Mechanistic studies of C- and Y-family DNA polymerases

Indrajit Lahiri
*University at Albany, State University of New York*, ilahiri@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](https://scholarsarchive.library.albany.edu/legacy-etd)

Part of the Biochemistry Commons

**Recommended Citation**
[https://scholarsarchive.library.albany.edu/legacy-etd/1170](https://scholarsarchive.library.albany.edu/legacy-etd/1170)

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.
MECHANISTIC STUDIES OF C- AND Y-FAMILY DNA POLYMERASES

by

Indrajit Lahiri

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 express my deepest gratitude to my mentor Dr. Janice Pata. Thank you Janice for giving me the opportunity to work in your lab. Thank you for believing in me and for all your guidance and support when I needed it the most. These last three years has been a dream come true for me.

I am grateful to all my committee members, Drs. Joachim Jaeger, Marlene Belfort, Paul Masters and Richard Cunningham. Thank you for all the insightful comments and invaluable suggestions.

I would like to thank the Pata and Jaeger lab members. Your encouragement helped me keep going even when things were not going my way. It was a pleasure to share all the highs and lows of my PhD work with you all.

During my final years of PhD I received financial support from the Wadsworth Center and the department of Biomedical Sciences at University at Albany. I am grateful for that.

All my friends in Albany, especially Aymen Yassin and Takeshi and Kae Yokoyama, made my stay here a memorable one. I am indeed lucky to have you all as friends. I would like to thank all my friends back in Kolkata, Debdeep Ghosh, Bodisatwa Das, Pooja Gupta Saraf and Anubhab Palit. Thank you all and all and the very best to you.

I thank my family, baba, maa and didi. I am grateful for your unwavering love and encouragement. This PhD would not have been possible without your support. Thank you for never stopping to believe in me even when times were difficult.
Finally, I want to thank my wife and colleague Purba Mukherjee. Thank you for your unconditional love and support. Working with you was an amazing experience and I hope that we get to work together in the future. Thank you for being with me through all the highs and lows and I look forward to sharing our lives together.

Indrajit Lahiri

April 2014
DISSERTATION ABSTRACT

In this dissertation, I have described my work with the Staphylococcus aureus replicative DNA polymerase, PolC and the human translesion synthesis (TLS) polymerase Polk. Although both carry out DNA polymerization, they have dramatically different characteristics. PolC belongs to the C-family of DNA polymerases and catalyzes very fast and accurate DNA synthesis. On the other hand, Polk belongs to the Y-family and has a slow DNA synthesis rate. However, unlike PolC, Polk efficiently replicates certain types of damaged DNA and this very feature makes the polymerase error-prone on undamaged DNA template. In this dissertation I have performed detailed pre-steady-state kinetic characterization of PolC. Notably this is the first detailed kinetic study of a bacterial replicative polymerase. Further, as PolC has been identified as a novel target for bactericidal antibiotics, the kinetic studies will help in developing more effective antibiotics against low G+C content Gram positive bacteria. We found that although PolC follows the same overall enzymatic pathway described for other DNA polymerases, there are some unique features (described in chapters 2 and 3). Most notably bond formation step was reversible and in equilibrium with dNTP binding for single nucleotide incorporation. Further, this equilibrium resulted from the presence of a slow post-chemistry step, presumably pyrophosphate (P Pi) release. Surprisingly, we have found that during processive synthesis, the presence of the next correct incoming deoxynucleotide triphosphate (dNTP) leads to a dramatic increase in the rate of P Pi release thereby leading to fast DNA synthesis. Modulation of P Pi release rate by the presence of the dNTP substrate might act as a novel mechanism to ensure fast and
accurate replication. Polκ is a eukaryotic TLS polymerase and belongs to the DinB subfamily of the Y-family polymerases. This is the only subfamily found in all domains of life. It has been found that DinB polymerases generate single base deletion mutations on repetitive deletion hotspot DNA sequences (a run of pyrimidines followed by a purine). Three mechanisms have been suggested for generating single base deletions, namely, Streisinger slippage, misincorporation misalignment and dNTP stabilized misalignment. Previous reports showed that the bacterial and archaeal DinB polymerases use Streisinger slippage mechanism to generate single base deletions on the hotspot sequence. Here we have found that like its bacterial and archaeal counterparts, Polκ uses Streisinger slippage mechanism to generate single base deletions on the hotspot sequence (chapter 4). Moreover, after generating a single base deletion, in the next round of catalysis Polκ can realign the bulged base to rectify the deletion and in the process generates a less deleterious misincorporation mutation. The implications of these findings will be discussed in the following chapters.
# TABLE OF CONTENTS

**Acknowledgements**  

**Dissertation Abstract**  

**Chapter 1: Introduction**  

1.1. Replication and DNA polymerases  
   1.1.1. DNA polymerase structure  
   1.1.2. Kinetics of DNA polymerases  
1.2. Processivity factor  
1.3. General organization of the replication fork  
1.4 C-family polymerases  
1.5 Y-family polymerases  


2.1. Description of contribution  
2.2. Summary  
2.3. Introduction  
2.4. Materials and methods  
   2.4.1. Materials.  
   2.4.2. Duplex DNA formation.  
   2.4.3. Expression and purification of *Sau*-PolC-ΔNΔExo.  
   2.4.4. Assays for optimization of enzymatic activity of *Sau*-PolC-ΔNΔExo.  
   2.4.5. Steady-state assays.  
   2.4.6. Measurement of DNA dissociation rate from binary complex.  
   2.4.7. Active site titration of *Sau*-PolC-ΔNΔExo.  
   2.4.8. $K_D^{dNTP}$ determination.
Chapter 3: Lahiri I, Mukherjee P, Pata JD. Catalysis by Staphylococcus aureus PolC is regulated by the next correct (but not incorrect) nucleotide. (Manuscript in preparation)

3.1. Description of contribution 72
3.2. Summary 73
3.3. Introduction 74
3.4. Methods 78
   3.4.1. Protein purification 78
   3.4.2. DNA annealing 80
   3.4.3. Primer extension assays 80
   3.4.4. Data analysis 81
   3.4.5. Simulation 82
   3.4.6. Measurement of pyrophosphate release rate 83
3.5. Results 84
   3.5.1. N-terminal and exonuclease domains slow DNA release without improving \( K_{D,A}^{DNA} \) 84
   3.5.2. \( \beta \) clamp increases the affinity of the binary complex. 89
3.5.3. Incoming dNTP stabilizes the pre-chemistry ternary complex 90
3.5.4. Nucleotide incorporation is reversible both in the presence and absence of the clamp 91
3.5.5. Processive synthesis by PolC is fast 98
3.5.6. Presence of next correct incoming dNTP allows rapid pyrophosphate release 101

3.6. Discussion 104
3.7. Acknowledgements 112
3.8. Supplementary information 113
  3.8.1. Choice of DNA substrate 113
  3.8.2. PolC stall products 114

Chapter 4: Mukherjee P, Lahiri I, Pata JD. Human polymerase kappa uses a template-slippage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. Nucleic Acids Res. 2013 May; 41(9):5024-35. 116

4.1. Description of contribution 116
4.2. Summary 117
4.3. Introduction 118
4.4. Materials and methods 124
  4.4.1. Protein purification 124
  4.4.2. Duplex DNA formation 125
  4.4.3. Pre-steady-state primer extension assays 127
4.5. Results 130
  4.5.1. The template sequence 3’-CCCCG-5’ is a deletion hotspot for hPolκ. 130
  4.5.2. hPolκ predominantly uses a template slippage deletion mechanism on iterative sequences. 134
4.5.3. hPolk readily misincorporates dC when the deletion hotspot is changed to 3’-CCCTG-5’

4.5.4. hPolk efficiently realigns primer-template DNA after initiating a deletion.

4.5.5. Sequence context influences the rate of nucleotide incorporation.

4.6. Discussion

4.7. Funding

4.8. Acknowledgements

Chapter 5: Summary and future directions

5.1. C-family polymerases
   5.1.1. Results summary
   5.1.2. Significance of the results
   5.1.3. Future directions

5.2. Y-family replicative polymerases
   5.2.1. Results summary
   5.2.2. Future directions

5.3. Concluding remarks

Bibliography


A-1. Description of contribution
A-2. Summary
A-3. Introduction
A-4. Experimental procedures
A-4.1. Protein expression and purification. 200
A-4.2. Primer-template DNA. 200
A-4.3. Deletion assays. 200
A-4.4. $K_D^{dNTP}$ and $k_{pol}$ determination. 201
A-4.5. Crystallization and Structure Determination. 204
A-5. Results 205
   A-5.1. Linker identity determines -1 deletion frequency on a repetitive sequence. 205
   A-5.2. Linker identity determines efficiency of mispair extension. 213
   A-5.3. Amino acid trio in the linker determines overall polymerase conformation. 213
   A-5.4. dNTP selection: linker-dependent alteration of nucleotide incorporation rate. 219
A-6. Discussion 223
A-7. Acknowledgements 226
A-8. References 227
# LIST OF FIGURES

**Figure 1.1.** Classification of DNA polymerase.  

**Figure 1.2.** Domain architecture of DNA polymerases.  

**Figure 1.3.** Minimal kinetic pathway of DNA polymerases.  

**Figure 1.4.** Two metal ion mechanism  

**Figure 1.5.** Organization of the replication fork  

**Figure 1.6.** Structure of Gka-PolC ternary complex  

**Figure 2.1.** Overview of PolC structure and domain organization.  

**Figure 2.2.** Minimal single-nucleotide incorporation reaction pathway for DNA polymerases.  

**Figure 2.3.** SDS-PAGE of *Sau*-PolC-ΔNΔExo.  

**Figure 2.4.** Primer extension assays for optimizing enzymatic activity of *Sau*-PolC-ΔNΔExo.  

**Figure 2.5.** Steady-state kinetic analysis of *Sau*-PolC-ΔNΔExo.  

**Figure 2.6.** Determination of the DNA dissociation rate from polymerase • DNA binary complex (k\text{off}).  

**Figure 2.7.** Pre-steady-state kinetics and active site titration of *Sau*-PolC-ΔNΔExo.  

**Figure 2.8.** Determination of K\text{D}^{\text{dNTP}} of *Sau*-PolC-ΔNΔExo.  

**Figure 2.9.** Simulation of kinetic pathway of *Sau*-PolC-ΔNΔExo.  

**Figure 2.10.** Minimal enzymatic pathway for *Sau*-PolC-ΔNΔExo.  

**Figure 3.1.** Reaction pathway of DNA polymerases.  

**Figure 3.2.** Pre-steady state kinetics of PolC-Exo\textsuperscript{Mut}.
Figure 3.3. Affinity of dNTP for the pre-chemistry binary complex.  
Figure 3.4. Simulation of the reaction pathway of PolC-Exo^{Mut}.  
Figure 3.5. Multiple nucleotide extension by PolC.  
Figure 3.6. Simulation of the processive synthesis reaction pathway for PolC.  
Figure 3.S1. Processive synthesis by PolC in the presence of trap DNA.  
Figure 4.1. Single base deletion mechanisms.  
Figure 4.2: Nucleotide incorporation by hPol^{K} on repetitive sequence containing deletion hotspot.  
Figure 4.3. Mechanism of second dGTP incorporation on 4C-G substrate.  
Figure 4.4. Mechanism of multiple dCTP incorporations.  
Figure 4.5. hPol^{K} misincorporates efficiently on 1T-G substrate.  
Figure 4.6. Structure of an hPol^{K} ternary complex.  
Figure 5. Domain movement of C-family polymerases.  
Figure A-1. Parental Y-family polymerases and chimeric constructs.  
Figure A-2. Single-base deletion efficiency is dependent on the identity of the linker.  
Figure A-3. Chimeric polymerases use a template-slipage deletion mechanism.  
Figure A-4. DbhDpo4^{RKS}Dbh adopts a Dpo4-like conformation.
LIST OF TABLES

**Table 2.1.** Comparison of kinetic parameters of different polymerases. 60

**Table 3.1.** Kinetic parameters obtained with different constructs of PolC. 107

**Table 4.1.** DNA Substrate Sequences. 126

**Table 4.2.** Summary of observed nucleotide incorporation rates. 129

**Table A-1.** DNA substrate sequences. 201

**Table A-2.** Summary of observed nucleotide incorporation rates. 206

**Table A-3.** Data collection and refinement statistics. 218

**Table A-4.** Kinetic parameters for correct and incorrect nucleotide incorporation on a deletion hotspot sequence (substrate 4C-G). 222
Chapter 1: Introduction

In this dissertation I will describe the studies that I have performed with bacterial C-family and human Y-family DNA polymerases. Our studies are the first pre-steady state kinetic analysis of C-family polymerases and helped define the mechanism of action of these enzymes. This study shows that the kinetics of the nucleotide incorporation cycle for PolC is modulated by the presence of the next correct incoming deoxynucleoside triphosphate (dNTP). Kinetic studies of the human Y-family polymerase, Polκ, helped to establish the single-base deletion mechanism used by this polymerase and we further found that Polκ can rearrange a bulged template nucleotide resulting in the generation of a mispair as opposed to the more harmful single base deletion.

In this chapter, I will provide a brief description of the replication machinery focusing in particular on the DNA polymerase and the processivity factor. I will summarize what is known about DNA polymerase structure and enzymatic mechanism. Then I will discuss C-family DNA polymerases, their structures and function. This will be followed by a brief description of the Y-family translesion synthesis polymerases. I will conclude this chapter with a short description of the results that I have obtained. The results will be elaborated upon in the later chapters.
1.1 Replication and DNA polymerases:

Genetic information of all cellular organisms is stored and transferred between generations in the form of DNA. For transferring genetic information from one generation to the next, the DNA needs to be duplicated and the duplicate copy is passed on to the next generation. DNA synthesis takes place semi-conservatively with one of the existing strands of the DNA double helix acting as the template for generation of the new strand. This process of semi-conservative DNA duplication is called replication. It is a highly regulated process and for this purpose most organisms possess a complex machinery called the replisome, which is spatially as well as temporally dynamic [1]. For instance, even a simple bacteriophage like T7 has a replisome consisting of at least four proteins while the bacterial replisome of *B. subtilis* is made up of at least thirteen [2, 3]. The main purpose of this macromolecular machine is to ensure fast and accurate replication such that the genetic information can be transmitted from one generation to the next.

Perhaps the most important catalytic component of this large complex is the nucleotidyl transferase called DNA-dependent DNA polymerase or simply DNA polymerase. This enzyme is responsible for synthesizing the new DNA strand, also called the primer strand, by catalyzing the phosphodiester bond formation between a monomeric dNTP and the 3’ hydroxyl of the nascent primer strand. Incoming dNTP selection is made based on
the identity of the base of templating nucleotide such that the nascent basepair (bp) can adopt a Watson-Crick geometry.

Based on their function we can broadly classify DNA dependent DNA polymerases into three major categories:

1) **Replicative polymerases:** These polymerases are responsible for replication of the bulk of the genome. The replicative DNA polymerases are very fast and generally have very high accuracy [4-8]. However, the very same features of the polymerase that ensure accurate replication make them highly inefficient at replicating damaged DNA template [9-11].

