Spring 2021

Mechanistic insights into the roles and activities of polymerases in host and viral replication

Vincent Nghi Duong
Yale University Graduate School of Arts and Sciences, vincent.duong@yale.edu

Abstract

Mechanistic Insights Into the Roles and Activities of Polymerases in Host and Viral Replication

Vincent Nghi Lam Duong

2021

Polymerases are vital enzymes in the continuation of life, responsible for the replication of genetic material and the conversion of genetic information to necessary products. A large subset of these polymerases is dedicated to the high-fidelity replication and repair of DNA in the cell cycle of organisms. In addition, viruses utilize polymerases in order to produce DNA or RNA used to synthesize products for virion assembly. With such an important role, polymerases have been a focus in many therapeutic studies of cancer and antiviral treatments. This dissertation focuses on three different polymerases, PrimPol, human immunodeficiency virus (HIV) reverse transcriptase (RT), and DNA polymerase α (Polα). The goal of this work was to understand their overall mechanisms and roles not only in the context of replication and repair, but also in antiviral therapies.

HIV treatment, typically referred to as highly active antiretroviral therapy (HAART), consists of drugs that target various enzymes important for viral life cycle. A major fraction of these compounds, which target RT, can be classified into nucleoside (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs).

One prevailing issue with NRTIs is that administration of these drugs may cause off-target toxicity within patients, affecting adherence to treatment regimens. This off-target toxicity can be attributed to the incorporation of NRTIs by host polymerases, such as the mitochondrial polymerase γ (Polγ). To this end, I investigated the possibility of PrimPol, a recently characterized polymerase, in mediating the mitochondrial toxicity

effects seen in HIV+ patients taking tenofovir (TFV)-containing treatments. Using gel-based kinetic assays, I validated that the active metabolite form of tenofovir is a substrate for PrimPol. Cellular-based assays using overexpression and knockdown PrimPol renal cells suggests that PrimPol likely plays a protective role against tenofovir-induced toxicity through its repriming activity, despite the *in vitro* incorporation evidence. Given this potential role of PrimPol in TFV toxicity, I biochemically assessed a PrimPol active site mutant in an HIV+ patient taking TFV. The mutant appears to have drastically reduced polymerase activity and complete loss of priming activity, which may predispose this patient to TFV toxicity.

With NNRTIs, there are continuous development efforts to improve pharmacokinetic properties and combat drug resistance. To this end, a series of 2-naphthyl phenyl ether compounds were developed to target the Y181C mutation of RT. Interestingly, early structures of RT with these class of compounds showed two different binding modes that affected potency against the mutant. By solving structures of 2-naphthyl phenyl ether derivatives with WT and Y181C RT, we determined that the compounds that interact with W229 retain potency against the mutant. These studies will be important to consider in the development process of next generation NNRTIs.

Polα, in complex with Primase, is similar to PrimPol by possessing the ability to carry out *de novo* synthesis of nucleic acid primers. The primary role of the Polα-Pri complex in the primosome is to produce Okazaki fragments during DNA replication in a coordinated manner. Where primase initiates the primer with ribonucleotides, Polα continues the initial primer with deoxyribonucleotides. Interestingly, recent evidence shows that after replication mutations are left over from Polα, which is low-fidelity and

lacks a proofreading mechanism. To gain insight on Polα's activity during replication, we solved the structure of Polα with two replication-like substrates (Polα:dNTP:RNA/DNA or DNA/DNA) and kinetically characterized its activity with these substrates. We observed that with the RNA/DNA structure, a kink in n-4 sugar on the RNA primer correlated to a decrease in activity of the enzyme. Our kinetic characterization also revealed that with the DNA/DNA strand, Polα had increased incorporation efficiency but lower processivity. Our studies provide evidence of how different nucleotide substrates may regulate polymerase activity during replication.

Taken together, the studies of three different polymerases presented here provide a mechanistic and functional understanding of these polymerases in diseases and potential treatments. Ultimately, these findings will contribute to the development of therapies in diseases where polymerases play a vital role.

Mechanistic Insights Into the Roles and Activities of Polymerases in Host and Viral Replication

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Vincent Nghi Lam Duong

Dissertation Director: Karen S. Anderson, Ph.D.

June 2021

Table of Contents

Table of Figures

Table of Tables

Abbreviations

8-oxo-G	8-Oxoguanine
Å	aangstrom
AA354	$PrimPol_{1-354}$
ABC	abacavir
AEP	Archaeo-Eukaryotic Primase
AIDS	acquired immune deficiency syndrome
ANOVA	analysis of variance
AP	apurinic/apyrimidinic site
ART	antiretroviral therapy
AZT	azidothymidine
BER	base excision repair
Chk1	checkpoint kinase 1
CCDC111	coiled-coiled domain containing 111
cDNA	complimentary DNA
CPD	cyclobutane pyrimidine dimer
DHFR	dihydrofolate reductase
DMEM	Dulbecco's modified Eagle media
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dRP	5´-deoxyribose-5-phosphate
DSF	differential scanning fluorimetry
DTT	dithiothreitol
EC_{50}	efficacious concentration
EDTA	ethylenediaminetetraacetic acid
EFV	efavirenz
EMSA	electrophoretic mobility shift assay
FBS	fetal bovine serum

FDA	Food Drug Administration
FPLC	fast protein liquid chromatography
HAART	highly active antiretroviral therapy
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography
HU	hydroxyurea
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl β- d-1-thiogalactopyranoside
Kb	kilobase
kDa	kilodaltons
K_d	dissociation constant
k_{obs}	observed rate constant
k_{off}	off rate
k_{pol}	maximum rate of incorporation
k_{ss}	steady-state rate
M	molar
mM	millimolar
ms	milliseconds
mtDNA	mitochondrial DNA
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
NHEJ	non-homologous end joining
nM	nanomolar
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
NTP	ribonucleotide triphosphate
nucDNA	nuclear DNA

NVP	nevirapine
OCR	oxygen consumption rate
p180 core	human polymerase alpha catalytic domain
PBMC	peripheral blood mononuclear cell
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
Pol	polymerase
PP_i	pyrophosphate
Pri	primase
qPCR	real-time PCR
RNA	ribonucleic acid
RNase	ribonuclease
rNTP	ribonucleotide triphosphate
RPTEC	renal proximal tubule epithelial cells
RT	HIV-1 reverse transcriptase
s	seconds
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
shRNA	short hairpin RNA
TCEP	tris(2-carboxyethyl)phosphine
TDF	tenofovir disoproxil fumarate
TFV	tenofovir
TFV-DP	tenofovir diphosphate
TRAP1	tumor necrosis factor receptor associated protein 1
TLS	translesion synthesis
µCi	microcurie

µM	micromolar
UV	ultraviolet
WT	wild-type
ZnF	zinc finger

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Karen S. Anderson for her mentorship and leadership throughout the five years I have been at Yale for my graduate work. With her guidance, I can say without a doubt that I would not have developed into the scientist I am today. I truly appreciate her for respecting my need for independence during my work, allowing me to take control of my dissertation work and consulting her when I needed her advice and expertise. She has pushed me to become a better scientist, mentor, writer, and speaker both in and out of the lab. I am incredibly thankful to have her as my mentor.

I would like to thank my thesis committee: Dr. Elias Lolis, Dr. Elijah Paintsil, and Dr. Joann Sweasy (also as an outside reader). My meetings have always been productive and provided me insights into troubleshooting experiments and future directions of my project. I would also like to thank our research collaborators: Dr. Tahir Tahirov, Dr. Youri Pavlov, and Dr. Andrey Baranovskiy on the polymerase alpha project; Dr. Luis Blanco, Dr. Maria Martínez-Jiménez, and Dr. Lei Zhou on the PrimPol project. I have learned so much from your expertise and your contributions to these projects led to their successes. I would like to thank Dave Harden at Bristol-Myers Squibb for your guidance and the opportunity to learn about life in the pharmaceutical industry. Finally, I am grateful for my undergraduate mentors, Dr. Christian P. Whitman and Dr. Jamison Huddleston, for setting me up for success in the biochemistry field in graduate school.

The members of the Anderson Lab have been my family for my time at Yale. I have had the opportunity to make so many memories in and out of the lab with members that have left while I was here and with those that have come after me. Everyone has been helpful, whether it be troubleshooting through a problem or just needing someone to

commiserate with my hardships. Krasimir, you can never get enough credit as our lab manager and my graduate work would never have been successful without you. The postdocs – Albert, Nicole, Shalley, and Dan – have been great sources of knowledge and expertise and most importantly, mentorship. The graduate students – Andrea, Tomo, Molly, Zach, Victor, Zechen, Alex, Christina – have been my best friends in the lab and sources of inspiration when lab and life was difficult. Lastly, the undergraduates – Mylinh, Cory, Brandon, Vikram – were always bright-eyed and eager, which helped me remain positive throughout the latter years of grad school.

I have to thank my friends, old and new, that have kept me sane throughout graduate school. I would like to thank Danny, Jessica, Jen, Tony, my friends from my time as an undergraduate at UT Austin. We spent so many late nights studying and eating together, and the distance created by graduate school never stopped us from talking about our issues. I cannot forget to mention my graduate school friends Nakeirah, Dani, Dahyana, Jen, Valerie, Cathy. I am certainly forgetting some off of this list, but I value the time that I have spent with all of you. Of all of the things that I will miss about New Haven, I will miss you all the most when we go our separate ways. Lastly, I would also like to thank my Dungeons & Dragons groups and Elm City Games, as D&D has become the hobby that I could look forward to every week. Thus, thank you Peter, Jim, Max, Christoph, Hannah, and Travis for dealing with my roleplaying antics.

Most importantly, thank you to my family for your continuous support despite the fact that I rarely tell you about anything related to my life and graduate school. Thank you, Mom, Dad, Nina, Sarah, and my uncles. I know that I have been difficult as a kid (and adult) but I know that you still love me. I appreciate how you all raised me and instilled an

importance of hard work and education in me. Without this blessing, I would not have been able to persevere throughout the hardships of graduate school. I love you all and I hope you all realize that my thesis work is a product of your sacrifices and care for me.

Chapter 1. Introduction

Chapter 1.1 Overview of polymerases in replication and repair

DNA polymerases are the vital enzymes within the cell, responsible for proper maintenance of the genome during DNA replication, repair, recombination, and lesion bypass.[1,2] In eukaryotes, DNA polymerase α (Polα) was the first major polymerase discovered in 1957.[3] More than a decade later and thereafter Pols β, γ, δ, and ε were discovered and characterized.[4-7] Together with Polα, these five polymerases can be deemed as "classical" polymerases due to the importance of their roles within the cell and their indispensability.[2] In the last two decades, additional polymerases have been discovered that are involved in other processes, which will be briefly mentioned below (Table 1.1).

Polα-Primase (Polα-Pri) is the primary complex responsible for Okazaki fragment synthesis.[8] Primase initiates *de novo* synthesis on a DNA template using ribonucleotide triphosphates (rNTPs or NTPs) for approximately 10 nucleotides, followed by polymerase switching to Polα, which continues synthesis for about 20-30 deoxyribonucleotides (dNTPs).[9] Due to the nature of leading and lagging strand replication, the necessity of Polα is more apparent with lagging strand replication due to the increased amount of origin firing needed.[10] Canonically, Polδ is the primary polymerase for processive lagging strand synthesis and Polε for leading strand synthesis.[11-14] Unlike Polα, Pols δ and ε possess an exonuclease domain that allows for a proofreading mechanism.[15] In combination with additional factors that increase processivity such as proliferating cell nuclear antigen (PCNA), Pols δ and ε are able to carry out processive, high-fidelity nucleotide incorporation.[16,17]

Polβ characterization studies have demonstrated its role in base excision repair (BER) with its polymerase and deoxyribose phosphate (dRP) lyase activities.[4,18] BER allows the cells to combat lesions and apurinic/apyrimidinic (AP) sites that arise due to cellular and exogenous agents that significantly cause damage to DNA.[19] The importance of Polβ has been underscored by mouse experiments where knockout of the gene leads to embryonic lethality.[20]

Polγ has long been thought to be the sole polymerase in the mitochondria, responsible for replication, repair, and recombination.[5,21-23] Biochemical assays have revealed that the holoenzyme possesses polymerase, 3'->5' exonuclease, and dRP lyase activities that support these functions within the mitochondria.[24,25] With its exonuclease capability and accessory subunits, Polγ unsurprisingly has high-fidelity and processive activity.[26] Polγ has been a topic of discussion regarding nucleoside analog inhibitors, in which off-target effects have led to mitochondrial toxicity, as discussed in section 1.5.[27,28]

In the past decade, the identification of another polymerase within the mitochondria challenged the existing idea that Polγ was solely responsible for mitochondrial DNA (mtDNA) maintenance. Unlike other polymerases discussed herein, PrimPol does not belong in the overarching DNA polymerases families but in the archaeo-eukaryotic primase (AEP) superfamily.[29,30] Biochemical characterization of PrimPol revealed the highly versatile nature of PrimPol, possessing polymerase, primase with both dNTPs and rNTPs, translesion synthesis (TLS), template scrunching, and terminal transferase activities.[31] Despite all of these activities however, further cellular studies have suggested that the primary role of PrimPol is repriming downstream of stalled replication forks.[32-35]

There are polymerases dedicated to replication across DNA lesions, also known as TLS polymerases. Examples of these lesions include abasic sites, cyclobutene pyrimidine dimers (CPDs), and 8-oxoguanine adducts (8-oxoG).[36,37] Interestingly, these polymerases are typically low-fidelity. The explanation for this activity is likely being able to use a variety of damaged DNA substrates as a trade-off. In this sense, the cell risks the chance for mutagenesis in exchange for undisrupted DNA replication and cell survival. In humans, key TLS polymerases include Pols η, ι κ, REV3 (ζ), and REV1.[2,37]

Other polymerases have been characterized aside from those involved in high-fidelity replication or TLS. Polθ plays a role in theta-mediated end joining which is a double-strand break (DSB) repair mechanism.[38] Polλ has proposed roles in non-homologous end joining (NHEJ), another DSB repair pathway, as well as BER.[39] Similar to Polλ, Polμ plays a role in NHEJ but additional evidence also shows a role for Polμ in V(D)J recombination.[40]

Although not a eukaryotic polymerase, human immunodeficiency virus (HIV) reverse transcriptase (RT) is a significant family member of the DNA polymerase enzymes. RT is the overall enzyme responsible for the conversion of its single stranded RNA genome into double stranded DNA that is able to be integrated within the host genome.[41] Two activities enable RT to do this: polymerase activity using either a RNA or DNA template, and an RNaseH domain that allows RT to degrade RNA on an RNA/DNA duplex. Due to the necessity of RT in viral replication, therapeutic efforts have led scientists to target the enzyme as a means of reducing viral load, as discussed in section 1.3.[42]

Pol	Functional tasks
α	Initiator pol
β	Base excision repair pol
γ	Mitochondrial replication pol
δ	Main pol at lagging strand
ε	Main pol at leading strand
ζ	Non-redundant developmental pol
η	Xeroderma pigmentosum variant pol
θ	Repair of interstrand cross-links
ι	Meiosis pol
κ	Deletion and base substitution pol
λ	Repair in meiosis
μ	Terminal deoxynucleotide transferase homology, V(D)J recombination

Table 1.1 Table of human polymerases. Table of polymerases discussed here and proposed or validated functions. Modified from Hübscher *et al. Annual Review of Biochemistry* (2002).[2]

Chapter 1.2 Conservation of structure and mechanisms of polymerases

Polymerases, while conserved in a number of aspects, can be separated into families based on sequence and structural homology: families A, B, C, X, and Y.[2] While these differences exist, one of these conserved aspects include the overall structure of polymerases. Seminal work from Tom Steitz described polymerases as having a right-hand structure, with thumb, fingers, and palm domains.[43] The primary roles of the palm domain are to carry out the phosphoryl transfer reaction of nucleotide incorporation, the fingers domain to interact with the incoming nucleotide and template, and the thumb domain to play a role in positioning and potentially processivity.[43] While the palm domain is conserved across the families, the fingers and thumbs domain are diverse. Based on the primary sequence, there are conserved regions in polymerases numbered I-VI. These regions are responsible for important functions of these polymerases, such as harboring the important catalytic triad of aspartic or glutamic acids required for polymerase activity.[2,43,44]

Addition of a dNTP into a growing strand requires the 3'-OH of the strand to attack the α-phosphate of the incoming nucleotide.[45] The carboxylate side chains of these amino acids are crucial for recruiting two divalent metal ions required for catalysis and binding of the incoming nucleotide, typically Mg^{2+}. An aspartate near the 3'-OH group is proposed to act as a base, taking a proton from the group and making it a stronger nucleophile. The activated nucleophile then attacks the α-phosphate of the incoming dNTP, with the transition state stabilized by a metal ion. The intermediate is then resolved by the release of pyrophosphate ($PP_{i,}$ formerly the β- and γ- phosphate of the dNTP) which is stabilized by the metal ion.

The utilization of steady state and pre-steady state kinetic techniques enabled a detailed evaluation of the kinetic scheme of polymerases.[45-47] First, the enzyme binds, in a sequential manner, the DNA and then the incoming dNTP. The presence of a rate limiting step prior to nucleotide incorporation was suggested through thio-analogue studies and spectroscopic experiments using 2-aminopurine. Once the incoming dNTP is incorporated into the growing strand, a subsequent conformational change allows for release of pyrophosphate. The enzyme can then translocate for another addition of a dNTP or dissociate from the DNA substrate.

Key pre-steady state experiments using rapid chemical quench or stopped-flow instruments have been utilized to characterize polymerases are burst and single turnover experiments (Figure 1.1).[45,48-50] Burst experiments with polymerases are done where the primer/template strand are in slight excess of the enzyme, and single turnover experiments with enzyme in greater excess over primer/template. A number of polymerases, when assessed under burst experiment conditions, display a biphasic kinetic profile of a rapid product formation followed by a rate-limiting step.[51] The initial fast step corresponds to a fast, chemical step while the slower, steady-state rate corresponds to product release. If a burst phase is observed with the polymerase, the amplitude of the burst is indicative of the active site concentration with the assumption that one polymerase is able to catalyze one nucleotide incorporation event. Single turnover experiments are conducted so that all the enzyme present in the reaction undergoes only one incorporation event and cycling does not occur due to limited primer/template availability. This experiment gives a clearer assessment of the rapid chemical step. Conducting these experiments with various dNTP concentrations, then plotting the rates versus [dNTP] provides two variables: k_{pol} and K_d.

The variable k_{pol} is defined as the maximal rate of incorporation of the polymerase, and the K_d as the apparent binding constant for the incoming nucleotide. Dividing these two variables yields the incorporation efficiency, k_{pol}/K_d, which is a useful convention for comparing the catalytic activities of different polymerases and nucleotide substrates

Structural studies have correlated the conformational changes mentioned above to movements in the fingers domain as a result of dNTP binding and an induced fit mechanism to allow for phosphoryl transfer. Several experiments with polymerases such as T7, Taq, RT, and Polβ demonstrate that when a dNTP is bound, the fingers move like a hinge towards the palm domain.[52] The purpose of the "open" and "closed" conformations may be a mechanism of checking the geometry of the base pairing in the active site to preclude mismatches from occurring. If a mismatch occurs, then the polymerase may switch to the open conformation and release the mismatch. High-fidelity polymerases may use this mechanism in addition to 3'->5' exonuclease activity to achieve consistent, correct base-pairing.[53]

It is important to note that while these structural and mechanistic findings apply to a large number of studied polymerases, there are exceptions. We will discuss an example in Chapter 5 regarding the mechanism of Polα.

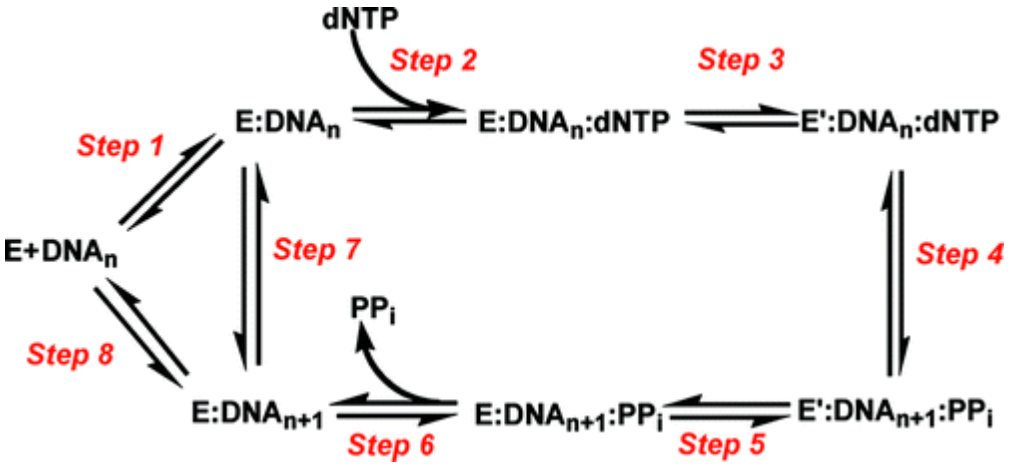


Figure 1.1 General mechanism of polymerases. A general overview for the binding events and catalytic steps in nucleotide incorporation of polymerases. Step 1. The polymerase binds a DNA substrate. Step 2. An incoming nucleotide binds after DNA. Step 3. A conformational change shift prior to catalysis may induce the polymerase into an optimal position for chemistry. Step 4. Phosphoryl transfer by nucleophilic attack of the 3'OH of the primer strand on the α-phosphate of the incoming nucleotide. Step 5. Relaxation of the polymerase to release products. Step 6. Release of pyrophosphate. Step 7. Translocation to incorporate another nucleotide for processive activity. Step 8. Release of DNA for low processivity polymerases. Reprinted with permission from Berdis, AJ. *Chem Rev.* (2009). Copyright 2009. ACS Publications.[45]

Chapter 1.3 DNA replication and polymerases as druggable targets

DNA replication is an absolute need for viability and survival of organisms from viruses to humans. Targeting DNA replication is a viable strategy to treat hyperproliferative conditions, such as in the case of cancer or viral infection where active DNA replication is prioritized.[54,55] A number of strategies have been developed to inhibit DNA replication for therapeutic purposes. One strategy is the usage of DNA damaging agents to permanently affect the overall integrity of the DNA so that nucleic acid enzymes may not recognize the substrate or act as efficiently.[56] One common medication is cisplatin, which is used as a chemotherapeutic agent. A second strategy is to target enzymes involved in DNA replication such as topoisomerase, which is responsible for the unwinding of DNA during replication. Another chemotherapy drug that targets topoisomerase is etoposide, which eventually causes dsDNA breaks.[57] A third strategy to indirectly target DNA replication is to reduce the available pool of dNTPs within the cell.[58] Antimetabolite drugs such as methotrexate function by inhibiting enzymes in the pathway responsible for the production of dNTPs. In the case of methotrexate, dihydrofolate reductase (DHFR) is inhibited. Lastly, one strategy to inhibit DNA replication is to target the enzymes responsible for the catalysis of replication, which is the topic of the work presented herein.

Given their direct roles in viral replication and cell division, polymerases have been explored as potential targets for therapy in antiviral medication and cancer treatments. Nucleoside analogue compounds are a mainstay in antiviral treatments due to the dependency of viruses on polymerases to replicate or reverse transcribe their genetic material. Key antivirals have been developed to treat HIV, Hepatitis B and C, and Ebola.[59] Nucleoside analogue drugs function by mimicking the natural nucleotide substrates

(dNTPs or rNTPs) and will be further discussed in section 1.4 using HIV RT as a case study.

Nucleoside analogues have also been developed for the treatment of cancers due to unrestrained DNA replication.[60] One difference between nucleoside analogues for antiviral versus cancer treatment is that chemotherapeutic agents such as fludarabine or gemcitabine have modified sugars that allow them to also act as antimetabolites by targeting enzymes such as ribonucleotide reductase.[55] Other drugs, like cytarabine, have been developed to inhibit both DNA and RNA polymerases to affect replication and transcription simultaneously.

One prevailing issue with chemotherapeutic treatments is that they are non-discriminatory and may affect non-tumor cells, leading to toxicity and severe patient side effects.[54,55] This effect applies to the examples mentioned above, including the polymerase-inhibiting nucleoside analogues. Despite the increased dependency of cancer cells on replication, the dosing of these nucleoside compounds must still be closely monitored to manage toxic effects. Thus, there have been efforts to look for polymerase inhibitors that may have an improved safety profile. One example includes dehydroaltenusin, a non-nucleoside Polα inhibitor.[61]

Chapter 1.4 Inhibitors of Reverse Transcriptase

HIV RT is the polymerase responsible for the conversion of the viral RNA genome to DNA. RT is a heterodimer consisting of a 66 and 51 kDa subunit (p66 and p51).[62] The p66 subunit is the functional subunit containing the polymerase active site and RNaseH domain.[63,64] The p51 acts as a scaffolding subunit primarily to maintain the active form of RT, while enhancement of other functions such as tRNA primer binding, primer/template loading, and strand displacement have been suggested.[65] In addition to targeting other machinery such as the viral integrase or protease, antiviral compounds targeting RT have been developed for lifelong treatment of HIV.

There are two classes of RT inhibitors: nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs the primary components of antiretroviral therapy (ART), being the first class of HIV compounds to gain regulatory approval with the usage of zidovudine in the late 1980s. NRTIs are nucleotide analogues that compete with natural nucleotides and function as chain-terminators. NRTIs lack a 3'-OH group that is necessary for the nucleophilic attack of the α-phosphate of the incoming nucleotide. The current eight FDA-approved NRTIs are zidovudine (AZT-TP), didanosine (ddI), zalcitabine (ddCTP), stavudine (d4T-TP), lamivudine ((-)-3TC-TP), abacavir (ABC), tenofovir (TFV-DP), and emtricitabine ((-)-FTC-TP) (Figure 1.2).

NNRTIs are allosteric inhibitors that bind in a largely hydrophobic site approximately 10 Å from the active site (Figure 1.3).[66] Interestingly, in comparison of unliganded to ligand-bound RT, the NNRTI binding pocket is unobservable if the compound is not present.[67] The mechanism of inhibition of NNRTIs is proposed to be the

induction of conformational changes that affect the positioning of catalytic residues in the active site upon binding.[68] First generation NNRTIs include nevirapine, delavirdine, and efavirenz, and second generation compounds include etravirine and rilpivirine.[69] Numerous structural studies have enabled structure-guided drug design of next generation NNRTIs. Novel NNRTIs are desired due to suboptimal pharmacokinetic properties in early generation compounds. In addition, the development of resistance of RT to NNRTIs is a prevalent issue, and creating drugs that are effective against common resistant mutations with a higher genetic barrier to resistance is desired. We will discuss our efforts in understanding a potential new class of NNRTIs in Chapter 4.

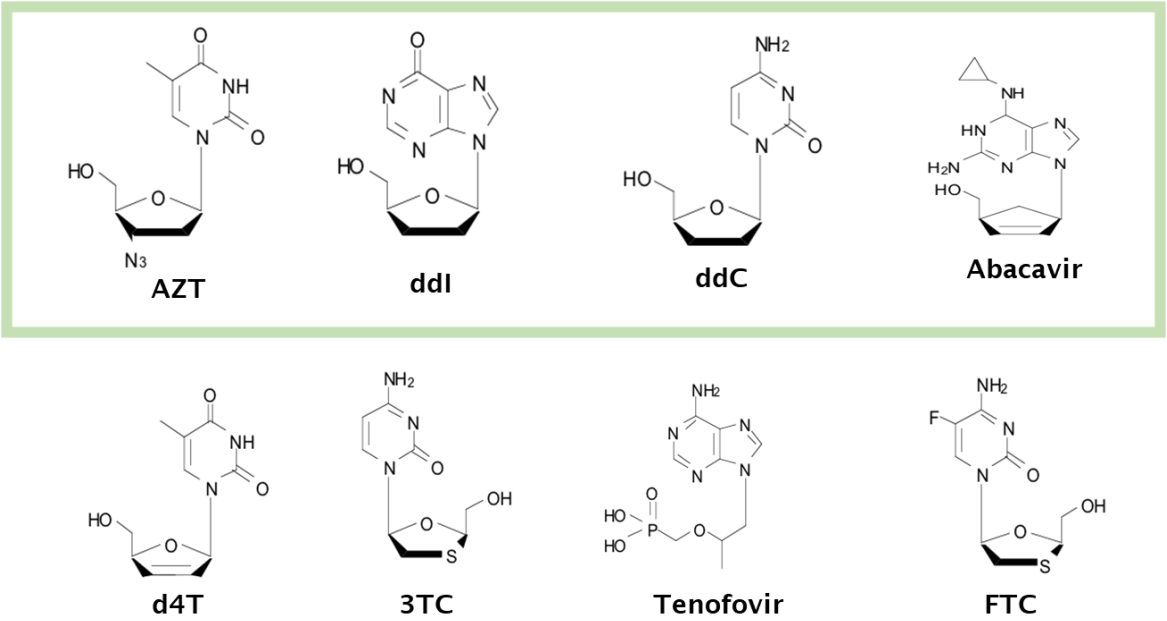


Figure 1.2 Current FDA approved NRTIs. Structures of the nucleoside reverse transcriptase inhibitors (NRTIs) that function by mimicking natural nucleotide substrates. (Green box) AZT, ddI, ddC, and ABC were previously demonstrated to be incorporated by PrimPol.

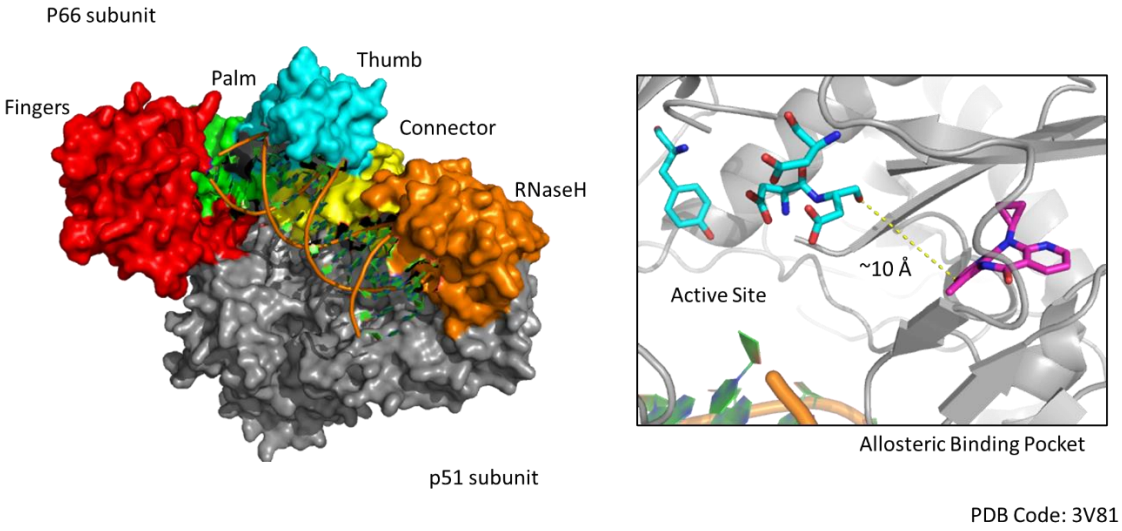


Figure 1.3 Structure of HIV RT and NNRTI binding pocket. Overall structure of the viral reverse transcriptase (RT). *Left.* RT exists as a heterodimer with a catalytic p66 subunit (color) and a scaffolding p51 subunit (gray). RT resembles a right hand with a palm (green), fingers (red), and thumb (cyan) domain. A connector region (yellow) connects the polymerase domain with the RNase H domain (orange). *Right.* The allosteric binding pocket for non-nucleoside reverse transcriptase inhibitors only appears in the presence of compound approximately 10 Å away (nevirapine, an NNRTI in magenta) from the active site (cyan). PDB: 3V81, Das *et al. Nat Struct Mol Biol* (2012).[70]

Chapter 1.5 Off-target toxicity of antiviral nucleoside analogues

While the introduction of AZT as one of the first HIV therapeutics was a significant step in treating HIV, it did not take long for the observation of negative side effects due to nucleoside analogue treatment. A connection was made that the myopathies, lactic acidosis, hepatic steatosis, and pancreatitis caused by NRTIs were similar to the effects observed in mitochondrial diseases.[71-73] With attention drawn to the mitochondria and the fundamental knowledge that nucleoside analogues could target host polymerases, the Polγ hypothesis was proposed.

The initial hypothesis was first supported with evidence showing direct inhibition of Polγ and concomitant depletion of mtDNA.[74] The hypothesis was further developed, suggesting the inhibition of the polymerase results in mtDNA depletion and accumulation of truncated mtDNA products.[75] The human mtDNA genome is approximately 17 kb, encoding 22 tRNAs, 2 rRNAs, and 13 proteins that are involved in oxidative phosphorylation. Thus, depleting mtDNA is linked to reducing levels of important metabolic enzymes, leading to pathologies similarly observed in other mitochondrial pathologies. Additional evidence suggests that other factors may play a role in NRTI-associated toxicity, such as oxidative damage to mtDNA due to free radical generation due to impaired electron transport chain processes.[72,73]

Although the Polγ hypothesis has been the prevailing explanation for NRTI-induced toxicity, there have been discrepancies between the inhibition of Polγ *in vitro* and clinical side effects.[27,76] One example is the case of tenofovir, where early clinical trials did not observe significant negative effects. However, case reports and observational studies soon demonstrated that there were cases of impartial or complete Fanconi

syndrome, a disorder of the renal proximal tubule and decline of kidney function.[77] This discrepancy between biochemical and clinical findings suggest that there are alterative mechanisms for NRTI-induced toxicity.

PrimPol, encoded by the *CCDC111* gene, is a 67 kDa polymerase that was identified through *in silico* analyses as a potential member of the AEP family.[29,30] PrimPol and its homologues consist of two conserved domains: an N-terminal AEP domain containing the active site, and a C-terminal zinc finger (ZnF) domain. PrimPol is unique to other eukaryotic polymerases in that it contains both polymerase and primase activities (Figure 1.4). Characterization of PrimPol revealed that PrimPol is able to replicate across lesions like 8-oxo-G in addition to normal 5'->3' polymerase activity. The ability for PrimPol to prime using dNTPs or NTPs is dependent on the presence of the ZnF domain. Studies have established PrimPol is able to reprime downstream of chain terminators, R-loops, and G-quadruplexes.[31-35] One structure of PrimPol has been solved, revealing that PrimPol does not resemble the canonical right-hand structure that describes many of the characterized polymerases.[78] As such, one interesting feature of PrimPol is that few contacts are observed in the primer strand. The lack of contacts in the primer/template may explain the low-fidelity nature of PrimPol with an error rate of approximately $1x10^{-4}$.[79] Contributing to this high error rate is the lack of an exonuclease domain that other replicative polymerases possess. In addition, PrimPol is a fairly low processive enzyme, incorporating 1-4 nucleotides per binding event.[35] Cellular knockout studies provide evidence that PrimPol acts as a DNA damage tolerance polymerase.[29,30,33,80]

Given evidence that PrimPol is localized in mitochondria and potentially plays a vital role in DNA damage tolerance mechanisms, emphasizing the importance of PrimPol

in mitochondrial genome maintenance.[29] Thus, our lab has explored the possibility that PrimPol could mediate NRTI-associated toxicity. Previous work in our lab demonstrated the ability of PrimPol to incorporate AZT, ddI, ddC, and ABC *in vitro* (Figure 1.2).[81] However, with the prevalent use of tenofovir as a standard component of ART, we aimed to assess the potential for PrimPol to mediate tenofovir-associated toxicity in Chapter 2 and 3.

