



University of
Sheffield

**Mechanistic understanding of multitasking
apicoplast DNA polymerase from
*Plasmodium falciparum***

By Anamika Kumari

Submitted for the degree of Doctor of Philosophy

School of Biosciences
The University of Sheffield, United Kingdom

April 2026

DECLARATION

I, Anamika Kumari, confirm that this Thesis and the work presented herein is my own work, except where duly acknowledged in the text and/or publication declaration. I am aware of the University Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university.

Anamika Kumari, April 2026

ABSTRACT

Plasmodium falciparum is an eukaryotic pathogen responsible for the majority of malaria-related fatalities. *Plasmodium* belongs to the phylum Apicomplexa and, like most members of this phylum, *Plasmodium* contains an essential non-photosynthetic plastid called the apicoplast. Apicoplast harbours its own genome that is replicated by a streamlined replisome built around a single A-family DNA polymerase, apPol, which is expected to perform both replicative and lesion bypass (or translesion) synthesis. While preliminary biochemical and structural studies have provided an overall characterization of apPol, the mechanism behind its multitasking behaviour remains unknown. Building on this problem, this work combines transient-state kinetics with single-particle cryo-electron microscopy (cryo-EM) to define how apPol's catalytic cycle, substrate affinities and conformational landscape enable the enzyme to multitask. Kinetic analyses show that apPol exhibits weak DNA and dNTP binding, low efficiency and processivity on undamaged templates and an intrinsically error-prone bypass of oxidative lesions, implying that additional, as-yet-uncharacterized accessory factors are required to achieve faithful apicoplast genome duplication. Our cryo-EM structures capture five pre-chemistry states of apPol showing that apPol can bind DNA and nucleotide independently. Interestingly, apPol forms the nascent base pair with the active site in an open configuration, which might facilitate the lesion bypass activity. The nascent base pair then undergoes large thumb-driven movements and the formation of a more restrictive active site to reach a catalytically poised state. Moreover, our structures reveal a novel checkpoint that ensures the integrity of the nascent base pair. Taken together, this work provides a mechanistic framework for how a single organellar DNA polymerase integrates replication and lesion bypass functionalities. The structural and kinetics experiments described here set the stage for future structure-function dissection of apPol. Other than expanding our understanding of the principles of genome maintenance in a reduced organelle, the structures reported in this work can facilitate downstream structure-based inhibitor design targeting apPol.

ACKNOWLEDGEMENTS

I would like to dedicate this thesis to my father, without whom I would never have reached this point. This work is a reflection of the values, strength, and belief in me that he has given throughout my life.

My PhD would certainly not have been possible without the support of a huge number of people, and I am deeply grateful to all of them. First and foremost, I would like to thank my supervisor, Dr. Indrajit Lahiri, for giving me the opportunity to work on organellar replication proteins from *Plasmodium falciparum*. It is hard to believe how long it has been since I first joined this lab, and even harder to express how grateful I am for your guidance, knowledge, ideas, and enthusiasm over these six years. It has truly been a privilege to work with you. Thank you for always being supportive and encouraging, and for teaching me how to become a more focused and productive version of myself. I have improved so much during my PhD, and that is in no small part because of you. Most of all, thank you for always reassuring me that I was doing a good job, your encouragement has helped me overcoming self-doubt at every point during these years. It is difficult to imagine this journey without your support. I would also like to thank the BBSRC for funding my PhD project.

I am equally grateful to my second supervisor, Dr. Tim Craggs, for giving me the opportunity to learn single-molecule experiments and for all the insightful discussion and suggestions throughout these years. I would also like to thank my advisor, Prof. Per Bullough, for always being available whenever I needed advice, you have been my academic problem solver on many occasions. A special thank you goes to Dr. Janice Pata (Wadsworth Centre, NY, USA), for being a wonderful collaborator and for generously sharing your expertise and knowledge of DNA polymerases. I would also like to express my gratitude to Dr. Purba Mukherjee (University of York), for helping me in many different ways, especially during the early days of my PhD. My sincere thanks also to Dr. Sevt Tzokov for all the training and support in every aspect of electron microscopy.

I am grateful to all the lab members I have worked with over the years- Theo, Georgia and Simran- and especially to Dr. Anjali Yadav, who has been a fantastic lab mate and friend. I am also thankful to my best friend back in India, Ismat for being constant sources of

emotional support through the highs and lows of this journey. Aruna, thank you for always staying in touch with me despite the distance. Your presence has given warmth, comfort and long talks felt close to home during these years. My PhD experience was also made brighter by the many wonderful friends I met along the way. To Devki, Anirudh, Raj, Ankita, Najma, Hannah, Becky and everyone in the EM group- thank you for the support, lab help, lunches, coffee breaks, conversations, and laughter.

On a more personal note, I would like to thank my parents for their unconditional love and support throughout my life. I would also like to thank my brothers, Aditya and Abhishek, for always encouraging me and for finding ways to make me laugh! This PhD would not have been possible without the support of my family.

Last, but certainly not least, I would like to thank my husband, Ajay, for your love, patience, and constant support throughout these years. Thank you for standing by me through all the ups, downs, and changes along the way. I feel very fortunate to have you in my life, and I look forward to continuing to build our lives together.

PUBLICATIONS

Anamika Kumari, Anjali Yadav, and Indrajit Lahiri, 2022. **Transient State Kinetics of *Plasmodium falciparum* Apicoplast DNA Polymerase Suggests the Involvement of Accessory Factors for Efficient and Accurate DNA Synthesis.** *Biochemistry* 61 (21), 2319-2333. <http://doi.org/10.1021/acs.biochem.2c00446>

Anamika Kumari, Theodora Enache, Timothy D Craggs, Janice D Pata, Indrajit Lahiri, 2025. **Structural basis of multitasking by the apicoplast DNA polymerase from *Plasmodium falciparum*,** *Nucleic Acids Research*, Volume 53, Issue 19, gkaf1005, <https://doi.org/10.1093/nar/gkaf1005>

PUBLICATIONS DECLARATION

Kumari et al., 2022 and Kumari et al., 2025 are published under a Creative Commons CC BY 4.0 licence. This permits copying, distribution, adaptation and remixing, providing the work is cited. <https://creativecommons.org/licenses/by/4.0/deed.en>

Chapters 2, 3, 4, and 5 represent expanded versions of Kumari et al., 2022 (chapter 2, 3 and 5) and Kumari et al., 2025 (chapters 4 and 5).

The work in Kumari et al., 2025 was performed in collaboration with Dr. Janice Pata and Theodora Enache. And the work in Kumari et al., 2022 was performed in collaboration with Dr. Anjali Yadav. The majority of the experimental work (with the exception of phosphate release assay in section 3.2.5, which was performed by Anjali and primer extension assays and analysis in chapter 4, which was performed in collaboration with Theodora) was performed by Anamika Kumari. This includes all of the protein purifications, kinetic assays, gel quantitation, electron microscopy, data processing, model building and analysis.

Further, all the figures presented in this thesis were designed and produced by Anamika Kumari and edited by Dr. Indrajit Lahiri. All the movies are produced in pyMOL by Dr. Janice Pata.

For the most part, the text herein is either original to this thesis or taken from the published works listed above and was written by Anamika Kumari in collaboration with Dr. Indrajit Lahiri.

ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
%	Percent
Å	Ångström
A	Adenine
A ₂₈₀	Absorbance at 280 nm
apDNA	apicoplast DNA
Asp	Aspartate
ATPases	Adenosine triphosphatases
BSA	Bovine serum albumin
°C	Degree Celsius
cryoSPARC	cryo-EM Single Particle Ab-Initio Reconstruction and Classification
Ca ²⁺	Calcium ion
CaCl ₂	Calcium chloride
Cm	Centimeter
CTF	Contrast transfer function
CV	Column Volume
dTTP	Deoxythymidine triphosphate
dNTP	Deoxyribonucleoside triphosphate
D	Aspartate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
eLBOW	electronic Ligand Builder and Optimization Workbench
E	Glutamate
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EMDB	Electron microscopy data bank
FAM	6-Carboxyfluorescein
h	Hour

IR	Infrared
ISOLDE	Interactive Structure Optimisation by Local Density Exploration
kb	Kilobase
kDa	Kilodalton
kV	Kilovolt
LASER	Light amplification by stimulation emission of radiation
ml	Millilitre
mM	Millimolar
M	Molar
MeA	Metal A
MeB	Metal B
MeC	Metal C
Min	Minute
Mg ²⁺	Magnesium ion
MgCl ₂	Magnesium chloride
nM	Nanomolar
nm	Nanometer
N	Asparagine
NTase	Nucleotidyltransferase
Ni ²⁺	Nickel ion
NaCl	Sodium chloride
-OH	Hydroxyl group
ODN	Oxidatively damaged nucleotide
pETDuet1	Plasmid expression vector pETDuet-1
pH	Potential of hydrogen
pKa	Acid dissociation constant
pyMOL	Python-based Molecular Graphics System
PDB	Protein Data Bank
PPi	Inorganic pyrophosphate
Pol	Polymerase
Pi	Inorganic phosphate
PSF	Point spread function
P/T	Primer/Template

Q	Glutamine
rNTP	Ribonucleoside triphosphate
RNA	Ribonucleic acid
ROS	Reactive oxygen species
Rosetta 2 (DE3)	Rosetta 2 DE3 <i>E. coli</i> expression strain
RQF-3	Reaction Quench Flow-3
ROS	Reactive oxygen species
ssDNA	Single-stranded DNA
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE	Standard error
S _N 2	Bimolecular nucleophilic substitution
SPA	Single-particle analysis
SSB	Single-stranded DNA binding protein
T	Thymine
T4	Bacteriophage T4 (one of four lytic bacteriophages)
T7	Bacteriophage T7 (one of seven lytic bacteriophages)
TEM	Transmission electron microscopy
TEV	Tobacco etch virus
TLS	Translesion synthesis
Tris-Cl	Tris hydrochloride
μM	Micrometer
UCSF	University of California, San Francisco
UV	Ultraviolet
Y	Tyrosine

Table of Contents

DECLARATION	I
ABSTRACT.....	II
ACKNOWLEDGEMENTS	III
PUBLICATIONS.....	V
PUBLICATIONS DECLARATION	VI
ABBREVIATIONS	VII
Table of Contents	X
List of figures.....	XIII
List of tables.....	XVI
List of movies.....	XVII
1. INTRODUCTION	1
1.1 DNA Replication	1
1.1.1 Key proteins of a minimal replisome.....	2
1.1.2 Accuracy during DNA replication.....	4
1.1.3 Errors during DNA replication	4
1.1.4 Replication fork (The trombone model)	5
1.2 DNA Polymerase	8
1.2.1 DNA polymerase classification	8
1.2.2 DNA polymerase structure	13
1.2.3 DNA polymerase catalytic pathway and reaction mechanism	15
1.2.4 Multiple DNA polymerases within a system.....	19
1.2.5 Conformational dynamics of DNA polymerases	23
1.2 Malaria and Plasmodium	24
1.4 Apicoplast and its origin.....	26
1.4.1 The apicoplast genome	26
1.4.2 The apicoplast replisome	27
1.4.3 Apicoplast DNA polymerase (apPol)	27
1.5 Techniques used to study enzymatic mechanisms of DNA polymerases	28
1.5.1 Kinetics of DNA polymerase.....	29
1.5.2 Cryo-EM	33
2. MATERIALS & METHODS	44
2.1 DNA substrate	44
2.2 Software	44
2.3 apPol Constructs	45

2.4	Cloning, Overexpression, and Purification of apPol	49
2.5	Determination of optimal conditions for apPol activity	50
2.6	Single Nucleotide Primer Extension Assays	53
2.7	Double Mixing Experiment to Determine the apPol-DNA Binary Complex Dissociation Rate.....	57
2.8	Multiple Nucleotide Primer Extension Assay.....	58
2.9	Primer Degradation Assay	61
2.10	Determination of Kinetic Constants from Data Fitting Using Numerical Integration	61
2.11	Negative stain grid preparation	62
2.12	CryoEM sample preparation	63
2.13	Cryo-EM grid preparation.....	63
2.14	Cryo-EM imaging and 3D reconstruction	64
2.15	Atomic model building and refinement	67
2.16	Structure and sequence analysis.....	68
3.	Transient State Kinetic of Apicoplast DNA Polymerase from Plasmodium Falciparum.....	70
3.1	INTRODUCTION	70
3.2	RESULTS	71
3.2.1	apPol purification.....	71
3.2.2	Optimal reaction conditions for apPol.	72
3.2.3	Kinetic Mechanism of Correct Nucleotide Incorporation by apPol.	74
3.2.4	Phosphodiester Bond Formation is Followed by Slow Dissociation of the Product DNA.....	74
3.2.5	ApPol Performs Pyrophosphorolysis Only in the Presence of High Concentration of Pyrophosphate.	77
3.2.6	apPol Binds Weakly to DNA Due to Rapid Dissociation of the Binary Complex.....	81
3.2.7	apPol Incorporates Nucleotides with Low Efficiency and Processivity.....	85
3.2.8	apPol Translocates Rapidly along the DNA Substrate.	90
3.2.9	Kinetics of Lesion Bypass by apPol	94
3.2.10	apPol Can Bypass Oxidatively Damaged Nucleotides (ODNs)	94
3.2.11	apPol Preferentially Adds dATP Opposite an Abasic Site, but the Bypass Is Inefficient.....	96
3.2.12	apPol Performs Distributive Error-Prone Synthesis opposite 8-Oxo-dGMP	102
3.2.13	apPol's Proofreading Activity Does Not Discriminate between Accurate and Error-Prone Bypass of 8-oxo-dGMP	110
4.	Structural Basis of Multitasking DNA Polymerase from Plasmodium Falciparum	113
4.1	INTRODUCTION	113
4.2	RESULTS	114
4.2.1	apPol complex behaves differently under negative stain and vitrified ice.	114
4.2.2	apPol binds DNA and nucleotide independent of each other	118

4.2.3 apPol-DNA binary complex base pairs with the incoming nucleotide with the fingers open.....	127
4.2.4 apPol-specific O ₁ helix extension acts as a nascent base pair checkpoint.....	134
4.2.5 Catalytically poised apPol has fingers closed around the nascent base pair.....	140
4.2.6 Steric gate residue in apPol.....	146
5. DISCUSSION	150
5.1 Kinetic signature and multitasking behaviour of apicoplast DNA polymerase...150	
5.1.1. Kinetics of apPol mediated DNA synthesis and its implication for apicoplast genome duplication.....	151
5.1.2 Translesion synthesis by apPol might be a source of mutation for the apicoplast genome.....	152
5.1.3 Pre-steady state kinetics define the mechanism of processive synthesis.....	154
5.2 Structural and mechanistic insights into multitasking by apicoplast DNA polymerase	155
5.2.1 apPol uses novel mechanism to perform TLS activity.	156
5.2.2 apPol uses two fidelity checkpoints to perform high fidelity replicative synthesis	160
5.2.3 ApPol thumb is a master manipulator of DNA position.....	161
5.2.4 Why do the fingers close?.....	162
5.2.5 What happens after bond formation?	163
5.3 Conclusions.....	164
5.4 Future Perspectives.....	165
BIBLIOGRAPHY	168
Appendix: A1.....	180
Appendix: A2.....	181
Appendix: A3.....	182
Appendix: A4.....	183
Appendix: A5 - Movie description.....	184

List of figures

- Figure 1.1: Replication fork (Trombone model)
- Figure 1.2: Classification of DNA polymerase
- Figure 1.3: Right-handed architecture of DNA polymerase
- Figure 1.4: Minimum catalytic pathway for DNA polymerase
- Figure 1.5: Two metal ion mechanism of nucleotide incorporation
- Figure 1.6: Multiple DNA polymerases in different biological system
- Figure 1.7: Active site of DNA polymerases
- Figure 1.8: Open and closed fingers conformations of T7 DNA polymerase
- Figure 1.9: Steady state vs pre-steady state
- Figure 1.10: Types of electrons scattering in cryo-EM
- Figure 1.11: An Argand diagram of unscattered and scattered waves in cryo-EM
- Figure 1.12: Contrast transfer function at two different defocus values
- Figure 1.13: Point spread function (PSF) in electron microscope
- Figure 2.1: apPol construct
- Figure 2.2: pET Duet-1 vector map
- Figure 2.3: Primer extension assay
- Figure 2.4: DNA trap experiment
- Figure 2.5: Double mixing experiment
- Figure 2.6: Cryo-EM sample preparation
- Figure 3.1: SDS-PAGE of apPol
- Figure 3.2: Reaction condition optimization
- Figure 3.3: Detection of the presence of slow post-chemistry step in apPol's catalytic cycle.
- Figure 3.4: Pyrophosphorolysis catalysed by apPol
- Figure 3.5: Phosphate release assay and nucleotide analogue
- Figure 3.6: Determination of K_D^{DNA} and active site titration
- Figure 3.7: Global fit of the minimal kinetic pathway governing the correct nucleotide incorporation by apPol opposite to an undamaged template.
- Figure 3.8: Confidence contour analysis for the kinetics of correct nucleotide incorporation opposite undamaged template by apPol.
- Figure 3.9: Kinetics of processive synthesis by apPol

Figure 3.10: Confidence contour analysis for the kinetics of processive synthesis on undamaged template by apPol

Figure 3.11: Translesion synthesis by apPol in the presence of multiple nucleotides

Figure 3.12: Single nucleotide incorporation by apPol against abasic site

Figure 3.13: Global fitting of the time courses from abasic site bypass by apPol

Figure 3.14: Confidence contour analysis for the kinetics of abasic site bypass by apPol

Figure 3.15: Single nucleotide extension against 8-oxo-dGMP

Figure 3.16: Global fitting of the time courses against 8-oxo-dG addition

Figure 3.17: Confidence contour analysis for the kinetics of 8-oxo-dGMP bypass by apPol

Figure 3.18: Primer extension assay performed using DNA substrates S5 and S6

Figure 3.19: Exonuclease assay of apPol^{WT}

Figure 4.1: apPol complex in Negative stain and vitreous ice

Figure 4.2: Atomic model of apPol complex

Figure 4.3: apPol's activity in presence of CHAPSO

Figure 4.4: Cryo-EM image processing and 3D reconstruction for dataset1 and dataset2.

Figure 4.5: Different conformational states of apPol's complex.

Figure 4.6: apPol-DNA binary complex (binary1)

Figure 4.7: Binary complex of apPol fitted model in cryo-EM maps

Figure 4.8: O-helix movement in different conformational states

Figure 4.9: DNA movement towards catalytic aspartates

Figure 4.10: Primer strand movement with respect to R390

Figure 4.11: DNA movement with respect to thumb domain

Figure 4.12: Distance between 3'primer end and D579

Figure 4.13: Closeup of the active site region of ternary1

Figure 4.14: DNA-apPol contacts in ternary2

Figure 4.15: Closeup of the active site region of ternary2

Figure 4.16: O₁ checkpoint in ternary2

Figure 4.17: Triple mutant primer extension assay

Figure 4.18: Protein sequence alignment of selected A-family DNA polymerases

Figure 4.19: Catalytically poised active site of apPol.

Figure 4.20: DNA-apPol contacts in ternary3

Figure 4.21: Thumb domain interactions with DNA minor groove

Figure 4.22: Q387 and K344 sensing the minor groove geometry

Figure 4.23: Single nucleotide primer extension assay to validate steric gate residue

Figure 4.24: Single nucleotide primer extension assay to validate steric gate residue at different time points

Figure 5.1: apPol's pre-insertion site is not accessible to T₀ base

Figure 5.2: apPol backtracks during the pre-chemistry steps

List of tables

Table 1: DNA substrates.

Table 2: Cryo-EM map and atomic model refinement statistics.

Table 3: Kinetic parameters of different DNA polymerases.

Table 4: Kinetic parameters 8-oxo-dG bypass by A-family DNA polymerase

List of movies

Movie 1: Overall domain motion of apPol as it traverses the pre-chemistry steps.

Movie 2: Motion of the thumb as apPol traverses the pre-chemistry steps.

Movie 3: Corkscrew motion of the DNA as apPol traverses the pre-chemistry steps.

Movie 4: Formation and dissolution of the pre-insertion checkpoint.

Movie 5: Interaction of the template strand with the palm domain.

Movie 6: DNA slippage with respect to the thumb.

1. INTRODUCTION

1.1 DNA Replication

DNA replication is the central process for genome duplication in all organisms. It is fundamental for faithful transfer of genetic information from one generation to the next during cell division. After the discovery of the DNA double-helix structure in 1953 [1] [2] [3], it took five more years to prove that replication takes place in a semi-conservative manner, such that each strand of the duplex DNA acts as a template (parental strand) for the synthesis of a new (daughter) strand [4]. This process is evolutionarily conserved throughout different species. Replication is carried out through the coordinated action of a dedicated, multi-component replication machinery known as the replisome, which carries out the highly regulated process of DNA replication. The replisome consists of many different proteins with specific jobs, all working in a highly coordinated manner. Even the simplest replisome characterised to date, the bacteriophage T7 replisome, requires a minimum of four proteins [5] [6], while the *E. coli* replisome contains as many as 20 proteins [7]. Coordinated action of each of these proteins is crucial for rapid and accurate replication.

1.1.1 Key proteins of a minimal replisome

A replisome comprises several proteins, which interact with each other and coordinate their activities for efficient and accurate replication. Four proteins are necessary and sufficient to carry out replication in a minimal T7 replisome. These proteins are helicase, primase, single-stranded DNA-binding protein (SSB), and DNA polymerase. A brief description of the four enzymes is provided below.

- i. **DNA helicase:** DNA helicases are ATPases that translocate along the DNA, and as the helicase moves along the DNA, it unwinds the two DNA strands by breaking the hydrogen bonds between the base pairs [8]. There are several families of helicases, and the directionality of translocation (5' to 3' or 3' to 5') depends on the individual family. Interestingly, while the prokaryotic replicative helicase translocates in a 5' to 3' direction, the archaeal and eukaryotic replicative helicases translocate from 3' to 5' [9].
- ii. **Primase:** The two single strands generated by the helicase act as templates for new DNA synthesis. The primase synthesizes a short RNA primer complementary to the template strand. This primer is essential for DNA polymerases to initiate DNA synthesis. A primase has three functional domains: a catalytic domain responsible for RNA primer synthesis, a helicase binding domain that tethers the primase with the helicase, and a zinc binding domain that recognizes the ssDNA and stabilizes primase activities [10, 11].

- iii. Single-stranded DNA binding protein:** SSBs are responsible for binding and protecting single-stranded DNA (ssDNA) from nucleases present in the cellular environment and prevent it from forming secondary structures. These single-stranded DNA are created by helicase while unwinding DNA during replication. SSB binds very tightly to single stranded DNA to protect ssDNA from damage. Despite the tight binding, SSB can reposition itself so that ssDNA can be accessed by other proteins involved in replication [12].
- iv. DNA Polymerase:** DNA dependent DNA polymerase or DNA polymerase is the central enzyme of the replisome. Using the parental DNA strand as a template, the polymerase incorporates complementary nucleotides into the nascent strand, leading to the semi-conservative nature of DNA replication. A DNA polymerase catalyses phosphodiester bond formation between the 3' hydroxyl (-OH) group of the DNA primer strand and an incoming deoxynucleoside triphosphate (dNTP) through a phosphoryl transfer reaction. The chemical reaction involves an acid-base catalysis and employs two or three divalent metal ions [13] [14]. Most DNA polymerases also harbour an intrinsic proofreading activity, mediated by a 3'-5' exonuclease domain that excises mismatched nucleotides generated by the polymerase active site, thereby enhancing overall fidelity of DNA synthesis [15].

1.1.2 Accuracy during DNA replication

Accurate DNA replication is crucial for faithful propagation of the genetic information. DNA polymerases ensure replication accuracy, or fidelity through multiple factors including 1) correct **dNTP selection** during nucleotide incorporation. An accurate DNA polymerase can have an error frequency as low as 10^{-5} errors per base per round of replication; 2) the **proofreading activity** of a DNA polymerase, which removes rare misincorporations can further reduce the error frequency by a factor of 10^{-2} ; and 3) post-replicative **repair pathways** can correct residual errors, further decreasing misincorporation by about 10^{-3} errors per base per round of replication. Together, these mechanisms can lead to an overall replication fidelity of approximately 10^{-10} errors per base per replication cycle, making the process extraordinarily accurate [16, 17].

1.1.3 Errors during DNA replication

DNA replication is a high-fidelity process that is safeguarded by multiple checkpoints, including the intrinsic fidelity of a DNA polymerase complemented with the 3'–5' exonuclease activity of DNA polymerases and post-replicative DNA repair pathways such as mismatch repair and base excision repair. These systems correct the vast majority of misincorporated nucleotides; however, any error that escapes detection can be converted into a permanent mutation in subsequent rounds of replication [18]. To maintain genome stability, the replisome must therefore sense, bypass, or repair lesions encountered at the fork to minimize replication stalling and incomplete chromosome duplication. DNA lesions can arise from both environmental and endogenous sources. Reactive oxygen species, ionizing and ultraviolet radiation, and numerous chemical mutagens can oxidize, alkylate, or crosslink

bases, as well as induce single- and double-strand breaks, all of which perturb the template prior to or during replication [19]. In parallel, the DNA polymerase itself can be a source of replication errors, introducing mispairs through base tautomerisation, non-Watson–Crick pairing, and template strand slippage events that generate insertions or deletions, especially within repetitive sequences [20, 21]. If not efficiently corrected, the accumulation of such lesions and misincorporations compromises genomic integrity and contributes to cellular dysfunction, genome instability syndromes, and carcinogenesis [22].

1.1.4 Replication fork (The trombone model)

The two strands of DNA are anti-parallel to each other, and DNA synthesis proceeds exclusively in a 5' to 3' direction because DNA polymerases can only add nucleotides to the 3' end of the growing strand. To maintain the directionality of DNA synthesis and the unidirectional movement of the replication fork, replication cannot occur continuously on both template strands (Figure 1.1). DNA synthesis proceeds continuously on the 3' to 5' strand, termed the leading strand. Synthesis on the template running in the 5' to 3' direction occurs discontinuously, producing short DNA fragments known as Okazaki fragments. This template strand is called the lagging strand [23, 24].

To maintain unidirectional replication fork progression, it is essential for the overall directionality of both leading and lagging strand synthesis to be the same. This is achieved by the lagging strand DNA template forming a loop at the replication fork, known as the replication loop. This looping arrangement allows the lagging strand polymerase to synthesize short fragments in the same overall direction as fork progression. The dynamic looping and release of DNA during lagging strand synthesis resembles the extension and

contraction of a trombone, and this model of replication fork progression is called the **trombone model** of replication (Figure 1.1) [25].

Despite fundamentally different modes of synthesis on leading strands and lagging strands, replication remains a highly coordinated process. This coordination is achieved by the physical tethering of the DNA polymerase synthesizing both strands [7]. Such coordination has been demonstrated in diverse systems, including *E. coli*, T4, and T7 bacteriophages, using different approaches like gel-based assays and electron microscopy [26, 27]. For over four decades, it has been proposed that two polymerases function simultaneously at the replication fork—one dedicated to the leading strand and the other to the lagging strand, a concept first proposed by A. Kornberg in 1992 [7, 27]. More recently, evidence from several systems has suggested that up to three polymerases might be operating at a single replication fork, providing additional flexibility and efficiency (a measure of how effectively the enzyme converts substrate into product during DNA synthesis, which depends on two parameters the nucleotide binding affinity and the rate of chemistry). Other replisome components are also recruited to the replication fork, where they act in a highly coordinated manner [28, 29] [30].

Despite extensive research, the detailed molecular mechanisms of replisome coordination, DNA polymerase exchange, and replication fork progression remain poorly understood. DNA replication is an extraordinarily complex and dynamic process and studying the replisome as a complete functional assembly remains technically challenging even in a relatively simple model system such as the bacteriophage T7. In this dissertation, I aim to elucidate the mechanistic aspects of DNA polymerase, the key enzyme of the replisome, and its role in organellar replication.

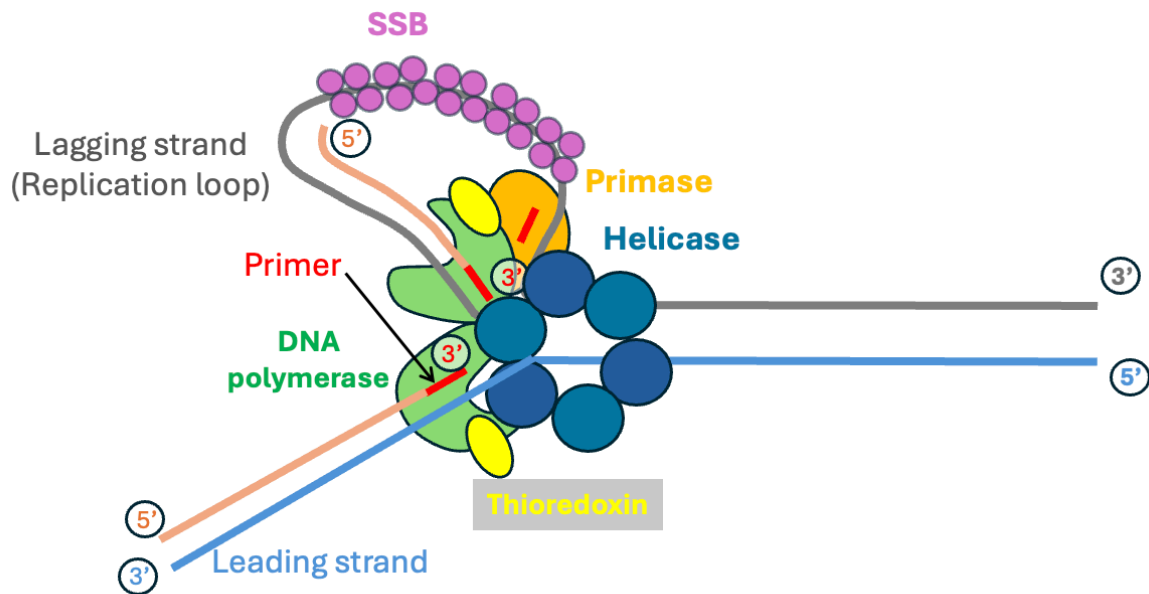


Figure 1.1: Replication fork (Trombone model). Schematic of replication fork based on cryo-EM structure of bacteriophage T7 replisome (EMD: 8565). The helicase (dark blue spheres) encircles the leading-strand template and unwinds the parental DNA duplex, while the primase domain (orange) synthesizes short RNA primers (red) on the lagging-strand template. Unlike most replisomes, in the T7 gp4 the helicase and primase domains are fused together in the same polypeptide but for simplicity I have shown only the active primase domain. A DNA polymerase–thioredoxin complex (green with yellow) extends primers on both leading (blue) and lagging (grey) strands; the lagging strand is organised into a replication loop to allow simultaneous synthesis in the 5'→3' direction. Thioredoxin acts as the processivity factor in the T7 system and increases the number of nucleotides added by the DNA polymerases per DNA binding event. Single-stranded DNA ahead of lagging-strand synthesis is coated by single-stranded DNA-binding protein (SSB, purple). Directionality of each strand is indicated by 5' and 3' labels.

1.2 DNA Polymerase

DNA polymerase is the enzyme that catalyses the chemical step of phosphodiester bond formation, adding the nucleotides to the 3' end of the daughter (primer) strand complementary to the nucleotide sequence of the template strand. Enzymes can be broadly classified into six categories based on the types of reaction they catalyse, and DNA polymerases belong to the transferase class, which comprises of enzymes that transfer a functional group from one molecule to another [31]. More specifically, DNA polymerases carry out nucleotidyl-transfer reaction that results in phosphodiester bond formation between the 3' hydroxyl (-OH) group of DNA primer terminus and the α -phosphate of the incoming deoxynucleoside triphosphate, constituting the fundamental enzymatic reaction underlying DNA synthesis [13].

1.2.1 DNA polymerase classification

Based on their functions, DNA polymerases can be classified into three categories:

- i. **Replicative polymerase** – These polymerases are responsible for copying most of the genetic material. Typically, they are highly accurate and fast enzymes [32]. While replicative polymerases can rapidly copy undamaged DNA, they often stall on encountering DNA damage (lesion) [33]. The very structural features that enable high-fidelity synthesis on an undamaged template contribute to their inability to accommodate damaged bases (discussed later), leading to stalling at sites of DNA damage. Most replicative DNA polymerases possess two enzymatic activities: a

nucleotidyl transferase or polymerization activity and 3' to 5' exonuclease (proofreading) activity. The proofreading activity allows a polymerase to detect and remove misincorporated nucleotides, thereby correcting errors during synthesis and contributing to the overall replication fidelity.

ii. Repair polymerase –Despite the proofreading activity of a replicative polymerase, some errors may persist in the newly synthesized DNA, including mispaired nucleotides, single nucleotide insertions or deletions, or the incorporation of damaged nucleotides. In addition, lesions can arise spontaneously within the DNA double helix through endogenous (for instance, reactive oxygen species (ROS) generated as a metabolic byproduct) or exogenous (for instance, UV radiation or IR radiation) sources. These types of damages are predominantly corrected by DNA repair pathways [34]. Cells employ five major classes of repair mechanisms to preserve genomic integrity: base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), non-homologous end-joining (NHEJ), and homologous direct repair (HDR) [34] [35]. Each pathway involves multiple specialized proteins, but all repair pathways employ specific DNA polymerases called repair polymerases to fill in or replace damaged regions of DNA. Typically, these polymerases synthesize short stretches of DNA to restore the correct sequence [34]. Depending on the type of repair, the polymerase may perform strand-displacement DNA synthesis.

iii. Translesion synthesis (TLS) polymerase – When a replicative polymerase encounters a lesion on the template strand, DNA synthesis can stall. If this blockage is not resolved, it can lead to replication fork collapse, double-strand break, and eventually cell death [36]. To prevent such a catastrophic outcome, cells employ a

tolerance mechanism known as translesion synthesis (TLS), in which the stalled replicative polymerase is transiently replaced by a specialized TLS polymerase. TLS polymerases can efficiently bypass lesions, but the accuracy of the bypass depends on the identity of the lesion and the TLS polymerase [37-39]. While TLS polymerases enable replication to continue past lesions, they are generally error-prone when copying undamaged DNA [40, 41].

Overall, TLS polymerase serves as essential components of the DNA damage tolerance network, balancing the trade-off between continued DNA synthesis and mutational risk. Without this mechanism, cells would face severe genomic instability and replication-associated lethality [40, 41]. Mutagenesis by error-prone TLS polymerases is a major driver of genetic diversification and evolution of antimicrobial resistance [42, 43].

Based on the domain fold of their catalytic core DNA dependent DNA polymerases are classified into three superfamilies and further classified into six families based on their sequence similarity. The three superfamilies are the following:

1. **Classical polymerases** - These polymerases resemble the most common “right-handed” fold, with the polymerase domain subdivided into the fingers, thumb, and palm subdomains. Perhaps the most well-known member of this superfamily is the Klenow fragment of bacterial DNA polymerase I. The palm domain consists of an anti-parallel β sheet flanked by α helices. The β sheet houses the two or three conserved catalytic residues (typically Asp) that coordinate two divalent metal ions

during phosphodiester bond formation. A-family, B-family, and Y-family DNA polymerases belong to this superfamily. (Figure 1.2)

2. **β -nucleotidyl transferase** – This polymerase fold is also known as the NTase fold. Like the classical polymerases, the overall shape of the protein resembles a right hand. The palm domain consists of five mixed β strands rather than purely antiparallel strands connected by two α helices as seen for the classic polymerases [44, 45]. (Figure 1.2).
3. **Two-barrel polymerase** - Archaeal D-family polymerases belong to this superfamily, also known as the double-psi β -barrel or DPBB fold because of the structural homology with the double-psi β -barrel family of RNA polymerases [46]. The archaeal replicative polymerase D (PolD) was the first DNA polymerase where this DNA polymerase fold was discovered [47, 48]. (Figure 1.2).

Based on sequence similarity DNA polymerases are divided into six families:

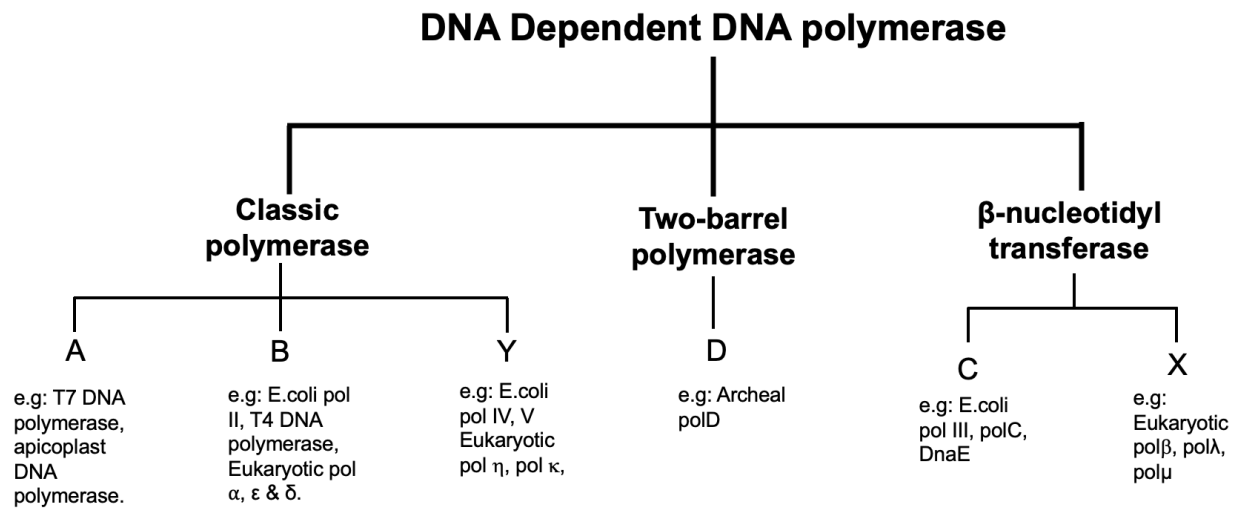


Figure 1.2: Classification of DNA polymerases. Based on sequence similarity. DNA polymerases are grouped into three major superfamilies according to conserved sequence motifs and evolutionary relationships: the classic polymerase family, the two-barrel polymerase family, and the β -nucleotidyl transferase family. The classical polymerase superfamily includes A-, B-, and Y-family polymerases. The two-barrel polymerase superfamily includes family D, while the β -nucleotidyl transferase includes C and X family polymerases. Few examples are also mentioned for each families.

1.2.2 DNA polymerase structure

The first structure of a DNA polymerase was reported in 1985, providing the foundation for understanding the molecular basis of DNA synthesis [49]. Since then, numerous structures from different DNA polymerase families have been determined, both in the *apo* state and in complex with DNA and/or dNTP, greatly expanding our knowledge of their mechanisms of action. The overall architecture of most DNA polymerases (with the exception of the two-barrel polymerases) resembles a right hand (Figure 1.3). The active site for the polymerization activity resides in the palm subdomain, whereas the thumb subdomain contacts and stabilizes the DNA duplex, and the fingers subdomain plays a key role in selecting the correct incoming dNTP and positioning the template strand accordingly [50].

In addition to the polymerization domain, most replicative DNA polymerases contain a 3'-5' exonuclease domain (Figure 1.3) harbouring the proofreading activity, which can increase the accuracy of DNA synthesis by approximately two orders of magnitude. The spatial arrangement of the exonuclease domain relative to the polymerase domain differs between polymerase families and even between polymerases from the same family [32, 51, 52]. However, in all known examples, the proofreading and the polymerization active sites are separated by a substantial distance, and we are only beginning to understand how the primer terminus is transferred between these sites to maximize the fidelity of DNA polymerases [53, 54].

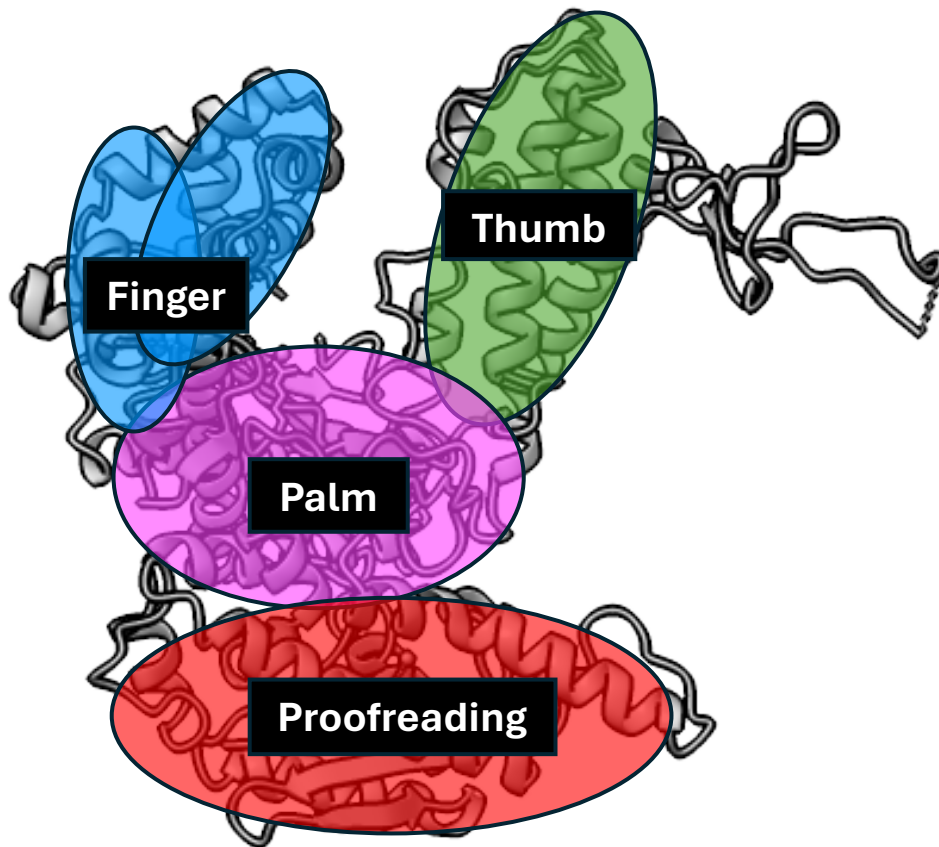


Figure 1.3: Right-handed architecture of DNA polymerase. Cartoon representation of the canonical right-handed fold of a DNA polymerase, with the fingers (blue), palm (magenta), thumb (green), and proofreading exonuclease (red) domains highlighted by colour and overlaid on the T7 DNA polymerase structure (grey) (PDB: 1T7P).

1.2.3 DNA polymerase catalytic pathway and reaction mechanism

Since the discovery of the first DNA polymerase [55] several biochemical and kinetic studies have characterised the catalytic mechanisms of diverse DNA polymerases. Pioneering work from the groups of Johnson, Benkovic, Andreson and Joyce, amongst others, have been instrumental in delineating the enzymatic cycle of nucleotide incorporation [56, 57]. Most DNA polymerases studied to date follow a conserved sequence of events (Figure 1.4).

The cycle begins with binding of the DNA polymerase to primer-template DNA, forming the pre-chemistry binary complex ($\text{Pol} \cdot \text{DNA}_n$) (Figure 1.4, step 1). The incoming dNTP then binds to form the pre-chemistry ternary complex or the initial collision ternary complex, capturing an initial contact of the incoming dNTP with the binary complex ($\text{Pol} \cdot \text{DNA}_n \cdot \text{dNTP}$) (Figure 1.4, step 2). This is followed by the chemical step of nucleotidyl transfer, and the DNA primer strand gets extended by one nucleotide ($\text{Pol} \cdot \text{DNA}_{n+1} \cdot \text{PPi}$) (Figure 1.4, step 3).

This chemical reaction is a $\text{S}_{\text{N}}2$ reaction (bimolecular nucleophilic substitution reaction) facilitated by two divalent metal ions (typically Mg^{2+}), and the mechanism is known as the **two-metal ion mechanism** (Figure 1.5) [58]. The byproduct of the $\text{S}_{\text{N}}2$ reaction is an inorganic pyrophosphate (PPi), which leaves during the fourth step of the minimal catalytic cycle leading to the formation of the post-chemistry binary complex ($\text{Pol} \cdot \text{DNA}_{n+1}$) (Figure 1.4, step 4). At this stage, the DNA polymerase can either dissociate from the product DNA (distributive synthesis) or translocate along the DNA and incorporate the next nucleotide (processive synthesis) [58, 59].

The two-metal ion mechanism was first proposed by Beese and Steitz based on the binary complex structure of the Klenow fragment [60]. Subsequent structural and biochemical work has shown that this mechanism is followed by most DNA polymerases, RNA polymerases, and many other nucleic acid processing enzymes. It is proposed that metal A (MeA, the catalytic metal) reduces the pKa of a nucleophile and activates the 3'OH nucleophile via the formation of a hydroxide ion (OH^-) and coordinates both the attacking oxygen and the α -phosphate of the incoming dNTP. The OH^- then performs an in-line nucleophilic attack on the α phosphate of the incoming dNTP. Metal A binds transiently, while metal B (MeB) is thought to bind stably to the incoming nucleotide and stabilizes the pentacovalent intermediate by dissipating the negative charge on the phosphates. Because of its stable binding, MeB is anticipated to remain bound throughout the reaction and is released with the PPi (Figure 1.5) [61, 62].

Recently, time-resolved crystallographic studies have revealed a transient involvement of a third metal ion (MeC) during the catalytic cycle of several DNA polymerases, including human DNA Pol η , Pol β , and Pol μ [63] [64, 65]. MeC appears during correct nucleotide incorporation, aiding precise alignment of substrate and enhancing catalysis, although its precise role is still debated [64]. To answer whether MeC is required for all DNA polymerases would require further structural and kinetic validation.

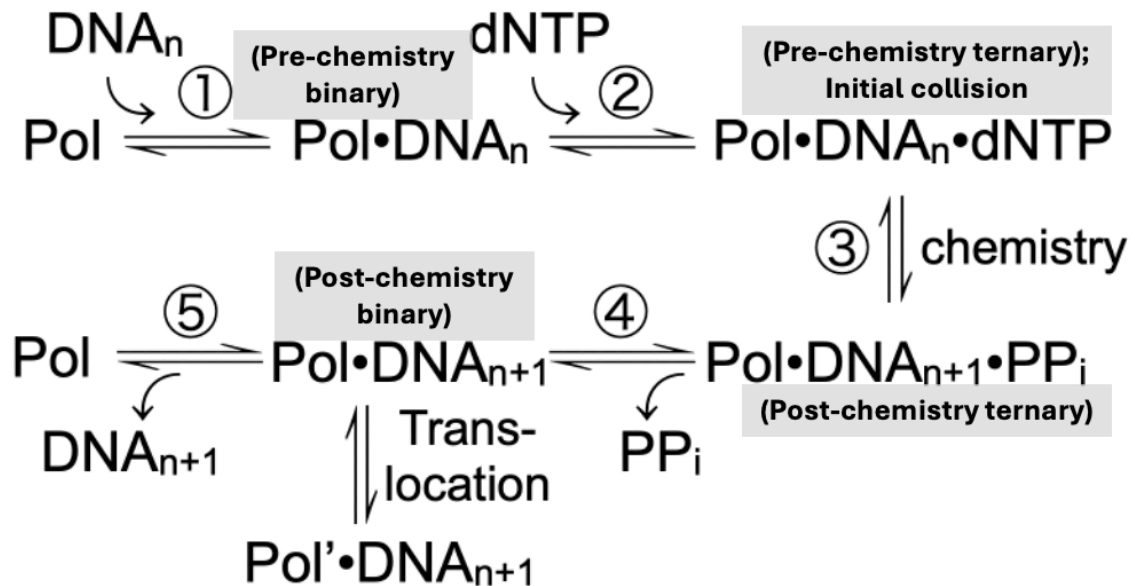


Figure 1.4: Minimum catalytic pathway for DNA polymerase. Minimal kinetic pathway for nucleotide incorporation by a DNA polymerase (Pol). DNA_n : DNA substrate with a primer strand n bases long, DNA_{n+1} : DNA substrate with a primer strand $n+1$ bases long, dNTP: incoming deoxynucleotide triphosphate, PP_i : inorganic pyrophosphate, and Pol' : DNA polymerase translocated by one base.

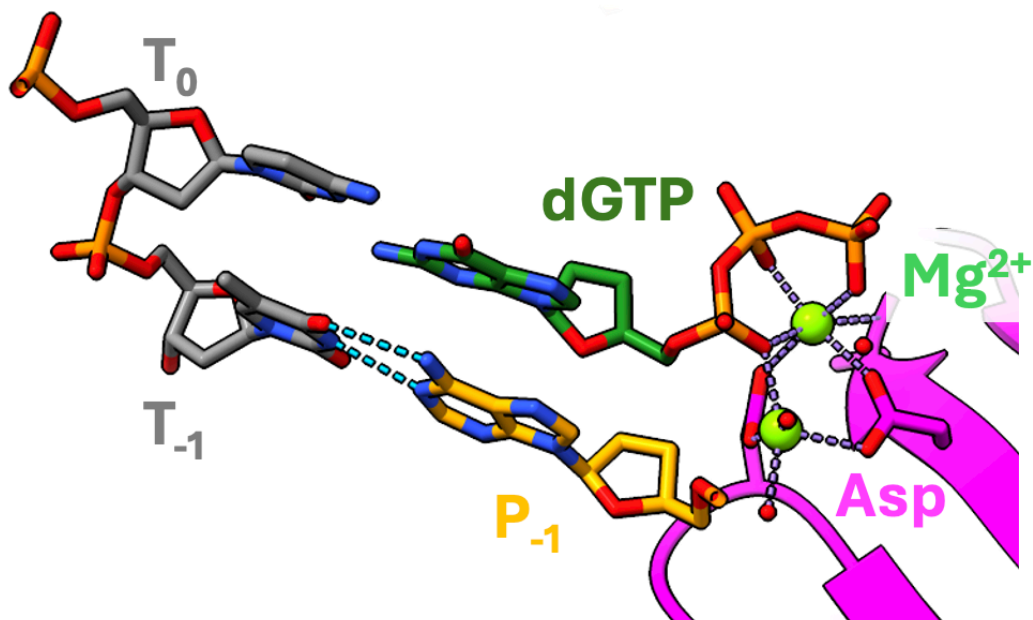


Figure 1.5: Two metal ion mechanism of nucleotide incorporation. Close-up view of the active site of bacteriophage T7 DNA polymerase (PDB: 1T7P) showing the incoming dGTP (green) positioned opposite the templating base T_0 and adjacent base T_{-1} (grey), with the 3' terminal primer base P_{-1} (orange) poised for the nucleophilic attack. Two catalytic Mg^{2+} ions (light green spheres) are coordinated by conserved aspartate residues in the palm domain (magenta) and by the triphosphate of the dGTP, stabilizing the pentavalent transition state and facilitating phosphodiester bond formation. Dashed lines indicate key hydrogen-bonding and metal-ligand interactions that support the canonical two-metal-ion mechanism used by replicative DNA polymerases. Red sphere: ordered water molecule.

1.2.4 Multiple DNA polymerases within a system

Faithful, efficient, and rapid copying of genomic DNA is essential for genome replication and maintenance. Genomic DNA presents a diverse set of chemical and structural challenges, including bulk replication, repair, lesion bypass, and end-joining, and no single DNA polymerase can optimally perform all of these functions simultaneously [66]. Consequently, most replication systems, including prokaryotic, eukaryotic, and organellar replisomes, use multiple DNA polymerases. For example, *E. coli* encodes five distinct DNA polymerases, while the human nuclear replisome uses more than 18 DNA polymerases and mitochondrial replication involves four different polymerases (Figure 1.6). These polymerases have specialized functions including replicative, repair, and translesion synthesis [17].

Most polymerases follow the same overall catalytic pathway, share a common reaction mechanism, and are structurally similar, leading to the question: How do DNA polymerases perform specialized roles. Typically, a major functional distinction lies in the architecture of the active sites. The active site of a **replicative polymerase** is highly restrictive or ‘closed’ such that it can only accommodate correct Watson-Crick (WC) base pairs (Figure 1.7 A). This strict geometric selection underlies high-fidelity replication, enabling efficient incorporation of the correct dNTP opposite the templating base, and makes these enzymes responsible for copying the bulk of the genome. However, because of this restrictive active site, replicative polymerases cannot accommodate damaged bases and are unable to bypass lesions [17].

In contrast, the active site of a **TLS polymerase** is typically more open and solvent-exposed (Figure 1.7 B). This open architecture allows accommodation of a wide variety of DNA lesions and non-WC base pairs within the active site. While this feature enables efficient

lesion bypass, it also renders TLS polymerases intrinsically error-prone when copying undamaged DNA [67]. TLS polymerases therefore act as specialized damage-tolerance factors: they prevent replication fork collapse and permit synthesis to continue past lesions, after which a high-fidelity replicative polymerase can re-engage and resume accurate DNA synthesis. This modular toolkit of DNA polymerases allows cells to balance genome stability (via high-fidelity replicative polymerases) with flexibility and damage tolerance (via TLS polymerases), which is essential for both genome maintenance and adaptation to environmental stress [68, 69].