2) **Translesion synthesis (TLS) polymerases:** DNA damage occurs frequently and can arise from both natural as well as man-made sources. Specialized polymerases called TLS polymerases are responsible for replicating past damaged template DNA. Depending on the particular TLS polymerase and the type of the lesion, the damaged DNA can be copied either correctly or in an error-prone manner [12-16], [17]. Due to the nature of their function, TLS polymerases are error-prone on undamaged DNA and they have a low DNA synthesis rate [18-24]. Further, their affinity for DNA is generally much lower compared to replicative polymerases [22, 25, 26]. As a result, their access to the DNA substrate is restricted and highly regulated, a necessary condition for faithful genome duplication.
3) **Repair polymerases:** These polymerases are involved in synthesizing new undamaged DNA in place of a damaged stretch of DNA. Often, repair polymerases and TLS polymerases function in concert along with other DNA damage removing pathways including nucleotide excision repair, base-excision repair etc.[27-32].

Based on sequence similarity DNA polymerases can be classified into six families with the TLS polymerases mostly belonging to the Y-family (Figure 1.1) [33, 34]. The eukaryotic replicative polymerases belong to the B-family while the archaeal counterparts belong to the B- and D-family. On the other hand, bacterial replicative polymerases belong to the C-family [33].
Figure 1.1 Classification of DNA polymerases. Based on sequence similarity DNA dependent DNA polymerases are grouped into six groups while based on the topology of the palm domain these six groups are placed into two super-families. Till date there have been no structural studies on the D-family polymerases. Therefore it is not known whether this family belongs to the classical polymerase or the β nucleotidyl transferase superfamily.
1.1.1 DNA polymerase structure:

Although it has been almost 60 years since the discovery of the first DNA polymerase [35], the atomic structures of these enzymes were solved about three decades ago [36]. To date, all the polymerases show a universal multi-domain architecture of the catalytic subunit, that resembling the topology of a right hand [37, 38]. Accordingly, the polymerase catalytic domain has been divided into thumb, palm and fingers sub-domains (Figure 1.2A and B) with the active site situated at the junction between the fingers and palm. The thumb is primarily involved in contacting the duplex DNA, the fingers recognize and stabilize the cognate dNTP at the active site while the crucial active site residues (discussed below) are located in the palm sub-domain [37]. Often DNA polymerases possess extra domains in addition to the catalytic ones. For instance, T7 DNA polymerase contains a thioredoxin-binding domain (Figure 1.2A) that is involved in improving processivity (a measure of the length of the DNA stretch copied by a DNA polymerase before dissociating from the template strand) [39-41].
Figure 1.2: Domain architecture of DNA polymerases. Structures of polymerase-DNA-dNTP ternary complexes of (A) T7 DNA polymerase (PDB ID: 1T7P) and (B) Dbh (PDB ID: 3BQ1) with the subdomains color coded as follows, palm: magenta; thumb: green; fingers: blue; exonuclease domain: red and LF/PAD: light orange. Primer strand is shown in grey and the template strand in orange. Thioredoxin on T7 DNA polymerase is colored cyan. The catalytic metals are shown as green spheres, while the incoming dNTP and the catalytic aspartates are shown in stick representation. (C and D) Zoomed in views of the active sites (marked by black squares in A and B) of (C) T7 DNA polymerase and (D) Dbh. Color coding is same as in A and B with the amino acids shown in the sphere representation.
Many polymerases, especially the replicative ones, contain an additional 3’ to 5’ exonuclease domain (Figure 1.2A) that is responsible for proofreading the newly synthesized DNA strand in case of an error by the polymerase, thereby increasing the accuracy of DNA synthesis by ~ two orders of magnitude [42]. Another crucial feature for increasing the accuracy is the geometric constraint imposed by the active site of the enzyme. Generally, replicative polymerases have a sterically constrained active site (Figure 1.2C), thus, allowing the addition of only the cognate dNTP, which forms a Watson-Crick base pair with the templating base [41, 43, 44]. On the other hand, Y-family active sites are more unconstrained/accessible and are able to accommodate a variety of lesions, distortions in DNA backbone and alternate base pairing geometry such as Hoogsteen pairs. This characteristic makes TLS enzymes very error-prone [19, 45-47] (Figure 1.2D).

As different DNA polymerase structures were being solved yet another classification of these enzymes arose. The six polymerase families could be grouped into two super-families based on the fold of the palm sub-domain, the classical polymerases and the β nucleotidyl transferases (β-nt) (Figure 1.1) [48]. Due to this different topology, the DNA substrates for these two super-families are oriented differently with respect to the palm, but similar active site geometry is ultimately maintained in both cases.
1.1.2 Kinetics of DNA polymerases:

Although biochemical studies on DNA polymerase have been carried out for a long time, detailed pre-steady state kinetic studies are relatively recent [4, 49, 50]. Kinetic studies have helped delineate the enzymatic pathway followed by DNA polymerases (Figure 1.3) [4, 49] and it has been found that all polymerases (with one exception [51]), to date, follow the same overall sequence of events. The distinguishing features between different polymerases lie in the rate constants governing these steps, giving rise to variations in speed, efficiency and fidelity.

The minimal kinetic pathway consists of an ordered binding of the DNA to the polymerase forming the pre-chemistry binary complex followed by the binding of the dNTP to the binary complex to form the pre-chemistry ternary complex. This is followed by the chemical step of phosphodiester bond formation. In general, for cognate dNTP addition, chemistry is thought to occur fast compared to the catalytic cycle however for incorrect nucleotide addition the situation is different and chemistry is often the rate-limiting step of the catalytic cycle [4, 5, 21, 49, 50, 52-54]. Some TLS polymerases are exceptional such that on particular DNA sequences they can add incorrect dNTPs as fast as, or even faster than, the correct dNTP [20, 55].
Figure 1.3: Minimal kinetic pathway of DNA polymerases. The different stages of the enzyme-substrate complex are written in red. Abbreviations used are as follows, E: DNA polymerase; DNA_n: unextended DNA substrate; dNTP: deoxynucleoside triphosphate; PPI: pyrophosphate; DNA_{n+1}: DNA extended by one nucleotide.
The chemical SN2 reaction is facilitated occurs by the two metal ion mechanism [56] (Figure 1.4). As the name suggests, crucial for this mechanism is the presence of two divalent metal ions (typically Mg$^{2+}$) at the active site (denoted as A and B). Metal B is thought to enter the active site with the incoming dNTP and leaves with the PPI generated from the reaction (discussed later). Presence of metal A at the active site is more transient in nature and it is thought to rapidly diffuse in and out of the active site. The first step of this mechanism is the deprotonation of the 3’ hydroxyl (OH) group of the ribose sugar of the nascent primer DNA strand. The metal ions help in the deprotonation by reducing the pK$_a$ of the OH group, while a recent time resolved crystallographic study on a Y-family DNA polymerase indicates a transient water molecule at the active site to be at least partially responsible for this event [57]. Then the deprotonated 3’ OH performs a nucleophilic attack on the $\alpha$ phosphate of the incoming dNTP. This configuration is stabilized by the two metal ions and the two conserved catalytic aspartates. Often, there is a third catalytic acidic residue involved. Eventually metal B leaves with the $\beta$-$\gamma$ phosphate leaving the primer DNA extended by one nucleotide. However, the exact signal for the departure of the PPI is not well known. Time resolved crystal structures of the X-family polymerase Polβ indicates that the solvation of metal ion B eventually leads to PPI release [58]. On the other hand, kinetic studies of A-family polymerases suggest a conformational change preceeding PPI release may be responsible for the departure of the pyrophosphate [4, 49]. However, it must be taken into account that for Polβ, unlike the A-family polymerases, there is no strong evidence supporting the presence of a rate
limiting pre-chemistry step [59, 60] thereby indicating that the kinetic pathways for A- and X-family polymerases might be different.

Figure 1.4: Two metal ion mechanism. Structural model of the arrangement of DNA polymerase active site during the nucleotidyl transfer reaction. The model is based on the pre-chemistry ternary complex structure of *Geobacillus kaustophilus* PolC (PDB ID: 3F2B). The universally conserved catalytic aspartates (D963 and D965) as well as the third catalytic aspartate (D1088) are labeled by the Gka-PoIC residue number. The two
Mg$^{2+}$ ions are shown as green spheres while the black arrow indicates the direction of the nucleophilic attack by the 3’OH of the primer strand.
Recently, time resolved crystallographic studies of Polθ and Polβ have suggested the transient presence of a third metal ion during the addition of the correct dNTP, although its function is not clear [57, 58]. Both these studies suggest that the third metal might help in stabilizing the newly formed PPi. However, it remains to be seen whether this observation can be applied generally to all DNA polymerases.

Often there are several intermediate steps in between the pre-chemistry ternary complex formation and the chemistry step [61-63]. The precise nature of these steps is not always clear but they appear to be crucial for enzyme fidelity. One such step is the closing of the fingers on recognizing the cognate dNTP [41, 64-66] and this leads to completion of the active site formation. Initially fingers-closing was thought to be the rate-determining step of the entire pathway [62] but recent data has suggested that this event is fast compared to the overall rate of catalysis [63, 67, 68]. Nonetheless, it appears to be an important step for discrimination between cognate and non-cognate dNTPs [69, 70]. Recent viewpoints suggest the rate-limiting step might be a minor rearrangement of the DNA substrate, deprotonation of the 3’ OH or even the insertion of metal ion A (it has been indicated that metal ion B is coordinated by the incoming dNTP while metal ion A diffuses in later) [60, 63]. Further, it is quite possible that the rate-limiting step differs for different polymerases. For correct dNTP incorporation by T7 DNA polymerase, it has been shown that there is no clear rate-limiting step. Rather, there are several reversible steps leading up to chemistry and both the forward and reverse rates of these steps determine fidelity[69].
As mentioned before, chemistry is followed by the release of pyrophosphate (PPI) resulting in the formation of the post-chemistry binary complex (Figure 1.3). For most polymerases PPI release is considered much faster than the rate-limiting step leading up to chemistry. However, for mitochondrial DNA poly it was found that while incorporating the damaged nucleotide 8-oxo-dGTP or nucleotide analogs including AZTTP, PPI release was slow, leading to equilibrium between dNTP binding and product formation [71, 72]. Again there can be several steps between chemistry and PPI release.

After PPI release the polymerase is faced with two choices. For distributive synthesis, the polymerase dissociates from the elongated DNA and takes part in another round of catalysis. This dissociation is generally slow and is the rate-limiting step of a single correct nucleotide incorporation cycle. In case of processive synthesis however, the post-chemistry binary complex binds the next correct dNTP and undergoes another round of catalysis.

1.2 Processivity factor:

One major criterion for fast DNA synthesis is high processivity such that the polymerase can synthesize long stretches of DNA without dissociation. Although the slow dissociation rate of the polymerase-DNA binary complex helps in this regard, an additional protein called the processivity factor often results in a substantial increase in processivity [73]. The nature of this protein varies drastically between different polymerases. For instance, the bacteriophage T7 utilizes a host protein, thioredoxin as its
processivity factor while the monomeric large subunit of mitochondrial Polγ improves its processivity [2, 74]. For bacterial polymerases, processivity factor is a ring-shaped homodimer called the β clamp while for eukaryotes it is a homo-trimer called proliferating cell nuclear antigen (PCNA) [75]. Archaeal processivity factor is also called PCNA, but, depending on the species, it can be either a homo- or a hetero-trimer [75].

An interesting scenario is presented by the eukaryotic leading strand replicative polymerase Polɛ. Unlike the lagging strand polymerase Polδ, Polɛ was found to be highly processive even in the absence of PCNA [76]. A recent structural study has shown that using a unique P-domain, the polymerase wraps itself around the DNA thereby providing its own processivity factor [77].

Just the presence of the processivity factor does not necessarily lead to an enhanced processivity (Wu Y and Pata JD, Manuscript under preparation), [78], [79]. For instance, the B-family Sulfolobus polymerase Dpo2 shows very low processivity and even with the addition of the processivity factor and the loader (RFC) there is no significant increase in processivity [78]. On the other hand, in the presence of the processivity factor some replicative polymerases can synthesize kilobase lengths of DNA before falling off [73, 80]. This would indicate that instead of being a passive component, processivity factor increases processivity selectively. This might be a mechanism of modulating the arrangement of the polymerases at the replication fork.
Other than increasing processivity, processivity factors have been shown to regulate replication in a number of different ways. For different polymerases these factors have been shown to influence different steps of the polymerization cycle [81]. For instance, in vitro experiments show that in case of Dbh, an archaeal TLS polymerase, PCNA increases fidelity, at least on certain DNA substrates, by increasing the efficiency of addition of the correct nucleotide. This is achieved by increasing both the rate of dNTP incorporation and the affinity of the correct dNTP for the pre-chemistry binary complex (Wu Y and Pata JD, unpublished data).

In the context of the entire replisome, processivity factors are thought to serve as a scaffold (or “tool belt”) to bring different types of polymerases into close spatial proximity [82]. Extending the toolbelt model, a recent report suggests that the different polymerases compete with each other leading to a polymerase switch at the replication fork [83]. Moreover, it has been shown that other than increasing the processivity of DNA polymerases and co-ordinating their function at the replication fork, processivity factor are involved in a variety of different DNA transactions [84-86].

