Mechanisms of DNA synthesis and fidelity by Y-family translesion and C-family replicative polymerases

Purba Mukherjee

University at Albany, State University of New York, pmukherjee@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Biochemistry Commons

Recommended Citation


This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
MECHANISMS OF DNA SYNTHESIS AND FIDELITY BY Y-FAMILY TRANSLESION AND C-FAMILY REPLICATIVE POLYMERASES

by

Purba Mukherjee

A Dissertation
Submitted to the University at Albany,
State University of New York

in
Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy
Department of Biomedical Sciences
School of Public Health
2014
ACKNOWLEDGEMENT

I am most grateful to my mentor Dr. Janice Pata, whose guidance and advice made these last 5 years possible. I am thankful to her for giving me the opportunity to work in her lab and for believing in me when I needed it the most. Thank you Janice for helping me achieve a life-long dream.

I am also extremely grateful to my committee members, Drs. Joachim Jaeger, Marlene Belfort, Haixin Sui and Graham Walker. Your insight and constant guidance helped me stay on the right track.

To the Pata and Jaeger lab members: it was great fun working (and playing) together. I would like to make a special mention of a past member of our lab, Ryan Wilson. He taught me most of the techniques, including crystallography. I will always remember the fun synchrotron trips and long conversations about protein purification, assays and good food.

I appreciate the contribution of the support staff at Wadsworth Center and Center for Medical Sciences in making my life easier at work. I am especially grateful for the financial support received from the Wadsworth Center and the Department of Biomedical Sciences at University at Albany during my final year of PhD.

My friends Aymen Yassin and Kae and Takeshi Yokoyama made the stay in Albany both enjoyable and memorable, I am very grateful to them. I am also lucky to have Neha Sharma Sabharwal, Sharmin Naaz, Bodhisatwa Das and Pooja Gupta Saraf as my friends. Talking to you kept me sane. Thank you all and I wish you the very best.

This dissertation only represents the successes of the last five years. I express my deepest gratitude for the people who were there beside me even during the failures. To my family—Ma and Bon— you have shared with me some of the most
difficult times, supported my every decision and given unwavering encouragement, while being several thousand miles away. I owe this to you. Finally, I was lucky to have my husband, Indrajit Lahiri as my colleague for the last three years. Indra, your unconditional love and encouragement, even mere presence has seen me through some very rough times. We made the best team ever, sharing both elation and disappointment. I look forward to our life together and sincerely hope we have another chance at working together. This PhD would not have been possible without you. Thank you always.

Purba Mukherjee
April 2014.
Since the discovery of the DNA polymerase by Arthur Kornberg nearly 60 years ago, there have been great advances in understanding the involvement of polymerases in replication and repair. Years of genetic, biochemical and structural studies have lead to the classification of DNA-dependent DNA polymerases into six families: A, B, C, D, X and Y. In this work, I have focussed on two different families, C and Y. Hence this work is divided into two parts. Part one discusses the studies on Y-family polymerases. All Y-family polymerases are involved in replicating past DNA lesions. The ability to tolerate unnatural nucleotides also makes Y-family enzymes extremely error-prone. Here I specifically focus on Y-family members belonging to the DinB category. DinB homologs are known to generate frameshift mutations and single base deletions at high frequencies, particularly on repetitive undamaged DNA sequences. I have determined the mechanism that is used by pol kappa (κ), the DinB homolog in humans, to generate single-base deletions on a homopolymeric run. We observe that DinB polymerases realign their substrate to convert deletions to base-substitution mutations. The importance of this observation is discussed. To better understand which element of Y-family polymerase architecture determines the error-generation specificity of these polymerases, I have used two closely related DinB homologues in archaea (Dbh and Dpo4) to generate chimeric constructs. Our biochemical and crystallographic evidence indicates that a linker connecting the polymerase domain to the C-terminal little finger/polymerase-associated domain (LF/PAD) in both these polymerases is crucial for determining conformation and mutation signature. The altered conformation specifically influenced the rate of correct nucleotide incorporation, thereby determining polymerase fidelity. The relevance of this result is also discussed. The second part of this thesis is aimed at filling the distinct gap that exists in our understanding of the kinetic pathway of bacterial replicative DNA polymerases. I have addressed this by using detailed pre-steady state and steady state assays to determine the kinetic parameters for the DNA polymerization pathway followed by the bacterial C-family replicative
polymerase, PoIC from the Gram positive pathogen *Staphylococcus aureus*. We find PoIC exhibiting several unusual characteristics, the most interesting one being slow pyrophosphate release. Our studies lead to the surprising observation that PoIC can overcome its slow PPI release, a critically limiting step, when the next incoming nucleotide is present. This novel observation and its implications are further discussed.
TABLE OF CONTENTS

Acknowledgement ii
Dissertation abstract iv
List of Figures xi
List of Tables xiii

Chapter 1: Introduction 1

1.1. Replication 1
   1.1.1. DNA helicase 1
   1.1.2. Single strand binding protein (SSB) 2
   1.1.3. Primase 2
   1.1.4. Clamp and clamp loader 4
   1.1.5. DNA polymerase 4
   1.1.6. C-family DNA polymerases 10

1.2. Damage repair and tolerance 12
   1.2.1. Translesion synthesis (TLS) 12
   1.2.2. TLS Polymerases 14
   1.2.3. Y-family of DNA polymerases 15

PART ONE

Chapter 2: Mukherjee P, Lahiri I, Pata JD. Human polymerase kappa uses a template-slipage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. Nucleic Acids Res. 2013 May; 41(9):5024-35.

   2.1. Description of contribution 24
   2.2. Summary 25
3.4.4. \( K_d^{\text{dNTP}} \) and \( k_{\text{pol}} \) determination.  

3.4.5. Crystallization and Structure Determination.  

3.5. Results  

3.5.1. Linker identity determines -1 deletion frequency on a repetitive sequence  

3.5.2. Linker identity determines efficiency of mispair extension.  

3.5.3. Amino acid trio in the linker determines overall polymerase conformation  

3.5.4. dNTP selection: linker-dependent alteration of nucleotide incorporation rate  

3.6. Discussion  

3.7. Acknowledgements  

PART TWO  


4.1. Description of contribution  

4.2. Summary  

4.3. Introduction  

4.4. Materials and methods  

4.4.1. Materials.  

4.4.2. Duplex DNA formation.  

4.4.3. Expression and purification of Sau-PolC-\( \Delta N\Delta \)Exo.  

4.4.4. Assays for optimization of enzymatic activity of Sau-PolC-\( \Delta N\Delta \)Exo.  

4.4.5. Steady-state assays.  

4.4.6. Measurement of DNA dissociation rate from binary complex.  

4.4.7. Active site titration of Sau-PolC-\( \Delta N\Delta \)Exo.
4.4.8. $K_{\text{dNTP}}^{\text{app}}$ determination. 100
4.4.9. Simulation. 101

4.5. Results 102
4.5.1. PolC purification. 102
4.5.2. Optimal reaction conditions. 102
4.5.3. Michaelis-Menten kinetics. 105
4.5.4. Rate of dissociation of p/t DNA from the binary complex. 105
4.5.5. Pre-steady-state burst kinetics. 109
4.5.6. Active site titration. 112
4.5.7. Nucleotide binding affinity. 113
4.5.8. Dependence of active enzyme•DNA binary complex on [dNTP]. 116

4.6. Discussion 119
4.7. Acknowledgements 123

Chapter 5: Lahiri I, Mukherjee P, Pata JD. Catalysis by *Staphylococcus aureus* PolC is regulated by the next correct (but not incorrect) nucleotide. (Manuscript in preparation)

5.1. Description of contribution 124
5.2. Summary 125
5.3. Introduction 126
5.4. Methods 129
5.4.1. Protein purification 129
5.4.2. DNA annealing 130
5.4.3. Primer extension assays 130
5.4.4. Data analysis 131
5.4.5. Simulation 132
5.4.6. Measurement of pyrophosphate release rate 132

5.5. Results 133
5.5.1. N-terminal and exonuclease domains slow DNA release without improving $(K_{\text{DNA}}^{\text{app}})$. 133
5.5.2. β clamp increases the affinity of the binary complex. 137
5.5.3. Incoming dNTP stabilizes the pre-chemistry ternary complex 137
5.5.4. Nucleotide incorporation is reversible both in the presence and absence of the clamp 138
5.5.5. Processive synthesis by PolC is fast 144
5.5.6. Presence of next correct incoming dNTP allows rapid pyrophosphate release 146

5.6. Discussion 149
5.7. Acknowledgements 155
5.8. Supplementary information 156
   5.8.1. Choice of DNA substrate 156
   5.8.2. PolC stall products 156

Chapter 6: Summary and future directions 159

6.1. Y-family TLS polymerase 159
6.2. Human polymerase κ performs template-slippage but realigns 159
   6.2.1. Functional relevance 160
   6.2.2. Active vs passive realignment strategies 161
   6.2.3. Effect of the cellular machinery 160
   6.2.3. Post-realignment extension and handoff 165
6.3. Linker dependent conformation of archaean DinB polymerases 166
6.4. C-family replicative polymerases 169
   6.4.1. PolICΔΔExo displays slow PPI release 169
   6.4.2. PPI release is no longer slow when next correct nucleotide is present 170
   6.4.3. Functional relevance of “+1 dNTP-dependent PPI release” 172
   6.4.4. Implications for C-family members 174
6.5. Concluding remarks 174

Bibliography 176
LIST OF FIGURES

Figure 1.1. DNA replication machinery 3
Figure 1.2. DNA Polymerase architecture 7
Figure 1.3. DNA Polymerase families 9
Figure 1.4. Comparison of DNA polymerase palm topology 11
Figure 1.5. Y-family polymerases 16
Figure 2.1. Single base deletion mechanisms. 29
Figure 2.2: Nucleotide incorporation by hPolK on repetitive sequence containing deletion hotspot. 38
Figure 2.3. Mechanism of second dGTP incorporation on 4C-G substrate. 44
Figure 2.4. Mechanism of multiple dCTP incorporations. 46
Figure 2.5. hPolK misincorporates efficiently on 1T-G substrate. 50
Figure 2.6. Structure of an hPolK ternary complex. 57
Figure 3.1. Parental Y-family polymerases and chimeric constructs. 65
Figure 3.2. Single-base deletion efficiency is dependent on the identity of the linker. 71
Figure 3.3. Chimeric polymerases use a template-slippage deletion mechanism. 75
Figure 3.4. DbhDpo4RSDbh adopts a Dpo4-like conformation. 79
Figure 4.1. Overview of PolC structure and domain organization. 92
Figure 4.2. Minimal single-nucleotide incorporation reaction pathway for DNA polymerases. 95
Figure 4.3. SDS-PAGE of Sau-PolC-ΔNΔExo. 103
Figure 4.4. Primer extension assays for optimizing enzymatic activity of Sau-PolC-$\Delta$N$\Delta$Exo. 104

Figure 4.5. Steady-state kinetic analysis of Sau-PolC-$\Delta$N$\Delta$Exo. 107

Figure 4.6. Determination of the DNA dissociation rate from polymerase $\cdot$ DNA binary complex ($k_{\text{off}}$). 108

Figure 4.7. Pre-steady-state kinetics and active site titration of Sau-PolC-$\Delta$N$\Delta$Exo. 110

Figure 4.8. Determination of $K_D^{\text{dNTP}}$ of Sau-PolC-$\Delta$N$\Delta$Exo. 115

Figure 4.9. Simulation of kinetic pathway of Sau-PolC-$\Delta$N$\Delta$Exo. 119

Figure 4.10. Minimal enzymatic pathway for Sau-PolC-$\Delta$N$\Delta$Exo. 120

Figure 5.1. Reaction pathway of DNA polymerase. 128

Figure 5.2. Pre-steady state kinetics of PolC-Exo$^{\text{Mut}}$. 134

Figure 5.3. Affinity of dNTP for the pre-chemistry binary complex. 139

Figure 5.4. Simulation of the reaction pathway of PolC-Exo$^{\text{Mut}}$. 142

Figure 5.5. Multiple nucleotide extension by PolC. 145

Figure 5.6. Simulation of the processive synthesis reaction pathway for PolC. 147

Figure 5.5.1. Processive synthesis by PolC in the presence of trap DNA. 158
LIST OF TABLES

Table 2.1. DNA Substrate Sequences. 33
Table 2.2. Summary of observed nucleotide incorporation rates. 36
Table 3.1. DNA substrate sequences 66
Table 3.2. Summary of observed nucleotide incorporation rates 73
Table 3.3. Data collection and refinement statistics 81
Table 3.4. Kinetic parameters for correct and incorrect nucleotide incorporation on a deletion hotspot sequence (substrate 4C-G) 84
Table 4.1. Comparison of kinetic parameters of different polymerases. 114
Table 5.1. Kinetic parameters obtained with different constructs of PolC. 151
CHAPTER 1
INTRODUCTION

Genome duplication and its maintenance are two vital processes for every organism. The inability to perform either of these could lead to loss of genetic information that may result in cell death. Consequently, “replication” and “repair” are both elaborate, tightly regulated, complex pathways involving multiple participants and require highly coordinated functions. Although DNA replication broadly follows the same basic steps among all domains of life, mechanisms of DNA repair are varied and often do not overlap.

1.1. REPLICATION

Several multiprotein complexes are involved in DNA replication, and together the large macromolecular machinery that carefully and efficiently synthesizes DNA is often referred to as the “replisome”. Interestingly, there exists more homology between the eukaryotic and archaeal replisome machinery than with their bacterial counterparts [1, 2]. The commonly conserved players and their roles in DNA replication are described below in detail (Figure 1.1).

1.1.1. DNA helicase: DNA replication begins with initiation- a synchronized event that requires two essential events: origin recognition and helicase loading. Origin recognition refers to a little understood process by which specific protein(s) or “initiator(s)” bind to unique DNA sequences or “replicators”. In bacteria, bacteriophages and eukaryotic viruses, only a single replicator or origin has been identified while some archaea and all eukaryotes have multiple origins of replication per chromosome [3]. Initiator proteins such as bacterial DnaA [4] or eukaryotic/archaeal origin recognition complex, ORC specifically recognize the replicator DNA sequence (ori) and lead to the assembly of a nucleoprotein complex that unwinds DNA at the origin [4-8]. This is followed by recruitment of an ATP-dependent helicase loader belonging to the AAA+ ATPase
superfamily [5, 8, 9] and responsible for loading the ring-shaped ATP fueled helicase around the DNA. Bacterial helicase is a homohexamer like the CMG complex in eukaryotes (consisting of Cdc45, MCM2-7 and GINS)[8, 9]. Eukaryotic helicase recruitment is more complex and involves MCM2-7 to be loaded as an inactivated dodecamer [10] that requires activation before it can unwind DNA ahead of the fork [7, 11]. Helicase assembly on both DNA strands initiates the establishment of bi-directional replication forks.

1.1.2. Single-stranded DNA binding protein (SSB): DNA that is being unwound by the helicase is prevented from reannealing, nucleolytic digestion or adopting secondary structure through association with SSB [12]. Bacterial SSB is homotetrameric and its eukaryotic or archaeal counterpart is a heterotrimer also known as RPA (replication protein A). In addition to the roles mentioned above SSB is also known to interact with members of the replisome and influence their function [13].

1.1.3. Primase: While the helicase is responsible for the initiation of replication, the primase is crucial for the replisome to begin elongation. A primase is a DNA-dependent RNA polymerase that remains closely associated with the helicase (often through direct protein-protein contacts) [11] and performs de novo synthesis of short RNA primers, templated by the unwound genomic DNA. Synthesis has to proceed in a 5’- 3’ direction and generates a free 3’-hydroxyl end that can then be used by a DNA polymerase for extension. This restriction on the direction of polymerization results in one of the two DNA strands being replicated continuously (leading) while the other is synthesized discontinuously (lagging). The lagging strand is made in shorter segments known as “Okazaki fragments” (~1-2kb in bacteria and 100-200bp in eukaryotes) using primers that are laid down in close succession by the primase [12, 14, 15]. Close contacts between the helicase and primase is one of the ways by which the replisome ensures that synthesis on both strands remains coordinated and replication can proceed at high speed [11].
Figure 1.1. DNA replication machinery

A simplified representation of the different components comprising the replisome is shown. Adapted from Molecular Biology of the Gene, Pearson Education Inc.
In eukaryotes, the primase is found in complex with Polα, a DNA polymerase that is known to extend from the RNA primer-DNA template hybrid to a short duplex DNA before handing off to the replicative polymerases. Recent structural evidence indicates that this handoff may occur when Polα senses the transition from A-form (hybrid RNA/DNA duplex) to B-form DNA [15]. Additionally, SSB or RPA binding to ssDNA is thought to facilitate the switch between primase and replicative polymerase [16, 17]. Since Polα lacks proofreading activity, both the RNA primer and nucleotides laid down by Polα are eventually removed and replaced to generate a proofread, DNA-only genome.

1.1.4. Clamp and clamp loader: The primed, single-stranded DNA is the substrate for recruitment of the replicative DNA polymerase. Once bound, the DNA polymerase needs to form a stable complex that can perform fast and processive synthesis while retaining the ability to release DNA without itself dissociating from the replisome. To ensure this, all organisms have evolved the “sliding clamp”, also known as β-clamp in bacteria and proliferating cell nuclear antigen (PCNA) in both archaea and eukaryotes [18]. The clamp is a multisubunit complex with ring architecture (with some exceptions [19]) and is either homodimeric (bacteria) or -trimeric (eukaryotes). It tethers the DNA polymerase to the DNA through specific interactions with the PCNA interacting peptide (PIP) sequence present on the polymerase. PIP motifs have also been found in some DNA repair proteins [20].

The ring architecture of the clamp poses the problem of loading it on DNA since the latter needs to thread through the central channel of the clamp. This is achieved by the clamp loader: an ATP-dependent pentameric multisubunit complex that breaks the “ring” to then load the sliding clamp on DNA [21]. It is still unknown if the clamp loaded at this stage is already associated with a DNA polymerase.

1.1.5. DNA polymerase: The final, essential enzymatic function, without which replication would not occur is performed by the DNA polymerase. In E.
coli a single polymerase (Pol III or DnaE) performs this function for both leading and lagging strands (Figure 1.1) [22]. The initial view that only two polymerase molecules function at a fork has been replaced by recent data suggesting that a third polymerase molecule is involved that periodically exchanges with the lagging strand polymerase and is constantly a part of the replisome [23, 24]. In eukaryotes, Pol ε and Polδ are thought to be the leading and lagging strand replicative polymerases respectively that remain at the replication fork in addition to Polα [25-27].

The DNA polymerase is responsible for the principal chemical, or “nucleotidyl transfer”, step that results in a growing chain of DNA from the pre-synthesized primer. The process of nucleotide addition is thought to occur by a “two-metal ion” mechanism [28, 29]. Two divalent cations (Metal A and B, generally magnesium) coordinate the pentacovalent transition state between the 3’-OH end of the primer, the incoming deoxynucleoside triphosphate (dNTP) and the active site catalytic carboxylates of the polymerase itself. It has been suggested that metal A interacts with the 3’OH of the sugar and thereby reduces the pKa of the hydroxyl. The deprotonated 3’-OH so formed then initiates a nucleophilic attack on the α-phosphate of the incoming dNTP, which enters the active site while bound to metal B [28, 29]. After nucleotidyl transfer occurs, metal A diffuses out and the pyrophosphate generated is thought to leave with metal B.

Structural evidence collected for several polymerases over the last two decades has corroborated most of the steps mentioned above. While there is no crystal structure of the pentacovalent transition state to date, recent time-resolved crystallography has further confirmed the above-mentioned step-wise progression of nucleotide incorporation [30, 31]. These studies have also suggested the presence of a third, more transient metal ion at the active site during catalysis. For Polβ an error-prone polymerase, the third metal has been observed only when the correct and not the incorrect nucleotide is being added, implying a specific role in the cognate nucleotide incorporation pathway [31]. It should be noted that both the time-resolved studies were with error-prone
polymerases and the third metal is yet to be seen with a high-fidelity replicative DNA polymerase.

Early biochemical work and subsequent structural studies spanning nearly 40 years have additionally helped establish the kinetic pathway followed by a DNA polymerase. Briefly, the polymerase binds its DNA substrate to form an enzyme-DNA (E.DNA) binary complex. This is followed by nucleotide binding to form a pre-chemistry ternary complex (E.DNA.dNTP). The ternary complex formation results in both major and minor polymerase conformational changes. In most polymerases, the fingers domain closing around the incoming nucleotide causes a large-scale conformational change followed by a smaller conformational change that results in the proper positioning of substrates at the active site. It has been suggested that either this small-scale change in conformation or the binding of metal A are rate-limiting for polymerization [32]. The next step is when nucleotidyl transfer or “chemistry” occurs, leading to the post chemistry ternary state that involves the polymerase bound to its DNA substrate with a primer elongated by one nucleotide still bound to the $\beta$-$\gamma$ phosphates or pyrophosphate (PPi). The release of the pyrophosphate and a reverse conformational change regenerates a binary complex that is ready for the next round of nucleotide addition [33-35]. Although essentially the same steps occur for every DNA polymerase, key differences exist in the rates at which various steps take place thereby influencing polymerase function and fidelity.

Based on the general appearance of the initial structures of DNA polymerases and their resemblance to a “right hand” [36], the three major sub-domains that comprise the polymerase core have been named as fingers, palm and thumb (Figure 1.2) [29, 36, 37].
Figure 1.2. DNA Polymerase architecture

The right hand architecture of the core polymerase domains present in all DNA polymerases is shown. Domains are color coded as follows: fingers (blue), palm (magenta) and thumb (green). A. B-family replicative polymerase from bacteriophage RB69 (PDB ID: 4DU1) and B. Y-family error-prone polymerase from *Sulfolobus solfataricus* (PDB ID: 3AGQ) are shown. Polymerase specific domains are represented as: RB69- exonuclease (red) and NH$_2$-terminal (gray), Dpo4: little finger/polymerase-associated domain LF/PAD (orange) and linker (yellow). The active site of RB69 Pol is more constricted with lesser exposure to solvent than the “open” active site of Dpo4.
The active site of the polymerase lies in its palm domain, which contains up to three crucial catalytic residues, of which the two universally conserved ones are always aspartates. The position of these two carboxylates is also seen to be well-conserved among all polymerases. The third residue however, can be either an aspartate, glutamate or serine. The fingers domain of the polymerase is known to contact the incoming dNTP and in replicative polymerases, the fingers undergo a large conformational change during catalysis (“open” to “closed”) [38]. Until recently, this was thought to be the rate-limiting step in the polymerase kinetic pathway [32]. The thumb is thought to contact upstream duplex DNA [39-41] and maybe important for partitioning between polymerization and proofreading. In addition to the polymerase core, most polymerases have other domains; for example, all replicative polymerases have an exonuclease domain that performs the 3’-5’ proofreading activity that is essential for maintaining high fidelity (1 error in \(10^6\)–\(10^7\) bases) of DNA synthesis during replication [42-44]. However, the exact function of several such subdomains is currently unknown.

Based on primary sequence homology DNA polymerases have been classified into 6 major families: A, B, C, D, X and Y (Figure 1.3) [45, 46]. Interestingly, all replicative polymerases do not belong to the same family. All eukaryotic ones can be placed in the B-family, archaeal polymerases are found in both the families B and D while most bacterial replicative polymerases are from the C-family [45, 47]. The X and Y-family polymerases are typically involved in DNA repair and damage tolerance mechanisms [48]. This classification and the growing number of high-resolution structures have revealed that the catalytic domain (palm) of bacterial replicative DNA polymerases may have evolved independently of their eukaryotic/archaeal counterparts [49]. This allows the polymerase families to be further categorized as: (i) the “classical” (A, B and Y) and (ii) the “\(\beta\)-nucleotidyl transferase (\(\beta\)-NT)” (C and X) superfamilies.
Figure 1.3. DNA Polymerase families

Classification of DNA polymerases is shown. Superfamilies are depicted in blue and polymerase families are shown in green. Sub-groups for the C- and Y-families are also shown. This work will focus on the groups that are shown in red. Examples for each group are mentioned to the right.
1.1.6. C-family DNA polymerases: The C-family bacterial polymerases can be grouped into 4 broad categories based on sequence homology: PolC, DnaE1, DnaE2 and DnaE3 (Figure 1.3). Bacteria utilize different combinations of these enzymes with not all of them being involved in replication [50, 51]. In Gram-negative bacteria, DnaE1 is responsible for both leading and lagging strand synthesis (e.g. Pol III in \textit{E.coli}). However, low G+C Gram positive bacteria have both PolC and DnaE3, of which the former is mainly replicative while DnaE3 is thought to extend from RNA primers before handoff to PolC (e.g. \textit{Bacillus subtilis}). Finally, in high G+C Gram positive bacteria two different kinds of DnaE are present, DnaE1 and DnaE2. DnaE1 acts as the replicative polymerase while DnaE2 is thought to be error-prone (e.g. \textit{Mycobacterium tuberculosis}).

When the first crystal structures of the bacterial C-family polymerases from \textit{Thermus aquaticus} DnaE (4.6 Å) [52] and \textit{E.coli} Pol III (2.3 Å) [53] were solved, even at low resolution it became apparent that these polymerases were structurally distinct from the already known A- and B-family replicative enzymes [29, 37, 38, 40, 54] (Figure 1.4). Comparison with earlier structures of X-family polymerases such as human Pol β showed a clear resemblance in palm topology that was absent with B-family members [52]. Unlike the classical polymerases where the two key catalytic carboxylates lie on two opposite anti-parallel β-strands, the catalytic residues of the β-NT polymerases lie on the same strand and the third carboxylate is located on the opposite parallel β-strand [52]. These surprising differences to their eukaryotic counterparts make the study of bacterial replicative polymerases particularly interesting.
Figure 1.4. Comparison of DNA polymerase palm topology

The β-sheet of the palm domain containing the conserved catalytic residues as in (A) classical superfamily (PDB ID: 4DU1) and (B) β-nucleotidyl transferase superfamily (PDB ID: 3F2B) are shown. The incoming dNTP and all three catalytic residues are shown as sticks with the conserved aspartates being marked by an asterisk “*”. The divalent metal is shown as green spheres.
Despite detailed structural studies (including high resolution ones [39]), there has been no exhaustive kinetic characterization of any C-family polymerase including that for *E. coli* Pol III [39, 52, 53, 55, 56]. In Chapter 4 of this thesis, I have determined the kinetic pathway of a N-terminal and exonuclease deleted construct of PolC from the low G+C Gram-positive pathogen *Staphylococcus aureus*. My results show that while following the same basic kinetic pathway, PolC shows certain distinct features one of which being slow pyrophosphate release after chemistry. As this would pose a problem for processive synthesis in the cell, subsequently in Chapter 5 we have investigated how PolC overcomes such a kinetic barrier.

### 1.2. DAMAGE REPAIR AND TOLERANCE

Unlike DNA replication, all organisms have multiple pathways for dealing with DNA damage. These include damage tolerance mechanisms such as translesion synthesis (TLS) as well as repair pathways like base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double-strand break repair mechanisms encompassing non-homologous end-joining (NHEJ) and homologous recombination (HR) [57-59]. An average mammalian cell incurs up to 30,000 DNA lesions per day [60] caused by exposure to both endogenous (for e.g. oxidative free radicals or alkylating agents) and exogenous sources (ionizing radiation or environmental mutagens) [61]. Therefore, the combined ability to tolerate and eventually repair DNA lesions is crucial for the survival of every cell. The failure to do so could be detrimental, leading to cell death in bacteria and various severe diseases in humans.

#### 1.2.1. TRANSLESION SYNTHESIS (TLS):

DNA lesions encountered during replication are particularly detrimental since at such sites the replisome could easily be stalled either entirely unable to synthesize past the lesion or undergoing futile cycles of dNTP addition and removal. This entails the risk of replication becoming uncoupled from the DNA unwinding by the helicase leading to long stretches of single-stranded DNA (ssDNA) that could undergo
rearrangements/recombination or breaks, both leading to cell death [62, 63]. However, cells perform TLS under such circumstances to avoid stalled forks.

Translesion synthesis (TLS), the ability to synthesize past a DNA lesion, is crucial to cells. Most organisms including humans have more than one DNA polymerase for this purpose [57, 58, 64]. When replication forks stall at sites of damage, in eukaryotes, the following two modes of TLS has been proposed: (i) “polymerase switch” model and (ii) post-replicative gap-filling model [64]. According to the first model, TLS Pols transiently gain access to the replisome by displacing the replicative Pol through a process termed as “polymerase switching”. Once past the damage a switch back to the high fidelity replicative Pol would occur and normal DNA synthesis is resumed [22, 65]. The second model suggests that the replisome disassembles at points of stalling and reassembles further down to continue replication, leaving behind stretches of ssDNA. The point of damage is marked by the sliding clamp (PCNA) that recruits TLS Pols, which then bypass the lesion and fill up the intermediate gaps [66]. In eukaryotes, the damage response system involves ubiquitination and SUMOylation of PCNA that possibly ensures an additional layer of control and specificity [67]. It must be noted that while there is evidence for both models mentioned above, they may not be mutually exclusive and could function during different phases of the cell cycle [66].

In prokaryotes, a slightly different view has been suggested recently [68]. The SOS response to DNA damage in E.coli triggers the upregulation of >40 genes, including those encoding the TLS Pols (Pol II, Pol IV and Pol V). The increased number of TLS Pols promotes polymerase switch with replicative Pol III at the fork and, contrary to the general view, form stable complexes with the rest of the replisome machinery (clamp, helicase and primase) before the replisome can undergo disassembly [69]. This is essential since reassembly is known to be a complex process [12]. These error-prone polymerases are thought to slow down the rate of DNA synthesis considerably (50-500 fold), thus allowing time for the downstream damage to be repaired by other mechanisms (MMR, NER or BER). If needed, these enzymes would eventually bypass the lesion and continue slow
DNA synthesis, until a downregulation of the SOS response occurs. It must be noted that the slow rate of synthesis would act as a mechanism for reducing the number of errors generated during the presence of TLS Pols on the undamaged DNA.

Taken together these suggested mechanisms imply that TLS Pols have access to DNA that is not limited to lesions. The highly error-prone nature of these polymerases requires their presence on undamaged DNA to be both restricted and tightly controlled to reduce the probability of mutations. TLS polymerases have been long thought to be one of the major causes of spontaneous and damage-induced mutations [64, 70]. Thus, it becomes important to understand the types of errors frequently made by these enzymes, the characteristics of the protein itself that influence or define such specificities and to ultimately understand their regulation.