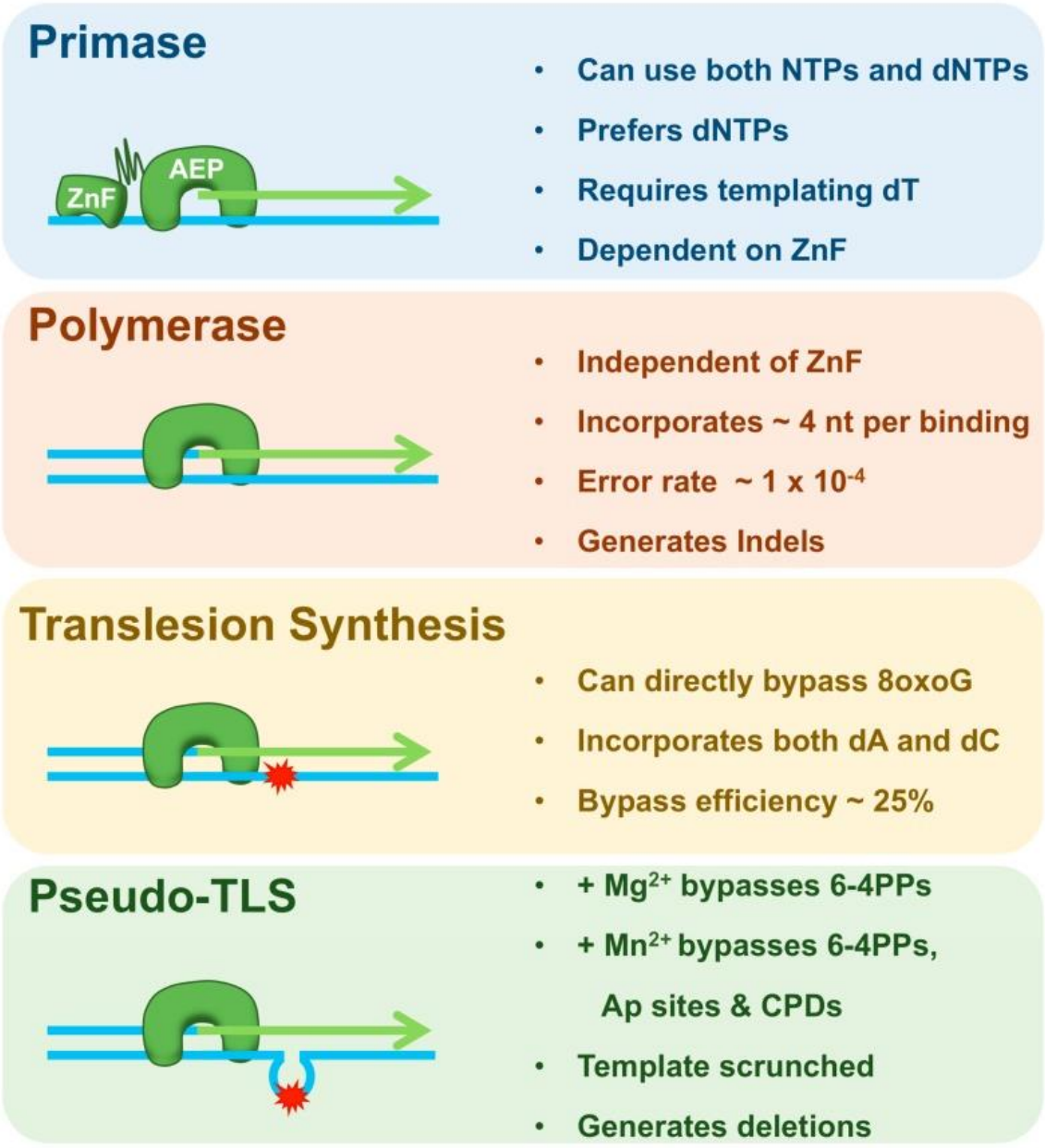


Figure 1.4 Activities of PrimPol. PrimPol is able to prime using its zinc finger (ZnF) domain, using either dNTPs or NTPs. PrimPol also has general and translesion synthesis (TLS) polymerase capabilities, at a low fidelity and low processivity. PrimPol is also able to bypass lesions through a template scrunching mechanism that results in deletions. Reprinted with permission Guilliam *et al. Genes* (2017).[31]

Chapter 1.6 Thesis objectives

There is strong evidence that polymerases are some of the most important enzymes to life. Polymerases are directly responsible for the maintenance of the genome in eukaryotes, prokaryotes, and viruses. In both viral and cancer treatments, targeting polymerases are viable treatment strategies. In order to develop therapeutics in diseases that target polymerases or are derived from polymerase activity, we must understand the mechanistic underpinnings of the processes at hand. In this dissertation, I will present experiments spanning three different polymerases, which exemplifies the prominent role of polymerases in various therapeutic endeavors. The goal of this work is to understand the role and mechanism of polymerases involved in the treatment of HIV and cancer.

Discerning the role of PrimPol in NRTI-derived toxicity

Adverse effects due to the administration of NRTIs are attributed to the off-target incorporation of NRTIs by Polγ. Inhibiting Polγ, the primary replicative polymerase in the mitochondria, leads to mitochondrial toxicity which may manifest as myopathy, lactic acidosis, or liver failure. These side effects greatly affect patient adherence to ART, necessitating the need to develop NRTIs that are not substrates for host polymerases. Tenofovir is a common NRTI in which cases of nephrotoxicity have been reported. Interestingly, there is a discrepancy between the toxicity observed in the clinic and *in vitro* data suggesting TFV is a weak substrate for Polγ. This raised the question if there were other polymerases within the mitochondria that could lead to toxicity with antiviral treatment.

Initial work in our lab validated the incorporation of four NRTIs by PrimPol, a recently discovered polymerase in the mitochondria, *in vitro*. However, due to the prevalence of tenofovir in preferred combination therapies and observed mitochondrial toxicity primarily in the kidney, we desired to understand the interactions of tenofovir with PrimPol. In Chapter 2, we assessed the potential for PrimPol to incorporate the active form of tenofovir (TFV-DP) *in vitro*. With the knowledge that PrimPol is able to reprime downstream of stalled replication forks, we proposed a model where PrimPol could either 1) mediate TFV-induced toxicity by incorporating TFV, or 2) protect against toxicity by repriming downstream of chain termination events caused by TFV. Using a cellular system, we assessed this model to discern the role of PrimPol in TFV-induced toxicity.

Previous work proposed that a Polγ mutation could predispose an HIV+ patient to toxic side effects by affecting how well the host polymerase is able to discriminate between NRTIs and natural nucleotides. Taking this possibility into account, we identified the D114N active site PrimPol mutation in an HIV+ patient experiencing mitochondrial toxicity under a tenofovir-based treatment. In Chapter 3, we biochemically and kinetically assessed the effects of the mutation on PrimPol and propose implications for possessing the mutation in the context of antiviral therapy.

Understanding the mechanism of a novel class of NNRTIs for RT

In efforts to design NNRTIs that target drug resistant mutations with improved pharmacological properties, the Anderson and Jorgensen lab at Yale have developed NNRTIs through structure-guided drug design. Initial characterization of the 2-naphthyl phenyl ether compounds revealed that the series of compounds could be distinguished into

two groups by efficacy against the common Y181C drug resistant mutation. In chapter 4, we structurally characterize a set of 2-naphthyl phenyl ether compounds with WT and Y181C RT to understand the discrepancy of inhibitory activity against Y181C. Understanding the mechanism of inhibition of these compounds will aid in the development of potent NNRTIs.

Distinguishing the effects of replication substrates on Polα activity

During replication, Polα is part of the primosome, a coordinated system of moving parts that allow for efficient Okazaki fragment synthesis and subsequent high-fidelity elongation. There are examples of this precise coordination: the primase accessory subunit regulates the length of the initial hybrid RNA/DNA duplex synthesized by the primase active site. Due to the low-fidelity of Polα, the primosome should limit the number of nucleotides Polα can incorporate onto the newly synthesized Okazaki fragment. Interestingly, there is evidence that after replication mutagenic DNA is left over from Polα activity, suggesting this coordination and regulation may be imperfect. If Polα activity is unrestrained, then incorrect incorporation could potentially lead to harmful mutations and tumor formation. While there are mechanisms to induce Polα switching to the higher fidelity replicative pols δ and ε, we investigated whether different nucleic acid substrates could act as an additional intrinsic regulation mechanism to limit Polα activity. In Chapter 5, we biochemically assessed the activity of Polα on two substrates that it encounters during replication. Revealing the effects of relevant nucleic acid substrates on Polα further progresses our understanding of replication and potential associated disease mechanisms.

Chapter 2. Delineating the role of PrimPol in TFV-mediated toxicity

This chapter is an excerpt from:

Duong VN, Zhou L, Martínez-Jiménez MI, He L, Cosme M, Blanco L, Paintsil E, Anderson KS. Identifying the role of PrimPol in TDF-induced toxicity and implications of its loss of function mutation in an HIV+ patient. Sci Rep. 2020 Jun 9;10(1):9343. doi: 10.1038/s41598-020-66153-z. PMID: 32518272; PMCID: PMC7283272.

2.1 Introduction

Human immunodeficiency virus (HIV) treatment and chemoprophylaxis regimens commonly consist of one or more nucleoside reverse transcriptase inhibitors (NRTIs) that target the HIV-1 reverse transcriptase enzyme. NRTIs are nucleoside analogs that lack a 3′-OH, thus acting as chain terminators of viral replication once incorporated by HIV reverse transcriptase (RT). Since the approval of zidovudine (AZT) in 1987, NRTIs have become the backbone of antiretroviral therapy (ART); the advent of ART has led to sustained HIV viral suppression, dramatic decrease in HIV-associated morbidity, and mortality.[82,83] Thus, contemporary NRTI-based regimens have significantly contributed to the health of HIV+ individuals allowing them to have near-normal or normal life expectancies as the general population in the absence of a cure for HIV.[82,84-86]

NRTI toxicity has widely been attributed to off-target inhibition of the primary polymerase responsible for mitochondrial genome replication, DNA Polymerase gamma (Polγ), sometimes termed the Polγ hypothesis.[27,73,87] Particularly with earlier generation NRTIs, these off-target effects could lead to lactic acidosis, lipodystrophy, peripheral neuropathies, cardiomyopathies, skeletal muscle myopathies, and pancytopenia.[88-90] However, there are discrepancies between the degree of in vitro inhibition of Polγ and the observed clinical toxicity of certain NRTIs. For instance, the NRTI tenofovir disoproxil

fumarate (TDF), a prodrug of tenofovir, is among the least toxic inhibitors of Polγ as determined by in vitro assays, however, there are reports of mitochondrial dysfunction and toxicity by TDF in the renal proximal tubules of the kidneys of HIV-infected individuals.[77,91,92] Mechanistic studies have shown that Polγ incorporates the natural dATP substrate much more efficiently and selects against the active tenofovir diphosphate (TFV-DP) metabolite leading to a very favorable in vitro discrimination factor, suggesting that the Polγ hypothesis cannot fully explain the proposed mitochondrial toxicity caused by TDF.[93,94] These discrepancies may be explained by factors such as differences in metabolism, binding affinity and rate of incorporation of the respective NRTIs by Polγ, ineffective exonuclease removal, and the role of additional host cell polymerases.[95-97]

PrimPol is the most recent enzyme involved in DNA replication that has been observed to be localized to the mitochondria apart from Polγ.[29,30,34,98,99] Characterization of PrimPol has revealed that it is a DNA and RNA primase as well as a DNA-dependent translesion synthesis polymerase.[29,30] Further evidence has implicated that the primary role of PrimPol in vivo is repriming stalled replication forks by hydroxyurea (HU) or UV light[33,34], rising from G-quadruplexes[32], R-loops[100], or chain-terminating nucleotides.[33] We have previously confirmed that PrimPol is able to incorporate a subset of NRTIs, establishing a potential role of PrimPol in NRTI-induced mitochondrial toxicity.[81]

Taking into consideration the repriming capabilities of PrimPol and the potential for off-target incorporation of NRTIs by host polymerases, we began by addressing the broader question of whether PrimPol may directly contribute to NRTI-induced mitochondrial toxicity with a focus on TDF. We validated that PrimPol was able to incorporate the active form of tenofovir (TVF-DP) in vitro, albeit with a relatively low

efficiency. Then we generated PrimPol overexpression and knockdown renal proximal tubular epithelial cells (RPTECs) to assess mitochondrial toxicity and respiration when the cells were treated with TDF. Under our experimental conditions, we propose that PrimPol plays a protective role against NRTI-induced toxicity.

2.2 Materials and Methods

Protein Purification of WT, AA354 PrimPol

PrimPol WT and AA354 (truncated protein without the zinc finger) was purified following a modified protocol of [81]. The pET28a-PRIMPOL expression vector was transformed into *E. coli* BL21(DE3)-pRIL cells. For 1 L of LB + kanamycin, 10 mL of overnight culture was added for protein production. Cells were grown at 37 °C until the OD_{600} 0.6, at which point the flasks were chilled at 4 °C, PrimPol induced with 1 mM IPTG, and allowed to induce overnight at 19 °C (approximately 16 hours). Cells were harvested at 12,000 × g for 15 min at 4 °C, and the pellet flash frozen in liquid nitrogen and stored at −80 °C. For every 1.5 g of pellet, 5–10 mL of lysis buffer (Buffer A: 50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM Imidazole, 10% v/v glycerol, EDTA-free protease inhibitor, 0.5 mM TCEP, 0.1% Triton X-100). The suspension of cells was lysed by passing the cells through a high-pressure homogenizer (Emulsiflex) 2–3 times. The lysate was centrifuged at 28,000 × g for 1 hour at 4 °C. The supernatant was loaded onto a 5 mL HisTrap FF crude column equilibrated in Buffer A and washed with buffer A after protein loading until the A_{280} was stabilized. The column was then equilibrated with buffer B (Buffer B: 50 mM Tris-HCl pH 8, 50 mM NaCl, 10% v/v glycerol, 0.5 mM TCEP) to reduce the salt concentration. The column was washed with 5 CV 95% Buffer B and 5% Buffer C (Buffer

C: 50 mM Tris-HCl pH 8, 50 mM NaCl, 600 mM imidazole, 10% v/v glycerol, 0.5 mM TCEP) or until A_{280} was stabilized. The protein was then eluted by a 0–100% gradient of buffer B and buffer C, separating the elution into 0.5 mL fractions across 5 column volumes (25 mL). The fractions were resolved by 4–20% Tris-glycine gel in SDS and proteins stained by Coomassie blue staining. The cleanest fractions were pooled and TEV protease was added to the pooled fractions (1:10 TEV:total protein w/w). The solution was dialyzed with buffer D (Buffer D: 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP) overnight at 4 °C using 25 kD dialysis membrane tubing. The protein was loaded onto a 5 mL HisTrap FF column equilibrated in Buffer B. Buffer B was used to wash the column until the A_{280} was stabilized. 5% buffer C was used to wash the column and then a 0–100% gradient of Buffer B and Buffer C was used in the same manner as the previous column. Because the His-tag was cleaved with TEV protease, PrimPol eluted in the 5% imidazole wash, although we did observe the presence of cleaved PrimPol in the 0–100% gradient fractions. The clean fractions were pooled, concentrated, and buffer exchanged to about 0.5–2 mL into buffer E (Buffer E: 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP) using a 10 K centrifugal filter. The protein was then separated using a Superdex 200 Increase 10/300 size exclusion column using Buffer E. Clean fractions were pooled, concentrated, aliquoted, and flash frozen and stored at −80 °C.

Oligonucleotide labeling and annealing

In general, primer oligonucleotides were labeled at the 5′ end with [γ−32P]ATP and T4-PNK with the provided reaction buffer for 30–45 minutes at 37 °C. The reaction was then stopped by heat shock for 5 minutes at 70–95 °C. For the primase assays, the labeled

oligos and unlabeled template were mixed at a 1:2 ratio in 50 mM Tris-HCl pH 7.5, 300 mM NaCl and heated for 10 minutes at 80 °C, and slowly cooled down to room temperature. In all other kinetic assays, primer and template were mixed in a 1:1.1 ratio in 10 mM Bis-Tris Propane, pH 7.0, 300 mM NaCl and annealed by heating to 90 °C for 5 minutes, 55 °C for 15 minutes, and 37 °C for 10 minutes.

Burst and single turnover kinetics

Kinetic assays to measure the activity of WT PrimPol were based on the single incorporation of dCTP on a radiolabeled DNA primer:
(5′-GCCTCGCAGCCGTCCAACCAACT-3′) annealed to a DNA template:
(5′-GGACGGCATTGGATCGAGGTTGAGTTGGTTGGACGGCTGCGAGGC-3′).

Burst and single turnover kinetics were conducted as previously done, with modifications to the protocol noted below[81]. All kinetics assays were carried out using reaction buffer (10 mM Bis-Tris Propane, pH 7.0, 300 mM NaCl) as reported previously. Briefly, PrimPol was incubated with the primer:template and mixed with dCTP and 10 mM $MnCl_2$ before quenching with 0.5 M EDTA. Products were collected in a tube with formamide dye [0.1% bromophenol blue (w/v), 0.1% xylene cyanol (w/v)] and separated by denaturing urea PAGE. The radiolabeled products were visualized by the Molecular Imager FX phosphorimager (Bio-Rad) and quantified by Quantity One, version 4.6.9 (Bio-Rad).

In the case of measuring the kinetics of wild-type PrimPol, pre-steady state kinetic assays were performed using the RQF-3 rapid chemical quench apparatus (KinTek) at room temperature. For burst reactions, final concentrations of 10 μM WT PrimPol was incubated with 30 μM annealed primer:template before mixing with 200 μM dCTP and

10 mM $MnCl_2$ at 37 °C. The [product] was plotted against time and the data points were fit to a burst equation, [product] = $A(1-e^{-kt}) + A(k_{ss})$ (t), where A is the burst phase amplitude, k_{obs} is the observed single exponential rate, k_{ss} is the steady-state rate, and t is the time.

For single turnover experiments, final concentrations of 10 μM WT PrimPol was mixed with 300 nM primer:template and mixed with 0–300 μM dCTP and 10 mM $MnCl_2$. The single data points were fit to a single turnover equation, [product] = $A(1-e^{-kt})$, where A is amplitude and k_{obs} is the observed single exponential rate, and t is the time using Kaleidagraph. The rates were then plotted against [dCTP] used and fit to a quadratic equation, [dNTP] = 0.5 (K_d + [dNTP] + [k_{pol}])− 0.5|(K_d + [dNTP] + [k_{pol}])2 − 4[dNTP] [k_{pol}]|1/2, in order to extract the k_{pol}, the maximal rate of incorporation, K_d, the apparent binding constant for the incoming nucleotide, and k_{pol}/K_d, the overall efficiency for nucleotide incorporation.[81] In both the burst and single turnover experiments, the values represent the fit estimate for the parameter ± one standard deviation.

For tenofovir diphosphate incorporation assays, the 3′-end of the primers were varied, D20A and D45:

5′-GCCTCGCAGCCGTCCAACCX$_1$–3′, where X_1 is A, C, G, or T.

The corresponding annealed oligo templates were:

5′-GGACGGCATTGGATCGAGGTTGAGTX$_2$GGTTGGACGGCTGCGAGGC-3′,

where X_2 is the natural base pair to X_1. Initial experiments used 200 μM TFV-DP under single turnover conditions, monitoring TFV-DP incorporation at 0, 2, 5, 30, 60, and 120 minutes. Then experiments were done under single turnover conditions using 0–1000 μM TFV-DP, and control dATP experiments used 0–100 μM. K_d curves were generated from each of these rates using a quadratic equation to estimate the k_{pol} and K_d.

In the case of TFV-DP, the rates did not vary with [TFV-DP], so k_{pol} was calculated by taking the average of all rates. The amplitude, however, did vary with TFV-DP concentration, so the K_d was calculated by plotting the amplitudes against [TFV-DP].

Confirmation experiments to validate the preceding nucleotide preference used a different oligo: D21A and D36:

5′-TCAGGTCCCTGTTCGGGCGCX1–3′ and

5′-TCTCTAGCAGTX2GCGCCCGAACAGGGACCTGAAAGC-3′,

using 200 μM TFV-DP under single turnover experiments at 0, 2, 5, 30, 60, and 120 minutes. Additional experiments to observe if the preference was present with the zinc-finger knockout (amino acids 1–354) used 10 μM $PrimPol_{1–354}$ and the D20A:D45 substrates at 0, 2, 5, 30, 60, and 120 minutes. Additional experiments were carried out to observe the preference effects using 200 μM d4T-TP, (−)-3TC-TP, or (−)-FTC-TP. For d4T-TP incorporation experiments, D22T:D45 was used:

5′-GCCTCGCAGCCGTCCAACCAAX1–3′ and

5′- GGACGGCATTGGATCGAGGTTGAX2TTGGTTGGACGGCTGCGAGGC.

For (−)-3TC-TP and (−)-FTC-TP experiments, D23C:D45 was used:

5′-GCCTCGCAGCCGTCCAACCAACX1–3′ and

5′-GGACGGCATTGGATCGAGGTTGX2GTTGGTTGGACGGCTGCGAGGC-3′.

The reactions were quenched for these experiments at 0, 30 seconds, 1, 3, 10, 30, and 120 minutes.

Cell culture maintenance and reagents

Immortalized renal proximal tubular epithelial cells (RPTECs) stably transfected with the OAT1 receptor were acquired from ATCC. RPTECs were maintained in

DMEM:F-12 (containing 2.5 mM L-glutamine, 15 mM HEPES, 0.5 mM sodium pyruvate, and 1200 mg/L sodium bicarbonate) supplemented with 0.3 µg/mL puromycin, 100 µg/mL G418, 25 ng/mL PGE1, 3 pg/mL triiodothyronine, 25 ng/mL hydrocortisone, 10 ng/mL hEGF, 3.5 ug/mL ascorbic acid, ten-fold diluted ITS-G (10x stock), and 5% HI-FBS. After transduction with lentivirus containing PrimPol shRNA or overexpression plasmids, 200 µg/mL hygromycin was supplemented to the media. Cells were passaged every 3 days by a 1:10 split.

Cell line transfection and transduction

HEK293T cells were cultured in DMEM with high glucose supplemented with 10% FBS and transfected with plasmids containing dR8.91, VsV-G, and a third with either the shRNA knockdown (pLKO.1) or overexpression construct (pLenti). Briefly, HEK293Ts were seeded to 70–80% confluency and allowed to adhere to a 6-well plate. A ratio of 1:0.1:1 (0.75 µg dR8.01, 75 ng VsV-G, 0.750 µg pLKO.1/pLenti) were combined in a tube. Approximately 1 mL of Serum-free DMEM was mixed with 4.7 µg PEI (3:1 PEI:total DNA ratio) and incubated at room temperature for 5 minutes. The media:PEI mixture was added dropwise to the three-plasmid solution and incubated for 15 minutes at room temperature. The final mixture was then added to HEK293T cells. After 24 hours, the media was replaced with complete growth media for RPTECs (DMEM/F12). After another 24 hours, the media containing the lentivirus was aliquoted into 1 mL volumes and flash frozen and stored at −80 °C until use.

For transduction of RPTECs, 6-well plates were plated to approximately 40–50% confluency on the day of transduction. The virus was thawed and added to the adherent cells at 37 °C overnight. In some cases, the virus was diluted 2-fold with complete growth

media. The media was replaced the next day with complete growth media for 24 hours at 37 °C. The following day, the media was replaced with selection growth media containing 200 µg/mL hygromycin. After seven days, the cells were immunoblotted for the overexpression or knockdown of PrimPol to confirm successful transduction.

Cell counting

6-well plates were seeded with 50,000 cells and allowed to adhere. On the second and fourth days, cells were trypsinized using 0.25% trypsin-EDTA and counted using trypan blue and the Countess II Automated Cell Counter (ThermoFisher). The cell counts were plotted against time after seeding. Each well containing cells was counted three times and each cell type consisted of three biological replicates.

Immunoblotting

TRAP1, PrimPol, or Chk1 levels in response to TDF treatment were measured through immunoblotting. 6-well plates were plated with 100,000 cells and allowed to adhere overnight. After treating RPTECs with 30 µM TDF for TRAP1 for 3 days or 1–60 µM TDF for 5 days for PrimPol, cells were harvested by washing with cold PBS two times followed by the addition of RIPA buffer supplemented with protease inhibitor (Roche) and cell scraping. Cells were incubated at 4 °C on a tube shaker for 30 minutes and centrifuged at 18,000xg for 10 minutes at 4 °C. The supernatant was collected and protein levels were measured through the BCA assay to normalize total protein loading. 10–30 µg of total protein were loaded onto a 4–20% Tris-Glycine gel and run in SDS buffer at 200V for 35 minutes. The loaded protein was transferred to a nitrocellulose membrane using the iBlot2 transfer system. The membrane was blocked with 5% milk for one hour, incubated with primary antibody overnight at 4 °C (1:1000 rabbit anti-TRAP1, 1:1000 rabbit anti-Chk1,

1:1000 rabbit anti-GAPDH, 1:500 rabbit anti-PrimPol), washed with 1x TBST three times for 5 minutes each, incubated with secondary antibody (anti-rabbit IGG, HRP-linked 1:1000) for 1 hour at room temperature, and washed with 1x TBST three times for 5 minutes each. The membranes were exposed to enhanced chemiluminescence reagent for TRAP1, Chk1, and GAPDH or SuperSignal West Femto Maximum Sensitivity Substrate for PrimPol for 1 minute. The membranes were then exposed to film for 5 seconds to 5 minutes and developed. Blots were quantified using ImageJ. In order to normalize across blots, the TRAP1/GAPDH or PrimPol/GAPDH levels of the treatment conditions were further normalized to the untreated control for each cell line, $n = 3$. Significance was determined by one-way ANOVA using GraphPad Prism. $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

Cell proliferation assays

5000 cells were plated into 96-well plates and allowed to adhere overnight. On the next day, 30 µM tenofovir disoproxil fumarate (TDF), 100 µM abacavir (ABC), or 20 µM efavirenz in complete media was used to replace the media. On the third day of treatment, the media was replaced with fresh media with TDF. On the fifth day, a solution of 6 mM MTT in PBS was diluted to 1 mM in growth media and 100 µL of the solution was added to each well. After 3 hours of incubation at 37 °C, 150 µL of stop solution was added to each well (10% H_2O v/v, 4% NP-40 v/v, 0.34% concentrated HCl v/v in isopropanol). The plates were protected from light and were shaken overnight at room temperature. The well solutions were resuspended by pipetting and absorbances were read at 590 nm. The absorbances were then subtracted from the background (no cells) and then normalized to the untreated cells. For each biological replicate (one 96-well plate), six technical replicates

were done (6 wells in each plate). Significance was determined by one-way ANOVA using GraphPad Prism, n = 3. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

mtDNA quantification

100000 cells were seeded into 6-well plates and allowed to adhere overnight. The cells were then treated with 30 µM TDF for 5 days, with replacement of media on the third day. On the fifth day, the cells were detached with 0.25% trypsin-EDTA for 5 minutes and pelleted at 150xg for 5 minutes. The pellet was resuspended with PBS and pelleted an additional two times and stored at −80 °C until lysis for DNA harvesting. Total DNA was then isolated from the cells using the Qiagen DNeasy Blood & Tissue Kit, lysing the cells by adding PBS and vortexing for at least 30 seconds. The DNA concentrations were quantified spectrophotometrically and then diluted to 3 ng/µL with sterile H_2O.

Quantification of mtDNA was followed using a qPCR-based method according to[101]. Each well in a 96-well plate contained 25 µL of reaction mixture: 2 µL of 3 ng/µL DNA (6 ng total), 2 µL of 400 nM qPCR primer pair (32 nM final), 12.5 µL of SYBR Green Master Mix, and 8.5 µL of nuclease-free water. The qPCR primer pairs target either the mitochondrial tRNA-Leu(UUR) gene or the nuclear B2-microglobulin gene. The qPCR reaction mixtures were heated to 95 °C for 3 minutes, then went through 50 cycles of 95 °C for 10 seconds and 60° for 30 seconds (reading the RFU at the end of each cycle using the SYBR channel), and finally 95 °C for 10 seconds, and then ramping up from 65 °C to 95 °C in 0.5° increments every 5 seconds (reading the RFU at every 5 seconds). After obtaining the C_T values for both mtDNA and nucDNA from the Bio-Rad CFX Manager software, the relative amount of mitochondrial DNA content was calculated by $2 \times 2^{\Delta CT}$, where ΔC_T is nucDNA C_T – mtDNA C_T. The relative mtDNA content values were then normalized to

the untreated control. For each biological replicate (one qPCR plate), three technical replicates were done (three wells in the plate). Significance was determined by one-way ANOVA using GraphPad Prism, $n = 3$. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

Mitochondrial respiration rate measurements

The respiration rate of the RPTECs were determined using the Agilent Seahorse XF-96 Extracellular Flux 374 Analyzer. The RPTECs were seeded in 96-well microplates at 35,000 cells per well and allowed to adhere overnight. The following day the cells were treated with either 30 µM TDF or in complete media for 2 days. One hour before plate reading, the media was switched to a media without bicarbonate or phenol red, 1 mM pyruvate, 2 mM glutamine, 10 mM glucose, 30 µM TDF and allowed to incubate at 37 °C without CO_2. The drug injection ports were filled with oligomycin (Port A, final 1.5 µM), FCCP (Port B, final 2.0 uM), and rotenone/antimycin A (Port C, final 0.5 uM). Respiration was measured three times before the first drug injection and after each drug injection to allow for reading stabilization. The exception was that six measurements were taken after oligomycin injection. Each biological replicate (one 96-well plate) contained 6 technical replicates. Oxygen consumption rates were obtained through the Wave software (Agilent). In order to calculate the basal respiration, M13 (first measurement after rotenone/antimycin A injection) was subtracted from M3 (measurement immediately before oligomycin injection). Proton leak was measured by subtracting M13 from M9 (last measurement after oligomycin injection). ATP-linked respiration was calculated by subtracting the proton leak from basal respiration. Maximal respiration was calculated by subtracting M13 from M10 (first measurement after FCCP injection). The spare reverse capacity was calculated by subtracting the basal respiration from the maximal respiration. The coupling efficiency

was calculated by diving the ATP-linked respiration by the basal respiration and multiplying by 100. The cell respiratory control ratio was calculated by dividing the maximal respiration by the proton leak. Lastly, the ATP-linked and maximal respiration ratio was calculated by diving the ATP-linked ratio by the maximal respiration. The values were averaged over separate experiments, $n = 3$. Significance was determined by one-way ANOVA using GraphPad Prism $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

2.3 Results

PrimPol has low efficiency of tenofovir diphosphate incorporation

The current study is focused on defining potential mechanisms of NRTI-mediated nephrotoxicity in HIV+ patients who are taking tenofovir-containing antiretroviral drug regimens. Since TDF has been shown to only be a weak inhibitor of mitochondrial Polγ[94], we examined a potential role of PrimPol, a primase-polymerase, found to have significant levels of expression in the kidney.[29]

A role of PrimPol is to reprime downstream of stalled replication forks, which may arise due to depletion in dNTPs, thymine-dimers formation by UV, G-quadruplexes, R-loops, or chain-terminating nucleotides, during both nuclear and mitochondrial DNA replication.[32-34,100,102] Thus, in the context of the mitochondrial toxicity associated with NRTI-based therapies, PrimPol can have a protecting role by repriming and rescuing replication forks that were stalled due to NRTI incorporation by Polγ (Fig. 2.1A, left). Alternatively, PrimPol could directly contribute to NRTIs-associated toxicity, taking into consideration these nucleotide analogues could also be valid substrates for PrimPol that could conceivably block its primase/polymerase activity (Fig. 2.1A, right). In this event,

PrimPol could increase toxicity via chain termination, by the synthesis of abortive primers and the consequent inability to rescue stalled forks. Our previous work has confirmed the incorporation of select NRTIs by PrimPol as ddATP or CBV-TP, the active metabolites of didanosine (ddI) and abacavir (ABC), with discrimination values in the efficiency of the incorporation ($efficiency_{dNTP}/efficiency_{NRTI}$) from 3 to 10^2-fold.[81] In the current study, we investigated the likelihood that tenofovir could be utilized as a substrate by PrimPol during elongation, taking into consideration that there is a rising, unexplained renal toxicity in patients taking TDF-based ART.

We tested the incorporation of the active form of TDF, tenofovir-diphosphate (TFV-DP), compared to the natural nucleotide dATP, using a defined labeled primer/template. In light of recent findings with other polymerases, indicating that the nucleotide directly adjacent to the incorporation site (n-1) may plausibly affect substrate binding[103], we used 4 variants of the primer/template, differing in the base pair (N:X) forming the primer-terminus (Fig. 2.1B). We currently term these substrates as PreA, PreC, PreG, and PreT, corresponding to a dA, dC, dG, and dT in the n-1 position of the primer. In conducting these biochemical assays, we observed that PrimPol can incorporate TFV-DP differently on these substrates, showing more efficient kinetics of incorporation in favor of PreT (Figs. 2.1C, D, and 2.2A). We validated these results with a different pair of primer and template, demonstrating that this effect is primarily due to the n-1 nucleotide (Figs. 2.2B). Furthermore, we also observed that this effect persists when the zinc finger domain of PrimPol is absent, which suggests that active site interactions in the polymerase domain may be able to explain the preceding nucleotide preference (Figs. 2.2C). Testing

other active triphosphate NRTIs that PrimPol incorporates, d4T, (−)-3TC, and (−)-FTC, shows that this effect is unique to tenofovir (Figs. 2.2D).

As a control to determine if this preceding nucleotide preference is unique to tenofovir or also shared with the natural nucleoside substrate, we generated full K_d curves through single-turnover kinetics for TFV-DP and dATP incorporation by PrimPol with all four preceding nucleotide variations in the template (Table 2.1, Fig. 2.3A-F). The efficiency of TFV-DP incorporation as a function of the preceding nucleotide was PreT> PreA> PreG ~ PreC, with a 10-fold variation between PreT and PreC. However, considering this preference effect with natural dATP incorporation shows a difference of only about 2-fold. This observation implies that the insertion of TFV-DP is facilitated somehow by a thymine base at the primer-terminus. However, even in this favored context the insertion of the TFV-DP is 2×10^4-fold lower than the insertion of the dATP counterpart. In comparison to previously determined incorporation efficiencies of other NRTIs, TFV-DP incorporation is weaker, being incorporated approximately three magnitudes less efficiently than CBV-TP (Table 2.2).[81]

Lastly, we addressed the possibility that TFV-DP could hinder the priming activity of PrimPol. We assessed the ability of TFV-DP to compete with ATP during dimer formation on a 3′-T_{20}G**TC**AGACAGCAT_{29}-5′ substrate. Even at concentrations of TFV-DP in excess of ATP, dimer formation was not disrupted (Fig. 2.4A). Next, we examined the ability of TFV-DP to interrupt primer elongation using various templates and substrates (Fig. 2.4B). Extremely high concentrations of TFV-DP were able to reduce the primer length of the products to a greater extent when competing against ATP compared to dATP. This is likely because PrimPol prefers to utilize dATP when elongating primers. We also

observe that TFV-DP is able to modestly reduce the length of products when dATP is in either the 3′ site, or 5′ site. Together, the weak interference of priming activity and modest efficiency of NRTI incorporation suggests that PrimPol would likely be a lesser contributor to TDF toxicity.

