indicative of an increase in dC:dT mismatch activity. Since the template sequence no longer provides a secondary binding location for the dG primer terminus, WT Pol κ must insert the dTTP opposite the template dC, resulting in an unfavorable mismatch. This explanation can be applied to the results that we see for the εdA:dG + dGTP experiment. The K_m and k_{cat} values of dGTP insertion are comparatively lower in value compared to the εdA: dG +

dTTP experiment, which is due to the fact that the dGTP being provided can be incorporated at two different positions. Additional steps, e.g., slipping back and forth, slows the overall incorporation rate down. We therefore suggest that mouse Pol κ catalyzes primer "realignment".

C. A different fidelity profile with LF mutants

In comparison to WT mouse Pol κ, the mutants display slightly altered extension preferences. The Open (Val/Val) mutant (Figure 3.8B) catalyzes correct dGTP insertion with all primer termini except εdA:dG, where again the fidelity is low, similar to the WT protein. However, the order of selection has changed to dATP>>dGTP≥dTTP, suggesting a profile that has shifted slightly towards correct extension. Though all of the LF mutants still employ template misalignment, they appear to be less dominant compared to the wild type protein in this qualitative analysis. For the Met/Val mutant (Figure 3.8C), the extension selection profile is comparable to the Open Val/Val mutant; however, half as much enzyme was required to obtain this profile, indicating that this mutant exhibits higher overall activity. Lastly, the Closed Met/Phe mutant displays the same characteristics compared to the Met/Val mutant activity, suggesting that the larger phenylalanine does not provide a positive effect through possible base stacking interactions or a negative effect through steric hindrance compared to a valine in this position. We therefore abandoned the Closed mutant and only considered the Open (Val/Val)and Met (Met/Val) mutants, and instead considered the last LF domain variant, the "Dpo4-revert" (Arg/Val) mutant for further analysis. To facilitate the comparison

among the mutants, we will refer to the mutants by their Leu-507-X mutation since in all remaining LF-constructs, residue Lys460 has been changed to a Val.

Based on the initial results, standing start steady-state primer extension assays were conducted. Substrates containing εdA were utilized in the following three contexts: error-free (εdA:dT+dGTP, the next correct extension base), lesion mispair/correct extension (εdA:dG + dGTP), and lesion mispair/mismatch extension (εdA:dG + dATP or dTTP).

D. Improved error-free extension upon modified DNA with LF mutants

Since the qualitative studies showed that the Pol κ mutants exhibited an altered extension fidelity profile, we calculated values the kinetic parameters of incorporation opposite the T_{+1} base after εdA damage with various incoming nucleotides. The final kinetic results of εdA:dT + dGTP error-free extension are summarized in Table 3.2A,B. Using undamaged DNA, the specificity constants (k_{cat}/K_m values) decrease as a result of the different LF mutations, dominated by sharp increases in K_m-*dGTP* values. These results show that turnover rates are virtually constant, but increasingly higher concentrations of dGTP are required to extend the primer, suggesting that these two residues are essential for DNA substrate interactions near the active site. The trend is that the Met- and Arg-mutants equally exhibit 2% the activity of the wild-type protein, with the Open Val-mutant being the most affected with an overall activity reduced over three orders of magnitude compared to the WT enzyme. This indicates the requirement of a bulkier residue in this location for efficient replication of undamaged DNA.

Extension studies after εdA:dT (correct basepair)

	enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dGTP}$ (μM)	k_{cat}/K_m	$F_{extn(dTTP)}$	F_{mut}
A. dA:dT + G	WT Leu/Lys	15 ± 0.5	0.43 ± 0.05	**35 ± 4**	**1.00**	
	Open Val/Val	9 ± 0.3	171 ± 19	**0.05 ± 0.01**		**0.001**
	Met/Val	9 ± 0.3	17 ± 2	**0.53 ± 0.06**		**0.015**
	Dpo4 Arg/Val	6 ± 0.2	15 ± 2	**0.38 ± 0.05**		**0.011**

	enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dGTP}$ (μM)	k_{cat}/K_m *1000	F_{extn} * 1000	F_{mut}
B. εdA:dT + G	WT Leu/Lys	0.06 ± 0.002	44 ± 4	**1.41 ± 0.13**	**0.04**	
	Open Val/Val	NM	NM	**0.08 ± 0.004**	**1.53**	**0.057**
	Met/Val	0.07 ± 0.007	999 ± 101	**0.07 ± 0.01**	**0.14**	**0.052**
	Dpo4 Arg/Val	0.17 ± 0.005	845 ± 62	**0.20 ± 0.02**	**0.52**	**0.140**

Extension studies after εdA:dG (mispair)

	enzyme	k_{cat} (min^{-1})	$K_{m\text{-}dNTP}$ (μM)	k_{cat}/K_m *1000	$F_{extn(dNTP)}$	F_{mut}
C. dA:dG + G	WT Leu/Lys	0.21 ± 0.01	46 ± 8	**4.53 ± 0.80**	**1.00**	
	Met/Val	NM	NM	**0.07 ± 0.003**	**1.00**	
	Dpo4 Arg/Val	0.32 ± 0.01	1286 ± 211	**0.25 ± 0.04**	**1.00**	
D. εdA:dG + G	WT Leu/Lys	0.05 ± 0.002	186 ± 23	**0.27 ± 0.03**	**0.06**	
	Met/Val	NM	NM	**0.01 ± 0.003**	**0.20**	**0.054**
	Dpo4 Arg/Val	NM	NM	**0.02 ± 0.001**	**0.08**	**0.078**
E. εdA:dG + A	WT Leu/Lys	0.83 ± 0.03	134 ± 16	**6.2 ± 0.8**	**1.36**	
	Met/Val	NM	NM	**0.09 ± 0.004**	**1.21**	**0.014**
	Dpo4 Arg/Val	0.34 ± 0.03	947 ± 184	**0.36 ± 0.08**	**1.46**	**0.059**
F. εdA:dG + T	WT Leu/Lys	0.18 ± 0.01	252 ± 47	**3.81 ± 0.76**	**0.84**	
	Met/Val	NM	NM	**0.01 ± 0.002**	**0.16**	**0.003**
	Dpo4 Arg/Val	NM	NM	**0.02 ± 0.002**	**0.08**	**0.005**
G. C-εdA:dG + T	WT Leu/Lys	0.82 ± 0.05	676 ± 106	**1.21 ± 0.20**	**0.27**	

Table 3.2 Values of kinetic parameters of post-εdA extension with WT and LF mutants in difference sequence contexts. Parameters were defined in Table 3.1. NM, not measurable (due to substrate inhibition).

In the presence of εdA-damaged DNA, some changes in the kinetic parameters are observed. First, the $K_{m\text{-}dGTP}$ for WT Pol κ increases 100-fold, confirming that extension past εdA in an error-free manner is limiting. As for the LF mutants, the Met-mutant now diverges from the Arg-mutant in overall activity, with a specificity constant that is more similar to the compromised Val-mutant. This suggests that a longer Met versus a Leu in this position as in the WT protein is not sufficient to improve εdA bypass. Though the Val-mutant displays a 50-fold increase in F_{mut} as compared to dGTP insertion after an unmodified dA:dT base pair, it cannot be described to be beneficial since its rates of extension are substantially lower when compared to its partners. The Arg-mutant (Arg/Val) shows the highest level of activity among the mutants, with an $F_{ext(dTTP)}$ of 0.52, 5-fold higher than WT Pol κ (Leu/Lys).

If we compared the final $F_{extn(dTTP)}$ values (F_{ext}*1000 column) between the enzymes upon εdA-damaged DNA, we see that WT Pol κ is strongly affected (~10^{-5}) by the presence of the lesion while the mutants are much less affected by εdA (~10^{-3}-10^{-4}) compared to the undamaged substrate. Therefore, in an error-free context, the mutants show improved activity with the following trend: Arg/Val > Met/Val > WT Leu/Lys > Val/Val. The (re-)introduction of a longer and positively charged residue at the 507 position increased the rate of error-free extension, in effect, thus mimicking the extension past εdA lesions as observed in Pol η. Additionally, these results suggest that two short hydrophobic residues in close vicinity to the DNA backbone, as seen in the Val-mutant, are unfavorable for activity.

E. Mutants show little improvement in dGTP extension after an εdA:dG mispair

We analyzed the extension activity of the WT protein and the Met- and Arg-mutants with an εdA:dG mispair substrate in the presence of the next correct base dGTP. The results of these experiments are summarized in Table 3.2C-D. In the control experiment (Table 3.2C, WT), the purine:purine mismatch caused an 8-fold reduction of the specificity constant for dGTP insertion compared to the WT protein, but its rate of insertion was still greater than after an εdA:dT terminus. In the presence of the lesion, however, the value of the specifcity constant has dropped 94%, dominated by both a decrease in k_{cat} and an increase in the $K_{m\text{-}dGTP}$ value. This suggests that our template is not stabilized by the protein, slipping back-and-forth as already suggested by Levine *et al*.(65), slowing the overall procession of the polymerase.

The LF mutants showed little improvement when inserting dGTP after the mispair. Correct extension with dGTP by the Met-mutant was again one order of magnitude lower than for the Arg-mutant on the undamaged substrate, but in the presence of εdA, both lost their overall activity with specificity constants in the range of 10^{-5}. Their $F_{extn(dGTP)}$ values were approximately equal to the F_{extn} value of the WT enzyme, suggesting that the mutants do not improve correct extension after a damaged mispair. One possible explanation for these results is that εdA:dT is energetically recognized as a mispair, hence the similar poor dGTP incorporation rates. The minimal extension activity of the Met-mutant suggests that a Met in this position is not favorable, which is emphasized in the error-free context. These combined experiments indicate that a positive charge is necessary to modulate mispair or lesion extension, and simple steric considerations along the LF DNA contact surface as in the error-free context are insufficient.

F. Mutants disrupt novel primer slippage activity

Within the sequence 3′-TCXCT-5′, primer slippage is the second dominant εdA bypass activity in the extension spectrum for wild-type Pol κ. However, changing the Leu to a bulkier Met or Arg completely inhibited this extension, causing greater than two order of magnitude reduction in activity (Table 3.2F). We observed this also in the $F_{extn(dGTP)}$ value, as it dropped from 0.8 for the WT to below 0.2 for the mutants. Therefore, the Leu/Lys combination in WT Pol κ is required for this novel primer slippage activity, and returning the positive charge as it is observed in Dpo4 polymerase stabilizes the substrate, knocking out this realignment activity. By inhibition of this primer realignment, the Dpo4-Revert Arg-mutant is shifting to a less error-prone profile, making dG extension more likely and leading to less error-prone variants of Pol κ overall.

G. Frameshift deletion mechanism moderately slowed by mutants

We tested our mutants to observe what effect, if any, the mutations would have on this well-established (-1) frameshift bypass mechanism (29, 60, 122, 131); results are summarized in Table 3.2E. As expected, the k_{cat}/K_m values show this activity to be the most frequent for the WT protein; the mutations do not seem to affect the predominance of this mechanism substantially. If we compare specificity constants of the enzymes' extension with dATP to the control dA:dG + dGTP extension experiment ($F_{extn(dGTP)}$), no obvious reduction in frameshift activity can be observed. However, the Dpo4 Arg-mutant in this context seems to impose some catalytic interference compared to the wild type protein, since its k_{cat} is reduced from 0.83 sec^{-1} to 0.34 sec^{-1}.

H. Discussion

Etheno-DNA adducts (ε-adducts) such as $1,N^6$-ethenodeoxyadenosine are common in tissues where steroidogenesis and/or reactive oxygen species are abundant, such as the liver, gonads, lungs, and adrenal glands (12, 85, 86, 105, 106). During normal hormonal biosynthesis, cytochrome P450 generates a large number of superoxide anions and free radicals that are typically counteracted by native antioxidant pathways(85). In cases of extreme oxidative stress, the inflammatory response triggers lipid peroxidation, resulting in electrophilic α, β-unsaturated crotonaldehyde and malondialdehyde, and reactive enals such as 4-hydroxy-2-nonenal (HNE) (85) that can directly adduct to deoxynucleotides. The resulting exocyclic adducts are miscoding and carcinogenic, and while these adducts are effectively excised by alkyl-N-purine glycosylases (8, 17, 36), not all of them can be repaired before the replicative cycle begins. Coincidently, expression of Pol κ is highest in these tissues (29), suggesting the essential role that Pol κ plays in the repair of oxidative damage. It has already been shown that Pol κ has evolved to extend preferentially from mispaired termini, and in the case of εdA lesions, extension is primarily preferred after a sequence-based misalignment of the primer terminus rather than an error-free base pair with thymine (65, 118, 122).

While several studies have continued to actively pursue residues that govern lesion specificity and bypass fidelity in Dpo4, this study is the first to pinpoint residues within the little finger domain of the error-prone eukaryotic Pol κ. Our mutagenesis studies show that the region analogous to Dpo4's Arg 332 (mouse Pol κ residues Leu 507 and Lys 460) mediate some critical role in Pol κ's error-prone lesion bypass. With the translocation of Lys 460 away from its canonical position to a farther strand, we

postulated that this local positive charge is now too far away from the DNA backbone to impart any stabilizing influence directly. With the human Pol κ-DNA complex structure now in place, we can see that this is indeed the case.

If we first examine the LF domain of Dpo4 in complex with unmodified DNA (1JX4 (67)), we observe that the T_0 phosphate is flanked by two positive charges, Arg 332 and Arg 331 (Figure 3.10). β-strand 14 is positioned parallel to the sugar–phosphate backbone and Arg 331 is within a distance of 3.0 Å to the T_0 phosphate group, while Arg 332 is within 2.8 Å to the phosphate group of the T_{-1} base. The β and γ methylene groups of Arg 332 are directly positioned beneath the nascent T_0 base, being on average 4.0 Å away from the T_0 phosphate group and creating an even interaction surface. Arg 332 has been observed to adopt different conformations in the presence of a lesion such as a thymidine dimer where it is forced in the direction away from the active site by the displacement of the phosphate due to the distorting effects of the linked bases.

In contrast, the human Pol κ ternary structure does not appear to provide a similar level of stabilization. In the same region, a conserved Arg (Arg 507, analogous to Arg 331) is positioned 2.6 Å away from the T_0 phosphate, but the other positively charged residue Lys 461 on β12, is more than 6 Å away from the T_0 and T_{-1} phosphate groups, and well out of direct hydrogen bonding range. While there is no evidence against the possibility of a water bridged hydrogen bonding network, the 3 Å resolution structure does not contain an ordered water in this position(70). The electron density maps are ambiguous in these areas, suggesting that a transient water molecule may occupy this position. Nevertheless, a water-bridged hydrogen bonding network coordinated by Lys 461, possibly in combination with Lys 459, may not be energetically equivalent to the

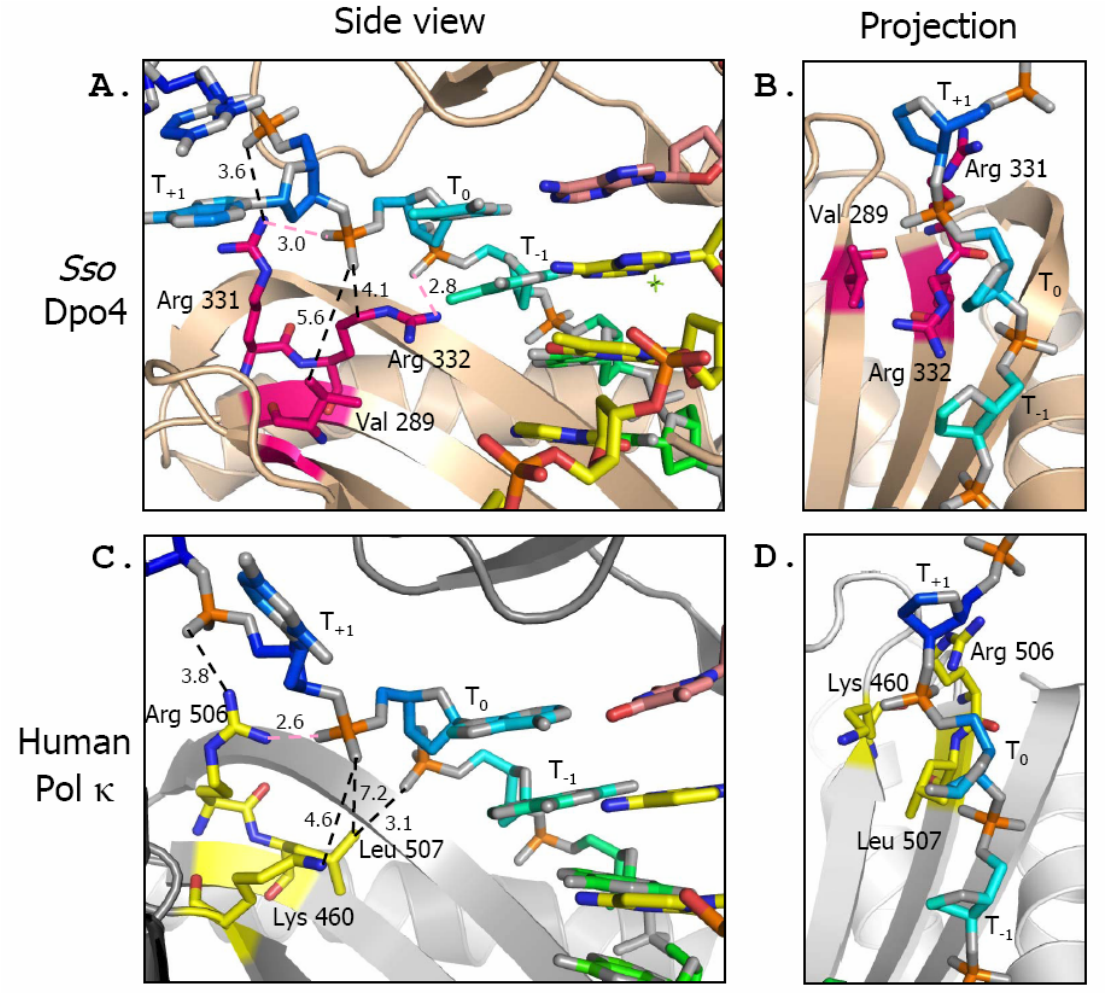


Figure 3.10 Comparison of distances between LF domain residues and DNA in Dpo4 and Pol κ. **(A)** and **(C)** View perpendicular to the plane of the bases in the active site. There are two electrostatic interactions within 3.0 Å to the template strand phosphate backbone oxygens in the major groove, while Pol κ only has one (Arg 506). **(B)** and **(D)** Top projection view of the template strand on β 14. Only the sugar-phosphate backbone is depicted for clarity. Note that Lys 460 in Pol κ is too far away to impose any direct effect on the template strand. Distances are reported in Å, and the electrostatic bonds are highlighted pink.

stabilizing effects of a direct salt bridge to the LF domain. Another residue that is located in proximity to the minor groove of the T_{-1} phosphate, Arg 420 in human Pol κ (Arg 419 in mouse Pol κ, not shown) has been modeled to aim away from the negatively charged backbone (70).

A closer look at the human Pol κ complex map shows that the electron density for this residue is not present, suggesting that the side chain may adopt multiple alternative conformations (70). In fact, there is negative (unallowed) difference density where they modeled their guanidinium group, suggesting that an alternate conformation exists for the side chain than the one which has been deposited. Since the structure is solved to 3.1 Å, the flexible arginine side chain is not visible. It is, therefore, reasonable to assume that Arg 420 places the guanidinyl group closer to the phosphate moieties, providing a movable "backrest" for the DNA binding channel rather than any other influential role. Overall, this lack of stable coordination of the T_{-1} base may be one of the reasons for the unique error-prone misalignment activities of Pol κ. When Lys 460 (mouse Pol κ) is replaced by a valine and an arginine is introduced in the position of Leu 507, Pol κ extends more correctly from an error-free εdA:dT base pair, and the rate of semi-productive complexes caused by an unstable primer terminus is reduced.

A significant disruption of the preferred template realignment schemes was not observed, despite the fact that the mutated residues are in direct contact with the template strand. In fact, it was the opposite primer strand which was mostly affected. The mutations abolished the error-prone primer slipping events and stabilized extension in an error-free manner. Does the human Pol κ complex structure provide any insight for this activity? In their discussion of their ternary complex structure, Lone *et al.* contend that

the amino acid composition of the region in proximity to the primer terminus is identical to that found in other Y-family DNA Pols (70). However, a comparison of the thumb domains of Pol κ and Dpo4 reveals that an unconserved β-turn between helices αM and αN (Pol κ) is within hydrogen bonding distance to the P_{-2} phosphate, and a concave pocket in the vicinity of the P_{-1} base. In Dpo4, a type II β turn is present with the sequence AD<u>VPGI</u>G (Dpo4 residues 181-187, Figure 3.11B), where the underlined sequence is the canonical 4-residue β-turn (Figure 3.11A). The turn harbors a divalent Mg^{2+} ion surrounded by 4 waters and the carbonyl oxygens of Ala 181 and Ile 186 in an octahedral arrangement, with one of the waters within hydrogen bonding range (3.3 Å) to the phosphate minor groove oxygen of the P_{-2} nucleotide, located one base downstream of the primer terminus. This oxygen is also located 2.8 Å from the backbone nitrogen of the strictly conserved Gly 185 in this β-turn.