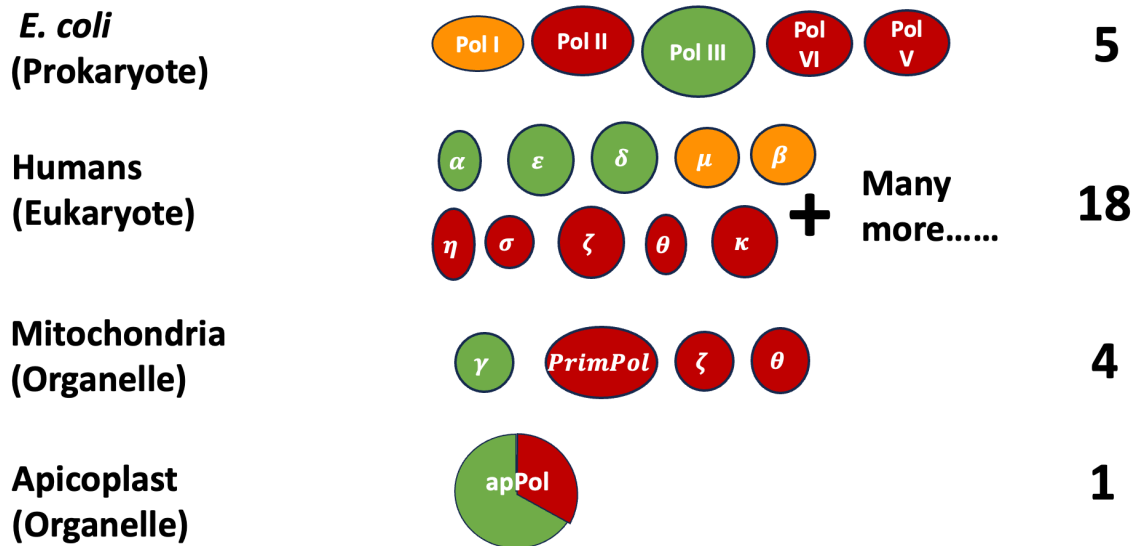


Figure 1.6: Multiple DNA polymerases in different biological system. Schematic overview of the number and functional classes of DNA polymerases presents in *E. coli* (prokaryote), humans (eukaryote), mitochondria, and the apicoplast. Replicative polymerases are shown in green, repair polymerases in orange, and translesion synthesis polymerases in red. The figure highlights the contrast between systems that utilize multiple specialized DNA polymerases and the apicoplast, which relies on a single DNA polymerase, apPol.

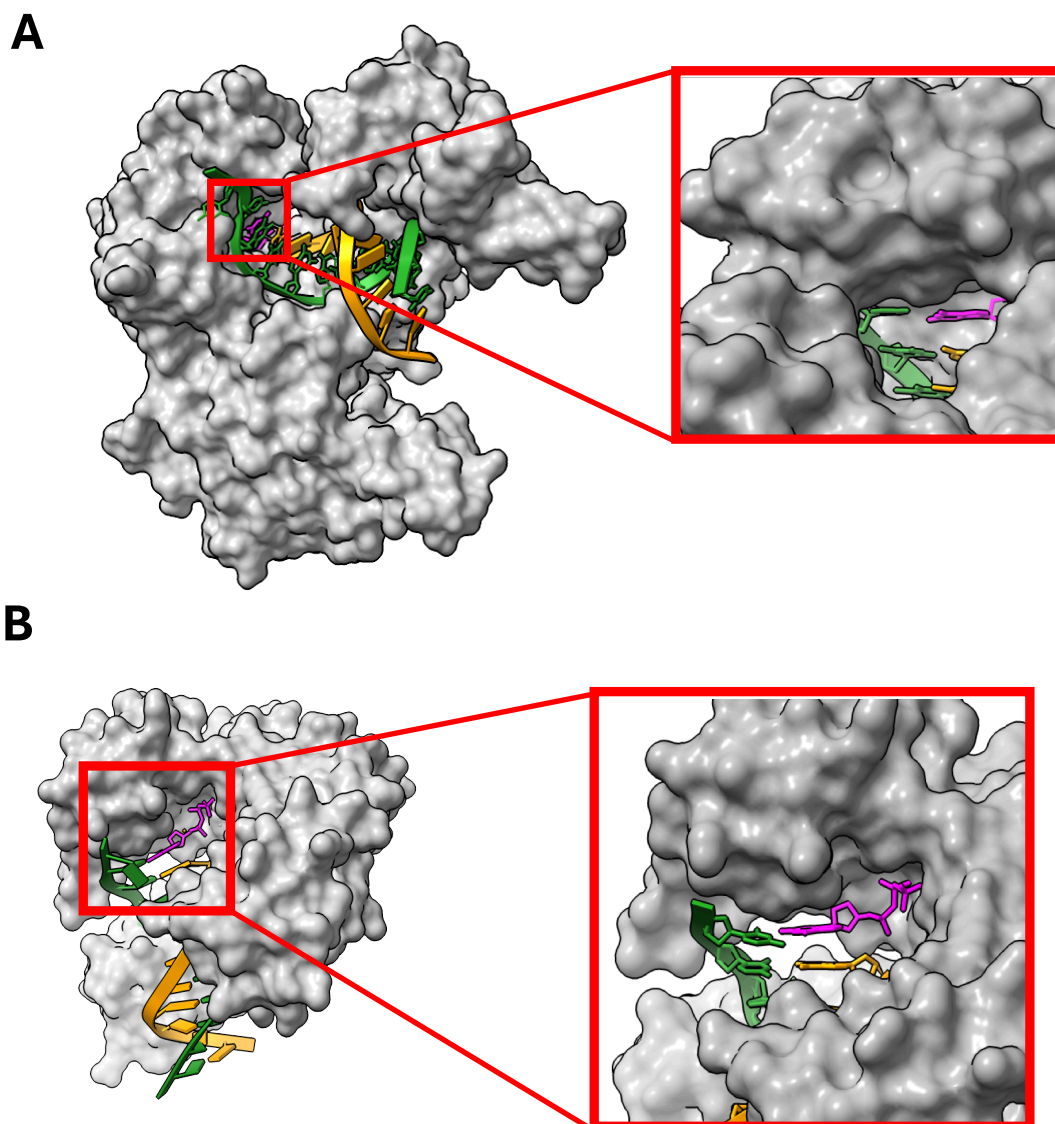


Figure 1.7: Active site of DNA polymerase. Surface and cartoon representations of replicative and TLS DNA polymerase active sites. **(A)** T7 DNA polymerase (PDB: 1T7P), which is a replicative polymerase and **(B)** Dbh (PDB: 3BQ1), which is a TLS polymerase. The zoomed-in panels highlight the active site, substrate-binding pocket and the positioning of the primer–template DNA and incoming nucleotide within the catalytic cleft. Where polymerase surface is shown in grey, primer strand shown in yellow, template strand shown in green and incoming nucleotide shown in magenta.

1.2.5 Conformational dynamics of DNA polymerases

DNA polymerase dynamics refer to the conformational changes a DNA polymerase undergoes during DNA synthesis, including substrate binding, fingers closing, catalysis, pyrophosphate release, substrate transition during proofreading, and translocation to the next templating position [70]. For instance, the major determinant of different active site geometries in most DNA polymerase families (A, B, C, and X) is the conformational changes in the fingers domain [71]. The two major unanswered questions in the DNA polymerase field are (a) what drives the catalytic cycle of nucleotide incorporation and (b) what's the major determinant of DNA polymerase fidelity. Soon after the initial structures of DNA polymerases were determined, it was suggested that conformational dynamics of a DNA polymerase might be responsible for both. Comparison of the apo DNA polymerase structure (**open state**) with that of the pre-chemistry ternary complex revealed a large-scale motion in the fingers domain for replicative polymerases [72] (Figure 1.8). As the polymerase binds to DNA and then dNTP, the fingers close on the substrates bringing them in precise alignment and occlude the active site from the external solvent in preparation for the chemistry step of bond phosphodiester bond formation. This conformational change is essential in most cases for the initial collision ternary complex to progress to a catalytically competent form, thus making fingers closing (**closed state**) a major driver of the nucleotide incorporation cycle, playing a major role in high-fidelity DNA synthesis [73] [74, 75].

In contrast, for some low-fidelity TLS polymerases, there is minimal motion of the fingers domain as the polymerase transitions from a pre-chemistry binary to a pre-chemistry ternary complex, and the fingers domain stays in a predominantly closed configuration throughout the nucleotide incorporation cycle [76]. Other than the large-scale motion of fingers closing,

several subtle conformational rearrangements of the active site residues have been implicated in DNA polymerase fidelity [72] [77, 78].

1.2 Malaria and *Plasmodium*

Malaria is a life-threatening disease affecting a major part of the world, and infection has increased between 2015 and 2024 by 8.5%. In 2024 there were an estimated of 282 million malaria cases and 610,000 deaths worldwide [79]. Malaria is caused by a parasite that belongs to the genus *Plasmodium*. The genus comprises many species but only five of them are responsible for human malaria: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi*. Among these, *Plasmodium falciparum* is responsible for the majority of severe diseases and deaths [80].

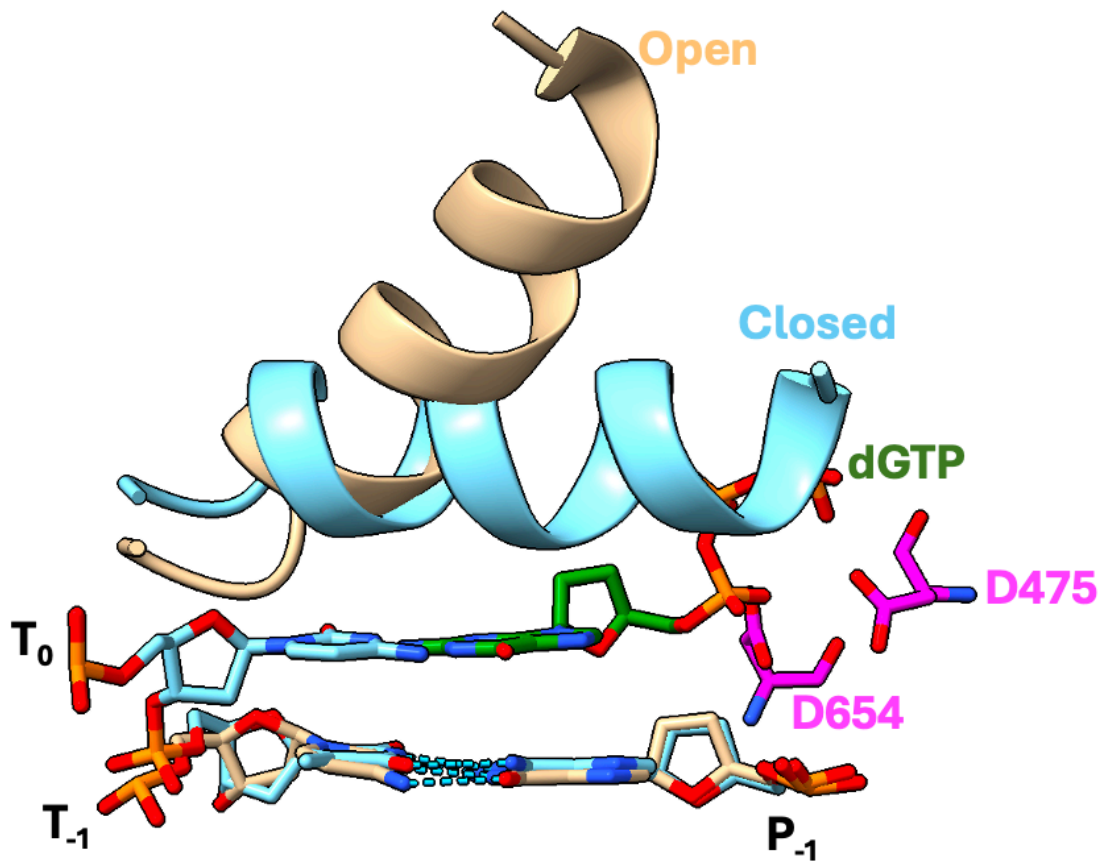


Figure 1.8: Open and closed fingers conformations of T7 DNA polymerase. Structural comparison of bacteriophage T7 DNA polymerase in the open (PDB 1TK5) and closed (PDB 1T7P) conformations, highlighting the movement of the fingers domain. In the open state (golden), the O-helix of the fingers is positioned away from the active site aspartates, leaving the nucleotide-binding pocket accessible for dGTP (green) sampling. In the closed state (light blue), the fingers curl over the nascent base pair. Where, incoming nucleotide is shown in green, paired with the T₀ base shown in cyan. The primer terminus (P₋₁) and template (T₋₁) of both structures superimpose with each other. The catalytic aspartates are shown in magenta.

1.4 Apicoplast and its origin

Plasmodium belongs to the phylum Apicomplexa, which comprises more than 6000 species of obligate parasitic protists [81]. Many members of this phylum (also called apicomplexans), including *Plasmodium*, possess a unique non-photosynthetic plastid called the apicoplast. Apicoplast harbours several essential biosynthetic pathways [82] including isoprenoid precursor (IPP) biosynthesis, heme biosynthesis, iron-sulfur cluster biogenesis among others, and consequently this organelle is essential for the *Plasmodium* viability [83, 84].

Similar to the mitochondrion and the chloroplast, the apicoplast originated via endosymbiosis, but in this case the organelle arose through a two-step (secondary) endosymbiosis process: a cyanobacterium was first engulfed by red algal ancestor, which was subsequently engulfed by a heterotopic eukaryote related to modern apicomplexans. As a result of this secondary endosymbiosis, the apicoplast is surrounded by four membranes, a feature clearly demonstrated in *Plasmodium* by cryo-electron tomography [82, 85].

1.4.1 The apicoplast genome

Other than the nuclear and mitochondrial genomes, *Plasmodium* contains a third genome, which is present inside the apicoplast. The apicoplast DNA (apDNA) is a 35 kb circular DNA with a highly repetitive sequence and an extreme skew (~86%) towards A and T [86]. apDNA codes for apicoplast housekeeping genes involved in transcription, translation, and several chaperones with unidentified substrates [87]. The proteins responsible for apicoplast replication are encoded by the *Plasmodium* nuclear genome and imported to the apicoplast through a poorly understood process [86]. Apicoplast genome duplication occurs during the

blood stage of *Plasmodium*'s life cycle, with apDNA replication initiating during the late trophozoite phase and continuing into the schizogony phase of the parasite's life cycle [88].

Faithful duplication and maintenance of the apicoplast DNA is essential for the viability of *Plasmodium*. Due to their essential nature, proteins involved in apicoplast DNA transactions could be potential drug targets against malaria [89].

1.4.2 The apicoplast replisome

Apicoplast replication is performed by a dedicated replisome with components that are divergent from other eukaryotic replisomes studied to date [90]. At the heart of the apicoplast replisome is a polyprotein called PREX (plastidic DNA replication/repair enzyme complex). PREX contains primase, helicase, and DNA polymerase activities in a single polypeptide [91]. After its import into the apicoplast, PREX is proteolytically cleaved into the three enzymatic components [92, 93].

1.4.3 Apicoplast DNA polymerase (apPol)

apPol belongs to the A-family of DNA polymerases, and within the A-family, apPol is placed in a poorly studied clade of viral DNA polymerases [94]. Apoenzyme structures of apPol provide an overview of this enzyme's architecture [95, 96]. Similar to other A-family polymerases, apPol resembles a right hand with two active sites, one for the polymerization activity and the other for 3' to 5' exonuclease proofreading activity [96]. apPol has an additional N-terminal domain (NTD) of unknown functions (Figure 4.2).

As explained before, both eukaryotic and prokaryotic cellular replisomes contain multiple DNA polymerases, each with a specialized function. In contrast, apPol is the only DNA polymerase identified in the apicoplast [90]. One possible scenario is that apPol performs the functions of multiple specialized DNA polymerases, including replicative and translesion synthesis. Indeed, it has been shown biochemically that apPol can bypass lesions. However, the mechanism of multitasking by apPol remains obscure. Despite earlier structural studies and steady-state kinetic analysis, the fundamental catalytic pathway of apPol and the mechanistic basis of its multitasking properties remain unclear. [93, 97] [98]

1.5 Techniques used to study enzymatic mechanisms of DNA polymerases

A variety of experimental techniques have been developed for studying the reaction pathway, catalytic mechanism, and different conformational states of DNA polymerases. Biochemical assays, particularly kinetic assays, remain fundamental for characterizing polymerase activity. Both steady-state and transient-state approaches provide valuable information about the catalytic cycle, as they monitor the rates of product formation, which are dependent on the covalent chemical transformation within the enzyme-substrate complex [75] [99, 100]. These methods have been instrumental in dissecting the sequence of chemical events underlying nucleotide incorporation and proofreading activities of various polymerases.

Biophysical techniques complement these biochemical assays by providing insights into the conformational states that accompanies catalysis. Fluorescence-based assays such as those employing 2-aminopurine and single molecule Förster resonance energy transfer (smFRET), which can monitor structural rearrangement in real time have been instrumental in capturing short-lived polymerase conformations. Similarly, force-based measurements, including

optical and magnetic tweezers, detect mechanical responses associated with polymerase movement along DNA, thereby offering a quantitative view of DNA-protein dynamics that depends on non-covalent interactions [101]. In addition to these approaches, sequencing-based assays—particularly long-read sequencing techniques such as SMRT sequencing—enable studies of polymerase fidelity at a single molecule resolution [102]. Finally, structural biology methods, including cryo-electron microscopy (cryo-EM) and X-ray crystallography, provide high-resolution snapshots of polymerase complexes at different stages of the catalytic cycle, linking structural states to functional outcomes [103].

In this dissertation, I have applied a combination of kinetic assays (chapter 3) and cryo-EM (chapter 4) to investigate the mechanistic features of the apicoplast DNA polymerase, aiming to elucidate its unique ability to multitask with the organellar replication system.

1.5.1 Kinetics of DNA polymerase

DNA polymerases operate through a multi-step catalytic cycle that includes DNA binding, nucleotide binding, the chemical step of bond formation, and pyrophosphate release (Figure 1.4). Then depending on whether the polymerase is performing processive or distributive synthesis, the enzyme will undergo either translocation (for processive synthesis) or dissociation from the DNA substrate (distributive synthesis). The same sequence of the catalytic steps are shared by most DNA polymerases, but the distinct catalytic signatures that differentiate one polymerase from another arise from the rate constants governing each step of the cycle [104]. By altering these rate constants, DNA polymerases can modulate their efficiency, fidelity (it refers to the accuracy with which the enzyme replicates DNA, reflecting its ability to select the correct incoming nucleotide and avoid misincorporation),

and processivity (it refers to the average number of nucleotides incorporated by a DNA polymerase per binding event), thereby becoming suited to specialized biological functions. Enzyme kinetics provides the only direct means to determine the rate constants for each step of the catalytic pathway [75].

A simple approach for measuring DNA polymerase activity is to provide a chemically synthesized prime/template DNA substrate of defined sequence and the incoming nucleotides and follow the rate of extension of the primer strand (Figure 2.3). Historically, most primer extension assays were carried out under steady-state conditions, in which the DNA is present in large excess over the polymerase. The detectable signal arises from multiple turnover of the enzyme, thus reporting mainly on the rate-limiting step of the cycle, which for most DNA polymerases is the release of the extended DNA product (Figure 1.9). Hence, steady-state assays are not informative for dissecting individual steps in the catalytic cycle. Moreover, if the steady-state assay is set up to measure the incorporation rate of a single nucleotide, then the assay essentially reports on the slow release of the DNA product extended by a single nucleotide, which might not be the most relevant step of processive DNA synthesis *in vivo* as explained below.

At the replication fork, replicative DNA polymerases do not dissociate from their DNA substrate frequently; instead, they translocate along the template and catalyse successive rounds of DNA synthesis (processive synthesis). Hence, the steady-state rate constant (k_{cat}) derived from single nucleotide incorporation assays does not reflect the rate of processive synthesis at the replication fork. [105]

To define the individual steps of the kinetic pathway of a DNA polymerase, experiments are typically performed under pre-steady state (also called transient state) conditions. One pre-

steady state approach is **single-turnover kinetics**, in which the polymerase is present in excess over DNA so that all DNA is pre-bound by the polymerase to form Pol•DNA binary complex prior to the nucleotide addition. Upon rapid mixing with the incoming nucleotide, the Pol•DNA complexes are converted to extended products, and interference from the slow post-chemistry DNA dissociation is minimized. Under this condition, the maximal rate of polymerisation (k_{pol}) reports on the rate-limiting step leading up to chemistry. (Figure 1.9) [106, 107]

A second pre-steady state approach employs **burst kinetics**, where the DNA is present in slight excess compared to the polymerase such that a large fraction of the DNA substrate pre-forms Pol•DNA binary complexes. The critical difference from a single-turnover assay is, under burst conditions, a small fraction of DNA remains unbound at the start of the assay. Following dNTP addition, the pre-bound binary complex generates an initial burst of product formation, whereas the remaining unbound DNA substrate is converted to product in the subsequent rounds of catalysis, thus reporting on both pre- and post-chemistry steps. Because DNA polymerases are very fast enzymes, transient-state kinetic experiments, including both single turnover and burst measurements, typically require quench-flow or stopped-flow instruments to capture reactions on a millisecond timescale (Figure 1.9) [107, 108].

1.5.2 Cryo-EM

Cryo-electron microscopy (cryo-EM) is a powerful technique for visualizing biological macromolecules in a close-to-native vitrified state. It is a transmission electron microscopy (TEM) method in which an electron wave is transmitted through a thin sample to form an image [109, 110]. Cryo-EM encompasses multiple different experimental approaches. In this dissertation, I will limit the discussion to single particle analysis (SPA). In SPA cryo-EM images of many individual macromolecules are computationally aligned in three-dimensional (3D) space and averaged to obtain a 3D reconstruction of the macromolecule under study [111]. Using modern cryo-EM hardware and image processing algorithms, 3D structures from SPA can be determined to a high resolution (better than 3Å) [112, 113]. In this dissertation, SPA has been used to study the mechanism of action of apicoplast DNA polymerase. This section introduces the basic principles of image formation in cryo-EM and outlines key concepts of image processing in SPA.

Theory of Image formation

▪ **Electron scattering**

Electrons provide the energy source used in electron microscopes and generate images through their interactions with the specimen, a phenomenon known as electron scattering. When the incident electron wave passes through a biological sample, individual electrons can remain unscattered, elastically scattered, or inelastically scattered (Figure 1.10). The scattered electrons and their interaction with the unscattered electrons lead to contrast. In TEM images there are two types of contrast: amplitude contrast and phase contrast. For

biological specimens visualized in cryo-EM, phase contrast is the primary source of contrast [114].

Unscattered electrons travel through the specimen without interacting with the sample and therefore, retain their original trajectory and energy. Electrons that interact with the sample but retain the original energy while being deflected to different angles are referred to as elastically scattered electrons. The interference between the elastically scattered and unscattered electrons generate phase contrast. On the other hand, inelastically scattered electrons transfer part of their energy to the specimen and are typically deflected by smaller angles, contributing predominantly to background noise in phase-contrast images and decreasing the signal-to-noise ratio (SNR) [115]. Inelastically scattered electrons can be removed or attenuated using an energy filter in modern electron microscopes.

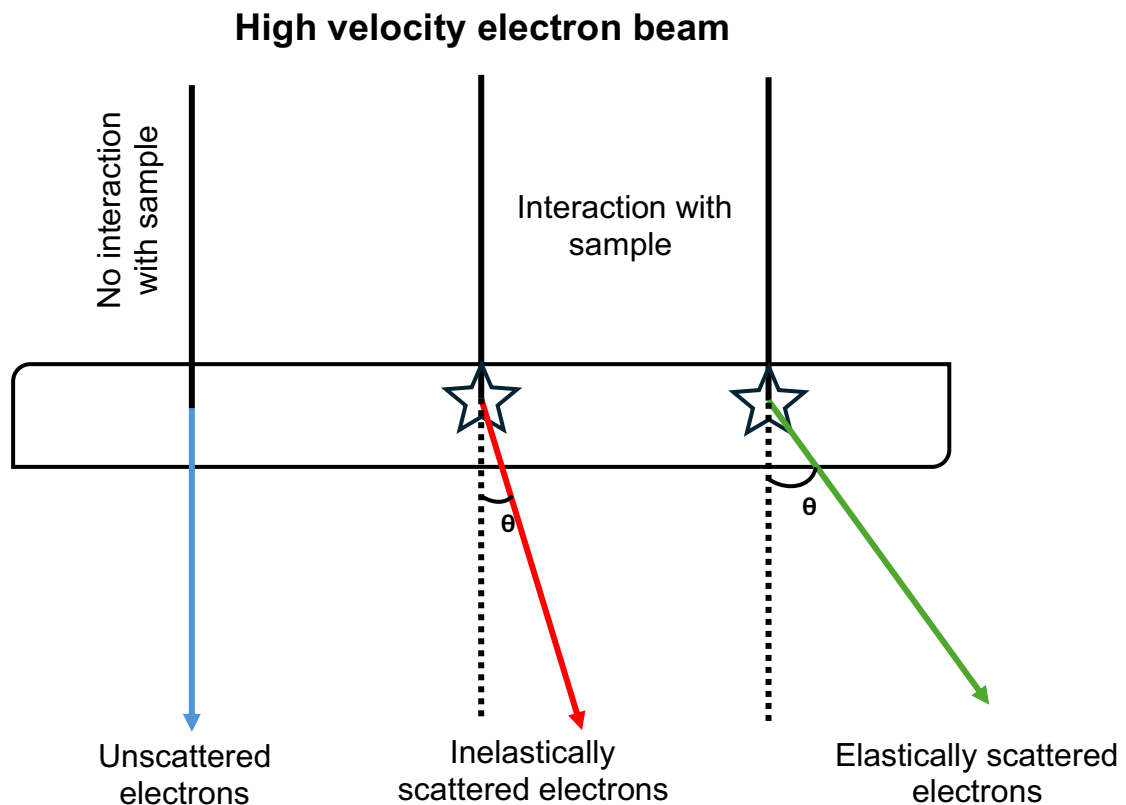


Figure 1.10: Types of electrons scattering in cryo-EM. Schematic representation of a high-energy electron beam interacting with a vitreous specimen. Electrons that pass through the sample without interacting are recorded as unscattered electrons (blue arrow). Electrons that interact with the specimen and lose energy, while being deflected by small angles, are shown as inelastically scattered electrons (red arrow); these primarily contribute to background noise. Electrons that interact with specimen, retaining their original energy but being scattered to larger angles, are depicted as elastically scattered electrons (green arrow) and carry the high-resolution structural information used for image formation in cryo-EM. θ : scattering angle.

▪ Contrast in cryo-EM

Contrast can be conceptualised as the difference in intensity within an image. For this discussion we will consider the wave property of electrons. When the incident electron wavefront interacts with the sample, scattering of the incident wave alters the amplitude and phase of the exiting wave. A substantially lower amplitude in the exit wave occurs when part of the incident wave is absorbed or blocked by the specimen, (for instance, in negative stain TEM) and generates strong amplitude contrast. In cryo-EM of frozen-hydrated biological samples, amplitude contrast has a negligible contribution to contrast [116]. Instead, most of the useful image contrast arises from phase contrast. Biological samples behave as weak phase objects, i.e., there is a single interaction between the incident wavefront and the object. Instead of absorbing or blocking the incident electron wave, weak phase objects introduce a phase shift in the elastically scattered wave. However, Detectors cannot directly record phase shift. They can only record intensity, which is proportional to the square of the amplitude of the wave resulting from the interference of the elastically scattered and unscattered waves. To convert the phase shifts into detectable intensity variations, specific imaging conditions and microscope settings are chosen, and intrinsic lens aberrations also play a role in enhancing contrast. Some of the main concepts are outlined below:

- 1. Weak phase object approximation**— This approximation applies to very thin samples (typically <50 nm) imaged at high accelerating voltage (>100 kV), such that the incident electron undergoes a single scattering event on interacting with the sample rather than multiple scattering events (which can happen in thicker samples). Under these conditions the scattered wave has a phase shift of 90° compared to the incident wave. Only a small fraction of the incident electron wave undergoes scattering, and thus the amplitude of the scattered wave is very small. The resulting

wave detected by the detector is the sum of the scattered and unscattered waves. The amplitude of the resulting wave is almost similar to the amplitude of the incident wave (Figure 1.11), leading to almost no change in detected intensity [116].

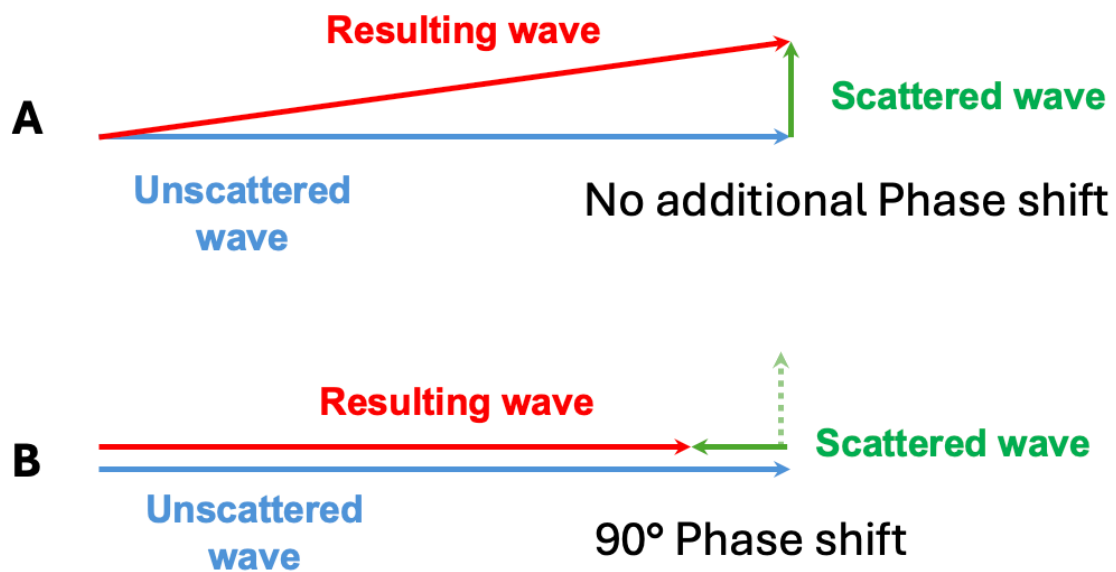


Figure 1.11: An Argand diagram of unscattered and scattered waves in cryo-EM.

Schematic representation of how the resulting wave amplitude depends on the relative phase between the unscattered reference wave (blue) and the specimen-scattered wave (green). **(A)** Without an additional phase shift, the scattered wave adds obliquely to the unscattered wave, producing a resulting wave (red) with only a small change in amplitude and therefore generating minimal phase contrast. **(B)** When the scattered wave undergoes an additional 90° phase shift relative to the unscattered wave, the two vectors add (or subtract) leading to a larger change in amplitude of the resulting wave thereby generating enhanced phase contrast in the image.

- 2. Defocus** – To increase the detectable contrast in the cryo-EM images, the objective lens is intentionally set to an under-focused condition. Changing the defocus modifies the pathlengths (and consequently the phase shifts) of the scattered electrons as a function of their scattering angle, while leaving the unscattered beam unaffected. Interference between the additionally phase-shifted scattered waves and the unscattered wave at the image plane leads to a resultant wave with a detectable amplitude difference compared to the incident wave. The scattering angle (θ) is related to the electron wavelength (λ) and the spatial frequency (inverse of real-space distance, d) of features in the specimen by equation 1. Hence, smaller interatomic distances (higher spatial frequencies) in the sample produce larger scattering angles, which in turn experience greater defocus-induced phase shifts and thus enhanced contrast.

$$\theta = \arcsin \frac{\lambda}{2d} \quad \text{Equation 1}$$

- 3. Spherical aberration (C_s)**—It is an intrinsic lens imperfection in which electrons that are further away from the optical axis are over-focused relative to those traveling near the axis. This introduces an additional spatial-frequency-dependent phase shift that contributes to phase contrast. This is particularly important at high resolution, because high-angle (high frequency) scattered electrons are more strongly over-focussed by C_s . Together with defocus, spherical aberration is essential for generating detectable phase contrast in the cryo-EM images.

▪ Contrast Transfer Function (CTF)

Contrast Transfer Function is a function that describes how contrast is detected as a function of spatial frequency. CTF captures the cumulative phase shifts introduced by defocus, spherical aberrations. Therefore, it determines whether the information at a given spatial frequency is transferred with positive contrast, inverted (negative contrast), or not transferred at all (zero crossing of the CTF). By combining the phase shifts from weak phase approximation, defocus, and spherical aberrations, we can calculate the CTF using the following equation:

$$CTF(f) = \sin(-\pi\lambda\delta f^2 + 2\pi C_s\lambda^3 f^4) \quad \text{Equation 2}$$

Where:

f is the spatial frequency

λ is the electron wavelength

δ is the defocus

C_s is the spherical aberrations

Most parameters in the equation are fixed for a given microscope setup, with the defocus being the principal variable. The CTF models the effect of defocus (Figure 1.12) and microscope aberrations on image formation and can be interpreted as the point spread function (PSF) in real space, which causes a single point in the sample to appear blurred out in the image (Figure 1.13). CTF oscillates between +1 and -1 and crosses zero multiple times. Spatial frequencies at which the CTF is zero are not transferred and therefore have no contrast and are not visible in the image (Figure 1.12). To recover information across the full range of spatial frequencies, images are collected at a series of defocus values so that

frequencies that are attenuated or missing at one defocus are transmitted at another (Figure 1.12), allowing all information to be adequately sampled for three-dimensional reconstruction during cryo-EM image processing [117].

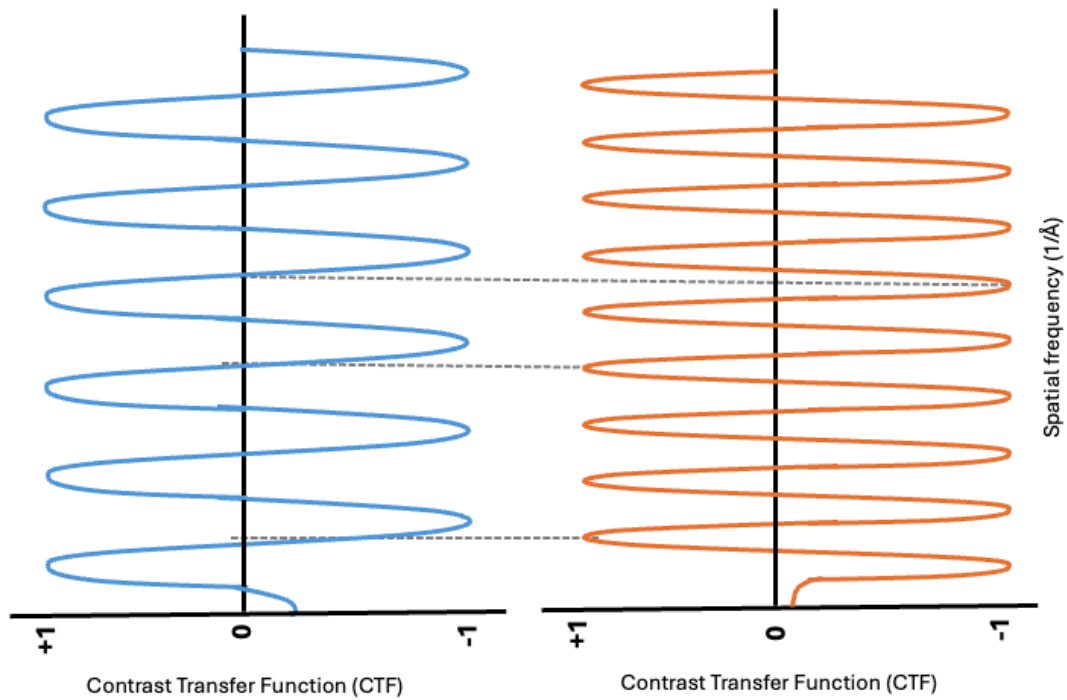


Figure 1.12: Contrast transfer function at two different defocus values. Plots of the contrast transfer function (CTF) as a function of spatial frequency for a lower defocus (top, blue) and a higher defocus (bottom, red) imaging condition. Oscillations about zero illustrate alternating bands of positive and negative image contrast, while the vertical dashed lines mark spatial frequencies at which the CTF crosses zero and information is not transferred to the image at one defocus value (top) but gained at different defocus value (bottom).

Theory of Image processing

Image processing aims to correct the various aberrations present in the raw cryo-EM images so that the final, corrected micrographs more accurately represent the underlying specimen and can be used for 3D reconstruction.

- **Movies to micrograph**

During data collection, cryo-EM samples are affected by several factors, leading to a blurring of the images. A major factor is beam-induced deformation of the vitreous ice and mechanical drift in the microscope. To overcome this blurring effect, a new scheme of image recording was introduced in 2012. In this scheme, instead of recording a single micrograph, the image is recorded as a movie split into multiple frames. Motion correction algorithms align the individual frames of the movie to compensate for image blurring, producing a motion-corrected micrograph preserving high-resolution details [118].

Electron irradiation damages biological specimens, with the high-resolution information being lost at higher cumulative doses compared to low-resolution information. To account for optionally a dose weighting scheme can be applied to the individual movie frames before summation. The high-resolution information from the later frames (higher accumulated dose) is downweighed compared to the high-resolution signal of the initial frames (lower accumulated dose). These corrections are only possible because images are recorded as movies rather than as a single exposure image, and the resulting motion-corrected, dose-weighted micrographs that form the basis for all subsequent processing steps [119].

▪ CTF Estimation and CTF Correction

As described in the previous section, the CTF captures the effect of defocus and other microscope aberrations that distort the recorded image relative to the true object. CTF estimation determines the true defocus and other CTF parameters for each micrograph, or even for different regions or individual particles within a micrograph. This helps address the **sample flatness** problem, an inherent feature of vitrified specimens in which a defocus gradient can arise from sample tilt and uneven ice thickness [116] [120].

Because the CTF is the Fourier space representation of PSF, it can be modelled and corrected computationally. The basic purpose of CTF correction is to compensate for the phase distortions introduced by the defocus and aberrations of the electron microscope so that the underlying specimen contrast is restored as accurately as possible. This improves the visibility and interpretability of particle images before downstream image processing [116] [120].

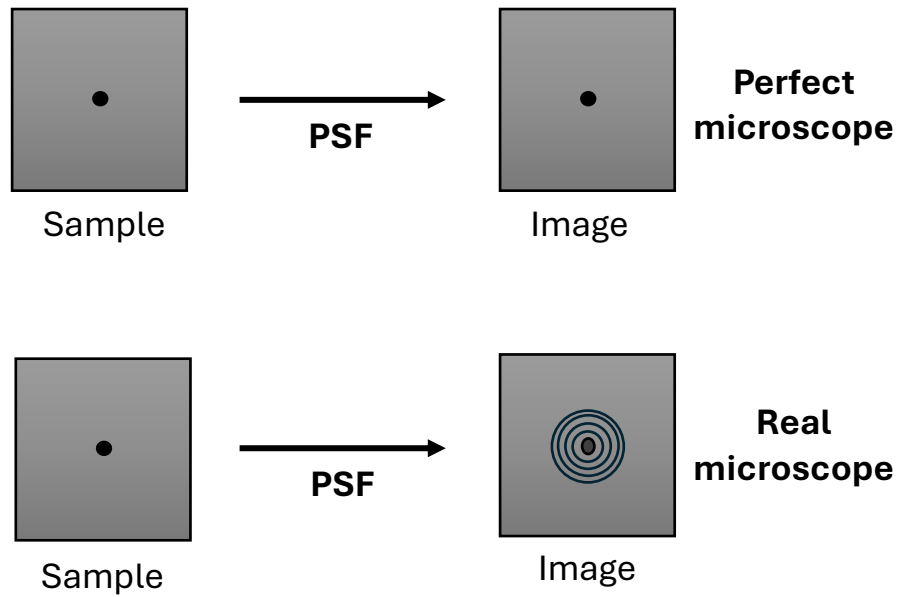


Figure 1.13: Point spread function (PSF). In an ideal microscope vs real microscope, Schematic illustration of how a single point object in the specimen (left) is imaged by a perfect, aberration-free microscope as a point of identical size and shape (top), whereas a real, microscope produces an extended blur pattern described by the point spread function (bottom).

2. MATERIALS & METHODS

2.1 DNA substrate

All DNA oligonucleotides with unmodified and dideoxy terminated nucleotides were purchased from Integrated DNA Technology (USA). Oxidative damage-containing DNA oligonucleotides were purchased from TriLink Biotechnologies (USA). DNA substrates (P/T DNA) were generated by annealing the primer and template DNA strands as shown in Table 1. All annealing reactions were performed using a 1.1-fold excess of template strands over primer strand in an annealing buffer (10 mM Tris-Cl (pH 7.5) and 50 mM NaCl). The sample was heated to 95°C for 5 min, followed by gradual cooling to 25°C.

2.2 Software

All urea acrylamide denaturing gels were analysed using ImageQuant TL version 10.1 (Cytiva, USA). Graph plotting and nonlinear regressions were performed using Prism version 9.3.1 (Graphpad Software, USA) unless otherwise mentioned. Global fitting of the primer extension data using numerical integration was performed using KinTek Explorer version 10.2.3 (KinTek Corp., USA) [121] [122]. Cryo-EM datasets were processed and analysed using the following software: CryoSPARC [123], Topaz [124], UCSF ChimeraX [125], Coot [126], ISOLDE [127], Phenix [128] and PyMOL Molecular Graphics System, (version 3.0 Schrödinger LLC). Protein sequence alignment was done using Clustal Omega [129] and the alignment was visualised using Jalview [130].

2.3 apPol Constructs

For all the cryo-EM and kinetic experiments except the exonuclease assays, we have used a construct of apPol harbouring two-point mutations (D82N and E84Q) at the active site of the proofreading exonuclease domain (Figure 2.1 A). The mutant construct, referred to as apPol in all chapters, showed no detectable DNA degradation (Figure 3.4 B) and allowed us to focus on the polymerization activity. For the exonuclease assays, the wild type apPol (apPol^{WT}) lacking the two-point mutations mentioned above has been used (Figure 2.1 B). For primer extension assays to determine the steric gate residue (which enables DNA polymerases to distinguish between dNTPs and rNTPs by sterically clashing with the 2'-OH group of ribonucleotides), an apPol construct with an additional point mutation (apPol^{E415A} or Q548A) (Figure 2.1 C and D) was used. For determining the effect of tyrosine 481, 485 and 486 on apPol's fidelity, an apPol construct where those three tyrosines were mutated alanine (apPol^{Y481A_Y485A_Y486A}) (Figure 2.1 E) was used.

All the constructs had an N-terminal hexa-histidine tag followed by a tobacco etch virus (TEV) protease cleavage site. The constructs were codon optimized for expression in *Escherichia coli* and were chemically synthesized and cloned into the pETDuet1 vector by GenScript Corp. (USA) (Figure 2.2).

Table 1: List of DNA substrates.

No.	Name	Sequence
1.	S1	5' *CAGGTGTCAGTCAGCTAGTGCTG 3' 3' GTCCACAGTCAGTCGATCACGACATGCA 5' T₀T₁
2.	S2	5' *CAGGTGTCAGTCAGCTAGTGCTG 3' 3' GTCCACAGTCAGTCGATCACGACXTGCATGCCATCG 5'
3.	S3	5' *CAGGTGTCAGTCAGCTAGTGCTG 3' 3' GTCCACAGTCAGTCGATCACGAC-TGCATGCCATCG 5'
4.	S4	5' *CAGGTGTCAGTCAGCTAGTGCTG 3' 3' GTCCACAGTCAGTCGATCACGACATGCATGCCATCG 5'
5.	S5	5' *CAGGTGTCAGTCAGCTAGTGCTGC 3' 3' GTCCACAGTCAGTCGATCACGACXTGCATGCCATCG 5'
6.	S6	5' *CAGGTGTCAGTCAGCTAGTGCTGA 3' 3' GTCCACAGTCAGTCGATCACGACXTGCATGCCATCG 5'
7.	S7	5' *CAGGTGTCAGTCAGCTAGTGCTGA 3' 3' GTCCACAGTCAGTCGATCACGACXTGCATGCCATCG 5'
8.	S8	5' *CAGGTGTCAGTCAGCTAGTGCTGC 3' 3' GTCCACAGTCAGTCGATCACGACXTGCATGCCATCG 5'
9.	S9	5' *CAGGTGTCAGTCAGCTAGTGCTGA 3' 3' GTCCACAGTCAGTCGATCACGACATGCATGCCATCG 5'
10.	S10	5' *CAGGTGTCAGTCAGCTAGTGCTGA 3' 3' GTCCACAGTCAGTCGATCACGACGTGCATGCCATCG 5'
11.	S11	5' CAGGTGTCAGTCAGCTAGTGCTGAC 3' 3' GTCCACAGTCAGTCGATCACGACTGGCATGCCATCG 5'
12.	S12	5' *CAGGTGTCAGTCAGCTAGTGCTG 3' 3' GTCCACAGTCAGTCGATCACGACCTGCATGCCATCG 5'

Top strand: Primer, Bottom strand: Template, Bold red: T₋₁, Bold black: T₀
Orange pair: Phosphorothioate bond between them
Red *: 6-carboxyfluorescein (FAM) label.
Bold green: 3' Dideoxy terminated

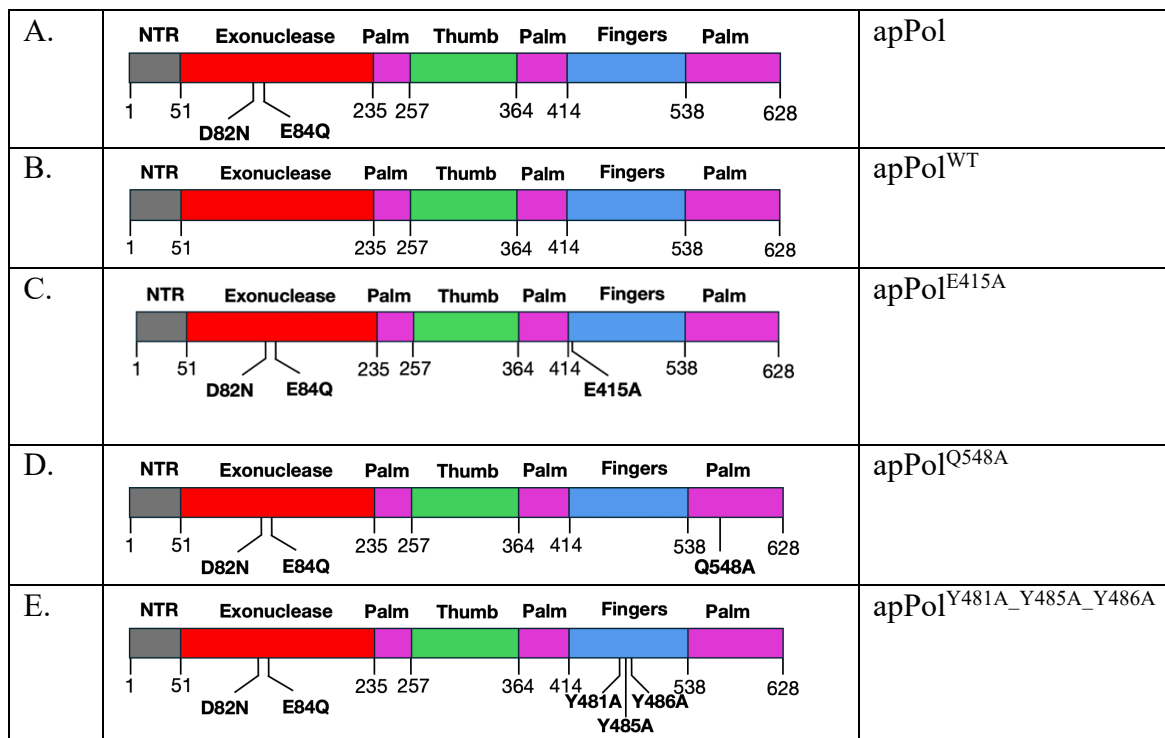


Figure 2.1: apPol constructs. Domain organization of *P. falciparum* apPol. Domain boundaries are based on the crystal structure of apPol (PDB: 5DKT). **(A)** The two residues (D82 and E84) of the proofreading exonuclease domain that were mutated (to N and Q, respectively) to eliminate the exonuclease activity, denoted as apPol in all chapters. **(B)** Wild type apPol without any mutations, denoted as apPol^{WT}. **(C)** One residue (E415) was mutated to A in addition to D82N and E84Q. **(D)** One residue (Q548) was mutated to A in addition to D82N and E84Q. **(E)** Three residues (Y481, Y486 and Y486) were mutated to A in addition to D82N and E84Q. NTD: N-terminal domain (grey), proofreading: 3' to 5' proofreading exonuclease domain (red). The palm, thumb, and fingers domains are coloured in magenta, green, and blue, respectively.

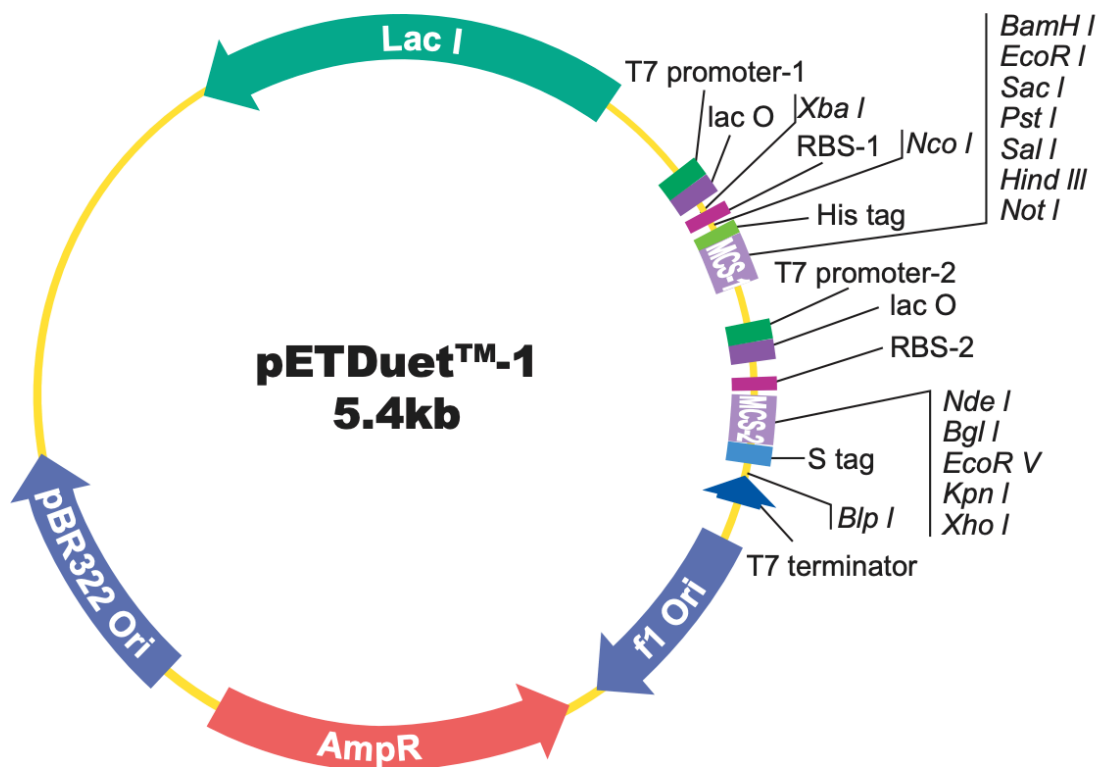


Figure 2.2: pET Duet-1 vector map. Map of the pETDuet™-1 expression vector (5.4 kb) (GenScript). The plasmid contains two T7 promoter-driven multiple cloning sites (MCS1 and MCS2) for co-expression of target genes, each preceded by a lac operator and ribosome-binding site. The vector also encodes an N-terminal His tag and an S tag for recombinant protein expression and purification, as well as a T7 terminator, a lacI repressor gene, an ampicillin resistance marker (AmpR), and the pBR322 origin of replication.

