2014

Y-family DNA polymerase architecture: three structural features control accurate deoxy CTP insertion opposite N2-deoxy-guanine-benzo-a-pyrene

Sholder, Gabriel D.

https://hdl.handle.net/2144/15225

Boston University
Y-FAMILY DNA POLYMERASE ARCHITECTURE:
THREE STRUCTURAL FEATURES CONTROL ACCURATE DEOXY CTP
INSERTION OPPOSITE N2-DEOXY-GUANINE-BENZO-A-PYRENE

by

GABRIEL D. SHOLDER

B.S., Bates College, 2002
M.A., Boston University, 2006

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
2014
Approved by

First Reader
_______________________________________________________
Edward Loechler, Ph.D.
Professor of Biology

Second Reader
_______________________________________________________
Dean Tolan, Ph.D.
Professor of Biology
Acknowledgments

I would like to dedicate this work to my mother for the endless guidance, love, patience, and support that she has given me throughout my life.
Y-FAMILY DNA POLYMERASE ARCHITECTURE: 
THREE STRUCTURAL FEATURES CONTROL ACCURATE DEOXY CTP INSERTION OPPOSITE N2-DEOXY-GUANINE-BENZO-A-PYRENE

GABRIEL D. SHOLDER
Boston University Graduate School of Arts and Sciences, 2014
Major Professor: Edward Loechler, Professor of Biology

ABSTRACT

Cells have lesion bypass DNA polymerases (DNAPs), often in the Y-Family, which synthesize passed DNA damage. One class of Y-Family DNAPs includes hDNAP k, EcDNAP IV and SsDbh, which insert accurately opposite N2-dG adducts, including BP-N2-dG formed from benzo[a]pyrene (BP). Another class includes hDNAP h, EcDNAP V and SsDpo4, which insert accurately opposite UV-damage. For correct Watson-Crick pairing between BP-N2-dG and dCTP, the BP moiety must be in the minor groove. On the minor groove side of the active site, k/IV/Dbh-class DNAPs have large openings that accommodate the BP moiety. Primer extension assays with purified proteins show that DNAP IV correctly inserts dCTP opposite BP more than 10-fold faster than it mis-inserts dATP, dGTP, or dTTP. In contrast, h/V/Dpo4-class DNAPs have small active site openings, which cannot accommodate BP and lead to a distorted structure and increased mutagenesis; e.g., Dpo4 has dGTP and dATP insertion rates that are 10-fold greater than those of dCTP. The opening in Dpo4 is plugged and bulky, whereas DNAP IV has a relatively spacious cavity. Consistent with this model, mutants of Dpo4 with a larger opening insert up to 10-fold more accurately opposite BP-N2-dG. Near the active site, Dpo4 has a single non-covalent
bridge (NCB) between the little finger domain and the thumb-palm-fingers domain. DNAP IV and Dbh have a second, distal NCB that is 8 angstroms away from the active site towards the 3’ end of the template DNA. Dpo4 becomes nearly 5-fold more accurate when mutated to carry a distal NCB, suggesting that NCB’s also help control mutagenesis. Lastly, the active site of Dpo4 has a cavity in the major groove side, which may allow base flipping and dGTP insertion opposite –BP, while k/IV/Dbh-type polymerases do not. When this cavity is plugged in Dpo4 by mutagenesis or the introduction of an N-clasp motif, dGTP rates increase by nearly 20-fold. In conclusion, this data suggests that three structural regions contribute to accurate dCTP insertion opposite BP-N2-dG by k/IV/Dbh-class DNAPs: a large opening on the minor groove side near the active site, a cavity on the major groove side, and the number of non-covalent bridges between the little finger domain and the thumb-palm-fingers domain.
# Table of Contents

## Chapter 1: Introduction

1-25

**Y-family DNA polymerases**

1-2

**Benzo[a]pyrene is a mutagen**

3-4

**E. coli Polymerases IV and V Bypass B[a]P adducts**

4-5

**Y-family DNA Polymerases Bypass Benzo[a] Pyrene**

5-6

**Opening in the Protein Surface on the Minor Groove Side**

6

**Non-Covalent Bridges**

7-9

**Rationale and Goals**

10-13

**Tables and Figures**

14-25

## Chapter 2: Materials and Methods

26-39

**Construction of Cells that Overproduce His-Tagged Dpo4 and DinB**

26-27

**DNAP IV Mutagenesis**

27-28

**Bacterial Transformations**

28-29

**Protein Purification**

29-30

**Protein Quantification**

30-31

**Construction of BP-Containing Templates for Primer Extension Assays**

31

**32P-radiolabeling of Extension Assay Primers**

33

**Primer Annealing**

33
Primer Extension Assays 33
Sequencing Gel and Analyzing Data 34
Tables and Figures 36-39

**Chapter 3: Accurately Quantitating Intensity of Minor 32P-DNA Bands in Gels** 40-62
Summary 40
Introduction 41
Results and Discussion 44
Tables and Figures 54

**Chapter 4: Y-family DNA Polymerase Architecture that Controls Accurate dCTP Insertion Opposite an N2-dG Adduct of Benzo[a]pyrene** 63-106
Summary 63
Introduction 64
Results and Discussion 72
Tables and Figures 93

**Chapter 5: Architecture of Y-family E. coli DNA Polymerase IV That Controls Accurate dCTP Insertion Opposite and N2-dG Adduct** 107-139
Summary 107
Introduction 108
Results and Discussion 113
List of Tables

Chapter 1

Table 1.1: Dominant dNTP insertions opposite various DNA adducts by Y-family Polymerases 14

Chapter 2

Table 2.1: Primers used for cloning DNAP reading frames into pET15b expression vectors 36
Table 2.2: Primers used for site-directed mutagenesis 37
Table 2.3: Oligonucleotides used for the Construction of DNA Templates 39

Chapter 3

Table 3.1: Relative Velocity of dNTP insertion opposite –BP for Dpo4 and Dpo4-mutants 54
Table 3.1: DNA Band Ratios Calculated by Five Different Methods 55

Chapter 4

Table 4.1: Relative Velocity of dNTP insertion opposite –BP for Dpo4 and mutant-Dpo4’s, and Dbh 93
Table 4.2: Dominant dNTP insertions opposite various DNA adducts by Y-family polymerases 95
Chapter 5

Table 5.1: Relative Velocity of dNTP insertion opposite –BP for DNAP IV and mutant DNAP IV’s 128

Table 5.2: Analysis of 434 DNAP IV sequences highlighting Coulombic interactions 129
List of Figures

Chapter 1

Figure 1.1 DNA polymerases resemble right hands 18
Figure 1.2 View from minor groove side of three Y-family DNA polymerases 19
Figure 1.3 Structures of BP adducts 20
Figure 1.4 DNAP IV family minor-groove openings 21
Figure 1.5 DNAP V family minor-groove openings 22
Figure 1.6 Active-site Non-Covalent Bridges 23
Figure 1.7 Rotated view of Figure 1.6 25

Chapter 3

Figure 3.1 Structures of +/- BP, and a phosphoimage of dNTP incorporation 58
Figure 3.2 Scans of dGTP Primer Extension reactions 59
Figure 3.3 A PAGE Gel of Purified Proteins 60
Figure 3.4 Scans of Different Primers and Extension Reactions 61
Figure 3.5 Alternative Methods for Quantifying DNA Bands 62

Chapter 4

Figure 4.1 Views of Y-family polymerases Dpo4, DNAP IV, DNAP kappa, and Dbh 98
Figure 4.2 Structures of +/- BP and phosphoimage of dNTP incorporation by Dpo4 and mutant-Dpo4’s

Figure 4.3 The impact on accuracy when amino acids of Y-family DNAP’s are changed

Figure 4.4 A different view of Figure 4.3 Highlighting dCTP and dGTP insertion rates

Figure 4.5 Amino Acids in the active site loop of Dpo4 and DNAP IV

Figure 4.6 The opening on the minor groove side of the active site in Dpo4

Chapter 5

Figure 5.1 Views of Y-family DNAP’s Dpo4, DNAP IV, DNAP kappa, and Dbh

Figure 5.2 The opening on the minor groove side of Dpo4 and DNAP IV

Figure 5.3 Structures of +/- BP and a phosphoimage of nucleotide insertion by DNAP IV and mutants

Figure 5.4 Amino acids in the active site loop of Dpo4 and DNAP IV

Figure 5.5 Pictoral presentation of the impact of amino acid changes on insertion rates of mutant-DNAP IV’s
List of Abbreviations

-BP: the minor stereoisomer of N2-dG Benzo[a]pyrene

+BP: the major stereoisomer of N2-dG Benzo[a]pyrene

A.A.: Amino Acid

AS: Active-Site

B[a]P: Benzo[a]pyrene

CPD: Cyclobutane Pyrimidine Dimer

DNAP: DNA polymerase

E(n)-band: A DNA band in a sequencing gel that represents extension product #n

EcDNAP: E. coli DNA Polymerase

hDNAP: Human DNA Polymerase

I-band: A DNA band in a sequencing gel that represents insertion opposite a lesion

LF-Domain: Little Finger Domain

NC: N-clasp domain

NCB: Non-Covalent Bridge

ScDNAP: Yeast DNA Polymerase

TLS: Trans-Lesion DNA Synthesis

TPF-Domain: Thumb/Palm/Fingers Domain

T^T: Thymine Dimer

UV: Ultra Violet
Chapter 1: Introduction

Y-family DNA Polymerases Fall in Two Categories

Cells possess many DNA polymerases (DNAPs); e.g., humans, yeast (S. cerevisiae) and E. coli have at least fifteen, eight and five, respectively (1-3). These DNAPs serve different purposes. For example, replicative DNAPs are often blocked by the DNA damage caused by chemicals and radiation. To avoid such lethal blockage, cells possess lesion-bypass DNAPs (1,4-17), which conduct translesion DNA synthesis (TLS). Most lesion-bypass DNAPs are in the Y-family (1-17), where human cells have three that are template-directed (hDNAPs η, ι and κ), yeast has one (ScDNAP η), and E. coli has two (EcDNAPs IV and V).

X-ray structures of Y-Family members hDNAP κ, hDNAP ι, yDNAP η, SsDpo4 and SsDbh show a conserved ~350aa core, which includes the polymerase active site (representative references: (18-35)). Figure 1.1 shows a crystal structure and simplistic model of a DNAP in the Y-family compared to the A family. Y-Family members resemble a right-hand with thumb, palm and fingers domains, which form a unit (TPF-Domains). The fingers are “stubby,” resulting in more solvent accessible surface around the template/dNTP-binding pocket (8), which can accommodate the bypass of bulky and/or deforming DNA adducts/lesions that typically protrude into these open spaces. Y-Family DNAPs grip DNA with an additional domain often called the “little finger” (LF-Domain), which is tethered to the TPF-Domains by a long (~10aa) linker (18-20). Figure
1.2A shows a representative Dpo4 structure as viewed from the minor groove side with the LF-Domain in yellow and the TPF-Domains in green. Homology modeled structures also exist for EcDNAP IV and EcUmuC (the polymerase subunit of EcDNAP V), and are shown in Figure 1.2B. Steps in the mechanism of Y-Family DNAPs have been proposed for both protein structural changes (14,23,25,26,32,36) and for chemical catalysis (37).

hDNAP κ was originally discovered because its sequence closely resembles *E. coli* DNAP IV (38-40), and dNTP insertion opposite a variety of adducts/lesions is remarkably similar for the DNAP IV/κ pair suggesting they are functional orthologs (discussed in (41)). These adducts/lesions and inserted bases are described in Table 1. The Archaea DNAP SsDbh behaves similarly (Table 1.1), which suggests it is also a member of the “κ/IV-class.” DNAPs IV and κ have been shown to accurately bypass a variety of N²-dG-adducts (42-49), including from endogenous sources, such as reactive derivatives of cellular trioses (47) or adducts formed from lipid peroxidation processes (49), which may be the main cellular rationale for the genesis of the IV/κ-class (10,42,47).

*E. coli* DNAP V and human DNAP η are also functional orthologs, based on their similarity of dNTP insertion opposite a variety of adducts/lesions (41). The Archaea DNAP SsDpo4 behaves similarly (Table 1.1), which suggests that it is also a member of the “η/V-class.” A case has been made that the main cellular rationale for the η/V-class is TLS of UV-damage (discussed in (41)).
Benzo[a]pyrene is a Potent Mutagen

Benzo[a]pyrene (B[a]P) is a polycyclic aromatic hydrocarbon that forms from incomplete combustion, and is found in smoke, charred foods, and soot. B[a]P metabolism involves enzymatic oxygenation usually to less toxic derivatives, but also to more toxic products such as (+)-anti-7,8-diol-9,10-expoxide-B[a]P (B[a]PDE) (50). B[a]PDE readily reacts with guanine at the N² position and forms the major adduct [+ta]-B[a]P-N²-dG (+BP). Another metabolic route produces [-ta]-B[a]P-N²-dG (-BP) to a lesser extent (51) (Figure 1.3).

While B[a]P adducts are removed by nucleotide excision repair (NER) (52,53), lesions that persist at the time of DNA synthesis pose a clear threat to the fidelity of replication. This is supported by early experiments which showed that B[a]PDE induces several types of mutations in Chinese hamster ovary cells (54). While G/C to T/A transversions (G→T) were the predominant mutation in these studies, other substitutions were observed approximately 6% of the time, and frameshifts were observed 14% of the time. Consistent with the existence of transcription-coupled NER, (55), 93% of all mutations in the above study were in the non-transcribed strand. Similarly, G→T transversions were the predominant mutation induced by B[a]PDE in cultured human fibroblasts (56). Although these studies used racemic mixtures of (+)- and (-)-B[a]PDE, the fact that –BP is known to be a relatively weak carcinogen suggests that the observed mutations were caused primarily by +BP (57).

As is the case in mammalian cells, +BP also induces a wide array of mutations in *E. coli*. Although no mutation occurs greater than 90% of the time in all contexts
described below, the errors that are introduced encompass nearly every category. For example, +BP has been shown to cause -1 frameshifts in a 5’GGG*A sequence where G* is the adducted base (45), as well as in runs of G/C basepairs (58). Alternatively, +BP induced mostly G→T substitutions in a 5’CG*G context (59), and in 5’TG*N sequences (58). In vivo experiments have also shown that +BP induces G→T mutations almost exclusively in a 5’TG*C sequence (60), but leads to G→A mutations in a 5’CG*T context (61), and equal numbers of G→T and G→A mutations in a 5’TG*T context (62).

The fact that mutational specificity is not dependent solely upon the identity of a lesion (e.g. +BP), but also upon its sequence context, has led to the hypothesis that each adduct can adopt several conformations in DNA (63). This suggests that the structural foundation of B[a]PDE-induced mutagenesis must be examined on a sequence-by-sequence basis. In order to simplify the interpretation of experimental results, the study herein will focus primarily on B[a]P adducts in a 5’TG*C sequence where G→T substitutions predominate.

**E. coli Polymerases IV and V Bypass B[a]P Adducts**

The requirements for DNAPs IV and V in the bypass of B[a]P adducts were most clearly shown in experiments with *E. coli* that are deficient in various combinations of Y-family and B-family DNAPs (46). Here, M13 single-stranded phage vectors that have been modified with site-specific –BP or +BP adducts in a 5’TG*C context were transformed into SOS-induced cells. Because single-stranded DNA was used, excision repair was not possible, and progeny phage was therefore a direct measure of relative
TLS proficiencies. It was shown that DNAP IV was both necessary and sufficient for the bypass of –BP. Alternatively, both DNAP IV and V were required for the bypass of +BP, in that cells without DNAP IV or V led to a 99% and 97% reduction progeny phage, respectively.

The fact that both DNAP IV and V are required for the bypass of +BP raises two questions: what is the specific role of each polymerase during synthesis opposite the lesion, and which DNAP inserts dATP in the G→T mutagenic pathway? Although these questions still remain unsolved, studies with purified proteins have alluded to their answers. It has been shown that DNAP V preferentially inserts dATP opposite +BP in a 5'CG*A context, and that DNAP IV inserts dCTP (43). The fact that no mutation occurs 90% of the time, but both Y-family polymerases are required to bypass +BP, suggests an enzyme switching model where DNAP IV performs correct insertion opposite the lesion and DNAP V carries out subsequent extension until DNAP III can resume replication. This hypothesis is not without precedent, as mammalian DNAP ι can insert opposite thymine dimers and B[a]P adducts, but requires other Y-family DNAPs for subsequent extension (64). By this same line of reasoning, it is almost certain that DNAP V does insertion and extension in the G→T pathway.

Y-family DNA Polymerase Structures and Benzo[a]pyrene

The differences in insertion patterns opposite B[a]P lesions by the κ/IV-class and the η/V-class of DNAPs must be due to structural differences (Table 1.1, (42)). Molecular modeling studies with DinB (65) and Crystal structures of Dpo4 (24,66,67) in
a complex with B[a]P adducted-DNA have given us insight as to the structural basis for accurate insertion by the κ/IV-class of DNAP’s, and for inaccurate insertion by the η/V-class. Below, we will discuss three structural features of Y-family DNAPs that dictate the accuracy by which they bypass B[a]P lesions.

**Opening in the Protein Surface on the Minor Groove Side**

For a proper Watson-Crick base pair to form between BP-N\(^2\)-dG and dCTP, the BP moiety must be in the minor groove. On the minor groove side of the active site, many DNAPs in the κ/IV-class have large openings (Figure 1.4), which accommodate the BP-moiety, thus allowing accurate dCTP insertion opposite BP-N\(^2\)-dG.

In contrast, η/V-class DNAPs have small openings (Figure 1.5), which cannot accommodate the BP-moiety without deforming adduct structure. X-ray structures of a +BP-adduct in Dpo4 show evidence for adduct deformation; e.g., in one structure, the minor groove in the vicinity of the active site is so plugged that both the BP-moiety and the adduct-dG moiety are forced into the minor groove (24). In another structure, the BP-moiety is on the major groove side of the active site following an anti-to-syn rotation of the adduct-dG-moiety (24). Near the Dpo4 active site, this opening is plugged by bulky amino acids V32 and M76. In contrast, EcDNAP IV, which is in the κ/IV-class, has a large opening because it has glycines in these equivalent positions.
Non-Covalent Bridges

Figure 1.2A shows a representative X-ray structure of Dpo4. In all of its many X-ray structures, Dpo4 shows a single non-covalent bridge between its TPF-domain and its LF-domain. This bridge forms near the active site, and is deemed the Active Site Non-Covalent Bridge (AS-NCB). As discussed below, other Y-family DNAPs, including DNAP IV, are also likely to have an AS-NCB.

In its only X-ray structure with DNA, Dbh has a single non-covalent bridge on the minor groove side, though it is ~8Å distal to the active site (Distal-NCB). However, this Dbh X-ray structure has a flaw: its duplex DNA region contains one unpaired, bulging base that is positioned such that Dbh is incapable of forming an AS-NCB, as found in Dpo4. Several observations suggest that the AS-NCB forms when Dbh has normal duplex (unbulged) DNA. In molecular dynamic simulations with normal DNA, the AS-NCB in Dbh reforms (Figure 1.2C). In studies involving little finger domain swaps (68), the chimera with the little finger of Dbh attached to Dpo4 (“Dpo4-LF/Dbh”) behaves like Dbh in its insertion properties opposite a variety of lesions, while the opposite chimera (“Dbh-LF/Dpo4”) behaves like Dpo4; these findings suggest that some structural element in the LF-Domain defines insertion specificity.

A comparison of the aa-sequence of the LF-Domains of Dpo4 and Dbh reveals that their greatest difference is in a nine amino acid stretch:


Dbh (positions 238-246): V E N K S K I P H
In the X-ray structure, the dominant contact in the Distal-NCB of Dbh involves a hydrophobic interaction between I244, which is in this region, and V101. Herein, we show strong evidence that these 9aa can play a role in decreasing incorrect dGTP and dATP insertion opposite BP-\(\text{N}^2\)-dG adducts. Molecular modeling

Furthermore, evidence suggests that EcDNAP IV also forms a Distal-NCB. Because no X-ray structure exists for DNAP IV, the Loechler laboratory did homology modeling and molecular dynamics to generate a model. The homology modeling began with a Dpo4 X-ray structure, which only contains a single non-covalent bridge (AS-NCB). In every molecular dynamics trajectory, the final DNAP IV structure had formed a Distal-NCB, while still retaining its original AS-NCB (Figure 1.2B). In the DNAP IV model, this Distal-NCB depends on P99/I100, which correspond to I101/A102 in Dpo4. In particular, the P99 proline causes a kink in the amino acid backbone of DNAP IV that thrusts the backbone-carbonyl on I100 ~3A toward the LF-Domain (compared to the equivalent backbone-carbonyl on A102 in Dpo4), thus allowing R240 in the LF-Domain of DNAP IV to form a Coulombic interaction with the backbone-carbonyl on I100. (The bulky hydrophobic R-group of I100 in DNAP IV also appears to be important.) R240 in DNAP IV is in approximately the same position as I244 in Dbh. The Loechler laboratory has evidence that P99/I100 can play a role in decreasing dGTP and dATP misinsertion opposite BP-\(\text{N}^2\)-dG adducts.

The two simplistic principles emerging from the work described above are the following. (1) The size of the opening on the minor groove side of the active site dictates the rate of dCTP insertion opposite BP-\(\text{N}^2\)-dG adducts. A larger opening allows the BP-
moiety to protrude into the minor groove, which allows the adduct-dG to adopt a conformation to Watson-Crick base pair with dCTP. (2) Non-covalent bridges between the LF-domain and the TPF-domain dictate the rate of misinsertion opposite BP-N²-dG adducts; i.e., an increase in quantity or quality of these NCBs leads to a decrease in rate of dATP/dGTP misinsertion. However, one nuance was noted: the AS-NCB in DNAP IV seems capable of decreasing dGTP/dATP misinsertion opposite BP-N²-dG adducts to a greater extent than the AS-NCB in Dpo4. In this study, we will develop a structural hypothesis to rationalize this difference, and investigate this hypothesis experimentally.

The Dpo4 AS-NCB involves two hydrogen bonds: R36/N254 and K252/carbonyl backbone of R36. Although no X-ray structure of DNAP IV exists, it is hypothesized that an AS-NCB is made via a different pair of hydrogen bonds (detailed below).

The R36 residue of Dpo4 resides in an active site loop (AS-loop), which stretches from V30 to T45 (Figures 1.6 and 1.7). Half of this loop (V30-R36) lies on the minor groove side of the active site and includes V32, which plugs the opening on the minor groove side of the active. The other half of this loop (R36-T45) lies on the major groove side and includes the amino acid that lies above the base in the template (A42) and the base in the dNTP.

The AS-loop in DNAP IV (I31-T43) has several notable differences (Figures 6 and 7). (1) The DNAP IV-AS-loop is approximately shaped like the loop on the end of a paper clip, while the Dpo4-AS-loop is three amino acids longer, which is accommodated by a bulge (E38-S40) which takes the aa-backbone away from the active site, thus providing more open space on the major groove side of the active site. (2) DNAP IV has
three arginines (R35, R37 and R38 at the tip of its AS-loop, whereas Dpo4 only has one (R36).

Rationale and Summary of this Project

In this study, we seek to determine which structural features of Y-family DNAPs determine the rates of mutagenic and non-mutagenic insertion, and therefore dictate overall fidelity, opposite BP-N²-dG. It is important to note that reaction rates are defined by transition state structures, and X-ray crystals contain ground state structures, which can provide important clues about mechanism, but may or may not be directly relevant. In fact, only structure/activity studies can provide insight into transition state structures (69). Herein, we take a structure/activity approach to investigate the transition state structure of BP-N²-dG adducts in several Y-Family DNAPs during dNTP incorporation. In particular, we take SsDpo4, which is in the h/V-class and has low fidelity during BP-adduct bypass, and make protein modifications based on structural elements in EcDNAP IV, SsDbh and hDNAP k that increase the rate of dCTP insertion and/or decrease the rate of dATP, dATP and/or dTTP misinsertion opposite –BP-dG. Accordingly, this structure/activity approach provides insights about key architectural features in these Y-Family DNAPs that dictate lesion insertional patterns in the transition state.

We chose SsDpo4, SsDbh and EcDNAP IV, because they are the simplest of the well-studied Y-family DNAPs, where each contains only core amino acids (352, 354 and 351 amino acids, respectively). Further, SsDpo4, SsDbh and EcDNAP IV share many structural features, making amino acid swaps feasible. Other Y-Family DNAPs are less
amenable to this approach. EcUmuC contains a catalytic core plus an extra domain of unknown function at its C-terminus (422aa); plus it is only active in the presence of regulatory subunits (i.e., UmuD’ and RecA). hDNAP κ contains a catalytic core, but also has an extra 100 amino acids on its N-terminus (“N-clasp”) and an extensive C-terminal domain whose function is incompletely understood (870aa). hDNAP η has a catalytic core, but it also has an extensive C-terminal domain whose function is incompletely understood (713aa). ScDNAP η has numerous amino acid inserts in its catalytic core, making it unique, in addition to having an extensive C-terminal domain whose function is incompletely understood (632aa). hDNAP ι has a catalytic core, but it also has a short extra N-terminal domain and an extensive C-terminal domain whose function is incompletely understood (715aa); plus its cellular function is not well understood.

By employing structure/activity assays, we discovered that no single structural feature accounted for DNAP IV’s accuracy in –BP-dG bypass vs. Dpo4’s inaccuracy. Rather, there were a number of structural features that each improved accuracy by 2-5-fold, but that collectively could account for the observation that EcDNAP IV is ~100-fold more accurate than SsDpo4 in its fidelity of insertion opposite –BP-dG.

EcDNAP IV has four structural elements that improve the fidelity for dNTP incorporation opposite –BP-dG compared to Dpo4 as outlined next.

(1) DNAP IV has a relatively large opening on the minor groove side of its active site that facilitates correct dCTP insertion, because the BP-moiety must be in the minor groove to form Watson-Crick-like adduct-dG:dCTP pairing, given that the N²-position of dG is in the minor groove (Chapter 4).
(2) DNAP IV has several structural features compared to Dpo4 that plug the major groove side of the active site, and, thereby, suppress dNTP misinsertions that occur when an adduct has bulk on the major groove side, such as we showed was the case for dGTP misinsertion with –BP-dG (Chapter 4).

(3) Y-Family DNAPs have three domains (thumb/palm/fingers), which operate as a unit (TPF-Domains), along with another domain, usually called the little finger domain (LF-Domain). The TPF-Domains and the LF-Domain are covalently connected by a long amino acid tether (~15aa). X-ray structures show that the TPF-Domains and the LF-Domain each interact with DNA, which brings them together, such that non-covalent bridges (NCBs) between the TPF-Domains and the LF-Domains can form. We note that several X-ray structures of Y-Family DNAPs without DNA do not show the proper arrangement of the TPF-Domains and the LF-Domain (31,70). We showed that both EcDNAP IV and Dbh have non-covalent bridge on the minor groove side and ~8Å distal from the active site (Distal-NCB), whose presence leads to the suppression of dATP, dGTP and dTTP misinsertions (Chapters 4 and 5). Dpo4 lacks a distal-NCB, which contributes to Dpo4 being less efficient at suppressing dATP, dGTP and dTTP misinsertions opposite –BP-dG.