Pol κ has a slightly different sequence, RK<u>VSGI</u>G (human Pol κ residues 352-358), and contains a Ser (Ser 355) in place of the constrained Pro. The non-canonical phi and psi angles also result in a loop which is wider than normal (3.4 Å instead of 3.1Å), though the carbonyl oxygens of Arg 352 and Ile 357 still coordinate a water molecule which occupies the same position as a water molecule in the Dpo4 Mg^{2+} coordination sphere. An additional carbonyl oxygen (from Val 354) points towards this water molecule, which also contributes to the widening of the turn. Due to the low resolution of the human Pol κ complex structure (70) (3.06 Å), no other water molecules are visible (though a highly ordered water molecule would be at this resolution), and it is evident that the Mg^{2+} ion is missing even though the DNA primer strand superimposes nearly identically in the Dpo4 structures. However, a well ordered Arg 352 guanidinium group

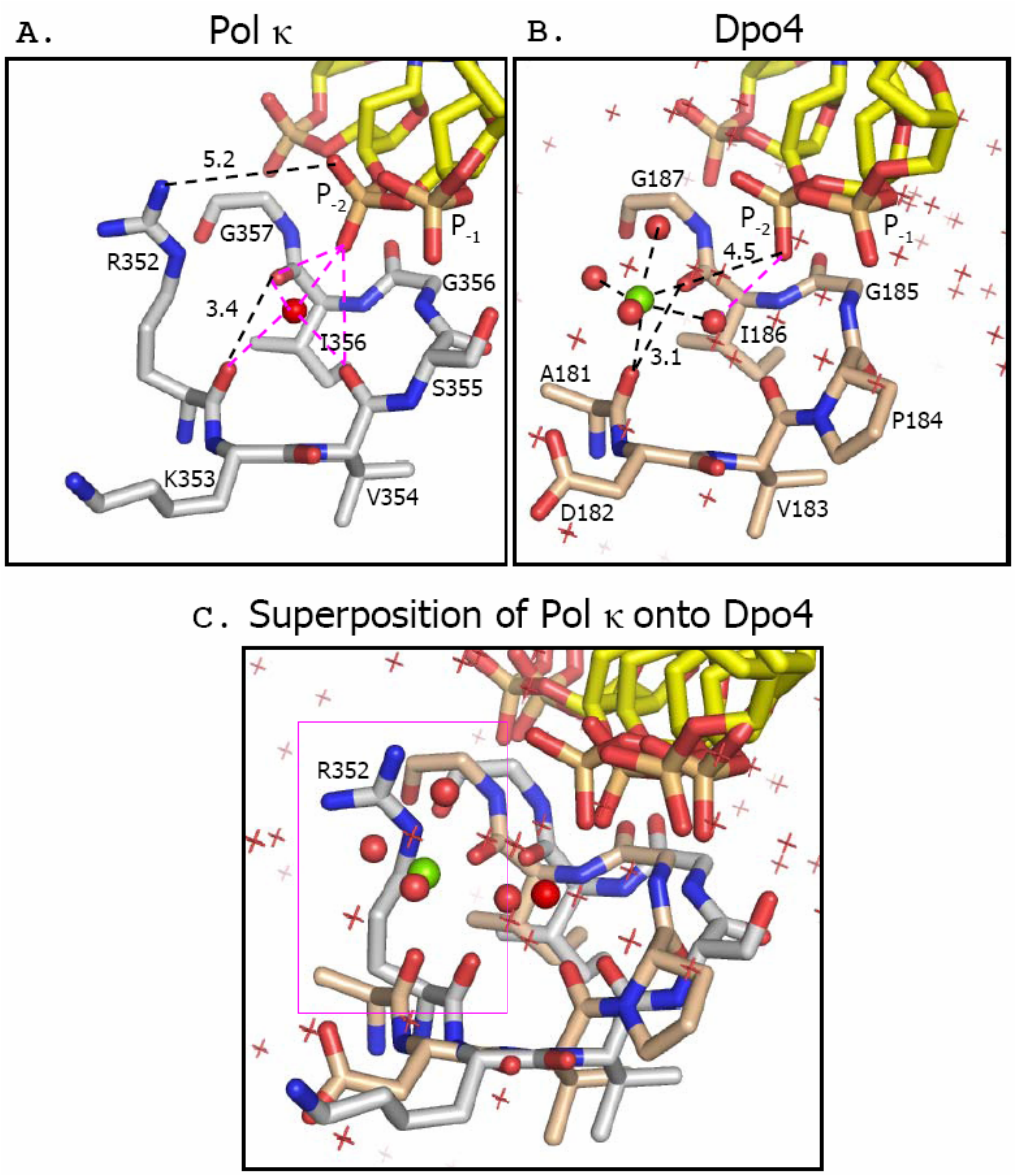


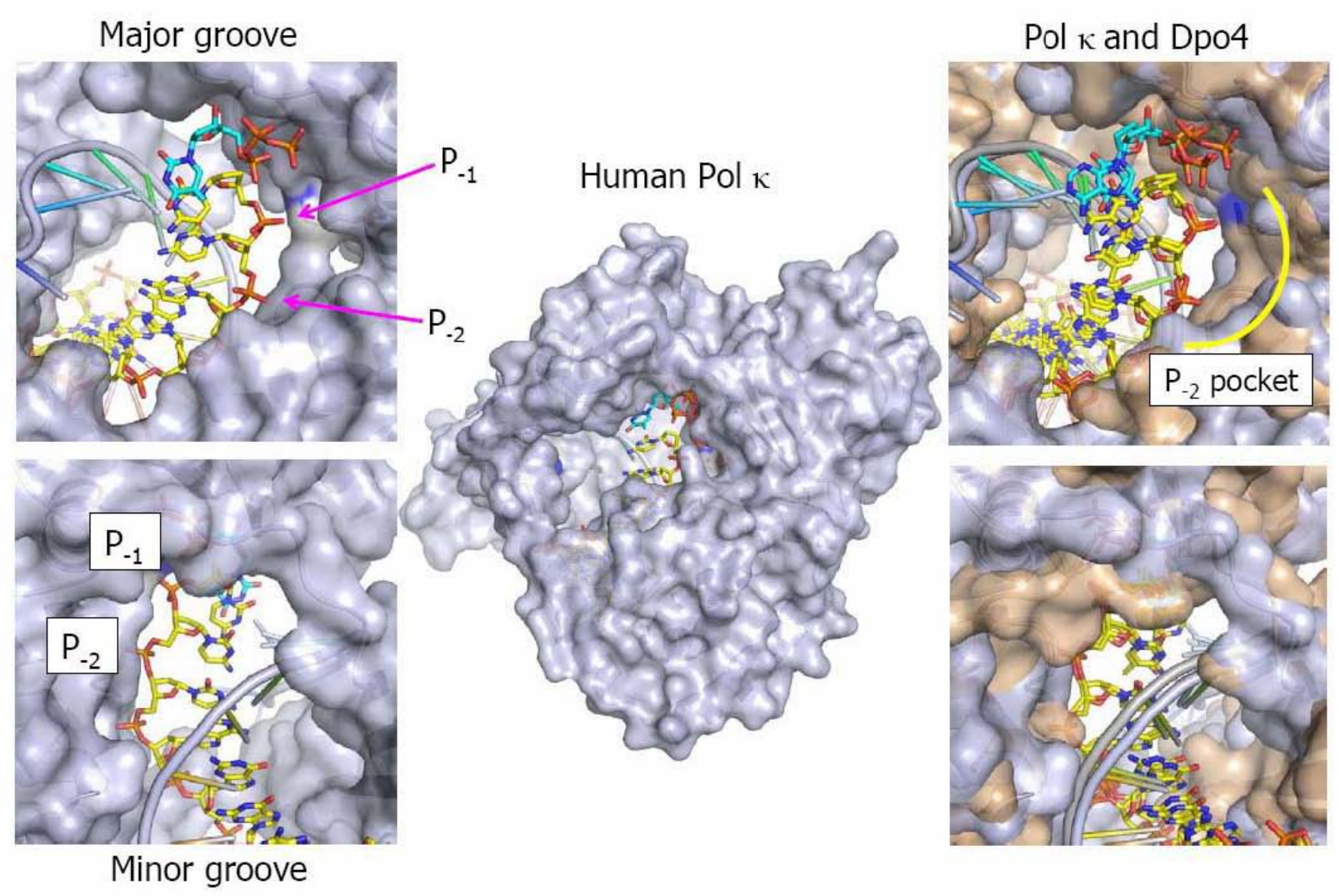
Figure 3.11 Unconserved β-turn in the thumb domain is sensitive to template destabilization. Stick representation of the β-loop and yellow primer strand bases P_{-1}–P_{-3} in **(A)** Pol κ (white) and **(B)** Dpo4 (beige). Pink dotted lines indicate hydrogen bonds, while black bonds provide distances for references purposes. Red spheres and stars, water molecules; green sphere, Mg^{2+}. **(C)** Pol κ contains an unconserved Arg (pink box) leading to the thumb domain β-loop which may replace or interfere with a similar coordination of Mg^{2+}atoms as observed in Dpo4, thus destabilizing the coordination of the primer strand.

is located directly in the area where the Mg^{2+} coordination network is positioned in Dpo4, and the arginine side chain is expected to interfere with the formation of a similar structural element in Pol κ. This Arg, while conserved in all eukaryotic Pol κ enzymes, is replaced by shorter residues in the other Y-family Pols such as alanine in Dpo4, Thr in yeast Pol η, and Asp in yeast Rev1, all of which preserve Mg^{2+} cation coordination. All of these observations imply that this loop may be less rigid in Pol κ than in Dpo4, loosening its restraint on the primer strand. Since there is only one observable contact with the P_{-1} phosphate in the minor groove (with a conserved Lys in the palm domain), it is also likely that the primer terminus is destabilized as well.

Therefore, it seems plausible to suggest that the combination of the generally open nucleotide pocket and a loosely held base in the T_{-1} position may provide enough freedom and entropy for the primer terminus to base pair promiscuously., there is also enough space available in the major and minor grooves for this movement, as shown in Figure 3.12, by comparing the superimposed solvent accessible surfaces of human Pol κ and Dpo4. This is best represented by comparing the two bottom panels, as the primer phosphate backbone is clearly visible in the human Pol κ model, but obscured in the Dpo4 model from this perspective of the minor groove.

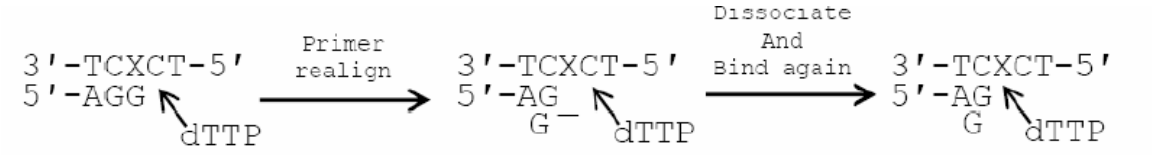
Once the primer is realigned, dTTP is incorporated opposite to the εdA lesion. The first assumption that has to be made is that the terminal guanine 3′-OH has to be in the correct position for dTTP incorporation, which can only be achieved if the loosely held P_{-2} is extruded. Otherwise, there is no primer terminus available to attack the incoming nucleotide. Since our kinetics studies showed that template strand coordination is weaker in the Pol κ LF domain as compared to Dpo4 (and yeast Pol η), the increased

Figure 3.12 (next page) Surface representations of human Pol κ and Dpo4 in proximity to the DNA primer strand. (Center) View of the complete human Pol κ surface (70) (grey) from the most exposed route of entry for incoming nucleotides. **(Top)** Close-up view of the major groove. Note the wider pocket to the right of the P_{-2} base. The conserved Lys 321, which creates a salt bridge to the P_{-1} phosphate is shown as a blue patch on the surface of the enzyme. **(Bottom)** View of the accessible minor groove towards the active site. **(Left)** Close-up views of human Pol κ; **(Right)**, close-up of same view superimposed onto the Dpo4 surface (beige, (67)). The surface that protrudes the most is visible, and differs between the two enzymes. Note how the VPGI β-turn in Dpo4 occupies more space than the concave P_{-2} pocket of Pol κ.


Major groove
P-1
P-2
Human Pol κ
Pol κ and Dpo4
P-2 pocket
P-1
P-2
Minor groove

entropy in the vicinity of the T_0 and T_{-1} sites may allow the εdA in the T_{-1} position to release the base pair with the dG terminus and orient itself towards the dTTP waiting in the active site. This may lead to a non-productive and non-processive complex, which is then released by the polymerase. In a second step, Pol κ binds to the realigned DNA with the εdA newly presented to the active site, allowing dTTP incorporation.

There is no evidence that other Y-family polymerases can process backwards along the template strand, as movement back along the DNA channel would require a movement of the T_{+1} dCTP out of the active site back into the Met134/Phe49 steric gate. This action would be energetically prohibitive and sterically not possible, and therefore dissociation seems to be the more likely event (below). When Arg 507 was substituted



for the hydrophobic WT Leu 507 in our mutagenesis study, the εdA:dG pair is stabilized in the T_{-1}/P_{-1} position by the Arg guanidinium group, decreasing the local entropy, and the εdA (in the T_{-1} position) is restricted from "sensing" the dTTP in the active site. Therefore, the primer cannot be released and realigned, making a slip-free process more likely.

In a recent crystal structure of Pol ι, it was shown that εdA:dC forms a stable Hoogsteen base pair, imposed by the human Pol ι active site (82-84) and leading to efficient extension. Correct extension past εdA:dC however, was the lowest activity observed in Pol κ among those qualitatively tested, suggesting that a Hoogsteen base pair is not an acceptable means of bypassing εdA for Pol κ. While εdA:dG and εdA:dT are

known to adopt a Hoogsteen configuration (15, 64), the sequence context in our experiments permitted misalignment mechanisms to extrude the lesion and facilitate extension. Pol κ is known to insert dAMP opposite 8-oxo-dG (a classic Hoogsteen configuration). However, Jaloszynski *et al*. also reported that this action was slightly inhibiting (46), but did not comment on the fact that their highest enzyme specificities occurred when a dT was neighboring the lesion. If a Hoogsteen base pair can not be well accommodated in the active site of Pol κ, as suggested by Wolfle *et al*. (121), the only way to achieve efficient bypass may therefore be a frameshift mechanism.

In conclusion, the unconserved little finger domain of Pol κ plays a major role in template strand stability, which then drives extension. Due to the relaxed energetic environment of the active and T_{-1} extension sites, Pol κ is prone to misalignment errors as supported in the literature. Further mutagenesis studies of the putative extended water-bridged hydrogen bonding network involving residues Lys 460, Lys 458, and Glu 419 may be necessary to shed light on the way the template slipping mechanism could be modulated.

CHAPTER FOUR

The Binary Complex Structure of Mouse Polymerase κ-I and DNA

Chapter 4: The Binary Complex Structure of Mouse Polymerase κ-I and DNA

I. Introduction

Y-family DNA Pols have spacious active sites that are able to accommodate and synthesize past DNA lesions, and the conservation of the TLS pathway through all organisms illustrates its indispensable role in sustaining life. While archaebacteria must rely on a single TLS Pol to catalyze bypass upon every lesion it encounters, higher eukaryotes have evolved Pols with specialized active sites that accommodate only specific lesions. This specialization is conferred by unconserved specific residues that form various intermolecular contacts with the DNA template strand to select only sterically favored lesions for entry into a largely conserved active site. The previous chapter identified residues in the Pol κ LF domain that influence the stability of the template base through important electrostatic and van der Waals interactions and thereby alter bypass fidelity and modulate frameshift mechanisms.

While the homology model provided very important information with respect to the catalytic core of mouse Pol κ, it is nevertheless "only" a model and was not a complete structure of the enzyme. Nearly 200 residues were missing between the N-terminus and the end of the LF domain. These missing regions are insertions in the eukaryotic Pol κ sequence which include the novel N-terminal clasp domain, but also a 70-residue disordered sequence known as the "tendril" domain, which are clearly not required in the minimized archaeal enzymes. It is plausible that these unmodeled regions may also govern lesion specificity in eukaryotic Y-family Pols. Therefore, to gain full insight into the catalytic mechanism of mouse Pol κ, a structure of the enzyme in complex with DNA was pursued. Subsequently, Uljon *et al.* determined the crystal

structure of the catalytic core of human Pol κ (residues 68-526) (110). It depicted a "classic" Y-family DNA Pol, comprised of the palm, fingers, and thumb domain, connected by a flexible 15-residue linker region to the LF domain. It also revealed one of the three helices found in the 100 amino acid N-terminal sequence, observed only in eukaryotic Pol κ, which folds into an α-helix and associates with the thumb domain. In the structure of the apoenzyme (Figure 4.1), the concave LF β-sheet, known to bind within the DNA major groove, was packed against the non-catalytic surface of the palm domain at two different angles, leaving the active site completely solvent exposed and prepared for DNA substrate binding. Another interesting feature of the apoenzyme structure was the visualization of part of the "tendril" region, a predominantly disordered 60-residue sequence within the palm domain. However, due to its flexibility, the resulting model of this part of the polypeptide chain was fragmented and disconnected from the main protein, which may be due to the fact that a key motif was not present since a truncated construct was used for crystallization.

A few observations posed concerns regarding the relevance of the apoenzyme structure. Primer extension assays conducted in the presence of all four dNTPs demonstrated that the Pol κ construct (analogous to residues 67-525 in mouse) had only 0.05% the activity of the wild type protein and a C-terminal truncation construct, Pol κ (19-526), which is equivalent to the WT protein in activity. Uljon *et al*. predicted the presence of only one α-helical motif which spans residues 19-65 within the 100-residue N-terminal domain and omitted it from their crystallization construct (110), but the secondary structure prediction algorithms nnPredict and DisEMBL suggest three folded elements spanning from residues 19-95. Beginning their construct at residue 68 preserved

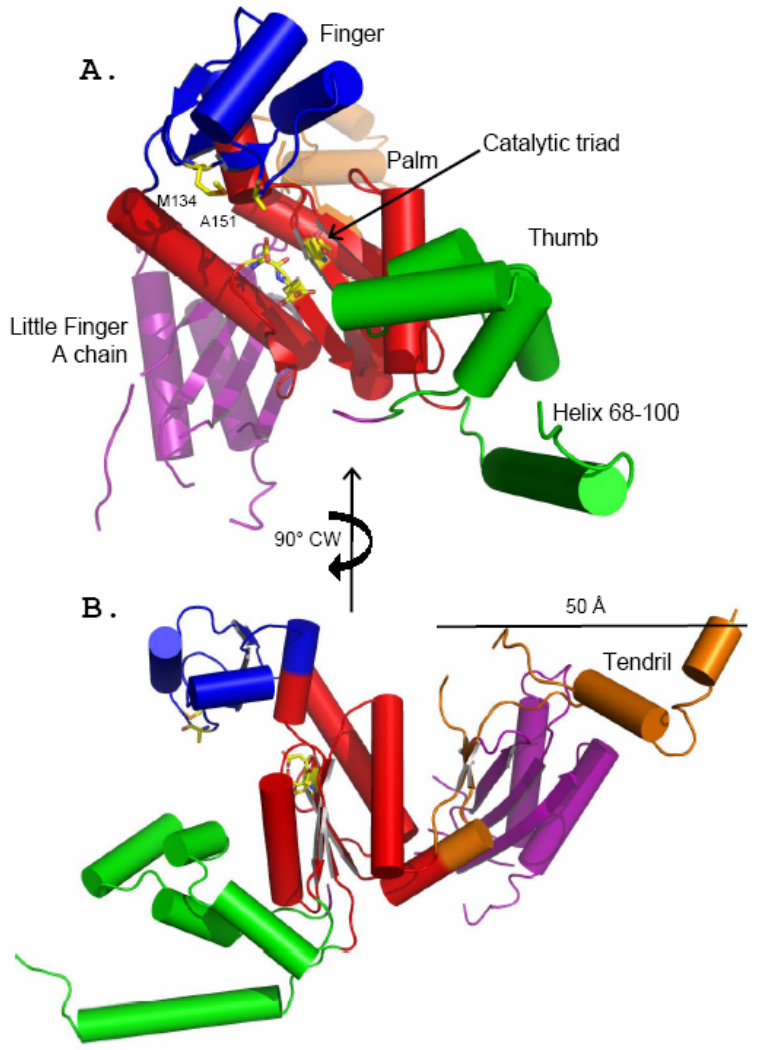


Figure 4.1 Cartoon representation of apo human Pol κ. α-helices are represented as cylinders and β-strands as arrows; domains are colored as described on page (xii). **(A)** In the apoenzyme, the LF domain (purple) stacks against the non-catalytic surface of the palm domain, completely exposing the active site. The acidic catalytic triad and roof residues M134 and A151 have been highlighted. The novel helix (residues 68-100) is also visible, and is part of the thumb domain. The orientation of the first molecule (chain A) is chosen, where the LF domain β-sheet comes in contact with the palm domain **(B)** Side view. The novel tendril region (orange) extends 50 Å from the palm domain. Image reproduced from Uljon *et al.*, Structure (2004) (110), and recolored using PyMol.

the third of the three helices, but it appears that this was not sufficient to retain full enzymatic activity. In addition to being catalytically compromised, the apoenzyme structure provides no information regarding protein-DNA interactions, and little was learned to explain Pol κ's lesion specificity or its error-prone catalytic mechanisms.

In order to better characterize the structural basis of lesion specificity catalyzed by eukaryotic Pol κ, a longer Pol κ construct, residues 36-524, that maintained activity that was comparable in activity to WT protein was used for co-crystallization studies with DNA. In the binary complex structure, the catalytic core appears to have dimerized, blocking the catalytic triad; however, the LF domain is still bound to the DNA major groove. Parallel to our structure solution, the same group that solved the apoenzyme structure published the structure of human Pol κ (residues 19-526) in a ternary complex with DNA and an incoming nucleotide (70). Their longer construct retained all three N-terminal helices, and thus retained full activity. Superposition of our structure with their apo- and DNA-bound models showed that our DNA-bound LF domain is located in between the LF domain locations of the apo- and DNA bound models, suggesting that we may have solved the structure of a transient loading intermediate.

II. Methods

A. Rationale of truncation constructs

To increase the probability of obtaining crystals of mouse Pol κ, eleven truncation constructs were designed in addition to the His-tagged (pET-16b) and tagless ($pTXB_1$) full length constructs. These truncation constructs, labeled A-K, were designed from secondary structure predictions (nnPredict) and disordered loop predictions (DisEMBL), and also from previous reports using active Polκ$\Delta C_{561-852}$ and Polκ$\Delta C_{617-852}$ proteins. In Figure 4.2, a schematic of the truncation constructs illustrates which domains were preserved. Briefly, all constructs included motifs I-V and the LF domain, but varied in length within the 100-residue N-terminal region and the C-terminal sequence after the LF domain.

B. Cloning and protein expression of Pol κ truncation constructs

Truncation constructs were created by introducing NdeI/SapI restriction endonuclease (RE) sites into the full length mouse Pol κ cDNA ($pTXB_1$) at the desired N- and C-termini, respectively (Appendix B.2). PCR amplification and restriction endonuclease digestion yielded the internal truncation product and two flanking sequences (or one sequence, if the truncation product began at the N-terminus) that were identified by size on an agarose DNA gel. The correct DNA fragment (as determined by length) was isolated from the gel by the HighPure PCR purification kit (Roche), subcloned into the NdeI/SapI sites of the $pTXB_1$ vector (IMPACT-CN^{TM}, New England Biolabs), and transformed into BL21(DE3)-codon plus-RIL competent cells (Stratagene). Cells harboring each construct were grown at 37°C in 50 ml 2xYT bacterial media to an

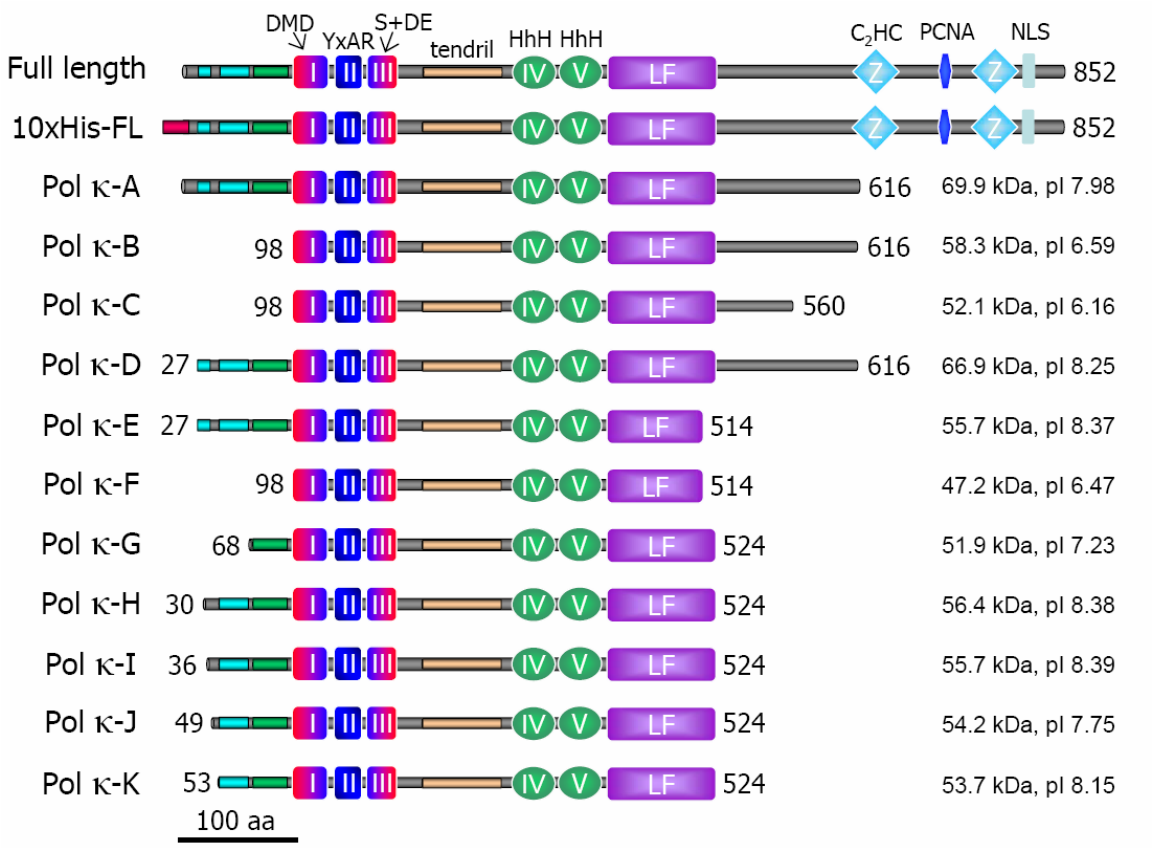


Figure 4.2 Schematic view of mouse Pol κ crystallization constructs including their molecular weight and pI. Motifs I-III contain the acidic catalytic triad (DM**D** S+**DE**) conserved in all DNA Pols, and these three motifs create the palm and finger domains. The thumb domain is comprised of two helix-hairpin-helix motifs (HhH). After the little finger domain, the sequence is significantly disordered except for two C2HC zinc fingers (cyan) whose presence are not required for full polymerase activity. Pol κ also contains a 100-residue N-terminal sequence that folds into three helices. The function of this domain was not known until recently. The tendril is a disorganized region that extends from the non-catalytic face of the palm domain, and its function is unknown. The domains match the coloring convention as described on page (xii).