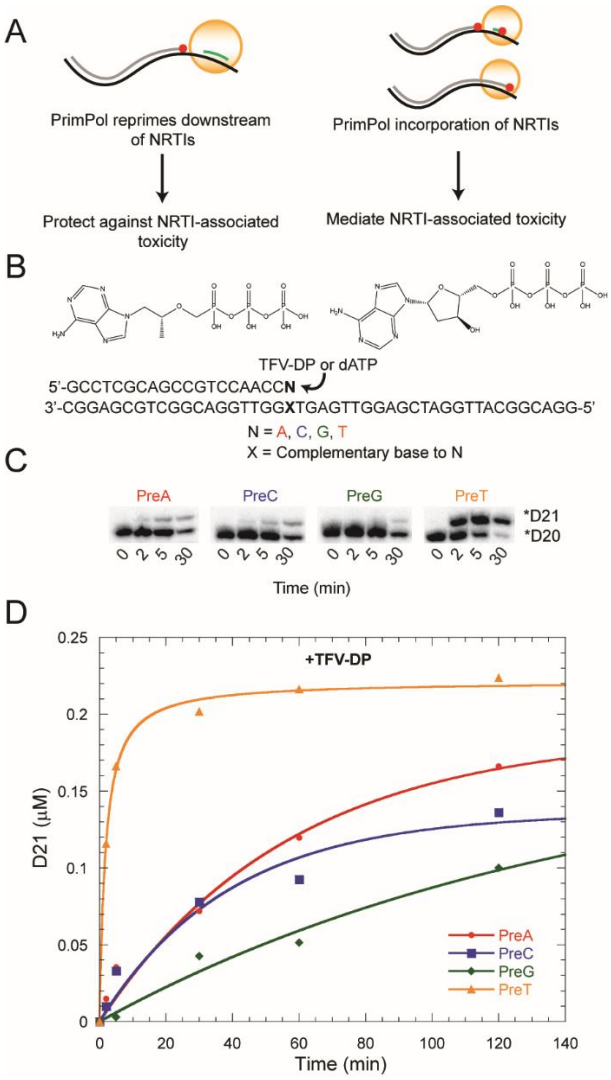


Figure 2.1. PrimPol modestly incorporates tenofovir-diphosphate *in vitro* with a preceding nucleotide preference.

A) Diagram depicting the potential roles of PrimPol in NRTI-associated toxicity. The left panel demonstrates the ability of PrimPol to alleviate toxicity by repriming downstream of a chain-terminated strand. In the right panel, PrimPol can mediate toxicity by incorporating NRTIs and thus stalling replication. Alternatively, the incorporation of NRTIs could prevent the ability of PrimPol to rescue replication by terminating priming. B) Experimental reaction set-up to demonstrate tenofovir-diphosphate incorporation by PrimPol. Generally, a radiolabeled dsDNA substrate with a template dT in the next incorporation position is extended by either dATP or TFV-DP. The n-1 nucleotide and its complimentary base was varied (referred to as PreA, PreC, PreG, PreT) to show the effect on efficiency of nucleotide incorporation. C) Denaturing PAGE of the TFV-DP incorporation reaction with varying nucleotides in the position preceding incorporation. The lower band is the initial substrate and the upper band is the TFV-DP-incorporated DNA. D) Graphical representation of the reaction shown in C).

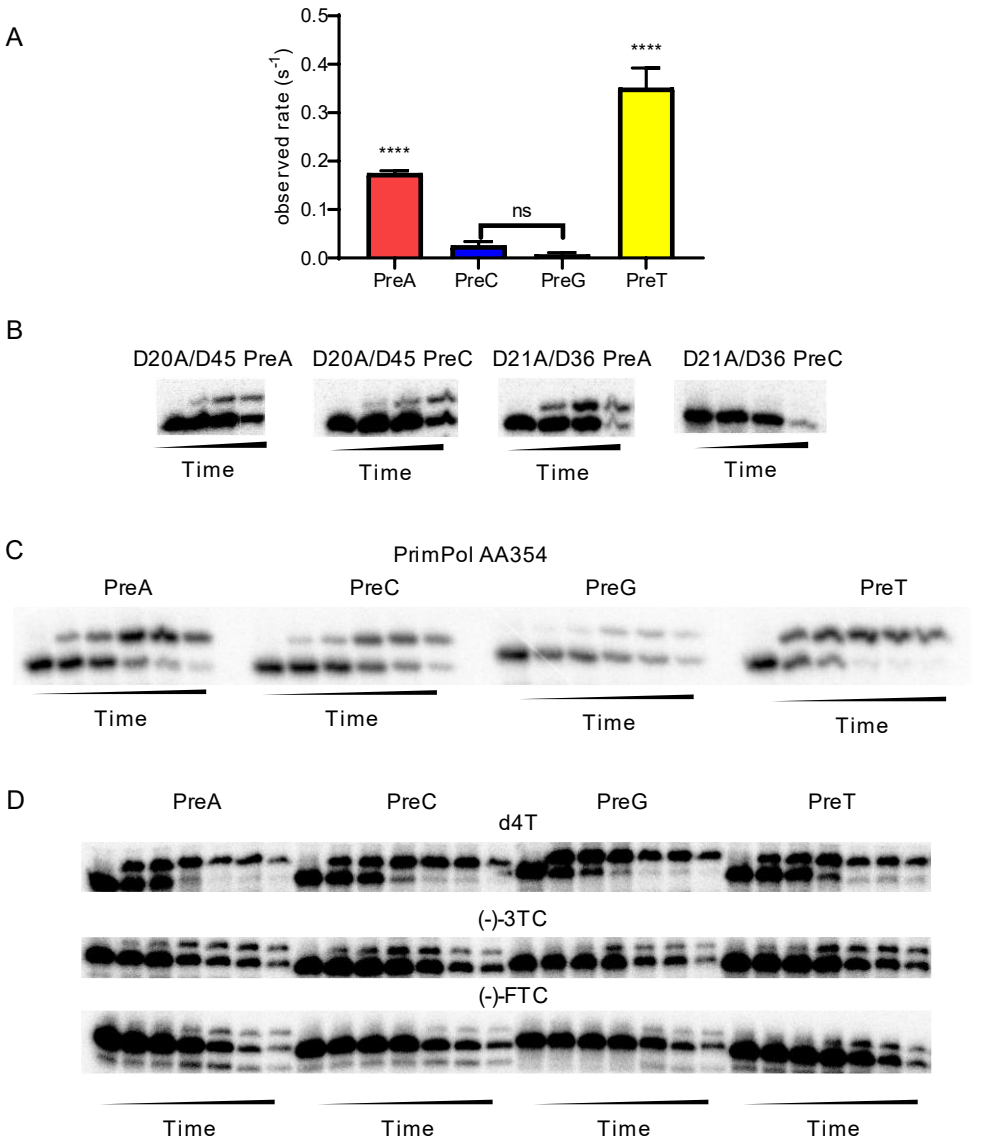
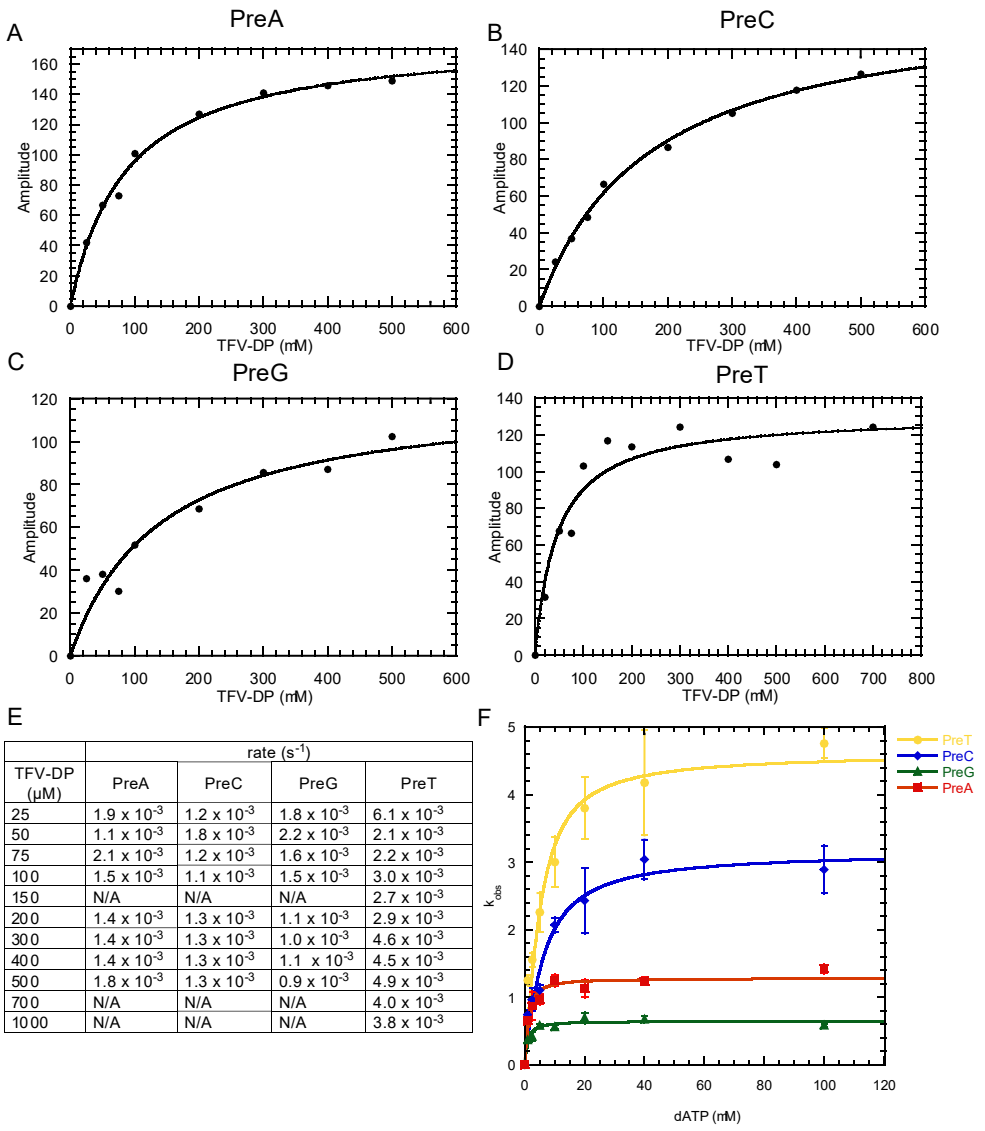


Figure 2.2. The preceding nucleotide preference is replicable in other sequences, occurs without the zinc finger, and is unique to TFV-DP.
A) Statistical significance of the estimated rates for TFV-DP incorporation in Fig. 1C and 1D using the one-way ANOVA test, **** = $p<0.0001$. B) Denaturing PAGE showing the preceding nucleotide preference for TFV-DP incorporation by PrimPol is present for two independent primer/template sequences, D20A/D45 (left two gels) and D21/D36 (right two gels). In both cases, PreA TFV-DP incorporation is favored over PreC. C) The preceding nucleotide preference also occurs with the isolated polymerase domain of PrimPol (amino acids 1-354) using the D20A/D45 primer/template. D) The preceding nucleotide preference was tested using different NRTIs, d4T (dT analog, D22T/D45, top), (-)-3TC (dC analog, D23C/D45, middle), and (-)-FTC (dC analog, D23C/D45, bottom). In all cases there was not a strong preceding nucleotide preference, demonstrating that the effect observed with TFV-DP is unique in the NRTIs tested.

TFV-DP	k_{pol} (s^{-1})	K_d (μM)	Inc. Eff. k_{pol}/K_d (s^{-1} μM^{-1})	Fold-difference to PreC
PreA	0.0016 ± 0.0003	85.6 ± 8.3	1.8×10^{-5}	2.4
PreC	0.0013 ± 0.0002	172.3 ± 15.3	7.6×10^{-6}	1
PreG	0.0014 ± 0.0005	138.1 ± 43.7	1.0×10^{-5}	1.3
PreT	0.0037 ± 0.001	45.0 ± 13.7	8.2×10^{-5}	10.8
dATP				
PreA	1.3 ± 0.06	0.7 ± 0.3	1.9	2.7
PreC	3.2 ± 0.2	4.6 ± 1.3	0.7	1
PreG	0.6 ± 0.03	0.7 ± 0.2	0.9	1.3
PreT	4.6 ± 0.3	2.9 ± 0.9	1.6	2.3

Table 2.1. Summary of TFV-DP and dATP incorporation by WT PrimPol dependent on the preceding nucleotide in the primer strand.

Pre-steady state kinetic parameters for both TFV-dP and dATP incorporation by WT PrimPol at at 37°C were determined by fitting the time course data to the following single exponential equation: $[\text{product}] = A(1- e^{-k}obs^{t})$, where A is amplitude and k_{obs} is the observed single exponential rate, and t is the time. In the case of dATP incorporation, the single exponential rates were then plotted against each concentration of [dNTP] using a quadratic equation in order to extract the k_{pol}, the maximal rate of incorporation, K_d, the apparent binding constant for the incoming nucleotide, and k_{pol}/K_d, the overall efficiency for nucleotide incorporation. In the case of TFV-DP, the k_{pol} was independent of [dNTP] the amplitude was plotted against [dNTP] and fit to the quadratic equation to determine the K_d value. The errors represent the standard error values of the parameters that corresponds to a confidence level of 68.3%, or to one standard deviation.



	rate (s^{-1})			
TFV-DP (μM)	PreA	PreC	PreG	PreT
25	1.9 x 10^{-3}	1.2 x 10^{-3}	1.8 x 10^{-3}	6.1 x 10^{-3}
50	1.1 x 10^{-3}	1.8 x 10^{-3}	2.2 x 10^{-3}	2.1 x 10^{-3}
75	2.1 x 10^{-3}	1.2 x 10^{-3}	1.6 x 10^{-3}	2.2 x 10^{-3}
100	1.5 x 10^{-3}	1.1 x 10^{-3}	1.5 x 10^{-3}	3.0 x 10^{-3}
150	N/A	N/A	N/A	2.7 x 10^{-3}
200	1.4 x 10^{-3}	1.3 x 10^{-3}	1.1 x 10^{-3}	2.9 x 10^{-3}
300	1.4 x 10^{-3}	1.3 x 10^{-3}	1.0 x 10^{-3}	4.6 x 10^{-3}
400	1.4 x 10^{-3}	1.3 x 10^{-3}	1.1 x 10^{-3}	4.5 x 10^{-3}
500	1.8 x 10^{-3}	1.3 x 10^{-3}	0.9 x 10^{-3}	4.9 x 10^{-3}
700	N/A	N/A	N/A	4.0 x 10^{-3}
1000	N/A	N/A	N/A	3.8 x 10^{-3}

Figure 2.3. Determining the kinetic parameters of TFV-DP and dATP incorporation by WT PrimPol dependent on the preceding nucleotide in the primer strand (see Table 1).

In the case of TFV-DP incorporation by PrimPol, the amplitude of product formation varied with TFV-DP concentration. A-D) The kinetic parameters of TFV-DP incorporation by PrimPol were calculated by plotting the amplitude of single turnover experiments with PrimPol and TFV-DP against the concentrations of TFV-DP used. The K_d was determined by the concentration at half of the amplitude max. E) The similar rates across TFV-DP concentrations were averaged to determine the k_{pol}. F) The k_{pol} and K_d of dATP incorporation by PrimPol was determined by plotting the rates of each single turnover experiment against the concentration of TFV-DP used in each experiment.

NRTI	k_{pol} (s^{-1})	K_d (μM)	Inc. Eff. k_{pol}/K_d (s^{-1} μM^{-1})
TFV-DP	0.0037 ± 0.001	45.0 ± 13.7	8.2×10^{-5}
ddATP[a]	0.0138 ± 0.0008	15 ± 3	9.0×10^{-4}
AZT-TP[a]	0.0040 ± 0.0008	38 ± 6	1.0×10^{-4}
ddCTP[a]	0.033 ± 0.002	18 ± 4	2.0×10^{-3}
CBV-TP[a]	0.013 ± 0.002	~1	1.3×10^{-2}

Table 2.2. Comparison of incorporation efficiencies of NRTIs by PrimPol.

Values for TFV-DP incorporation by PrimPol under optimal conditions (preceding dT in the primer) are compared to the incorporation efficiency values of other NRTIs by PrimPol. The errors represent the standard error values of the parameters that corresponds to a confidence level of 68.3%, or to one standard deviation. [a]Values obtained from Ref. [81]

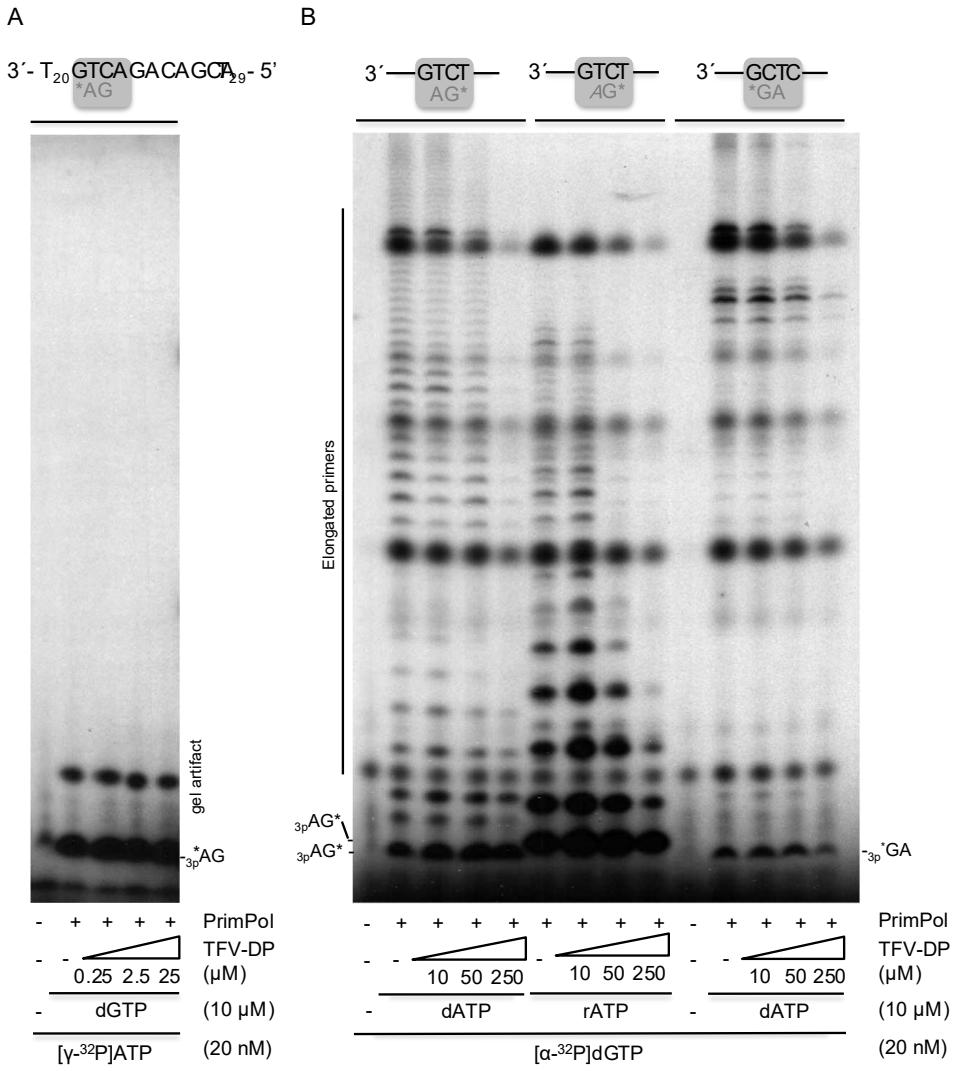


Figure 2.4. TFV-DP is able to reduce the product length of primers during elongation but unable to inhibit dimer formation.
A) Dimer formation during primer initiation by PrimPol was monitored in various concentrations of TFV-DP. TFV-DP did not compete with ATP even at excess concentrations. B) The interference of TFV-DP in primer elongation by PrimPol was monitored. Using a 3'-GTCT-5' and a 3'-GCTC-5' substrate, TFV-DP was added at different concentrations in competition with dATP or ATP.

Respiratory capacity is ablated by reducing PrimPol levels

Because PrimPol was able to incorporate TFV-DP *in vitro*, although with a low efficiency, the role of PrimPol either mediating or protecting against mitochondrial toxicity remained uncertain. We predicted that PrimPol overexpression or knockdown cell lines could address the role of PrimPol in tenofovir-associated toxicity. If PrimPol had a protective effect against tenofovir toxicity, then the overexpression cell lines would fare better compared to the knockdown strains. Conversely, if PrimPol actively incorporates tenofovir and stalls replication, then the knockdown cell line would exhibit less phenotypes related to toxicity when treated with tenofovir. Due to the presence of tenofovir-caused nephrotoxicity in the renal proximal tubules of the kidney, we generated stable cell lines with overexpressed wild-type PrimPol or knocked-down levels of PrimPol using immortalized renal proximal tubular epithelial cells (RPTECs) (Fig. 2.5A).[104]

Tenofovir treatment of cells can affect metabolism and reduce the respiratory capability of the mitochondria. Tenofovir is able to downregulate TRAP1, a regulator of glycolysis, which is accompanied by changes in cellular respiration.[105] As a first approach, we monitored TRAP1 levels through immunoblotting after treatment with tenofovir disoproxil fumarate (TDF) for 3 days to observe the potential of varying levels of TRAP1 as an indicator of corresponding changes in metabolism (Fig. 2.5B, C). We observed a downward trend in TRAP1 protein levels in both the scrambled and shRNA knockdown cell lines, but not in the overexpression strain. This result prompted us to further assess the potential of PrimPol to resist or enhance the effects of tenofovir treatment on metabolism through cellular respiration measurements using the Seahorse XF Analyzer (Fig. 2.5D).[106]

Upon treatment with tenofovir, we observed a universal reduction in basal respiration and maximal respiration (Fig. 2.6A, B). Interestingly, even in the untreated controls, the shRNA knockdown RPTECs showed a lower maximal respiration rate compared to both scrambled and overexpression cell lines, which is recapitulated in the TDF treatment conditions. Consequently, there was a sharp reduction in the spare respiratory capacity, or the ability of the cell to respond to respiratory needs which may arise in stress conditions (Fig. 2.6C). In order to control for effects of respiration based on changes in cell number, we calculated the internally normalized parameters of ATP-linked respiration/maximal respiration, cell respiratory control ratio, and coupling efficiency (Fig. 2.6D-F). Of the three internally normalized parameters, the ATP-linked respiration to maximal respiration ratio displayed a downward shift with the treatment of TDF, and is thus the appropriate parameter to compare cell strains (Fig. 2.6D). Although there were significant decreases in basal and maximal respiration and spare respiratory capacity with the shRNA knockdown strains, the difference in the ATP-linked-respiration to maximal respiration ratio was absent, which suggests that the differences in cellular respiration were due primarily to decreases in cell count.

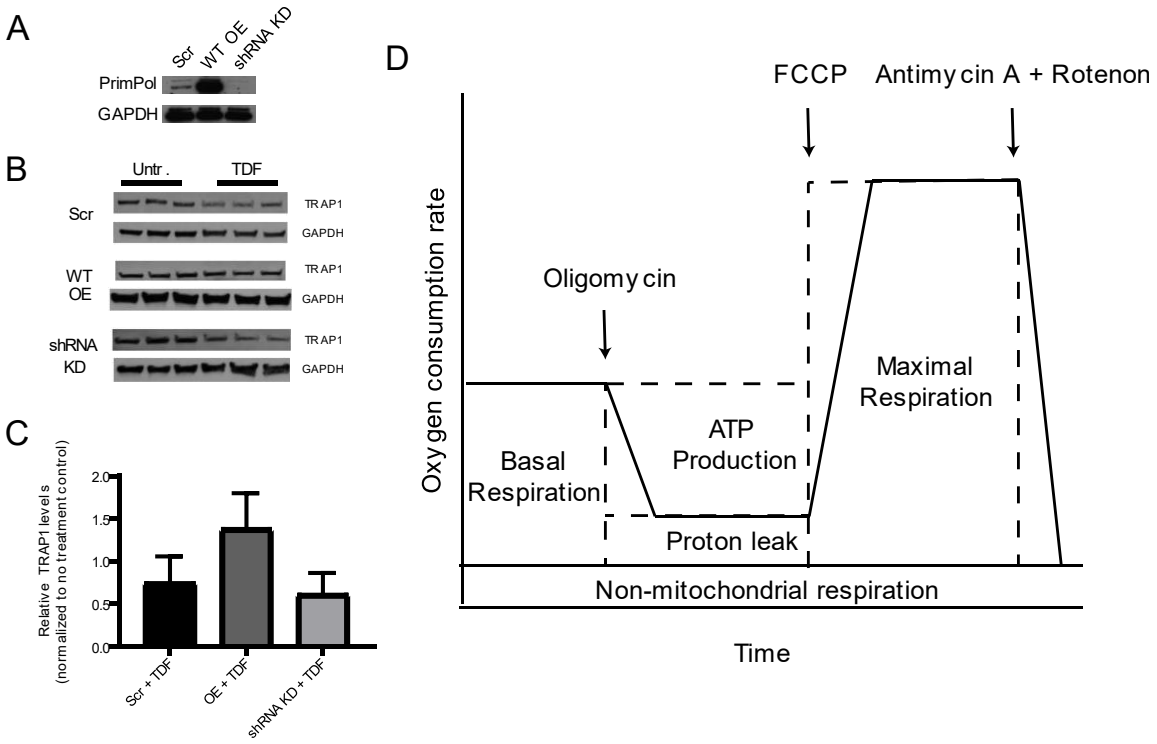


Figure 2.5. TRAP1 as an indicator of affected metabolism in PrimPol knockdown RPTECs.

A) Western blotting of PrimPol of the scrambled, wild-type PrimPol overexpression, and shRNA knockdown renal proximal tubule epithelial cell lines. B) Immunoblotting of TRAP1 in RPTECs after treatment of 30 μM TDF for 3 days. The lanes for each treatment represent a technical replicate, n=3. C) Quantified relative amounts of TRAP1 levels across 3 biological replicates. The normalized levels of TRAP1 to GAPDH were compared for each cell strain to the untreated control. The shRNA and scrambled cell lines show a downward, but statistically nonsignificant (one-way ANOVA), trend of TRAP1 levels compared to the PrimPol overexpression strain. D) Seahorse XF Analyzer MitoStress assay schematic to measure parameters related to mitochondrial respiration.

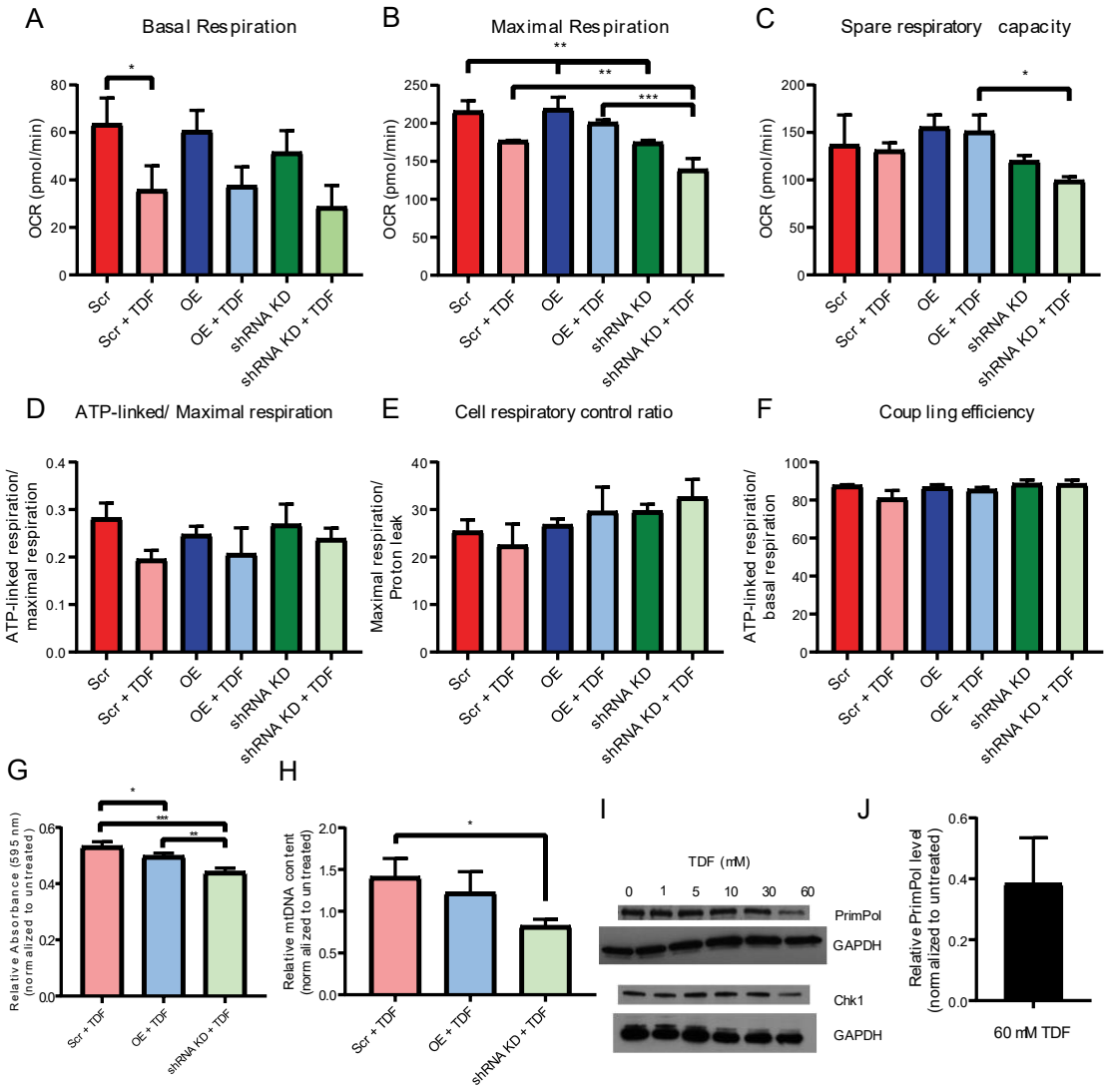


Figure 2.6. PrimPol knockdown RPTECs display reduced mitochondrial fitness and hypersensitivity to TDF treatment.

Scrambled, PrimPol overexpression, and shRNA knockdown strains of RPTECs were treated with 30 µM TDF for 3 days and oxygen consumption rate was monitored (A-F). The spare respiratory capacity, C), was calculated by subtracting the basal respiration from the maximal respiration. Internally normalized parameters were calculated to control for cell number (D-F). For toxicity experiments, RPTECs were treated with 30 µM TDF and parameters related to toxicity were measured after 5 days. G) Cell proliferation was measured through the MTT assay. Absorbances of the treatment conditions were normalized to the untreated control. H) Relative mtDNA content was quantified by qPCR and the treatment conditions were normalized to the mtDNA content of the untreated control. I) Scrambled cells were treated with varying concentrations of TDF for 5 days and then immunoblotted for PrimPol. J) The relative amounts of PrimPol at 60 µM in I) were quantified relative to the untreated control. Significance was determined by one-way ANOVA, n = 3. * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$.

PrimPol knockdown cells display increased sensitivity to tenofovir

In addition to examining the cellular respiration, it is essential to understand if PrimPol contributes to or alleviates mitochondrial toxicity in the context of tenofovir treatment. Cellular proliferation assays were utilized with the knowledge from the respiration experiments that suggested that the knockdown cell lines may be more sensitive to TDF treatment. After treatment of TDF for 5 days, cellular proliferation was monitored via the MTT assay and absorbances were normalized to untreated cells (Fig. 2.6G). Although a decrease in proliferation was observed in the overexpression cell lines compared to the scrambled strain, the shRNA knockdown RPTECs experienced a greater reduction in proliferation compared to both scrambled and overexpression cells. The apparent sensitivity of the shRNA knockdown cells to TDF treatment compared to the other cell lines corroborates our respiration data (Fig. 2.6A-D and 2.7A), implying that the reduction in cell number contributed to the decreased respiration. Altered mtDNA copy number is an additional phenotype related to NRTI-associated toxicity.[72,73,107] After cells were treated with TDF for 5 days, the relative mtDNA content was quantified using qPCR and normalized to untreated cells. In both the scrambled and overexpression cell lines, the amount of mtDNA was increased compared to the untreated cells (Figs. 2.6H). In contrast, the PrimPol shRNA knockdown cells experienced a decrease in mtDNA compared to the untreated control, which may be an indicator of toxicity.

To validate our results, we observed the effect of the NRTI abacavir and efavirenz, a non-nucleoside reverse transcriptase inhibitor (NNRTI) on the proliferation of the RPTEC strains. In the case of abacavir, we expected that we would observe a similar decrease in proliferation in the knockdown strains. However, because abacavir is

incorporated by PrimPol to a greater extent than tenofovir, the possibility that higher levels of PrimPol could mediate toxicity remained a possibility. We observed similar results to when the cells were treated with TDF in the proliferation assays, suggesting that PrimPol plays a protective role even in the case of ABC treatment (Fig. 2.7B). To confirm that our results are specific to NRTIs, we replicated the assays in the presence of efavirenz and observed no difference in proliferation between our overexpression and knockdown cell lines (Fig. 2.7C).

Consolidating both the cellular respiration and toxicity analyses provide the impression that the primary role of PrimPol in NRTI-associated toxicity is repriming downstream of chain-terminated nucleotides and preventing stalled replication forks. Given this putative role, we hypothesized that the cells may upregulate PrimPol levels in response to TDF treatment as a protective measure. Surprisingly, our findings were contrary to our predictions in that, under high doses of TDF (60 μM), we observed a downregulation of PrimPol protein levels in RPTECs that was similarly observed with Polγ (Fig. 2.6I, J).[105] Recently, it was shown that PrimPol is upregulated by the activation of the ATR pathway[108], perhaps through interactions with RPA.[109,110] It has also been previously described that NRTIs are able to downregulate Chk1, which is downstream of ATR. Thus, TDF treatment may concurrently downregulate Chk1 and PrimPol. Indeed, immunoblotting for Chk1 also shows a decrease of protein levels under high dosing of TDF (Fig. 2.6I) and that Chk1 downregulation is associated with PrimPol downregulation. Although the full mechanism to which PrimPol and Polγ is downregulated with treatment of high concentrations of TDF is unclear, it is possible that decreased amounts of these proteins may contribute to manifestation of toxicity.

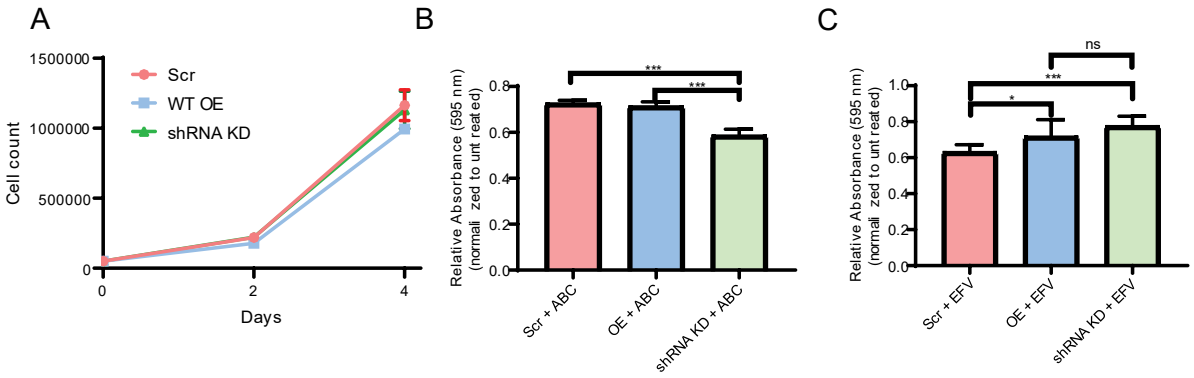


Figure 2.7. Modulating PrimPol does not affect cell viability and validating the effects of PrimPol levels on NRTI-caused toxicity.
A) Cells were seeded in 6-well plate at a low density and then counted after 2 and 4 days, n=3. B) The PrimPol RPTEC strains were treated with abacavir over 5 days and cellular proliferation was monitored using the MTT assay, n=3. C) As a control, the cells were treated with efavirenz, an NNRTI, over 5 days and the MTT assay was used to assess the toxicity, n=3. Significance was determined by one-way ANOVA, * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$.