(4) All three DNAPs have a non-covalent bridge near their active site (AS-NCB) that also suppresses dATP, dGTP and dTTP misinsertions opposite –BP-dG (Chapters 4 and 5). Evidence was presented that the AS-NCB in DNAP IV was superior to the AS-NCB in SsDpo4 at suppressing dATP, dGTP and dTTP misinsertions.
Collectively, this study has revealed the roles of several structural elements of Y-family DNA polymerases that have not previously been studied. It also illustrates the importance of structure/activity assays in determining the mechanisms and possible conformations that are involved in mutagenesis. Lastly, in studying DNA polymerase architecture, we were able to develop a highly sensitive assay for the quantitation of minor 32P-DNA bands in sequencing gels.
Table 1.1: Dominant dNTP insertions opposite various DNA adducts/lesions by *E. coli* DNAPs IV and V, *Sulfolobus* Dpo4 and Dbh, and human DNAPs κ and η.¹

<table>
<thead>
<tr>
<th>Lesion</th>
<th>DNAP V</th>
<th>DNAP η</th>
<th>Dpo4</th>
<th>DNAP IV</th>
<th>DNAP κ</th>
<th>Dbh</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+ta]-BP-N²-dG</td>
<td>A/C</td>
<td>A&gt;G</td>
<td>A&gt;C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AAF-C8-dG</td>
<td>C</td>
<td>C</td>
<td>-</td>
<td>C/T</td>
<td>C/T</td>
<td>-</td>
</tr>
<tr>
<td>AF-C8-dG</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>TT-CPD</td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>T(6-4)T</td>
<td>AG</td>
<td>nG</td>
<td>AG</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>AP site</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>n</td>
<td>A*</td>
<td>n</td>
</tr>
</tbody>
</table>

¹ Dominant dNTP insertion using purified DNAPs, where “n” indicates “no” or low activity, “A*” indicates bypass by an unusual mechanism, and “-” indicates data unavailable.
LEGENDS

Figure 1.1: DNA polymerases resemble right hands, with thumb, palm and finger domains, which are associated (TPF-domain). Y-Family DNAPs have an extra little finger domain (LF-domain) that is lacking in other DNAP Families, which tend to have big thumbs instead.

Figure 1.2: View from the minor groove side of three Y-Family DNAPs: Dpo4 (Panel A), DNAP IV (Panel B) and Dbh (Panel C). The TPF-domain (green) and LF-domain (yellow) are shown, along with the active site loop (blue). Amino acids with van der Waals radii are involved in the formation of non-covalent bridges (NCBs) between the TPF-domain and LF-domain, which are circled in red. All three DNAPs have NCBs near the active site (AS-NCBs), while DNAP IV and Dbh are likely to form a second Distal-NCB, as discussed in the text.

Figure 1.3: Structures. +BP ([+ta]-B[a]P-N²-dG) and -BP ([-ta]-B[a]P-N²-dG).

Figure 1.4: The DNAP IV model showing the large opening on the minor groove side (“chimney”) and the X-ray structure of DNAP κ showing its slot. A BP moiety (red) fits in the chimney of DNAP IV, while it is projected to fit in the slot in DNAP κ. The template (gray) and primer (brown) are also shown.
Figure 1.5 The DNAP V model and Dpo4 structure showing the plugged opening (chimney) on the minor groove side, along with the capped opening in the case of DNAP η. The template (gray) and primer (brown) are also shown.

Figure 1.6: Interactions crucial to the formation of the active site non-covalent bridge (AS-NCB) in Dpo4 and DNAP IV. In each case, the view is from the major groove, and the templating base and dNTP are indicated, where duplex DNA stretches downward. (Panel A) In Dpo4, R36 (purple), which is in the TPF-domain, forms a hydrogen bond (yellow arrow) with N254 (dark blue), which is in the LF-domain. In addition, K252 (turquoise) forms a hydrogen bond (yellow arrow) with the carbonyl-oxygen of R36 (purple). The active site loop (AS-loop) of Dpo4 is also shown, with V30-G35 (yellow) being on the minor groove side and with E38-A44 (brown) being on the major groove side. A42 lies above the templating base in the active site. Dpo4 has a three-amino acid bulge, which is highlighted by the arrow. (Panel B) In DNAP IV, R37 (purple), which is in the TPF-domain, is proposed to form a hydrogen bond (yellow arrow) with D251 (turquoise), which is in the LF-domain. In addition, R38 (red) in the TPF-domain is proposed to form a hydrogen bond (yellow arrow) with D252 (dark blue). The active site loop (AS-loop) of DNAP IV is shown, with I29-E34 (yellow) being on the minor groove side and with G39-S42 (brown) being on the major groove side. V40 lies above the templating base in the active site.
Figure 1.7: The structures in Figure 7 are identical to Figure 6 except the images have been rotated such that the view is from above, where the minor groove is up and the major groove is down.
Figure 1.1: DNA polymerases resemble right hands, with thumb, palm and finger domains, which are associated (TPF-domain). Y-Family DNAPs have an extra little finger domain (LF-domain) that is lacking in other DNAP Families, which tend to have big thumbs instead.
Figure 1.2: View from the minor groove side of three Y-Family DNAPs: Dpo4 (Panel A), DNAP IV (Panel B) and Dbh (Panel C). The TPF-domain (green) and LF-domain (yellow) are shown, along with the active site loop (blue). Amino acids with van der Waals radii are involved in the formation of non-covalent bridges (NCBs) between the TPF-domain and LF-domain, which are circled in red. All three DNAPs have NCBs near the active site (AS-NCBs), while DNAP IV and Dbh are likely to form a second Distal-NCB, as discussed in the text.
Figure 1.3: Structures. +BP ([+ta]-B[a]P-N^2-dG) and -BP ([-ta]-B[a]P-N^2-dG).
Figure 1.4: The DNAP IV model showing the large opening on the minor groove side ("chimney") and the X-ray structure of DNAP κ showing its slot. A BP moiety (red) fits in the chimney of DNAP IV, while it is projected to fit in the slot in DNAP κ. The template (gray) and primer (brown) are also shown.
Figure 1.5 The DNAP V model and Dpo4 structure showing the plugged opening (chimney) on the minor groove side, along with the capped opening in the case of DNAP η. The template (gray) and primer (brown) are also shown.
Figure 1.6: Interactions crucial to the formation of the active site non-covalent bridge (AS-NCB) in Dpo4 and DNAP IV. In each case, the view is from the major groove, and the templating base and dNTP are indicated, where duplex DNA stretches downward.
(Panel A) In Dpo4, R36 (purple), which is in the TPF-domain, forms a hydrogen bond (yellow arrow) with N254 (dark blue), which is in the LF-domain. In addition, K252 (turquoise) forms a hydrogen bond (yellow arrow) with the carbonyl-oxygen of R36 (purple). The active site loop (AS-loop) of Dpo4 is also shown, with V30-G35 (yellow) being on the minor groove side and with E38-A44 (brown) being on the major groove side. A42 lies above the templating base in the active site. Dpo4 has a three-amino acid bulge, which is highlighted by the arrow. (Panel B) In DNAP IV, R37 (purple), which is in the TPF-domain, is proposed to form a hydrogen bond (yellow arrow) with D251 (turquoise), which is in the LF-domain. In addition, R38 (red) in the TPF-domain is proposed to form a hydrogen bond (yellow arrow) with D252 (dark blue). The active site loop (AS-loop) of DNAP IV is shown, with I29-E34 (yellow) being on the minor groove side and with G39-S42 (brown) being on the major groove side. V40 lies above the templating base in the active site.
Figure 1.7: The structures in Figure 7 are identical to Figure 6 except the images have been rotated such that the view is from above, where the minor groove is up and the major groove is down.
Chapter 2:

Materials and Methods

Construction of Cells that Overproduce His-Tagged Dpo4 and DinB:

The wt-Dpo4 gene in a pET22 vector was obtained from Dr. Roger Woodgate and cloned into pET15b for subsequent overexpression of His-tagged proteins. The Dpo4 gene was PCR-amplified with forward and reverse primers (Table 2.1) which had 5’ NdeI sites and 3’ BamHI sites, respectively, in a reaction with 100 ng of each primer (Invitrogen, Carlsbad, CA), 100 ng of DNA template, and 1x concentration of Qiagen PCR Taq Master Mix (QIAGEN Inc., Valencia, CA) in the following thermocycler program: 30 cycles 95°C/1min, 56°C/1min, 72°C/45sec. PCR product was purified using a Qiagen Gel Extraction Kit (QIAGEN). This fragment and pET15b were cleaved with 10 units of NdeI and BamHI (New England Biolabs, Ipswich, MA (NEB)), mixed (120 µg Dpo4 with ~40 ng pET15b), and ligated overnight (16°C, 1mM Tris-HCl, 5mM KCl, 0.1mM DTT, 5% glycerol, pH 7.5) using 100 units of T4 DNA ligase (NEB). The ligated product was transformed into super competent XL1 blue cells (Agilent Technologies Inc., Santa Clara, CA) (procedure described below). Transformed colonies were grown in 10mL LB/Amp (LB/Amp refers to Lauria Broth or Lauria Agar containing 100 µg/ml ampicillin) overnight. The plasmid was isolated (Wizard Plus SV Promega miniprep, Promega, Madison, WI) and stored in DNAse-free water. Candidate plasmids were sequenced (Genewiz, Boston, MA) and the final vector containing Dpo4 was called “Dpo4- pET15b.”
Dbh and Dpo4-lf-Dbh were also obtained from Dr. Roger Woodgate (NIH, Bethesda, MD) and cloned into pET-15b in the same procedure as above, yielding “Dbh-pET15b” and “Dpo4-lf-Dbh-pET15b,” respectively.

DinB-pET16b was obtained from Dr. Robert Fuchs (CRCM, Marseille, France) and is a vector with wt-dinB, which encodes His-tagged DNAP IV. DinB-pET16b was transformed via electroporation (procedure described below) into BL-21DE3 cells (provided by Dr. John Celenze, Boston University, Boston, MA), which encode bacteriophage T7 polymerase that is induced from lacP with IPTG. Dpo4-pET15b also was also transformed via electroporation into BL-21DE3 cells.

The N-Clasp-Dpo4 chimeric protein was created by PCR amplifying the N-clasp from hDNAP κ (cloned gene provided by Drs. Louise and Satya Prakash University of Texas, Galveston, TX) with primers that include 5′- and 3′-NdeI sites (Table 2.1), and cloning it into the 5′ NdeI site of the Dpo4 open reading frame in Dpo4- pET15b. This was carried out in the same manner as above, using Dpo4-pET15b as a host plasmid. Candidate colonies were ultimately screened for insert orientation and sequence (Genewize), and the plasmid was called “NC-Dpo4.”

**DNAP IV Mutagenesis**

Several mutant constructs of Dpo4 and DNAP IV were created using the parent plasmid described above, and the Stratagene QuickChange II kit (Agilent) Appropriate mutant primers (Table 2.2) were used to amplify parent plasmids via PCR with the following protocol: 95°C for 30sec, 55°C for 1min, 68°C for 7min for 18 cycles. Each
reaction contained the following in a total volume of 25 µl: ~40 ng plasmid, 62.5 ng of each primer, 2.5 µL 10x dNTP’s, 2.5U Pfu, 2.5 µL 10x reaction buffer (Agilent) and 16.5 µl distilled water. PCR success was assessed via agarose gel electrophoresis (0.8%, TAE buffer, 4% ethidium bromide), and parent plasmid was digested by adding 1 µl DpnI (Agilent) into the PCR reaction mixture and incubating at 37°C for one hour. The mutant PCR product was heat-shocked into XL1 Blue cells (Agilent) as described below. Colonies were then picked and grown overnight in 10 mL LB/Amp media. The next day, plasmids were purified (Wizard Plus SV Miniprep, Promega) and sequenced (Genewiz).

**Bacterial Transformations**

**Heat Shock Transformation.** To a chilled 15mL tube (Falcon or conical) containing 30-40 µL of XL1 Blue cells (Agilent), 30-40 ng of the plasmid of interest was added and incubated on ice for 30min, followed by heat shock (42°C for 45sec) and incubation on ice for 2min. Pre-warmed (42°C) LB media was added to the cells and the mixture incubated (37°C, 45 min) with shaking (225 rpm). The cell suspension (150 µl) was plated on LB/Amp plates and incubated overnight at 37°C.

**Electroporation Transformation.** Cells competent to transformation via electroporation were prepared as follows. A 10 ml overnight culture of BL-21DE3 cells was added to fresh LB media (1L) and grown at 37°C to OD₆₀₀nm ~0.5, after which cells were twice pelleted (8000 rpm, 10min, 4°C), and then resuspended in 100mL ice-cold distilled water. This pelleting procedure was repeated a second time, followed by a final re-suspension in 500 µL 10% Glycerol. 100 µL aliquots were stored at -80°C.
Approximately 40 ng of the plasmid of interest was added to a thawed aliquot (100 µl), incubated on ice for 30 min, placed in a chilled 1mm electroporation cuvette and electroporated at 1.8kV (BioRad Gene Pulser). 800 µl of LB broth was added immediately afterword and the mixture was incubated at 37°C for 45min, plated (10 µl onto LB/Amp plates), and incubated overnight at 37°C.

**Protein Purification**

**Protein Induction.** BL-21DE3 cells that harbor the plasmid encoding the protein to be purified were grown overnight in 20 ml cultures (37°C, 250 rpm, 10 mL of LB/Amp). After removing 500 µl of the culture to create a permanent stock (500 µl cells and 500 µl 30% glycerol, stored at -80°C) the remaining sample was added to 1.5L LB/Amp, and then grown to OD_{600nm} ~0.5 (37°C, 275rpm). Following induction with 1mM IPTG, cells were incubated for 2hrs at 37°C, pelleted (9000rpm, 8min, 4°C), and stored at -20°C.

**Purification of Proteins.** All steps were carried at at 4°C. Cell pellets were thawed on ice and resuspended in 75mL Lysis Buffer (50mM NaH_{2}PO_{4}, 300mM NaCl, 10mM Imidazole, pH 8), and incubated on ice (30min) in the presence of lysozyme (7.5 mg). Following centrifugation (9000rpm, 45min), the supernatant was transferred into two 50ml conical tubes and the cellular debris was discarded. Proteins were designed with an N-Terminal His-Tag to facilitate protein purification via affinity chromatography. Ni^{2+}-NTA resin (500µl to each tube, ThermoScientific Inc., Waltham, MA) was added to the supernatant, and incubated (4°C, 30 min) with gentle shaking. The suspension was added to a 14in column at room temperature, flow-through (75mL) was discarded and
after washing (100ml of Wash Buffer: 50mM NaH$_2$PO$_4$, 300mM NaCl, 20mM Imidazole, pH 8) protein was eluted (10mL of Elution Buffer: 50mM NaH$_2$PO$_4$, 300mM NaCl, 250mM Imidazole, pH 8) in 1mL fractions. Fractions were stored (-80°C) after addition of glycerol (250µl). Working stocks of purified proteins were created by diluting eluted fractions to 50ng/µl with Elution Buffer containing 10% glycerol. Stocks were stored at -80°C.

**Protein Visualization and Quantification**

**SDS-PAGE.** Presence and relative concentration of purified proteins were analyzed via SDS-PAGE (acrylamide/bis-acrylamide, 4% stacking and 10% resolving). Loading Dye (4 µL 50mM Tris-HCl pH 6.8, 2% SDS, 10% Glycerol, 1% β-Mercaptoethanol, 12.5mM EDTA, and 0.02% Bromophenol Blue) was added to samples (10µl per sample), heated (95°C, 5min), and centrifuged gently before loading into a gel. Electrophoresis proceeded until the dye front migrated into the resolving layer (75V, ~15min), after which electrophoresis was continued until the dye front reached ~1cm above the bottom of the gel (150V, ~1hr).

Following electrophoresis, the resolving gel was rinsed with distilled water, placed in distilled water, and microwaved (1min). After repeating this step the gel was stained (Coomassie Brilliant Blue R Stain Sigma-Aldrich, St. Louis, MO, (0.2%w/v stain in 5:5:1 water: methanol: acetic acid) via microwaving (1min) followed by incubation (10min, room temperature) with gentle shaking. To destain, the gel was rinsed with distilled water and incubated (30min, room temperature) with gentle shaking in destain
solution (40% methanol, 7% acetic acid). Finally, the gel was sandwiched between cellophane pre-soaked in Drying Buffer (30% methanol, 5% glycerol) and allowed to dry between two frames clamped together.

**SDS-PAGE Quantification.** Relative amounts of protein were estimated by comparison to known amounts of a lysozyme standard.

**Bradford Protein Quantification.** Varying amounts (2µl, 5µl, and 10µl) of elution fractions were added to Elution Buffer (200µl total, Elution Buffer: 50mM NaH₂PO₄, 300mM NaCl, 250mM Imidazole, pH 8) to which was added 800µl of the Bradford Reagent (0.01% w/v Coomassie stain in 5% v/v 95% ethanol, 10% v/v 85% phosphoric acid). Following incubation (5min, room temperature) absorbance (595nm) was determined. Protein concentration was estimated by interpolation to a standard curve derived using varying amounts of BSA (1µg, 2µg, 5µg, 10µg and 20µg).

**+BP and –BP Containing Templates for Primer Extension Assays**

**Kinase Reactions.** An un-adducted 56-mer and an adducted 13-mer oligonucleotide (Invitrogen, Table 2.3) were both phosphorylated on the 5’ side using T4 Polynucleotide Kinase (NEB) in a reaction that contained the following in a total of 20µl: 100pmoles 13-mer, 100pmoles 56-mer, 2µl T4 Ligase Buffer (contains ATP used in the phosphorylation, NEB), 8µl distilled nuclease-free water, and 2 units T4 PNK (NEB). This reaction was performed in a thermocycler (37°C/60min, 70°C/20min). The same procedure was used in parallel to construct un-adducted, control DNA templates.
Primer Annealing Reaction. 100pmoles of a 90-mer scaffold and 100pmoles of an unphosphorylated 21-mer primer (Table 2.3) were added to the kinase reaction described above. The 21-mer, 13-mer, and 56-mer primers were annealed to the 90-mer scaffold (Figure 2.1) in a thermocycler program that started at 80°C and decreased by 0.42°C/min down to 4°C.

Primer Ligation Reaction. The 21-mer, 13-mer, and 56-mer were ligated together in the same reaction tube containing the kinase and annealing reactions with the addition of 100 units T4 Ligase Buffer (NEB), 2µl T4 DNA Ligase (NEB), and 8µl distilled water. The reaction mixture was incubated overnight (16°C).

Degradation of the Scaffold and Quality Check of the Adducted Template. The scaffold, which contained uracil bases in place of thymine, was degraded upon treatment with 2 units USER enzyme mixture (uracil DNA glycosylase and AP endonuclease, NEB) and incubation (37°C/60min). Templates were purified using a Nucleotide Removal Kit, following the manufacturer’s protocol (QIAGEN Inc., Valencia, CA), and quantified by comparing absorbance at 260 nm to that of an un-adducted control 90-mer (Invitrogen).

To check the quality of the adducted templates, the 5’ end of the template was P³²-radiolabeled (procedure described below) and loaded on a polyacrylamide (40% urea, 20% acrylamide, TBE buffer) gel along with a 90-mer standard of the same sequence (Invitrogen). Gels were then dried (80°C, 120 min), exposed on a phosphoimaging screen, and scanned on a Typhoon Trio fluorescent scanner.
32P-radiolabeling of Extension Assay Primers

The reaction primer (5’-CGACGGCCAGTGAATTCCTG-3’) was 5’-end 32P-radiolabeled with Polynucleotide Kinase (NEB). Reactions (10µL of 0.5µM primer, 5µL PNK buffer, 2 µL cold ATP (12.5 µM), 8µL γP32-ATP (3000Ci / mMol), 23 µL distilled water, and 2 units PNK (NEB)) were performed at 37ºC (60min) after which PNK was inactivated (65ºC, 20min).

Primer Annealing

The 32P- radiolabeled primer was annealed to the template oligonucleotide starting at 80ºC, after which the temperature was decreased by 1.5 ºC/min to 25 ºC. The mix contained 10µL radiolabeled primer (100nM), 15µL template (100nM), 5µL reaction buffer (500nM NaCl, 100mM Tris-HCl, 10mM EDTA, pH7.5), and 20µL DI water.

Primer Extension Assays

Primer extension assay reaction mixtures contained a total of 9 µL (2nM primer/template oligo, 20mM Tris, 4% glycerol, 8mM MgCl2, 8mM DTT, 100 µg/mL BSA, 50 ng DNAP for -BP reactions and 250 ng for +BP reactions), and were initiated by addition of 1µL dNTP (100 µM final concentration). Reactions were held at 37 ºC for the indicated times, quenched by addition of 10 µL stop buffer (95% formamide, 25mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol), and heated at 95 ºC for 10 minutes to denature the DNA prior to loading onto a sequencing gel.
Sequencing Gel and Analyzing Data

Primer extension reactions were separated in DNA sequencing gels (40% urea, 20% acrylamide, TBE buffer), run at 1200V for 6 hours. Each gel was then transferred to Whatman paper and dried at 80ºC for 90 minutes (BioRad Gel Dryer), placed in a phosphoimaging cassette overnight, and imaged in a Typhoon Trio Fluorescent Scanner (GE Healthcare Life Sciences, Pittsburgh, PA). Because the intensities of the bands assume a skewed Gaussian curve, we developed a novel method to quantify band intensity. This is described in detail in Chapter 3.

Once the relative band intensity has been determined, rate constants can be computed. Specifically, we focused on the extent to which any primer received one insertion event, which is merely the sum of the primers that have received one dNTP insertion (I), plus additional dNTP extensions (i.e., E1, E2, etc.). Total insertion (I_{tot}) is proportional to the sum of the area for these peaks (I + E1 + E2 + E3 + …), while total area for all peaks, including remaining unextended 20-mer primer is “T.”

The reactions in Figure 1 were conducted at 37ºC for 60m with dATP, dCTP and dTTP, while dGTP was for 20m. We were concerned that such long incubation times might lead to a decline in DNAP activity. To investigate protein stability, we preincubated each of these proteins at 37ºC for 0hr, 1hr, 2hr and 4hr, and found no significant decline in activity, which is sensible given that Dpo4 ad Dbh are from Sulfolobus solfataricus, which is a thermophile.
With several proteins, rate as a function of time was shown to be pseudo-first order; i.e., reaction rate decayed exponentially with time as primer/template was used up. Thus, rate constants were computed according to the following formula.

\[ k_{\text{obs}} \text{ (min}^{-1} \text{)} = -\ln\left[1 - \left(\frac{T - I_{\text{tot}}}{0.84}\right)\right]/\text{time(min)} \]

The value 0.84 is included because we determined that primer extension does not proceed to 100% completion at long times (i.e., > 10-half-lives), in that a small amount of primer (~16%) could not be extended. Reactions times at 37°C were chosen such that the extent of reaction was generally in the range of 10%-to-50%, though in some cases the extent of reaction was less (e.g., dATP, dGTP and dTTP with Dbh).
Table 2.1: Primers used for cloning DNAP reading frames into pET-15b expression vector

<table>
<thead>
<tr>
<th>Protein Reading Frame</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpo4</td>
<td>Dpo4_Ndel_For</td>
<td>5’-GATCGACATATGATTGTCTTTTCGTTGATTTTGACTAC-3’</td>
</tr>
<tr>
<td>Dpo4</td>
<td>Dpo4_BamHI_Rev</td>
<td>5’-TCGATCGGATCCTATTAAAGTGATCGAAGAAACTTGCTAATCCTAGTGC-3’</td>
</tr>
<tr>
<td>Dbh</td>
<td>Dbh_Ndel_For</td>
<td>5’-GATCGACATATGATAGTGATATTCGTTGATTTTGATTATTTC-3’</td>
</tr>
<tr>
<td>Dbh</td>
<td>Dbh_BamHI_Rev</td>
<td>5’-TCGATCGGATCCTATTAAATGTCGAAGAAATCAGATAATTGTC-3’</td>
</tr>
<tr>
<td>Dpo4-LF_Dbh</td>
<td>LF_Dbh_Ndel_For</td>
<td>5’-GATCGACATATGATGTTCTTTTCGTTGATTTTGACTAC-3’</td>
</tr>
<tr>
<td>Dpo4-LF-Dbh</td>
<td>LF_Dbh_BamHI_Rev</td>
<td>5’-TCGATCGGATCCTATTAAATGTCGAAGAAATCAGATAATTGTC-3’</td>
</tr>
<tr>
<td>N-Clasp</td>
<td>N-Clasp_Ndel_For</td>
<td>5’-GATCGACATATGGATAGCACAAAAGGAGAAGGTGTGACA GTTAC-3’</td>
</tr>
<tr>
<td>N-Clasp</td>
<td>N-Clasp_Ndel_Rev</td>
<td>5’-TCGATCCATATGGCTCAAATTTTCGCTTTTGTCTAATTCCATTGC-3’</td>
</tr>
</tbody>
</table>
Table 2.2: Primers used for Site-Directed Mutagenesis. The highlighted sequences represent the substitutions that will be introduced during the PCR.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mutation</th>
<th>Primers Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNAPIV(DinB)</td>
<td>R35A</td>
<td>5’-CTATTGGCGGCAGGCCGAACGTCGGGGGGTGATCA GCAC-3’</td>
</tr>
<tr>
<td>DNAPIV(DinB)</td>
<td>R37A</td>
<td>5’-CTATTGGCGGCAGGCCGAACGTCGGGGGGTGATCA GCAC-3’</td>
</tr>
<tr>
<td>DNAPIV(DinB)</td>
<td>R38A</td>
<td>5’-CTATTGGCGGCAGGCCGAACGTCGGGGGGTGATCA GCAC-3’</td>
</tr>
<tr>
<td>DNAPIV(DinB)</td>
<td>R37E</td>
<td>5’-CTATTGGCGGCAGGCCGAACGTCGGGGGGTGATCAG CAC-3’</td>
</tr>
<tr>
<td>DNAPIV(DinB)</td>
<td>E251K</td>
<td>5’-GCGGTGAAACGACAGTGGGCAAGATATTCATCAC TGGTC TGAATG-3’</td>
</tr>
<tr>
<td>Dpo4</td>
<td>V32G</td>
<td>5’-CCAGTTGTTTGTGGATTTCAAGGGAG-3’</td>
</tr>
<tr>
<td>Dpo4</td>
<td>M76G</td>
<td>5’-CCTAATGCGTATTACCTACCGGGATAGAAAGGTA TATCAG CAAG-3’</td>
</tr>
<tr>
<td>Dpo4</td>
<td>IV-Loop</td>
<td>5’-CCGTTCTTGAAGAAAACCAGTTGTTGTTGGTTGGA AGCCG CGAACGTCGGGGGGTGATCGCTACTGCAAACCTATG AAGCTAG AAAATTTGG-3’</td>
</tr>
<tr>
<td>[IV-Loop]-Dpo4</td>
<td>R35A</td>
<td>5’-GTTGGAAAGCGCCGAACGTCGGGGGGTGATCGCTACTGC-3’</td>
</tr>
<tr>
<td>[IV-Loop]-Dpo4</td>
<td>R37A</td>
<td>5'-GGTGGAAAGCCGCGAAAGCTCGGGGGGTGATCGCTAC</td>
</tr>
<tr>
<td>[IV-Loop]-Dpo4</td>
<td>R38A</td>
<td>5'-GGTGGAAAGCCGCGAACGTTGGGGTGATCGCTACT-3'</td>
</tr>
<tr>
<td>Dpo4</td>
<td>238-246/Dbh</td>
<td>5'-CTCTCTAGCTAGAGACGAGTATAACGACCTGTAGAATAAAATCAAAAAATCCCGCATGGGAGAATTGTAACGATGAAGAGAAATAGCAGG-3'</td>
</tr>
<tr>
<td>Dpo4</td>
<td>D3</td>
<td>5'-CCAGTTGTTGTTTGTATTTTCAGGGAGAAGCGG--TGCTG TGGCTACTGCAAACTATGAAGC-3'</td>
</tr>
<tr>
<td>Dpo4</td>
<td>UmuC-Loop</td>
<td>5'-CCGTCTTTGAAAGGAAAACCAGTTGTTGTCTATCGAATAATGACGGTTGCGTGCTACTGCAAACTATGAAGC-3'</td>
</tr>
<tr>
<td>Dpo4</td>
<td>Kappa-Loop</td>
<td>5'-CCGTCTTTGAAAGGAAAACCAGTTGTTGTGGTCAATGTCAATGTTAGCTACTGCAAACTATGAAGC-3'</td>
</tr>
</tbody>
</table>
Table 2.3: Oligonucleotides used for the construction of DNA templates

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-mer with +/- BP</td>
<td>5’-GAAGACCTG*CAGG-3’</td>
</tr>
<tr>
<td>21-mer</td>
<td>5’-AGCTATGACCATGATTCAGTG-3’</td>
</tr>
<tr>
<td>56-mer</td>
<td>5’-AATTCACTGCGCGTCTTTTACACGCGTACTGGG AA AACCCTGGCGTTACCCA-3’</td>
</tr>
<tr>
<td>Scaffold</td>
<td>5’-GUAAAACGACGGCCAGUGAAUCCUCGAGUGCUUC CAC UGAAUCAUGGUCAUAGCU-3’</td>
</tr>
</tbody>
</table>

1The underlined G is the base which is adducted with +/- BP.
CHAPTER 3
Accurately Quantitating Intensity of Minor $^{32}$P-DNA Bands in Gels

SUMMARY

In many types of studies, quantitation of relative $^{32}$P-band intensity of DNA fragments of different sizes in DNA sequencing-type gels is of interest, such as in our work on the relative rate of primer extension on DNA templates by DNA polymerases (DNAPs). Following (e.g.) phosphorimaging, multiple $^{32}$P-bands in a single lane often appear by eye to be well-resolved, simple Gaussian curves. However, when lanes are scanned, the resulting plot of intensity versus gel-position reveals that band intensity is actually skewed-Gaussian, with a sharp leading edge and a shallower trailing edge, and, furthermore, neighboring bands are overlapping and not well-resolved. Herein, we describe a method to accurately quantitate relative intensity of such adjacent $^{32}$P-bands. The method relies on scanning a lane (e.g., using ImageJ), which contains a single pure $^{32}$P-band (i.e., $^{32}$P-primer only) that becomes the standard for the shape of a single-component skewed-Gaussian curve for that particular experiment. This single-component curve/scan is introduced into a file (created using Adobe Illustrator) that contains an experimentally determined curve/scan with two (or more) overlapping skewed-Gaussian curves, which were generated by (e.g.) primer-extension by a DNAP on template-DNA. The single-component scan is superimposed on one of the components in the experimentally determined scan, and the intensity in excess of the single-component curve/scan is attributed to the other component(s) in the experimentally determined curve/scan. Relative area (intensity) of the components is determined via
pixel analysis (using Adobe Photoshop). In summary, a method is described to quantitate relative $^{32}\text{P}$-band intensity of DNA fragments of different sizes in a DNA sequencing-type gel; relative intensity of minor $^{32}\text{P}$-bands can be accurately assessed at the <5\% level.