1.2.2. TLS POLYMERASES: All TLS Pols share certain structural and functional characteristics. Unlike the fast (100-1000nt/s), processive and extremely accurate (1 error in 10⁵ without proofreading) polymerases that perform replication, the TLS and repair polymerases are slower (1-10nt/s), less processive (even with sliding clamp) and more error prone (1 error in 10⁻⁵ to 10⁻⁴) [42, 43, 48, 64, 68, 71-74]. Some of these functional variations are an outcome of the distinct structural features of the TLS enzymes. Replicative polymerases display high fidelity because of their ability to proofread [42, 44]. TLS polymerases, however, lack the 3'-5' exonuclease domain and are, thus, extremely error-prone on undamaged DNA. Compared to their replicative counterparts, the active sites of TLS polymerases are solvent exposed, unconstricted (Figure 1.2) and more accommodating to bulkier lesions or even non-cognate basepairs, thus allowing bypass [41, 75-79]. TLS polymerases are known to accurately bypass particular lesions while being error prone in other sequence or damage contexts. Thus, the lower processivity and slow speeds are additional controls to ensure reduced presence on undamaged DNA.
1.2.3. **Y-FAMILY OF DNA POLYMERASES:** Most TLS Polymerases belong to the Y-family. These polymerases share several structural features including a C-terminal domain known as the little finger/polymerase-associated domain (LF/PAD) [48]. The LF/PAD stays connected to the core polymerase domain (made of fingers, palm and thumb) via a 10-15 amino acid interdomain linker (Figure 1.2B). Some Y-family members such as Rev1 and Polx have additional N-terminal extensions. Furthermore, all eukaryotic TLS Pols have C-terminal regions that contain motifs for binding PCNA (PIP box), ubiquitin (UBZ/UBM) and/or other TLS Pols (Pol specific) [48, 57, 58, 64]. Although structural information about the core polymerase with the LF/PAD is available for most of these proteins [30, 77, 80-82], little is known about these other extensions.

Even the core polymerase of Y-family members is distinct from that of polymerases belonging to the A- or B-family. For instance, Y-family Pols have shorter fingers that have fewer contacts with the incoming dNTP and do not undergo the large conformational change during dNTP binding seen for replicative polymerases [37, 38, 40, 54, 75, 76, 78, 82, 83]. This results in an “open” and considerably preformed active site with the carboxylates positioned at or close to their final positions when in the presence of the nucleotide. Moreover, the thumb sub-domain is shrunken such that it does not track the minor groove of the DNA duplex very effectively; thus allowing distortions in the duplex to be overlooked (Figure 1.2) [48, 75, 78, 84]. However, several contacts exist between the DNA and the LF/PAD domain that holds the DNA duplex in position between itself and the thumb domain.

Y-family polymerases are divided into 6 subgroups (Figure 1.3 and Figure 1.5) that have different bypass abilities and generate specific errors [46].
Figure 1.5. Y-family polymerases

The six different subgroups of the Y-family polymerases are shown as an unrooted phylogenetic tree (adapted from [46]). Each of the different subfamilies is shown in a different color: DinB (red), Rev1(purple), Rad30B (blue), Rad30A (green), UmuC from Gram-negative bacteria (yellow) and UmuC from Gram-positive bacteria (orange).
The characteristics of every subgroup and their functional relevance in cells is described below in greater detail:

**UmuC:** Two of the six Y-family branches belong to the UmuC enzymes. These TLS Pols are found only in bacteria and are further divided into those from (i) Gram negatives and (ii) Gram positives [46]. The best-studied of these polymerases is *E.coli* UmuC. In *E.coli*, UmuC functions as a part of Pol V: a highly error-prone (1 in $10^3$-$10^4$), multisubunit polymerase that is known to function late in the SOS response [85]. The active form of Pol V comprises of one subunit of UmuC, two subunits of UmuD’ (auto-cleaved product of UmuD) and an ATP-associated RecA molecule derived from a RecA-ssDNA nucleoprotein complex [85, 86]. Pol V can bypass a wide range of damage including abasic sites and cyclobutane pyrimidine dimers (CPDs), however it does so in an error-prone manner [87, 88]. Owing to its highly mutagenic nature, Pol V is only upregulated late in the SOS response [86].

The Gram-positive UmuC branch has very little sequence similarity to Gram-negative UmuC [46, 89]. Additionally, it appears that there may not be an associated UmuD in these organisms suggesting that the polymerase may look and function quite differently. UmuC remains the only branch of the Y-family for which no structure currently exists.

**Rad30A:** Of the three subtypes that are found only in eukaryotes, the rad30A branch contains the first Y-family member to be directly associated with a human disorder: *Xeroderma pigmentosum variant* (XPV) [90, 91] [92]. XPV patients have a defect in the rad30 gene encoding the TLS polymerase Pol eta ($\eta$) and are consequently sensitive to exposure to sunlight. This has been attributed to Pol$\eta$’s intrinsic ability to accurately bypass UV-induced photoproducts such as CPDs. In its absence, other TLS Pols (Pol$\kappa$ and $\iota$) are thought to perform error-prone bypass and cause mutations that result in an increased incidence of skin cancer in XPV patients.
Crystal structures of Pol\textsubscript{\eta} show that it can successfully incorporate two consecutive dATPs opposite a cross-linked thymidine dimer. Moreover, specific polymerase-DNA contacts establish a “molecular splint” that prevents any distortion and slippage in the DNA duplex formed after lesion bypass [93]. However, well known to be accurate on CPDs, on undamaged DNA Pol\textsubscript{\eta} is found to be extremely error prone (1 error in 10) [64].

\textbf{Rad30B:} Also found only in eukaryotes, the gene product of rad30B in humans is known as Pol iota (Pol \textsubscript{\iota}). This polymerase family shares close sequence homology to Pol \textsubscript{\eta} but its functional relevance is less understood because of its being absent in \textit{Saccharomyces cerevisiae}, the best studied, genetically amenable eukaryote. Error bypass such as abasic sites and oxidative damage by Pol iota was generally thought to be inefficient [94] until more recently when it has been shown that Pol \textsubscript{\iota} functions better in the presence of low concentrations of Mn\textsuperscript{2+} (as low as 50\muM) instead of Mg\textsuperscript{2+} [95].

Unlike other Y-family members, the Pol \textsubscript{\iota} active site appears to constrain the width of the terminal base-pair, resulting in a preferred Hoogsteen configuration between the incoming nucleotide and the template base. Consequently, Pol \textsubscript{\iota} is both highly error prone in certain contexts while being accurate in others. For instance, it invariably adds a dTTP opposite a templating A while incorporating a dGTP opposite a templating T [80, 81]. Recently, Pol \textsubscript{\iota} has also been shown to incorporate rNTPs in vitro and it has been suggested that in the absence of the “steric gate” residue, that discriminates between the sugar moiety being incorporated, Pol \textsubscript{\iota} displays better base discrimination and favors Watson-Crick base pairs [96].

\textbf{Rev1:} This category of the Y-family is also found only in eukaryotes. Although one of the first genes to be discovered [97], the limited activity of the Rev1 proteins prevented their acceptance as Y-family polymerases until later [64]. Strictly, Rev1 cannot be classified as a DNA-dependent DNA polymerase and is instead a non-templated deoxycytidyl transferase. This is because Rev1
incorporates dCTP indiscriminately by using an arginine residue in its “N-digit” as the template, while ignoring the templating base [98, 99]. This makes it particularly adept at bypassing bulky guanosine adducts that can be flipped out of the active site during nucleotide incorporation. Conversely, it renders Rev1 extremely error prone on other sequences: damaged or not. However, it is possible that Rev1’s primary function in cells is not as a polymerase in cells because unlike any of the other Y-family members, Rev1’s C-terminal extension contains interaction modules for the other four TLS Pols in humans [57, 58, 64]. This suggests a role in coordinating and possibly recruiting that is not yet fully understood.

**DinB:** Named after the “damage induced” group of genes in bacteria, this last class of Y-family polymerases is also the only one found in all domains of life. The nomenclature of these enzymes is diverse: known as Pol IV or DinB in *E.coli*, Pol kappa (Pol κ) in eukaryotes and Dpo4 and Dbh in the archaea *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* respectively. Of all the Y-family polymerases studied to date, the most structures exist for Dpo4. This is because of the ease with which it crystallizes and the high-resolution to which these crystals diffract. Nonetheless, for each of these enzymes some extent of both biochemical and structural characterization exists [75, 77-79, 82-84, 100-135] such that several common features are already known. Structurally, DinB homologs share the typical Y-family architecture. In humans, Pol κ contains an additional N-terminal extension known as the “N-clasp” that covers the active site and a C-terminal extension that is not essential for polymerase activity but is involved in protein-protein interaction [77, 82].

Functionally, all DinB polymerases accurately bypass bulky DNA adducts in the N2-position of dG [48, 136] while typically generating single-base deletions on an abasic site [100, 137]. Despite the similarities, functional variations also exist. Various DinB polymerases tend to bypass 7,8-Dihydro-8-oxo-2'-deoxyguanosine (8-oxodG) lesions with different accuracies. 8-oxodG is the most common oxidative form of DNA damage in cells [111, 125, 138]. In the anti-conformation when proper Watson-Crick base pairing with dCTP can occur,
the oxygen at position ‘8’ of the damaged guanine base clashes with its own sugar. To relieve this strain the damaged base changes orientation to the syn-conformation. In this altered orientation of 8-oxo-dG, hydrogen bonds are preferentially formed with an incoming dATP [125]. If incorporated and left uncorrected, this error would generate a transversion mutation in the next cycle of replication. While Polκ tends to misincorporate a dATP; Dpo4 is known to perform accurate bypass of 8-oxodG by incorporating dCTP with high efficiency [120, 125].

TLS polymerases have been long thought to be important contributors to spontaneous and damage-induced mutations [70] caused through base substitutions and frameshift mutations. DinB enzymes from archaea and bacteria are known to generate single base deletions (−1 frameshift) at high frequencies (1 in 10^3) in a particular “hotspot” sequence where the template sequence contains a run of identical pyrimidines flanked by a guanosine on the 5’ side (PyrGC) [112, 119, 126]. The presence of specific mutation “hotspots” for polymerases is additional evidence for the susceptibility of a particular locus. DNA synthesis on repetitive and PyrGC-rich templates is a daunting task for replicative Pols that are known to pause at such sites in vitro [139]. The genomes of vertebrates are organized into GC-rich and GC-poor regions also known as “isochores” and for humans, the early-replicating genes are mostly found in the GC-rich isochores [140]. This suggests that even early in replication, encountering such PyrGC sequences could cause stall sites providing access to TLS enzymes. It has thus become important to understand the mechanism by which TLS Pols deal with such repetitive sequences and the consequences thereof.

Several mechanisms have been suggested for the generation of single base deletions on repetitive “hotspot” sequences (see Chapter 2). Earlier biochemical and structural studies with DinB homologs from archaea (Dbh and Dpo4) and bacteria (DinB or Pol IV) have generally suggested template slippage to be the primary mechanism [78, 84, 103, 105, 107, 119]. However, the same information is not present for eukaryotic Polκ. In Chapter 2 of this thesis, using
biochemical approaches I have identified the mechanism preferred by the human DinB ortholog (hPolx) for creating single base deletions. Not surprisingly, we find that hPolx utilizes the same mechanism as its homologs. However, after generating the deletion in this context it is capable of realigning the template and extending from a mispair. The ability to realign is surprising but not new, however the outcome is more interesting: the conversion of a single-base deletion to a mispair. As cells are better equipped to remove a mispair (using MMR, BER etc.) than to correct a deletion, the ability to realign suggests that the human DinB polymerase has evolved to reduce the number of deleterious mutations it generates. The implications of this are discussed further in chapters 2 and 6.

A little understood aspect of Y-family polymerases is the source of their specificity. More than a decade of work has enabled a better visualization of these enzymes, and extensive structural and biochemical characterization has established the type of lesion-bypass specificity displayed by each group. However, very little is known as to the structural and/or kinetic parameter that determines nucleotide selection and lesion bypass specificity.

In Chapter 3 of this thesis, I have addressed this aspect in two closely related DinB enzymes. Dbh and Dpo4 both belong to the same genus of thermophilic archeon: Sulfolobus. Surprisingly, however, these two polymerases have different functional characteristics (overall activity, base substitution vs deletion rates and abasic site bypass ability) [78, 84, 100, 102-104, 112, 124, 128]. Additionally, in ternary structures the LF/PAD domain of these two polymerases is positioned very differently [75, 78, 84, 137]. An earlier report suggested that the various distinctive characteristics of Dbh and Dpo4 are due to the high level of sequence divergence in their LF/PAD domain [100]. Recent work from our lab has shown that these features can be attributed to the short, 15-amino acid linker connecting the LF/PAD to the polymerase domain [128].

I have further narrowed the structural determinants of polymerase specificity within the interdomain linker and find that 3 amino acids in this 15 residue
stretch is enough to alter Dbh into a more Dpo4-like enzyme: both kinetically and structurally. We have also determined the kinetic parameters relevant to nucleotide binding and incorporation for parent and chimeric enzymes to identify the factor that most influences fidelity in these polymerases. The results and implications as such are discussed further in Chapters 3 and 6.
PART ONE
CHAPTER 2

Mukherjee P, Lahiri I, Pata JD. Human polymerase kappa uses a template-slipage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. Nucleic Acids Res. 2013 May; 41(9):5024-35.

2.1. DESCRIPTION OF CONTRIBUTION

For this manuscript, I have designed and performed the experiments and analyzed the results with guidance from my mentor JDP and help from IL. The final manuscript was primarily written by me.
2.2. SUMMARY

Polymerases belonging to the DinB class of the Y-family translesion synthesis DNA polymerases have a preference for accurately and efficiently bypassing damaged guanosines. These DinB polymerases also generate single-base (-1) deletions at high frequencies with most occurring on repetitive “deletion hotspot” sequences. Human DNA polymerase kappa (hPolκ), the eukaryotic DinB homolog, displays an unusually efficient ability to extend from mispaired primer termini, either by extending directly from the mispair or by primer-template misalignment. This latter property explains how hPolκ creates single-base deletions in non-repetitive sequences, but does not address how deletions occur in repetitive deletion hotspots. Here we show that hPolκ uses a classical Streisinger template-slippage mechanism to generate -1 deletions in repetitive sequences, as do the bacterial and archaeal homologs. After the first nucleotide is added by template slippage, however, hPolκ can efficiently realign the primer-template duplex before continuing DNA synthesis. Strand realignment results in a base-substitution mutation, minimizing generation of more deleterious frameshift mutations. On non-repetitive sequences, we find that the rate of nucleotide misincorporation is lower if the incoming nucleotide can correctly basepair with the nucleotide immediately 5’ to the templating base, competing with the mispairing with the templating base.
2.3. INTRODUCTION

During replication, DNA polymerases frequently encounter unrepaired DNA lesions. Owing to their restrictive active sites and 3’→5’ exonuclease proofreading activities, replicative DNA polymerases tend to stall at sites of damage, which can lead to replication fork collapse, and which, unless rescued, can eventually cause large-scale genomic rearrangements and cell death. As a preventative measure, most organisms have alternative polymerases that are capable of “damage tolerance” and can synthesize past DNA lesions by a mechanism called translesion synthesis (TLS). However, these specialized polymerases are generally much more error-prone than replicative polymerases[58]. Most polymerases involved in TLS belong to the Y-family of DNA polymerases. Phylogenetic data further subdivides this family into 6 classes: two UmuC subfamilies, found only in bacteria; Rad30A (pol eta), Rad30B (pol iota) and Rev1 proteins, found only in eukaryotes; and the DinB subfamily that is found in all domains of life [46].

The role of the DinB class of polymerases in cells has not been as clear as for some other Y-family polymerases, despite the fact that they are so widely conserved. However, bacterial, archaeal and eukaryotic DinB polymerases all have the ability to efficiently and accurately bypass N2-adducted guanosines in vitro [115, 121, 122, 141, 142], and their absence makes cells more sensitive to DNA damaging agents [117, 141]. Furthermore, all of these polymerases incorporate dCTP much more efficiently than other nucleotides; a property presumed to relate to a role in bypassing damaged guanosines [104, 110, 119].

A characteristic mutational feature of DinB polymerases is their ability to generate single base deletions at high frequencies (~10² to 10⁴) on undamaged DNA sequences [100, 113, 118, 119, 143]. Most deletions have been found to occur on “deletion hotspots” containing a short run of identical pyrimidines followed by a 5’ G, but many deletions are also found in non-repetitive sequences where other polymerases rarely make deletions.
The human DinB ortholog, polymerase kappa (hpolκ), shares the lesion-bypass and mutational characteristics with the bacterial (DinB) and archaeal (Dpo4 and Dbh) enzymes, but it also has some distinctive features. One key feature is its ability to efficiently extend from mispaired bases located at the primer terminus [127, 129]. Unlike other polymerases, hpolκ extends from mispairs at a higher frequency than it generates mispairs, with dC-C mispairs being extended most readily [127]. Mispair extension can occur either by direct extension from the mispair, or by a misalignment mechanism, where the misincorporated base at the primer terminus forms a correct pair with the next templating base. A single-base deletion results when DNA synthesis continues from this misaligned conformation [127, 129]. Efficient mispair extension, combined with an inability to incorporate nucleotides opposite the 3' base of intra-strand cross-linked bases (cis-syn TT dimers and 6-4 photoproducts), led to the proposal that pol kappa works in concert with other TLS polymerases to extend DNA synthesis after another polymerase has copied the lesion itself [110, 127].

Another characteristic that distinguishes hpolκ from the other DinB polymerases (and from other classes of Y-family polymerases) is that it has a moderate degree of processivity, about 25 nucleotides incorporated per DNA binding event compared to at most a few nucleotides added by other Y-family polymerases [118]. This moderate processivity is conferred on hpolκ by a C-terminal extension of the enzyme that is not found in other members of the polymerase family. Because of this processivity, hpolκ has a higher potential to introduce mutations in the genome. In vitro analysis shows that hpolκ residues 1-560 and full-length polymerase generate nearly identical mutational spectra at similar frequencies, with average base-substitution and deletion frequencies of $7 \times 10^{-3}$ and $2 \times 10^{-3}$, respectively [118].

To date three mechanisms have been suggested for generation of single base deletions (Figure 2.1). On repetitive/iterative sequences, Streisinger slippage is thought to occur, where the intrinsic nature of the substrate allows DNA duplexes to readily misalign, leaving an extrahelical nucleotide in one strand (Figure 2.1B). After a round of replication, this can lead to addition or deletion
of the unpaired nucleotide, depending on whether the unpaired nucleotide is in the primer or template strand. When this occurs in an open reading frame, Streisinger slippage results in +1 or -1 frameshift mutations [144, 145]. For deletions to occur by this mechanism, the polymerase must be able to tolerate the unpaired nucleotide that is skipped during replication in the template strand, causing the newly synthesized primer strand to be shorter than the template.

For deletions on non-repetitive sequences, both misincorporation-misalignment and dNTP-stabilized misalignment mechanisms have been proposed [44, 146-149]. Misincorporation-misalignment requires incorporation of the incorrect nucleotide opposite the templating base (0 position), followed by isomerization of the template to allow base pairing between the newly added nucleotide and the next correct templating base (+1 position), resulting in the original templating base becoming unpaired at what becomes the -1 position (Figure 2.1C). dNTP stabilized misalignment, however, directly utilizes the nucleotide next to the templating base (+1 position) for incorporation, again resulting in the templating nucleotide at the 0 position to become extrahelical in the -1 position (Figure 2.1D).
Figure 2.1. Single base deletion mechanisms.

On a deletion "hotspot" containing a homopolymeric run of pyrimidines followed by a 5′G, DinB polymerases could (A) add the correct nucleotide (dGTP) or generate single base deletions in the presence of the incorrect nucleotide (dCTP) by any of three proposed mechanisms: (B) template slippage (C) misincorporation-misalignment and (D) dNTP-stabilized misalignment (see text for more details). The incoming nucleotide and newly added base are shown in italics (bold). The base positions assigned are shown with respect to the templating position defined as 0. “●” represents base-pairing before phosphodiester bond formation.
Mutational data show that hpolκ makes single-base deletions at higher rates in repetitive sequences [118]. This is consistent with a template slippage mechanism, but this has not been tested directly. Since the deletion rate on non-repetitive sequences is also high, hpolκ could use one of the other deletion mechanisms on both repetitive and non-repetitive sequences. The most likely alternative mechanism would be misincorporation-misalignment, since the misalignment part of this mechanism is used by hpolκ when extending from a mispaired primer terminus that is complementary to the next templating base [129].

Here we investigate how hpolκ generates deletion mutations, finding that hpolκ does primarily use a template-slippage mechanism in repetitive sequences, as do the archaeal and bacterial DinB polymerases, but hpolκ is also highly proficient at realigning the slipped strands, a property that has not been investigated previously. Furthermore, the rate of nucleotide misincorporation is highly dependent on sequence context.
2.4. MATERIALS AND METHODS

2.4.1. Protein purification

The human DNA polymerase kappa (κ) construct encoding for amino acids 1-560 (hPol κ1-560) was cloned into pGEX vector with a N-terminal GST tag and was expressed in Rosetta2 E.coli cells (EMD Millipore). For simplicity, we will refer to hPol κ1-560 as hPolκ, since earlier work has shown that constructs containing residues 1-560 and 1-526 have the same nucleotide incorporation activity and fidelity as the full-length polymerase [77, 118, 150]. Cells expressing the construct were grown at 37 °C until OD600 reached 0.7 and then induced with 0.5 mM IPTG overnight at 20 °C. All of the remaining steps were performed at 4 °C. Cell pellets were resuspended in Buffer A (50 mM Tris-Cl (pH 8.0), 300 mM NaCl, 0.5 mM EDTA, 10 mM β-mercaptoethanol and 10%(w/v) sucrose) containing lysozyme. Resuspended cells were lysed by sonication followed by centrifugation at 20,000 x g for 1 hr. Ammonium sulfate was added to the clear supernatant to a final concentration of 35% (0.208 g/ml) of supernatant to precipitate hPolκ. The precipitate was collected after centrifugation, resuspended in 35 ml of Buffer A and dialyzed overnight against Buffer A. The dialyzed protein was then loaded on (2x5 ml) GST columns (GE Healthcare) pre-equilibrated with Prescision protease buffer (50 mM Tris-Cl (pH 8.0), 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 10% glycerol). Columns were washed again before treating overnight with 1 column volume of Prescision protease buffer containing 1 mg/ml of HRV3C protease. The untagged hPolκ was obtained the next day by eluting with Prescision protease buffer and the eluate was immediately applied to (2x5 ml) HiTrap SP-columns (GE Healthcare), followed by gradient elution with 0.05-1 M NaCl in 20 mM HEPES (pH 7.25), 0.1 mM EDTA, 1 mM DTT and 10% glycerol. The eluted fractions were then dialyzed overnight into storage buffer containing 25 mM HEPES (pH 7.5), 50 mM NaCl, 0.5 mM EDTA, 5 mM DTT and 10% glycerol. After concentrating the protein using Amicon Ultra-15 filters with 10 kDa MW cutoff (Milipore) to achieve a desired concentration (calculated using a theoretical extinction coefficient of 28,880 M⁻¹ cm⁻¹).
1cm), the aliquots were frozen at -80 °C.

2.4.2. Duplex DNA formation

All DNA oligonucleotides (shown in Table 2.1) were obtained from Integrated DNA Technologies and primer oligonucleotides were synthesized with a 5' 6-carboxyfluorescein (6FAM) label. Primer and template oligonucleotides were annealed in buffer containing 10 mM Tris-Cl (pH 7.5) and 50 mM NaCl by heating to 95 °C followed by slow cooling to 25 °C.
Table 2.1: DNA Substrate Sequences.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 4C-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCC GCA TT -5’</td>
</tr>
<tr>
<td>2 1T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCT GCA TT -5’</td>
</tr>
<tr>
<td>3 2T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CTC GCA TT -5’</td>
</tr>
<tr>
<td>4 3T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC TCC GCA TT -5’</td>
</tr>
<tr>
<td>5 4T-G</td>
<td>5’-(FAM)-AGG CAC TGA TC GGG -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGT CCC GCA TT -5’</td>
</tr>
<tr>
<td>6 4C-A</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCC ACA TT -5’</td>
</tr>
<tr>
<td>7 1T-A</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCT ACA TT -5’</td>
</tr>
<tr>
<td>8 C-C mispair</td>
<td>5’-(FAM)-AGG CAC TGA TCG GGC -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCC GCA TT -5’</td>
</tr>
<tr>
<td>9 4C-GA</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCC GAA TT -5’</td>
</tr>
<tr>
<td>10 4G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G -3’</td>
</tr>
<tr>
<td></td>
<td>3’-CC GTG ACT AGC CCC GCA TT -5’</td>
</tr>
</tbody>
</table>

* Primers are labeled on the 5’ end with 6-carboxyfluorescein (FAM). Nucleotides in bold vary between substrates.
2.4.3. Pre-steady-state primer extension assays

All assays were done using a KinTek RQF-3 rapid quench instrument (KinTek Corp.). To achieve single turnover conditions, where all of the DNA substrate is prebound by polymerase, an excess of hPolκ (30 µM) was preincubated with 100 nM DNA at room temperature (25 °C) in reaction buffer (25 mM MES-Tris (pH 7.5), 25 mM NaCl, 8 mM MgCl₂, 2 mM DTT and 10% glycerol). Reactions were initiated by mixing with an equal volume of 2 mM dNTP in the same buffer (final concentrations were 15 µM hPolκ, 50 nM DNA substrate and 1 mM dNTP) and quenched with 250 mM EDTA at appropriate time points (0.02 to 60s). The extended product was separated from unextended primer on a 17.5% (19:1) acrylamide-(1x)TBE gel containing 7.5M urea. The gel was imaged using a Typhoon 9400 scanner (GE Healthcare) and bands were quantitated using ImageQuant software (GE Healthcare). Percentage of primer extension was determined by measuring the relative intensity of the band corresponding to the extended primer with respect to the total labeled DNA (i.e. both extended and unextended primer strands) (Equation 2.1).

\[
\%\text{extension} = \frac{100 \times \text{[extended]}}{\text{[unextended]} + \text{[extended]}} \tag{2.1}
\]

Where applicable, the amount of extended primer includes sum of primer extension seen as multiple bands when extended by more than one nucleotide.

Most data were fit to an exponential equation (Equation 2.2) unless a biphasic nature was observed in which case the burst equation (Equation 2.3) was used:

\[
y = A(1 - e^{-k_{\text{obs}}t}) + c \tag{2.2}
\]

where \(A\) is amplitude, \(k_{\text{obs}}\) is the observed rate of the reaction, \(t\) is the time after which reaction was quenched and \(c\) is a constant.

\[
y = A(1 - e^{-k_1 t} + k_2 t) + c \tag{2.3}
\]
where $A$ is the amplitude of the fast phase, $k_1$ is rate of the fast phase of the reaction and $k_2$ is the rate of the slow phase, $t$ is the time after which reaction was quenched and $c$ is a constant.

Experiments were performed in duplicate with two different protein preparations and error bars represent the standard deviation of the data collected. All of the graphs and nonlinear regressions were done using GraphPad Prism, version 6.0a (GraphPad Software Inc.). Observed nucleotide incorporation rates for all experiments are summarized in Table 2.2.
Table 2.2: Summary of observed nucleotide incorporation rates.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Observed rate ($s^{-1}$)</th>
<th>dCTP addition</th>
<th>dGTP addition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td>4C-G</td>
<td></td>
<td>2 ± 0.5</td>
<td>0.146 ± 0.08</td>
</tr>
<tr>
<td>1T-G</td>
<td></td>
<td></td>
<td>0.44 ± 0.03</td>
</tr>
<tr>
<td>2T-G</td>
<td></td>
<td></td>
<td>0.38 ± 0.04</td>
</tr>
<tr>
<td>3T-G</td>
<td></td>
<td></td>
<td>3.47 ± 0.34</td>
</tr>
<tr>
<td>4T-G</td>
<td></td>
<td></td>
<td>1.59 ± 0.14</td>
</tr>
<tr>
<td>4C-A</td>
<td></td>
<td></td>
<td>0.12 ± 0.01</td>
</tr>
<tr>
<td>1T-A</td>
<td></td>
<td></td>
<td>1.14 ± 0.09</td>
</tr>
<tr>
<td>C-C mispair</td>
<td></td>
<td></td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>4C-GA</td>
<td></td>
<td>2.5 ± 0.5</td>
<td>0.153 ± 0.02</td>
</tr>
<tr>
<td>4G</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.5. RESULTS

2.5.1. The template sequence 3'-CCCCG-5’ is a deletion hotspot for hPolκ.

To determine if hPolκ is error prone on the “deletion hotspot” sequence characterized for other DinB polymerases [78, 84, 100, 113, 119], we used DNA substrate 4C-G which contains a homopolymeric run of four C's followed by a 5' G on the template strand (3'-CCCCG-5', Table 2.1). On this substrate, hPolκ adds the correct incoming nucleotide, dGTP, at a rate that is only 3.6-fold faster ($k_{obs}$ of 2.62 ± 0.39 s⁻¹) than the rate of incorrect nucleotide, dCTP ($k_{obs}$ of 0.72 ± 0.04 s⁻¹) addition (Figure 2.2A), indicating that deletions can be initiated on this substrate at a very high frequency.

Interestingly, the dGTP (correct) addition follows biphasic behavior with a burst amplitude of 56.2 ± 3.2 nM and a slow second phase rate of 0.036 ± 0.012 s⁻¹. This was unexpected because the assay was designed to be performed under single-turnover conditions, where all of the DNA is initially bound by polymerase so that the DNA would all be extended in one cycle without polymerase having to rebind. Under these identical conditions, dCTP (incorrect nucleotide) addition by hPolκ displays only a single phase of product formation, which is consistent with single turnover conditions and accordingly fits well to the single exponential equation (Equation 2.2) with an amplitude of 77.41 ± 1.63 nM.
Figure 2.2: Nucleotide incorporation by hPolκ on repetitive sequence containing deletion hotspot.
Figure 2.2. Nucleotide incorporation by hPolκ on repetitive sequence containing deletion hotspot.

15 µM of hPolκ was preincubated with 50 nM DNA before initiating the reaction with addition of 1 mM dNTP (final concentrations). (A) On 4C-G (substrate 1), containing a run of four C’s followed by a 5’ G, hPolκ adds dGTP (○) ~4-fold faster than dCTP (●). (B) Modification of the nucleotide 5’ to the templating position (i.e. +1 position) from G to A (4C-A(□), substrate 6) results in reduction in overall $k_{\text{obs}}$ for incorporation of dCTP by 6-fold from $0.72 \pm 0.04 \text{ s}^{-1}$ to $0.12 \pm 0.01 \text{ s}^{-1}$. (C) Pre-steady-state assays were performed under single turnover conditions using DNA templates containing unpaired T’s at every C position in the homopolymeric run of 4C-G (substrates 2-5) to test if hPolκ can tolerate the presence of an extrahelical nucleotide at these positions. dCTP incorporation by hPolκ on 1T-G(▲), 2T-G(△), 3T-G(●) and 4T-G(○) substrates is shown. hPolκ extends the various bulged templates with the -3T bulge (3T-G, substrate 4) being utilized most efficiently. (D) Modification of the +1 G position of the 1T-G substrate (▲) to an A (1T-A (▽), substrate 7) resulted in a 3-fold increase in $k_{\text{obs}}$ from $0.44 \pm 0.03 \text{ s}^{-1}$ to $1.14 \pm 0.09 \text{ s}^{-1}$.
Hence, the biphasic nature of product formation for dGTP addition can not be explained as simply as a result of the presence of initial unbound DNA. Because the two different reactions used the same primer-template (p/t) DNA and polymerase preparation, the difference in amplitudes (56 nM for dG vs. 77 nM for dC) is dependent on the identity of the incoming dNTP, which must influence the fraction of active enzyme complex. Previously, the active fraction of hPolκ was found to be ~8-fold larger when extending from a mispair than from a perfectly base paired p/t DNA junction [151]. It was suggested that this characteristic might imply the preference for mispaired rather than correctly paired DNA substrates. Here we find that even when supplied with a perfectly base paired p/t DNA, the active enzyme concentration of hPolκ varies based on the incoming dNTP. When dCTP is added to initiate a deletion, the nucleotide skipped is expected to adopt an extra-helical conformation which could thereby disrupt the B-form conformation of the DNA. This alteration of DNA conformation may be the key difference that could favorably influence the active enzyme concentration of hPolκ.