2.4 Discussion

Nucleotide sequence context can influence NRTI incorporation by PrimPol and other polymerases

Upon testing the "preceding nucleotide" effect on TFV-DP incorporation by PrimPol, we did not expect the striking differences observed. Although there are numerous examples of enzymes with nucleic acid sequence preferences[111-113], whether the current observations might extend to other polymerases requires further investigation. Possible explanations for this effect may be due to additional binding interactions of TFV-DP with PreT, or the PreT dsDNA substrate may have an altered nucleic acid structure[114] that may affect the active site conformation and subsequently TFV-DP incorporation that may be related to the acyclic nature of the chemical structure.[115] Determining the ternary structure of PrimPol:dsDNA:TFV-DP would be valuable to reveal the mechanism behind this nucleotide preference.

Understanding the biochemical and structural mechanism(s) underlying such a preference for inserting TFV-DP next to a thymine at the primer terminus could prove to be useful in drug design efforts.[116] For example, if a nucleoside analog inhibitor was being developed as an antiviral therapy, an important consideration would be avoidance of off-target effects for host polymerases such as Polγ. Based upon our findings of the influence of sequence specificity this assessment should be carried out using a variety of DNA substrates. While we have confirmed the preceding nucleotide effect in the specific case of PrimPol and TFV-DP, it may be crucial to extend these experiments to other NRTIs and the respective target and host polymerases. In our experiments, we were able to observe discrimination differences up to 10-fold with different nucleotides in the n-1 position

(Table 2.1). Thus, it is highly likely that NRTI discrimination values in the current literature may be under- or overestimated. Assessing the potential for sequence effects on NRTI incorporation by Polγ would be essential to provide more accurate estimates of the contribution of the enzyme to NRTI-associated toxicity.

We speculate that our observations may arise from the acyclic nature of tenofovir compared to other NRTIs. We assessed the potential for other NRTIs to display sequence dependence effects but only observed a significant difference in incorporation with different preceding nucleotides when using tenofovir. Interestingly, we observed faster incorporation of tenofovir with dT or dA in the n-1 position compared to dC and dG. It is possible that an A-T base pair enhances tenofovir incorporation compared to a G-C base pair. Solving the complexes of PrimPol with tenofovir and these duplex substrates will allow us to understand how the atypical structure of tenofovir and base-pairing may affect its incorporation efficiency.

PrimPol plays a key role in protection against tenofovir-associated toxicity

In the present study, we addressed the possibility that PrimPol can alleviate tenofovir-associated toxicity. In light of our cellular experiments, we propose that the benefits of the repriming ability outweigh any possible toxicity due to tenofovir incorporation by PrimPol. One caveat to our experimental setup was that the RPTECs were treated with TDF for a short amount of time (3–5 days). However, if nephrotoxicity arises in patients after steady, low, long-term exposure to antiretroviral therapies, then the possibility that PrimPol could incorporate tenofovir at a low level over years of treatment still exists. Thus, it would be desirable to recapitulate our assays under longer treatment periods of TDF to more appropriately mirror a clinical situation.

Intriguingly, we observed decreased amounts of PrimPol when cells were treated with a high amount of TDF. A previous study demonstrated that treating cells with a high amount of TFV decreased the protein levels of Polγ and the authors suggest that the downregulation of Polγ may lead to toxicity.[105] While it is unknown if PrimPol and Polγ cause toxicity through downregulation or are downregulated as a result of toxicity, PrimPol appears to be regulated by a similar pathway as Polγ. In light of recent findings showing that PrimPol is regulated in part by the ATR pathway, we also determined that PrimPol regulation by TDF is associated with Chk1 regulation. Our current findings support these previous studies and suggest a potential pathway to investigate to determine the mechanism of regulation of PrimPol by tenofovir. It will also be of value to examine the levels of Polγ and PrimPol in our cohort of patients to identify any changes in proteins level that may stem from long-term exposure to NRTIs in antiviral therapy.

Chapter 3. Biochemical investigation of the PrimPol D114N active site mutation identified in a HIV+ patient with mitochondrial toxicity

This chapter is an excerpt from:

Duong VN, Zhou L, Martínez-Jiménez MI, He L, Cosme M, Blanco L, Paintsil E, Anderson KS. Identifying the role of PrimPol in TDF-induced toxicity and implications of its loss of function mutation in an HIV+ patient. Sci Rep. 2020 Jun 9;10(1):9343. doi: 10.1038/s41598-020-66153-z. PMID: 32518272; PMCID: PMC7283272.

3.1 Introduction

In the previous chapter, our cell culture experiments suggest that PrimPol plays a protective role against TDF-toxicity, perhaps through repriming catalytic activity. The possible involvement of mitochondrial polymerases in toxicity could be magnified by mutations in PrimPol or Polγ that impair catalytic function. In fact, prior studies from our lab identified a Polγ R953C mutant in an HIV+ patient, which may predispose the patient to NRTI-induced mitochondrial toxicity by altering the ability of Polγ to discriminate between natural nucleotides and NRTI nucleotides.[117] We postulated that if variants of PrimPol that impair the function of PrimPol existed in individuals, then these mutations could predispose these individuals to possible NRTI-induced toxicity.

Based upon the earlier finding that a mutation in Polγ may predispose patients on NRTI-regimens, we sought to identify possible mutations in the *PRIMPOL* gene in a cohort of HIV+ patients experiencing mitochondrial toxicity under tenofovir-containing antiretroviral drug regimens. We identified an HIV+ patient in this cohort who had a D114N active site mutation in PrimPol. In the current study, we characterized the effects of D114N PrimPol mutation at the molecular level and found that this amino acid substitution substantially impairs the primase and polymerase catalytic activities.

Taking into consideration of our results in chapter 2 that PrimPol plays a protective role against NRTI-induced toxicity, we surmise that the presence of inactivating mutations in PrimPol such as D114N might contribute to the mitochondrial toxicity associated renal toxicity in some patients on TDF-based ART.

3.2 Materials and Methods

Protein Purification of D114N and ZnF PrimPol

D114N and the isolated zinc finger domain of PrimPol were purified identically to the WT and AA354 constructs detailed in Chapter 2.

Oligonucleotide labeling and annealing

Oligonucleotide labeling and annealing were done identically as detailed in Chapter 2.

Burst and single turnover kinetics

Kinetic assays to measure the activity of D114N PrimPol were done with identical conditions and oligos as detailed in Chapter 2, with exceptions. Reactions for the comparison of D114N to WT were done at room temperature due to relative protein instability at 37 °C. In addition, the incorporation of the D114N variant was slow enough to allow for manual mixing reactions to be employed rather than rapid chemical quench flow methodology.

Study participants and procedures

Study participants were enrolled at the Yale-New Haven Hospital from April 2011 to March 2013. The details of the study design for this cohort have been described previously.[118] In brief, for this PrimPol sub-study, cases (n = 13) comprised HIV-infected

individuals on ART for at least 12 months with clinical and/or laboratory toxicities associated with mitochondrial toxicity. Cases were matched by age, sex, and race/ethnicity to HIV-negative controls (n = 19). All participants gave their written informed consent before participation in the study. The study protocol was approved by the Institutional Review Board of the Yale School of Medicine and all the research was performed in accordance to the relevant guidelines and regulations.

At study enrollment, participants answered a brief survey comprised of demographic characteristics and past medical history. Medical records of HIV-infected participants were reviewed, and disease characteristics and laboratory data (complete blood count, serum chemistries, liver function test, lipid profile, urinalysis, HIV RNA copy number, and CD4 + T-cell count) were extracted. Each participant gave about 20 ml of venous blood at the time of enrollment. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood within 2 hours of collection using Ficoll gradient (Ficoll-Hypaque; ICN) as described previously. Aliquots of PBMCs were stored at −80 °C until DNA extraction for the experiments.

DNA extraction and sequencing

Genomic DNA was extracted from PBMCs using TRIzol Reagent (Invitrogen, Carlsbad, CA, Cat.No.15596026) according to the manufacturer's instructions. This sub-study included only study participants with sufficient archived DNA for the analysis (cases, n = 13, and controls, n = 19). Conversed active site and zinc finger coding exons were amplified with target specific primers:

3th exon forward primer: 5′-TGGGCAACAGAGCTGACTC-3′,

3th exon reverse primer: 5′-GAAAAACTTGAGTTGGCCATT-3′;

5th exon forward primer: 5′-TAAGATGCGGTGTGTGGAGA-3′,

5th exon reverse primer: 5′-CGGTCTGATGGAGAAAGCTG-3′;

9th exon forward primer: 5′-GTGAATAAAGATGGCATTAAAGGAGG-3′,

9th exon reverse primer: 5′-ATTTTTAAAACAAAATAGTTTTCATATTCGCAAC-3′.

All primers were synthesized from Keck biotechnology resource laboratory of Yale university. PCR products were collected using QIAquick gel extraction kit (Qiagen,Germany, Cat.No.28704) according to the manufacturer's instructions. Samples were sent to Keck biotechnology resource laboratory of Yale university for further sequencing. SnapGene Viewer was used for sequence alignment.

Site-directed Mutagenesis of D114N and D114A

The wild-type construct was previously subcloned into a pet28a vector.[81] A TEV-cleavage site was introduced using the megaprimer method to replace the FKBP12 protein in an N-terminal His-tag-TEV-FKBP12 expression construct.[119] Site directed mutagenesis to introduce the D114N mutation was carried out using the New England Biolabs Q5 Site-Directed Mutagenesis Kit using the forward and reverse primers:

5′-GTGTGCAAGCTTTATTTTAACTTGGAATTTAACAAACC-3′ and

5′-GGTTTGTTAAATTCCAAGTTAAAATAAAGCTTGCACAC-3′.

The PCR products were transformed into *E. coli* XL10-Gold cells and then the isolated plasmid DNA was sequenced to confirm successful cloning or mutagenesis.

The plasmid pET16::*CCDC111* containing the gene coding for WT PrimPol was used as template to generate the D114A mutation by the QuikChange Site-Directed Mutagenesis protocol (Stratagene). Oligonucleotides used to introduce the mutation were synthesized by Sigma Aldrich (St Louis, MO, USA): D114A-sense:

5′ GTGCAAGCTTTATTTTGCTTTGGAATTTAACAAACCTGCCAACCC 3′ and D114A-antisense:

5′ GGGTTGGCAGGTTTGTTAAATTCCAAAGCAAAATAAAGCTTGCAC 3′. The specific D114A mutation and the absence of other mutations in the *PRIMPOL* gene was confirmed by sequencing the recombinant plasmid that was kept in *E. coli* DH5α.

Protein Purification of D114A PrimPol

Note that the protocol for D114A varies slightly to the WT and D114N purification as described in these current methods. PrimPol D114A protein expression and purification was performed identically to the WT PrimPol[29] in *E. coli* BL21(DE3)-pRIL cells were transformed with the expression plasmid pET16::*CCDC111*-D114A. An overnight culture (50 ml) was incubated at 37 °C and used to inoculate 2 L of LB + ampicillin and chloramphenicol. Cells were grown at 30 °C up to $O.D._{600}$ 0.8, and then PrimPol production was induced with 1 mM IPTG (Ref. 367–93–1, Sigma Aldrich, St Louis, MO, USA) for 2,5 h. Cells were harvested at 12,000 g for 5 min at 4 °C, and the resulting bacterial pellet (≈3 g/L) was frozen at −20 °C. Approximately 6 g of cells were resuspended in 100 mL lysis buffer (Buffer A: 50 mM Tris-HCl pH 8, 1 M NaCl, 10% glycerol, 1 mM PMSF, 2 mM β-mercaptoethanol, 10 mM imidazole and 400 mM $AcNH_4$) and lysed by 10 min sonication pulses. The lysate was centrifuged at 27,000 g for 30 min at 4 °C to remove cell debris. Supernatant was incubated in batch with 2 mL HisPur Ni-NTA Resin (Ref. 88222; Thermo Scientific, Waltham, MA, USA) during 2 h at 4 °C. Resin was washed in batch with Buffer A and packed into a column. The resin was washed with 40 CV (column volume) of Buffer A containing 20 mM imidazole. Then washed with Buffer B (50 mM Tris-HCl pH 8, 10% glycerol, 1 mM PMSF, 2 mM β-mercaptoethanol) and 1 M

NaCl, and later on with Buffer B and 50 mM NaCl. Finally, the protein was eluted with buffer B supplemented with 50 mM NaCl and 200 mM imidazole. Fractions containing PrimPol were loaded in a Heparin Sepharose 6 Fast Flow (Ref. 17–0998 from GE Healthcare, Chicago, IL, USA), washed with 10 CV of Buffer B with 50 mM NaCl, then with 10 CV of Buffer B with 100 mM NaCl, and finally eluted with Buffer B and 1 M NaCl. Fractions containing 99% pure PrimPol were dialysed in a Slide-A-Lyzer Dialysis Cassette (Ref. 66380 from Thermo Scientific, Waltham, MA, USA) against Buffer C (25 mM Tris-HCl pH 8, 50% glycerol, 500 mM NaCl and 1 mM DTT) for at least 2 h at 4 °C and stored at −20 °C.

Full extension polymerase assay on a specific primer:template molecule

Full extension of a specific sequence substrate used a primer:

5′-CTGCAGCTGATGCGC-3′ and template:

5′ -GTACCCGGGGATCCGTACGGCGCATCAGCTGCAG-3′

(in a 1:2 ratio). Reaction mixtures (in 20 µL) contained Buffer R [50 mM Tris–HCl pH 7.5, 40 mM NaCl, 2.5% (w/v) glycerol, 1 mM DTT, 0.1 mg/mL BSA, 1 mM $MnCl_2$], 2.5 nM [γ-^{32}P]-labeled primer:template, 200 nM purified PrimPol and dNTPs (1, 10 100 µM). Reactions were incubated during 30 min at 30 °C, and stopped by adding 8 µL of formamide loading buffer, then loaded onto 8 M urea-containing 20% polyacrylamide sequencing gels of 30 cm long and run 2 h at 30 W. Following denaturing electrophoresis, primer extension was detected by autoradiography using AGFA CP-BU NEW Healthcare NV Medical X-RAY films blue (Ref. EWPKK, Mortsel, Belgium) and developed by a Kodak X-OMAT 2000 Processor (Rochester, NY, USA).

Primase assay using M13mp18 template

Single-stranded M13mp18 DNA (5 nM) was used as template to assess priming activity of PrimPol (400 nM) in the presence of indicated dNTPs (10 μM) and 16 nM [α-^{32}P]dGTP (250 μCi; 3000 Ci/mmol). Reaction mixtures (20 μL) in Buffer R, were incubated 30 min at 30 °C then stopped by adding 8 μL of formamide loading buffer, and loaded onto 8 M urea-containing 20% polyacrylamide sequencing gels (60 cm) and run 2 h at 50 W. After electrophoresis, products were detected by autoradiography using AGFA CP-BU NEW Healthcare NV Medical X-RAY films blue (Ref. EWPKK, Mortsel, Belgium) and developed by a Kodak X-OMAT 2000 Processor (Rochester, NY, USA).

Primase assay on specific oligonucleotide templates

Primase assays were carried out using the following unlabeled ssDNA oligonucleotides as templates: 3′-$(T)_{20}$G**TC**$C(T)_{36}$–5′ or 3′-$(T)_{20}$G**TC**$AGACAGCA(T)_{29}$–5′. The reaction mixture (20 μL) in Buffer R contained 1 μM ssDNA template, 400 nM PrimPol, 16 nM [γ-^{32}P]*ATP* or [α-^{32}P]dGTP and indicated dNTPs at 10 μM. Dimer synthesis experiment was measured using *ATP* as a 5′nucleotide (1, 10, 100 μM). Pre-made $_{3P}$*A*GT primer (10 μM) (synthesized by IDT, Coralville, IA, USA) was used to measure the elongation capacity of PrimPol variants. After an incubation time of 30 min at 30 °C, reactions were stopped adding 8 μL of formamide loading buffer. Synthesized primers were resolved in a 8 M urea-containing 20% polyacrylamide sequencing gels (60 cm) and run 2 h at 50 W. After electrophoresis, products were detected by autoradiography using AGFA CP-BU NEW Healthcare NV Medical X-RAY films blue (Ref. EWPKK, Mortsel, Belgium) and developed by a Kodak X-OMAT 2000 Processor (Rochester, NY, USA).

Electrophoretic mobility shift assay (EMSA)

5% polyacrylamide gels were cast and prerun at 150V in 0.5x tris-borate-EDTA (TBE) buffer for at least an hour at 4°C. Various total concentrations of WT or D114N (0-15 µM final) PrimPol were mixed with 3 nM primer:template substrate and incubated at room temperature for 40 minutes. Ficoll was added to a final percentage of 2% (wt/vol) and samples were loaded onto the polyacrylamide gel. Samples were run at 150V for 33 minutes at 4°C, with the gel cassette surrounded by ice. Reactions were visualized by phosphorimaging (Molecular Imager FX; Bio-Rad)

Differential scanning fluorimetry (Thermal shift assay)

The thermal shift assay was carried out using SYPRO orange dye to monitor protein unfolding. 5 µM PrimPol WT, D114N, AA354, or ZnF alone or in combination with each other, were mixed with additional combinations of 5 µM dsDNA (D20A/D45), and 10 µM $MgCl_2$ or $MnCl_2$ in 50 mM Tris pH 7.5, 300 mM NaCl, 0.5 µM TCEP, 5% glycerol, and a final concentration of 5x SYPRO orange dye in a 96-well PCR plate. The plates were placed in a BioRad CFX connect real time system and held at 4 °C for 5 minutes, then raised to 95 °C by 0.5 °C steps every 30 seconds. At each 30 second step, the relative fluorescence units were measured using the FAM channel. The derivatives of the melting curves were obtained from the Bio-Rad CFX Manager software and plotted against temperature. Each qPCR plate contained three (combined zinc finger and polymerase) or six (full length protein) technical replicates and to determine the melting temperatures, each value was averaged and the standard deviation was calculated. For the full length PrimPol thermal shift assays, two biological replicates were conducted.

3.3 Results

Identifying the D114N PrimPol mutation in an HIV+ patient with mitochondrial toxicity

Our previous studies identified a Polγ mutation in an HIV+ patient that appeared to confer an increased susceptibility to mitochondrial-associated toxicity due to NRTI-based antiretroviral therapy, in particular, 3TC (lamivudine).[117] We sought to identify possible mutations in the *PRIMPOL* gene in a cohort of HIV+ patients with ART-induced mitochondrial toxicity. The current PrimPol study is a subanalysis of mitochondrial toxicity study that enrolled participants at the Yale-New Haven Hospital from April 2011 to March 2013. The details of the study design for this cohort have been described previously.[118] In brief, for this PrimPol sub-study, cases (n = 13) comprised HIV-infected individuals on ART for at least 12 months with clinical and/or laboratory toxicities associated with mitochondrial toxicity. Cases were matched by age, sex, and race/ethnicity to HIV-negative controls (n = 19). All participants gave their written informed consent before participation in the study. The study protocol was approved by the Institutional Review Board of the Yale School of Medicine and all the research was performed in accordance to the relevant guidelines and regulations.

The demographic and HIV disease characteristics of study participants are illustrated in Fig. 3.1A. Archived peripheral blood mononuclear cells (PBMCs) of the participants were used for sequencing of *PRIMPOL (*also named as *Ccdc111)*. At enrollment CD4 counts, viral load, and duration of exposure to NRTI-containing therapies were extracted from their medical records. Upon Sanger sequencing of the conserved active site and zinc finger domain regions of *PRIMPOL* of the cohort, we observed heterogeneity

within the genomic sequence of a patient compared to healthy individuals (Fig. 3.1B, C). We concluded that this patient, referred to as individual 001, possesses a heterozygous g.340 transition mutation to a.340, translating into a D114N mutation at the protein level. Protein sequence alignment of PrimPol across different species and other individuals in the cohort emphasize the conservation of D114 (Fig. 3.1D). Given the importance of D114 as a key residue in the catalytic triad of PrimPol that coordinates a divalent metal ion (Fig. 3.1E)[78,120], we hypothesized that D114N is a hindering mutation that impairs the overall activity of PrimPol in individual 001.

A

Variable	HIV-infected participants with toxicity[a]	HIV-uninfected participants[b]
Gender (n)		
Female	4	9
Male	9	10
Age (range) (yrs)	38–58	38–66
Race (%)		
White non-Hispanic	2	5
White Hispanic	2	3
African American	9	11
Mean CD4 count (range) (count/µl)	697(86–1311)	NA[c]
Mean viral load (range) (copies/ml)	2,389(20–29297)	NA
Mean duration of exposure to nucleoside analogs (range) (yrs)	4.3 (0.5–10)	NA

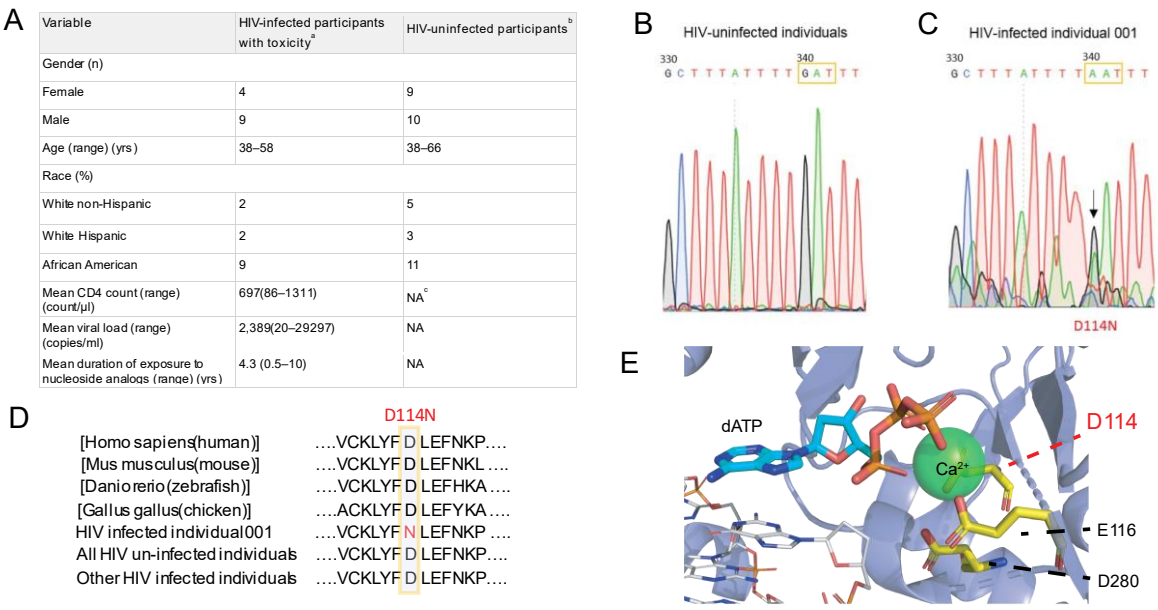


Figure 3.1. Identification of the D114N active site mutation in an HIV-positive patient experiencing nephrotoxicity.
A) Patient data table of a cohort of 13 HIV-infected patients experiencing toxicity under a tenofovir-containing therapy compared to 19 HIV-uninfected patients. *a.* n = 13, *b.* n = 19, *c.* NA, not applicable. B) Sanger sequencing results of the PrimPol gene in a healthy control and a C) HIV-infected individual 001 showing a heterozygous mutation of g.340->a.340 resulting in a D114N mutation in the protein. D) Protein sequence alignment of HIV patient 001 compared to the PrimPol gene of various species and to other patients in the cohort study. E) Active site of the PrimPol crystal structure demonstrating the role of D114 as a catalytic residue (yellow) that coordinates a divalent metal ion (green). PDB: 5L2X, Rechkoblit, O, et al. *Sci Adv* (2016).

The PrimPol D114N mutation is deficient in primase activity

With mounting evidence of the role of PrimPol as primarily a repriming enzyme[31,34,35,102,121], we first examined the ability of PrimPol to synthesize primers, using a M13 ssDNA template to validate our hypothesis that the D114N mutation would hinder PrimPol catalysis (Fig. 3.2A). In comparison to wild-type PrimPol, we did not observe nascent primer production with the mutant in a heterogeneous sequence context. We then assessed D114N activity in a single template context [3′(T_{20})-G**TC**AGACAGCA-(T_{29})5′] by providing [γ-^{32}P]ATP and the indicated dNTPs, to test sequentially the ability to initiate and elongate primers (Fig. 3.2B). Again, we observed a complete lack of primer initiation and the subsequent elongation by the D114N mutant, which parallels the null activity of a more drastic change of Asp[114] to alanine (D114A in Fig. 2B[120]; Calvo *et al*., 2019). Next, to boost the formation of the initial dimer, we provided higher concentrations of the rate limiting nucleotide ATP and [α-^{32}P]dGTP, on the template sequence 3′(T_{20})-G**TC**A-(T_{36})5′ (Fig. 3.2C). Even in the presence of high concentrations of the 5′ ATP nucleotide (100 μM), we failed to observe dimer formation using the D114N mutant protein. Although D114N PrimPol cannot form the initial dinucleotide for primer synthesis, it may retain the ability to elongate preexisting primers. To test that, we supplied the reaction with a synthetic 3-mer primer with a 5′-triphosphate, (3p*A*GT), which has previously been demonstrated to be important for the binding of PrimPol to the initiated primer.[122] In these conditions, extension of the primer was efficiently carried out by wild type PrimPol but not by the D114N mutant (Fig. 3.2D). Lastly, we examined the conventional DNA polymerase activity of the D114N PrimPol mutant, by using a mature 5′ radiolabeled ssDNA 15-mer (which is a valid primer despite the lack of the 5′ triphosphate) annealed to a ssDNA 34-

mer. Although we observed a prominent reduction in polymerase activity, D114N still retained some ability to incorporate dNTPs onto a primer (Fig. 3.2E). Even taking into consideration that the D114N mutant incorporates nucleotides to lesser degree compared to wild-type, the residual catalytic activity was surprising in contrast to the complete loss of primase function.

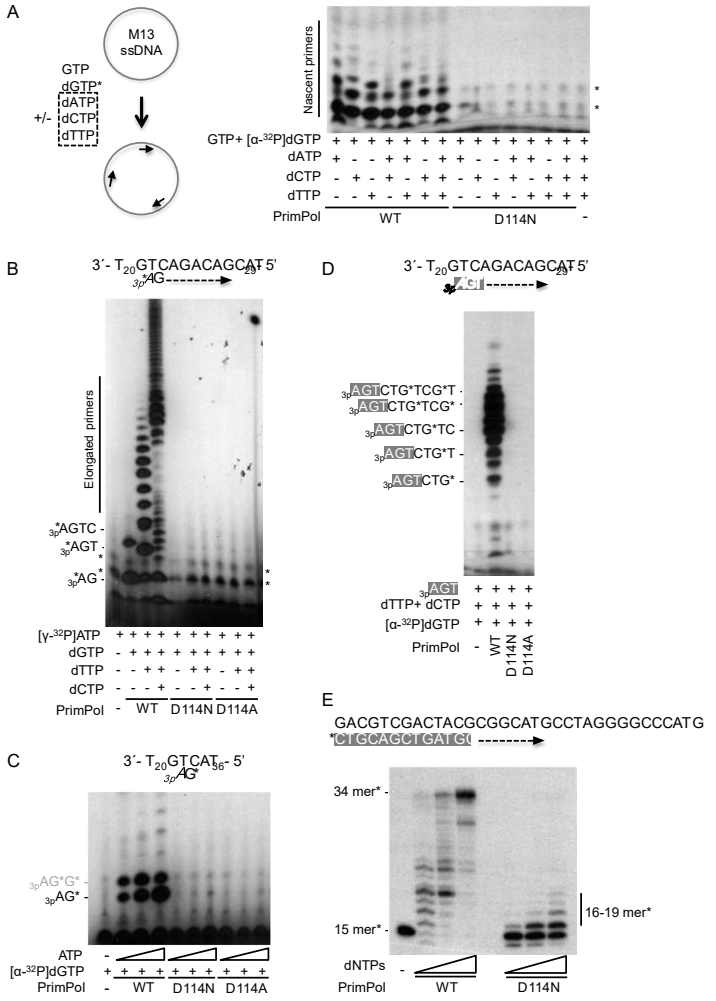


Figure 3.2. The PrimPol D114N mutation is deficient in primase activity.
A) WT, D114N, or D114A PrimPol and Pol γ are mixed with the M13 ssDNA plasmid. In the event that PrimPol primes the plasmid, Pol γ is able to extend the mature primer. The D114N mutant cannot form mature primers for Polγ extension. B) The initiation and elongation of primers by PrimPol alone with limiting 5' nucleotide. The D114N mutation is unable to initiate a primer compared to WT. C) Dimer synthesis of PrimPol. Using the preferred priming sequence of 5'-GTCA-3', the D114N PrimPol is unable to form the dinucleotide for primer initiation. D) PrimPol extension of a supplied primer. The ability of the D114N mutant to extend a supplied primer with a 3'-triphosphate was assayed. Compared to the WT, the D114N mutant is unable to utilize the supplied 3'-triphosphate primer as a substrate. E) Full elongation by PrimPol under standard polymerase conditions. When supplied with all dNTPs and a radiolabeled 15-mer annealed to a templating DNA, the WT is able to fully elongate the primer while the D114N mutant is able to catalyze a limited number of insertions.

The D114N mutation retains catalytic activity but drastically reduces the kinetics of polymerization

In order to further probe the magnitude of the hindering D114N mutation on catalytic activity, we measured the kinetics of natural nucleotide incorporation under pre-steady-state burst and single-turnover conditions to compare the mutant and wild-type proteins. Under burst conditions where the DNA substrate is in slight excess of protein, there have been many examples of DNA polymerases exhibiting biphasic kinetics.[123-125] These kinetic observations indicate that in the overall kinetic mechanism of catalysis, product release of the DNA from the protein is slower in relation to the chemical catalysis of incorporation. We previously established that PrimPol exhibits a burst phase[81], and we observed that the D114N mutation displayed similar biphasic kinetics (Fig. 3.3A, B). Compared to wild type, the burst rate of the mutant was approximately 60-fold slower, and the linear steady-state rate approximately 20-fold slower (Table 3.1).

To further examine the catalytic mechanism, we carried out single turnover experiments to calculate the k_{pol}, the maximal rate of incorporation, K_d, the apparent binding affinity of the incoming nucleotide, and k_{pol}/K_d, the overall incorporation efficiency (Fig. 3.3C, D). Single-turnover experiments provide a clearer comparison of the chemical catalysis steps that may be obscured by the linear phase under burst conditions. Importantly, the binding of the incoming nucleotide may also be compromised because the D114N mutant could potentially affect divalent metal coordination, which may appear as a change in the K_d. Overall, we saw a 100-fold decrease in the overall incorporation efficiency (k_{pol}/K_d) with the mutant (Table 3.1). Examining the individual kinetic parameters that define the incorporation efficiency revealed that this 100-fold difference

was reflected solely in the k_{pol}, while the K_d for the incoming nucleotide remained nearly identical between the mutant and wild-type proteins. This drop of k_{pol} in D114N during polymerization activity explains perfectly its incapacity of primer synthesis.

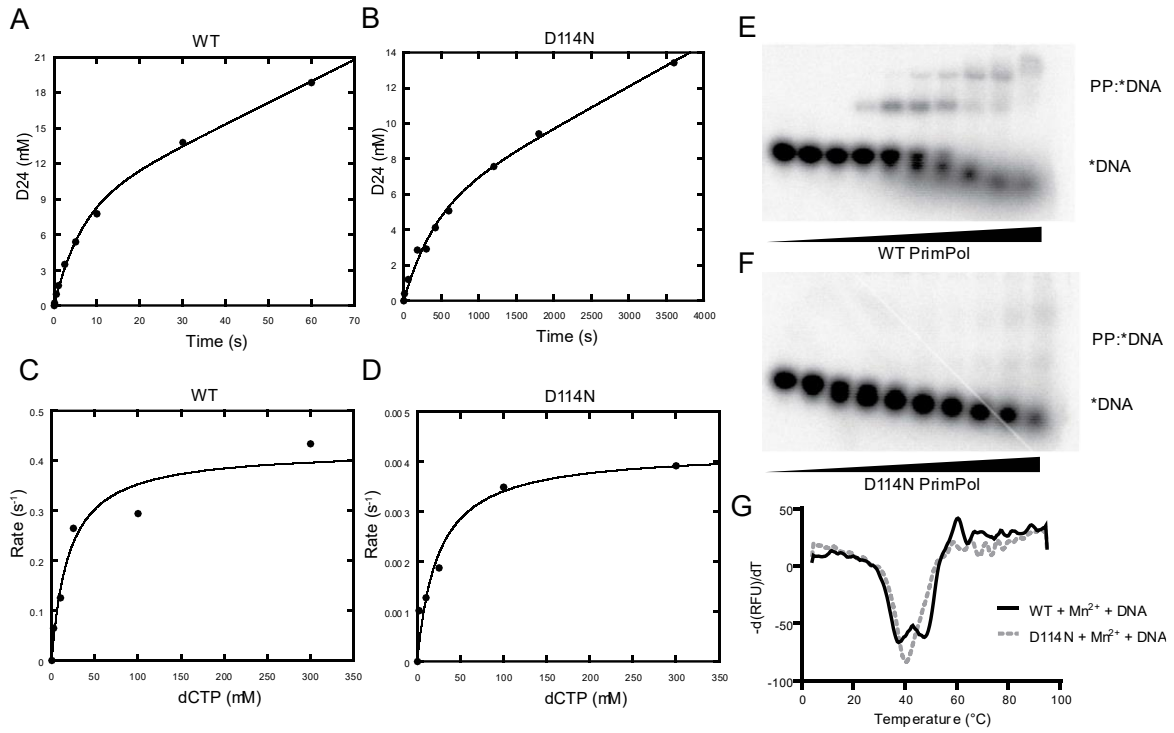


Figure 3.3. Kinetic and biochemical characterization of the PrimPol D114N mutation. A) Pre-steady state burst kinetics of WT and B) D114N PrimPol. The concentration of a 23-mer annealed to a 45-mer (DNA:DNA) was held in slight excess to protein and the incorporation of dCTP was plotted against time and fit to a burst equation. C) K_d curves of WT and D) D114N mutant PrimPol. Rates of dCTP incorporation were measured under single-turnover conditions where PrimPol is in excess of DNA:DNA and plotted against varying concentrations of incoming dCTP. E) Electrophoretic mobility shift assays of WT and F) D114N PrimPol with radiolabeled dsDNA substrates in the presence of Mn^{2+} where the upper bands correspond to the formation of the PrimPol:DNA complex. G) Comparison of the stability of WT and D114N PrimPol by a thermal shift assay using SYPRO orange. The derivative of the melting curves were plotted and the local minima correspond to melting temperatures of the protein.

Burst			
	k_{ss} (s^{-1})	k_{burst} (s^{-1})	
WT	$9.9 \times 10^{-3} \pm 4.8 \times 10^{-3}$	$1.5 \times 10^{-1} \pm 6.8 \times 10^{-2}$	
D114N	$4.7 \times 10^{-4} \pm 1.1 \times 10^{-4}$	$2.5 \times 10^{-3} \pm 5.6 \times 10^{-4}$	
Single Turnover			
	K_d (μM)	k_{pol} (s^{-1})	k_{pol}/K_d (μM^{-1} s^{-1})
WT	19.7 ± 7.5	$4.2 \times 10^{-1} \pm 4.2 \times 10^{-2}$	2.1×10^{-2}
D114N	23.9 ± 8.7	$4.2 \times 10^{-3} \pm 4.2 \times 10^{-4}$	1.8×10^{-4}

Table 3.1. Summary of pre-steady state kinetics of wild-type (WT) and D114N PrimPol. Burst pre-steady state kinetic parameters for both WT and D114N PrimPol at at 25°C were determined by fitting the time course data to the burst equation: $[product] = A(1 - e^{-k}obs^{t}) + A(k_{ss})(t)$, where A is the burst phase amplitude, k_{obs} is the observed single exponential rate, k_{ss} is the steady-state rate, and t is the time.
Single turnover kinetic parameters were determined by fitting the time course data with the following single exponential equation: $[product] = A(1 - e^{-k}obs^{t})$, where A is amplitude and k_{obs} is the observed single exponential rate, and t is the time. The single exponential rates were then plotted against each concentration of [dNTP] using a quadratic equation in order to extract the k_{pol}, the maximal rate of incorporation, K_d, the apparent binding constant for the incoming nucleotide, and k_{pol}/K_d, the overall efficiency for nucleotide incorporation. The errors represent the standard error values of the parameters that corresponds to a confidence level of 68.3%, or to one standard deviation.