OD_{600} of 0.5-0.6 was obtained and then induced with 1 mM IPTG, whereupon the temperature was lowered to 30°C and incubated for an additional 10 hours for protein synthesis. Test expressions showed that only two of the constructs expressed readily in the soluble fraction: Construct A (residues 1-616, Pol κ-A), and construct I (36-524, Pol κ-I). Interestingly, our attempt to copy the Uljon construct (68-524, construct G) did not yield soluble protein with the $pTXB_1$ expression vector, suggesting that their truncated construct remained soluble due to an uncleaved Glutathione S-transferase (GST)-tag (110).

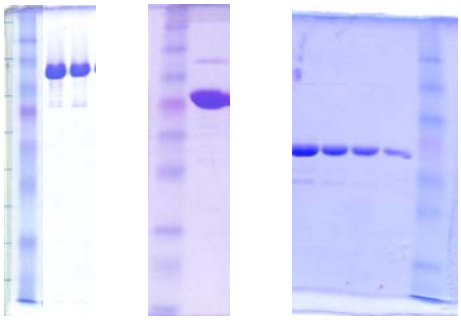
For crystallization trials, full length Pol κ ($pTXB_1$) was purified as described in Chapter 3, except that a 20 l cell culture (4 l batch x 5) was incubated in a BioFlo 4500 bacterial fermentor (New Brunswick Scientific, Edison, NJ). Since the fermentor provided a better growth environment for the transformed cells by regulating dissolved oxygen and temperature, average yields increased to 2.5 mg Pol κ/l culture. Samples were dialyzed against a storage buffer that contained 200 mM NaCl because of the high protein concentration and concentrated by PEG bath to >20 mg/ml. All samples were frozen at -80°C and thawed later for crystallization trials.

Since Pol κ-A has a calculated pI of ~8, a different purification protocol was required. Most of the steps were preserved with the following changes. Cells (wet pellet weight 3.5 g) harvested from a 4 l culture were resuspended and purified with 25 ml acidic chitin column buffer (CCB-A; 20 mM Na Phosphate, pH 6.0, 400 mM NaCl, 1 mM EDTA, pH 8.0). The intein cleavage reaction with 75 mM DTT was extended to 24-30 hours at 4°C. The cleaved protein was eluted from the column with 90 ml CCB-A, pooled, and dialyzed overnight against 1 l of 20 mM Bis-Tris, pH 6.0, 100 mM NaCl, and

3 mM DTT, with two buffer exchanges. The dialyzed sample was then centrifuged to remove particulate material and applied to a 25 ml Source S cation exchange (CE) FPLC column (Amersham) with buffers CE-A (20 mM Bis-Tris, pH 6.0, 3 mM DTT) and CE-B (CE-A + 1 M NaCl), where Pol κ-A eluted at 180 mM NaCl. Fractions containing Pol κ-A were pooled (average concentration of pool ~2.3 mg/ml) and quality of the sample was characterized by Dynamic Light Scattering (DLS), which measures the polydispersity of the hydrodynamic radii of dominant species of a sample. Finally, the protein was concentrated by PEG bath (20% PEG 35000, 20 mM Bis-Tris, pH 6.0, 200 mM NaCl) to ~20 mg/ml, and flash frozen for later use. Typical final yields of Pol κ-A were 4-5 mg protein/l culture using this protocol.

Pol κ-I had a calculated pI of 8.39. Its purification protocol was identical to the wild type, except that the clarified cell supernatant was applied to only 10 ml chitin resin (instead of doubling the instructions from before), and that the Na Phosphate and Tris-HCl buffers throughout the protocol were at pH 7.0 instead of 8.0. The intein cleavage reaction required 18 hours at 4°C with CCB supplemented with 75 mM DTT, similar to the wild type protein. Typical yields using this protocol were 6-8 mg Pol κ-I/l culture. In each case, the protein was aliquoted in 15 µl drops, flash frozen in liquid N_2 and stored at -80°C for later use. Representative PAGE gels show purity of samples below.

Representative gels from purification of the three mouse Pol κ constructs, WT, Pol κ-A and Pol κ-I, respectively.



C. Activity of crystallization constructs

To confirm the activity of the Pol κ-A and the Pol κ-I constructs, primer extension assays with full length Pol κ-A and Pol κ-I were conducted in the presence of either all four dNTPs or the next incoming nucleotide; dGTP, using the same reaction conditions as published for the apoenzyme structure construct. The DNA substrate was composed of a 29-mer oligonucleotide primer (5′-GAATTCCTGCAGCCCAGGATCGACTCGTA-3′) annealed to a 54-mer template (5′-ATTCCAGACTGTCAATAACACTC-GGTACCAGTCGATCCTGGGCTGCAGGAATTC-3′). Increasing concentrations of enzyme (10 nM - 100 nM) and 10 nM radiolabeled P/T were mixed with 25 mM Tris-HCl buffer, pH 7.5, 5 mM $MgCl_2$, and 1 mM DTT in 9 µl volume. The extension reaction was started by adding 1 µl of 10x 4-dNTP stock (final concentration, 250 µM of each deoxynucleotide) and incubating it at 37°C for 10 min. Extension reactions were quenched by 20 µl Stop solution, and products were resolved on a 20% D-PAGE gel.

Since only the Pol κ-I led to crystals in the presence of DNA, the extension product catalyzed by mouse Pol κ-I was analyzed by time course primer extension experiments with either the next incoming nucleotide or all four dNTPs using ^{32}P-radiolabeled crystallization substrate under the same time and temperature conditions as that prior to crystallization. Samples were taken directly after the addition of nucleotides (~5 sec), and then 1, 2, 5, and 10 minute post addition. All samples were quenched and analyzed as described previously.

D. Co-crystallization and data collection of the Pol κ-DNA complex

Since Pol κ has a tendency to frameshift in a sequence-dependent manner, the crystallization oligos were designed to prohibit slippage in either direction. The 16-mer primer 05-154 (5′-GTCCTGTTCGTCGTGC-3′) and 23-mer template 05-155 (5′-ACCGCACGCACGACGAACAGGAC-3′) were obtained from Dr. Charles Iden's Laboratory, Stony Brook University. Oligos were dialyzed against 1 l 0.22 μm filtered water overnight (with at least one dialysate exchange) and concentrated by Speedvac (ThermoSavant). The two samples were resuspended in 0.22 μm sterile filtered water to a final concentration of 10 mM primer and 11 mM template. The oligos were mixed in the following ratio: primer : template : buffer = 1 : 1 :1.33 with the buffer containing 30 mM Tris-HCl buffer, pH 7.5 resulting in a final annealed P/T concentration of 3 mM (10x P/T stock) in 10 mM Tris-HCl buffer, pH 7.5. The oligos were heated to 90°C for 10 minutes and allowed to cool slowly in a heat block removed from its heat source to maximize annealing.

The truncation constructs of mouse Pol κ were pursued after attempts at crystallizing the full length protein alone or in complex with different lengths of DNA were unsuccessful (Appendix B.3). For co-crystallization of Pol κ-A with the 16/23 P/T, protein (277 μM) and DNA (3 mM P/T stock) were admixed in a molar ratio of protein:DNA = 1:1.1, then diluted with Pol κ-A storage buffer to a final concentration of 138 μM (9.8 mg/ml). The sample was dialyzed in a microdialysis button with a 1 kDa MWCO dialysis membrane against 20 mM Bis-Tris buffer, pH 6.0, 100 mM NaCl, 2.5 mM $MgCl_2$ and 1 mM DTT for 1 hour at 4°C to lower the salt concentration and promote complex formation. To create the ternary complex, which stabilizes the active

polymerase for crystallization, the sample was mixed with 15-fold molar excess of the next incoming nucleotide dGTP and incubated at 30°C for 10 min. The sample was then centrifuged at 25000 x g at 4°C for 30 min to settle any debris (dust or precipitate), and used immediately for crystallization. The Pol κ-I ternary complex was prepared identically (protein stock, 497 μM), except that Tris-HCl, pH 7.0 buffer was used instead of Bis-Tris 6.0 in the dialysis buffer.

Fourteen promising conditions for ternary complexes with both constructs were initially identified with incomplete factorial screens Optimix I and II in a micro-fluidic chambered Topaz® 4.96 interface diffusion chip (Fluidigm, San Francisco, CA). This method screens nanoliter quantities of enzyme (up to 4 samples) by injecting it into an etched hydrated silicone chip pre-filled with 100 potential crystallization reagents; this permits controlled diffusion of the two solutions. When these conditions were translated to hanging drop vapor diffusion experiments, only 4 conditions reproduced microcrystals. In hanging drop setups, 1 μl of protein was overlaid with an equal volume of reservoir solution on a silanized glass cover slide. The 2 ul drop was quickly inverted and sealed with vacuum grease over a reservoir of 1 ml crystallization reagent, and then incubated at the desired temperature in a minimally vibrating environment until crystals formed. Both constructs led to crystals in solutions refined from Optimix-II #35 (25% PEG 1500, 100 mM Tris-HCl, pH 8.5, and 100 mM $MgCl_2$). Hexagonal plate crystals of a putative Pol κ-A ternary complex grew at 15°C in a reservoir solution containing 17-25% PEG 1500, 100 mM Tris-HCl, pH 8.5, and 100 mM $MgCl_2$, the largest crystals growing in 20% PEG 1500. Small plates first appeared after a week, and grew to their full size (120 x 120 x 20 μm^3) in a month (Figure 4.3A). Pol κ-A microcrystals also formed in 25% PEG 1000

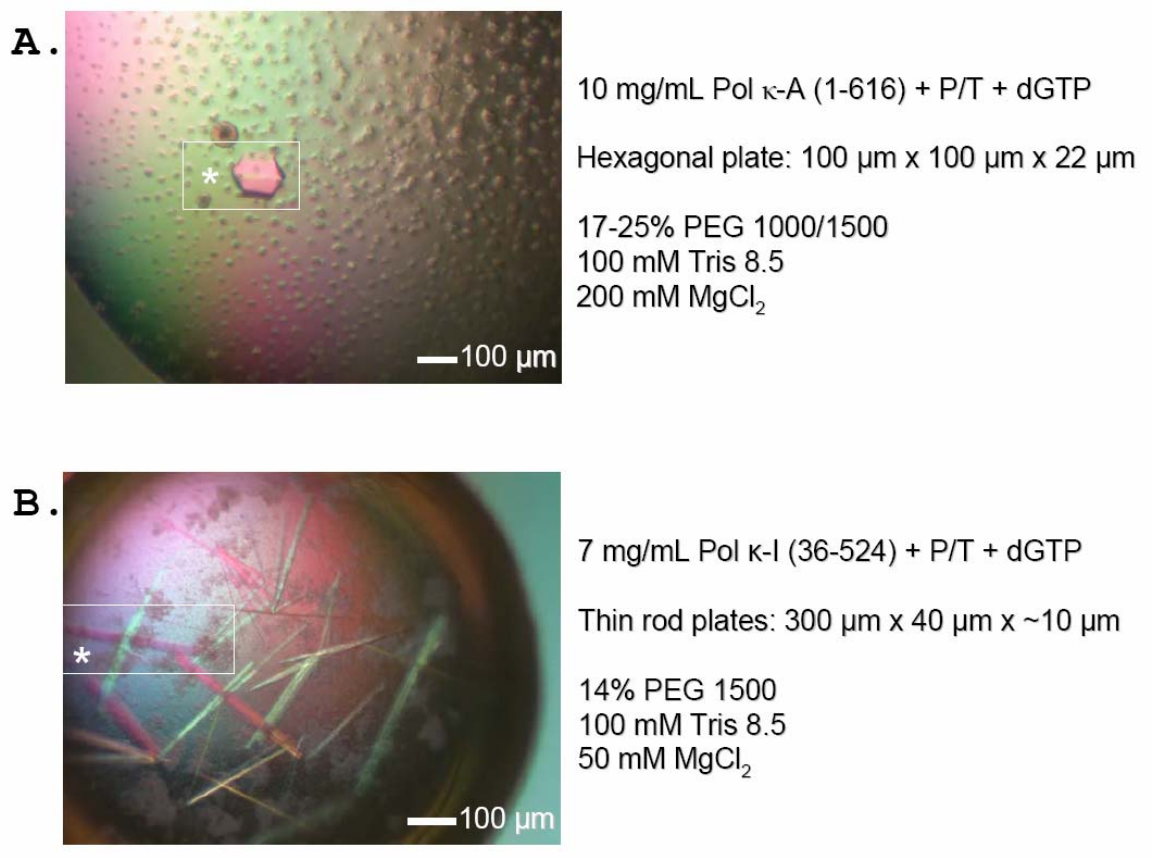


Figure 4.3. Crystals of putative Pol κ-A and Pol κ-I ternary complex. Asterisks indicate the crystal used for diffraction studies. Pictures were taken using a circular polarizer, which created the colored effect in these crystals. **(A)** Hexagonal crystal of Pol κ-A. The best crystal diffracted only to ~9 Å. **(B)** Thin rod-shaped crystals of the Pol κ-I ternary complex. A dataset was collected to 3.0 Å resolution.

and 300 mM Ca Acetate and in 30% PEG 4000, 100 mM Tris-HCl, pH 8.5, and 200 mM $MgCl_2$, but the crystal size could not be improved despite many optimization attempts such as microseeding, macroseeding, or changes in environmental parameters. Subsequent follow-ups in the absence of DNA did not yield crystals, which suggested that it could be a binary complex at minimum.

Long, very thin rod-like plates of Pol κ-I ternary complex crystals were grown at 22°C in a reservoir solution containing 14% PEG 1500, 100 mM Tris-HCl, pH 8.5, and 50 mM $MgCl_2$. These crystals (>300 x 22 x ~10 μm^3, Figure 4.3B) grew suddenly from small clouds of precipitate after incubation for three weeks, and were unfortunately never reproduced after numerous attempts.

For data collection, individual crystals were cryoprotected by transferring the crystals into solutions consisting of their respective mother liquors supplemented with 1 mM dGTP and glycerol in 5% steps up to 20% glycerol, and flash cooled in liquid N_2. Initial diffraction studies of the Pol κ-A and Pol κ-I crystals were conducted at beamline X26C at the National Synchrotron Light Source at Brookhaven National Laboratory at a wavelength of 1.1 Å and 100 K on a Quantum IV ADSC CCD detector. The Pol κ-A crystals only diffracted to 9-10 Å; crystals also did not perform well after a flash thaw/freeze cycle (annealing). Other growth variations, such as microseeding or changing the incubation temperature, or simply larger crystals did not improve crystal quality or diffraction. The Pol κ-I crystal diffracted to 3.0 Å, but the reflections were very weak. Final diffraction data for the Pol κ-I crystal were collected at a more intense x-ray beam at beamline X29 and 100 K at a wavelength of 1.1 Å on an ADSC Quantum 315 3x3 detector. Due to the alignment of the thin crystal along the axis of the X-ray beam, the

goniometer was tilted to κ=50° so that the beam could be directed diagonally to pass through as much of the crystal as possible. During data collection, the reflections displayed a strong diffraction anisotropy (non-spherical diffraction), diffracting to 3.0 Å along **a*** (vector in reciprocal space perpendicular to the **b** x **c** plane) and **b***, but only to 4.4 Å along **c***, suggesting fewer crystal contacts along that vector. Radiation damage from the intense x-rays also decreased the quality of the crystals and subsequent diffraction over time.

E. Structure determination and refinement of the Pol κ-I/DNA complex

The Pol κ-I dataset was indexed, integrated, and scaled using the HKL2000 suite (94). Due to the severe diffraction anisotropy, the dataset was truncated to include reflections only up to the 3.5 Å spherical resolution shell to attain the highest data completeness (96%) and to maintain a signal-to-noise ratio (I/σI) cutoff above 2.0.

Since it was very likely that the position of the little finger domain relative to the remainder of the protein was altered in comparison to the apoenzyme structure, only the palm, fingers, and thumb domains (residues 100-118, 180-215, and 290-400) of the apoenzyme structure (PDB entry 1T94 (110)) were used as an initial search model to obtain the phase information by molecular replacement (MR). Both MR programs PHASER(76) and MOLREP (111) did not identify reasonable results, but led to models with overlapping thumb and finger domains for the two monomers in the asymmetric unit in a manner reminiscent of a "yin-yang" shape. New search models excluding either the thumb or finger domains were used for the next round of molecular replacement calculations, and a visual comparison of the electron density maps with the Cα trace

obtained by molecular replacement confirmed that the thumb domain was present. A second round of MR was performed to find the two LF domains, but only one could be located. This LF domain, however, was connected to a symmetry-related neighbor outside the asymmetric unit. In order to obtain the position of the little finger domain for one of the two monomers in the asymmetric unit, a symmetry operator had to be applied using the program PDBSET. A third round of MR to identify the position of the finger domains and the missing LF domain in the second monomer was unsuccessful.

Additional electron density, which was not explained by the search model, identified the position of the novel second N-terminal helix (residues ~50-72) which was not present in the previously published apoenzyme structure. Initial rigid body, simulated annealing, and translation/libration/screw (TLS) motion determination refinement was performed with the program suite CNS (11). Rigid and TLS domains were defined by subdomain architecture: TLS group 1 (palm), residues 100-117, 170-210 and 290-339; TLS group 2 (thumb), residues 74-99 and 340-399; and TLS group 3 (LF-A), residues 414-514. The N-terminal helix and thumb/LF linker were built and loops within the LF domain rebuilt using the program O (51), and the human Pol κ residues were replaced by the mouse Pol κ residues and the new side chains fitted into the electron density using the program COOT (19). By employing non-crystallographic symmetry restraints, significant improvement of the electron density map revealed difference density for 5 primer strand phosphates and 8 template strand phosphates. An idealized model of the P/T extended product was generated by the MAKE-NA server (http://structure.usc.edu/make-na/), manually adjusted to fit the phosphate density, and refined with the program REFMAC5 (79, 120). Refinement of the mouse Pol κ complex was achieved when the R_{work} (R_{free})

values remained constant at 28.0% (35.0%) and no longer improved or began to increase as a result of over-modeling.

When the human Pol κ ternary complex structure (PDB entry 2OH2 (70)) became available, all of the above steps were repeated using the palm, thumb, and LF domains of the new structure as a search model for MR calculations. This more complete model helped identify the true termini of our electron density, and also improved the model of the N-clasp αN2 helix. Several more residues were resolved at the finger junction and in the DNA substrate, but the model still gave high statistical errors due to the low resolution and the anisotropic nature of the diffraction dataset.

The high R_{work} (R_{free}) values are a result of two assumptions: (1) The crystal diffracted equally in each dimension, and so expected reflections are evenly distributed through the spherical resolution shell during scaling (depending on the space group), and (2) Reflections were more or less equally intense (again, depending on the space group) and averaged within the spherical resolution shell. However, in the Pol κ-I dataset, the sharp anisotropy (3.0 x 3.0 x 4.4 Å) resulted in an elliptically shaped dataset with clear areas of missing or weak reflections. Therefore, when the reflection intensities were averaged within these spherical shells, the higher resolution reflections were diluted because they are averaged against the virtually empty shell and the quality of the electron density map decreases.

To account for the anisotropic diffraction, a novel secondary elliptical truncation algorithm (http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/)(101) was used to further truncate the dataset by calculating the true resolution limits and removing the weakest reflections ($I/\sigma I<3.0$) outside of the new elliptical resolution shells. The ANISOSCALE

server then provided a new reflection list from which the new structure factors could be used for continued model refinement. The previous built mouse Pol κ model was refined against this new output, and the R_{work} (R_{free}) decreased from 28.0% (35.0%) to 27.7% (30.8%) within one round (15 cycles) of TLS and restrained refinement by REFMAC. The final refined structural model of Pol κ-I in complex with DNA had a final R_{work} (R_{free}) of 22.6% (29.6%).

Because of the low resolution of the structure, only 4 water molecules and one $H_2PO_4^-$ molecule were located manually. The stereochemistry of the final structure was analyzed with PROCHECK (62), and secondary structure elements were assigned by PROMOTIF v3 (44). The refinement statistics are summarized in Table 4.1. All figures were generated with PyMOL (16) (Delano Scientific LLC, Palo Alto, CA).

	spherical dataset	elliptical dataset
Data Collection		
Space Group	$P2_12_12_1$	
Resolution limits (Å)	10-3.5	15-3.51
R_{sym} [a]	0.204 (0.468)	0.204 (ND)
Signal-to-noise $\langle I/\sigma I \rangle$ [b]	6.2 (2.1)	6.2 (2.0)
Unit cell dimensions		
a, b, c (Å)	65.5, 129.6, 137.4	
α, β, γ (*)	90.0, 90.0, 90.0	
Completeness	0.976 (0.873)	0.808 (0.220)
Mean Redundancy	6.3 (5.1)	6.3 (ND)
Refinement		
Resolution (Å)	3.5	3.5
Number of reflections	13,426	11,439
R_{work} (R_{free}) [c]	0.280 (0.350)	0.226 (0.296)
No. of atoms	4807	5150
Protein	4560	4722
DNA	247	428
B-factors		
Protein	61.6	40.5
DNA	~70	~83
rms deviation		
bond lengths (Å)	0.015	0.010
bond angles (°)	1.66	1.39

Table 4.1 Data collection and Refinement statistics of the original and truncated mouse Pol κ-I datasets. [a] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_i - \langle I \rangle|/\Sigma_{hkl}\Sigma_i$, where I_i is the i^{th} measurement and $\langle I \rangle$ is the weighted mean of all measurements of I. [b] $\langle I/\sigma I \rangle$ indicates the average of the intensity divided by its standard deviation. [C] $R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$, and R_{free} is the same calculation on 5% of the reflections excluded from the original refinement to assess the potential of model bias. Numbers in parentheses refer to the highest resolution data shell in each data set unless otherwise defined. ND, not defined in the elliptical output file.

III. Results and discussion

A. Activity of the Pol κ A and I constructs

Since the Pol κ-I construct lacks part of the αN1 helix, it was necessary to analyze its enzymatic activity. Primer extension assays in the presence of all four dNTPs were conducted to compare the activities of full length, Pol κ-A and Pol κ-I to the apoenzyme truncation construct (68-524). The reaction conditions for human Pol κ differed from those described in the kinetics study. We therefore subjected the mouse Pol κ constructs to the same reaction conditions as published for the apo human construct to compare the enzymatic activities of the different constructs as closely as possible.