2.4 Cloning, Overexpression, and Purification of apPol

All apPol proteins (Figure 2.1) were overexpressed, and purified following similar strategies. The pET-Duet1 plasmid with the relevant apPol construct was transformed into Rosetta 2 (DE3) *E. coli* cells (Merck, USA). The cells were grown in autoinduction terrific broth (Formedium) at 37°C for ~5 h, and then growth was maintained at 20°C for ~20 h before harvesting the cells by centrifugation. All subsequent steps were carried out at 4°C. Cell pellets were resuspended in lysis buffer containing 50 mM Tris–HCl (pH 7.5), 800 mM NaCl, 25 mM imidazole, and 10% glycerol. To prevent proteolytic degradation and improve lysis efficiency, ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablet (Roche, USA) and lysozyme were added to the resuspended cells. Cells were lysed by sonication and clarified by centrifugation. The clarified lysate was loaded onto a 5 ml Hi-Trap Chelating HP column (Cytiva, USA) charged with Ni²⁺ and pre-equilibrated with lysis buffer. The unbound protein was washed with 10 column volumes (CVs) of lysis buffer, followed by an additional 10 CV wash with low-salt wash buffer (50 mM Tris–HCl (pH 7.5), 200 mM NaCl, 25 mM imidazole, and 10% glycerol). apPol was eluted over a 10 CV linear gradient of imidazole from 25 mM to 1 M. The elution fractions were analyzed using SDS-PAGE, and fractions containing apPol were pooled and diluted with no-salt buffer (50 mM Tris–HCl (pH 7.5), 25 mM imidazole, and 10% glycerol) to decrease the NaCl concentration to 100 mM. The diluted sample was loaded on a 5 ml Hi-Trap SP column (Cytiva, USA) pre-equilibrated with Buffer A (50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 5 mM EDTA, 5 mM β-mercaptoethanol, and 10% glycerol). The unbound samples were washed with 10 CVs of Buffer A, and apPol was eluted with a linear gradient of 0.1 to 1 M NaCl (10 CVs). For further purification, the eluent from the S-column was concentrated and then loaded onto a Superdex 200 Increase 10/300 size-exclusion chromatography column (Cytiva, USA) that

had been pre-equilibrated with storage buffer (50 mM Tris-Cl (pH 7.5), 200 mM NaCl, 2 mM DTT, and 20% glycerol). Fractions containing pure apPol were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80°C . Protein concentration was determined based on the theoretical extinction coefficient of $64180\text{ M}^{-1}\text{ cm}^{-1}$. The N-terminal hexa-histidine tag was not removed prior to using the protein.

2.5 Determination of optimal conditions for apPol activity

Single nucleotide primer extension assays (Figure 2.3) were performed to determine the reaction condition for optimum enzymatic activity of apPol. The baseline reaction was performed at 37°C in a buffer containing 20 mM Tris-Cl (pH 7.5), 10 mM MgCl_2 , 20 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, and 5% glycerol. All the other components were kept constant except the component whose effect was being tested. Varying conditions of MgCl_2 (2-20 mM), NaCl (20-100 mM), and pH (6.5-8.5) were tested. The assays were set up under burst conditions such that there was a slight excess of the DNA substrate after pre-forming the pre-chemistry binary complex (Figure 1.4, step 1). A final concentration of 60 nM apPol was incubated with 100 nM DNA (Table 1, S1), and the reaction was started by adding 500 μM dTTP. After incubating the reaction for various time intervals (0-1 s), reactions were quenched using 250 mM EDTA in 100% formamide and analyzed on 15% polyacrylamide (19:1) urea (7.5 M) gels run at $45\text{--}50^{\circ}\text{C}$. The gels were imaged on a Typhoon FLA7000 LASER-based scanner (Cytiva, USA) using an excitation wavelength of 488 nm (blue LASER) and an emission cutoff at 525 nm. This allowed the detection of the fluorescence signal from the FAM-labeled primer DNA (Table 1, top strands). The intensities of the gel

bands were quantitated using Image Quant software (Cytiva, USA), and the concentration of extended DNA was calculated using the following formula.

$$\text{extended product} = \frac{I^E - I^B}{(I^E - I^B) + (I^U - I^B)} \quad \text{Equation 3}$$

where I_E is the intensity of the extended primer band, I_B is the intensity of the gel background, and I_U is the intensity of the unextended primer band.

All primer extension assays were performed using an RQF-3 quench flow instrument (KinTek Corp., USA). The biphasic time courses were fit to the burst equation-

$$Y = A[(1 - e^{-(k_{fast}t)} + (k_{slow}t)] + C \quad \text{Equation 4}$$

where Y is the product concentration, A is the amplitude of the fast phase, k_{fast} and k_{slow} are the rates for the fast and slow phases, respectively, t is the time interval and C is a constant.

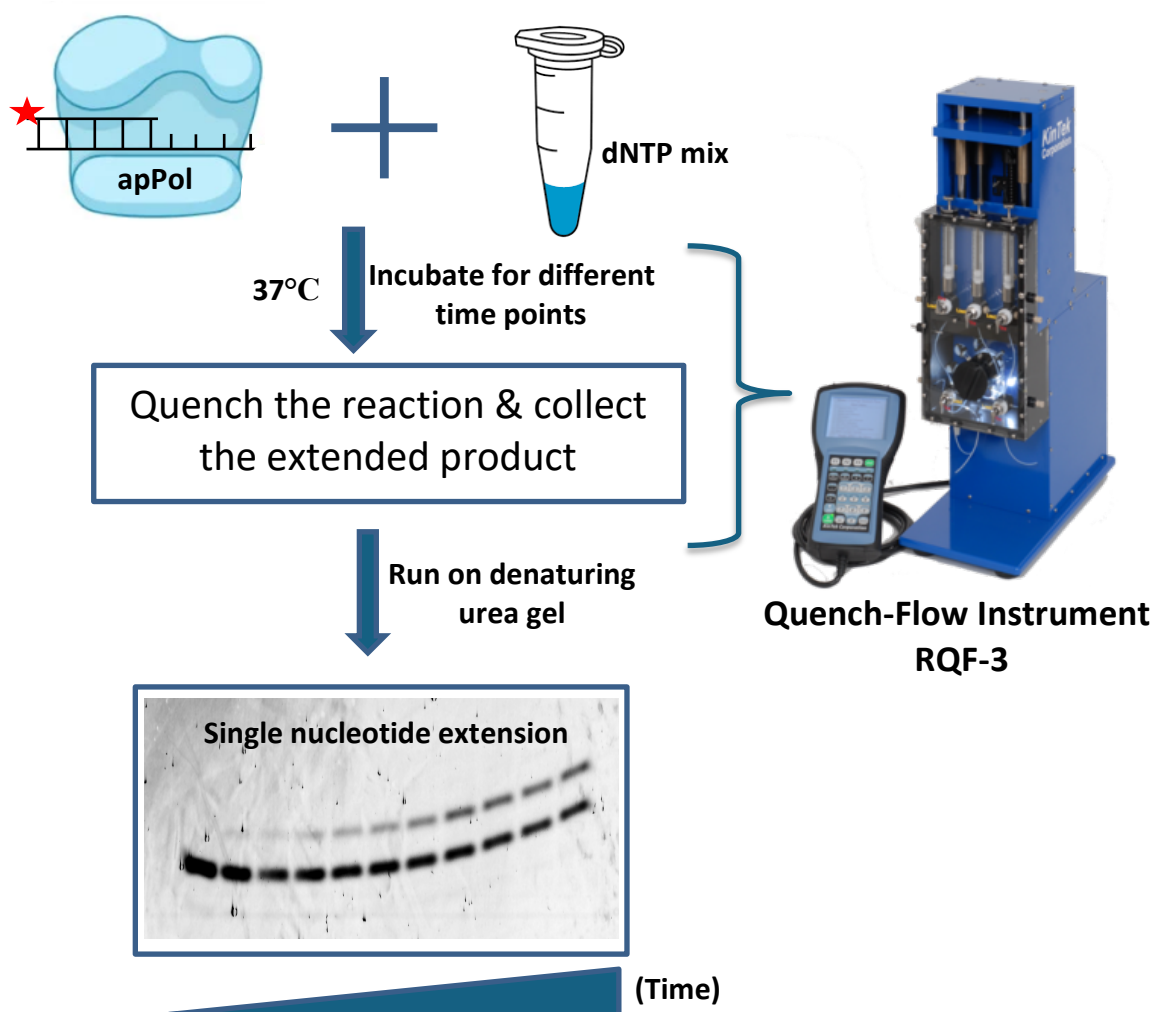


Figure 2.3: Primer extension assay: Schematic of primer extension assay setup, where pre-formed apPol•DNA binary complex was incubated with incoming nucleotide (either one in single nucleotide extension assay or with multiple nucleotides in multiple nucleotide extension assay) for various time intervals. The quench-flow Instrument RQF-3 was used for quenching the reaction after varying time intervals (milliseconds time range). The longer time intervals (>5 seconds) were quenched manually. The extended products were run on denaturing urea acrylamide gel providing single nucleotide resolution.

2.6 Single Nucleotide Primer Extension Assays

Single nucleotide primer extension assays (Figure 2.3) were performed at 37 °C in reaction buffer containing 20 mM Tris-Cl (pH 7.5), 10 mM MgCl₂, 30 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, and 5% glycerol. The assays were set up under burst conditions such that there was a slight excess of the DNA after pre-forming the pre-chemistry binary complex (Figure 1.4, step 1). A final concentration of 480 nM (104 nM was active) apPol was incubated with various DNA and nucleotide substrates as mentioned in the corresponding figure legends. The active fraction of apPol was estimated as described in the next section (Figure 3.6). The reactions were quenched using 250 mM EDTA and analyzed as described above (section 2.5).

All primer extension assays were performed using an RQF-3 quench-flow instrument (KinTek Corp., USA) except those with DNA substrate S3, which were performed by manual mixing of the reagents. The biphasic time courses were fit to the full burst equation-

$$Y = A \left[\left(\frac{k_{fast}}{k_{fast} + k_{slow}} \right)^2 (1 - e^{-(k_{fast} + k_{slow})t}) + \frac{k_{fast}k_{slow}t}{k_{fast} + k_{slow}} \right] + C \quad \text{Equation 5}$$

where Y is the product concentration, A is the amplitude of the fast phase, k_{fast} and k_{slow} are the rates for the fast and slow phases, respectively, t is the time interval, and C is a constant.

Time courses that did not show a biphasic behaviour were fit to the single exponential equation.

$$Y = A(1 - e^{-k_{obs}t}) + C \quad \text{Equation 6}$$

where Y is the product concentration, A is the amplitude, k_{obs} is the rate of product formation, t is the time interval, and C is a constant.

Active site titration assay -

Primer extension assays were performed with a final concentration of 480 nM apPol, 500 μM dTTP, and varying concentrations of DNA substrate S1 (50, 100, 200, 400, and 800 nM). The reactions were incubated for different time intervals (0 to 0.8 s) and quenched with EDTA. Product concentrations were plotted as a function of time, and data were fit to a full burst equation (Equation 5). The amplitudes calculated from the time courses were plotted as a function of DNA concentration, and the data were fit to the quadratic equation-

$$A = \frac{(K_{D,app}^{DNA} + E + D) - \sqrt{(K_{D,app}^{DNA} + E + D)^2 - 4ED}}{2} \quad \text{Equation 7}$$

where A is the amplitude, $K_{D,app}^{DNA}$ is the apparent equilibrium dissociation constant of the apPol•DNA binary complex, E is the active apPol concentration, and D is the DNA concentration.

Determination of slow post-chemistry step -

To determine whether product DNA release was the slow step in apPol's catalytic cycle, primer extension assay was performed under burst conditions with a final concentration of 480 nM apPol, 200 nM of labelled DNA substrate S1 (Table 1) with or without 1 μM of trap DNA (unlabelled DNA substrate, Table 1, S1) along with 500 μM dTTP, and The reactions were incubated for different time intervals (0 to 0.8 s) and quenched with EDTA (Figure 2.4). Product concentrations were plotted as a function of time, and data were fit either to a full

burst equation (Equation 5; when no trap DNA was present) or to a single exponential equation (Equation 6; when trap DNA was present).

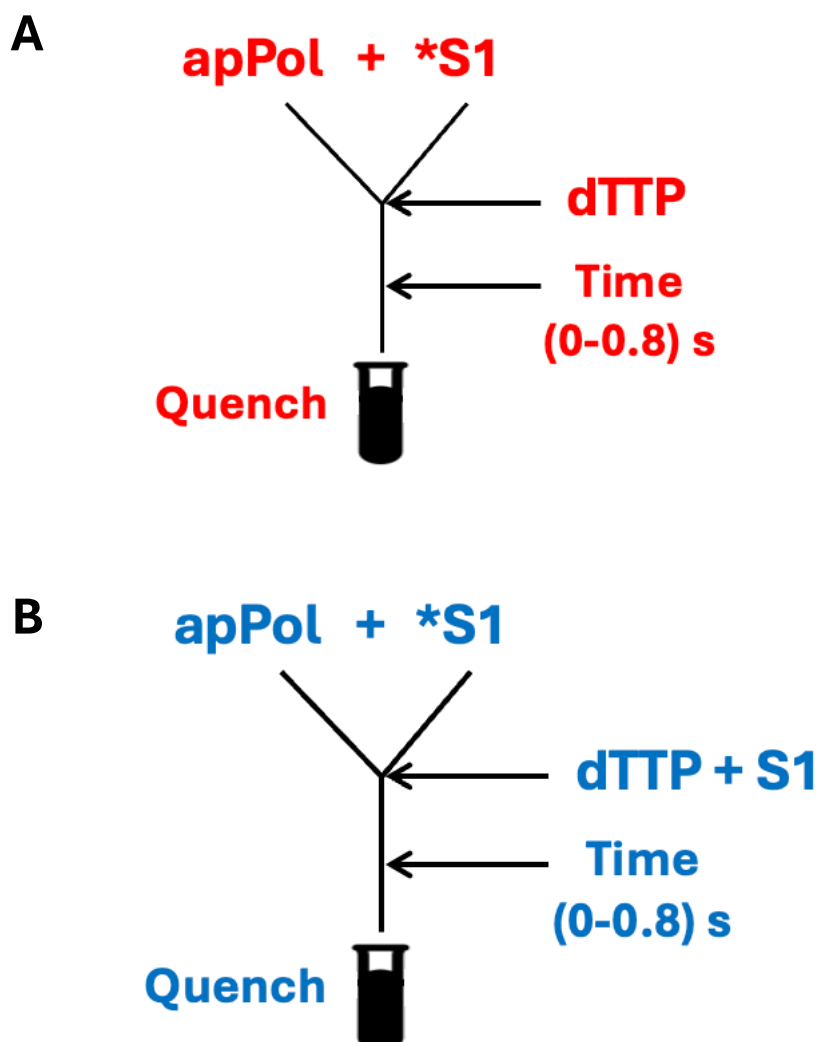


Figure 2.4: DNA trap experiment. Schematic of the experimental setup to ascertain the presence of a slow post-chemistry step. A pre-formed apPol-S1 complex, where * denotes FAM label, was rapidly with (A) 500 μ M dTTP (red) or (B) 500 μ M dTTP and 1 μ M trap DNA (unlabelled S1) (blue). The reactions were quenched with EDTA and analysed by gel electrophoresis. The red and blue colour are consistent with the corresponding plots in Figure 3.3.

Determination of rate of nucleotide incorporation by apPol –

To assess the rate of nucleotide incorporation (k_{pol}) and the apparent equilibrium dissociation constant ($K_{D,app}^{dNTP}$) of nucleotide binding to the pre-chemistry binary complex, primer extension assays were performed under burst condition with final concentrations of 104 nM active apPol, 100 nM DNA (the identity of the DNA substrate varied between S1, S2, and S3 depending on the experiment) (Table 1), and varying concentrations of the incoming dNTP (described in respective figures and figure legends). The reactions were incubated at 37°C for varying time intervals and quenched with EDTA. The time intervals were varied depending on the identity of the DNA and are mentioned in the respective figure legends. Extended products were plotted as a function of time, and data were fitted to either the burst equation (Equation 4) or the single exponential equation (Equation 6). k_{fast} (for biphasic time courses) or k_{obs} (for single exponential time courses) was plotted as a function of dNTP concentration, and the data were fit to the hyperbolic equation—

$$k = \frac{k_{pol} dNTP}{K_{D,app}^{dNTP} + dNTP} \quad \text{Equation 8}$$

where k is either k_{fast} or k_{obs} (depending on the nature of the time course), k_{pol} is the maximal rate of polymerization, $K_{D,app}^{dNTP}$ is the apparent dissociation constant of nucleotide binding, and $dNTP$ is the nucleotide concentration.

Identification of steric gate residue –

Single nucleotide primer extension assays were performed to ascertain the identity of the steric gate residue of apPol. A final concentration of 1 μ M apPol or apPol^{E415A} or apPol^{Q548A} (Figure 2.1A, C and D) was incubated with 50 nM DNA substrate S12 (Table 1). 125 μ M

dGTP or GTP was added to initiate the reaction. The reactions were incubated for varying time intervals ranging from 0 to 120 s and then quenched with EDTA. The products (primer strand extended by multiple or a single nucleotide) were analysed using the same protocol as mentioned above (section 2.8 and 2.6).

Effect of O1 helix checkpoint on primer extension activity -

To investigate the role of Y481, Y485, and Y486 in the primer extension activity of apPol, single nucleotide primer extension assays were performed with apPol (Figure 2.3) and apPol^{Y481A_Y485A_Y486A} (Figure 2.1 E). The assay condition is similar to as described above in this section with the following changes. CHAPSO was not included in the reaction buffer, and instead of multiple nucleotides, only a single nucleotide (either dGTP or dATP) was added. The extended products were analysed using a similar protocol as mentioned above (section 2.6).

2.7 Double Mixing Experiment to Determine the apPol-DNA Binary Complex Dissociation Rate

To measure the rate of dissociation of the apPol•DNA binary complex (Figure 1.4, step 5), a preincubated mixture of 208 nM active apPol and 400 nM DNA substrate S1 was mixed with an equal volume of 100 μ M trap DNA (S1 without the FAM label) such that the final concentrations of active apPol, S1, and trap were 104 nM, 200 nM, and 50 μ M, respectively. The mixture was incubated for various time intervals ranging from 0 to 6 s. This was followed by the addition of 500 μ M dTTP and a second incubation for 170 milliseconds before being quenched with excess EDTA (Figure 2.5). The product concentration was plotted as a function of the first incubation time, and the data was analysed by global fitting.

2.8 Multiple Nucleotide Primer Extension Assay

Multiple nucleotide primer extension assays were performed and analysed using a protocol similar to the one for single nucleotide primer extension assays (section 2.6) with a few differences. The different multiple nucleotide primer extension experiments performed in this thesis are described below.

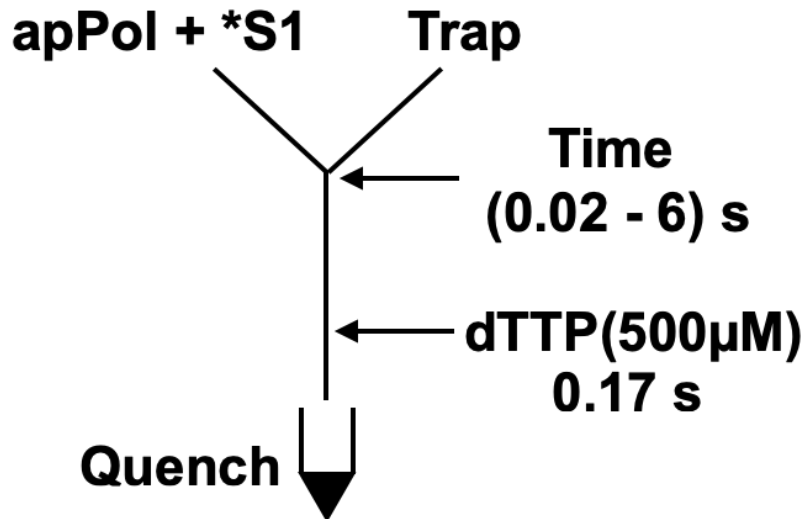


Figure 2.5: Double mixing experiment. Schematic of the experimental setup of double-mixing assay used to monitor the dissociation rate of apPol•DNA binary complex. A pre-formed apPol•*S1 complex, where * denotes FAM label, was rapidly mixed with a trap solution containing unlabelled DNA S1 and incubated for the indicated time interval before the addition of dTTP (500 µM). The reaction was further incubated for 0.17 seconds and then quenched with EDTA and analysed by gel electrophoresis.

Assay to calculate the average rate of processive synthesis -

Primer extension assays were performed with a final concentration of 104 nM active apPol; 200 nM DNA substrate S1; 250 μ M each of dTTP, dATP, and dCTP; and 50 μ M trap DNA (unlabeled S1). Reactions were incubated at 37°C for different time intervals ranging from 0 to 4 s before quenching with excess EDTA. The products were visualised using denaturing gel electrophoresis and the data were analysed by global fitting using the KinTek Explorer.

Assay to assess the lesion bypass capability of apPol against oxidative damages -

Multiple nucleotide extension assays were performed using a final concentration of 104 nM active apPol, 200 nM DNA substrate (S2, S3, or S4), and 250 μ M of each of the four dNTPs. Reactions were incubated at 37°C for 0 to 20 minutes and then quenched with excess EDTA. Extended products were visualised on 15% acrylamide-urea gel.

apPol's activity in presence of CHAPSO –

A multiple nucleotide primer extension assay was performed to verify apPol's activity in the presence of 8 mM CHAPSO. A final concentration of 480 nM apPol was incubated with 200 nM DNA substrate S12 (Table 1). The reaction was initiated by adding 250 μ M of each dNTP. The reaction was quenched after 10 or 20 seconds by adding 250 mM EDTA.

2.9 Primer Degradation Assay

Primer degradation assays were performed to investigate apPol's pyrophosphorolysis and 3' to 5' proofreading exonuclease activities. For assessing pyrophosphorolysis, a final concentration of 104 nM active apPol and 100 nM DNA substrate S1 was incubated with varying concentrations of inorganic pyrophosphate (2000, 1000, 500, 250, and 125 μ M) in apPol reaction buffer for 0 to 2 min followed by quenching with EDTA.

For assessing the proofreading activity of apPol, final concentrations of 104 nM apPol^{WT} and 200 nM DNA substrate (S7, S8, S9, or S10) were incubated with a final concentration of 10 mM MgCl₂ for 0 to 30 s before quenching with EDTA. Degraded products were visualised on a 15% acrylamide-urea gel, and the concentration of the shortened primer (by a single nucleotide) was plotted as a function of time. The time course was fit to the burst equation (Equation 4), and the rate of the fast exponential phase (k_{fast}) was approximated as the exonuclease rate (k_{exo}).

2.10 Determination of Kinetic Constants from Data Fitting Using Numerical Integration

Global fitting of the kinetic data using numerical integration was performed using the KinTek Explorer [131]. The kinetic pathways used for data fitting are described in the “Results” section. In all the cases, the association rate constant governing the binding of the incoming dNTP (Figure 1.4, step 2) to the pre-chemistry binary complex was locked at a diffusion-limited macromolecular association rate constant of 10 μ M⁻¹s⁻¹ [132]. 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 1D and 2D confidence contour

plots for all possible pairs of rate constants using the FitSpace function in KinTek Explorer [122]. The chi-square threshold suggested by the software was used to calculate the 95% confidence interval.

While fitting the primer extension data for the abasic site bypass (section 3.2.11), for the second dATP addition, the rate constant governing the chemistry step and the $K_D^{\text{dATP}_{2^{\text{nd}} \text{ addn.}}}$ could not be independently determined with high precision. However, the efficiency of the second addition given by the ratio of these two kinetic constants (rate constant governing second chemistry/ $K_D^{\text{dATP}_{2^{\text{nd}} \text{ addn.}}}$) was well constrained by the data. Accordingly, these two parameters were linked at their best fit values during global fitting and one-dimensional (1D) and two-dimensional (2D) confidence contour analysis.

2.11 Negative stain grid preparation

Ternary complex of apPol-DNA-dGTP was prepared by incubating 2 μM apPol with 4 μM of P/T DNA (3'-dideoxy-terminated primer DNA, S11, Table 1) and 1 mM dGTP at 37°C for 10 minutes in complex-formation buffer (20 mM Tris-Cl (pH 7.5), 10 mM MgCl_2 , 65 mM NaCl, and 1 mM DTT), complex was immediately used for negative stain grid preparation. 400 mesh copper grids with continuous carbon foil were glow discharged (using Cressington 208 carbon, High Vacuum Carbon Coater, Cressington Scientific Instruments) for 10 seconds to make the carbon foil hydrophilic. 4 μl of 2 μM complex was applied to the grid and incubated for 5 seconds. The grid was stained using 0.75% uranyl formate and then the excess stain was blotted using a filter paper. All negative stain grids were screened on a Philips CM 100 TEM at a nominal magnification of 28000X leading to a pixel size of 6.84 Å/pixel at the object level.

2.12 CryoEM sample preparation

A ternary complex of apPol, DNA substrate S11 (Table 1), and dGTP was prepared by incubating 20 μ M apPol with 30 μ M primer DNA and 6 mM dGTP at 37°C for 10 minutes in complex-formation buffer (20 mM Tris-Cl (pH 7.5), 10 mM MgCl₂, 65 mM NaCl, and 1 mM DTT). This complex was used for cryo-EM dataset 1. The ternary complex for collecting cryo-EM dataset 2 was formed similar to dataset 1 with the following three alterations. First, the primer strand was not 3' dideoxy-terminated (Table 1, S11); second, the complex-formation buffer had 10 mM CaCl₂ instead of 10 mM MgCl₂; and third, 4 mM of dGTP was added instead of 6 mM. 8 mM CHAPSO was added to the samples immediately before grid preparation (Figure 2.6).

2.13 Cryo-EM grid preparation

All cryo-EM grids were prepared using a Leica EM GP automatic plunge freezer (Leica Microsystems). Quantifoil R1.2/1.3 400 mesh holey carbon grids (SPT Labtech) were glow-discharged using Cressington 208 carbon coater (Cressington Scientific Instruments) and 4 μ L of sample (described in section 2.11) was applied to a glow-discharged grid. The excess liquid was blotted manually using a Whatman number 1 filter paper. Then an additional 4 μ L sample was applied to the grid, and the grid was put in a Leica EM GP for blotting and vitrification.

Before settling onto the above-mentioned grid preparation condition, several grid types and preparation conditions were tested to improve particle distribution and reduce sample aggregation and denaturation. These included Quantifoil 0.6/1 Au 300 mesh carbon grids

(SPT Labtech), Quantifoil R2/2 Cu 400 mesh graphene oxide (GO) grids (SPT Labtech), 0.1% poly L-lysine hydrobromide-coated grids, and addition of 0.01% Tween 20 to the complex formation buffer.

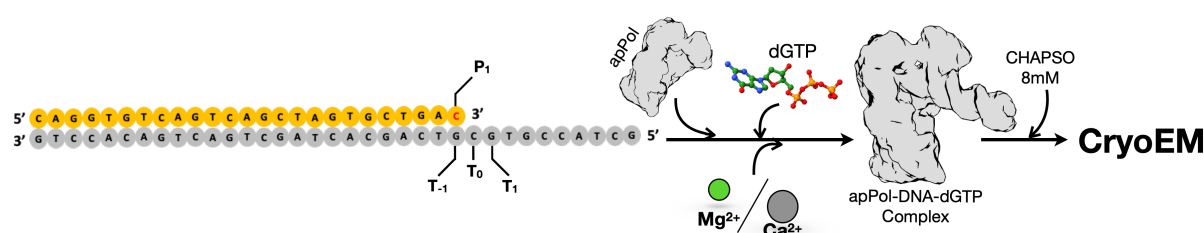


Figure 2.6: Cryo-EM sample preparation. Schematic representation of apPol complex assembly. A synthetic primer/template DNA (primer: orange, template: grey) was incubated with apPol and dGTP in the presence of either Mg^{2+} or Ca^{2+} to form the apPol pre-chemistry complex. 8 mM CHAPSO was added just before cryo-EM grid preparation. The templating base (T_0) and the ones immediately 5' and 3' to it (T_1 and T_{-1} , respectively), along with the base at the 3' end of the primer strand (P_{-1}), are highlighted. The P_{-1} nucleotide (dC; red) is dideoxy terminated when Mg^{2+} is used as the divalent metal.

2.14 Cryo-EM imaging and 3D reconstruction

Cryo-EM grids were screened on an FEI Tecnai Arctica electron microscope (FEI, USA) operated at 200kV and the images were recorded using a Falcon III direct electron detector (DED) operated in linear mode. The final high-resolution cryo-EM datasets were collected on a Titan Krios (ThermoFisher Scientific) at the Electron Bio-Imaging Centre (eBIC) operated

at 300 kV. Dose fractionated movies were recorded on a K3 direct electron detector (Gatan Inc.) operated in super-resolution mode. Data collection was automated using EPU (ThermoFisher Scientific) and the images were collected at a nominal magnification of 81,000X such that the object level pixel size was 1.06 Å/pixel (super-resolution pixel size: 0.53 Å/pixel). The images were recorded as 1 second movies divided into 70 frames. The total dose and fluence were 70 electron/Å² and 1 electron/Å²/frame, respectively for dataset 1 and 60 electron/Å² and 1.2 electron/Å²/frame, respectively, for dataset 2.

All image processing and three-dimensional (3D) reconstruction jobs (Figure 4.4) were performed using CryoSPARC (version 4.5.1) [123] (Structura Biotechnology Inc.). The individual super-resolution movie frames were binned by 2, and the frames were aligned using alignparts_lmbfgrs [133] as implemented within CryoSPARC. The contrast transfer function (CTF) of the micrographs were estimated using the patch CTF routine of CryoSPARC.

For dataset 1, after removing unsuitable micrographs (micrographs with a defocus higher than -3.8 µm, 0.5 CTF fit resolution worse than 6 Å, and relative ice thickness higher than 1.1 were rejected), 14,519 micrographs were retained out of 16,197 for further processing. Particles were picked using Topaz [124]. The training set contained ~1400 manually picked particles from micrographs with varying defocus. At this stage, 1,922,870 particles were selected. These particles were subjected to multiple rounds of 2D classification, and 898,336 particles contributing to optimal 2D class averages were retained for further processing. These particles were used for reference-free 3D reconstruction using Ab-initio reconstruction [123]. One of the classes (354,534 particles) showed clear features consistent with apPol and was taken forward for further processing. These particles were refined using the non-uniform

refinement routine [134] resulting in a consensus map with a nominal resolution of 3.2 Å. Multiple rounds of 3D classification (without alignment) were done, including masked classification [135], focusing on the DNA and nucleotide substrates to resolve the heterogeneity in the dataset. This resulted in four different classes defined as binary1, binary2, ternary1 and ternary2. Residual beam-induced motion and defocus estimation errors of the particles contributing to these classes were rectified using the reference-based motion correction and per-particle CTF estimation routines of CryoSPARC. This was followed by 3D refinement (using the non-uniform refinement routine of CryoSPARC) of each of the 3D classes resulting in the final maps.

For dataset 2, after removing unsuitable micrographs (micrographs with average defocus greater than -3.8 µm, 0.5 CTF fit resolution worse than 5 Å, full frame motion greater than 27.4 pixels and relative ice thickness greater than 1.0 were rejected). 11,231 micrographs were retained out of 11,944 for further processing. Particles were picked using Topaz. The training set contained ~1400 manually picked particles from micrographs with varying defocus. At this stage, 1,993,221 particles were selected. These particles were subjected to multiple rounds of 2D classification and 238,125 particles contributing to optimal 2D class averages were retained for further processing. These particles were used for reference-free 3D reconstruction using Ab-initio reconstruction and one class resembling apPol (102,768 particles) was taken forward for further processing. Multiple rounds of heterogeneous refinements were performed to further eliminate particles that do not contribute to high-resolution reconstruction. and 92,787 particles were taken forward. The residual beam-induced motion was rectified, and the defocus of these particles were re-estimated using the reference-based motion correction and per-particle CTF estimation routines of CryoSPARC

respectively. This was followed by 3D refinement (using non-uniform refinement) leading to a 3.5 Å map of the ternary₃ complex.

The resolutions of all cryo-EM maps are reported based on the 0.143 cutoff of the gold-standard Fourier shell correlation (FSC) curves [136] [137]. For better map interpretability, the refined maps were sharpened either by applying a negative B factor [138], estimated during non-uniform refinement, or using DeepEMhancer [139]. For DeepEMhancer-based sharpening, the tightTarget model was used.

2.15 Atomic model building and refinement

The published apo apPol structure [96] (PDB ID: 5DKT) was used as the starting model for apPol. The starting model for the primer/template DNA was generated from the structure of *B. stearothermophilus* DNA polymerase I (BSt Pol I) (PDB ID: 1LV5) [140]. The starting models were fit into the sharpened cryo-EM maps using the ‘Fit in Map’ tool in ChimeraX [141]. The fitted models were combined to generate the apPol•DNA model for binary₁; the combined models were relaxed through a molecular dynamics based simulation run in ISOLDE [127], which uses energy minimization and restrained dynamics to relax strained geometry, clashes, and unrealistic conformations to settle into a physically more plausible state while being guided by the map density.

The fitted and relaxed models were examined in Coot [142] and fits of side chain rotamers and main chain were optimized either by real space refinement within Coot or in ISOLDE and subjected to real space refinement within Phenix [143]. The models were refined using ‘real space refinement’ routine in Phenix. During refinement secondary structure restraints were used to restrain protein and nucleic acid geometry. For ternary_{1,2} and 3 ligand restraint for dGTP was generated using eLBOW ligand restraint generation [144]. The refined structures were analysed in Coot and rebuilt where necessary to optimize model geometry

(assessed by MolProbity [145]) and the fit of the model within the cryo-EM map (Table 3). To estimate over-fitting, FSC curves were calculated between the cryo-EM map and the atomic model (map-to-model FSC [138]) in Phenix.

2.16 Structure and sequence analysis

All maps and models were visualized in UCSF ChimeraX, and unless otherwise stated, the map density depicted in the figures 4.5 corresponds to the maps sharpened using DeepEMhancer. Map segmentation was performed using Segger [146]. For comparison between different apPol complexes, residues 405-411 and 572-585 of apPol's palm domain were used for alignment. The movies depicting morphs between the different states of apPol were prepared in PyMOL (Schrödinger). Protein sequence alignment was performed using Clustal Omega [147] and the alignment was visualized using Jalview [148].

Table 2: Cryo-EM map and atomic model refinement statistics.

	Binary1 (EMDB- 53378) (PDB 9QUJ)	Binary2 (EMDB- 53379) (PDB 9QUN)	Ternary1 (EMDB- 53376) (PDB 9QUA)	Ternary2 (EMDB- 53374) (PDB 9QU8)	Ternary3 (EMDB- 53391) (PDB 9QV9)	Consensus (EMDB- 53335) (PDB 9QSC)
Data collection and processing						
Magnification	81000X	81000X	81000X	81000X	81000X	81000X
Voltage (kV)	300	300	300	300	300	300
Electron exposure (e ⁻ /Å ²)	70	70	70	70	60	70
Defocus range (μm)	-0.8 to -3.8	-0.8 to -3.8	-0.8 to -3.8	-0.8 to -3.8	-1 to -3.8	-0.8 to -3.8
Pixel size (Å)	1.06	1.06	1.06	1.06	1.06	1.06
Symmetry imposed	None	None	None	None	None	None
Initial particle images (no.)	1,922,870	1,922,870	1,922,870	1,922,870	1,993,221	1,922,870
Final particle images (no.)	53,452	20,205	9,780	10,241	92,787	354,534
Map resolution (Å)	3.7	3.9	4.2	4.2	3.5	3.2
FSC threshold						
Map resolution range (Å)	2.8 to 5.0	3.0 to 5	3.5 to 6.0	3.5 to 6.0	2.7 to 6.0	2.5 to 5.0
Refinement						
Initial model used (PDB code)	5DKT and 1LV5	5DKT	5DKT and 1LV5	5DKT and 1LV5	5DKT and 1LV5	5DKT and 1LV5
Model resolution (Å)	4	4.2	4.4	4.6	4.1	3.4
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5
Model resolution range (Å)						
Map sharpening <i>B</i> factor (Å ²)	-133	-114	-124	-134	-155	-169
Model composition						
Non-hydrogen atoms	6022	5243	6072	6032	6006	5601
Protein residues	627	628	627	627	621	626
Ligands	N/A	1	1	1	2	N/A
<i>B</i> factors (Å ²)						
Protein	72.04	100.6	109.66	120.11	78.18	80.89
Nucleotide	151.78	N/A	178.25	173.81	99.19	175.19
Ligand	N/A	114.94	124.95	135.29	76.81	
R.m.s. deviations						
Bond lengths (Å)	0.005	0.004	0.004	0.004	0.005	0.004
Bond angles (°)	0.585	0.600	0.756	0.757	0.709	0.489
Validation						
MolProbity score	1.42	1.43	1.47	1.70	1.64	1.39
Clashscore	4.42	4.17	4.9	6.45	5.54	5.58
Poor rotamers (%)	0.51	0.17	0.17	1.02	0.00	0.51
Ramachandran plot						
Favored (%)	96.8	96.49	96.64	95.04	95.12	97.60
Allowed (%)	3.04	3.35	3.2	4.8	4.72	2.24
Disallowed (%)	0.16	0.16	0.16	0.16	0.16	0.16

3. Transient State Kinetic of Apicoplast DNA Polymerase from Plasmodium Falciparum

3.1 INTRODUCTION

Preliminary steady state kinetics of apPol showed that this enzyme can copy oxidative lesions including 8-oxodG, 2-hydroxy-dA and thymine glycol in addition to undamaged DNA.

However, as explained in section 1.5.1, single nucleotide incorporation assays performed under steady state condition do not provide an accurate picture of the catalytic cycle of a DNA polymerase [106]. We determined the kinetic mechanism of action of apPol using pre-steady state kinetics assays to provide the first description of the kinetic mechanism of apPol.

We discovered that apPol catalyses DNA extension with low efficiency and processivity, properties that might prevent this polymerase from performing replicative synthesis.

Moreover, while apPol can perform robust TLS against oxidatively damaged nucleotides (ODNs), this bypass is error-prone. This work lays down a foundation for understanding the mechanism of organellar genome duplication in *Plasmodium* apicoplast.

3.2 RESULTS

3.2.1 apPol purification

All constructs of apPol were purified using Ni-NTA, cation exchange, and size-exclusion chromatography. SDS-PAGE analysis of the purified protein is shown in Figure 3.1. apPol migrates slightly lower than its theoretical molecular weight of 76 kDa. The final concentration of purified protein was measured based on the absorbance of the protein solution at 280 nm (A_{280}) and a theoretical extinction coefficient of $64180 \text{ M}^{-1}\text{cm}^{-1}$.

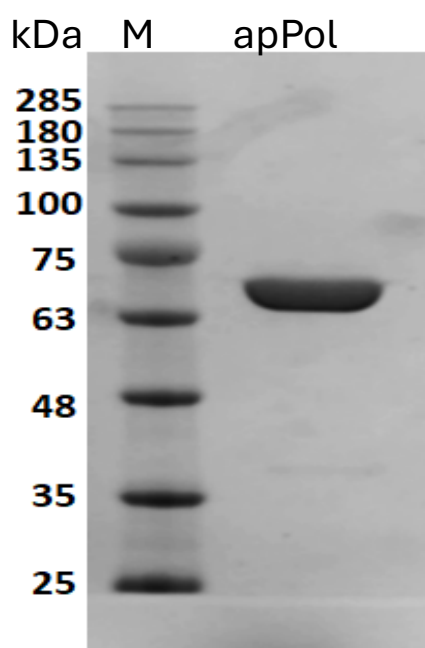


Figure 3.1: SDS-PAGE of apPol. Purified apPol obtained after size-exclusion chromatography was resolved on a 10% SDS-polyacrylamide gel and stained with Coomassie Brilliant Blue. The pre-stained molecular weight marker is shown in lane M, and purified apPol is shown in the apPol lane. The protein migrates slightly below its theoretical molecular mass of 75 kDa.

3.2.2 Optimal reaction conditions for apPol.

Reaction conditions for apPol were optimized by quantifying the incorporation of the correct nucleotide on DNA substrate S1 (Table 1). All reaction conditions were kept constant, except the one whose effect was being tested. The variable reaction conditions were concentration of MgCl_2 , concentration of NaCl , and pH of the buffer (Figure 3.2 A-C). apPol display activity over a wide range of NaCl concentration with the maximum rate of primer extension observed for 30 mM NaCl (Figure 3.2 A). apPol displayed a wide activity range for MgCl_2 , and within a concentration range of 5 to 20 mM MgCl_2 there was minimal variation in apPol's activity (Figure 3.2 B). Dependence of primer extension on the buffer pH increased with increasing pH, which is consistent with an acid-base reaction. The activity at pH 8.5 was slightly higher than that at pH 7.5 (Figure 3.2 C). Based on these results and considering the physiological conditions, all subsequent reactions were performed in a reaction buffer containing 20 mM Tris-Cl (pH 7.5), 30 mM NaCl , and 10 mM MgCl_2 .

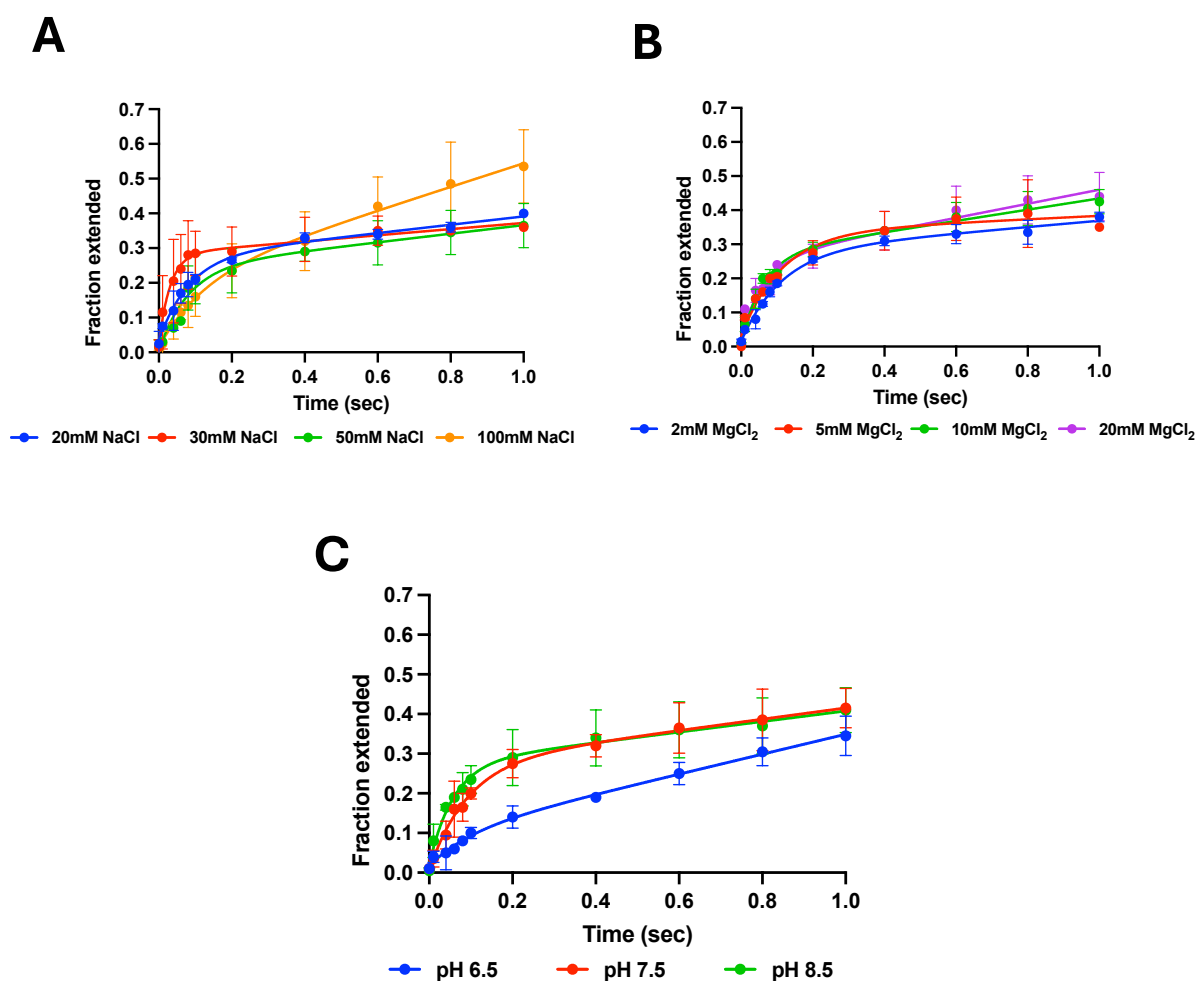


Figure 3.2: Reaction condition optimization for primer extension assays. Time courses of single nucleotide primer extension by apPol. (A) Effect of NaCl concentration (B) Effect of Mg²⁺ concentration, and (C) Effect of pH on primer extension. Primer extension assays were carried out under pre-steady-state burst condition by adding 500 μ M dTTP (the correct incoming dNTP) to a pre-incubated solution of 100 nM P/T DNA and 60 nM apPol. Reactions were quenched after varying time intervals (0-1 s) by addition of excess EDTA. Unextended and extended primers were separated by gel electrophoresis on a 15% denaturing urea-acrylamide gel. Fraction of primer DNA extended was determined from equation 3.

3.2.3 Kinetic Mechanism of Correct Nucleotide Incorporation by apPol.

We performed transient-state kinetic assays under burst condition to describe apPol's catalytic cycle of correct nucleotide incorporation opposite undamaged DNA. In these assays, comparable concentrations of apPol and DNA were pre-incubated, and then the correct incoming nucleotide was added rapidly to initiate the reaction. The DNA or dNTP concentration was varied, and the reactions were quenched using EDTA after varying time intervals depending upon the different experimental setup. The extended product was visualised on denaturing urea-acrylamide gels and quantified as described in the Methods section.

We initially performed data fitting using traditional non-linear regression and used the resulting kinetic constants as a starting point for global fitting. Non-linear regression relies on simplifying assumptions about the rate equations and may often neglect intrinsic relationships among observables from different experiments. To address these limitations, primer extension data from different experiments were fit globally using KinTek Explorer, which allowed us to fit all the experiments to a single kinetic model using numerical integration. This global fit yielded the kinetic constants that define the minimal catalytic pathway for correct nucleotide incorporation by apPol.

3.2.4 Phosphodiester Bond Formation is Followed by Slow Dissociation of the Product DNA.

To determine if the catalytic step of bond formation (Figure 1.4; step 4) is followed by a slow post-chemistry step we performed a single nucleotide primer extension assay under burst condition (Figures 1.9, 2.3 and 2.4) using DNA substrate S1 (Table 1). Such an assay allowed

us to monitor the kinetics of both the first round and a few subsequent rounds of primer extension. A biphasic time course of nucleotide incorporation was observed, and the data was fit to the full burst equation (Equation 5) with a fast exponential phase ($k_{\text{fast}} = 11.2 \pm 2.9 \text{ s}^{-1}$), followed by a slower linear phase ($k_{\text{slow}} = 0.37 \pm 0.3 \text{ s}^{-1}$) (Figure 3.3, red curve). These observations can be explained by a preformed apPol•DNA binary complex rapidly forming the elongated product (Figure 1.4, DNA_{n+1}) in the first round of catalysis, followed by a slow post-chemistry step that would limit the rate of subsequent turnover of apPol and result in the slow linear phase. Such biphasic time courses have been observed for most DNA polymerases.

While the biphasic nature of product formation indicates the presence of a slow post-chemistry step, it does not identify the rate-limiting step. For most DNA polymerases, dissociation of the post-chemistry binary complex (Figure 1.4, step 5) is the rate-limiting step and thus dominates the rate of the slow linear phase. To determine whether product DNA release was the slow step in apPol's catalytic cycle, primer extension assay was performed under burst conditions, but an excess of unlabelled trap DNA (substrate S1 without the FAM label; Table1) was added along with the incoming nucleotide (dTTP) (Figure 2.4 B). As before, the pre-bound labelled DNA would rapidly convert to product, but upon dissociation from the post-chemistry binary complex (apPol•DNA_{n+1}) (Figure 1.4, step 4) apPol would be 'trapped' by the excess unlabelled DNA, preventing any subsequent rounds of catalysis from being detected. In the presence of a trap, the slow phase was absent (Figure 3.3, blue curve), indicating that dissociation of the post-chemistry binary complex is indeed the rate-limiting step in apPol's catalytic cycle. The rate of product formation was $16.9 \pm 1.9 \text{ s}^{-1}$, which is a close match with k_{fast} in the absence of a trap.

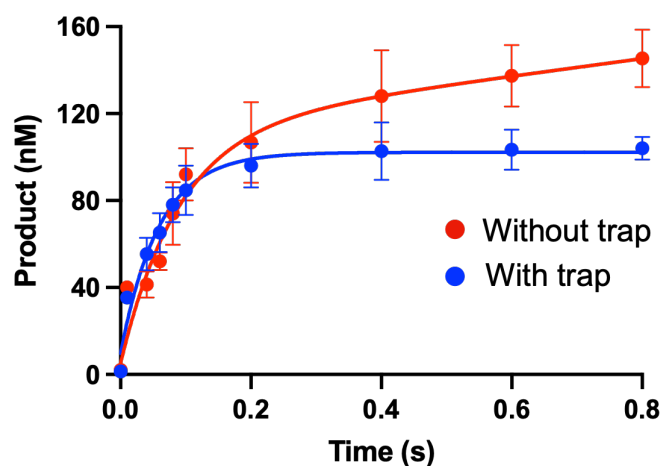


Figure 3.3: Detection of the presence of slow post-chemistry step in apPol's catalytic cycle. Time course of product formation from single nucleotide primer extension under pre-steady state condition in the presence (blue) or absence (red) of unlabelled DNA acting as trap for apPol. 480 nM apPol (corresponding to an active concentration of 104 nM) was pre-incubated with 200 nM of labelled DNA substrate S1. The reaction was started with 500 μ M dTTP (with or without 1 μ M unlabelled P/T DNA). The reaction were quenched after various time intervals ranging from 0 to 0.8 s with 250 mM EDTA. In the absence of the trap, the time course was biphasic in nature, and the data were fit to the full burst equation (Equation 3). In the presence of trap, the time course was monophasic, and the data were fit to the single exponential equation (Equation 4).

3.2.5 ApPol Performs Pyrophosphorolysis Only in the Presence of High Concentration of Pyrophosphate.

Rapid enzyme turnover is a necessity for efficient catalysis. Therefore, generally the product(s) of an enzyme-catalysed reaction have low affinity for the enzyme itself. Consistent with this, when DNA polymerases catalyse nucleotide incorporation, the PPi byproduct typically gets released rapidly from the post-chemistry ternary complex (Figure 1.4, step 4), and this step generally occurs prior to or in conjunction with the translocation step of the polymerization cycle [149, 150], ensuring that bond formation (Figure 1.4, step 3) is essentially irreversible. The elongated DNA (Figure 1.4, DNA_{n+1}) is an exception to the general rule of fast product release because the product DNA from one round of nucleotide incorporation becomes the substrate for the next round, and hence, the extended DNA typically has a high affinity for the DNA polymerase. *In vitro*, the build-up of a high concentration of PPi in the presence of a DNA polymerase has been shown to result in pyrophosphorolysis, that is, the reversal of the phosphodiester bond formation [151].

We performed primer degradation assays in the presence of varying concentrations of PPi, and primer degradation could be detected with 250 μ M PPi, with the degradation increasing at higher PPi concentrations, indicating that apPol can catalyse pyrophosphorolysis but only when there is a large buildup of PPi (Figure 3.4 A). We could not detect any pyrophosphorolysis in the presence of 800 nM PPi (Figure 3.4 B).

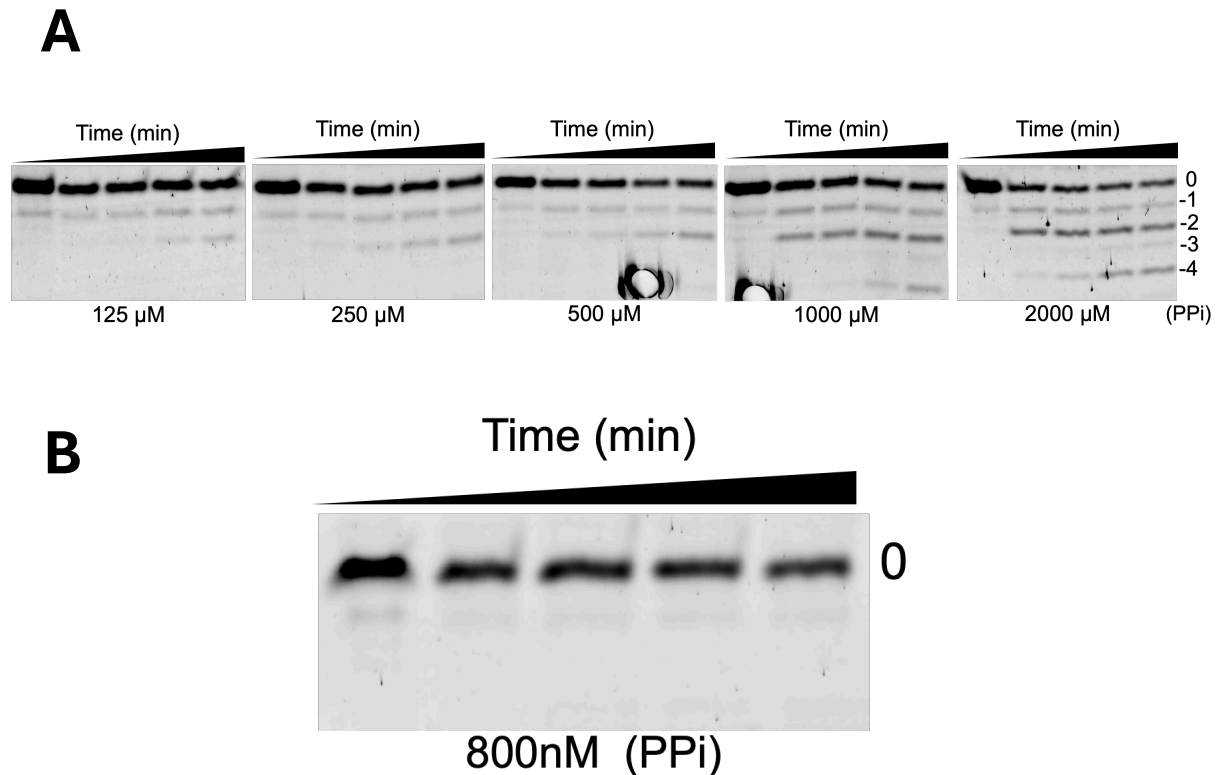


Figure 3.4: Pyrophosphorolysis catalyzed by apPol. (A) Acrylamide-urea denaturing gels showing primer degradation in the presence of increasing (from left to right) concentrations of inorganic pyrophosphate (PPi). A final concentration of apPol, DNA substrate and reaction conditions were similar to panel A and varying concentrations of PPi ranging from 125 to 2000 μ M in apPol reaction buffer. 0: the starting primer without any degradation. -1, -2, -3, and -4: primer strand degraded by one, two, three, and four nucleotides, respectively. **(B)** 15% acrylamide-urea gel depicting primer degradation assay in the presence of 800 mM PPi. A final concentration of 104 nM active apPol was incubated with 100 nM of DNA substrate S1 and 800 nM PPi in apPol reaction buffer. The reactions were incubated at 37°C for various time intervals ranging from 0 to 2 min (0, 0.33, 0.5, 1 and 2 min) and then quenched with excess EDTA. The samples were then analysed on a 15% acrylamide-urea denaturing gel. 0: undegraded primer strand.