**INTRODUCTION**

We have been investigating how lesion-bypass DNA polymerases (DNAPs) insert and extend dNTPs opposite DNA damage (46,71). We have focused on DNAPs in the Y-Family (1-17), where (e.g.) human cells have three that are template-directed (hDNAPs h, i and k), while yeast has one (ScDNAPs h) and *E. coli* has two (EcDNAPs IV and V). Structurally Y-Family DNAPs have a conserved ~350aa core, which resembles a right-hand with thumb, palm and fingers domains that form a unit (TPF-Domains). Y-Family DNAPs grip DNA with an additional domain (18-20), often called the “little finger” (LF-Domain), which is tethered to the TPF-Domains by a long (~15aa) linker.

We principally study lesions derived from benzo[a]pyrene (B[a]P), which is a well-studied DNA damaging agent that is a potent mutagen and carcinogen, and an example of a polycyclic aromatic hydrocarbon (PAH), a class of ubiquitous environmental substances produced by incomplete combustion (50). PAHs in general and B[a]P in particular induce the kinds of mutations that are thought to be relevant to carcinogenesis and may be important in human cancer (51, and references therein). The
major B[a]P DNA adduct forms at N²-dG (+BP-dG) and is shown in Figure 3.1, which also shows its mirror image adduct (–BP-dG).

In our recent experiments, purified DNAPs are being used in primer extension reactions, in which a single dNTP is added to a 20-mer ³²P-primer annealed to a template containing a BP-adduct. Reactions are stopped, separated on a DNA sequencing-type gel, and visualized by phosphorimaging. Figure 3.1 shows typical results in an experiment involving dATP, dCTP, dGTP or dTTP incorporation opposite -BP-dG by several DNAPs, where the same concentrations of [dNTP] and [DNAP] were used. The DNAPs are: Sulfolobus solfataricus Dpo4 and Sulfolobus solfataricus Dbh, along with two Dpo4 derivatives: one with the little finger domain of Dbh fused to the thumb/palm/fingers-domains of Dpo4 (Dbh/LF-Dpo4), and a second with nine amino acids from Dbh substituted into the equivalent positions in Dpo4 (aa238-246/Dbh-Dpo4).

\[
\text{Dbh:} \quad V_{238} E N K S K I P_{246} \\
\text{Dpo4:} \quad I_{237} R T R V R K S I_{245}
\]

The amount of primer extension for correct dCTP insertion opposite –BP-dG is approximately the same, while dATP, dGTP and dTTP misinsertions decreased in the order: Dpo4 > aa238-246/Dbh-Dpo4 > Dbh/LF-Dpo4 > Dbh (Figure 3.1). An analysis of the mechanistic implications of these findings will be published elsewhere. Herein, we describe the method we developed to quantitatively assess the extent of primer extension in experiments such as shown in Figure 3.1.

In cases where multiple ³²P-bands are present in a single lane in Figure 3.1, the bands have the appearance of being reasonably well-resolved and simple Guassian...
curves. However, when the individual lanes for (e.g.) dGTP in Figure 3.1 were scanned (using ImageJ), the tracings (Figure 3.2, in black) show that, in fact, neighboring bands are not well-resolved and, furthermore, their shape is actually skewed-Gaussian, with a sharp leading edge and a shallower trailing edge. Primer alone is shown as the red tracing in each lane in Figure 3.2 and it is also skewed-Gaussian. To determine relative rates of dNTP incorporation, the fraction of 20-mer primer converted to the first insertion product opposite the –BP-dG adduct (I), as well as subsequent extensions (E1, E2 and E3), must be determined. Because the scanned bands are overlapping and skewed-Gaussians, quantitating relative intensity is not straightforward.

We searched the literature and found no method that, to us, seemed adequate to quantitate relative intensity of overlapping skewed-Gaussian curves, especially when band intensity for insertion (I) is relatively low, for example, as is the case for both Dbh and Dbh/LF-Dpo4 (Figure 3.2). Herein, we describe a method, which is both accurate and easy.
RESULTS AND DISCUSSION

The gene for each DNAP was expressed using a vector that added an N-terminal His-tag to facilitate protein purification (Materials and Methods). The DNAPs are reasonably pure (Figure 3.3). Though it is hard to detect by eye, the Dbh/LF-Dpo4 and Dbh proteins contained a low level (~1% based on staining intensity following scanning) of a high molecular weight protein (~200kD). Dpo4 and aa238-246/Dbh-Dpo4 contained no detectable contaminant. In Figure 3.2, we attempted to load equal amounts of protein, and band intensity is reasonably consistent (aa238-246/Dbh-Dpo4/100%, Dbh/98%, Dbh/LF-Dpo4/97% and Dpo4/80%). Primer extension reactions were performed using a 20-mer $^{32}$P-primer annealed to a template containing a –BP-dG. DNAPs were preincubated with the primer/template at 37°C for >5 minutes, and polymerization initiated by dNTP addition. All experiments used the same concentration of DNAP (50nM) and dNTP (100mM), which is relevant because [dNTP] in cells is in the 10-100mM range. Reactions were stopped, separated on a DNA sequencing gel and subjected to phosphorimaging; e.g., as shown for the dGTP reaction in Figure 3.1.

The phosphorimages for all four dGTP reactions were scanned using ImageJ and saved in a single TIF-file. (Likewise the dATP, dCTP and dTTP reactions were scanned and each set of four was saved in its own TIF-file.) The 20-mer primer was alone scanned again (using ImageJ) and saved as a second TIF-file. Both the TIF-file with the four dGTP reactions and the TIF-file with the 20-mer primer alone were imported into Adobe Illustrator and both images were turned into an “Inked Drawing” (using the “Live Trace” function), which transforms the scans into lines on an invisible, see-through
background. This file (with two TIF-images) was saved as a Scalable Vector Graphics image (SVG-file), and then reopened in Adobe Illustrator. The scanned-lines were made as thin as possible (0.25mm). The 20-mer primer alone was made red and was superimposed on each of the other four scans to give images that appeared like in Figure 3.2. During superimposition, only the amplitude of the 20-mer primer was varied; by this, we mean that the height of the 20-mer primer alone was varied, but not its width (variance), for reasons discussed below.

Because of the spacing between peaks and the sharp leading edge of these skewed-Gaussian curves, the height of the insertion band (I, which is a 21-mer in Figure 2) is virtually zero at the position of the primer band, which meant that the peak height of the 20-mer primer alone (red curve) could be stretched all the way to the top of the experimental curve for the 20-mer (black curve) during superimposition.

In the next phase of our analysis, the excess intensity associated with the non-20-mer peaks (i.e., the peaks attributable to I, E1, E2, etc. in Figure 2) was estimated. In each scan, a baseline was drawn. Two files were exported (as TIF-files). The first TIF-file looked like Figure 3.2 and was used to estimate the excess relative area due to the sum of the peaks attributable to I, E1, E2, etc. (as explained below), which appear as peaks in the case of Dpo4 and aa238-246-Dpo4, but appear as bulges in the case of Dbh/LF-Dpo4 and Dbh. In the second TIF-file, the 20-mer primer (red line) was removed, and the scans in this second TIF-file were used to estimate the total area for all of the peaks. Each TIF-file was opened in Adobe Photoshop, and the relevant areas were
quantitated as total pixels (using the “Magic Wand” function), from which the fraction of 20-mer primer converted to longer products was computed (see below).

Whereas it is apparent in the phosphorimages in Figure 3.1 that dGTP misinsertion has occurred with Dpo4 and aa238-246/Dbh-Dpo4, no primer extension is obvious in the lane for Dbh, while the band for Dbh/LF-Dpo4 is weak. However, excess intensity in the region expected for a single insertion (21-mer) is apparent in Figure 3.2 as a bulge above the 20-mer primer alone red lines for both Dbh/LF-Dpo4 and Dbh. But are these bulges really attributable to primer extension? Duplicate samples were run in the same gel, and similarly shaped bulges were observed once again for Dbh/LF-Dpo4 and Dbh (Figure 3.4). Two duplicate 20-mer primers were also run in the same gel and scanned independently (Figure 3.4); they superimpose closely on the original 20-mer primer (red line), which suggests that bulges for Dbh/LF-Dpo4 and Dbh are not due to vagaries in the scanning. Furthermore, the relative area (intensity) for the bulges in Figures 3.2 and 3.3 agree closely: (1) For Dbh/LF-Dpo4, the extended primer was 4.4% in Figure 3.2 and 5.0% in Figure 3.4, and (2) for Dbh, the extended primer was 1.7% in Figure 3.2 and 1.8% in Figure 3.4. For all of these reasons, we attribute the bulges in the lanes for Dbh/LF-Dpo4 and Dbh in Figures 3.2 and 3.3 to primer extension by the DNAPs.

In Figure 3.2, dGTP misinsertion (I) opposite –BP-dG can be detected for all four DNAPs (Dpo4, aa238-246-Dpo4, Dbh/LF-Dpo4 and Dbh), but significant extension beyond the incorporation of a single dGTP was also apparent, most notably with Dpo4 and aa238-246-Dpo4-Dpo4. For our purposes, we were interested only in kinetics
relevant to the rate of insertion of the first dNTP, so we simply assessed the sum of the relative areas for all peaks that included the first insertion event, which included \([I + E1 + E2 + \ldots]\). Using this basic approach, an observed rate constant \(k_{obs}\) was computed (Table 3.1) for correct dCTP insertion, and for dATP, dGTP and dTTP misinsertion for Dpo4, KIP-Dpo4, Dbh/LF-Dpo4 and Dbh (as detailed in Materials and Methods).

The values for relative velocity for all of the duplicate samples are given in Table 3.1. Among the sixteen combinations of DNAPs/dNTPs, the largest discrepancy for a duplicate was 28\% (i.e., \(0.68 \times 10^{-3}\) min\(^{-1}\) versus \(0.51 \times 10^{-3}\) min\(^{-1}\) for dTTP with Dbh/LF-Dpo4). The average discrepancy was 13\%. This finding suggests that the method describe herein gives consistent results.

Deconvoluting the kinetics of the insertion step (I) versus the extensions steps (E1, E2, etc.) requires a more sophisticated approach, which we did not pursue, though we have shown it to be feasible. Using curves for pure 20-mer, 21-mer, 22-mer, etc, the positioning and amplitude of each pure scan can be adjusted, such that the sum of each pure component at any X-value adds up to the observed X-value in the experimental curve of a scan. The accuracy of the curve amplitude assignments can be evaluated by making sure that the area for each individual component curve (i.e., for the 20-mer + 21-mer + 22-mer, etc.), when added together, agrees with the observed total area.

We note that frequently quantitation of bands in gels is done via other methods, often involving the drawing of boxes around bands using ImageJ and then assigning intensity in the boxed region to the major band in that region. The scans in Figures 3.2 and 3.3 suggest that such an approach may be problematic, but how inaccurate is it? We
surveyed a number of colleagues who described various methods, which we tried to reproduce in Figure 3.5, using the data for dGTP misinsertion with aa238-246/Dbh-Dpo4 as an example. Using the method described herein, the fraction of extended primer was 0.11 (= \( \frac{[I + E1 + E2 + \ldots]}{\text{total}} \)).

(Alternative Method 1): The first alternative approach has each band tightly boxed (Panel A, Lane 1), and the corresponding scan is shown in Panel B/Lane 1 with the blue lines showing what the boxed regions correspond to in the scan. Alternative Method 1 gives a fractional value of 0.23 (= \( \frac{a}{[a + b]} \)). It is worth noting that our eyes deceive us, as it looks like the majority of (e.g.) the primer band intensity is in the b-box, while by scanning/quantitation only 48% of the primer band is actually in the b-box. Alternative Method 1 is also flawed, because values are not corrected for background.

(Alternative Method 2): The second approach (Panel A/Lane 2) shows one method to correct for background, where a relatively less intense region between the two peaks (“x-box”) is used to subtract “background” from the relative areas associated with both the a-box and the b-box. Alternative Method 2 gives a fractional value of 0.03 (= \( \frac{a-x}{[a + b – 2x]} \)). Alternative Method 1 and 2 are flawed because they ignore too much overall band intensity in the diffuse trailing edges and because too little intensity is ascribed to the primer band.

(Alternative Method 3): The third approach (Panel A/Lane 3) shows one method to try to correct for these flaws, where larger boxes are drawn. Indeed Panel B/Lane 3 shows that more of the trailing edge is included. Alternative Method 3 gives a fractional
value of 0.25 (= \([a'/a' + b']\)). Approach 3 is flawed in part because background is not corrected.

(Alternative Method 4): A fourth approach adds a background correction step to Approach 3, where areas in the primer only band are used to correct for background (Panel A/Lane P). In particular the insertion/extension band (a’-box) is corrected by subtracting the corresponding regions from the primer-only band (z-box), while remaining primer (b’-box) is corrected by subtracting the y-box region in Lane 4. Approach 4 gives a fractional value of 0.15 (= \([a'-z]/[a' - z + b' - y]\)). Approach 4 comes closest to giving the correct value. However, it should be obvious that drawing boxes around bands is inherently problematic because it cannot solve the real dilemma, which is accurately quantitating overlapping skewed-Gaussian curves.

We compared these four alternatives methods to our approach in another way. In a preliminary (and flawed) experiment, we purposely spiked the 20-mer primer with a 21-mer (\(^{32}\)P-radiolabeled), such that the latter was either present at 3.4% or 23.3%. Thereafter, we used our method and the four alternative methods to estimate the percentage of the 21-mer. Table 3.2 shows that our method over-estimated the lower percentage (5.6% vs. 3.4%), but is quite close in the base of the larger percentage (26.5% vs. 23.3%). Alternative Methods 1-3 are inferior, where (e.g.) Alternative Method 2 gives a negative number for the lower percentage (-20.4% vs. 3.4%). Alternative Method 4 is the best alternative, though it is not as accurate as our method.

It is also important to note that, in order for Method 4 to be accurate, all of the lanes must run an equal distance in the gel and have the same amount of radiolabeled
DNA. If the lanes are not perfectly flat, then an incorrect portion of the Primer will be subtracted from the background of any given lane and the results will be flawed. Similarly, if the Primer Lane is under- or over-loaded due to human error, the results will be skewed towards over- or under-estimating the extension proportions throughout the entire gel, respectively. These common variables do not affect the validity of our method because the Primer scan can be moved to the left or right in Adobe Illustrator, which accounts for uneven lanes, and can be adjusted for height, which accounts for loading errors. Prior to submitting this work for publication, this experiment must be repeated and expanded.

There were a variety of wrinkles that we discovered while doing this work.

(1) Since all scans in a gel were ultimately analyzed in comparison to a single 20-mer primer (i.e., the red scans in Figure 3.2), we always included at least three 20-mer primer lanes in a gel, and then chose the most representative (best) scan as the standard for the rest of the analysis. Usually, these three primers were virtually indistinguishable (Figure 3.4).

(2) To ensure that the $^{32}$P-20-mer primer standard was as matched as possible to experimental samples, it was also mixed with a DNAP (usually Dpo4), along with all other reagents except no dNTP, and the sample was kept on ice. Practically, there was not a significant difference between standards that contained or did not contain DNAP.

(3) For the sake of consistency, all scans done in ImageJ used the same scanning width (0.38cm) and length (10.0cm). In our DNA sequencing gels, a comb with 64 lanes was used, where each lane was ~0.4cm wide. In our phosphoimages, bands appeared to
be \( \sim 0.42 \text{cm} \) wide, so our choice of scan width (0.38cm) gave us the intensity of the centermost \( \sim 90\% \) of the lane. We did not use a full-width scan (i.e., 0.42cm), because small amounts of band intensity in adjacent lanes diffuse sideways during electrophoresis; using a 0.38 band-width virtually eliminates the impact of sideways-diffusion.

(4) When scanning lanes to be included in a single ImageJ output TIF-file, we always grouped lanes according to what was most relevant for comparative purposes. dATP, dCTP, dGTP and dTTP reactions were run for each DNAP (Figure 3.1); however, we did not group these four lanes in an output TIF-file from ImageJ for the following reason. Because ultimately we are interested in comparing how (e.g.) the rate of dGTP insertion varies for different DNAPs, we virtually always grouped together all of the dGTP scanned lanes in a single SVG-file using Adobe Illustrator; e.g., as shown in Figure 3.2.

(5) Multiple scans for multiple lanes were usually included in a single output TIF-file from ImageJ, though never more than eight scans. For the sake of consistency, TIF-file scans imported into Adobe Illustrator were always sized identically: The four centermost scanned lanes were sized to fit in the large box found in the center of the Adobe Illustrator rectangular window.

(6) In some cases, the width of a band in a tracing for a scanned lane (i.e., the variance) was not identical to the width of the 20-mer primer alone; e.g., the width of the 20-mer primer is slightly broader than the width for the dGTP reaction with Dbh/LF-Dpo4 in Figure 3.4. (In some cases, the discrepancy was as great as \( \sim 10\% \).) For the sake
of consistency, we never vary the width of the 20-mer primer differentially in one lane compared to all other lanes, unless there is a compelling reason to do so, which is rare. In cases where the widths do not match, we superimpose the left side (trailing edge) of 20-mer primer with the line for remaining primer. In cases where we tested the effect of varying the width of the 20-mer-primer, the computed relative area is not affected dramatically (<~20%).

(7) In some cases when a peak shape was relatively complex and the area relatively large (e.g., for Dpo4 in Figure 2), the “Magic Wand” function in Adobe Photoshop sometimes gave a value for total pixels that were unrealistically low. To circumvent this problem, we sub-divide large peak-areas using vertical lines (typically three) and then summed the area for each of the four sub-regions.

(8) When a scan is converted to an “Inked Drawing” (using the “Live Trace” function in Adobe Illustrator), there are often small gaps in the line tracing, which were filled in with lines.

(9) Even though the 20-mer primer was purified, there is a small amount of what looks to be a 19-mer contaminant (estimated to be ~2-3% in all experiments). If this 19-mer were extendable, it would be lower in intensity in cases where more reaction is observed (e.g., Dpo4 versus Dbh), which is not the case. We ignore the small area of the 19-mer by drawing a line on the leading edge of the 20-mer portion of the curve, as shown in the tracings in Figures 3.3 and 3.4.

(10) In each experiment, we attempted to load the same amount of primer in each lane, and, following analysis, the total number of pixels per lane was often very similar
from lane-to-lane, with the lowest and highest values typically varying by <1.5-fold. We purposely loaded different amounts of the same reaction in different lanes to test if this had an effect on the computed extent of reaction. A typical dCTP/Dpo4 reaction was conducted, and 1.0mL, 1.5mL and 2.0mL of sample were loaded in different lanes in a sequencing gel. Following analysis, the fraction of extended primer was computed to be: 0.16, 0.16 and 0.14, respectively, in one set of three loadings, and 0.16, 0.15 and 0.16, respectively, in a second set of three loadings. Thus, the computed fractional extent of reaction is not significantly affected by having different amounts of total radioactivity in a lane.

In collaboration with Dr. Daniel Segre, we explored a variety of digital computational approaches to analyze our data. Using MATLAB, four different mathematical functions, which each describe a skewed-Gaussian curve, were evaluated. Three of the skewed-Gaussian functions failed to give a good fit to the particular shape of the skewed-Gaussian curves associated with our experimental data. The fourth function fit well, but the method involved an iterative approach to compute an integral, which was too time intensive to be practical. Furthermore, during the work on the approach described herein, we discovered that considerable judgment is required to accurately do curve-fitting, which made an automated digital computational approach less tenable.

In conclusion, the method described in this chapter made it possible to accurately quantitate the amount of primer extension in reactions with a BP-containing template, a dNTP and a DNAP (e.g., Figure 3.1), especially in cases where the extent of extension was relatively low, such as with dGTP misinsertion with Dbh/LF-Dpo4 and Dbh.
Table 3.1: Relative Velocity of dNTP insertion opposite –BP-dG for Dpo4 and mutant-Dpo4s, as well as Dbh.\(^1\)

<table>
<thead>
<tr>
<th>Protein</th>
<th>dATP</th>
<th>dCTP</th>
<th>dGTP</th>
<th>dTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpo4</td>
<td>10.0(^1)</td>
<td>6.3</td>
<td>48.2</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>8.6</td>
<td>5.0</td>
<td>53.5</td>
<td>3.8</td>
</tr>
<tr>
<td>aa238-246/Dbh-Dpo4</td>
<td>4.1 (0.40)</td>
<td>5.2 (0.82)</td>
<td>7.0 (0.15)</td>
<td>5.0 (1.02)</td>
</tr>
<tr>
<td></td>
<td>3.6 (0.41)</td>
<td>4.8 (0.94)</td>
<td>6.9 (0.13)</td>
<td>4.3 (1.12)</td>
</tr>
<tr>
<td>LF-Dbh</td>
<td>2.0 (0.20)</td>
<td>2.2 (0.35)</td>
<td>2.6 (0.056)</td>
<td>0.68 (0.14)</td>
</tr>
<tr>
<td></td>
<td>2.5 (0.28)</td>
<td>2.4 (0.49)</td>
<td>3.1 (0.058)</td>
<td>0.51 (0.13)</td>
</tr>
<tr>
<td>Dbh(^2)</td>
<td>0.78 (0.078)</td>
<td>2.5 (0.39)</td>
<td>1.0 (0.021)</td>
<td>0.44 (0.09)</td>
</tr>
<tr>
<td></td>
<td>0.72 (0.084)</td>
<td>2.4 (0.48)</td>
<td>1.1 (0.020)</td>
<td>0.35 (0.09)</td>
</tr>
</tbody>
</table>

\(^1\) Rates are for dNTP insertion opposite a –BP-dG adduct in a primer/template using methods described in Chapter 3. Purified proteins included Dpo4, along with two mutant-Dpo4s, as well as Dbh. Reactions were pseudo-first order (Materials and Methods) and (e.g.) the number 10.0 for Dpo4 with dATP means the velocity was 10.0 x 10\(^{-3}\) min\(^{-1}\) using standard conditions, which were dNTP concentration [100µM] and protein concentration 50ng/mL in a 60 minute reaction.

\(^2\) Numbers in parenthesis are the velocity (non-parenthesis on the left) divided by the velocity of the corresponding Dpo4 reaction; e.g., for aa238-246/Dbh-Dpo4 dATP is 0.40 = (4.1/10.0).
Table 3.2: DNA Band Ratios as Calculated by 5 Different Methods\(^1\).

<table>
<thead>
<tr>
<th>Actual Percentage(^2)</th>
<th>Our Method</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Method 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4%</td>
<td>5.6%</td>
<td>18.1%</td>
<td>-20.4%</td>
<td>18.7%</td>
<td>8.2%</td>
</tr>
<tr>
<td>23.3%</td>
<td>26.5%</td>
<td>38.5%</td>
<td>25.0%</td>
<td>38.2%</td>
<td>32.3%</td>
</tr>
</tbody>
</table>

\(^1\) Radiolabeled 20-mer DNA Oligos were spiked with known amounts of Radiolabeled 21-mer DNA Oligos, and were resolved on a Sequencing Gel. The band ratios were then quantified using five different methods, as described in Chapter 3.

\(^2\) Actual Percentage was calculated by taking the known amount of the 21-mer Oligo, and dividing it by the sum of the 21-mer and 20-mer that was loaded in that lane.
LEGENDS

Figure 3.1: Structures of +BP-dG and –BP-dG, along with the primer/template sequence and a phorphorimage showing dATP, dCTP, dGTP and dTTP incorporation by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4), which are explained in the text.

Figure 3.2: Scans of the dGTP primer extension reactions shown in Figure 1 by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4), as explained in the text. Primer only is shown as the red scan/line in each of the four panels.

Figure 3.3: SDS-PAGE of the purified DNAPs: DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4). The position of a trace contaminant is also indicated (see text). Either 10ng (Lanes 1-4) or 40ng (Lanes 5-8) were added.