DNA binding ability and protein stability are hindered by the D114N mutation

We predicted that the D114N mutation was unlikely to have significant effects on DNA binding or overall structure, as described for a catalytically inactive mutation (D114A) of the same residue[120]; however, it could not be discounted that the change of Asp^{114} to Asn could have a significant effect on the stability of the whole protein, as described in other studies on PrimPol mutants.[126]

The binding of wild-type and mutant proteins to a template/primer DNA was compared through electrophoretic shift mobility assays (EMSAs). While the wild-type protein exhibited a similar DNA binding affinity to that described in previous work, a striking reduction in the DNA-binding capabilities of the mutant was evident (Fig. 3.3E, F). Taking into consideration that DNA binding and correct positioning of the template DNA is a prerequisite for nucleotide incorporation activity and that EMSAs may be unable to detect weak binding complexes[127], we concluded that the D114N mutation diminished the DNA binding affinity of the protein but did not completely attenuate it. The reduction in activity or dsDNA binding may stem from the importance of Asp^{114} as a catalytic residue or a contributor to key interactions that coordinate with the DNA or stabilize the protein. In order to examine the latter effects of the mutation, we utilized differential scanning fluorimetry (DSF)[128,129] to observe changes in the melting temperatures of the wild-type and mutant proteins as shown in Fig. 3.3. In comparison of the melting temperatures of wild-type to D114N PrimPol in the presence of dsDNA and Mn^{2+}, we observed the presence of two melting temperatures in the wild-type protein and one in the mutant (Fig. 3.3G). The presence of the 47.17 °C ± 0.26 melting peak in the wild-type compared to the singular 40.17 °C ± 0.26 with D114N PrimPol appeared to indicate that the D114N

mutation caused some instability within the protein (Table 3.2). In addition, due to the proximity of the melting temperature of the mutant to physiological temperature (37 °C), we conducted our kinetic assays at 25 °C.

We further analyzed if the two peaks found in wild-type PrimPol correspond to the polymerase domain (AA354) and the zinc finger domain (ZnF).[35] We conducted the thermal shift assay with the isolated polymerase domain, the zinc finger domain, or their combination in the presence of either Mg^{2+} or Mn^{2+} (Fig. 3.4A, B). The lower melting temperature peak appears to correspond to the ZnF and the more stable peak to the polymerase domain (Table 3.2). Taking this information into account in reference to the D114N mutant, the absence of the higher temperature minimum may represent the destabilization of the polymerase domain. Interestingly, further characterization demonstrates that both WT and D114N PrimPol were slightly more stable in the presence of Mg^{2+} than Mn^{2+} (Fig. 3.4C, D). Moreover, addition of dsDNA stabilized the wild-type PrimPol and Mn^{2+} complex, possibly by stabilizing the polymerase domain, while there was no effect on the Mg^{2+} complex (Fig. 3.4E, F). The dsDNA stabilization effect was notably absent in the mutant, which may be reflective of the reduced ability of the mutant to bind dsDNA as observed in the EMSAs in Fig. 3.3F (Fig. 3.4G, H).

Exp. 1A	WT + Mg^{2+}	WT + Mn^{2+}	D114N + Mg^{2+}	D114N + Mn^{2+}	WT + Mg^{2+} + DNA	WT + Mn^{2+} + DNA	D114N + Mg^{2+} + DNA	D114N + Mn^{2+} + DNA		
Min 1	40.08 ± 0.20	36.08 ± 0.20	42.83 ± 0.26	38.67 ± 0.26	40.00 ± 0.00	37.83 ± 0.41	42.58 ± 0.20	40.17 ± 0.26		
Min 2	48.83 ± 0.26	47.83 ± 0.41			48.92 ± 0.20	47.17 ± 0.26				
Exp. 1B	WT + Mg^{2+}	WT + Mn^{2+}	D114N + Mg^{2+}	D114N + Mn^{2+}	WT + Mg^{2+} + DNA	WT + Mn^{2+} + DNA	D114N + Mg^{2+} + DNA	D114N + Mn^{2+} + DNA		
Min 1	39.92 ± 0.20	36.25 ± 0.27	42.75 ± 0.27	38.75 ± 0.27	40.42 ± 0.20	38.00 ± 0.55	42.67 ± 0.26	40.00 ± 0.00		
Min 2	48.83 ± 0.26	47.83 ± 0.41			48.92 ± 0.20	47.17 ± 0.26				
Exp. 2	1-354 + Mg^{2+}	1-354 + Mn^{2+}	ZnF + Mg^{2+}	ZnF + Mn^{2+}	1-354 + Mg^{2+} + DNA	1-354 + Mn^{2+} + DNA	ZnF + Mg^{2+} + DNA	ZnF + Mn^{2+} + DNA	1-354 + ZnF + Mg^{2+} + DNA	1-354 + ZnF + Mn^{2+} + DNA
Min 1	47.33 ±0.29	47.33 ± 0.29	35.50 ± 0.00	33.50 ± 0.00	47.17 ± 0.29	47.33 ± 0.29	36.00 ± 0.00	34.00 ± 0.00	38.67 ± 0.58	38.50 ± 0.71
Min 2									35.15 ± 0.00	34.94 ± 0.29

Table 3.2. Melting temperatures of WT and D114N PrimPol dependent on the presence of Mg^{2+} or Mn^{2+} and DNA.
WT or D114N PrimPol (5 μM) was mixed with either $MgCl_2$ or $MnCl_2$ (10 μM) and dsDNA (5 μM) and the unfolding of the protein was monitored through fluorescence by SYPRO orange. The derivatives of the relative fluorescence units were plotted against temperature, the local minima identified, and the corresponding temperatures were recorded. Numbers are shown as the mean ± SD. For experiment 1A and 1B, n=6, and for experiment 2, n=3.

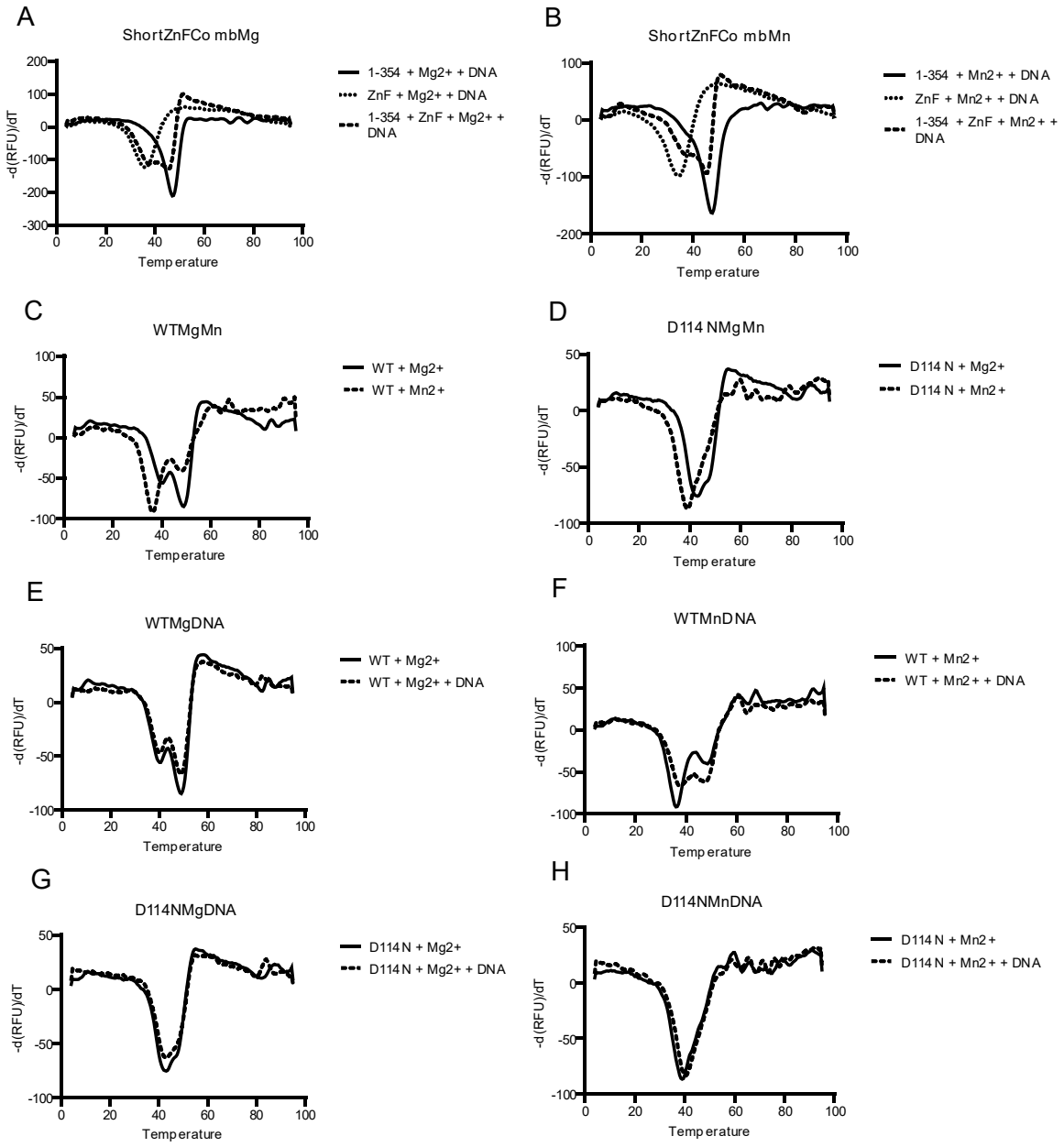


Figure 3.4. Thermal shift assay of WT and D114N PrimPol dependent on the presence of Mg^{2+} or Mn^{2+} and DNA (see Fig. 4G).
WT or D114N PrimPol (5 µM) was mixed with either $MgCl_2$ or $MnCl_2$ (10 µM) and dsDNA (5 µM) and the unfolding of the protein was monitored through fluorescence by SYPRO orange (A-H). A) Comparison of the polymerase domain (1-354) alone, zinc finger domain alone, and a stoichiometric mixture of both complexes in the presence of DNA with either Mg^{2+} or B) Mn^{2+}. C) Comparison of the melting temperatures with Mg^{2+} to Mn^{2+} present without DNA of WT or D) D114N PrimPol. E) Observing the effects of the presence of DNA on the melting temperature of WT PrimPol with either Mg^{2+} or F) Mn^{2+}. G) Observing the effects of the presence of DNA on the melting temperature of D114N PrimPol with either Mg^{2+} or H) Mn^{2+}.

3.4 Discussion

Structural basis for the loss of primase activity but residual polymerase activity in D114N PrimPol

It is vital to understand how mutations present in individuals may contribute to disease or in this particular case, how PrimPol mutations could influence the phenotype predisposing towards mitochondrial toxicity. Biochemical studies investigating how PrimPol variants function at a molecular level is beneficial not only in considering treatment options and strategies in patients, but also contributes to the basic mechanistic understanding of the enzyme. This study is the first to identify a PrimPol mutation in an individual in the context of HIV ART: a single G to A transition in the PrimPol gene implicated an Asp to Asn mutation at amino acid position 114. Asp^{114} is part of the invariant motif A (**D**x**E**), and is one of the 3 catalytic carboxylates forming the active site of AEP-like enzymes as PrimPols.[29,99] Mutation of any of these residues to alanine abolishes PrimPol activities.[120] Like the other two carboxylates, Asp^{114} is involved in metal coordination at the active site, which is critical for nucleotide incorporation.

Our initial hypothesis was that the D114N mutation would similarly compromise both primase and polymerase catalytic activities. However, we were surprised that PrimPol mutant D114N was able to retain some polymerase activity, but completely lacked primase activity. Some remaining activity can be explained as the substitution of a charged residue as Asp114 to Asn is more conservative than its change to a hydrophobic Ala. Kinetic analysis of the D114N mutant in DNA polymerase assays indicated a strong reduction in the k_{pol} which suggests a deficient metal coordination at the active site, which compromises (but still allows) nucleotide incorporation. Interestingly, we observed that a change in K_d

is not observed with the Asn mutant. We had predicted that loss of the metal B ion in the active site would affect the K_d for the incoming nucleotide, since the metal ion is partially responsible for binding the phosphate moieties of the incoming nucleotide. This evidence contradicts our hypothesis that metal coordination is altered in the mutant. It is possible that the mutant is able to coordinate the metal to allow binding of the incoming nucleotide, but the optimal geometry of the active site may be altered to decrease the rate of catalysis.

On the other hand, a similar drop of k_{pol} in PrimPol D114N provokes a more dramatic consequence for dimer formation to start primer synthesis. This could be explained by the stricter and perhaps specific metal coordination required for binding the two initiating nucleotides. The crystal structure of PrimPol in priming mode with two catalytically competent metal ions would be critical in understanding the nuances of PrimPol activity when compared to the current structure.[78]

The extensive comparison of wild-type and D114N PrimPol through the thermal shift assay also revealed interesting phenomena. Although Mn^{2+} has been demonstrated to be crucial for priming activity, we observed that PrimPol tended to be more stable in the presence of Mg^{2+}. Assuming that PrimPol may utilize either Mg^{2+} or Mn^{2+} in the cell[120,130], the stabilization/destabilization effect may favor a particular mode for PrimPol. In agreement with this idea, a recent paper[131] demonstrated that in the presence of Mn^{2+}, a conformational transition step from non-productive to productive PrimPol:DNA complexes limits the enzymatic turnover, whereas, in the presence of Mg^{2+}, the chemical step becomes rate limiting. The appearance of two melting minima for wild-type PrimPol, which correspond to the ZnF and the catalytic polymerase domain, may also aid in understanding the primase mechanism of PrimPol. Because there are two distinct peaks for

each domain, it is possible that the two domains may be able to function somewhat independently from one another at a structural level. It is worth mentioning that PrimPol lacking the ZnF is polymerase competent but is not able to start primer synthesis.[34,35] The lack of two melting temperature minima for D114N PrimPol reveals that the mutation destabilizes the polymerase domain. Interestingly, analysis of the Y89D mutation in PrimPol shows a similar destabilization.[126] To discern whether the single amino acid mutation causes the destabilization of the catalytic domain, or if coordination of a metal divalent ion assists in stabilization of the protein would require further study.

It is noteworthy that the patient identified with the D114N PrimPol mutation was heterozygous for the mutation. It may be interesting to determine if partially reducing the total amount of functional PrimPol is substantial enough to predispose a patient to off-target toxicity, considering the knockdown of PrimPol in our cell lines was highly efficient. In addition, because many factors could contribute to the complex toxicity caused by NRTIs, it is difficult to attribute this patient's toxicity solely to this active site mutation. The potential for other polymerases or other molecules involved in oxidative stress or maintaining dNTP pools to contribute to mitochondrial toxicity may explain why we do not observe other mutations in PrimPol in our patient cohort.[95]

Chapter 4. Structural investigation of 2-naphthyl phenyl ether compounds that target HIV replication

This chapter is an excerpt from:

Duong VN, Ippolito JA, Chan AH, Lee WG, Spasov KA, Jorgensen WL, Anderson KS. Structural investigation of 2-naphthyl phenyl ether inhibitors bound to WT and Y181C reverse transcriptase highlights key features of the NNRTI binding site. Protein Sci. 2020 Jul 8;29(9):1902–10. doi: 10.1002/pro.3910. Epub ahead of print. PMID: 32643196; PMCID: PMC7454559.

4.1 Introduction

Human immunodeficiency virus (HIV) is a worldwide health issue, with 37.9 million people infected as of 2018.[132] Untreated HIV leads to acquired immunodeficiency syndrome (AIDS), rendering the infected person susceptible as a host to other deadly diseases. Although there is not a cure for HIV, life-long drug-based therapies exist to suppress the virus to prevent progression of HIV into AIDS. Highly active anti-retroviral therapy (HAART) is a combination therapy consisting of various drugs that target different stages of the HIV-1 infection life cycle, improving the quality of life and life expectancy of patients.[86,133,134] Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are vital components of HAART that target HIV-1 reverse transcriptase (RT), which is responsible for the production of viral DNA.[41,135] Unlike nucleoside reverse transcriptase inhibitors (NRTIs) that act as replication chain terminators, NNRTIs inhibit RT activity by binding an allosteric site approximately 10 Å from the active site mediated through a conformational change that alters the rate limiting step in chemical catalysis.[66,68,136]

HIV-1 is highly mutable and drug-resistant strains have emerged because RT is error-prone, necessitating the need to develop next generation drugs that remain efficacious

in the rise of resistance.[137,138] One such mutation is Y181C in the allosteric pocket of RT, which has been demonstrated to reduce the potency of first-generation NNRTIs such as nevirapine or delavirdine.[139-141] To this end, second generation NNRTIs such as etravirine (ETV) and rilpivirine (RPV) were developed that are effective against WT and an array of resistance-associated mutations.[135,142] Even with the improved second generation of compounds, further development of NNRTIs is needed as compounds such as ETV and RPV have poor pharmacological properties and the development of resistant mutations have been reported.[143-146]

In our previous work, we identified a set of catechol diether compounds with a 7-cyano-2-naphthyl substituent (2-naphthyl phenyl ethers) as potential drug candidates through computational modeling using the *BOMB* and *MCPRO* programs.[147,148] After biochemical and structural characterization, we observed that our synthesized compounds demonstrated two distinct binding modes signified by the orientation of the 2-naphthyl ring (Figure 4.1). The nitrile group of our parent compound (Figure 4.1A) projects toward Y181 while the nitrile of a derivative compound (Figure 4.1B) projects towards W229, likely due to a 1-position substitution. Interestingly, compound **1** has an EC_{50} of 22 and 2600 nM to WT and Y181C respectively, while compound **2**, containing 1-methyl and 4-chloro substituents has an EC_{50} of 6.2 and 58 nM (Table 4.1). We hypothesized that the differences observed in Y181C EC_{50} values are determined by whether the 2-naphthyl ether largely interacts with W229 or Y181.

In this study, we have solved the crystal structures of Y181C RT in complex with a series of 2-naphthyl phenyl ether compounds (Figure 4.2, compounds **1** and **2**) and compare them to wild type structures. These structures explain the mechanism by which

certain 2-naphthyl ethers may overcome the prevalent Y181C RT mutation. In addition, we solved structures of RT with derivatives of the parent compound (Compounds **3**-**6**) to improve efficacy against the Y181C mutation, based on our previous results. Our structural analysis highlights key interactions that should be considered in the development of 2-naphthyl ethers and future NNRTIs that combat drug resistant mutants of HIV-1 while maintaining optimal pharmacokinetic properties.

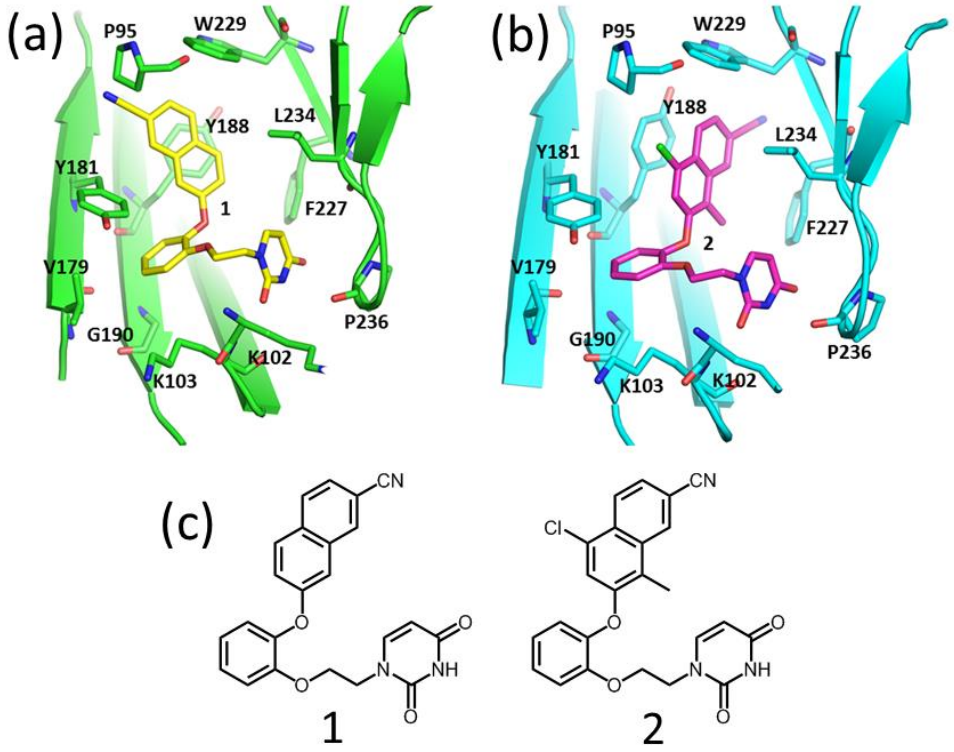


Figure 4.1. Structures of compound 1 and 2 with WT RT.

Crystal structures of compound 1 (a) colored in yellow and compound 2 (b) colored in magenta in complex with to WT RT. (c) Chemical structures of compounds 1 and 2.

Compound	WT	Y181C	K103N/Y181C
1[a]	22	2600	4000
2[a]	6.2	58	280
3[a]	7.8	60	890
4[a]	1.9	28	410
5[a]	18	900	1200
6	23	2100	1700

Table 4.1. Inhibitory Activity (EC_{50}, nM) for HIV-1 in MT-2 Cell Assays.

[a]See ref. [148]

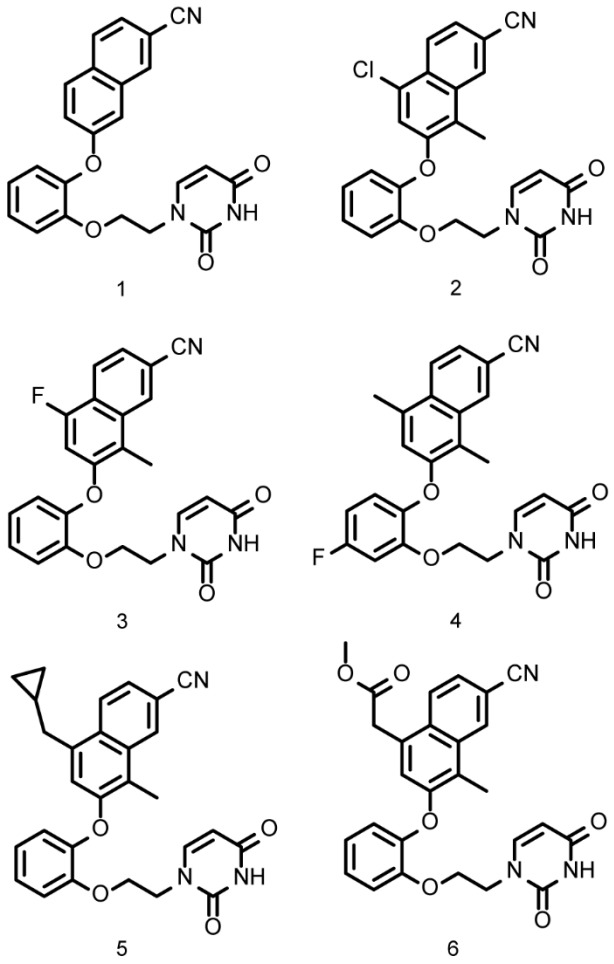


Figure 4.2. Chemical structures of 2-naphthyl phenyl ether compounds. Chemical structures of Compounds 1-6.

4.2 Materials and Methods

Expression, Purification, and Crystallization

WT and Y181C RT was expressed, purified, and crystallized as described previously from a plasmid obtained from Stephen Hughes, Paul Boyer, and Andrea Ferris expressed in expressed in *E. coli* BL21(DE3) pLysS cells.[124,149,150]

Data collection, processing, and structure determination and refinement

Apo crystals used for inhibitor soaking were grown by hanging drop vaporization in 50 mM imidazole or HEPES (pH 6.5-7.0), 16–20% (wt/vol) PEG 8000, 100 mM ammonium sulfate, 15 mM magnesium sulfate, and 5 mM spermine with an initial protein concentration of 10 mg/mL. Inhibitor was added to final concentration of 0.5 mM to drops containing suitable crystals for overnight soaking. Crystals of the WT:3 complex were obtained by cocrystallization, in which 20 mg/mL of protein was first incubated with 0.5 mM inhibitor on ice for 1 h before crystallization. Cocrystals of WT:3 were then grown in 50 mM MES pH 6.0, 14% (wt/vol) PEG 8000, 100 mM ammonium sulfate, 15 mM magnesium sulfate, and 5 mM spermine. For cryoprotection, all crystals were transferred to a solution containing 27% (vol/vol) ethylene glycol and 0.5 mM inhibitor before being flash frozen in liquid nitrogen. X-ray diffraction data sets on frozen crystals were collected at the Advanced Photon Source on NE-CAT beam line 24-ID-E. Data sets were processed with either HKL2000[151] or XDS.[152] Phases were determined by molecular replacement with the program PHASER[153] using the previously solved structure PDB 5TER[148] as the initial search model. All model building into electron density was performed with COOT[154] and the structures were refined using Phenix Refine. [155155] For each refinement, 5% of all reflections were omitted and used for the calculation of R_{free}. Successive rounds

of simulated annealing, XYZ coordinate, and individual B-factor refinement were performed until acceptable R factors, geometry statistics, and Ramachandran statistics were achieved. Data collection, diffraction, and refinement statistics can be found in Table 4.2. Iterative build, composite omit electron density maps shown in Figure 4.3 were calculated using Phenix Autobuild.[156] All atomic coordinates and structure factors and have been deposited in the Protein Data Bank with PDB ID codes 6X47, 6X49, 6X4A, 6X4B, 6X4C, 6X4D, 6X4E, and 6X4F. All figures were prepared by PYMOL. Crystallography programs were compiled by SBGrid.[157]

MTT Cellular Viability Assays

A tetrazolium-based colorimetric assay was used to determine the *in vitro* anti-HIV-1 activity of compounds in MT-2 cells as described previously.[158]

	WT:1	Y181C:1	Y181C:2	WT:3	WT:4	WT:5	WT:6	Y181C:6
PDB Code	6X47	6X49	6X4A	6X4B	6X4C	6X4D	6X4E	6X4F
Data collection								
X-Ray Source	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-E
Wavelength, Å	0.97915	0.97915	0.97915	0.97915	0.97915	0.97915	0.97915	0.97915
Space Group	C2	C2	C2	C2	C2	C2	C2	C2
Cell dimensions								
a, b, c (Å)	163.06, 74.11, 108.43	162.14, 73.99, 107.86	161.77, 73.98, 107.90	221.25, 67.40 102.91	162.19, 74.11, 108.36	162.74, 74.27, 108.73	162.73, 74.08, 108.39	162.00, 73.71, 108.48
α, β, γ (˚)	90, 99.48, 90	90, 99.34, 90	90, 100.05, 90	90, 107.77, 90	90, 100.13, 90	90, 100.02, 90	90, 99.93, 90	90, 100.01, 90
Resolution (Å)	50 - 2.75 (2.80-2.75)	50 - 2.75 (2.80-2.75)	50 - 2.53 (2.57-2.53)	50 - 2.50 (2.54-2.50)	50 - 2.85 (2.90-2.85)	50 - 2.65 (2.71-2.65)	50 - 2.60 (2.66-2.60)	50 - 2.72 (2.78-2.72)
R_{merge}	0.071 (0.538)	0.074 (0.477)	0.056 (0.508)	0.089 (0.520)	0.057 (0.517)	0.054 (0.595)	0.054 (0.614)	0.060 (0.610)
$I / \sigma I$	31.3 (3.5)	35.7 (3.7)	34.9 (3.4)	27.8 (3.2)	32.3 (3.3)	19.0 (2.2)	20.2 (2.4)	16.1 (2.3)
Completeness (%)	99.3 (99.2)	99.8 (100.0)	99.7 (99.6)	99.9 (100.0)	99.8 (99.7)	99.7 (99.9)	99.7 (99.6)	99.5 (99.9)
Redundancy	3.4 (3.5)	3.7 (3.7)	3.7 (3.7)	3.8 (3.8)	3.7 (3.8)	3.7 (3.8)	3.7 (3.8)	3.7 (3.6)
Refinement								
Resolution (Å)	43.4 - 2.77	43.3 - 2.75	38.3 - 2.54	41.4 - 2.50	43.2 - 2.86	43.4 - 2.65	43.3 - 2.60	43.1 - 2.72
No. reflections	32,413	33,075	41,528	50,154	29,301	37,292	39,236	33,958
R_{work} / R_{free}	0.2340, 0.2671	0.2283, 0.2509	0.2317, 0.2685	0.2252, 0.2591	0.2201, 0.2584	0.2236, 0.2582	0.2263, 0.2594	0.2225, 0.2518
No. atoms								
Protein	7718	7672	7732	7708	7692	7734	7734	7730
Inhibitor	30	30	32	32	33	35	36	36
Ion	0	5	10	21	0	10	15	5
Water	0	7	44	44	0	71	27	53
B -factors								
Protein	78.35	82.1	73.76	56.01	82.93	62.61	59.93	64.14
Inhibitor	69.26	62.96	60.09	51.6	71.07	52.45	50.07	60.67
Ion		116.85	98.89	81.21		90.82	87.1	104.98
Water		56.63	54.65	48		44.59	38.27	49.05
R.m.s deviations								
Bond lengths (Å)	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.002
Bond angles (˚)	0.64	0.61	0.61	0.68	0.63	0.48	0.5	0.5
Ramachandran								
Favored (%)	97.44	97.23	96.7	97.3	97.1	96.8	97.3	97.2
Allowed (%)	2.56	2.77	3.3	2.7	2.9	3.2	2.7	2.8
Outliers (%)	0	0	0	0	0	0	0	0

Table 4.2. Crystallographic statistics table.

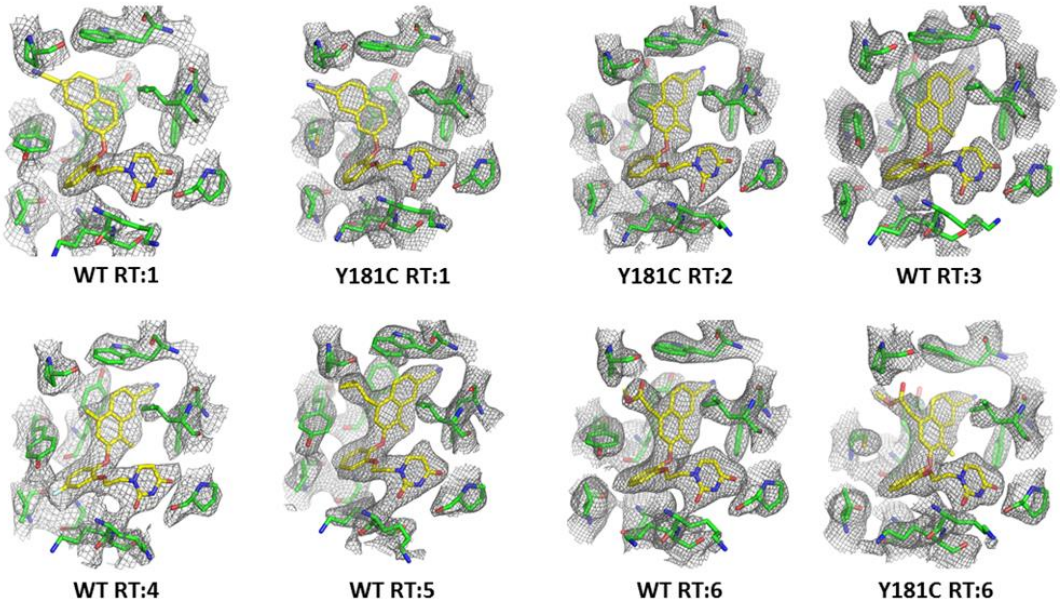


Figure 4.3. σA –weighted 2mFo -Fc electron density maps.
RT-inhibitor complexes determined for this study superimposed upon final refined structures (protein in green, inhibitor in yellow). The maps are contoured at 1.0 σ. Coordinates for the compounds were omitted from the model for the calculation of iterative-build omit maps from the initial structure factors

4.3 Results

To deduce the structural consequences that result in the pronounced differences in activity between WT and Y181C RT observed for the 2-naphthyl phenyl ethers, we have determined the structures of compounds **1**, **2**, and **6** (Fig. 4.2) bound to the Y181C mutant of RT using x-ray crystallography (Fig. 4.4). We have also solved the crystal structures of compounds **3**, **4**, **5**, and **6** bound to WT RT to examine the effects that various ring substitutions have on the binding of the 2-naphthyl scaffold (Fig. 4.5). Additionally, we have redetermined the structure of 1 bound to WT RT to a higher resolution than the structure reported previously[148] for a more accurate comparison with the Y181C RT:1 complex structure presented here. Crystal structures were solved to resolutions of 2.50 Å – 2.86 Å and refined to final Free R-factors between 0.2509 and 0.2685 and displayed good stereochemical and geometry statistics (Table 4.2). All structures have been deposited into the PDB. Calculated 2*F*o-*F*c and *F*o-*F*c omit electron density maps allowed for unambiguous modelling of inhibitors in both WT and Y181C protein structures (Figure 4.3). Consistent with previously reported structures of NNRTI complexes, no significant changes in the overall structure of RT were observed in either the WT or Y181C structures.

The structure of compound **1** bound to Y181C RT shows **1** binds in a similar conformation as it does in WT RT. The cyano group of the naphthyl ring projects over Y181 and away from W229. As seen in the superposition with the WT structure, when bound to Y181C RT, both the naphthyl and catechol rings of **1** move 1.4 Å towards the empty space created by the vacated Y181 side chain (Figure 4.4A). As a result of this shift, the naphthyl ring of **1** moves further away from W229 compared to its position in WT RT. Interestingly, the side chain of Y188 in the mutant compensates for this movement by

turning towards C181 along with compound **1**. This compensatory movement allows Y188 to re-optimize its stacking interaction with the displaced **1** in the Y181C RT mutant. Likewise, compound **2** binds to the Y181C mutant in the same overall conformation previously observed in WT RT, in which the cyano group of the naphthyl ring points toward the W229 side chain.[148] However, unlike compound **1**, the superposition of the WT and mutant structures clearly shows the binding of compound **2** is identical in both WT and Y181C RT (Figure 4.4B). This suggests that the interaction with W229, which results primarily from the different conformation of the 2-naphthyl ring of **2**, contributes much more to the binding of this compound than it does for compound **1**.