Based on the lack of product formation in primer extension assays performed with a nucleotide analogue where the bridging oxygen between the β and γ phosphates was replaced by a nitrogen (Figure 3.5 A), it has been proposed that the PPi generated during apPol-mediated dNTP incorporation gets hydrolysed to inorganic phosphates (Pi), followed by dissociation of the Pi from the post-chemistry complex [152]. This hypothesis is in contradiction to our observation. If it is necessary for PPi to be hydrolysed to Pi before product release from the apPol post-chemistry complex, then Pi, and not PPi, should bind to the apPol•DNA post-chemistry binary complex. Consequently, PPi would not be able to drive apPol-catalysed pyrophosphorolysis. However, as mentioned above, apPol performs robust pyrophosphorolysis in the presence of PPi. Moreover, our attempts to measure the amount of Pi released during primer extension by apPol could not detect any Pi above the background (Figure 3.5 B) Thus, we conclude that apPol does not hydrolyse PPi to Pi. However, it remains possible that such hydrolysis is required for the functioning of certain TLS polymerases, which has been reported previously. [152] [153, 154].

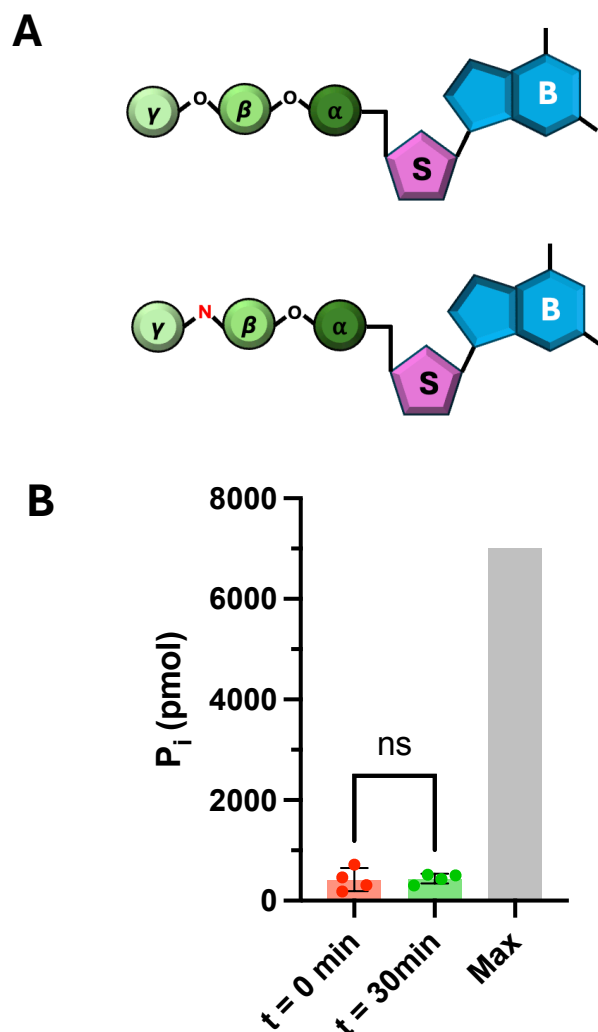


Figure 3.5: Phosphate release assay. (A) Schematic representation of a nucleotide analogue in which the bridging oxygen between the β and γ phosphates is replaced by nitrogen. The upper panel shows the native triphosphate linkage, whereas the lower panel shows the corresponding non-hydrolyzable β, γ -imido analogue. B, nucleobase; S, sugar; α , β , and γ , phosphate groups. (B) Detection of the amount of P_i released during replication of the M13mp18 single stranded (ss) DNA by apPol. The red bar ($t = 0$ min) shows the amount of phosphate present in the assay before the start of the polymerization reaction. The green bar ($t = 30$ min) shows the amount of phosphate generated after 30 minutes of the polymerization reaction. The grey bar (Max) shows the maximum amount of phosphate expected to be generated if P_i was the product released from the post-chemistry complex. The experiments were performed four times, and the average of the four data sets are plotted as translucent bars while the error bars represent the standard deviation of the data. ns: Not significant based on paired parametric t test.

3.2.6 apPol Binds Weakly to DNA Due to Rapid Dissociation of the Binary Complex.

Replicative DNA polymerases that copy the bulk of a genome need to remain bound to the substrate DNA for multiple cycles of DNA synthesis. This would require a high affinity for the DNA substrate and A-family replicative polymerases such as the bacteriophage T7 DNA polymerase and the mitochondrial DNA polymerase gamma tightly bind DNA in the presence of their respective processivity factors (Table 3). However, for apPol, no processivity factor has been identified to date, raising the possibility that apPol might have an intrinsically high affinity for DNA, allowing the enzyme to perform replicative synthesis without a processivity factor, as is seen for the bacteriophage phi29 DNA polymerase [155].

We performed primer extension assays under burst condition with varying concentrations of DNA substrate S1 (Table 1) to determine apPol's affinity for DNA. During the fast phase of product formation, only the fraction of DNA that pre-forms the binary complex (Figure 1.4, step 2) gets extended. Therefore, the amplitude of the fast phase is indicative of the concentration of the productive pre-formed binary complex. From the relation between the amplitudes of the fast phase and S1 concentration we determined the apparent affinity of apPol for DNA (K_{Dapp}^{DNA}) and the active fraction of apPol. We found that out of 480 nM apPol, 104 ± 6.4 nM was active, indicating that ~22% of the enzyme was active which indicates only 22% of the total enzyme population produced extension product under these assay conditions (Figure 3.6). Using this active fraction, global fitting of all product formation time courses (described in all subsequent sections) was performed to determine the kinetic constants governing the minimal kinetic pathway for correct dNTP incorporation by apPol (Figures 3.7 and 3.8).

Table 3: Kinetic parameters of different DNA polymerase. Kinetic parameters for nucleotide incorporation opposite undamaged dNTP by different DNA polymerase. *: replicative polymerase, **: TLS polymerase, ***: DNA repair polymerase. K_D^{dNTP} : the affinity of the nucleotide for the binary complex, k_{pol} : the rate of dNTP incorporation, k_{off} : the rate of dissociation of DNA from binary complex.

DNA polymerase	$K_D^{DNA}(\text{nM})$	$K_D^{dNTP}(\mu\text{M})$	$k_{pol}(\text{s}^{-1})$	$k_{off}(\text{s}^{-1})$	Processivity (k_{pol}/k_{off})	Efficiency (k_{pol}/K_D^{dNTP}) ($\mu\text{M}^{-1}\text{s}^{-1}$)	Ref.
<i>P. falciparum</i> apPol*	78	131	30	1.9	16	0.23	This study
T7 Pol-thioredoxin *	18	18	287	0.2	1435	16	[151]
Human Pol gamma-PCNA*	9.9	0.78	45	0.02	2250	58	[156]
Mammalian Pol delta-PCNA*	64	0.93	21	0.006	3500	22.6	[157]
<i>S. cerevisiae</i> Pol epsilon holoenzyme*	22	11	411	0.0058	70862	37.4	[158]
<i>E. coli</i> Pol IV**	50	441	12	0.18	67	0.03	[159]
Rat pol beta***	49	110	10	0.3	163	0.09	[160]

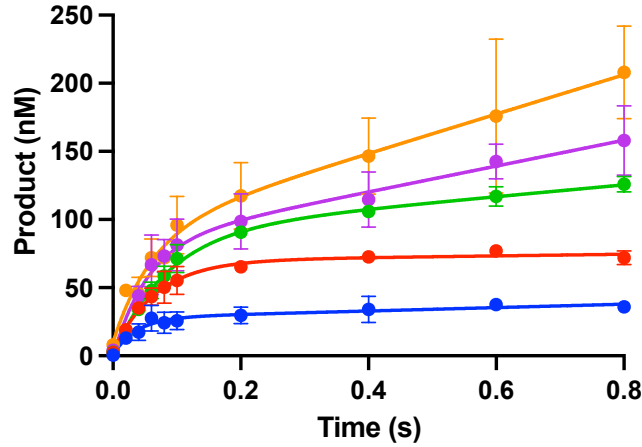
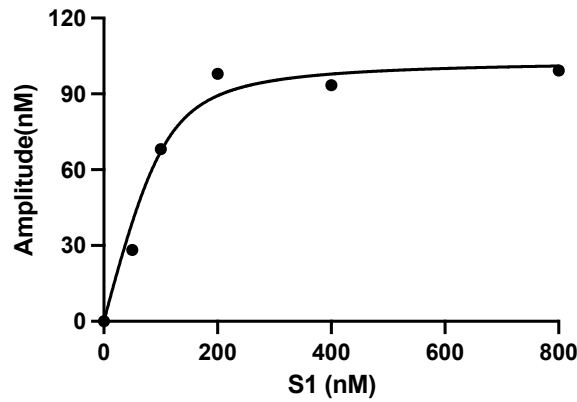
A**B**

Figure 3.6: Determination of K_D^{DNA} and active site titration. Active site titration of apPol analysed by non-linear regression. **(A)** Time course of product formation from single nucleotide primer extension under pre-steady state condition using 480 nM apPol and varying concentration of labelled DNA substrate S1 ranging from 50 to 800 nM (50 nM: blue, 100 nM: red, 200 nM: green, 400 nM: magenta, and 800 nM: orange), then 500 μM dTTP was added to start the reaction. The reaction were quenched at various time points ranging from 0 to 0.8 s with 250 mM EDTA. Time courses were fit to the full burst equation. **(B)** The amplitudes of the fast exponential phases of product formation plotted as a function of DNA substrate S1 concentration. The data was fit to a quadratic equation and from the fit $K_{D,\text{app}}^{\text{DNA}}$ was calculated to be 77.9 ± 9 nM and the active fraction of apPol was 104 ± 6.4 nM, which is $\sim 22\%$ of the total apPol (480 nM) added.

The sequence of events for the minimal pathway (Figure 3.7 A) is similar to the general kinetic scheme (Figure 1.4). PPi release was assumed to be fast and irreversible such that this step was not explicitly modelled. Pyrophosphorolysis, that is, the reverse of the phosphodiester bond formation, was not detected in the presence of 800 nM PPi (Figure 3.4 B) (the highest concentration of PPi that could be generated in the single nucleotide primer extension assays included in the global fit) (Figure 3.7), which supported the assumption of PPi release being essentially unidirectional under these reaction conditions. With rapid PPi release, no build-up of the post-chemistry ternary complex (Figure 1.4, Pol•DNA_{n+1}•PPi) should occur, and it may be anticipated that any slow reversibility of the chemistry step will not become apparent under these experimental conditions. Therefore, the chemical step of bond formation (Figure 1.4, step 3) was modelled to be irreversible. We note that the kinetic pathway used for global fitting is perhaps the minimal catalytic cycle of apPol. One or more conformational steps might be required to lead the initial-collision pre-chemistry ternary complex (Figure 3.7 A, Pol•S1_n•dTTP) to a catalytically poised state. However, these steps are not modelled explicitly since the kinetic experiments performed in this study would not be able to constrain the rate constants governing the conformational changes. The kinetic constants governing the catalytic pathway were determined by numerical integration and readily explained all the pre-steady state experimental data (Figure 3.7, A-C) with 1D and 2D confidence contour analysis indicating that all rate constants are well-constrained (Figure 3.8).

Based on global fitting, the dissociation constant (K_D^{DNA}) is defined as the ratio of the dissociation rate constant (k_{-1}) and the association rate constant (k_1), for the pre-chemistry binary complex, it was determined to be 77.9 nM (Figure 3.7 B; k_{-1}/k_1). Compared to other replicative DNA polymerases, apPol has a somewhat weaker affinity for DNA (Table 3). The weak K_D^{DNA} can result from a rapid dissociation of the binary complex or a slow association

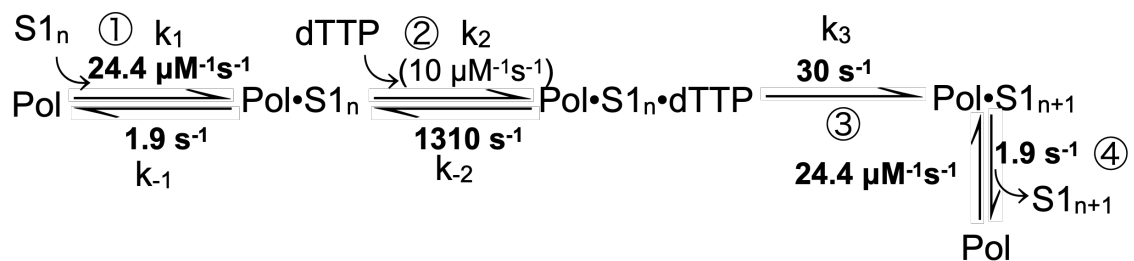
of apPol and DNA. To determine the cause of the weak K_D^{DNA} , we directly measured the dissociation rate of the apPol•DNA binary complex using a double-mixing experiment (Figure 2.5). apPol and fluorescently labelled DNA substrate S1 were preincubated to form the apPol•S1 binary complex. This sample was then incubated with an excess unlabelled DNA trap for varying time intervals. Any polymerase molecule that dissociated from the labelled DNA would bind to the unlabelled trap. Thus, with increasing incubation time, the concentration of the apPol•S1 complex would decrease as a function of the dissociation rate of the binary complex. After incubation, a saturating amount of dTTP was added to extend S1 still bound to apPol by a single nucleotide, and the reaction was terminated after 170 milliseconds. The concentration of the extended labelled DNA product was plotted as a function of the first incubation time (Figure 3.7 C), and the data were globally fit to the minimal pathway. The dissociation rate constant (k_{-1}) was 1.9 s^{-1} (Figure 3.7, k_{-1}), which is 1 to 2 orders of magnitude faster than the k_{-1} for most replicative DNA polymerases in the presence of their corresponding processivity factors (Table 3). Based on K_D^{DNA} and k_{-1} we determined the association rate constant for the apPol•DNA binary complex (k_1) to be $24.4 \mu\text{M}^{-1} \text{ s}^{-1}$ (Figures 3.7 A; k_1), indicating a diffusion-limited association of the binary complex. Taken together, our results indicate that apPol has a weak affinity for its DNA substrate primarily due to the rapid dissociation of the apPol•DNA binary complex.

3.2.7 apPol Incorporates Nucleotides with Low Efficiency and Processivity.

During genome duplication, a high efficiency of dNTP incorporation, combined with high processivity, ensures that replicative DNA polymerases can rapidly add tens of thousands of nucleotides to the growing DNA strand. The efficiency of nucleotide incorporation is determined by two kinetic parameters, the affinity of the nucleotide for the binary complex (K_D^{dNTP}) and the rate of dNTP incorporation (k_{pol}). To determine these two parameters single

nucleotide primer extension assays we performed under burst condition using DNA substrate S1 (Table 1) while the concentration of the incoming nucleotide (dTTP) was varied. As expected, the time course of product formation was biphasic, and k_{fast} increased with increasing concentration of dTTP and saturated hyperbolically (Figure 3.7 D and Appendix A1). This would be consistent with the nucleotide binding to the pre-chemistry binary complex coming to a rapid equilibrium, followed by a slower catalytic step. From global fitting, the dissociation constant for the apPol•DNA•dNTP pre-chemistry ternary complex (K_D^{dNTP}) was determined to be 131 μM (Figures 3.7 D; k_{-2}/k_2) and the rate constant governing nucleotide addition was 30 s^{-1} (Figures 3.7 D, k_3). Compared to other replicative polymerases, K_D^{dNTP} of apPol was weaker by almost 1 or 2 orders of magnitude (Table 3). For example, the mitochondrial DNA polymerase gamma holoenzyme (Pol gamma) has a K_D^{dNTP} of 0.78 μM , while the K_D^{dNTP} for the mammalian DNA polymerase delta-PCNA complex (Pol delta-PCNA) is 1 μM (Table 3). The weak K_D^{dNTP} of apPol, combined with a modest rate of bond formation, resulted in a low efficiency (k_3/K_D^{dNTP}) of correct nucleotide incorporation (0.23 $\mu\text{M}^{-1}\text{s}^{-1}$) (Table 3).

A



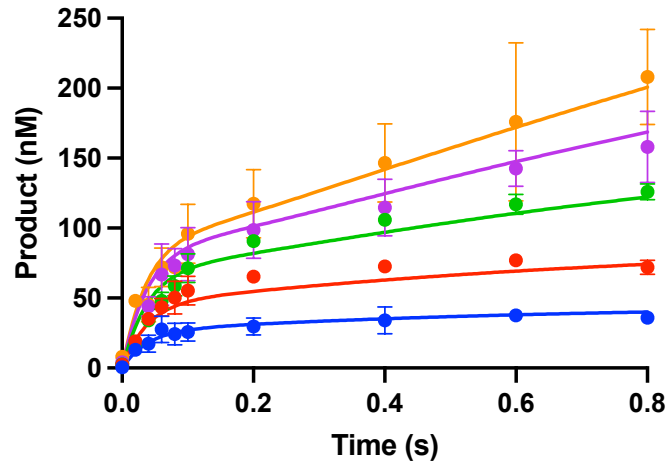
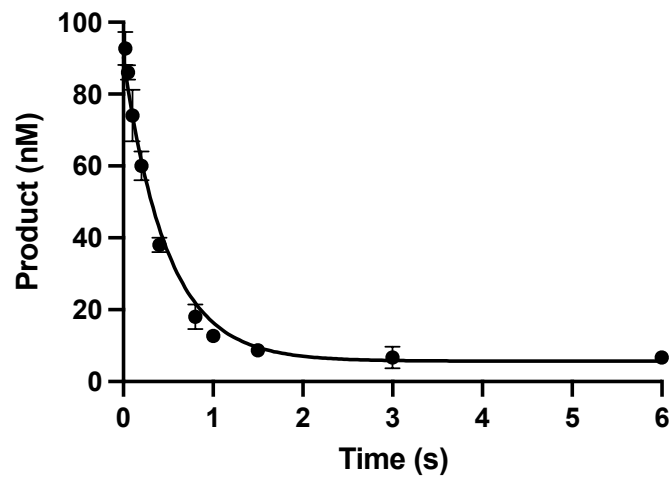
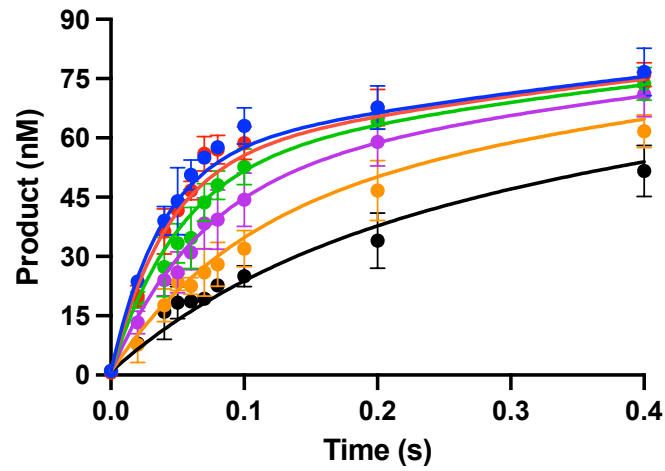
B**C****D**

Figure 3.7: Global fit of the minimal kinetic pathway governing the correct nucleotide incorporation by apPol opposite to an undamaged template. (A) Catalytic pathway for dTTP addition against undamaged DNA. Kinetic model used for global fitting. The best fit values of the rate constants governing the different steps are highlighted in bold. The rate

constants in brackets were not allowed to vary during global fitting. S1_n: DNA substrate S1 with 23 nucleotides long primer strand. S1_{n+1}: S1 with a primer strand extended by a single nucleotide. **(B)** Active site titration of apPol. Single nucleotide primer extension assays were set up as described in Figure 3.6 and the same data were fit by numerical integration. Colour scheme was similar to figure 3.6 A. **(C)** Determination of the dissociation rate of the prechemistry binary complex. 400 nM DNA substrate S1 was preincubated with 208 nM active apPol. This sample was incubated with an equal volume of 100 μ M trap DNA (S1 without the FAM label) for various time intervals ranging from 0.02 to 6 s. Therefore, the final concentrations of active apPol, S1, and trap DNA were 104 nM, 200 nM, and 50 μ M, respectively. After incubation, an equal volume of 1 mM dTTP was added to the reaction and incubated for another 0.17 s. The reaction was quenched with 250 mM EDTA, and the sample was analysed on a denature acrylamide-urea gel. The product formed (substrate S1 extended by one nucleotide) was plotted as a function of the first incubation time, and the data were fit by numerical integration. Based on the fit, the rate constant governing dissociation of the prechemistry binary complex (k_{-1}) was calculated to be 1.9 s⁻¹. **(D)** Primer extension assays for determination of the rate of polymerization and the affinity of dTTP for the prechemistry binary complex (K_D^{dTTP}). Single nucleotide primer extension assays were set up as described in Appendix A1. and the time courses were fit by numerical integration. Labelled as similar to figure 3.6 A. K_D^{dTTP} was 131 μ M, while k_3 , the rate constant governing the chemistry step, was 30 s⁻¹. All the experiments were performed in triplicate, and the average of the three independent datasets is plotted in the graphs shown in (B-D), while the error bars represent the standard deviation (SD) of datasets. The smooth lines overlaying the data represent the best fit based on global fitting.

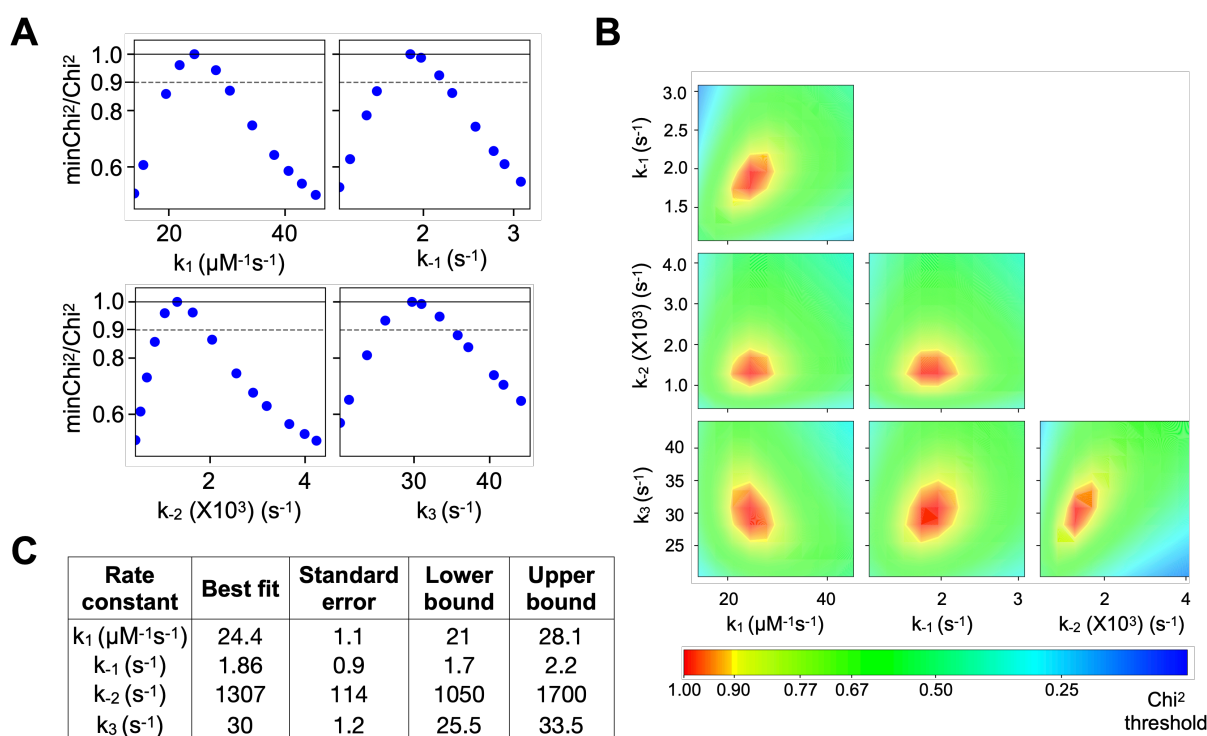


Figure 3.8: Confidence contour analysis for the kinetics of correct nucleotide incorporation opposite undamaged template by apPol. (A) One-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in Figure 3.7. The forward rate constant for nucleotide binding was fixed at a diffusion limited rate constant of $10 \mu\text{M}^{-1}\text{s}^{-1}$. The dashed lines represent the minimum $\text{chi}^2/\text{chi}^2$ threshold value of 0.9 (recommended by the FitSpace routine in KinTek Explorer) based on which the lower and upper bounds of the rate constants were determined. A ratio of 1 represent the best-fit value. (B) Two-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in Figure 3.7. The heat maps display the confidence intervals at the indicated chi^2 thresholds for each pair of rate constants. (C) Table listing the best fit values, standard errors and the upper and lower bounds (determined from the 1D and 2D confidence contour analysis) of the rate constants governing the kinetic pathway shown in Figure 3.7.

Processivity of a DNA polymerase is a measure of the number of nucleotides added by the enzyme per DNA binding event and can be expressed as the ratio of the rate constants governing the steps of bond formation and DNA dissociation (Figure 3.7; k_3/k_{-1}). For apPol

processivity was calculated to be ~16 (Figure 3.7, Table 3), which is 2 to 3 orders of magnitude lower compared to the corresponding ratio for typical replicative DNA polymerases in the presence of their corresponding processivity factors (Table 3). Taken together, this study shows that under our experimental conditions, apPol performs nucleotide incorporation with low efficiency and processivity when compared to most other replicative DNA polymerases. This raises the possibility that the interaction of apPol with other components of the apicoplast replisome might influence these kinetic parameters and convert apPol into an efficient replicative DNA polymerase.

3.2.8 apPol Translocates Rapidly along the DNA Substrate.

In vivo, replicative polymerases perform processive synthesis by translocating along the DNA substrate and moving to the next base on the template strand (Table 1, substrate S1, position T₁) from the templating position (Table 1, substrate S1, position T₀). In this scenario, the slow rate-limiting step of post-chemistry binary complex dissociation (Figure 1.4, step 5) does not occur after every dNTP addition and is instead replaced by the translocation step (Figure 1.4, Translocation). The rate of translocation has been estimated directly or indirectly for several DNA polymerases, and it is typically a rapid process. [161, 162]. On the other hand, a recent kinetic study found PPi release and translocation to be partially rate-limiting for the T7 DNA polymerase–thioredoxin complex (T7 Pol) at low temperatures [163]. To assess apPol’s speed of translocation, processive synthesis was mimicked by performing a primer extension assay with substrate S1 in the presence of multiple nucleotides. 250 μ M each of dTTP, dATP, and dCTP was added along with an excess of the unlabelled DNA trap (Figure 3.9). The templating sequence of S1 (Table 1 and Figure 3.9 A) is such that the

addition of these three dNTPs should result in the elongation of the primer strand by exactly three nucleotides. The presence of trap DNA ensured that apPol molecules that dissociate from the fluorescently labelled DNA are sequestered and cannot rebind to labelled S1. The three extension products (Figure 3.9 A) were quantitated and plotted individually as a function of time (Figure 3.9 B). This data could be explained by a simplified mechanism describing processive synthesis (Figure 3.10 A), where fast translocation is assumed, and therefore, this step was not explicitly modelled. The average rate of nucleotide addition was 8.4 s^{-1} with the rates of all three individual nucleotide additions being similar (Figure 3.10 A). This rate is comparable to the k_{fast} of 18.8 s^{-1} recorded from the single nucleotide incorporation experiment in the presence of $250 \text{ } \mu\text{M}$ dTTP (Figure 3.3). Taken together, these results indicate that the rate of translocation is at least comparable to the rate of bond formation and is not rate-limiting during processive synthesis by apPol.

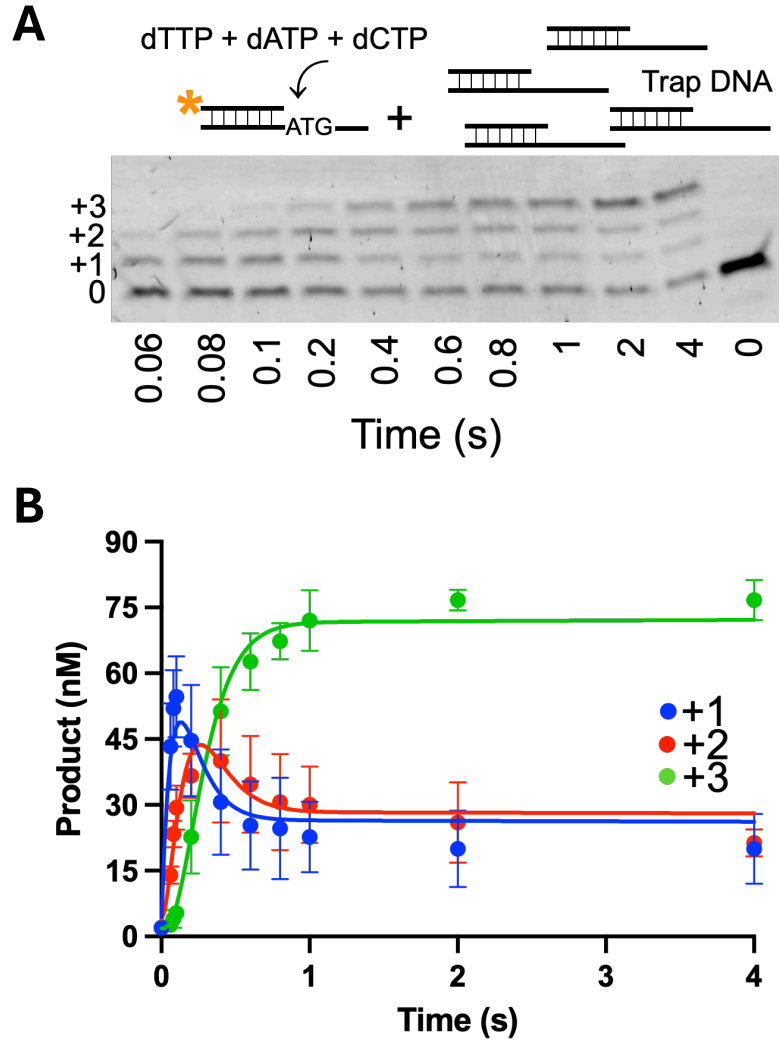


Figure 3.9: Kinetics of processive synthesis by apPol. (A) Representative acrylamide-urea gel showing primer extension in the presence of multiple nucleotides. The schematic of the reactants (substrate S1, the nucleotides, and the trap DNA) added is shown above the gel. A final concentration of 200 nM DNA substrate S1 was incubated with 104 nM active apPol; 250 μ M each of dTTP, dATP, and dCTP; and 50 μ M trap DNA at 37°C for various time intervals ranging from 0 to 4 s. The reactions were quenched with 250 mM EDTA and analysed on an acrylamide-urea gel. Orange *: FAM label. 0: unextended primer. +1, +2, and +3: primer strand extended by one, two, and three nucleotides, respectively. (B) Concentrations of the DNA primer strands extended by one (blue), two (red), and three (green) nucleotides from (A) plotted as a function of reaction time. The smooth lines over laying the datapoints represent the best global fit of the data to the kinetic scheme shown in Figure 3.10 A. The experiment was performed in triplicate, and the average of the three independent datasets is plotted in (B). Error bars represent standard deviation.

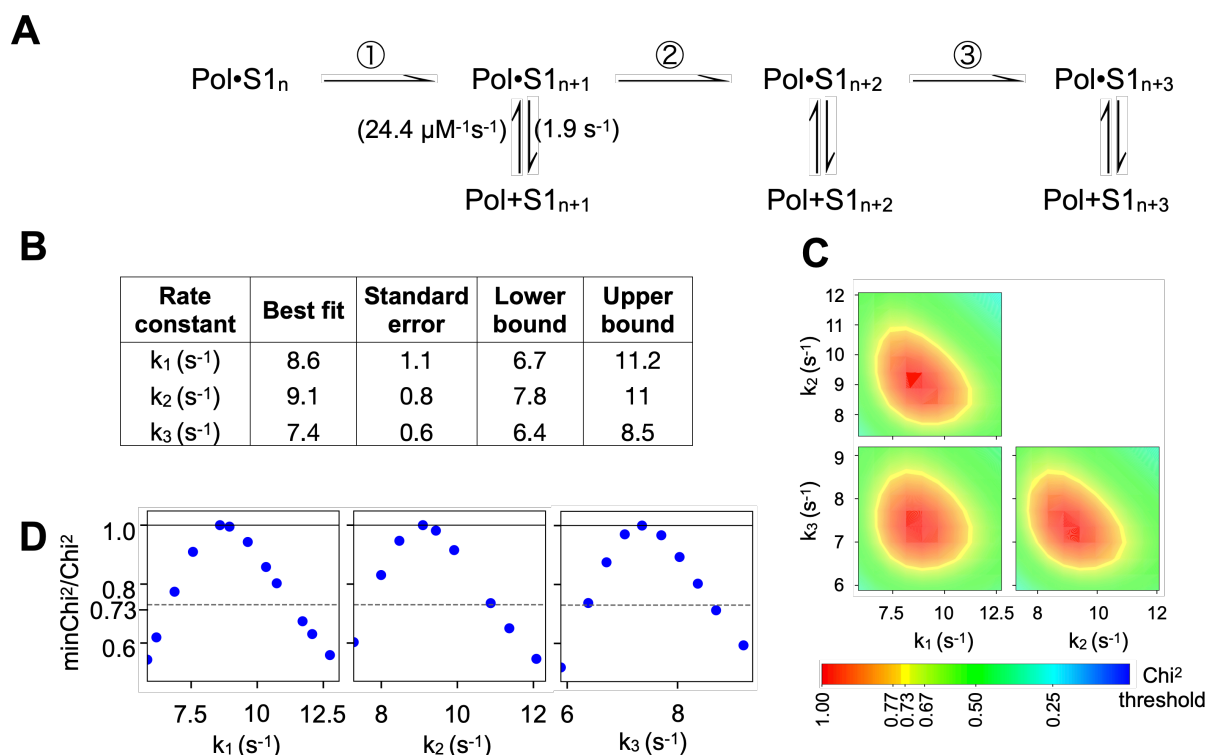


Figure 3.10: Confidence contour analysis for the kinetics of processive synthesis on undamaged template by apPol. **(A)** Kinetic scheme used for global fitting of the time courses shown in Figure 3.9 B. The forward and reverse rate constants governing the formation of the binary complexes were fixed at $24.4 \mu\text{M}^{-1}\text{s}^{-1}$ and 1.9s^{-1} respectively, based on the corresponding rate constants determined from the single nucleotide incorporation kinetics of apPol (Figure 3.6 and 3.7). For clarity, the fixed rate constants are shown (in brackets) for only one binary complex ($\text{Pol} \cdot \text{S1}_{n+1}$) formation step. S1_n : Substrate S1 with 23 nucleotide long unextended primer strand. S1_{n+1} , S1_{n+2} , S1_{n+3} : Substrate S1 with primer strand extended by one nucleotide, two and three nucleotides respectively. **(B)** Table listing the best fit values, standard errors and the upper and lower bounds (determined from the 1D and 2D confidence contour analysis) of the rate constants governing the kinetic pathway shown in panel A. **(C)** Two-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in panel A. The heat maps display the confidence intervals at the indicated χ^2 thresholds for each pair of rate constants. **(D)** One-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in panel A. The dashed lines represent the minimum χ^2/χ^2 threshold value of 0.73 (recommended by the FitSpace routine in KinTek Explorer) based on which the lower and upper bounds of the rate constants were determined. A ratio of 1 represent the best-fit value.

3.2.9 Kinetics of Lesion Bypass by apPol

The oxidizing environment of the apicoplast indicates that apDNA and the apicoplast nucleotide pool will encounter high levels of reactive oxygen species (ROS) [164]. ROS are potent mutagens and cause oxidative damage to nucleobases [165]. As the only DNA polymerase within the apicoplast, we anticipate apPol to encounter oxidatively damaged nucleotides (ODNs). Using pre-steady state primer extension assays, I have examined whether apPol can perform efficient TLS on encountering an ODN in the template strand.

3.2.10 apPol Can Bypass Oxidatively Damaged Nucleotides (ODNs)

Multiple nucleotide primer extension assays were performed to determine whether apPol can bypass ODNs containing DNA substrates and can replicate without stalling at those damages. Two critical lesions: 8-oxo-7,8-dihydroguanosine monophosphate (8-oxo-dGMP) and apurinic/apyrimidinic site (abasic site) (Table 1, substrates S2 and S3, respectively), were tested. We found that apPol can bypass both the ODNs, albeit to different extents (Figure 3.11). apPol performed a robust bypass of 8-oxo-dGMP and could extend till the end of the DNA template (Figure 3.11 A). However, the bypass of the abasic site was considerably less efficient with extension past the lesion being negligible (Figure 3.11 B). Moreover, upon reaching the end of the DNA, apPol added an extra nucleotide to the nascent DNA strand (Figure 3.11, A, C, "•"), and this non-templated nucleotide addition was independent of the presence of an ODN.

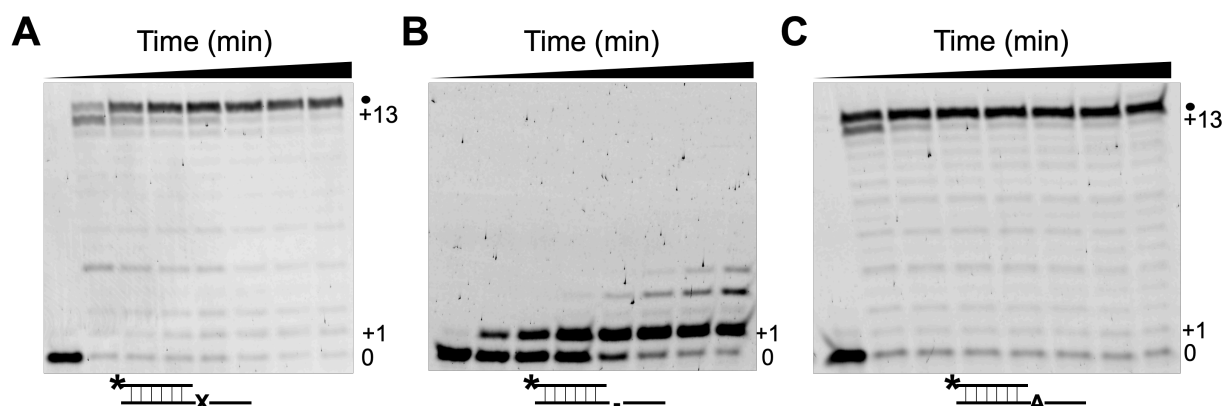


Figure 3.11: Translesion synthesis by apPol in the presence of multiple nucleotides.

(A–C) Acrylamide-urea gels depicting multiple nucleotide extensions of DNA substrate S2 (8-oxo-dGMP at the templating position) (A) S3 (abasic site at the templating position) (B) and S4 (undamaged DNA) (C). The schematics of the DNA substrates are shown below the respective gels. In all three cases, a final concentration of 104 nM active apPol was incubated with 200 nM DNA and 250 μ M of each of the four dNTPs. The reactions were incubated at 37°C for various time intervals ranging from 0 to 20 minutes (0, 0.5, 1, 2, 5, 10, 15, and 20 minutes) and then quenched by adding excess EDTA. The reactions were analysed on a 15% acrylamide-urea gel. 0: 23 nucleotide long primer strand (this is the starting length of the primer), +1: primer strand extended by one nucleotide, +13: primer strand extended by 13 nucleotides, which is the length of the 5' overhang of the template strand, and •: 37 nucleotides long primer strand. The length is one nucleotide more than the length of the template strand and is formed due to one non-templated nucleotide addition by apPol. *: the FAM label on the primer strand, X: 8-oxo-dGMP, -: abasic site, and A: dAMP.

3.2.11 apPol Preferentially Adds dATP Opposite an Abasic Site, but the Bypass Is Inefficient.

Abasic sites are particularly detrimental as they present a unique challenge to polymerases due to the complete loss of base information from the templating strand. To determine which nucleotide is incorporated by apPol opposite the abasic site, we performed single nucleotide incorporation assays using DNA substrate S3 (Table 1) while varying the identity of the incoming nucleotide. We found that apPol preferentially adds dATP opposite an abasic site but can also add dGTP at a much slower rate (Figure 3.12). dATP addition can either occur by the so-called A-rule [166], a mechanism shared by other A-family DNA polymerases for abasic site bypass [167, 168] or, owing to the templating sequence in S3, through dNTP-stabilized misalignment [169]. apPol was unable to incorporate any of the pyrimidines in this context (Figure 3.12).

To understand the relevance of an abasic site bypass by apPol, the efficiency of dATP incorporation opposite this lesion was determined. Primer extension assays were performed with substrate S3 and varying concentrations of dATP (Appendix A2). Although the experiments were performed under burst conditions, no burst of product formation was observed. This would indicate that for abasic site bypass apPol performs distributive synthesis such that the post-chemistry steps were not rate-limiting. The rate of incorporation saturated hyperbolically as a function of dATP concentration (Appendix A2; B), consistent with nucleotide binding to the binary complex to come to a rapid equilibrium, followed by the rate-limiting chemistry step.

The primer extension data was fitted using the KinTek Explorer (Figures 3.13). The efficiency of dATP incorporation opposite an abasic site lesion was $8.1 \times 10^{-5} \mu\text{M}^{-1} \text{s}^{-1}$ (Figures 3.13 and 3.14; k_3k_2/k_{-2}), which is more than 3 orders of magnitude worse than the efficiency

of dNTP incorporation opposite an undamaged nucleotide ($0.23 \mu\text{M}^{-1}\text{s}^{-1}$, Table 3). The low efficiency results primarily from a 750-fold slower rate of bond formation opposite the abasic site, with a rate constant of 0.04 s^{-1} (compare k_3 from Figure 3.14 and 3.8). In contrast, the incoming nucleotide's affinity for apPol•DNA binary complex ($K_D^{\text{dATP:abasic}}$) is only 4-fold weaker ($524 \mu\text{M}$; Figure 3.13) than the corresponding K_D^{dNTP} for undamaged DNA ($131 \mu\text{M}$; Table 3). Taken together, these results indicate that, by itself, apPol can only perform inefficient, distributive synthesis while bypassing an abasic site.

In substrate S3 (Table 1), the T_1 base after the abasic site is a thymine, resulting in the possibility of a second A being incorporated after bypassing the abasic site lesion. Indeed, while performing primer extension assays with dATP, two nucleotide incorporations was observed at longer incubation times (Figure 3.12 B): dATP incorporation opposite the abasic site (Figure 3.12 B; +1) followed by dATP incorporation opposite the dTMP at position T_1 (Figure 3.12 B; +2). Having quantitated the two extension products separately, the data could be readily fit to a kinetic model with two bond formation steps (Figure 3.13). We verified if the model could sufficiently explain the data without overfitting using 1D and 2D confidence contour analysis (Figure 3.14). While the accurate determination of the K_D^{dNTP} and rate of incorporation for the second nucleotide could not be made, the efficiency of dATP addition opposite the T_1 base could be estimated (Figures 3.13 A-C and 3.14). We found that the second addition was extremely inefficient ($\sim 1 \times 10^{-6} \mu\text{M}^{-1} \text{ s}^{-1}$), over 70-fold worse than the incorporation efficiency opposite the abasic site itself, hinting that apPol-mediated bypass may not be the major route for negotiating abasic sites in the apicoplast. Alternatively, interaction with other components of the apicoplast replisome might make the abasic site bypass more robust.

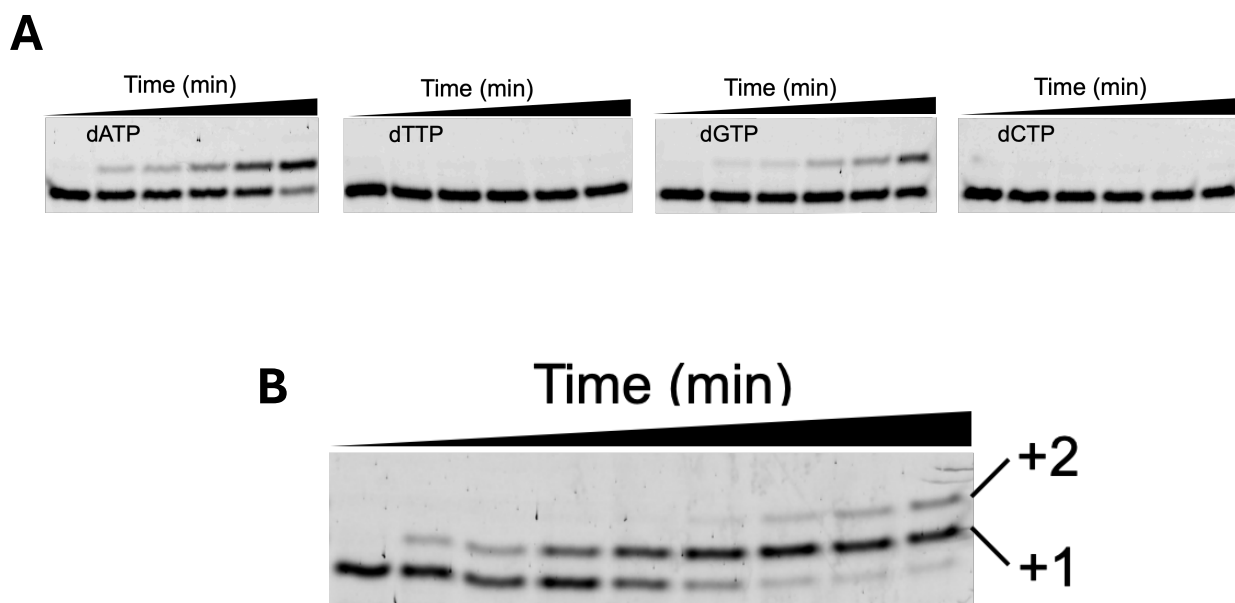
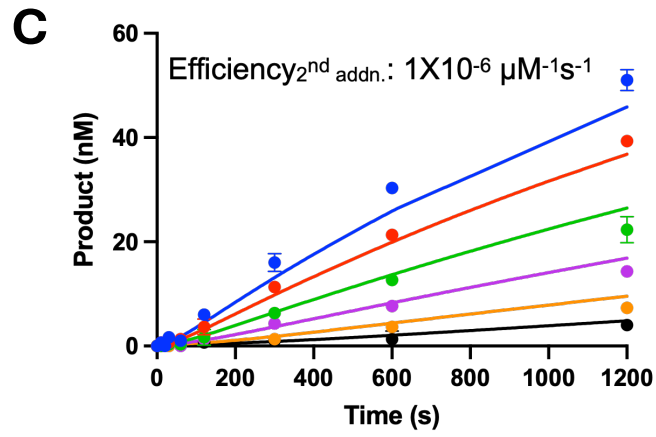
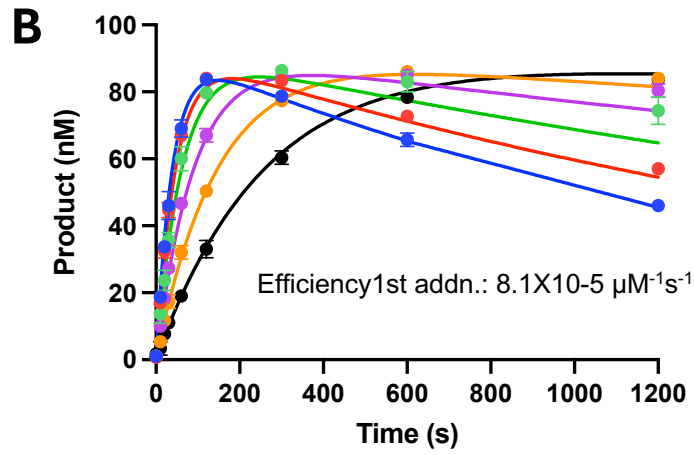
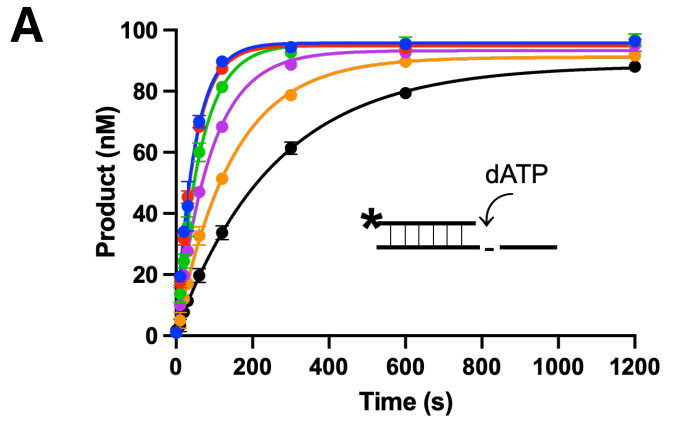


Figure 3.12: Single nucleotide incorporation by apPol against abasic site. (A)

Acrylamide-urea gels depicting apPol-mediated single nucleotide incorporations in DNA substrate S3. Reactions were set up at 37°C with a final concentration of 104 nM active apPol, 100 nM substrate S3, and 250 μ M incoming dNTP (from left: dATP, dTTP, dGTP, and dCTP) and incubated for varying time intervals (0, 0.33, 0.5, 1, 2, and 5 minutes) following which an excess of EDTA was added to quench the reactions. The products were run on a 15% acrylamide-urea gels. **(B)** Acrylamide-urea gel depicting the time course of primer extension with 500 μ M dATP. +1 and +2: primer strand extended by one and two nucleotides respectively.



D

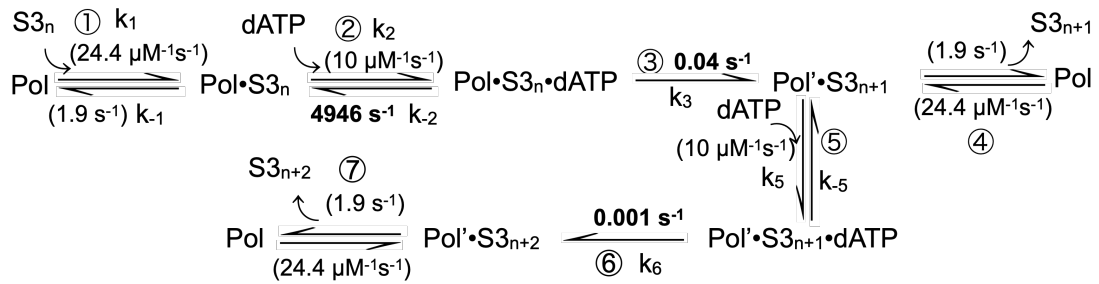


Figure 3.13: Global fitting of the time courses for abasic site bypass by apPol. Primer extension assays to determine the efficiencies of dATP incorporation by apPol opposite an abasic site and extension from the dAMP: abasic site basepair. **(A)** Single nucleotide primer extension assay was performed as described in figure 3.12 with following changes. The incoming nucleotide was dATP for all the assays, and the concentration of the nucleotide (62.5 μ M: black, 125 μ M: orange, 250 μ M: magenta, 500 μ M: green, 1mM: red, and 2 mM: blue), incubated for 0.17 to 20 minutes (0, 0.17, 0.33, 0.5, 1, 2, 5, 10, and 20 minutes). Time courses were fit by numerical integration using the kinetic scheme shown in panel D. The smooth lines overlaying the datapoints are the best global fit of the data. **(B, C)** Global fitting of the +1 (B) and +2 (C) products of (A) using model shown in panel D. Experiments were performed in triplicate. Error bars represent standard deviation. **(D)** Kinetic pathway used for global fitting of dATP incorporation opposite the abasic site and extension past the lesion. The best fit rate constants are shown next to the corresponding steps. Rate constants shown in brackets were held constant during global fitting. k_{-5} and k_6 were not individually constrained by the data; however, their ratio was well-constrained; therefore, only k_6 is shown. $S3_n$: Substrate S3 with 23 nucleotide long unextended primer strand. $S3_{n+1}$ and $S3_{n+2}$: S3 with a primer strand extended by one and two nucleotides respectively. Pol': apPol translocated by one nucleotide.