Figure 3.4: Duplicate runs were included in the same gel as shown in Figure 1 and scans for the dGTP primer extension reactions for LF/Dbh-Dpo4 and Dbh are shown. Multiple primer runs were also included in Figure 1, and they are also shown. Details are explained in the text.
Figure 3.5: Phosphorimage (Panel A) and scans (Panel B) of the dGTP primer extension reaction with aa238-246/Dbh-Dpo4 (Lanes 1-3), using the same data as shown in the seventh lane of Figure 1, along with primer only (Lane “P”). The boxes in each of the four lanes in Panel A show a region (labeled: a, b, a’, b’, x, y and z) that was quantitated as described in the text. The corresponding boxed regions are shown in the scans in Panel B, as indicated by the blue lines.
Figure 3.1: Structures of +BP-dG and –BP-dG, along with the primer/template sequence and a phophorimage showing dATP, dCTP, dGTP and dTTP incorporation by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4), which are explained in the text.
Figure 3.2: Scans of the dGTP primer extension reactions shown in Figure 1 by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4), as explained in the text. Primer only is shown as the red scan/line in each of the four panels.
Figure 3.3: SDS-PAGE of the purified DNAPs: DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and Dbh/LF-Dpo4). The position of a trace contaminant is also indicated (see text). Either 10ng (Lanes 1-4) or 40ng (Lanes 5-8) were added.
Figure 3.4: Duplicate runs were included in the same gel as shown in Figure 1 and scans for the dGTP primer extension reactions for LF/Dbh-Dpo4 and Dbh are shown. Multiple primer runs were also included in Figure 1, and they are also shown. Details are explained in the text.
Figure 3.5: Phosphorimage (Panel A) and scans (Panel B) of the dGTP primer extension reaction with aa238-246/Dbh-Dpo4 (Lanes 1-3), using the same data as shown in the seventh lane of Figure 1, along with primer only (Lane “P”). The boxes in each of the four lanes in Panel A show a region (labeled: a, b, a’, b’, x, y and z) that was quantitated as described in the text. The corresponding boxed regions are shown in the scans in Panel B, as indicated by the blue lines.
CHAPTER 4  
Y-Family DNA Polymerase Architecture  
That Controls Accurate dCTP Insertion  
Opposite an $N^2$-dG Adduct of Benzo[a]pyrene

SUMMARY

To synthesize past DNA damage, cells have lesion bypass DNA polymerases (DNAPs), often in the Y-Family. One class of Y-Family DNAPs includes hDNAP $\kappa$ and EcDNAP IV, which accurately insert dCTP opposite $N^2$-dG adducts, including from the environmental carcinogen benzo[a]pyrene (BP-$N^2$-dG). Another class includes hDNAP $\eta$ and EcDNAP V, which insert accurately opposite UV-damage, but inaccurately opposite BP-$N^2$-dG. Herein we make protein modifications in SsDpo4, which is a canonical $\eta/V$-class member, to make it more like $\kappa/IV$-class members. We identify three structural elements that contribute to insertion fidelity opposite a BP-$N^2$-dG adduct (-BP-dG). (1) To form a Watson-Crick-type base pair between dCTP and BP-$N^2$-dG, the BP moiety must be on the minor groove side of the active site. We show that correct dCTP insertion opposite –BP-dG is facilitated by having a large opening in the protein surface on the minor groove side. (2) Non-covalent bridges (NCBs) allow the little finger domain (LF-Domain), which is a structural feature found uniquely in DNAPs in the Y-Family, to interact with the thumb/palm/fingers domains (TPF-Domains), which contains the active site. Y-Family DNAPs can have a NCB in the minor groove near the active site (AS-NCB) and in the minor groove distal to the active site (Distal-NCB). We show that dATP/dTTP misinsertion is suppressed by both the quantity and quality of NCBs.
between the TPF-domains and the LF-domain. Findings suggest that the BP-moiety is also in the minor groove during dATP and dTTP misinsertion, though the BP-dG conformation is different than during correct dCTP insertion. (3) In contrast, dGTP misinsertion opposite –BP-dG by Dpo4 is affected by the size of the opening on the major groove side of the active site, which suggests that the bulk of the BP-N2-dG adduct is on the major groove side, implying syn-adduct-dG:dGTP Hoogsteen pairing occurs during misinsertion. In summary, three structural elements contribute to the fidelity of Y-Family DNAP insertion opposite a BP-N2-dG adduct: (1) A large opening on the minor groove side of the active site facilitates correct dCTP insertion, because the BP-moiety is in the minor groove; (2) The quality and quantity of non-covalent bridges between the TPF-domains and the LF-domain suppresses dATP and dTTP misinsertion; and (3) Protein features that that plug the major groove side of the active site suppresses dGTP misinsertion, which is also suppressed by NCBs. Furthermore, this work suggests what adduct conformation is found during mutagenic misinsertion: the BP-moiety is in the major groove during dGTP misinsertion, while it is in the minor groove during dATP and dTTP misinsertion, though in different conformations than the minor groove conformation used during correct dCTP insertion.

**INTRODUCTION**

Cells possess many DNA polymerases (DNAPs); e.g., humans, yeast (S. cerevisiae) and *E. coli* have at least fifteen, eight and five, respectively (1-3) (72). These DNAPs serve many functions; e.g., replicative DNAPs are often blocked by the DNA
damage caused by chemicals and radiation, and to avoid such lethal blockage, cells possess lesion-bypass DNAPs (1-17), which conduct translesion DNA synthesis (TLS). Most lesion-bypass DNAPs are in the Y-family (1-17), where human cells have three that are template-directed (hDNAPs η, ι and κ), while yeast has one (ScDNAPs η) and E. coli has two (EcDNAPs IV and V).

X-ray structures of Y-Family members hDNAP κ,, hDNAP η, hDNAP ι, SsDpo4, SsDbh and EcDNAP IV show a conserved ~350aa core, which includes the polymerase active site (representative references: (18-35), (73-76)). As with all DNA polymerases, Y-Family members resemble a right-hand with thumb, palm and fingers domains, which form a unit (TPF-Domains). The fingers are “stubby,” resulting in more solvent accessible surface around the template/dNTP-binding pocket (8), which can accommodate the bypass of bulky and/or deforming DNA adducts/lesions that typically protrude into these open spaces. Y-Family DNAPs grip DNA with an additional domain (18-20), often called the “little finger” (LF-Domain), which is tethered to the TPF-Domains by a long (~15aa) linker. The best studied Y-Family DNAP is Dpo4 from Sulfolobus solfataricus, and a representative Dpo4 X-ray structure is shown in Figure 4.1A, which is a view from the minor groove side with the LF-Domain in yellow and the TPF-Domains in green. Figure 4.1 also shows representative X-ray structures for EcDNAP IV (Figure 4.1B, (73)) and hDNAP κ (Figure 1C, (31,76)). A modeled structure for SsDbh is also shown (Figure 4.1D), which we developed based on an X-ray structure (76) for reasons described below. Steps in the mechanism of Y-Family DNAPs
have been proposed for both protein structural changes (15, 22, 24, 25, 31, 36) and for chemical catalysis (37).

Our work has focused on benzo[a]pyrene (B[a]P), which is a well-studied DNA damaging agent that is a potent mutagen and carcinogen, and an example of a polycyclic aromatic hydrocarbon (PAH), a class of ubiquitous environmental substances produced by incomplete combustion (50). PAHs in general and B[a]P in particular induce the kinds of mutations that are thought to be relevant to carcinogenesis and may be important in human cancer (51, and references therein). The major B[a]P DNA adduct (+BP) forms at N²-dG and is shown in Figure 4.2, which also shows its mirror image –BP.

Based on the insertion preference opposite a variety of lesions, Y-Family DNAPs fall into different classes. The IV/κ-class insert correct dCTP opposite a variety of N²-dG adducts, including adducts from environmental carcinogen (e.g. BP-dG) (77-80) as well as from endogenous sources, such as reactive derivatives of cellular trioses (47) or adducts formed from lipid peroxidation processes (49), which may be the main cellular rationale for the genesis of the IV/κ-class. Y-Family DNAPs in the IV/κ-class also correctly insert opposite a variety of other lesions, including PhIP-C8-G, TG, 8oxoG and some interstrand crosslinks (81-84). DNAPs in the IV/κ-class do not insert opposite UV-damage and tend to be relatively accurate in the replication of undamaged DNA. The κ/IV-class includes hDNAP κ, EcDNAP IV and SsDbh. The V/η-class tends to be relatively inaccurate in the replication of undamaged DNA and inaccurate in insertion opposite N²-dG adducts, but accurate with UV-damage, which may be the main cellular
rationale for the genesis of the V/η-class. EcDNAP V, hDNAP η, and SsDpo4 are in the V/η-class.

The differences in insertional preferences by DNAPs in the κ/IV-class versus the η/V-class must reflect an interface between protein structural differences and adduct structural/conformational differences. Regarding BP-adducts, five X-ray structures have been observed with Dpo4 (67,85). Of these, only one is an insertion structure with the adduct-dG interacting with a dNTP, while the other four are extension structures where the base on the 5’-side of the BP-adduct is in the active site. (1) In the insertion structure (2IA6/A-peptide), the dG-moiety of +BP-dG is in the syn-conformation and is paired via two hydrogen bonds to dATP, which is also in the syn-conformation. The +BP-pyrene moiety is intercalated below the syn-dG:syn-dATP base pair and the adduct bond is on the major groove side of the active site. Overall the structure appears distorted, where the adduct-syn-dG is not perpendicular to the helix axis and is not well stacked. (2) The B-peptide in the 2IA6 crystal shows an extension structure with both the +BP-moiety and the dG-moiety in the minor groove. In the active site, the template-dC:dGTP appears more-or-less like a normal Watson-Crick base pair. (3) Another crystal (2IBK) shows a similar structure, except the dGTP had been incorporated into the primer. (4 and 5) The 1SOM crystal has a different BP-adduct (i.e., a –BP-dG adduct with a cis-adduct bond) and shows two extension structures with the adduct-dG adopting the syn-orientation, such that the pyrene-moiety is on the major groove side of the active site. The A-peptide has a well-stacked intercalated BP-moiety, while the B-peptide has the BP-moiety in the major groove.
These five X-ray structures show that a BP-pyrene moiety in Dpo4 can be intercalated or in the major groove or the minor groove. In addition, the adduct-dG can be either in the anti-orientation or the syn-orientation, in which case the BP-moiety is in the minor or major groove, respectively. Furthermore, the adduct-dG-moiety can be either paired with a base in the opposite strand or extra-helical. Finally, only one of the five structures had the dG-moiety in the insertion position, while the other four had extension structures. Given the complexity in these X-ray crystals, it is hard to draw conclusions about what structural elements might relate either to correct or to incorrect dNTP incorporation.

Furthermore, it is important to note that reaction rates are defined by transition state structures, and X-ray crystals contain ground state structures, which can provide important clues about mechanism, but may or may not be directly relevant. In fact, only structure/activity studies can provide insight into transition state structures (69). Herein, we take a structure/activity approach to investigate the transition state structure of BP-dG adducts in several Y-Family DNAPs during dNTP incorporation. In particular, we take SsDpo4, which is in the η/V-class and has low fidelity during BP-adduct bypass, and make protein modifications based on structural elements in EcDNAP IV, SsDbh and hDNAP k that increase the rate of dCTP insertion and/or decrease the rate of dATP, dATP and/or dTTP misinsertion opposite –BP-dG. Accordingly, this structure/activity approach provides insights about key architectural features in these Y-Family DNAPs that dictate lesion insertional patterns in the transition state.
We chose SsDpo4, SsDbh and EcDNAP IV, because they are the simplest of the well-studied Y-family DNAPs, where each contains only core amino acids (352, 354 and 351 amino acids, respectively). Further, SsDpo4, SsDbh and EcDNAP IV share many structural features, making amino acid swaps feasible. Other Y-Family DNAPs are less amenable to this approach. EcUmuC contains a catalytic core plus an extra domain of unknown function at its C-terminus (422aa); plus it is only active in the presence of regulatory subunits (i.e., UmuD’ and RecA). hDNAP κ contains a catalytic core, but also has an extra 100 amino acids on its N-terminus (“N-clasp”) and an extensive C-terminal domain whose function is incompletely understood (870aa). hDNAP η has a catalytic core, but it also has an extensive C-terminal domain whose function is incompletely understood (713aa). ScDNAP η has numerous amino acid inserts in its catalytic core, making it unique, in addition to having an extensive C-terminal domain whose function is incompletely understood (632aa). hDNAp τ has a catalytic core, but it also has a short extra N-terminal domain and an extensive C-terminal domain whose function is incompletely understood (715aa); plus its cellular function is not well understood.

Based on molecular modeling studies, we have proposed that a large opening on the minor groove side of the active site, as found in κ/IV-class Y-Family DNAPs, should facilitate correct dCTP insertion opposite BP-adducts, given that the BP-moiety must be in the minor groove in order to form more-or-less normal adduct-dG: dCTP Watson-Crick base pairs (65). Herein, we provide evidence that a large opening on the minor groove side of the active site does indeed facilitate dCTP insertion opposite –BP-dG.
We also describe findings that speak to long-standing questions concerning adduct conformation during mutagenic misinsertion. For example, a role for Hoogstein base pairing, which requires anti-to-syn base rotation, has often been considered as a possible mechanism for adduct:dNTP misinsertion. In fact, evidence suggests that DNAP ι promotes Hoogsteen base pairing because of its cramped active site, which preferentially accommodates a shorter \([N9:N1]\) distance found in [syn-purines;pyrimidine] base pairs (75). However, in the case of adduct bypass, the major dG-adduct of 1-nitropyrene forms a Watson-Crick-type base pair with dCTP in DNAP ι though this is a ground state structure (74,75). In other Y-Family DNAPs, which do not have a cramped active site like DNAP ι, structure/activity studies have not been undertaken to investigate a possible role of Hoogsteen base pairing during bypass.

Herein, we show that the –BP-dG adduct has bulk in the major groove during dGTP misinsertion, which is most consistent with syn-adduct-dG:dGTP suggesting Hoogsteen pairing. Studies have shown that the Y-Family DNAPs in the κ/IV-class are less efficient in the bypass of adducts with bulk in the major groove (86), and herein using the dGTP misinsertion pathway opposite –BP-dG, we identify some key structural elements that κ/IV-class DNAPs use to suppress misinsertion when adducts have bulk in the major groove.

We also show that –BP-dG adduct bulk is in the minor groove during dATP and dTTP misinsertion, though evidence suggests that the –BP-dG adduct conformation during dATP/dTTP misinsertion must be different than the minor groove conformation associated with correct dCTP insertion. Finally, we establish that non-covalent bridges
(NCBs) formed between the TPF-domains and the LF-domain help to suppress the rate of dATP and dTTP, as well as dGTP misinsertion opposite –BP-dG.
RESULTS

The *dpo4*, *dinB* (for DNAP IV) and *dbh* genes were incorporated into vectors that included a N-terminal His-Tag to facilitate protein purification (Materials and Methods). Primer extension reactions were performed using a 20-mer $^{32}$P-primer annealed to a template containing a –BP-dG adduct, as described previously (Materials and Methods). DNAPs were preincubated with the primer/template at 37°C for >5 minutes, and polymerization initiated by dNTP addition. Reactions were stopped, separated on a DNA sequencing gel, and--following phosphorimaging (Figure 2)--individual reactions were scanned and quantitated (Chapter 3). Polymerization often proceeded beyond the first insertion; e.g., Dpo4 inserts dATP opposite –BP (I), and then extends by adding dATP opposite the 5’-T (E1) and the next C-base (E2). For kinetic analysis, the intensity of all extension products was summed (I + E1 + E2) and compared to remaining primer to determine kinetics. All experiments used the same concentration of dNTP (100mM) and DNAP (50nM) unless noted otherwise (Chapter 3), and Table 1 lists observed rate constants ($k_{obs}$) in units of min$^{-1}$ for a variety of wild type and mutant DNAPs. These data are biologically relevant given that [dNTP] concentrations in cells are in the 10-100mM range. dNTP concentrations of 100mM are below Km, which is in the ~1mM range, and, accordingly, kinetic data are most relevant to Vmax/Km considerations. Determining Km accurately was not feasible, because dNTP insertion rate becomes progressively inhibited at higher [dNTP].

It is hard to evaluate the large amount of data in Table 4.1 as a whole. We found the easiest way to interpret trends requires normalization of rate by dividing [mutant-
Dpo4/wt-Dpo4] for a particular dNTP. For example, the rate of dGTP misinsertion was 
0.00178 min$^{-1}$ for A42V-Dpo4 and was 0.0635 min$^{-1}$ for wt-Dpo4; thus, normalized 
dGTP insertion for A42V-Dpo4 was ~0.28 (= 0.00178/0.0681). This ratio shows that the 
A42V modification decreased dGTP misinsertion ~3.8-fold compared to wt-Dpo4. Such 
normalizations are meaningful, because the reactions used the same concentrations of 
dNTP (100mM) and DNAP (50nM).

Frequently in the literature a different normalization is reported, namely 
“fidelity,” which is the ratio of [misincorporation/correct incorporation]. For example, 
dGTP fidelity for A42V-Dpo4 is ~3.1 (= 0.00178/0.00058) and for wt-Dpo4 is 10.9 (= .00635/.00058). However, trying to understand why the A42V-mutation improves 
fidelity ~3.5-fold is difficult, because fidelity comparisons reflects how a protein 
structural change influences both dCTP and dGTP incorporations, where the structural 
difference might affect dCTP versus dGTP differently. In contrast, the [mutant-Dpo4/wt-
Dpo4] ratio for (e.g.) the dGTP reaction involves trying to interpret how a particular 
structural change affects only one dNTP insertion mechanism at a time. Ultimately, we 
do reflect on fidelity, as it is the ultimate measure of the [advantage/disadvantage] of a 
structural difference.

dATP and dTTP Misinsertions Involve Similar –BP-dG Adduct Conformations

Figure 4.3A shows normalized dCTP misinsertions versus normalized dATP 
misinsertion as a Log-Log plot for a variety of mutant-Dpo4s using a –BP-dG template. 
Given that the rate of all mutant-Dpo4 proteins are normalized to the rate of wt-Dpo4, the
data point for wt-Dpo4 itself by definition is located at the origin (red point at 0, 0). Data for DNAP IV (Table 4.1), along with several of its derivatives (open circles in Figures 4.3 and 4.4), and Dbh are also in the plot (open squares).

The two large arrows in Figure 4.3A are a reminder that fidelity is improved for a mutant-Dpo4 if the mutation either increases the rate of the dCTP reaction (motion upward) or decreases the rate of the dATP reaction (motion leftward). Figure 4.3A shows considerable scatter, which implies that protein mutational changes do not affect correct dCTP insertion in the same way as incorrect dATP misinsertion. In fact, the data in Figure 4.3A show why we abandoned the use of fidelity normalizations to try to understand how protein structural changes affect activity/mechanism.

Figure 4.3B shows normalized dTTP misinsertions versus normalized dATP misinsertion for a variety of mutant-Dpo4s (as a Log-Log plot). The majority of protein modifications decreased both normalized dATP misinsertion and normalized dTTP misinsertion, both of which contribute to improved fidelity. Unlike the data in Figure 4.3A, the points in Figure 4.3B tend to cluster around a line with a slope of 1.0, which implies that protein modifications tend to affect dATP and dTTP misinsertions approximately equally and suggest that dATP and dTTP misinsertions occur from the same or a similar adduct conformation, as discussed further below.

Figure 4.3C shows normalized dCTP misinsertions versus normalized dTTP misinsertion as a Log-Log plot for a variety of mutant-Dpo4s. As expected from the data in Figures 4.3A and 4.3B, no simple relationship exists.
Figure 4.4A shows normalized dCTP insertion plotted versus normalized dGTP insertion, and, once again, no simple relationship exists, which suggests that the BP-adduct is in a different conformation during dCTP insertion versus dGTP misinsertion.

**The –BP-dG adduct has bulk in the Major Groove during dGTP Misinsertion**

hDNAP κ has an extra 100aa at its N-terminus, which forms a unique U-shaped structure called the N-clasp (shown in scarlet in the structure in Figure 4.1C) that closes off the opening on the major groove side of the active site. In hDNAP κ, the N-clasp has been shown to inhibit bypass of adducts with bulk on the major groove side of the active site (86,87), and molecular modeling studies have shown why (88). When the N-clasp was added to Dpo4, dGTP misinsertion decreased from $\sim 64 \times 10^{-3} \text{ min}^{-1}$ to $13.7 \times 10^{-3} \text{ min}^{-1}$ (~4.7-fold). This decrease suggests that -BP-adduct bulk is in the major groove during dGTP misinsertion. This effect is also seen in the plots in Figures 4.4A and 4.4B, where the latter is provided to show a subset of the data in order to more easily describe certain aspects of the data. The effect of adding the N-clasp to Dpo4 is unique to the dGTP reaction, as the N-clasp had little effect on dCTP insertion ($\sim 5.8 \times 10^{-3} \text{ min}^{-1}$ vs. $\sim 6.0 \times 10^{-3} \text{ min}^{-1}$), dATP misinsertion ($\sim 10.9 \times 10^{-3} \text{ min}^{-1}$ vs. $\sim 10.6 \times 10^{-3} \text{ min}^{-1}$) or dTTP misinsertion ($\sim 4.3 \times 10^{-3} \text{ min}^{-1}$ vs. $\sim 4.9 \times 10^{-3} \text{ min}^{-1}$), suggesting that -BP-adduct bulk is not in the major groove in these three cases. This can also be seen in the plots in Figures 4.3 and 4.4.

Y-Family DNAPs have a loop above their active site (AS-loop), which is shown from the major groove side for Dpo4 and DNAP IV in Figure 4.5. This loop begins on
the minor groove side (yellow amino acids in Figure 4.5) and loops back to the major
groove side (brown amino acids in Figure 4.5). The AS-loop contains a number of key
structural features, such as the amino acid that sits above the templating base (A42 in
Dpo4 and V40 in DNAP IV) and the amino acid that sits above the base in the dNTP
(A44 in Dpo4 and S42 in DNAP IV).

V40 in DNAP IV is aligned with A42 in Dpo4, and the R-group at this position
protrudes into the major groove just above the templating base in the active site (Figure
4.5). The rate of dGTP misinsertion for mutant A42V-Dpo4 (~18 x 10^{-3} \text{ min}^{-1}) was ~3.6-
fold lower than for wt-Dpo4 itself (~64 x 10^{-3} \text{ min}^{-1}), which also shows that increasing
protein bulk on the major groove side of the active site suppresses the dGTP reaction,
which is consistent with the –BP-moiety being in the major groove. In contrast, the
A42V-Dpo4 modification had little effect on dCTP insertion (~5.8 x 10^{-3} \text{ min}^{-1} vs. ~5.8 x
10^{-3} \text{ min}^{-1}), dATP misinsertion (~12 x 10^{-3} \text{ min}^{-1} vs. ~11 x 10^{-3} \text{ min}^{-1}), or dTTP
misinsertion (~4.6min^{-1} vs. ~4.9min^{-1}), suggesting that the –BP moiety is not in the major
groove in these cases.

The AS-loop in DNAP IV (I31-I41) has three fewer amino acids than in Dpo4
(V30-V43). The three extra amino acids are accommodated as a small bulge in the active
site loop of Dpo4 (brown arrow in Figure 4.5A), which moves the protein backbone away
from the active site, thus reducing bulk on the major groove side of Dpo4. The three
amino acids in this bulge were removed (F37, E38 and D39) to give D3-Dpo4, whose
dGTP misinsertion rate (~11 x 10^{-3} \text{ min}^{-1}) is ~5.6-fold lower than wt-Dpo4 (~64 x 10^{-3}
\text{ min}^{-1}), which also suggests that the –BP adduct is in the major groove. The D3-Dpo4
modification had little effect on dCTP insertion (\(-6.9 \times 10^{-3} \text{ min}^{-1} \) vs. \(-5.8 \times 10^{-3} \text{ min}^{-1}\)) and dTTP misinsertion (~4.6 min\(^{-1}\) vs. ~4.9 min\(^{-1}\)), and a modest effect on dATP misinsertion (~6.7 x 10\(^{-3}\) min\(^{-1}\) vs. ~11 x 10\(^{-3}\) min\(^{-1}\)), suggesting that the –BP moiety is not in the major groove in these cases.

Replacing the entire Dpo4 AS-loop (V30-V43) with the smaller AS-loop from UmuC (V29-V37), which is the polymerase subunit of EcDNAP V, also decreased dGTP misinsertion (~3.2-fold), suggesting that the –BP-adduct is in the major groove. Adding the UmuC AS-loop to Dpo4 had little effect on dCTP insertion, dTTP misinsertion or dATP misinsertion, suggesting that the –BP-adduct is not in the major groove in these cases.

Replacing the entire Dpo4 AS-loop with the AS-loop from DNAP IV (I31-I41) decreased dGTP misinsertion (~5.6-fold), suggesting that the –BP-adduct is in the major groove. Adding the DNAP IV AS-loop to Dpo4 has a complicated impact on dCTP, dTTP and dATP insertion, which is addressed below.

In conclusion, a variety of data suggest that the –BP-adduct has bulk that projects into the major groove during dGTP misinsertion, which implies that an \textit{anti}-to-\textit{syn} base rotation of the dG-moiety has occurred and, thus, dGTP misinsertion involves some kind of \textit{syn}-adduct-dG:dGTP Hoogstein base pair. In contrast, the majority of findings suggest that the –BP-adduct is not in the major groove during incorporation of dATP, dCTP or dTTP.
dCTP, dATP and dTTP Incorporation Rates Depend on the Size of an Opening on the Minor Groove Side of the Active Site

For a proper Watson-Crick base pair to form between a BP-N2-dG adduct and dCTP, the BP-moiety must be in the minor groove, since the N2-atom is in the minor groove in a Watson-Crick base pair. Based on our molecular models, we proposed that EcDNAP IV has a large opening on the minor groove side of the active site, which could accommodate the BP-moiety (Figure 4.1E) and would seemingly be able to facilitate accurate dCTP insertion opposite BP-N2-dG (65). Recent EcDNAP IV X-ray structures with undamaged DNA confirm the existence of a large minor groove opening (73). The DNAP IV opening is large in part, because of two glycines (G32 and G73 in Figure 6), which lack an R-group to plug the opening. In contrast, Dpo4’s opening is plugged by V32 and M76 in the equivalent positions. Compared to wt-Dpo4, the double mutant V32G/M76G-Dpo4 increases incorporation rate with dATP (~2-fold), dCTP (~2.2-fold) and dTTP (~2.3-fold) (Table 4.1, Figure 4.4C). These findings suggest that the –BP-moiety is in the minor groove during dATP, dCTP and dTTP incorporation. In contrast, V32G/M76G-Dpo4 leads to a very slight decrease in dGTP misinsertion (0.9-fold), suggesting that the –BP-adduct is not in the minor groove during dGTP insertion.

For wt-DNAP IV, the rate of dATP misinsertion (0.29 x10^{-3} \text{ min}^{-1}) is ~30-fold lower than for wt-Dpo4 (10.6 x10^{-3} \text{ min}^{-1}), and the rates for dTTP misinsertion are similarly lower (0.13 x10^{-3} \text{ min}^{-1} vs. 4.9 x10^{-3} \text{ min}^{-1}). This shows that DNAP IV has structural elements that counteract the enhanced rate of dATP/dTTP misinsertion brought about by V32G/M76G in DNAP IV (addressed in the next subsection). The dATP and
dTTP misinsertion rates with V32G/M76G-Dpo4 are ~2-fold greater than with wt-Dpo4, while dGTP misinsertion is slightly lower (~0.9-fold). This same trend is observed for wt-DNAP IV vs. wt-Dpo4, where dATP/dTTP misinsertion suppression (~0.03-fold) is less than dGTP misinsertion suppression (~0.01-fold).

The dCTP insertion rate was enhanced ~4.4-fold for wt-DNAP IV (~26 x 10^{-3} min^{-1}) vs. wt-Dpo4 (~5.8 x 10^{-3} min^{-1}), which is about twice as great as the ~2-fold enhancement for V32G/M76G-Dpo4 vs. wt-Dpo4 observed with dCTP, as well as dATP and dTTP incorporations. This difference suggests that structural elements other than V32G/M76G are important and prompted us to insert the entire AS-loop of DNAP IV (I31-I41) into Dpo4 (V30-V43). Contrary to our naïve expectation, this change did not further boost correct dCTP insertion when comparing IVloop-Dpo4 with wt-Dpo4 (8.4 x 10^{-3} min^{-1} vs. 4.9 x 10^{-3} min^{-1}). In work to be described at greater length elsewhere, we have shown that the presence of multiple arginines in the AS-loop of DNAP IV (i.e., R35, R37 and R38) allow the IV-loop to function properly in the DNAP IV context, but not in the Dpo4 context, which normally has only one arginine (R36) in its AS-loop. For example, when an arginine was removed with the R37A mutation to give R37A-IV-loop-Dpo4, dCTP insertion rate (21 x 10^{-3} min^{-1}) was much closer to the rate for wt-DNAP IV (26 x 10^{-3} min^{-1}). dATP misinsertion is also increased with R37A-IVloop-Dpo4 (14 x 10^{-3} min^{-1}) vs. IVloop-Dpo4 (5.0 x 10^{-3} min^{-1}). Similarly dTTP misinsertion is increased with R37A-IVloop-Dpo4 (11 x 10^{-3} min^{-1}) vs. IV-loop-Dpo4 3.3 x 10^{-3} min^{-1}). Thus, the addition of the IV-loop has complex effects on Dpo4 activity, although the fact that the rate of dCTP insertion was similar for Dpo4 (5.8 x 10^{-3} min^{-1}), D3-Dpo4 (6.9 x 10^{-3} min^{-1})
and Vloop-Dpo4 (6.0 x 10^{-3} \text{ min}^{-1}) shows that smaller loop size is not a contributing factor.