The structures of 2-naphthyl ether analogs **3** and **4** bound to WT RT have also been determined. The superposition of these structures with the compound **2** complex show all three compounds bind to WT RT very similarly (Figure 4.6). The fluorine substitution at the 4-position in **3** and the methyl group substitution on **4** make no impact on binding conformation of the 2-naphthyl ring compared to **2**, nor do they significantly alter activity of the compounds against either WT or mutant HIV (Table 4.1).[148] The nearly identical bound conformations of **2**, **3**, and **4** most likely result from the shared methyl group substitution at the 1-position of the naphthyl ring which severely restricts the conformational freedom of the 2-naphthyl compounds.[148] The increased activity measured for **4** may result from the addition of the fluorine on the catechol ring as seen in other classes of catechol diether inhibitors.[159,160]

A series of 2-naphthyl compounds containing larger substitutions off the naphthyl 4-position have been previously reported.[148] For our current analysis, we have determined the structures of two examples from this series, **5** and **6**, bound to WT RT. The

superpositions of the structures of **5** and **6** with **2** show the ligands bind similarly (Figures 4.5A, B). The 4-cyclopropyl group of **5** and 4-methyl ester group of **6** both bind within a small pocket surrounded by residues P95, Y181, and Y188 of RT. Interestingly, these three residues all shift from their relative positions in the compound **2** bound structure (Figure 4.5A, 4.5B). Whereas P95 and Y188 move away from the bulky 4-position substituents, Y181 moves toward the substituent to assist in forming the sub-pocket that accommodates the cyclopropyl and methyl ester groups. In addition to these protein residue rearrangements, the 2-naphthyl rings of **5** and **6** must also slightly shift toward F227, away from P95 and Y181, to create room for the 4-position groups of each compound. Within its position in the P95 pocket, the methyl ester group of compound 6 is unable to form any hydrogen bonds with the protein. Furthermore, the superposition of 6 bound to Y181C RT with that of the WT structure shows the binding of the compound is nearly identical in both structures (Figure 4.5C). The large methyl ester group off the 4-positon continues to point towards P95 and does not move toward the empty space created by the missing tyrosine in the Y181C RT mutant.

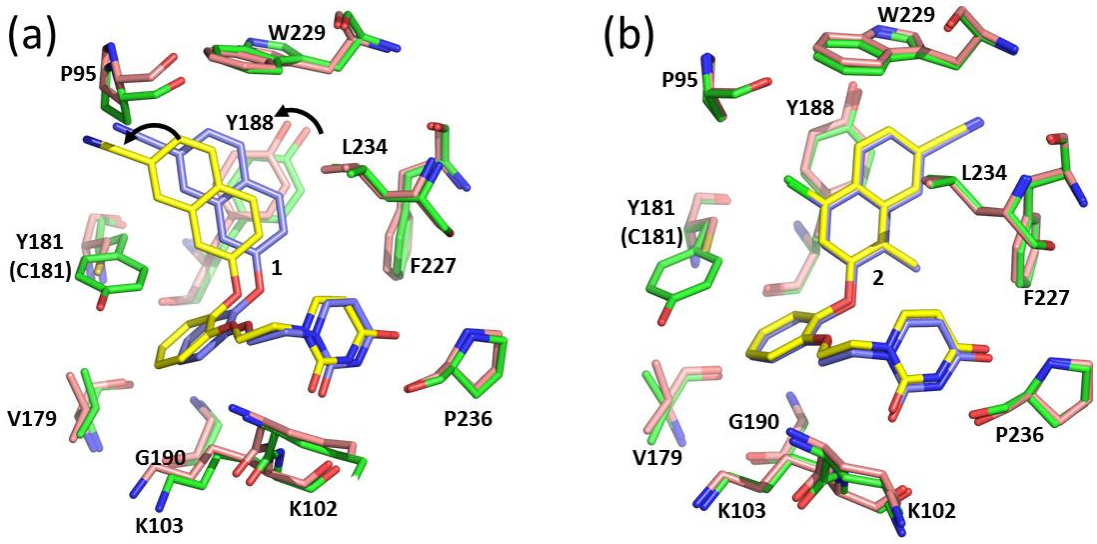


Figure 4.4. Superposition of 2-naphthyl compounds with WT and Y181C RT.
(a) Superposition of 1 bound to WT RT (1 in lilac, protein in green) and Y181C RT (1 in yellow, protein in pink) (b) Superposition of 2 bound to WT RT (2 in lilac, protein in green) and Y181C RT (2 in yellow, protein in pink).

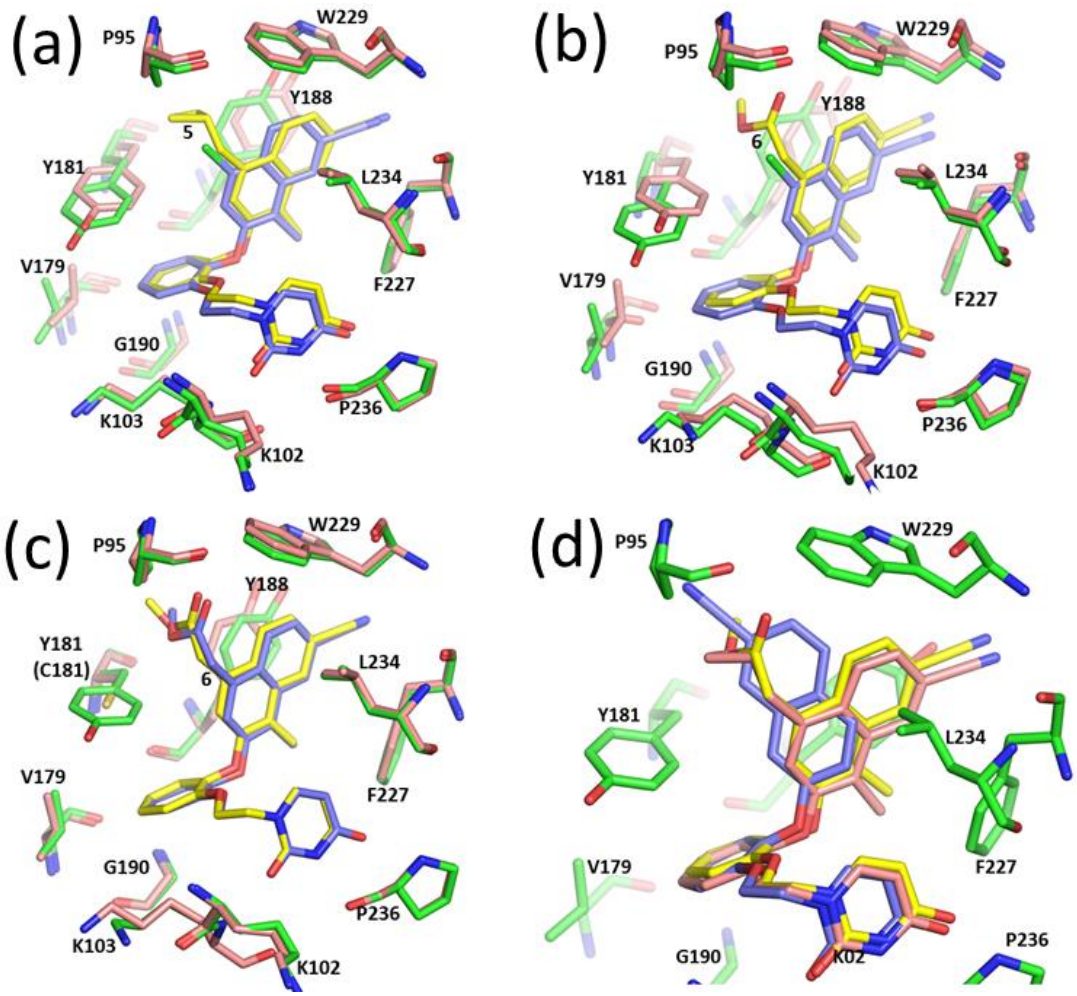


Figure 4.5. Structures of 2-naphthyl compounds with large substitutions on the 4-position. (a) Superposition of 5 bound to WT RT (5 in yellow, protein in pink) with 2 bound to WT RT (2 in lilac, protein in green). (b) Superposition of 6 bound to WT RT (6 in yellow, protein in pink) with 2 bound to WT RT (2 in lilac, protein in green). (c) Superposition of 6 bound to WT RT (1 in lilac, protein in green) and Y181C RT (1 in yellow, protein in pink). (d) Superposition of bound structures of 5 (pink), and 6 (yellow) with the structure of the WT:1 complex (1 in lilac, protein in green).

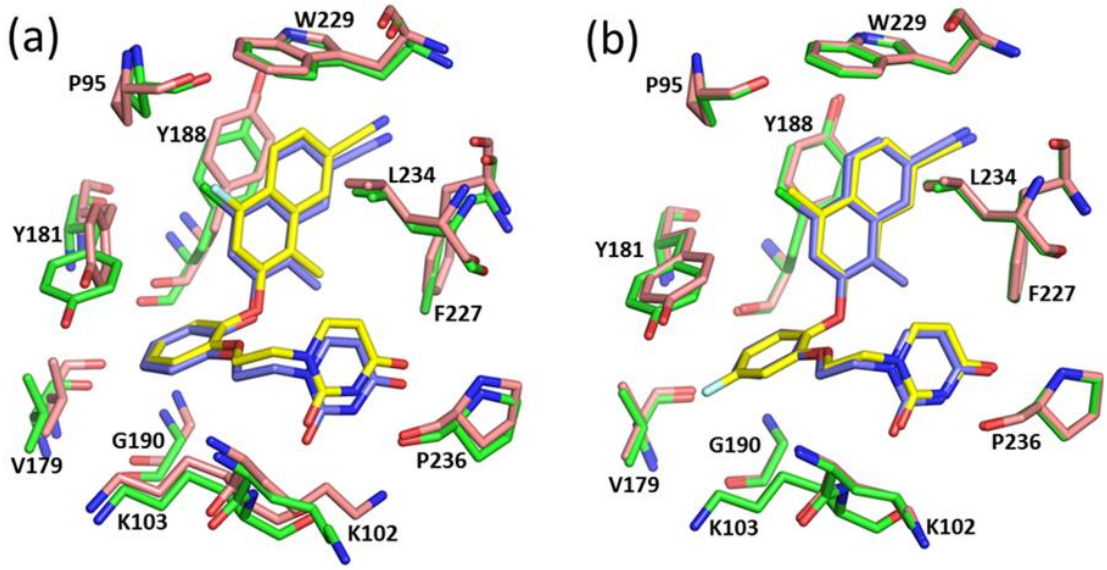


Figure 4.6. Structures of 2-naphthyl compounds with small substitutions on the 4-position.
(a) Superposition of 3 bound to WT RT (3 in yellow, protein in pink) with 2 bound to WT RT (2 in lilac, protein in green). (b) Superposition of 4 bound to WT RT (4 in yellow, protein in pink) with 2 bound to WT RT (2 in lilac, protein in green).

4.4 Discussion

The 2-naphthyl phenyl ethers represent a promising class of NNRTIs that have been shown to have 1 – 10 nM potencies against WT HIV strains.[148] However, initial analogs from this class, such as those presented here, showed weaker activity against clinical HIV strains that contain the Y181C RT mutant. Whereas compound **2** displays a 10-fold decrease in activity between WT and Y181C strains, compounds **1** and **6** show up to 500- to 1000-fold diminished activity. The structures determined here provide insight into the structural basis that result in these reduced activities in the presence of the Y181C mutation in the NNRTI binding pocket.

Compounds **1** and **2** have previously been shown to bind WT RT with different conformations of the 2-naphthyl ring dependent upon the substitution at the naphthyl 1-position. When the 1-position is substituted with a methyl group as in compound **2**, the 2-naphthyl ring prefers to bind with a conformation such that the cyano group projects towards the vicinity of W229. Since this binding conformation takes advantage of a greater interaction between the naphthyl group and W229, it has been previously postulated that compound **2** would show less of an impact when Y181 is substituted by cysteine.[148] The structure of **2** bound to Y181C RT reported here does indeed reveal the interaction with W229 to be substantial enough to maintain compound **2** in its WT position even in the absence of Y181. In contrast, we show the position of compound **1** to be significantly altered in the presence of the Y181C mutation, and given the larger role Y181 plays in the binding of **1** to WT RT, this result is not unexpected. In the absence of Y181, the catechol and 2-naphthyl rings of **1** shift toward C181 and away from W229. Whereas the interaction made by **1** with W229 is already suboptimal in WT RT, it becomes even less optimal in

the presence of C181, as the naphthyl ring of **1** is displaced further away from W229 in order to occupy the space created by the vacated Y181. Thus, compound **1** loses key contacts with both W229 as well as the absent Y181 in the Y181C RT mutant. Although Y188 slightly compensates for this loss by rotating toward the shifted naphthyl ring of **1** in the mutant, the activity of **1** still diminishes 1000-fold against Y181C HIV strain. In comparison, compound **2** only loses interaction with Y181 and not W229 in the Y181C mutant and as a result, displays only a 10-fold decrease in activity against the mutant strain. Four additional 2-naphthyl analogs that did not have a substitution at the 1-positon were also previously examined, and they similarly show 46-163-fold worse activities against the variant HIV strain.[148]

The 2-naphthyl analogs that contain larger substituents off the naphthyl ring 4-position were initially designed with the expectation that the larger groups would occupy a portion of the space created in the Y181C mutant and have improved potency against mutant HIV strains. The initial compounds, however, were less active against both WT and variant HIV strains compared to the parent compound that possesses only a methyl group substitution at the 4-position[148]. The structural rearrangement of residues P95, Y181, and Y188 that are necessitated to accommodate the larger 4-position groups most likely incurs an entropic penalty that results in the decreased activity exhibited by these compounds. Furthermore, in the presence of the Y181C mutation, the bulky group of **6** maintains its WT-bound conformation and does not move toward the vacated space of Y181. This inflexibility may explain why these compounds do not fare better against HIV variants as initially hypothesized. In the case of compound **6**, the methyl ester group is unable to form any hydrogen bonds with the surrounding residues, and the resulting

desolvation penalty may further contribute to its weaker activity. Interestingly, the superposition of the **5** and **6** structures with the structure of **1** shows the cyclopropyl and methyl ester substituents to occupy the same space as that of the 2-naphthyl ring of **1** (Figure 4.5D). The poorer activities displayed by both the 4-position substituted analogs and **1** against WT HIV (Table 4.1) suggest that the P95 sub-pocket can be a difficult region of the NNRTI binding site to target successfully in design studies.

Although the initially designed set of 2-naphthyl compounds containing large 4-position substituents described here did not show improved activity against either WT or mutant HIV strains, subsequent analogs have been successfully utilized in the design of covalent inhibitors against Y181C RT.[158] These compounds contain either an α-halo amide or acrylamide group off the naphthyl 4-position that act as an electrophilic warhead, and they provide an interesting comparison to the 4-position substituted examples presented here. Upon binding to WT RT, the catechol ring of these warhead-containing compounds shifts away from Y181 by 1.2 Å to provide room for the large acrylamide group from the 4-position, much like what is observed in the structures of **5** and **6** RT complexes. As a result, these warhead-containing 2-naphthyls similarly display a 100-fold decrease in activity against WT HIV compared to compound **1**. However, the electrophilic functional groups contained in these compounds have been shown to covalently modify the C181 side chain and irreversibly inhibit activity of the resistant Y181C RT mutant.[158] This special feature allows these 4-substituted 2-naphthyl inhibitors to retain their activity against Y181C mutant strains. Whereas placing bulky groups off the 4-position that look to solely replace the vacated Y181 residue may not be sufficient to overcome Y181C resistance,

success can be achieved by instead utilizing functional groups at this position that are capable of reacting with the substituted C181 side chain.

Based on the structural results presented here, the interaction between NNRTI compounds with W229 appears to be a major driving force for overcoming Y181C RT strains. Compound **1** does not interact as strongly with W229 in WT RT (Figure 4.7A) and displays a 1000-fold reduction of activity against Y181C RT. In comparison, compound **2** interacts much more strongly with W229 (Figure 4.7B) and only shows a 10-fold reduction against the mutant strain. Furthermore, a compound from the 1-naphthyl class of catechol diethers, which has been shown to make an even more extended edge to face interaction with W229 than that observed for compound **2** (Figure 4.7C)[160], displays only a 3-fold reduction in activity against Y181C HIV.[159] This comparison suggests that the more a compound interacts with W229 within the NNRTI binding site, the more likely it will have the ability to overcome the Y181C resistance mutation in HIV. Given that W229 does not appear to be a very mobile residue when comparing the many RT structures solved and that W229 mutations that yield resistance have yet to be identified in the clinic further suggest that targeting W229 in drug design studies may lead to the discovery of more efficacious NNRTI drugs.

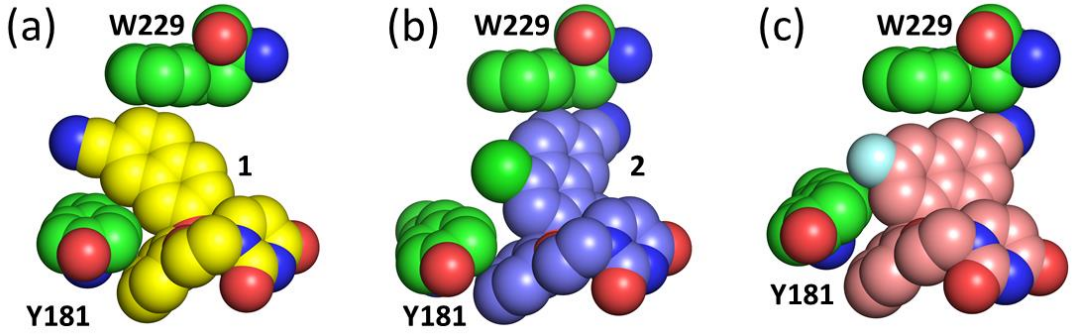


Figure 4.7. Space filling models of 2-naphthyl compounds.
Depicting the interactions between compounds 1 (a, 1 in yellow), 2 (b, in lilac), and a 1-naphthyl analog (c, in pink, PDB code: 6OE3) and Y181 and W229 (in green).

Chapter 5. Distinguishing the activity of Polα with two replication substrates

This chapter is an excerpt from:

Baranovskiy AG, **Duong VN**, Babayeva ND, Zhang Y, Pavlov YI, Anderson KS, Tahirov TH. Activity and fidelity of human DNA polymerase α depend on primer structure. J Biol Chem. 2018 May 4;293(18):6824-6843. doi: 10.1074/jbc.RA117.001074. Epub 2018 Mar 19. PMID: 29555682; PMCID: PMC5936803.

5.1 Introduction

Three DNA polymerases are required for genome replication in eukaryotes: DNA polymerase α (Polα), Polε, and Polδ.[161] All of them belong to the B-family and fulfill the different tasks. Polε synthesizes most of the leading strand, Polδ is mainly involved in synthesis of the lagging strand, and Polα generates the primers for both polymerases.[162,163] Polα alone cannot synthesize DNA primers *de novo* and relies on RNA primers created by primase. Polα works in a tight complex with primase, called the primosome.[164,165] Synthesis of the chimeric RNA–DNA primers by the primosome is highly coordinated by autoregulation through the alternating activation/inhibition of two catalytic centers, which is mediated by the C-terminal domain of the primase accessory subunit.[166] Relatively low fidelity of Polα, which does not possess a proofreading activity, results in mutational hot spots predominantly on the lagging strand.[167] In addition to the established role of primosome in nuclear replication, it is involved in formation of hybrid DNA:RNA duplexes in the cytosol, which are important for regulation of the type I interferon response.[168] Polα is a direct target of an anti-tumor toxin CD437, an attractive anti-cancer lead molecule, which induces apoptosis selectively in cancer cells.[169]

Human Polα (hPolα) is composed of two polypeptides: the catalytic subunit (p180) and the accessory B-subunit (p70), with calculated molecular masses of 166 and 66 kDa, respectively. p180 contains two domains, the catalytic (residues 338–1250) and the C-terminal (Polα CTD, residues 1266–1462) domains, which are flexibly connected by a 15-residue-long linker.[166] The catalytic domain possesses DNA-polymerizing activity but has no proofreading exonuclease activity, in contrast to other replicative DNA Pols, δ and ε.[170] Polα CTD connects the catalytic domain with p70 and primase and contains two conserved zinc-binding modules, where each zinc ion is coordinated by four cysteines.[166,171,172] The N terminus of p180 (residues 1–337) is predicted to be poorly folded and does not participate in primer synthesis. The structural information for this region is limited to a small peptide in the catalytic subunit of yeast Polα (yPolα; residues 140–147) that interacts with the replisome.[173]

In this chapter, we use structural and kinetic approaches to analyze hPolα interaction with the template:primer and dNTP and the effect of the primer structure on hPolα catalysis, processivity, and fidelity. Understanding these interactions will give us insight into the mechanism of Polα replication.

5.2 Materials and Methods

Protein expression and purification

Cloning, expression, and purification to homogeneity of p180core (residues 335–1257), which contains the catalytic domain of hPolα, have been described elsewhere.[174] Peak fractions obtained from Heparin HP HiTrap column (GE Healthcare) were combined and dialyzed to the buffer specific for each application (see below).

Crystallization

DNA:RNA and DNA:DNA duplexes were obtained at 0.2 mM concentration by annealing at 43 °C for 30 min (after heating at 70 °C for 1 min) in buffer containing 10 mM Tris-HCl, pH 7.9, and 70 mM KCl. In the case of DNA:RNA duplex, the sequences for DNA template and RNA primer were 5′-ATTACTATAGGCGCTCCAGGC (the region complementary to a primer is underlined) and 5′-rGrCrCrUrGrGrArGrCrG/ddC/, respectively (/ddC/is a dideoxycytidine, which prevents polymerization during crystallization). In the case of DNA:DNA duplex, the template and primer were 5′-ATAGGCGCTCCAGGC and 5′-GCCTGGAGCG/ddC/, respectively. Dialyzed protein sample (15 μM p180core) was diluted 1.5-fold with dialysis buffer (10 mM Tris-HCl, pH 7.7, 100 mM KCl, 1% glycerol, and 1 mM DTT) containing 36 μM template:primer, 3.6 mM $MgCl_2$, and 12 mM dCTP. The ternary complex was concentrated 10-fold and flash-frozen in liquid nitrogen. Before crystallization experiments, the aliquots of protein/template:primer/dCTP solutions were defrosted and centrifuged to remove the precipitate, and the sample monodispersity was verified with the dynamic light scattering. The screening of crystallization conditions was performed with the sitting-drop vapor diffusion method at 295 K by mixing 1 μl of ternary complex solution with 1 μl of reservoir

solution. Initial screen solutions producing tiny crystals were optimized to produce well-shaped crystals at 295 K with reservoir solutions containing 0.8 mM zinc sulfate, 8.8% (v/v) PEG MME 550, and 50 mM MES, pH 6.5, for p180core/DNA:RNA/dCTP and 0.8 mM cobalt chloride, 2 mM tris(2-carboxyethyl) phosphine (TCEP), pH 7.5, 250 mM 1,6-hexanediol, and 50 mM sodium acetate, pH 4.6, for p180core/DNA:DNA/dCTP.

Data collection

For data collection, each crystal was soaked in cryoprotectant solution (100 mM potassium chloride, 0.8 mM zinc sulfate, 1.1 mM magnesium chloride, 15% (v/v) PEG MME 550, 15% (v/v) ethylene glycol, and 50 mM MES, pH 6.5, for p180core/DNA:RNA/dCTP crystals, and 0.8 mM cobalt chloride, 2 mM TCEP, pH 7.5, 230 mM 1,6-hexanediol, 26% (v/v) ethylene glycol, and 50 mM sodium acetate, pH 4.6, for p180core-DNA:DNA-dCTP crystals) for a few seconds, scooped in a nylon-fiber loop, and flash-cooled in a dry nitrogen stream at 100 K. All initial diffraction data were obtained on a Rigaku R-AXIS IV imaging plate using Osmic VariMax^{TM} HR mirror-focused CuKα radiation from a Rigaku FR-E rotating anode operated at 45 kV and 45 mA. Complete diffraction data sets were collected using synchrotron X-rays on the Argonne National Laboratory Advanced Photon source beamline 24-ID-E using ADSC Quantum 315 detector. All intensity data were indexed, integrated, and scaled with DENZO and SCALEPACK from the HKL-2000 program package.[151] The crystals of p180core/DNA:RNA/dCTP belong to trigonal space group $P3_221$ and diffract up to 2.2 Å resolution, and the crystals of p180core/DNA:DNA/dCTP belong to tetragonal space group $P4_22_11$ and diffract up to 2.95 Å resolution. Both crystals contained only one copy of the ternary complex in the

asymmetric unit. The crystal parameters and data-processing statistics are summarized in Table 5.1.

Crystal structure determination

Initial phases for p180core/DNA:RNA/dCTP structure were determined by the molecular replacement method using the coordinates of backbone atoms of the yeast Polδ catalytic core derived from its ternary complex with template:primer and dCTP (PDB code 3IAY) as a search model. The positions of magnesium and zinc ions were determined using an anomalous difference Fourier map. Molecular replacement and initial automated model rebuilding with Phenix[155] revealed over 70% of correctly built protein structure. The model building was continued and completed manually with Turbo-Frodo, and the structure was refined using standard protocols of CNS version 1.1.[175] After the addition of solvent molecules, the model was refined at 2.2 Å resolution to an R_{cryst} of 21% and an R_{free} of 23.9%. The structured region of protein starts with Glu^{338} and ends with Thr^{1244}. The electron density maps were poor or were missing also for the internal regions 674–677, 809–833, and 883–895; therefore, these regions were excluded from the structure.

The p180core/DNA:DNA/dCTP structure was determined by the molecular replacement method using the coordinates of p180core in the complex with DNA:RNA and aphidicolin (PDB code 5Q5V) as a search model. The model was adjusted manually with Turbo-Frodo. The positions of magnesium and cobalt ions were determined using an anomalous difference Fourier map. The structure was refined at 2.95 Å resolution using standard protocols of CNS version 1.1[175]to an R_{cryst} of 25.9% and an R_{free} of 30.1%. The traceable electron density starts with Glu^{338} and ends with Val^{1248}. The regions 674–677, 810–833, and 883–895 with missing electron density were excluded from the model. The

electron density was relatively weak for the portion of the thumb closer to the thumb tip, especially for residues 1137–1154, indicating their partial disorder and/or elevated mobility.

The refinement statistics for both structures are summarized in 5.1. The figures containing molecular structures were prepared with the PyMOL Molecular Graphics System (version 1.8, Schrödinger, LLC).

Pre-steady-state kinetic assays

Pre-steady-state kinetic assays were performed using the RQF-3 rapid chemical quench apparatus (KinTek) at 37 °C. All kinetics assays were carried out using reaction buffer. Polα was incubated with template:primer and rapidly mixed with dNTPs before quenching with 0.5 M EDTA. Products were collected in a tube with formamide dye (0.1% bromphenol blue (w/v), 0.1% xylene cyanol (w/v)) and separated by denaturing urea PAGE. The radiolabeled products were visualized by the Molecular Imager FX phosphor imager (Bio-Rad) and quantified by Quantity One, version 4.6.9 (Bio-Rad).

For optimal separation of RNA primer-elongated products, a modified denaturing urea PAGE protocol was employed. A solution of acrylamide mix was made with the addition of 10% formamide (19% acrylamide, 1% bisacrylamide, 10% formamide, 7 M urea, 1× TBE buffer (pH 8.3)). The acrylamide polymerization reaction was initiated with 0.05% ammonium persulfate and 0.1% TEMED. Before loading the gel with sample, the gel was run at 3000 V for at least 2 h to allow the gel to heat up. The elongated products containing equal parts of quenched product and formamide dye were then loaded onto the gel (<10 μl) and allowed to run for ∼6 h.

Burst incorporation kinetics

Burst incorporation kinetics were conducted with final concentrations of excess template:primer (9 μM) to Polα (3 μM). To ensure that binding of the incoming nucleotide was not rate-limiting, a high concentration of dATP (200 μM) was used in the polymerization reaction. The reaction products were quantified, and the concentration of product was plotted against time and fit to both a linear and a burst equation, [product] = $A(1-e^{-kt}) + k_{ss}t$, where *A* is the burst phase amplitude corresponding to the active site concentration, k_{obs} is the observed rate, k_{ss} is the steady-state rate, and *t* is the time. In the event of misincorporation after a correct incorporation, 500 μM dATP was used to ensure nucleotide saturation.

Single- and double-nucleotide incorporation assays

Single- and double-nucleotide incorporation assays were performed with final concentrations of excess Polα (3 μM) to template:primer (100 nM). In the case of single-nucleotide incorporation assays, the concentration of dATP was varied (1–300 μM). For each experiment at a particular dATP concentration, the concentration of product was plotted against time and fit to a single-exponential equation, [product] = $A(1-e^{-kt})$, where *A* is the amplitude, k_{obs} is the observed rate for the incorporation of the incoming dNTP, and *t* is the time. The k_{obs} was then plotted against dATP concentration and then fit to a quadratic equation to obtain the K_d, the apparent binding constant for the incoming nucleotide for the Polα/template:primer binary complex, and k_{pol}, the maximum rate of nucleotide incorporation.

In the case of double-nucleotide incorporation assays, Polα first incorporates a dATP followed by a subsequent dTTP incorporation. The concentration of dATP was held constant at saturating amounts based on single-turnover experiments (100 μM for DNA

primer and 300 μM for RNA primer), and the concentration of dTTP was varied (1–300 μM). The data were fit using KinTek Global Explorer[176] to provide an estimate for the K_d and k_{pol} for the second incorporation event.

Processivity assays

Processive polymerization assays were performed under similar single-turnover conditions as the single- and double-nucleotide incorporation assays with final concentrations of excess Polα (3 μM) to template:primer (100 nM). The Polα/template:primer complex was mixed with saturating concentrations of all dNTPs (240 μM each) to allow for complete extension of the primer substrate. The processivity data were modeled in KinTek Global Explorer to determine the k_{pol} and k_{off}, the rate of dissociation of the template:primer substrate from Polα, for each incorporation event.

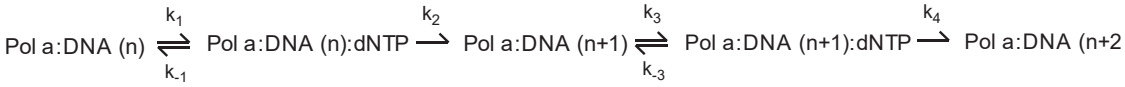
In fitting the data to the models, the following assumptions were made: 1) all incorporation events were irreversible, 2) the K_d and k_{pol} values obtained from the single-incorporation experiments were manually fixed for the first incorporation in the double-incorporation modeling, 3) the k_{on} values for template:primer binding to Polα were assumed to be the same, disregarding primer length differences.

Data analysis and kinetic models

Analysis of the radiolabeled products through the phosphor imager provides raw data in count values. For each time point, the counts of each band present were summed to calculate the total amount of counts and subsequently the percentage of each product length present. These values were then normalized to add up to 100% and then converted into concentration amounts based on the total amount of nucleic acid substrate used in the assay.

The KinTek Global Explorer program requires user-inputted models to fit kinetic data for the estimation of rates (Fig. 5.1)[176,177]. Briefly, the data from double-incorporation assays were fit to a two-step mechanism of dNTP binding followed by incorporation that was modified to account for two incorporations. The processivity data were fit using a model that allows for binding or dissociation of each DNA product and combines the dNTP binding and incorporation event into one rate.

Double incorporation



Processivity

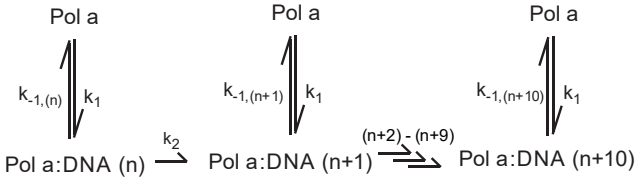


Figure 5.1. Kinetic models utilized to fit kinetic data in KinTek Global Explorer for the estimation of rates.

Various models were employed to fit the different kinetic experiments, using Polα to extend a DNA or RNA primer. For the double incorporation assays, a two-step mechanism was utilized, where the binding of the incoming dNTP is followed by incorporation of the dNTP to elongate the strand. In the processivity modeling, a model where the primer could either elongate or dissociate from Polα at each primer length was utilized.

5.3 Results

Interaction of human Polα with DNA:DNA and RNA:DNA double helices

The structures of the p180 core of hPolα with DNA:RNA (PDB: 4QCL) or DNA:DNA (PDB: 6AS7) and dCTP were solved (Table 5.1). Due to potential crystal packing effects seen in the DNA:DNA ternary structure, in order to compare the two substrates the binary complex of DNA:DNA (closed form, PDB: 5IUD) was overlapped with the RNA:DNA structure. Binding of a hybrid duplex results in potential steric hindrance between Arg^{1081} and the 2′-OH of the P_4 ribose (Fig 5.2A) and in a local change of RNA conformation (Fig 5.2B), which includes the unusual *syn*-conformation of P_4 (torsion angle $\chi = -70.3°$), 2′-endo pucker of its ribose, and unstacking between the bases of P_4 and P_5. This bending of the RNA primer was also observed in the yPolα/DNA:RNA/dGTP complex.[178] It is notable that the orientation of the P_3 phosphate in the RNA primer is more favorable for hydrogen bonding with Arg^{1082} and Arg^{702}, compared with the P_3 phosphate of the DNA primer. The same position of Arg^{1081} analog in ternary complexes of yeast Pols δ and ε indicates a common mechanism of RNA sensing in replicative DNA polymerases.

Polα does not exhibit typical polymerase burst kinetics

A pre-steady-state burst experiment was performed to provide insight into the relative rates of DNA polymerization steps in the kinetic mechanism of Polα. It has been frequently observed that under burst conditions (slight excess of template:primer over enzyme), many polymerases exhibit biphasic kinetics.[123-125] This biphasic pattern indicates that in the overall mechanism, the release of products is rate-limiting and slow compared with steps governing chemical catalysis. p180core (3 μM) was incubated with either a 15-

nucleotide DNA (D_{15}) or RNA primer (R_{15}) annealed to a 25-nucleotide DNA template (T_{25}; 9 μM) to allow for the formation of the binary complex (Fig 5.3A). The preincubated solution was rapidly mixed with dATP (200 μM) using an RQF-3 rapid chemical quench apparatus. When the concentrations of products were plotted *versus* reaction time, an apparent biphasic curve was not observed but rather a linear formation of product for the elongation of both the DNA (k_{ss} = 104.3 ± 3.1 s^{-1}) and RNA (k_{ss} = 31.7 ± 1.7 s^{-1}) primer substrates (Fig. 5.3B, C). Therefore, it appears that the Polα catalytic core does not exhibit typical biphasic kinetics observed in reactions with main replicative polymerases, Polδ and Polε.[179,180] This kinetic behavior indicates that for Polα, the rate of product release is equal to or faster than the rate of chemistry.

Polα incorporates the first and second nucleotides into a DNA primer more efficiently than into an RNA

Single-nucleotide incorporation experiments were done under single-turnover conditions (excess Polα over template:primer) to quantify the maximal rate of catalysis, k_{pol}, and the apparent binding constant for the incoming nucleotide, K_d. p180core (3 μM) was incubated with a template:primer (100 nM) and rapidly mixed with varying concentrations of dATP (1–300 μM) under rapid chemical quench conditions (Fig. 5.2D). For the DNA primer substrate, the k_{pol} is 33.8 ± 3.7 s^{-1}, the K_d is 9.2 ± 3.4 μM, and the incorporation efficiency (k_{pol}/K_d) is 3.7 μM^{-1} s^{-1} (Table 5.2). For the RNA primer substrate, the k_{pol}, K_d, and k_{pol}/K_d values are 48.0 ± 2.7 s^{-1}, 62.2 ± 10.4 μM, and 0.8 μM^{-1} s^{-1}, respectively. It appears that Polα incorporates a nucleotide more efficiently to a DNA primer by approximately a factor of 4.5. Strikingly, this discrepancy is primarily due to differences in the K_d for the incoming nucleotide.