1.3 General organization of the replication fork:

DNA synthesis takes place in a 5’ to 3’ direction. Thus, due to the antiparallel nature of the DNA double helix, synthesis occurs in opposite directions for the two template strands. In order to maintain directionality of DNA synthesis and at the same time ensure that the replication fork moves unidirectionally, DNA synthesis cannot occur
continuously for both the template strands. The synthesis of the primer strand that uses the 3’ to 5’ strand as the template is continuous and this strand is called the leading strand. On the other hand, the strand that is synthesized using the 5’ to 3’ template strand is made discontinuously as fragments (Okazaki fragments) and is called the lagging strand. The discontinuous nature of lagging strand synthesis allows the lagging strand to loop back onto itself thereby allowing the replication machinery to move unidirectionally [87-89] (Figure 1.5A). Moreover, the loop grows and shrinks with each round of Okazaki fragment synthesis. Similar rates of leading and lagging strand synthesis indicate that replication events for these two strands are co-ordinated [90, 91]. However, the detailed molecular mechanisms of co-ordinating this complex process are still unclear.
Figure 1.5: Organization of the replication fork. (A) A schematic diagram of the orientation of the leading and lagging stands at the replication fork. The arrows indicate the direction of nascent strand synthesis. For simplicity, the replisome is not shown. (B) A schematic diagram of the arrangement of DNA polymerases at the bacterial replication fork. For simplicity the accessory proteins are not shown. Color coding for both parts are the same.
Extensive biochemical and fluorescence microscopy studies have provided us with a wealth of information about the organization of the DNA polymerases at the replisome with recent views suggesting that there are two polymerases for the synthesis of the two DNA strands while a third polymerase is present at the replication fork, which exchanges continuously with the lagging strand polymerase [89, 92]. For eukaryotes, leading and lagging strand DNA synthesis is carried out by Polε and Polδ respectively [93]. For both leading and lagging strand Polα synthesizes a short stretch of DNA extending from the RNA primer laid down by the primase and then hands over to the replicative polymerases [93, 94]

In case of bacteria, for Gram negatives and high G+C content Gram positives, the replicative polymerase DnaE1 is responsible for both extending from the RNA primer as well as replicating the bulk of the genome (Figure 1.5B) [95]. For low G+C content Gram positives, it was thought until recently that PolC carried out leading strand synthesis while DnaE3 replicated the lagging strand [95]. However, a recent in vitro study of the B. subtilis replisome has changed that view [3]. Now it is thought that PolC is responsible for replicating the bulk of the genome both for leading and lagging strand synthesis while DnaE3 synthesizes a short stretch of DNA, extending from a RNA primer and then handing replication to PolC, making the function of DnaE3 analogous to that of eukaryotic Polα (Figure 1.5B).
1.4 C-family polymerases:

Bacterial replicative polymerases belong to the C-family. They are encoded either by the polC gene in low G+C content Gram positive bacteria or by the dnaE gene in high G+C content Gram positive and Gram negative bacteria [95]. DnaE can be sub-divided into three classes, DnaE1 is the replicative polymerase in Gram negatives and DnaE2 occurs in some high G+C content Gram positives along with DnaE1 [96]. The exact role of this polymerase is unknown, although genetic studies suggest that DnaE2 is an error-prone polymerase [96]. Finally, DnaE3 occurs in low G+C Gram positives and is thought to be responsible for synthesizing DNA from the RNA primer synthesized by primase [3]. After DnaE3 has synthesized a short stretch of DNA, PolC takes over and synthesizes the rest. It is estimated that the switch to PolC from DnaE3 takes place early in the replication cycle; however, the exact mechanism of this handoff is unknown. One possibility is that like Polα [94], DnaE3 can distinguish between the A- and B-forms of DNA and a change from the A- to the B-form might act as a trigger for the transfer.

Although C-family polymerases have been studied biochemically for a long time [97], their structures have been solved very recently. Now X-ray crystallographic structures of both PolC and DnaE1 exist in apo enzyme, pre-chemistry binary and pre-chemistry ternary forms [43, 98-101]. However, only the pre-chemistry ternary complex of PolC is at a high enough resolution to allow detailed investigation of the interaction of the enzyme with the DNA and dNTP substrates (Figure 1.6A) [43]. After some initial controversy, it was established that, surprisingly, unlike the eukaryotic and archaeal
replicative B-family polymerases, C-family polymerases belong to the β-nt superfamily [98-100]. This suggests that evolutionarily, the bacterial replicative polymerases are distant from their archaeal and eukaryotic counterparts. Hence, the reaction mechanism of C-family polymerases cannot be assumed to be similar to the B-family ones. The other DNA polymerase family belonging to the β-nt superfamily is the X-family of repair polymerases [48].

The structure of the PolC ternary complex led to the identification of several unique structural features that might play an important role in polymerization mechanism and fidelity [43]. One such observation suggesting a structural basis for high fidelity of PolC was the extensive contacts between the thumb domain and the DNA. It is common for the thumb to contact the bases of the minor groove of the DNA up to one or two base pairs (bp) upstream of the templating base, and PolC is no exception. However, unique to PolC are two β strands that contact the pyrophosphate backbone of the DNA as far as 8 bp upstream of the templating base. It has been hypothesized that the thumb, via these β strands, can sense any distortion in the DNA geometry even up to 8 bp upstream and might act as a signal to trigger exonuclease proofreading activity (Figure 1.6B). In the absence of detailed kinetic characterization of the enzyme validity of such predictions could not be verified experimentally.
Figure 1.6: Structure of Gka-PolC pre-chemistry ternary complex. (A) Overview of the Gka-PolC ternary complex (PDB ID: 3F2B). The protein is shown in surface representation while the DNA is shown as ribbon. Primer strand is colored grey while the template strand is orange. Color coding of the various domains of the enzyme is shown below. (B) Interaction of the thumb domain with the DNA substrate. The DNA is shown in sphere representation. Color coding for the protein domains is the same in both parts of the figure.
A large part of our knowledge of the process of replication has been derived from the studies of the *E. coli* replisome [102]. Replisomes of low G+C Gram positives like *S. aureus*, *B. subtilis* and *S. pyogenes* have also been studied, but to a lesser extent [103-106]. While these studies are very informative about the overall behavior and organization of the replisome, they do not provide information about the mechanism of action of the polymerase. A combination of detailed pre-steady state and steady state kinetic studies would be required to gain a mechanistic insight into the mode of polymerase action. Such knowledge will be crucial for a better mechanistic understanding of bacterial replication including the mechanisms used to maintain high fidelity. Moreover, PolC has been identified as a target for different bactericidal compounds [107, 108] and a detailed knowledge of the mechanism of PolC action from a kinetic and structural perspective will help in the development of more potent antibiotic strategies. Although there have been some preliminary steady-state kinetic studies [109, 110], pre-steady state kinetic characterization of C-family polymerases is lacking. This is in stark contrast with most other polymerase families, for which detailed kinetic characterizations of several of the family members are available [59].

In this dissertation, I have performed detailed steady state and pre-steady state kinetic analysis of a truncated version as well as the full-length PolC. I wanted to understand the reaction mechanism of C-family polymerases, with a long-term goal of mathematically defining fidelity for these enzymes such that the predictions made about the structural sources of fidelity could be verified.
I found that overall PolC follows the same reaction pathway that has been delineated for other DNA polymerases [49]. However, PolC exhibits a few unique features. I found that the phosphodiester bond formation is reversible for PolC and is in equilibrium with dNTP binding thereby indicating that a step after chemistry, presumably PPI release, is slower than the reverse of chemistry. Further, our data shows that presence of the next correct incoming nucleotide increases the PPI release rate thus making processive DNA synthesis fast (Chapters 2 and 3).

1.5 Y-family polymerases and TLS synthesis:

Genomic DNA of all organisms suffer damage on a daily basis. The sources of these damages can be natural as well as man-made. As mentioned before replicative polymerases copy the genome faithfully and this very property typically prevents them from efficiently duplicating damaged DNA. Consequently, when the replisome encounters a DNA lesion, replication stalls. Such a situation will lead to the collapse of the replication fork, which can be very detrimental for the cell and can even lead to cell death. There are several different pathways to counter this situation. These pathways can be broadly divided into two major categories:

- **Damage repair pathways**, in which the lesion is repaired. For instance, base excision repair, nucleotide excision repair, etc.
• **Damage tolerance pathways**, where the damage remains in the template strand, but specialized pathways ensure that replication can proceed past these damages. For instance, homologus recombination, TLS synthesis, etc. In case of TLS synthesis, specialized TLS polymerases take over and replicate past the damaged DNA.

Although TLS synthesis is more error-prone compared to the repair pathways, TLS synthesis has the advantage that it can take place at the replication fork thereby preventing a stall in replication [111]. Often TLS synthesis and different repair pathways like nucleotide excision repair (NER), base excision repair (BER) etc. work in concert such that multiple pathways are involved and exhibit overlap in substrate specificity [112].

Central to TLS synthesis are the specialized polymerases called TLS polymerases, which can replicate past DNA lesions. These enzymes are found in all domains of life and with a few exceptions (e.g. Polζ, Pol II) belong to the Y-family of polymerases [34]. Different TLS polymerases show preference for bypassing different DNA lesions. For instance, eukaryotic Polκ can replicate past damaged guanosine bases like N2-dG adducts while Polη specializes in bypassing cis-syn TT dimers [113, 114]. Moreover, depending on the particular TLS polymerase, the damage can be replicated either accurately or inaccurately.

One of the major questions that arose immediately with the discovery of TLS polymerases was how these enzymes tolerated DNA lesions and could replicate past
them. The initial structural studies of TLS polymerases answered this question. It was found that unlike the highly “restrictive” active site of a replicative polymerase, TLS polymerases have a short fingers and thumb domain leading to a permanently “open” active site (Figure 1.2B and D), which allows them to accommodate different bulky DNA lesions or distorted DNA [19, 45]. However, another puzzling question remained unanswered: what gives rise to the preference shown by different TLS polymerases for bypassing different lesions? Recent data from our lab suggests that the position of the LF/PAD domain relative to the polymerase domain influences the orientation of the substrates and the formation of the active site and this is at least partially responsible in determining the mutational signature [55]. Further, it was found that this overall conformation was mainly determined by a short stretch of amino acids in the linker between the polymerase and the LF/PAD domain (see Figure 1.2B for the domain arrangement of a Y-family polymerase). However, the general applicability of this observation for all the Y-family polymerases needs to be verified.

While the “open” active site of TLS polymerases allows them to bypass lesions, it makes them very error-prone on undamaged DNA template [20, 21]. For instance, over-expression of E. coli DinB, a bacterial TLS polymerase, leads to increased single-base deletion frequency [115, 116]. The high error rate of TLS enzymes makes it essential to understand the mutation mechanisms used by them. In order to prevent widespread mutations, access of the TLS polymerases to the DNA is highly restricted. However, recent data suggest that at least for bacteria, TLS polymerases might have access to the DNA for longer than was thought before [117].
Y-family TLS polymerases can be divided into six sub-families with the DinB sub-family found in all domains of life [34]. DinB polymerases are known to cause single base deletion mutations on undamaged templates, especially on “deletion hotspot” sequences containing a run of pyrimidines followed by a purine [20, 118, 119]. There have been three major mechanisms proposed for single base deletion: Streisinger slippage, dNTP stabilized misalignment and misinsertion misincorporation misalignment [19]. While our lab and others have shown that *E. coli* DinB and archaeal Dbh used Streisinger slippage mechanism [19, 118, 120], Dpo4 (an archaeal TLS polymerase with 54% identity to Dbh) was thought to use the dNTP stabilized misalignment mechanism based on a structural study [121]. However, more recently our lab has demonstrated that on deletion hotspot sequences Dpo4, like its bacterial and archaeal counterparts, uses the Streisinger slippage mechanism and not the dNTP stabilized mechanism [122]. Polk is the eukaryotic counterpart of DinB polymerases and is the most distant relative in the DinB family [123]. Unlike other DinB polymerases, Polk possesses an elaborate N-terminal elongation that forms the N-clasp, a structural feature important for tighter DNA binding [26]. Moreover, some residues of the N-clasp are hypothesized to extend into the active site and are vital for catalysis [123]. So far the mechanism used by this polymerase for single base deletion was unclear. However, it has been shown that Polk can extend from a mispaired primer terminus by misaligning the primer/template junction [124], leading to the possibility that it might generate a single base deletion by misincorporation-misalignment mechanism.
In the second part of the thesis I have identified the mechanism by which the human DinB polymerase, Polκ, makes single base deletion mutations. We have found that like its archaeal and bacterial counterparts, Polκ uses Streissinger slippage to generate single base deletions on repetitive sequences. Further, we found that Polκ can actively realign the primer-template DNA and convert the single base deletion in a less deleterious base mismatch mutation (Chapter 4).

“*” Contributed equally

2.1. Description of contribution

For this manuscript, I have contributed equally while designing and performing the experiments and analyzing the results. The final manuscript was primarily written by me.
2.2. 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 ($k_{pol}$ 330 s$^{-1}$) but also dissociates from DNA rapidly ($k_{off}$ ~150 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.
2.3. 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 [33, 125], 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 [126]. 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 [3, 92, 103, 104, 106, 127-132].

Recently, several crystal structures of C-family polymerases have been reported, including DnaE from Escherichia coli and Thermus aquaticus and PolC from Geobacillus kaustophilus [43, 98-100]. Other than non-conserved N- and C-terminal
extensions, the individual domains of PolC and DnaE are structurally conserved [43], as expected from sequence conservation, but differ somewhat in their linear organization in the protein (Figure 2.1).
Figure 2.1. Overview of PolC structure and domain organization. (A) Structure of Gka-PolC-ΔNΔExo shown (PDB ID: 3F2B). Domains are color coded as in Figure 2.1B. 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 [102] 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 [102], 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 2.2) has been established for correct nucleotide incorporation [4, 49, 61, 62, 133]. 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 [61, 62]. Earlier structural studies suggested that this slow step might correspond to the large-scale domain movement associated with nucleotide binding [41, 134], but more recent studies have shown that motion to be too fast to be rate limiting [63, 67]. Although the conformational change accompanying nucleotide binding is faster than chemistry, it still controls specificity of nucleotide addition [69, 135]. 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 [61] 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) [43]. Furthermore, this construct has all the domains that are conserved between the PolC and DnaE polymerases (Figure 2.1). 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 2.2. 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.
2.4. MATERIALS AND METHODS

2.4.1. 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.) [136, 137].

2.4.2. 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.

2.4.3. 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 °C. All subsequent steps were carried out at 4 °C. 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$_{280}$ of the protein sample using a theoretical extinction coefficient of 87,100 M$^{-1}$cm$^{-1}$, and was stored at -80 °C.

2.4.4. 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 MgCl$_2$, 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 °C). 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 (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). All reactions were performed in triplicate.

2.4.5. 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:

$$ Y = [ED]_A k_{obs} t + C \quad \text{(Equation 2.1)} $$

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:

$$ k_{obs} = \frac{k_{cat} [dTTP]}{K_M + [dTTP]} \quad \text{(Equation 2.2)} $$

where $k_{cat}$ is the maximum steady-state rate and $K_M$ is the Michaelis constant for dNTP.
2.4.6. Measurement of DNA dissociation rate from binary complex. Kintek RQF3 rapid quench device was used to perform this experiment. 300 nM Sau-PoIC-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 2.3) and the rate of decrease of product formation was interpreted as the rate of dissociation of Sau-PoIC-ΔNΔExo from the preformed Sau-PoIC-ΔNΔExo • p/t DNA binary complex.

\[ Y = Ae^{-kt} + C \]  \hspace{1cm} \text{(Equation 2.3)}

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.