As shown in Figure 4.4A-B, the enzyme titration experiments qualitatively showed that WT and Pol κ-A can extend the full primer at a 1:1 molar ratio, while Pol κ-I reached the same level of extension at a 4:1 protein:DNA molar ratio. Furthermore, Pol κ-I is unable to catalyze the last incorporation event opposite the terminal base of the template strand, since a majority of the products appear to lack the last 1-2 nucleotides. WT and Pol κ-A, however, incorporate bases beyond the end of the template, perhaps by employing template-directed primer slippage mechanisms described earlier; previous studies have shown that Pol κ lacks terminal deoxytransferase activity (97). The longer protein constructs are much more active than the Pol κ (68-524) construct, which was used to obtain the apo enzyme structure (110). The loss of the first two helices results in a sharp decrease in activity, as demonstrated by the 60:1 molar ratio that was needed to obtain significant extension.

From these observations, it is clear that the second helix (residues ~50-72) in the N-terminal region plays an essential role in primer elongation, while the first helix plays

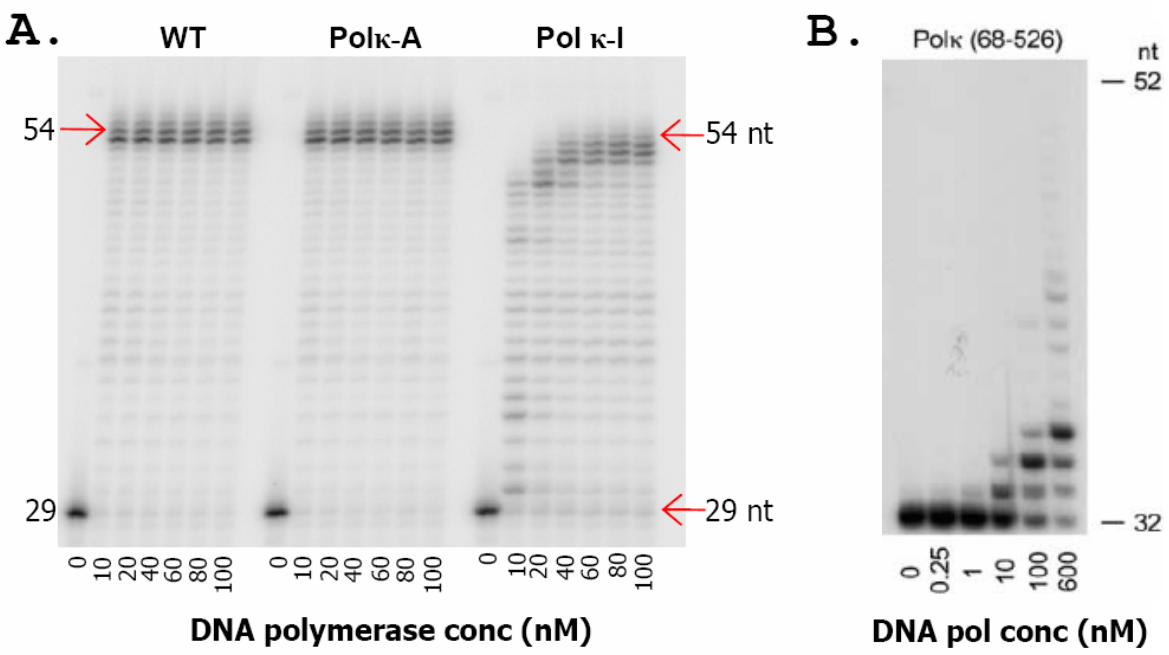


Figure 4.4 Relative extension activities of various Pol κ constructs. (A) wild type mouse Pol κ (1-852) and Pol κ-A (1-616) are equal in activity, but Pol κ-I (36-524) activity is impaired by the deletion of the first helix as shown by the decrease in relative extension activity. **(B)** Deletion of the first two helices in Pol κ results in a significant loss in extension activity. Image reproduced from Uljon *et al*. (2004) (110).

a supportive role in extension efficiency. The recently published structure of the full catalytic core (residues 19-526) by the same group in which all three N-terminal helices retained exhibited full enzymatic activity (70). The roles of the helices were explained by their structure, which will be described below.

B. Structure Determination of mouse Pol κ in complex with DNA

The mouse Pol κ-I – DNA binary complex (monomer A: residues 51-120, 170-223, and 283-517, monomer B: 51-116, 170-224, 282-410; primer strand bases 1-12 and template strand bases 2-9) crystallized in a primitive orthorhombic $P2_12_12_1$ space group with unit cell dimensions a=65.5 Å, b=129.6 Å, c=137.4 Å, $\alpha = \beta = \gamma = 90°$. From the space group and the combined molecular weights of Pol κ-I and the DNA, the solvent content analysis (Matthew's coefficient) estimated 2 monomers in the asymmetric unit with solvent content of 58%. The 3.5 Å resolution structure, shown in Figure 4.5A, was determined by molecular replacement using the palm, thumb, and LF subdomains of the human Pol κ – DNA complex structure (PDB entry 2OH2) as a search model (70), and was refined to a final R_{work} (R_{free}) of 22.8% (29.5%). The final model revealed two monomers in the asymmetric unit, which unexpectedly form a dimeric structure. The A monomer which lacks the finger domain but contains a LF domain that is bound within the major groove of the DNA (Figure 4.5B), and the B monomer, which consists only of the palm and thumb subdomains. It is not immediately clear why the finger domains are not observed, but the thumb domain of the neighboring monomer occupies the exact location where the finger domain should have been. This result was unprecedented

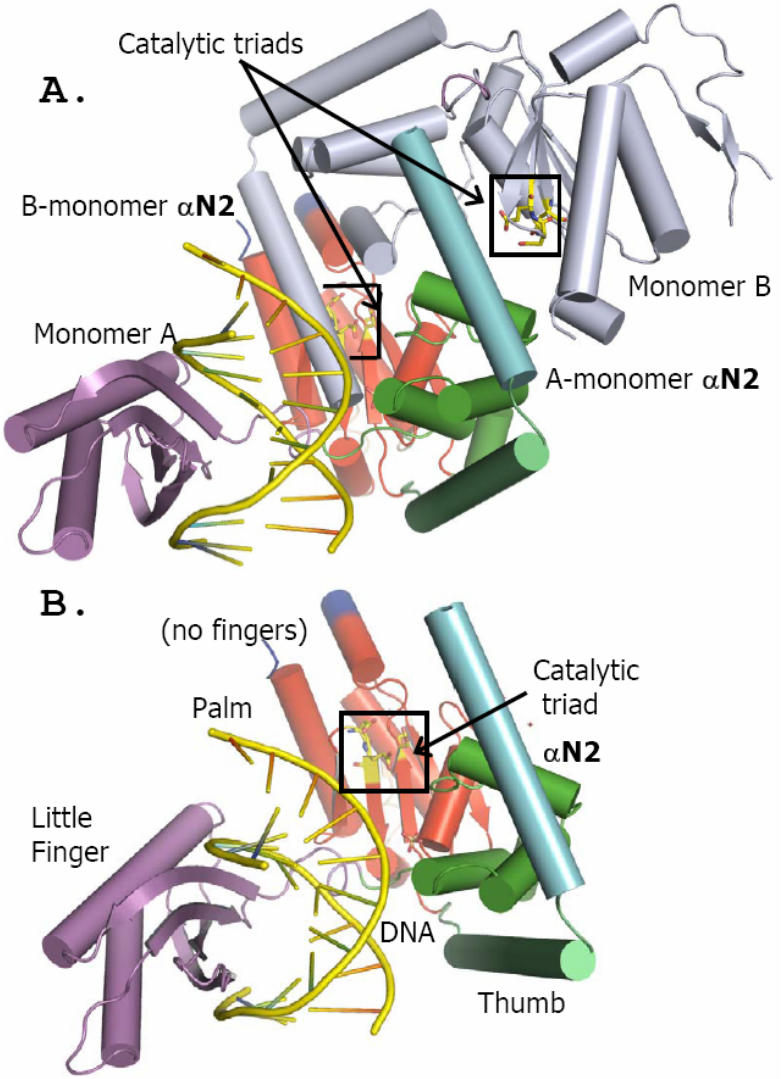


Figure 4.5 Domain arrangement and dimerization in the mouse Pol κ-I structure. **(A)** Model of the asymmetric unit, with the dimerized B-monomer chain colored grey. The DNA duplex cannot bind inside the catalytic pocket due to the presence of the αN2-terminal helix of chain B. **(B)** Cartoon representation of molecule A, showing the novel αN2 helix, palm, thumb, LF subdomain, and DNA substrate. The text (no fingers) is located where the finger domain should have been present. The catalytic triad (in stick representation) has been boxed, and the novel αN2 helix is colored cyan. All structural motifs are colored as described on page (xii), and the red-blue mixed helix indicates the junction between the palm and finger domain.

because DLS and gel filtration experiments using various salt concentrations and pH failed to reproduce the dimer in solution.

Only half of the 16/23-mer DNA substrate is visible in the structure; 7 base pairs are clearly resolved and electron density for DNA backbone phosphates extending along both directions can be observed. Interestingly, the P/T is double stranded along the entire length of the visible segment with no evidence of a template overhang. In both monomers, the novel second N-terminal helix αN2 (residues 50-74, henceforth called αN2) was clearly visible as an extension emerging from the thumb domain towards the active site. As illustrated in Figure 4.5A, it appears that the neighboring monomer dimerized in such a way that this novel helix blocks the catalytic triad of the neighboring palm domain, thus reducing the potential for nucleotide binding and subsequent polymerization. Figure 4.6 also shows that dimerization hinders the LF and blunt-ended DNA from binding to the active site of Pol κ, as shown in the superposition of the palm domain of Pol κ-I construct and the ternary human Pol κ complex. What drives dimerization as seen in the model, and what physiological relevance does this have for Pol κ? Since monomer A is the more complete model of the two, the detailed discussion will be limited to this monomer with the invading B-monomer domains involved in dimerization being referenced as needed.

C. DNA binding mode within Y-family Pols

Before the mouse Pol κ structure can be evaluated, a reminder of canonical DNA binding within Y-family DNA polymerases and the recently solved human Pol κ – DNA ternary complex is provided. DNA ternary complexes were already been solved with Y-family Pols *Sso* Dpo4 (67), yeast Pol η (109), human Rev1 (81) and Pol ι (84), and also

human Pol κ (70) (Figure 4.6B), revealing domains and a conserved DNA binding mode similar to that of replicative DNA Pols. The palm domain is composed of 3 α-helices (αE, αJ and αK) which pack against the non-catalytic surfaces of a β-sheet made of 5 non-sequential antiparallel β strands. This conserved β-sheet contains three invariant acidic residues, Asp 107, Asp 198, and Glu 199 (human Pol κ numbering) located on β-strands β1 and β6 that are required for chelation of a catalytically essential divalent Mg^{2+} cation. This Mg^{2+} ion correctly orients the α and β phosphates of the incoming nucleotide for nucleophilic attack by the terminal 3′-OH group of the primer, while the glycosidic bond is prevented from rotating the base moiety from *anti* to *syn* by van der Waals contacts by the finger domain.

The primer terminus is positioned for nucleophilic attack by a salt bridge between the P_{-1} minor groove phosphate and the amine group of the palm residue Lys 321. Primer strand contact is continued through the thumb domain (green), which "holds" onto the minor groove of the double stranded DNA substrate. It is composed of two (+)–polarized helix-hairpin-helix (HhH) motifs whose hairpins stack one on top of the other, and the distance between the two hairpin motifs equals the average width of the DNA minor groove (~9-12 Å). Therefore, when Pol κ is correctly bound to the primer/template DNA as in the human ternary complex, the primer strand phosphates in the minor groove slide past the first HhH, which contains a water bridged type IV β-turn near the N-terminus of αN in eukaryotic Pol κ (not to be confused with αN1 and αN2 of the N-clasp domain), while the N-terminus of helix αP in the second HhH motif interacts with the template strand phosphates, as illustrated below in Figure 4.8A (70).

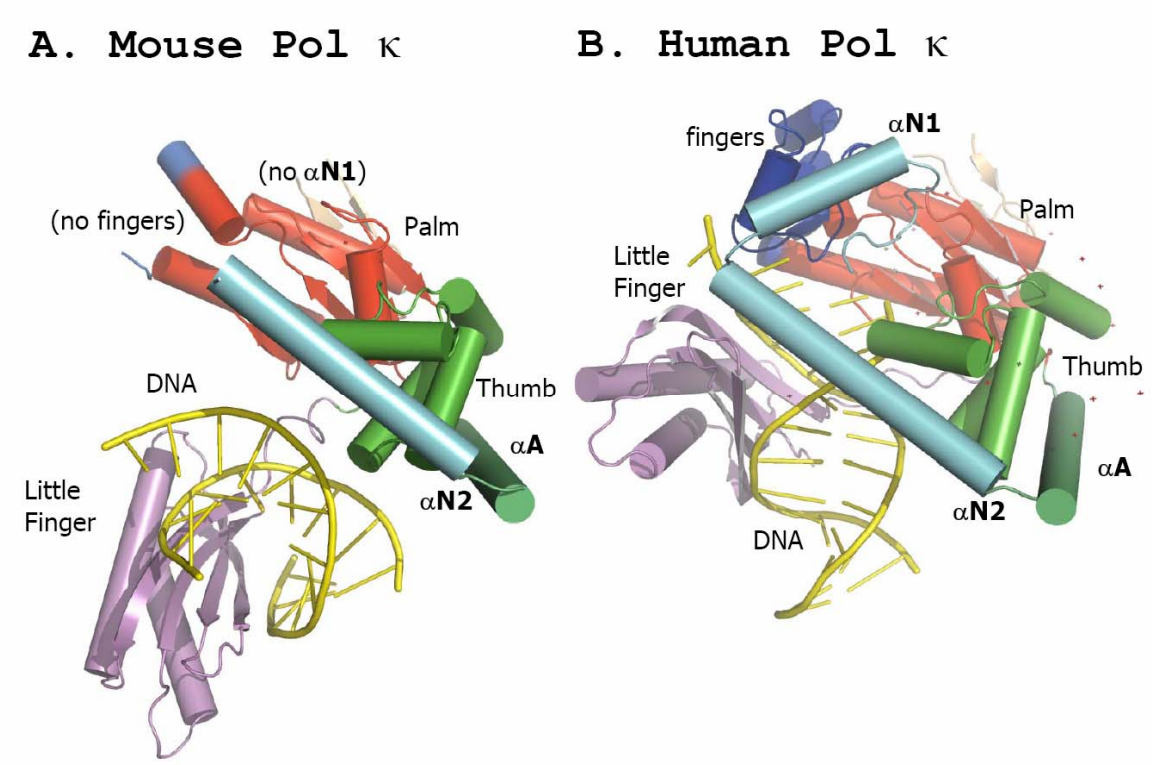


Figure 4.6 Comparison of mouse and human Pol κ complex structures. Superposition of the palm domains from the **(A)** mouse and **(B)** human Pol κ complex structure (70) highlights differences in domain arrangement. The αN1 helix is not present in the mouse structure because it was truncated and too flexible to observe. Note that the LF domain and DNA substrate are excluded from the mouse Pol κ active site due to the presence of the B-monomer αN2 helix (not shown for clarity). Domains are colored as described on page (xii).

Stabilizing contacts to the template strand, especially the single-stranded substrate and nascent base pair, are provided by the LF domain, the novel and distinguishing feature among Y-family Pols. It is composed of a 4-stranded β-sheet folded into a "shortened jelly roll" motif, with two α-helices spanning the non-catalytic face of the domain (Figure 4.6B). The major groove oxygen atoms of the template and primer strand phosphate backbones are coordinated by the amine backbone of the outermost β-strands β11 and β13, whereas residues in the concave portion of the domain provide a surface for the bases to slide upon as the template bases enter the active site. The non-specific interaction of the major groove phosphates upon the LF domain can be compared to the real world example of train cars (the LF domain) sliding along both rails of a train track (the DNA major groove). These bases are also supported from the nascent minor groove face by the impinging finger domain, and from the major groove side by unconserved residues in the LF domain that are Y-subfamily-specific. The network of these protein-DNA interactions illustrates the minimum structural requirement to promote nascent base pair formation in the Pol κ and the shared Y-family Pol active site (70, 110).

Pol κ contains one additional domain called the N-Clasp. Composed of the first 100 N-terminal residues, it precedes β1 of the palm domain, which is the consensus start of Y-family Pols (with the exception of Rev1 which is preceded by a BRCT domain). It folds into three amphipathic α-helices, αN1, αN2, and αA, which "wrap" the 100 residues around the core of the protein (Figure 4.6B). These helices are stabilized by hydrophobic interactions with nonpolar residues: helix αA interacts with helix αP of the HhH2 motif, the C-terminus of αN1 with helix αN2 of HhH-I, and then the N-terminus of αN2 and αN1 with Phe 464 of the LF domain and αD of the finger domain and even the

hydrophobic T_{+1} base. This latter interaction forms half of the single-base steric gate that was analyzed in Chapter 3. The structural arrangement of the three N-terminal helices explains why the catalytic activity is compromised upon loss of the N-terminal domain, since the predominantly nonpolar faces of the catalytic core would then suddenly be exposed to the solvent. Coincidently, the helices that form the hydrophobic interactions with the N-terminal helices are present in the other Y-family-Pols, but their surface residues are polar and therefore conducive to interaction with the solvent.

D. The Mouse Pol κ-I dimer

Measurements by dynamic light scattering confirmed that mouse Pol κ-I is a monomer in solution, but our crystal structure revealed two monomers in the asymmetric unit that have unexpectedly dimerized into a "yin-yang" shape (Figure 4.7). One monomer has rotated a full 180° in relation to the other, allowing the αN2 helix that typically extends across the active site from the thumb to the finger domain, to pack against the catalytic triad, thereby blocking potential activity (Figure 4.7B). It forms stable contacts to the palm-A domain from both ends: (i) by hydrogen bonds that are formed between ε-carboxylate oxygen (OE1) of the N-terminal Glu 56 residue of αN2-B with its donor, the hydroxyl group of Tyr 200 of the palm-A domain 3.0 Å away, and (ii) through the formation of a salt bridge between the terminal amine of Lys 70 from αN2-B and the δ-carboxylate of Glu 115 of αB-A 2.5 Å away on the palm/finger-A junction. However, these associations are only possible if the finger domain is displaced and thereby allows the N-clasp and thumb domains to occupy this position.

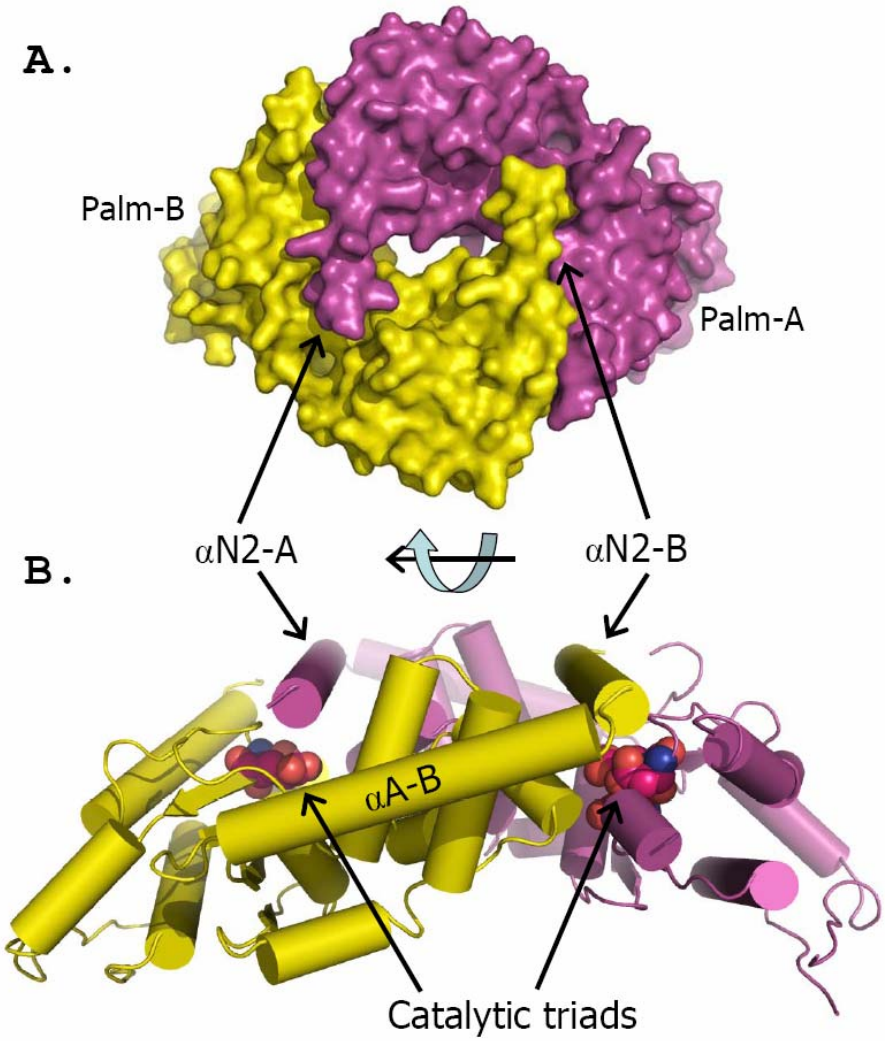


Figure 4.7 Association of the two Pol κ-I monomers. The two monomers are intertwined through their αN2 helices, which interact with the neighboring protein's palm domain. **(A)** Solvent accessible surface view of the dimer, with the protruding αN2 helices positioned on top of the palm where the incoming nucleotide would bind. **(B)** Cartoon representation of N-clasp binding interaction. Helices αN2 and the thumb domain displace the finger domain, as it is not present in the structure. The active site residues are displayed with their van der Waals radii to emphasize their proximity to the αN2 helices. Panel A has been rotated into the plane of the paper to generate the side view in Panel B. Molecule A, magenta; molecule B, yellow.

Additional difference electron density was observed at a map contour of 5σ. These were located in a positively charged pocket within 3.1 Å of the polarized N-terminal ends of both αN2-B and palm domain helix αK-A (the HhH-1 interaction), thereby providing additional stability for the dimer. This density was assigned as a $H_2PO_4^-$ anion due to its electronic environment, the pH of the crystallization reagent, and a plausible source of its introduction into the mother liquor. The exact source of this monophosphate anion is uncertain, since polymerization of bases onto the primer strand releases inorganic pyrophosphate rather than monophosphate ions.

During the first purification steps, however, Na Phosphate buffer was utilized and the anion may have remained bound to the protein throughout the entire purification protocol. Superposition of the mouse thumb domain with the human Pol κ ternary structure showed that this anion is too far away from the position where the primer strand would normally associate with the HhH-1 motif, and no solitary anions or waters occupy that location in the human Pol κ complex structure. This phosphate is therefore only present because of the position of the αN2 helix.

While the αN2 helix is mostly responsible for dimer stabilization, none of the observed interactions support the displacement of the finger domain. With the exception of the disordered finger domain, the other domains maintain their position and superimpose well, which is emphasized by the low rms deviation of 1.06 Å for all main-chain atoms. As shown in Figure 4.5A, the αN2 C-terminal end remains tightly bound to the thumb domain via van der Waals contacts to non-polar residues Met 363, Ala 366, and the Leu-rich patch from residues 381-385. Its position has not changed in comparison to the the catalytic core of the human protein. When the palm domains of both proteins

were superimposed, the αN2 helix of the mouse Pol κ-I construct was shifted only 4° away from the palm domain. This observation implies that the thumb-based stabilization of the N-clasp is very strong, perhaps even solid enough to "push" the fingers out of the way. Therefore, dimerization has occurred between two rigid bodies composed of the palm, thumb, and N-clasp domains, while the LF domain and finger domains have moved significantly due to the dimerization event.