2.5.2. hPolκ predominantly uses a template slippage deletion mechanism on iterative sequences.

Since hPolκ has been reported to use a misalignment mechanism to make deletions on non-repetitive sequences [129], we first investigated if the incorporation of dC on the deletion hotspot sequence also occurred by this mechanism. Of the three mechanisms for single base deletion (Figure 2.1), only misincorporation-misalignment does not depend on the nucleotide on the 5' side of the templating position (+1 position) for templating dC addition, since in this mechanism the first nucleotide would be added by mispairing with the C at position 0 (Figure 2.1C). In the 4C-G substrate, a G occupies this position. To determine if hPolκ utilizes this +1G when generating deletions, we changed it from G to A (Table 2.1, 4C-A). On testing for addition of dCTP on this template, we observed that modifying the base at the +1 position reduced the $k_{\text{obs}}$ 6-fold, from $0.72 \pm 0.04$ s$^{-1}$ to $0.12 \pm 0.01$ s$^{-1}$ (Figure 2.2B). This suggested a clear dependence on the +1G, thus ruling out misincorporation-misalignment as the
major deletion mechanism used by hPolκ on this repetitive sequence.

DNA templates with a repetitive sequence can exist in multiple conformations with the same primer, resulting in alternate base pairing situations that could lead to deletions or additions after a round a replication [144, 145]. In an attempt to localize and stabilize an extrahelical nucleotide at specific positions in the homopolymeric run of C’s of the 4C-G substrate (3’-CCCCG-5’), we systematically modified each of the four C’s on the template strand to a T one at a time (Table 2.1, substrates 2-5). The higher energy penalty of forming a G-T mispair compared to a G-C pair makes the former less favored, biasing the DNA to adopt a conformation with an unpaired nucleotide at the altered position [78]. We used this assay to test if hPolκ is able to tolerate and extend p/t DNA with an extrahelical nucleotide at positions -4, -3, -2 and -1 of the template (Figure 2.1B and D; Figure 2.2C). Efficient use of the templates with T at positions -4, -3 and -2 would indicate use of template slippage. Efficient use of the template with the T at the -1 position would indicate use of either dNTP-stabilized misalignment or misincorporation-misalignment deletion mechanism. Having already ruled out misincorporation-misalignment for the 4C-G template, we used these altered substrates to distinguish template slippage from dNTP-stabilized misalignment. Similar DNA substrates have been used to assess deletion mechanisms for other DinB polymerases [78, 84, 107].

We found that hPolκ showed a strong preference for unpaired nucleotides present at positions -3 and -4 (substrates 3T-G and 4T-G) with $k_{\text{obs}}$ of $3.47 \pm 0.34$ s$^{-1}$ and $1.59 \pm 0.14$ s$^{-1}$ respectively (Figure 2.2C). Interestingly, these data showed a biphasic behavior, so were fit to a burst equation (Equation 2.3), as was done for the dGTP (correct) addition on the 4C-G substrate. An unpaired base at the -2 position is utilized ~10 fold less efficiently than at the -3 position, with a $k_{\text{obs}}$ of $0.38 \pm 0.04$ s$^{-1}$ suggesting that proximity to the active site reduces tolerance for an extrahelical nucleotide. This preference for an unpaired base further away from the active site has been observed for other DinB polymerases as well [78, 84, 107].
The ability to efficiently utilize DNA substrates containing an unpaired nucleotide at the -4, -3 and -2 positions (substrates 3-5, Table 2.1) indicates that hPolκ predominantly uses a template slippage mechanism to generate single base deletions. Compared to the 4C-G substrate, the dC incorporation rate is faster on the 4T-G and 3T-G substrates but slower on the 2T-G substrate. This suggests that the 4C-G template adopts multiple conformations and the observed nucleotide incorporation rate results from a combination of different rates on substrates with different conformations.

Although we found that hPolκ is able to use slipped template DNA sequences efficiently, it is also able to add dCTP on 1T-G (Table 2.1), which has a T at the 0 position, with a $k_{obs}$ of $0.44 \pm 0.03$ s$^{-1}$ (Figure 2.2C), comparable to the rate of addition on the 2T-G substrate and only ~50% more slowly than on the 4C-G substrate. This left open the possibility that hpolκ could potentially use dNTP-stabilized misalignment or misinsertion-misalignment on this substrate, even though the latter had been ruled out on the original deletion hotspot sequence.

2.5.3. hPolκ readily misincorporates dC when the deletion hotspot is changed to 3’-CCCTG-5’

We then considered the possibility that hPolκ used entirely different nucleotide incorporation mechanisms on substrates 4C-G and 1T-G, despite differing by only a single nucleotide. To test this possibility, we altered the template sequence on the 1T-G substrate such that the +1G was modified to A (1T-A, substrate 7). If dCTP addition on this substrate is through dNTP-stabilized misalignment, then the +1G on the template would prove crucial for extension and a considerable reduction in rate of addition may be expected. However, if incorrect nucleotide addition was independent of the +1G no such decrease would occur, suggesting that dCTP addition on the 4C-G substrate occurs by a different mechanism than on the 1T-G substrate.

Surprisingly, on 1T-A (i.e. 1T-G with +1G modified to A) a 2.6-fold increase in $k_{obs}$ from $0.44 \pm 0.03$ s$^{-1}$ to $1.14 \pm 0.09$ s$^{-1}$ was observed (Figure 2.2D), in contrast to
the 6-fold decrease observed on the 4C-A substrate compared to the 4C-G substrate (Figure 2.2B). Absence of a rate decrease when the +1G is altered excludes the possibility that dNTP-stabilized misalignment occurs in the sequence context found in the 1T-G substrate. Earlier genetic evidence has shown that hPolκ produces T-dCMP mispairs at a very high frequency of $8.2 \times 10^{-3}$ [118]. On the 1T-G substrate, therefore, dCTP addition most likely occurs by misincorporation opposite the templating T. We conclude that dC addition occurs by misincorporation on both the 1T-G and 1T-A substrates, but the sequence context influences the efficiency of misincorporation, as evidenced by the 1T-A sequence context stimulating incorporation. The mechanism for the increased rate of dC incorporation on the 1T-A substrate is discussed later.

2.5.4. **hPolκ efficiently realigns primer-template DNA after initiating a deletion.**

We observed that hPolκ performs multiple additions on most DNA substrates examined, even when provided with just a single nucleotide. On the 4C-G substrate this was true for both dGTP (correct) and dCTP (incorrect) incorporations (Figures 2.3A, panel (i), and 2.4B, panel (i)), but to differing extents. Interestingly, we found that multiple addition of dGTP followed a very different pattern than that observed for dCTP, indicating dissimilarity in mechanism. To better understand the relevance of this observation and ascertain its mechanistic implications, we decided to investigate how hPolκ performs multiple nucleotide additions on the 4C-G substrate.
Figure 2.3. Mechanism of second dGTP incorporation on 4C-G substrate.
**Figure 2.3. Mechanism of second dGTP incorporation on 4C-G substrate.**

(A) Pre-steady-state assays under single turnover conditions were performed. (i) and (ii) show sequencing gels depicting multiple dG additions on substrates 4C-G and 4G respectively. (B) Plot of % extension versus time for both 4C-G (○) and 4G (x) substrates. Addition of first dG (○) plateaus at ~60% followed by slow misincorporation of the 2nd dGTP (●). Rate of second dG incorporation on the 4C-G corresponds well with misincorporation of dGTP on the 4G substrate (x) that has the 1st dGTP already added. (C) Schematic showing the probable mechanism of 2nd dGTP addition on 4C-G substrate. The incoming nucleotide and newly added base are shown in italics (bold). The base positions assigned are shown with respect to the templating position defined as 0. “● ” represents base-pairing before bond formation.
Figure 2.4. Mechanism of multiple dCTP incorporations.
**Figure 2.4. Mechanism of multiple dCTP incorporations.**

(A) Schematic representation of the two possible ways in which hPolκ can efficiently add a second dCTP on 4C-G (substrate 1). The incoming nucleotide and newly added base are shown in italics (bold). The base positions assigned are shown with respect to the templating position defined as 0. “●” represents base-pairing before bond formation. “★” indicates the +2 templating position. (B) Gels showing the unique pattern of multiple incorporations of dCTP on (i) 4C-G (substrate 1) and (ii) 4C-GA (substrate 9) to test for mispair formation. (iii) Gel showing the incorporation of dC to extend from a C-C mispair (substrate 8) suggests hPolκ’s ability to realign. (C) Plot shows that mechanism of multiple dC additions is distinct from dG additions on 4C-G substrate. First dCTP addition (●) reaches a maximum of ~40% extension, after which increase in second dCTP incorporation (●) occurs exponentially, implying that most of the primer extended by one nucleotide is rapidly extended by two. (D) Overall dCTP incorporation on the 4C-G (●), 4C-GA (□) and C-C mispair substrates (■) shown with respect to time along with percent of second dCTP addition as seen for the 4C-G (●) and 4C-GA (□) substrates. Second dCTP incorporation traces are shown as dotted lines.
We found that for dGTP addition on the 4C-G substrate, the second dG is added slowly ($k_{\text{obs}} \sim 0.09 \text{ s}^{-1}$), after rapid addition of the first dG ($k_{\text{obs}}$ of $1.6 \pm 0.3 \text{ s}^{-1}$) (Figure 2.3A, panel (i)), suggesting a gradual misincorporation opposite the +1G after the correct nucleotide at position 0 had been added (Figure 2.3C). To test this possibility we designed substrates that already had the first nucleotide incorporated as a G (Table 2.1, substrate 4G) and tested the rate of insertion of the next nucleotide. The rate for the second dG addition ($0.09 \text{ s}^{-1}$) corresponds well with the misincorporation rate ($k_{\text{obs}} \sim 0.11 \text{ s}^{-1}$) (Figure 2.3A(ii) and 2.3B), confirming this interpretation.

In contrast to the slow rate of addition of a second dG, we found that addition of a second dC is considerably faster and followed an entirely different pattern (Figure 2.3A and B vs. Figure 2.4B and C). As the data above indicate, the first addition of dC occurs predominantly by template slippage, allowing the proper base pairing of the dCTP opposite the +1G. We can envision two possibilities for the addition of the second dC (Figure 2.4A). One possibility is that the next dCTP is misincorporated opposite the templating C at +2 to generate a C-dCMP mispair. A second possibility is that the extrahelical nucleotide generated during the first dCTP addition is rearranged such that now there is a C-C mispair generated at the p/t junction. As hPolκ is an efficient extender of mispairs [127], it could tolerate the mispair to add the second dCTP correctly paired opposite the +1G.

To distinguish between these mechanisms, two different substrates were designed: one with the +2C modified to an A (Table 2.1, 4C-GA) and another where the primer contained a 3'-C to generate a mispair at the p/t junction (Table 2.1, C-C mispair). Modification of the +2C to an A (4C-GA) did not alter the overall rate of nucleotide addition; $k_{\text{obs}}$ for 4C-G and 4C-GA were $0.72 \pm 0.04 \text{ s}^{-1}$ and $0.79 \pm 0.03 \text{ s}^{-1}$ respectively (Figure 2.4B and D). Under steady-state conditions, dC-A incorporation is 3.8-fold less efficient than dC-C (comparing $V_{\text{max}}/K_m$ of $8 \times 10^4$ for dC-A and $3 \times 10^3$ for dC-C) [110]. Thus, the unchanged incorporation rate upon substituting the +2C to an A indicates that the base in the +2 position is not used to template the second dC addition.
Dissecting this data further, the rates of second dC incorporation on 4C-G and 4C-GA substrates were also found to be essentially identical (4C-G: 0.146 s\(^{-1}\) and 4C-GA: 0.153 s\(^{-1}\)), but dCTP incorporation from a C-C mispair was two-fold faster (0.33 ± 0.02 s\(^{-1}\)) (Figure 2.4D). This observation can be explained since time taken for rearrangement after the first nucleotide addition might slow the rate of second nucleotide addition when compared to a substrate with a preformed C-C mispair at the p/t junction. These data strongly suggest that after the first dCTP addition, rearrangement at the p/t junction occurs and allows extension from a C-C mispair for the second incorporation. From these experiments, we conclude that after initiating a deletion by incorporation of dCTP on the deletion hotspot sequence 3'-CCCCG-5', hPolκ can efficiently realign the slipped p/t junction to correctly incorporate a second dC from a C-C mispaired p/t junction. Notably, this is the mispair that previously was found to be most efficiently extended by hpolκ [110].

**2.5.5. Sequence context influences the rate of nucleotide incorporation.**

The ability to incorporate a single nucleotide multiple times was also observed on the 1T-G and 1T-A substrates (Figure 2.5A and D). As demonstrated above, the addition of the first dCTP on 1T-G substrate occurred by misincorporation, since modification of the +1G to A in the 1T-A substrate resulted in ~3-fold increase in overall rate. Although this clearly indicated that dNTP-stabilized misalignment was not the predominant incorporation mechanism, the increase in rate was puzzling. To understand this better, we calculated the amount of extension for the first and second additions separately (Figure 2.5B and E) and found that different patterns and rates of incorporation on the two substrates provided further information to decipher the mechanism.
Figure 2.5. hPolκ misincorporates efficiently on 1T-G substrate.
Figure 2.5. hPolκ misincorporates efficiently on 1T-G substrate.

(A) Gel and (B) graph showing multiple incorporations of dCTP on the 1T-G substrate. First addition (▲), second addition (▲) and overall addition (▲). (C) Schematic showing possible mechanisms of dCTP addition on the 1T-G substrate. (D) Gel and (E) graph showing multiple incorporations of dCTP on the 1T-A substrate. First addition (), second addition () and overall addition (◆). (F) Schematic showing possible mechanisms of dCTP addition on the 1T-A substrate. In (C) and (F) the incoming nucleotide and newly added base are shown in italics (bold). The base positions assigned are shown with respect to the templating position defined as 0. “◆” represents base-pairing before bond formation.
On the 1T-G substrate, the amount of primer extended by just one nucleotide reached a maximum level of ~30%, at which point the level decreased as the second dC was being added rapidly (Figure 2.5A and B). On the 1T-A substrate, in contrast, the amount of primer extended by one nucleotide reached a maximum level of ~80% (Figure 2.5D and F) before dropping as the second dC was being added slowly. For each substrate, the overall rate of primer extension (1.14 s⁻¹ for 1T-A; 0.44 s⁻¹ for 1T-G) was dominated by the rate of the first nucleotide incorporation.

We interpret these results as shown in Figure 2.5C and F. On both substrates, the first dC addition is templated by the T in position 0. On the 1T-G substrate (Figure 2.5C), the second dC is templated by the G at the +1 position, and the rate of this incorporation is fast because hpolκ can efficiently add the next correct nucleotide even from a mispaired primer terminus [127]. On the 1T-A substrate (Figure 2.5F), the second dC is templated by the A at the +1 position, but the rate is slow because dC-A is a mispair and previous steady-state data show that it is the slowest mispair formed by hpolκ [110]. The fast rate of second nucleotide addition on 1T-G would lower the amplitude of the primer extended by one nucleotide as this serves as the substrate for the second nucleotide addition. On the 1T-A substrate, the amplitude of the first addition would be higher because the primer extended by one nucleotide would only slowly be extended by a second nucleotide.

The lower overall rate of nucleotide incorporation (the sum of the first and second additions) on the 1T-G substrate can be explained if the dCTP can stably pair both with the T at the 0 position and with the G at the +1 position, and if the rate of addition is negligible for the nucleotide paired at the +1 position. In this case, only a fraction of the substrate initially has the nucleotide paired at the 0 position where it can be incorporated efficiently. Because ~80% of the primer eventually becomes extended, the nucleotide paired with the +1G must slowly shift to mispair with the T at position 0, thus lowering the overall rate of addition. For the 1T-A substrate, the dCTP is unlikely to be able to pair stably with the +1A and not compete for binding at position 0. Thus more of the
substrate will initially have the nucleotide paired with the T at position 0 where it can be incorporated efficiently, increasing the overall rate of nucleotide addition on the 1T-A substrate.
2.6. DISCUSSION

Since hpolκ is the most processive of the human Y-family polymerases, it has the greatest potential to introduce mutations into the genome [118]. Investigating the mutational mechanisms of hpolκ is therefore important for understanding the mutagenic consequences of DNA damage tolerance by translesion synthesis. Through the studies reported here, we have found that template slippage is strongly preferred as the mechanism of deletion by hpolκ on a repetitive deletion hotspot 3′-CCCG-5′ template sequence. Strikingly, however, our results indicate that hpolκ is able to efficiently realigning the DNA strands, resulting in a mispaired primer terminus rather than an extrahelical template base. Extension from the mispaired primer terminus, which hpolκ is able to do quite proficiently [127], would result in a less deleterious base-substitution mutation being generated rather than a frameshift mutation.

A distinctive feature of the mutation spectrum produced by hpolκ when replicating undamaged DNA is that deletions are formed quite readily, with only a 2- to 3-fold difference existing between single base deletion error rates observed on iterative (1.6-3.1x10⁻³) versus non-iterative (1.0x10⁻³) sequences [118]. Our findings help explain why there is such a small difference in single base deletion rates. On repetitive sequences, hPolκ can convert some of the deletions that are initiated into base substitution by realigning the slipped sequence and extending from the resulting mispaired primer terminus. Additionally, on non-iterative sequences, hPolκ has been found to extend from mispaired primer-template termini by initiating template-primer misalignment when direct extension does not occur [129]. This would result in an increase in single base deletions on non-repetitive sequences. Taken together these biochemical experiments corroborate the earlier mutational studies showing high rates of T•dC misincorporation and of deletion of Ts that are located in a 3′-TG-5′ sequence context [118].

We have not found any evidence that hpolκ is able to use a dNTP-stabilized misalignment mechanism to create deletions, even on non-repetitive sequences.
In fact, the preponderance of data indicates that none of the DinB polymerases is able to efficiently use this mechanism [78, 84, 103, 105, 107]. The first suggestion that Y-family polymerases might use a dNTP-stabilized misalignment deletion mechanism came from one of the first two Y-family structures bound to DNA: the Dpo4 ‘type II’ complex, which was crystallized in an attempt to determine the structure of a mispair in the active site. Instead of forming a mispair, however, the incoming nucleotide was found paired with the next templating base, which happened to be complementary to the dNTP used [75]. This pairing configuration is the hallmark of dNTP-stabilized misalignment, however it was puzzling that the skipped template base was stacked within the DNA duplex instead of adopting an extrahelical conformation [75]. Fluorescence measurements on both Dbh and DinB have shown that a skipped template base becomes unstacked from the surrounding DNA duplex when a single-base deletion is made [103, 105]. An extrahelical conformation has also been observed in crystal structures of both Dbh and Dpo4 with unpaired bases at positions -3 and -4 in templates that contain deletion hotspots [78, 84].

Our data demonstrate that misincorporation is slower when the incoming nucleotide can pair correctly with the base in the +1 position rather than pairing incorrectly with the base in the 0 position (Figure 2.5). Unless the skipped base adopts an extrahelical conformation, the primer terminus and alpha phosphate would have difficulty approaching each other closely enough for efficient catalysis. Thus, rather than showing a deletion being formed by dNTP-stabilized misalignment, we believe that the Dpo4 type II structure shows instead how the rate of nucleotide misincorporation can be slowed depending on the sequence context. This seems to be the case in the type II structure, since nucleotide addition did not occur even though neither the primer terminus nor incoming nucleotide were altered to prevent catalysis during crystallization.

Most polymerases show a strong dependence of deletion frequency on homopolymeric run length. Y-family polymerases differ from other polymerase families in not having a very strong dependence of deletion frequency on run length, with a run of just two nucleotides already resulting in a high deletion
frequency. The structure of hpolκ (Figure 2.6) shows that nucleotides 3’ of the templating base are adjacent to a large solvent-accessible gap between the polymerase and polymerase-associated domains of the enzyme, suggesting that bulged bases can readily be accommodated in this area. Crystal structures of the archaeal DinB polymerases Dbh and Dpo4 show how extrahelical nucleotides at positions -3 and -4 interact with the protein in this region [78, 84]. In contrast, polymerases from other families have much tighter constraints around the DNA duplex, which would suppress nucleotides from adopting extrahelical conformations.
Figure 2.6. Structure of an hPolκ ternary complex.
Figure 2.6. Structure of an hPolκ ternary complex.

(A) View looking into the active site of the polymerase. (B) View looking at the template strand of DNA entering the active site, between the polymerase and polymerase-associated domains. The polymerase (PDB code 3IN5 [128]) is shown in surface representation, except for the N-clasp (yellow), which is shown in ribbons representation. The polymerase domain is composed of fingers (blue), palm (magenta) and thumb (green) subdomains and is connected to the polymerase-associated domain (orange) by a relatively unstructured polypeptide linker (white). DNA is colored white, except for the templating base and nucleotides at positions -1 through -4 on the 3’ side of the templating base. The first residue visible in the structure (amino acid 25) is marked with a yellow asterisk.
Recent data indicate that the overall polymerase conformation strongly influences differences in Y-family polymerase fidelity. Structural and biochemical comparisons of Dbh, Dpo4 and chimeric constructs of the two polymerases demonstrated that close contacts between the catalytic and polymerase-associated domains next to the templating base led to a higher rate of nucleotide misincorporation and a higher propensity for adding multiple nucleotides [125]. In the case of hpolκ (Figure 2.6), the polymerase-associated domain does not directly contact the fingers subdomain, but the two domains are bridged by the N-clasp as mentioned above, which could stabilize the protein in a conformation that favors nucleotide misincorporation. The N-clasp is only found in the eukaryotic DinB polymerases. Deletion of the first 19 amino acids of the N-clasp reduces the ability of hpolκ to extend from mispaired primer termini without significantly reducing overall polymerase activity [150]. The precise structural role of these residues remains to be determined, since they were not included in any of the constructs of hpolκ that have yielded crystals, but they are likely to reach into the polymerase active site, on the major groove side of the nascent basepair and primer terminus where they could directly contact and stabilize mispairs (Figure 2.6A).

Overall, hpolκ displays a remarkable ability to create deletion and base substitution mutations using both template slippage and nucleotide misincorporation mechanisms. As for other DinB polymerases, hpolκ has a strong preference for creating deletions in repetitive sequences by template slippage, but is exceptionally proficient at realigning slipped DNA strands and extending DNA synthesis from mispaired primer termini. These unique properties give hpolκ a mechanism for suppressing deletion mutations at the expense of increasing base-substitution mutations.

2.7. FUNDING

National Institutes of Health [GM080573 to J.D.P.]
2.8. ACKNOWLEDGEMENTS

We thank Xiaojun Huang for constructing the hPol $\kappa_{1560}$ clone, Zhigang Wang for providing the full-length clone of hPol $\kappa$, and Ryan Wilson for careful proofreading of the manuscript.

Conflict of interest statement. None declared.
CHAPTER 3


3.1. DESCRIPTION OF CONTRIBUTION

For this manuscript, I designed the experiments, analyzed the data and performed the structure refinement. I also helped with data collection. The final manuscript was written by me.
3.2. SUMMARY

Dpo4 and Dbh are from two closely related *Sulfolobus* species and are well-studied archaeal homologs of Pol IV, an error prone Y-family polymerase from *E. coli*. Despite sharing 54% amino acid identity, these polymerases display distinct mutagenic and translesion specificities. Structurally, Dpo4 and Dbh adopt different conformations owing to the difference in relative orientation of their N-terminal catalytic and C-terminal DNA binding domains. Using chimeric constructs of these two polymerases, we have previously demonstrated that the interdomain linker is a major determinant of polymerase conformation, base substitution fidelity and abasic site translesion synthesis. Here we find that the interdomain linker also affects the single-base deletion frequency and the mispair extension efficiency of these polymerases. Exchanging just three amino acids in the linkers of Dbh and Dpo4 is sufficient to change the fidelity by up to 30-fold, predominantly by altering the rate of correct (but not incorrect) nucleotide incorporation. Additionally, from a 2.4Å resolution crystal structure, we have found that the three linker amino acids from Dpo4 are sufficient to allow Dbh to adopt the standard conformation of Dpo4. Thus, a small region of the interdomain linker, located more than 11 Å away from the catalytic residues, determines the fidelity of these Y-family polymerases, by controlling the alignment of substrates at the active site.
3.3. INTRODUCTION

DNA polymerases are the enzymatic workhorses that ensure effective genome duplication through multiple rounds of nucleotide addition. Translesion synthesis (TLS) polymerases have the ability to effectively bypass sites of DNA damage, an activity that is crucial to cell survival [58]. Functionally, TLS polymerases lack a 3’ to 5’ proofreading activity and, structurally, they lack the extensive protein-DNA contacts found in their replicative counterparts that ensure a tightly constrained active site. Thus, these polymerases are unusually mutagenic on undamaged DNA.

Most TLS polymerases belong to the Y-family of DNA polymerases [46]. These can be further categorized based on sequence similarity into six types that include the ubiquitous DinB family, the two UmuC families found only in bacteria and the Rad30A (pol eta), Rad30B (pol iota) and Rev1 families found only in eukaryotes. These enzymes display individual substrate specificities in the preferential bypass of certain types of DNA damage [58]. TLS can be error-free or error-prone, depending on the polymerase and the lesion. Even when replicating undamaged DNA, the Y-family polymerases each display unique mutagenic signatures (see for example [110, 151-155]).

In the past decade, a large number of studies have focused on understanding the correlation between structure and function in the Y-family TLS polymerases [156]. These enzymes share a core structure consisting of an N-terminal catalytic domain, with palm, fingers and thumb sub-domains (as are found in other families of DNA polymerases), and a C-terminal domain (unique to the Y-family) that is known as the “little finger” or polymerase-associated domain (LF/PAD) [75, 157]. While much is known about which specific DNA lesions are bypassed by individual polymerases, we still do not have a comprehensive understanding of which features of polymerase architecture are responsible for providing lesion bypass selectivity and mutagenic specificity. This knowledge is highly significant, since mutations made by Y-family polymerases predispose cells to developing cancer or antibiotic resistance.
In 2004, Boudsocq et al. reported that many of the lesion-bypass and mutational activities of two closely related archaeal DinB homologues, Dbh and Dpo4, were largely dependent on sequences outside of the catalytic domain [100]. Dbh and Dpo4 are from two different strains of *Sulfolobus* and share 54% sequence identity, yet Dpo4 is able to bypass abasic sites and thymidine dimers, while Dbh cannot. Dpo4 makes more base-substitution mutations, while Dbh has a higher propensity for making single-base deletion mutations [100, 113, 119].

We have shown recently that the 15 residue linker connecting the polymerase and LF/PAD domains of Dbh and Dpo4 (Fig. 3.1) can alone influence the enzymatic activity and selectivity of these polymerases [128]. A ternary complex structure of a chimera with the Dbh polymerase core and LF/PAD but the Dpo4 linker (Dbh-Dpo4-Dbh) adopts a Dpo4-like conformation. Like Dpo4, the LF/PAD is in contact with the fingers domain, docking into the major groove of the DNA duplex and positioning the primer-template junction at the active site for efficient catalysis. This chimera was also found to bypass an abasic site and display single nucleotide incorporation fidelity similar to Dpo4.

Here we extend our previous studies of chimeric constructs of Dbh and Dpo4 and find that the single-base deletion activity of these polymerases is also dependent on the linker identity. Furthermore, just three residues in the interdomain linker control the enzyme conformation and influence fidelity by affecting the rate of nucleotide incorporation, without being in the vicinity of the active site.
Both Dbh and Dpo4 have a polymerase core consisting of palm (magenta), fingers (blue) and thumb (green) domains in addition to a C-terminal little finger/polymerase-associated domain (LF/PAD) (orange). The amino acid sequence of the linker connecting the polymerase core to the LF/PAD is shown. Residues highlighted in red have been interchanged to create chimeras. Chimeric polymerases are named by the parental source of each domain, in the order polymerase-linker-LF/PAD. Superscript indicates the swapped residues present in the chimeric polymerase.
3.4. EXPERIMENTAL PROCEDURES

3.4.1. Protein expression and purification. Dbh expression and purification were performed as described before [78]. Dpo4 and all the chimeric constructs used had C-terminal 6x-His tags and growth and purification steps were performed as described before [84, 128].

3.4.2. Primer-template DNA. All DNA substrates used in this study are listed in Table 3.1 and were synthesized from Integrated DNA Technologies. The primer used for extension assays was synthesized with a 5'-6-carboxyfluorescein (FAM) label; the DNA used for crystallization was unlabeled. Primer was annealed to template in annealing buffer containing 10 mM HEPES (pH 7.5) and 50 mM NaCl.

3.4.3. Deletion assays. Reaction mixtures contained a final concentration of 40 nM annealed primer-template (p/t) DNA, 4 µM polymerase, 20 mM HEPES (pH 7.0), 85 mM NaCl, 5 mM MgCl₂, 1 mM DTT, and 1 mM dCTP or dGTP. Reactions were incubated at room temperature, and quenched after 1, 2, 4, 8, 12, or 20 minutes by mixing them with an equal volume of stopping solution (80% formamide, 100 mM EDTA, with bromophenol blue and xylene cyanol dyes). Samples were separated by electrophoresis on a 17.5% polyacrylamide (19:1)–7.5 M urea–1x Tris-borate-EDTA (TBE) sequencing gel. Gels were imaged using a Typhoon 9400 scanner and the fluorescence intensity of the unextended and extended primer bands was quantified with ImageQuant software (GE Healthcare). Percentage of primer extension was determined by measuring the relative intensity of the band corresponding to the extended primer with respect to the total labeled DNA. Data were fit to an exponential equation (Equation 3.1):

\[ y = A(1 - e^{-k_{obs}t}) + c \]  

where \( A \) is amplitude, \( k_{obs} \) is the observed rate of product formation, \( t \) is the time after which reaction was quenched and \( c \) is a constant. All the graphs and
nonlinear regressions were done using GraphPad Prism, version 6.0a (GraphPad Software Inc.). Experiments were performed in triplicate with error bars representing the standard error of the data collected.