2.4.7. Active site titration of Sau-PoIC-ΔNΔExo. 300 nM Sau-PoIC-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:

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

where \( Y \) is the concentration of the extended primer, \([ED]_A\) is the concentration of the preformed active enzyme \( \cdot \) 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:

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

(Equation 2.5)

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 \([DNA]_T\) is the concentration of total p/t DNA at the beginning of the assay.
2.4.8. $K_D^{dNTP}$ determination. Primer extension assays were performed with a RQF-3 rapid quench instrument using final concentrations of 804 nM active $Sau$-PoIC-$\Delta N\Delta Exo$, 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 2.4). The rate, $k_1$, and [ED]$_A$ thus obtained were further plotted against [dTTP] and then fit to the appropriate hyperbolic equation:

$$k_1 = \frac{k_{pol}[dTTP]}{K_M^{dNTP} + [dTTP]} \quad \text{(Equation 2.6)}$$

$$[ED]_A = \frac{[ED]_{A\ max}^{\max} [dTTP]}{K_D^{dNTP} + [dTTP]} \quad \text{(Equation 2.7)}$$

where $k_{pol}$ is the maximum rate of the burst of product formation, $K_D^{dNTP}$ is the dissociation constant for dNTP binding to the $Sau$-PoIC-$\Delta N\Delta Exo$・p/t DNA binary complex and [ED]$_A^{\max}$ 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 $Sau$-PoIC-$\Delta N\Delta Exo$. All data were combined and analyzed together.
2.4.9. Simulation. The reaction mechanism of Sau-PoIC-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 [136] for a detailed description of how to interpret these plots.
2.5. RESULTS

2.5.1. PolC purification. Recombinant Sau-PolC-ΔNΔExo was purified using Ni-NTA, anion exchange, heparin and size-exclusion chromatography. Figure 2.3 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 [103], inducing protein expression at temperatures below 20 °C was critical for obtaining purified protein, estimated to be ~95% homogeneous, that did not have significant levels of proteolytic products.
Figure 2.3. SDS-PAGE of \textit{Sau-PolC-ΔNΔExo}. A 10\% SDS-polyacrylamide gel stained with Coomassie R-250 showing purified \textit{Sau-PolCΔNΔExo} obtained after size exclusion chromatography. (a) Kaleidoscope pre-stained marker. (b) 2.5 mM purified \textit{Sau-PolC-ΔNΔExo}. 
2.5.2. 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 2.4A). 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 Mg$^{2+}$ and reaction temperature (Figure 2.4B-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 2.4B). 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 2.4C). A concentration of 8 to 12 mM Mg$^{2+}$ was found to be optimal for enzymatic activity of Sau-PolC-ΔNΔExo (Figure 2.4D). No primer extension was observed in the absence of Mg$^{2+}$, as expected for a polymerase using a two-metal-ion mechanism. Very little primer extension occurred at 4 °C and 50 °C, but, for all other temperatures tested (25 °C, 30 °C and 37 °C), the enzyme performed well (Figure 2.4E). Based on these results, all subsequent reactions were performed at 25 °C at pH 8 with 25 mM NaCl and 8 mM Mg$^{2+}$. 
Figure 2.4. 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).
2.5.3. 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-PoIC-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 2.5A), and were plotted as a function of [dTTP] (Figure 2.5B). The data were fit to the Michaelis-Menten equation (Equation 2.2) and gave a $k_{cat}$ of $17 \pm 1 \text{s}^{-1}$ and $K_M^{dNTP}$ of $43 \pm 7 \text{μM}$. 
Figure 2.5. Steady-state kinetic analysis of *Sau*-PolC-ΔNΔExo.

**A**

Product (nM) vs. Time (s)

**B**

$k_{cat} = 17 \pm 0.9$ s$^{-1}$

$K_{M_{dNTP}} = 43 \pm 7$ μM
Figure 2.5. Steady-state kinetic analysis of Sau-PolC-ΔNΔExo. 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-ΔNΔExo. 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-ΔNΔExo. 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 ± 1 s^{-1} and Michaelis constant for dNTP (K_{M}^{dNTP}) was determined to be 43 ± 7 μM.
2.5.4. 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 2.6A. 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 2.6B). 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 2.3) and the rate of decrease in product formation (k_{off}) was found to be 150 ± 30 s\(^{-1}\). This indicated that for *Sau*-PolC-ΔNΔExo, 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 2.6. 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 2.3). 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 $150 \pm 30 \text{ s}^{-1}$. 
2.5.5. Pre-steady-state burst kinetics. To determine if Sau-PolC-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 2.7A). The fast phase represents the initial burst of dTTP incorporation by the pre-formed Sau-PolC-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 2.4) [138]. The rates of the fast and slow phases obtained were 150 ± 30 s⁻¹ and 8.5 ± 1 s⁻¹, respectively. Product formed during the fast burst phase was 12 ± 1 nM, indicating that out of 150 nM of Sau-PolC-ΔNΔExo, only 12 nM formed active enzyme • DNA binary complex that got converted to product.
Figure 2.7. Pre-steady-state kinetics and active site titration of Sau-PolC-ΔNΔExo.
Figure 2.7. 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 2.4). The rate of the fast phase was $150 \pm 30$ s$^{-1}$ and that of the slower phase was $8.5 \pm 1$ s$^{-1}$, [ED]$A$ was found to be $12 \pm 1$ 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 $300 \pm 14$ s$^{-1}$ and an amplitude of $11.5 \pm 0.5$ 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 2.4). 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 2.5). $K_D^{DNA}$ was determined to be $390 \pm 70$ nM and the concentration of active Sau-PolC-ΔNΔExo was found to be $100 \pm 8$ nM.
Since the rate of dissociation of the DNA substrate from the binary complex (Figure 2.2, 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 2.2, 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 2.7A), 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.

2.5.6. 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 $K_{D}^{DNA}$ and the concentration of active $Sau$-PolC-$\Delta N\Delta$Exo. For these assays, the final concentration of total $Sau$-PolC-$\Delta N\Delta$Exo 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 2.7B. 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]$_A$), obtained from the amplitudes of the fast phase, were plotted as a function of DNA concentration (Figure 2.7C). The data were fit to a quadratic equation (Equation 2.5). From the fit, the apparent $K_{D}^{DNA}$ was determined to be $390 \pm 70$ nM, indicating a relatively weak binding to DNA (Table 2.1), and the concentration of active $Sau$-PolC-$\Delta N\Delta$Exo was $100 \pm 8$ nM, implying that $\sim 70\%$ of the $Sau$-PolC-$\Delta N\Delta$Exo 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-$\Delta N\Delta$Exo associates with DNA with a rate constant ($k_{on}$) of $\sim 4 \times 10^8$ M$^{-1}$s$^{-1}$, which suggests that the rate of DNA binding is limited by diffusion.
Table 2.1. 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>[81,4]</td>
</tr>
<tr>
<td><em>Sau-PolC-ΔNΔExo</em> (observed)</td>
<td>C</td>
<td>390</td>
<td>4</td>
<td>180</td>
<td>150</td>
<td>-</td>
</tr>
<tr>
<td><em>Sau-PolC-ΔNΔExo</em> (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>[51,4]</td>
</tr>
<tr>
<td><em>T7DNA Polymerase</em></td>
<td>A</td>
<td>(18)</td>
<td>(18)</td>
<td>(287)</td>
<td>(0.2)</td>
<td>[4,5]</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>[139,8]</td>
</tr>
<tr>
<td><em>Mammalian Polô</em></td>
<td>B</td>
<td>300</td>
<td>-</td>
<td>7</td>
<td>0.005</td>
<td>[9,9]</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><em>Yeast Polô</em></td>
<td>B</td>
<td>30</td>
<td>24</td>
<td>1</td>
<td>0.045</td>
<td>[52,0]</td>
</tr>
<tr>
<td><em>Human mitochondrial Poly</em></td>
<td>B</td>
<td>39</td>
<td>14</td>
<td>3.5</td>
<td>0.03</td>
<td>[74,140,2]</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>[141,3]</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 $k_{cat}$.
2.5.7. Nucleotide binding affinity. Primer extension assays were again performed under burst conditions to determine the apparent $K_D^{dNTP}$ and $k_{pol}$. 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 2.8A. Rates of the fast phase ($k_1$), 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 2.6). From the fit, the apparent $K_D^{dNTP}$ was determined to be $3.2 \pm 0.9$ μM and $k_{pol}$ was $180 \pm 9$ s$^{-1}$ (Figure 2.8B). Although these parameters were reasonable compared to other replicative polymerases, the overall fit to the data was not good ($R^2$ 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 2.8C). The deviation of the observed nucleotide incorporation rates appeared to be due to the lower amplitudes of the fast phase at low nucleotide concentrations.

2.5.8. 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 [61]. Hence there is no equilibrium between dNTP binding (Figure 2.2, 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 2.8D) and fit well (R^2 of 0.98) to a hyperbolic equation (Equation 2.7). 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 2.2, step 3) and is in equilibrium with ground state dNTP binding (Figure 2.2, 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.
Figure 2.8. 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 ($1.17 \mu M$, $4.69 \mu M$, $9.4 \mu M$, $18.75 \mu M$, $30.14 \mu M$, $50 \mu M$ and $75 \mu M$). The concentrations of extended primers were plotted against time and the plots were fit to the full burst equation (Equation 2.4). (B) A plot of the observed rates of the fast phase ($k_1$) versus [dTTP]. The data were fit to a hyperbolic equation (Equation 2.6). From the fit, $K_D^{dNTP}$ was determined to be $3.2 \pm 0.9 \mu M$ and maximum rate of the burst ($k_{pol}$) was found to be $178 \pm 9 \text{ s}^{-1}$. $R^2$ value for this fit was 0.75. (C) An enlarged view of panel (B) of up to 15 $\mu M$ 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 $4.0 \pm 0.3 \mu M$ and the maximum $[ED]_A$ was $36 \pm 0.6 \text{ nM}$. $R^2$ value for this fit was 0.98.
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 2.9A). 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 2.7A, 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 μM⁻¹ 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 $K_D^{DNA}$ and DNA release rate ($k_{off}$) from the binary complex were used to determine the concentration of the preformed Sau-PolC-ΔNΔExo • 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 2.8A (Figure 2.9B). 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 $K_D^{dNTP}$ was 7.5 $\mu$M. From 3-D confidence contour analysis, all the calculated parameters appeared to be well constrained by the data (Figure 2.9C). 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 Sau-PolC-$\Delta\text{NAExo}$, suggesting that PPi release (or a conformational change required for PPi release) may govern the steady-state rate.
**Figure 2.9. 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.
2.6. DISCUSSION

We have determined the minimal kinetic pathway (Figure 2.10) 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 2.1). 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 $330 \text{ s}^{-1}$, but also binds DNA weakly, with a $K_D^{DNA}$ of $390 \text{ nM}$, and dissociates rapidly from DNA, with a $k_{off}$ of $150 \text{ s}^{-1}$ 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 replicative polymerases being non-processive in the absence of accessory protein factors like the β clamp [103, 142]. Furthermore, weak DNA binding is important for rapid, high fidelity DNA synthesis by bacterial replicative polymerases, as evidenced by an *E. coli* DnaE mutant, *dnaE173*, which has a single amino acid substitution of E612K [143]. 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 [143]. 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 [109, 115, 144].
Figure 2.10. 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 s\textsuperscript{-1} and is likely to be the main determinant of $k_{\text{cat}}$ (17 s\textsuperscript{-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 [71, 145], [72]. 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 showed 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 [146, 147]. 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 [103]. 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 [148]. 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 [103].

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⁻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⁻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 [43, 98-100]. 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 [108, 149]. 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.

2.7. 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.
CHAPTER 3

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

(MANUSCRIPT IN PREPARATION)

3.1. Description of contribution

For this manuscript, I contributed equally while designing and performing all the experiments. The fluorescence based pyrophosphate release assays that were entirely performed by Indrajit. I was involved equally in analyzing the results and am primarily responsible for writing the manuscript.
3.2. 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 \(\beta\) clamp processivity factor corroborate the earlier observation of equilibrium. Measuring the rate of pyrophosphate (PPi) release following nucleotide incorporation confirms PPi 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 PPi 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) [23, 42, 103, 104, 106] belong to the C-family [33]. In contrast, their replicative counterparts in eukaryotes and archaea mostly belong to the B-family [33]. 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 [48] 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 [96]. 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 [3]. 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 [43, 98]. A structure for the ternary complex of *Geobacillus kaustophilus* PolC (Gka-PolC) with DNA and incoming dNTP exists [43] but most of the biochemical work on C-family polymerases has focused on reconstituting the replisome [3, 104, 117, 132, 150] with kinetic studies of this family being long confined to preliminary steady-state assays [109, 110, 144]. 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 [151]. 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 [4, 49] (Figure 3.1A), several intriguing differences were observed.

PolC was found to bind DNA weakly (\(K_D^{\text{DNA}} \approx 400 \text{ nM}\)) mainly due to a rapid dissociation of the enzyme-DNA binary complex (\(k_{\text{off}} \approx 150 \text{ s}^{-1}\)) and the rate of chemistry (\(\approx 330 \text{ s}^{-1}\)) 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 (\(\approx 26 \text{ s}^{-1}\)) following chemistry being much slower (\(\approx 5\) 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 [71, 72, 145], 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 (~500 to 1000 nt/s) that is nearly 20-40 fold faster than PPi release?
**Figure 3.1. 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.

3.4. METHODS

3.4.1. 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 [151] 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 280 nm using a theoretical extinction coefficient of 108,000 M$^{-1}$cm$^{-1}$ and then stored at -80°C in storage buffer.

C-terminally hexa-histidine tagged $S$. aureus $\beta$ clamp monomer was cloned in pET23a [103] (a gift from Dr. Mike O’Donnell, The Rockefeller University) and transformed into Rosetta(DE3) pLysS E. coli cells. Cells were grown to OD$_{600}$ 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 (2 X 5 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$^{-1}$cm$^{-1}$ and stored at -80°C in storage buffer.
3.4.2. DNA annealing: Chemically synthesized 5’6-FAM labeled DNA primers were annealed to unlabeled DNA templates to form duplex DNA substrates (Figure 3.1B) 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.

3.4.3. 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-Exo\textsuperscript{Mut} was pre-incubated with the appropriate DNA substrate (Figure 3.1B) 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-Exo\textsuperscript{Mut}, 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-Exo\textsuperscript{Mut}, 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 MgCl\textsubscript{2}, 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 [151]. Briefly, pre-incubated PolC-Exo$^{\text{Mut}}$ 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.

3.4.4. 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 3.1) or to a single exponential equation (Equation 3.2) as shown below:

\[ Y = A(1 - e^{-k_1 t} + k_2 t) + c \]  \hspace{1cm} \text{(Equation 3.1)}

where \( Y \) is the concentration of the product formed, \( A \) is amplitude of the reaction, \( k_1 \) and \( k_2 \) are observed rates of the first and second phases respectively and \( c \) is a constant.

\[ Y = A(1 - e^{-kt}) + c \]  \hspace{1cm} \text{(Equation 3.2)}

where \( k \) 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 3.3).
\[ A = \frac{((K_D^{DNA})_{App} + E_A + [DNA]_T) - \sqrt{((K_D^{DNA})_{App} + E_A + [DNA]_T)^2 - 4E_A[DNA]_T}}{2} \]  

(Equation 3.3)

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 (\(k_1\)) were plotted as a function of \([dNTP]\) and the data were fit to equations 3.4 and 3.5 respectively.

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

(Equation 3.4)

\[ k_1 = \frac{k_\text{pol}[dNTP]}{((K_D^{dNTP})_{App} + [dNTP])} \]  

(Equation 3.5)

Where \(A_{max}\) is the maximum amplitude of the fast phase, \(k_\text{pol}\) 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.

3.4.5. Simulation: The reaction mechanisms of PolC-Exo\textsuperscript{Mut} (both in the presence and absence of beta clamp) were simulated using KinTek Explorer software [137]. 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
certainty contour plots for all possible pairs of rate constants [136].