E. LF domain coordinates the DNA substrate

Despite the rigid dimerization of the catalytic core, the mouse LF domain still possesses the freedom to bind to the DNA substrate. In the mouse model, it was observed to bind to the major groove phosphates with the peptide of the outermost LF β-strands β11 and β13. These are the same non-specific interactions as observed in the human ternary complex, except that the mouse LF domain was packed against the catalytic core in a different orientation as compared with the human ternary complex. In Figure 4.8, the DNA appears that while the mouse LF domain (magenta) has bound to the DNA substrate, it was excluded from the active site, and its position midpoint to both apo- and bound orientations suggests that this model is likely an intermediate form.

Only 12 primer strand bases and 8 template strand bases of the 16/23 primer-template duplex are visible in our model. Its interaction with the LF domain stabilized 8 base pairs, but the primer strand is considerably longer and nearly completes one full turn of a double helix. The dimensions of the major groove (measured between the $P_{T(n)}$:$P_{P(n+8)}$ atoms) ranges from 19-24 Å, while the minor groove (measured between the

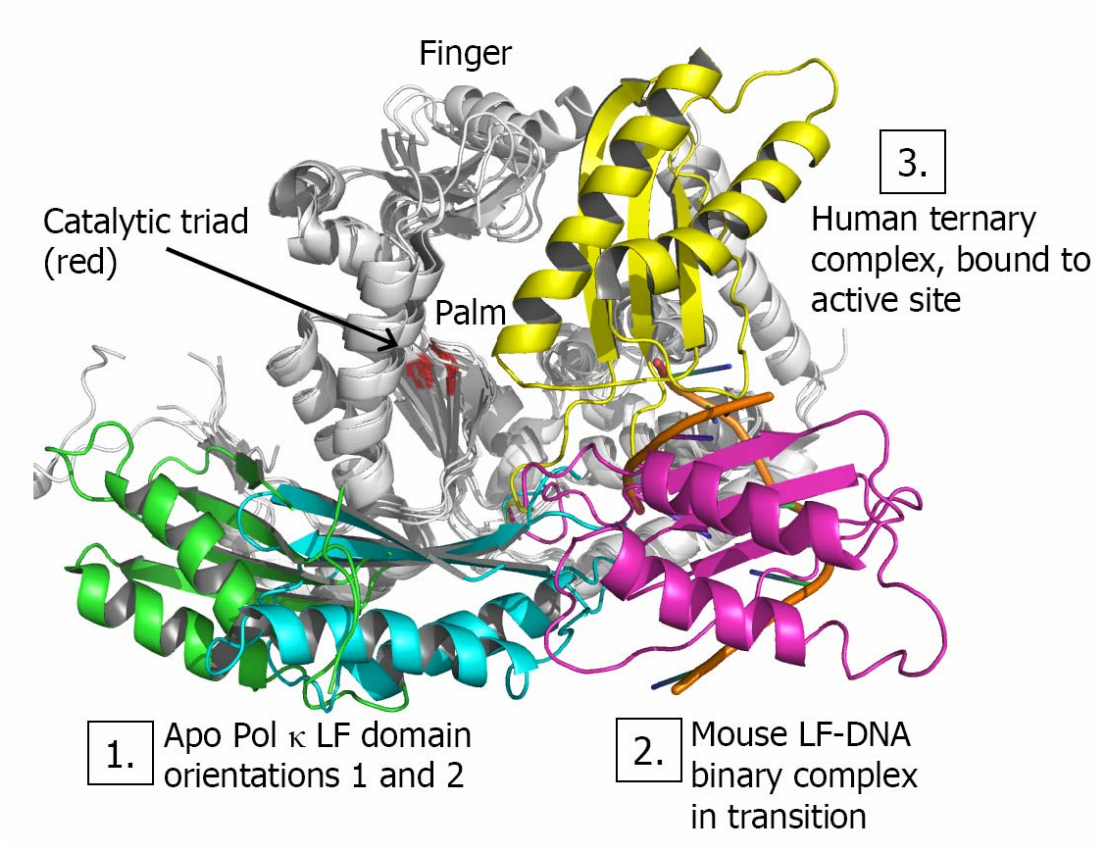


Figure 4.8 Superposition of Pol κ palm domain in various stages of DNA binding shows flexibility of the LF domain. The flexibility of the 15-residue linker provides a lot of mobility thus permitting packing and binding of the LF domain in different orientations in proximity to the catalytic core of the protein. In the apoenzyme models, **(1)** the LF domains (green and cyan) have associated with the non-catalytic face of the palm domain in two positions suggesting a "waiting" position. **(2)** While the mouse LF domain (magenta) has bound to the DNA substrate away from the primer terminus, excluding it from the active site but suggesting that it may be in transition. **(3)** The little finger, (yellow) which correctly holds the T_0 along with the finger domain, presents the primer terminus to the active site for polymerization. This proposed scheme may explain how the primer terminus is delivered into the Pol κ active site. Images recolored from Uljon *et al.*(110) and Lone *et al.*(70) using PyMol.

$C1^*_{P(n)}$:$C4^*_{T(n+3)}$ atoms) is 9-11 Å wide, roughly half the width of the major groove and indicating that it conforms to B-form DNA.

From the results obtained from parallel primer extension assays (Figure 4.9B), Pol κ-I only extends the primer by one base after a ten-minute reaction at 30°C. However, the primer strand in our model did not appear to be the primer terminus. As discussed in Chapter 3, Arg 506 forms essential electrostatic contacts with the T_0 and T_{+1} phosphates in order to stabilize the nascent base pair. As shown in Figure 4.9A, attempts to assign the sequence of the 17/23 product duplex were made by assuming the template phosphate in proximity to Arg 506 (template Cytosine-4 in our structure) was the T_0 nucleotide of the nascent base pair.

However, the proposed primer sequence did not fit the observed density. The "T_0" base did not have a Watson-Crick binding partner, as indicated by the absence of electron density, and the presence of unallowed "negative" difference density in Figure 4.9A. And yet the sugar-phosphate backbone is clearly present. The primer strand continues upstream for 2-3 nucleotides longer than expected, implying that the LF domain bound to double-stranded DNA rather than to the template:primer-terminus junction.

There are two sources from which this terminus-less substrate could result: First, either the addition of high concentrations of PEG did not sufficiently inhibit polymerase activity, permitting mouse Pol κ-I to replicate erroneously with dGTP to the end of the template strand, or, second, this is the double-stranded portion of the extra long P/T provided for crystallization, with the terminus junction being upstream and closer to the other Pol κ-I molecule. To test this latter idea, the coordinates for the DNA-bound LF domain were superimposed into the analogous position at the end of the other Pol κ-I

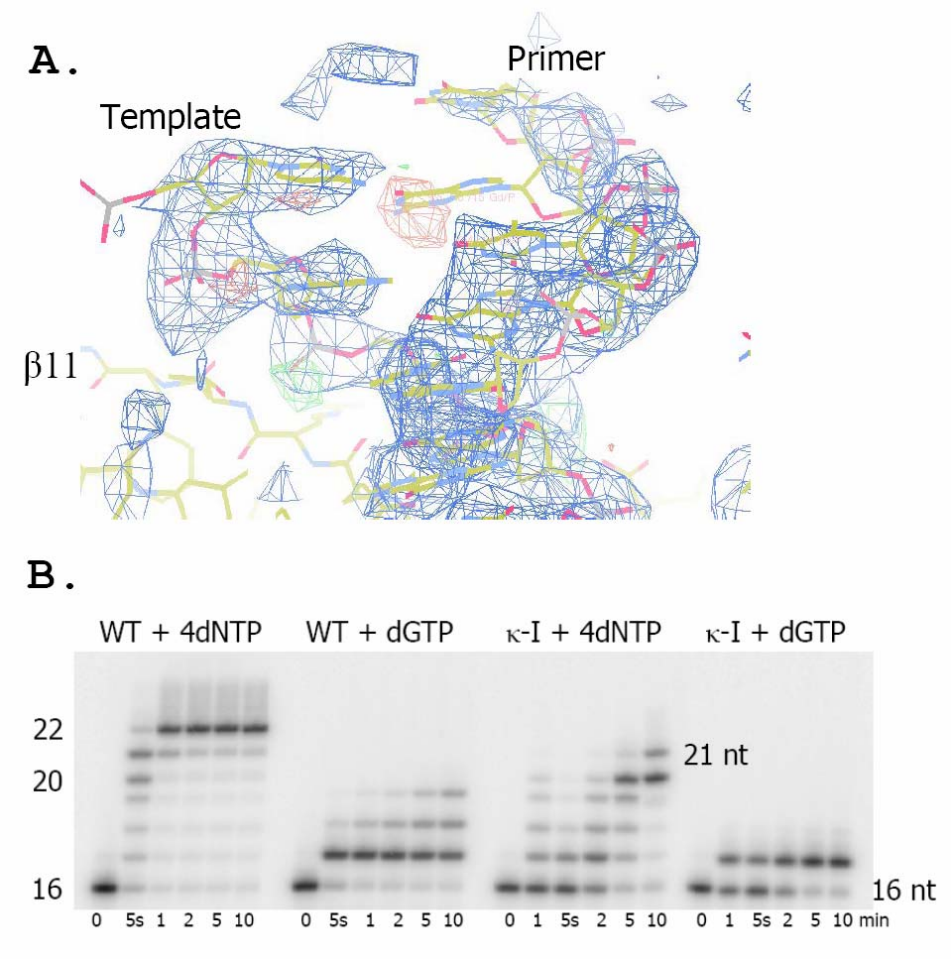


Figure 4.9 Structural and enzymological analysis of the DNA substrate in the model. **(A)** $2F_o$-F_c electron density (blue mesh) and F_o-F_c difference density maps (red mesh) of the DNA substrate (drawn in line representation and colored by element) contoured at 0.9σ. The 5′ phosphate of the template terminus and the nucleotide opposite this position are missing, as indicated by the presence of negative difference electron density at this map contour level. **(B)** Primer extension time course experiment with WT and Pol κ-I when provided either four dNTPs or only the next base dGTP using the pre-crystallization reaction conditions. After 10 min at 30°C, Pol κ-I was only able to extend the primer by one base when given the next incoming nucleotide.

molecule. As shown in Figure 4.10, the proposed location of the second LF domain (called LF′) is rotated 180° relative to the LF-A domain, giving the same relative orientation as the dimerized catalytic cores. The highly symmetrical DNA almost appears as one molecule, except that the wrong termini (5′-template to 5′-primer) are oriented towards one another due to the symmetrical operation. Nonetheless, it is easy to visualize this DNA model as showing one continuous DNA duplex. Since our template is 23 bases long, there is enough clearance in the crystal lattice for the entire structural form of the substrate, whether fully extended or a substrate with a template overhang. A closer look at the crystal contacts in proximity to the theoretical LF′ domain also suggests that no major clashes with nearby symmetry-related atoms are present.

Based on this model, we can hypothesize why the Pol κ-I enzymes dimerized. It is because while one molecule of Pol κ-I is able to bind to a DNA molecule, the LF domains of the second monomer attempts to interact with the DNA major groove at opposite ends of the same molecule during crystallization. Since the LF domain is connected to a rigidly packed thumb domain, the two enzymes are straining to bring their thumb domains as close as possible to the primer strand, and in doing so, push the other monomer's finger domain out of the way with the αN2 helix. The αA-B chain thumb helix then squeezes in between the finger-A domain and αN1-A helix, which explains why both structural elements are missing. Why then is the LF′ domain not resolved then? This may verify that the DNA substrate is, in fact, the 17/23 product duplex, making 1/3rd of the available DNA substrate single-stranded, as well as being oriented in the inverted 3′→5′ orientation for the LF′ domain to bind. A major groove turn must be present for both monomers to compete for the substrate, thereby forming the dimer.

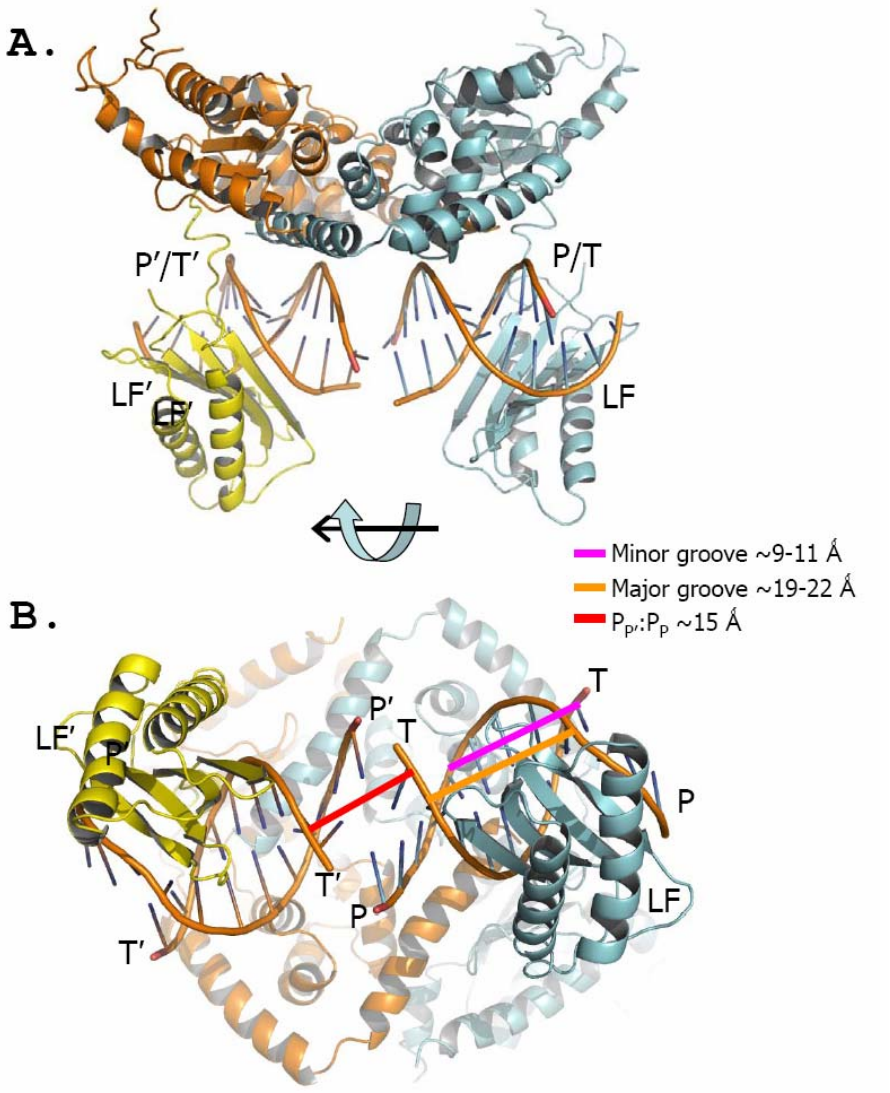


Figure 4.10 Putative location of the Pol κ-I, molecule B LF′ domain and DNA. Cartoon representation of the mouse Pol κ-I dimer where the LF domain and DNA substrate belonging to the A molecule were translated and affixed to the B-molecule. Distance between the two DNA substrates and the widths of the grooves are reported. **(A)** perpendicular view to the DNA axis. Pol κ-I(A) molecule and its resolved LF domain, cyan.; Pol κ-I (B) molecule, orange; superimposed LF domain (LF′), yellow. **(B)** View along the axis of symmetry. Note that the nucleic acid termini do not match, but the double helix pattern seems to be conserved, suggesting the orientation of the true duplex.

Based on the observations, the dimer appears to be a crystallization artifact – however, one that may imply physiological relevance.

F. The DNA packs against a selective thumb motif

It is interesting that the LF domain is a promiscuous binder of DNA major grooves, since the binary complex appears to be bound to double-stranded DNA. While the major groove is the required motif for LF domain binding, the duplex DNA backbones pose a steric problem since the thumb domain and encircled active site have evolved to accommodate only the primer terminus. Therefore, primer terminus accommodation by the thumb domain can be viewed as "step two" of the DNA binding model.

In the mouse binary complex, however, this event is prevented in two ways: (i) the αN2 helix in bound in the active site, preventing the minor groove from entering the active site, and (ii) our double stranded product lacks a "local" primer terminus, and the double helix (specifically the presence of the primer strand's sugar-phosphate backbone) appears to be excluded from the active site by the N-clasp. While the DNA substrate has been significantly displaced by the αN2-B helix (87.6° rotation, 18.55 Å translation away from the catalytic core), the primer strand managed to make one weak polar contact with the apical end of the HhH-2 motif in the thumb domain.

As shown in Figure 4.11A,B, the minor groove of the human and other Y-family Pol ternary complexes (66, 67, 70, 81, 83) were shown to associate with these HhH motifs. The minor groove phosphate oxygens of the primer strand are within 4 Å of the outer second HhH motif (Figure 4.11A). However, in the mouse structure, the αN2-B

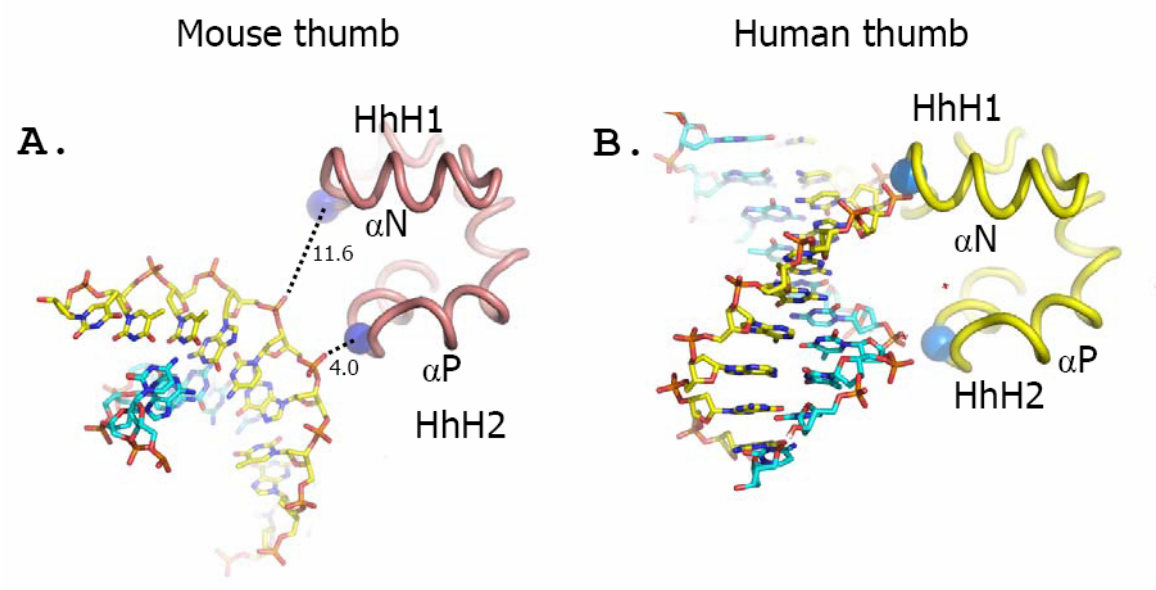


Figure 4.11 Cartoon and stick representation of the thumb:DNA interaction. (A) The mouse DNA primer phosphate backbone is 11.6 Å away from the its canonical binding partner, the HhH-1 motif. It is instead interacting through polar contacts with the HhH-2 motif, leaving the template strand uncoordinated by the protein. Blue spheres indicate the (+)-polarized end (apical backbone N atom) of the HhH motifs. Therefore, the minor groove is not engaged by the catalytic core in the mouse binary structure. **(B)** The human ternary complex demonstrates the correct thumb:DNA minor groove interaction required for catalysis. DNA is colored as described on page (xii).

helix “pushes” the DNA helix further from the palm domain, causing it to interact with the second HhH turn instead of the intended VSGI β-turn of the first HhH motif. If the DNA substrate used in our co-crystallization experiments was bound in the LF domain such that it presented the primer terminus to the catalytic core, it would easily enter past the HhH motifs and bind within the active site.

Though our dimerization artifact renders these Pol κ-I molecules inactive, one small detail about the binding event is present – the availability of the HhH2 for DNA coordination. As shown in the putative binding model in Figure 4.12, since coordination to this relatively exposed motif was conserved (even with the wrong DNA strand), it suggests that the apical turn of the HhH2 would be the first potential contact for template:primer-terminus forks. Our double-stranded substrate cannot enter since the elongated primer backbone is too large, and so the duplex DNA remains stacked against the N-clasp domain. In other words, the HhH motifs also act as a two-part substrate recognition pocket where only a successfully oriented primer-terminus would rotate/tilt into the active site, like a cantilever, allowing the minor groove to bind to both HhH motifs. If this thumb pocket could not accommodate both minor groove backbones, then the substrate would not be in the correct position.

The third and final step of the DNA binding event would be the introduction of the template strand between the LF domain and the finger domain. In the binary Pol κ structure, it appears that the LF domain can bind into the major groove promiscuously. Proper positioning of the T_0 and T_{-1} phosphates along the β11 strand are crucial for primer strand elongation. How is proper alignment achieved? None of the structures in our model yield clear conclusions. However, one can speculate that it could occur in one

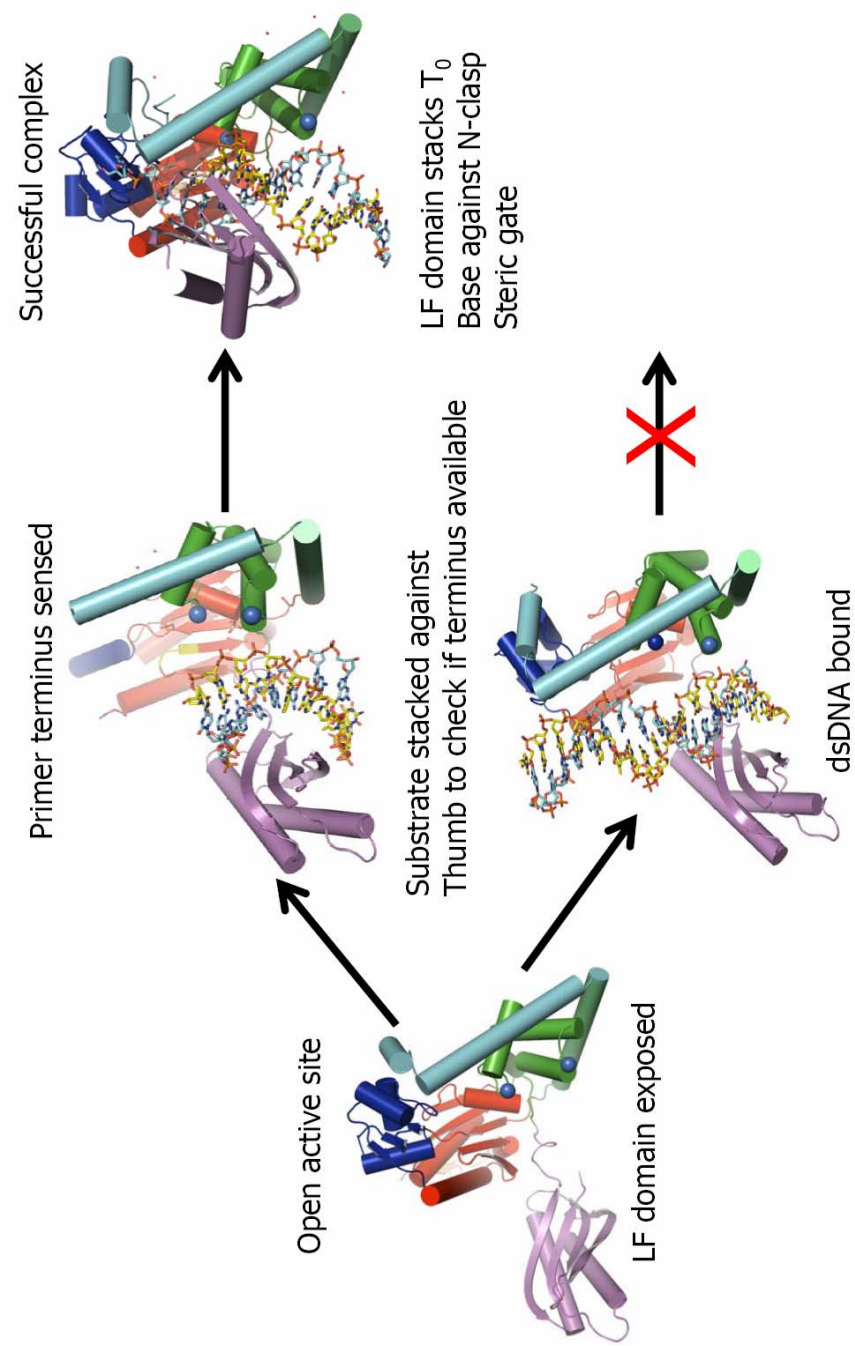


Figure 4.12 Putative model for a Pol κ pre-loading complex. The dimerization event may have isolated some of the earliest substrate recognition steps, alluding to an order of events which have never before been observed. Step 1 (**left**): a flexible LF domain is free to bind to the major groove of DNA. Step 2a (**top center**) if in proximity to the primer terminus, the last of a continued primer strand backbone will permit it to interact with the HhHs motifs and (Step 3, **top right**) tilt into the encircled active site. Otherwise (Step 2b, **bottom**) the duplex will be excluded from the catalytic core. The domains are colored as described on page (xii), and blue spheres indicate the apical ends of the HhH motifs.

of two ways. First, the LF domain "follows the train tracks to their end." In other words, since the T_{+1} phosphate is unpaired and precedes the major groove by one position, the nascent base pair (T_0) + 1 base is simply the farthest the LF domain can process along the template strand. However, due to the non-specific interactions between DNA and LF domain, this event cannot be catalyzed by the LF domain alone and must be driven by another source. Or second, this "sliding" event occurs upon primer terminus binding to the thumb domain. As mentioned previously, there is an 87° tilt and 18 Å translocation the primer terminus must make to enter the active site based on our model, which suggests that the LF domain could tilt and translocate with it, packing the exposed T_{+1} and T_0 bases on β11 against the hydrophobic finger domain and N-terminal clasp steric gate. This interaction is stabilized through various electrostatic (Arg 506) and hydrophobic (Phe 464, Phe 48, Phe 154) contacts.

