Structure and function of the Plasmodium falciparum apicoplast DNA polymerase: the first look at an "atypical" A-family polymerase and its potential in antimalarial drug discovery

Morgan Eilise Milton

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Biochemistry Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/15049

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Structure and function of the *Plasmodium falciparum* apicoplast DNA polymerase: The first look at an “atypical” A-family polymerase and its potential in antimalarial drug discovery

by

**Morgan Eilise Milton**

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

Major: Biochemistry

Program of Study Committee:
Scott Nelson, Major Professor
   Amy Andreotti
   Mark Hargrove
   Richard Honzatko
   Sanjeevi Sivasankar

Iowa State University

Ames, Iowa

2016

Copyright © Morgan Eilise Milton, 2016. All rights reserved.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>Dissertation Organization</td>
<td>20</td>
</tr>
<tr>
<td>References</td>
<td>22</td>
</tr>
<tr>
<td>CHAPTER 2 CRYSTALLIZATION AND PRELIMINARY X-RAY ANALYSIS OF THE PLASMODIUM FALCIPARUM APICOPLAST DNA POLYMERASE</td>
<td>28</td>
</tr>
<tr>
<td>Abstract</td>
<td>28</td>
</tr>
<tr>
<td>Introduction</td>
<td>28</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>30</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>37</td>
</tr>
<tr>
<td>CHAPTER 3 CRYSTAL STRUCTURE OF THE APICOPLAST DNA POLYMERASE FROM PLASMODIUM FALCIPARUM: THE FIRST LOOK AT AN “ATYPICAL” A-FAMILY DNA POLYMERASE</td>
<td>40</td>
</tr>
<tr>
<td>Abstract</td>
<td>40</td>
</tr>
<tr>
<td>Introduction</td>
<td>40</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>43</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>49</td>
</tr>
<tr>
<td>References</td>
<td>68</td>
</tr>
<tr>
<td>CHAPTER 4 A HIGH-THROUGHPUT ASSAY TO IDENTIFY INHIBITORS OF THE APICPLAST DNA POLYMERASE FROM PLASMODIUM FALCIPARUM</td>
<td>74</td>
</tr>
<tr>
<td>Abstract</td>
<td>74</td>
</tr>
<tr>
<td>Introduction</td>
<td>74</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>77</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>82</td>
</tr>
<tr>
<td>References</td>
<td>87</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

First and foremost, I would like to thank my major professor, Dr. Scott Nelson, for all of his support and guidance. He has taught me to persevere and work hard to achieve my goals, but to always save a little room for fun. He allowed me to take on many responsibilities and helped me develop a wide range of skill sets. Scott has provided me with numerous opportunities to grow as a scientist and a person, and for that I am ever thankful.

I would like to thank my thesis committee: Dr. Amy Andreotti, who recruited me to Iowa State and showed me what an amazing department BBMB is; Dr. Mark Hargrove, who has been my family away from home; Dr. Richard Honzatko; and Dr. Sanjeevi Sivasankar. Their support, suggestions, and guidance during my graduate career has helped me overcome many obstacles. My many hours talking with Dr. Honzatko has provided me with countless new perspectives on life. Dr. Honzatko has taught me what it means to be a crystallographer, and without his dedication to teaching, my work would not have been able to reach its full potential. Thank you for being such an irreplaceable mentor.

Dr. Jun-Yong Choe has been an invaluable asset to my work and personal growth. His generously offered me numerous hours of his synchrotron time during which I learned immensely from him. His guidance and suggestions for crystallization, data collection, and data processing have taught me so much about crystallography. Dr. Choe taught me that many things are difficult to accomplish, but “its ok, that is life” and to just keep pushing forward, to try a new approach. Thank you so much for all of the hours of your time you have spent with me at the beamline screening disordered crystal after disordered crystal. Watching you work has always inspired me.
Thank you to the staff at GM/CA-CAT at Argonne National Laboratories for their all of their assistance. Without the excellent service and advanced technology they provide at Sector 23, I would not have been able to solve my structure.

I would also like to thank Dr. Gaya Amarasinghe and Dr. Daisy Leung for starting off my career at Iowa State and teaching me the basics of biochemistry when I had never so much as seen an affinity chromatography column before. Thank you for generously allowing me to accompany you to the synchrotron and all you have taught me along the way.