Experiments were conducted to examine whether subsequent incorporations to a DNA primer were more efficient as compared with an RNA primer after the initial dNTP incorporation event. During the second incorporation, a 3′-hydroxyl of deoxyribonucleotide attacks the α-phosphate of incoming dNTP in both the DNA and RNA primer extension assays. Double-incorporation assays were conducted to estimate the kinetic parameters of a subsequent incorporation after the first incorporation. Experiments were designed to observe a dATP followed by a dTTP incorporation under single-turnover conditions. Preincubated solutions of p180core (3 μM) and template:primer (100 nM) were mixed with a solution containing both dATP and dTTP. Because the kinetic parameters of the second incorporation were desired, concentrations of dATP were held at saturating amounts (100 μM for DNA primer and 300 μM for RNA, based on single-incorporation experiments), and concentrations of dTTP were varied from 1 to 300 μM (Fig. 5.4A).

The appearance of the second incorporation product is complex due to its dependence on the first product being made; product concentration cannot simply be fit to a single-exponential curve like the single-incorporation assays. Thus, the KinTek Global Explorer program[176] was utilized to fit the data and provide an estimate for the k_{pol} and K_d of the dTTP incorporation (Fig. 5.5). For the second incorporation to a DNA primer, the k_{pol} is estimated to be 65.8 s^{-1}, and the K_d is 5.4 μM. The estimates for the RNA second incorporation are k_{pol} = 23.1 s^{-1} and K_d = 13.4 μM. This leads to incorporation efficiencies of 10.4 and 1.7 $\mu m^{-1} s^{-1}$ in favor of DNA primer extension (Table 5.3). This 6-fold difference in incorporation efficiency for DNA *versus* RNA for sequential dNTPs is comparable with the single-incorporation results. Contrary to the single-incorporation experiments, a major difference in k_{pol} values is observed in the double-incorporation

experiments. We hypothesize that the ~3-fold difference in k_{pol} values may stem from the primer bending at P_4 in the hybrid duplex. The altered helix structure possibly affects the rate of any step in subsequent incorporations after the initial incorporation, such as translocation or necessary protein conformational changes. The reason for similar k_{pol} values during the first incorporation is the preformation of the enzyme/template:primer complex before the incorporation reaction is initiated.

Comparison of Polα processivity on DNA:DNA and DNA:RNA duplexes

A processivity experiment was employed to examine multiple rounds of dNTP incorporation and look at the full extension of the primers and determine whether the rates of each incorporation were different over time between DNA and RNA primers. This analysis observes the k_{pol} for each incorporation event as well as k_{off}, which characterizes the dissociation rate of the Polα/template:primer complex. The assay was conducted under single-turnover conditions and saturating concentrations (240 μM) of all dNTPs (Fig. 5.4B, C). The products were plotted against time and fit to a processive mechanism using KinTek Global Explorer (Fig. 5.4D, E; for visual clarity, only the first five incorporations are shown). The k_{pol} and k_{off} values for the first seven incorporations are shown in Table 5.4. Although this experiment provides only estimates of these rates, it is clear that there is a higher average k_{pol} for the second to fourth incorporations during DNA primer extension, with an average value of 26.8 s^{-1} compared with 12.0 s^{-1} for RNA. As was shown above, the first incorporation does not show the real k_{pol} values due to preformation of Polα/template:primer complex; that is why it was omitted from averaging.

Interestingly, we observe an increase in the rate of incorporation after the fourth incorporation into the RNA primer, with an average k_{pol} for the fifth to seventh

incorporations of 24.4 s^{-1}, which is similar to the value obtained on a DNA primer (22.2 s^{-1}). This is particularly interesting in light of our structural results pertaining to the kink at P_4 in the hybrid duplex structure. This observation can be explained in structural terms by the fact that after four incorporations, the rigid and conserved DNA-binding cleft of Polα no longer makes contacts with the ribonucleotides in the hybrid duplex. This increase in k_{pol} agrees with our previous experiments demonstrating that DNA primer elongation is more efficient. Furthermore, to confirm that the increase from 12.0 to 24.4 s^{-1} after the fourth incorporation is significant, we simulated how the time course of processive elongation of an RNA primer would change if the fourth incorporation rate was modified to match the rate observed in the DNA experiment (Fig. 5.4E, *dotted lines*). Changing 8.3 s^{-1} to 19.3 s^{-1} results in a drastically different product distribution. The most prominent effects are the decreased accumulation of R_{18} and faster production of R_{20} because the rate of R_{19} formation is no longer significantly rate-limiting.

Additionally, the processivity of a polymerase can be qualitatively assessed by inspection of separated products on a gel. Comparing the processivity gels from Polα (Fig. 5.4B, C) and T7 DNA polymerase[123], the striking difference is the presence of primers that have not been fully extended with dNTPs at later points in the experimental time course. Through our experiments, it is observed that there is a significant population of primers that have been only singly elongated (D_{16}) at later time points even when the final product is finally produced (D_{25}). In contrast, once T7 DNA polymerase incorporates a dNTP into a primer, it continues to incorporate nucleotides into the same primer. In the direction of increasing time, the pattern on the gel for the Polα can be described as extending a ladder upward, whereas T7 can be described as shifting a ladder upward. Processivity experiments

have also been done with Polε[180], and the pattern of processive polymerization is similar to Polα. Although Polε is considered more processive than Polα, the striking difference between T7 and the former polymerases highlights the highly variable kinetic characteristics of polymerases.

Pre-steady state experiments demonstrate the low fidelity activity of Polα

Initial burst experiments, performed on a broad time scale to establish the relative rate of dNTP incorporation by the p180core, revealed that Polα incorporates an incorrect dNTP at a notable rate (Fig. 5.6). This is illustrated at later time points by the appearance of a second band that corresponds to a dATP incorporation across from a template adenine. This observation was unexpected due to the prevalence of misincorporation within a time frame relatively close to correct incorporation. Although previous studies[181,182] have indicated that Polα is a low-fidelity polymerase, to the best of our knowledge, this is the first direct observation of significant Polα misincorporation in a millisecond time scale. These results further emphasize the benefits of utilizing a pre-steady-state kinetic approach.

Table 5.1. Data collection and refinement statistics.

Crystal	p180core/DNA:RNA/dCTP (PDB ID: 4QCL)	p180core/DNA:DNA/dCTP (PDB ID: 6AS7)
Data collection		
Space group	*P* 3_221	*P* 4_22_12
Cell dimensions		
a = *b*, *c* (Å)	140.76, 181.32	151.81, 113.3
Resolution (Å)	40 – 2.2 (2.24 – 2.2)[a]	50 – 2.95 (3 – 2.95)
R_{merge}	0.055 (0.531)	0.072 (0.47)
$I/\sigma I$	19.76 (1.9)	18.2 (2.2)
Completeness (%)	91 (81)	99.6 (99.3)
Unique reflections	95874 (4196)	28565 (1406)
Redundancy	2.7 (1.9)	9.6 (5.5)
Refinement		
Resolution (Å)	29.51 – 2.2 (2.34 – 2.2)	46.2 – 2.95 (3.13 – 2.95)
Number of reflections		
Working set	88726 (12002)	26970 (4410)
Test set	4683 (643)	1380 (209)
R_{work} /R_{free}	0.21/0.239 (0.388/0.404)	0.259/0.301 (0.38/0.402)
Number of atoms		
Protein	6947	6996
DNA and RNA	496	487
Ligands/ions	48/5	28/4
Water	414	1
Mean B-factors (Å^2)	51.1	57.6
R.m.s deviations		
Bond lengths (Å)	0.007	0.01
Bond angles (°)	1.3	1.6
Ramachandran plot (%)		
Core	89.7	82.1
Allowed	9.7	16.8
Generously allowed	0.5	1.0
Disallowed	0.1	0.1

[a] Numbers in parentheses refer to highest-resolution shell.

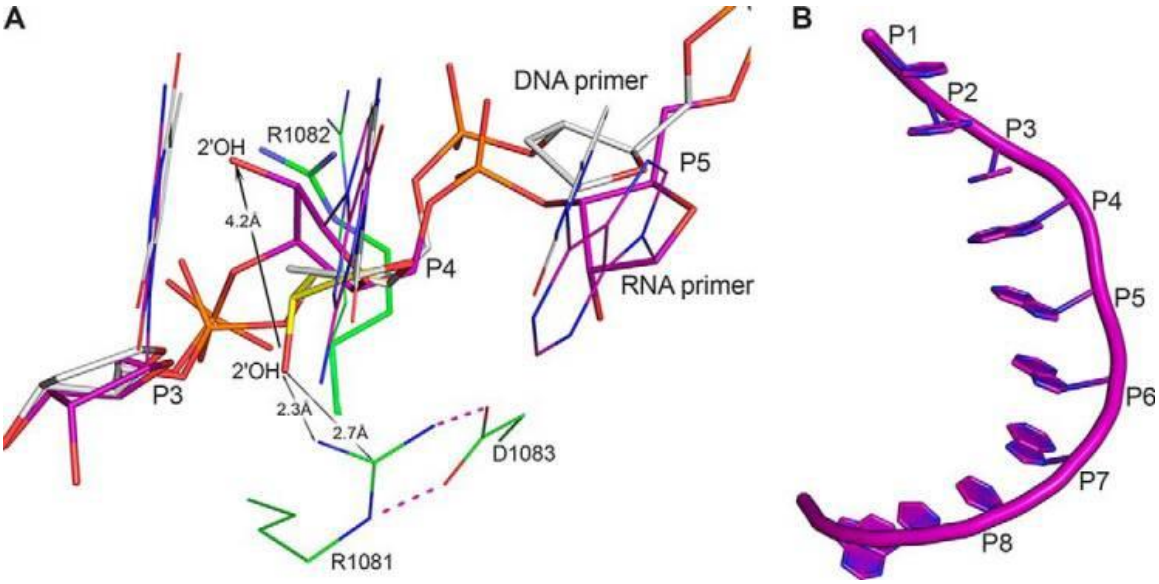


Figure 5.2. RNA primer bending.
A, primer conformation in the complexes of hPolα with DNA:RNA and DNA:DNA. The p180core/DNA:RNA/dCTP and p180core/DNA:DNA complexes (PDB codes 4QCL and 5IUD, respectively) were aligned with RMSD of 0.5 Å using the palm and thumb. In the complex containing DNA:DNA, the carbons of a DNA primer and amino acids are *colored gray* and *green*, respectively. In the complexes containing RNA or DNA primer, amino acids are represented as *sticks* or *lines*, respectively. The modeled ribose with the 3′-endo pucker was superimposed on the P_4 sugar of the DNA primer and has the same position of carbons (*colored yellow*) except C2. The *arrow* depicts the difference in position of P_4 2′-OH in the RNA primer and the modeled ribose. *Pink dashed lines* and *double arrows* depict H-bonds and distances between atoms, respectively. *B*, overall view of RNA primer in the complex p180core/DNA:RNA/dCTP.

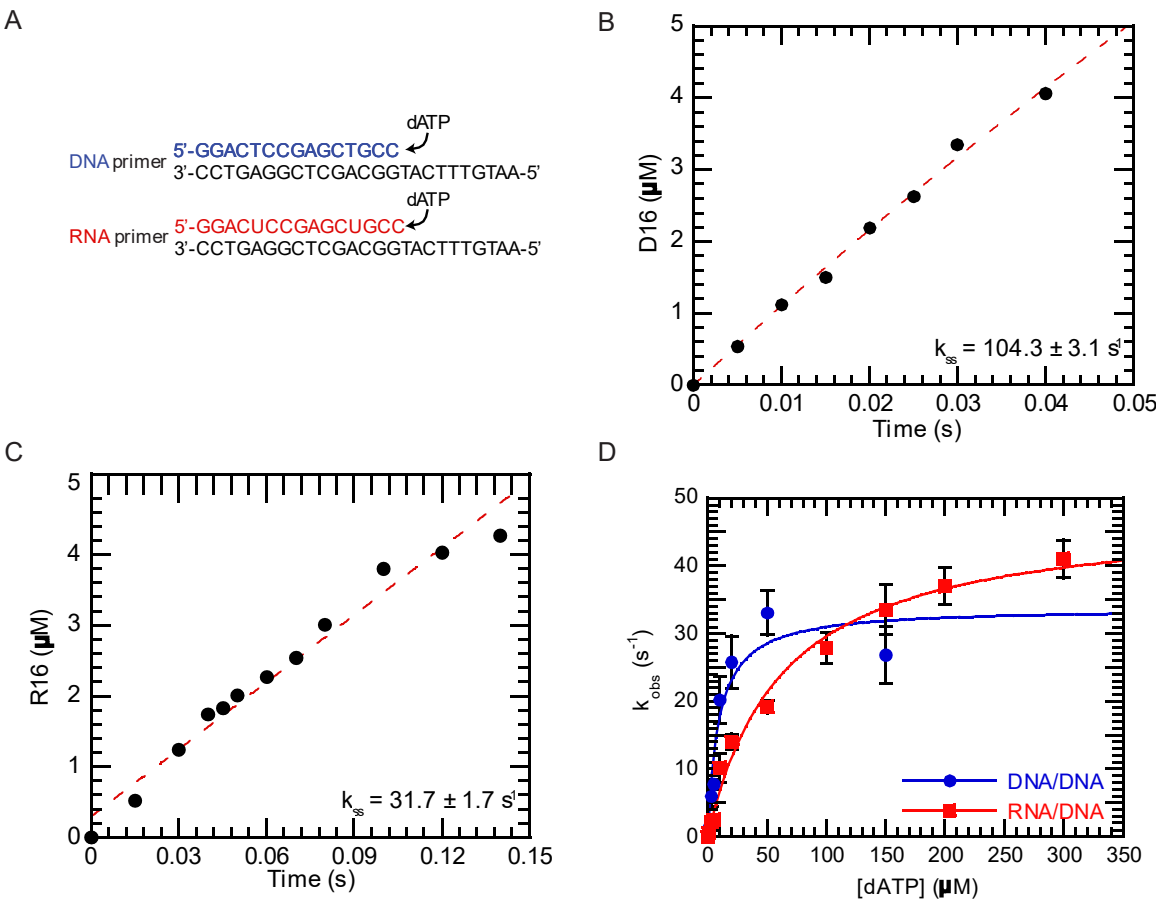


Figure 5.3. Single-incorporation kinetics of hPolα.
A, sequence of template:primers used for transient kinetic assays. *B* and *C*, burst kinetics assay on DNA and RNA primers, respectively. p180core (3 μm) was incubated with radiolabeled DNA (*B*) or RNA (*C*) primer (9 μm) annealed to a DNA template and rapidly mixed with dATP (200 μm). Product formation was plotted *versus* time, and a linear equation was fit to the data (*red dashed line*). *D*, single-nucleotide incorporation assay. p180core (3 μm) was incubated with radiolabeled DNA or RNA primer (100 nm) annealed to a DNA template and rapidly mixed with various concentrations of dATP (1–300 μm). Product formation was plotted *versus* time and fit to a single-exponential equation to obtain the observed rates of product formation. The rates were plotted against concentration of dATP and fit to a hyperbolic equation for both DNA primer (*blue*) and RNA primer (*red*) elongation to obtain K_d and k_{pol} values.

Table 5.2. Incorporation efficiencies of dATP from single incorporation assays.

Primer	k_{pol} (s^{-1})	K_d (μM)	Efficiency (μM^{-1} s^{-1})[a]
DNA	33.8 ± 3.7	9.2 ± 3.4	3.7
RNA	48.0 ± 2.7	62.2 ± 10.4	0.8

[a]Efficiency values are calculated by dividing k_{pol} by K_d

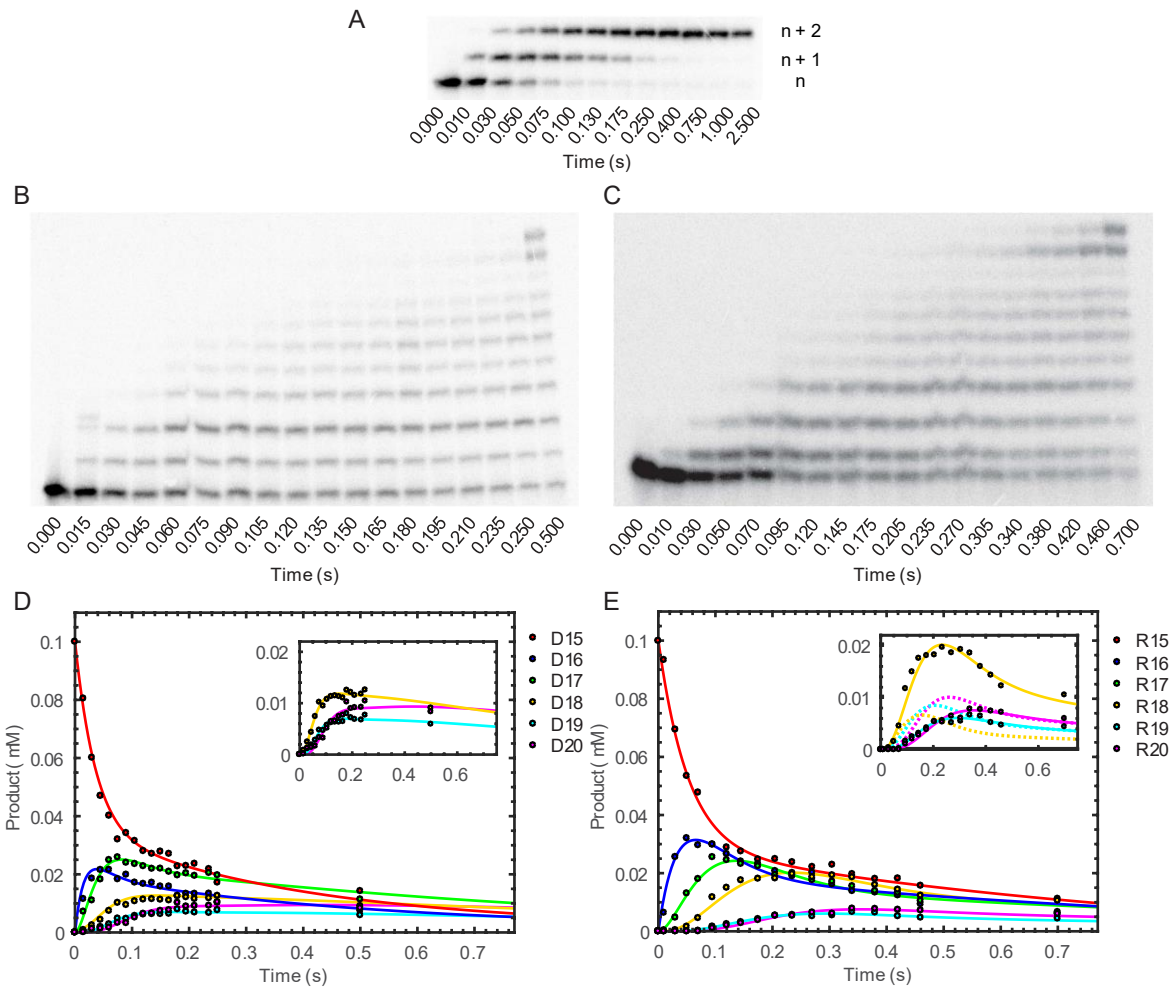


Fig 5.4. Sequential incorporation modeling of hPolα kinetics.
A, a representative gel of the elongation of the DNA primer substrate in the presence of 100 μM dATP and 2 μM dTTP. The *n* band represents the 15-mer primer; *n* + 1, the primer elongated by dATP across template dT; *n* + 2, the primer elongated first by dATP followed by dTTP across template dA. p180core (3 μM) was incubated with radiolabeled DNA template:primer (100 nM) and rapidly mixed with saturating concentrations of dATP (100 μM) and various concentrations of dTTP (1–100 μM). *B*, the time course of DNA primer elongation. p180core (3 μM) was incubated with radiolabeled DNA template:primer (100 nM) and rapidly mixed with saturating concentrations of dNTPs (240 μM). *C*, time course of RNA primer elongation. *D*, processivity modeling for DNA primer elongation. Product formation was plotted *versus* time, and the incorporations were fit to a processive kinetic model using KinTek Global Explorer to provide estimates of the k_{pol} and k_{off} values. The first five incorporations are shown. The *red line* represents the unelongated primer (15-mer). The *inset* shows a *zoomed in* view of the third through fifth incorporations. *E*, processivity modeling for RNA primer elongation. The *dashed line* shown in the *inset* represents the model if the rate of the fourth incorporation of the RNA assay was simulated to be identical to the rate of the fourth incorporation from the DNA processivity assay.

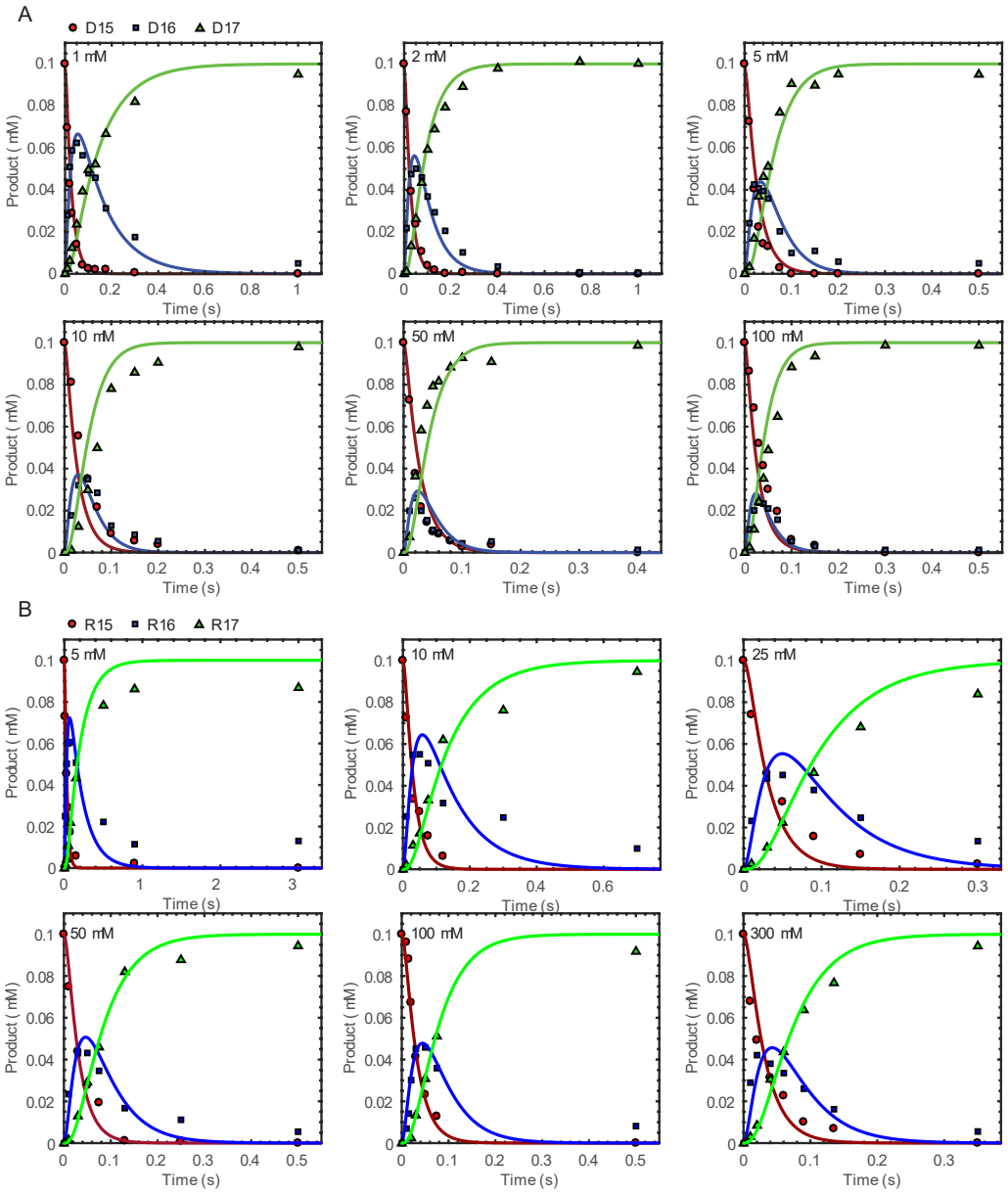


Fig 5.5. Kinetic modeling of a dATP and subsequent dTTP incorporation into a primer by Polα.

A, Modeling data for double incorporation into a DNA primer. Product formation was plotted versus time and the data were fit simultaneously to a kinetic model using KinTek Global Explorer to provide estimates of the Kd and kpol values for dTTP incorporation into a DNA primer. The red line corresponds to the unelongated primer (n), the blue corresponds to the dATP elongated primer (n + 1), and the green corresponds to the primer elongated by dATP and dTTP (n + 2). B, Modeling data for double incorporation into an RNA primer.

Table 5.3. Incorporation efficiencies of dTTP from double incorporation assays.

Primer	k_{pol} (s^{-1})	K_d (μM)	Efficiency (μM^{-1} s^{-1})[a]
DNA	65.8 ± 11.6	5.4[b]	12.2
RNA	23.1 ± 3.5	13.4[b]	1.7

[a]Efficiency values are calculated by dividing k_{pol} by K_d.
[b]K_d values estimated from kinetic modeling do not provide error measurements.

Table 5.4. Kinetic parameters from processivity assays.

Incorporation	**k_{pol} (s^{-1})**		**k_{off} (s^{-1})**	
	DNA	RNA	DNA	RNA
1^{st}	14.0 ± 0.9	14.1 ± 0.7	4.5 ± 1.2	4.5 ± 0.8
2^{nd}	49.9 ± 7.1	16.2 ± 1.1	11.0 ± 3.6	4.0 ± 0.8
3^{rd}	11.2 ± 1.3	11.4 ± 1.0	4.2 ± 1.6	2.1 ± 0.7
4^{th}	19.3 ± 4.5	8.3 ± 1.0	5.8 ± 3.7	0.9 ± 0.8
5^{th}	38.0 ± 16.8	28.1 ± 8.9	6.5 ± 7.6	1.8 ± 2.5
6^{th}	14.6 ± 4.8	20.6 ± 6.2	2.4 ± 3.8	1.8 ± 2.5
7^{th}	14.1 ± 7.4	24.6 ± 9.9	1.2 ± 5.0	2.0 ± 3.4
Average (2^{nd} – 4^{th})	26.8	12.0	7.0	2.3
Average (5^{th} – 7^{th})	22.2	24.4	3.4	1.9

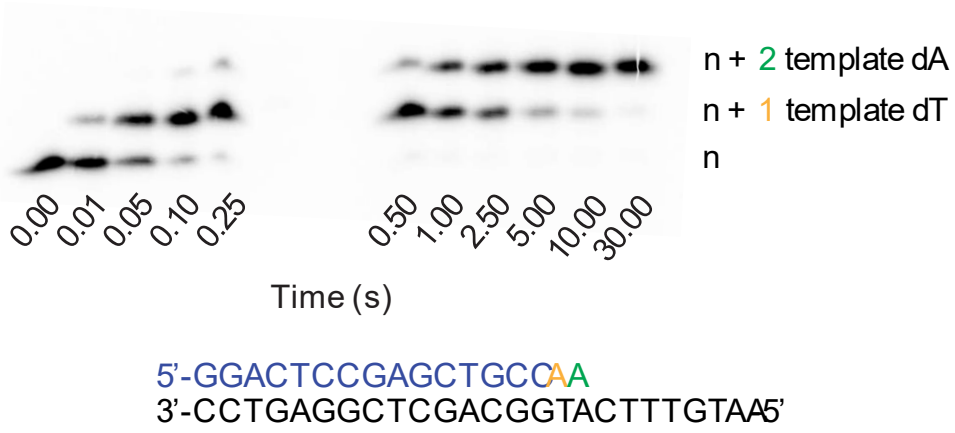


Figure 5.6. Polα displays misincorporation on a rapid time scale.
p180core (3 μM) was incubated with radiolabeled DNA template:primer substrate (10 μM) and rapidly mixed with saturating dATP (500 μM). The gel shows the elongation of the DNA primer in the presence of only dATP. The *n* + 1 *band* corresponds to a correct incorporation of dATP across from the templating dT. The *n* + 2 *band* corresponds to a misincorporation of dATP across from a templating dA. The sequence shows where the incoming dATP is incorporated onto the substrate.

5.4 Discussion

Interestingly, a number of DNA polymerases exhibit burst phase kinetics. The presence of a burst implies that product release is the rate limiting step for a polymerase. In the case of Polα, a burst was not observed, making it a unique case among many characterized polymerases. The fast product release may have implications for its activity; if fast product release is correlated to k_{off} rate, it may be beneficial for Polα to readily release DNA in the context of high misincorporation rates.

In this study, we solved the first B-family polymerase ternary structure in the open conformation. We posit that this may be due to weak interactions of the fingers in the closed form. One potential consequence of this is that the overall step of nucleotide incorporation could be rate limited if the fingers domain is not stable in the closed conformation, which is a prerequisite for catalysis. This could potentially explain the lack of a burst phase.

We have also observed a major difference between the RNA:DNA and DNA:DNA structures of Polα: the RNA primer has a kink in the 4^{th} nucleotide from the 3' end. In addition, processivity experiments showed that the first four nucleotide incorporations on an RNA:DNA substrate are slower compared to latter incorporations, implying that the overall structure of the nucleic acid could affect the activity of Polα and may possibly act as a "sensing" mechanism. Our single turnover experiments also corroborated our processivity assays, demonstrating that overall, incorporation on a DNA:DNA substrate is more efficient. The processivity assays also showed that our estimated k_{off} rates for the DNA:DNA substrate were higher, in the absence of other replication factors. Thus, there is evidence that the identity of the nucleic acid strand could regulate the incorporation of Polα.

Further experiments should be done with additional subunits added to the p180 core, as additional factors may affect kinetic activity. Thus, it would be interesting to add the accessory subunit, primase subunits, or other replication proteins and conduct the burst experiment again.

As mentioned previously, Polα has a significant contribution to the mutation rate *in vivo*, despite the fact that only 1.5% of Polα-synthesized DNA has been shown to be retained in the mature genome. It is astounding that a replication enzyme would display significant misincorporation on a millisecond timescale. While other low-fidelity polymerases exist, having this activity may be beneficial for evolutionary purposes. Further work must be done to understand why nucleotides incorporated by Polα are not completely removed during Okazaki fragment maturation.

Chapter 6. Conclusions

Polymerases continue to play a major role on the therapeutic stage of modern antiviral and chemotherapeutic drugs. It is unsurprising that the first FDA approved COVID-19 treatment, remdesivir, targets the viral RNA-dependent RNA polymerase, given the rich history of nucleoside analogues as antiviral compounds.[183,184] With such a vital role in replication, polymerases play a crucial role in viral reproduction and hyperproliferative conditions like cancer. In order to develop drugs that target these polymerases or understand diseases that arise from dysfunctional polymerase activity, we must closely examine how these polymerases function at a molecular level. To this end, this dissertation examined three different polymerases from a mechanistic and therapeutic perspective.

In Chapter 2, we first examined the role of a novel mitochondrial polymerase, PrimPol, in TFV-derived toxicity. Tenofovir is one of the most prominent HIV drugs, being a component of Truvada (or PrEP). Although initial clinical trials did not demonstrate a severe cause for concern in terms of safety, clinical cases thereafter showed mitochondrial toxicity in a set of HIV+ patients that manifested in the kidneys. In addition, TFV did not appear to significantly inhibit Polγ at an *in vitro* level, challenging the prevailing hypothesis that the major replicative polymerase could cause off-target toxicity. We hypothesized the PrimPol could mediate TFV-toxicity by incorporating the drug into mtDNA. Using gel-based kinetic assays, we showed that PrimPol could use TFV as a substrate, albeit at a low incorporation efficiency. Interestingly, we observed a sequence-dependent effect for TFV incorporation that we are further investigating through crystallography. Despite showing *in vitro* incorporation of TFV, we conducted cellular based assays which showed that PrimPol likely plays a protective role against TFV-derived

toxicity. One explanation for this is the ability of PrimPol to reprime downstream of stalled replication forks.

In Chapter 3, we extended our study conducted in chapter 2. Previous work in our lab showed that a mutation in Polγ affected how well the polymerase could discriminate between NRTIs and natural nucleotides. This set the precedent that mutations in host polymerases could play a role in NRTI-derived toxicity. This study was the impetus for our efforts looking at mutations of PrimPol in HIV+ patients taking tenofovir with mitochondrial toxicity. If PrimPol protects cells against NRTI-toxicity, then loss-of-function mutants may predispose patients against these off-target effects. In our study, we identified the D114N active site mutant in one of our patients. Further biochemical analysis showed complete loss of primase activity and ablated polymerase activity, likely due to overall protein stability and weakened binding to substrates.

In Chapter 4, we sought to understand a discrepancy in the binding mode and efficacy against the Y181C mutant of a class of NNRTIs, the 2-naphthyl phenyl ether compounds. There is a need to continue developing NNRTIs, as pharmacological properties can be improved and the potential of drug resistance to current generation compounds is a threat. Our structural work using crystallography revealed that a subset of these 2-naphthyl compounds bind in a conformation that takes advantage of interactions with an immutable W229. These compounds retain efficacy against the common Y181C mutation compared to 2-naphthyl ethers that bind in an alternative mode. We will take these interactions into consideration in our continued efforts to develop improved NNRTIs.

In Chapter 5, we desired to dissect the mechanism of Polα replication using two replication substrates. It is somewhat surprising that Polα incorporates nucleotides with

low fidelity. This could cause concern if nucleotides inserted into the genome by Polα remain after Okazaki fragment maturation. In fact, recent sequencing efforts have shown that approximately 1.5% of DNA after replication has been synthesized by Polα. In order to fully understand how Polα may contribute to potential disease by misincorporation, we quantitatively assessed its activity on two substrates it encounters during replication: a RNA/DNA and DNA/DNA duplex. Structures of Polα with both substrates revealed a kink in the RNA/DNA duplex, revealing a possible "sensing" mechanism that is dependent on the overall architecture of the duplex substrate. Our kinetic analysis revealed that incorporation (k_{pol}/K_d) on a DNA/DNA substrate is more efficient compared to RNA/DNA, although Polα is more likely to dissociate (k_{off}) from the DNA/DNA as well. We also provided the first pre-steady state analysis of Polα, revealing that Polα does not exhibit a burst, unlike most polymerases. Our initial burst experiments also revealed the astounding rate of misincorporation of Polα on a millisecond timescale, reinforcing the notion that it is a low fidelity enzyme. Further studies will need to be done to understand how other replication factors and proteins affect Polα activity.

In conclusion, these functional and mechanistic studies will contribute to our understanding of the role of polymerases in viral infections and diseases derived from polymerase dysfunction. In furthering our knowledge of these interactions, we can then develop appropriate therapeutics for these diseases.

Publication Record

1. **Duong VN***, Ippolito JA*, Chan AH, Lee WG, Spasov KA, Jorgensen WL, Anderson KS. Structural investigation of 2-naphthyl phenyl ether inhibitors bound to WT and Y181C reverse transcriptase highlights key features of the NNRTI binding site. Protein Sci. 2020 Jul 8;29(9):1902–10. doi: 10.1002/pro.3910. Epub ahead of print. PMID: 32643196; PMCID: PMC7454559.

2. **Duong VN**, Zhou L, Martínez-Jiménez MI, He L, Cosme M, Blanco L, Paintsil E, Anderson KS. Identifying the role of PrimPol in TDF-induced toxicity and implications of its loss of function mutation in an HIV+ patient. Sci Rep. 2020 Jun 9;10(1):9343. doi: 10.1038/s41598-020-66153-z. PMID: 32518272; PMCID: PMC7283272.

3. Baranovskiy AG*, **Duong VN***, Babayeva ND, Zhang Y, Pavlov YI, Anderson KS, Tahirov TH. Activity and fidelity of human DNA polymerase α depend on primer structure. J Biol Chem. 2018 May 4;293(18):6824-6843. doi: 10.1074/jbc.RA117.001074. Epub 2018 Mar 19. PMID: 29555682; PMCID: PMC5936803.

*Denote co-first authorship

References

1 Bebenek, K. & Kunkel, T. A. Functions of DNA polymerases. *Adv Protein Chem* **69**, 137-165, doi:10.1016/S0065-3233(04)69005-X (2004).