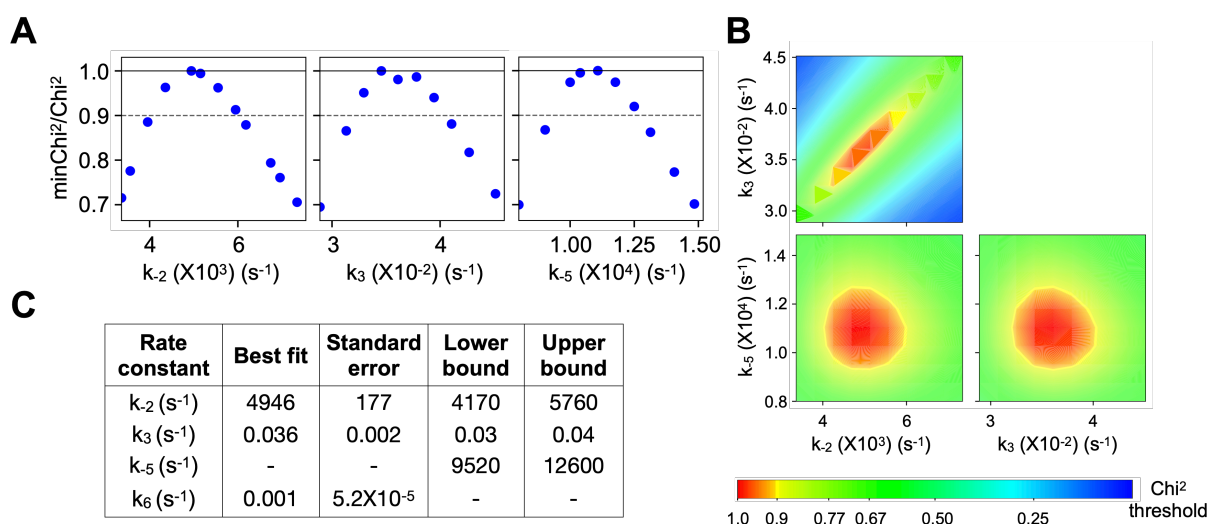


Figure 3.14: Confidence contour analysis for the kinetics of abasic site bypass by apPol.

(A) One-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in Figure 3.13. The forward rate constants for nucleotide binding were fixed at a diffusion limited rate constant of $10 \mu M^{-1} s^{-1}$ and the forward and reverse rate constants governing the binding of apPol to the DNA substrates were fixed at $24.4 \mu M^{-1} s^{-1}$ and $1.9 s^{-1}$ respectively (best-fit values derived from the kinetics of single nucleotide incorporation opposite an undamaged template (Figure 3.8)). The dashed lines represent the minimum χ^2/χ^2 threshold value of 0.9 (recommended by the FitSpace routine in KinTek Explorer) based on which the lower and upper bounds of the rate constants were determined. A ratio of 1 represent the best-fit value. **(B)** Two-dimensional confidence contour analysis of the rate constants governing the kinetic scheme shown in Figure 3.13. The heat maps display the confidence intervals at the indicated χ^2 thresholds for each pair of rate constants. **(C)** Table listing the best fit values, standard errors and the upper and lower bounds (determined from the 1D and 2D confidence contour analysis) of the rate constants governing the kinetic pathway shown in Figure 3.13. The primer extension data shown in Figures 3.13 A, B, and C could not constrain both k_6 and k_5 independently. However, the ratio of the two rate constants was well-constrained. Hence, during global fitting k_6 and k_5 were linked at their individual best-fit values and then the two rate constants were allowed to float as a pair. Thus, the best-fit value and standard error of only k_6 is shown in panel C and only k_5 was included in the confidence contour analysis shown in panels A and B.

3.2.12 apPol Performs Distributive Error-Prone Synthesis opposite 8-Oxo-dGMP

8-oxo-dG is the most abundant of the oxidative lesions [170]. A large number of DNA polymerases, including several A-family polymerases, can incorporate either dCTP or dATP opposite this lesion, with varying degrees of efficiency and fidelity [157, 171]. The versatility arises from the ability of 8-oxo-dGMP to efficiently pair with either dCTP (Watson–Crick pairing) or dATP (Hoogsteen pairing) (Figure 3.15 A) [172]. From single nucleotide primer extension assays with DNA substrate S2 (Table 1), we found that apPol could incorporate both dCTP and dATP opposite 8-oxo-dGMP (Figure 3.15 B). No dTTP addition was detected, while dGTP was added very slowly (Figure 3.15).

To determine how efficiently apPol incorporates dCTP or dATP opposite 8-oxo-dGMP, single nucleotide primer extension assays was performed with varying concentrations of incoming dCTP or dATP (Appendix A3 and A4 respectively). Although both sets of assays were set up under pre-steady state burst conditions, no burst of product formation could be detected (Appendix A3; A and A4; A). This indicates that unlike nucleotide incorporation opposite an undamaged template, chemistry (or a step preceding chemistry) is rate-limiting in apPol's catalytic cycle of nucleotide incorporation opposite 8-oxo-dG.

For both dCTP and dATP, increasing the concentration of the incoming nucleotide increased the rate of primer extension (k_{obs}), and the rate saturated hyperbolically as a function of the dNTP concentration (Appendix A3; B and A4; B). Consequently, we performed global fitting of the primer extension data to a kinetic scheme characterized by the rapid equilibrium of dNTP binding, followed by slow chemistry (Figure 3.16). We verified if the model could sufficiently explain the data without overfitting using 1D and 2D confidence contour analysis

(Figure 3.17). Based on this fit, we found that apPol inserts dCTP and dATP with almost equal efficiencies ($k_3/K_D^{\text{dNTP:8oxodG}}$) of $1.8 \times 10^{-3} \mu\text{M}^{-1}\text{s}^{-1}$ and $1.1 \times 10^{-3} \mu\text{M}^{-1}\text{s}^{-1}$ respectively (Figures 3.16 A, B and 3.17 C, F and Table 4). This indicates that apPol does not strongly discriminate between correct and incorrect incorporations opposite 8-oxo-dG. However, these efficiencies were 2 orders of magnitude less than the efficiency of correct incorporation opposite an undamaged template, suggesting that apPol can sense the damaged templating base. The lower efficiency was the combined effect of a nearly 30-fold reduced rate of incorporation (1.6 s^{-1} and 0.78 s^{-1} for dCTP and dATP, respectively, compared to 30 s^{-1} for dTTP incorporation opposite dAMP) and a 7-fold weaker dNTP binding ($872 \mu\text{M}$ and $713 \mu\text{M}$ for dCTP and dATP incorporation opposite 8-oxo-dGMP respectively, compared to $131 \mu\text{M}$ for dTTP incorporation opposite dAMP) (Tables 3 and 4).

Table 4: Kinetic parameters of 8-oxo-dG bypass by A-family DNA polymerase.

$\text{Efficiency}^{\text{dCTP:8oxodG}} = (k_{\text{pol}}^{\text{dCTP:8oxodG}}/K_D^{\text{dCTP:8oxodG}})$, $\text{Efficiency}^{\text{dATP:8oxodG}} = (k_{\text{pol}}^{\text{dATP:8oxodG}}/K_D^{\text{dATP:8oxodG}})$, Fidelity = $(\text{Efficiency}^{\text{dCTP:8oxodG}}/\text{Efficiency}^{\text{dATP:8oxodG}})$, $\text{Efficiency}^{\text{Ud}}$: Efficiency of nucleotide incorporation opposite undamaged dNMP.

DNA polymerase	$K_D^{\text{dCTP:8oxodG}}$ (μM)	$k_{\text{pol}}^{\text{dCTP:8oxodG}}$ (s^{-1})	Efficiency ^{dCTP:8oxodG} ($\mu\text{M}^{-1}\text{s}^{-1}$)	$K_D^{\text{dATP:8oxodG}}$ (μM)	$k_{\text{pol}}^{\text{dATP:8oxodG}}$ (s^{-1})	Efficiency ^{dATP:8oxodG} ($\mu\text{M}^{-1}\text{s}^{-1}$)	Fidelity	Efficiency ^{Ud} ($\mu\text{M}^{-1}\text{s}^{-1}$)	Ref.
<i>P. falciparum</i> apPol	872	1.6	0.0018	713	0.78	0.0011	1.6	0.23	This study
T7 Pol-thioredoxin	11	1.6	0.14	110	0.2	0.0018	78	27.6	[173]
Human Pol gamma holoenzyme	100	10	0.1	300	3.4	0.011	9	40.8	[174]
<i>E. coli</i> Pol I	19	5.5	0.3	190	0.3	0.0016	187.5	10.3	[175]

TLS consists of two parts: incorporation of a nucleotide opposite the lesion, followed by extension past the lesion. To investigate the kinetics of extension past 8-oxo-dGMP, we used substrates S5 and S6 (Table 1). S5 has dCMP already added to the primer strand opposite 8-oxo-dGMP, while S6 has dAMP in place of the dCMP. The kinetics of dATP incorporation was monitored by primer extension under the pre-steady state burst condition. In both cases, the time courses were biphasic, with k_{fast} of $\geq 10 \text{ s}^{-1}$ (Figure 3.18), over 5 times faster than the rate of DNA dissociation from the post-chemistry binary complex (Figure 3.8), suggesting that apPol resumes processive synthesis immediately after incorporation opposite 8-oxo-dGMP. To determine the efficiency of extension past the lesion, we performed pre-steady state assays with substrates S5 and S6 and varied the dATP concentration. However, we found that the affinities of dATP for apPol•S5 and apPol•S6 binary complexes are very weak (data not shown) and prevent accurate determination of efficiencies.

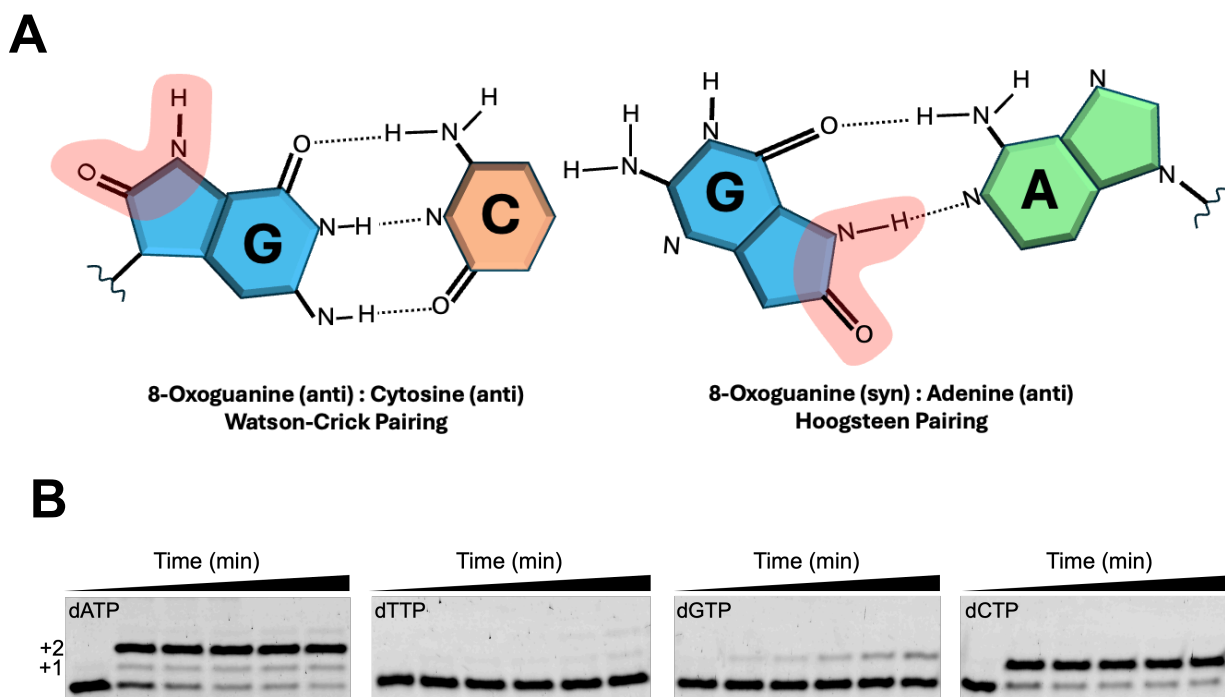


Figure 3.15: Single nucleotide primer extension against 8-oxo-dGMP. (A) Base-pairing modes of 8-oxoguanine. 8-Oxoguanine can adopt either the syn or anti conformation to form alternative base pairs. In the anti-conformation it pairs with cytosine through canonical Watson–Crick hydrogen bonding (Left). Whereas, In the syn conformation, 8-oxoguanine pairs with adenine through Hoogsteen hydrogen bonding (Right). **(B)** Acrylamide-urea gels depicting apPol-mediated single nucleotide incorporations to DNA substrate S2 in the presence of (from left) dATP, dTTP, dGTP, or dCTP. +1 and +2: primer strand extended by one and two nucleotides respectively. Single nucleotide primer extension assays were set up at 37°C with a final concentration of 104 nM active apPol, 100 nM DNA substrate S2, and 250 μ M of incoming dNTP. The reactions were incubated for varying time intervals ranging from 0 to 5 minutes (0, 0.33, 0.5, 1, 2, and 5 minutes) following which an excess of EDTA was added to quench the reactions, and the samples were analysed on 15% acrylamide-urea gels.

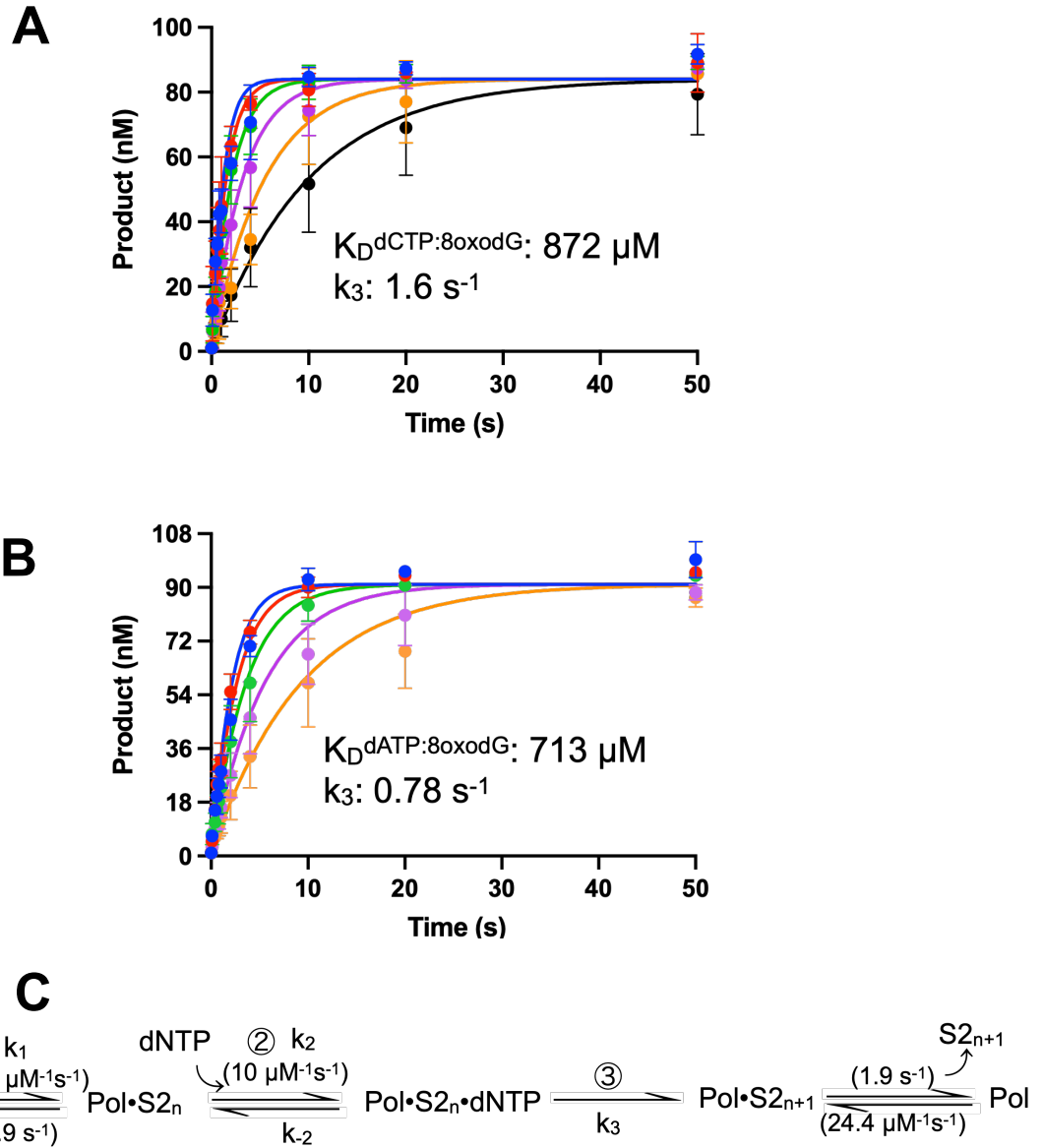
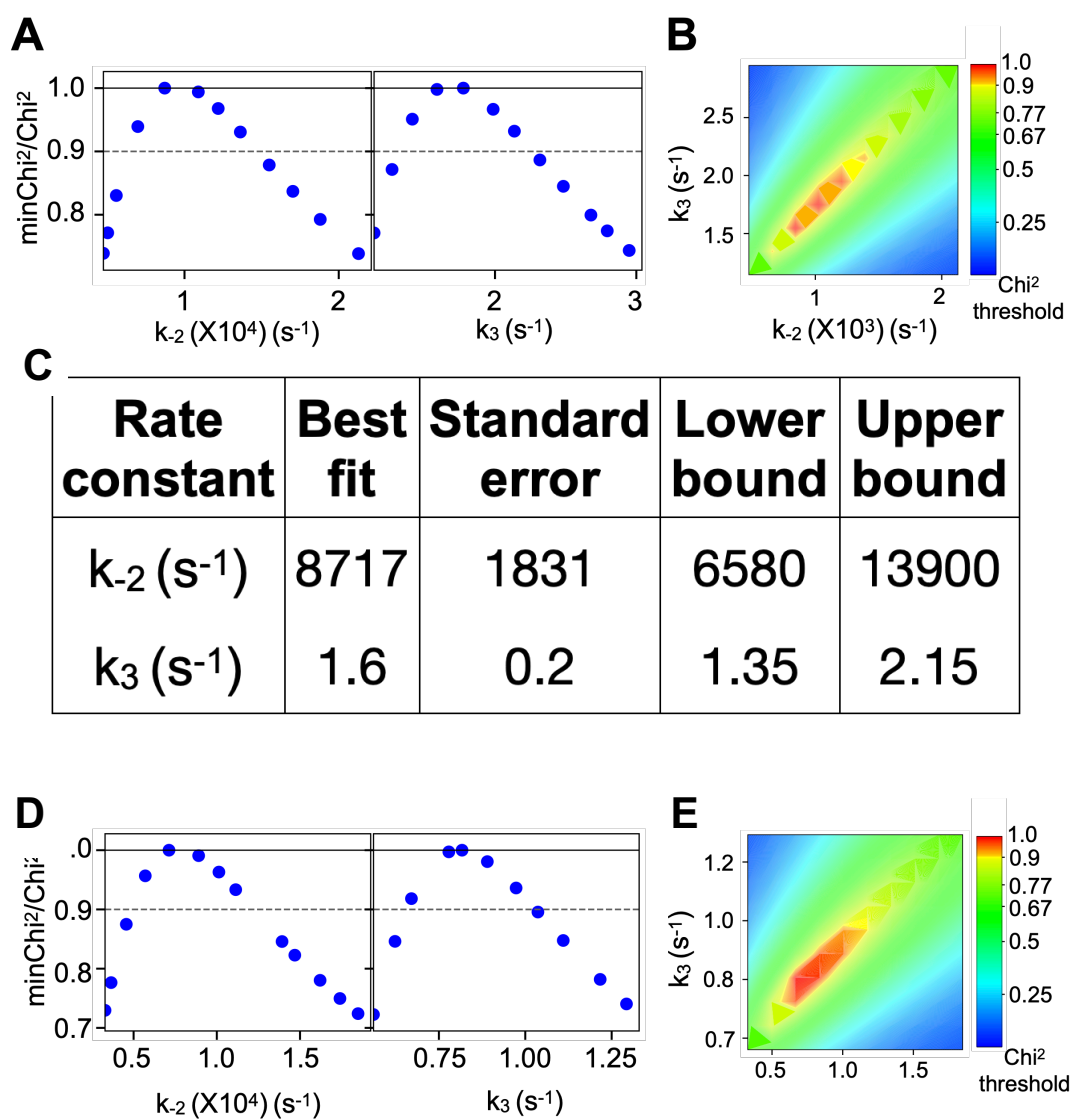


Figure 3.16: Global fitting of the kinetics of 8-oxo-dG bypass. (A) The time courses shown in A3; A overlaid with the best fit of the data from global fitting (smooth lines). **(B)** The time courses in show A4; A overlaid with the best fit of the data from global fitting (smooth lines), the combined product concentrations (i.e., the total of the +1 and +2 products). The rate constant governing dCTP or dATP addition (k_3) and the affinity of the nucleotide for the binary complex ($K_D^{\text{dNTP:8oxodG}}$, where dNTP is either dCTP or dATP) are shown as insets with the corresponding graphs. All experiments were performed in triplicate, and the average of the three independent data sets is plotted in the graph, while the error bars represent the standard deviation. **(C)** Kinetics pathway for dCTP or dATP incorporation against 8-oxo-dG. Kinetic pathway used for global fitting of the primer extension assays shown in panel A and B. Rate constants shown in brackets were kept constant during global

fitting. $S2_n$: Substrate S2 with 23 nucleotide long unextended primer strand. $S2_{n+1}$: Substrate S2 with the primer strand extended by one nucleotide.



F

Rate constant	Best fit	Standard error	Lower bound	Upper bound
$k_{-2} \text{ (s}^{-1}\text{)}$	7127	1251	5340	11700
$k_3 \text{ (s}^{-1}\text{)}$	0.78	0.06	0.67	0.97

Figure 3.17: Confidence contour analysis for the kinetics of 8-oxo-dGMP bypass by apPol. **(A)** One-dimensional confidence contour analysis of the rate constants governing the kinetics of dCTP incorporation opposite 8-oxo-dGMP (Figure 3.16 A). The kinetic scheme is shown in Figure 3.16. The forward rate constants for nucleotide binding were fixed at a diffusion limited rate constant of $10 \mu\text{M}^{-1}\text{s}^{-1}$ and the forward and reverse rate constants governing the binding of apPol to the DNA substrates were fixed at $24.4 \mu\text{M}^{-1}\text{s}^{-1}$ and 1.9 s^{-1} respectively (best-fit values derived from the kinetics of single nucleotide incorporation opposite an undamaged template (Figure 3.8)). The dashed lines represent the minimum χ^2/χ^2 threshold value of 0.9 (recommended by the FitSpace routine in KinTek Explorer) based on which the lower and upper bounds of the rate constants were determined. A ratio of 1 represent the best-fit value. **(B)** Two-dimensional confidence contour analysis of the rate constants governing dCTP incorporation opposite 8-oxo-dGMP. The heat maps display the confidence intervals at the indicated χ^2 thresholds for the pair of rate constants. **(C)** Table listing the best fit values, standard errors and the upper and lower bounds (determined from the 1D and 2D confidence contour analysis) of the rate constants governing the kinetic pathway of dCTP incorporation opposite 8-oxo-dGMP. **(D)** One-dimensional confidence contour analysis of the rate constants governing the kinetics of dATP incorporation opposite 8-oxo-dGMP (Figure 3.16 B). The kinetic scheme is shown in Figure 3.16. The forward rate constants for nucleotide binding were fixed at a diffusion limited rate constant of $10 \mu\text{M}^{-1}\text{s}^{-1}$ and the forward and reverse rate constants governing the binding of apPol to the DNA substrates were fixed at $24.4 \mu\text{M}^{-1}\text{s}^{-1}$ and 1.9 s^{-1} respectively (best-fit values derived from the kinetics of single nucleotide incorporation opposite an undamaged template (Figure 3.8)). The dashed lines represent the minimum χ^2/χ^2 threshold value of 0.9 (recommended by the FitSpace routine in KinTek Explorer) based on which the lower and upper bounds of the rate constants were determined. A ratio of 1 represent the best-fit value. **(E)** Two-dimensional

confidence contour analysis of the rate constants governing dATP incorporation opposite 8-oxo-dGMP. The heat maps display the confidence intervals at the indicated chi2 thresholds for the pair of rate constants. **(F)** Table listing the best fit values, standard errors and the upper and lower bounds (determined from the 1D and 2D confidence contour analysis) of the rate constants governing the kinetic pathway of dATP incorporation opposite 8-oxo-dGMP.

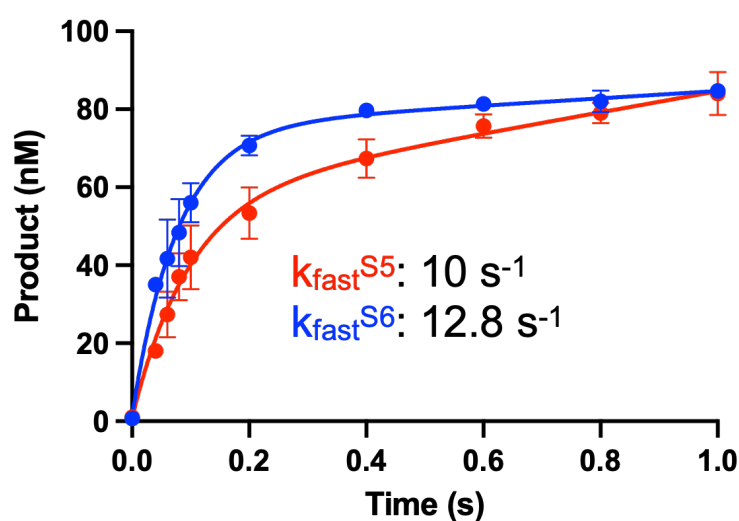


Figure 3.18: Primer extension assay performed using DNA substrates S5 and S6. Time courses from primer extension assays performed with DNA substrates S5 (red) and S6 (blue). Single nucleotide primer extension assays were performed as described in Figure 3.15 with the following changes. The incoming nucleotide was dATP (final concentration 2 mM), and the maximum incubation time was 1 s. The time courses were fit to the full burst equation, and the rates of the fast phases are mentioned in the inset. All experiments were performed in triplicate, and the average of the three independent data sets is plotted in the graph, while the error bars represent standard deviation.

3.2.13 apPol's Proofreading Activity Does Not Discriminate between Accurate and Error-Prone Bypass of 8-oxo-dGMP

In addition to the polymerisation domain, apPol contains a proofreading domain like other replicative DNA polymerases (Figure 2.1). The 3' to 5' exonuclease activity of apPol might act as a checkpoint to discriminate between correct and incorrect incorporation opposite 8-oxo-dGMP by efficiently excising out the dAMP mispaired with 8-oxo-dG. We investigated if this was the case by performing apPol-catalysed exonuclease assays with DNA substrates S7–S10 (Table 1). For these experiments, wild-type apPol (apPol^{WT}) lacking the D82N and E84Q point mutations was used (Figure 2.1 B), and all DNA substrates had a phosphorothioate linkage between the P₋₂ and P₋₃ positions of the primer (Table 1, S7-S10, orange pair), which limited any exonucleolytic degradation to a single nucleotide.

The exonuclease assays were performed under the pre-steady state burst condition, and we detected a biphasic time course of primer degradation indicating the presence of a slow step following the exonucleolytic cleavage. On comparing the rates of the fast phase of the primer degradation time courses, we found that the primer strands of the DNA substrates containing a dAMP:8-oxo-dGMP or dCMP:8-oxo-dGMP pair at the 3' end of the primer (Table 1, substrates S7 and S8, respectively) were cleaved at comparable rates of 0.9 ± 0.1 and 1.5 ± 0.2 s⁻¹ respectively (Figure 3.19, magenta and green time courses), indicating that the proofreading activity of apPol^{WT} does not discriminate between correct and incorrect pairing opposite 8-oxo-dGMP. Since the rate of extension past the 8-oxo-dGMP lesion is > 10-fold faster than the rate of excision (comparing k_{obs} of Table 4 with k_{exo} of Figure 3.19), there is a high probability of a dAMP:8-oxo-dGMP mispair becoming integrated into the apDNA.

The rates of cleavage of undamaged DNA substrates S9 and S10 by apPol^{WT} were also measured. In substrate S9, the primer and template strands are complementary to each other, while in S10, there is a mispair at the 3' end of the primer. The primer degradation rate for S9 was almost 10-fold lower than the corresponding rate for S10 ($0.12 \pm 0.05 \text{ s}^{-1}$ for S9 and $1.14 \pm 0.2 \text{ s}^{-1}$ for S10; Figure 3.19), indicating that the proofreading activity of apPol^{WT} can discriminate between correctly paired and mispaired DNA substrates. Interestingly, the exonuclease rates for substrates S10 (undamaged DNA with mispair) and S8 (dCMP:8-oxo-dGMP at the primer/template junction) are comparable and >10-fold higher than the exonuclease rate for S9 (perfectly paired undamaged DNA), suggesting that the dCMP:8-oxo-dGMP pair does not adopt a conventional Watson–Crick base pair geometry and is recognized as a mispair by apPol^{WT}.

In conclusion, the kinetic characterization presented in this chapter provides the first comprehensive kinetic analysis of *P. falciparum* apicoplast DNA polymerase (apPol) and investigates the lesion bypass activity of this enzyme. This work also provides a foundation for future structure-function correlation studies. The implication of these results in the context of apicoplast genome duplication are summarized in chapter 5.

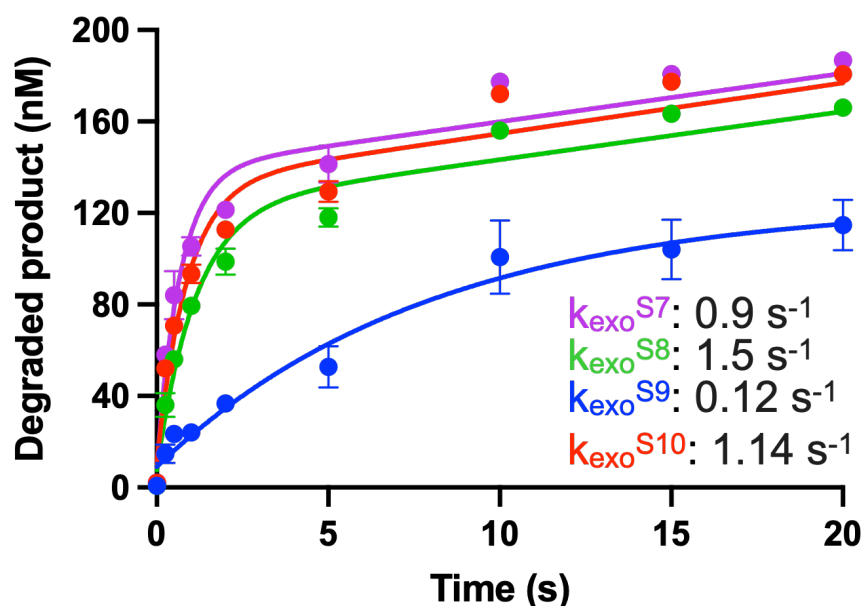


Figure 3.19: Exonuclease activity of apPol^{WT}. Time courses of primer degradation by apPol^{WT}. The concentration of the primer strand shortened by a single nucleotide (degraded products) is plotted as a function of time. A final concentration of 104 nM active apPol^{WT} was incubated with 200 nM of DNA (S7 magenta, S8 green, S9 blue, or S10 red), and the reaction was initiated by adding a final concentration of 10 mM Mg²⁺. Reactions were incubated at 37°C for various time intervals (0, 0.25, 0.5, 1, 2, 5, 10, 15, and 20 s) and then quenched with excess EDTA. Samples were analysed on a 15% acrylamide-urea denaturing gel, and the concentration of the degraded product was plotted as a function of time. The data were fit to the burst equation, and the rate of the fast phase was approximated as the rate of the exonucleolytic cleavage (k_{exo}). k_{exo} for S7, S8, S9, and S10 were 0.9 ± 0.1 , 1.5 ± 0.2 , 0.12 ± 0.05 , and 1.14 ± 0.2 s⁻¹ respectively. All experiments were performed in triplicate, the average of the three independent datasets was plotted, and the error bars represent the standard deviation of data sets.

4. Structural Basis of Multitasking DNA Polymerase from *Plasmodium Falciparum*

4.1 INTRODUCTION

Pre-steady state kinetic studies of apPol on both undamaged and damaged DNA substrates demonstrated that the enzyme is capable of catalysing both replicative and lesion bypass synthesis, although with different efficiencies. These observations indicate that apPol, the sole DNA polymerase in the apicoplast, exhibits multitasking behaviour; however, the structural basis underlying this versatility remains unclear. We used single particle cryo-EM (SPA) to provide the first structural insight into the molecular mechanism of multitasking by apPol. Two pre-chemistry ternary complexes of apPol (Figure 2.1 A) were assembled bound to a primer/template (P/T) DNA and the corresponding incoming nucleotide (dGTP) (Figure 2.6). To prevent the chemical step of bond formation (Figure 1.4, step 3) either a 3' dideoxy-terminated primer strand with Mg^{2+} as the divalent cation (dataset 1) or a non-dideoxy-terminated primer strand with Ca^{2+} as the cation was used (dataset 2). In both datasets, 8 mM CHAPSO was used to achieve uniform particle distribution over a holey carbon grid (Figure 2.6). We performed multiple nucleotide incorporation assays to ascertain that apPol retained its activity in the presence of CHAPSO (Figure 4.3). Structures of these complexes were solved using SPA.

For dataset 1 the consensus structure was determined to an overall resolution of 3.2 Å (Figure 4.2). Most of the apPol density was well-resolved, and we could trace the entire polypeptide backbone (Figure 4.2) and identify densities for several amino acid sidechains. Overall, the apPol structure was similar to the previously reported structure of the apo enzyme [176]. The

DNA density was fragmented, and there was no clear density for the incoming nucleotide (Figures 4.2 and 4.5). The particles contributing to the consensus map were further classified based on the alignment parameters from the consensus refinement and resolved into four structures (global resolution ranging from 3.8 Å to 4.1 Å) corresponding to apPol-DNA, apPol-dGTP, and two conformational states of apPol-DNA-dGTP complexes (Figures 4.4 and 4.5). From dataset 2, an apPol-DNA-dGTP-Ca²⁺ ternary complex structure was solved to a global resolution of 3.5 Å (Figures 4.4 and 4.5). Here both the DNA and nucleotide substrates were clearly visible. By combining all these structures, the events leading up to the chemical step of bond formation by apPol is described below.

4.2 RESULTS

4.2.1 apPol complex behaves differently under negative stain and vitrified ice.

Negative-stain EM analysis was used to assess the overall quality of the assembled apPol-DNA-dGTP ternary complex, and the resulting micrographs show homogenous particle distribution, a representative micrograph is shown in figure 4.1 A. The complex assembly was similar for both negative stain EM and cryo-EM (as described in chapter 2, section 2.13) except the concentration of the complex was higher for cryo-EM (~200 nM for negative stain and ~4µM for cryo-EM). However, the cryo-EM micrographs do not show homogenous particle distribution. This might indicate one or more of the following issues: 1) denaturation or disassembly or aggregation of the complex potentially due to adsorption at air-water interface or 2) adsorption on carbon foil of the Quantifoil R1.2/1.3 400 mesh holey carbon grids (SPT Labtech). A representative cryo-EM micrograph is shown in figure 4.1 B.

Multiple approaches were tried to improve particle homogeneity, reduce aggregation and improve complex integrity under vitreous ice (described in chapter 2, section 2.13). We found that addition of the detergent CHAPSO at its critical micelle concentration (CMC) of 8mM dramatically improved particle distribution over open holes. [177]. Accordingly, the grids for large scale data collection we prepared in the presence of 8mM CHAPSO (Figure 4.1 C).

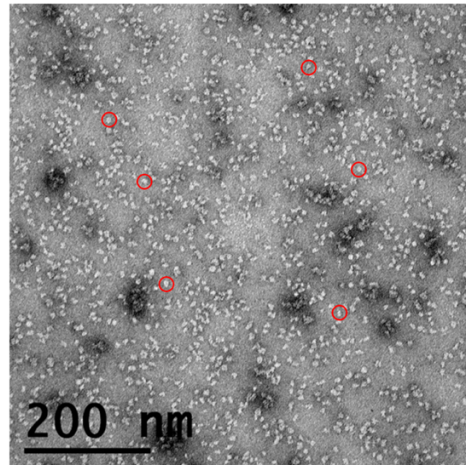
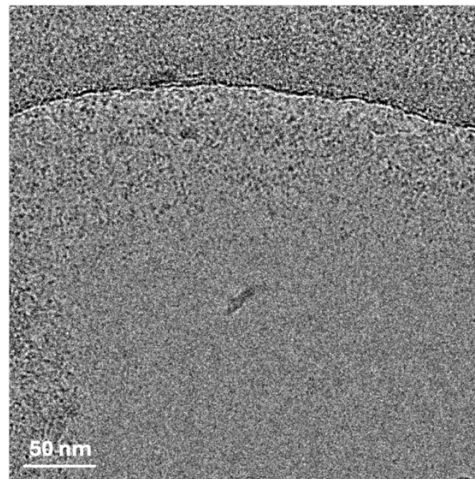
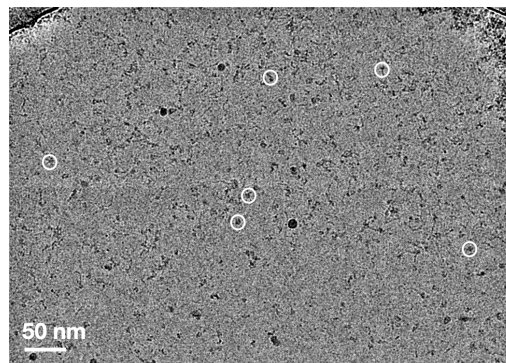
A**B****C**

Figure 4.1. apPol complex in Negative stain and vitreous ice. (A) Representative micrographs from negative stain imaging with particles highlighted with red circles. **(B)** Representative cryo-EM micrograph without CHAPSO. **(C)** Representative cryo-EM micrograph with CHAPSO with evenly distributed particles highlighted with white circles.

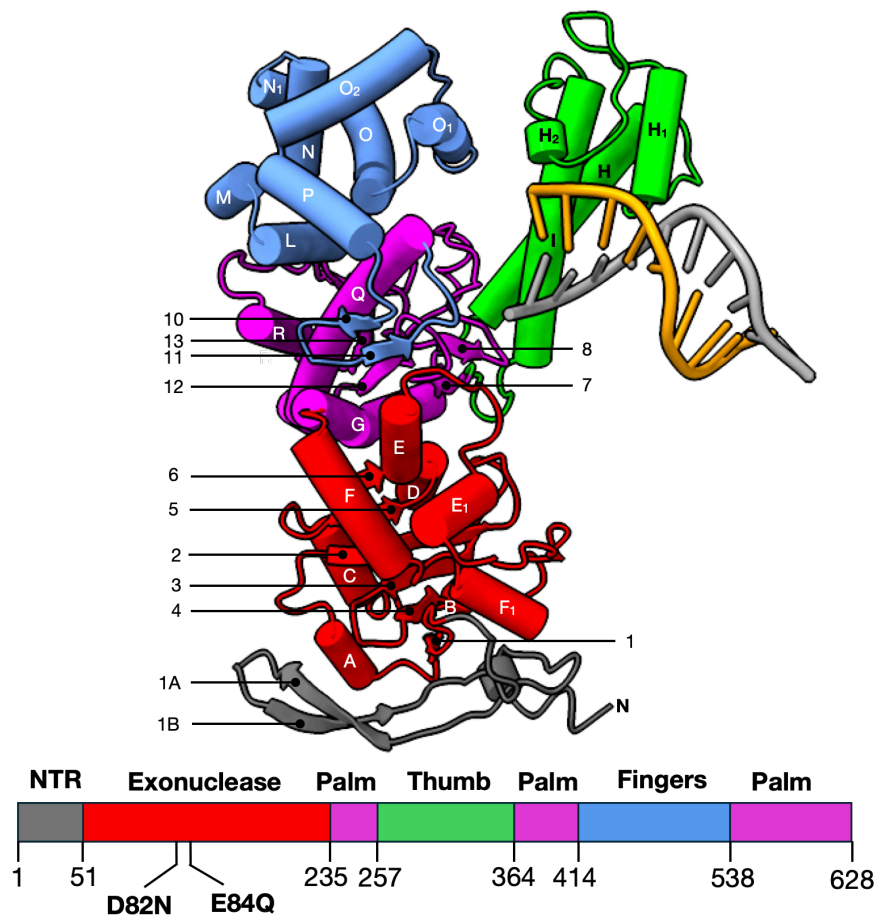


Figure 4.2: Atomic model of apPol complex. Atomic model (tube and arrow representation) of the 3.2 Å consensus map from dataset 1 along with the domain organization of apPol. The domains are coloured as follows (N-terminal region (NTR): grey; 3'–5' proofreading exonuclease: red; palm: magenta; thumb: green and fingers: blue). The exonuclease active site residues (D82 and E84) that have been mutated to N and Q, respectively, in all the apPol constructs used in this study are highlighted. The secondary structural elements are labelled based on structure alignment with Klenow fragment (PDB id: 1KLN) [178].

4.2.2 apPol binds DNA and nucleotide independent of each other

The 3D classification of the particles contributing to the consensus structure of dataset 1 resulted in two binary complexes; one had apPol bound to DNA (binary1), and the other had apPol bound to the incoming dGTP (binary2) (Figures 4.4 and 4.5). In binary1, density for 20 out of the 25 base pairs of the DNA duplex was detected (Figures 4.5, 4.6 A and 4.7). Of the 20 base pairs, only 6 were contacted by apPol (Figure 4.6 A). In the sharpened map, no clear density was visible for the template overhang, including the T₀ base. However, in the unsharpened map, the density for the T₀ base could be detected (Figure 4.6 A) and that guided the register of the DNA duplex. Similar to other A family polymerases, the DNA is mainly contacted by the thumb domain of apPol. Helices H, H₁ and H₂ form a positively charged groove that accommodates the minor groove phosphate backbone of the DNA duplex (Figure 4.6 A). However, consistent with a recent report [179], unlike other A family polymerases, the interactions between the DNA and the thumb is limited in apPol with a total buried surface area of only 607 Å². Two specific contacts were detected: Q387 forms a hydrogen bond with the 3' base of the primer strand (P₋₁) potentially determining the position of the DNA in binary1 and the backbone nitrogen of T319 interacts with the phosphate backbone of P₋₃ on the primer strand (Figure 4.6 B). The template strand does not make any significant contact with apPol with a buried surface area of only 174 Å². At the resolution range of the structures reported here (3.2 Å to 4.1 Å) it is not possible to reliably detect density for ordered water molecules, and thus any potential water-mediated interactions have not been analysed throughout this work.

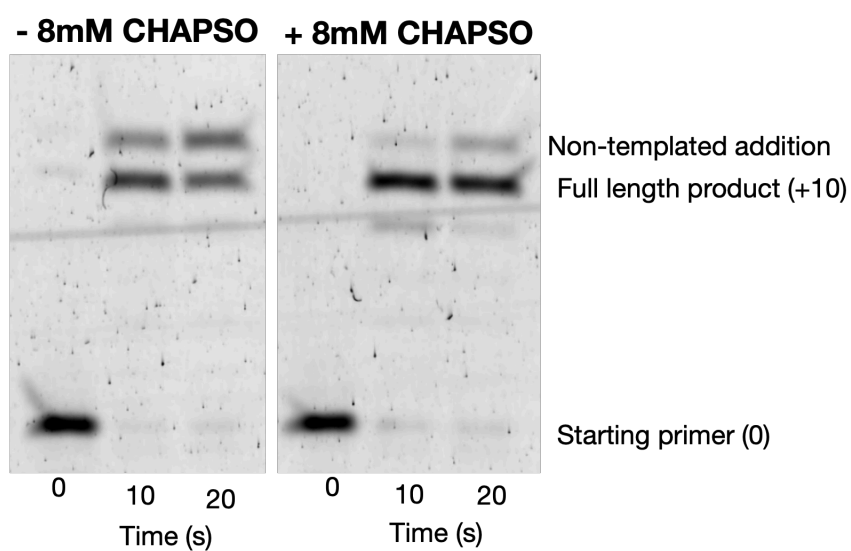


Figure 4.3: apPol's activity in presence of CHAPSO. Multiple nucleotide incorporation by apPol in the absence (left) and presence (right) of 8mM CHAPSO.

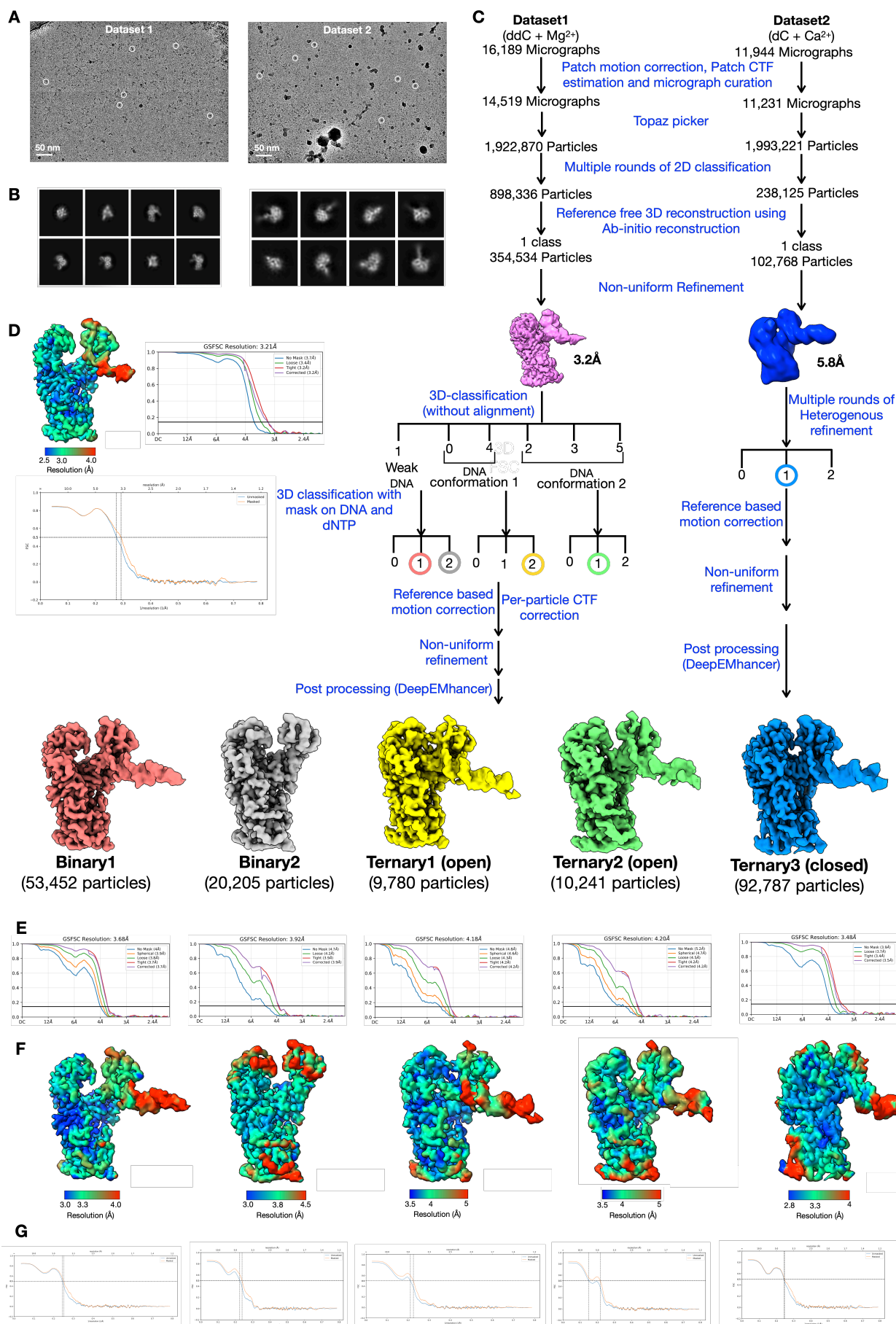


Figure 4.4: Cryo-EM image processing and 3D reconstruction from datasets1 and 2. (A) Representative micrographs from dataset 1 (left) and dataset 2 (right). Representative particles are highlighted with white circles. (B) Representative 2D class averages for dataset 1 (left) and dataset 2 (right). (C) Schematic representation of the image processing pipeline for datasets1 and 2. The post processed maps are coloured as follows: binary1 (salmon pink), binary2 (grey), ternary1 (yellow), ternary2 (green) and ternary3 (blue). Coloured circles: classes that were taken forward along the pipeline. Colour coding of the circles is the same as that of the maps. (D) The 3.2Å consensus map of dataset 1 (pink in panel C) coloured according to local resolution (top left). Gold standard FSC curve of the consensus map (top right) and FSC curve between the consensus map and the corresponding atomic model (bottom). (E, F, G) Gold standard FSC curves (E), maps coloured according to local resolution (F) and FSC curve between the cryo-EM maps and the corresponding atomic models (G) of binary1, binary2, ternary1, ternary2 and ternary3 (left to right).

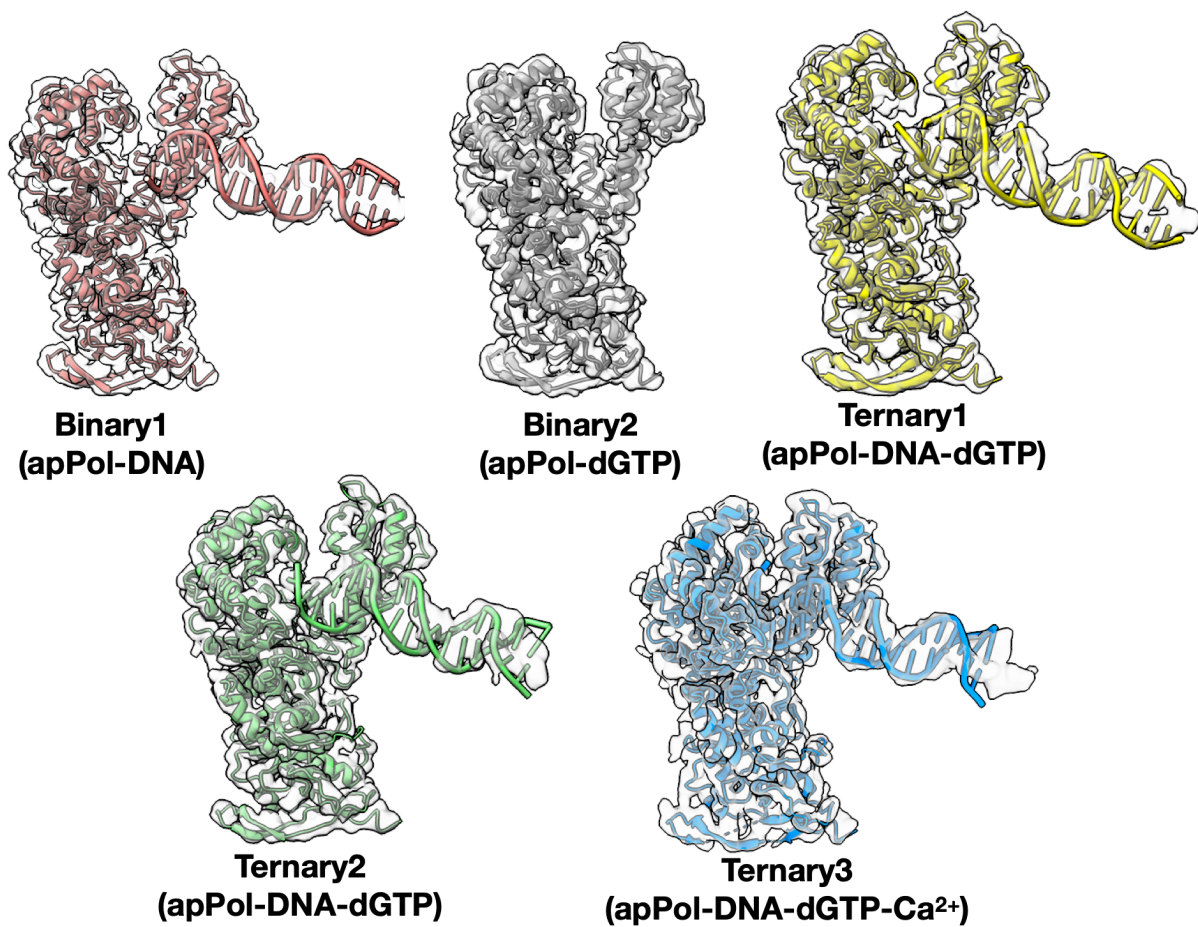


Figure 4.5: Different conformational states of apPol's complex. Atomic models of the two binary complexes (apPol–DNA (binary1): salmon pink; apPol–dGTP (binary2): grey) and three ternary complexes of apPol–DNA–dGTP (ternary1, ternary2, and ternary3: yellow, green, and blue, respectively) fitted to their respective cryo-EM maps (transparent grey).