3.4.6. Measurement of pyrophosphate release rate: Pyrophosphate (PPI) release rates
of PolC-ExoMut in the presence and absence of beta clamp and (where applicable) the
next correct incoming nucleotide, dATP, was determined using a previously published
protocol [71, 152]. A SF-120 stopped-flow apparatus (KinTek Corp.) was used for the
assay. PolC-ExoMut and DNA were pre-incubated in the presence or absence of the β
clamp (concentrations of reactants identical to primer extension assays) along with 1.5
mM E. coli phosphate binding protein mutant labeled at Cys197 with 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.
3.5. RESULTS

3.5.1. N-terminal and exonuclease domains slow DNA release without improving (K_D^{DNA})_{App}

Extending from our earlier study [151] with the PolC\Delta N\Delta 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 3.2A) 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_D^{DNA})_{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\Delta N\Delta Exo, PolC-Exo^Mut displayed weak DNA affinity ((K_D^{DNA})_{App}) of ~ 530 ± 280 nM (Figure 3.2C).
Figure 3.2. Pre-steady state kinetics of PolC-Exo$^{\text{mut}}$.

- **PolC-Exo$^{\text{mut}}$**
  - **A**: Product (nM) vs. Time (s)
  - **C**: Amplitude (nM) vs. [DNA] (nM)
    - $K_{\text{DNA}}^D = 530 \pm 280$ nM
  - **E**: Product (nM) vs. Time (s)
    - $k_{\text{off}} = 7.4 \pm 3$ s$^{-1}$

- **PolC-Exo$^{\text{mut}} + \beta$-clamp**
  - **B**: Product (nM) vs. Time (s)
  - **D**: Amplitude (nM) vs. [DNA] (nM)
    - $K_{\text{DNA}}^D = 116 \pm 40$ nM
  - **F**: Product (nM) vs. Time (s)
    - $k_{\text{off}} = 0.7 \pm 0.07$ s$^{-1}$
  - **H**: Product (nM) vs. Time (s)
    - (+ trap DNA) vs. (- trap DNA)
**Figure 3.2. 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 3.1). 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 3.3) and the \((K_{D_{DNA}})^{App}\) 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 \(k_{off}\) 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^{Mut}. This was surprising since Sau-PolC-ΔNΔExo preparations had been ~70% active on the same substrate [151]. Similar low active enzyme concentrations have been reported previously for the eukaryotic replicative polymerases, ε (~15%) and δ (~35%) [6, 153]. Despite an inactivated exonuclease domain, with alanine substitutions for its key catalytic residues (D426 and E428), the exo-domain in PolCExo^{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 _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 [154], indicating that the DNA can sample the exonuclease site easily. During _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 _Sau_-PolC-ΔNΔExo we found that a very high dissociation rate of the pre-chemistry binary complex (k_{off}) was responsible for the weak (K_{D,DNA}^{DNA})_{app}. For PolCExo^{Mut}, a similar situation holds true since the binary complex dissociates at a rate of 7.4 ± 3 s^{-1} (Figure 3.2E), comparable to that for the bacterial Y-family polymerase Pol IV [81]. Although fast compared to other replicative DNA polymerases, k_{off} for PolCExo^{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 $K_{D}^{DNA} = k_{off}/k_{on}$, where $k_{on}$ is the second order association rate constant, we calculated a $k_{on}$ of $1.25 \times 10^7$ M$^{-1}$s$^{-1}$. This was ~30 fold slower than the corresponding $k_{on}$ for Sau-PolC-ΔNΔExo and suggested that although the binary complex is more stable for full length PolC a slower association rate prevents tighter DNA binding.

### 3.5.2. $\beta$ 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 $\beta$ clamp increases both the processivity and activity of the enzymes [103], although the exact mechanism remains unclear. We hypothesized that, similar to the eukaryotic DNA Pol$\delta$ and mitochondrial Pol$\gamma$ [74, 153], 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 PolCExo$^{\text{Mut}}$ in the presence of the S. aureus $\beta$ clamp. For all experiments that included the sliding clamp, a longer substrate (Figure 3.1B, substrate S2) was used to accommodate the $\beta$ clamp on the DNA along with PolC.

We found that the presence of the clamp did improve DNA binding affinity by ~5-fold ($\langle K_{D}^{DNA} \rangle_{\text{App}} 116 \pm 40$ nM) (Figure 3.2B and D) but the active enzyme concentration remained unaltered at ~18%. The improved ($K_{D}^{DNA}$)$_{\text{App}}$ in the presence of the $\beta$ clamp could be a result of slower dissociation of the binary complex or a faster association of
the DNA and PolC. Therefore, we measured $k_{\text{off}} (0.7 \pm 0.07 \text{ s}^{-1})$ (Figure 3.2F) and then calculated the corresponding $k_{\text{on}} (0.6 \times 10^7 \text{ M}^{-1} \text{s}^{-1})$ 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 PolCExo$^{\text{Mut}}$, we used substrate S2 to obtain $k_{\text{off}}$ in the absence of $\beta$ clamp (data not shown) and found this rate comparable to the one measured using S1 (7 vs 18 s$^{-1}$).

### 3.5.3. 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 2G). Primer extension was performed in the presence of excess unlabeled 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 3.2H), 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 3.2G-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.

3.5.4. 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, \((K_D^{dNTP})_{\text{App}}\), the rate of the fast phase of product formation \((k_1)\) is seen to depend hyperbolically on dNTP concentration while the amplitude of this phase does not. However, PolC-Exo\textsuperscript{Mut} showed both rate and amplitude dependence on dNTP concentration (Figure 3.3C and 5.3E). Further, we obtained an apparent maximum rate of polymerization \((k_{\text{pol}})\) of 350 ± 19 s\(^{-1}\) and \((K_D^{dNTP})_{\text{App}}\) of 3.5 ± 0.7 and 1.3 ± 0.2 \(\mu\)M, 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 Sau-PolC-ΔNΔExo [151].
Figure 3.3. Affinity of dNTP for the pre-chemistry binary complex.

PolC-Exo$^{\text{mut}}$

A

Product (nM)

Time (s)

B

Product (nM)

Time (s)

C

$K_{D}^{\text{dNTP}}_{\text{app}} = 3.5 \pm 0.7 \, \mu\text{M}$

$k_{\text{pol}} = 350 \pm 19 \, \text{s}^{-1}$

D

$K_{D}^{\text{dNTP}}_{\text{app}} = 6 \pm 1 \, \mu\text{M}$

$k_{\text{pol}} = 470 \pm 28 \, \text{s}^{-1}$

E

Amplitude (nM)

[fTTP] (µM)

F

Amplitude (nM)

[fTTP] (µM)
Figure 3.3. 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 3.1). [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-ExoMut and those in (B) were 50 nM S2, 500 nM PolC-ExoMut and 10 μM clamp. (C and D) Secondary plots of the rate of the fast phase of product formation (k1) versus [dTTP] in the absence (C) and presence (D) of the clamp. Data were fit to a hyperbolic equation (Equation 3.4) and the (Kd dNTP)App and kp1 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 3.5) and (Kd dNTP)App 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 [71, 152]. The stopped-flow fluorescence traces and the corresponding quench-flow single-nucleotide incorporation data from the \((K_D^{dNTP})_{\text{App}}\) determination experiments were fit together globally using KinTek Explorer software (Johnson 2009) (Figure 3.4B and D). From this global fit it was evident that PPI release was indeed slower (10 s\(^{-1}\)) (Figure 3.4A) than chemistry (350 s\(^{-1}\)) for PolC-Exo\(^{\text{Mut}}\).

Although the results are consistent with our previous observations for PolC\(^{\Delta\text{N} \Delta\text{Exo}}\), 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 [103]. 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 3.3B 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 3.4C 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 3.4A). Furthermore, the \((K_d^{dNTP})_{App}\) obtained was comparable in all cases,
ranging between 2-5 \(\mu\text{M}\).
Figure 3.4. Simulation of the reaction pathway of PolC-Exo\textsuperscript{Mut}.
Figure 3.4. Simulation of the reaction pathway of PolC-Exo<sup>Mut</sup>. (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 3.3A 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<sup>Mut</sup> and 100 μM dTTP for (D) and 50 nM S2, 500 nM PolC-Exo<sup>Mut</sup>, 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.
3.5.5. 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 3.5). 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 3.5A). 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 3.5B and supplementary figure 3.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 3.5B). 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 basepairs (data not shown). Rate of processive synthesis calculated from this assay was also ~ 1000 s\(^{-1}\) (5 nt extended in 5 ms) and again, the synthesis rate remained independent of the presence of the processivity factor.
Figure 3.5. 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 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 alongside 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$^{\text{Mut}}$, 10 μM β clamp (dimer) and 100 μM of each dNTP. Final concentrations of the reactants for (B) were 50 nM S1, 1 μM PolC-Exo$^{\text{Mut}}$ and 100 μM of each dNTP.
3.5.6. 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<sup>Mut</sup> in the presence of the first two incoming nucleotides (1<sup>st</sup>: dTTP and 2<sup>nd</sup>: 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 3.6).

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 3.6B). Although the rate of the second PPI release (i.e. after dATP addition) was slow (~50 s<sup>-1</sup>) and similar to our previously determined rate of 40 s<sup>-1</sup>, the first PPI release event (i.e. after dTTP addition) was distinctly fast (~800 s<sup>-1</sup>) and comparable to the rate of both chemistry and processive synthesis (Figures 3.4A and 3.5). 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 3.6. 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-Exo\textsuperscript{Mut}. (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-Exo\textsuperscript{Mut}, 10 μM β clamp (dimer) and 100 μM each of dTTP (the first incoming nucleotide) and dATP (the second incoming nucleotide).
3.6. DISCUSSION

In this study, we have performed the pre-steady state kinetic characterization of exonuclease-inactivated full-length construct of *S. aureus* PolC (PolC-Exo$^{Mut}$) (Table 3.1). We find that post-chemistry pyrophosphate release by PolC-Exo$^{Mut}$ is slow (10 to 50 s$^{-1}$) irrespective of the presence of the β clamp and within the predicted range for PolCΔNΔExo. 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 [4] and viruses [50] led to the understanding that replicative DNA polymerases bind DNA very tightly (K$_D^{DNA}$ in low nanomolar range). However, more recent biochemical characterizations of cellular replicative polymerases from bacteria (*E.coli* DnaE) and eukaryotes (Pol δ and Pol ε) have revealed weak DNA binding [6, 153, 155] that is in agreement with our observation for *S. aureus* PolC ([K$_D^{DNA}$_]$_{App}$$\sim$500 nM). Five-fold tighter DNA binding by PolC was achieved through association with the processivity factor as has been shown for DnaE, Polδ and mitochondrial replicative DNA polymerase Poly [74, 153, 155]. Notably, the presence of the N-terminal and exonuclease domains in this construct did not appear to influence DNA binding when compared to *Sau*-PolC-ΔNΔExo (Table 3.1).
In contrast to the similarity observed for DNA binding affinity of both PolC constructs, the DNA dissociation rate ($k_{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 it is 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 $k_{off}$ [156]. Moreover, *E. coli* DnaE is known to show a significant improvement in polymerase activity when associated with its proofreading subunit ($\varepsilon$) [147, 154]. The crystal structure of Gka-PolC and subsequent 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 [43, 154]. Addition of the $\beta$ clamp reduces the $k_{off}$ ~ 10-fold more (Table 3.1), becoming the major reason for the improvement observed in DNA binding affinity. This has been reported for other DNA polymerases [74, 153]; 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 $k_{off}$ for PolC-Exo$^{\text{Mut}}$ compared to the deletion construct is offset by the corresponding ~30-fold slower association rate constant ($k_{on}=1.25\times10^7 \text{ M}^{-1}\text{s}^{-1}$) 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 3.1. Kinetic parameters obtained with different constructs of PolC.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PolCΔNΔExo</th>
<th>PolC-Exo$^{\text{Mut}}$</th>
<th>PolC-Exo$^{\text{Mut} + \beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D^\text{DNA}$</td>
<td>390 nM</td>
<td>530 nM</td>
<td>116 nM</td>
</tr>
<tr>
<td>$K_D^\text{dNTP}$</td>
<td>7.5 μM</td>
<td>2 μM</td>
<td>4.6 μM</td>
</tr>
<tr>
<td>$k_{\text{off}}$</td>
<td>150 s$^{-1}$</td>
<td>7.4 s$^{-1}$</td>
<td>0.7 s$^{-1}$</td>
</tr>
<tr>
<td>$k_{\text{forward}}$</td>
<td>220 s$^{-1}$</td>
<td>270 s$^{-1}$</td>
<td>540 s$^{-1}$</td>
</tr>
<tr>
<td>$k_{\text{reverse}}$</td>
<td>110 s$^{-1}$</td>
<td>120 s$^{-1}$</td>
<td>83 s$^{-1}$</td>
</tr>
<tr>
<td>$k_{\text{ppj}}$</td>
<td>27 s$^{-1}$</td>
<td>10 s$^{-1}$</td>
<td>40 s$^{-1}$</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 may be its importance in ensuring fast and high-fidelity DNA synthesis as is seen for \textit{E.coli} Pol III, where a mutant (\textit{dnaE173}) with tighter $K_D^{\text{DNA}}$ compromises both speed and accuracy of replication [143].

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 Poly 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) [71, 72]. For Pol\textbeta, a X-family repair polymerase, it has been demonstrated that pyrophosphorolysis enhances its fidelity such as in the 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 [157]. 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 [58]. 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 ([151], 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 showed 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 of 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 indicate different rates of PPi release after each nucleotide addition (Figure 3.6). We believe that the fast (~800 s\(^{-1}\)) 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\(^{-1}\)) 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 [158, 159], 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 the case of a dNTP pool imbalance, when replication by PolC would slow down due to a scarcity of the next correct dNTP thereby 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 sites 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.

3.7. ACKNOWLEDGEMENTS

We thank Dr. Ken Johnson for advice on the kinetic simulations, An Li for help with the pyrophosphate release assays and Dr. Mike O’Donnell for gifting us the clone of *S. aureus* β clamp.
3.8. SUPPLEMENTARY INFORMATION

3.8.1. 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. Substrate S1 was identical to that used for PolΔNΔExo [151] 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 [43]. 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.
3.8.2. 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 $k_{\text{off}}$ observed for the binary complex without the clamp (Figure 3.5B). However, it was surprising that the intermediate products were present even with the clamp (Figure 3.5A), since the binary complex in this case is relatively stable. To verify that PolC-Exo$^{\text{Mut}}$ does fall off midway through synthesis we repeated the above experiment in the presence of a DNA trap (Figure 3.S1). 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 3.S1B) could not extend very far (~5 to 10 nt), confirming the distributive nature of the enzyme. However, in the presence of the clamp (Figure 3.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 3.S1. Processive synthesis by PolC in the presence of trap DNA. Multiple nucleotides 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 3.5A and B, respectively, with the exception that for both Figure 3.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 4: Mukherjee P*, Lahiri I*, Pata JD. Human polymerase kappa uses a template-slippage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. Nucleic Acids Res. 2013 May; 41(9):5024-35.

“*” Contributed equally

4.1. DESCRIPTION OF CONTRIBUTION

For this manuscript, I have contributed equally while designing and performing the experiments and analyzing the results. I have played a major role in editing the manuscript.
4.2. 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.
4.3. 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 [160]. 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 [34].