The AS-loop in hDNAP κ is exceedingly small (G131-L136), and G131 adopts \(\phi/\theta\) -angles unique to glycines (89), and both of these features ensure that the opening on the minor groove side of the active site in hDNAP κ is large. In fact, hDNAP κ has a large open slot in its minor groove. When κloop was added to Dpo4 to give κloop-Dpo4 the rate of dCTP insertion (14 x 10^{-3} \text{ min}^{-1}) was ~2.5-fold greater than with wt-Dpo4 (5.8 x 10^{-3} \text{ min}^{-1}). κloop-Dpo4 also shows a higher dTTP misinsertion rate (12 x 10^{-3} \text{ min}^{-1}) than Dpo4 (4.9 x 10^{-3} \text{ min}^{-1}). These results also suggest that –BP-dG has bulk in the minor groove during dCTP and dTTP incorporation. In contrast, dGTP misinsertion is not significantly affected for κloop-Dpo4 (58 x 10^{-3} \text{ min}^{-1}) versus wt-Dpo4 (64 x 10^{-3} \text{ min}^{-1}), suggesting that –BP-dG does not have bulk in the minor groove during dGTP incorporation. The one finding that is not consistent with simple expectations is that dATP misinsertion did not increase for κloop-Dpo4 versus wt-Dpo4 (value based on a single experiment).

In summary, a variety of experiments show that increasing the size of the opening on the minor groove of the active site increased the rate of dCTP, dATP and dTTP incorporation, which suggests that –BP-dG has bulk on the minor groove side during these incorporation events.

dATP, dGTP and dTTP Misinsertion Rates Depend on a Non-Covalent Bridge Between the LF-Domain and the TPF-Domains
In *Sulfolobus solfataricus*, Dpo4 is a η/V-class member, while Dbh is a κ/IV-class member, based on insertion patterns opposite a variety of lesions, including cyclopyrimidine dimers, [6-4] photoproducts and AP sites (68). Our work with –BP-dG supports this classification, as (e.g.) dGTP fidelity is much better for Dbh than Dpo4, given that dGTP misinsertion is ~50-fold slower for Dbh than Dpo4, while dCTP insertion is only ~2.5-fold slower. Several studies have shown that a chimera with the Dbh LF-domain joined to the Dpo4 TPF-domains (Dbh/LF-Dpo4) behaved more like Dbh than like Dpo4 in many respects, including insertion opposite a CPD and an AP-site (68,76). In our work, dGTP misinsertion was ~21-fold lower for the chimera Dbh/LF-Dpo4 (2.9 min⁻¹) than for Dpo4 (64 min⁻¹).

A comparison of the aa-sequence of the LF-Domain of Dpo4 vs. Dbh reveals that their greatest difference is in a nine amino acid region.

Dpo4: I₂₃⁷ R T R V R K S I₂₄⁵

Dbh: V₂₃⁸ E N K S K I P H₂₄⁶

DNAP IV: N₂₃⁵ S E R L R K S V₂₄₃

These nine amino acids were placed in Dpo4 to give aa238-246/Dbh-Dpo4, whose dGTP misinsertion rate (17 x 10⁻³ min⁻¹) is ~3.7-fold lower than wt-Dpo4 (64 x 10⁻³ min⁻¹). The protein aa238-aa246/Dbh-Dpo4 also has lower dATP misinsertion (~2-fold) and dTTP misinsertion (~1.4-fold) compared to Dpo4. In contrast, correct dCTP insertion is unaffected for aa238-246/Dbh-Dpo4 (6.0 x 10⁻³ min⁻¹) vs. Dpo4 (5.8 x 10⁻³ min⁻¹).

While our work was in progress, another group showed that a Dpo4 derivative containing this region of Dbh behaved more like Dbh in several respects (76). For
example, Dpo4 misinserts dATP/dGTP/dTTP more readily than Dbh on undamaged DNA, while Dpo4 with aa232-246 from Dbh has dATP/dGTP/dTTP misinsertion properties more akin to Dbh. Furthermore, the both Dbh and its aa232-246/Dbh-Dpo4 derivative bypassed an Abasic site less efficiently than Dpo4 (76). In the same study a less extensive substitution of only three amino acids (aa243-245/Dbh-Dpo4) behaved almost the same as aa232-246/Dbh-Dpo4.

A variety of X-ray structures show that Dbh forms a non-covalent bridge (NCB), which is on the minor groove side of the active site and is ~8Å distal to the active site (76). In the Dbh X-ray structure, the dominant contact in the Distal-NCB involves a hydrophobic contact between V101 and I244, the latter being in this region. No Dpo4 X-ray structure shows an analogous Distal-NCB. It seems likely that this Distal-NCB is the structural basis for suppressing some of the activity differences between Dbh and Dpo4, including the suppression of dATP, dGTP and/or dTTP misinsertions opposite –BP-dG.

We believe that DNAP IV is also likely to have a Distal-NCB to suppress misinsertion opposite –BP-dG. When we developed our DNAP IV molecular models, we took a homology modeling approach using Dpo4 as a template (65,80,90). Since Dpo4 does not have a Distal-NCB, our starting DNAP IV structures also did not have a Distal-NCB. Nevertheless, in all of our molecular dynamics trajectories, DNAP IV formed a Distal-NCB. What drove the formation of this Distal-NCB? We do not believe it reflects an element in the analogous region of DNAP IV, which more closely resembles Dpo4 than Dbh. In fact, DNAP IV’s Distal-NCB depends on P99, which enforces a kink in a b-strand in the TPF-domains, thus, forcing this region in DNAP IV to be ~3Å closer to the
LF-domain than the corresponding region in Dpo4. (L102 also appears to be important because it fills a hydrophobic pocket.) To test this notion, we studied the double mutant I101P/A102L-Dpo4, which resembles DNAP IV in this region. dGTP misinsertion rate is ~2.4-fold slower for I101P/A102L-Dpo4 (26 x 10^{-3} \text{ min}^{-1}) than for Dpo4 (64 x 10^{-3} \text{ min}^{-1}), while dCTP is slightly affected (4.4 x 10^{-3} \text{ min}^{-1} vs. 5.8 x 10^{-3} \text{ min}^{-1}). dATP misinsertion also decreases (~2.7-fold), though dTTP is decreased less (~1.4-fold). In fact, I99P/A100L-Dpo4 is remarkably similar to aa238-246/Dbh-Dpo4 for all dNTP incorporations.

In our model R240 in the LF-Domain reaches across to the TPF-Domains to form the Distal-NCB (Figure 4.1B). In contrast, the equivalent residue in Dpo4 (R242) interacts with DNA. Ten X-ray structures of DNAP IV with undamaged DNA were recently reported (76). R240 adopts seven different positions in these ten structures. In three structures, R240 interacts with both template-DNA and P99(C=O). In two structures, R240 interacts with E98 and P99(C=O). In one structure each, R240 interacts with P99(C=O), E237, R240(C=O), template-DNA and E98/P99(C=O)/template-DNA. Thus, these X-ray structures show heterogeneity for R240 positioning, but in 8/10 cases some kind of Distal-NCB is evident.

Findings presented above suggested that -BP-dG has bulk in the minor groove during dATP and dTTP misinsertion, so it is easy to rationalize why the Distal-NCB, which is in the minor groove, might suppress dATP and dTTP misinsertion. In contrast, the Distal-NCB has little effect on dCTP insertion, even though findings discussed above suggest that -BP-dG has bulk in the minor groove during dCTP insertion, which is another
piece of evidence that the −BP-dG conformation during dATP/dTTP misinsertion is different than during correct dCTP insertion.

All of the protein changes that suppressed dGTP misinsertion (i.e., NC-Dpo4, A42V-Dpo4, D3-Dpo4, etc.) were in the TPF-Domains in Dpo4 and on the major groove side of the active site. However, the presence of a Distal-NCB, which is in the minor groove also suppresses dGTP misinsertion, as noted above. This finding suggests that the formation of non-covalent bridges between the LF-Domain and the TPF-Domains help position the LF-Domain to block aberrant DNA structures from entering the active site whether the aberrant structures have bulk in the minor or major grooves.

**An Active Site Non-Covalent Bridge Also Suppresses dATP, dGTP and dTTP Misinsertion**

X-ray structures reveal that Dpo4 has an active site non-covalent bridge (AS-NCB), which involves two Coulombic interactions: (1) R36 in the AS-loop interacting with N254 in the LF-domain, and (2) K253 in the LF-domain interacting with the carbonyl-oxygen of R36 (Figure 4.5A, yellow arrows). R36 is the only arginine in Dpo4’s AS-loop, while DNAP IV has an arginine in approximately the equivalent position (R37), along with two nearby arginines (R35 and R38), whose significance was investigated. dGTP misinsertion was much lower in wt-DNAP IV ($0.63 \times 10^{-3} \text{ min}^{-1}$) than in either R37A-DNAP IV ($17.3 \times 10^{-3} \text{ min}^{-1}$) or R38A-DNAP IV ($16.2 \times 10^{-3} \text{ min}^{-1}$). dATP misinsertion also increased significantly for R37A-DNAP IV (33-fold) and R38A-DNAP IV (13-fold, respectively), as did dTTP misinsertion (21-fold and 80-fold, respectively).
(R35A-DNAP IV is currently being examined, and will not be discussed herein due to conflicting results that we have at this time.)

In contrast, correct dCTP insertion for R37A-DNAP IV (53.6 \times 10^{-3} \text{ min}^{-1}) and R38A-DNAP IV (33.1 \times 10^{-3} \text{ min}^{-1}) were only slightly greater than for wt-DNAP IV (25.6 \times 10^{-3} \text{ min}^{-1}), which shows that the structural purpose of R37 and R38 in DNAP IV is primarily to suppress misinsertions.

All ten DNAP IV X-ray structures show a Coulombic interaction between R38 and D252 that contributes to an AS-NCB. In contrast, in most of these X-ray structures R37 appears to be interacting with solvent and is not involved in the AS-NCB. However, we have biochemical evidence to support the notion that R37 in DNAP IV forms a Coulombic interaction with E251 that is part of the AS-NCB, which is addressed below, but will be described in detail in Chapter 5.

The evidence in this section shows that protein modifications that are likely to weaken the AS-NCB in DNAP IV (i.e., 37A and R38A) significantly increase incorrect dATP, dGTP and dTTP, implying that its AS-NCB is designed to suppress these misinsertion. Once again, the AS-NCB, which is in the minor groove, also suppresses the dGTP misinsertion, whose bulk is in the major groove, which suggests that the formation of non-covalent bridges between the LF-Domain and the TPF-Domains help position the LF-Domain to block aberrant DNA structures from entering the active site whether the aberrant structures have bulk in the minor or major grooves.
DISCUSSION

BP-N²-adducts have an intact hydrogen bonding face, so they can in principle form a more-or-less normal Watson-Crick-type base pair with dCTP. However, the BP-moiety itself is large, and it is intuitively reasonable to imagine that any DNAP that could correctly incorporate dCTP must have special structural elements to accommodate this minor groove bulk. The fact that Dpo4 is able to incorporate opposite –BP-dG shows that Y-Family DNAPs have this capacity, though the Dpo4 structure is not ideal, given its inaccuracy, as the order of incorporation opposite –BP-dG is dGTP > dATP > dCTP ≈ dTTP. This is reflected in the five X-ray structures of BP-adducts in Dpo4, which ostensibly all appear distorted and are vastly different from each other as summarized in the introduction. As we have noted in modeling studies of BP-adducts in DNAPs (65), +BP-dG, because of its adduct bond geometry, has its BP-moiety protruding almost perpendicularly to the helix axis leading one to predict that it will be particularly difficult to bypass. In contrast the adduct bond in –BP-dG allows the bulky BP-moiety to lie flat against the minor groove, such that it should be relatively easier to bypass. In studies to be published elsewhere this is observed: Dpo4, Dbh and DNAP IV each incorporate dNTPs more rapidly opposite -BP-dG than +BP-dG.

The fact that dGTP misinsertion opposite –BP-dG with wt-Dpo4 is much faster than the other dNTPs suggests that its bypass might involve a special conformation. Three different structural changes in Dpo4, where each added protein bulk in the major groove, all decreased dGTP misinsertion: N-clasp addition (N-clasp-Dpo4), addition of the A42V mutation (A42V-Dpo4), and removing the small bulge in Dpo4 that moves the
AS-loop away from the major groove opening (D3-Dpo4). These three changes did not significantly affect dATP, dCTP or dTTP misincorporation rate, which serves as a kind of internal control and suggests that BP-moiety bulk is not in the major groove during their incorporation. Others have noted that having the bulk of the N-clasp inhibits bypass of adducts with bulk on the major groove side of the active site (86,87), and molecular modeling studies have shown why (88).

On the minor groove side of the active site, Y-Family DNAPs have an opening, which is relatively large in κ/IV-class DNAPs, but is plugged in η/V-class DNAPs; e.g., by V32 and M76 in Dpo4, and by large aa-inserts in hDNAP η (aa81-87) and ScDNAP η (aa93-126). We have pointed out that the BP-moiety would have to protrude from an opening on the minor groove side of the active site in order to form a more-or-less normal Watson-Crick base pair, and, thus, a large opening on the minor groove side of the active site should facilitate this (65,80). Our data are consistent with this notion. dCTP insertion is slower for Dpo4 \( (5.8 \times 10^{-3} \text{ min}^{-1}) \), in which V32 and M76 plug the minor groove opening, than with DNAP IV \( (26 \times 10^{-3} \text{ min}^{-1}) \), which has glycines in the equivalent positions (G32 and G72). Consistent with this notion is the fact that—even though Dbh is nominally in the κ/IV-class of Y-Family DNAPs—it has a small minor groove opening, because it also has V32 and M76, and Dbh’s dCTP insertion rate \( (2.5 \times 10^{-3} \text{ min}^{-1}) \) is even lower than with Dpo4 \( (5.8 \times 10^{-3} \text{ min}^{-1}) \). dCTP insertion is \( \sim2.2 \)-fold higher for V32G/M76G-Dpo4 compared to Dpo4, though this is less than the \( \sim4.4 \)-fold higher rate for DNAP IV compared to Dpo4. We suspected that there were other elements in the DNAP IV AS-loop that were also important, but this seemed not to be the
case given that placing the entire DNAP IV AS-loop into Dpo4 did not further increase dCTP insertion, as dCTP insertion for IVloop-Dpo4 was ~1.4-fold greater than wt-Dpo4. Based on other lines of thinking we hypothesized that this might be due the fact that the AS-loop in DNAP IV has three arginines (R35, R37 and R38), which might keep the IVloop from functioning correctly in the Dpo4 context, which has only one arginine (R36). This hypothesis proved correct; e.g., dCTP insertion for R37A-IVloop-Dpo4 (21 x 10^{-3} \text{ min}^{-1}), which has one fewer arginines, was close to the rate for wt-DNAP IV (26 x 10^{-3} \text{ min}^{-1}).

While having a large minor groove opening does increase dCTP insertion rate, any benefit to fidelity is negated by the similar increase for dATP and dTTP misinsertions; e.g., adding the V32G/M76G mutations to Dpo4, which increase the size of its minor groove opening, also increased misincorporation rates for both dATP (~2-fold) and dTTP (~2.3-fold) relative to Dpo4. While this finding does suggest that the BP-moiety is likely to have bulk in the minor groove during dATP/dTTP misinsertion, it also means that DNAP IV must have additional structural elements to suppress dATP/dTTP misinsertion. Furthermore, Dbh, whose minor groove opening is also plugged by V32/M76 must have mechanisms to suppress dATP/dTTP misinsertion. In the case of Dbh, dATP/dTTP suppression was traced to residues in the LF-Domain, given that the dATP and dTTP misinsertion rates for the chimera Dbh/LF-Dpo4 (1.6 x 10^{-3} \text{ min}^{-1} and 0.7 x 10^{-3} \text{ min}^{-1}) are close to those for Dbh itself (0.9 x 10^{-3} \text{ min}^{-1} and 0.6 x 10^{-3} \text{ min}^{-1}). Additional experiments indicated that aa238-246 in Dbh is one structural component, since aa238-246/Dbh-Dpo4 had dATP and dTTP misinsertion rates (5.3 x 10^{-3} \text{ min}^{-1} and
3.4 x 10^{-3} \text{ min}^{-1} \) that were lower than for Dpo4 itself (11 x 10^{-3} \text{ min}^{-1} and 4.9 x 10^{-3} \text{ min}^{-1}).

However, other parts of the LF-Domain of Dbh must also be important, since suppression of dGTP misinsertion is less for aa238-246/Dbh-Dpo4 (~3.7-fold) than for the chimera Dbh/LF-Dpo4 (21-fold). Dpo4 has no Distal-NCB, and R242 in this region of Dpo4 interacts with template-DNA. Dbh has no arginine at this position to interact with DNA; however, Dbh appears to compensate for this lost DNA interaction via R283 and R325, both of which interact with DNA, whereas the equivalent residues in Dpo4 (A283 and E325) do not interact with DNA. Thus, R283 and R325, which are present in the chimera Dbh/LF-Dpo4, but are lacking in aa238-246/Dbh-Dpo4, might explain why Dbh/LF-Dpo4 suppresses dGTP misinsertion more than aa238-246/Dbh-Dpo4.

X-ray structures reveal that Dbh forms a Distal-NCB between the TPF-Domains and the LF-Domain that depends on residues in the aa238-246 region. Our data suggest that the Distal-NCB is important for suppressing dATP/dTTP misinsertion with –BP-dG, while the work of others has shown that this Distal-NCB is also likely to be important for proper functioning of Dbh vis-à-vis other bypass events, such as suppressing misinsertion on undamaged DNA templates (76). DNAP IV may also have a Distal-NCB that helps suppress dATP/dTTP misinsertion. The DNAP IV Distal-NCB depends in part on P99/L100, which cause a structural kink that is absent in Dpo4, which has I101/A102 in the equivalent positions. The mutations in I101P/A102L-Dpo4 suppressed dATP and dTTP misinsertion (4 x 10^{-3} \text{ min}^{-1} and 3.5 x 10^{-3} \text{ min}^{-1}) relative to Dpo4 itself (11 x 10^{-3} \text{ min}^{-1} and 4.9 x 10^{-3} \text{ min}^{-1}).
In addition, a second non-covalent bridge between the TPF-Domains and the LF-Domain forms near the active site (AS-NCB) and also seems to be important for suppressing dATP/dTTP misinsertions. Evidence for this comes most clearly from studies in DNAP IV, where both R37A and R38A mutations significantly elevated misinsertion of both dATP (33- and 13-fold, respectively) and dTTP (21- and 80-fold, respectively) compared to wt-DNAP IV. DNAP IV X-ray structures show that R38 interacts with D252 in the LF-Domain to help form an AS-NCB. We have biochemical evidence to support the notion that R37 interacts with E251 in the LF-Domain (to be presented elsewhere). In X-ray structures of DNAP IV, R37 is not interacting with E251, but rather looks to be in contact with solvent in the minor groove (76). How can this difference be reconciled? The most obvious possibility relates to the fact that DNAP IV X-ray structures are with undamaged DNA, and accommodating the –BP-dG adduct in the active site might require a conformational change, which would allow/require R37 to interact with E251. A second possibility is that X-ray structures reveal structural information about ground state structures, while structure-activity approaches, which include studies on how amino acid substitutions affect reaction rates, reveal information about transition state structures, where ground state and transition state structures could be different (69).

Both the Distal-NCB and the AS-NCB are in the minor groove, and yet they also suppressed dGTP misinsertion, which is a reaction where –BP-dG bulk is likely in the major groove. This observation argues that non-covalent bridges between the LF-Domain and the TPF-Domains help position the LF-Domain to block aberrant DNA
structures from entering the active site whether the aberrant structures have bulk in the minor or major grooves.

Many other structure-activity studies have revealed the roles of important amino acids for Y-Family DNAP function. For example, the “steric gate” is a conserved Phe or Tyr residue that is present to ensure that rNTPs are excluded from the active site (42,77,91-93). Regarding BP-adduct bypass, F171 in hDNAP κ has been called a “molecular break” because it slows down bypass of BP adducts (94) we note that EcDNAP IV has F76 in the equivalent position, and, in fact, most Y-Family DNAPs have F, Y or another bulky hydrophobic residue at this site. We have published evidence (95) that Y-Family DNAPs fall into two categories with respect to the roof amino acid, which lies above the nucleobase on the dNTP, and the next amino acid: The [roof-aa/next-aa] is preferentially [I/A] in η/V-class DNAPs, while the κ/IV-class prefers [A/T] or [S/T]. In DNAP IV, the hydroxyl group of the roof-aa Ser42 was shown to improve fidelity via its –OH group forming H-bonds in the minor groove with the template:dNTP base pair (73). Ten distal amino acids in EcDNAP IV were shown to be important for promoting effective dCTP insertion opposite a N2-furfuryl-dG adduct (96).

A role for Hoogstein base pairing, which requires anti-to-syn base rotation, has often been proposed as a possible mechanism for how adduct:dNTP mispairing might lead to mutagenesis; e.g., with DNAP i (75). DNAP τ promotes Hoogsteen base pairing because its cramped active site preferentially accommodates the shorter [N9:N1] distance found in [syn-purines;pyrimidine] base pairing. Interestingly, evidence suggests that the major dG-adduct of 1-nitropyrene forms a Watson-Crick base pair with dCTP in DNAP τ.
though this is a ground state X-ray structure. A role for Hoogsteen base pairing in other Y-Family DNAPs, which do not have a cramped active site, has not been established. Our work herein strongly suggests that indeed a BP-adduct-\textit{syn}-dG:dGTP Hoogsteen base pair is likely to be mechanistically important. Template looping out, such that misincorporation is directed by the \([n+1]\) base, has also been considered (97).

In summary we have identified three types of structural elements that Y-Family DNAPs in the \(\kappa/IV\)-class use to enhance fidelity in the bypass of \(N^2\)-dG adducts such as –BP-dG. (1) A large opening on the minor groove side of the active site facilitates correct dCTP insertion, because the BP-moiety must be in the minor groove for Watson-Crick-like adduct-dG:dCTP pairing. (2) Having protein structural features that plug the major groove side of the active site suppresses dNTP misinsertions when a lesion is present on the major groove side, such as during dGTP misinsertion with –BP-dG. (3) Non-covalent bridges between the TPF-domains and the LF-domain suppress dNTP misinsertions by excluding aberrant conformations from the active site, whether adduct bulk is in the minor or major groove, such as in the suppression of dATP, dGTP and dTTP misinsertion opposite–BP-dG.
**Table 4.1:** Relative Velocity of dNTP insertion opposite –BP-dG for Dpo4 and mutant-Dpo4s, as well as Dbh.

<table>
<thead>
<tr>
<th>Modification of dATP</th>
<th>Velocity (x 10^3 min^-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpo4</td>
<td>Derivative</td>
</tr>
<tr>
<td>wt</td>
<td>Dpo4</td>
</tr>
</tbody>
</table>

**Major Groove Considerations**

<table>
<thead>
<tr>
<th>Modification</th>
<th>Dpo4</th>
<th>NC</th>
<th>A42V</th>
<th>Δ3</th>
<th>Vloop</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>Dpo4</td>
<td>10.9 (1.02)</td>
<td>5.8 (1.0)</td>
<td>13.7 (0.22)</td>
<td>4.3 (0.88)</td>
</tr>
<tr>
<td>A42V</td>
<td>Dpo4</td>
<td>12.2 (1.14)(^2)</td>
<td>5.8 (1.0)</td>
<td>7.8 (0.28)</td>
<td>4.6 (0.95)</td>
</tr>
<tr>
<td>Δ3</td>
<td>Dpo4</td>
<td>6.7 (0.63)</td>
<td>6.9 (1.17)</td>
<td>11.2 (0.18)</td>
<td>4.5 (0.93)</td>
</tr>
<tr>
<td>Vloop</td>
<td>Dpo4</td>
<td>9.2 (0.87)</td>
<td>6.0 (1.03)</td>
<td>19.8 (0.31)</td>
<td>4.7 (0.96)</td>
</tr>
</tbody>
</table>

**Minor Groove Opening Size Considerations**

<table>
<thead>
<tr>
<th>Modification</th>
<th>Dpo4</th>
<th>V32G/M74G</th>
<th>Kloop</th>
<th>NC-Kloop</th>
<th>IVloop</th>
<th>R35A-IVloop</th>
<th>R37A-IVloop</th>
<th>R38A-IVloop</th>
</tr>
</thead>
<tbody>
<tr>
<td>V32G/M74G</td>
<td>Dpo4</td>
<td>22.4 (2.11)</td>
<td>12.7 (2.18)</td>
<td>54.4 (0.86)</td>
<td>11.1 (2.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kloop</td>
<td>Dpo4</td>
<td>10.3 (0.97)</td>
<td>14.5 (2.49)</td>
<td>58.3 (0.92)</td>
<td>11.6 (2.37)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC-Kloop</td>
<td>Dpo4</td>
<td>3.5 (0.33)</td>
<td>5.9 (1.02)</td>
<td>18.0 (0.28)</td>
<td>4.1 (0.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVloop</td>
<td>Dpo4</td>
<td>5.0 (0.47)</td>
<td>8.4 (1.44)</td>
<td>12.0 (0.19)</td>
<td>3.3 (0.67)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R35A-IVloop</td>
<td>Dpo4</td>
<td>2.9 (0.28)</td>
<td>1.7 (0.29)</td>
<td>8.6 (0.14)</td>
<td>3.3 (0.74)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R37A-IVloop</td>
<td>Dpo4</td>
<td>13.7 (1.29)</td>
<td>20.8 (3.56)</td>
<td>12.3 (0.19)</td>
<td>10.6 (2.16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R38A-IVloop</td>
<td>Dpo4</td>
<td>10.4 (0.98)</td>
<td>13.5 (2.31)</td>
<td>7.0 (0.11)</td>
<td>3.2 (0.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Distal-NCB Considerations**

<table>
<thead>
<tr>
<th>Modification</th>
<th>Dpo4</th>
<th>wt</th>
<th>LF/Dbh</th>
<th>aa238-246/Dbh</th>
<th>I101P/A102L</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>Dbh(^2)</td>
<td>0.85 (0.080)</td>
<td>2.5 (0.42)</td>
<td>1.26 (0.020)</td>
<td>0.55 (0.11)</td>
</tr>
<tr>
<td>LF/Dbh</td>
<td>Dpo4</td>
<td>1.6 (0.16)</td>
<td>2.5 (0.43)</td>
<td>2.9 (0.047)</td>
<td>0.71 (0.15)</td>
</tr>
<tr>
<td>aa238-246/Dbh</td>
<td>Dpo4</td>
<td>5.3 (0.50)</td>
<td>6.0 (1.03)</td>
<td>17.1 (0.27)</td>
<td>3.4 (0.70)</td>
</tr>
<tr>
<td>I101P/A102L</td>
<td>Dpo4</td>
<td>4.0 (0.37)</td>
<td>4.4 (0.76)</td>
<td>26.2 (0.41)</td>
<td>3.5 (0.73)</td>
</tr>
</tbody>
</table>

**AS-NCB Considerations**

<table>
<thead>
<tr>
<th>Modification</th>
<th>Dpo4</th>
<th>wt</th>
<th>R35A</th>
<th>R37A</th>
<th>R38A</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>DNAP IV</td>
<td>0.29 (0.028)(^2)</td>
<td>25.6 (4.38)</td>
<td>0.63 (0.010)</td>
<td>0.13 (0.026)</td>
</tr>
<tr>
<td>R35A</td>
<td>DNAP IV</td>
<td>0.21 (0.020)</td>
<td>39.3 (6.73)</td>
<td>1.27 (0.020)</td>
<td>0.92 (0.19)</td>
</tr>
<tr>
<td>R37A</td>
<td>DNAP IV</td>
<td>9.7 (0.91)</td>
<td>53.6 (9.19)</td>
<td>17.3 (0.27)</td>
<td>2.71 (0.55)</td>
</tr>
<tr>
<td>R38A</td>
<td>DNAP IV</td>
<td>3.8 (0.36)</td>
<td>35.1 (6.02)</td>
<td>16.2 (0.26)</td>
<td>10.4 (2.14)</td>
</tr>
</tbody>
</table>

\(^1\) Rates are for dNTP insertion opposite a –BP-dG adduct in a primer/template using methods described in Chapter 3. Purified proteins included Dpo4, along with fifteen mutant-Dpo4s, as well as Dbh. Reactions were pseudo-first order (Chapter 3) and (e.g.)
the number 10.5 for Dpo4 with dATP means the velocity was $10.6 \times 10^{-3}\text{min}^{-1}$ using standard conditions, which were dNTP concentration [100µM] and protein concentration 50ng/mL.