3.4.4. $K_{D}^{dNTP}$ and $k_{pol}$ determination. Primer extension assays were performed using final concentrations of 4 µM polymerase and 40 nM p/t DNA in reaction buffer (25 mM HEPES (pH 7.5), 85 mM NaCl, 10 mM MgCl$_2$, 2 mM DTT). Reactions were started by adding various deoxynucleotide triphosphate (dNTP) concentrations ranging from 5-2000µM (depending on polymerase) and allowed to proceed for appropriate time intervals. Experiments were performed in duplicate. Time courses of primer extension reactions were fit to a single exponential equation (Equation 3.1). The observed rates ($k_i$) thus obtained were further plotted as a function of dNTP concentration and then fit to a hyperbolic equation:

$$k_i = \frac{k_{pol}[dNTP]}{K_{D}^{dNTP} + [dNTP]}$$  \hspace{1cm} (3.2)$$

where $k_{pol}$ is the maximum rate of product formation, $K_{D}^{dNTP}$ is the dissociation constant for dNTP binding to the enzyme · DNA binary complex and [dNTP] is the concentration of dNTP added.
Table 3.1: DNA substrate sequences

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Sequence&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 4C-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG GG-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGC CCC GCA TT -5’</td>
</tr>
<tr>
<td>2 1T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG GG-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGC CCT GCA TT -5’</td>
</tr>
<tr>
<td>3 2T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG G G-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGC CTC GCA TT -5’</td>
</tr>
<tr>
<td>4 3T-G</td>
<td>5’-(FAM)-AGG CAC TGA TCG GG-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGC TCC GCA TT -5’</td>
</tr>
<tr>
<td>5 4T-G</td>
<td>5’-(FAM)-AGG CAC TGA TC GGG-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGT CCC GCA TT -5’</td>
</tr>
<tr>
<td>6 C-C mispair</td>
<td>5’-(FAM)-AGG CAC TGA TCG GGC-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGC CCC GCA TT -5’</td>
</tr>
<tr>
<td>7 X-4T-G</td>
<td>5’- GG CAC TGA TC GGG-3’</td>
</tr>
<tr>
<td></td>
<td>3’- CC GTG ACT AGT CCC GCA TT -5’</td>
</tr>
</tbody>
</table>

<sup>a</sup> FAM refers to the 6-carboxyfluorescein label at the 5’ ends of primers for substrates 1-6, used for primer-extension assays. Substrate 7, used for crystallization, was unlabelled.

The deletion hotspot sequence and variations thereof are shown in bold type.
3.4.5. **Crystallization and Structure Determination.** Complexes were prepared at room temperature by combining 200 µM DbhDpo4^{465}Dbh (Fig. 3.1) and 240 µM DNA (X-4T-G substrate, Table 3.1) in 25 mM HEPES (pH 7.0), 5 mM Ca(OAc)$_2$, 85 mM NaCl, 1 mM DTT, and 1 mM dCTP (final concentrations). Catalysis was prevented by including Ca$^{2+}$ as the divalent metal ion instead of Mg$^{2+}$. Crystals were grown at room temperature by hanging-drop vapor diffusion by mixing equal volumes of complex and well solution containing 14% PEG-3350, 100 mM MES-Tris (pH 6.0), 100 mM Ca(OAc)$_2$, 2.5% glycerol, and 250 mM sucrose. Crystals were stabilized and cryoprotected by the addition of a solution containing 20% PEG-3350, 100 mM MES-Tris (pH 6.5), 100 mM Ca(OAc)$_2$, 20% w/v sucrose, and 1 mM dCTP. Crystals were flash cooled in liquid nitrogen. X-Ray diffraction data were collected at Brookhaven National Laboratory (BNL), National Synchrotron Light Source (NSLS) beamline X25, and were processed and scaled using HKL2000 [158]. The structure was solved by molecular replacement using the DbhDpo4Dbh ternary complex (PDB 4F4Y, [128]) as a search model and was refined using PHENIX [159], alternating with cycles of manual rebuilding in Coot [160]. The geometry of the DNA was analyzed using Curves+ [161] and protein conformations were analyzed using DynDom [162]. Structure figures were made using PyMol (Version 1.5.0.4 Schrödinger, LLC). Refined coordinates and structure factors have been deposited as PDB ID: 4NLG.
3.5. RESULTS

3.5.1. Linker identity determines -1 deletion frequency on a repetitive sequence

Both Dbh and Dpo4 generate single base deletions on repetitive sequences using a template-slippage mechanism, but they do so at different frequencies [100, 113, 119], with Dbh being substantially more error prone in this respect. To determine if the linker is the major determinant of this characteristic, we performed single nucleotide incorporation assays on a DNA substrate with a sequence that has previously been shown to be a deletion hotspot (3'-CCCCG-5' in the template strand) for not only Dbh and Dpo4, but also other DinB homologues [107, 113, 118, 119]. Primer extension assays using the 4C-G substrate (Table 3.1) included either dGTP or dCTP as the incoming nucleotide, to give either correct extension (dGTP) or to initiate a single-base deletion (dCTP). Nucleotide incorporation rates were determined for both parental enzymes and for six chimeras constructed from all possible combinations of the polymerase domain, 15-amino acid linker and LF/PAD (Fig. 3.2A-H and Table 3.2). Chimeric polymerases are named by the source enzyme in the order: polymerase core-linker-LF/PAD.
**Figure 3.2: Single-base deletion efficiency is dependent on the identity of the linker.**

A. Dbh

B. DbhDbhDpo4

C. Dpo4DbhDbh

D. Dpo4DbhDpo4

E. Dpo4

F. Dpo4Dpo4Dbh

G. DbhDpo4Dpo4

H. DbhDpo4Dbh

I. C:G mispair

J. 5'—GGG—dCTP—3' → 5'—C—GCCC—dCTP—3' → 5'—GGG—dCTP—3' → 5'—GGGG—dCTP—3'

2nd addition

Realignment

2nd addition
Figure 3.2. Single-base deletion efficiency is dependent on the identity of the linker.

Primer extension by Dbh (A), Dpo4 (E) and the chimeric polymerases (Dbh linker: B-D; Dpo4 linker: F-H) on the 4C-G substrate, which contains the repetitive hotspot sequence (Table 3.1). Incorporation of dGTP (●) shows the correct product formation while dCTP addition (○) initiates a single-base deletion. 40 nM annealed primer-template DNA was preincubated with 4 µM protein before 1mM of the appropriate nucleotide was added to start the reaction. All concentrations given are final. Reactions were quenched after varying time intervals. Lower panels show primer-extension products over time (0-20 min) for each of the nucleotides. (I) Gel shows efficient mispair extension by Dpo4 on the hotspot sequence with a C-C mispair at the primer template junction (sequence shown below gel), through dCTP incorporation. (J) Schematic representation of possible mechanism of mispair formation and extension is shown.
Table 3.2. Summary of observed nucleotide incorporation rates *

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>dGTP (correct) $k_1$ (min$^{-1}$)</th>
<th>dCTP (incorrect) $k_1$ (min$^{-1}$)</th>
<th>$k_{1}^{dc} / k_{1}^{ac}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbh</td>
<td>~0.02**</td>
<td>~0.01**</td>
<td>-</td>
</tr>
<tr>
<td>Dbh(DbhDpo4)</td>
<td>0.04±0.02</td>
<td>0.11±0.02</td>
<td>0.36</td>
</tr>
<tr>
<td>Dpo4(DbhDbh)</td>
<td>0.20±0.04</td>
<td>0.21±0.03</td>
<td>0.95</td>
</tr>
<tr>
<td>Dpo4(DbhDpo4)</td>
<td>0.06±0.01</td>
<td>0.10±0.01</td>
<td>0.60</td>
</tr>
<tr>
<td>Dpo4</td>
<td>0.77±0.20</td>
<td>0.10±0.02</td>
<td>7.70</td>
</tr>
<tr>
<td>Dpo4(Dpo4Dbh)</td>
<td>1.29±0.09</td>
<td>0.40±0.11</td>
<td>3.23</td>
</tr>
<tr>
<td>Dbh(Dpo4Dpo4)</td>
<td>0.92±0.21</td>
<td>0.13±0.03</td>
<td>7.08</td>
</tr>
<tr>
<td>Dbh(Dpo4Dbh)</td>
<td>0.88±0.09</td>
<td>0.23±0.05</td>
<td>3.83</td>
</tr>
</tbody>
</table>

* For data shown in Fig. 3.2, using the 4C-G substrate.

** These values are only approximate because of the large errors associated with these measurements.
We find that the rates of correct and incorrect nucleotide incorporation are dependent on the identity of the linker. As we have observed previously [78, 84], Dbh incorporates both nucleotides (correct, dG and incorrect, dC) at near equal rates (Fig. 3.2A) while Dpo4 adds the correct nucleotide 8-fold faster than the incorrect nucleotide (Fig. 3.2E and Table 3.2). The ratios of incorrect to correct nucleotide incorporation rates are consistent with Dbh having a higher single-base deletion frequency than Dpo4. Similarly, all chimeras containing the Dpo4 linker add the correct dG nucleotide 3- to 8-fold faster than the incorrect dC (Fig. 3.2, panels F-H; Table 3.2) whereas the chimeras containing the Dbh linker incorporate dG up to 3-fold more slowly than dC (Fig. 3.2, panels B-D; Table 3.2). Thus, the linker is a major determinant of single-base deletion activity.

Using variations on the deletion hotspot sequence, with a T individually substituting for each C (Fig. 3.3 and Table 3.1, substrates 1T-G, 2T-G, 3T-G and 4T-G), we confirmed that all the chimeras predominantly use a template-slippage deletion mechanism, as do the parental enzymes [78, 84, 103]. Since the deletion hotspot sequence can adopt multiple conformations, with each of the C’s potentially being unpaired, these variant substrates are designed to favor an unpaired base at a single position, as shown in Table 3.1. The 1T-G substrate is used least efficiently in all cases (Fig. 3.3), ruling out both dNTP-stabilized misalignment and misincorporation-misalignment as the major mechanism by which single base deletions are made. The 3T-G and 4T-G substrates are used most efficiently by all the enzymes (Fig. 3.3), indicating use of a template slippage mechanism with a preference for the unpaired base being 3 or 4 nucleotides upstream of the templating base. In fact, incorporation of dCTP on the 3T-G and 4T-G substrates by each polymerase is at least as rapid as the incorporation of the correct dGTP on the 4C-G substrate (Figs. 3.2 and 3.3).
Figure 3.3: Chimeric polymerases use a template-slipage deletion mechanism.
Figure 3.3: Chimeric polymerases use a template-slippage deletion mechanism.

(A-H) Incorporation of dCTP by Dbh, Dpo4, and chimeras on primer-template DNA containing modified hotspot sequences (1T-G, 2T-G, 3T-G and 4T-G; Table 3.1). Fraction of primer DNA extended, as a function of time, is displayed graphically for the different DNA substrates shown: 1T-G (■), 2T-G (□), 3T-G (▲) and 4T-G (△). The 2T-G, 3T-G and 4T-G substrates are designed to favor an unpaired base at positions 2, 3 and 4 nucleotides 3’ of the templating base, mimicking the multiple conformations that can occur during template slippage. The 1T-G substrate is designed to inhibit slippage and mimic the substrate present if either a dNTP-stabilized misalignment or a misincorporation-misalignment mechanism was used to create deletions in repetitive sequences. The extremely slow nucleotide incorporation on the 1T-G substrate compared to the 2T-G, 3T-G and 4T-G substrates indicates template slippage as the major deletion mechanism.
3.5.2. Linker identity determines efficiency of mispair extension.

Another difference among the polymerases is evident on inspection of the pattern of product formation: the enzymes containing the Dpo4-linker tend to efficiently incorporate a second dCTP (Fig. 3.2E-H and Fig. 3.3E-H), but those containing the Dbh-linker do not (Fig. 3.2A-D and Fig. 3.3A-D). Because the polymerases use a template-slippage deletion mechanism, misincorporation of the first dC nucleotide occurs when the incoming dCTP pairs with the +1G as templating base in the hotspot sequence, skipping over one of the C's in the template strand (Fig. 3.2J). We suspected that the second nucleotide added is likely to be templated by the same G, after isomerization of the primer-template DNA to form a C-C (or C-T) mispair at the primer terminus (Fig. 3.2J). To test this idea, we examined the ability of Dpo4 to extend from a C-C mispair substrate (Table 3.1), which would be the substrate for addition of the second nucleotide, and found that it can efficiently add dC to this primer-template junction (Fig. 3.2I). Altogether, these results suggest that Dpo4 and the chimeras containing the Dpo4 linker have a greater ability to realign slipped DNA and extend from a mispair than do those that contain the Dbh linker.

3.5.3. Amino acid trio in the linker determines overall polymerase conformation

We have previously shown that the key determinant of the polymerase conformation for Dbh and Dpo4 is the interdomain linker [128]. Furthermore, only 3 out of the 15 amino acids in the linker, residues 242-244 (Arg-Lys-Ser) in Dpo4 and 243-245 (Lys-Ile-Pro) in Dbh (Fig. 3.1), are responsible for the base-substitution and abasic-site bypass properties of the two polymerases, suggesting that those 3 amino acids might be sufficient to control the polymerase conformation.

To test this hypothesis, we determined a 2.4 Å ternary complex crystal structure of Dbh containing just the three linker residues from Dpo4 (Fig. 3.4, Table 3.3). Here we refer to this chimeric polymerase as DbhDpo4RKS Dbh. We were unable to
crystallize the complementary chimera, Dpo4Dbh<sup>κε</sup>Dpo4. In the crystal structure, an incoming dCTP is correctly paired to the G in the deletion hotspot sequence (Fig. 3.4A) and the template contains an unpaired T four nucleotides 3’ to the templating G (Fig. 3.4B). The template slippage, enforced by the T replacing a C in the deletion hotspot sequence, allows the “incorrect” incoming nucleotide to form a standard Watson-Crick pair. DbhDpo4<sup>κε</sup>Dpo4 superimposes on Dbh (PDB code 3BQ1; [78]) with an RMSD of 5.08 Å (Fig. 3.4C), even though the polymerase and LF/PAD domains individually align with RMSDs of 1.62 and 0.5 Å, respectively. In contrast, DbhDpo4<sup>κε</sup>Dpo4 superimposes on Dpo4 (PDB code 3QZ7; [84]) with an overall RMSD of 1.34Å (Fig. 3.4D).

Thus, the three amino acids are indeed sufficient to allow the Dbh LF/PAD to adopt the same conformation as in structures of Dpo4. With the polymerase domains superimposed, a rotation of ~50° around an axis roughly parallel to the helical axis of the DNA would be required to bring the LF/PAD domain of Dbh into alignment with that of DbhDpo4<sup>κε</sup>Dpo4 (Fig. 3.4C). In the structure of DbhDpo4<sup>κε</sup>Dpo4, the LF/PAD is positioned so that it is in contact with the β2-3 loop in the fingers domain, causing this loop to become ordered. This contrasts with the β2-3 loop being poorly ordered or completely disordered in all published Dbh structures [78, 163, 164].
Figure 3.4: DbhDpo4$^{RKS}$Dbh adopts a Dpo4-like conformation.
Figure 3.4: DbhDpo4RKS Dbh adopts a Dpo4-like conformation.

Ternary complex of DbhDpo4RKS Dbh chimeric polymerase (A) is shown with domains color coded as in Fig. 3.1 (linker is shown in yellow-orange). Both DNA substrate (white) and incoming dCTP (yellow) are shown as sticks and the divalent Ca²⁺ ions are shown as green spheres. (B) Primer-template DNA around the intercalated -4T nucleotide is shown (white). Linker residues that are of Dpo4 origin (ArgLysSer 243-245) in the chimera are shown in yellow and a 2F₀-Fₐ map calculated from the final refined structure is shown contoured at 1.5 σ (grey mesh). Superposition of DbhDpo4RKS Dbh (color coded as in (A)) on the polymerase core of (C) Dbh (PDB ID: 3BQ1, chain A) and (D) Dpo4 (PDB ID: 3QZ7, chain A). Parental enzymes are shown in grey. The LF/PAD of the chimeric polymerase would need to undergo a ~50° rotation [162] compared to Dbh to adopt the Dpo4-like domain orientation. Linker residues that are of Dpo4 origin in the chimeric Dbh polymerase are shown in red. Carboxylates in the chimera palm are shown (magenta sticks). Close up view of the Dbh (E) and DbhDpo4RKS Dbh (F) active site with domains colored as before. The steric gate residue (Phe 12) is shown as sticks to reflect the stacking interaction established when substrates are positioned correctly in DbhDpo4RKS Dbh, allowing selection against ribonucleotides.
Table 3.3: Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Ternary complex of:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymerase</td>
<td>DbhDpo4&lt;sup&gt;rs&lt;/sup&gt;Dbh</td>
</tr>
<tr>
<td>DNA substrate</td>
<td>X-4T-G</td>
</tr>
<tr>
<td>Incoming nucleotide</td>
<td>dCTP</td>
</tr>
<tr>
<td>PDB ID</td>
<td>4NLG</td>
</tr>
</tbody>
</table>

**Data collection**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>X25</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.1</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>30.0 - 2.4 (2.43-2.40)</td>
</tr>
<tr>
<td>Space group</td>
<td>C 2</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a,b,c (Å)</td>
<td>120.19, 103.07, 53.05</td>
</tr>
<tr>
<td>β (°)</td>
<td>105.4</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>23003 (1042)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.7 (1.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>91.7 (80.7)</td>
</tr>
<tr>
<td>R-merge (%)</td>
<td>4.4 (19.1)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>23.2 (2.9)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>28.7 - 2.4</td>
</tr>
<tr>
<td>Reflections</td>
<td>22579</td>
</tr>
<tr>
<td>R-work (%)</td>
<td>17.4</td>
</tr>
<tr>
<td>R-free (%)</td>
<td>23.1</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>3822</td>
</tr>
<tr>
<td>Number of:</td>
<td></td>
</tr>
<tr>
<td>Protein residues</td>
<td>344</td>
</tr>
<tr>
<td>Nucleotides (including dCTP)</td>
<td>32</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt; ions</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>41</td>
</tr>
<tr>
<td>RMS (bonds)</td>
<td>0.011</td>
</tr>
<tr>
<td>RMS (angles)</td>
<td>1.19</td>
</tr>
<tr>
<td>Ramachandran region</td>
<td></td>
</tr>
<tr>
<td>most favored (%)</td>
<td>90.6</td>
</tr>
<tr>
<td>additional allowed (%)</td>
<td>8.7</td>
</tr>
<tr>
<td>generously allowed (%)</td>
<td>0.6</td>
</tr>
<tr>
<td>disallowed (%)</td>
<td>0.0</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>69.4</td>
</tr>
<tr>
<td>macromolecules</td>
<td>69.8</td>
</tr>
<tr>
<td>ligands</td>
<td>45</td>
</tr>
<tr>
<td>solvent</td>
<td>53.2</td>
</tr>
</tbody>
</table>
Interestingly, the unpaired T in the template DNA is not in an extrahelical conformation. Instead, it intercalates into the DNA duplex (Fig. 3.4B), causing a tilt of 18° between the flanking basepairs. In our previous structures of chimeric polymerases containing this same primer-template DNA, the unpaired T was also intercalated into the duplex when bound to the DbhDpo4Dbh chimera, whereas it was in an extrahelical conformation when bound to the DbhDpo4Dpo4 chimera [128].

3.5.4. dNTP selection: linker-dependent alteration of nucleotide incorporation rate

Next, we wanted to understand how the enzyme conformation conferred by the Dbh and Dpo4 linkers influence the steps involved in nucleotide incorporation. To investigate this we performed single-turnover primer extension assays using the 4C-G substrate (Table 3.1) to determine the nucleotide binding affinity ($K_{D_{dNTP}}$) and rate of polymerization ($k_{pol}$) for the parental proteins, Dbh and Dpo4, as well as for the two chimeras, DbhDpo4*RKS*Dbh and Dpo4Dbh*KIP*Dpo4. The results are summarized in Table 3.4.

Comparison of the four polymerases shows that the linker residues strongly influence the overall fidelity of the enzyme. Dpo4’s fidelity is reduced by 6.5-fold (from 39 to 6) when Dbh residues are present in the linker of Dpo4 (Dpo4Dbh*KIP*Dpo4). Conversely, Dpo4 residues in the linker of Dbh (DbhDpo4*RKS*Dbh) increase fidelity of Dbh by ~30-fold (from 1.4 to 42).

The major contribution to fidelity comes from differences in the maximum rate of nucleotide incorporation, $k_{pol}$ (Table 3.4). Comparing the rates of correct vs. incorrect nucleotide incorporation ($k_{pol}^{dCTP}/k_{pol}^{dCTP}$), the Dpo4 linker causes a 14-fold increase in the selectivity of nucleotide incorporation (1 vs. 14), while the Dbh linker causes a 13-fold decrease (8.5 vs. 0.67). Thus, the linker-dependent change in nucleotide incorporation rates contributes to the differences in single-base deletion fidelity, and this occurs by substantially influencing the rate of correct, but not incorrect, nucleotide incorporation (Table 3.4).
Nucleotide binding affinity also contributes to fidelity, but in a more complex way (Table 3.4). The ratio of the dissociation constants for correct and incorrect nucleotides ($K_{d}^{dGTP}/K_{d}^{dCTP}$) is decreased by ~2-fold for both the Dbh and Dpo4 linkers (0.7 vs. 0.3 for Dbh vs. DbhDpo4RKSDbh; 0.2 vs. 0.1 for Dpo4 vs. Dpo4DbhRKS*Dpo4). Since increased affinity results in a lower $K_{d}$, the linker sequences increase the preference of the enzyme for binding the correct nucleotide, which contributes to a modest increase in fidelity for both polymerases.

Overall, the linker-dependent changes to nucleotide binding and incorporation rate act in the same direction with the Dpo4 linker, increasing fidelity, but act in opposite directions with the Dbh linker, increasing the fidelity based on nucleotide binding but decreasing the fidelity based on nucleotide incorporation rate. It also appears that residues from the polymerase and/or LF/PAD domains contribute to nucleotide binding affinity, but this is difficult to dissect because of the large errors in measuring weak nucleotide binding.
Table 3.4: Kinetic parameters for correct and incorrect nucleotide incorporation on a deletion hotspot sequence (substrate 4C-G)

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>dGTP (correct)</th>
<th>dCTP (incorrect)</th>
<th>K_{dNTP}^{dGTP} (µM)</th>
<th>k_{pol}^{dGTP} (min⁻¹)</th>
<th>Efficiency* (x10⁻³) (µM⁻¹min⁻¹)</th>
<th>K_{dNTP}^{dCTP} (µM)</th>
<th>k_{pol}^{dCTP} (min⁻¹)</th>
<th>Efficiency* (x10⁻³) (µM⁻¹min⁻¹)</th>
<th>K_{dCTP}^{dGTP} / K_{dCTP}^{dCTP}</th>
<th>k_{pol}^{dGTP} / k_{pol}^{dCTP}</th>
<th>Fidelity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbh</td>
<td>822±397</td>
<td>0.17±0.004</td>
<td>0.21</td>
<td>1135±367</td>
<td>0.17±0.03</td>
<td>0.15</td>
<td>0.7</td>
<td>1</td>
<td>1.4</td>
<td>0.67</td>
<td>6</td>
</tr>
<tr>
<td>Dpo4Dbh^KIPDpo4</td>
<td>92±29</td>
<td>1.9±0.13</td>
<td>20.8</td>
<td>818±198</td>
<td>2.85±0.31</td>
<td>3.5</td>
<td>0.1</td>
<td>0.67</td>
<td>6</td>
<td>39</td>
<td>8.5</td>
</tr>
<tr>
<td>Dpo4</td>
<td>199±49</td>
<td>20.4±2.2</td>
<td>102</td>
<td>893±121</td>
<td>2.39±0.1</td>
<td>2.6</td>
<td>0.2</td>
<td>8.5</td>
<td>39</td>
<td>39</td>
<td>8.5</td>
</tr>
<tr>
<td>DbhDpo4^KKS Dbh</td>
<td>74±29</td>
<td>16.5±1.9</td>
<td>222</td>
<td>222±84</td>
<td>1.17±0.13</td>
<td>5.3</td>
<td>0.3</td>
<td>14</td>
<td>42</td>
<td>42</td>
<td>14</td>
</tr>
</tbody>
</table>

*: Efficiency = k_{pol} / K_{dNTP}^{dNTP}

†: Fidelity = Efficiency dGTP / Efficiency dCTP
3.6. DISCUSSION

Using two closely related Y-family DNA polymerases from *Sulfolobus* and their chimeric counterparts, we show here that three residues in the interdomain linker have a major impact on the replication of a repetitive sequence that is a deletion hotspot, changing the single-base deletion fidelity by up to 30-fold. The linker identity correlates with the maximal rate of correct versus incorrect nucleotide incorporation but not with nucleotide binding affinity (Table 3.4). On the deletion hotspot sequence, enzymes with the Dbh linker catalyze addition of both dGTP (the correct nucleotide) and dCTP (the nucleotide that initiates a single-base deletion) at approximately the same rates, while enzymes with the Dpo4 linker have a distinct preference for incorporating the correct nucleotide. Significantly, the rates of incorrect nucleotide incorporation do not vary much between the enzymes (either parent or chimera). Thus, the linker specifically alters the rate of correct nucleotide addition, which accounts for the changes in single-base deletion fidelity.

A correlation between polymerase fidelity and the efficiency of correct, but not incorrect, nucleotide incorporation has been noted previously for a wide range of high and low fidelity polymerases [165]. Interestingly, the linker sequences affect single-base deletion and base-substitution fidelity in opposite directions: the Dpo4 linker increases the base substitution frequency [128] but decreases the single-base deletion fidelity (Fig. 3.2) while the Dbh linker does the opposite.

Information about the identity of the linker sequences is propagated to the active site via the overall conformation of the polymerase. The structure of DbhDpo4<sup>res</sup>Dbh demonstrates that replacing three amino acids in the Dbh linker with the equivalent residues from Dpo4 allow the Dbh LF/PAD to move into a position where it contacts the β2-3 loop of the catalytic domain (Fig. 3.4C-D). This is in contrast to structures of Dbh that show a gap between the two domains at this location. Contacts between the two domains tightly constrain the width of the nascent basepair binding pocket. This brings the primer
terminus and incoming nucleotide into alignment at the active site (Fig. 3.4E-F), providing an explanation for the increased rate of correct nucleotide incorporation in enzymes containing the Dpo4 linker. Even though the crystallized substrate contains a misaligned primer-template, the unpaired base is located at a position where nucleotide incorporation occurs at a rate as fast as on correctly aligned primer-template with the correct incoming nucleotide.

The ability to extend from a C-C mismatched primer terminus, which is enhanced by the Dpo4 linker sequences, is comparable to what has been reported for human polymerase kappa [151, 166]. For pol kappa, extension from the mismatch involves realigning the slipped DNA strands [166] and we presume that this is the case here. It remains to be determined if the realignment is actively performed by the polymerase or occurs due to transient unpairing of the template DNA from the primer terminus. In either case, realignment and mispair extension abilities act to reduce the frequency of deletion mutations at the expense of increasing base-substitution mutations.

The results reported here, combined with our previous data [128], show that base-pair substitution frequency, single-base deletion frequency, mispair extension and abasic site bypass are all strongly affected by the linker-dependent conformation of the polymerase. These observations have two significant implications. First, they highlight how different substrate specificities can evolve as a result of just a few amino acid changes remote from the site of catalysis. The Y-family polymerases display an extraordinary variety of lesion-bypass and mutagenic activities. It is easy to envision how these enzymes could readily diverge if, after a gene duplication event, just a few mutations could produce a polymerase with a new specificity that provides a selective advantage. Second, they suggest that Y-family polymerase fidelity and specificity could be regulated by altering polymerase conformation. This could, for example, be the mechanism by which the single-base deletion activity of *E. coli* DNA polymerase IV (DinB) is suppressed when forming a complex with UmuD and RecA [167]. Thus, it will be interesting to study the influence of protein partners on both the activity and structural conformation of TLS.
polymerases, which may shed light on the mechanism of regulating these polymerases in the cell.

3.7. ACKNOWLEDGEMENTS

We thank Jessica Olson for assistance in the early stages of this project and Joachim Jaeger for helpful advice throughout. This work was supported by NIH grant R01-GM080573 to J.D.P. and by fellowship support from the Wadsworth Center to P.M. and I.L.
PART TWO
CHAPTER 4


4.1. Description of contribution

For this manuscript, I contributed equally while designing and performing the experiments as well as analyzing the results. I also helped in writing and editing the manuscript.
4.2. SUMMARY

PolC is the C-family replicative polymerase in low G+C content Gram-positive bacteria. To date several structures of C-family polymerases have been reported, including a high resolution crystal structure of a ternary complex of PolC with DNA and incoming deoxynucleoside triphosphate (dNTP). However, kinetic information needed to understand the enzymatic mechanism of C-family polymerases is limited. For this study we have performed a detailed steady-state and pre-steady-state kinetic characterization of correct dNTP incorporation by PolC from the Gram-positive pathogen *Staphylococcus aureus*, using a construct lacking both the non-conserved N-terminal domain and the 3'-5' exonuclease domain (Sau-PolC-ΔNΔExo). We find that Sau-PolC-ΔNΔExo has a very fast catalytic rate ($k_{pol} \approx 330 \text{ s}^{-1}$) but also dissociates from DNA rapidly ($k_{off} \approx 150 \text{ s}^{-1}$), which explains the low processivity of PolC in the absence of sliding clamp processivity factor. Although Sau-PolC-ΔNΔExo follows the overall enzymatic pathway defined for other polymerases, some significant differences exist. The most striking feature is that the nucleotidyl transfer reaction for Sau-PolC-ΔNΔExo is reversible and is in equilibrium with dNTP binding. Simulation of the reaction pathway suggests that rate of pyrophosphate release, or a conformational change required for pyrophosphate release, is much slower than rate of bond formation. The significance of these findings is discussed in the context of previous data showing that binding of the β-clamp processivity factor stimulates the intrinsic nucleotide incorporation rate of the C-family polymerases, in addition to increasing processivity.
4.3. INTRODUCTION

DNA replication is the complex process of genome duplication involving several different proteins that form the “replisome”. A key enzyme of the replisome is the DNA polymerase, a nucleotidyl transferase that catalyzes the addition of a deoxynucleoside triphosphate (dNTP) to the nascent DNA chain. All organisms have several types of DNA polymerases of which the ones responsible for duplicating most of the genome are known as replicative polymerases. These are characterized by being highly efficient enzymes that can select the next correct nucleotide with extraordinarily high accuracy in a template-dependent manner.