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 \textit{in vitro} [161-165], and their absence makes cells more sensitive to DNA damaging agents [161, 166]. 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 [20, 21, 167].
A characteristic mutational feature of DinB polymerases is their ability to generate single base deletions at high frequencies (~$10^{-2}$ to $10^{-4}$) on undamaged DNA sequences [20, 115, 119, 168, 169]. 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 [124, 170]. Unlike other polymerases, hpolκ extends from mispairs at a higher frequency than it generates mispairs, with dC-C mispairs being extended most readily [170]. 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 [124, 170]. 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 [167, 170].
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 [119]. 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 [119].

To date three mechanisms have been suggested for generation of single base deletions (Figure 4.1). 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 4.1B). 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 [171, 172]. 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 [173-176]. 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 4.1C). 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 4.1D).
Figure 4.1. 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 hpol makes single-base deletions at higher rates in repetitive sequences (15). 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, hpol 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 hpol when extending from a mispaired primer terminus that is complementary to the next templating base [124].

Here we investigate how hpol generates deletion mutations, finding that hpol does primarily use a template-slippage mechanism in repetitive sequences, as do the archaeal and bacterial DinB polymerases, but hpol 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.
125

4.4. MATERIALS AND METHODS

4.4.1. 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 [26, 119, 123]. 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 (2 x 5 ml) GST columns (GE Healthcare) pre-equilibrated with Prescission 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 Prescission protease buffer containing 1 mg/ml of HRV3C protease. The untagged hPolκ was obtained the next day by eluting with Prescission 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⁻¹cm⁻¹), the aliquots were frozen at -80 °C.

4.4.2. Duplex DNA formation

All DNA oligonucleotides (shown in Table 4.1) 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 4.1: DNA Substrate Sequences.**

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

* Primers are labeled on the 5’ end with 6-carboxyfluorescein (FAM).

Nucleotides in bold vary between substrates.
4.4.3. 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 hPolk (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 MgCl$_2$, 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 hPolk, 50 nM DNA substrate and 1 mM dNTP) and quenched with 250 mM EDTA at appropriate time points (0.02 to 60 s). 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).

$$\%_{\text{extension}} = \frac{100 \times [\text{extended}]}{[\text{unextended}]+[\text{extended}]}$$  \hspace{1cm} (Equation 4.1).

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

Most data were fit to an exponential equation (Equation 4.2) unless a biphasic nature was observed, in which case the burst equation (Equation 4.3) was used:
\[ y = A(1 - e^{-k_{\text{obs}}t}) + c \]  \hspace{1cm} (Equation 4.2)

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.

\[ y = A(1 - e^{-k_1t} + k_2t) + c \]  \hspace{1cm} (Equation 4.3)

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 4.2.
TABLE 4.2: Summary of observed nucleotide incorporation rates.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>dCTP addition</th>
<th>dGTP addition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>4C-G</td>
<td>2 ± 0.5</td>
<td>0.146 ± 0.08</td>
</tr>
<tr>
<td>1T-G</td>
<td></td>
<td>0.44 ± 0.03</td>
</tr>
<tr>
<td>2T-G</td>
<td></td>
<td>0.38 ± 0.04</td>
</tr>
<tr>
<td>3T-G</td>
<td></td>
<td>3.47 ± 0.34</td>
</tr>
<tr>
<td>4T-G</td>
<td></td>
<td>1.59 ± 0.14</td>
</tr>
<tr>
<td>4C-A</td>
<td></td>
<td>0.12 ± 0.01</td>
</tr>
<tr>
<td>1T-A</td>
<td></td>
<td>1.14 ± 0.09</td>
</tr>
<tr>
<td>C-C mispair</td>
<td></td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>4C-GA</td>
<td>2.5 ± 0.5</td>
<td>0.153 ± 0.02</td>
</tr>
<tr>
<td>4G</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.5. RESULTS

4.5.1. 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 [19, 20, 122, 168, 169], 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 4.1). On this substrate, hPolκ adds the correct incoming nucleotide, dGTP, at a rate that is only 3.6-fold faster ($k_{obs}$ of $2.62 \pm 0.39 \text{ s}^{-1}$) than the rate of incorrect nucleotide, dCTP ($k_{obs}$ of $0.72 \pm 0.04 \text{ s}^{-1}$) addition (Figure 4.2A), indicating that deletions can be initiated on this substrate at a very high frequency.
FIGURE 4.2: Nucleotide incorporation by hPolκ on repetitive sequence containing deletion hotspot.
**Figure 4.2. 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 $k_{obs}$ for incorporation of dCTP by 6-fold from $0.72 ± 0.04 \text{ s}^{-1}$ to $0.12 ± 0.01 \text{ s}^{-1}$. (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 $k_{obs}$ from $0.44 ± 0.03 \text{ s}^{-1}$ to $1.14 ± 0.09 \text{ s}^{-1}$. 
Interestingly, the dGTP (correct) addition follows biphasic behavior with a burst amplitude of $56.2 \pm 3.2$ nM and a slow second phase rate of $0.036 \pm 0.012$ 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 4.2) with an amplitude of $77.41 \pm 1.63$ nM.

Hence, the biphasic nature of product formation for dGTP addition cannot 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 $\sim 8$-fold larger when extending from a mispair than from a perfectly base paired p/t DNA junction [22]. 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 hPolk.

4.5.2. hPolk predominantly uses a template slippage deletion mechanism on iterative sequences.

Since hPolk has been reported to use a misalignment mechanism to make deletions on non-repetitive sequences [124], 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 4.1), 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 4.1C). In the 4C-G substrate, a G occupies this position. To determine if hPolk utilizes this +1G when generating deletions, we changed it from G to A (Table 4.1, 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 $0.72 \pm 0.04 \text{ s}^{-1}$ to $0.12 \pm 0.01 \text{ s}^{-1}$ (Figure 4.2B). This suggested a clear dependence on the +1G, thus ruling out misincorporation-misalignment as the major deletion mechanism used by hPolk 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 [171, 177]. 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 4.1, 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 [19]. We used this assay to test if hPolk is able to tolerate and extend p/t DNA with an extrahelical nucleotide at positions -4, -3, -2 and -1 of the template (Figure 4.1B and D; Figure 4.2C). 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 [19, 120, 122].

We found that hPolk showed a strong preference for unpaired nucleotides present at positions -3 and -4 (substrates 3T-G and 4T-G) with $k_{obs}$ of $3.47 \pm 0.34 \text{ s}^{-1}$ and $1.59 \pm 0.14 \text{ s}^{-1}$ respectively (Figure 4.2C). Interestingly, these data showed a biphasic behavior, so were fit to a burst equation (Equation 4.3), 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 $k_{obs}$ of $0.38 \pm 0.04 \text{ s}^{-1}$ 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 [19, 120, 122].

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 hPolk 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 hPolk is able to use slipped template DNA sequences efficiently, it is also able to add dCTP on 1T-G (Table 4.1), which has a T at the 0 position, with a $k_{obs}$ of 0.44 ± 0.03 s$^{-1}$ (Figure 4.2C), 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 hpolk 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.
4.5.3. hPolX readily misincorporates dC when the deletion hotspot is changed to 3’-CCCTG-5’

We then considered the possibility that hPolX 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 $0.44 \pm 0.03 \text{ s}^{-1}$ to $1.14 \pm 0.09 \text{ s}^{-1}$ was observed (Figure 4.2D), in contrast to the 6-fold decrease observed on the 4C-A substrate compared to the 4C-G substrate (Figure 4.2B). 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 hPolX produces T-dCMP mispairs at a very high frequency of $8.2 \times 10^{-3}$ [119]. 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.

4.5.4. 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 4.3A, panel (i), and 4.4B, 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 4.3. Mechanism of second dGTP incorporation on 4C-G substrate.
Figure 4.3. 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 4.4. Mechanism of multiple dCTP incorporations.
**Figure 4.4. Mechanism of multiple dCTP incorporations.** (A) Schematic representation of the two possible ways in which hPolk 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 hPolk’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 4.3A, panel (i)), suggesting a gradual misincorporation opposite the +1G after the correct nucleotide at position 0 had been added (Figure 4.3C). To test this possibility we designed substrates that already had the first nucleotide incorporated as a G (Table 4.1, 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 4.3A(ii) and B), 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 4.3A and B vs. Figure 4.4B 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 4.4A). 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 hPolx is an efficient extender of mispairs [170], 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 4.1, 4C-GA) and another where the primer
contained a 3’-C to generate a mispair at the p/t junction (Table 4.1, 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 4.4B 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) [167]. 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: $0.146 \text{ s}^{-1}$ and 4C-GA: $0.153 \text{ s}^{-1}$), but dCTP incorporation from a C-C mispair was two-fold faster ($0.33 \pm 0.02 \text{ s}^{-1}$) (Figure 4.4D). This observation can be explained since the 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 [167].
4.5.5. 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 4.5A 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 4.5B and E) and found that different patterns and rates of incorporation on the two substrates provided further information to decipher the mechanism.
Figure 4.5. hPolk misincorporates efficiently on 1T-G substrate.
Figure 4.5. hPolk 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 4.5A and B). On the 1T-A substrate, in contrast, the amount of primer extended by one nucleotide reached a maximum level of ~80% (Figure 4.5D 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 4.5C and F. On both substrates, the first dC addition is templated by the T in position 0. On the 1T-G substrate (Figure 4.5C), the second dC is templated by the G at the +1 position, and the rate of this incorporation is fast because hpol\(x\) can efficiently add the next correct nucleotide even from a mispaired primer terminus [170]. On the 1T-A substrate (Figure 4.5F), 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 hpol\(x\) [167]. 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.

4.6. DISCUSSION

Since hpolκ is the most processive of the human Y-family polymerases, it has the greatest potential to introduce mutations into the genome [119]. 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’-CCCCG-5’ template sequence. Strikingly, however, our results indicate that hpolκ is able to efficiently realign 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 [166], 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 hpolk 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 (1.6-3.1 x 10^{-3}) versus non-iterative (1.0 x 10^{-3}) sequences [119]. Our findings help explain why there is such a small difference in single base deletion rates. On repetitive sequences, hPolk 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, hPolk has been found to extend from mispaired primer-template termini by initiating template-primer misalignment when direct extension does not occur [124]. 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 [119].

We have not found any evidence that hpolk 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 [19, 118, 120, 122, 178]. 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 [121]. 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 [121]. 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 [118, 178]. 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 [19, 122].

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 4.5). 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 hpolk (Figure 4.6) 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 [19, 122]. In contrast, polymerases from other families have much tighter constraints around the DNA duplex, which would suppress nucleotides from adopting extrahelical conformations.
Figure 4.6. Structure of an hPolk ternary complex.
Figure 4.6. Structure of an hPolk 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 [55]) 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 [179]. In the case of hpolκ (Figure 4.6), 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 [26]. 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 3.6A).

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.

3.7. FUNDING

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

3.8. ACKNOWLEDGEMENTS

We thank Xiaojun Huang for constructing the hPol κ1-560 clone, Zhigang Wang for providing the full-length clone of hPol κ, and Ryan Wilson for careful proofreading of the manuscript.

*Conflict of interest statement.* None declared.
Chapter 5: Conclusion and Future Direction

In this dissertation I have worked with two different families of DNA polymerases, namely the C- and the Y-family. In this chapter I will briefly describe the results of those studies, discuss their significance and finally talk about some of the unanswered questions and the future directions for these projects.

5.1. C-family polymerases:

Bacterial replicative polymerases belong to the C-family [33]. This family has been studied biochemically and genetically for a long time but only recently has there been structural information available [43, 98-100]. Based on the structure of the low G+C content Gram positive replicative polymerase PolC, several interesting suggestions about the structural basis for fidelity were made [43]. However, due to lack of kinetic data these suggestions could not be verified. In this study I have defined the minimal reaction pathway of PolC from *S. aureus*. Notably, my studies are the first detailed kinetic characterization of a C-family polymerase.

5.1.1. Results summary:

The initial characterization was done using a truncated version of PolC (PolC-ΔNΔExo) lacking a non-conserved N-terminal domain and a 3’ to 5’ proofreading exonuclease domain. This construct is identical to the Gka-PolC construct that was used for the
crystallographic study. Kinetic characterization of this construct showed that although PolC follows the same general enzymatic pathway that has been described for other polymerases, there are some unique features. Firstly, we found that PolC binds DNA weakly with an apparent dissociation constant ((K_{D_{DNA}})^{App}) of ~ 400 nM. This was in contrast to other replicative polymerases that generally have a (K_{D_{DNA}})^{App} in the low nM range [4] [50] [74]. Further experiments showed that this weak binding resulted from a very rapid dissociation of the polymerase - DNA binary complex with a dissociation rate constant (k_{off}) of ~ 150 s^{-1}. Most surprisingly, we found that the bond formation step for PolC was reversible and in equilibrium with dNTP binding. This would indicate the presence of a slow post-chemistry step, presumably PPi release, such that the PPi release rate is slower than the reverse rate of chemistry. Indeed, simulation of the reaction pathway predicted a slow PPi release rate of 26 s^{-1}, forward rate of chemistry to be 220 s^{-1} with a reverse rate of 110 s^{-1}.

Although this study provided us with the first glimpse of the kinetic mechanism of PolC, several questions remained unanswered. Firstly, is the kinetics of the full-length PolC similar to that of PolC-ΔNΔExo? The second question was more puzzling. PPi release needs to occur after addition of every nucleotide, i.e., after every round of chemistry. Thus, if PPi release is slower than chemistry, then the processive synthesis rate will be guided by PPi release and will never be faster than 26 s^{-1}. However, it is well documented that bacterial replication is a fast process with rates ranging between 500 to 1000 s^{-1} [3, 103, 180]. Comparison of our results with the previously published reports indicates that
PPi release rate for processive synthesis is faster than that for distributive synthesis. In that case what causes PPi release to be fast for processive synthesis?

In order to use full-length PolC in our kinetic study we mutated two of the catalytic aspartate residues (D426A and E428A) of the exonuclease domain (PolC-Exo\textsuperscript{Mut}). This significantly reduced the exonuclease activity of PolC and prevented non-specific degradation of the substrate DNA. Kinetic studies with PolC-Exo\textsuperscript{Mut} clearly showed that other than having slower association and dissociation rates of the binary complex, full-length PolC exhibited similar kinetic trends compared to the truncated version of PolC.

Most importantly, like PolC-\Delta\Delta\text{Exo}, full-length PolC showed reversible bond formation followed by slow PPi release. Based on a previously published report [103], we initially hypothesized that addition of the $\beta$ clamp processivity factor would increase PPi release rate, thereby making processive synthesis fast. However, from our single nucleotide incorporation assays, we found that although the clamp increases the stability of the binary complex by reducing $k_{\text{off}}$, it does not have any significant effect on increasing the PPi release rate.

On the other hand, based on the processive synthesis and related PPi release assays, we came to the surprising conclusion that the presence of the next correct dNTP is necessary and sufficient for increasing the rate of PPi release, thereby making the processive synthesis rate comparable to the rate of bacterial replication. Further, we found that at
least under our assay conditions, the clamp did not play any major role in this phenomenon. The implications of this finding are discussed below.