G. Broken "knuckles"

The finger domain, unresolved in the mouse Pol κ structure, is a ~55 residue globular domain composed of two short α-helices and a small 3-stranded anti-parallel β-sheet. The core of this domain is made up of one continuous sequence, which is identical between human and mouse proteins and emerges from the palm domain and returns to it through a hinge-like connector. Superposition of the finger domains from the human apoenzyme and complex structures revealed an rms deviation of 0.83 Å, indicating very little change in position upon binding of the incoming nucleotide. However, its interactions with the rest of the protein are surprisingly few, considering its rigidity. It forms no direct side chain interactions with the palm domain. In the apoenzyme structure,

the finger domain remained in position as the "roof" of the active site due to an extensive water-bridged hydrogen bonding network, which was substituted by the incoming nucleotide. This indicates that the restraints required to keep the domain in position are already contained within the primary sequence. A closer look at the hinge residues (amino acids 115-120 between palm/finger helix αB and the finger domain and 167-175 between the finger domain and palm helix αE) reveals several proline residues (Pro 120 and 167-168), which could restrain the loops for proper positioning of the domain.

To allow dimerization as seen in my mouse Pol κ model, both finger domains had to be displaced so that the neighboring thumb could occupy the same space. This can be accomplished in one of two ways: proteolytic cleavage of one or both solvent-exposed connecting loops, or rigid body relocation of the domain, which depends on some level of flexibility of the hinge loops despite the presumed rigidity. Currently, there seems to be enough evidence to suggest that the latter is true.

First, the density map clearly shows the presence of continuous protein backbone until the prolines, despite several potential endoprotease cleavage sites. The presence of prolines in the primary sequence. The first hinge contains the sequence N-116-MRDN||**P**|EL, and the second hinge is composed of the sequence N-165-IIV||**P**|**P**N, where || indicates the end of the modeled density, and | indicates the starting position in which the electron density was present but refined with low occupancy. Notably, these sequences are not conserved between the Y-family subfamilies, with Pol κ being the most proline-rich. The modeled sequences terminate directly preceding a proline, whose peptide bond is known to isomerize between *trans* and *cis* configurations at room temperature(14). If the backbones of the two hinge sequences adopted a *cis*-peptide bond

in these areas, then it could be imagined that the finger domains would "open up" and point away from the intruding αA and αN2 helices – which is exactly what is required to accommodate these intervening helices. When the electron density map of the binary complex structure was inspected, there is room in the lattice for this rearrangement to occur, and sparse disconnected regions of electron density are scattered in this region at a map contour level of 0.7σ. Due to the low resolution limit, however, I was unable to build the finger domain.

There are other clues to the presence of the finger domain in the lattice. As shown in Figure 4.13A, Helix αB and αE, which border the connecting loops are each tilted ~7° away from the palm domain in the mouse model, indicating that some movement of the finger domain can already be observed by these slight shifts at the palm-finger domain junction. Furthermore, the last residue that could be modeled, the residues preceding the proline, appear to adopt the cis-conformation, supporting this hypothesis.

Another interesting observation from the dimerization artifact is that the interaction between the finger domain and N-clasp αN1 helix had to be disrupted. This interaction is normally stabilized by numerous hydrophobic contacts, and provides the dominant interface that holds the N-clasp across the active site (70). It could also be speculated that such a hydrophobic cleft between the N-clasp and finger domains could provide a binding pocket for bulky hydrophobic adducts such as BPDE or other exocyclic lesions passing through the steric gate for entry to the Pol κ active site. However, this hydrophobic interaction was disrupted by the dimerization of the two monomers, the neighboring charged αA helix wedged itself in between the finger and αN1 helices (not observed in the Pol κ-I:DNA complex structure). Since both the finger and αN-1

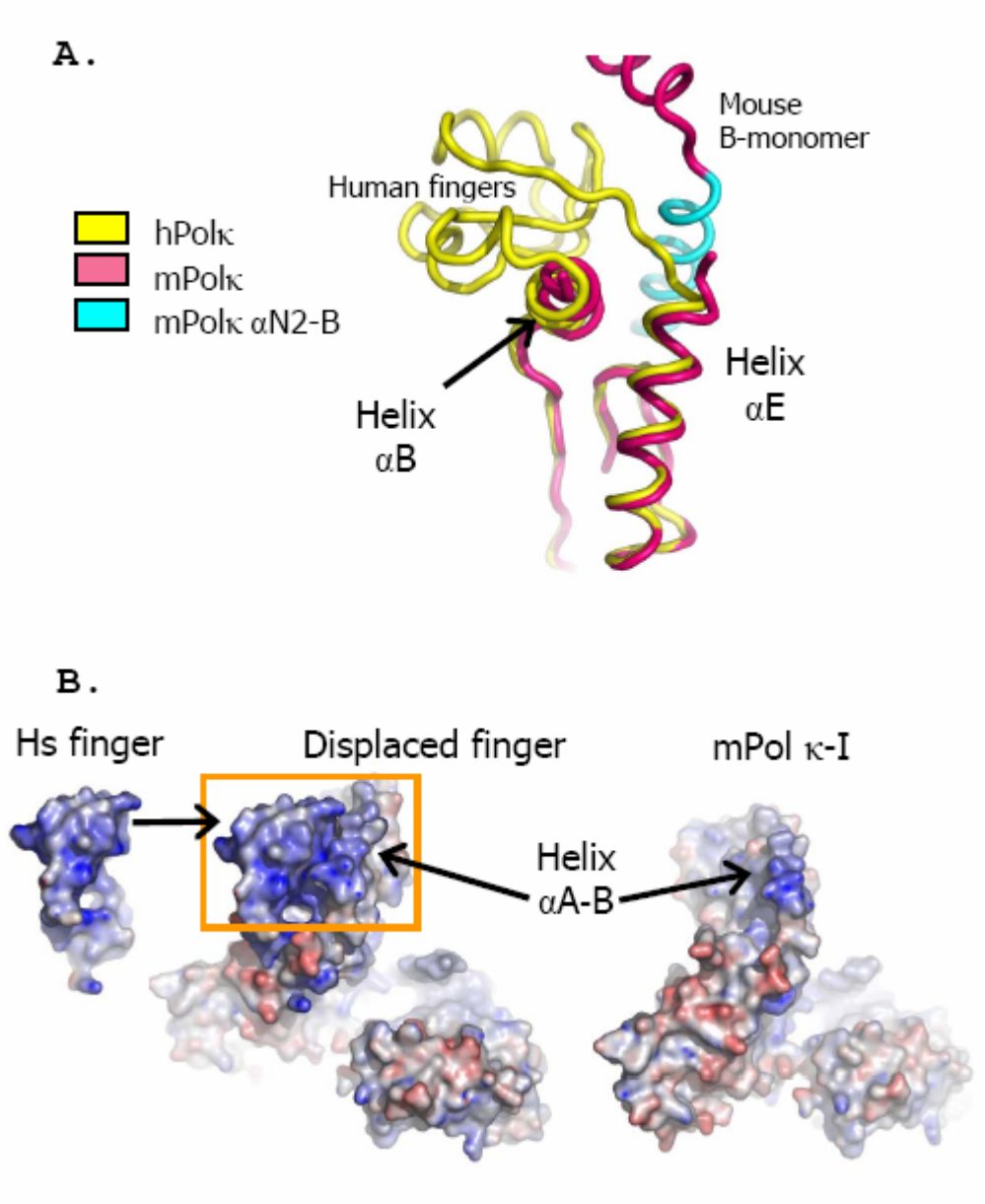


Figure 4.13 The "Broken" Finger domain. **(A)** Slight 7° rotations of the α-helices leading to and from the finger domain indicate some response to a global movement of the finger domain. Parts of the superimposed palm β-sheet are visible in the rear of the image, along with the αA and αN2 (cyan) helices of the dimerized monomer; the mouse thumb-B domain has been omitted for clarity. **(B)** Electrostatic surface potential maps where the spectrum of color relative to charge is [-10e,0,10e] = [red,white,blue]. These images show that the human finger is very positively charged, with extended patches of hydrophobic regions (shown as paler blue/white). The mPol κ-I αA helix, which displaces the finger domain, also is rich in basic residues. These two surfaces are hypothesized to repel each other strongly (finger 10e, αA 6e), and may explain the reason the finger domain is disordered in the final structure.

hydrophobic faces would now be pointing towards the polar side chains of the αA helix, it is not surprising to find that these secondary structures, though present (in part) in our constructs, are not visible in the final structure. A theoretical model was created to estimate the location of the finger domain if proper crystal contacts could be made to the αA-B helix, and notably, it is not only the hydrophobic face, but a region of basic finger residues that impinge on the basic residues of the αA-B helix. This charge repulsion, and therefore absence of regular residue contacts, could also explain why the finger domain is disordered in dimerized structure.

In the discussion of the human Pol κ ternary complex (70), Lone *et al.* concluded that since the fingers did not move in comparison with the apoenzyme structure, the N-clasp would have to move in order to accommodate lesions. However, our structure does not support this hypothesis. The rigid coordination of the αN2-helix, combined with the malleable finger hinges suggest that it is the finger domain which moves in response to steric stressors such as larger lesions, or in this case, the dimerization artifact. While dimerization was facilitated by the truncated αN1 helix in our construct, comparison to the human Pol κ 19-526 construct suggests that only one critical hydrophobic interaction, Ile 37, is missing to the hydrophobic interface of the finger domain. The rest of the leading N-terminal residues (19-27) are modeled as a poly-Ala chain and believed to be folded non-natively, since amino acids Asn 24 through Leu 31 would block access of incoming nucleotides to the active site. Notably, the electron density map of the second monomer did not resolve residues 19-32, betraying their true flexibility at 3.0 Å resolution. This suggests that the interaction between full length Pol κ (either mouse or human) and the mouse Pol κ-I construct may be dependent on these few αN1 helix

interactions, meaning that separation of the N-terminal clasp and fingers domain prevented by only one helical turn. It is still possible that residues 1-18 finally position the true N-terminus of the protein, it is impossible to interpret where that may be.

Therefore, to continue using the structural analogy that DNA Pols are similar to the human right hand, it appears that in the binary complex structure of mouse Pol κ-I, the "knuckles" of the finger domains have been broken to produce the dimer.

H. Final comments

While nearly all aspects of polymerase activity and processivity have been studied for these enzymes, there is still much speculation about how the Y-family Pols initially bind to the primer terminus. It is clear that the active sites of template-directed DNA Pols are designed for a DNA substrate with a template strand overhang, but access to Pol κ's active site is partially obstructed by the N-clasp domain. While it has been shown that PCNA plays a role in Y-family Pol recruitment to stalled replication forks (13, 27, 37-39, 56, 57, 71), the same has not been conclusively proven for Pol κ (6, 89). Also, Pol κ, among the other Y-family Pols, is completely capable of loading itself onto the template:primer junction, since many enzymological assays (including those discussed in Chapter 3 of this dissertation), were conducted in the absence of PCNA.

While this model is speculative, it combines experimental observations of domain movements as seen in the various static structures. Pol κ has only one domain with a wide range of mobility (the LF domain on its 15-residue flexible linker), and the start and end points of the domain movements are defined. The binary structure presented here attempts to bridge the apoenzyme and ternary complex structure together to provide

possible scenarios for initial substrate recognition and accommodation. The mechanism discussed here may explain how Pol κ is able to bind to the template:primer junction without the need for PCNA to prepare the terminus.

CONCLUDING SUMMARY

Chapter 5: Concluding summary

A. Kinetics of lesion bypass

Translesion Synthesis (TLS) is essential to the survival of the cell, since hindrance of the replication machinery stalls cell cycle progression and can lead to chromosomal aberrations or cell death. The success of TLS depends on how the damaged base is accommodated within the bypass polymerase active site and to what degree the adduct interferes with normal Watson-Crick base pairing. The higher eukaryotic Y-family Pols have evolved active sites that exclude lesions based on their stereochemistry thereby accommodating only a subset of lesions, which can be efficiently bypassed. This dissertation explored the structure/function relationships in several protein-DNA interaction sites to explain the discriminatory and catalytic mechanisms unique to DNA polymerase κ.

The mutational analysis of the fingers and LF domain revealed how Pol κ coordinates (or rather, does not coordinate) not only the nascent base pair but also the bases upstream and downstream of the active site compared to the archaeal homologue Dpo4 (67). Only a few conserved interactions can be observed in Pol κ, such as the LF residue Arg 506 (Dpo4 Arg 331) which binds to the single-stranded template, and the palm residues Asp 106, Asp 197, and Glu 198 required for nucleotide triphosphate coordination, and Tyr 111, the ribonucleotide exclusion sensor (Figure 5.1), but compared with the other residues in the active site, all substitutions in the LF domain in comparison to Dpo4 appear to have "relaxed" the constraints imposed on the DNA substrate. This loose coordination provides enough space for the incorporation of

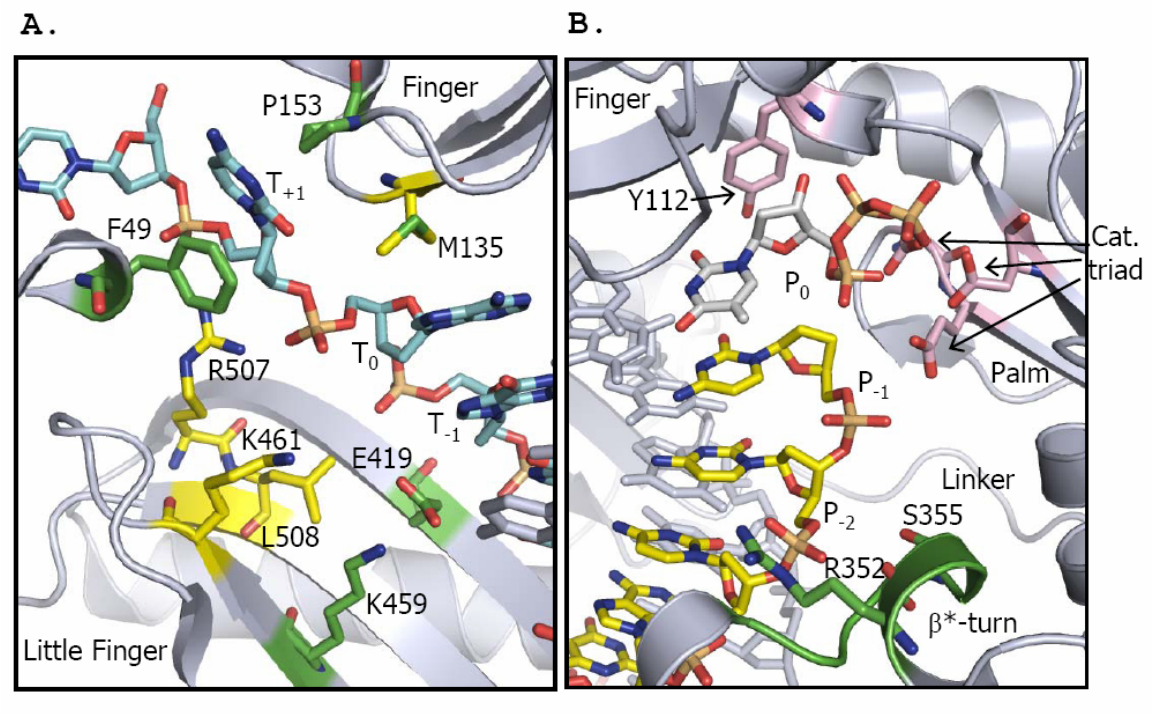


Figure 5.1 Localization of key protein-DNA interaction sites in Pol κ. Cartoon representation of the human Pol κ-DNA ternary complex(70) which highlights key amino acids (in stick representation) discussed in this dissertation Yellow residues indicate amino acids mutated in this study, while green residues are promising targets for further study. **(A)** Close up view of the finger and LF domains in the vicinity of the T_0 template base. The N-clasp has mostly been removed for clarity; a small portion of the N-clasp helix was retained to highlight the location of Phe 49 of the steric gate. The potential "speed bump" which may extrude template bases, Lys 459 and Glu 419, is also shown in green. The mouse Pol κ equivalent residue positions are Phe 48, Met 134, Pro 152, Glu 418, Lys 458, Lys 460, Arg 506, Leu 507, and Arg 506. **(B)** Close up view of the palm and thumb domains in the vicinity of the nascent base pair and primer terminus. The invariable catalytic triad and Tyr 112 ribonucleotide sensor are colored in pink. The disorganized β-turn (β*-turn) in the thumb domain is colored green and is also a potential target for understanding primer terminus stabilization. The mouse Pol κ equivalent residue positions are Tyr 111, Arg 351, and Ser 354.

damaged bases into the double stranded product, assuming they can translocate through the "gate".

This study was based on the results of a homology model of mouse Pol κ before the DNA-bound structure of the human protein was available(110). As previously discussed, the role of the first 100 N-terminal residues was not well understood, beyond the fact that they were essential for full activity. Once this structure was solved, the steric gate against linked-base lesions became clear: the steric gate was not only a "groove" created by the terminal side chain atoms of Met 134, but the novel N-clasp domain (composed of the 100 N-terminal residues) contained an aromatic residue, Phe 48 (Phe 49 in human Pol κ), which stacked against the T_{+1} base from the other side (70). It is these residues, in combination with Pro 152, that most likely create the linked-base lesion steric gate. Since the N-clasp was unchanged in this study, the presence of cisplatin still significantly decreased the k_{cat} of insertion and showed only very mild improvement with respect to the Pol κ variants. Pol η and Dpo4 can easily bypass these lesions because they do not have secondary structures analogous to the N-clasp in Pol κ and therefore their active site is far more solvent exposed and less restricted.

In light of these new structural and kinetic data, mutational analysis of other putative discriminating interactions may provide further insight into substrate accommodation and lesion bypass catalyzed by mouse Pol κ. For example, kinetic studies with the LF mutants were unable to disrupt a template misalignment mechanism significantly, and the search continues for the residues responsible for this error-prone action. Since template slippage is a sequence-dependent mechanism, which occurs even upon undamaged DNA, it follows that the impetus for slipping comes from interactions

with the major groove rather than sequence independent minor groove interactions. Based on this assumption, the LF domain remains a good candidate for further analysis. One hypothesis presented in this dissertation involves β12-strand residues Lys 460 and Lys 458 (human residues Lys 461 and 459 in Figure 5.1) and how they may alter the support of the T_0 and T_{+1} bases as the substrate moves through the active site. Lys 458 is not conserved among the other Y-family Pols, except for Rev1 (Lys 667), which notably is the only other Y-family Pol capable of -1 frameshift deletions. In both Pol κ and Rev1, the lysine (mouse Pol κ Lys 458) is within hydrogen bonding distance to a β11-strand residue Glu (mouse Pol κ Glu 418), a residue that is generally conserved among the other Y-family Pols and is directed into the major groove of the DNA substrate. Mutation of this Lys to shorter residues such as those found in Pol η (Ser) or Dpo4(Val) will destabilize the Glu as well, but this rigidly held Glu may be the reason that T_{+1} phosphate groups "jump" the course and become extruded. Relaxing this "speed bump" may alter frameshift propensity. Kinetic experiments similar to those performed earlier can monitor template frameshift activity, but in the interest of cost and time, unmodified oligos that still promote slippage can be utilized.

The enzymological studies uncovered a novel primer realignment activity. Changes in the LF domain in proximity to the template strand altered the stability of the primer strand, since it appears that Pol κ lacks significant interactions with the growing primer strand. We hypothesized that the thumb residue Arg 352 is preventing the formation of a Mg^{2+} binding pocket which in turn positions the primer terminus for nucleotide insertion by interactions with the base beneath it. While we cannot assume that replacing the Arg with something smaller will permit a magnesium ion to bind in that

location, the rest of the elements required to create the coordination sphere are present and may prevent the primer strand from letting go of the template base opposite it. Kinetic experiments similar to those performed earlier will monitor changes in the kinetic parameters of nucleotide incorporation.)

B. The binary complex structure of mouse Pol κ-I and DNA

The structure of a putative complex of mouse Pol κ as it initially loads the DNA substrate into its active site was solved. The dimerized core of Pol κ consists of only the palm, thumb, and little finger domains, indicating that the finger domain was more flexible than previously observed. It is unfortunate that the crystals were not reproducible, but this might attest to the sensitive and transient environment which triggered the conformational changes necessary to adopt the dimer in solution.

Since a structure of human Pol κ in complex with undamaged has been solved DNA (70), observing Pol κ in the presence of damaged DNA is of great interest. Since Pol κ is very efficient at bypassing the bulky adducts (+/-) trans-BPDE-dG, it would be interesting to perform a step-wise study with this damaged base in the $T_{-1,}$ T_o, and T_{+1} positions to observe how Pol κ is able to accommodate, incorporate, and extend past this lesion. The physiological relevance of this particular complex is high because BPDE is a potent food and cigarette smoke carcinogen, and because to date, no structure has been solved of a DNA polymerase with BPDE-dG in its active site. However, any of the lesions described previously, preferred or not, would be interesting as well.

Another interesting experiment to pursue would be the structure of Pol κ in complex with (-1) frameshifted DNA. The kinetic analyses discussed earlier showed that

Pol κ misaligns template strands preferentially and reproducibly, and so a suitable sequence context can trigger template misalignment, and appropriate rate limiting concentrations of substrates should trap this slipped ternary complex for crystallization purposes. If this structure can be obtained it would provide a clear view of the residues that are responsible for stabilizing the slipping product, and may itself suggest further targets for mutational analysis. Because of the number of mechanisms at work within the Pol κ active site, including correct base pairing, mismatched base pairing, and template frameshifts, there is still much to learn about this enigmatic Y-family DNA polymerase.