Numbers in parenthesis are the velocity (non-parenthesis on the left) divided by the velocity of the corresponding Dpo4 reaction; e.g., for dATP/A42V-Dpo4 is $1.14 = (12.2/10.6)$.

Dbh was studied identically.
Table 4.2: Dominant dNTP insertions opposite various DNA adducts/lesions by *E. coli* DNAPs IV and V and human DNAPs κ and η.¹

<table>
<thead>
<tr>
<th>Lesion</th>
<th>DNAP V</th>
<th>DNAP η</th>
<th>Dpo4</th>
<th>DNAP IV</th>
<th>DNAP κ</th>
<th>Dbh</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+ta]-BP-N²-dG</td>
<td>A/C</td>
<td>A&gt;G</td>
<td>A&gt;C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AAF-C8-dG</td>
<td>C</td>
<td>C</td>
<td>-</td>
<td>C/T</td>
<td>C/T</td>
<td>-</td>
</tr>
<tr>
<td>AF-C8-dG</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>TT-CPD</td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>T(6-4)T</td>
<td>AG</td>
<td>nG</td>
<td>AG</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>AP site</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>n</td>
<td>A*</td>
<td>n</td>
</tr>
</tbody>
</table>

¹ Dominant dNTP insertion using purified DNAPs, where “n” indicates “no” or low activity, “A*” indicates bypass by an unusual mechanism, and “-” indicates data unavailable.
LEGENDS

Figure 4.1: Views of Y-Family DNAPs SsDpo4 (Panel A), hDNAP κ (Panel C) and SsDbh (Panel D), based on X-ray structures (31,37,70) along with DNAP IV (Panel B), based on a homology model (65). The view is from the minor groove and the TPF-Domains are shown in green and the LF-Domain is shown in yellow. Noncovalent bridges form between the TPF-Domains and the LF-Domain, which can be of three types: near the active site (AS-NCB) and distal to the active site on the minor groove side, or on the major groove side near the active site that is present uniquely in hDNAP κ, because of its N-clasp (scarlet structure). These NCBs are shown pictorially in Panel F. Panel E contains another representation of DNAP IV showing its minor groove opening, which is large enough to accommodate the BP moiety of BP-dG adducts.

Figure 4.2: Structures of +BP-dG and –BP-dG, along with the primer/tem,late sequence and a phophorimage showing dNTP incorporation by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and LF/Dbh-Dpo4), which are explained in the text.

Figure 4.3: Pictorially presentation of the impact on dNTP incorporation rate when various amino acid changes are made in a variety of Y-Family DNAPs. The plots show data after normalization of rate by dividing [mutant-Dpo4/wt-Dpo4] for different dNTP. For example, the average rate of dGTP misinsertion was 0.00178 min⁻¹ for A42V-Dpo4 and was 0.0635 min⁻¹ for wt-Dpo4, so normalized dGTP insertion for A42V-Dpo4 was
~0.28 (= 0.00178/0.0681). Panel A shows normalized dTTP misinsertions versus normalized dATP misinsertion (as a Log-Log plot) for a variety of mutant-Dpo4s, using data from Table 1. The point for Dpo4 itself is by definition at the origin (0, 0). Data for DNAP IV (Table 2), along with several of its derivatives, as well as Dbh, are also in the plot. Symbols change for Dpo4 (closed circles), DNAP IV (open circles) and Dbh (open square). Log-Log plots for dCTP versus dATP (Panel B) and dCTP versus dTTP (Panel C) are also shown.

Figure 4.4: Figure 4.4 is similar to Figure 4.3, except for dCTP versus dGTP. All of the data is shown in Panel A, while selected data from Panel A is shown in Panels B-D.

Figure 4.5: Amino acids in the active site loop (AS-loop) of Dpo4 (Panel A) and DNAP IV (V31-V40, Panel B). The view is from the major groove, with the template base and dNTP included for reference. The AS-loops in Y-Family DNAPs begin on the minor groove side of the active site (yellow amino acids) and then turns back to the major groove side of the active site (brown amino acids). Certain key amino acids in the AS-loop are shown as described in the text.

Figure 4.6: The opening on the minor groove side near the active site for Dpo4 (Panel A) and DNAP IV (Panel B), showing a key difference: the Dpo4 opening is plugged by M75 and V32, while the DNAP IV opening is more spacious because its equivalent amino acids are glycines (G73 and G32).
Figure 4.1: Views of Y-Family DNAPs SsDpo4 (Panel A), hDNAP κ (Panel C) and SsDbh (Panel D), based on X-ray structures (31,37,70) along with DNAP IV (Panel B),
based on a homology model (65). The view is from the minor groove and the TPF-Domains are shown in green and the LF-Domain is shown in yellow. Noncovalent bridges form between the TPF-Domains and the LF-Domain, which can be of three types: near the active site (AS-NCB) and distal to the active site on the minor groove side, or on the major groove side near the active site that is present uniquely in hDNAP κ, because of its N-clasp (scarlet structure). These NCBs are shown pictorially in Panel F. Panel E contains another representation of DNAP IV showing its minor groove opening, which is large enough to accommodate the BP moiety of BP-dG adducts.
Figure 4.2: Structures of +BP-dG and –BP-dG, along with the primer/template sequence and a phosphorimage showing dNTP incorporation by the Y-Family DNAPs SsDpo4 and SsDbh, along with two Dpo4 derivatives (aa238-246/Dbh-Dpo4 and LF/Dbh-Dpo4), which are explained in the text.
Figure 4.3: Pictorially presentation of the impact on dNTP incorporation rate when various amino acid changes are made in a variety of Y-Family DNAPs. The plots show
data after normalization of rate by dividing [mutant-Dpo4/wt-Dpo4] for different dNTP. For example, the average rate of dGTP misinsertion was 0.00178 min$^{-1}$ for A42V-Dpo4 and was 0.0635 min$^{-1}$ for wt-Dpo4, so normalized dGTP insertion for A42V-Dpo4 was \(~0.28 (= 0.00178/0.0681)\). Panel A shows normalized dTTP misinsertions versus normalized dATP misinsertion (as a Log-Log plot) for a variety of mutant-Dpo4s, using data from Table 1. The point for Dpo4 itself is by definition at the origin (0, 0). Data for DNAP IV (Table 2), along with several of its derivatives, as well as Dbh, are also in the plot. Symbols change for Dpo4 (closed circles), DNAP IV (open circles) and Dbh (open square). Log-Log plots for dCTP versus dATP (Panel B) and dCTP versus dTTP (Panel C) are also shown.
Figure 4.4: Figure 4.4 is similar to Figure 4.3, except for dCTP versus dGTP. All of the data is shown in Panel A, while selected data from Panel A is shown in Panels B-D.
Figure 4.5: Amino acids in the active site loop (AS-loop) of Dpo4 (Panel A) and DNAP IV (V31-V40, Panel B). The view is from the major groove, with the template base and dNTP included for reference. The AS-loops in Y-Family DNAPs begin on the minor groove side of the active site (yellow amino acids) and then turns back to the major
groove side of the active site (brown amino acids). Certain key amino acids in the AS-loop are shown as described in the text.
Figure 4.6: The opening on the minor groove side near the active site for Dpo4 (Panel A) and DNAP IV (Panel B), showing a key difference: the Dpo4 opening is plugged by M75 and V32, while the DNAP IV opening is more spacious because its equivalent amino acids are glycines (G73 and G32).
CHAPTER 5
Architecture of Y-Family E. coli DNA Polymerase IV
That Controls Accurate dCTP Insertion
Opposite an N²-dG Adduct (Benzo[a]pyrene)

SUMMARY

To synthesize past DNA damage, cells have lesion bypass DNA polymerases (DNAPs), often in the Y-Family. One class of Y-Family DNAPs includes hDNAP κ and EcDNAP IV, which accurately insert dCTP opposite N²-dG adducts, including from the environmental carcinogen benzo[a]pyrene (BP-N²-dG). Another class includes hDNAP η and EcDNAP V, which insert accurately opposite UV-damage, but inaccurately insert opposite BP-N²-dG adducts. Herein we make protein modifications in EcDNAP IV that either decrease correct dCTP insertion or increase dGTP/dATP/dTTP misinsertion, thereby identifying structural elements that contribute to insertion fidelity opposite a BP-N²-dG adduct. Y-Family DNAPs have a little finger domain (LF-domain) that is tethered via an amino acid linker to the thumb, palm and fingers domains, which operate as a unit (TPF-domains). We show that non-covalent bridges (NCBs) between the LF-domain and the TPF-domains help suppress dATP, dGTP and dTTP misinsertions. Evidence is presented that R37 and R38, which are in a loop near the active site in the TPF-domains, interact with E251 and D252 in the LF-domain, respectively, and when these interactions are lost dATP, dGTP and dTTP misinsertion rate increases opposite BP-N²-dG. The structure of the AS-NCBs in DNAP IV is compared to the AS-NCB in Dpo4, which has
R36:N252 and K250:R36(C=O) interactions and is less efficient at suppressing dATP, dGTP and dTTP misinsertions. Evidence is also presented that DNAP IV suppresses dATP, dGTP and dTTP misinsertions via a second non-covalent bridge between the LF-domain and the TPF-domain, which is ~8Å from the active site (Distal-NCB) and depends on a P99-enforced kink in a β-strand in the TPF-Domains of DNAP IV. Dpo4 does not have this kink and lacks a distal-NCB, making Dpo4 less efficient at suppressing dATP, dGTP and dTTP misinsertions. These findings are considered in the context of other Y-Family DNAPs.

INTRODUCTION

Cells possess many DNA polymerases (DNAPs); e.g., humans, yeast (S. cerevisae) and E. coli have at least fifteen, eight and five, respectively (1-3, 72). These DNAPs serve many functions; e.g., replicative DNAPs are often blocked by the DNA damage caused by chemicals and radiation, and to avoid such lethal blockage, cells possess lesion-bypass DNAPs (1-17), which conduct translesion DNA synthesis (TLS). Most lesion-bypass DNAPs are in the Y-family (1-17), where human cells have three that are template-directed (hDNAPs η, ι and κ), while yeast has one (ScDNAPs η) and E. coli has two (EcDNAPs IV and V).

X-ray structures of Y-Family members hDNAP κ, hDNAP ι, hDNAP η, ScDNAP η, SsDpo4, SsDbh and EcDNAP IV show a conserved ~350aa core, which includes the polymerase active site [representative references: 18-35, (73-76)]. As with all
DNA polymerases, Y-Family members resemble a right-hand with thumb, palm and fingers domains, which form a unit (TPF-Domains). The fingers are “stubby,” resulting in more solvent accessible surface around the template/dNTP-binding pocket (8), which can accommodate the bypass of bulky and/or deforming DNA adducts/lesions that typically protrude into these open spaces. Y-Family DNAPs grip DNA with an additional domain (18-20), often called the “little finger” (LF-Domain), which is tethered to the TPF-Domains by a long (~15aa) linker. The best studied Y-Family DNAP is Dpo4 from Sulfolobus solfataricus, and a representative Dpo4 X-ray structure is shown in Figure 5.1A, which is a view from the minor groove side with the LF-Domain in yellow and the TPF-Domains in green. Figure 5.1 also shows representative X-ray structures for EcDNAP IV (Figure 5.1B, (73)) and hDNAP κ (Figure 5.1C, (88)). A modeled structure for SsDbh is also shown (Figure 5.1D), which we developed based on an X-ray structure (76) for reasons described below. Steps in the mechanism of Y-Family DNAPs have been proposed for both protein structural changes (15, 22, 24, 25, 31, 36) and for chemical catalysis (37).

Our work has focused on benzo[a]pyrene (B[a]P), which is a well-studied DNA damaging agent that is a potent mutagen and carcinogen, and an example of a polycyclic aromatic hydrocarbon (PAH), a class of ubiquitous environmental substances produced by incomplete combustion (50). PAHs in general and B[a]P in particular induce the kinds of mutations that are thought to be relevant to carcinogenesis and may be important in human cancer (51, and references therein). The major B[a]P DNA adduct (+BP) forms at N²-dG and is shown in Figure 5.2, which also shows its mirror image –BP.
Based on the insertion preference opposite a variety of lesions, Y-Family DNAPs fall into different classes. The IV/κ-class insert correct dCTP opposite a variety of N²-dG adducts, including adducts from environmental carcinogen (e.g. BP-dG (77-79)), as well as from endogenous sources, such as reactive derivatives of cellular trioses (47) or adducts formed from lipid peroxidation processes (49), which may be the main cellular rationale for the genesis of the IV/κ-class. Y-Family DNAPs in the IV/κ-class also correctly insert opposite a variety of other lesions, including PhIP-C8-G, TG, 8oxoG and some interstrand crosslinks (81-84). DNAPs in the IV/κ-class do not insert opposite UV-damage and tend to be relatively accurate in the replication of undamaged DNA. The κ/IV-class includes hDNAP κ, EcDNAP IV and SsDbh. The V/η-class tend to be relatively inaccurate in the replication of undamaged DNA and inaccurate in insertion opposite N²-dG adducts, but accurate with UV-damage, which may be the main cellular rationale for the genesis of the V/η-class. EcDNAP V, hDNAP η, and SsDpo4 are in the V/η-class.

The differences in insertional preferences by DNAPs in the κ/IV-class versus the η/V-class must reflect an interface between protein structural differences and adduct structural/conformational differences. In Chapter 4, we presented evidence that the pyrene moiety of the –BP-dG adduct is in the minor groove during dCTP insertion, which is sensible, because the pyrene moiety is attached to the N²-position, which must be in the minor groove to form a more-or-less normal adduct-dG:dCTP Watson-Crick base pair. Based on molecular modeling work, we pointed out that DNAP IV has a large opening in its protein surface on he minor groove side (Figure 5.1E), which can accommodate the
bulk of the adduct-BP moiety and allow adduct-dG:dCTP pairing. The DNAP IV opening is large in part because of two glycines, G32 and G74, which lack R-group bulk (Figure 5.2B). In contrast, Dpo4, which is in the η/V-class, has two bulky aa in the equivalent positions (V32/M76) that plug the opening making it smaller (Figure 5.2A), such that the adduct-BP structure is deformed (65,67), this deformation can rationalize why dCTP insertion is slower for Dpo4 than DNAP IV. In Chapter 4, we showed that the rate of dCTP insertion was greater for the double mutant V32G/M76G-Dpo4 than for Dpo4, which indicates the significance of these glycines. However, dCTP insertion rate is slower for V32G/M76G-Dpo4 than for DNAP IV, showing that other structural elements are also important.

In Chapter 4, we showed that dATP and dTTP misinsertion also occurs from conformations with adduct-BP bulk being in in the minor groove, though these minor groove conformation(s) are different than the minor groove conformation for dCTP insertion. In contrast, evidence presented in Chapter 4 suggested that adduct-BP moiety bulk is in the major groove during dGTP misinsertion. We also showed that the formation of non-covalent bridges (NCBs) between the LF-domain and the TPF-domain help suppress dATP, dGTP and dTTP misinsertions opposite BP-adducts. Herein, we investigate more closely the nature of two non-covalent bridges in DNAP IV.

Y-Family DNAPs have a loop above their active site (AS-loop), which is shown from the major groove side for Dpo4 and DNAP IV in Figure 5.5. This loop begins on the minor groove side (yellow amino acids in Figure 5.5) and loops back to the major groove side (brown amino acids in Figure 5.5). The AS-loop contains a number of
important structural features, such as the amino acid that sits above the templating base (A42 in Dpo4 and V40 in DNAP IV) and the amino acid that sits above the base of the dNTP (A44 in Dpo4 and S42 in DNAP IV). R37 and R38 are at the tip of DNAP IV’s AS-loop, and we present evidence that they interact with E251 and D252, respectively, in the LF-domain to form an active site non-covalent bridge (AS-NCB). When these Coulombic interactions are lost via mutations (e.g., with R37A- and R38A-DNAP IV), dATP, dGTP and dTTP misinsertion rates increase opposite BP-adducts, while correct dCTP insertion is relatively unaffected. The structure of the AS-NCB in DNAP IV is compared to the AS-NCB in Dpo4, which depends on R36:N252 and K250:R36(C=O) interactions and is much less efficient at suppressing dATP, dGTP and dTTP misinsertions.

While our work was in progress, ten X-ray structures of DNAP IV with undamaged DNA were published (76) which show the R38/D252 interactions. However, R37 seems to be interacting with solvent in the minor groove and not with E251. Our case for the R37/E251 interaction is strong, and we discuss possible sources for this discrepancy, such as the possibility that R37/E251 exists during bypass of –BP-dG but not with undamaged DNA.

We also present evidence that DNAP IV has a second non-covalent bridge, which is on the minor groove side and ~8Å distal from the active site (Distal-NCB) and also suppresses dATP, dGTP and dTTP misinsertions. Dpo4 lacks a distal-NCB, which contributes to Dpo4 being less efficient at suppressing dATP, dGTP and dTTP misinsertions.
RESULTS

The *dpo4* and *dinB* (for DNAP IV) genes were incorporated into vectors that included a N-terminal His-Tag to facilitate protein purification (Materials and Methods). Primer extension reactions were performed using a 20-mer $^{32}$P-primer annealed to a template containing a –BP-dG adduct (Chapter 2). DNAPs were preincubated with the primer/template at 37°C for >5 minutes, and polymerization initiated by dNTP addition. Reactions were stopped, separated on a DNA sequencing gel, and--following phosphorimaging (Figure 5.2)--individual reactions were scanned and quantitated (Chapter 3). Polymerization often proceeded beyond the first insertion; e.g., Dpo4 inserts dATP opposite –BP (I), and then extends by adding dATP opposite the 5’-T (E1) and the next C-base (E2). For kinetic analysis, the intensity of all extension products was summed (I + E1 + E2) and compared to remaining primer to determine kinetics. All experiments used the same concentration of dNTP (100µM) and DNAP (50nM) unless noted otherwise (Chapter 3), and Table 1 lists observed rate constants ($k_{obs}$) in units of min$^{-1}$ for a variety of wild type and mutant DNAPs. These data are biologically relevant given that [dNTP] concentrations in cells are in the 10-100µM range. dNTP concentrations of 100µM are below Km, which is in the ~1mM range, and, accordingly, kinetic data are most relevant to Vmax/Km considerations. Determining Km accurately was not feasible, because dNTP insertion rate becomes progressively inhibited at higher [dNTP] (data not shown).

In DNAP IV, R37 and R38 Suppresses dGTP Misinsertion Opposite –BP-dG
Numerous Dpo4 X-ray structures show that the TPF-domains and LF-domain interact via a non-covalent bridge (Figure 5.1A), which forms near the active site (AS-NCB). This AS-NCB involves a Coulombic interaction between the guanidinium moiety of R36, which is in a loop near the active site (AS-loop), and N254 in the LF-domain (Figure 5.4). In most X-ray structures, Dpo4’s AS-NCB is reinforced by a second Coulombic interaction between the amino-moiety of K252 and the backbone carbonyl of R36. In contrast, DNAP IV has A248 in the position equivalent to K252, which suggests that the AS-NCBs in DNAP IV and Dpo4 must be different. While Dpo4 has a single arginine in its AS-loop (i.e., R36), DNAP IV has three (R35, R37 and R38), and we started by mutating each arginine to alanine to investigate their importance.

Both R37A-DNAP IV and R38A-DNAP IV showed an apparent increase in rate of dGTP misinsertion opposite –BP-dG; e.g., the intensity of the insertion band (I) in Figure 3 for dGTP misinsertion is greater for R37A-DNAP IV compared to wt-DNAP IV itself. Quantitatively, the rate of dGTP misinsertion was greater for R37A-DNAP IV (17 x 10^{-3} min^{-1}) and R38A-DNAP IV (16 x 10^{-3} min^{-1}) than for wt DNAP IV (0.63 x 10^{-3} min^{-1}). dCTP insertion was also greater for both R37A-DNAP IV (54 x 10^{-3} min^{-1}) and R38A-DNAP IV (35 x 10^{-3} min^{-1}) than for wt-DNAP IV (25.6 x10^{-3} min^{-1}), though the fold-change was smaller than for dGTP, showing that on balance the presence of R37 and R38 helps improve dGTP fidelity in DNAP IV. As discussed at greater length below, we believe that R37 and R38 interact with residues in the LF-Domain to secure the AS-NCB in DNAP IV. (Work on R35 is incomplete.)
To further investigate R37 and R38 and their role in the AS-loop, the entire AS-loop from DNAP IV (V31-V41) was inserted into Dpo4 to give IVloop-Dpo4. If the IVloop in the IVloop-Dpo4 context behaves the same as it does when IVloop is in DNAP IV itself, then dGTP misinsertion should increase when the R37A and R38A mutations are placed in IVloop-Dpo4, which was not the case: dGTP misinsertion was about the same for R37A-IVloop-Dpo4 (12 x 10$^{-3}$ min$^{-1}$) and R38A-IVloop-Dpo4 (9 x 10$^{-3}$ min$^{-1}$) as for IVloop-Dpo4 (12 x 10$^{-3}$ min$^{-1}$). These findings suggest that R37 and R38 in the IV-loop behave differently when they are interacting with the LF-Domain of Dpo4 vs. the LF-Domain of DNAP IV.

**Evidence that R37 Interacts with E251 in DNAP IV**

In Dpo4, the AS-NCB involves R36/N254 and K252/R36(C=O), while DNAP IV has A248 at the K252 position in Dpo4. Furthermore, DNAP IV has no other positively charged amino acids in the vicinity of A248. For a variety of reasons (below and in Discussion), we hypothesized that a R37/E251 interaction reinforces DNAP IV’s AS-NCB, which also includes R38/D252, where D252 in DNAP IV is in the analogous position to N254 in Dpo4.

First, Molecular modeling showed that the R37/E251 and R38/D252 interactions are indeed plausible in principle (Figure 5.4B).

The plausibility of R37/E251 and R38/D252 interactions was also evaluated by considering 434 DNAP IV sequences, which we compiled previously (80). In DNAP IV, R38 is highly conserved (421/434 of sequences or 97%), and a hydrogen bond-accepting
amino acid, is usually found at position 252, where D252 (77%) and N252 (11%) are prominent. Arginine is also the consensus amino acid at position 37 in DNAP IV, though it is less conserved (161/434 or 37%). Nevertheless, the data is revealing. [In the following analysis, only a subset of DNAP IVs are considered (Table 5.2).] When DNAP IV sequences have R37, 113 have the combination R37/E251 and six have R37/N251 or R37/Q251 (Table 5.2), all of which can interact via hydrogen bonding; in contrast, combinations that cannot form hydrogen bonds (e.g., R37/K251) are found in only 21 sequences for a positive interaction percentage of ~85% (= 119/140). (A variety of amino acids at position 251 are ambiguous and are considered “uncertain.”) Lysine can also be present at position 37 (Table 2), and combinations that can form hydrogen bonds (e.g., K37/D251, K37/E251 or K37/N251) outnumber those that cannot (27/34 = 79%). Interestingly, some DNAP IVs have a negatively charged amino acid (i.e., Asp or Glu) at position 37, in which case hydrogen bond donating amino acids are found at position 251 in a high percentage of cases (24/31 = 77%). Thus, a high fraction of DNAP IV sequences with D37 or E37 (instead of R37) have a reciprocal change, including R251, K251, Q251 or H251 (instead of E251).

If indeed an interaction between amino acids at positions 37/251 is important, then (e.g.) both single mutants R37E-DNAP IV and E251K-DNAP IV should have higher dGTP misinsertion rates opposite –BP-dG compared to Dpo4, while the double mutant R37E/E251K-DNAP IV should then decrease dGTP misinsertion rate compared to single mutant. This trend is indeed observed (Table 5.1). Compared to the dGTP misinsertion rate for DNAP IV (0.6 x 10^{-3} \text{ min}^{-1}), the rates are higher for R37E-DNAP IV
(17 x 10^{-3} \text{ min}^{-1}) and E251K-DNAP IV (16 x 10^{-3} \text{ min}^{-1}), and similar, as if both mutations lead to the loss of the same interaction. Furthermore, the rate for the double mutant R37E/E251K-DNAP IV (5.2 x 10^{-3} \text{ min}^{-1}) is ~3-fold lower than either single mutant. These results strongly suggest that a Coulombic interaction forms, which can be either R37/E251 in EcDNAP IV or E37/K251 in other DNAP IVs.

**The Effect of the R37/E251 and R38/D252 on dATP and dTTP Misinsertion**

**Opposite –BP-dG**

In addition to suppressing dGTP misinsertion, R37 and R38 also suppress dATP and dTTP misinsertions.

Compared to the dATP misinsertion rate with wt-DNAP IV (0.3 x 10^{-3} \text{ min}^{-1}), dATP misinsertion was higher for R37A-DNAP IV (9.7 x 10^{-3} \text{ min}^{-1}), R38A-DNAP IV (3.8 x 10^{-3} \text{ min}^{-1}), R37E-DNAP IV (14 x 10^{-3} \text{ min}^{-1}), and E251K-DNAP IV (11 x 10^{-3} \text{ min}^{-1}). Compared to the dTTP misinsertion rate with wt-DNAP IV (0.13 x 10^{-3} \text{ min}^{-1}), dTTP misinsertion was higher for R37A-DNAP IV (2.7 x 10^{-3} \text{ min}^{-1}), R38A-DNAP IV (10.4 x 10^{-3} \text{ min}^{-1}), R37E-DNAP IV (7.1 x 10^{-3} \text{ min}^{-1}) and E251K-DNAP IV (4.5 x 10^{-3} \text{ min}^{-1}).

Furthermore, the rate of dATP misinsertion for the double mutant R37E/E251K-DNAP IV (3.4 x 10^{-3} \text{ min}^{-1}) was similarly decreased with respect to both R37E-DNAP IV and E251K-DNAP IV (14 x 10^{-3} \text{ min}^{-1} and 11 x 10^{-3} \text{ min}^{-1}). In addition, the rate of dTTP misinsertion for R37E/E251K-DNAP IV (1.6 x 10^{-3} \text{ min}^{-1}) was similarly decreased
compared to both R37E-DNAP IV and E251K-DNAP IV (7.1 x 10^{-3} \text{ min}^{-1} and 4.5 \times 10^{-3} \text{ min}^{-1}).

Collectively, these findings suggest that the formation of an effective AS-NCB in DNAP IV is able to minimize aberrant structures from entering the active site, independent of whether adduct bulk is in the major, which is the case of dGTP misinsertion, or the minor groove, which is the case for dATP and dTTP misinsertions.