5.1.2. Significance of the results

Firstly, the increase of the PPi release rate in the presence of the next correct incoming dNTP may act as a means to ensure accurate DNA synthesis at rapid speeds of 500-1000 $s^{-1}$. As we find that only the next correct dNTP can speed up processive synthesis, we can expect that the $t_{-1}$ (the next templating position) base has to sample the next correct incoming nucleotide for speeding up PPi release. Hence, this might act as a checkpoint such that rapid synthesis can only be achieved once the presence of the next correct incoming dNTP has been ensured. Additionally, if dNTP selection takes place simultaneously with catalysis, then the efficiency of selection process will be greatly improved and the time required for each round of catalysis might get reduced.

Alternatively, this may act as a sensory mechanism to detect the overall cellular levels of dNTP or any imbalance in the dNTP pool. Once any imbalance is sensed, polymerization will slow down due to a reduction in the concentration of the next correct dNTP, thereby preventing a large number of errors from being generated, and at the same time buying time for the cell to rectify the imbalance.

It has been shown that an elevated level of cellular dNTPs leads to error prone DNA synthesis by PolC, while a lower dNTP level has been associated with accurate DNA
synthesis [181]. This may be a direct outcome of our observation. It is possible that once the concentration of the incorrect dNTP increases beyond a certain level, the sheer high concentration will allow pairing of the $t_{-1}$ base to an incorrect nucleotide and thereby even the next incorrect dNTP may increase the speed of processive synthesis leading to error-prone replication.

5.1.3. Future directions

We have demonstrated that the next correct incoming dNTP increases the rate of processive synthesis by PolC by speeding up PPi release. However, the mechanism by which this happens is still unclear. The most obvious hypothesis is that the next correct nucleotide pairs with the $t_{-1}$ base and then this new base pair “swings into” the active site of the polymerase. If such is the case, then it is possible that the triphosphate moiety of the next correct nucleotide displaces the PPi generated during the first round of nucleotidyl transfer thereby speeding up PPi release.

Although the ternary complex structure of Gka-PolC shows that the $t_{-1}$ base is solvent exposed and can base pair with a nucleotide, the “swinging in” motion of the basepair is hard to imagine. Also, from the pre-chemistry ternary complex structure available, it is difficult to envision the constraints that would ensure correct Watson-Crick pairing between the next incoming dNTP and the $t_{-1}$ nucleotide. Further, there remains yet another question, if our hypothesis is true then what is the function of the conventional dNTP entry channel of PolC? It remains a possibility that the dNTP that samples the $t_{-1}$
position is not the one that actually gets added. Instead, the nucleotide that will finally be added enters the active site through the conventional dNTP binding channel.

Nonetheless, if our hypothesis about the next incoming dNTP base-pairing with the $t_{+1}$ base is true, then the presence of the $t_{+1}$ base should be vital for speeding up the PPi release. This can be verified by making the $t_{+1}$ position abasic. Under this condition, our hypothesis would predict that PPi release would never become fast. Moreover, if the triphosphate of next nucleotide displaces the PPi, then the second round of catalysis should not be required for fast PPi release. Adding a non-hydrolyzable dNTP analog, like dNMP-C-PP, as the next correct nucleotide, can be useful to test this hypothesis. A non-hydrolyzable dNTP analog will behave like a regular dNTP with regard to base pairing and binding to the enzyme but will prevent the second round of phosphoryl transfer. So, if our hypothesis is correct, then PPi release will speed up even in the presence of a non-hydrolyzable dNTP analog.

In order to recognize the structural properties of the next incoming dNTP that are responsible for fast PPi release, instead of adding the next correct dNTP, we can add different components of the dNTP separately like base, deoxyribose, mono-, di- or tri-phosphate. If a single component is responsible for speeding up PPi release, then adding that component instead of the entire dNTP should lead to fast PPi release. However, it is quite possible that one part of the dNTP, presumably the base, is required for recognition of the $t_{+1}$ nucleotide while a different component is needed for the displacement of the
PPi. To test for such a possibility different combinations of the components, like deoxynucleoside monophosphate or deoxynucleoside diphosphate, can be added.

It will be very useful to be able to visualize how the next correct dNTP samples the t_{-1} nucleotide. Crystallization conditions for the pre-chemistry ternary complex of Gka-PolC are already known. Hence, it might be possible to grow the Gka-PolC ternary complex crystals and then soak them with the next incoming dNTP. However, one possible caveat in this plan is that we do not know the step of the catalytic cycle where the sampling of the t_{-1} nucleotide occurs. In order to grow the pre-chemistry ternary complex crystals the DNA substrate had a 3’ dideoxy ribose thereby preventing catalysis. So, if the sampling occurs after catalysis then it might be difficult to obtain the crystals.

As discussed above, we need to know at which point of the catalytic cycle does the next correct dNTP sample the t_{-1} base. Three major possibilities open up. First, the next correct dNTP is sequestered by the t_{-1} nucleotide even before chemistry or possibly pre-chemistry ternary complex formation takes place for the first incoming dNTP. A somewhat similar situation has been indicated for RNA PolII [158, 159]. Second, the next correct dNTP can potentially pair up with the t_{-1} nucleotide right after chemistry for the first dNTP addition and this pairing might act as a signal for flipping in of the t_{-1} base from the solvent exposed configuration, which will be required for its orientation at the active site. Finally, it is possible that the flipping of the t_{-1} base initiates after the first chemistry event, without the second incoming dNTP being present. But during the flipping the next correct dNTP samples the t_{-1} nucleotide and accelerates the base flip.
In order to experimentally verify these possibilities, using a fluorescent label on the next incoming dNTP might be useful. In such an experiment we can monitor the fluorescence anisotropy of the labeled nucleotide and thus can follow the sampling of $t_{+1}$ base by the next incoming dNTP. Any pairing should lead to a reduced tumbling motion of the second incoming nucleotide and thus will lead to an increase in fluorescence anisotropy. Base flipping for the $t_{+1}$ nucleotide can be followed in the same experiment by using the fluorescent dNTP analog, 2-amino purine (2-AP), as the $t_{+1}$ nucleotide. There is an increase in the fluorescence signal of 2-AP when the base is stacked and vice versa [118]. When the $t+1$ base is flipped out towards the solvent side then stacking interaction will be weaker. However, when the base flips in, then it stacks against the neighboring bases leading to strong stacking interactions. Hence, monitoring the fluorescence of 2AP will allow us to follow base flipping. However, there is a possibility that using a fluorescently labeled dNTP and a nucleotide analog simultaneously might significantly alter the enzyme kinetic pathway. In such a situation, sampling of the $t_{+1}$ base by the second incoming dNTP and the base flipping events need to be followed separately and the data from the two experiments can be analyzed globally.

Finally, it will be interesting to know whether all the C-family polymerases show slow PPI release in the absence of the next correct dNTP or whether it is just PolC that shows this surprising behavior. Moreover, we need to find out if the DNA sequence plays any significant role in modulating the PPI release rate. Our preliminary studies with *S. aureus* DnaE3 indicate that this polymerase does exhibit slow PPI release but only on an RNA
primer – DNA template hybrid duplex. On a pure DNA duplex we do not have any compelling evidence of slow PPi release. Notably, a recent study has indicated that the main function of DnaE3 is to generate a short stretch of DNA from an RNA primer thereby making the hybrid duplex its natural substrate. One possibility is that all C-family polymerases show slow PPi release on their natural substrates. However, more C-family polymerases, including *E. coli* DnaE1, need to be kinetically studied before we can arrive at any conclusion.

The kinetic study of PolC was started with the goal of defining the enzymatic pathway for this polymerase such that we could mathematically define fidelity and verify the predictions made about the different structural sources of fidelity for PolC. In the studies with PolC we have described a minimal kinetic pathway for the correct dNTP incorporation by the polymerase. However, in order to clearly define fidelity, a detailed picture of the enzymatic pathway for both correct and incorrect dNTP incorporation is required.

Fidelity is defined as the ratio between the efficiencies of correct and incorrect dNTP incorporation, while efficiency is defined as $k_{\text{cat}}/K_M$ where $k_{\text{cat}}$ is the steady state rate constant and $K_M$ is the Michaelis constant [138]. Comparison of the apo enzyme, binary and ternary complex structures of C-family polymerases indicates that PolC undergoes a series of large conformational changes before chemistry (Figure 5A). Moreover, there can be several post-chemistry conformational changes and, as we know, in addition to chemistry being reversible, PPi release is slow for PolC. All these make the enzymatic
pathway for PolC very complex and hence the simple assumption that $k_{cat}/K_M = k_{pol}/K_D$, which holds true for the following simple minimalist pathway, may not hold true for PolC.

\[
\begin{align*}
\text{K}_D & \quad k_{pol} & \quad \text{fast} \\
E + D & \quad E.D & \quad E.D.dNTP & \quad E.D_1.PPi \rightarrow ED_1 + PPi & \\
\uparrow & \quad & \quad & \quad & \\
\text{dNTP}
\end{align*}
\]

(where E is the polymerase, D and $D_1$ are unextended and extended DNA respectively, $K_D$ is the dissociation constant for dNTP binding and $k_{pol}$ is the rate of chemistry).

Further, recent data for T7 DNA polymerase suggests that incorporation of the correct and incorrect dNTPs can follow completely different pathways \cite{69} and since we do not have any idea about the misincorporation pathway of PolC, we cannot define the efficiency of misincorporation.
Figure 5: Domain movement of C-family polymerases. (A) Global superposition of the apo and ternary structures reveals large domain motions. Gka-PolC ternary complex (PDB ID: 3F2B) and *E. coli* DnaE (PDB ID: 2HNH) superimposed in PyMol. Lines represent vectors joining Ca atoms that aligned between the two structures (color coded as shown below). Arrows indicate direction of rotation/movement of the domains going from the apo to the ternary state. (B) Superposition through alignment of the middle finger of apo conformation of *E.coli* DnaE (shown in grey) with the Gka-PolC ternary complex (color coded as shown below). Enlarged view of the active site shows motion of the N-palm and index finger residues that accommodate and stabilize nucleotide binding in the ternary complex (shown in bold for Gka-PolC and italicized for *E.coli* DnaE).
One solution to these difficulties might be performing steady state assays for both correct and incorrect single nucleotide incorporation and experimentally determining $k_{\text{cat}}$ and $K_M$. In most cases these assays have the caveat of the binary complex dissociation being the slowest step, thereby dominating $k_{\text{cat}}$ and when compared to processive synthesis, artificially lowering the steady state rate constant. However, for PolC-ΔNΔExo, this might be a feasible method because the slowest step of the pathway is PPi release, which occurs after every cycle of catalysis, while DNA release is extremely fast. For full length PolC however, it seems both PPi release and DNA dissociation are comparable and there is no clear rate limiting step, while in the presence of the β clamp DNA dissociation is rate limiting. Therefore, for these cases full enzymatic pathways need to be described before $k_{\text{cat}}/K_M$ can be calculated mathematically. In the cases of T7 DNA polymerase, Polβ and Klenow fragment, FRET experiments and studies with fluorescent nucleotide analogs proved vital for delineating the detailed pathway [63, 68, 69] and similar studies might be useful for PolC.

Describing the enzymatic pathway will allow us to define the mechanistic roles of different unique structural features of PolC. Perhaps the most surprising of these features was the uncommon architecture of the palm domain of PolC [43]. Structurally the palm is usually a continuous domain, but for apo enzyme form of the C-family polymerases it is split into two parts, the N-palm and C-palm [43]. Comparison of PolC ternary complex with the apo enzyme structures of *E. coli* and *Thermus aquaticus* suggests that the palm remains in the split conformation in the apo state, while in the ternary state the N- and C-palms join together, thereby completing the formation of the active site by bringing the
catalytic aspartates to the required positions (Figure 5B). Although the implications of this phenomenon are unclear, this might act as a checkpoint to ensure the presence of the cognate dNTP at the active site before catalysis can take place. One major question is when do the N- and C-palms join together? Does it occur during the pre-chemistry binary complex formation or after the binding of the cognate dNTP and the formation of the ground state pre-chemistry ternary complex? In addition to the kinetic pathway a high-resolution pre-chemistry binary complex structure of PolC will be instrumental in defining the mechanistic role of the palm joining.

One major bottleneck faced during the execution of this project was experimentally verifying the PPi release rate of PolC. We used an assay developed by Dr. Ken Johnson (UT Austin, TX) and it depends on a coupled reaction [152]. Briefly, the PPi released is broken down into inorganic phosphate by the enzyme pyrophosphatase and then the phosphate reacts with a MDCC-labeled phosphate binding protein leading to an increase in the fluorescence. While this is the best time-resolved PPi release assay available, there are a few problems. Firstly, since the assay actually measures phosphate and phosphate contamination can occur easily, a “phosphate mop” is required in order to absorb any contaminating phosphate. However, using this mop limits the slower end of the PPi release rates that can be accurately measured by this assay. More importantly, the coupled reactions make data analysis extremely complicated. In order to obtain reliable data, simulation of the reaction pathway is required. Recently another time-resolved PPi release assay has been developed but this assay also depends on a coupled reaction system [182].
It would be far simpler if there was an assay that directly measured the PPI released instead of breaking it down to Pi and then measuring the reporter signal. Then not only would the data analysis become simpler, the phosphate mop would also not be required. While there are standard assay kits (PhosphoWorks Fluorimetric Pyrophosphate Assay Kit, AAT Bioquest, Inc.) that do not require the breakdown of PPI, the main problem is that it requires at least 10 to 20 minutes incubation time before the signal can be recorded. This would indicate that the PPI detecting reaction is very slow and this limits the time-resolved aspect of the assay.

5.2. Y-family polymerases:

The study with the Y-family polymerase was aimed at understanding the single-base deletion mechanism used by human Polκ on a repetitive deletion hotspot sequence. Polκ is a eukaryotic TLS polymerase that belongs to the DinB subfamily, the only subfamily of Y-family polymerase that is found in all domains of life [34].

Previous studies have demonstrated that on repetitive sequences DinB polymerases generate single-base deletion errors [20, 119]. Single-base deletions are particularly harmful because if occurring in an open reading frame, they result in frameshift mutation. There are three mechanisms of single-base deletion, namely Streissinger slippage, dNTP stabilized misalignment and misincorporation misalignment [19]. It has been shown by our lab and others that bacterial and archaeal DinB polymerases used the Streissinger
slippage mechanism for generation of single base deletion on repetitive deletion hotspot sequence [19, 118, 120, 122]. However, the mechanism employed by the eukaryotic counterpart, Polk, was unknown.

5.2.1. Results summary

Experiments described in this thesis proved conclusively that on repetitive deletion hotspot sequence human Polk uses the Streigssinger slippage mechanism to generate a single-base deletion. This indicates that DinB polymerases across all domains of life universally use Streissinger slippage mechanism on the deletion hotspot. Moreover, based on different DNA constructs used, we found that like the bacterial and the archaeal DinB polymerases, Polk prefers to have the bulged base away from the active site, with the most preferred position being t₃ (three nucleotides upstream of the templating base). However, for further insight into the process structural data will be necessary.