The replicative polymerases of all bacteria are grouped by sequence similarity into the C-family of DNA polymerases [45, 168], but subdivide into two branches. The polC gene encodes the replicative polymerase of Gram-positive bacteria with low G+C content, while the dnaE gene encodes for the same in Gram-negative bacteria and in Gram-positive bacteria with high G+C content [169]. The replisomes of both Gram-positive and Gram-negative bacteria have been reconstituted and studied in vitro providing a wealth of knowledge about how replication occurs inside the bacterial cell [23, 24, 71, 72, 170-177].

Recently, several crystal structures of C-family polymerases have been reported, including DnaE from Escherichia coli and Thermus aquaticus and PolC from Geobacillus kaustophilus [39, 52, 53, 55]. Other than non-conserved N- and C-terminal extensions, the individual domains of PolC and DnaE are structurally conserved [39], as expected from sequence conservation, but differ somewhat in their linear organization in the protein (Figure 4.1).
Figure 4.1. Overview of PolC structure and domain organization.

A. Structure of Gka-PolC-ΔNΔExo shown (PDB ID: 3F2B). Domains are color coded as in Figure 4.1B. OB domain is not shown for simplicity. The linker connecting the N-palm and PHP domains is shown (yellow).

B. Sequence alignment showing domain organization of C-family polymerases.

A. Structure of Gka-PolC-ΔNΔExo shown (PDB ID: 3F2B). Domains are color coded as in Figure 4.1B. OB domain is not shown for simplicity. The linker connecting the N-palm and PHP domains is shown (yellow). B. Sequence alignment showing domain organization of C-family polymerases.
The OB domain of PolC is located just after the non-conserved N-terminal domain, while the OB domain of DnaE is located just before the non-conserved C-terminal domain. Additionally, PolC contains an intrinsic 3' to 5' exonuclease proofreading domain that is absent in DnaE. In Gram-negative bacteria, this function is performed by the epsilon (ε) subunit [178] of the replisome, which is homologous to the PolC exonuclease domain. A remarkable finding from the crystallographic studies is that the bacterial replicative polymerases are not related to the replicative polymerases from either eukaryotes or the archaea.

Despite the fact that in vitro studies of the bacterial replisome have been key to delineating the main features of DNA replication in all forms of life [178], comparatively little is known about the kinetic mechanism of polymerization by C-family polymerases. This contrasts with the extensive kinetic information available for other polymerase families, including the B-family enzymes that are the replicative polymerases in eukaryotes and most archaea. Thus, the foundation for detailed structure-function studies of C-family polymerases has not yet been laid.

For all polymerases studied to date, the same overall enzymatic pathway (Figure 4.2) has been established for correct nucleotide incorporation [33-35, 179, 180]. The minimal pathway involves substrates binding to the polymerase in an ordered manner, with DNA binding first (step 1), followed by binding of the incoming dNTP (step 2). This is succeeded by the chemical step of bond formation (step 3). Typically, this step is preceded by a slower step along the pathway, which has been interpreted as a conformational change of the polymerase [33, 180]. Earlier structural studies suggested that this slow step might correspond to the large-scale domain movement associated with nucleotide binding [38, 40], but more recent studies have shown that motion to be too fast to be rate limiting [32, 181]. Although the conformational change accompanying nucleotide binding is faster than chemistry, it still controls specificity of nucleotide addition [182, 183]. Currently, the slower conformational change is thought to be a later, smaller-scale movement, but the precise nature of this non-covalent change is not known and may vary among
different polymerases. Bond formation is followed by the release of the pyrophosphate (PPI) (step 4) generated during phosphoryl transfer. This step is generally presumed to be rapid [33] and is likely to be accompanied by the reverse of the conformational change that is induced by dNTP binding. Finally, release of the product DNA occurs (step 5), allowing the polymerase to perform subsequent rounds of catalysis. During processive synthesis, DNA would translocate along the polymerase rather than dissociate.

In this study we have performed both steady-state and pre-steady-state kinetic characterization of correct dNTP incorporation by an N-terminal and exonuclease domain deficient mutant of *Staphylococcus aureus* PolC (Sau-PolC-ΔNΔExo). This construct has the identical domain organization as the *G. kaustophilus* PolC used in the crystallographic study (Gka-PolC-ΔNΔExo) [39]. Furthermore, this construct has all the domains that are conserved between the PolC and DnaE polymerases (Figure 4.1). Hence, kinetic data obtained using this construct can be utilized directly for making structure-function correlations among the C-family polymerases and establish a foundation for future mechanistic studies of this polymerase family.
Figure 4.2. Minimal single-nucleotide incorporation reaction pathway for DNA polymerases.

Abbreviations used were: E, DNA polymerase; D₀, unextended DNA; D₁, DNA extended by one base-pair; PPᵢ, inorganic pyrophosphate. Dashed arrow indicates the polymerase entering another round of catalysis.
4.4. MATERIALS AND METHODS

4.4.1. Materials. 5'-6FAM labeled primer and unlabeled template oligonucleotides were purchased from Integrated DNA Technologies, Inc. Unlabeled ultrapure grade dTTP was purchased from GE Healthcare Biosciences. All the graphs and nonlinear regressions were done using GraphPad Prism, version 6.0a (GraphPad Software Inc.). Simulation of the reaction mechanism of Sau-PolC-ΔNΔExo was performed using KinTek Explorer, version 3.0 (KinTek Corp.) [184, 185].

4.4.2. Duplex DNA formation. The primer and template DNA oligonucleotides were incubated together in annealing buffer (10 mM Tris-HCl (pH 8), 1 mM EDTA and 100 mM NaCl) and heated to a temperature of 95 °C followed by gradual cooling to room temperature.

4.4.3. Expression and purification of Sau-PolC-ΔNΔExo. S. aureus PolC lacking the N-terminal domain (amino acids 1-207) and the exonuclease domain (amino acids 415-609) and containing a C-terminal hexahistidine tag (Sau-PolC-ΔNΔExo) was expressed from a pET32A vector and was a generous gift from Thale Jarvis (Crestone Inc.). The plasmid was transformed into Rosetta(DE3)pLysS E. coli cells. Cells were grown to an OD$_{600}$ of ~0.65 and then induced with 0.5 mM IPTG for ~16 hrs at 17 ºC. All subsequent steps were carried out at 4 ºC. Cell pellets were resuspended in IMAC buffer (50 mM Tris-HCl (pH 7.5), 800 mM NaCl, 10 mM imidazole and 10% glycerol). In order to prevent proteolytic degradation of Sau-PolC-ΔNΔExo, EDTA-free protease inhibitor tablet (Roche) was added to IMAC buffer at a concentration of 1 tablet/10 g of cells. Cells were lysed by sonication and the clarified cell lysate was passed through Ni-NTA columns (3x5 ml). In order to reduce the NaCl concentration to 100 mM for later steps, the columns were washed with 10 column volumes of low salt IMAC buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM imidazole and 10% glycerol). The protein was eluted using a linear gradient of imidazole from 10 to 400 mM in low salt IMAC buffer. During
elution, two proteins with molecular weights of ~75 kD eluted before Sau-PolC-ΔNΔExo. These are likely to be partial proteolytic products of Sau-PolC-ΔNΔExo and care was taken to remove these contaminants during elution from the Ni-NTA columns. Intact Sau-PolC-ΔNΔExo obtained from Ni-NTA chromatography was loaded onto Q-sepharose columns (HiTrap Q XL, 2x5 ml) pre-equilibrated in Buffer A (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM EDTA, 10% glycerol and 1 mM DTT). The protein was eluted from the Q column using a linear gradient of NaCl from 100 mM to 1 M in Buffer A. Eluent of the Q column was diluted ~7 fold in Buffer A, to a NaCl concentration of ~100 mM, and was subjected to heparin column chromatography (HiTrap Heparin HP, 2x5 ml). Buffer A was used to pre-equilibrate the heparin columns and protein was eluted using a linear gradient of NaCl ranging from 100 mM to 1 M in Buffer A. As a final step of purification, Sau-PolC-ΔNΔExo eluted from the heparin column was subjected to size exclusion chromatography using a Superdex 200 column (HiLoad 16/60 Superdex 200 pg) pre-equilibrated in storage buffer (50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 5 mM EDTA, 10% glycerol and 1 mM DTT). The purified protein obtained from the size exclusion column was concentrated to ~150-200 µM, calculated from the OD of the protein sample using a theoretical extinction coefficient of 87,100 M⁻¹cm⁻¹, and was stored at -80 ºC.

4.4.4. Assays for optimization of enzymatic activity of Sau-PolC-ΔNΔExo. Primer extension assays were performed in order to determine the optimum buffer conditions for the enzymatic activity of Sau-PolC-ΔNΔExo. 400 nM p/t DNA was incubated with 1 nM Sau-PolC-ΔNΔExo in PolC reaction buffer (25 mM MES-Tris (pH 8), 25 mM NaCl, 8 mM MgCl₂, 2 mM DTT and 5% glycerol). All the components of the PolC reaction buffer were kept fixed except the component whose effect was being tested. Unless mentioned otherwise, all the assays were carried out at room temperature (25 ºC). Assays were initiated by addition of 1 mM (final concentration) dTTP to the reaction mix. After 2 minutes, an equal volume of 250 mM EDTA was added to quench the assays. The extended and unextended primers were separated on a 17% acrylamide / 7 M urea denaturing 1xTBE gel. The gel was imaged using a Typhoon 9400 scanner (GE Healthcare) and bands were quantitated using ImageQuant software
Percentage of primer extension was determined by measuring the relative intensity of the band corresponding to the extended primer with respect to the total labeled DNA (i.e. both extended and unextended primer strands). All reactions were performed in triplicate.

4.4.5. **Steady-state assays.** Primer extension assays done for determining the steady-state parameters were performed using a KinTek RQF-3 rapid quench instrument (KinTek Corp.). Reactions were initiated by mixing a pre-equilibrated solution of 5 µM p/t DNA and 50 nM total Sau-PolC-ΔNΔExo in PolC reaction buffer (this corresponded to an active enzyme concentration of 33 nM, as described in the “Active site titration” section of the Results) to an equal volume of various concentrations of dTTP (18.76 to 600 µM) in the same buffer. Hence, the final p/t DNA and active Sau-PolC-ΔNΔExo concentrations in the reactions were 2.5 µM and 16.5 nM respectively and the final concentration range of dTTP was 9.38 to 300 µM. The assays were quenched at various time intervals by addition of 250 mM EDTA. The time intervals were adjusted such that primer extension was between 5-15%. Separation and quantitation of the extended primers was performed as described above. The concentration of primers extended for different concentrations of dTTP were plotted as a function of time and the data were fit to the steady-state rate equation:

\[
Y = [ED]_A k_{obs} t + C
\]

where \( Y \) is the concentration of primer extended, \([ED]_A\) is the concentration of active Sau-PolC-ΔNΔExo · p/t DNA binary complex that gets converted to product, \( k_{obs} \) is the observed rate of primer extension, \( t \) is time interval after which the reaction was quenched, and \( C \) is a constant. The observed rates were plotted as a function of dTTP concentration ([dTTP]) and the data were fit to the Michaelis-Menten equation:

\[
k_{obs} = \frac{k_{cat}[dTTP]}{K_M^{dNTP} + [dTTP]}
\]

where \( k_{cat} \) is the maximum steady-state rate and \( K_M \) is the Michaelis constant for dNTP.
4.4.6. Measurement of DNA dissociation rate from binary complex.

Kintek RQF3 rapid quench device was used to perform this experiment. 300 nM Sau-PolC-ΔNΔExo (this corresponded to an active enzyme concentration of 200 nM, as described in the “Active site titration” section of the Results) was preincubated with 160 nM p/t DNA in PolC reaction buffer in a 16 µl reaction volume. This was mixed with an equal volume of 96 µM unlabelled p/t DNA in the same buffer and incubated for various time intervals (0.005-0.05 s). Finally, ~80 µl of 200 µM dTTP was added for primer extension (~140 µM final concentration). At this stage the reaction was allowed to proceed for 0.028 s and quenched by collection of the sample in a microfuge tube containing 100 µl of 250 mM EDTA. Concentration of the extended primer was plotted as a function of time. Data were fit to the following exponential equation (Equation 4.3) and the rate of decrease of product formation was interpreted as the rate of dissociation of Sau-PolC-ΔNΔExo from the preformed Sau-PolC-ΔNΔExo • p/t DNA binary complex.

\[ Y = Ae^{-kt} + C \]  

(4.3)

where, Y is the concentration of the product formed, A is the amplitude, k is the rate of product formation, t is the first incubation time (ranging from 0.005 to 0.05 s) and C is a constant. Reactions were performed in triplicate.

4.4.7. Active site titration of Sau-PolC-ΔNΔExo.

300 nM Sau-PolC-ΔNΔExo was pre-equilibrated with various concentrations of p/t DNA (20 to 1800 nM) in PolC reaction buffer. The reactions were initiated by rapid mixing of this solution with an equal volume of PolC reaction buffer containing 2 mM dTTP in a KinTek RQF-3 rapid quench instrument. Final concentrations were 10-900 nM p/t DNA, 150 nM polymerase and 1 mM dTTP. The reactions were terminated at different time intervals by addition of 250 mM EDTA. The time-courses of primer extension for different p/t DNA concentrations were fit to the full burst equation:

\[ Y = [ED]_a \left( \frac{k_1}{k_1 + k_2} \right)^2 \left( 1 - e^{-\left(\frac{k_1 + k_2}{k_1 + k_2}\right) t} \right) + \left( \frac{k_1 k_2 t}{k_1 + k_2} \right) + C \]  

(4.4)

where Y is the concentration of the extended primer, [ED]_a is the concentration
of the preformed active enzyme · DNA binary complex that gets converted to product before turnover, \( k_1 \) is the rate of the fast phase, \( k_2 \) is the rate of the slow phase, \( t \) is the time interval after which the reaction was quenched and \( C \) is a constant. \( [ED]_A \) for different DNA concentrations thus obtained were plotted out as a function of p/t DNA concentration and fit to the following quadratic equation:

\[
[ED]_A = \frac{(K_{DNA}^D + [E]_A + [DNA]_r) - \sqrt{(K_{DNA}^D + [E]_A + [DNA]_r)^2 - 4[E]_A[DNA]_r}}{2} \quad (4.5)
\]

where \( K_{DNA}^D \) is the dissociation constant for binding of Sau-PolC-ΔNΔExo to p/t DNA, \( [E]_A \) is the concentration of active Sau-PolC-ΔNΔExo and \( [DNA]_r \) is the concentration of total p/t DNA at the beginning of the assay.

### 4.4.8. \( K_{dNTP}^D \) determination.

Primer extension assays were performed with a RQF-3 rapid quench instrument using final concentrations of 804 nM active Sau-PolC-ΔNΔExo, 50 nM p/t DNA and various [dTTP] (1.17 to 100 µM). Reactions were quenched by addition of 250 mM EDTA. Time courses of primer extension reactions were plotted as a function of [dTTP] and the data were fit to the full burst equation (Equation 4.4). The rate, \( k_1 \), and \( [ED]_A \) thus obtained were further plotted against [dTTP] and then fit to the appropriate hyperbolic equation:

\[
k_1 = \frac{k_{pol}[dTTP]}{K_{dNTP}^M + [dTTP]} \quad (4.6)
\]

\[
[ED]_A = \frac{[ED]_{A\text{max}}[dTTP]}{K_{dNTP}^D + [dTTP]} \quad (4.7)
\]

where \( k_{pol} \) is the maximum rate of the burst of product formation, \( K_{dNTP}^D \) is the dissociation constant for dNTP binding to the Sau-PolC-ΔNΔExo · p/t DNA binary complex and \( [ED]_{A\text{max}} \) is the maximum concentration of the enzyme · DNA binary complex that gets converted to product before turnover.
Unless mentioned otherwise, all reactions were done in at least three independent experiments, using two different preparations of Sau-PolC-ΔNΔExo. All data were combined and analyzed together.

4.4.9. Simulation. The reaction mechanism of Sau-PolC-ΔNΔExo was simulated using KinTek Explorer software. Details about the mechanism used for the simulation is discussed under “Results” section. The software was used to fit data to the simulation using an iterative procedure until a “best fit” was achieved. The simulated curves and the raw data were exported from the software and final plots overlaying the raw data with the simulated curves were prepared using GraphPad Prism. To determine the range within which each of the rate constants was constrained by the model, and to investigate the relationships between different rate constants, we computed 3-D confidence contour plots for all possible pairs of rate constants. See reference [184] for a detailed description of how to interpret these plots.
4.5. RESULTS

4.5.1. PolC purification. Recombinant Sau-PolC-ΔNΔExo was purified using Ni-NTA, anion exchange, heparin and size-exclusion chromatography. Figure 4.3 shows SDS-PAGE analysis of the final purified protein. Sau-PolC-ΔNΔExo migrates as expected for a protein with a theoretical molecular weight of 120 kD. As has been noted previously for full-length PolC [175], inducing protein expression at temperatures below 20 ºC was critical for obtaining purified protein, estimated to be ~95% homogeneous, that did not have significant levels of proteolytic products.

4.5.2. Optimal reaction conditions. Reaction conditions for Sau-PolC-ΔNΔExo were optimized by quantitating incorporation of the next correct dNTP on a p/t DNA with an 18-bp duplex region and a 19-nt single stranded template region (Figure 4.4A). All reaction conditions were kept constant, except for the one whose effect was being tested. The reaction conditions varied were: pH of the buffer, concentration of NaCl, concentration of Mg²⁺ and reaction temperature (Figure 4.4B-E). Dependence of primer extension on the buffer pH followed a bell shaped curve typical of an acid-base reaction, with an optimum pH of 8 (Figure 4.4B). The rate of primer extension was found to decrease with an increase in the concentration of NaCl, with the maximum activity occurring at 25 mM NaCl (Figure 4.4C). A concentration of 8 to 12 mM Mg²⁺ was found to be optimal for enzymatic activity of Sau-PolC-ΔNΔExo (Figure 4.4D). No primer extension was observed in the absence of Mg²⁺, as expected for a polymerase using a two-metal-ion mechanism. Very little primer extension occurred at 4 ºC and 50 ºC, but, for all other temperatures tested (25 ºC, 30 ºC and 37 ºC), the enzyme performed well (Figure 4.4E). Based on these results, all subsequent reactions were performed at 25 ºC at pH 8 with 25 mM NaCl and 8 mM Mg²⁺.
Figure 4.3. SDS-PAGE of Sau-PoIc-ΔNΔExo.

A 10% SDS-polyacrylamide gel stained with Coomassie R-250 showing purified Sau-PoIcΔNΔExo obtained after size exclusion chromatography. (a) Kaleidoscope pre-stained marker. (b) 2.5 mM purified Sau-PoIc-ΔNΔExo.
Figure 4.4. Primer extension assays for optimizing enzymatic activity of Sau-PolC-ΔNΔExo.

(A) Duplex DNA sequence used for all primer extension assays performed in this study. "*/'' at 5' end of primer indicates 6-FAM label. (B) Effect of pH (C) Effect of NaCl concentration (D) Effect of Mg2+ concentration and (E) Effect of temperature on Sau-PolC-ΔNΔExo activity. Primer extension assays were carried out under steady-state conditions by adding 1 mM dTTP (the correct incoming dNTP) to a pre-incubated solution of 400 nM p/t DNA and 1 nM Sau-PolC-ΔNΔExo. Reactions were quenched after 2 minutes by addition of an equal volume of 250 mM EDTA. Unextended and extended primers were separated by gel electrophoresis on a 17% denaturing TBE-acrylamide gel. Fraction of primer DNA extended was determined by measuring the relative intensity of the extended primer band with respect to the total labeled DNA (extended and unextended primer).
4.5.3. **Michaelis-Menten kinetics.** Primer extension assays were performed under steady-state conditions, with substrates present in excess of enzyme, using a final concentration of 2.5 µM p/t DNA, 16.5 nM active Sau-PolC-ΔNΔExo (see result of “Active site titration” for calculation of active enzyme concentration) and various concentrations of dTTP ranging from 9.38 to 300 µM. Observed rates of nucleotide incorporation were calculated from the concentration of product formed vs. time, using the linear portion of progress curves (Figure 4.5A), and were plotted as a function of [dTTP] (Figure 4.5B). The data were fit to the Michaelis-Menten equation (Equation 4.2) and gave a $k_{cat}$ of $17 \pm 1$ s$^{-1}$ and $K_{M}^{dNTP}$ of $43 \pm 7$ µM.

4.5.4. **Rate of dissociation of p/t DNA from the binary complex.** Typically, in single-nucleotide incorporation assays such as the one used here, dissociation of the DNA from the polymerase · DNA binary complex is the rate limiting step of the catalytic cycle, and the rate of this step ($k_{off}$) governs the steady-state rate ($k_{cat}$). To determine if this was the case for Sau-PolC-ΔNΔExo, we directly measured $k_{off}$ using the experimental design shown in Figure 4.6A. For this experiment Sau-PolC-ΔNΔExo was preincubated with p/t DNA and the preformed Sau-PolC-ΔNΔExo · p/t DNA binary complex was rapidly mixed with an equal volume of unlabelled p/t DNA. The resulting reaction (containing final concentrations of 150 nM Sau-PolC-ΔNΔExo, 80 nM p/t DNA, and 48 µM unlabelled p/t DNA) was incubated for various time intervals ranging from 0.005 to 0.05 s. The unlabelled p/t DNA trapped any Sau-PolC-ΔNΔExo that dissociated from the preformed binary complex during this time. Next, dTTP was added to initiate the reaction and a further incubation of 0.028 s was performed, to allow extension of p/t DNA bound to PolC. Finally, the reaction was quenched with EDTA. A plot of labeled product formed versus the variable incubation time (0.005 to 0.05 s) showed a clear reduction in product concentration with increasing time (Figure 4.6B). This was attributed to the decrease in the concentration of the preformed Sau-PolC-ΔNΔExo · p/t DNA due to the dissociation of the labeled p/t DNA from the complex and rebinding of the enzyme to the excess unlabeled p/t DNA. The data fitted well to an exponential equation (Equation 4.3) and the rate of decrease in product
formation ($k_{off}$) was found to be $150 \pm 30$ s$^{-1}$. This indicated that for Sau-PolC-$\Delta$N$\Delta$Exo, DNA dissociation is approximately 9-fold faster than $k_{cat}$ and, surprisingly, is not the rate-limiting step of the single-nucleotide incorporation cycle.
Figure 4.5. Steady-state kinetic analysis of Sau-PolC-$\Delta$N$\Delta$Exo.

Primer extension assays were performed by adding dTTP (final concentration range 9.38 to 300 µM) to a final concentrations of 2.5 µM p/t DNA and 16.5 nM active Sau-PolC-$\Delta$N$\Delta$Exo. The reactions were quenched at different time intervals with 250 mM EDTA. (A) A typical time course of primer extension followed during the steady-state kinetic assays (final [dTTP] was 300 µM). The concentration of primer extended was plotted against time and fit to the steady-state equation (Equation 4.1). (B) Michaelis-Menten plot for Sau-PolC-$\Delta$N$\Delta$Exo. The observed rates of primer extension were plotted as a function of the dTTP concentration. The resulting plot was fit to the Michaelis-Menten equation (Equation 4.2). From the fit, steady-state rate constant ($k_{cat}$) was calculated to be $17 \pm 1$ s$^{-1}$ and Michaelis constant for dNTP ($K_{M,dNTP}$) was determined to be $43 \pm 7$ µM.
Figure 4.6. Determination of the DNA dissociation rate from polymerase · DNA binary complex ($k_{off}$).

(A) Schematic representation of the experimental procedure. (B) Plot of product formed vs time. The data were fit to a single exponential equation (Equation 4.3). The rate of decrease in product formation (which is equivalent to the rate of dissociation of p/t DNA from Sau-PolC-ΔNΔExo · p/t DNA binary complex ($k_{off}$)) was $150 \pm 30$ s$^{-1}$. 

200nM active Sau-PolC-ΔNΔExo + 160 nM labeled DNA

96uM unlabelled DNA

Incubation for different time intervals

200μM dTTP

Incubation for 0.028s

Sample collected in tubes containing 100ul of 250 mM EDTA

B

Product (nM)

$15$

$10$

$5$

$0$

0.00 0.01 0.02 0.03 0.04 0.05

Time(s)

$k_{off} = 150 \pm 30$ s$^{-1}$
4.5.5. Pre-steady-state burst kinetics. To determine if Sau-PoIC-ΔNΔExo displayed a rate-limiting step after chemistry, primer extension assays were performed under pre-steady-state conditions with a total enzyme concentration of 150 nM and 80 nM p/t DNA (final concentrations). After pre-incubation to form the binary complex, reactions were started by the addition of dTTP to a final concentration of 35 µM and product formation was followed up to 0.2 s. Plot of the concentration of product formed with respect to time was biphasic in nature (Figure 4.7A). The fast phase represents the initial burst of dTTP incorporation by the pre-formed Sau-PoIC-ΔNΔExo • p/t DNA binary complex, while the slow phase represents dTTP incorporation in subsequent rounds of primer extension, after the enzyme dissociates from the first p/t DNA substrate and rebinds another. The data were fit using the full burst equation (Equation 4.4) [186]. The rates of the fast and slow phases obtained were $150 \pm 30$ s$^{-1}$ and $8.5 \pm 1$ s$^{-1}$, respectively. Product formed during the fast burst phase was $12 \pm 1$ nM, indicating that out of 150 nM of Sau-PoIC-ΔNΔExo, only 12 nM formed active enzyme • DNA binary complex that got converted to product.
Figure 4.7. Pre-steady-state kinetics and active site titration of Sau-PoIC-ΔNΔExo.
Figure 4.7. Pre-steady-state kinetics and active site titration of Sau-PolC-ΔNΔExo.

(A) A time course of primer extension under pre-steady-state condition in the presence (●) and absence (◆) of unlabelled p/t DNA acting as an enzyme trap. 35 µM dTTP (with or without 48 µM of unlabelled p/t DNA) was added to 150 nM Sau-PolC-ΔNΔExo (corresponding to an active Sau-PolC-ΔNΔExo concentration of 100 nM) and 80 nM p/t DNA (all concentrations are final). In the absence of the trap, the time course was biphasic in nature and the data were fit to the full burst equation (Equation 4.4). The rate of the fast phase was 150 ± 30 s⁻¹ and that of the slower phase was 8.5 ± 1 s⁻¹, [ED]ₐ was found to be 12 ± 1 nM. In the presence of the trap, the time course was monophasic and the data were fit to a single exponential equation, with a rate of 300 ± 14 s⁻¹ and an amplitude of 11.5 ± 0.5 nM. The data can also be fit equally well to the full burst equation, but the data were not sufficient to justify using the more complex model. (B) A representative set of primer extension assays performed during active site titration. Time resolved primer extension assays were performed using 150 nM Sau-PolC-ΔNΔExo, 1 mM dTTP and varying concentrations of p/t DNA (● 40 nM, ■ 80 nM, ▲ 160.1 nM, ▼ 284.76 nM, ◆ 379.69 nM, ○ 506.25 nM, + 675 nM and x 900 nM). The concentration of extended primer was plotted versus time and data were fit to the full burst equation (Equation 4.4). For ease of understanding, the background primer extension has been deducted from each time course. (C) A plot of the concentrations of pre-formed active enzyme-DNA complex getting converted to product before turnover ([ED]ₐ) versus DNA concentration was fit to a quadratic equation (Equation 4.5). KᵩDNA was determined to be 390 ± 70 nM and the concentration of active Sau-PolC-ΔNΔExo was found to be 100 ± 8 nM.
Since the rate of dissociation of the DNA substrate from the binary complex (Figure 4.2, step 5) was very fast, it was possible that the DNA did not form a stable ternary complex even in the presence of the correct incoming dNTP (Figure 4.2, step 3). In order to test whether such was the case, we repeated the above burst experiment in the presence of 48 µM of unlabelled p/t DNA that was added at the same time as the dTTP. Any Sau·PolC-ΔNΔExo that dissociated from the labeled p/t DNA would be trapped by the excess unlabelled DNA, which would eliminate the slower phase. Additionally, any unstable ternary complex having a dissociation rate comparable to the rate of chemistry or faster would result in lower amplitude of product formation in the presence of the trap. Our result shows that, in the presence of the DNA trap, the slow phase was eliminated, as expected, and the amplitude of product formation was 11.5 ± 0.5 nM (Figure 4.7A), identical to the amplitude in the absence of the trap.

These results indicate that although DNA dissociation from the binary complex is rapid, disassembly of the ternary complex is not rapid and, during a single nucleotide-incorporation cycle, DNA does not dissociate from the enzyme after nucleotide binds. The difference in the rates of product formation for the first and subsequent rounds of enzyme turnover, as observed in the burst experiment, indicates the presence of a slow and at least partially rate-limiting step after dNTP incorporation. The low burst amplitude suggested that binding of Sau·PolC-ΔNΔExo to p/t DNA was weak and/or only a fraction of the enzyme was active. A third possibility is the presence of an internal equilibrium in the pathway leading to a reduction in product formation. The following experiments indicate that all three of these possibilities contribute to the low burst amplitude.

4.5.6. Active site titration. The formation of a stable ternary complex and the presence of a slow, rate-limiting step after chemistry allowed us to perform burst kinetic assays to determine the apparent $K_{o}^{DNA}$ and the concentration of active Sau·PolC-ΔNΔExo. For these assays, the final concentration of total Sau·PolC-ΔNΔExo was 150 nM and the final DNA concentration was varied between 10 and 900 nM. Product formation for a representative set of DNA
concentrations is shown in Figure 4.7B. The time courses were fit to the full burst equation, and the concentrations of the initial active enzyme · DNA complex that was converted into product during the first round of catalysis ([ED]₀), obtained from the amplitudes of the fast phase, were plotted as a function of DNA concentration (Figure 4.7C). The data were fit to a quadratic equation (Equation 4.5). From the fit, the apparent KₐDNA was determined to be 390 ± 70 nM, indicating a relatively weak binding to DNA (Table 4.1), and the concentration of active Sau-PolC-ΔNΔExo was 100 ± 8 nM, implying that ~70% of the Sau-PolC-ΔNΔExo was active. The active enzyme concentration was lower than expected given the purity of the preparation, but the result was consistent for different preparations. From the apparent DNA binding affinity and the rate of DNA dissociation, we estimate that Sau-PolC-ΔNΔExo associates with DNA with a rate constant (kₐ) of ~4x10⁸ M⁻¹s⁻¹, which suggests that the rate of DNA binding is limited by diffusion.

4.5.7. Nucleotide binding affinity. Primer extension assays were again performed under burst conditions to determine the apparent KₐdNTP and kₚol. For these assays the final concentrations of active Sau-PolC-ΔNΔExo and p/t DNA were 804 nM and 50 nM respectively. The concentration of dTTP was varied from 1.17 to 100 µM and a representative range of data is shown in Figure 4.8A. Rates of the fast phase (k₁), obtained by fitting the time course to the full burst equation, were plotted against [dTTP] and the data were fit to a hyperbolic equation (Equation 6). From the fit, the apparent KₐdNTP was determined to be 3.2 ± 0.9 µM and kₚol was 180 ± 9 s⁻¹ (Figure 4.8B). Although these parameters were reasonable compared to other replicative polymerases, the overall fit to the data was not good (R² of 0.75), primarily because the observed rates for lower nucleotide concentrations did not match well with the rates predicted by the hyperbolic equation (Figure 4.8C). The deviation of the observed nucleotide incorporation rates appeared to be due to the lower amplitudes of the fast phase at low nucleotide concentrations.
Table 4.1. Comparison of kinetic parameters of different polymerases.