In DNAP IV, a Distal Non-Covalent Bridge Helps Suppress dATP/dGTP/dTTP Misinsertions

Dbh is an *Archaea* Y-Family DNAP that inserts dCTP opposite –BP with relatively high fidelity (Chapter 4). Dbh X-ray structures with DNA show a non-covalent bridge between the LF-domain and the TPF-domains that is on the minor groove side, that is ~8Å distal from the active site (Distal-NCB) and that involves a hydrophobic interaction. All Dpo4 X-ray structures show a single AS-NCB and no Distal-NCB. This prompted us to test whether the Dbh Distal-NCB might affect fidelity, and we showed that it suppresses dATP, dGTP and dTTP misinsertion opposite –BP-dG (Chapter 4). We speculated that the Distal-NCB also helps to keep the LF-domain positioned in order to exclude certain aberrant –BP-dG adduct structures from entering the active site, from which misinsertion can occur.

To make our homology model of DNAP IV, we began with a Dpo4 X-ray structure, and in every molecular dynamics trajectory the final DNAP IV structure retained an AS-NCB (Figure 5.1B), which depends on the R37/E251 and R38/E252
interactions based on the data described above. In addition, in every trajectory a second non-covalent bridge formed in DNAP IV on the minor groove side of the active site in a position analogous to the Distal-NCB in Dbh. This Distal-NCB in DNAP IV depends on P99/L100. The P99 proline kinks the amino acid backbone and thrusts this region of the protein ~3 Å closer to the LF-Domain of DNA IV compared to the analogous region of Dpo4, which has I101/A102 in the equivalent positions. The bulky hydrophobic R-group of L100 in DNAP IV appears important for filling a hydrophobic pocket that forms because of this kink.

To test the importance of P99/L100, the mutant P99I/L100A-DNAP IV was studied. Considering dGTP misinsertion, the rate for wt-DNAP IV (0.63 x 10^{-3} \text{ min}^{-1}) is much less than P99I/L100A-DNAP IV (11 x 10^{-3} \text{ min}^{-1}). The P99I/L100A double mutation also increases dATP misinsertion (0.29 x 10^{-3} \text{ min}^{-1} vs. 5.0 x 10^{-3} \text{ min}^{-1}) and dTTP misinsertion (0.13 x 10^{-3} \text{ min}^{-1} vs. 2.2 x 10^{-3} \text{ min}^{-1}). The rate of dCTP insertion was also greater for both P99I/L100A-DNAP IV (39 x 10^{-3} \text{ min}^{-1}) than for wt-DNAP IV (26 x10^{-3} \text{ min}^{-1}), though the fold-change was much smaller for dCTP insertion, showing that on balance the presence of P99/L100 helps improve fidelity in DNAP IV.

This region was also evaluated in the Dpo4 context. The reverse mutation I101P/A102L-Dpo4 decreased the misinsertion rates compared to wt-Dpo4 by ~2.7-fold for dATP, ~2.4-fold for dGTP and ~1.4-fold for dTTP.

All of these findings show that the presence of P99/I100 tends to suppress dATP, dGTP and dTTP misinsertions, and a plausible structural mechanism is the formation of a Distal-NCB.
DISCUSSION

BP-N²-adducts have an intact hydrogen bonding face, so they can in principle base pair with dCTP. However, the BP-moiety itself is large, and it is reasonable to imagine that any DNAP that could correctly incorporate dCTP must have special structural elements to accommodate this bulk. We have pointed out that the BP-moiety would have to protrude from an opening on the minor groove side of the active site in order to form a more-or-less normal Watson-Crick base pair, since the N2-position is in the minor groove, and, thus, a large opening on the minor groove side of the active site should facilitate this (65). In fact, the opening on the minor groove side of the active site is relatively large in κ/IV-class DNAPs, such as EcDNAP IV, which has a hole (Figure 5.1E, (65,76)), which has a slot (Figure 5.1C). In contrast, the minor groove opening is plugged in η/V-class DNAPs; e.g., by V32 and M76 in Dpo4 (Figure 5.2A), and by large aa-inserts in hDNAP η (aa81-87) and ScDNAP η (aa93-126). The work described in Chapter 4 suggested that a larger opening on the minor groove side of the active site does indeed enhance the dCTP insertion opposite –BP-dG.

Regarding misinsertion, two structural elements seem important. (1) Protein structural features that plug the opening on the major groove side of the active site suppress dNTP misinsertions when a lesion is present on the major groove side, as has been investigated in the case of the N-clasp of DNAP κ (86-88), and as we have shown for dGTP misinsertion opposite –BP-dG, which has bulk in the major groove based on evidence presented in Chapter 4. (2) Non-covalent bridges between the TPF-domains and the LF-domain suppress dNTP misinsertions by excluding aberrant conformations
from the active site, whether adduct bulk is in the major groove (dGTP with –BP-dG), or
the minor groove (dATP and dTTP with –BP-dG).

The first set of experiments described in Chapter 5 were focused on trying to
understand the nature of the AS-NCB in DNAP IV in comparison to Dpo4. The AS-
NCB in Dpo4 includes an interaction between R36 in the TPF-Domains and N254 in the
LF-Domain. While our work was in progress, ten X-ray structures of DNAP IV with
undamaged DNA emerged and all ten showed a R38/D252 interaction that stabilized its
AS-NCB. D252 in DNAP IV is in the analogous position to N254 in Dpo4. As will be
discussed below, R38 in DNAP IV is about one amino acid closer to the major groove
than R36 in Dpo4 (Figure 5.4). The AS-NCB in Dpo4 is reinforced by an interaction
between K252 and R36(C=O). Where Dpo4 has K252, DNAP IV has A248, and,
furthermore, no other positively charged amino acids are in the vicinity of A248 in the
LF-Domain of DNAP IV. In fact, in all ten X-ray structures R38/D252 is the sole
interaction in the AS-NCB in DNAP IV. In all ten X-ray structures, R37 appears to be
interacting with solvent above the minor groove and is not interacting with E251 as we
hypothesize. Nevertheless, for the following reasons we believe that the R37/E251
interaction is important during –BP-dG bypass.

(1) Though the R37/E251 interaction is not present in the X-ray structures, it can
form in principle based on molecular modeling studies (Figure 5.4B). Furthermore, an
R37/E251 interaction can be formed via dihedral bond rotations in some of the DNAP IV
X-ray structures (Data not shown).
(2) Both R37A-DNAP IV and R38A-DNAP IV misinsert dGTP at similar rates ($17 \times 10^{-3}$ min$^{-1}$ and $16 \times 10^{-3}$ min$^{-1}$), which are both greater than with wt-DNAP IV ($0.6 \times 10^{-3}$ min$^{-1}$), which suggests that dGTP misinsertion is suppressed by some structural element that each of these arginines can provide with approximately equal ability. A sensible possibility is that the AS-NCB depends upon two similar Coulombic interactions R37/E251 and R38/D252.

(3) In its AS-loop, Dpo4 has only one arginine (R36), which interacts with N254 in the LF-Domain. The fact that the rate of misinsertion was similar for R37A-IVloop-Dpo4 ($12 \times 10^{-3}$ min$^{-1}$) and IVloop-Dpo4 ($12 \times 10^{-3}$ min$^{-1}$) suggests that R38 alone in R37A-IVloop-Dpo4 is capable of interacting with N254 in the Dpo4 LF-Domain. Likewise, the fact that misinsertion rate was similar for R38A-IVloop-Dpo4 ($7 \times 10^{-3}$ min$^{-1}$) and IVloop-Dpo4 ($12 \times 10^{-3}$ min$^{-1}$) suggests that R37 alone in R38A-IVloop-Dpo4 is also capable of interacting with N254 in the Dpo4 LF-Domain. Collectively, these findings suggest that R37 is as capable as R38 to form a Coulombic interaction with a residue (i.e., N254) in the LF-Domain. In fact, R37 in DNAP IV is actually closer than R38 to being in the same position as R36 in Dpo4 (Figure 4).

(4) We analyzed 434 DNAP IV sequences, which we compiled previously (80). We noticed that a number of DNAP IVs had either Asp or Glu at position 37, in which case position 251 had a hydrogen bond donating amino acid in a high percentage of cases (77%). This analysis suggested that--if the 37/251 positions interact--then dGTP misinsertion rate should increase for both R37E-DNAP IV and E251K-DNAP IV, which was the case ($17 \times 10^{-3}$ min$^{-1}$ and $16 \times 10^{-3}$ min$^{-1}$), while it should then decrease for the
double mutant R37E/E251K-DNAP IV (5.2 x 10^{-3} \text{ min}^{-1}), which was also the case. These results strongly suggest that a Coulombic interaction does form, which can be R37/E251 in many DNAP IVs (e.g., from \textit{E. coli}) or E37/K251 in other DNAP IVs.

These findings strongly implicate the formation of an interaction between R37 in the TPF-Domains and E251 in the LF-Domain of DNAP IV. How can this finding be reconciled with the ten X-ray structures of DNAP IV, all of which show R37 interacting with solvent in the minor groove (76). (1) The most obvious possibility relates to the fact that the DNAP IV X-ray structures are with undamaged DNA, while our findings are for –BP-dG adduct bypass, which might require a conformational change that either allows or requires R37 to interact with E251. (2) X-ray structures reveal information about ground state structures, while structure-activity approaches, which include studies on how amino acid substitutions affect reaction rates, reveal information about transition state structures (69). Ground state and transition state structures could be different. (3) X-ray coordinates show a unit cell and do not show if a particular surface residue might be interacting with a residue in a neighboring unit cell, which might be the case for R37 in DNAP IV.

The dGTP misinsertion rate is remarkably similar for the DNAP IV derivatives: R37A-DNAP IV (17.3 x 10^{-3} \text{ min}^{-1}), R37E-DNAP IV (17.4 x 10^{-3} \text{ min}^{-1}) and E251K-DNAP IV (16.1 x 10^{-3} \text{ min}^{-1}), which is sensible given that each of these mutations is hypothesized to result in the loss of the same interaction: R37/E251. Furthermore, the dGTP misinsertion rate for R38A-DNAP IV (16.2 x 10^{-3} \text{ min}^{-1}) is also similar,
presumably because of the loss of the R38/D252 interaction. In each of these four cases, the supposition is that one of two Coulombic interactions is lost, leaving only one.

The dGTP misinsertion rates for the Dpo4 derivatives: IVloop-Dpo4 (12 x 10^{-3} min^{-1}), R37A-IVloop-Dpo4 (12 x 10^{-3} min^{-1}) and R38-IVloop-Dpo4 (7.0 x 10^{-3} min^{-1}) are all similar to each other. Furthermore they are similar to the rates for the four DNAP IV derivatives mentioned in the previous paragraph (i.e., R37A-DNAP IV, R38A-DNAP IV, R37E-DNAP IV and E251R-DNAP IV). It is as if all seven of these proteins have a single Coulombic interaction in their AS-NCB between their TPF-Domains and their LF-Domain, which makes sense for the following reason. Remembering that Dpo4 has a R36/N254 Coulombic interaction, it seems likely that R37A-IVloop-Dpo4 has R38 in its IVloop interacting with N254 in the LF-Domain of Dpo4, while R38A-IVloop-Dpo4 has R37 in its IVloop interacting with N254 in the LF-Domain of Dpo4. The fact that IVloop-Dpo4, which has both R37 and R38, has virtually the same dGTP rate (7.0 x 10^{-3} min^{-1}) as these seven mutants suggests that either R37 or R38, but not both, can form one Coulombic interaction with a residue in the LF-Domain (presumably N254). There is one wrinkle in this analysis. We also proposed that DNAP IV has a Distal-NCB, which is not present in Dpo4. Why then does this putative Distal-NCB not result in a lower dGTP misinsertion rate for the four DNAP IV derivatives (i.e., R37A-DNAP IV, R38A-DNAP IV, R37E-DNAP IV and E251R-DNAP IV, compared to the three Dpo4 mutants (i.e., IVloop-Dpo4, R37A-IVloop-Dpo4 and R38-IVloop-Dpo4)? This can be reconciled if the Distal-NCB cannot form in DNAP IV unless the AS-NCB is secure. Additional experiments will have to be done to investigate this possibility.
The double mutant R37E/E251K-DNAP IV does not decrease dATP, dGTP and dTTP misinsertion as well as wt-DNAP IV, which suggests that R37/E251 functions better than E37/K251 in the *E. coli* DNAP IV context. It is worth noting that DNAP IVs with E37/K251 have other differences in the vicinity of both position 37 and position 251.

<table>
<thead>
<tr>
<th>Amino acids near the tip of</th>
<th>AS-Loop in</th>
<th>TPF-Domain</th>
<th>LF-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> DNAP IV</td>
<td>RERRG</td>
<td>MAEDI</td>
<td></td>
</tr>
<tr>
<td>Other DNAP IVs</td>
<td>-GERG</td>
<td>LKKEI</td>
<td></td>
</tr>
</tbody>
</table>

DNAP IVs with E37/K251 typically are missing an amino acid at position 35 and also contain a glycine. There are also differences in the LF-Domain (e.g., A vs. K at position 250). These other changes might help DNAP IVs containing E37/K251 further suppress dATP, dGTP and dTTP misinsertions.

Findings described above suggested that DNAP IV forms a Distal-NCB, which is dependent on a kink induced by P99. In our model, this Distal-NCB is formed by R240 in the LF-Domain reaching across to the TPF-Domains to form the Distal-NCB (Figure 5.2B). In contrast, the equivalent residue in Dpo4 (R242) interacts with template-DNA. In the ten X-ray structures of DNAP IV (76), R240 adopts seven different positions. In three structures, R240 interacts with both template-DNA and P99(C=O). In two structures, R240 interacts with E98 and P99(C=O). In one structure each, R240 interacts with either: P99(C=O), E237, R240(C=O), template-DNA or E98/P99(C=O)/template-DNA. Thus, these X-ray structures show heterogeneity for R240 positioning, but in 8/10
cases some kind of Distal-NCB is evident. If R240 in DNAP IV is involved in the Distal-NCB and not interacting with DNA, then DNAP IV would have lost one DNA-interacting arginine and a compensatory DNA interaction might have evolved. Indeed, we projected that R285 in DNAP IV should be able to interact with DNA given that it is in the analogous position to R283 in Dbh, which is known to interact with DNA and, presumably, helps to compensate for the loss of a DNA interacting residue because of the Dbh Distal-NCB (discussed in Chapter 4). Dpo4 has A283 at the equivalent position. In 9/10 DNAP IV X-ray structures, R285 does interact with DNA.

While both DNAP IV and Dpo4 appear to have an AS-NCB to suppress dNTP misinsertions, DNAP IV’s AS-NCB bridge seems to be more effective, as can be assessed by comparing findings for P99I/L100A-DNAP IV, which has its AS-NCB intact but lacks its Distal-NCB, to Dpo4, which has only an AS-NCB. P99I/L100A-DNAP IV has a lower rate for dATP (~2-fold), for dGTP (~5.7-fold) and dTTP (~2.2-fold) compared to Dpo4. In fact the fold-decreases in the rate of dATP and dTTP misinsertions are lower limits, given that we showed that the larger opening on the minor groove side of the active site in DNAP IV enhances dATP and dTTP ~2-fold each (Chapter 4). Why might the AS-loop in DNAP IV be better at suppressing dNTP misincorporation? The R37/E251 and R38/D252 interactions in the AS-NCB of DNAP IV are shifted more toward the major groove side of the active site compared with the Dpo4 AS-NCB, which depends on R36:N252 and K250:R36(C=O) (Figure 5.4), which might be germane. We note that the only non-covalent bridge between the TPF-Domians and the LF-Domain in DNAP κ involves interactions between residues in the N-clasp, which
is on the major groove side of the active site (Figure 5.1C); i.e., DNAP κ has no non-
covalent bridges in its minor groove, and dATP, dGTP and dTTP misinsertions are
relatively slow (88).

In summary, DNAP IV has four structural elements that improve fidelity for
dNTP incorporation opposite –BP-dG compared to Dpo4. (1) DNAP IV has a relatively
larger opening on the minor groove side of the active site that facilitates correct dCTP
insertion, because the BP-moiety must be in the minor groove for Watson-Crick-like
adduct-dG:dCTP pairing. (2) DNAP IV has several structural features compared to Dpo4
that plug the major groove side of the active site (i.e., V40 instead of A42 and a smaller
AS-loop), which suppresses dNTP misinsertions when adduct bulk is present on the
major groove side, such as during dGTP misinsertion with –BP-dG. (3) DNAP IV has a
Distal-NCB, which depends on P99/L100, that is lacking in Dpo4. (4) The AS-NCB in
DNAP IV, which involves R37/E251 and R38/D252, appears to suppress dNTP
misincorporation better than the AS-NCB in Dpo4, which depends on R36:N252 and
K250:R36(C=O). Finally, we note that these non-covalent bridges between the TPF-
domains and the LF-domain suppress dATP and dTTP misinsertions, which have –BP-
dG bulk in the minor groove, as well as dGTP misinsertion, which has –BP-dG bulk in
the major groove; this suggests that NCBs act by excluding aberrant conformations from
the active site, whether adduct bulk is in the minor or major groove.
Table 5.1: Relative Velocity of dNTP insertion opposite –BP-dG by Dpo4 and DNAP IV, as well as Mutant-DNAP IVs.¹

<table>
<thead>
<tr>
<th>Modification</th>
<th>Derivative of dATP</th>
<th>Velocity (x 10⁻³ min⁻¹)</th>
<th>dCTP</th>
<th>dGTP</th>
<th>dTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>Dpo4</td>
<td>10.6¹</td>
<td>5.8</td>
<td>63.5</td>
<td>4.9</td>
</tr>
<tr>
<td>I99P/A100L</td>
<td>Dpo4</td>
<td>4.0 (0.37)</td>
<td>4.4  (0.76)</td>
<td>26.2 (0.41)</td>
<td>3.5 (0.73)</td>
</tr>
<tr>
<td>wt</td>
<td>DNAP IV</td>
<td>0.29 (0.028)²</td>
<td>25.6 (4.38)</td>
<td>0.63 (0.010)</td>
<td>0.13 (0.026)</td>
</tr>
<tr>
<td>R37A</td>
<td>DNAP IV</td>
<td>9.7 (0.91)</td>
<td>53.6 (9.19)</td>
<td>17.3 (0.27)</td>
<td>2.71 (0.55)</td>
</tr>
<tr>
<td>R38A</td>
<td>DNAP IV</td>
<td>3.8 (0.36)</td>
<td>35.1 (6.02)</td>
<td>16.2 (0.26)</td>
<td>10.4 (2.14)</td>
</tr>
<tr>
<td>R37E</td>
<td>DNAP IV</td>
<td>14.0 (1.3)</td>
<td>41.2 (7.06)</td>
<td>17.4 (0.27)</td>
<td>7.08 (1.45)</td>
</tr>
<tr>
<td>E251K</td>
<td>DNAP IV</td>
<td>10.9 (1.03)</td>
<td>34.9 (6.0)</td>
<td>16.1 (0.25)</td>
<td>4.51 (0.92)</td>
</tr>
<tr>
<td>R37E/E251K</td>
<td>DNAP IV</td>
<td>3.4 (0.32)</td>
<td>36.5 (6.3)</td>
<td>5.2 (0.082)</td>
<td>1.6 (0.33)</td>
</tr>
<tr>
<td>*P99I/L100A</td>
<td>DNAP IV</td>
<td>5.0 (0.47)</td>
<td>39.53 (6.72)</td>
<td>11.0 (0.17)</td>
<td>2.2 (0.45)</td>
</tr>
</tbody>
</table>

¹ Rates are for dNTP insertion opposite a –BP-dG adduct in a primer/template using methods described in Chapter 3. Purified proteins included Dpo4 and DNAP IV, along with six mutant-DNAP IVs. Reactions were pseudo-first order (Chapter 3) and (e.g.) the number 10.6 for Dpo4 with dATP means the velocity 10.6 x 10⁻³ min⁻¹ using standard conditions, which were dNTP concentration [100µM] and protein concentration 50ng/mL, except in the following cases. When a value for dATP/dGTP/dTTP was <2.0 x 10⁻³ min⁻¹ higher protein concentrations were used and corrected based on the observed rate of the dCTP reaction compared to the typical dCTP rate. For example, the value for dGTP/DNAP IV was actually 4.69 x 10⁻³ min⁻¹ higher in an experiment in which dCTP/DNAP IV was 190 x 10⁻³ min⁻¹, and, thus, 0.63 x 10⁻³ min⁻¹ = 4.69 (25.6/190) x 10⁻³ min⁻¹.

² Numbers in parenthesis are the velocity (non-parenthesis on the left) divided by the velocity of the corresponding Dpo4 reaction; e.g., for dATP/R37A-DNAP IV is 0.020 = (0.21/10.5).
Table 5.2: Analysis of 434 DNAP IV Sequences with Either R37, K37 or D/E37 in Combination with Either R38/D252 or R38/N252

<table>
<thead>
<tr>
<th>Arg at 37</th>
<th>Lys at 37</th>
<th>Asp or Glu at 37</th>
<th>Gln at 37</th>
</tr>
</thead>
<tbody>
<tr>
<td>37/251</td>
<td>30/252</td>
<td>37/251</td>
<td>30/252</td>
</tr>
<tr>
<td>Positive Interactions (Green)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>E</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>R</td>
<td>Q</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Negative Interactions (Red)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>G</td>
<td>K</td>
<td>R</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Uncertain Interactions (Yellow)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>S</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>R</td>
<td>T</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

**Percentage Positive**

<table>
<thead>
<tr>
<th>37/251</th>
<th>30/252</th>
<th>37/251</th>
<th>30/252</th>
<th>37/251</th>
<th>30/252</th>
</tr>
</thead>
<tbody>
<tr>
<td>(119/140)</td>
<td>84%</td>
<td>(27/34)</td>
<td>79%</td>
<td>(24/31)</td>
<td>77%</td>
</tr>
</tbody>
</table>

1 434 DNAP IV sequences were collected from literature as reported in reference (80) which are analyzed for potential evidence that positions 37/251 might interact.
LEGENDS

Figure 5.1: Views of Y-Family DNAPs SsDpo4 (Panel A), hDNAP κ (Panel C) and SsDbh (Panel D), based on X-ray structures (31, 37, 96, respectively), along with DNAP IV (Panel B), based on a homology model (65). The view is from the minor groove and the TPF-Domains are shown in green and the LF-Domain is shown in yellow. Noncovalent bridges form between the TPF-Domains and the LF-Domain, which can be of three types: near the active site (AS-NCB) and distal to the active site on the minor groove side, or on the major groove side near the active site that is present uniquely in hDNAP κ, because of its N-clasp (scarlet structure). These NCBs are shown pictorially in Panel F. Panel E contains another representation of DNAP IV showing its minor groove opening, which is large enough to accommodate the BP moiety of BP-dG adducts.

Figure 5.2: The opening on the minor groove side near the active site for Dpo4 (Panel A) and DNAP IV (Panel B), showing a key difference: the Dpo4 opening is plugged by M75 and V32, while the DNAP IV opening is more spacious because its equivalent amino acids are glycines (G73 and G32).

Figure 5.3: Structures of +BP-dG and –BP-dG, along with the primer/template sequence and a phosphorimage showing dNTP incorporation by the Y-Family DNAPs EcDNAP IV and SsDpo4, along with three DNAP IV derivatives (R35A-, R37A- and R38A-DNAP IV), which are explained in the text.
Figure 5.4: Amino acids in the active site loop (AS-loop) of Dpo4 (Panel A) and DNAP IV (V31-V40, Panel B). The view is from the major groove, with the template base and dNTP included for reference. The AS-loops in Y-Family DNAPs begin on the minor groove side of the active site (yellow amino acids) and then turns back to the major groove side of the active site (brown amino acids). Certain key amino acids in the AS-loop are shown as described in the text.

Figure 5.5: Pictorially presentation of the impact on dNTP incorporation rate when various amino acid changes are made in a variety of Y-Family DNAPs. The plots show data after normalization of rate by dividing [mutant-Dpo4/wt-Dpo4] for different dNTP. For example, the average rate of dGTP misinsertion was 0.0169 min$^{-1}$ for R37A-Dpo4 and was 0.0635 min$^{-1}$ for wt-Dpo4, so normalized dGTP insertion for R37A-Dpo4 was ~0.27 (= 0.00169/0.0681). Panel A shows normalized dCTP misinsertions versus normalized dGTP misinsertion (as a Log-Log plot) for a variety of mutant-DNAP IVs, using data from Table 5.1, as well as Dpo4 (and derivatives). Symbols change for Dpo4 (closed circles), DNAP IV (open circles) and Dbh (open square). The point for Dpo4 itself is by definition at the origin (0, 0). Log-Log plots for dCTP versus dATP (Panel B) and dCTP versus dTTP (Panel C) are also shown.
Figure 5.1: Views of Y-Family DNAPs SsDpo4 (Panel A), hDNAP κ (Panel C) and SsDbh (Panel D), based on X-ray structures (31, 37, 96, respectively), along with DNAP
IV (Panel B), based on a homology model (65). The view is from the minor groove and the TPF-Domains are shown in green and the LF-Domain is shown in yellow. Noncovalent bridges form between the TPF-Domains and the LF-Domain, which can be of three types: near the active site (AS-NCB) and distal to the active site on the minor groove side, or on the major groove side near the active site that is present uniquely in hDNAP κ, because of its N-clasp (scarlet structure). These NCBs are shown pictorially in Panel F. Panel E contains another representation of DNAP IV showing its minor groove opening, which is large enough to accommodate the BP moiety of BP-dG adducts.
Figure 5.2: The opening on the minor groove side near the active site for Dpo4 (Panel A) and DNAP IV (Panel B), showing a key difference: the Dpo4 opening is plugged by M75 and V32, while the DNAP IV opening is more spacious because its equivalent amino acids are glycines (G73 and G32).
Figure 5.3: Structures of +BP-dG and –BP-dG, along with the primer/template sequence and a phophorimage showing dNTP incorporation by the Y-Family DNAPs EcDNAP IV and SsDpo4, along with three DNAP IV derivatives (R35A-, R37A- and R38A-DNAP IV), which are explained in the text.
Figure 5.4: Amino acids in the active site loop (AS-loop) of Dpo4 (Panel A) and DNAP IV (V31-V40, Panel B). The view is from the major groove, with the template base and dNTP included for reference. The AS-loops in Y-Family DNAPs begin on the minor groove side of the active site (yellow amino acids) and then turns back to the major
groove side of the active site (brown amino acids). Certain key amino acids in the AS-loop are shown as described in the text.
Figure 5.5: Pictorially presentation of the impact on dNTP incorporation rate when various amino acid changes are made in a variety of Y-Family DNAPs. The plots show
data after normalization of rate by dividing [mutant-Dpo4/wt-Dpo4] for different dNTP. For example, the average rate of dGTP misinsertion was 0.0169 min\(^{-1}\) for R37A-Dpo4 and was 0.0635 min\(^{-1}\) for wt-Dpo4, so normalized dGTP insertion for R37A-Dpo4 was ~0.27 (= 0.00169/0.0681). Panel A shows normalized dCTP misinsertions versus normalized dGTP misinsertion (as a Log-Log plot) for a variety of mutant-DNAP IVs, using data from Table 5.1, as well as Dpo4 (and derivatives). Symbols change for Dpo4 (closed circles), DNAP IV (open circles) and Dbh (open square). The point for Dpo4 itself is by definition at the origin (0, 0). Log-Log plots for dCTP versus dATP (Panel B) and dCTP versus dTTP (Panel C) are also shown.
CHAPTER 6
Perspective and Future Directions

Summary of Findings

In this thesis, we took a structure-activity approach to examine the structural features of the Y-family DNAPs SsDpo4, SsDbh and EcDNAP IV in an attempt to understand what controls both mutagenic and non-mutagenic insertion opposite –BP-dGTP. In particular the structure-activity approach involved making modifications (amino acid substitutions) in these three proteins and determining how these changes affected the rates of correct dCTP insertion, as well as incorrect dATP, dGTP and dTTP misinsertion opposite –BP-dG. Structure-activity studies are the only way to gain insight into the structure of transition states, which in turn determine why a reaction has the kinetic rates that it does (69).