2 Hubscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. *Annu Rev Biochem* **71**, 133-163, doi:10.1146/annurev.biochem.71.090501.150041 (2002).

3 Bollum, F. J. & Potter, V. R. Incorporation of thymidine into deoxyribonucleic acid by enzymes from rat tissues. *J Biol Chem* **233**, 478-482 (1958).

4 Chang, L. M. & Bollum, F. J. Low molecular weight deoxyribonucleic acid polymerase in mammalian cells. *J Biol Chem* **246**, 5835-5837 (1971).

5 Kalf, G. F. & Ch'ih, J. J. Purification and properties of deoxyribonucleic acid polymerase from rat liver mitochondria. *J Biol Chem* **243**, 4904-4916 (1968).

6 Boulet, A., Simon, M., Faye, G., Bauer, G. A. & Burgers, P. M. Structure and function of the Saccharomyces cerevisiae CDC2 gene encoding the large subunit of DNA polymerase III. *EMBO J* **8**, 1849-1854 (1989).

7 Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K. & Sugino, A. A third essential DNA polymerase in S. cerevisiae. *Cell* **62**, 1143-1151, doi:10.1016/0092-8674(90)90391-q (1990).

8 Lehman, I. R. & Kaguni, L. S. DNA polymerase alpha. *J Biol Chem* **264**, 4265-4268 (1989).

9 Conaway, R. C. & Lehman, I. R. A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos. *Proc Natl Acad Sci U S A* **79**, 2523-2527, doi:10.1073/pnas.79.8.2523 (1982).

10 Nick McElhinny, S. A., Kissling, G. E. & Kunkel, T. A. Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}. *Proc Natl Acad Sci U S A* **107**, 21070-21075, doi:10.1073/pnas.1013048107 (2010).

11 Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E. & Kunkel, T. A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127-130, doi:10.1126/science.1144067 (2007).

12 Garg, P., Stith, C. M., Sabouri, N., Johansson, E. & Burgers, P. M. Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev* **18**, 2764-2773, doi:10.1101/gad.1252304 (2004).

13 Jin, Y. H., Ayyagari, R., Resnick, M. A., Gordenin, D. A. & Burgers, P. M. Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick. *J Biol Chem* **278**, 1626-1633, doi:10.1074/jbc.M209803200 (2003).

14 Jin, Y. H. *et al.* The 3'-->5' exonuclease of DNA polymerase delta can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proc Natl Acad Sci U S A* **98**, 5122-5127, doi:10.1073/pnas.091095198 (2001).

15 Tran, H. T., Gordenin, D. A. & Resnick, M. A. The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae. *Mol Cell Biol* **19**, 2000-2007, doi:10.1128/mcb.19.3.2000 (1999).

16 Chilkova, O. *et al.* The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res* **35**, 6588-6597, doi:10.1093/nar/gkm741 (2007).

17 Langston, L. D. & O'Donnell, M. DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. *J Biol Chem* **283**, 29522-29531, doi:10.1074/jbc.M804488200 (2008).
18 Sobol, R. W. & Wilson, S. H. Mammalian DNA beta-polymerase in base excision repair of alkylation damage. *Prog Nucleic Acid Res Mol Biol* **68**, 57-74, doi:10.1016/s0079-6603(01)68090-5 (2001).
19 Boiteux, S. & Guillet, M. Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair (Amst)* **3**, 1-12, doi:10.1016/j.dnarep.2003.10.002 (2004).
20 Sugo, N., Aratani, Y., Nagashima, Y., Kubota, Y. & Koyama, H. Neonatal lethality with abnormal neurogenesis in mice deficient in DNA polymerase beta. *EMBO J* **19**, 1397-1404, doi:10.1093/emboj/19.6.1397 (2000).
21 Kaguni, L. S. DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem* **73**, 293-320, doi:10.1146/annurev.biochem.72.121801.161455 (2004).
22 Arnold, J. J., Smidansky, E. D., Moustafa, I. M. & Cameron, C. E. Human mitochondrial RNA polymerase: structure-function, mechanism and inhibition. *Biochim Biophys Acta* **1819**, 948-960, doi:10.1016/j.bbagrm.2012.04.002 (2012).
23 Bolden, A., Noy, G. P. & Weissbach, A. DNA polymerase of mitochondria is a gamma-polymerase. *J Biol Chem* **252**, 3351-3356 (1977).
24 Longley, M. J., Prasad, R., Srivastava, D. K., Wilson, S. H. & Copeland, W. C. Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro. *Proc Natl Acad Sci U S A* **95**, 12244-12248, doi:10.1073/pnas.95.21.12244 (1998).
25 Macao, B. *et al.* The exonuclease activity of DNA polymerase gamma is required for ligation during mitochondrial DNA replication. *Nat Commun* **6**, 7303, doi:10.1038/ncomms8303 (2015).
26 Longley, M. J., Nguyen, D., Kunkel, T. A. & Copeland, W. C. The fidelity of human DNA polymerase gamma with and without exonucleolytic proofreading and the p55 accessory subunit. *J Biol Chem* **276**, 38555-38562, doi:10.1074/jbc.M105230200 (2001).
27 Johnson, A. A. *et al.* Toxicity of antiviral nucleoside analogs and the human mitochondrial DNA polymerase. *J Biol Chem* **276**, 40847-40857, doi:10.1074/jbc.M106743200 (2001).
28 Hudson, G. & Chinnery, P. F. Mitochondrial DNA polymerase-gamma and human disease. *Hum Mol Genet* **15 Spec No 2**, R244-252, doi:10.1093/hmg/ddl233 (2006).
29 Garcia-Gomez, S. *et al.* PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell* **52**, 541-553, doi:10.1016/j.molcel.2013.09.025 (2013).
30 Bianchi, J. *et al.* PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol Cell* **52**, 566-573, doi:10.1016/j.molcel.2013.10.035 (2013).
31 Guilliam, T. A. & Doherty, A. J. PrimPol-Prime Time to Reprime. *Genes (Basel)* **8**, doi:10.3390/genes8010020 (2017).
32 Schiavone, D. *et al.* PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. *Mol Cell* **61**, 161-169, doi:10.1016/j.molcel.2015.10.038 (2016).
33 Kobayashi, K. *et al.* Repriming by PrimPol is critical for DNA replication restart downstream of lesions and chain-terminating nucleosides. *Cell Cycle* **15**, 1997-2008, doi:10.1080/15384101.2016.1191711 (2016).
34 Mouron, S. *et al.* Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat Struct Mol Biol* **20**, 1383-1389, doi:10.1038/nsmb.2719 (2013).

35 Keen, B. A., Jozwiakowski, S. K., Bailey, L. J., Bianchi, J. & Doherty, A. J. Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res* **42**, 5830-5845, doi:10.1093/nar/gku214 (2014).
36 Waters, L. S. *et al.* Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev* **73**, 134-154, doi:10.1128/MMBR.00034-08 (2009).
37 Goodman, M. F. & Woodgate, R. Translesion DNA polymerases. *Cold Spring Harb Perspect Biol* **5**, a010363, doi:10.1101/cshperspect.a010363 (2013).
38 Wood, R. D. & Doublie, S. DNA polymerase theta (POLQ), double-strand break repair, and cancer. *DNA Repair (Amst)* **44**, 22-32, doi:10.1016/j.dnarep.2016.05.003 (2016).
39 Garcia-Diaz, M. *et al.* DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J Biol Chem* **277**, 13184-13191, doi:10.1074/jbc.M111601200 (2002).
40 Nick McElhinny, S. A. & Ramsden, D. A. Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol Cell Biol* **23**, 2309-2315, doi:10.1128/mcb.23.7.2309-2315.2003 (2003).
41 Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a006882 (2012).
42 De Clercq, E. The nucleoside reverse transcriptase inhibitors, nonnucleoside reverse transcriptase inhibitors, and protease inhibitors in the treatment of HIV infections (AIDS). *Adv Pharmacol* **67**, 317-358, doi:10.1016/B978-0-12-405880-4.00009-3 (2013).
43 Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *J Biol Chem* **274**, 17395-17398, doi:10.1074/jbc.274.25.17395 (1999).
44 Gangurde, R., Kaushik, N., Singh, K. & Modak, M. J. A carboxylate triad is essential for the polymerase activity of Escherichia coli DNA polymerase I (Klenow fragment). Presence of two functional triads at the catalytic center. *J Biol Chem* **275**, 19685-19692, doi:10.1074/jbc.M002307200 (2000).
45 Berdis, A. J. Mechanisms of DNA polymerases. *Chem Rev* **109**, 2862-2879, doi:10.1021/cr800530b (2009).
46 Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. & Benkovic, S. J. Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry* **24**, 4010-4018, doi:10.1021/bi00336a031 (1985).
47 Bryant, F. R., Johnson, K. A. & Benkovic, S. J. Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry* **22**, 3537-3546, doi:10.1021/bi00284a001 (1983).
48 Johnson, K. A. Advances in transient-state kinetics. *Curr Opin Biotechnol* **9**, 87-89, doi:10.1016/s0958-1669(98)80089-x (1998).
49 Joyce, C. M. Techniques used to study the DNA polymerase reaction pathway. *Biochim Biophys Acta* **1804**, 1032-1040, doi:10.1016/j.bbapap.2009.07.021 (2010).
50 Raper, A. T., Reed, A. J. & Suo, Z. Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem Rev* **118**, 6000-6025, doi:10.1021/acs.chemrev.7b00685 (2018).
51 Johnson, K. A. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim Biophys Acta* **1804**, 1041-1048, doi:10.1016/j.bbapap.2010.01.006 (2010).
52 Doublie, S., Sawaya, M. R. & Ellenberger, T. An open and closed case for all polymerases. *Structure* **7**, R31-35, doi:10.1016/S0969-2126(99)80017-3 (1999).
53 Joyce, C. M. *et al.* Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* **47**, 6103-6116, doi:10.1021/bi7021848 (2008).
54 Lange, S. S., Takata, K. & Wood, R. D. DNA polymerases and cancer. *Nat Rev Cancer* **11**, 96-110, doi:10.1038/nrc2998 (2011).

55 Berdis, A. J. DNA polymerases as therapeutic targets. *Biochemistry* **47**, 8253-8260, doi:10.1021/bi801179f (2008).
56 Havelka, A. M., Berndtsson, M., Olofsson, M. H., Shoshan, M. C. & Linder, S. Mechanisms of action of DNA-damaging anticancer drugs in treatment of carcinomas: is acute apoptosis an "off-target" effect? *Mini Rev Med Chem* **7**, 1035-1039, doi:10.2174/138955707782110196 (2007).
57 Baldwin, E. L. & Osheroff, N. Etoposide, topoisomerase II and cancer. *Curr Med Chem Anticancer Agents* **5**, 363-372, doi:10.2174/1568011054222364 (2005).
58 McGuire, J. J. Anticancer antifolates: current status and future directions. *Curr Pharm Des* **9**, 2593-2613, doi:10.2174/1381612033453712 (2003).
59 De Clercq, E. & Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **29**, 695-747, doi:10.1128/CMR.00102-15 (2016).
60 Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat Rev Drug Discov* **12**, 447-464, doi:10.1038/nrd4010 (2013).
61 Murakami-Nakai, C. *et al.* The effects of dehydroaltenusin, a novel mammalian DNA polymerase alpha inhibitor, on cell proliferation and cell cycle progression. *Biochim Biophys Acta* **1674**, 193-199, doi:10.1016/j.bbagen.2004.06.016 (2004).
62 Lowe, D. M. *et al.* HIV-1 reverse transcriptase: crystallization and analysis of domain structure by limited proteolysis. *Biochemistry* **27**, 8884-8889, doi:10.1021/bi00425a002 (1988).
63 Le Grice, S. F., Naas, T., Wohlgensinger, B. & Schatz, O. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J* **10**, 3905-3911 (1991).
64 Restle, T., Muller, B. & Goody, R. S. Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. *J Biol Chem* **265**, 8986-8988 (1990).
65 Harris, D., Lee, R., Misra, H. S., Pandey, P. K. & Pandey, V. N. The p51 subunit of human immunodeficiency virus type 1 reverse transcriptase is essential in loading the p66 subunit on the template primer. *Biochemistry* **37**, 5903-5908, doi:10.1021/bi9728452 (1998).
66 Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790, doi:10.1126/science.1377403 (1992).
67 Hsiou, Y. *et al.* Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4**, 853-860, doi:10.1016/s0969-2126(96)00091-3 (1996).
68 Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* **267**, 988-993, doi:10.1126/science.7532321 (1995).
69 de Bethune, M. P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res* **85**, 75-90, doi:10.1016/j.antiviral.2009.09.008 (2010).
70 Das, K., Martinez, S. E., Bauman, J. D. & Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* **19**, 253-259, doi:10.1038/nsmb.2223 (2012).

71 Margolis, A. M., Heverling, H., Pham, P. A. & Stolbach, A. A review of the toxicity of HIV medications. *J Med Toxicol* **10**, 26-39, doi:10.1007/s13181-013-0325-8 (2014).
72 Kohler, J. J. & Lewis, W. A brief overview of mechanisms of mitochondrial toxicity from NRTIs. *Environ Mol Mutagen* **48**, 166-172, doi:10.1002/em.20223 (2007).
73 Lewis, W., Day, B. J. & Copeland, W. C. Mitochondrial toxicity of NRTI antiviral drugs: an integrated cellular perspective. *Nat Rev Drug Discov* **2**, 812-822, doi:10.1038/nrd1201 (2003).
74 Lewis, W. & Dalakas, M. C. Mitochondrial toxicity of antiviral drugs. *Nat Med* **1**, 417-422, doi:10.1038/nm0595-417 (1995).
75 Lewis, W., Copeland, W. C. & Day, B. J. Mitochondrial dna depletion, oxidative stress, and mutation: mechanisms of dysfunction from nucleoside reverse transcriptase inhibitors. *Lab Invest* **81**, 777-790, doi:10.1038/labinvest.3780288 (2001).
76 Apostolova, N., Blas-Garcia, A. & Esplugues, J. V. Mitochondrial interference by anti-HIV drugs: mechanisms beyond Pol-gamma inhibition. *Trends Pharmacol Sci* **32**, 715-725, doi:10.1016/j.tips.2011.07.007 (2011).
77 Fernandez-Fernandez, B. *et al.* Tenofovir nephrotoxicity: 2011 update. *AIDS Res Treat* **2011**, 354908, doi:10.1155/2011/354908 (2011).
78 Rechkoblit, O. *et al.* Structure and mechanism of human PrimPol, a DNA polymerase with primase activity. *Sci Adv* **2**, e1601317, doi:10.1126/sciadv.1601317 (2016).
79 Guilliam, T. A. *et al.* Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. *Nucleic Acids Res* **43**, 1056-1068, doi:10.1093/nar/gku1321 (2015).
80 Stojkovic, G. *et al.* Oxidative DNA damage stalls the human mitochondrial replisome. *Sci Rep* **6**, 28942, doi:10.1038/srep28942 (2016).
81 Mislak, A. C. & Anderson, K. S. Insights into the Molecular Mechanism of Polymerization and Nucleoside Reverse Transcriptase Inhibitor Incorporation by Human PrimPol. *Antimicrob Agents Chemother* **60**, 561-569, doi:10.1128/AAC.02270-15 (2016).
82 Collaboration, H.-C. *et al.* The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *AIDS* **24**, 123-137, doi:10.1097/QAD.0b013e3283324283 (2010).
83 Fischl, M. A. *et al.* The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med* **317**, 185-191, doi:10.1056/NEJM198707233170401 (1987).
84 Palella, F. J., Jr. *et al.* Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* **338**, 853-860, doi:10.1056/NEJM199803263381301 (1998).
85 Palella, F. J., Jr. *et al.* Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study. *J Acquir Immune Defic Syndr* **43**, 27-34, doi:10.1097/01.qai.0000233310.90484.16 (2006).
86 Antiretroviral Therapy Cohort, C. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet* **372**, 293-299, doi:10.1016/S0140-6736(08)61113-7 (2008).
87 Bailey, C. M. & Anderson, K. S. A mechanistic view of human mitochondrial DNA polymerase gamma: providing insight into drug toxicity and mitochondrial disease. *Biochim Biophys Acta* **1804**, 1213-1222, doi:10.1016/j.bbapap.2010.01.007 (2010).
88 Moyle, G. Clinical manifestations and management of antiretroviral nucleoside analog-related mitochondrial toxicity. *Clin Ther* **22**, 911-936; discussion 898, doi:10.1016/S0149-2918(00)80064-8 (2000).

89 Montaner, J. S. *et al.* Nucleoside-related mitochondrial toxicity among HIV-infected patients receiving antiretroviral therapy: insights from the evaluation of venous lactic acid and peripheral blood mitochondrial DNA. *Clin Infect Dis* **38 Suppl 2**, S73-79, doi:10.1086/381449 (2004).
90 Brinkman, K., ter Hofstede, H. J., Burger, D. M., Smeitink, J. A. & Koopmans, P. P. Adverse effects of reverse transcriptase inhibitors: mitochondrial toxicity as common pathway. *AIDS* **12**, 1735-1744, doi:10.1097/00002030-199814000-00004 (1998).
91 Tourret, J., Deray, G. & Isnard-Bagnis, C. Tenofovir effect on the kidneys of HIV-infected patients: a double-edged sword? *J Am Soc Nephrol* **24**, 1519-1527, doi:10.1681/ASN.2012080857 (2013).
92 Kohler, J. J. *et al.* Tenofovir renal toxicity targets mitochondria of renal proximal tubules. *Lab Invest* **89**, 513-519, doi:10.1038/labinvest.2009.14 (2009).
93 Lee, H., Hanes, J. & Johnson, K. A. Toxicity of nucleoside analogues used to treat AIDS and the selectivity of the mitochondrial DNA polymerase. *Biochemistry* **42**, 14711-14719, doi:10.1021/bi035596s (2003).
94 Birkus, G. *et al.* Tenofovir diphosphate is a poor substrate and a weak inhibitor of rat DNA polymerases alpha, delta, and epsilon*. *Antimicrob Agents Chemother* **46**, 1610-1613, doi:10.1128/aac.46.5.1610-1613.2002 (2002).
95 Young, M. J. Off-Target Effects of Drugs that Disrupt Human Mitochondrial DNA Maintenance. *Front Mol Biosci* **4**, 74, doi:10.3389/fmolb.2017.00074 (2017).
96 Parker, W. B. *et al.* Metabolism of carbovir, a potent inhibitor of human immunodeficiency virus type 1, and its effects on cellular metabolism. *Antimicrob Agents Chemother* **37**, 1004-1009, doi:10.1128/aac.37.5.1004 (1993).
97 Chang, C. N., Skalski, V., Zhou, J. H. & Cheng, Y. C. Biochemical pharmacology of (+)- and (-)-2',3'-dideoxy-3'-thiacytidine as anti-hepatitis B virus agents. *J Biol Chem* **267**, 22414-22420 (1992).
98 Wan, L. *et al.* hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *EMBO Rep* **14**, 1104-1112, doi:10.1038/embor.2013.159 (2013).
99 Iyer, L. M., Koonin, E. V., Leipe, D. D. & Aravind, L. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res* **33**, 3875-3896, doi:10.1093/nar/gki702 (2005).
100 Svikovic, S. *et al.* R-loop formation during S phase is restricted by PrimPol-mediated repriming. *EMBO J* **38**, doi:10.15252/embj.201899793 (2019).
101 Rooney, J. P. *et al.* PCR based determination of mitochondrial DNA copy number in multiple species. *Methods Mol Biol* **1241**, 23-38, doi:10.1007/978-1-4939-1875-1_3 (2015).
102 Torregrosa-Munumer, R. *et al.* PrimPol is required for replication reinitiation after mtDNA damage. *Proc Natl Acad Sci U S A* **114**, 11398-11403, doi:10.1073/pnas.1705367114 (2017).
103 Maffioli, S. I. *et al.* Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **169**, 1240-1248 e1223, doi:10.1016/j.cell.2017.05.042 (2017).
104 Wieser, M. *et al.* hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. *Am J Physiol Renal Physiol* **295**, F1365-1375, doi:10.1152/ajprenal.90405.2008 (2008).
105 Zhao, X. *et al.* Tenofovir and adefovir down-regulate mitochondrial chaperone TRAP1 and succinate dehydrogenase subunit B to metabolically reprogram glucose metabolism and induce nephrotoxicity. *Sci Rep* **7**, 46344, doi:10.1038/srep46344 (2017).

106 Zhang, J. *et al.* Measuring energy metabolism in cultured cells, including human pluripotent stem cells and differentiated cells. *Nat Protoc* **7**, 1068-1085, doi:10.1038/nprot.2012.048 (2012).
107 White, A. J. Mitochondrial toxicity and HIV therapy. *Sex Transm Infect* **77**, 158-173, doi:10.1136/sti.77.3.158 (2001).
108 Quinet, A. *et al.* PRIMPOL-Mediated Adaptive Response Suppresses Replication Fork Reversal in BRCA-Deficient Cells. *Mol Cell* **77**, 461-474 e469, doi:10.1016/j.molcel.2019.10.008 (2020).
109 Martinez-Jimenez, M. I., Lahera, A. & Blanco, L. Human PrimPol activity is enhanced by RPA. *Sci Rep* **7**, 783, doi:10.1038/s41598-017-00958-3 (2017).
110 Guilliam, T. A. *et al.* Molecular basis for PrimPol recruitment to replication forks by RPA. *Nat Commun* **8**, 15222, doi:10.1038/ncomms15222 (2017).
111 Tian, L., Kim, M. S., Li, H., Wang, J. & Yang, W. Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid. *Proc Natl Acad Sci U S A* **115**, 507-512, doi:10.1073/pnas.1719746115 (2018).
112 Salter, J. D. & Smith, H. C. Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends Biochem Sci* **43**, 606-622, doi:10.1016/j.tibs.2018.04.013 (2018).
113 Hoehener, C., Hug, I. & Nowacki, M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell* **173**, 234-247 e237, doi:10.1016/j.cell.2018.02.029 (2018).
114 Chuprina, V. P. *et al.* Sequence effects on local DNA topology. *Proc Natl Acad Sci U S A* **88**, 9087-9091, doi:10.1073/pnas.88.20.9087 (1991).
115 De Clercq, E. Clinical potential of the acyclic nucleoside phosphonates cidofovir, adefovir, and tenofovir in treatment of DNA virus and retrovirus infections. *Clin Microbiol Rev* **16**, 569-596, doi:10.1128/cmr.16.4.569-596.2003 (2003).
116 Anderson, A. C. The process of structure-based drug design. *Chem Biol* **10**, 787-797, doi:10.1016/j.chembiol.2003.09.002 (2003).
117 Li, M. *et al.* The DNA Polymerase Gamma R953C Mutant Is Associated with Antiretroviral Therapy-Induced Mitochondrial Toxicity. *Antimicrob Agents Chemother* **60**, 5608-5611, doi:10.1128/AAC.00976-16 (2016).
118 Selvaraj, S. *et al.* Antiretroviral therapy-induced mitochondrial toxicity: potential mechanisms beyond polymerase-gamma inhibition. *Clin Pharmacol Ther* **96**, 110-120, doi:10.1038/clpt.2014.64 (2014).
119 Tropea, J. E., Cherry, S. & Waugh, D. S. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* **498**, 297-307, doi:10.1007/978-1-59745-196-3_19 (2009).
120 Calvo, P. A. *et al.* The invariant glutamate of human PrimPol DxE motif is critical for its Mn(2+)-dependent distinctive activities. *DNA Repair (Amst)* **77**, 65-75, doi:10.1016/j.dnarep.2019.03.006 (2019).
121 Pilzecker, B. *et al.* PrimPol prevents APOBEC/AID family mediated DNA mutagenesis. *Nucleic Acids Res* **44**, 4734-4744, doi:10.1093/nar/gkw123 (2016).
122 Martinez-Jimenez, M. I., Calvo, P. A., Garcia-Gomez, S., Guerra-Gonzalez, S. & Blanco, L. The Zn-finger domain of human PrimPol is required to stabilize the initiating nucleotide during DNA priming. *Nucleic Acids Res* **46**, 4138-4151, doi:10.1093/nar/gky230 (2018).
123 Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511-525, doi:10.1021/bi00216a029 (1991).

124 Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J Biol Chem* **267**, 25988-25997 (1992).
125 Graves, S. W., Johnson, A. A. & Johnson, K. A. Expression, purification, and initial kinetic characterization of the large subunit of the human mitochondrial DNA polymerase. *Biochemistry* **37**, 6050-6058, doi:10.1021/bi972685u (1998).
126 Keen, B. A., Bailey, L. J., Jozwiakowski, S. K. & Doherty, A. J. Human PrimPol mutation associated with high myopia has a DNA replication defect. *Nucleic Acids Res* **42**, 12102-12111, doi:10.1093/nar/gku879 (2014).
127 Ryder, S. P., Recht, M. I. & Williamson, J. R. Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol Biol* **488**, 99-115, doi:10.1007/978-1-60327-475-3_7 (2008).
128 Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221, doi:10.1038/nprot.2007.321 (2007).
129 Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* **79**, 28 29 21-28 29 14, doi:10.1002/0471140864.ps2809s79 (2015).
130 Zafar, M. K., Ketkar, A., Lodeiro, M. F., Cameron, C. E. & Eoff, R. L. Kinetic analysis of human PrimPol DNA polymerase activity reveals a generally error-prone enzyme capable of accurately bypassing 7,8-dihydro-8-oxo-2'-deoxyguanosine. *Biochemistry* **53**, 6584-6594, doi:10.1021/bi501024u (2014).
131 Xu, W., Zhao, W., Morehouse, N., Tree, M. O. & Zhao, L. Divalent Cations Alter the Rate-Limiting Step of PrimPol-Catalyzed DNA Elongation. *J Mol Biol* **431**, 673-686, doi:10.1016/j.jmb.2019.01.002 (2019).
132 Mahy, M. *et al.* HIV estimates through 2018: data for decision-making. *AIDS* **33 Suppl 3**, S203-S211, doi:10.1097/QAD.0000000000002321 (2019).
133 Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* **2**, a007161, doi:10.1101/cshperspect.a007161 (2012).
134 Moore, R. D. & Chaisson, R. E. Natural history of HIV infection in the era of combination antiretroviral therapy. *AIDS* **13**, 1933-1942, doi:10.1097/00002030-199910010-00017 (1999).
135 Usach, I., Melis, V. & Peris, J. E. Non-nucleoside reverse transcriptase inhibitors: a review on pharmacokinetics, pharmacodynamics, safety and tolerability. *J Int AIDS Soc* **16**, 1-14, doi:10.7448/IAS.16.1.18567 (2013).
136 Holec, A. D., Mandal, S., Prathipati, P. K. & Destache, C. J. Nucleotide Reverse Transcriptase Inhibitors: A Thorough Review, Present Status and Future Perspective as HIV Therapeutics. *Curr HIV Res* **15**, 411-421, doi:10.2174/1570162X15666171120110145 (2017).
137 Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **84**, 9864-9878, doi:10.1128/JVI.00915-10 (2010).
138 Reynolds, C., de Koning, C. B., Pelly, S. C., van Otterlo, W. A. & Bode, M. L. In search of a treatment for HIV--current therapies and the role of non-nucleoside reverse transcriptase inhibitors (NNRTIs). *Chem Soc Rev* **41**, 4657-4670, doi:10.1039/c2cs35058k (2012).
139 Casado, J. L. *et al.* Non-nucleoside reverse transcriptase inhibitor resistance among patients failing a nevirapine plus protease inhibitor-containing regimen. *AIDS* **14**, F1-7, doi:10.1097/00002030-200001280-00001 (2000).

140 Mackie, N. in *Antiretroviral Resistance in Clinical Practice* (ed A. M. Geretti) (2006).
141 Shafer, R. W. *et al.* HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* **21**, 215-223, doi:10.1097/QAD.0b013e328011e691 (2007).
142 Minuto, J. J. & Haubrich, R. Etravirine: a second-generation NNRTI for treatment-experienced adults with resistant HIV-1 infection. *Futur HIV Ther* **2**, 525-537, doi:10.2217/17469600.2.6.525 (2008).
143 Jorgensen, W. L. Computer-aided discovery of anti-HIV agents. *Bioorg Med Chem* **24**, 4768-4778, doi:10.1016/j.bmc.2016.07.039 (2016).
144 Saravanan, S. *et al.* Etravirine and Rilpivirine Drug Resistance Among HIV-1 Subtype C Infected Children Failing Non-Nucleoside Reverse Transcriptase Inhibitor-Based Regimens in South India. *AIDS Res Hum Retroviruses* **33**, 567-574, doi:10.1089/AID.2016.0133 (2017).
145 Maiga, A. I. *et al.* Resistance-associated mutations to etravirine (TMC-125) in antiretroviral-naive patients infected with non-B HIV-1 subtypes. *Antimicrob Agents Chemother* **54**, 728-733, doi:10.1128/AAC.01335-09 (2010).
146 Scholler-Gyure, M. *et al.* A pharmacokinetic study of etravirine (TMC125) co-administered with ranitidine and omeprazole in HIV-negative volunteers. *Br J Clin Pharmacol* **66**, 508-516, doi:10.1111/j.1365-2125.2008.03214.x (2008).
147 Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc Chem Res* **42**, 724-733, doi:10.1021/ar800236t (2009).
148 Lee, W. G., Chan, A. H., Spasov, K. A., Anderson, K. S. & Jorgensen, W. L. Design, Conformation, and Crystallography of 2-Naphthyl Phenyl Ethers as Potent Anti-HIV Agents. *ACS Med Chem Lett* **7**, 1156-1160, doi:10.1021/acsmedchemlett.6b00390 (2016).
149 Das, K. *et al.* High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: strategic flexibility explains potency against resistance mutations. *Proc Natl Acad Sci U S A* **105**, 1466-1471, doi:10.1073/pnas.0711209105 (2008).
150 Frey, K. M. *et al.* Structure-based evaluation of non-nucleoside inhibitors with improved potency and solubility that target HIV reverse transcriptase variants. *J Med Chem* **58**, 2737-2745, doi:10.1021/jm501908a (2015).
151 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).
152 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
153 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
154 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
155 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
156 Terwilliger, T. C. *et al.* Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr D Biol Crystallogr* **64**, 515-524, doi:10.1107/S0907444908004319 (2008).
157 Morin, A. *et al.* Collaboration gets the most out of software. *Elife* **2**, e01456, doi:10.7554/eLife.01456 (2013).

158 Chan, A. H. *et al.* Covalent inhibitors for eradication of drug-resistant HIV-1 reverse transcriptase: From design to protein crystallography. *Proc Natl Acad Sci U S A* **114**, 9725-9730, doi:10.1073/pnas.1711463114 (2017).
159 Kudalkar, S. N. *et al.* Structural and pharmacological evaluation of a novel non-nucleoside reverse transcriptase inhibitor as a promising long acting nanoformulation for treating HIV. *Antiviral Res* **167**, 110-116, doi:10.1016/j.antiviral.2019.04.010 (2019).
160 Lee, W. G. *et al.* Picomolar Inhibitors of HIV-1 Reverse Transcriptase: Design and Crystallography of Naphthyl Phenyl Ethers. *ACS Med Chem Lett* **5**, 1259-1262, doi:10.1021/ml5003713 (2014).
161 Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA Replication Fork. *Annu Rev Biochem* **86**, 417-438, doi:10.1146/annurev-biochem-061516-044709 (2017).
162 Lujan, S. A., Williams, J. S. & Kunkel, T. A. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol* **26**, 640-654, doi:10.1016/j.tcb.2016.04.012 (2016).
163 Pavlov, Y. I. & Shcherbakova, P. V. DNA polymerases at the eukaryotic fork-20 years later. *Mutat Res* **685**, 45-53, doi:10.1016/j.mrfmmm.2009.08.002 (2010).
164 Nunez-Ramirez, R. *et al.* Flexible tethering of primase and DNA Pol alpha in the eukaryotic primosome. *Nucleic Acids Res* **39**, 8187-8199, doi:10.1093/nar/gkr534 (2011).
165 Baranovskiy, A. G. & Tahirov, T. H. Elaborated Action of the Human Primosome. *Genes (Basel)* **8**, doi:10.3390/genes8020062 (2017).
166 Baranovskiy, A. G. *et al.* Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome. *J Biol Chem* **291**, 10006-10020, doi:10.1074/jbc.M116.717405 (2016).
167 Reijns, M. A. M. *et al.* Lagging-strand replication shapes the mutational landscape of the genome. *Nature* **518**, 502-506, doi:10.1038/nature14183 (2015).
168 Starokadomskyy, P. *et al.* DNA polymerase-alpha regulates the activation of type I interferons through cytosolic RNA:DNA synthesis. *Nat Immunol* **17**, 495-504, doi:10.1038/ni.3409 (2016).
169 Han, T. *et al.* The antitumor toxin CD437 is a direct inhibitor of DNA polymerase alpha. *Nat Chem Biol* **12**, 511-515, doi:10.1038/nchembio.2082 (2016).
170 Pavlov, Y. I., Shcherbakova, P. V. & Rogozin, I. B. Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *Int Rev Cytol* **255**, 41-132, doi:10.1016/S0074-7696(06)55002-8 (2006).
171 Suwa, Y. *et al.* Crystal Structure of the Human Pol alpha B Subunit in Complex with the C-terminal Domain of the Catalytic Subunit. *J Biol Chem* **290**, 14328-14337, doi:10.1074/jbc.M115.649954 (2015).
172 Klinge, S., Nunez-Ramirez, R., Llorca, O. & Pellegrini, L. 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. *EMBO J* **28**, 1978-1987, doi:10.1038/emboj.2009.150 (2009).
173 Simon, A. C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature* **510**, 293-297, doi:10.1038/nature13234 (2014).
174 Baranovskiy, A. G. *et al.* Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Res* **42**, 14013-14021, doi:10.1093/nar/gku1209 (2014).
175 Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921, doi:10.1107/s0907444998003254 (1998).
176 Johnson, K. A., Simpson, Z. B. & Blom, T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal Biochem* **387**, 20-29, doi:10.1016/j.ab.2008.12.024 (2009).

177 Johnson, K. A., Simpson, Z. B. & Blom, T. FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal Biochem* **387**, 30-41, doi:10.1016/j.ab.2008.12.025 (2009).
178 Perera, R. L. *et al.* Mechanism for priming DNA synthesis by yeast DNA polymerase alpha. *Elife* **2**, e00482, doi:10.7554/eLife.00482 (2013).
179 Dieckman, L. M., Johnson, R. E., Prakash, S. & Washington, M. T. Pre-steady state kinetic studies of the fidelity of nucleotide incorporation by yeast DNA polymerase delta. *Biochemistry* **49**, 7344-7350, doi:10.1021/bi100556m (2010).
180 Zahurancik, W. J., Klein, S. J. & Suo, Z. Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase epsilon. *Biochemistry* **52**, 7041-7049, doi:10.1021/bi400803v (2013).
181 Kunkel, T. A., Roberts, J. D. & Sugino, A. The fidelity of DNA synthesis by the catalytic subunit of yeast DNA polymerase alpha alone and with accessory proteins. *Mutat Res* **250**, 175-182, doi:10.1016/0027-5107(91)90174-m (1991).
182 Copeland, W. C., Lam, N. K. & Wang, T. S. Fidelity studies of the human DNA polymerase alpha. The most conserved region among alpha-like DNA polymerases is responsible for metal-induced infidelity in DNA synthesis. *J Biol Chem* **268**, 11041-11049 (1993).
183 Gordon, C. J. *et al.* Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* **295**, 6785-6797, doi:10.1074/jbc.RA120.013679 (2020).
184 Beigel, J. H., Tomashek, K. M. & Dodd, L. E. Remdesivir for the Treatment of Covid-19 - Preliminary Report. Reply. *N Engl J Med* **383**, 994, doi:10.1056/NEJMc2022236 (2020).