The binary2 structure is similar to that of binary1 except that there is no clear DNA density (Figures 4.5 and 4.4). Instead, the density for the incoming dGTP near the O helix of the fingers domain can be detected (Figure 4.7 B). The base moiety is stabilized by H-bond to Y486 of the O₁ helix while the triphosphate forms hydrogen bonds with R390. The configuration of dGTP in binary2 is not conducive for catalysis.

For most DNA polymerases, a transition from the apo to the catalytically poised pre-chemistry ternary complex is accompanied by the closure of the fingers domain [180] [74, 181] [182] [183] constraining the active site by forming a highly restrictive space for the nascent base pair. For A-family polymerases, fingers closure is achieved by a large-scale rotation of the O and N helices towards the nascent base pair [181] [183] [184]. In both binary1 and 2, the O helix is in a predominantly open position with the N-terminal tip of the O helix rotating towards the active site by only $\sim 11^\circ$ (Figure 4.8 A). We denote this state of the O-helix as quasi-open. Taken together, binary1 and 2 complexes indicate that apPol can engage with the DNA and nucleotide substrates independent of each other.

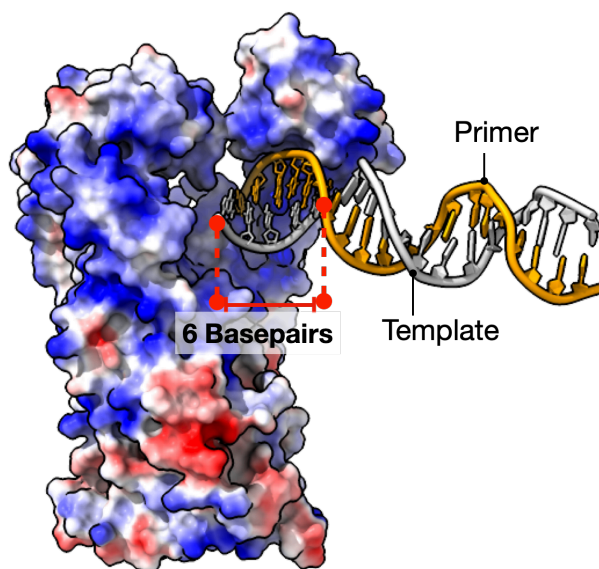
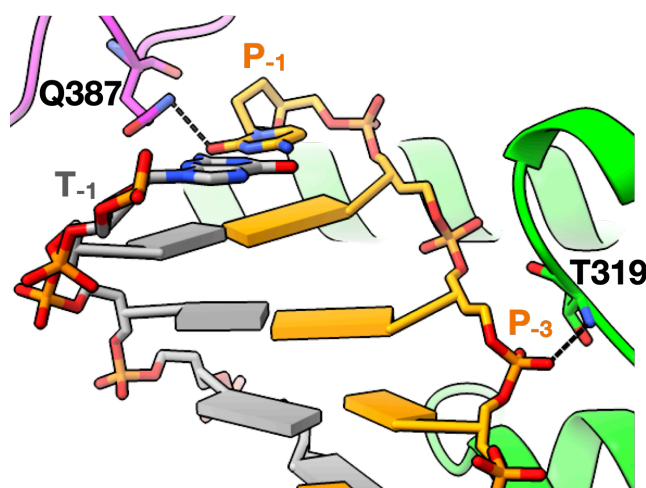
A**B**

Figure 4.6: apPol-DNA Binary complex (Binary1). (A) Six base pairs (stick representation; red dashed lines) of DNA interact with a positively charged groove (blue) of apPol (surface representation; coloured according to coulombic potential; red: negative charge to blue: positive charge). (B) Specific interactions of apPol with the DNA duplex. Colour coding is the same as in Figure 4.2. Potential hydrogen bonds are shown as dashed lines.

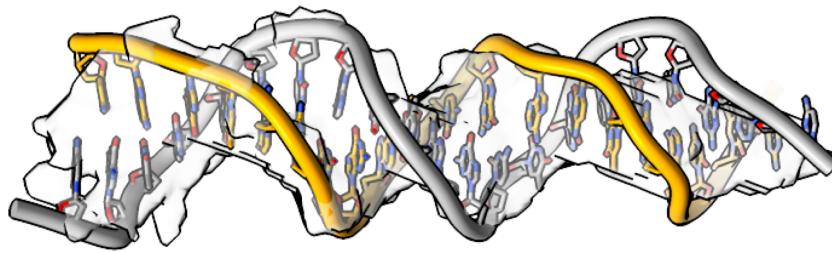
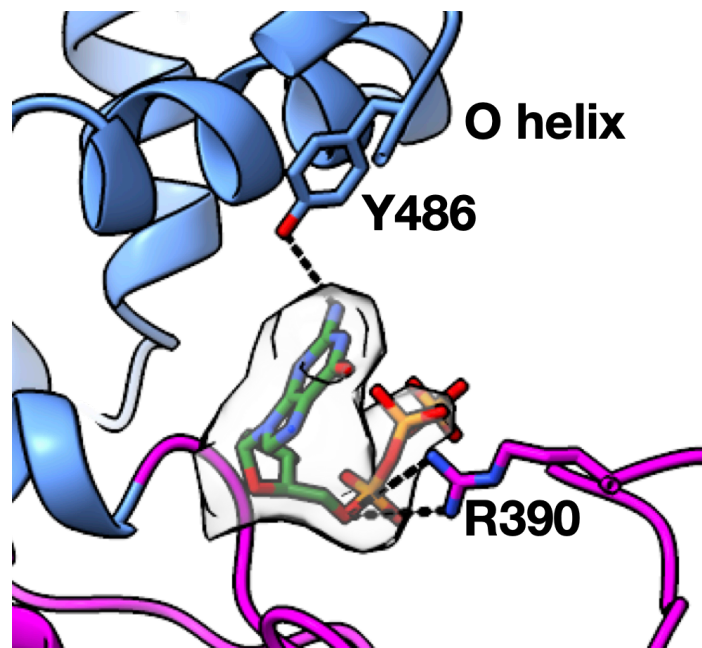
A**B**

Figure 4.7: Binary complexes of apPol fitted model in cryo-EM maps. (A) Primer/template DNA of binary1 fitted into the cryo-EM density for the DNA (transparent grey) segmented out from the unsharpened map. **(B)** Interaction of dGTP with apPol in binary2. The cryo-EM density for the nucleotide is shown in transparent grey.

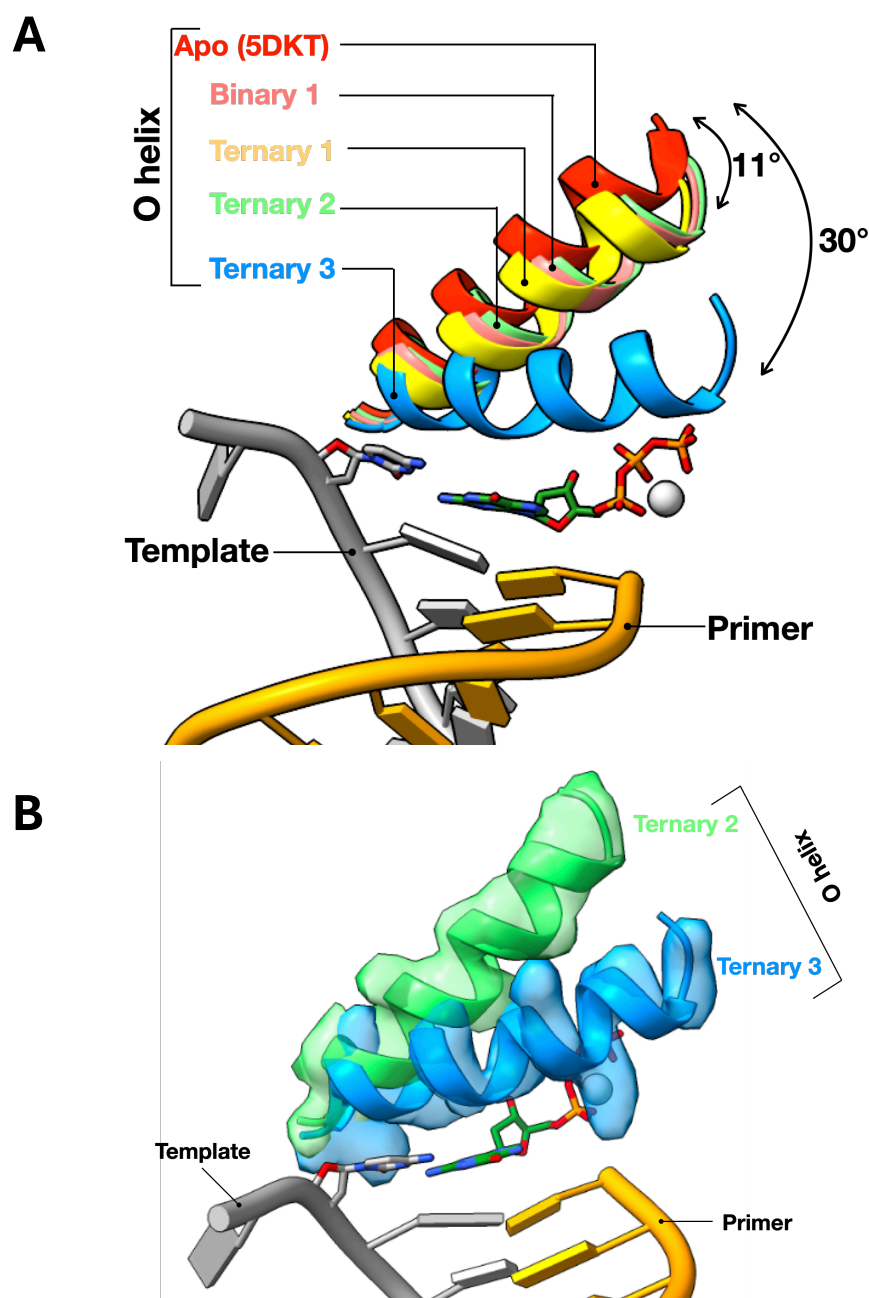


Figure 4.8: O-helix movement in different conformational states. (A) Position of the O-helix of apPol pre-chemistry complexes with respect to the O helix position of apo apPol (PDB id: 5DKT; red). The primer (orange)/template (grey) DNA, incoming dGTP (green), and Ca^{2+} (white) occupying the position of metal B in ternary3 are shown for reference. The T_0 base is depicted in stick representation, while all other bases of the DNA are shown as slabs. For angle measurement the Ca atom of Y471 of apo apPol was selected as the vertex. (B) Segmented cryo-EM map density was shown in transparent green for ternary2 and transparent blue for ternary3 overlaid on their respective atomic model. The colour scheme is as panel A.

4.2.3 apPol-DNA binary complex base pairs with the incoming nucleotide with the fingers open

The structure of apPol-DNA-dGTP ternary complex was solved in three conformational states (ternary1, 2, and 3) (Figure 4.4 and 4.5). Ternary1 and 2 were resolved through 3D classification of dataset 1, while ternary3 is from dataset 2 (Figure 4.4 and 4.5). The overall protein structure is similar among these states (average C α RMSD between ternary1 and 2 is 1.5 Å and between ternary1 and 3 and ternary2 and 3 are 2.8 Å and 2.9 Å respectively). In ternary1 and 2 the O helix is quasi-open, while in ternary3 the fingers is in a closed conformation with the O helix rotating towards the active site by 30° (Figure 4.8; Movie 1). By comparing the DNA positions and active site configurations of the ternary complexes with those of binary1 (Figures 4.9 and 4.10), the complexes were arranged on a reaction pathway starting from the initial interaction of dGTP with the apPol-DNA binary complex and ending with the catalytically poised pre-chemistry ternary complex.

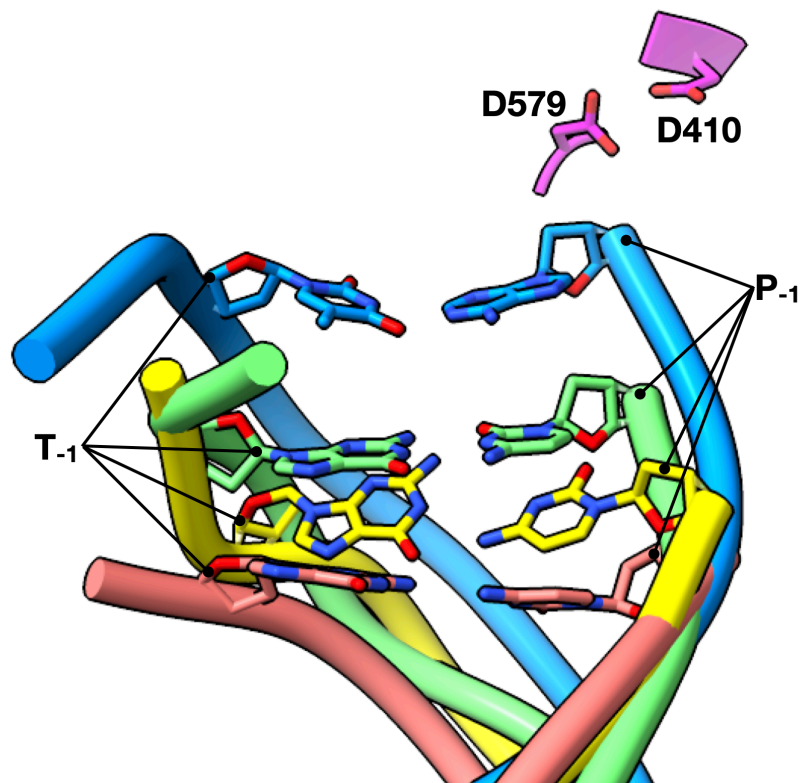


Figure 4.9: DNA movement towards the catalytic aspartates. Position of the T₋₁-P₋₁ base pair of binary1 (salmon), ternary1 (yellow), ternary2 (green), and ternary3 (blue) with respect to the catalytic aspartates (D579 and D410; magenta; only the aspartates of ternary3 are shown for clarity). The DNA backbones are shown as ribbons with the T₋₁ and P₋₁ bases shown in stick representation. All apPol complexes are aligned on the palm domain. We note that the peptide backbones of D579 and D410 of all the apPol complexes were also superimpose to each other.

The ternary1 DNA position is the closest to that of binary1 (Figure 4.9), and it is assigned as the initial collision ternary complex capturing an initial contact of the incoming dGTP with the apPol-DNA pre-chemistry binary complex. Going from binary1 to ternary1 the H₁ and H₂ helices of the thumb push down on the DNA such that the DNA pivots around P₋₃, and the T₋₁-P₋₁ base pair swings in deeper into the palm domain getting closer to the catalytic aspartates (D410 and D579) (Figures 4.9 and 4.11; Movies 2 and 3).

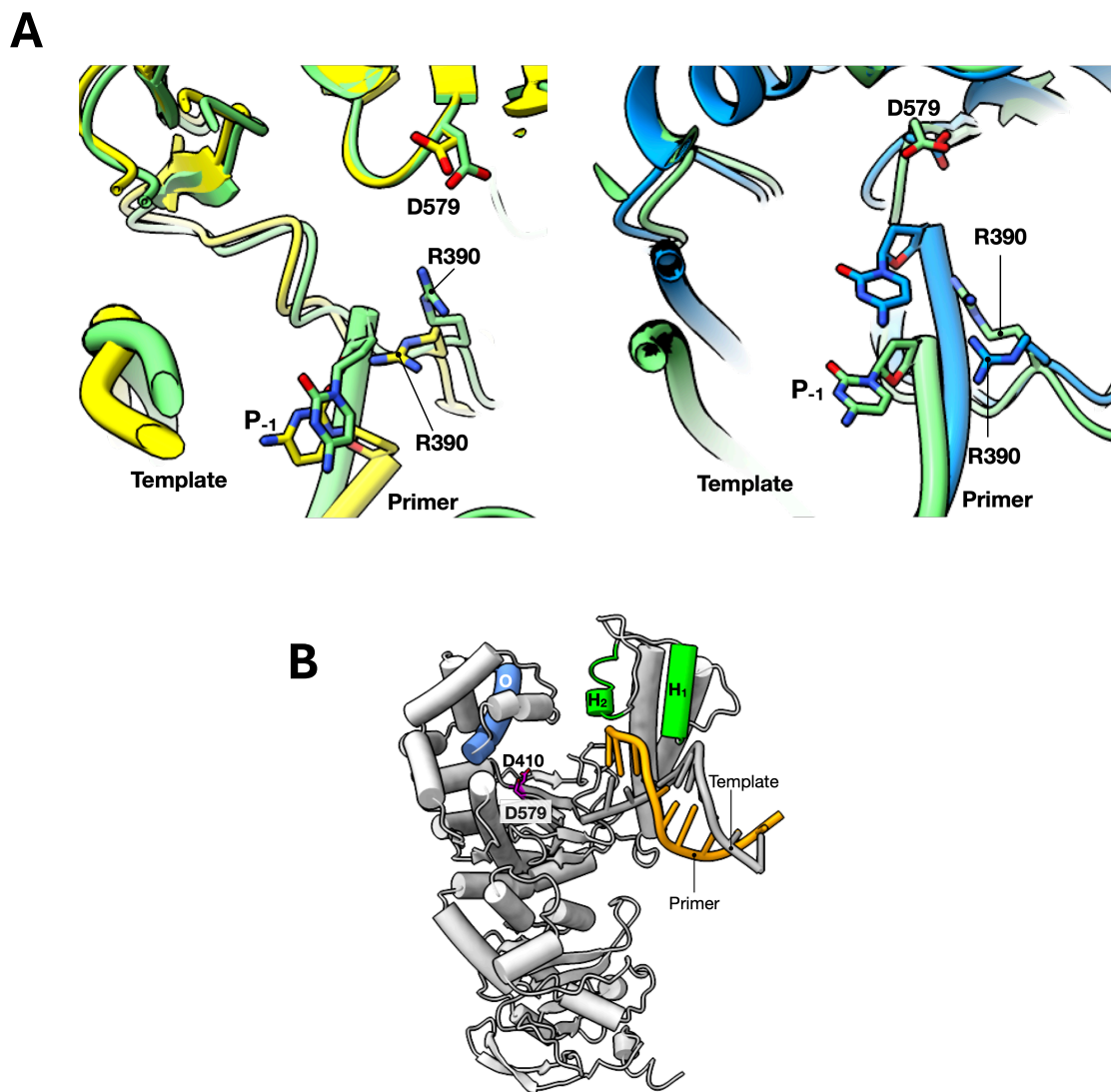


Figure 4.10: Primer strand movement with respect to R390. (A) Position of R390 in ternary1 (yellow) and ternary2 (green) (left) and ternary2 and ternary3 (blue) (right) with respect to the 3' end of the primer. R390, D579, and the P₋₁ nucleotide are shown in stick representation. **(B)** apPol model from Figure 4.2. The protein is coloured white with the O (blue), H1 and H2 (green), helices and D410 and 579 (magenta) highlighted.

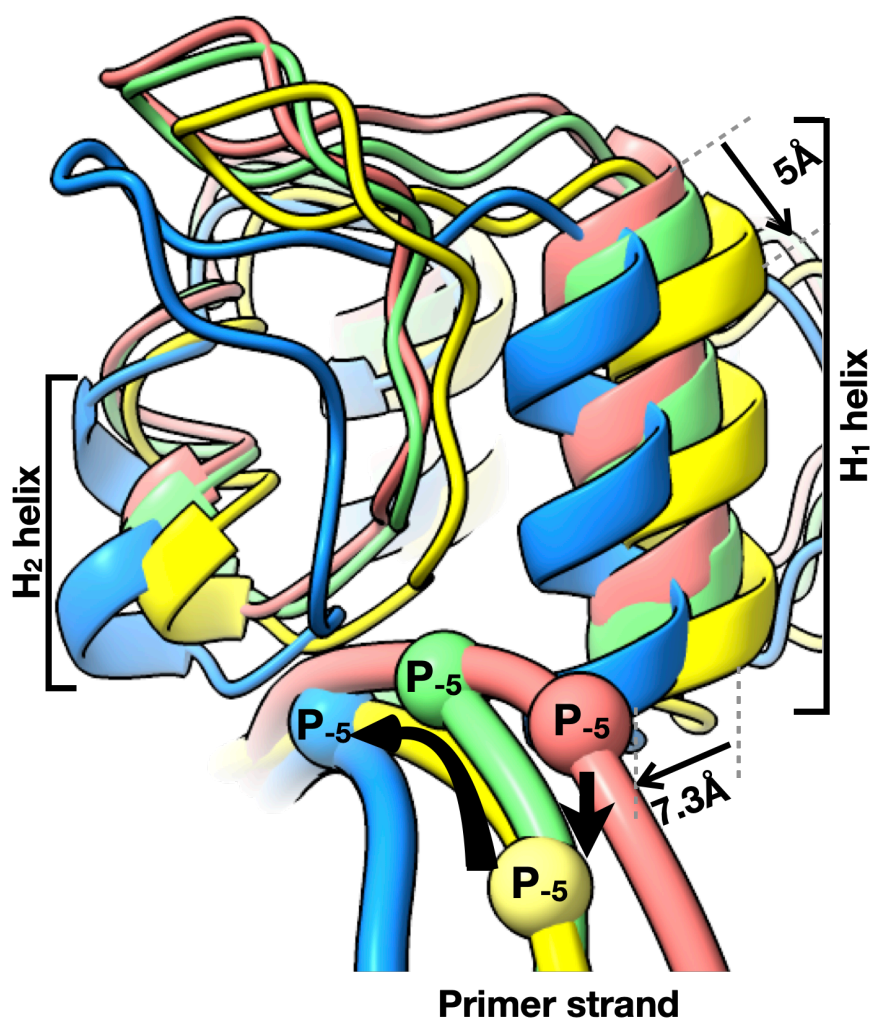


Figure 4.11: DNA movement with respect to thumb domain. Alterations in thumb and DNA positions among the different apPol complexes. For clarity, only H1 and H2 helices (highlighted in green in figure 4.10 B) of the thumb, along with the connecting loop and the DNA primer strands, are shown. The spheres on the primer strands denote the P-5 positions and are used to illustrate the DNA motion. Black arrows show the potential thumb (thin arrows) and DNA (thick arrows) motions. Colour coding: salmon pink: binary1; yellow: ternary1; green: ternary2; blue: ternary3.

In this state the 3' end of the primer strand is ~ 17 Å away from D579 (Figure 4.12). Density for 20 base pairs in the duplex DNA was detected, but no density was present for the template overhang except for the T₀ base. The density for the incoming dGTP was detected in between

the fingers and palm domains (Figure 4.13). The triphosphate moiety of the dGTP is in an extended configuration and is stabilised by Q413 of the fingers domain (Figures 4.13). Compared to the triphosphate, the densities for the ribose and base are weak and could only tentatively model their orientations. We postulate that in ternary1 the incoming dNTP can sample multiple conformations, and once favourable pairing occurs with the T_0 base, the complex moves to ternary2. In addition to the apPol-DNA contacts of binary1, a hydrogen bond between Y485 of the O_1 helix of the fingers domain and the T_0 base was also noted (Figure 4.13). Interestingly, the sidechain of R390 is sandwiched between the 3' end of the primer and the incoming dGTP and acts as a roadblock for the DNA moving towards the fingers (Figures 4.13 and 4.10).

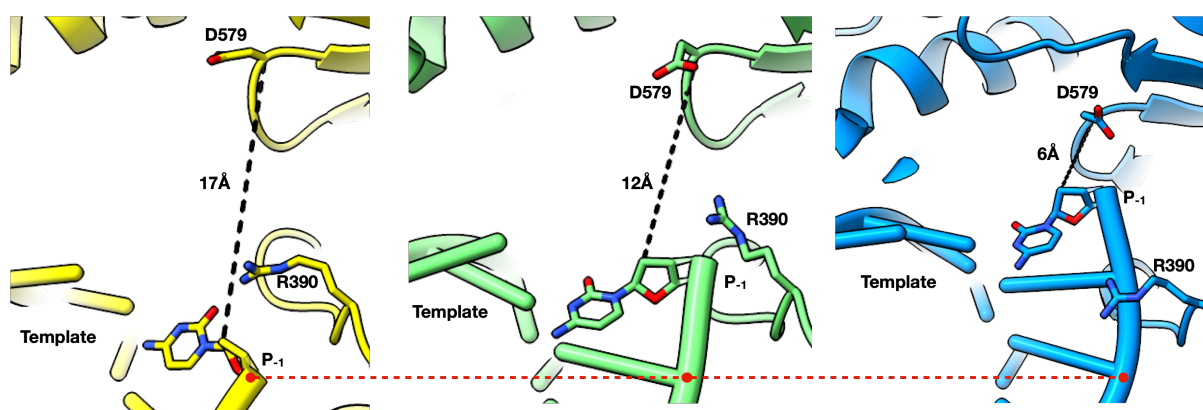


Figure 4.12: Distance between 3' primer end and D579. Distance between the 3' end of the primer strand and D579 (both shown in stick representation) in ternary1 (yellow; left), ternary2 (green; middle) and ternary3 (blue; right). For clarity the incoming dGTP has not been shown. Red dashed line indicates the position of the 3' end of the primer in ternary1. Position of R390 with respect to the primer strand is shown for reference.

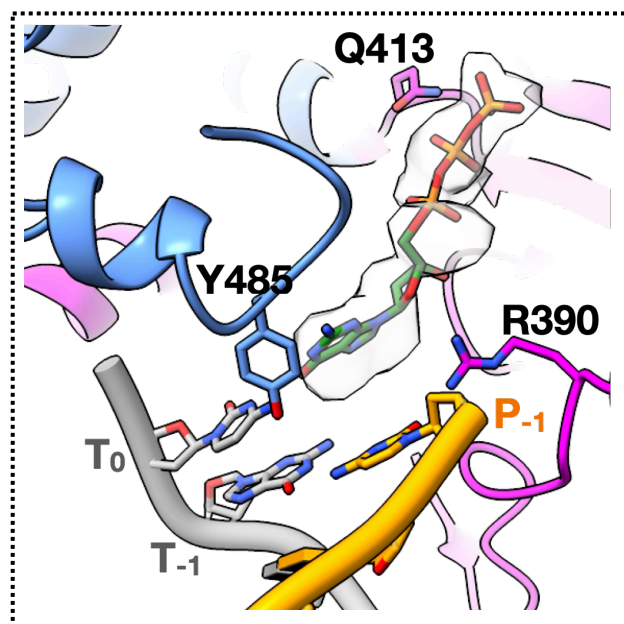
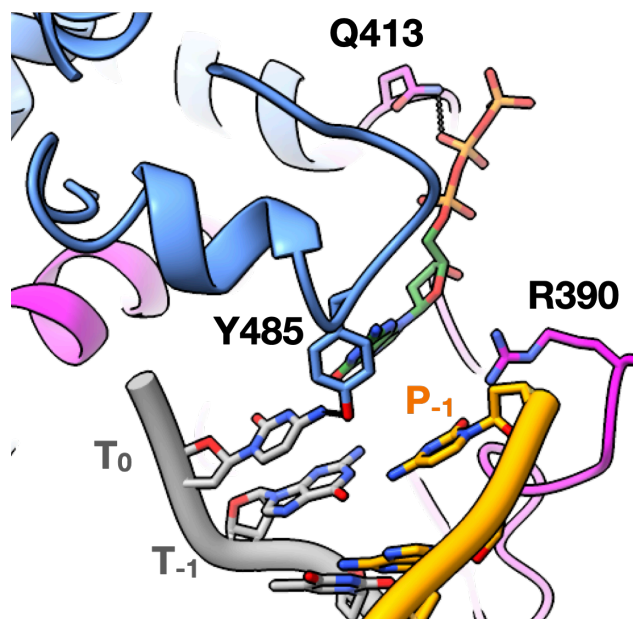


Figure 4.13: Closeup of the active site region of ternary1. The incoming dGTP (green) and the residues interacting with it, Q413 and R390 are shown in stick representation. Dotted line shows hydrogen bond interaction between Q413 and dGTP and between Y485 and T₀. Inset: The cryo-EM density for the dGTP (transparent grey) has been segmented out from the postprocessed map.

4.2.4 apPol-specific O₁ helix extension acts as a nascent base pair checkpoint

Going from ternary1 to ternary2, R390 adopts an altered conformation, no longer restricting the 3' end of the primer terminus (Figures 4.10 A, 4.12). This altered R390, combined with a rigid body motion of the thumb towards the fingers domain (Figure 4.11), guides the DNA towards the fingers and moves the 3' end of the primer closer to the catalytic aspartates by ~6 Å (Figures 4.10, 4.11 and 4.12). In ternary2, the template strand forms multiple hydrogen bonds with the thumb and palm domain residues (Figure 4.14). The incoming dGTP pairs with the T₀ dC and the pairing is consistent with a Watson-Crick base pair (Figure 4.15). The nascent base pair is tucked in a snug cavity formed by Y481, Y485 and Y486 of the O₁ helix (Figures 4.16 A and B; Movie 4). Compared to DNA polymerase I (Pol I), the O₁ helix of apPol is longer by ~1 turn at the C-terminal end (Figure 4.17) and Y481, Y485 and Y486, conserved within the apPol clade of A-family polymerases (Schoenfeld et al., 2013), are positioned in this additional helical turn. Mutating these tyrosines to alanines made the mutant apPol (apPol^{Y481A_Y485A_Y486A}) error prone.

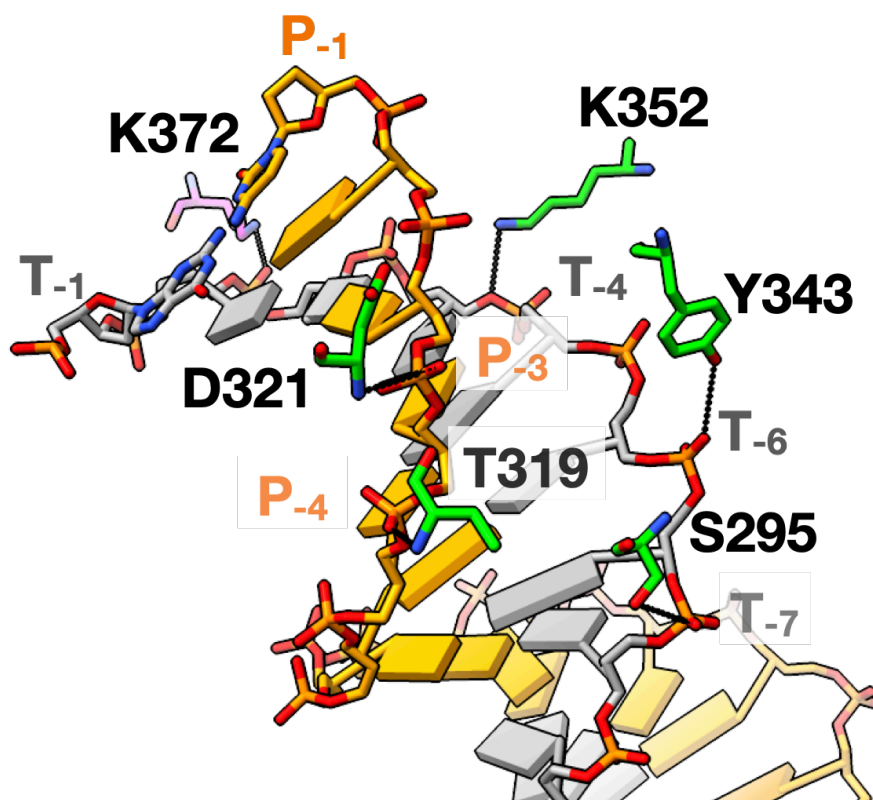


Figure 4.14: DNA-apPol contacts in ternary2. Specific interactions of apPol with the DNA duplex in ternary2. The bases are shown as slabs except for T₋₁ and P₋₁, which, along with the interacting residues are shown in stick representation. Dotted black line shows hydrogen bond interaction between DNA backbone and protein residues.

In addition to the correct incorporation of a dGTP, apPol^{Y481A_Y485A_Y486A} could also misincorporate dATP opposite a templating dC (Figure 4.18). Unlike the incoming nucleotides of the catalytically poised closed ternary complexes of DNA polymerases, the triphosphate moiety of the incoming dGTP in ternary2 adopts a different orientation with the oxygens of the gamma phosphate curled away from the catalytic aspartates, which is referred as goat tail orientation (Figures 4.15 and 4.16) [185]. The dGTP is stabilised by H439 and the backbone of Q413 (Figure 4.15). We could not detect any metal ion density with the incoming dGTP in ternary1 or 2.

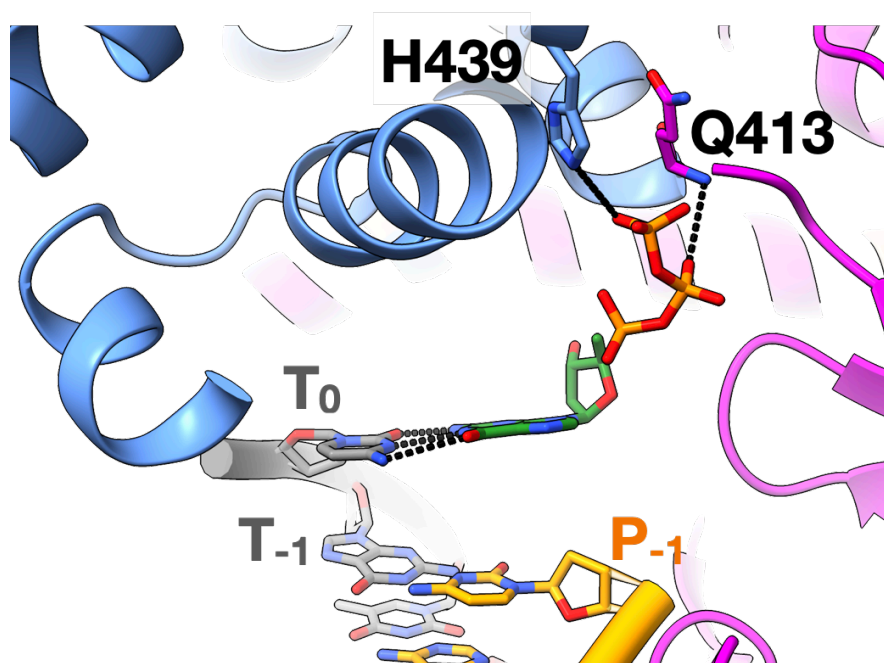


Figure 4.15: A close-up of the nascent base pair of ternary2. Residues interacting with dGTP (green) are shown in stick representation. Dotted black line: hydrogen bonds. Colour coding is similar to figure 4.2.

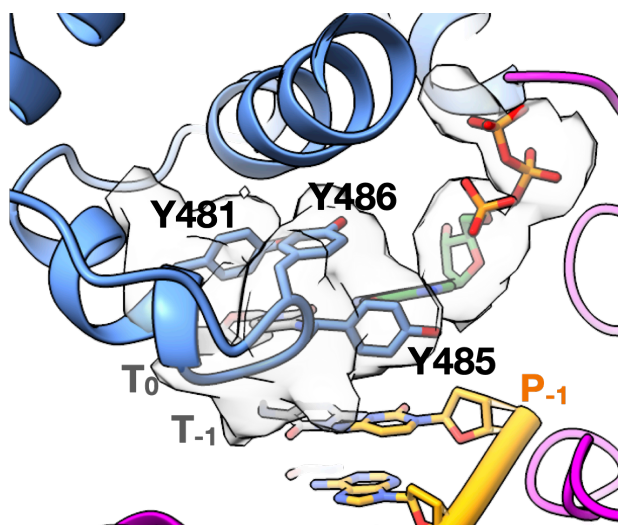
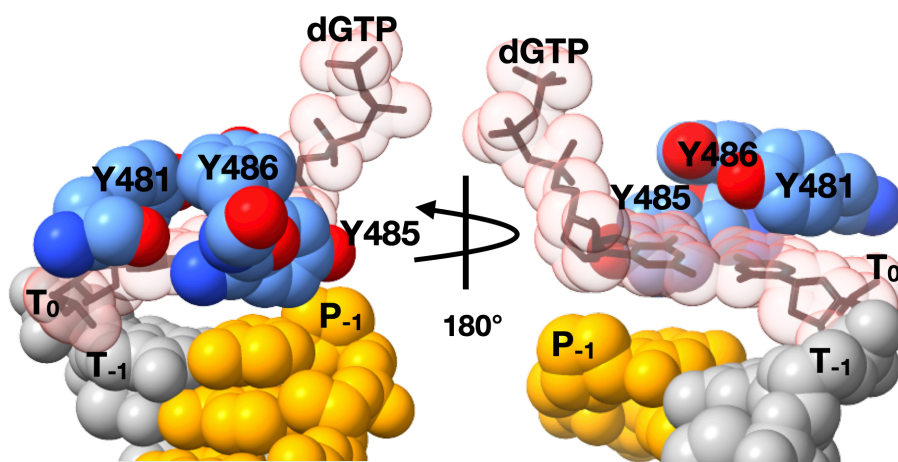
A**B**

Figure 4.16: Pre-insertion checkpoint in ternary2. Pre-insertion checkpoint formed by Y481, Y485, and Y486 of the O1 helix in ternary2. **(A)** Cartoon representation of the pre-insertion checkpoint with the tyrosines shown in stick form. Colour coding is the same as Figure 4.2 with the incoming dGTP shown in dark green. Cryo-EM map density (transparent grey) around the tyrosines and incoming dGTP is shown in transparent grey. **(B)** Space filling sphere representation of the pre-insertion checkpoint. Only the three tyrosines, the DNA duplex and the incoming dGTP are shown. The incoming dGTP and the T₀ nucleotide are shown both in sphere (transparent pink) and stick representation (black).

apPol	1	-----DEITKKYIKDNIINVDNIIKKKDIIFKLKNNENNEITECAFEYFE	44
Pol theta	1	1GFKDNSPI SDT SFSLQLSQDGLQLTPASSSESLSI IDVASDQNLQFTFIKEWRCKKRFSI-----SLACE-KIRSLTSSKTATI-	79
T7 pol	1	-----MIVSDIEANALLESTV-KFH	19
Klentaq	1	-----ALEEAP-----W-----PPPE-----	11
Klenow	1	-----VISYDNYVTIL-----DEETLKAWIAKLEKAP-----VFADTETDSDLNISI-----	42
apPol	45	SKKKFDDDI ESRRFI INDNNYNENINLIYKDIKYCGLDIETTGLE-----VF DENIRL-----IQIAVENYPV I IYDMFNINKK--	118
Pol theta	80	GSRFKQASSPQEIP IRDDGFP IK--GCDDTLVVGLAVCWGGRDAYYFSLQKE-----QKHSEISASLVPPSLDPSLT LKDRMW	155
T7 pol	20	CGVIYD-YSTA EYVSYP RSDFGAYLDALAEAVARGGLIVFHN-GHKYDVPALTKLAKL-----QLNREFHLP----	84
Klentaq	12	YVDGMEWNNFNEEMMDYNVQDVVVTKALLEKLLSDKHYPF--PEIDFTDVGYTTFWSESLE-----AVDI EHRAAWL LAKQER	48
Klenow	43	ANLVGLSFAI EPGVAAYIPVAHDYLDAPDQISRERALELLKPLLEDEKALK-----	93
apPol	119	DILDGLRKVLENKN--IKI I QNGKFDAKFLHNNFKIEN-I- FDTYIASK--LLDKNNMYGFKLNNIVEKYNVILD-----	191
Pol theta	156	YLGSCLRKESDKESVVIYDFIQS--YK- ILLLSGCI SLEQSY-EDPKVACW--LLDPDSQE--PTLSHIVTSFLPHELPLLEGME	233
T7 pol	85	RENCI-----DTLVLSRLIHSNLKDTDMGLLRSGKLPGRKFGSHAL EAWGYRLGEMKG EYKDDFKRML EEQ-----GEE	153
Klentaq	49	ALRDL-----KEA-----RGLLAK--DL SVLALREGLGLPP-G-DPMLLAY--LLDP SNTT-----HRAPEPYK-----	96
Klenow	94	VGQNL-----KYD--RGILAN--YGIE--LRGI-----A- FDTMLESY--ILNSVAGR--HDMDSLAEERWLKHKTITFEEIA	154
apPol	192	KQQQNS- VWNNSLLNNN-----QLFY--AARDSSCLLKLY- KKLKEEIKKE- NLHIVNDI ENKCILP ICDMEL	254
Pol theta	234	TSQGIQSLGL- NAGSEHS-----GRYRASVE--SILIFNSMNQNSLLQKENLQDVFRKVE MPSSQYCLALL EL	298
T7 pol	154	YVDGMEWNNFNEEMMDYNVQDVVVTKALLEKLLSDKHYPF--PEIDFTDVGYTTFWSESLE-----AVDI EHRAAWL LAKQER	229
Klentaq	97	GVARRYG-----GEWTEEAGERAAALSERLFANLWGRLEGEERLLWL YREV ERPL SAVLAHMEA	154
Klenow	155	GK- GKNQLTFNQI ALEEA-----GRY--AAEDADVT LQLHLKMWPD LQKHGKPLNVFENI EMLPVPVLSRI ER	219
apPol	255	NGIKVDLENLQKSTNEILNELNIEKDNLK-----KK-----	285
Pol theta	299	NGIGFSTACESEKQHIMQAKLDAIE-TQA-----YQ-----	328
T7 pol	230	NGFPDTKAI EELYVELAARRSELLRKL TETFGSWYQPKGGT EMFCHPRTGKPLPKYPRIKTPKVGGI FKKPKNKAQREGREPCEL	315
Klentaq	155	TGVRLDVAYLRALSLEVAEEIARLE-AEV-----FR-----	184
Klenow	220	NGVKIDPKVLHNSSEELTLRLAELE-KKA-----HE-----	249
apPol	286	-----LKDENINVNSSQQQVLKALQK-----NNVRDISN-----KL-----IENTSDSNLKNF- LNHE	331
Pol theta	329	LAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKT LGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHP	401
T7 pol	316	DTRYEVAGAPYTPVEHVVFNPSSRDHIQKKLQEA GWVP-T-----KYTDKGAPVVDDEVLEGVVRVDDP	377
Klentaq	185	LAGHPFNLSRDQLERVL FDELGLPAIGKTEK-----TGKRSTSAVLEALREAH	235
Klenow	250	IAGEEFNLSSTKQLQTI LFEKQGIKPLKKT-----GGAPSTSEEVLEELALDYP	299
apPol	332	EII-----SLRNYRRLYKLYSAFYLKLP--LHINTKTNKI-HTTFNQLKTFSGRFSSEKPNLQQIPRQKN-----	393
Pol theta	402	LPG-----LLEWRRIITNAITKVVPFQREKCLNPF LGMERIYPVQSSTATGRITFTFEPNIQNVPRDFEIKMPTLVGESPPSQAVG	483
T7 pol	378	EKQAAIDLKEYLMIQKRI GQSAEGDKAWLRYVAEDGKI-HGSVNPNGAVTGRATHAFPNLAQIPGVRSPY-	447
Klentaq	236	IVE-----KILQYRELTKLKSTYIDPLP--DLIHPRTGRL-HTRFNQTATATGR LCCCDPNLQNI PVRTPL-----	298
Klenow	300	LPK-----VILEYRGLAKLKSTYTDKLP--LMINPKTGRV-HTSYHQA V TGRLSSTDPNLQNI PVRNEE-----	362
apPol	394	-----IREIFI PNDDNFI IADFKQIELKIAAEITNDEIMLKAYNNNIDLHT	440
Pol theta	484	KGLLP MGRGKYKKGF SVNPRCQAQMEERAADRGMPFSI SMRHA FVPFPGGSI LAADY S QLELRILAHLSHRRRLIQV LNTGADVFR	569
T7 pol	448	-----GEQCRAA-----FGAEHL DGTGKPPWQAGIDASGLELRCLAHFMA-----RFDNGEYAE	499
Klentaq	299	-----GQRIRRGFI AEEGWLLVALDY S QIELRVLALHLSG DENLIRV FQEGRD IHT	348
Klenow	363	-----GRRIRQA FI APEDYVIVSADYSQIELRIMAHLSRDKGLLTAFAEGKDIHR	412
apPol	441	LTASII TKKNIPDINKERHIAKAINFGLIYGMNYVNLKNYANTYYGLNMSLDQCLYFYNSSFFEHYKGIYK-----	512
Pol theta	570	SIAAEWKMI EPESVGDDLRQQAKQICYGII YGMGAKSLGEQ-----MGIKENDAACYIDSFKSRYTGINQ-----	635
T7 pol	500	ILNGDIHTKNQIAAELPTRDNAKTFIYGFLYGAGDEKIGQI-----VAGAKERGKELKKKFL ENTPAIAALRESIQQTLV ESSQW	579
Klentaq	349	ETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHL SQE-----LAIPYEEAQAFIERYFQSFPKVRA-----	414
Klenow	413	ATAAEVFGLPLETVT SEORRSAKAINFGLIYGM S A FGLARQ-----LNIPRKEAQK YMDLYFERYPGVLE-----	478
apPol	513	HNQYKQK--RALQYSTLSNRKVIFPY-----FSFTKALNYPVQGTCA DTLK LALVDLYDNLK-----	567
Pol theta	636	MTETVKNCKRQDGFVQTI LGRRRYLPGIKDNNPYRKAHARQAINTIVQGSADIVKIATVN IQKQLET FHSTFKSHGHR EGM LQSD	721
T7 pol	580	VAGEQVQVKKRRWIKGLDGRKVHV-----RSPHAALNTLLQ SAGALICKLWI I KTEEMLVE-----KGLKHG-----	641
Klentaq	415	I EKTLEEGRRRGYVETLFGRRRYVPDL E ARVKSVREAAERMAFNMPVQGTADLMK LAMVKL--FPRL-----	480
Klenow	479	MERTRAQAK EQGYVETLDGRRLYLPDI KSSNGARRAAAEAAI NAPMGTAA D I I KRAMI AVDAWLQA-----	546
apPol	568	-----DINGKI ILCVHDEIIIEVNKKFQEAL-KILVQSMENSASYFLKKYKCEVSVKIAENWGS-----	626
Pol theta	722	RTGLSRKRKLQGMFCPI RGGFFI LQLHDELLYEVVEEDVQVA-QIVKNEMESAV--KLSVKLKVVKIGASW GELKDFDV	799
T7 pol	642	-----DGDFA YMAWVHDEIQVGCRTEEI AQVVIET AQEAMRWVGDHWNFRCL LDT EGKMGPNWAI CH-----	704
Klentaq	481	-----EEMGARM LQVHDELVL EAPKERA EAVA-RLAKEVMEGVY--PLAVPLEVEVGI GEDWLSAKE--	540
Klenow	547	-----EQPRVRMIMQVHDELVEVHKDDVDVAVA-KQIHQLMENCT--RLDVPLLEVGVSGENWQDAH----	605

Figure 4.17: Protein sequence alignment of selected A-family DNA polymerases. The aligned sequences are coloured based on sequence conservation with residues having 100% conservation shown in dark blue background. Pol theta: human DNA polymerase theta (GenBank id: AAC33565.1; the helicase domain was excluded), T7 pol: DNA polymerase from bacteriophage T7 (Genbank id: UJQ71113.1), Klentaq: Klenow fragment of *Thermus aquaticus* DNA polymerase I (PDB id: 1TAQ), Klenow: Klenow fragment of *E. coli* DNA polymerase I (GenBank id: EGH36700.1). The two catalytic aspartates (D410 and D579) of apPol are highlighted with red circles and Y481, Y485 and Y486 of apPol are highlighted with red *. The region around the O1 helix is highlighted with a black box.

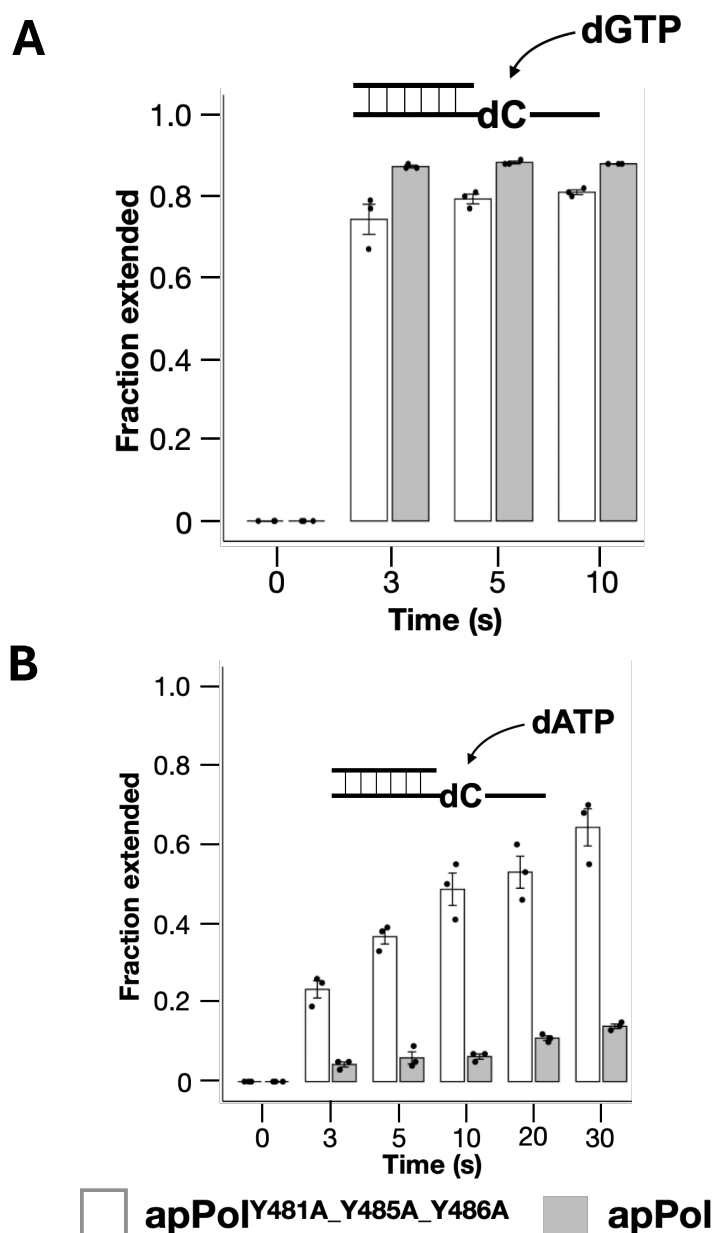


Figure 4.18: Primer extension assay to assess the influence of Y481, Y485 and Y486 on apPol's fidelity. Single nucleotide addition by apPol^{Y481A_Y485A_Y486A} (open bar) or apPol (filled bar) incorporating (A) dGTP opposite dC (correct addition) or (B) dATP opposite dC (misincorporation). 480 nM enzyme was pre-incubated with 200 nM FAM-P/T (Table 1, S12) and 250 μ M dGTP or dATP was added to initiate the reaction. The reactions were quenched after various time intervals with 250 mM EDTA and the fraction of starting primer extended by one nucleotide has been plotted as bars. The experiments were performed in triplicate and the individual replicates are shown as filled black circles with the average value shown as bars. Error bars show the standard deviation.

4.2.5 Catalytically poised apPol has fingers closed around the nascent base pair

To capture ternary3, a primer strand with 3'OH was used but instead of Mg^{2+} , Ca^{2+} was used as the divalent metal ion (dataset 2) (Figures 2.6, 4.4). It has been shown that while Ca^{2+} allows the formation of the catalytically poised ternary complex, catalysis for most polymerases is prevented in part due to the subtle differences in coordination geometries and the size difference between the calcium and magnesium ions [77] [186]. Based on the active site arrangement of ternary3 (Figure 4.19), we concluded that this state is the catalytically poised state of apPol. Going from ternary2 to ternary3, the thumb undergoes a rigid body motion towards the fingers, leading to a stronger interaction with the DNA and a corkscrew movement of the DNA into the apPol active site (Figures 4.9, 4.10, 4.11 and 4.12; Movies 2, 3 and 5). The DNA duplex in ternary3 is stabilised through extensive contacts with apPol (Figure 4.20) with a total buried surface area of $1798 \text{ \AA}^2$. In this state the thumb tracks both the backbone geometry and the base pair orientation of the DNA minor groove (Figures 4.21, 4.22; Movie 2), and their interactions might provide additional checks for potential misincorporations. In ternary3, the 3' primer terminal base is stabilised by a minor groove hydrogen bond with R377 (Figure 4.19), placing the P_{-1} C3' $4.2 \text{ \AA}$ away from alpha phosphate (P_{α}) of the incoming dGTP. Notably, the position of the 3' end of the primer strand of ternary3 is occupied by R390 in ternary2 (Figure 4.10 A). Thus, as the DNA moves into the active site, R390 adopts an altered conformation to accommodate the primer strand (Figures 4.10 A and 4.12).