Our kinetic analysis relied on being able to accurately assess the amount of primer extension that a particular DNAP could carry out on a –BP-dG-containing template, as assayed in a DNA sequencing gel, where the primer was 32P-radiolabeled. Ultimately, this required the quantitation of over-lapping, skewed-Gaussian curves. These kinetics could not have been assessed properly without developing a reliable and accurate method, as described in Chapter 3. This method relied on generating in each experiment a standard that showed the proper shape of a skewed-Gaussian for a single-component curve; purified 32P-primer alone was used as the standard. We showed that the method could accurately determine as little as ~2% primer extension. Furthermore, we showed that alternative approaches, which have become standard practice, are not accurate.
For our studies, the Y-Family DNAPs SsDpo4, SsDbh and EcDNAP IV were chosen, because they are the simplest of the well-studied Y-family DNAPs, where each contains only core amino acids (~350aa), and they share many structural features, making amino acid swaps feasible. Other Y-Family DNAPs, including EcUmuC, hDNAP k, hDNAP h, ScDNAP h and hDNAP i were deemed less appropriate for a variety of reasons as discussed in Chapter 4.

Y-Family DNAPs fall into two classes. The k/IV-class insert accurately opposite N²-dG adducts such as –BP-dG. The h/V-class inserts accurately opposite UV-damage but inaccurately opposite N²-dG adducts. SsDbh and EcDNAP IV are in the k/IV-class, while SsDpo4 is in the h/V-class. Simple amino acid sequence comparisons are not helpful in finding likely structural differences that give rise to functional differences, given that Dpo4 and Dbh, which are in different classes, are more similar (54% identity/79% conserved) than are DNAP IV and Dbh (28% identity/52% conserved), which are in the same class (90). Instead, we spent considerable time looking at X-ray structures and models of Y-Family DNAPs, searching for structural hypotheses that could be tested by doing amino acid swaps. In particular, we hypothesized that a particular amino acid (or collection of amino acids) in (e.g.) DNAP IV, when conferred on Dpo4, might make Dpo4 more DNAP IV-like in its properties toward insertion opposite –BP-dG. Dpo4 itself incorporates dNTPs in the rate order dGTP > dATP > dCTP ≈ dTTP opposite –BP-dG (Chapter 4), while both SsDbh and DNAP IV preferentially insert dCTP. Thus, we made amino acid substitutions in Dpo4 and assessed whether the dCTP insertion rate increased, or dATP, dGTP or dTTP misinsertion rate decreased (Chapter 4).
Similarly, we made amino acid subsitutions in DNAP IV and assessed whether dCTP insertion rate decreased or dATP, dGTP or dTTP misinsertion rate increased (Chapter 5).

Taking the approach outlined in the previous paragraph, we discovered that no single structural feature accounted for DNAP IV’s accuracy in –BP-dG bypass vs. Dpo4’s inaccuracy. Rather, there were a number of structural features that each improved accuracy by 2-5-fold, but that collectively could account for the observation that EcDNAP IV is ~100-fold more accurate than SsDpo4 in its fidelity of insertion opposite –BP-dG.

EcDNAP IV has four structural elements that improve the fidelity for dNTP incorporation opposite –BP-dG compared to Dpo4 as outlined next.

1. DNAP IV has a relatively large opening on the minor groove side of its active site that facilitates correct dCTP insertion, because the BP-moiety must be in the minor groove to form Watson-Crick-like adduct-dG:dCTP pairing, given that the N2-position of dG is in the minor groove (Chapter 4).

2. DNAP IV has several structural features compared to Dpo4 that plug the major groove side of the active site, and, thereby, suppress dNTP misinsertions that occur when an adduct has bulk on the major groove side, such as we showed was the case for dGTP misinsertion with –BP-dG (Chapter 4).

3. Y-Family DNAPs have three domains (thumb/palm/fingers), which operate as a unit (TPF-Domains), along with another domain, usually called the little finger domain (LF-Domain). The TPF-Domains and the LF-Domain are covalently connected by a long amino acid tether (~15aa). X-ray structures show that the TPF-Domains and the LF-
Domain each interact with DNA, which brings them together, such that non-covalent bridges (NCBs) between the TPF-Domains and the LF-Domains can form. We note that several X-ray structures of Y-Family DNAPs without DNA do not show the proper arrangement of the TPF-Domains and the LF-Domain (29). We showed that both EcDNAP IV and Dbh have non-covalent bridge on the minor groove side and ~8Å distal from the active site (Distal-NCB), whose presence leads to the suppression of dATP, dGTP and dTTP misinsertions (Chapters 4 and 5). Dpo4 lacks a distal-NCB, which contributes to Dpo4 being less efficient at suppressing dATP, dGTP and dTTP misinsertions opposite –BP-dG.

(4) All three DNAPs have a non-covalent bridge near their active site (AS-NCB) that also suppresses dATP, dGTP and dTTP misinsertions opposite –BP-dG (Chapters 4 and 5). Evidence was presented that the AS-NCB in DNAP IV was superior to the AS-NCB in SsDpo4 at suppressing dATP, dGTP and dTTP misinsertions.

We showed that the AS-NCB in DNAP IV requires two Coulombic interactions involving R37 and R38 in the TPF-Domains contacting E251 and D252, respectively, in the LF-Domain (Chapter 5). While the R38/D252 interaction is observed in the ten X-ray structures of DNAP IV (Rf5), the R37/E251 interaction is observed in none of them. However, the biochemical data supporting the R37/E251 interaction are compelling (Chapter 5), and we suggested three possible reasons for this discrepancy, the most obvious being that the X-ray structures contained undamaged DNA, while we studied insertion opposite a –BP-dG adduct, which--because of its bulk--almost certainly must cause changes in DNAP IV structure in the vicinity of the active site where the AS-NCB
is located. These findings show that, while X-ray structures are valuable, they cannot be viewed as providing definitive information about structures relevant to transition state structures in enzymatic reactions.

In the course of our work, a number of other mechanistic aspects of Y-family DNAPs were investigated, though none were sufficiently finished to be included as a separate chapters. These preliminary findings are summarized in the next four subsections of Chapter 6.

**Y-Family DNAP Insertion opposite the adduct +BP-dG**

In most of the experiments, side-by-side experiments were performed with +BP-N\(^2\)-dG (+BP-dG), whose structure is shown in Figure 3.1 in Chapter 3. +BP-G and –BP-dG are similar in a number of respects, including having the same atom composition, notably with the same bulky pyrenyl ring. However, the stereochemistry of their adduct bonds are opposite, as is the stereochemistry of the three hydroxyl groups. +BP-dG and –BP-dG are in fact almost mirror images. Though these stereochemical differences might seem insignificant, it is known that they affect how the adducts sit in ds-DNA. In NMR studies (98), both –BP-dG and +BP-dB have been shown to lie in the minor groove when paired with cytosine. However, the p-face of the pyrenyl moiety in -BP-dG lies flat against the minor groove and points toward the 3’-base, thereby causing minimal structural disturbance, while the pyrenyl moiety in +BP-dG sits almost perpendicular to the minor groove and protrudes into solvent. These structural differences are imposed by the stereochemistry of the adduct bond, and the bulky nature of the pyrenyl moiety does
not allow flexibility of positioning. Molecular modeling studies have shown that Y-family DNAPs face different challenges during the bypass of -BP-dG vs. +BP-dG adducts (65,90).

While our experiments with +BP-dG are not completed, we can describe some of the emerging patterns.

First, SsDpo4 and its derivatives in general incorporate dNTPs more slowly opposite +BP-dG than –BP-dG. For example, to get comparable dNTP rates, 5-fold more [DNAP] had to used with +BP-dG than with –BP-dG. The relative rates with SsDpo4 is dATP ≈ dGTP > dCTP ≈ dTTP opposite +BP-dG; i.e., +BP-dG does not show dGTP having a much greater misinsertion rate as was observed with –BP-dG, whose order was dGTP > dATP > dCTP ≈ dTTP. Chapter 4 shows a particular kind of plot involving normalization to the Dpo4 reaction (Figures 4.3 and 4.4) for insertions opposite –BP-dG. Figure 4.3B showed a correlation between dATP and dTTP misinsertions suggesting a similar adduct structure during misinsertion. In contrast, with +BP-dG the correlation is between dATP and dGTP misinsertion rates, suggesting a similar adduct structure during their misinsertions opposite +BP-dG.

In the case of EcDNAP IV, to get comparable rates of dATP, dGTP and dTTP misinsertion also required 5-times more protein with +BP-dG than is required with–BP-dG, which indicates that EcDNAP IV suppresses dATP, dGTP and dTTP misinsertion more effectively with +BP-dG than with –BP-dG. This result suggests that EcDNAPs can suppress the use of aberrant structures, which lead to dATP, dGTP and dTTP misinsertion, more effectively with +BP-dG than with –BP-dG. Stated in another way,
this implies that the AS-NCB and Distal-NCB in EcDNAP IV suppress incorrect dNTP misinsertion better with +BP-dG than with –BP-dG. In contrast, the rate of dCTP insertion opposite +BP-dG and –BP-dG is approximately equal when the same amount of EcDNAP IV is used. This implies that even though the pyrenyl moiety in the structure with +BP-dG paired with a dC moiety looks more problematic than for –BP-dG paired with a dC moiety, DNAP IV protein structure is able to accommodate each adduct approximately equally well. Stated in another way, this implies that the opening on the minor groove side of the active site in EcDNAP IV is able to accommodate the pyrenyl moiety equally well with +BP-dG and –BP-dG, in spite of their dramatically different conformations.

When the opening on the minor groove side of the active site of Dpo4 is enlarged (e.g., with V32G/M76G-Dpo4, IIVLoop-Dpo4, and kLoop-Dpo4) the rate of dCTP and dTTP incorporation opposite +BP-dG increased compared to wt-Dpo4, while dATP and dGTP misinsertion rate was unaffected. This finding can be rationalized if the +BP-dG adduct has bulk in the minor groove during dCTP and dTTP incorporations. This is reminiscent of our observations with –BP-dG, except that increasing the size of the opening on the minor groove side of the active site increased not only with dCTP and dTTP but also with dATP (though not dGTP). This pattern suggests that +BP-dG bulk might be in the major groove in the case of both dATP and dGTP misinsertions. When a Distal-NCB is added to Dpo4 (e.g., with aa248-246/Dbh-Dpo4 and I99P/A100L) the rate of dATP and dGTP misinsertions opposite +BP-dG decreased compared to wt-Dpo4, while dCTP and dTTP incorporation rates were unaffected. This finding can be
rationalized if the +BP-dG adduct has bulk in the major groove, and the Distal-NCB helps position the LF-Domain to exclude these kinds of structures from the active site. This is reminiscent of our observations with –BP-dG, except adding a Distal-NCB decreased not only dATP and dGTP but also dTTP (though not dCTP). Thus, in the case of +BP-dG, dTTP misinsertion is behaving like correct dCTP insertion, as if dTTP misinsertion does not arise from an aberrant adduct structure. This could be rationalized if dTTP misinsertion opposite +BP-dG is occurring from an adduct conformation that is close to the conformation for correct dCTP insertion, such as a wobble base pair between dTTP and the dG moiety of +BP-dG. More work needs to be done to solidify these observations and these tentative conclusions.

**Single-Stranded Binding Protein (SSB) Suppresses dATP and dGTP Misinsertion Opposite +BP-dG**

Historically, some groups studying Y-family DNAPs have included SSB in their assays and some have not, and in fact, there have been disagreements about whether SSB has an effect on the properties of some Y-Family DNAPs (99). In the early phase of our work we were uncertain whether to include SSB or not, so we did preliminary studies to investigate its effect. At the time, our findings in experiments +/-SSB were dumbfounding and complex, so we decided against adding SSB, which also made sense because the inclusion of SSB just added another variable/complication. However, now that we have a better understanding of these reactions, we have developed a hypothesis for what SSB might be doing.
When SSB was added to Dpo4, the rate of dATP and dGTP misinsertions opposite +BP-dG decreased, while dCTP and dTTP rates were unaffected. This finding follows the same pattern as when a Distal-NCB was added to Dpo4, as described in the previous subsection; namely, the dATP and dGTP misinsertions rates opposite +BP-dG decreased with aa248-246/Dbh-Dpo4 and I99P/A100L compared to wt-Dpo4, while dCTP and dTTP incorporation rates were unaffected. This striking correlation suggests that SSB might be able to interact with Dpo4 to help secure the LF-Domain to the TPF-Domains in a fashion analogous to having a Distal-NCB. More experiments will have to be done to investigate this hypothesis. For example, one prediction is that a Dpo4 derivative that already has a Distal-NCB (e.g., aa248-246/Dbh-Dpo4 or I99P/A100L-Dpo4) will not show a difference in the rate of dATP and dGTP misinsertions -SSB vs. +SSB.

UV-Damage and Abasic Sites

As mentioned several times, Y-Family DNAPs fall into two classes: the k/IV-class, which insert accurately opposite N$^2$-dG adducts such as –BP-dG, and the h/V-class, which inserts accurately opposite UV-damage but not N$^2$-dG adducts. SsDbh and EcDNAP IV are in the k/IV-class, while SsDpo4 is in the h/V-class. Much of the work in Chapter 4 describes Dpo4 derivatives that insert dCTP faster opposite –BP-dG (e.g., V32G/M76G-Dpo4) and slower in the case of dATP, dGTP and dTTP misinsertion (e.g., aa248-246/Dbh-Dpo4 and I99P/A100L-Dpo4, which have a Distal-NCB). Dpo4 is known to preferentially insert dATP at both the 5’-T and 3’-T of a thymine-thymine
cyclopyrimindine dimer (T^T), which is the dominant lesion formed following UV irradiation of DNA. One question of interest is whether Dpo4 derivatives that improve fidelity during bypass with –BP-dG might be worse in terms of bypassing a T^T.

In preliminary experiments, a template containing a T^T was synthesized and studied with Dpo4 and several Dpo4 derivatives. Interestingly, V32G/M76G-Dpo4 was much slower at inserting dATP opposite a T^T than wt-Dpo4. While this finding is preliminary, it does make sense: it seems likely that k/IV-class DNAPs are optimized for the bypass of N^2-dG adducts and h/V-class DNAPs are optimized for the bypass of UV-damage, and any attempt to improve the ability of a h/V-class DNAP to bypass N^2-dG adducts would be detrimental to its ability to bypass UV damage.

Future experiments could include a survey of dATP insertion opposite T^T by all of the Dpo4 derivatives described herein, which could be done fairly readily as it would require a single reaction lane (i.e., for dATP) with 20-30 DNAPs and could be accomplished in a single DNA sequencing gel. Furthermore, the h/V-class DNAPs also bypass abasic sites (Introduction), so a survey of how various Dpo4 derivatives insert opposite abasic sites would also be of interest.

**Other Studies**

The AS-NCB in EcDNAP IV involves two interactions R37/E251 and R38/D252, as we showed in a series of studies involving mutating these residues (Chapter 5). DNAP IV X-ray structures show the R38/D252 interaction but not the R37/E251 interaction, which indicates that X-ray structures may not be definitive in permitting an assessment of
what interactions are key in transition state structures. As discussed in Chapters 4 and 5, numerous X-ray structures show that Dpo4’s AS-NCB involves two interactions: R36/N254 and K252/R36(C=O), and we would like to assess whether these interactions are indeed important during –BP-dG bypass. Several preliminary experiments showed that the mutant K252A-Dpo4 did not affect dNTP insertion patterns opposite –BP-dG. A very preliminary study suggested that R36A-Dpo4 also does not have a significant impact on dNTP insertion patterns. However, the double mutant R36A/K252A-Dpo4 misinserted dATP, dGTP and dTTP even faster than wt-Dpo4. We intend to complete this study in the near future.

In Chapter 4, we discussed the fact that the addition of a portion of the LF-Domain of Dbh suppressed dGTP misinsertion, namely, the derivative aa238-246/Dbh-Dpo4 inserted dGTP ~3.7-fold slower than wt-Dpo4. This was evidence that a Distal-NCB can suppress misinsertion opposite –BP-dG. However, this decrease was less than the decrease observed when the entire LF-Domain of Dbh was added to Dpo4: dGTP misinsertion was ~21-fold slower for the chimera Dbh/LF/Dpo4 compared to wt-Dpo4. This comparison suggested that residues other than aa238-244 in the Dbh LF-Domain must also be important for suppressing dGTP misinsertion. We noted that by substituting aa238-244 from Dbh into Dpo4, the resulting protein (aa238-244/Dbh-Dpo4) has lost one DNA-interacting residue present in Dpo4, namely R242. Dbh itself appears to compensate for this lost DNA interaction via R283 and R325, both of which interact with DNA, whereas the equivalent residues in Dpo4 (A283 and E325) do not interact with DNA. Thus, we propose that adding A283R and E325R to aa238-244/Dbh-Dpo4 might
give a triple mutant whose rate of dGTP misinsertion would be closer to the rate for Dbh/LF-Dpo4.

**Origin of the N-clasp in hDNAP k**

As shown in the work in this thesis, non-covalent bridges are important for suppressing dATP, dGTP and dTTP misinsertion reactions opposite –BP-dG. The AS-NCBs and Distal-BCBs in the case of the bacterial EcDNAP IV, and Archaea SsDpo4 and SsDbh are on the minor groove side of the active site, which seems intuitively undesirable given that the goal of these DNAPs is to correctly insert opposite N²-dG-adducts, which must have bulk in the minor groove during dCTP insertion. hDNAP k seems to have solved this issue, since its only non-covalent bridge is on the major groove side of the active site (Chapter 4, Figure 4.1C), which involves interactions between residues in the N-clasp and the LF-Domain.

Interestingly, when the N-clasp from human DNAP k was added to Archaea SsDpo4 (giving N-clasp-Dpo4), it worked, as evidenced by the finding that dGTP misinsertion was inhibited ~5-fold compared to wt-Dpo4. Why did the N-clasp work in Dpo4? The R-groups of I36 and I39 in the N-clasp form a hydrophobic pocket with L160 in DNAP k, which allows a portion of the N-clasp to stick to a region of the TPF-Domains. In the equivalent position to L160 DNAP k, Dpo4 has I67, which could in principle interact with I36 and I39 in N-clasp-Dpo4 and, thus, allow the N-clasp to function in Dpo4 to suppress dGTP misinsertion.
The fact that the N-clasp works in Dpo4 hints at how the N-clasp might have evolved in DNAP κ proteins. If a Dpo4-like (or DNAP IV-like protein) picked up an insertion mutation at its N-terminus from some heterologous genetic source, it might have functioned to suppress certain kinds of adverse misinsertions, analogous to the suppression of dGTP misinsertions opposite –BP-dG, as observed in the case of N-clasp-Dpo4. Thereafter, secondary mutations could have occurred to perfect the N-clasp, including the evolution of an effective AS-NCB between the TPF-Domains and the LF-Domain on the major groove side of the active site, which would preclude the need for an AS-NCB and Distal-NCB on the minor groove side, such as found in Dbh and DNAP IV.

What might have been the source of the N-clasp aa-sequence? In a BLAST search, the N-clasp is most similar to a region of the large subunit of a Clostridium cellulolyticum phage terminase.

<table>
<thead>
<tr>
<th>Human DNAP κ</th>
<th>Clostridium cellulolyticum phage terminase</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 GLDKEKINKIIIMEATKGSRFYGNELK--KEKQVQRIENMMQQAAQITSQQLRK 81</td>
<td></td>
</tr>
<tr>
<td>60 GL KEK+++IIM KG+++ Y N+++ + + N + K IT +Q +K</td>
<td></td>
</tr>
<tr>
<td>212 GLSKEKIDQIIMNVPKGKTLYKNKIQGQLGRATGLIFGNFTRGKNVITREQAKK 265</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>230</td>
</tr>
</tbody>
</table>

Importantly, a number of key amino acids (underlined) are conserved between hDNAP κ and the Clostridium phage terminase. I36 and I39 in hDNAP κ are conserved in the Clostridium phage terminase (I218 and I221) in a region that is highly conserved. In X-ray structures of hDNAP κ, the N-clasp-NCB appears to involve several interactions:
R48/N464, F49/F465 and N52/N464(C=O). When DNAP k and the Clostridium phage terminase sequences are compared, one of these residues is identical (N234 for N52), one is charge conserved (K230 for R48) and one is hydrophobic conserved (L231 for F49).

In this analysis, we do not mean to imply that we can conclude that some precursor to hDNAP k evolved in a Clostridium, but the analysis does provide a vision of the kind of evolutionary process that might have occurred.
**List of Journal Abbreviations**

<table>
<thead>
<tr>
<th>Journal Abbreviation</th>
<th>Full Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adv Protein Chem</td>
<td>Advances in Protein Chemistry</td>
</tr>
<tr>
<td>Annu Rev Biochem</td>
<td>Annual Review of Biochemistry</td>
</tr>
<tr>
<td>Annu Rev Microbiol,</td>
<td>Annual Review of Microbiology</td>
</tr>
<tr>
<td>Biochim Biophys Acta</td>
<td>Biochimica et Biophysica Acta</td>
</tr>
<tr>
<td>Cancer Res</td>
<td>Cancer Research</td>
</tr>
<tr>
<td>Cell Mol Life</td>
<td>Cellular and Molecular Life Sciences</td>
</tr>
<tr>
<td>Cell Res</td>
<td>Cell Research</td>
</tr>
<tr>
<td>Chem Res Toxicol</td>
<td>Chemical Research in Toxicology</td>
</tr>
<tr>
<td>Curr Opin Struct Biol,</td>
<td>Current Opinion in Structural Biology</td>
</tr>
<tr>
<td>Embo J</td>
<td>The EMBO Journal</td>
</tr>
<tr>
<td>Environ Mol Mutagen</td>
<td>Environmental and Molecular Mutagenesis</td>
</tr>
<tr>
<td>Genes Cells</td>
<td>Genes to Cells</td>
</tr>
<tr>
<td>Genes Dev</td>
<td>Genes and Development</td>
</tr>
<tr>
<td>J Am Chem So</td>
<td>Journal of the American Chemistry Society</td>
</tr>
<tr>
<td>J Bacteriol</td>
<td>Journal of Bacteriology</td>
</tr>
<tr>
<td>J Mol Biol</td>
<td>Journal of Molecular Biology</td>
</tr>
<tr>
<td>J Mol Graph Model</td>
<td>Journal of Molecular Graphics and Modeling</td>
</tr>
<tr>
<td>J Nucleic Acid</td>
<td>Journal of Nucleic Acids</td>
</tr>
<tr>
<td>Microbiol Mol Biol Rev</td>
<td>Microbiology and Molecular Biology Reviews</td>
</tr>
<tr>
<td>Mol Cell</td>
<td>Molecular Cell</td>
</tr>
<tr>
<td>Mol Microbio</td>
<td>Molecular Microbiology</td>
</tr>
<tr>
<td>Mutat Res</td>
<td>Mutation Research</td>
</tr>
<tr>
<td>Nat Rev Cancer,</td>
<td>Nature Reviews Cancer</td>
</tr>
<tr>
<td>Nat Struct Biol,</td>
<td>Nature Structural Biology</td>
</tr>
<tr>
<td>Nat Struct Mol Biol</td>
<td>Nature Structural and Molecular Biology</td>
</tr>
<tr>
<td>Nucleic Acids Re</td>
<td>Nucleic Acids Research</td>
</tr>
<tr>
<td>Proc Natl Acad Sci U S A</td>
<td>Proceedings of the National Academy of Sciences</td>
</tr>
<tr>
<td>Trends Microbiol</td>
<td>Trends in Microbiology</td>
</tr>
</tbody>
</table>
References

14. Pata, J. D. *Biochim Biophys Acta* **1804**(5), 1124-1135


Curriculum Vitae
Gabriel Sholder
January 2014

24 Cummington St. 111 Wallace St.
Room 748 Newton, MA 02461
Boston, MA 02215 (617) 283-4434
(617) 353-6997 gsholder@bu.edu

EDUCATION
Ph.D., Molecular/Cellular Biology and Biochemistry, Boston University, January, 2014
M.A., Biochemistry, Boston University, 2006
B.S., Biochemistry and Spanish, Bates College, 2002

RESEARCH EXPERIENCE
Ph.D. Dissertation, Molecular/Cellular Biology and Biochemistry, Boston University, 2014
Y-Family DNA Polymerase Architecture: Three Structural Features Control Accurate
dCTP Insertion Opposite an N2-dG Adduct (Benzo[a]pyrene). Designed a novel project
wherein specialized trans-lesion DNA polymerases from E. coli and Sulfolobus sp. were
cloned, purified, and assayed in vitro for their ability to replicate damaged DNA.
Homology modeling and site-directed mutagenesis were then used to determine the
specific structural features of Y-family polymerases that govern the accurate replication
of damaged DNA.

Master’s Research Thesis, Biology Department, Boston University, 2006
Employed Microarray and RT-PCR technology in conjunction with small-molecule
protein inhibitors to elucidate the signaling networks that mediate the transcriptional
response of quiescent mammalian cells to growth factor stimulation.

Associate Biochemist, Genzyme Inc., MA, 2003
Improved in-house plasmid chromatography techniques in a GMP-compliant
environment.

Research Fellow, Bates College, 2002
Carried out an independent endeavor funded by Center for Environmental BioInorganic
Chemistry in which the alkane-hydroxylase from R. erythropolis was cloned.

Undergraduate Research Thesis, Bates College, 2002
Administered analytical probes to bacterial that over-expressed metalloenzymes and
examined products with GC-MS in order to elucidate the different mechanisms of alkane-
hydroxylation.
TEACHING EXPERIENCE
As a TF at Boston University, I taught the following sections:
Introductory Biology (Laboratory Instructor) 2006
Cellular Biology (Discussion Instructor) 2003
Genetics (Discussion Instructor) 2004
Cellular Biology and Biotechnology (Discussion Instructor, Guest Lecturer) 2008
Molecular Biology (Discussion Instructor) 2008, 2009, 2010
Biochemistry Laboratory (Laboratory Instructor) 2006, 2007, 2012
Molecular Biology Laboratory (Laboratory Instructor) 2007

PUBLICATIONS

G. Sholder, A. Creech, E. Loechler. Multiple Non-Covalent Interactions Between Active Site and Little Finger Domains Control the Accuracy of DinB DNA Polymerase. (In Preparation)

G. Sholder, E. Loechler. Quantitating Minor 32P-DNA Bands in Electrophoresis Gels. (In Preparation)


CONFERENCE PARTICIPATION
2012, Gordon Conference in Genetic Toxicity
Y-Family DNA Polymerase Architecture: Two Structural Features Control Accurate dCTP Insertion Opposite an N2-dG Adduct (Benzo[a]pyrene)

2009, Gordon Conference in Genetic Toxicity
Amino Acid Architecture that Influences Insertion Efficiency by Y-Family DNA Polymerases

2006, Boston University Science and Engineering Symposium
Architecture of DNA Polymerase V Based on Studies with Benzo[a]pyrene

2005, Boston University Science and Engineering Symposium
Roles of Sustained Erk Signaling in Immediate-Early and Delayed Primary Response Genes

2002, Bates College Mt. David Summit of Academic Achievement
Di-Iron Metalloenzymes that Hydroxylate Hydrocarbons

GRANTS AND FELLOWSHIPS
National Institute of Environmental Health Science, Research Grant ES003775, 2012
Boston University, Teaching Fellowships, 2006-2012
Princeton Center for Bio-Inorganic Chemistry, Research Grant, 2002

DEPARTMENTAL PRESENTATIONS
As a Research Fellow at Boston University, departmental talks in the form of PowerPoint presentations that outline our personal research are an annual requirement.

DEPARTMENTAL SERVICE
Graduate Student Organization Representative, Biology Department, Boston University, 2005