Works Cited:

1. Albertella, M.R., Lau, A., and O'Connor, M.J., "The overexpression of specialized DNA polymerases in cancer". DNA repair, 2005. **4**(5): p. 583-593.

2. Alberts, B., *Molecular biology of the cell*. 1994, New York: Garland Pub.

3. Artymiuk, P.J., Poirrette, A.R., Rice, D.W., and Willett, P., "A polymerase I palm in adenylyl cyclase?" Nature, 1997. **388**(6637): p. 33-34.

4. Baynton, K. and Fuchs, R.P., "Lesions in DNA: hurdles for polymerases". Trends in biochemical sciences, 2000. **25**(2): p. 74-79.

5. Beard, W.A. and Wilson, S.H., "Structure and mechanism of DNA polymerase Beta". Chemical reviews, 2006. **106**(2): p. 361-382.

6. Bi, X., Barkley, L.R., Slater, D.M., Tateishi, S., Yamaizumi, M., Ohmori, H., and Vaziri, C., "Rad18 regulates DNA polymerase kappa and is required for recovery from S-phase checkpoint-mediated arrest". Molecular and cellular biology, 2006. **26**(9): p. 3527-3540.

7. Bi, X., Slater, D.M., Ohmori, H., and Vaziri, C., "DNA polymerase kappa is specifically required for recovery from the benzo[a]pyrene-dihydrodiol epoxide (BPDE)-induced S-phase checkpoint". J Biol Chem, 2005. **280**(23): p. 22343-22355.

8. Borys, E. and Kusmierek, J.T., "Endogenous and exogenous DNA lesions recognized by N-alkylpurine-DNA glycosylases". Acta biochimica Polonica, 1998. **45**(2): p. 579-586.

9. Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R., "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB- like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta". Nucleic acids research, 2001. **29**(22): p. 4607-4616.

10. Boudsocq, F., Kokoska, R.J., Plosky, B.S., Vaisman, A., Ling, H., Kunkel, T.A., Yang, W., and Woodgate, R., "Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication". J Biol Chem, 2004. **279**(31): p. 32932-32940.

11. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L., "Crystallography & NMR system: A new software suite for macromolecular structure determination". Acta crystallographica, 1998. **54**(Pt 5): p. 905-921.

12. Byrns, M.C., Vu, C.C., and Peterson, L.A., "The formation of substituted 1,N6-etheno-2'-deoxyadenosine and 1,N2-etheno-2'-deoxyguanosine adducts by cis-2-butene-1,4-dial, a reactive metabolite of furan". Chemical research in toxicology, 2004. **17**(12): p. 1607-1613.

13. Cleaver, J.E., "Mechanisms by which human cells bypass damaged bases during DNA replication after ultraviolet irradiation". ScientificWorldJournal, 2002. **2**: p. 1296-1305.

14. Creighton, T.E., *Proteins : structures and molecular properties*. 1993, New York: W.H. Freeman.

15. de los Santos, C., Kouchakdjian, M., Yarema, K., Basu, A., Essigmann, J., and Patel, D.J., "NMR studies of the exocyclic 1,N6-ethenodeoxyadenosine adduct (epsilon dA) opposite deoxyguanosine in a DNA duplex. Epsilon dA(syn).dG(anti) pairing at the lesion site". Biochemistry, 1991. **30**(7): p. 1828-1835.

16. DeLano, W.L., *The PyMol Molecular Graphics system*. 2002, Delano Scientific LLC: Palo Alto, CA.

17. Dosanjh, M.K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., and Singer, B., "All four known cyclic adducts formed in DNA by the vinyl chloride metabolite chloroacetaldehyde are released by a human DNA glycosylase". Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(3): p. 1024-1028.

18. Easterby, J.S., *HYPER*. 1993: Liverpool, England. p. Hyperbolic Regression Analysis of Enzyme Kinetic Data.

19. Emsley, P. and Cowtan, K., "Coot: model-building tools for molecular graphics". Acta crystallographica, 2004. **60**(Pt 12 Pt 1): p. 2126-2132.

20. Fischhaber, P.L., Gerlach, V.L., Feaver, W.J., Hatahet, Z., Wallace, S.S., and Friedberg, E.C., "Human DNA polymerase kappa bypasses and extends beyond thymine glycols during translesion synthesis in vitro, preferentially incorporating correct nucleotides". J Biol Chem, 2002. **277**(40): p. 37604-37611.

21. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R., and Steitz, T.A., "Cocrystal structure of an editing complex of Klenow fragment with DNA". Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(23): p. 8924-8928.

22. Friedberg, E.C., "Xeroderma pigmentosum: recent studies on the DNA repair defects". Archives of pathology & laboratory medicine, 1978. **102**(1): p. 3-7.

23. Friedberg, E.C., "Xeroderma pigmentosum, Cockayne's syndrome, helicases, and DNA repair: what's the relationship?" Cell, 1992. **71**(6): p. 887-889.

24. Friedberg, E.C., "Why do cells have multiple error-prone DNA polymerases?" Environmental and molecular mutagenesis, 2001. **38**(2-3): p. 105-110.

25. Friedberg, E.C., "DNA damage and repair". Nature, 2003. **421**(6921): p. 436-440.

26. Friedberg, E.C., Fischhaber, P.L., and Kisker, C., "Error-prone DNA polymerases: novel structures and the benefits of infidelity". Cell, 2001. **107**(1): p. 9-12.

27. Garg, P., Stith, C.M., Majka, J., and Burgers, P.M., "Proliferating cell nuclear antigen promotes translesion synthesis by DNA polymerase zeta". J Biol Chem, 2005. **280**(25): p. 23446-23450.

28. Gerlach, V.L., Aravind, L., Gotway, G., Schultz, R.A., Koonin, E.V., and Friedberg, E.C., "Human and mouse homologs of Escherichia coli DinB (DNA polymerase IV), members of the UmuC/DinB superfamily". Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(21): p. 11922-11927.

29. Gerlach, V.L., Feaver, W.J., Fischhaber, P.L., and Friedberg, E.C., "Purification and characterization of pol kappa, a DNA polymerase encoded by the human DINB1 gene". J Biol Chem, 2001. **276**(1): p. 92-98.

30. Goodman, M.F., "Error-prone repair DNA polymerases in prokaryotes and eukaryotes". Annual review of biochemistry, 2002. **71**: p. 17-50.

31. Goodman, M.F., Creighton, S., Bloom, L.B., and Petruska, J., "Biochemical basis of DNA replication fidelity". Crit Rev Biochem Mol Biol, 1993. **28**(2): p. 83-126.

32. Goodman, M.F. and Tippin, B., "Sloppier copier DNA polymerases involved in genome repair". Current opinion in genetics & development, 2000. **10**(2): p. 162-168.

33. Guex, N. and Peitsch, M.C., "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling". Electrophoresis, 1997. **18**(15): p. 2714-2723.

34. Guo, C., Fischhaber, P.L., Luk-Paszyc, M.J., Masuda, Y., Zhou, J., Kamiya, K., Kisker, C., and Friedberg, E.C., "Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis". Embo J, 2003. **22**(24): p. 6621-6630.

35. Hall, T., "BioEdit: a user-friendly biological sequence alignment editor and analysis. " 1999.

36. Hang, B., Chenna, A., Rao, S., and Singer, B., "1,N6-ethenoadenine and 3,N4-ethenocytosine are excised by separate human DNA glycosylases". Carcinogenesis, 1996. **17**(1): p. 155-157.

37. Haracska, L., Acharya, N., Unk, I., Johnson, R.E., Hurwitz, J., Prakash, L., and Prakash, S., "A single domain in human DNA polymerase iota mediates interaction with PCNA: implications for translesion DNA synthesis". Molecular and cellular biology, 2005. **25**(3): p. 1183-1190.

38. Haracska, L., Johnson, R.E., Unk, I., Phillips, B.B., Hurwitz, J., Prakash, L., and Prakash, S., "Targeting of human DNA polymerase iota to the replication machinery via interaction with PCNA". Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(25): p. 14256-14261.

39. Haracska, L., Kondratick, C.M., Unk, I., Prakash, S., and Prakash, L., "Interaction with PCNA is essential for yeast DNA polymerase eta function". Molecular cell, 2001. **8**(2): p. 407-415.

40. Haracska, L., Prakash, L., and Prakash, S., "Role of human DNA polymerase kappa as an extender in translesion synthesis". Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(25): p. 16000-16005.

41. Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., and Prakash, S., "Efficient and accurate replication in the presence of 7,8-dihydro-8- oxoguanine by DNA polymerase eta". Nat Genet, 2000. **25**(4): p. 458-461.

42. Hopfner, K.P., Eichinger, A., Engh, R.A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B., "Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius". Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 3600-3605.

43. Hubscher, U., Maga, G., and Spadari, S., "Eukaryotic DNA polymerases". Annual review of biochemistry, 2002. **71**: p. 133-163.

44. Hutchinson, E.G. and Thornton, J.M., "PROMOTIF--a program to identify and analyze structural motifs in proteins". Protein Sci, 1996. **5**(2): p. 212-220.

45. Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Jr., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., and et al., "Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA". Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(13): p. 6320-6324.

46. Jaloszynski, P., Ohashi, E., Ohmori, H., and Nishimura, S., "Error-prone and inefficient replication across 8-hydroxyguanine (8-oxoguanine) in human and mouse ras gene fragments by DNA polymerase kappa". Genes Cells, 2005. **10**(6): p. 543-550.

47. Johnson, R.E., Prakash, L., and Prakash, S., "Distinct mechanisms of cis-syn thymine dimer bypass by Dpo4 and DNA polymerase {eta}". Proceedings of the

National Academy of Sciences of the United States of America, 2005. **102**(35): p. 12359-12364.

48. Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S., and Prakash, L., "Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions". Nature, 2000. **406**(6799): p. 1015-1019.

49. Johnson, R.E., Washington, M.T., Prakash, S., and Prakash, L., "Fidelity of human DNA polymerase eta". J Biol Chem, 2000. **275**(11): p. 7447-7450.

50. Johnson, S.J., Taylor, J.S., and Beese, L.S., "Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations". Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(7): p. 3895-3900.

51. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M., "Improved methods for building protein models in electron density maps and the location of errors in these models". Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-119.

52. Kai, M. and Wang, T.S., "Checkpoint activation regulates mutagenic translesion synthesis". Genes Dev, 2003. **17**(1): p. 64-76.

53. Kai, M. and Wang, T.S., "Checkpoint responses to replication stalling: inducing tolerance and preventing mutagenesis". Mutation research, 2003. **532**(1-2): p. 59-73.

54. Kannouche, P. and Lehmann, A., "Localization of y-family polymerases and the DNA polymerase switch in mammalian cells". Methods in enzymology, 2006. **408**: p. 407-415.

55. Kannouche, P. and Stary, A., "Xeroderma pigmentosum variant and error-prone DNA polymerases". Biochimie, 2003. **85**(11): p. 1123-1132.

56. Kannouche, P.L. and Lehmann, A.R., "Ubiquitination of PCNA and the Polymerase Switch in Human Cells". Cell cycle (Georgetown, Tex, 2004. **3**(8).

57. Kannouche, P.L., Wing, J., and Lehmann, A.R., "Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage". Molecular cell, 2004. **14**(4): p. 491-500.

58. Kim, T.W., Brieba, L.G., Ellenberger, T., and Kool, E.T., "Functional evidence for a small and rigid active site in a high fidelity DNA polymerase: probing T7 DNA polymerase with variably sized base pairs". J Biol Chem, 2006. **281**(4): p. 2289-2295.

59. Kim, T.W., Delaney, J.C., Essigmann, J.M., and Kool, E.T., "Probing the active site tightness of DNA polymerase in subangstrom increments". Proceedings of the

National Academy of Sciences of the United States of America, 2005. **102**(44): p. 15803-15808.

60. Kokoska, R.J., Bebenek, K., Boudsocq, F., Woodgate, R., and Kunkel, T.A., "Low fidelity DNA synthesis by a y family DNA polymerase due to misalignment in the active site". J Biol Chem, 2002. **277**(22): p. 19633-19638.

61. Krissinel, E. and Henrick, K., "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions". Acta crystallographica, 2004. **60**(Pt 12 Pt 1): p. 2256-2268.

62. Laskowski, R.A., McArthur, M.W., Moss, D.S., and Thornton, J.M., "PROCHECK: a program to check the stereochemical quality of protein structures." J Appl Crystallogr, 1993. **26**: p. 283-291.

63. Lee, C.H., Chandani, S., and Loechler, E.L., "Homology modeling of four Y-family, lesion-bypass DNA polymerases: the case that E. coli Pol IV and human Pol kappa are orthologs, and E. coli Pol V and human Pol eta are orthologs". Journal of molecular graphics & modelling, 2006. **25**(1): p. 87-102.

64. Leonard, G.A., McAuley-Hecht, K.E., Gibson, N.J., Brown, T., Watson, W.P., and Hunter, W.N., "Guanine-1,N6-ethenoadenine base pairs in the crystal structure of d(CGCGAATT(epsilon dA)GCG)". Biochemistry, 1994. **33**(16): p. 4755-4761.

65. Levine, R.L., Miller, H., Grollman, A., Ohashi, E., Ohmori, H., Masutani, C., Hanaoka, F., and Moriya, M., "Translesion DNA synthesis catalyzed by human pol eta and pol kappa across 1,N6-ethenodeoxyadenosine". J Biol Chem, 2001. **276**(22): p. 18717-18721.

66. Ling, H., Boudsocq, F., Plosky, B.S., Woodgate, R., and Yang, W., "Replication of a cis-syn thymine dimer at atomic resolution". Nature, 2003. **424**(6952): p. 1083-1087.

67. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W., "Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication". Cell, 2001. **107**(1): p. 91-102.

68. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W., "Snapshots of Replication through an Abasic Lesion; Structural Basis for Base Substitutions and Frameshifts". Molecular cell, 2004. **13**(5): p. 751-762.

69. Ling, H., Sayer, J.M., Plosky, B.S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D.M., and Yang, W., "Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase". Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(8): p. 2265-2269.

70. Lone, S., Townson, S.A., Uljon, S.N., Johnson, R.E., Brahma, A., Nair, D.T., Prakash, S., Prakash, L., and Aggarwal, A.K., "Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass". Molecular cell, 2007. **25**(4): p. 601-614.

71. Maga, G. and Hubscher, U., "Proliferating cell nuclear antigen (PCNA): a dancer with many partners". J Cell Sci, 2003. **116**(Pt 15): p. 3051-3060.

72. Masuda, Y., Takahashi, M., Fukuda, S., Sumii, M., and Kamiya, K., "Mechanisms of dCMP transferase reactions catalyzed by mouse Rev1 protein". J Biol Chem, 2002. **277**(4): p. 3040-3046.

73. Masutani, C., Kusumoto, R., Iwai, S., and Hanaoka, F., "Mechanisms of accurate translesion synthesis by human DNA polymerase eta". Embo J, 2000. **19**(12): p. 3100-3109.

74. Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F., "The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta". Nature, 1999. **399**(6737): p. 700-704.

75. Mathews, C.K. and Sinha, N.K., "Are DNA precursors concentrated at replication sites?" Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(2): p. 302-306.

76. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J., "Likelihood-enhanced fast translation functions". Acta crystallographica, 2005. **61**(Pt 4): p. 458-464.

77. McDonald, J.P., Rapic-Otrin, V., Epstein, J.A., Broughton, B.C., Wang, X., Lehmann, A.R., Wolgemuth, D.J., and Woodgate, R., "Novel human and mouse homologs of Saccharomyces cerevisiae DNA polymerase eta". Genomics, 1999. **60**(1): p. 20-30.

78. McDonald, J.P., Tissier, A., Frank, E.G., Iwai, S., Hanaoka, F., and Woodgate, R., "DNA polymerase iota and related rad30-like enzymes". Philosophical transactions of the Royal Society of London, 2001. **356**(1405): p. 53-60.

79. Murshudov, G.N., Vagin, A.A., and Dodson, E.J., "Refinement of macromolecular structures by the maximum-likelihood method". Acta crystallographica, 1997. **53**(Pt 3): p. 240-255.

80. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K., "Human DNA polymerase iota incorporates dCTP opposite template G via a G.C + Hoogsteen base pair". Structure, 2005. **13**(10): p. 1569-1577.

81. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K., "Rev1 employs a novel mechanism of DNA synthesis using a protein template". Science (New York, N.Y, 2005. **309**(5744): p. 2219-2222.

82. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K., "Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase iota". Nat Struct Mol Biol, 2006. **13**(7): p. 619-625.

83. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K., "An incoming nucleotide imposes an anti to syn conformational change on the templating purine in the human DNA polymerase-iota active site". Structure, 2006. **14**(4): p. 749-755.

84. Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L., and Aggarwal, A.K., "Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing". Nature, 2004. **430**(6997): p. 377-380.

85. Nair, J., "Lipid peroxidation-induced etheno-DNA adducts in humans". IARC Sci Publ, 1999(150): p. 55-61.

86. Nair, J., Barbin, A., Velic, I., and Bartsch, H., "Etheno DNA-base adducts from endogenous reactive species". Mutation research, 1999. **424**(1-2): p. 59-69.

87. Napolitano, R., Janel-Bintz, R., Wagner, J., and Fuchs, R.P., "All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis". Embo J, 2000. **19**(22): p. 6259-6265.

88. O-Wang, J., Kawamura, K., Tada, Y., Ohmori, H., Kimura, H., Sakiyama, S., and Tagawa, M., "DNA polymerase kappa, implicated in spontaneous and DNA damage-induced mutagenesis, is overexpressed in lung cancer". Cancer research, 2001. **61**(14): p. 5366-5369.

89. Ogi, T., Kannouche, P., and Lehmann, A.R., "Localisation of human Y-family DNA polymerase kappa: relationship to PCNA foci". J Cell Sci, 2005. **118**(Pt 1): p. 129-136.

90. Ogi, T. and Lehmann, A.R., "The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair". Nat Cell Biol, 2006. **8**(6): p. 640-642.

91. Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., Ohmori, H., and Kunkel, T.A., "Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene". J Biol Chem, 2000. **275**(50): p. 39678-39684.

92. Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F., and Ohmori, H., "Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa". Genes Dev, 2000. **14**(13): p. 1589-1594.

93. Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G.C., Wang, Z., and Woodgate, R., "The Y-family of DNA polymerases". Molecular cell, 2001. **8**(1): p. 7-8.

94. Otwinowski, Z. and Minor, W., eds. *Processing of X-ray Diffraction Data Collected in Oscillation Mode*. Methods in enzymology, ed. C.J. Carter and R. Sweet. Vol. 276. 1997, Academic Press New York. 307-326.

95. Pan, Q., Fang, Y., Xu, Y., Zhang, K., and Hu, X., "Down-regulation of DNA polymerases kappa, eta, iota, and zeta in human lung, stomach, and colorectal cancers". Cancer Lett, 2005. **217**(2): p. 139-147.

96. Prakash, S., Johnson, R.E., and Prakash, L., "Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function". Annual review of biochemistry, 2005. **74**: p. 317-353.

97. Rattray, A.J. and Strathern, J.N., "Error-prone DNA polymerases: when making a mistake is the only way to get ahead". Annu Rev Genet, 2003. **37**: p. 31-66.

98. Rechkoblit, O., Zhang, Y., Guo, D., Wang, Z., Amin, S., Krzeminsky, J., Louneva, N., and Geacintov, N.E., "Translesion synthesis past bulky benzo[a]pyrene diol epoxide N2-dG and N6-dA lesions catalyzed by DNA bypass polymerases". J Biol Chem, 2002. **277**(34): p. 30488-30494.

99. Silverthorn, D.U., *Human physiology : an integrated approach*. 1998, Upper Saddle River, N.J.: Prentice Hall.

100. Silvian, L.F., Toth, E.A., Pham, P., Goodman, M.F., and Ellenberger, T., "Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus". Nat Struct Biol, 2001. **8**(11): p. 984-989.

101. Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D., "Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis". Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(21): p. 8060-8065.

102. Suzuki, N., Ohashi, E., Hayashi, K., Ohmori, H., Grollman, A.P., and Shibutani, S., "Translesional synthesis past acetylaminofluorene-derived DNA adducts catalyzed by human DNA polymerase kappa and Escherichia coli DNA polymerase IV". Biochemistry, 2001. **40**(50): p. 15176-15183.

103. Suzuki, N., Ohashi, E., Kolbanovskiy, A., Geacintov, N.E., Grollman, A.P., Ohmori, H., and Shibutani, S., "Translesion synthesis by human DNA polymerase kappa on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-anti-N(2)-BPDE (7,8-dihydroxy-anti-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene)". Biochemistry, 2002. **41**(19): p. 6100-6106.

104. Suzuki, N., Yasui, M., Santosh Laxmi, Y.R., Ohmori, H., Hanaoka, F., and Shibutani, S., "Translesion synthesis past equine estrogen-derived 2'-deoxycytidine DNA adducts by human DNA polymerases eta and kappa". Biochemistry, 2004. **43**(35): p. 11312-11320.

105. Swenberg, J.A., Bogdanffy, M.S., Ham, A., Holt, S., Kim, A., Morinello, E.J., Ranasinghe, A., Scheller, N., and Upton, P.B., "Formation and repair of DNA adducts in vinyl chloride- and vinyl fluoride-induced carcinogenesis". IARC Sci Publ, 1999(150): p. 29-43.

106. Swenberg, J.A., Fedtke, N., Ciroussel, F., Barbin, A., and Bartsch, H., "Etheno adducts formed in DNA of vinyl chloride-exposed rats are highly persistent in liver". Carcinogenesis, 1992. **13**(4): p. 727-729.

107. Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., and Goodman, M.F., "UmuD'(2)C is an error-prone DNA polymerase, Escherichia coli pol V". Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(16): p. 8919-8924.

108. Thompson, J.D., Higgins, D.G., and Gibson, T.J., "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". Nucleic acids research, 1994. **22**(22): p. 4673-4680.

109. Trincao, J., Johnson, R.E., Escalante, C.R., Prakash, S., Prakash, L., and Aggarwal, A.K., "Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis". Molecular cell, 2001. **8**(2): p. 417-426.

110. Uljon, S.N., Johnson, R.E., Edwards, T.A., Prakash, S., Prakash, L., and Aggarwal, A.K., "Crystal structure of the catalytic core of human DNA polymerase kappa". Structure (Camb), 2004. **12**(8): p. 1395-1404.

111. Vagin, A. and Teplyakov, A., "An approach to multi-copy search in molecular replacement". Acta crystallographica, 2000. **56**(Pt 12): p. 1622-1624.

112. Vaisman, A., Frank, E.G., McDonald, J.P., Tissier, A., and Woodgate, R., "poliota-dependent lesion bypass in vitro". Mutation research, 2002. **510**(1-2): p. 9-22.

113. Vaisman, A., Masutani, C., Hanaoka, F., and Chaney, S.G., "Efficient translesion replication past oxaliplatin and cisplatin GpG adducts by human DNA polymerase eta". Biochemistry, 2000. **39**(16): p. 4575-4580.

114. Velasco-Miguel, S., Richardson, J.A., Gerlach, V.L., Lai, W.C., Gao, T., Russell, L.D., Hladik, C.L., White, C.L., and Friedberg, E.C., "Constitutive and regulated expression of the mouse Dinb (Polkappa) gene encoding DNA polymerase kappa". DNA repair, 2003. **2**(1): p. 91-106.

115. Wang, J., Sattar, A.K., Wang, C.C., Karam, J.D., Konigsberg, W.H., and Steitz, T.A., "Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69". Cell, 1997. **89**(7): p. 1087-1099.

116. Wang, J., Smerdon, S.J., Jager, J., Kohlstaedt, L.A., Rice, P.A., Friedman, J.M., and Steitz, T.A., "Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer". Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(15): p. 7242-7246.