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>Family</th>
<th>$K_D^{DNA}$ (nM)</th>
<th>$K_D^{dNTP}$ (µM)</th>
<th>$k_{pol}$ (s⁻¹)</th>
<th>$k_{off}$ † (s⁻¹)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli Pol IV</td>
<td>Y</td>
<td>50</td>
<td>441</td>
<td>12</td>
<td>0.18</td>
<td>[187]</td>
</tr>
<tr>
<td>Sau-PolC-ΔNΔExo (observed)</td>
<td>C</td>
<td>390</td>
<td>4</td>
<td>180</td>
<td>150</td>
<td>-</td>
</tr>
<tr>
<td>Sau-PolC-ΔNΔExo (simulated)</td>
<td>C</td>
<td>-</td>
<td>7.5</td>
<td>330</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pol I (Klenow)</td>
<td>A</td>
<td>5</td>
<td>5.5</td>
<td>50</td>
<td>0.06</td>
<td>[34]</td>
</tr>
<tr>
<td>T7DNA Polymerase</td>
<td>A</td>
<td>(18)</td>
<td>(18)</td>
<td>(287)</td>
<td>(0.2)</td>
<td>[35]</td>
</tr>
<tr>
<td>E. coli Pol II</td>
<td>B</td>
<td>21</td>
<td>4.4</td>
<td>13</td>
<td>0.05</td>
<td>[188]</td>
</tr>
<tr>
<td>Mammalian Polδ</td>
<td>B</td>
<td>300</td>
<td>-</td>
<td>7</td>
<td>0.005</td>
<td>[189]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(64)</td>
<td>(0.93)</td>
<td>(21)</td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td>Yeast Polδ</td>
<td>B</td>
<td>30</td>
<td>24</td>
<td>1</td>
<td>0.045</td>
<td>[190]</td>
</tr>
<tr>
<td>Human mitochondrial Polγ</td>
<td>B</td>
<td>39</td>
<td>14</td>
<td>3.5</td>
<td>0.03</td>
<td>[191, 192]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.9)</td>
<td>(0.78)</td>
<td>(45)</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Polβ</td>
<td>X</td>
<td>49</td>
<td>110</td>
<td>10</td>
<td>0.3</td>
<td>[193]</td>
</tr>
</tbody>
</table>

Data obtained from this study are highlighted in bold.

Replicative polymerases are shown in italics.

Values shown in parentheses were measured in the presence of the corresponding processivity factor.

† For all DNA polymerases shown here except for Sau-PolC-ΔNΔExo $k_{off}$ is equivalent to $k_{cat}$. 
Figure 4.8. Determination of $K_{D}^{dNTP}$ of Sau-PolC-$\Delta$N$\Delta$Exo.

(A) A representative set of primer extension assays performed during $K_{D}^{dNTP}$ determination for Sau-PolC-$\Delta$N$\Delta$Exo. The reactions were performed with 804 nM active Sau-PolC-$\Delta$N$\Delta$Exo, 50 nM p/t DNA and various concentrations of dTTP ($\diamond$ 1.17 µM, $\bigcirc$ 4.69 µM, $\blacklozenge$ 9.4 µM, $\blacktriangledown$ 18.75 µM, $\blacktriangle$ 30.14 µM, $\blacksquare$ 50 µM and $\bullet$ 75 µM). The concentrations of extended primers were plotted against time and the plots were fit to the full burst equation (Equation 4.4). (B) A plot of the observed rates of the fast phase ($k_1$) versus [dTTP]. The data were fit to a hyperbolic equation (Equation 4.6). From the fit, $K_{D}^{dNTP}$ was determined to be 3.2 ± 0.9 µM and maximum rate of the burst ($k_{pol}$) was found to be 178 ± 9 s$^{-1}$. $R^2$ value for this fit was 0.75. (C) An enlarged view of panel (B) of up to 15 µM of [dTTP]. (D) A plot of $[ED]_A$ versus [dTTP] was fit to a hyperbolic equation (Equation 4.7). From the fit, $K_{D}^{dNTP}$ was found to be 4.0 ± 0.3 µM and the maximum $[ED]_A$ was 36 ± 0.6 nM. $R^2$ value for this fit was 0.98.
4.5.8. Dependence of active enzyme·DNA binary complex on [dNTP].

Typically, bond formation (Figure 4.2, step 3) is irreversible, because pyrophosphate release (Figure 4.2, step 4) is fast, and the binding of the incoming dNTP is a rapid equilibrium process [33]. Hence there is no equilibrium between dNTP binding (Figure 4.2, step 2) and bond formation. Therefore, an increase in the concentration of the incoming dNTP does not influence the concentration of the preformed active enzyme·DNA binary complex that gets converted to product before turnover ([ED]_A). This is observed as the lack of correlation in a plot of [ED]_A versus [dNTP]. Upon closer inspection of our data, however, we observed that [ED]_A obtained from the burst amplitude of the fast phase was dependent on the dTTP concentration, saturating at higher concentrations (Figure 4.8D) and fit well (R^2 of 0.98) to a hyperbolic equation (Equation 4.7). From the fit, the apparent K_D^{dNTP} was found to be 4.0 ± 0.3 µM and the maximum concentration of [ED]_A ([ED]_A^{max}) was 36 ± 0.5 nM. This dNTP concentration dependence of [ED]_A suggests that bond formation is reversible (Figure 4.2, step 3) and is in equilibrium with ground state dNTP binding (Figure 4.2, step 2). As a result, [ED]_A increases when increasing concentrations of dTTP drive the equilibrium towards product formation. This observation suggests that there is a slow step after catalysis but prior to PPi release that allows chemistry to be reversible.

We therefore turned to using KinTek Explorer to calculate the kinetic parameters for the forward and reverse steps of chemistry accurately by numerical integration, and also to define a rate constant for the slow step immediately after chemistry. The enzymatic pathway of Sau-PolC-ΔNΔExo was simulated through a three-step mechanism: (1) dNTP binding to the polymerase·DNA binary complex, (2) dNTP incorporation, and (3) PPi release (Figure 4.9A). We have used this model for simplicity because, for chemistry to be reversible, PPi must be positioned at the active site to cause pyrophosphorolysis. Although PPi release is modeled as a simple binding interaction, it is likely that this step in the pathway also involves a conformational change of the polymerase-DNA complex. Thus, it is important to keep in mind that the kinetic parameters defined for PPi release may actually describe more than one elementary step.
that occurs immediately after chemistry. Since the ternary complex is stable, as shown in Figure 4.7A, we assumed that DNA would not dissociate from the enzyme when either dNTP or PPI were bound. For the simulation, the rate of dNTP association was considered to be limited by diffusion and accordingly the second order association rate constant was fixed at 100 \( \mu \text{M}^{-1}\text{s}^{-1} \). Also, PPI release was assumed to be irreversible, since the concentration of PPI in solution during the reaction period would be negligible. The experimentally determined \( K_d^{\text{DNA}} \) and DNA release rate \( (k_{\text{off}}) \) from the binary complex were used to determine the concentration of the preformed \( \text{Sau-PolC-\DeltaN\DeltaExo} \cdot \text{p/t DNA complex} \) and the second order association rate constant for the formation of the binary complex. The simulated curves were generated through iterative steps that used kinetic parameters obtained from nonlinear regression as initial values.

Simulated curves were superimposed on the representative time course data shown in Figure 4.8A (Figure 4.9B). Through numerical integration, the rate of chemistry was determined to be \( \sim 220 \text{ s}^{-1} \) for the forward reaction and \( \sim 110 \text{ s}^{-1} \) for the reverse reaction, yielding a net maximum rate \( (k_{\text{pol}}) \) of \( \sim 330 \text{ s}^{-1} \). Also, the rate of pyrophosphate release following chemistry was determined to be \( 26 \text{ s}^{-1} \) and the \( K_d^{\text{dNTP}} \) was \( 7.5 \mu \text{M} \). From 3-D confidence contour analysis, all the calculated parameters appeared to be well constrained by the data (Figure 4.9C). From the simulation, the calculated rate of PPI release is much lower than the calculated rate of catalysis and is very close to the \( k_{\text{cat}} \) of \( \text{Sau-PolC-\DeltaN\DeltaExo} \), suggesting that PPI release (or a conformational change required for PPI release) may govern the steady-state rate.
Figure 4.9. Simulation of kinetic pathway of Sau-PolC-ΔNΔExo.

(A) The three-step reaction mechanism used for the simulation. Values obtained for the different rate constants are shown alongside the appropriate step. Rate of dNTP association to the E-D₀ binary complex was assumed to be diffusion limited and accordingly the second order rate constant for this step was fixed at 100 µM⁻¹s⁻¹. (B) Simulated curves generated for the representative dataset shown in Figure 4.8A superimposed on the raw data (concentrations of dTTP shown are ○ 1.17 µM, ◆ 4.69 µM, ● 9.4 µM, ▽ 18.75 µM, ▲ 30.14 µM, ■ 50 µM and × 75 µM). (C) 3-D confidence contours for the various rate constants determined from the simulation. For each case the search was carried out up to a sum of squares error (SSE) that is 2-fold higher than the minimum SSE. The upper and lower limits of each parameter were determined using an SSE threshold of 1.2.
4.6. DISCUSSION

We have determined the minimal kinetic pathway (Figure 4.10) for Sau-PolC-ΔNΔExo and defined parameters for individual steps within the pathway using both steady-state and pre-steady-state kinetic approaches. The kinetic steps fit the same pathway used by other polymerases, but PolC exhibits several distinguishing features (Table 4.1). To our knowledge, this is the first comprehensive kinetic study of the catalytic subunit of the bacterial replisome. As described below, our findings provide deeper insight into several activities of the bacterial C-family polymerases that have been previously observed.

Our data show that Sau-PolC-ΔNΔExo is a fast enzyme, with a maximum nucleotide incorporation rate $k_{pol}$ of 330 s$^{-1}$, but also binds DNA weakly, with a $K_D^{DNA}$ of 390 nM, and dissociates rapidly from DNA, with a $k_{off}$ of 150 s$^{-1}$ when dNTP is not bound. These parameters indicate that Sau-PolC-ΔNΔExo would have a very low processivity, which is consistent with previous reports about C-family replisome polymerases being non-processive in the absence of accessory protein factors like the β-clamp [175, 194]. Furthermore, weak DNA binding is important for rapid, high fidelity DNA synthesis by bacterial replisome polymerases, as evidenced by an E. coli DnaE mutant, dnaE173, which has a single amino acid substitution of E612K [195]. The mutant enzyme binds DNA more tightly and simultaneously becomes more processive (even in the absence of clamp), but also shows a reduced rate of DNA elongation and is more error-prone [195]. The data we have obtained for Sau-PolC-ΔNΔExo are also consistent with the little kinetic data previously available for full-length C-family polymerases [196-198].
Figure 4.10. Minimal enzymatic pathway for Sau-PolC-ΔNΔExo.

The kinetic parameters determined are shown alongside the corresponding steps of the pathway. $K_D^{dNTP}$, forward and reverse rates of chemistry and rate of PPi release were derived from simulation of the reaction pathway. The rate of enzyme-DNA association ($k_{on}$) was calculated from the $K_D^{DNA}$ and $k_{off}$ using the relation that $K_D^{DNA} = k_{off}/k_{on}$. It should be noted that although nucleotide binding and PPi release are each shown as single steps, they may in fact be comprised of more than one elementary step, such as a conformational change in the polymerase accompanying substrate binding and product release.
The most surprising result from these studies is that nucleotide binding is in equilibrium with the chemical step of the phosphoryl transfer reaction. For this to be the case, the step immediately after chemistry must be slow enough so that there is a build up of the enzyme-DNA-PPI complex resulting from chemistry, thus driving the reverse of nucleotide addition (pyrophosphorolysis). We hypothesize that PPI release after catalysis, or a conformational change that precedes PPI release, might act as a rate-limiting step in the catalytic cycle, allowing the polymerase to retain a conformation favorable to reversal of bond formation. From simulating the reaction mechanism, the rate of this slow step was calculated to be $26 \text{ s}^{-1}$ and is likely to be the main determinant of $k_{\text{cat}} (17 \text{ s}^{-1})$.

Equilibrium between chemistry and ground state dNTP binding has not, to our knowledge, been observed previously for standard dNTPs, but has been reported for mitochondrial DNA polymerase gamma incorporating the damaged nucleotide 8-oxo-dGTP and the nucleotide analog AZT-TP [199-201]. Since this was not observed with standard nucleotides, it was proposed to be a mechanism for removing non-natural nucleotides. If the slow step after chemistry plays the same role for PolC, it could provide a rudimentary proofreading function.

Previous studies show that the sliding β-clamp processivity factor increases not only the processivity of the C-family polymerases but also the intrinsic rate of nucleotide incorporation. E. coli pol III core, for example, synthesizes DNA at a rate of 20 nt/sec, but the rate increases to at least 500 nt/sec in the presence of clamp [202, 203]. Similarly, the intrinsic nucleotide incorporation rate of S. aureus PolC on primed circular single-stranded DNA was estimated to be ~80 nt/sec, and this increased to ~480 nt/sec in the presence of the sliding clamp [175]. Association with processivity factor, however, does not always stimulate the rate of nucleotide incorporation in this way. The processivity of E. coli pol II increases from 5 to ~1600 nucleotides in the presence of β-clamp, but nucleotides are incorporated at a rate of ~30 nt/sec in both cases [13]. These observations led to the proposal that β-clamp not only increases the processivity of the C-family replicative polymerases, but that it also increases the rate of a
limiting step in the catalytic cycle [175].

Our kinetic data suggest a mechanism by which β-clamp may increase the rate of nucleotide incorporation: by increasing the rate of the slow step immediately after chemistry. β-clamp increases processivity by encircling the DNA duplex and eliminating DNA dissociation as a step in the reaction pathway. However, since Sau-PolC-ΔNΔExo has a slow kinetic step (26 s	extsuperscript{-1}) between catalysis and DNA dissociation, nucleotide incorporation would not be stimulated if this was the only step in the reaction pathway that was changed. We hypothesize that β-clamp binding to PolC would speed up the rate of the slow step after chemistry, perhaps by stimulating PPI release. If the step immediately after chemistry was no longer rate limiting, we would expect the rate of processive DNA synthesis to increase to the rate of chemistry (330 s	extsuperscript{-1}).

Overall, the results presented here establish the kinetic foundation for future structure-function studies of the C-family polymerases by allowing a quantitative comparison of enzyme activities. Structural studies indicate that DnaE and PolC are members of the β-nucleotidyl transferase superfamily, making the bacterial replicative polymerases more closely related to eukaryotic repair polymerases (such as pol β from the X-family) than to eukaryotic or archaeal replicative polymerases [39, 52, 53, 55]. Some questions that can now be addressed include: How do the C-family polymerases incorporate nucleotides more than 100-fold faster than the X-family polymerases? What contributes to the substantial differences in substrate binding between the two polymerase families? How do the replicative C-family polymerases achieve such a high fidelity of DNA synthesis? PolC has been identified as a novel drug target for antibiotics against Gram-positive bacteria [204, 205]. Beyond increasing our fundamental understanding of bacterial DNA replication, addressing these questions may help in identifying novel features of PolC that could help in developing new antibiotics against Gram-positive pathogens.
4.7. ACKNOWLEDGEMENTS

We thank Thale Jarvis for the clone of *S. aureus* PolC, Ken Johnson for advice on the kinetic simulations, Karen Chave for assistance optimizing protein expression, and Ryan Wilson for careful proofreading of the manuscript. We acknowledge the use of the Wadsworth Center’s Protein Expression Core.
Lahiri I, Mukherjee P, Pata JD. Catalysis by *Staphylococcus aureus* PolC is regulated by the next correct (but not incorrect) nucleotide.

(MANUSCRIPT IN PREPARATION)

5.1. Description of contribution

For this manuscript, I contributed equally while designing and performing all the experiments except the fluorescence based pyrophosphate release assays that were entirely performed by Indrajit. I was involved equally in analyzing the results and writing the manuscript.
5.2. SUMMARY

Bacterial replication is carried out by C-family DNA polymerases. PolC, the first C-family member to be kinetically characterized is the major replicative polymerase in low G-C content Gram-positive bacteria. The kinetic mechanism of a N-terminal and exonuclease deficient construct of *Staphylococcus aureus* PolC revealed that nucleotidyl transfer (or chemistry) was fast (≈ 330 s⁻¹) and yet reversible. This surprising feature was suggested to be a consequence of the equilibrium that resulted between nucleotide binding and incorporation, since the reverse rate of chemistry was comparable or slower than a step following chemistry. Our results with an exonuclease inactivated full-length PolC, both alone and in the presence of the β clamp processivity factor corroborate the earlier observation of equilibrium. Measuring the rate of pyrophosphate (P Pi) release following nucleotide incorporation confirms P Pi release as the post-chemistry slow step (≈10-40 s⁻¹). This direct evidence conflicted with the ~25-100 fold faster rates reported for processive DNA synthesis by PolC. We find that presence of the next correct incoming nucleotide increases the measured rate of P Pi release, making it comparable to chemistry itself and enabling fast DNA synthesis. Our findings provide insight into a novel mechanism that could be employed by PolC to maintain accuracy at high speeds of DNA synthesis.
5.3. INTRODUCTION

The last 60 years have led to a detailed structural and biochemical understanding of the constituents and interactions involved in bacterial replication. The DNA polymerases that duplicate the bulk of the genome in bacteria with high accuracy (1 error/10^5 bases copied) and speed (500-1000 s⁻¹) [43, 71, 72, 74, 175] belong to the C-family [45]. In contrast, their replicative counterparts in eukaryotes and archaea mostly belong to the B-family [45]. In addition to the lack of sequence homology, recent structural studies have revealed that the topology of the catalytic palm domain of C-family polymerases places them in the β nucleotidyl transferase (β·nt) superfamily, which also includes error-prone X-family repair polymerases like Polβ [49] and is distinct from the classical superfamily members from the A-, B- and Y-families.

C-family replicative polymerases subdivide into the DnaE and PolC branches. While DnaE is found in Gram-negative bacteria such as *Escherichia coli* and high G-C Gram-positives like *Mycobacterium tuberculosis*, low G-C Gram-positive bacteria like *Staphylococcus aureus* and *Bacillus subtilis* use PolC as their primary replicative polymerase [50]. In addition to PolC, low G-C Gram positive organisms also have a DnaE homolog that, from a recent *in vitro* study, is suggested to be involved in initiating DNA synthesis from RNA primers laid down by primase [206]. After DnaE has synthesized a short stretch of DNA PolC is suspected to take over and replicate the bulk of the genome.

PolC and DnaE differ in their domain arrangement while sharing several accessory domains that have little-known functions [39, 53]. A structure for the ternary complex of *Geobacillus kaustophilus* PolC (Gka-PolC) with DNA and incoming dNTP exists [39] but most of the biochemical work on C-family polymerases has focused on reconstituting the replisome [24, 68, 71, 177, 206] with kinetic studies of this family being long confined to preliminary steady-state assays [196-198]. Only recently has there been a detailed pre-steady state kinetic study of a truncated version of *S. aureus* PolC (SaPolCΔNΔExo) from our lab describing the minimal reaction pathway for this enzyme [207]. This
truncated protein was identical in domain arrangement to the construct used for the crystal structure of Gka-PolC. Although the overall kinetic pathway was found to be similar to those described for other DNA polymerases [34, 35] (Figure 5.1A), several intriguing differences were observed.

PolC was found to bind DNA weakly ($K_{DNA} \sim 400\text{nM}$) mainly due to a rapid dissociation of the enzyme-DNA binary complex ($k_{off} \sim 150\text{ s}^{-1}$) and the rate of chemistry ($\sim 330\text{ s}^{-1}$) was comparable to the rate of DNA synthesis by the replisome. Unexpectedly, chemistry was found to be reversible and in equilibrium with dNTP binding. This equilibrium appeared to be established due to pyrophosphate (PPi) release ($\sim 26\text{ s}^{-1}$) following chemistry being much slower ($\sim 5$ fold) than the reverse rate of chemistry. Although a similar situation has been observed earlier for mitochondrial Polγ with some non-natural nucleotide analogs and lesion-containing nucleotide substrates [199-201], so far PolC appears to be the only DNA polymerase to exhibit this behavior for cognate dNTPs. This result posed an interesting question: if PPi release is slower than chemistry itself, how is this rate-limiting step overcome by PolC to achieve rapid processive synthesis ($\sim 500$ to $1000\text{ nt/s}$) that is nearly 20-40 fold faster than PPi release?
Figure 5.1. Reaction pathway of DNA polymerase.

(A) The minimal kinetic pathway followed by DNA polymerases is shown. Abbreviations used, E: DNA polymerase, DNA_n: unextended DNA, DNA_{n+1}: DNA extended by one nucleotide. (B) DNA substrates used in this study. Unless otherwise mentioned S1 is used in the absence of the clamp and S2 in the presence. "*" indicates the FAM label at the 5' end of the primer.
Here we have performed the kinetic analysis of full-length PolC with an inactivated 3’ to 5’ exonuclease domain (D426A and E428A) both alone and in the presence of the β-clamp processivity factor, with an aim to begin to understand the contribution of components of the replisome towards the kinetics of DNA synthesis. We find that the presence of the N-terminal and exonuclease domain slows down DNA dissociation from the binary complex, however, like the truncated version, full-length PolC exhibits reversible chemistry and slow PPI release (~ 10-40 s⁻¹) making it difficult to envision fast processive synthesis. To our surprise, we find that the presence of the next correct nucleotide stimulates PPI release to achieve processive synthesis at rates close to 1000 s⁻¹. Taken together our results indicate that modulation of the PPI release rate by the next correct incoming dNTP might act as an additional layer of regulation to ensure accurate DNA synthesis at high speed.

5.4. METHODS

5.4.1. Protein purification: Full length S. aureus PolC with point mutations at two amino acid residues (D426A and E428A) and a cleavable C-terminal hexa-histidine tag (PolC-ExoMut) was cloned into pET23a vector and transformed into Rosetta (DE3) pLysS E. coli cells. Protein expression and purification were performed using the same procedure described for the truncated PolC [207] with some modifications. Briefly, cells were grown to OD₆₀₀ of ~ 0.6 and induced with 0.5 mM IPTG overnight at 17°C. The C-terminal hexa-histidine tag was cleaved using TEV protease after initial purification on a Ni-NTA column. Untagged PolC-ExoMut was further purified through subsequent chromatographic purification steps as described earlier. Final concentration of the protein was calculated from absorbance at 280nm using a theoretical extinction coefficient of 108,000 M⁻¹cm⁻¹ and then stored at -80°C in storage buffer.

C-terminally hexa-histidine tagged S. aureus β clamp monomer was cloned in pET23a [175] (a gift from Dr. Mike O’Donnell, The Rockefeller University) and transformed into Rosetta(DE3) pLysS E. coli cells. Cells were grown to OD₆₀₀ of ~ 0.6 and induced overnight with 0.5 mM IPTG at 15°C. Protein purification was
carried out using three chromatographic steps. Clarified cell lysate resuspended in lysis buffer (25 mM HEPES (pH 7.5), 500 mM NaCl, 50 mM imidazole, 10% glycerol) was passed through (2X5 ml) nickel-charged columns (GE Bioscience) pre-equilibrated in lysis buffer and the protein was eluted using a linear gradient of 50-1000 mM imidazole. NaCl concentration in the collected peak was reduced to ~ 250 mM by diluting in buffer (25 mM HEPES (pH 7.5), 0.1 mM EDTA, 5 mM DTT, 10% glycerol). The diluted protein was passed over (2X5 ml) anion exchange columns (Q-column) (GE Bioscience) pre-equilibrated in Buffer A (25 mM HEPES, 250 mM NaCl, 0.1 mM EDTA, 5 mM DTT, 10% glycerol). Protein was eluted using a linear gradient of 250-1000 mM NaCl. For the final step the eluate from the anion exchange columns were subjected to size exclusion chromatography using a Superdex S-75 column (GE Bioscience) pre-equilibrated in storage buffer (25 mM HEPES (pH 7.5), 250 mM NaCl, 10% glycerol, 0.1 mM EDTA, 5 mM DTT). Final concentration of the protein was estimated from absorbance at 280 nm using a theoretical extinction coefficient of 15,930 M⁻¹cm⁻¹ and stored at -80°C in storage buffer.

5.4.2. DNA annealing: Chemically synthesized 5’6-FAM labeled DNA primers were annealed to unlabeled DNA templates to form duplex DNA substrates (Figure 5.1B) by heating to 95°C and slowly cooling to 25°C in annealing buffer containing 10 mM Tris-HCl (pH 8), 1 mM EDTA and 100 mM NaCl.

5.4.3. Primer extension assays: All primer extension assays were performed using the rapid chemical quench device RQF-3 (Kintek Corp.). The following protocol was followed for all the primer extension assays in this study except those performed to obtain the dissociation rate from the binary complex. PolC-ExoMut was pre-incubated with the appropriate DNA substrate (Figure 5.1B) and, when included, β clamp was added during the pre-incubation. Reactions were initiated by adding one or more deoxynucleoside triphosphates (dNTP)s. The final concentrations of reactants in the presence of the clamp was 500 nM PolC-ExoMut, 10 µM β clamp (i.e. 20 µM β monomer, 25-400 nM DNA substrate S2, and 0.8-1000 µM dNTP while the concentrations of reactants in the absence of the clamp were 1 µM PolC-ExoMut, 50-800 nM DNA substrate S1 and 0.8-1000
µM dNTP. Where appropriate, 48 µM unlabelled trap DNA (final concentration) was added along with the dNTP. All reactions were performed at room temperature in PolC reaction buffer (25 mM MES-Tris (pH 7.5), 25 mM NaCl, 8 mM MgCl₂, 2 mM DTT, 5% glycerol and 5 µM BSA). After different time intervals 250 mM EDTA was used to quench the reactions. Quenched samples were run on a 17% denaturing urea-polyacrylamide gel and imaged with a Typhoon 9400 scanner. The primer bands were quantified using ImageQuant software and the amount of product formed was determined by measuring the intensity of the extended primer band with respect to the total intensity of extended and unextended primers.

Dissociation rate of PolC from the binary complex was determined by performing primer extension assays as described previously [207]. Briefly, pre-incubated PolC-Exo Mut and labeled DNA was rapidly mixed with 96 µM trap DNA and incubated for different time intervals. Reactions were initiated by adding 500 µM dNTP and allowed to continue for 28 ms before quenching with 250 mM EDTA. Quantitation of the product formed was performed as mentioned above.

5.4.4. Data analysis: All non-linear regression analysis was performed using GraphPad Prism (v6.0). Primary data providing product concentration was plotted as a function of time and the reaction time courses were fit to the burst equation (Equation 5.1) or to a single exponential equation (Equation 5.2) as shown below:

\[ Y = A(1 - e^{-kt}) + c \] (5.1)

where \( Y \) is the concentration of the product formed, \( A \) is amplitude of the reaction, \( k \) is the observed single exponential rate and all other parameters are as mentioned above.

For active site titration assays the amplitudes (A) were further plotted as a function of substrate DNA concentration and the data was fit to a quadratic
equation (Equation 5.3).

\[
A = \frac{((K_D^{DNA})_{App} + E_A + [DNA]_T)}{2} - \sqrt{((K_D^{DNA})_{App} + E_A + [DNA]_T)^2 - 4E_A[DNA]_T}
\]  
(5.3)

Where \((K_D^{DNA})_{App}\) is the apparent dissociation constant for the binary complex, \(E_A\) is the maximum concentration of the pre-chemistry binary complex that gets converted to product during the first round of catalysis and \([DNA]_T\) is the initial DNA concentration at the beginning of the reaction.

For assays to determine \(K_D^{dNTP}\), both the amplitudes \(A\) and rates of the fast phase \(k_i\) were plotted as a function of \([dNTP]\) and the data were fit to equations 5.4 and 5.5 respectively.

\[
A = \frac{A_{max}[dNTP]}{((K_D^{dNTP})_{App} + [dNTP])}
\]  
(5.4)

\[
k_i = \frac{k_{pol}[dNTP]}{((K_D^{dNTP})_{App} + [dNTP])}
\]  
(5.5)

Where \(A_{max}\) is the maximum amplitude of the fast phase, \(k_{pol}\) is the maximum rate of bond formation and \((K_D^{dNTP})_{App}\) is the apparent dissociation constant of the nucleotide from the pre-chemistry ternary complex.

**5.4.5. Simulation:** The reaction mechanisms of PolC-Exo\(^{Mut}\) (both in the presence and absence of beta clamp) were simulated using KinTek Explorer software [185]. The reaction schemes used for the simulation are discussed in the “Results” section. Simulated curves and the raw data were exported from the software and final plots overlaying the raw data on the simulated curves were prepared using GraphPad Prism. To determine the range within which each rate constant was constrained by the model, and to investigate the relationships between different rate constants, we computed 3-D confidence contour plots for all possible pairs of rate constants [184].

**5.4.6. Measurement of pyrophosphate release rate:** Pyrophosphate (PPi) release rates of PolC-Exo\(^{Mut}\) in the presence and absence of beta clamp and
(where applicable) the next correct incoming nucleotide, dATP, was determined using a previously published protocol [199, 208]. A SF-120 stopped-flow apparatus (KinTek Corp.) was used for the assay. PolC-Exo\textsuperscript{Mut} and DNA were pre-incubated in the presence or absence of the \( \beta \) clamp (concentrations of reactants identical to primer extension assays) along with 1.5 mM \textit{E. coli} phosphate binding protein mutant labeled at Cys197 with \text{N-[[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC-PBP), 100 mM 7-methylguanosine, 0.02 U/ml purine nucleotide phosphorylase and 0.005 U/ml yeast inorganic pyrophosphatase. The solution was then rapidly mixed at 25°C with one or more dNTPs (200 µM each) and 100 mM 7-methylguanosine and 0.02U/ml purine nucleotide phosphorylase. The fluorophore was excited at 425 nm and a 450 nm long pass filter was used to observe emission. The reaction was performed in PolC reaction buffer. The fluorescence data thus obtained was fit globally by computer simulation along with relevant primer extension assay data.