Further, our experiments showed that once Polk generates a bulged base by causing a single base deletion, during the next round of catalysis it can realign the bulged nucleotide and therefore correct the single-base deletion mutation at the cost of generating a less deleterious base substitution mutation. This result explains why in vitro genetic assays indicate that there is a very small difference in the frequencies of generating deletions on repetitive versus non-repetitive sequences [119]. Some of the deletions generated by Streisinger slippage on the repetitive sequence will get converted into base substitution mutation, thereby reducing the frequency of generation of single
base deletion. Similar observations about realignment of the bulged base have been made for Dpo4 and also for Dbh, albeit on shorter repeats [20, 55]. Taken together, it appears that rearrangement of the bulged base generated during single-base deletion in favor of generating a mispair, is a common property of DinB polymerases.

5.2.2. Future directions

As mentioned before our results show that, consistent with the behavior of other DinB polymerases, human Polκ generates single-base deletions on repetitive deletion hotspot sequence using Streissinger slippage mechanism. However, in the absence of structural data on how a bulge is tolerated at the active site of Polκ, it is difficult to understand the DNA realignments and the novel contacts between the enzyme and the DNA that are required for this process. Although a structural model can be created based on the crystallographic structures of the pre-chemistry ternary complexes of archaeal Dbh and Dpo4 with a bulged base at t₃ position, it may not provide us with the real picture. Both Dbh and Dpo4 are related archaeal DinB polymerases with 54% sequence identity and yet they tolerate the bulged t₃ base in completely different ways [19, 122]. Hence, in order to better understand the structural details of the Streissinger slippage mechanism employed by Polκ it is of vital importance to obtain high-resolution structures of Polκ in complex with the bulge containing DNA at different points of the enzymatic cycle, most importantly the pre-chemistry ternary and the pre-chemistry binary stages. Polκ has already been crystallized in both the apo and ternary forms, albeit with different DNA substrates [26, 123]. These crystallizing conditions can be used as the starting conditions
for the crystallographic trials. If they prove unsuccessful, different commercially available crystallization screens can be tried.

One major unanswered question regarding the DNA rearrangement after generation of the single base deletion is whether Polκ actively rearranges the bulged base or whether, due to the distributive nature of the enzyme, it dissociates from the DNA after generating the single base deletion and the now-free DNA breathes and rearranges the bulged base.

One possible experiment to answer this question is described here. For this we need a DNA substrate with a 5’ FAM label and 2-AP as a bulged base at a suitable position. If 2-AP is used to mimic the bulged base then the rearrangement of the 2-AP to form a mispair (thereby stack in against the neighboring bases) can be followed by monitoring the fluorescence intensity of the analog while simultaneously, we can determine whether the polymerase is still loaded on the DNA by measuring any change in fluorescence anisotropy of the 5’ FAM labeled DNA substrate. If the polymerase releases the substrate, then there should be a noticeable drop in anisotropy. This experiment can be performed using bulk kinetics approach. However, if data analysis proves difficult, single molecule kinetics might be useful.

Finally, it will be interesting to study the effect of PCNA on the fidelity of Polκ on the hotspot sequence. For the archaeal DinB homologue Dhb, it has been seen that PCNA improves the fidelity by increasing the efficiency of the correct dNTP addition (Wu Y and Pata JD, manuscript under preparation). It has been hypothesized that PCNA changes
the overall conformation of the polymerase, which may in turn lead to an increase in fidelity. It remains to be seen whether such a situation holds true for Polk. If we find that PCNA does alter the fidelity of the polymerase on the hotspot sequence, then it might be very informative to visualize the PCNA-Polk-DNA-dNTP complex.

Crystallographic approach promises the highest resolution, but if unsuccessful, cryo-electron microscopy (Cryo-EM) with single particle reconstruction techniques can be used. Previously negative stain EM has been used successfully to determine the overall orientation of the PCNA-Dbh-DNA-dNTP complex, albeit at a low resolution. Although the size of the complex will be rather small for cryo-EM, recent advances in data collection techniques [183] might be able to circumvent this problem. Further, cryo-negative EM can be used to improve the contrast of the particles while preventing the undesirable dehydration effects of negative stain EM.

5.3. Concluding Remarks

In this dissertation I have worked with two of the six DNA polymerase families. For the C-family polymerase PolC I have described the minimal enzymatic pathway and have identified the minimum requirement for fast processive synthesis, while for the Y-family polymerase Polk I have identified the deletion mechanism used on repetitive deletion hotspot DNA sequences. In both the cases enzyme kinetics provided us with significant mechanistic insights into the polymerase’s workings and my results indicate that although DNA polymerases follow the same overall pathway each polymerase has its own
speciality and needs to be studied separately. These results have set the stage for future structure-function correlation studies for both these polymerases. Moreover, my results provide some interesting suggestions about the function of the replication machinery and in future, it will be informative to study the dynamic interactions of the polymerase with the replisome and impact of such interactions on the polymerase kinetics. In conclusion, work presented in this dissertation significantly increases our knowledge about DNA polymerase function and brings us a step closer to understanding how replication is regulated by these enzymes.
BIBLIOGRAPHY


91. Salinas, F. and S.J. Benkovic, Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate. Proceedings of


APPENDIX 1:


A-1. DESCRIPTION OF CONTRIBUTION

For this manuscript, I helped in designing some of the experiments and analyzed the data. I also helped with X-ray diffraction data collection. I played a major role in editing the final manuscript.
A-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.
A-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 [1]. 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 [2]. 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 [1]. 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 [3-8]).

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 [9]. 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) [10,11]. 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 [12]. 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 [12-14].

We have shown recently that the 15 residue linker connecting the polymerase and LF/PAD domains of Dbh and Dpo4 (Fig. A-1) can alone influence the enzymatic activity and selectivity of these polymerases [15]. 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.
**Figure A-1: Parental Y-family polymerases and chimeric constructs.** 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.
A-4. EXPERIMENTAL PROCEDURES

A-4.1. Protein expression and purification. Dbh expression and purification were performed as described before [16]. Dpo4 and all the chimeric constructs used had C-terminal 6x-His tags and growth and purification steps were performed as described before [15,17].

A-4.2. Primer-template DNA. All DNA substrates used in this study are listed in Table A-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.

A-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 MgCl$_2$, 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 A-1):

\[ y = A(1 - e^{-k_1 t}) + c \]  

(Equation A-1)

where \( A \) is amplitude, \( k_1 \) 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.

**A-4.4. \( K_D^{dNTP} \) and \( k_{pol} \) determination.** Primer extension assays were performed using final concentrations of 4 \( \mu \)M 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 5-2000\( \mu \)M (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 A-1). The observed rates (\( k_1 \)) thus obtained were further plotted as a function of dNTP concentration and then fit to a hyperbolic equation:

\[ k_1 = \frac{k_{pol}[dNTP]}{K_D^{dNTP} + [dNTP]} \]  

(Equation A-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 A-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.
A-4.5. Crystallization and Structure Determination. Complexes were prepared at room temperature by combining 200 μM DbhDpo₄⁺锵啟 Dbh (Fig. A-1) and 240 μM DNA (X-4T-G substrate, Table A-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 Ca²⁺ as the divalent metal ion instead of Mg²⁺. 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 [18]. The structure was solved by molecular replacement using the DbhDpo₄Dbh ternary complex (PDB 4F4Y, [15]) as a search model and was refined using PHENIX [19], alternating with cycles of manual rebuilding in Coot [20]. The geometry of the DNA was analyzed using Curves+ [21] and protein conformations were analyzed using DynDom [22]. 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.
A-5. RESULTS

A-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 [12-14], 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 [13,14,23,24]. Primer extension assays using the 4C-G substrate (Table A-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. A-2A-H and Table A-2). Chimeric polymerases are named by the source enzyme in the order: polymerase core-linker-LF/PAD.
Figure A-2: Single-base deletion efficiency is dependent on the identity of the linker.
Figure A-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 1). Incorporation of dGTP (closed circles) shows the correct product formation while dCTP addition (open circles) 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 A-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>$k_1^{dG}/k_1^{dC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbh</td>
<td>~0.02**</td>
<td>~0.01**</td>
<td>–</td>
</tr>
<tr>
<td>DbhDbhDpo4</td>
<td>0.04±0.02</td>
<td>0.11±0.02</td>
<td>0.36</td>
</tr>
<tr>
<td>Dpo4DbhDbh</td>
<td>0.20±0.04</td>
<td>0.21±0.03</td>
<td>0.95</td>
</tr>
<tr>
<td>Dpo4DbhDpo4</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>Dpo4Dpo4Dbh</td>
<td>1.29±0.09</td>
<td>0.40±0.11</td>
<td>3.23</td>
</tr>
<tr>
<td>DbhDpo4Dpo4</td>
<td>0.92±0.21</td>
<td>0.13±0.03</td>
<td>7.08</td>
</tr>
<tr>
<td>DbhDpo4Dbh</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. A-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 [16,17], Dbh incorporates both nucleotides (correct, dG and incorrect, dC) at near equal rates (Fig. A-2A) while Dpo4 adds the correct nucleotide 8-fold faster than the incorrect nucleotide (Fig. A-2E and Table A-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. A-2, panels F-H; Table A-2) whereas the chimeras containing the Dbh linker incorporate dG up to 3-fold more slowly than dC (Fig. A-2, panels B-D; Table A-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. A-3 and Table A-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 [16,17,25]. 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 1. The 1T-G substrate is used least efficiently in all cases (Fig. A-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. A-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. A-2 and A-3).
Figure A-3: Chimeric polymerases use a template-slippage deletion mechanism.
Figure A-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 A-1).

Fraction of primer DNA extended, as a function of time, is displayed graphically for the different DNA substrates shown: 1T-G (closed squares), 2T-G (open squares), 3T-G (closed triangles) and 4T-G (open triangles). 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.
A-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. A-2E-H and Fig. A-3E-H), but those containing the Dbh-linker do not (Fig. A-2A-D and Fig. A-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. A-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. A-2J). To test this idea, we examined the ability of Dpo4 to extend from a C-C mispair substrate (Table A-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. A-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.

A-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 [15]. 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. A-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. A-4, Table A-3). Here we refer to this chimeric polymerase as DbhDpo4RKSDbh. We were unable to crystallize the complementary chimera, Dpo4DbhKIPDpo4. In the crystal structure, an incoming dCTP is correctly paired to the G in the deletion hotspot sequence (Fig. A-4A) and the template contains an unpaired T four nucleotides 3’ to the templating G (Fig. A-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. DbhDpo4RKSDbh superimposes on Dbh (PDB code 3BQ1; [16]) with an RMSD of 5.08 Å (Fig. A-4C), even though the polymerase and LF/PAD domains individually align with RMSDs of 1.62 and 0.5 Å, respectively. In contrast, DbhDpo4RKSDbh superimposes on Dpo4 (PDB code 3QZ7; [17]) with an overall RMSD of 1.34Å (Fig. A-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 DbhDpo4RKSDbh (Fig. A-4C). In the structure of DbhDpo4RKSDbh, 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 [16,26,27].
Figure A-4: DbhDpo$^{RKS}$Dbh adopts a Dpo4-like conformation.
Figure A-4: DbhDpo4<sup>RKS</sup>Dbh adopts a Dpo4-like conformation. Ternary complex of DbhDpo4<sup>RKS</sup>Dbh chimeric polymerase (A) is shown with domains color coded as in Fig. A-1 (linker is shown in yellow-orange). Both DNA substrate (white) and incoming dCTP (yellow) are shown as sticks and the divalent Ca<sup>2+</sup> 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 2F<sub>b</sub>-F<sub>c</sub> map calculated from the final refined structure is shown contoured at 1.5σ (grey mesh). Superposition of DbhDpo4<sup>RKS</sup>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 [22] 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 DbhDpo4<sup>RKS</sup>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 DbhDpo4<sup>RKS</sup>Dbh, allowing selection against ribonucleotides.
Table A-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;RK5&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**

<table>
<thead>
<tr>
<th>Beamline</th>
<th>X25</th>
</tr>
</thead>
<tbody>
<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>120.19, 103.07, 53.05</td>
</tr>
<tr>
<td></td>
<td>β (°) 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/sigma(I)</td>
<td>23.2 (2.9)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>28.7 - 2.4</th>
</tr>
</thead>
<tbody>
<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>
Contact between the LF/PAD and β2-3 loop brings the substrates into alignment at the active site for catalysis, by constraining the width of the nascent basepair binding pocket. This is consistent with the DbhDpo4Dbh chimera adding a C very efficiently to the 4T-G variant of the deletion hotspot sequence (Fig. A-3H). In comparison to the Dbh ternary complex structure (Fig. A-4E), the chimera (Fig. A-4F) has two metal ions bound at the active site and the ribose of the dNTP sits directly on top of the steric gate residue (Phe 12). Additionally the α-phosphate of the nucleotide is 3.7 Å from the 3’-OH of the primer terminus, compared to 5.3 Å in the parental Dbh structure. Thus, the active site in this structure has the characteristics associated with correct dNTP incorporation.

Interestingly, the unpaired T in the template DNA is not in an extrahelical conformation. Instead, it intercalates into the DNA duplex (Fig. A-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 [15].

A-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
A-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 A-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 A-4). Comparing the rates of correct vs. incorrect nucleotide incorporation ($k_{pol}^{dGTP}/k_{pol}^{dCTP}$), 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 A-4).

Nucleotide binding affinity also contributes to fidelity, but in a more complex way (Table A-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. DbhDpo4$^{RKS}$Dbh; 0.2 vs. 0.1 for Dpo4 vs. Dpo4Dbh$^{KIP}$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.
Table A-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>$K_D^{dNTP}$</th>
<th>$k_{pol}$</th>
<th>Efficiency*</th>
<th>$K_D^{dNTP}$</th>
<th>$k_{pol}$</th>
<th>Efficiency*</th>
<th>$K_D^{dGTP}$</th>
<th>$k_{pol}^{dGTP}$</th>
<th>$k_{pol}^{dCTP}$</th>
<th>Fidelity†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_D^{dNTP}$</td>
<td>$k_{pol}$</td>
<td>Efficiency*</td>
<td>$K_D^{dNTP}$</td>
<td>$k_{pol}$</td>
<td>Efficiency*</td>
<td>$K_D^{dCTP}$</td>
<td>$k_{pol}^{dGTP}$</td>
<td>$k_{pol}^{dCTP}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dbh</td>
<td>822±397</td>
<td>0.17±0.04</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></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dpo4Dbh^KIP-Dpo4</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></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DbhDpo4^RK5-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></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: Efficiency = $k_{pol}/K_D^{dNTP}$; †: Fidelity = Efficiency dGTP/Efficiency dCTP
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.

A-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 A-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 [28]. Interestingly, the linker sequences affect single-base deletion and base-substitution fidelity in opposite directions: the Dpo4 linker increases the base substitution frequency [15] but decreases the single-base deletion fidelity (Fig. A-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>RKS</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. A-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. A-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 [3, 29]. For pol kappa, extension from the mismatch involves realigning the slipped DNA strands [29] 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 [15], 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 [30]. 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.

A-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.
A-7. REFERENCES