117. Wang, Z., "Translesion synthesis by the UmuC family of DNA polymerases". Mutation research, 2001. **486**(2): p. 59-70.

118. Washington, M.T., Johnson, R.E., Prakash, L., and Prakash, S., "Human DINB1-encoded DNA polymerase kappa is a promiscuous extender of mispaired primer termini". Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 1910-1914.

119. Washington, M.T., Prakash, L., and Prakash, S., "Yeast DNA polymerase eta utilizes an induced-fit mechanism of nucleotide incorporation". Cell, 2001. **107**(7): p. 917-927.

120. Winn, M.D., Murshudov, G.N., and Papiz, M.Z., "Macromolecular TLS refinement in REFMAC at moderate resolutions". Methods in enzymology, 2003. **374**: p. 300-321.

121. Wolfle, W.T., Washington, M.T., Kool, E.T., Spratt, T.E., Helquist, S.A., Prakash, L., and Prakash, S., "Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase kappa". Molecular and cellular biology, 2005. **25**(16): p. 7137-7143.

122. Wolfle, W.T., Washington, M.T., Prakash, L., and Prakash, S., "Human DNA polymerase kappa uses template-primer misalignment as a novel means for extending mispaired termini and for generating single-base deletions". Genes Dev, 2003. **17**(17): p. 2191-2199.

123. Wozniak, K. and Blasiak, J., "Recognition and repair of DNA-cisplatin adducts". Acta biochimica Polonica, 2002. **49**(3): p. 583-596.

124. Yagi, Y., Ogawara, D., Iwai, S., Hanaoka, F., Akiyama, M., and Maki, H., "DNA polymerases eta and kappa are responsible for error-free translesion DNA synthesis activity over a cis-syn thymine dimer in Xenopus laevis oocyte extracts". DNA repair, 2005.

125. Yang, W., "Damage repair DNA polymerases Y". Current opinion in structural biology, 2003. **13**(1): p. 23-30.

126. Yasui, M., Laxmi, Y.R., Ananthoju, S.R., Suzuki, N., Kim, S.Y., and Shibutani, S., "Translesion synthesis past equine estrogen-derived 2'-deoxyadenosine DNA adducts by human DNA polymerases eta and kappa". Biochemistry, 2006. **45**(19): p. 6187-6194.

127. Yu, S.L., Johnson, R.E., Prakash, S., and Prakash, L., "Requirement of DNA polymerase eta for error-free bypass of UV-induced CC and TC photoproducts". Molecular and cellular biology, 2001. **21**(1): p. 185-188.

128. Zhang, Y., Wu, X., Guo, D., Rechkoblit, O., Geacintov, N.E., and Wang, Z., "Two-step error-prone bypass of the (+)- and (-)-trans-anti-BPDE-N2-dG adducts by human DNA polymerases eta and kappa". Mutation research, 2002. **510**(1-2): p. 23-35.

129. Zhang, Y., Wu, X., Guo, D., Rechkoblit, O., and Wang, Z., "Activities of human DNA polymerase kappa in response to the major benzo[a]pyrene DNA adduct: error-free lesion bypass and extension synthesis from opposite the lesion". DNA repair, 2002. **1**(7): p. 559-569.

130. Zhang, Y., Yuan, F., Wu, X., Rechkoblit, O., Taylor, J.S., Geacintov, N.E., and Wang, Z., "Error-prone lesion bypass by human DNA polymerase eta". Nucleic acids research, 2000. **28**(23): p. 4717-4724.

131. Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J.S., Geacintov, N.E., and Wang, Z., "Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro". Nucleic acids research, 2000. **28**(21): p. 4138-4146.

132. Zhang, Y., Yuan, F., Xin, H., Wu, X., Rajpal, D.K., Yang, D., and Wang, Z., "Human DNA polymerase kappa synthesizes DNA with extraordinarily low fidelity". Nucleic acids research, 2000. **28**(21): p. 4147-4156.

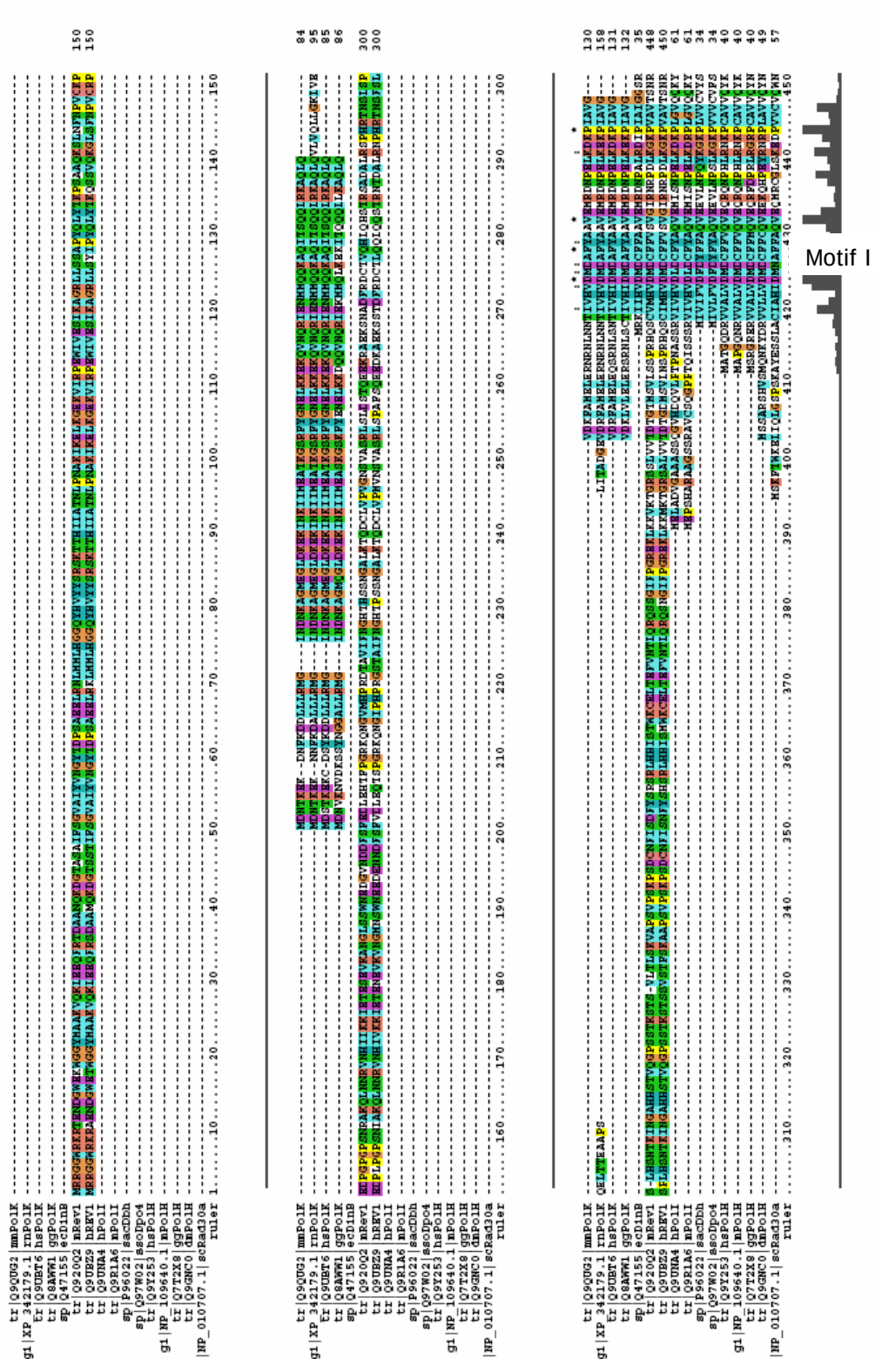
133. Zhou, B.L., Pata, J.D., and Steitz, T.A., "Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain". Molecular cell, 2001. **8**(2): p. 427-437.

Appendices

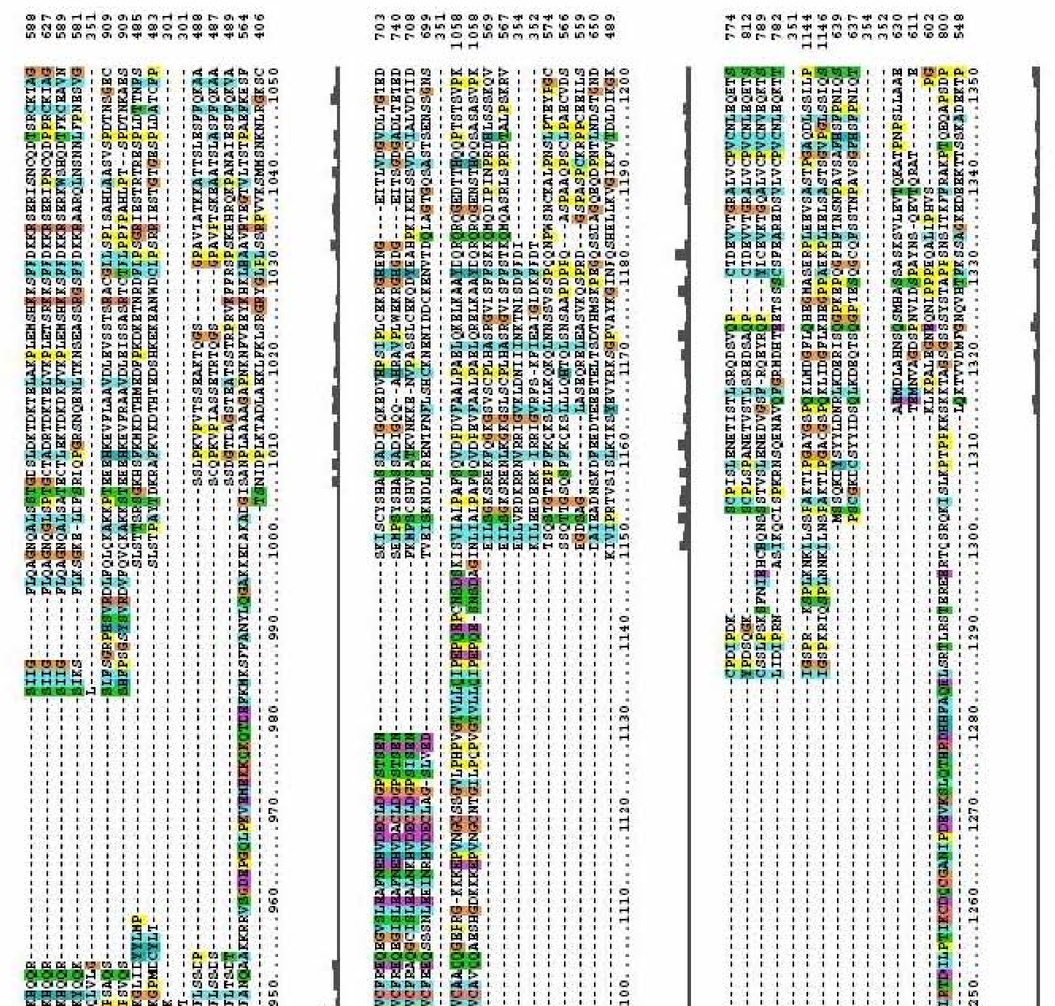
Supplemental material

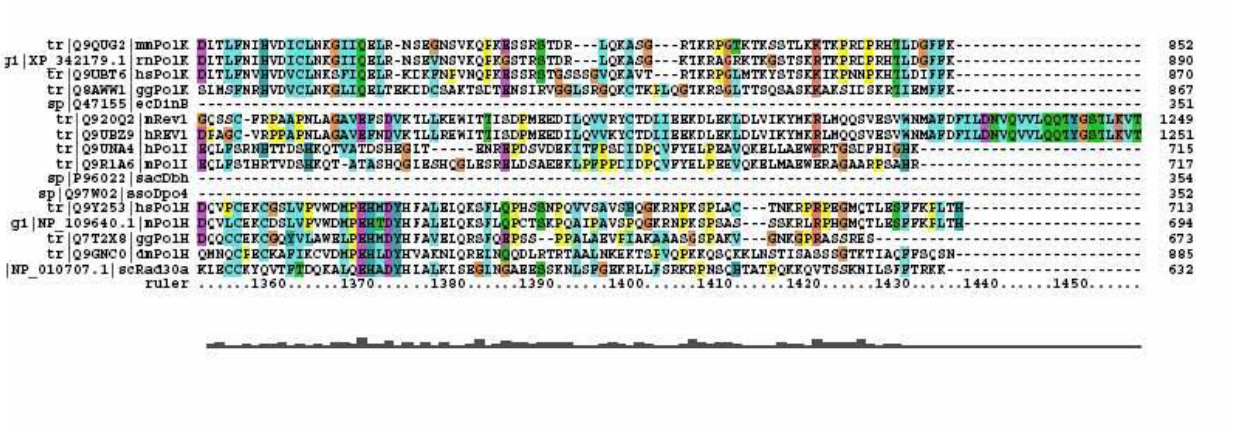
Appendix **A.1: CLUSTALX multiple sequence alignment output.**

Motifs I-V have been indicated under their representative sequences.

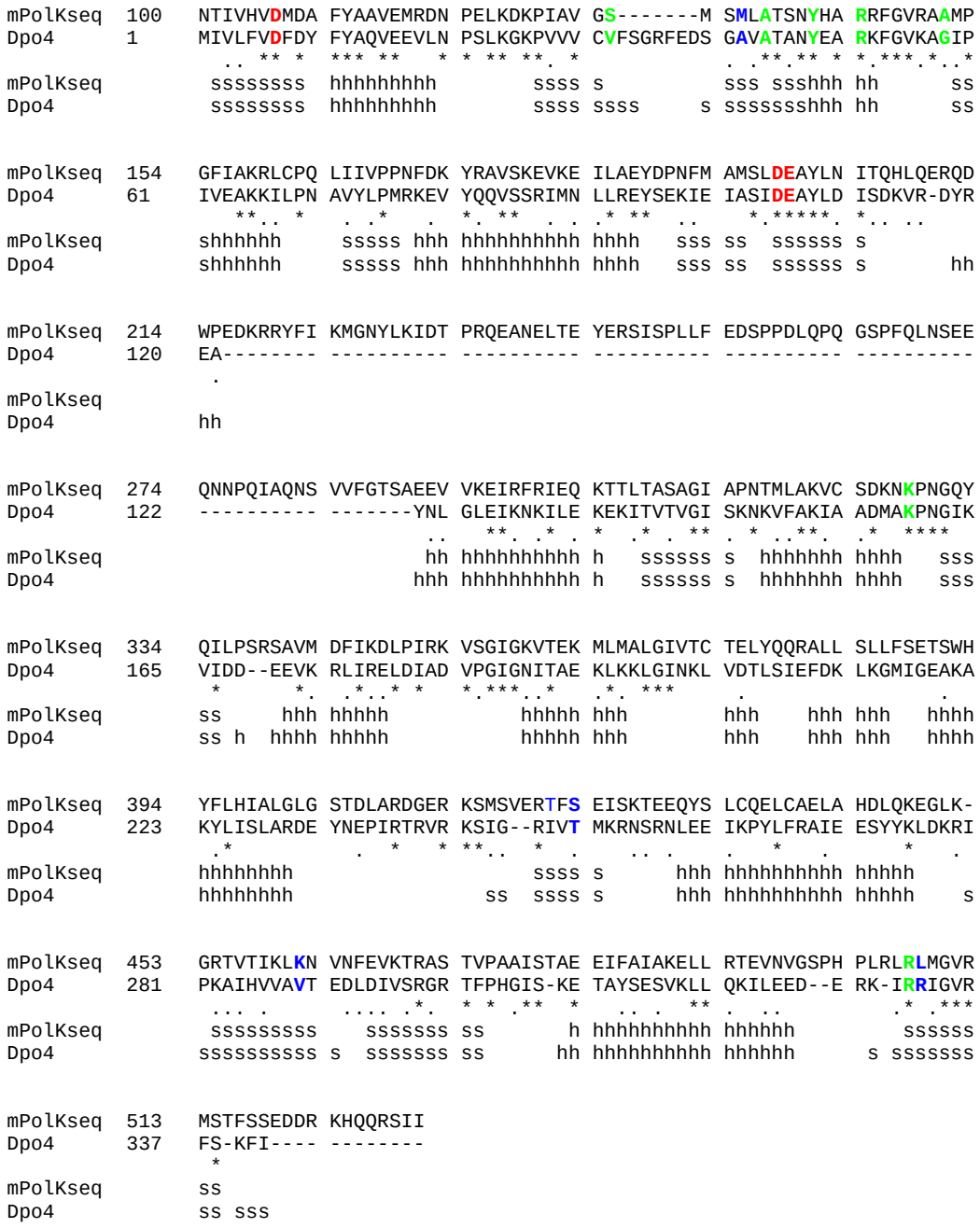


Motif II
Motif III
Motif IV
Motif V




tr|Q9QUG2|mmPolK
gi|XP_342179.1|rnPolK
tr|Q9UBT6|hsPolK
tr|Q8AWW1|ggPolK
sp|Q47155|ecDinB
tr|Q920Q2|mRev1
tr|Q9UBZ9|hREV1
tr|Q9UNA4|hPolI
tr|Q9R1A6|mPolI
sp|P96022|sacDbh
sp|Q97W02|ssoDpo4
tr|Q9Y253|hsPolH
gi|NP_109640.1|mPolH
tr|Q7T2X8|ggPolH
tr|Q9GNC0|dmPolH
|NP_010707.1|scRad30a
ruler
852
890
870
867
351
1249
1251
715
717
354
352
713
694
673
885
632

Appendix A.2 Sequence alignment of mouse Pol κ (100 – 524) on full-length Dpo4



Similarity symbols: dash (-), insertion; asterisk (*), identical residue; period (.), similar residue. Structural alignment: (s), beta strand; (h), alpha helix. Highlighted residues: strictly conserved catalytic triad, red; conserved substrate coordination residues, green; and unconserved substrate coordination residues, blue.

Appendix B: All oligonucleotides used in this dissertation with their purpose.

B.1

Mutation	Mutagenesis Primer	length	Tm*
Finger			
M134A	5'-GCTGTAGGATCCATGAGC**GCG**TTGGCTACTTCGAATTAC	39	79
M134Q	5'-GCTGTAGGATCCATGAGC**CAG**TTGGCTACTTCGAATTAC	39	77
Little Finger			
L507M	5'-CACCCCCTGCGGTTAAGA**ATG**ATGGGTGTCCGAATGTC	38	74.4
L507R	5'-CACCCCCTGCGGTTAAGA**CGG**ATGGGTGTCCGAATGTC	38	76.5
L507V	5'-CACCCCCTGCGGTTAAGA**GTG**ATGGGTGTCCGAATGTC	38	75.4
K460F	5'-CCGTCACCATTAAGCTG**TTC**AACGTGAATTTTGAAG	36	68
K460V	5'-CCGTCACCATTAAGCTG**GTG**AACGTGAATTTTGAAG	36	69.2

* (C°, +50 mM NaCl)

B.2

N/C truncation	PCR primer (sense)	length	Tm*
NdeI sites			
N-1	5'-GGAGATATA**CATATG**GAT	18	50.8
N-27	5'-CGATAACAAAGCA**CATATG**GAAGGGTTGG	29	60.0
N-30	5'-GCAGGCATG**CATATG**TTGGATAAGGAG	27	66.1
N-36	5'-GGATAAGGAG**CATATG**AACAAAATTATC	28	60.2
N-49	5'-CAAAGGGGTCC**CATATG**TATGGAAATG	27	64.6
N-54	5'-GATTTTATGGA**CATATG**CTCAAGAAGG	27	61.6
N-68	5'-CGGATTGAAAAT**CATATG**CAACAAAAAGCTC	32	58.0
N-98	5'-GAACGGAACCGG**CATATG**AACAATACC	27	66.1
SapI sites			
C-515	5'-CCTGTCATCT**GCTCTTCA**GCAAGTAGACATTCGG	34	66.0
C-525	5'-GTAAGAAACCAATG**GCTCTTCT**GCATTGGTGTTTCCTG	38	66.0
C-560	5'-CTGATCGCTTTTTATCA**GCTCTTCT**GCACTTATGAGACATTTC	43	69.5
C-616	5'-GCAAACTGGACAT**GCTCTTCT**GCAACTGTCATTTG	35	69.4

* (C°, +50 mM NaCl)

B.3

Crystallization Oligonucleotides		length	Tm*
primer	5'-GTCCTGTTCGTCGTGC	16	64.9
templates			
5 base overhang	5'-CGCACGCACGACGAACAGGAC	21	68.4
6 base overhang	5'-CCGCACGCACGACGAACAGGAC	22	78.8
7 base overhang	5'-ACCGCACGCACGACGAACAGGAC	23	78.9

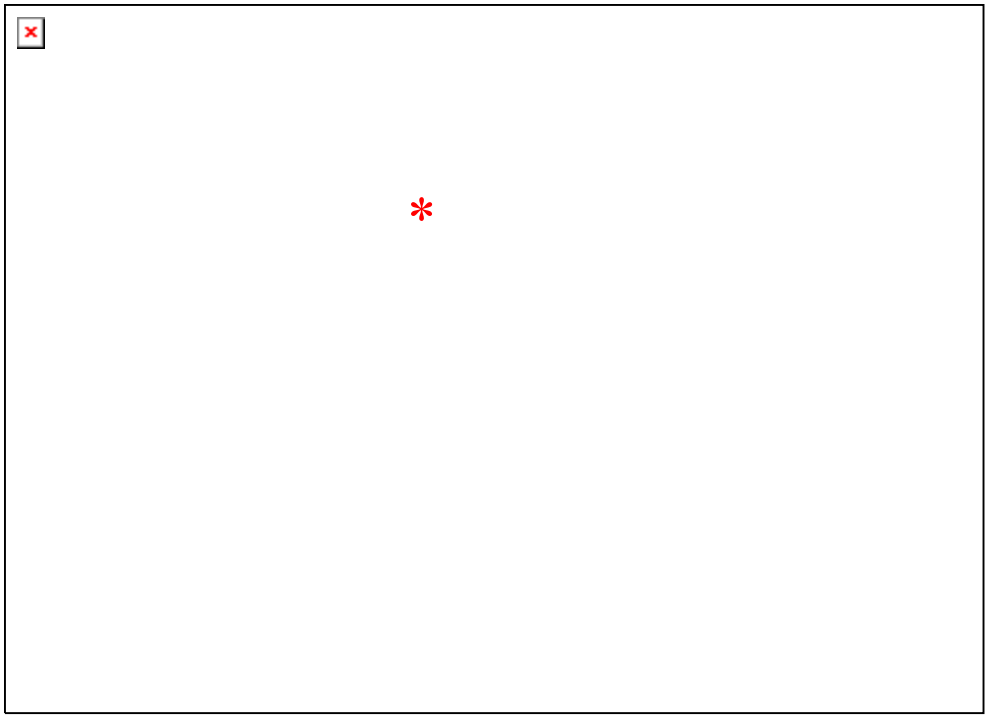
* (C°, +50 mM NaCl)

B.4 Incomplete factorial crystallization screens used for co-crystallization trials

All constructs were with ~5-20 mg/ml protein (if available), with the following screens: Hampton Screens I & II and Lite I & II, Wizard Screens I & II, Topaz Screens Optimix I-III, Nextal Screens Classic PEG suite and pHClear suites I & II. The Clear Strategy Screen was used as well with Tris 7.0, Tris 8.0, and Tris 8.5 as the buffer fraction.

Appendix C: MALDI-TOF analysis of cisplatinated DNA

Top: Cisplatinated 36-mer template contaminated with 2% unmodified template to monitor potential non-adducted contamination products



Bottom: MALDI-TOF result of cisplatinated 36-mer template alone