5.5. RESULTS

5.5.1. \textit{N-terminal and exonuclease domains slow DNA release without improving (K\textsubscript{D\textsuperscript{DNA}})\textsubscript{App}}

Extending from our earlier study (Ref) with the PolC\textDelta N\textDelta Exo construct, we first investigated if full-length PolC also displayed biphasic product formation under pre-steady state conditions. The biphasic nature of product formation (Figure 5.2A) confirmed the presence of a slow step after chemistry. The biphasic nature of product formation was exploited to obtain an apparent dissociation constant for the binary complex ((K\textsubscript{D\textsuperscript{DNA}})\textsubscript{App}). As expected, the amplitude of the burst increased with an increase in [DNA] and on plotting the amplitude against [DNA], we found that similar to PolC\textDelta N\textDelta Exo, PolC-Exo\textsuperscript{Mut} displayed weak DNA affinity ((K\textsubscript{D\textsuperscript{DNA}})\textsubscript{App}) of \( \sim 530 \pm 280 \) nM (Figure 5.2C).
Figure 5.2. Pre-steady state kinetics of PolC-Exo$^{\text{Mut}}$.
Figure 5.2. Pre-steady state kinetics of PolC-Exo<sup>Mut</sup>.

(A and B) Time courses of product formation for primer extension assays in the (A) absence and (B) presence of the β clamp. The time courses were fit to a burst equation (Equation 1). For (A) final concentrations of reactants were 50 nM S1, 1 µM PolC-Exo<sup>Mut</sup> and 1 mM dTTP. For (B) final concentration of reactants were 50 nM S2, 500 nM PolC-Exo<sup>Mut</sup>, 10 µM β clamp (dimer) and 1 mM dTTP. (C and D) Secondary plots of Amplitude v/s [DNA] for active site titration assay in the (C) absence and (D) presence of the clamp. Data were fit to a quadratic equation (Equation 5.3) and the (K<sub>D<sub>DNA</sub></sub>)<sub><sub>App</sub></sub> thus obtained is mentioned in the graph. Final concentrations of reactants for (C) were 1 µM PolC-Exo<sup>Mut</sup>, 1 mM dTTP and S1 being varied from 50 to 800 nM, and those for (D) were 500 nM PolC-Exo<sup>Mut</sup>, 10 µM β clamp, 1 mM dTTP and S2 varied from 25 to 400 nM. (E and F) Dissociation rates of the pre-chemistry binary complex in the (E) absence and (F) presence of the clamp. Data were fit to a single exponential equation (Equation 2) and the resulting k<sub>off</sub> are mentioned in the graph. Final concentrations of the reactants for (E) were 50 nM S1, 1 µM PolC-Exo<sup>Mut</sup>, 48 µM trap DNA (S1 without 5' FAM label) and 500 µM dTTP and those for (F) were 50 nM S2, 500 nM PolC-Exo<sup>Mut</sup>, 48 µM trap DNA (S2 without the 5' FAM label), 10 µM β clamp and 500 µM dTTP. (G and H) primer extension assays to determine stability of the pre-chemistry ternary complex in the (G) absence and (H) presence of the clamp. Primer extension assays were done in both the presence and absence of the trap DNA and the time courses of product formation were fit to the burst equation. Comparison of the amplitudes of the fast phase in the presence (○) and absence (●) of the trap indicated the stability of the ternary complex. Final concentrations of the reactants for (G) and (H) were identical to that used for (E) and (F) except dTTP concentration used was 100 µM.
Interestingly, only ~18% of the total enzyme appeared active for PolC-Exo\textsuperscript{mut}. This was surprising since PolCΔNΔExo preparations had been ~70% active on the same substrate [207]. Similar low active enzyme concentrations have been reported previously for the eukaryotic replicative polymerases, ε (~15%) and δ (~35%) [209, 210]. Despite an inactivated exonuclease domain, with alanine substitutions for its key catalytic residues (D426 and E428), the exo-domain in PolCExo\textsuperscript{mut} still retains its DNA binding ability. Since only the fraction of enzyme that has DNA bound to the polymerase domain can be detected through primer extension assays, a large subpopulation may remain undetected when bound to the exo-domain, thus explaining the apparent low active enzyme concentration observed. Indeed, for \textit{E. coli} Pol III holoenzyme it has been shown recently, that when not undergoing polymerization the DNA substrate gets rapidly degraded by the ε proofreading subunit [211], indicating that the DNA can sample the exonuclease site easily. During \textit{in vivo} DNA synthesis however, presence of the entire replication machinery could provide additional constraint on the relative orientation of the polymerase and DNA, thereby preventing the substrate from prematurely binding the exonuclease site.

In case of PolCΔNΔExo we found that a very high dissociation rate of the pre-chemistry binary complex (k\textsubscript{off}) was responsible for the weak (K\textsubscript{DNA}\textsuperscript{App}) for PolCExo\textsuperscript{mut}, a similar situation holds true since the binary complex dissociates at a rate of 7.4 ± 3 s\textsuperscript{-1} (Figure 5.2E), comparable to that for the bacterial Y-family polymerase Pol IV [187]. Although fast compared to other replicative DNA polymerases, k\textsubscript{off} for PolCExo\textsuperscript{mut} was ~20 fold slower than for PolCΔNΔExo, indicating that the presence of the N-terminal and the exonuclease domains does stabilize the binary complex. Based on the relation K\textsubscript{DNA} = k\textsubscript{off}/k\textsubscript{on}, where k\textsubscript{on} is the second order association rate constant, we calculated a k\textsubscript{on} of 1.25 X 10\textsuperscript{7} M\textsuperscript{-1}s\textsuperscript{-1}. This was ~ 30 fold slower than the corresponding k\textsubscript{on} for PolCΔNΔExo and suggested that although the binary complex is more stable for full length PolC a slower association rate prevents tighter DNA binding.
5.5.2. β clamp increases the affinity of the binary complex:

The sliding clamp is the closest associated protein-partner for a DNA polymerase in the replisome. For replicative C-family polymerases, it has been shown that the β clamp increases both the processivity and activity of the enzymes [175] although the exact mechanism remains unclear. We hypothesized that similar to the eukaryotic DNA Polδ and mitochondrial Polγ [192, 209], the sliding clamp increases the affinity of PolC for its DNA substrate, thus contributing to an improvement in both processivity and overall activity. We tested this by measuring the kinetic parameters for PolCExoMut in the presence of the S. aureus β clamp. For all experiments that included the sliding clamp, a longer substrate (Figure 5.1B, substrate S2) was used to accommodate the β-clamp on the DNA along with PolC.

We found that the presence of the clamp did improve DNA binding affinity by ~5-fold ((K_D^{DNA})_{App} 116 ± 40 nM) (Figure 5.2B and D) but the active enzyme concentration remained unaltered at ~ 18%. The improved (K_D^{DNA})_{App} in the presence of the β clamp could be a result of slower dissociation of the binary complex or a faster association of the DNA and PolC. Therefore, we measured k_{off} (0.7 ± 0.07 s⁻¹) (Figure 5.2F) and then calculated the corresponding k_{on} (0.6 X 10⁷ M⁻¹s⁻¹) and find that the clamp stabilizes the binary complex by reducing the DNA dissociation rate by ~ 10 fold.

In order to ensure that the DNA substrate did not play a significant role in determining stability of the binary complex for PolCExoMut, we used substrate S2 to obtain k_{off} in the absence of β clamp (data not shown) and found this rate comparable to the one measured using S1 (7 vs 18 s⁻¹).

5.5.3. Incoming dNTP stabilizes the pre-chemistry ternary complex:

The relatively high rate of DNA dissociation from the binary complex (7 s⁻¹) in case of PolC alone led us to ascertain whether the pre-chemistry ternary complex was indeed stable (Figure 5.2G). Primer extension was performed in
the presence of excess unlabelled DNA trap to ensure that polymerase that
dissociated from the pre-chemistry ternary complex did not rebind labeled DNA.
Amplitudes of the fast phase of product formation (a measure of the initial
ternary complex that formed product) were identical irrespective of the trap
being present. Furthermore, this was not influenced by the presence of the
clamp (Figure 5.2H) indicating that despite the formation of an unstable binary
complex with the polymerase alone, a stable pre-chemistry ternary complex was
formed when incoming nucleotide was present.

Surprisingly, product formation in the presence of the DNA trap showed
biphasic behavior and was true irrespective of the clamp (Figure 5.2G-H).
Separate control experiments confirmed that this was not due to an ineffective
trap (data not shown). One possible explanation for the biphasic product
formation observed with the trap could be the presence of a slow internal
equilibrium.

5.5.4. Nucleotide incorporation is reversible both in the presence and
absence of the clamp:

Typically, when the correct nucleotide concentration is varied to determine the
dissociation constant of dNTP, \( (K_{d\text{dNTP}}^{\text{App}}) \), the rate of the fast phase of product
formation \( (k_f) \) is seen to depend hyperbolically on dNTP concentration while the
amplitude of this phase does not. However, PolC-Exo\textnormal{Mut} showed both rate and
amplitude dependence on dNTP concentration (Figure 5.3C and 5.3E). Further,
we obtained an apparent maximum rate of polymerization \( (k_{pol}^{\text{App}}) \) of \( 350 \pm 19 \text{ } \text{s}^{-1} \)
and \( (K_{d\text{dNTP}}^{\text{App}}) \) of \( 3.5 \pm 0.7 \) and \( 1.3 \pm 0.2 \) \( \text{µM} \) respectively from the rate and
amplitude dependence on [dNTP]. Although uncommon, amplitude dependence
on [dNTP] was not surprising since it had been noted earlier for PolCΔNΔExo
[207].
Figure 5.3. Affinity of dNTP for the pre-chemistry binary complex.
Figure 5.3. Affinity of dNTP for the pre-chemistry binary complex.

(A and B) Time course of product formation for various [dTTP] in the (A) absence and (B) presence of the β clamp. The time courses were fit to the burst equation (Equation 5.1). [dTTP] and the symbols used to represent the corresponding time course are shown in the figure. Concentrations of reactants in (A) were 50 nM S1 and 1 µM PolC-Exo<sup>mut</sup> and those in (B) were 50 nM S2, 500 nM PolC-Exo<sup>mut</sup> and 10 µM clamp. (C and D) Secondary plots of the rate of the fast phase of product formation (k<sub>1</sub>) versus [dTTP] in the absence (C) and presence (D) of the clamp. Data were fit to a hyperbolic equation (Equation 5.4) and the (K<sub>D_dNTP</sub>)<sub>App</sub> and k<sub>pol</sub> thus obtained are shown in the graph. (E and F) Secondary plots of amplitude of the fast phase of product formation versus [dTTP] in the (E) absence and (F) presence of the clamp. The data fit well to a hyperbola (Equation 5.5) and (K<sub>D_dNTP</sub>)<sub>App</sub> obtained from the fit are shown in the graph.
Amplitude dependence on [dNTP] indicates the presence of a slow post-chemistry step, presumably PPI release, such that the reverse rate of chemistry is faster than the rate of pyrophosphate release. In order to verify if such was the case, we experimentally determined the PPI release rate using a previously described fluorescence based assay [199, 208]. The stopped-flow fluorescence traces and the corresponding quench-flow single-nucleotide incorporation data from the \((K_{d}^{dNTP})_{app}\) determination experiments were fit together globally using KinTek Explorer software (Johnson 2009) (Figure 5.4B and D). From this global fit it was evident that PPI release was indeed slower (10 s\(^{-1}\)) (Figure 5.4A) than chemistry (350 s\(^{-1}\)) for PolC-Exo\(^{Mut}\).

Although the results are consistent with our previous observations for PolC\(^{\Delta N\Delta Exo}\), it poses a puzzle: PPI release needs to occur after every round of dNTP incorporation; if it is slower than chemistry then the rate of processive synthesis should be governed by the rate of PPI release. If this is true then how does processive synthesis by PolC achieve the well-documented speed of 500 – 1000 s\(^{-1}\) for high fidelity DNA replication?

A previous study suggested that addition of the \(\beta\) clamp processivity factor significantly increases the rate of DNA synthesis by \textit{S. aureus} PolC on linear M13mp18 DNA [175]. Based on this we wanted to ascertain if the presence of the \(\beta\) clamp could increase the rate of PPI release and consequently increase the rate of processive synthesis by PolC. Unexpectedly though, even in the presence of the clamp amplitude of the fast phase remained hyperbolically dependent on [dNTP] (Figure 5.3B and F), indicating that PPI release was still slow. Experimental measurement of the PPI release rate and global fit of the data confirmed this (Figure 5.4C and E). We determined a PPI release rate of \(~40\) s\(^{-1}\) while the forward and reverse rates of chemistry were 540 s\(^{-1}\) and 83 s\(^{-1}\) respectively (Figure 5.4A). Furthermore, the \((K_{d}^{dNTP})_{app}\) obtained was comparable in all cases, ranging between 2-5 \(\mu\)M.
Figure 5.4. Simulation of the reaction pathway of PolC-Exo$^{\text{Mut}}$. 
Figure 5.4. Simulation of the reaction pathway of PolC-Exo$^\text{Mut}$.

(A) The four-step reaction mechanism used for the simulation. Values obtained for the different rate constants in the absence of the clamp are shown in green and those obtained in the presence of the clamp are shown in maroon with the common rate constants shown in black. Rate constants shown in parentheses were not allowed to vary during the simulation. (B and C) Simulated curves generated for the datasets shown in Figure 5.3A and B respectively superimposed on the raw data. (D and E) Fluorescence traces measuring PPi release in the (D) absence and (E) presence of the β clamp with the corresponding simulated curve superimposed on them. The final concentrations of the reactants were 50 nM S1, 1 µM PolC-Exo$^\text{Mut}$ and 100 µM dTTP for (D) and 50 nM S2, 500 nM PolC-Exo$^\text{Mut}$, 10 µM clamp and 100 µM dTTP for (E). (F and G) 3-D confidence contours for the various rate constants determined from the simulation in the (F) absence and (G) presence of the β clamp. In (F) the search was carried out up to a sum of squares error (SSE) that is 1.05-fold higher than the minimum SSE. The upper and lower limits of each parameter were determined using an SSE threshold of 1.01. In (G) the search was carried out up to a sum of squares error (SSE) that is 1.1-fold higher than the minimum SSE. The upper and lower limits of each parameter were determined using an SSE threshold of 1.02.
5.5.5. Processive synthesis by PolC is fast:

In order to verify that our PolC preparations support fast processive synthesis, we determined the rate of processive DNA synthesis by PolC-Exo\textsuperscript{Mut} in the presence of all four nucleotides (Figure 5.5). The rate of processive synthesis was estimated from time taken to generate the full-length product (S1:19 nts and S2:38 nts) using the following equation:

\[
\text{Rate} = \frac{\text{Number of dNTPs added to generate the full length product}}{\text{Time taken to form the full length product}}
\]

In the presence of the clamp, PolC-Exo\textsuperscript{Mut} successfully generated a full-length product of 38 nucleotides within ~ 50 and 60 ms suggesting a rate of ~ 600 to 750 nt/s (Figure 5.5A). This was in good agreement with the speed of bacterial DNA synthesis and nearly 20-fold faster than the expected rate if PPI release remained slow. For extension by the polymerase alone we observed several prominent stall sites (Figure 5.5B and supplementary figure 5.S1) that made analysis difficult. To circumvent this problem, the DNA synthesis rate was estimated from the number of nucleotides that were added within the first time interval (10 ms) and again, we obtained a rate of ~ 1000 nt/s, comparable to rapid synthesis (Figure 5.5B). Repeating this experiment for PolC-Exo\textsuperscript{Mut} alone with both substrates used in this study (S1 and S2) gave comparable results (data not shown) indicating that the DNA substrate did not contribute significantly to the synthesis rate. Furthermore, both methods of calculation provided consistent results for extension in the presence of the clamp.

We further confirmed our observation that processive synthesis by PolC is fast by performing the processive synthesis assay in the presence of only three instead of all four dNTPs. The nature of the DNA template and the selection of nucleotides provided result in extension of the primer by only 5 base pairs (data not shown). Rate of processive synthesis calculated from this assay was also ~ 1000 s\(^{-1}\) (5 ntds extended in 5 ms) and again, the synthesis rate remained independent of the presence of the processivity factor.
Figure 5.5. Multiple nucleotide extension by PolC.

Multiple nucleotide extension assays were performed in the presence of all four dNTPs both in the (A) presence and (B) absence of the $\beta$ clamp. The arrow in (A) indicates the full-length product obtained by addition of 38 nucleotides. In (B) the number of dNTPs inserted is written along side the corresponding extended primer band. Adding 19 dNTPs forms full-length product. For (A), final concentrations of reactants were 50 nM S2, 500 nM PolC-Exo$^{\text{Mut}}$, 10µM $\beta$ clamp (dimer) and 100 µM of each dNTP. Final concentrations of the reactants for (B) were 50 nM S1, 1µM PolC-Exo$^{\text{Mut}}$ and 100 µM of each dNTP.
5.5.6. **Presence of next correct incoming dNTP allows rapid pyrophosphate release:**

Fast processive synthesis by PolC in the presence of all four nucleotides suggests the surprising possibility that the PPI release rate increases (~20 to 100 fold) when the next incoming dNTP is present. To validate this observation, we experimentally measured the PPI release rate of PolC-Exo\textsuperscript{Mut} in the presence of the first two incoming nucleotides (1\textsuperscript{st}: dTTP and 2\textsuperscript{nd}: dATP). The choice of the nucleotides allowed extension of the primer strand by two nucleotides only and represents the simplest case of processive synthesis. To reduce complications that may arise from distributive synthesis, the β clamp processivity factor was present in these assays. Under these conditions we performed both quench-flow primer extension and stopped-flow PPI release assays and the data were fit globally using KinTek Explorer (Figure 5.6).

The rates determined for the two PPI release events that occur during incorporation of dTTP and dATP were found to differ by ~20-fold (Figure 5.6B). Although, the rate of the second PPI release (i.e. after dATP addition) was slow (~50 s\textsuperscript{-1}) and similar to our previously determined rate of 40 s\textsuperscript{-1}, the first PPI release event (i.e. after dTTP addition) was distinctly fast (~800 s\textsuperscript{-1}) and comparable to the rate of both chemistry and processive synthesis (Figures 5.4A and 5.5). Our observations suggest that for the first dNTP incorporation, the presence of dATP (the next correct nucleotide) ensured that PPI release was rapid while the absence of the next correct dNTP (i.e. dGTP) from the reaction mix during the second nucleotide addition (dATP) resulted in the second PPI release event following dATP addition to be slow. This surprising observation clearly implicates the presence of the next correct incoming nucleotide as a crucial factor for influencing pyrophosphate release and altering PolC kinetics during processive DNA synthesis.
Figure 5.6. Simulation of the processive synthesis reaction pathway for PolC.
Figure 5.6. Simulation of the processive synthesis reaction pathway for PolC.

(A) Time course of product formation by PolC in the presence of two dNTPs and the β clamp. The appearance and disappearance of primer bands is shown with the first (blue) and second (red) nucleotide additions depicted separately. Superimposed on the raw data are the curves generated from the simulation of the processive synthesis pathway of PolC-ExoMut. (B) Fluorescence trace measuring PPi release in the presence of two incoming dNTPs and the clamp. The corresponding simulated curve is superimposed on the raw data. The rates of the first and second PPi release events as determined from the simulation of the processive synthesis pathway are mentioned in the graph. Concentrations of the reactants are same for both parts of this figure: 50 nM S2, 500 nM PolC-ExoMut, 10 µM β clamp (dimer) and 100 µM each of dTTP (the first incoming nucleotide) and dATP (the second incoming nucleotide).
5.6. DISCUSSION

In this study, we have performed the pre-steady state kinetic characterization of exonuclease inactivated full-length construct of \textit{S. aureus} PolC (PolC-ExoM) (Table 5.1). We find that post-chemistry pyrophosphate release by PolC-ExoM is slow (10 to 50 s\(^{-1}\)) irrespective of the presence of the \(\beta\)-clamp and within the predicted range for PolC\(\Delta\!\!\!\Delta\)Exo. Despite this slow step, PolC efficiently performs processive synthesis at rates comparable to bacterial replication fork progress (~700-1000 s\(^{-1}\)), but does so only when the next correct nucleotide is present. Taken together our results provide a thorough understanding of the enzymatic mechanism of the bacterial replicative polymerase and provide novel insight into its regulation.

The early characterized enzymes from bacteriophages [35] and viruses [212] led to the understanding that replicative DNA polymerases bind DNA very tightly (\(K_D^{\text{DNA}}\) in low nanomolar range). However, more recent biochemical characterizations of cellular replicative polymerases from bacteria (\textit{E.coli} DnaE) and eukaryotes (Pol \(\delta\) and Pol \(\varepsilon\)) have revealed weak DNA binding [209, 210, 213] that is in agreement with our observation for \textit{S. aureus} PolC ((\(K_D^{\text{DNA}}\)\text{App} \text{~500 nM}). 5-fold tighter DNA binding by PolC was achieved through association with the processivity factor as has been shown for DnaE, Pol\(\delta\) and mitochondrial replicative DNA polymerase PolY [192, 209, 213]. Notably, the presence of the N-terminal and exonuclease domains in this construct did not appear to influence DNA binding when compared to PolC\(\Delta\!\!\!\Delta\)Exo (Table 5.1).

In contrast to the similarity observed for DNA binding affinity of both PolC constructs, the DNA dissociation rate (\(k_{\text{off}}\)) from the pre-chemistry binary complex is reduced by nearly 20-fold in the presence of the N-terminal and exonuclease domains indicating that these additional domains help stabilize the binary complex once formed. This is not unexpected since the N-terminal domain of PolC has been predicted to consist of \(\tau\)- and DNA binding domains that may be involved in reducing \(k_{\text{off}}\) [214]. Moreover, \textit{E. coli} DnaE is known to
show a significant improvement in polymerase activity when associated with its proofreading subunit (ε) [203, 211]. The crystal structure of Gka-PolCΔNΔExo and subsequently chemical crosslinking experiments with DnaE have also suggested the involvement of the exonuclease domain in making contacts with duplex DNA, hence reducing fall off from the substrate [39, 211]. Addition of the β-clamp reduces the $k_{\text{off}} \sim 10$-fold more (Table 5.1), becoming the major reason for the improvement observed in DNA binding affinity. This has been reported for other DNA polymerases [192, 209]; however, it must be noted that the DNA dissociation rate obtained with the processivity factor is still faster than for most replicative DNA polymerases suggesting that other factors contributing to complex stability must exist. The results reported here could be an underestimation of the effects of the clamp because the DNA substrate used in this study is linear.

The slower $k_{\text{off}}$ for PolC-ExoMut compared to the deletion construct is offset by the corresponding $\sim 30$-fold slower association rate constant ($k_{\text{on}} = 1.25 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$) and resulted in the nearly unchanged $K_D^{\text{DNA}}$ for full-length PolC. The reduction in the on-rate may be attributed to the presence of an inactivated exo-domain and/or the N-terminal domain that can still bind DNA. It remains unclear how the DNA bound to the exonuclease domain transitions to the polymerase active site, what facilitates this switch and how this influences the DNA association rate.
Table 5.1. Kinetic parameters obtained with different constructs of PolC.

<table>
<thead>
<tr>
<th></th>
<th>PolCΔNΔExo</th>
<th>PolC-Exo\textsuperscript{Mut}</th>
<th>PolC-Exo\textsuperscript{Mut + β}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D^{DNA}$</td>
<td>390 nM</td>
<td>530 nM</td>
<td>116 nM</td>
</tr>
<tr>
<td>$K_D^{dNTP}$</td>
<td>7.5 µM</td>
<td>2 µM</td>
<td>4.6 µM</td>
</tr>
<tr>
<td>$k_{off}$</td>
<td>150 s\textsuperscript{-1}</td>
<td>7.4 s\textsuperscript{-1}</td>
<td>0.7 s\textsuperscript{-1}</td>
</tr>
<tr>
<td>$k_{\text{forward}}$</td>
<td>220 s\textsuperscript{-1}</td>
<td>270 s\textsuperscript{-1}</td>
<td>540 s\textsuperscript{-1}</td>
</tr>
<tr>
<td>$k_{\text{reverse}}$</td>
<td>110 s\textsuperscript{-1}</td>
<td>120 s\textsuperscript{-1}</td>
<td>83 s\textsuperscript{-1}</td>
</tr>
<tr>
<td>$k_{PPi}$</td>
<td>27 s\textsuperscript{-1}</td>
<td>10 s\textsuperscript{-1}</td>
<td>40 s\textsuperscript{-1}</td>
</tr>
</tbody>
</table>
Taken together our findings and the above-mentioned reports indicate that unlike their replicative counterparts in bacteriophages and viruses, the cellular replicative polymerases form a relatively weak pre-chemistry binary complex with DNA that is stabilized by interaction with their respective processivity factors. Replicative DNA polymerases function as part of a large multiprotein complex and the requirement for the presence of the processivity factor for tighter substrate binding suggests several possibilities. It could indicate that stabilization of the binary complex by the clamp acts as a signal to the replication machinery that it is ready for synthesis. Formation of the weak Pol-DNA binary complex might thus provide a means to regulate both the activities of these enzymes and their access to the fork. Another advantage to a weak DNA binding affinity maybe its importance in ensuring fast and high-fidelity DNA synthesis as is seen for *E.coli* Pol III, where a mutant (*dnaE173*) with tighter $K_{d, DNA}$ compromises both speed and accuracy of replication [195].

To our knowledge PolC is the only replicative DNA polymerase that has been reported to display slow PPI release for its cognate dNTP. Typically, pyrophosphate release is expected to occur rapidly, thereby pulling the reaction equilibrium towards product formation. However, some instances of DNA polymerases displaying slow PPI release have been reported before. Mitochondrial Polγ is known to show slow PPI release while adding non-cognate dNTPs such as a modified nucleotide (8-oxo-dGTP) or a nucleotide analog (AZT-TP) [199, 200]. For Polδ, a X-family repair polymerase, it has been demonstrated that pyrophosphorolysis enhances its fidelity such as in case of dATP incorporation opposite a templating 8-oxo-dGMP damage, implying that this polymerase can easily reverse bond formation at least in specific situations [215]. In contrast, time-resolved crystallography of Polβ, an X-family member, has hinted at slow PPI release occurring only when in the presence of the correct nucleotide [31]. It is intriguing that so far reports of slow PPI release occurring with a cognate dNTP are limited to C- and X-family members, both of which belong to the β-nt superfamily. It remains to be seen whether the unique fold of the palm sub-domain of β-nt polymerases plays a role in encouraging this phenomenon.
Two unusual events were identified on performing the detailed kinetic analysis of PolC: (i) bond formation is reversible and (ii) a post-chemistry step, which is not DNA dissociation, is slower than the reverse rate of chemistry and results in an internal equilibrium ([207], this study). These observations remained unchanged for PolC-Exo\textsuperscript{mut} even when its processivity factor was present. Experimentally measured rates for pyrophosphate release (10 to 40 s\textsuperscript{-1}) corroborated that the β-clamp does not influence PPI release, suggesting that the replication machinery is not involved in altering this particular step of PolC’s kinetic pathway. However, this result left a key question unanswered. If PPI release after each nucleotide addition is slow (~40 s\textsuperscript{-1}), then how does processive DNA synthesis in bacteria achieve rates that are 20 to 100 fold faster (~ 500 to 1000 s\textsuperscript{-1})?

Here, we experimentally show that our preparation of PolC readily performs processive DNA synthesis in the presence of the clamp and all four nucleotides at rates that are in agreement with earlier reports on the speed of bacterial replication fork progress. Our measurement of pyrophosphate release in the presence of incoming and next correct nucleotide further confirmed that PolC achieves PPI release at rates comparable to chemistry in the presence of β clamp. We believe that the role of the processivity factor in this case is limited to tethering the polymerase to DNA; ensuring better binding and reduced fall off.

Both the above-mentioned experiments share the presence of the next correct incoming nucleotide as a common denominator. Our results from the global fit of data obtained for the simplest scenario of processive synthesis, where only the first two incoming nucleotides are present, clearly indicates different rates of PPI release after each nucleotide addition (Figure 5.6). We believe that the fast (~800 s\textsuperscript{-1}) rate of pyrophosphate release after the first nucleotide incorporation is an outcome of the next correct nucleotide being present. Consequently, the second nucleotide incorporation is followed by slow (~50 s\textsuperscript{-1}) pyrophosphate release because of the absence of the next correct dNTP during the second incorporation step. Our observations indicate that the identity of the nucleotide
present in addition to the incoming nucleotide acts as an important determinant for pyrophosphate release. Although reported for RNA polymerase II [216, 217], to our knowledge this is the first time that such an observation has been made for a DNA-dependent DNA polymerase.

Our results lead to the intriguing question of the functional relevance and mechanism of this phenomenon. On considering the advantage gained by the ability to sense the presence of the next correct nucleotide several possibilities present themselves. The most obvious one being that PPI release acts as a checkpoint that regulates the kinetic pathway to ensure that fast and accurate synthesis occurs only when the next cognate dNTP is around. Moreover, the modulation of PPI release rate by the nucleotide substrate could act as a sensor to control replication in response to different cellular environments such as in case of a dNTP pool imbalance, when replication by PolC would slow down due to a scarcity of the next correct dNTP thus preventing misinsertions. Since the next correct dNTP is required to hasten PPI release, it is tempting to hypothesize that the mechanism of sensing the next incoming base involves base pairing to the +1 templating base. In such a situation the β-γ phosphate of this next incoming nucleotide could compete for binding with the PPI generated from the current round of catalysis, thereby hastening its release. It must be remembered that the active site of replicative polymerases are constricted and enclosed in the ternary complex conformation and would require rearrangement at the active site, maybe even translocation of the base-pair just formed, to accommodate the next incoming pair. However, further experiments need to be performed to determine the nature and timing of the sampling step for the next correct dNTP.

Our observations with the processivity factor do not rule out the possibility that other replication components influence PolC kinetics. Indeed, it will be of interest to determine the polymerase kinetic pathway while as a part of a reconstituted replisome.

In conclusion, the results presented here provide evidence for a novel
mechanism of regulation that has not been considered for DNA polymerases. It is possible that the ability to sense the next correct nucleotide is common to other C-family members and it remains to be seen if this also holds true for the X-family repair polymerases that are already known to display slow PPi release. The presence of a kinetic mechanism that is common to unrelated processes like bacterial replication and transcription, suggests a shared mode of regulation at the “nucleotidyl transfer” step that has been overlooked so far. Future studies directed at ascertaining the mechanism of next cognate nucleotide recognition including structural and biochemical approaches will help provide better insight into this intriguing problem.

5.7. ACKNOWLEDGEMENTS

We thank Ken Johnson for advice on the kinetic simulations and An Li for help with the pyrophosphate release assays.
5.8. SUPPLEMENTARY INFORMATION

5.8.1. Choice of DNA substrate:

Owing to the longer stretch of duplex DNA required when the polymerase is tethered to the clamp, DNA substrates S1 (18bp) and S2 (30bp) were used for experiments with PolC alone and with clamp respectively (Refs). Substrate S1 was identical to that used for PolΔNΔExo (Lahiri et al.) and thus allows direct comparison between the kinetics of the truncated and full-length enzymes. The long ssDNA overhang of S2 (30-mer) was designed to test if its presence would improve DNA binding through interactions with the OB-domain as predicted previously (Evans et al.). However, on both S1 and S2 in the absence of the clamp, PolC displayed similar kinetics, including dissociation rate of the binary complex, processive synthesis and K_d^dNTP (data not shown). Therefore, we conclude that the changes in kinetic parameters observed with addition of the β clamp are specific to its presence and not an artifact of altered DNA length. It must be noted that since synthetic linear DNA instead of circular DNA was used in this study the observed effects may well be an under-prediction of the influence that the clamp has on polymerase activity. However, the successful observation of processive synthesis when the clamp is present convincingly indicates that the effects observed are indeed specific.