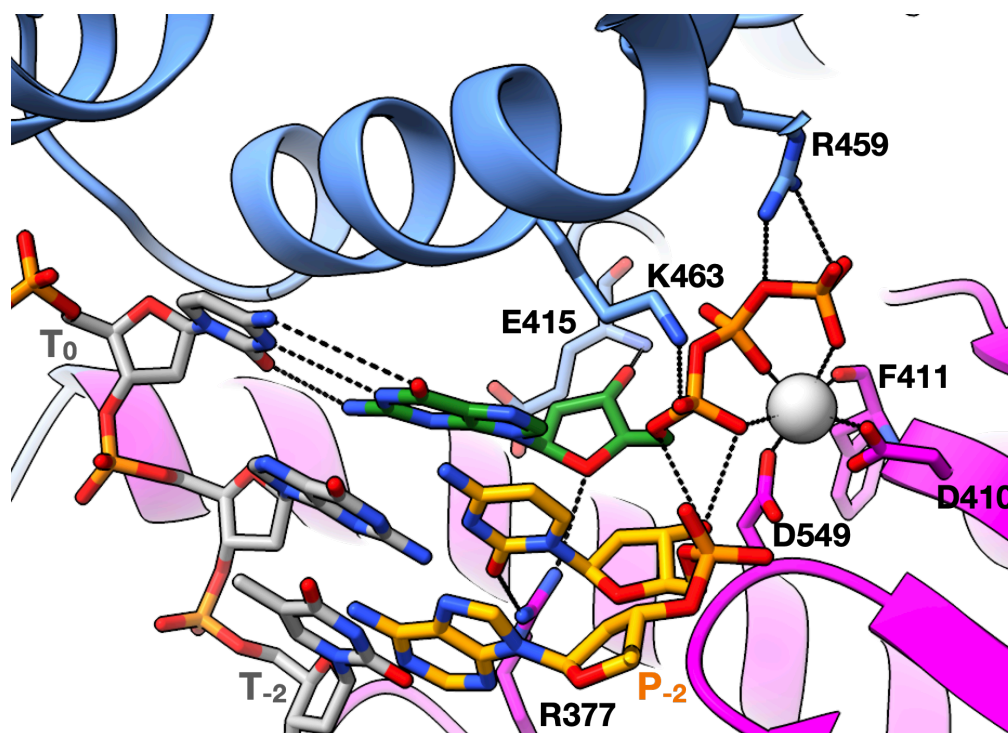


Figure 4.19: Catalytically poised active site of apPol. Residues interacting with the incoming dGTP are shown in stick representation. Dashed black lines denote hydrogen bonds. Colour coding is similar to figure 4.2, with the incoming dGTP shown in dark green and Ca²⁺ is represented as a white sphere.

In ternary³ the O helix of the fingers domain clamp down on the nascent base pair, restricting the active site of apPol in preparation for phosphodiester bond formation (Figures 4.8 and 4.19). The closure of the O helix is accompanied with a concomitant movement of the O1 helix away from the catalytic aspartates (Movies 1 and 4). The O helix closure pushes the dGTP down into the active site with the triphosphate in a configuration conducive for catalysis (Figure 4.19). The triphosphate is stabilised through interactions with R459 and K463 of the fingers. We could detect density for metal B, coordinated by the two catalytic aspartates (D579 and 410), F411 backbone and the triphosphate of the incoming dGTP (Figure 4.19). Interestingly, density for metal A could not be detected, indicating that metal A might have weaker binding affinity compared to metal B. The 3'OH of the primer strand is within hydrogen bonding distance from the incoming dGTP.

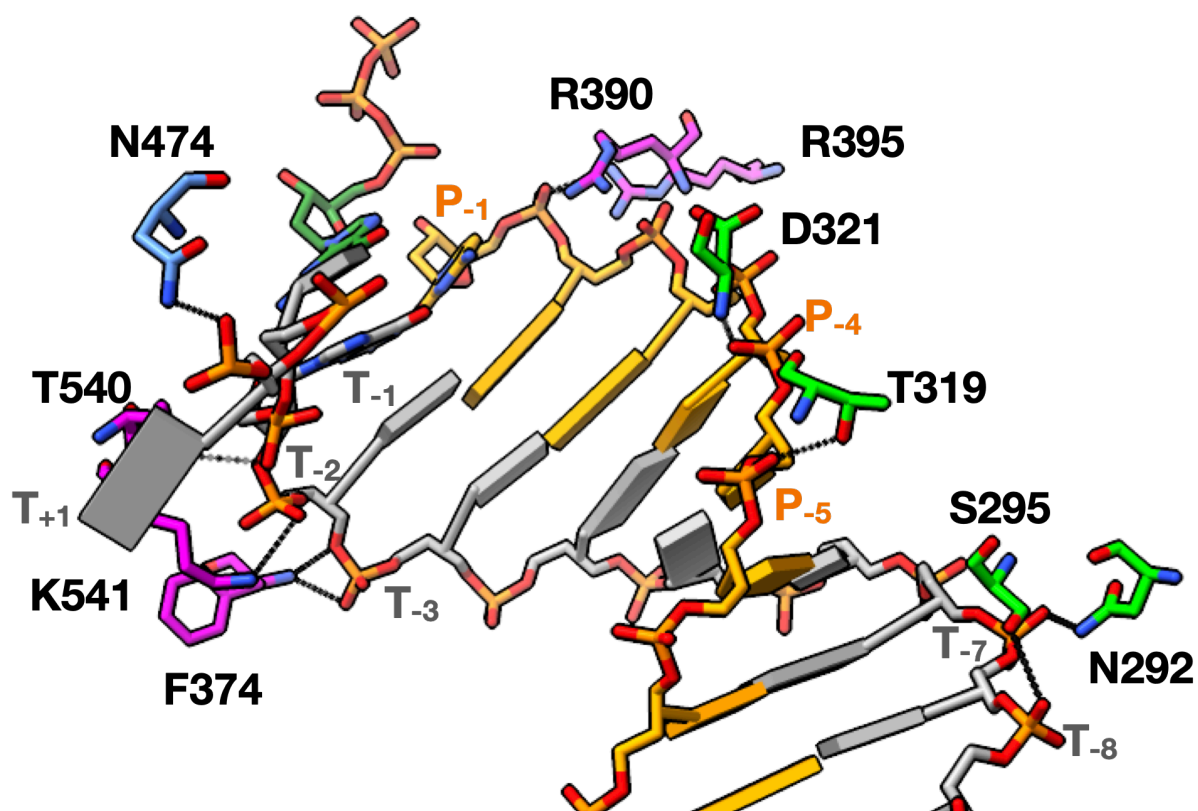


Figure 4.20: DNA-apPol contacts in ternary3. Specific interactions of apPol with the DNA duplex in ternary3. The bases are shown as slabs except for T₋₁ and P₋₁. T₋₁, P₋₁, the incoming dGTP, and the interacting residues are shown in stick representation. Dashed black lines denote hydrogen bonds. Colour coding is same as figure 4.2.

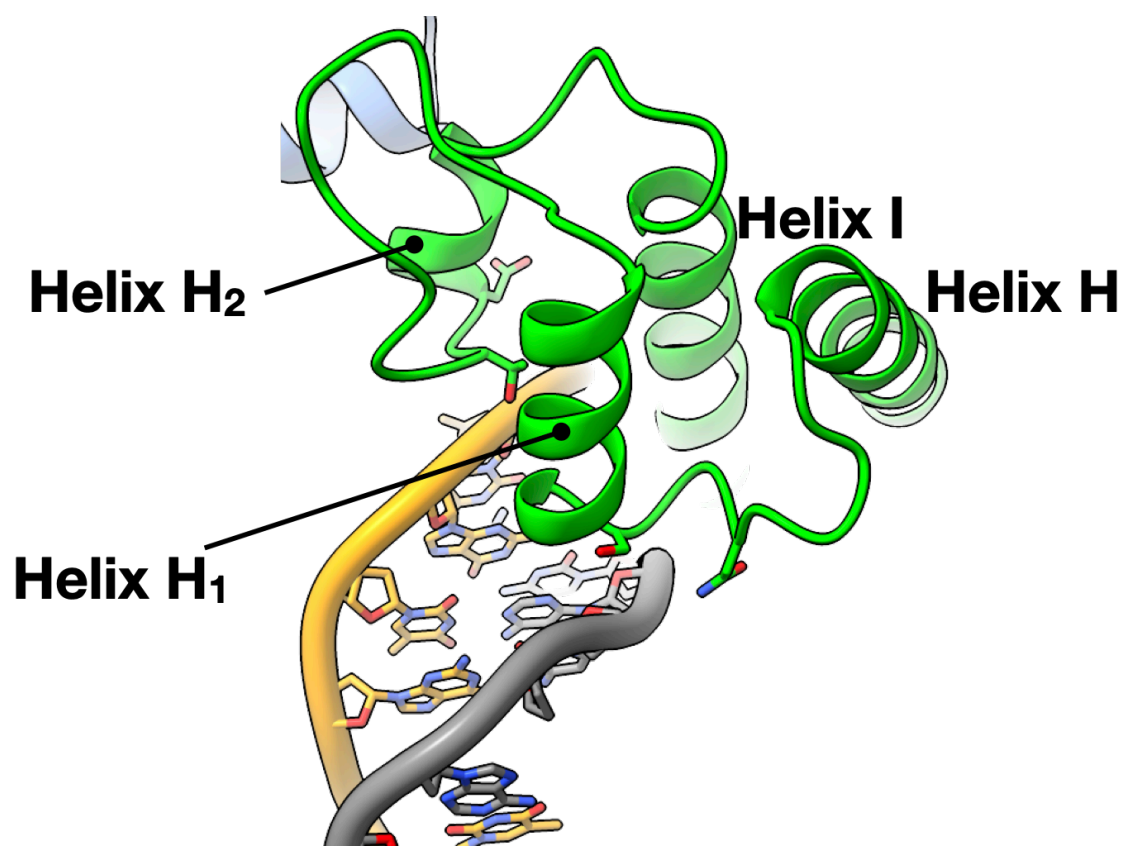


Figure 4.21: Thumb domain interactions with DNA minor groove. Thumb elements of ternary3 interacting with the minor groove backbone of the DNA. The residues involved in specific contacts are shown in stick representation. Colour scheme is similar to figure 4.2.

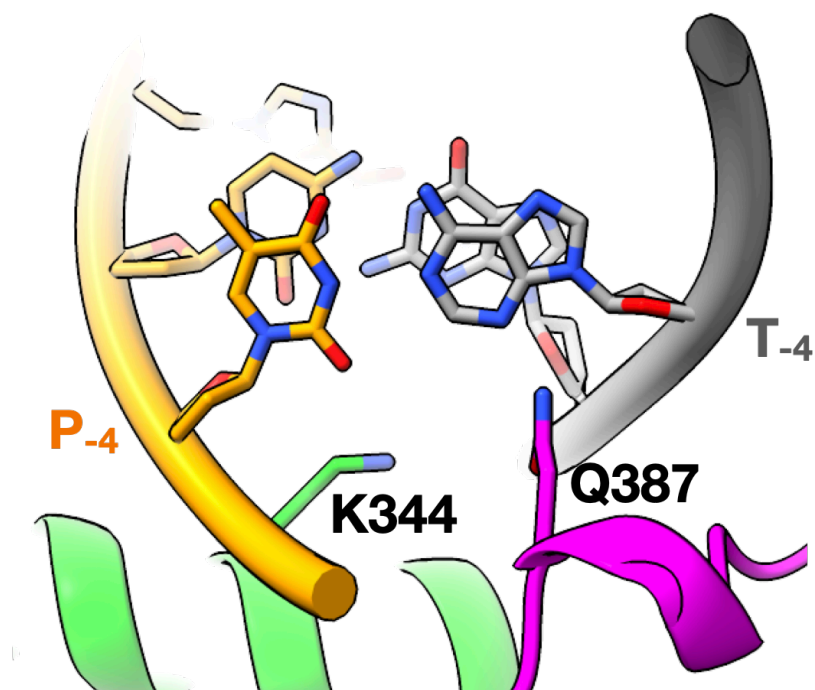


Figure 4.22: Q387 and K344 sensing the minor groove geometry in ternary3. Top-down view of the primer/template DNA duplex highlighting Q387 and K344 sensing the minor groove geometry of the T₋₄-P₋₄ base pair. Colour coding is the same as in Figure 4.2.

4.2.6 Steric gate residue in apPol

The steric gate residue of a DNA polymerase discriminates against ribonucleotide incorporation by sterically excluding the 2'-OH group [187]. From the ternary³ structure, we identified E415 as the steric gate residue [188]. The C2' of the dGTP is positioned 3.7 Å above E415 (Figure 4.19) and we anticipate that this residue will discriminate against an incoming NTP through steric clash with the 2'OH. In fact, this glutamate is conserved in A-family polymerases and acts as the steric gate [178] [188, 189] [190]. We performed single nucleotide primer extension assays with both apPol and a mutant version of apPol where E415 was mutated to alanine (apPol^{E415A}) with either dGTP or GTP as the incoming nucleotide (Figures 4.23, 4.25 A and B).

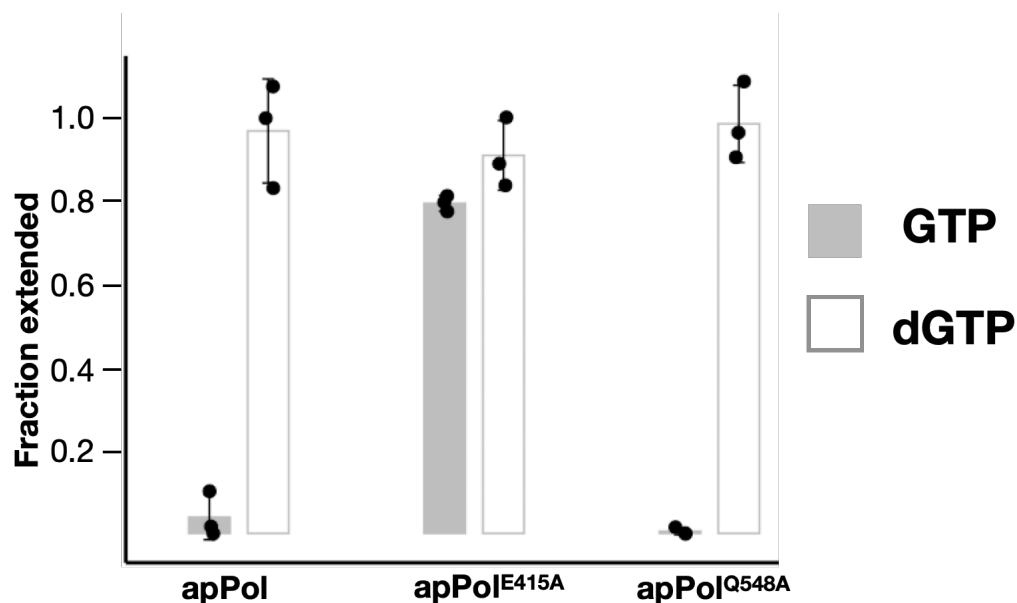


Figure 4.23: Single nucleotide primer extension assay to validate steric gate residue. 1 μ M apPol (either apPol or apPol with an additional point mutation (E415A or Q548A)) (Figure 2.1) was incubated with 50 nM FAM-P/T. 125 μ M GTP (filled bars) or dGTP (open bars) was added to initiate the reaction. The reactions were quenched after 2 minutes with 250 mM EDTA and the fraction of starting primer extended by one nucleotide has been plotted as bars. The experiments were performed in triplicates, and the individual replicates are shown as black circles with the average value shown as bars. Error bars show the standard deviation.

Unlike apPol, apPol^{E415A} can incorporate both GTP and dGTP opposite dC (Figure 4.23), although GTP incorporation was slower than dGTP (Figures 4.24). This indicates that E415 indeed plays a key role in discriminating GTP over dGTP as the incoming nucleotide. Recently it has been proposed that Q548 might act as the steric gate residue for apPol [179]. However, primer extension assays with an apPol^{Q548A} shows that Q548 does not discriminate against GTP (Figure 4.24). Ternary3 is the only state where we could detect the template overhang density for the T₊₁ position. N474 forms a hydrogen bond with the phosphate of T₊₁ thus making this base ordered (Figure 4.20). The T₊₁ base is in a flipped-out configuration, excluding it from the active site (Figures 4.19 and 4.20). This flipped-out configuration is essential to avoid a steric clash with Y481.

In conclusion, the cryo-EM structures described in this chapter provide a comprehensive picture of the pre-chemistry steps of nucleotide incorporation by apPol and provide the first insight into how this enzyme might multitask to perform both replicative and lesion bypass synthesis. The potential multitasking mechanism and the broader implications of these structures are discussed in Chapter 5.

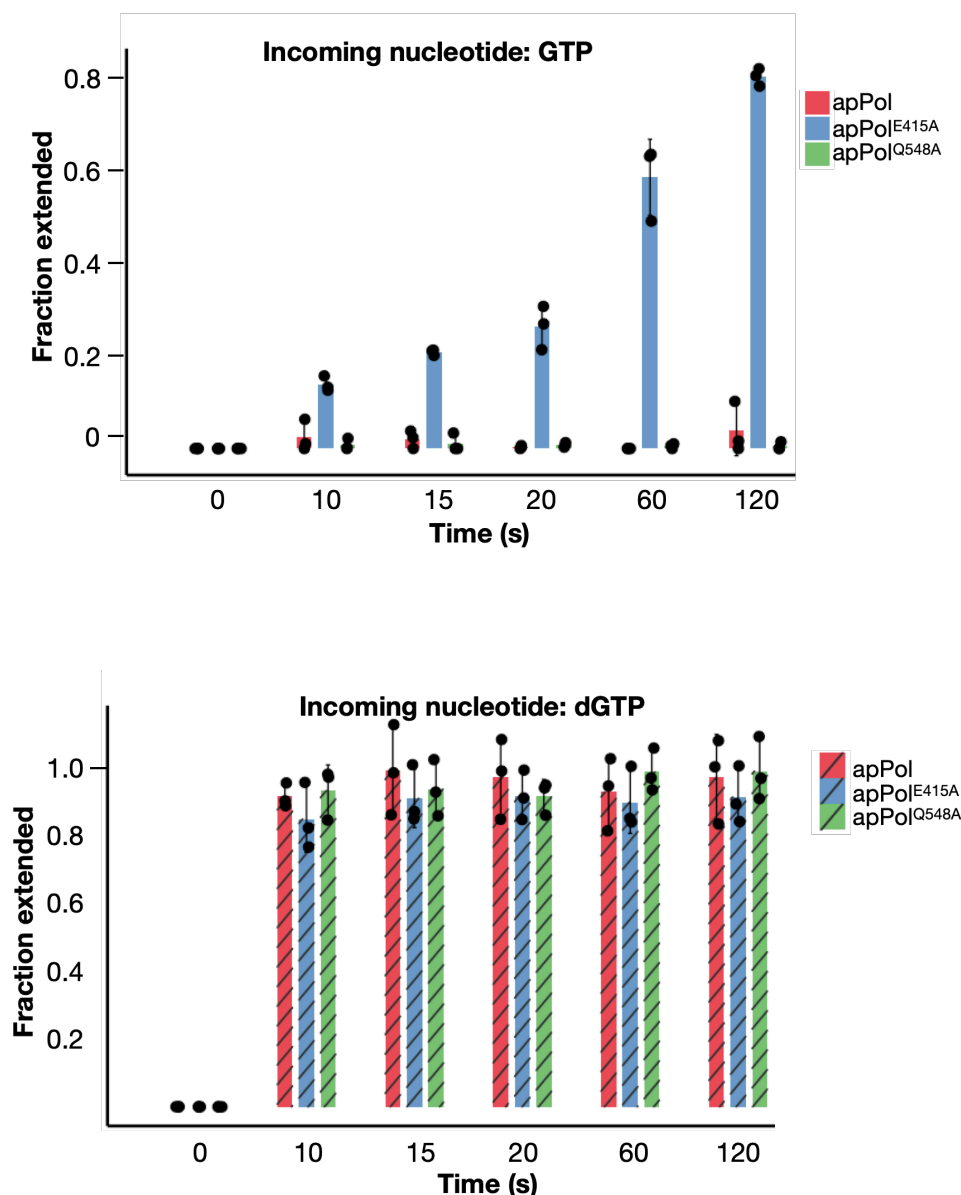


Figure 4.24: Single nucleotide primer extension assay to validate steric gate. Time courses of GTP (top; solid colours) or dGTP (bottom; hashed bars) incorporation by apPol (red), apPol^{E415A} (blue) and apPol^{Q548A} (green). 1 μ M enzyme was incubated with 50 nM FAM-P/T (Table 1, S12) and 125 μ M of dGTP or GTP was added to initiate the reaction. The reactions were quenched after various time intervals by adding 250 mM EDTA and the fraction of starting primer that has been extended by one nucleotide has been plotted as bars. The experiments were performed in triplicates, and the individual replicates are shown as black circles on the respective bars (representing average of the three replicates). Error bars show the standard deviation.

5. DISCUSSION

Apicoplast DNA polymerase (apPol) is the sole DNA polymerase within the apicoplast, an organelle inside *Plasmodium falciparum*. In contrast to other replisomes that deploy separate polymerases for high-fidelity genome duplication and lesion bypass, the apicoplast appears to rely on a single polymerase to perform these distinct tasks (Figure 1.6). This indicates that apPol must multitask and balance accurate replication of the organellar genome with the ability to tolerate and bypass DNA lesions that are likely to arise due to the oxidative stress within the organelle. In the first part of this dissertation, I have defined the minimal kinetic pathway of apPol on both undamaged and lesion-containing DNA, thereby establishing its intrinsic capacity to support both replicative and translesion synthesis. In the second part, I used cryo-EM to solve five pre-chemistry complexes of apPol in different compositional and conformational states. By analysing these five structures of apPol engaged on undamaged DNA, I identified unique features of apicoplast DNA polymerase compared with other A-family DNA polymerases and proposed a mechanism for apPol's multitasking behaviour. Taken together, our findings reveal how a single DNA polymerase can execute multiple functions to support a minimal replisome.

5.1 Kinetic signature and multitasking behaviour of apicoplast DNA polymerase

In this work, I present the first comprehensive kinetic analysis of the *P. falciparum* apicoplast DNA polymerase, including detailed characterization of its lesion bypass activity using pre-steady state kinetic assays. The implications of these results for apicoplast genome duplication are summarized below.

5.1.1. Kinetics of apPol mediated DNA synthesis and its implication for apicoplast genome duplication.

We found that compared to other replicative DNA polymerases, apPol has surprisingly low efficiency and processivity, more in line with DNA polymerases involved in DNA repair and TLS (Table 3). In our primer extension assays to analyse the kinetics of copying undamaged DNA, the incoming dNTP was a dTTP. Hence, we have only determined the efficiency of dTTP incorporation by apPol. Given that ~43% of the apicoplast genome comprises of dATP, poor dTTP incorporation efficiency of apPol indicate that its processivity within the organelle might be even lower than what we have determined unless the pool of dTTP within the apicoplast is significantly higher than the dTTP concentration within prokaryotic or eukaryotic cytosol [124] [24] [101]. A recently published work investigating apPol's kinetics of copying undamaged DNA validates our findings and shows that the efficiency of incorporation can vary over an order of magnitude depending on the identity of the incoming nucleotide [93]. For instance, efficiency of dATP incorporation is 4 times better than the efficiency of dTTP incorporation. Given that individual strands of the apicoplast DNA do not have any significant skew towards dA or dT, we anticipate that the poor dTTP incorporation efficiency might lead to frequent stall or pause during apicoplast replication.

If apPol copies the apicoplast genome on its own, then it must do so using a non-processive mechanism, unlike other replicative polymerases. As a part of the replisome, a DNA polymerase interacts with various components of the replication machinery. Most replicative DNA polymerases (with the exception of certain phage DNA polymerases like phi29 phage polymerase) [191] interact with a partner protein called the processivity factor [192] which increases the processivity of a polymerase either by reducing the dissociation rate of the pre-chemistry binary complex or by increasing the rate of chemistry [156] [193]. In certain cases,

a processivity factor may even increase the affinity of the incoming nucleotide for the pre-chemistry binary complex [159]. However, to date, no processivity factor for apPol has been identified, and it has been proposed that the relatively small size of the apicoplast genome (~35 kb) might allow apPol to perform replicative synthesis without any associated factors [90]. It remains to be seen if components such as the helicase or SSB of the apicoplast replisome or an as-yet-undiscovered protein influence apPol's efficiency and processivity. Alternatively, it is possible that the bulk of the apicoplast genome is copied by a DNA polymerase that has not been identified yet. Future studies looking at the activity of apPol as part of the apicoplast replisome will shed more light on our understanding of how this polymerase duplicates the apicoplast DNA.

5.1.2 Translesion synthesis by apPol might be a source of mutation for the apicoplast genome.

Inside a human host, apicoplast replication occurs during the blood stage of *Plasmodium*'s life cycle [90]. In addition to the high levels of ROS in erythrocytes, the biochemical pathways of the apicoplast are anticipated to generate ROS as a byproduct. *Plasmodium* infection leads to further ROS production as part of the host's defence mechanism [194] and many conventional antimalarials like quinolones utilize a ROS-mediated mechanism to kill *Plasmodium* [195]. ROS are a potent mutagen that generates ODNs [165]. In its role as the only known DNA polymerase within the apicoplast, we expect apPol to encounter and effectively copy oxidatively damaged DNA.

We found that apPol can bypass both 8-oxo-dGMP and abasic site lesions with varying efficiencies. 8-oxo-dGMP is the most abundant of the ODNs and can be present in the DNA through in situ oxidation of dGMP or due to incorporation of the lesion by a DNA polymerase [165]. Abasic sites are formed as a result of hydrolysis of the base moiety of a

nucleotide and are arguably the most deleterious of the ODNs [196]. In addition to ROS-mediated generation, abasic sites can be formed via a variety of different routes including spontaneous hydrolysis of the DNA [197]. We found that, unlike the robust bypass of 8-oxo-dGMP, incorporation of a nucleotide opposite the abasic site was much slower, with ~3000-fold lower efficiency of bypass ($8.1 \times 10^{-5} \mu\text{M}^{-1} \text{s}^{-1}$) compared to the efficiency of incorporation opposite an undamaged DNA (Table 4). Our data indicate that in the apicoplast, apPol might not bypass the majority of the abasic sites and that a separate pathway is probably used to handle this lesion. Similar to apPol, Pol gamma is inefficient at bypassing abasic sites, and in the mitochondria, the base excision repair (BER) pathway repairs the majority of abasic site lesions and prevents mitochondrial DNA degradation [198]. An intact long patch BER pathway has been identified in apicoplast [90], and it is possible that this pathway is the primary route for abasic site removal. Unlike apPol and Pol gamma, the A-family chloroplast DNA polymerases can bypass and extend past an abasic site with high efficiency [199], underscoring the inherent differences in TLS mechanisms among the organellar replicative polymerases.

In contrast to helix-distorting lesions like an abasic site, 8-oxo-dG do not block replication. Rather, this lesion is a potent mutagen [200]. If left unrepaired, an 8-oxo-dGMP: dA misincorporation can lead to a G to T transversion in subsequent rounds of replication. To counter this, cellular replisomes use specific TLS polymerases that preferentially incorporate dCTP opposite 8-oxo-dGMP [201]. We show that apPol bypasses 8-oxo-dGMP with an almost equal propensity of incorporating a dCTP or a dATP opposite this lesion. This promiscuity, reflected in the low fidelity of 8-oxo-dG bypass (1.6; Table 4), is a distinguishing characteristic of apPol that is in contrast with other A-family polymerases studied to date. Although the accuracy of bypass varies between different A-family

polymerases (Table 4), these polymerases preferentially incorporate dCTP opposite 8-oxo-dG.

Our results further show that the proofreading activity of apPol cannot discriminate between dCMP:8-oxo-dGMP and dAMP:8-oxo-dGMP base pairs, treating them instead as mispairs to be removed by the exonuclease site, suggesting that even the ‘correct’ dCMP:8-oxo-dGMP pair does not adopt the canonical Watson–Crick geometry. Our observations are consistent with previous structural studies investigating the molecular mechanism of 8-oxo-dGMP bypass by the A- family *Bacillus stearothermophilus* DNA polymerase I [202], where the authors showed that a dCMP:8-oxo-dGMP base pair in a postinsertion binary complex resembles a DNA mismatch. Unlike the bypass of an abasic site, apPol bypasses 8-oxo-dGMP with relatively high efficiency (Table 4), making it possible that a significant fraction of 8-oxo-dGMP lesions are bypassed by apPol during apicoplast DNA replication. Given the low fidelity of apPol’s nucleotide incorporation opposite this lesion, such bypass might be a major source of mutations within the apicoplast genome. Alternatively, these misincorporations might be rectified post-replication by an apicoplast DNA repair pathway like base excision repair. For viral and cellular replisomes from all domains of life, it has been reported that partner proteins such as the processivity factor can increase the fidelity and/or lesion bypass activity of a DNA polymerase, [203] [204] [205] [206] [207] [208] and it remains to be seen if this holds true for apPol-mediated TLS.

5.1.3 Pre-steady state kinetics define the mechanism of processive synthesis.

While this work reports the first transient state kinetic analysis of nucleotide incorporation and TLS activities of apPol, there have been previous reports on the multiple turnover kinetics of apPol, including a recent study exploring the TLS capacity of this polymerase

(albeit with a slightly different protein construct) [93] [97, 98]. Based on the steady-state analysis, the authors concluded that the rate of dNTP incorporation opposite an undamaged base by apPol is the same as the rate of adding a nucleotide opposite an oxidative lesion [98]. This conclusion is inconsistent with our results and arises from the limitations of steady-state primer extension assays. We found that apPol copies an undamaged nucleotide over an order of magnitude faster compared to replicating 8-oxo-dGMP (Table 3 and 4). In fact, the rates of nucleotide incorporation determined by us are over 1 to 3 orders of magnitude faster than the corresponding steady-state rates reported previously [93] [97, 98]. Moreover, the efficiencies of incorporation of dCTP and dATP opposite 8-oxo-dGMP calculated in this work are an order of magnitude higher than the corresponding efficiencies reported based on steady-state experiments [98].

In conclusion, our biochemical work provides the first detailed glimpse of the kinetic mechanism of DNA synthesis by apPol and lays down the foundation for further kinetic characterisation of the apicoplast replisome.

5.2 Structural and mechanistic insights into multitasking by apicoplast DNA polymerase

In chapter 4 we report a series of cryo-EM structures showing multiple pre-chemistry states of apPol in complex with its DNA and nucleotide substrates. These structures explain how apPol transitions through the initial stages of interactions with DNA and dNTP, ultimately forming a catalytically poised pre-chemistry ternary complex while maintaining the Watson–

Crick geometry of the nascent base pair. We place our results within the context of A-family polymerases and propose a mechanism by which apPol might accommodate DNA lesions within its active site while biasing the nascent base pair towards Watson–Crick geometry, a prerequisite for performing both replicative and lesion bypass synthesis within the same DNA polymerase.

5.2.1 apPol uses novel mechanism to perform TLS activity.

Pol I has been a model A-family replicative polymerase and based on a series of structures of *B. stearothermophilus* Pol I (Bst Pol I) performing nucleotide addition *in crystallo*, a conformational coupling mechanism was proposed for high-fidelity replication [140, 209, 210]. In the pre-chemistry binary complex, the O helix remains open with the insertion site (position of T₀ base during catalysis) occupied by a tyrosine residue at the C-terminus of the O helix (Y714 in BSt Pol I) (Figure 5.1). This tyrosine is conserved in all A-family polymerases and occupies the insertion site in most apo or binary A-family polymerase structures [211] [212-214], potentially preventing premature base pair formation with the fingers in the open configuration. In the fingers-open state of Bst Pol I, the T₀ base remains lodged within a pocket formed by the O and O₁ helix (pre-insertion site) (Figure 5.1).

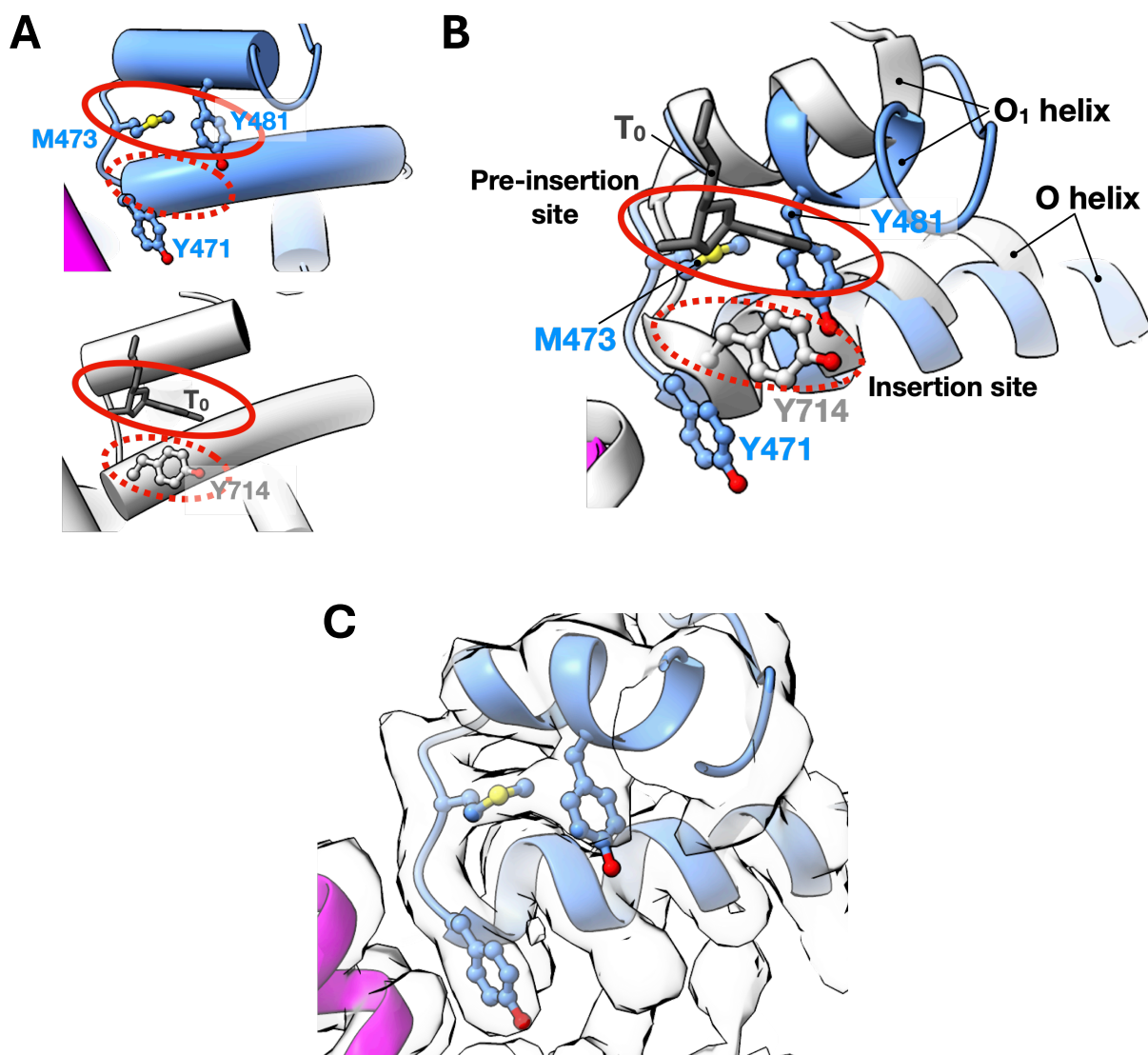


Figure 5.1: apPol's pre-insertion site is not accessible to T₀ base. (A) Close-up view of the pocket between the O and O₁ helices of the fingers domain of the apPol binary1 complex (top) and BSt Pol I (bottom; PDB id: 1L3S). For clarity, the DNA is not shown, and the helices are shown as tubes. (B) Superposition of the pockets shown in panel (A) on the BSt Pol I binary structure (PDB id: 1L3S; white). apPol colour coding is the same as Fig. 1B. BSt Pol I is shown in white and the T₀ nucleotide of BSt Pol I is shown in dark grey. The pre-insertion and insertion sites are highlighted with solid red and dashed red ovals, respectively. Y481, M473, and Y471 of apPol and Y714 of BSt Pol I are shown in ball-and-stick representation. (C) Map density of apPol's pre-insertion site. Atomic model of binary1 in the orientation shown in panel B with the sharpened cryo-EM map (transparent grey) superimposed on it.

As the polymerase moves from binary to ternary state, the O helix closes, leading to a conformational change in the loop connecting the O and O₁ helices, which (a) blocks the pre-insertion site and (b) alters the orientation of Y714 and frees the insertion site. Thus, the T₀ base is forced out of the pre-insertion site and into the newly available insertion site. Moreover, the altered conformation of the loop connecting the O and O₁ helices forces the T₊₁ base to adopt a flipped-out orientation to avoid any steric clashes. These coupled conformational changes ensure that (a) no premature bond formation occurs with the fingers open, thus biasing the nascent base pair towards Watson–Crick geometry (since in the fingers-closed state the active site can accommodate predominantly Watson–Crick base pair), and (b) only a single templating base gets access to the insertion site during a single round of the catalysis, thereby reducing the chance of a frameshift error.

apPol does not undergo the coupled conformational changes noted for Pol I. In the binary1 complex, the pre-insertion site is blocked by M473 and Y481, excluding the T₀ base from this position (Figure 5.1). Moreover, the tyrosine at the C terminus of apPol's O helix (Y471) does not occupy the insertion site in the binary1 complex (Figure 5.1) and consequently apPol can form the nascent base pair with the fingers in the quasi-open state as seen in ternary1 and 2 (Figures 4.13 and 4.15, Movie 1). Since there are very few contacts between the templating base or the nascent base pair and apPol in binary1 and ternary1, respectively (Figures 4.6 and 4.13), the nascent base pair formation with the fingers in the open state is anticipated to be independent of the identity of the bases. This might allow apPol to accommodate a nascent base pair that deviates from the Watson–Crick geometry, a feature that could aid in apPol's TLS activity.

Base pairing in fingers-open configuration is unique to apPol and has not been observed in any other known A-family DNA polymerase (Figure 5.1), raising the question what allows base pairing in the fingers-open state. A potential explanation is that, in apPol nascent base

pairing can occur at a position relatively distant from the active site, as captured in the ternary1 and ternary2 states. After the base pair formation the large DNA movement facilitates the repositioning of the pre-formed nascent base pair towards the catalytic aspartates at the active site. These coordinated rearrangements are visualized across successive conformational states (Movie 4 and in Figure 5.2).

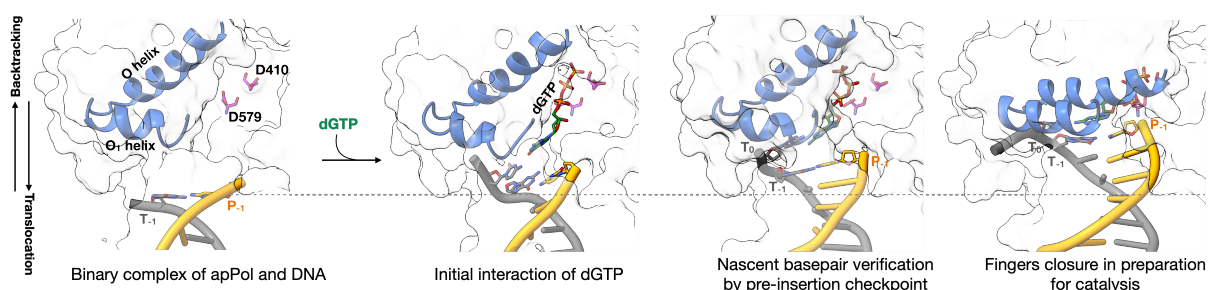


Figure 5.2: apPol backtracks during the pre-chemistry steps. Zoomed-in view of 3' primer terminus in binary1, ternary1, ternary2 and ternary3 complexes (left to right) with the DNA primer (orange)/template (grey), O and O1 helices (blue) shown in ribbon representations. The dGTP (green), catalytic aspartates (magenta), and the P₋₁, T₋₁ and T₀ bases are shown in stick representations. The rest of the apPol is shown as a translucent grey surface. The grey dashed line marks the position of P₋₁ base in binary1. The left black arrow indicates the direction of apPol movement relative to the DNA, highlighting backtracking and forward translocation. All structures were aligned on the palm domain.

5.2.2 apPol uses two fidelity checkpoints to perform high fidelity replicative synthesis

Base pairing in open configuration raises the question: how does apPol ensure fidelity during nucleotide incorporation? Our structures suggest that apPol might use two fidelity checkpoints for accurate synthesis. The first checkpoint is captured by ternary2, where the nascent base pair fits within the snug pocket formed by Y481, Y485, and Y486 at the tip of the O₁ helix (Figure 4.16 A, B; Movie 4). This pocket is in fact a modified pre-insertion site. By modifying this site, apPol has converted the pre-insertion site into a pre-insertion checkpoint to assess the nascent base pair geometry. Consistent with this hypothesis, mutations in these three tyrosines to alanines reduced the accuracy of DNA synthesis (Figure 4.17 A and B). Moreover, using an assay measuring multiple nucleotide incorporation and strand displacement abilities of apPol, it has been reported that apPol^{Y481A_Y485A_Y486A} shows an overall reduced activity [176]. Notably, Y481, Y485, and Y486 are conserved in the apPol clade of the A-family but absent in other A-family polymerases studied to date (Figure 4.18) [176]. Another question is how this pre-insertion checkpoint might play a role in apPol's TLS activities. Further structure–function correlation studies with the triple-mutant apPol will be essential to answer this question. The second checkpoint is captured by ternary3, where the nascent base pair is poised for catalysis and the fingers is in a closed conformation (Figure 4.19). In this state the active site of apPol is similar to the tight active site of a replicative polymerase (Figure 1.7), which will favour the Watson–Crick base pair geometry. These two fidelity checkpoints might allow apPol to perform accurate replicative synthesis.

5.2.3 ApPol thumb is a master manipulator of DNA position

In all A-family polymerases studied to date, the DNA undergoes minimal movement between the binary and ternary complexes [140] [184, 215]. However, apPol is an exception to the rule with the enzyme backtracking with respect to the DNA [211] during transition from pre-chemistry binary to the catalytically poised pre-chemistry ternary state (Figure 5.2, Movie 3). Going from binary1 to ternary1 a downward motion of the thumb facilitates DNA movement towards the active site (Figure 4.11, Movie 2), pushing the 3' end of the primer towards the active site. Further, as apPol transitions from ternary1 to ternary2 and finally to ternary3, the DNA undergoes additional corkscrew motion, positioning the 3' end of the primer into the active site (Figure 4.10 and 4.12; Movie 3).

This large-scale DNA movement is orchestrated in part by a rigid body motion of the thumb towards the finger's domain. In addition, the DNA slips with respect to the thumb to achieve the full extent of the movement. For instance, T319 of the thumb domain contacts the phosphate backbone of P₋₃ in binary1 and ternary1, P₋₄ in ternary2, and P₋₅ in ternary3 (Figures 4.6, 4.14 and 4.20; Movie 6). As apPol progresses from binary1 to ternary3 the template strand forms contacts with residues from the palm and fingers and with the incoming nucleotide (Movie 5, Figures 4.19 and 4.20); alternating breakage and re-formation of the contacts with the thumb and residues in the vicinity of the active site leads to the DNA slippage.

Unlike other A-family polymerases, the thumb of apPol makes few contacts with the DNA, especially in the binary state, and this might explain the relatively weak affinity of DNA for apPol [179, 216]. For instance, Bst Pol I and T7 DNA polymerase binary complexes have polymerase–DNA interfaces of 1782 Å² (PDB id: 1L3T) and 2009 Å² (PDB id: 2AJQ), respectively, while the corresponding interface for apPol is only 607 Å². For the apPol-specific pre-insertion checkpoint to function, the DNA and the incoming nucleotide need to

undergo large-scale motions as apPol transitions from the pre-chemistry binary complex to the catalytically poised complex. The weaker contact between the apPol thumb and the DNA might have evolved to facilitate the large DNA motion enabling this apPol-specific checkpoint.

5.2.4 Why do the fingers close?

In most catalytically poised DNA polymerase ternary complexes, including apPol, the fingers domain stays in a closed conformation [181] [186]. On the other hand, in the apo or binary states, the fingers are typically open. Closed fingers restrict the polymerization active site and sequester the 3'OH of the primer and incoming nucleotide from the bulk solvent, facilitating the phosphodiester bond formation. Nucleotide binding has typically been assigned as the trigger for fingers closure [75]. However, it has been shown that the Klenow fragment of *E. coli* Pol I and the herpes simplex virus-1 DNA polymerase can sample both the open and closed conformations in the absence of the incoming nucleotide [180, 212, 217]. This indicates that dNTP binding is not essential for fingers closing. It is possible that the closed state is stabilized in the presence of the correct incoming nucleotide and/or the two divalent metal ions associated with the nucleotide [180, 218, 219]. However, recent structures of the human leading strand DNA polymerase, pol ϵ , show that the pre-chemistry ternary complex with a correct nascent base pairing can exist with the fingers in open, closed, and intermediate states [54], indicating that even in the presence of the correct nucleotide, the fingers may stay open. Similarly, in case of apPol, the fingers can be in open (ternary1 and 2), partially closed (9CAY) and fully closed (ternary3) in presence of the correct incoming dNTP. Interestingly, it has recently been observed that unlike high fidelity A-family polymerases, a specialised A-family DNA polymerase involved in double strand break repair and translesion synthesis, Pol θ , can have its fingers in a fully closed conformation in

presence of both correct and incorrect nucleotides, indicating that closure is not restricted to the cognate pairing [220]. However, it is possible that for high fidelity replicative polymerases, the fingers do not close in the presence of the incorrect incoming nucleotide. Taken together, these new observations raise the question: what triggers fingers closing? A combination of kinetic and time-resolved structural measurements investigating the short-lived pre-chemistry states of a DNA polymerase would be essential to answer this question, which is central to our understanding of the catalytic mechanisms of DNA polymerases.

5.2.5 What happens after bond formation?

The series of structures presented here elucidate how apPol interacts with DNA and the incoming nucleotide to catalyse phosphodiester bond formation; however, we do not have direct structural insight into the post-chemistry events. While a post-chemistry binary complex (Figure 1.4 Step 4) structure of apPol (generated through addition of a dideoxyATP to the 3' end of the primer strand) has been reported [179], the modest resolution of 5.2 Å precludes drawing any molecular-level inferences. In fact, at this resolution, it is difficult to ascertain whether the structure is of a post-chemistry binary state or whether apPol has translocated (Figure 1.4, Step 5) to a pre-chemistry binary form.

Nonetheless, based on the DNA position in binary1 and the published post-chemistry binary structure, we predict that, after chemistry apPol over-translocates, leading to the apPol–DNA pre-chemistry binary complex (binary1). Following this apPol backtracks to correct for the over-translocation and positions the templating DNA base in the active site (Figure 5.2).

Among the well-studied replication systems, the T7 bacteriophage replisome is perhaps the closest to the apicoplast replisome [90]. *In vitro* studies indicate that in T7 replisome the leading strand gp5 T7 DNA polymerase follows the gp4 helicase-primase closely with only a

single nucleotide of the DNA template separating the two proteins [221] [5]. However, the over-translocation mechanism for apPol would require a larger slack in the DNA between apPol and the apicoplast helicase or any other partner protein positioned downstream of the polymerase. Alternatively, the over-translocation might be used by apPol only under special situations like lesion bypass, and during bulk DNA synthesis, a close contact with a partner protein like the helicase might prevent apPol from over-translocating. Understanding the molecular mechanism of apPol's post-chemistry substrate reorientation and its implications for the apicoplast replisome organization would be vital for developing a clear picture of how the apicoplast genome is copied.

5.3 Conclusions

In summary, the kinetic and structural analyses presented in this dissertation establish apPol as a multitasking A-family DNA polymerase that combines intrinsically low-efficiency, low-processivity and error-prone TLS activity with unique fidelity checkpoints encoded in a highly dynamic pre-chemistry conformational cycle. Transient-state kinetics shows that apPol copies undamaged DNA with modest efficiency and weak DNA and dNTP affinities hinting at the involvement of additional replisome proteins for an efficient and highly processive DNA synthesis inside apicoplast. Our kinetic studies also show that apPol bypasses oxidative lesions such as 8-oxo-dG and abasic sites in a distributive manner, with low efficiency and fidelity, suggesting that this might be a source of mutation within the apicoplast genome, and repair pathways might play a critical role in the faithful propagation of the apicoplast DNA. The five cryo-EM structures of apPol's pre-chemistry conformational states reveal a multistep mechanism that might help the polymerase perform TLS and replicative synthesis. The nascent base pairing in the fingers-open configuration, which, among the A-family polymerases is unique to apPol, can only be possible because a) the nascent base pair forms

far away from the active site, which provides wide space mimicking the open active site of a TLS polymerase and b) after base pair formation, nascent base-pair together with the primer-template DNA move towards the active site. This orchestrated mechanism might help apPol perform translesion synthesis.

These structures also reveal that the apPol-specific O1 helix tyrosine triad formed by Y481, Y485 and Y486 act as a novel checkpoint in the fingers-open state. This checkpoint in combination with the restrictive active site resulting from the fingers closed state provide sequential checks that bias the nascent base pair towards Watson-Crick geometry.

These findings define a mechanistic framework in which extensive DNA ‘backtracking’ and large thumb-driven DNA movements, combined with weak DNA binding, might allow apPol to shuttle between replication and lesion-bypass synthesis, providing a plausible solution for how a single organellar polymerase supports both replication and damage tolerance in the oxidative environment of the apicoplast.

5.4 Future Perspectives

While the work on apPol presented in this thesis forms a starting point for understanding how apicoplast replication functions, we are far away from developing a comprehensive mechanistic insight into the workings of the apicoplast replisome. Future work should reconstitute larger sub-complexes of the apicoplast replisome to test how the apicoplast replication proteins like primase, helicase, SSB, putative processivity factors, and 5’-3’ exonuclease interact with and modulate apPol’s processivity, efficiency, nucleotide selectivity, and lesion-bypass kinetics on long, A/T-rich DNA templates that more closely mimic the apicoplast genome. Future structural work focusing on the conformational landscape of apPol during TLS will be essential for a detailed mechanistic understanding of

apPol's lesion bypass activities thereby completing the picture of multitasking by this polymerase.

Most of the structural work on DNA polymerases, including the work presented here, relies on capturing conformational intermediates based on trapping the enzyme using substrate analogues or non-cognate divalent metal ions. While this approach has provided rich information on polymerase mechanism, the use of analogs or non-cognate metals prevent us from visualising the enzyme performing catalysis and thus the steps immediately before chemistry remain beyond our reach. Moreover, when we tease out multiple conformational states from the trapped polymerase datasets we do not have any experimental evidence regarding the chronological order of these states. Taken together, these limitations have prevented a complete structural description of the catalytic cycle of a DNA polymerase.

Time-resolved cryo-EM is emerging as a powerful approach for visualizing transient catalytic intermediates and linking structural rearrangements to reaction kinetics, thereby providing direct insight into the molecular basis of nucleotide incorporation. This can be used for apPol with defined oxidative lesions to map the conformational landscape of apPol during TLS, elucidating how apPol-specific structural features like the pre-insertion tyrosine checkpoints and thumb-mediated DNA movements are altered when the enzyme engages with damaged template bases.

Like most replicative DNA polymerases, apPol contains a proofreading exonuclease domain that helps correct rare misincorporation events, thereby enhancing replication fidelity.

However, the work in this thesis focused primarily on the polymerization domain, while providing only limited insight into the mechanism of apPol's proofreading. Future work should therefore combine kinetic and structural approaches to determine how apPol recognizes a mismatched primer terminus, unwinds the DNA and transfers the primer strand

from the polymerase site to the exonuclease site, excises the error, and then returns the primer for continued DNA synthesis.

DNA polymerase fidelity is often reduced in repetitive sequences because these regions are especially susceptible to strand slippage, leading to increased insertion or deletion errors [222]. The risk of errors generally increases with repeat length and can be further influenced by the intrinsic accuracy of the polymerase and its proofreading activity. Thus, it is essential to look at the accuracy of synthesis within the context of the DNA sequence. Traditionally this has been assessed using a LacZ synthesis assay. However, such experiments can underestimate misincorporations by orders of magnitude. Recent advances in next-generation sequencing has overcome this barrier and in future, the fidelity of apPol should be assessed within the sequence context of A/T-rich DNA using long-read next-generation sequencing [223].