5.8.2. PolC stall products:

Despite the fast rate of processive synthesis, we noted a number of intermediate products that might indicate the distributive nature of PolC and would be consistent with the high k_off observed for the binary complex without the clamp (Figure 5.5B). However, it was surprising that the intermediate products were present even with the clamp (Figure 5.5A), since the binary complex in this case is relatively stable. To verify that PolC-ExoMut does fall off midway through synthesis we repeated the above experiment in the presence of a DNA trap (Figure 5.51). This ensured that enzyme dissociating from the labeled DNA substrate at any step during multiple nucleotide extension would
instead bind the unlabeled trap. The effectiveness of the trap was verified using appropriate control experiments (data not shown). In the presence of the trap, PolC-Exo$^{\text{Mut}}$ (Figure 5.S1B) could not extend very far (~5 to 10 ntds), confirming the distributive nature of the enzyme. However, in the presence of the clamp (Figure 5.S1A), the pattern of product formation remained unchanged irrespective of the trap, indicating that PolC-Exo$^{\text{Mut}}$ does indeed become processive. The continued presence of the intermediate products is puzzling and suggests that PolC-Exo$^{\text{Mut}}$ stalls on the substrate. However, further experiments need to be performed to confirm their source.
Figure 5.S1. Processive synthesis by PolC in the presence of trap DNA.

Multiple nucleotide extension assays were performed in the presence of all four dNTPs both in the (A) presence and (B) absence of the β clamp. The arrow in (A) indicates the full-length product obtained by addition of 38 nucleotides. In (B) the number of dNTPs inserted is written along side the corresponding extended primer band. Adding nineteen dNTPs forms the full-length product. Concentrations of the reactants are the same as in Figure 5.5A and B respectively with the exception that for both Figure 5.S1A and B, a final concentration of 48 µM trap DNA (S1 or S2 without the FAM label) was added along with the dNTP in order to prevent rebinding of the polymerase to the labeled DNA.
CHAPTER 6
SUMMARY AND FUTURE DIRECTIONS

In this final chapter, I aim to summarize the findings reported in the previous chapters and discuss both the implications of the results and the various possibilities that open up with this information.

6.1. Y-family TLS polymerase

Chapters 2 and 3 of this thesis investigate two important aspects of DinB polymerases: (i) mechanisms for generating -1 deletions and (ii) elements of polymerase architecture that determine specificity.

6.2. Human polymerase \( \kappa \) performs template-slippage but realigns

In Chapter 2 we found that Streisinger's "template slippage" (Figure 2.1) is the principal mechanism used by human Pol \( \kappa \) for generating single base deletions in a repetitive sequence context. Single base deletions are one of the most frequent errors generated by DinB polymerases [103]. Our study brings together biochemical and structural observations for DinB homologs, from all three domains of life [78, 84, 103, 105, 107, 112, 113, 119], to present the unified view that, although diversified through evolution, these enzymes have retained a common mechanism for generating -1 deletions. It must be noted that single base deletions by Pol \( \kappa \) are generated not only on undamaged repetitive sequences but also when the enzyme encounters DNA damages such as abasic sites, thymine-thymine (T-T) dimers, 3-methyl cytosines (3-meC) and other lesions [129, 133, 218]. Template slippage is generally relevant in homopolymeric runs while in a non-repetitive sequence context Pol \( \kappa \) has been shown to use misinsertion-misalignment (Figure 2.1) for creating -1 deletions
Interestingly, we also observed that Pol κ realigns its template after the first round of nucleotide incorporation to be able to add a second nucleotide. This ability to realign on homopolymeric runs has been observed before for archaeal DinB homologues, Dbh [119] and Dpo4 [84, 128] and thus, appears to be a more universal ability for DinB enzymes than appreciated before.

The ability to realign the duplex poses several intriguing questions. What is the functional relevance of realignment? Is this a sequence specific observation: G-C vs A-T rich sequences? Is the realignment an active or passive process? Will realignment occur even when PCNA and/or other components of the replisome machinery are present? Below I have addressed each of these concerns.

**6.2.1. Functional relevance:**

The obvious advantage of realigning the substrate immediately after a deletion is that it would help reduce the number of deletions that are generated when the polymerase gains access to DNA, whether damaged or undamaged. Cells are thought to lack dedicated repair systems that can selectively correct deletions, although the ubiquitous presence of mismatch repair mechanisms [219] indicates that most organisms are better equipped at dealing with base-substitution errors.

The ability to realign is possibly one of several mechanisms that influence the frequency of deletions. In *E.coli* strains overexpressing a DinB variant (Y79L) that is inefficient at extending slipped DNA intermediates formed on a repetitive sequence, frameshift frequency is considerably reduced [107]. In such a strain, the proofreading subunit epsilon (ε) of replicative Pol III (encoded by dnaQ) is thought to initiate realignment, thereby reducing deletion frequency. It is still unclear if this function is performed by ε alone or when associated with Pol III. In a separate study, DinB’s propensity for -1 frameshifts on a homopolymeric run is seen to decrease when it forms a complex with monomeric RecA and a dimer.
of uncleaved UmuD (UmuD$^\text{2}$) [105, 167]. The authors suggest that these associated proteins help constrict the “open” active site of DinB [167] and, thus, prevent tolerance of any bulge in the template strand. It remains to be seen if such a complex further interacts with $\varepsilon$. These recent studies of $E.\text{coli}$ DinB provide evidence for mechanisms of realignment that involve other protein partners and suggest the existence of a little understood and complex mode of regulation.

Realignment of the primer-template junction is also known to occur when a polymerase encounters a DNA lesion and has been observed for several repair Pols including B- ($E.\text{coli}$ Pol II), X- (Pol $\mu$) and other Y-family (Pol $\eta$) members [220-222]. Taken together, these studies and the results presented in this work suggest that DNA polymerases may have a specific mechanism for realignment both before incorporation i.e. when generating the deletion and after (as seen here for Pol $\kappa$). It is particularly interesting for the DinB enzymes since the extrahelical base that may need to be reincorporated into the duplex could be up to 3 or 4 ntds upstream [78, 84, 105, 107]. This leads to the question of the polymerase’s involvement in the realignment process and the possibility of substrate rearrangement occurring actively or passively.

6.2.2. Active vs passive realignment strategies:

“DNA breathing” or “fraying” refers to spontaneous, local conformational changes that happen due to thermal fluctuations occurring at temperatures below the temperature of melting ($T_m$) and cause transient unpairing of existing base-pairs in duplex DNA, resulting in their becoming exposed to solvent [223, 224]. Both ensemble and single-molecule measurements place such changes in the 20-100$\mu$s time range [223, 225]. When base pairs form again after a DNA breathing event, error in how the base-pairing takes place could result in a duplex with bases in an extrahelical conformation. Here I refer to such a rearrangement as “passive realignment”. However, if the DNA rearrangement involves the participation of DNA binding proteins, it is termed “active realignment”. For DNA polymerases, it would mean that the polymerase is
responsible for \textit{initiating the unpairing} and subsequent \textit{re-pairing} of the bases to smooth out the extruded base.

For both DinB and Dbh, active realignment before nucleotide insertion has been observed on repetitive sequences \cite{103, 105}. On binding DNA that has a homopolymeric run, the polymerase initiates slippage and stabilizes one of the many thermodynamically possible conformations of the DNA (i.e. intermediates with an extrahelical nucleotide in different positions). It must be noted that the presence of the appropriate dNTP is required to shift the equilibrium in favor of the slipped intermediate. For DinB, stopped-flow fluorescence data suggests that the interaction with RecA and UmuD influences this particular pre-chemistry step of realignment, thereby reducing -1 frameshifts \cite{105}.

Realignment of a substrate post-chemistry could occur by a similar polymerase-dependent pathway and is possibly also influenced by the identity of the nucleotide present. Unlike the GC-rich DNA sequence used in chapter 2 and other studies of DinB enzymes \cite{78, 84, 107, 119}, the fluorescence-based assays mentioned above used an AT rich homopolymer \cite{103, 105}. This confirms that template slippage only requires a repetitive sequence context and occurs independent of the identity of the sequence itself. It remains to be seen if the post chemistry realignment observed in chapter 3 also occurs on an AT-rich substrate. However, this can be expected to be highly likely since AT stretches can easily melt and/or distort and, thus, should have a higher propensity to misalign thereby increasing the chances of frame shifts.

For all the DinB enzymes studied to date the preferred position for tolerating an extrahelical nucleotide is either 3 or 4 bases upstream from the primer/template junction \cite{78, 84, 107}. DNA unwinding up to four nucleotides has been reported for replicative polymerases when the transition between polymerization and proofreading modes occur and the primer terminus moves from the polymerase active site to that of the exonuclease domain \cite{47, 226}. In the stopped-flow experiments mentioned above \cite{103, 105}, the probe used was 2-aminopurine (2AP), an adenine analog whose fluorescence signal is quenched
on base stacking. Since the probe was positioned either at the templating base or 1 base upstream, no direct evidence of the exact position of the unpaired upstream base could be obtained from this study. Although ternary crystal structures of Dbh and Dpo4 bound to DNA with a bulged base exist [78, 84], a real-time measurement as the base becomes extruded on DNA polymerase binding is still lacking. Such evidence would provide information about the order of events that lead to pre-chemistry bulge formation.

Several aspects of realignment are intriguing and remain unclear. The rate at which the process occurs is not known. Moreover, it is important to consider how the polymerase compensates for the energetic cost of realignment and overcomes the barrier encountered when breaking several hydrogen bonds. The energy required to unpair three or four G-C base pairs is considerably more than that of a similar number of A-T pairs. It is clear from DNA breathing studies that under physiological conditions GC-rich regions are not prone to unpairing easily even at high temperatures, thus making it a particularly daunting task [225]. However, it must be noted that cellular DNA remains in close contact with associated proteins and the effect of such complexes on DNA breathing has remained unaccounted.

Recent development of single-molecule Förster resonance energy transfer (smFRET) and single-molecule fluorescence linear dichroism (smFLD) allows the design of experiments that could provide answers to some of the gaps in our knowledge [223, 227, 228]. Using these techniques and by placing rigid fluorescent probes “internally” into DNA backbones at positions near a synthetically designed fork junction, Phelps et.al could detect “DNA breathing” events at the forks that occur at micro-millisecond timescales [227]. Additionally, in the presence of the T4 DNA helicase they were able to observe an increase in breathing upon protein binding followed by larger changes in unwound duplex DNA when the helicase function was activated. Such methods will be interesting to apply to the study of DNA Pols and their propensity to realign.
The dual ability to detect passive and active realignment events owing to both sensitivity and time range for detection makes this assay an advantageous choice. Using this system it will be possible to detect the sequence of events at the primer-template junction/replication fork from before the polymerase binds, to when the binary complex is formed and then finally the ternary complex formation occurs in the presence of individual nucleotides (correct and incorrect). By placing probes at different positions (even 3-4 bases) upstream of the fork, it will be possible to detect specific fluctuations in the phosphate backbone. For the same assay system, it will be possible to vary the sequence context between AT- and GC-rich ones. Finally, it will become possible to better understand the contribution of passive realignment to the results obtained through ensemble kinetics.

6.2.3. Effect of the cellular machinery:

In light of evidence suggesting the “polymerase switching” model it becomes important to evaluate the effect that other components of the replisome have on TLS Pols and their mutation spectra. Even for the “gap-filling” model the role of the sliding clamp remains vital [64]. Thus, it would be interesting to experimentally determine if full-length Pol κ generates single base deletions at frequencies comparable to that when associated with PCNA and how the realignment ability of Pol κ is affected by such an interaction. Full-length Pol κ can be expected to display deletion efficiencies similar to the construct used in our study based on \textit{in vitro} gap-filling DNA synthesis assays [118], however this may change when the clamp is present. For Dbh association with PCNA is seen to improve correct over incorrect nucleotide addition thereby affecting deletion efficiency (Wu and Pata JD, unpublished data). It will be important to test similar conditions for Pol κ. Another interesting possibility to explore is to assess the effect of Rev1 on the mutational specificities of eukaryotic TLS Pols. The physical interaction that is expected because of the Rev1 binding motifs present in the eukaryotic TLS polymerases may involve more than a mere fork-recruitment and targeting function and may be important for the polymerase specificity (including DNA binding and lesion bypass).
6.2.3. Post-realignment extension and handoff

In chapter 2 we found that post-realignment the next dNTP addition by Pol $\kappa$ was performed while extending from a mispair at the primer/template junction. Human Pol $\kappa$ has been dubbed a “mispair extender” before [127, 229] and is thought to compete with DNA polymerase zeta (Pol $\zeta$), a B-family eukaryotic TLS member [127, 135, 152]. However, this characteristic of Pol $\kappa$ is unlike other Y-family Pols such as $\eta$ and $\iota$, which generate and extend mispairs at comparable frequencies [129, 154, 230]. A preference for mispair extension suggests Pol $\kappa$’s involvement in taking over from replicative polymerases stalled at sites of unresolved mispairs.

How TLS Pols are recruited to replication forks is still an unanswered question. It has been suggested that prokaryotes depend on mass-action i.e. higher numbers of TLS Pol molecules during SOS to displace the lower concentrations of replicative Pols [12, 57, 68]. The situation however, is more complex for eukaryotes where post-translational modifications and ubiquitination targets these polymerases to bind PCNA [66, 231]. There is also evidence that Rev1 mediates such interactions [64, 232, 233]. Pol $\zeta$ has been recently shown to form a more active complex known as $\zeta_4$ that is now able to interact with PCNA since it complexes with specific subunits of the lagging strand replicative polymerase, Pol $\delta$ [234, 235]. Taken together these findings and the observation that *E.coli* TLS Pols interact with RecA at the fork [69] all suggest that recruitment to the fork may have multiple signals including association with specific protein partners before this happens.

The large number of structures that exist for Y-family polymerases has made it possible to visualize how individual enzymes bypass particular lesions (most of them Dpo4 structures). However, a large gap in the field exists in understanding the structural and functional influence of several *in vivo* protein partners of these enzymes. Most crystallographic studies till date are with protein fragments, often particular interacting domains, and do not provide a
comprehensive view of these associations [232]. Crystallographic constructs for
the eukaryotic Pols have lacked the extensive C-terminal extensions [30, 77, 80,
82, 108] regions that are expected to be disordered, except when interacting
with a cognate cellular partner. Thus, there is need for obtaining high-purity
stable complexes between the full-length polymerase and its protein partner
that may include sliding clamp, Rev1 (eukaryotes only). Structure determination
of such complexes can be pursued through a combination of alternative
methods, such as small-angle X-ray scattering (SAXS) and nuclear magnetic
resonance (NMR). The overall envelope of the entire complex can be obtained
from SAXS data and detailed information that can be revealed through NMR of
smaller domains can be fit to this envelope using previously known constraints
to generate a model. Such structure-function correlations will be important for
understanding how TLS Pols interact with their protein partners, bypass
damage, generate mutations and also undergo switching.

6.3. Linker dependent conformation of archaeal DinB polymerases

In chapter 3 of this thesis, I have described structural and biochemical
experiments that provide evidence for the importance of three amino acid
residues present in the interdomain linker of Dbh (Lys-Ile-Pro) and Dpo4 (Arg-
Lys-Ser) that determine the ability of the polymerase to generate single base
deletions. Swapping these 3 residues between Dbh and Dpo4 creates chimeras
that have characteristics determined by the new amino acids. We determined
the structure of DbhDpo4\textsuperscript{RKS}Dbh (at 2.4Å), the chimeric polymerase that is
mostly Dbh except for three residues in its linker region and found that this
minimal change alters the conformation of Dbh, positioning its LF/PAD domain in
a Dpo4-like orientation in close contact with the fingers domain. Combining the
results from this study and earlier work from our lab [128], we are able to
conclude that these amino acids in the linker collectively influence abasic site
bypass, single-base deletion frequency, base-substitution frequency and
mispair extension ability of the polymerase.
The β2-3 loop is a generally disordered part of the fingers domain in Dbh structures [78]. In DbhDpo4\(^{\text{RKS}}\)Dbh the β2-3 loop is well-resolved and seen to make contacts with the Dbh LF/PAD in the chimera. This is possible because of the Dpo4-like conformation of this particular chimera. In this altered conformation, the Dbh active site resembles Dpo4 being both constrained and well formed, with the incoming nucleotide and primer terminus positioned correctly. Therefore, this loop region >11Å away from the active site of the polymerase is clearly important for the specificity of these DinB enzymes.

The observation that the Dbh LF/PAD is inherently capable of adopting the position as seen in Dpo4 leads to the possibility that during chemistry Dbh transiently adopts the Dpo4 ternary complex conformation, thus aligning the substrates correctly at the active site. The short-lived nature of the well-formed active site may be key to the high deletion frequency observed for Dbh and future experiments directed at capturing this transient state would help understand this event better.

We determined the $K_d^{\text{dNTP}}$ and $k_{\text{pol}}$ for both the correct and incorrect nucleotides (dGTP and dCTP respectively) in a hotspot sequence containing template, to understand the influence of the conformational change in the chimera and identify its effect on the kinetic parameters representing nucleotide binding and incorporation. Our findings show that the Dpo4 linker can selectively improve correct over incorrect nucleotide incorporation by specifically altering the $k_{\text{pol}}$ for correct dNTP insertion and this result is well explained by the constrained and well-formed active site in the Dpo4-like chimera. There is no in vivo data suggesting that Dbh is more error-prone than Dpo4, since \textit{Sulfolobus} is not very tractable genetically. Therefore, it is possible that the transient Dpo4-like state (conformation and kinetic parameters) that Dbh needs to achieve during catalysis is stabilized through interaction with other protein partners in the cell. Indeed, there is some evidence that in the presence of PCNA, Dbh incorporates the correct over the incorrect nucleotide more efficiently (Wu Y and Pata JD, unpublished data). DinB has also been reported to
bind nucleotide better in the presence of the sliding clamp [187] although this study did not report on the effect of this interaction on DinB’s fidelity.

Pol I, a eukaryotic error-prone polymerase is the only other Y-family member for which a clear correlation between its structural architecture and its distinct incorporation specificity is known [58, 64, 80, 81]. Crystal structures show that the short β2-3 loop of Pol I results in the LF/PAD domain to be tilted and positioned close to the fingers such that the polymerase active site is constricted and unable to accommodate standard Watson-Crick base-pairs in B-form DNA (~10.5 Å), resulting in Hoogsteen base pairs (~9 Å) being preferred [80, 81]. Similar to the Pol I study our results show that the LF/PAD conformation is important for incorporation specificity. Additionally, we find that the interdomain linker is crucial for achieving this conformation. Interdomain linkers are present in all Y-family polymerases and it remains to be seen if the linker region is found to be a common determinant for mutagenic and lesion bypass signature for other sub-families. However, it is still formally possible that Y-family polymerases do not share such a common determinant and each sub-family merely retains its own characteristic structural feature for this purpose.

The amino acid modifications made in our study occur in a region of the enzyme that is neither structurally nor functionally important for polymerization activity itself [79], but result in the information being conveyed to the active site through the altered positioning of the LF/PAD and cause the proper alignment of the substrates (DNA and nucleotide) (Figure 3.4). Reduced polymerase activity observed in a construct lacking the linker and LF/PAD regions of Dbh provides further evidence for the functional importance of these regions of the enzyme [79]. Our observation that very few amino acid changes could result in an altered phenotype has an important implication for the evolution of these enzymes. The phenotypic change that resulted from these minimal changes, as is seen for these two Sulfolobus strains, was selected for possibly because it represented a survival advantage for the organism.
Taken together our results provide a better understanding of what determines the fidelity and characteristics of DinB enzymes and lead the way to future studies that correlate this information with that for other error-prone polymerases.

6.4. C-family replicative polymerases

Chapters 4 and 5 of this thesis focus on the kinetic mechanism followed by the bacterial replicative polymerase from the Gram-positive pathogen *S. aureus*. Using pre-steady state and steady state conditions, we were able to define the key kinetic parameters of a simplified polymerization pathway for PolCΔExo and PolC-Exo\textsuperscript{mut}. We also determined the effect of the sliding clamp on PolC-Exo\textsuperscript{mut}. The parameters determined include DNA binding affinity (K\textsubscript{D}DNA), nucleotide binding affinity (K\textsubscript{D}dNTP), maximum rate of nucleotide incorporation (k\textsubscript{pol}), DNA off-rate from binary (k\textsubscript{off}) and steady state rate (k\textsubscript{cat}), a parameter that is dominated by the slowest step in the pathway. Using these experimentally determined rates and binding constants, we simulated the polymerization cycle as a simplified 3-step process to obtain the forward and reverse rates of several important steps. Below I have discussed both the significance of our findings and the direction in which future studies may be directed.

6.4.1. PolCΔExo displays slow PPi release

On completion of the studies with PolC detailed in chapter 4, we became aware of the unusual kinetic pathway that is followed by this replicative polymerase when incorporating the correct nucleotide. Both the weak DNA binding (K\textsubscript{D}DNA ≈ 400 nM) and fast binary dissociation rate (k\textsubscript{off} ≈ 150 s\textsuperscript{-1}) are uncharacteristic for most replicative polymerases. However, the most unusual feature was the clear dependence observed of product amplitude on [dNTP] and the absence of the same in the rates of product formation (k\textsubscript{i}) (Figure 4.8). Furthermore, k\textsubscript{cat} appeared to be much slower (~30 s\textsuperscript{-1}) than k\textsubscript{off}. This was in contrast to other replicative Pols where the fall-off from DNA dominates the rate of the slowest
step when multiple-turnover is allowed [209]. From the simulation we could predict that our observations were best explained if the step after chemistry i.e. pyrophosphate release or a conformational change associated with it is slow and comparable to the $k_{cat}$.

As discussed already (chapter 4), this was the first time such an observation had been made for a replicative polymerase when incorporating its cognate dNTP. Several major points of discussion arose from this study. Is slow PPI release an aberrant feature observed only because of the deletion construct being used? If not, then how does PolC overcome this crucial kinetic step and undergo fast processive synthesis to achieve rates of $\sim 500$-1000nt/s? Finally, does the PPI release rate become faster when other components of the replication machinery (such as sliding clamp and helicase) are present? Also, is this a characteristic common to the kinetics of all C-family members? Some of these questions were then addressed in chapter 5.

**6.4.2. PPI release is no longer slow when next correct nucleotide is present**

In chapter 5 we characterized full-length PolC with an inactivated exonuclease domain (PolC-Exo$^{mut}$), both in the presence and absence of the $\beta$ clamp processivity factor, to find that the kinetic pathway in each case resembles that for the deletion construct with some minor variations. $k_{off}$ appeared to be the most affected parameter going from the deletion construct (150 s$^{-1}$) to full-length PolC (7 s$^{-1}$) to PolC associated with $\beta$ clamp (0.7 s$^{-1}$). For each of these changes there was an order of magnitude reduction in the off-rate. When compared to the influence of the processivity factor on replicative Pol $\delta$ and mitochondrial Pol $\gamma$ (chapter 5), this improvement in stability of the binary complex for PolC-Exo$^{mut}$ with the $\beta$-clamp is to be expected [192, 209]. We then experimentally measured the PPI release rate for PolC-Exo$^{mut}$ (in presence and absence of clamp) and confirmed that the slow rates calculated earlier from the simulation (chapter 4) were indeed true. Although, these observations
answered the concerns regarding the construct and the influence of other replication components, the slow rate of pyrophosphate release remained unexplained.

Slow pyrophosphate release by a replicative DNA polymerase will interfere with processive synthesis, a term that refers to polymerization of long stretches (often several kilobases) of DNA by the polymerase before it dissociates from the substrate [24]. In single nucleotide incorporation assays done in vitro, DNA dissociation remains the rate-limiting step for most polymerases. However, in vivo this scenario is altered by the need for processive synthesis, which prevents DNA dissociation with the result that a pre-chemistry step in the polymerization cycle becomes rate limiting. As the slowest step in PolC’s kinetic pathway, post-chemistry PPi release cannot be avoided in vivo, thus posing a problem for fast DNA synthesis.

We used an in vitro setup to test the rate at which PolC could perform processive synthesis. Our results indicate that processive synthesis by PolC is not limited by its slow PPi release, and occurs at speeds comparable to both its \(k_{pol}\) (chapters 4 and 5) and reported rates for replication fork progress [71-73]. This result was true even with just two or three of all four nucleotides provided, suggesting that presence of the next correct dNTP is enough to alter PolC’s kinetics.

We believe that the presence of the cognate dNTP that will be next incorporated is a requirement and cannot be substituted by any other nucleotide. This is because of two reasons. First, measuring the PPi release rates during the incorporation of the first two nucleotides showed two distinct pyrophosphate release rates; fast for the 1\(^{st}\) nucleotide incorporation and slow for the 2\(^{nd}\). Second, in our experimental setup for the single nucleotide incorporations (such as \(K_{d}^{dNTP}\) measurements) where only the first incoming nucleotide (dTTP) was present, [dTTP] provided was 20-100 fold above \(K_{d}^{dNTP}\) (~1-5µM). Under these conditions, the concentration of dTTP present was in enough excess to act as the incorrect nucleotide for the second incorporation.
(ideally, dATP). Yet PPI release observed was slow, suggesting that the next incorrect nucleotide is insufficient to cause fast release of pyrophosphate. Further verification for this evidence might be obtained by varying the 1st nucleotide concentration (as for determining the $K_D^{dNTP}$) in the presence of a constant concentration of the 2nd correct nucleotide (dATP for substrates used in chapters 4 and 5). We would expect the amplitude dependence on [dTTP] concentration to be removed, implying faster PPI release such that no equilibrium is setup to drive the bond formation step backward.

6.4.3. Functional relevance of “+1 dNTP-dependent PPI release”

It is important to consider the mechanistic relevance of the phenomenon observed by us. Recognizing the presence of the next correct nucleotide and utilizing it to modulate the rates of essential steps in the kinetic pathway could play a role in improving DNA polymerase fidelity. Moreover, this mechanism could be crucial for sensing environmental signals, such as dNTP pool imbalances that would then slow down the speed of the replication fork. Such a situation could explain why the presence of the clamp did not alter the PPI release rate. However, this does not eliminate the possibility of the influence of other proteins at the replication fork. It also remains to be seen if the slow release of pyrophosphate triggers a switch from the polymerase to the exonuclease domain thus buying time for the next correct nucleotide to be present.

A particular aspect that remains vital to identify is the component - base, sugar, or phosphates- of the next correct nucleoside triphosphate that is required for inducing fast pyrophosphate release and if this occurs due to a competitive or non-competitive activation mechanism.

Although it is possible to suggest several scenarios where the presence of the next correct dNTP may prove to be mechanistically important, it is difficult to envision how such an interaction could be accommodated by the polymerase. A particularly tempting possibility is that the binding of the incoming dNTP and
the next correct nucleotide occur allosterically to induce a conformational change in the polymerase that helps trigger fast release of the pyrophosphate. The *G. kaustophilus* PolCΔNΔExo ternary structure shows the +1 templating base to be swung outwards from the active site and wedged between the fingers and N-palm domains [39]. Hypothetically, in this position the +1 base might be able to recognize the next correct nucleotide by Watson-Crick pairing between the two, which could cause a conformational change desired for faster PPi release. Correct Watson-Crick base pairing would however require additional constraints that ensure the correct orientation of the bases themselves. There also needs to be discrimination against incorrect nucleotides. Sampling for the next correct nucleotide could happen in parallel with the polymerase proceeding towards chemistry, however the post-chemistry PPi release would only occur if the next correct nucleotide induces the right fit/conformational change.

It is still unclear how DNA polymerases select for the correct nucleotide out of 4 dNTP and 4 rNTP choices. Recent smFRET experiments with the Klenow fragment of Pol I suggest that the polymerase selects for the correct nucleotide amongst other incorrect options while in the “open” state and does not undergo the conformational change to the “closed” state till the cognate dNTP is already chosen [236]. It remains to be seen if this holds true for other DNA polymerases. For PolC this could mean a more complex situation where sampling of the correct nucleotide for two consecutive template positions (0 and +1) occurs at the same time. To better answer questions about nucleotide selection it becomes important to visualize the players at the active site during polymerization. Although the structure of a PolC ternary complex exists, crystallization of a stable binary complex is difficult for PolC (and other C-family members) owing to the weak DNA binding. Moreover in the one structure of Taq Pol III-DNA binary the active site appears disordered [56]. Thus, alternative methods may be needed to initially study the dynamics of these enzymes.
6.4.4. Implications for C-family members

Before our studies with *S. aureus* PolC, detailed kinetic characterization of the polymerization pathway of a C-family DNA polymerase did not exist. Thus, it will be interesting to test if the unusual characteristics shown by PolC hold true for the other C-family members since not all of them are replicative DNA polymerases. DnaE, the other C-family enzyme in low G-C Gram-positive bacteria, is known to be an error-prone polymerase in *S. aureus* [237]. In *Bacillus subtilis*, another Gram positive, DnaE is thought to extend from RNA primers before handoff to PolC [206], similar to eukaryotic Pol α. Our initial studies (Lahiri I, Mukherjee P, Kosoy G and Pata JD, unpublished data) with *S. aureus* DnaE show that on a DNA substrate there is no dependence of the amplitude of product formation on [dNTP] concentration, as would be expected if PPi release were fast. However, we find that on a hybrid substrate with a RNA primer and DNA template, the amplitude dependence returns, signifying that PPi release is now slow. Based on our preliminary data we believe that C-family polymerases may distinguish between cognate and incorrect substrates using this mechanism.

6.5. Concluding remarks

The results reported here greatly enhance our understanding of the mechanistic, structural and functional aspects of two of the six known DNA polymerase families. For the Y-family enzymes, I have identified template slippage as the mechanism of generating single-base deletions on hotspot sequences in human Pol κ and determined the interdomain linker as a major contributor to lesion bypass specificity for archaeal Dbh and Dpo4. In conjunction, these studies provide a deeper understanding of the functioning of the large family of DinB-type TLS polymerases. For the replicative C-family polymerase, PolC from *S. aureus* I have identified a novel kinetic mechanism of regulation and my results indicate the importance of nucleotide identity in determining the rate of processive synthesis. Taken together, these studies
highlight the existence of a broad spectrum of regulatory mechanisms adopted by DNA polymerases that aid in the performance of their assigned roles in DNA replication or damage tolerance. The observations reported in this thesis emphasize the gap that still exists in our understanding of the mechanistic, functional and structural workings of these simple molecular machines and prompt for future studies that will address these challenging issues.
BIBLIOGRAPHY


177


195. Yanagihara, F., et al., *The dnaE173 mutator mutation confers on the alpha subunit of Escherichia coli DNA polymerase III a capacity for highly...