BIBLIOGRAPHY

1. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737–738.
2. Wilkins, M.H., A.R. Stokes, and H.R. Wilson, *Molecular structure of deoxypentose nucleic acids*. Nature, 1953. **171**(4356): p. 738–40.
3. Franklin, R.E. and R.G. Gosling, *Molecular Configuration in Sodium Thymonucleate*. Nature, 1953. **171**(4356): p. 740–741.
4. Meselson, M. and F.W. Stahl, *The replication of DNA in Escherichia coli*. Proceedings of the National Academy of Sciences, 1958. **44**(7): p. 671–682.
5. Gao, Y., et al., *Structures and operating principles of the replisome*. Science, 2019. **363**(6429).
6. Kulczyk, A.W. and C.C. Richardson, *The Replication System of Bacteriophage T7*. Enzymes, 2016. **39**: p. 89–136.
7. Chastain, P.D., 2nd, et al., *Architecture of the replication complex and DNA loops at the fork generated by the bacteriophage t4 proteins*. J Biol Chem, 2003. **278**(23): p. 21276–85.
8. Raney, K.D., A.K. Byrd, and S. Aarattuthodiyil, *Structure and mechanisms of SF1 DNA helicases*. Advances in Experimental Medicine and Biology, 2013. **767**: p. 17–46.
9. Li, Y. and H. Araki, *Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication*. Genes To Cells, 2013. **18**(4): p. 266–277.
10. Lee, S.J., et al., *Zinc-binding domain of the bacteriophage T7 DNA primase modulates binding to the DNA template*. J Biol Chem, 2012. **287**(46): p. 39030–40.
11. Pan, H. and D.B. Wigley, *Structure of the zinc-binding domain of Bacillus stearothermophilus DNA primase*. Structure, 2000. **8**(3): p. 231–239.
12. Maffeo, C. and A. Aksimentiev, *Molecular mechanism of DNA association with single-stranded DNA binding protein*. Nucleic Acids Res, 2017. **45**(21): p. 12125–12139.
13. Nakamura, T., et al., *Mechanism of the nucleotidyl-transfer reaction in DNA polymerase revealed by time-resolved protein crystallography*. Biophysics (Nagoya-shi, Japan), 2013. **9**: p. 31–36.
14. Chang, C., C. Lee Luo, and Y. Gao, *In crystallo observation of three metal ion promoted DNA polymerase misincorporation*. Nature Communications, 2022. **13**(1): p. 2346.
15. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity*. Annu Rev Biochem, 2000. **69**: p. 497–529.
16. Schaaper, R.M., *Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli*. J Biol Chem, 1993. **268**(32): p. 23762–5.
17. Fijalkowska, I.J., R.M. Schaaper, and P. Jonczyk, *DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair*. FEMS Microbiol Rev, 2012. **36**(6): p. 1105–21.
18. Widlak, W., *DNA Replication, Mutations, and Repair*, in *Molecular Biology: Not Only for Bioinformaticians*, W. Widlak, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 49–69.

19. Lindahl, T., *Instability and decay of the primary structure of DNA*. Nature, 1993. **362**(6422): p. 709–715.
20. Sehn, J.K., *Chapter 9 - Insertions and Deletions (Indels)*, in *Clinical Genomics*, S. Kulkarni and J. Pfeifer, Editors. 2015, Academic Press: Boston. p. 129–150.
21. Viguera, E., D. Canceill, and S.D. Ehrlich, *Replication slippage involves DNA polymerase pausing and dissociation*. Embo j, 2001. **20**(10): p. 2587–95.
22. Reddy, E.P., et al., *A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene*. Nature, 1982. **300**(5888): p. 149–52.
23. Okazaki, R., et al., *Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains*. Proc Natl Acad Sci U S A, 1968. **59**(2): p. 598–605.
24. Burgers, P.M., *Solution to the 50-year-old Okazaki-fragment problem*. Proc Natl Acad Sci U S A, 2019. **116**(9): p. 3358–3360.
25. Kulczyk, A.W., et al., *Cryo-EM structure of the replisome reveals multiple interactions coordinating DNA synthesis*. Proceedings of the National Academy of Sciences, 2017. **114**(10): p. E1848–E1856.
26. Lee, J., et al., *Lagging strand synthesis in coordinated DNA synthesis by bacteriophage t7 replication proteins*. J Mol Biol, 2002. **316**(1): p. 19–34.
27. Lee, J., et al., *Coordinated Leading and Lagging Strand DNA Synthesis on a Minicircular Template*. Molecular Cell, 1998. **1**(7): p. 1001–1010.
28. Krishna, T.S., et al., *Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA*. Cell, 1994. **79**(7): p. 1233–1243.
29. Lovett, S.T., *Polymerase switching in DNA replication*. Mol Cell, 2007. **27**(4): p. 523–6.
30. Geertsema, H.J., et al., *Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery*. Proc Natl Acad Sci U S A, 2014. **111**(11): p. 4073–8.
31. McDonald, A.G. and K.F. Tipton, *Enzyme nomenclature and classification: the state of the art*. The FEBS Journal, 2023. **290**(9): p. 2214–2231.
32. Zahurancik, W.J., S.J. Klein, and Z. Suo, *Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase ϵ* . Biochemistry, 2013. **52**(40): p. 7041–7049.
33. Cortez, D., *Replication-Coupled DNA Repair*. Molecular Cell, 2019. **74**(5): p. 866–876.
34. Chatterjee, N. and G.C. Walker, *Mechanisms of DNA damage, repair, and mutagenesis*. Environmental and Molecular Mutagenesis, 2017. **58**(5): p. 235–263.
35. Roos, W.P., A.D. Thomas, and B. Kaina, *DNA damage and the balance between survival and death in cancer biology*. Nature Reviews. Cancer, 2016. **16**(1): p. 20–33.
36. Marians, K.J., *Lesion bypass and the reactivation of stalled replication forks*. Annual Review of Biochemistry, 2018. **87**: p. 217–238.
37. Wilson, R.C., M.A. Jackson, and J.D. Pata, *γ -family polymerase conformation is a major determinant of fidelity and translesion specificity*. Structure, 2013. **21**(1): p. 20–31.
38. Kokoska, R.J., et al., *Low fidelity DNA synthesis by a γ family DNA polymerase due to misalignment in the active site*. The Journal of Biological Chemistry, 2002. **277**(22): p. 19633–19638.

39. Boudsocq, F., et al., *Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication*. The Journal of Biological Chemistry, 2004. **279**(31): p. 32932–32940.
40. Ohashi, E., et al., *Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene*. The Journal of Biological Chemistry, 2000. **275**(50): p. 39678–39684.
41. DeLucia, A.M., N.D. Grindley, and C.M. Joyce, *Conformational changes during normal and error-prone incorporation of nucleotides by a Y-family DNA polymerase detected by 2-aminopurine fluorescence*. Biochemistry, 2007. **46**(38): p. 10790–803.
42. Cirz, R.T., et al., *Inhibition of Mutation and Combating the Evolution of Antibiotic Resistance*. PLOS Biology, 2005. **3**(6): p. e176.
43. Fujii, S. and R.P. Fuchs, *A Comprehensive View of Translesion Synthesis in Escherichia coli*. Microbiol Mol Biol Rev, 2020. **84**(3).
44. Bailey, S., R.A. Wing, and T.A. Steitz, *The Structure of T. aquaticus DNA Polymerase III Is Distinct from Eukaryotic Replicative DNA Polymerases*. Cell, 2006. **126**(5): p. 893–904.
45. Lamers, M.H. and M. O'Donnell, *A consensus view of DNA binding by the C family of replicative DNA polymerases*. Proceedings of the National Academy of Sciences, 2008. **105**(52): p. 20565–20566.
46. Ruprich-Robert, G. and P. Thuriaux, *Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases*. Nucleic Acids Res, 2010. **38**(14): p. 4559–69.
47. Sauguet, L., et al., *Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography*. Nature Communications, 2016. **7**(1): p. 12227.
48. Raia, P., et al., *Structure of the DP1-DP2 PolD complex bound with DNA and its implications for the evolutionary history of DNA and RNA polymerases*. PLoS Biol, 2019. **17**(1): p. e3000122.
49. Ollis, D.L., et al., *Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP*. Nature, 1985. **313**(6005): p. 762–766.
50. Baranovskiy, A.G., et al., *Activity and fidelity of human DNA polymerase α depend on primer structure*. The Journal of Biological Chemistry, 2018. **293**(18): p. 6824–6843.
51. Swan, M.K., et al., *Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta*. Nat Struct Mol Biol, 2009. **16**(9): p. 979–86.
52. Dodd, T., et al., *Polymerization and editing modes of a high-fidelity DNA polymerase are linked by a well-defined path*. Nature Communications, 2020. **11**(1): p. 5379.
53. Ganai, R.A., G.O. Bylund, and E. Johansson, *Switching between polymerase and exonuclease sites in DNA polymerase ϵ* . Nucleic Acids Research, 2015. **43**(2): p. 932–942.
54. Roske, J.J. and J.T.P. Yeeles, *Structural basis for processive daughter-strand synthesis and proofreading by the human leading-strand DNA polymerase Pol ϵ* . Nat Struct Mol Biol, 2024. **31**(12): p. 1921–1931.
55. Lehman, I.R., et al., *Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli*. The Journal of Biological Chemistry, 1958. **233**(1): p. 163–170.

56. Kuchta, R.D., et al., *Kinetic mechanism of DNA polymerase I (Klenow)*. Biochemistry, 1987. **26**(25): p. 8410–8417.
57. Joyce, C.M. and S.J. Benkovic, *DNA polymerase fidelity: kinetics, structure, and checkpoints*. Biochemistry, 2004. **43**(45): p. 14317–14324.
58. Gong, S., et al., *Kinetic and thermodynamic analysis defines roles for two metal ions in DNA polymerase specificity and catalysis*. J Biol Chem, 2021. **296**: p. 100184.
59. Pata, J.D., Y.W. Yin, and I. Lahiri, *Editorial: Nucleic Acid Polymerases: The Two-Metal-Ion Mechanism and Beyond*. Front Mol Biosci, 2022. **9**: p. 948326.
60. Beese, L.S. and T.A. Steitz, *Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism*. Embo j, 1991. **10**(1): p. 25–33.
61. Xia, S., et al., *Structural Insights into Complete Metal Ion Coordination from Ternary Complexes of B Family RB69 DNA Polymerase*. Biochemistry, 2011. **50**(42): p. 9114–9124.
62. Yang, W., J.Y. Lee, and M. Nowotny, *Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity*. Mol Cell, 2006. **22**(1): p. 5–13.
63. Tsai, M.D., *Catalytic mechanism of DNA polymerases-Two metal ions or three?* Protein Sci, 2019. **28**(2): p. 288–291.
64. Gao, Y. and W. Yang, *Capture of a third Mg²⁺ is essential for catalyzing DNA synthesis*. Science, 2016. **352**(6291): p. 1334–7.
65. Nakamura, T., et al., *Watching DNA polymerase η make a phosphodiester bond*. Nature, 2012. **487**(7406): p. 196–201.
66. Garcia-Diaz, M. and K. Bebenek, *Multiple functions of DNA polymerases*. CRC Crit Rev Plant Sci, 2007. **26**(2): p. 105–122.
67. Maiorano, D., et al., *Translesion Synthesis or Repair by Specialized DNA Polymerases Limits Excessive Genomic Instability upon Replication Stress*. Int J Mol Sci, 2021. **22**(8).
68. Zhao, L. and M.T. Washington, *Translesion Synthesis: Insights into the Selection and Switching of DNA Polymerases*. Genes (Basel), 2017. **8**(1).
69. Lehmann, A.R., et al., *Translesion synthesis: Y-family polymerases and the polymerase switch*. DNA Repair (Amst), 2007. **6**(7): p. 891–9.
70. Millar, D.P., *Conformational Dynamics of DNA Polymerases Revealed at the Single-Molecule Level*. Front Mol Biosci, 2022. **9**: p. 826593.
71. Xu, C., et al., *Global conformational dynamics of a Y-family DNA polymerase during catalysis*. PLoS Biol, 2009. **7**(10): p. e1000225.
72. Tsai, Y.-C. and K.A. Johnson, *A New Paradigm for DNA Polymerase Specificity*. Biochemistry, 2006. **45**(32): p. 9675–9687.
73. Meli, M., et al., *DNA Polymerase Conformational Dynamics and the Role of Fidelity-Confering Residues: Insights from Computational Simulations*. Front Mol Biosci, 2016. **3**: p. 20.
74. Evans, G.W., et al., *Real-time single-molecule studies of the motions of DNA polymerase fingers illuminate DNA synthesis mechanisms*. Nucleic Acids Res, 2015. **43**(12): p. 5998–6008.
75. Johnson, K.A., *The kinetic and chemical mechanism of high-fidelity DNA polymerases*. Biochim Biophys Acta, 2010. **1804**(5): p. 1041–8.
76. Wilson, R.C. and J.D. Pata, *Structural insights into the generation of single-base deletions by the Y family DNA polymerase dbh*. Mol Cell, 2008. **29**(6): p. 767–79.

77. Freudenthal, B.D., et al., *Observing a DNA polymerase choose right from wrong*. Cell, 2013. **154**(1): p. 157–68.
78. Dangerfield, T.L. and K.A. Johnson, *Conformational dynamics during high-fidelity DNA replication and translocation defined using a DNA polymerase with a fluorescent artificial amino acid*. J Biol Chem, 2021. **296**: p. 100143.
79. Venkatesan, P., *WHO world malaria report 2024*. Lancet Microbe, 2025. **6**(4).
80. Zekar, L. and T. Sharman, *Plasmodium falciparum Malaria*, in StatPearls. 2025, StatPearls Publishing: Treasure Island (FL).
81. Morrison, D.A., *Evolution of the Apicomplexa: where are we now?* Trends Parasitol, 2009. **25**(8): p. 375–82.
82. McFadden, G.I. and E. Yeh, *The apicoplast: now you see it, now you don't*. Int J Parasitol, 2017. **47**(2-3): p. 137–144.
83. Elahi, R. and S.T. Prigge, *New insights into apicoplast metabolism in blood-stage malaria parasites*. Curr Opin Microbiol, 2023. **71**: p. 102255.
84. Okada, M., et al., *Critical role for isoprenoids in apicoplast biogenesis by malaria parasites*. Elife, 2022. **11**.
85. Lemgruber, L., et al., *Cryo-electron tomography reveals four-membrane architecture of the Plasmodium apicoplast*. Malar J, 2013. **12**: p. 25.
86. Kennedy, S.R., et al., *The biochemistry and fidelity of synthesis by the apicoplast genome replication DNA polymerase Pfpex from the malaria parasite Plasmodium falciparum*. J Mol Biol, 2011. **410**(1): p. 27–38.
87. Sato, S., *The apicomplexan plastid and its evolution*. Cell Mol Life Sci, 2011. **68**(8): p. 1285–96.
88. Williamson, D.H., et al., *The plastid DNA of the malaria parasite Plasmodium falciparum is replicated by two mechanisms*. Mol Microbiol, 2002. **45**(2): p. 533–42.
89. Mukherjee, A. and G.C. Sadhukhan, *Anti-malarial Drug Design by Targeting Apicoplasts: New Perspectives*. J Pharmacopuncture, 2016. **19**(1): p. 7–15.
90. Milton, M.E. and S.W. Nelson, *Replication and maintenance of the Plasmodium falciparum apicoplast genome*. Mol Biochem Parasitol, 2016. **208**(2): p. 56–64.
91. Seow, F., et al., *The plastidic DNA replication enzyme complex of Plasmodium falciparum*. Mol Biochem Parasitol, 2005. **141**(2): p. 145–153.
92. Bhowmick, K. and S.K. Dhar, *Plasmodium falciparum single-stranded DNA-binding protein (PfSSB) interacts with PfPrex helicase and modulates its activity*. FEMS Microbiology Letters, 2014. **351**(1): p. 78–87.
93. Wingert, B.M., E.E. Parrott, and S.W. Nelson, *Fidelity, Mismatch Extension, and Proofreading Activity of the Plasmodium falciparum Apicoplast DNA Polymerase*. Biochemistry, 2013. **52**(44): p. 7723–7730.
94. Schoenfeld, T.W., et al., *Lateral Gene Transfer of Family A DNA Polymerases between Thermophilic Viruses, Aquificae, and Apicomplexa*. Molecular Biology and Evolution, 2013. **30**(7): p. 1653–1664.
95. Chheda, P.R., et al., *Promising antimalarials targeting apicoplast DNA polymerase from Plasmodium falciparum*. Eur J Med Chem, 2022. **243**: p. 114751.
96. Milton, M.E., et al., *Crystal Structure of the Apicoplast DNA Polymerase from Plasmodium falciparum: The First Look at a Plastidic A-Family DNA Polymerase*. Journal of Molecular Biology, 2016. **428**(20): p. 3920–3934.

97. Sharma, M., N. Narayanan, and D.T. Nair, *The proofreading activity of Pfp_{prex} from Plasmodium falciparum can prevent mutagenesis of the apicoplast genome by oxidized nucleotides*. Scientific Reports, 2020. **10**(1): p. 11157.
98. Sharma, M. and D.T. Nair, *Pfp_{prex} from Plasmodium falciparum can bypass oxidative stress-induced DNA lesions*. The FEBS Journal, 2022. **289**(17): p. 5218–5240.
99. O'Flaherty, D.K. and F.P. Guengerich, *Steady-state kinetic analysis of DNA polymerase single-nucleotide incorporation products*. Curr Protoc Nucleic Acid Chem, 2014. **59**: p. 7.21.1–13.
100. Li, Q.-S., et al., *Kinetic assays of DNA polymerase fidelity: A theoretical perspective beyond Michaelis-Menten kinetics*. Physical Review E, 2021. **104**(1): p. 014408.
101. Fijen, C., et al., *Using single-molecule FRET to probe the nucleotide-dependent conformational landscape of polymerase β -DNA complexes*. J Biol Chem, 2020. **295**(27): p. 9012–9020.
102. Betancurt-Anzola, L., et al., *The A, B, C, D's of replicative DNA polymerase fidelity: utilizing high-throughput single-molecule sequencing to understand the molecular basis for DNA polymerase accuracy*. Nucleic Acids Res, 2025. **53**(21).
103. Joyce, C.M., *Techniques used to study the DNA polymerase reaction pathway*. Biochim Biophys Acta, 2010. **1804**(5): p. 1032–40.
104. Johnson, K.A., *Conformational coupling in DNA polymerase information transfer*. Philos Trans R Soc Lond B Biol Sci, 1992. **336**(1276): p. 107–12.
105. Bertram, J.G., et al., *DNA polymerase fidelity: comparing direct competition of right and wrong dNTP substrates with steady state and pre-steady state kinetics*. Biochemistry, 2010. **49**(1): p. 20–8.
106. Dieckman, L.M., et al., *Pre-steady state kinetic studies of the fidelity of nucleotide incorporation by yeast DNA polymerase delta*. Biochemistry, 2010. **49**(34): p. 7344–50.
107. Johnson, K.A., *1 Transient-State Kinetic Analysis of Enzyme Reaction Pathways*. The Enzymes, 1992. **20**: p. 1–61.
108. Renders, M., et al., *Full Pre-Steady-State Kinetic Analysis of Single Nucleotide Incorporation by DNA Polymerases*. Curr Protoc Nucleic Acid Chem, 2019. **78**(1): p. e98.
109. Cheng, Y., et al., *A primer to single-particle cryo-electron microscopy*. Cell, 2015. **161**(3): p. 438–449.
110. Thompson, R.F., et al., *An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology*. Methods, 2016. **100**: p. 3–15.
111. Cheng, Y., *Single-particle cryo-EM—How did it get here and where will it go*. Science, 2018. **361**(6405): p. 876–880.
112. Nakane, T., et al., *Single-particle cryo-EM at atomic resolution*. Nature, 2020. **587**(7832): p. 152–156.
113. Yip, K.M., et al., *Atomic-resolution protein structure determination by cryo-EM*. Nature, 2020. **587**(7832): p. 157–161.
114. Vulović, M., et al., *Image formation modeling in cryo-electron microscopy*. Journal of Structural Biology, 2013. **183**(1): p. 19–32.
115. Dickerson, J.L., et al., *Imaging biological macromolecules in thick specimens: The role of inelastic scattering in cryoEM*. Ultramicroscopy, 2022. **237**: p. 113510.

116. Frank, J., *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. 2006: Oxford University Press.
117. Glaeser, R.M., et al., *Electron Crystallography Of Biological Macromolecules*. 2007: Oxford University Press.
118. Brilot, A.F., et al., *Beam-induced motion of vitrified specimen on holey carbon film*. Journal of Structural Biology, 2012. **177**(3): p. 630–637.
119. Grant, T. and N. Grigorieff, *Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6*. eLife, 2015. **4**: p. e06980.
120. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs*. J Struct Biol, 2015. **192**(2): p. 216–21.
121. Johnson, K.A., Z.B. Simpson, and T. Blom, *Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data*. Analytical Biochemistry, 2009. **387**(1): p. 20–29.
122. Johnson, K.A., Z.B. Simpson, and T. Blom, *FitSpace Explorer: An algorithm to evaluate multidimensional parameter space in fitting kinetic data*. Analytical Biochemistry, 2009. **387**(1): p. 30–41.
123. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination*. Nat Methods, 2017. **14**(3): p. 290–296.
124. Bepler, T., et al., *Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs*. Nat Methods, 2019. **16**(11): p. 1153–1160.
125. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and developers*. Protein Sci, 2021. **30**(1): p. 70–82.
126. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486–501.
127. Croll, T.I., *ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps*. Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 519–530.
128. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix*. Acta Crystallographica Section D, 2019. **75**(10): p. 861–877.
129. Sievers, F. and D.G. Higgins, *Clustal Omega for making accurate alignments of many protein sequences*. Protein Sci, 2018. **27**(1): p. 135–145.
130. Waterhouse, A.M., et al., *Jalview Version 2—a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189–1191.
131. Johnson, K.A., *Chapter 23 Fitting Enzyme Kinetic Data with KinTek Global Kinetic Explorer*, in *Methods in Enzymology*, M.L. Johnson and L. Brand, Editors. 2009, Academic Press. p. 601–626.
132. Northrup, S.H. and H.P. Erickson, *Kinetics of protein-protein association explained by Brownian dynamics computer simulation*. Proceedings of the National Academy of Sciences, 1992. **89**(8): p. 3338–3342.
133. Rubinstein, J.L. and M.A. Brubaker, *Alignment of cryo-EM movies of individual particles by optimization of image translations*. J Struct Biol, 2015. **192**(2): p. 188–95.
134. Punjani, A., H. Zhang, and D.J. Fleet, *Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction*. Nature Methods, 2020. **17**(12): p. 1214–1221.

135. Scheres, S.H., *Processing of Structurally Heterogeneous Cryo-EM Data in RELION*. Methods Enzymol, 2016. **579**: p. 125–57.
136. Henderson, R., et al., *Outcome of the first electron microscopy validation task force meeting*. Structure, 2012. **20**(2): p. 205–14.
137. Scheres, S.H. and S. Chen, *Prevention of overfitting in cryo-EM structure determination*. Nat Methods, 2012. **9**(9): p. 853–4.
138. Rosenthal, P.B. and R. Henderson, *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy*. Journal of Molecular Biology, 2003. **333**(4): p. 721–745.
139. Sanchez-Garcia, R., et al., *DeepEMhancer: a deep learning solution for cryo-EM volume post-processing*. Communications Biology, 2021. **4**(1): p. 874.
140. Johnson, S.J., J.S. Taylor, and L.S. Beese, *Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations*. Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 3895–3900.
141. Meng, E.C., et al., *UCSF ChimeraX: Tools for structure building and analysis*. Protein Sci, 2023. **32**(11): p. e4792.
142. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126–32.
143. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography*. Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531–544.
144. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation*. Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074–80.
145. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation*. Protein Sci, 2018. **27**(1): p. 293–315.
146. Pintilie, G.D., et al., *Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions*. J Struct Biol, 2010. **170**(3): p. 427–38.
147. Madeira, F., et al., *The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024*. Nucleic Acids Res, 2024. **52**(W1): p. W521–w525.
148. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189–91.
149. Golosov, A.A., et al., *The mechanism of the translocation step in DNA replication by DNA polymerase I: a computer simulation analysis*. Structure, 2010. **18**(1): p. 83–93.
150. Berman, A.J., et al., *Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases*. Embo j, 2007. **26**(14): p. 3494–505.
151. Patel, S.S., I. Wong, and K.A. Johnson, *Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant*. Biochemistry, 1991. **30**(2): p. 511–25.
152. Kottur, J. and D.T. Nair, *Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction*. Nucleic Acids Res, 2018. **46**(12): p. 5875–5885.
153. Frevert, Z., et al., *Visualizing DNA polymerase I catalyze Hoogsteen-directed DNA synthesis*. Nat Commun, 2025. **16**(1): p. 5979.
154. Weaver, T.M., et al., *Visualizing Rev1 catalyze protein-template DNA synthesis*. Proc Natl Acad Sci U S A, 2020. **117**(41): p. 25494–25504.

155. Salas, M., et al., *The bacteriophage phi29 DNA polymerase*. IUBMB Life, 2008. **60**(1): p. 82–5.
156. Johnson, A.A., et al., *Human mitochondrial DNA polymerase holoenzyme: reconstitution and characterization*. Biochemistry, 2000. **39**(7): p. 1702–8.
157. Einolf, H.J. and F.P. Guengerich, *Kinetic analysis of nucleotide incorporation by mammalian DNA polymerase delta*. J Biol Chem, 2000. **275**(21): p. 16316–22.
158. Zahurancik, W.J. and Z. Suo, *Kinetic investigation of the polymerase and exonuclease activities of human DNA polymerase ϵ holoenzyme*. J Biol Chem, 2020. **295**(50): p. 17251–17264.
159. Bertram, J.G., et al., *Increased dNTP binding affinity reveals a nonprocessive role for Escherichia coli beta clamp with DNA polymerase IV*. J Biol Chem, 2004. **279**(32): p. 33047–50.
160. Werneburg, B.G., et al., *DNA polymerase beta: pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity*. Biochemistry, 1996. **35**(22): p. 7041–50.
161. Lieberman, K.R., et al., *Kinetic mechanism of translocation and dNTP binding in individual DNA polymerase complexes*. J Am Chem Soc, 2013. **135**(24): p. 9149–55.
162. Dahl, J.M., et al., *Direct observation of translocation in individual DNA polymerase complexes*. J Biol Chem, 2012. **287**(16): p. 13407–21.
163. Dangerfield, T.L. and K.A. Johnson, *Conformational dynamics during high-fidelity DNA replication and translocation defined using a DNA polymerase with a fluorescent artificial amino acid*. The Journal of Biological Chemistry, 2021. **296**: p. 100143.
164. Seeber, F. and D. Soldati-Favre, *Metabolic pathways in the apicoplast of apicomplexa*. Int Rev Cell Mol Biol, 2010. **281**: p. 161–228.
165. Cooke, M.S., et al., *Oxidative DNA damage: mechanisms, mutation, and disease*. Faseb j, 2003. **17**(10): p. 1195–214.
166. Strauss, B.S., *The 'A rule' of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions?* Bioessays, 1991. **13**(2): p. 79–84.
167. Liu, X., et al., *Bypass of an Abasic Site via the A-Rule by DNA Polymerase of Pseudomonas aeruginosa Phage PaP1*. Chem Res Toxicol, 2018. **31**(1): p. 58–65.
168. Obeid, S., et al., *Replication through an abasic DNA lesion: structural basis for adenine selectivity*. Embo j, 2010. **29**(10): p. 1738–47.
169. Kunkel, T.A. and A. Soni, *Mutagenesis by transient misalignment*. Journal of Biological Chemistry, 1988. **263**(29): p. 14784–14789.
170. Kumar, K., A.J. Fornace, Jr., and S. Suman, *8-Oxo-dG: A Potential Biomarker for Chronic Oxidative Stress Induced by High-LET Radiation*. DNA (Basel), 2024. **4**(3): p. 221–238.
171. McCulloch, S.D., et al., *The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases δ and η* . Nucleic Acids Research, 2009. **37**(9): p. 2830–2840.
172. Gannett, P.M. and T.P. Sura, *Base pairing of 8-oxoguanosine and 8-oxo-2'-deoxyguanosine with 2'-deoxyadenosine, 2'-deoxycytosine, 2'-deoxyguanosine, and thymidine*. Chem Res Toxicol, 1993. **6**(5): p. 690–700.
173. Furge, L.L. and F.P. Guengerich, *Analysis of nucleotide insertion and extension at 8-oxo-7,8-dihydroguanine by replicative T7 polymerase exo- and*

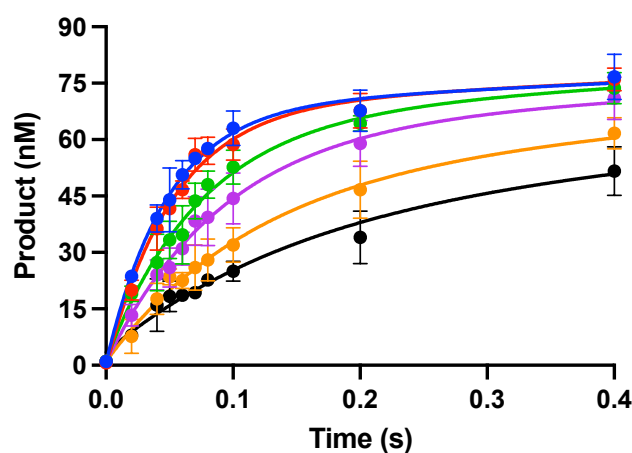
- human immunodeficiency virus-1 reverse transcriptase using steady-state and pre-steady-state kinetics*. Biochemistry, 1997. **36**(21): p. 6475–87.
174. Hanes, J.W., D.M. Thal, and K.A. Johnson, *Incorporation and replication of 8-oxo-deoxyguanosine by the human mitochondrial DNA polymerase*. J Biol Chem, 2006. **281**(47): p. 36241–8.
 175. Lowe, L.G. and F.P. Guengerich, *Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by Escherichia coli polymerases I exo- and II exo*. Biochemistry, 1996. **35**: p. 9840.
 176. Milton, M.E., et al., *Crystal Structure of the Apicoplast DNA Polymerase from Plasmodium falciparum: The First Look at a Plastidic A-Family DNA Polymerase*. J Mol Biol, 2016. **428**(20): p. 3920–3934.
 177. Xu, Y. and S. Dang, *Recent Technical Advances in Sample Preparation for Single-Particle Cryo-EM*. Front Mol Biosci, 2022. **9**: p. 892459.
 178. Doublie, S., et al., *Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution*. Nature, 1998. **391**(6664): p. 251–258.
 179. Lo, C.Y., et al., *Cryo-EM Structures of the Plasmodium falciparum Apicoplast DNA Polymerase*. J Mol Biol, 2024. **436**(23): p. 168842.
 180. Hohlbein, J., et al., *Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion*. Nat Commun, 2013. **4**: p. 2131.
 181. Li, Y., S. Korolev, and G. Waksman, *Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation*. Embo j, 1998. **17**(24): p. 7514–25.
 182. Brautigam, C.A. and T.A. Steitz, *Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes*. Curr Opin Struct Biol, 1998. **8**(1): p. 54–63.
 183. Park, J., et al., *Poly coordinates DNA synthesis and proofreading to ensure mitochondrial genome integrity*. Nat Struct Mol Biol, 2023. **30**(6): p. 812–823.
 184. Brieba, L.G., et al., *Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase*. Embo j, 2004. **23**(17): p. 3452–61.
 185. Chandani, S. and E.L. Loechler, *Y-Family DNA polymerases may use two different dNTP shapes for insertion: a hypothesis and its implications*. J Mol Graph Model, 2009. **27**(7): p. 759–69.
 186. Franklin, M.C., J. Wang, and T.A. Steitz, *Structure of the replicating complex of a pol alpha family DNA polymerase*. Cell, 2001. **105**(5): p. 657–67.
 187. Brown, J.A. and Z. Suo, *Unlocking the sugar "steric gate" of DNA polymerases*. Biochemistry, 2011. **50**(7): p. 1135–42.
 188. Astatke, M., et al., *A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides*. Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3402–7.
 189. Patel, P.H. and L.A. Loeb, *Multiple amino acid substitutions allow DNA polymerases to synthesize RNA*. J Biol Chem, 2000. **275**(51): p. 40266–72.
 190. Richardson, F.C., et al., *Polymerization of 2'-fluoro- and 2'-O-methyl-dNTPs by human DNA polymerase alpha, polymerase gamma, and primase*. Biochem Pharmacol, 2000. **59**(9): p. 1045–52.
 191. de Vega, M., et al., *Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs*. Proc Natl Acad Sci U S A, 2010. **107**(38): p. 16506–11.

192. Zhuang, Z. and Y. Ai, *Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis*. Biochim Biophys Acta, 2010. **1804**(5): p. 1081–93.
193. Fagan, S.P., et al., *Pyrophosphate release acts as a kinetic checkpoint during high-fidelity DNA replication by the Staphylococcus aureus replicative polymerase PolC*. Nucleic Acids Res, 2021. **49**(14): p. 8324–8338.
194. Percário, S., et al., *Oxidative stress in malaria*. Int J Mol Sci, 2012. **13**(12): p. 16346–72.
195. Egwu, C.O., et al., *Reactive Oxygen Species as the Brainbox in Malaria Treatment*. Antioxidants (Basel), 2021. **10**(12).
196. Halliwell, B. and O.I. Aruoma, *DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems*. FEBS Lett, 1991. **281**(1-2): p. 9–19.
197. Loeb, L.A. and B.D. Preston, *Mutagenesis by apurinic/apyrimidinic sites*. Annu Rev Genet, 1986. **20**: p. 201–30.
198. Kozhukhar, N., et al., *The efficiency of the translesion synthesis across abasic sites by mitochondrial DNA polymerase is low in mitochondria of 3T3 cells*. Mitochondrial DNA A DNA Mapp Seq Anal, 2016. **27**(6): p. 4390–4396.
199. Baruch-Torres, N. and L.G. Briebe, *Plant organellar DNA polymerases are replicative and translesion DNA synthesis polymerases*. Nucleic Acids Res, 2017. **45**(18): p. 10751–10763.
200. Fuchs, R.P. and S. Fujii, *Translesion DNA synthesis and mutagenesis in prokaryotes*. Cold Spring Harb Perspect Biol, 2013. **5**(12): p. a012682.
201. Rodriguez, G.P., J.B. Song, and G.F. Crouse, *In vivo bypass of 8-oxodG*. PLoS Genet, 2013. **9**(8): p. e1003682.
202. Hsu, G.W., et al., *Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase*. Nature, 2004. **431**(7005): p. 217–21.
203. Wu, Y., et al., *Heterotrimeric PCNA increases the activity and fidelity of Dbh, a Y-family translesion DNA polymerase prone to creating single-base deletion mutations*. DNA Repair (Amst), 2020. **96**: p. 102967.
204. Kunkel, T.A., S.S. Patel, and K.A. Johnson, *Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit*. Proc Natl Acad Sci U S A, 1994. **91**(15): p. 6830–4.
205. Haracska, L., et al., *Interaction with PCNA is essential for yeast DNA polymerase eta function*. Mol Cell, 2001. **8**(2): p. 407–15.
206. Paschalis, V., et al., *Interactions of the Bacillus subtilis DnaE polymerase with replisomal proteins modulate its activity and fidelity*. Open Biol, 2017. **7**(9).
207. Fujii, S. and R.P. Fuchs, *Biochemical basis for the essential genetic requirements of RecA and the beta-clamp in Pol V activation*. Proc Natl Acad Sci U S A, 2009. **106**(35): p. 14825–30.
208. Maga, G., et al., *8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins*. Nature, 2007. **447**(7144): p. 606–8.
209. Chim, N., et al., *Crystal structures of DNA polymerase I capture novel intermediates in the DNA synthesis pathway*. eLife, 2018. **7**: p. e40444.
210. Chim, N., et al., *Following replicative DNA synthesis by time-resolved X-ray crystallography*. Nature Communications, 2021. **12**(1): p. 2641.
211. Cheung, A.C. and P. Cramer, *Structural basis of RNA polymerase II backtracking, arrest and reactivation*. Nature, 2011. **471**(7337): p. 249–53.

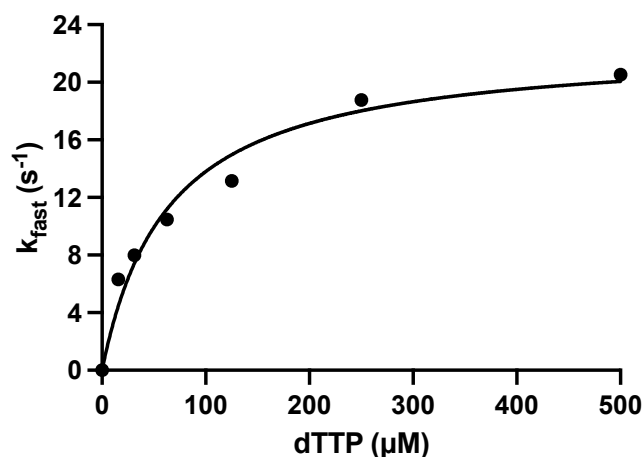
212. Botto, M.M., A. Borsellini, and M.H. Lamers, *A four-point molecular handover during Okazaki maturation*. *Nature Structural & Molecular Biology*, 2023. **30**(10): p. 1505–1515.
213. Lee, Y.S., W.D. Kennedy, and Y.W. Yin, *Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations*. *Cell*, 2009. **139**(2): p. 312–24.
214. Park, J., et al., *An interaction network in the polymerase active site is a prerequisite for Watson-Crick base pairing in Pol γ* . *Sci Adv*, 2024. **10**(21): p. eadl3214.
215. Park, J., et al., *An interaction network in the polymerase active site is a prerequisite for Watson-Crick base pairing in Pol γ* . *Science Advances*, 2024. **10**(21): p. eadl3214.
216. Kumari, A., A. Yadav, and I. Lahiri, *Transient State Kinetics of Plasmodium falciparum Apicoplast DNA Polymerase Suggests the Involvement of Accessory Factors for Efficient and Accurate DNA Synthesis*. *Biochemistry*, 2022. **61**(21): p. 2319–2333.
217. Shankar, S., et al., *Viral DNA polymerase structures reveal mechanisms of antiviral drug resistance*. *Cell*, 2024. **187**(20): p. 5572–5586.e15.
218. Freudenthal, B.D., W.A. Beard, and S.H. Wilson, *Structures of dNTP intermediate states during DNA polymerase active site assembly*. *Structure*, 2012. **20**(11): p. 1829–37.
219. Miller, B.R., 3rd, et al., *The Closing Mechanism of DNA Polymerase I at Atomic Resolution*. *Structure*, 2015. **23**(9): p. 1609–1620.
220. Li, C., L.M. Maksoud, and Y. Gao, *Structural basis of error-prone DNA synthesis by DNA polymerase θ* . *Nat Commun*, 2025. **16**(1): p. 2063.
221. Nandakumar, D., M. Pandey, and S.S. Patel, *Cooperative base pair melting by helicase and polymerase positioned one nucleotide from each other*. *Elife*, 2015. **4**.
222. Hite, J.M., K.A. Eckert, and K.C. Cheng, *Factors Affecting Fidelity of DNA Synthesis During PCR Amplification of d(C-A) $_n$ •d(G-T) $_n$ Microsatellite Repeats*. *Nucleic Acids Research*, 1996. **24**(12): p. 2429–2434.
223. Betancurt-Anzola, L., et al., *The A, B, C, D's of replicative DNA polymerase fidelity: utilizing high-throughput single-molecule sequencing to understand the molecular basis for DNA polymerase accuracy*. *Nucleic Acids Research*, 2025. **53**(21).

Appendix: A1

A

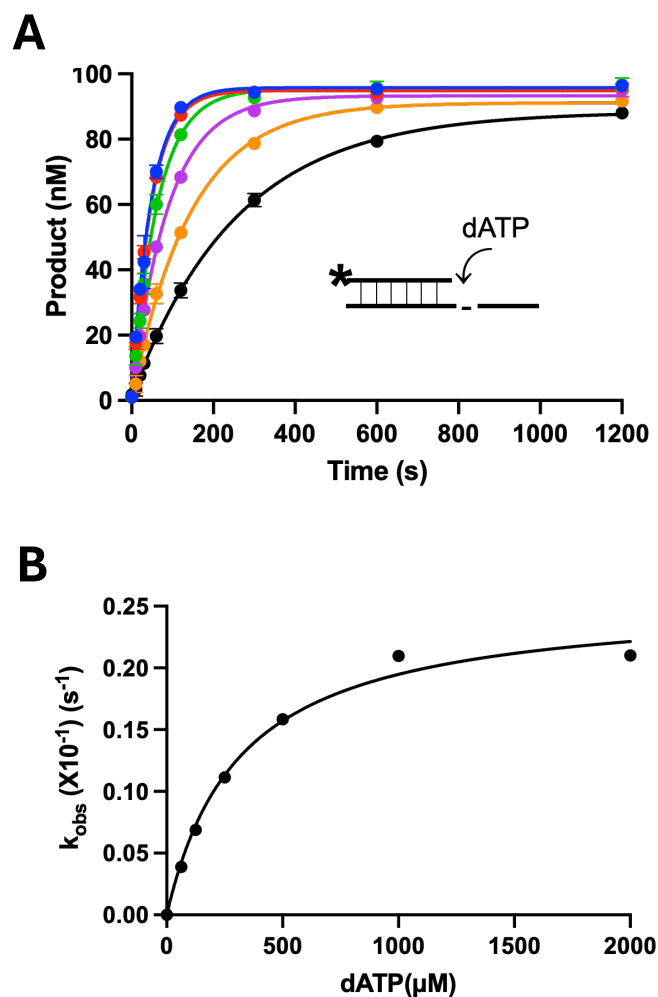


B



A1: Determination K_D^{dNTP} graph by non-linear regression. Single nucleotide primer extension assays for determining the affinity of dNTP for the pre-chemistry binary complex (K_D^{dNTP}) and the rate of bond formation. **(A)** Single nucleotide primer extension assays were setup, the final concentration of dTTP was varied from 15.6 to 500 μM (15.6 μM : black, 31.25 μM : orange, 62.5 μM : magenta, 125 μM : green, 250 μM : red, and 500 μM : blue), and the incubation time ranged from 0 to 0.4 s. The time courses of product formation were fit to full burst equation, and K_D^{dNTP} was 131 μM , while k_3 , the rate constant governing the chemistry step, was 30 s^{-1} . **(B)** The rates of the fast phase of the time courses from panel A (k_{fast}) plotted as a function of dTTP concentration. The data was fit to the hyperbolic equation and from the fit maximal rate of bond formation (k_{pol}) was calculated to be $22.6 \pm 1.6 \text{ s}^{-1}$ and the apparent dissociation constant for nucleotide binding ($K_{D,\text{app}}^{\text{dNTP}}$) was $64.2 \pm 14.3 \mu\text{M}$

Appendix: A2



A2: Analysis of abasic site bypass by apPol using nonlinear regression. (A) The time courses of dATP incorporation opposite an abasic site at varying dATP concentrations. Single nucleotide primer extension assay was performed as described in figure 3.13. The time courses were fit to a single exponential equation. The colour scheme is also similar to figure 3.13. Due to the sequence of the T_0 and T_1 nucleobases in substrate S3, we could detect two dATP addition events. The time courses show the combined product concentrations (i.e. the total of the +1 and +2 products). Experiments were performed in triplicate. Error bars represent standard deviation. (B) Rates of product formation (k_{obs}) determined from the time courses of panel A plotted as a function of dATP concentration. The data was fit to a hyperbolic equation and from there the maximal rate of nucleotide incorporation (k_{pol}) was determined to be $0.026 \pm 0.001 s^{-1}$ and the apparent K_D for dATP binding to the binary complex ($K_{D,app}^{dATP}$: abasic) was $316 \pm 44 \mu M$.

Appendix: A3

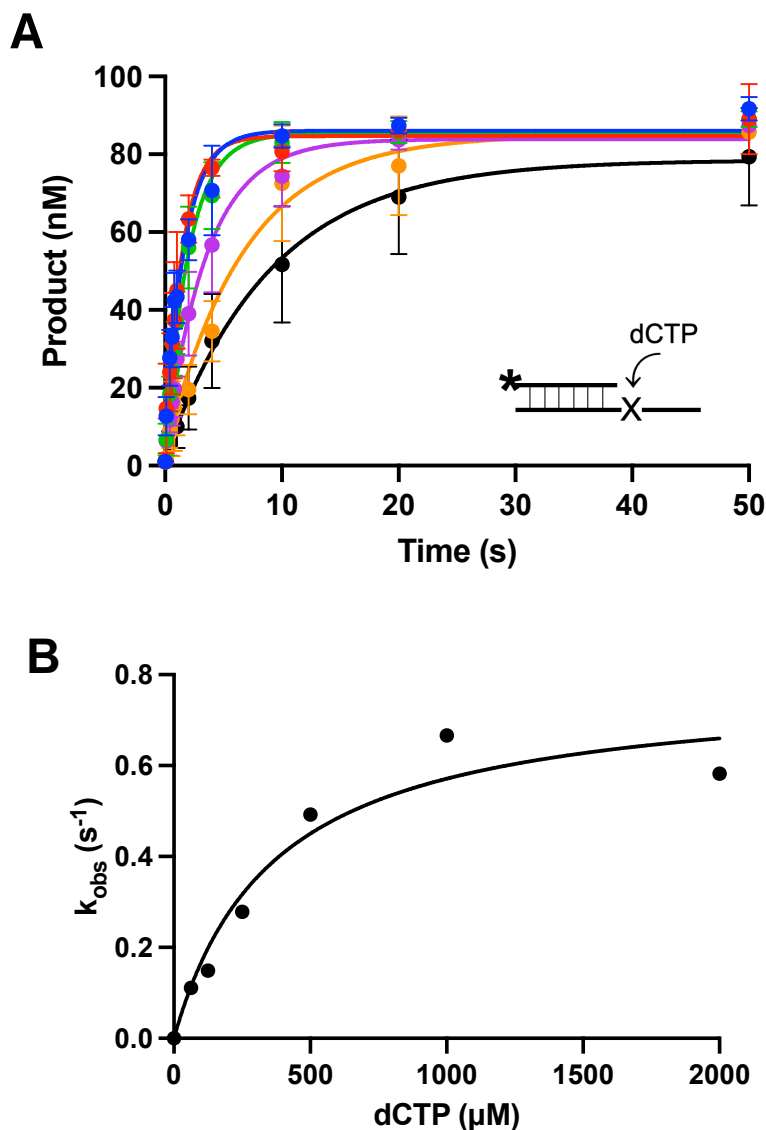


Figure A3: Determination $K_D^{dCTP:8-oxo-dG}$ by nonlinear regression. (A) The time courses from primer extension assays to determine the efficiency of dCTP incorporation by apPol opposite 8-oxo-dGMP. Single nucleotide primer extension assays were performed as described in figure 3.16. Colour scheme is also similar to figure 3.16. The time courses were fit to a single exponential equation. The experiment was performed in triplicate and mean of the three experiments are shown. The error bars represent the standard deviation of data sets. $\pm$: standard error (SE). Inset: Schematics of the reactants used for the respective primer extension assays. (B) Rates of product formation (k_{obs}) determined from the time courses of panel A plotted as a function of dCTP concentration. The data was fit to a hyperbolic

equation and from there the maximal rate of nucleotide incorporation (k_{pol}) was determined to be $0.78 \pm 0.09 \text{ s}^{-1}$ and the apparent K_D for dCTP binding to the binary complex ($K_{D,app}^{dCTP:8oxodG}$) was $367 \pm 130 \text{ } \mu\text{M}$.

Appendix: A4

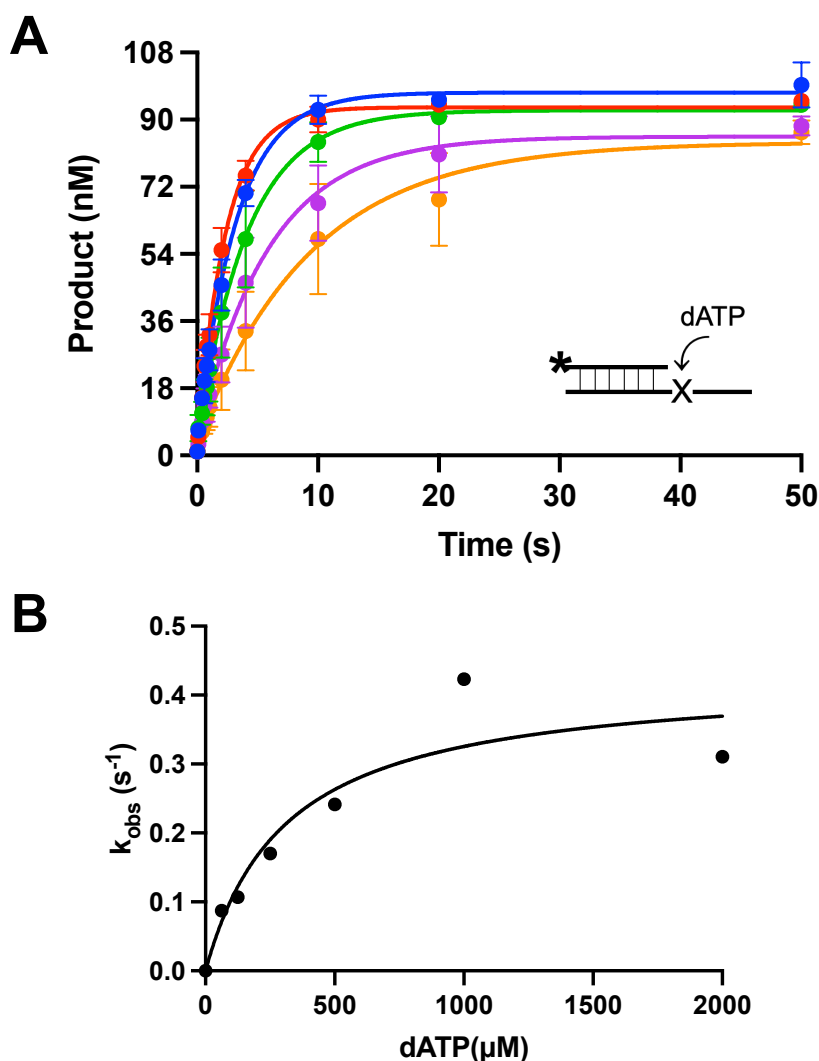


Figure A4: Determination $K_D^{dATP:8-oxo-dG}$ by nonlinear regression. (A) The time courses from primer extension assays to determine the efficiency of dATP incorporation by apPol opposite 8-oxo-dGMP. Single nucleotide primer extension assays were performed as described in Figure 3.16. Colour scheme is also similar to figure 3.16. The time courses were fit to a single exponential equation. The experiment was performed in triplicate and mean of

the three experiments are shown. The error bars represent the standard deviation of data sets. $\pm$: standard error (SE). Inset: Schematics of the reactants used for the respective primer extension assays. **(B)** Rates of product formation (k_{obs}) determined from the time courses of panel A plotted as a function of dATP concentration. The data was fit to a hyperbolic equation and from there the maximal rate of nucleotide incorporation (k_{pol}) was determined to be $0.43 \pm 0.07 \text{ s}^{-1}$ and the apparent K_D for dATP binding to the binary complex ($K_{D,\text{app}}^{\text{dATP:8oxodG}}$) was $312 \pm 160 \text{ }\mu\text{M}$.

Appendix: A5 - Movie description

Movie 1: Overall domain motion of apPol as it traverses the pre-chemistry steps. A morph between binary, ternary, ternary2 and ternary3 and back. The protein is in ribbon representation with the residues contacting the DNA strands in stick representation. In addition, Y481, 485 and 486 are shown in stick form. Colour coding is the same as 1B with dGTP shown in dark green and the P -3 position on the primer strand is highlighted in yellow.

Movie 2: Motion of the thumb as apPol traverses the pre-chemistry steps. A morph between binary1, ternary1, ternary2 and ternary3 and back, with a zoomed-in view of the thumb. Colour coding and residue representation are the same as Movie 1. Polar contacts between apPol and DNA are shown with black dashed lines.

Movie 3: Corkscrew motion of the DNA as apPol traverses the pre-chemistry steps. A morph between binary1, ternary1, ternary2 and ternary3 and back, with a zoomed-in view of the DNA duplex. The thumb domain has been omitted for clarity. Colour coding is the same as Movie 1. R390, D579 and D410 are shown in stick representation with colouring based on elements (oxygen: red, nitrogen: blue).

Movie 4: Formation and dissolution of the pre-insertion checkpoint. A morph between binary1, ternary1, ternary2 and ternary3 and back, with a zoomed-in view of the O1 helix. Colour coding and residue representation are the same as Movie 3. Polar contacts between apPol, dGTP and DNA are shown with black dashed lines.

Movie 5: Interaction of the template strand with the palm domain. A morph between binary1, ternary1, ternary2 and ternary3 and back, focusing on the interaction of the DNA template with the palm domain. Colour coding and residue representation are the same as Movie 1. Interactions between apPol, dGTP and DNA are shown with black dashed lines.

Movie 6: DNA slippage with respect to the thumb. A morph between binary1, ternary1, ternary2 and ternary3 and back, zooming in on the interactions of the DNA with the thumb helices H1, H2 and the loop connecting these two helices. Colour coding and residue representation are the same as Movie 1 with the following addition. Thumb residues predominantly interacting with the primer strand are colored based on the elements. Interactions between apPol, dGTP and DNA are shown with black dashed lines.