Thank you to all of my friends and colleagues, you have made working at Iowa State a very enjoyable experience. Thank you for sharing in my joys and sorrows and allowing me to share in yours. You all make the BBMB community such a wonderful place to be a graduate student.

Lastly, I am so grateful for my family who has continuously supported me. You taught me to always work hard, to love what I do, and enjoy the beauty in life. Most importantly, I would like to thank my husband, AJ Milton. Without you everything I have achieved would have been significantly more difficult. Thank you for proof reading all of my papers (including this one), for coming into lab with me late at night to do experiments, for assisting me on numerous synchrotron trips, but most importantly for following me to Iowa and accompanying me on this crazy journey. I promise I will never make you live in Iowa again, but it was definitely worth it.
The leading cause of malaria infections in humans is the parasite *Plasmodium falciparum*. The parasite contains a non-photosynthetic plastid-like organelle called the apicoplast, which is essential for its survival within the host. The apicoplast maintains its own genome, which must undergo replication and repair. The only DNA polymerase in the apicoplast (apPOL) is classified as an atypical A-family polymerase. apPOL shares no direct orthology to mammalian polymerases, making the *P. falciparum* apicoplast DNA polymerase an attractive anti-malarial drug target. We solved the crystal structure of *P. falciparum* apPOL, the first structural representative of the atypical A-family polymerases. We showed that apPOL diverges from typical members in two of three previously identified signature motifs and a region not implicated by sequence. Moreover, apPOL has an additional N-terminal subdomain that extends the exonuclease domain. This region may be involved in binding DNA, participating in protein-protein interactions, and/or stabilizing the proofreading domain. These structural variances may account for functional differences in polymerase activity. The crystal structure also provides a point of departure for structure-based anti-malarial drug design, and these atypical regions could be targeted for potent and specific inhibitors. We have developed a fluorescence-based high-throughput DNA polymerase assay to screen for compounds that inhibit apPOL activity. Analysis of validation experiments indicates that the assay is statistically robust. A pilot screen of a 2,880 compound library identified 62 possible inhibitors that cause at least 50% inhibition of polymerase activity. The simplicity and robustness of the assay provides a solid platform for screening apicoplast polymerase inhibitors that could serve as lead compounds in the efforts to discover and produce anti-malarial drugs. In addition to its potential as a potent drug
target, apPOL provides structural and biochemical insight into a poorly characterized subgroup of A-family polymerases.
CHAPTER 1
INTRODUCTION

Literature Review
Morgan E Milton and Scott W Nelson

Modified from a review paper to be submitted for publication in *Molecular and Biochemical Parasitology*.

**Malaria**

Malaria is a devastating disease that affects nearly 200 million people each year (1). In 2014, approximately 440 thousand children died due to this mosquito-borne disease before reaching their fifth birthday (1). The spread of drug-resistant parasites in central Africa, India, southeast Asia, and northern South America is a growing problem with common anti-malarial drugs such as chloroquine and sulfadoxin-pyrimethamine becoming ineffective (2). Currently no drug is available to which some level of resistance has not developed. Even artemisinin, discovered by Tu Youyou, who received a Nobel Prize in Medicine in 2015 for her findings, has become less effective across southeast Asia and it is only a matter of time before that strain makes its way to Africa. The development of new and potent anti-malarial treatments will be critical in our ability to continue to combat malaria in the future (1).

Due to the critical need for new anti-malarial therapeutics, many avenues are being explored for their potential as anti-malarial drug targets. The leading cause of malaria in humans is *Plasmodium falciparum*, a member of the phylum Apicomplexa. Apicomplexa houses some of the most common and deadly protistan parasites and many members contain an essential organelle called the apicoplast, which is exclusive to the phylum (3). Due to its exclusivity to
Apicomplexa and its strict requirement for parasite survival, drugs targeted towards apicoplast functions should be potent and specific.

**Origins of the Apicoplast**

The apicoplast has had a convoluted history. The presence of an unidentified organelle and an extrachromosomal DNA in malarial parasites bewildered parasitologists for over two decades. A 35 kb circular DNA was originally observed in 1974 and was thought to originate from the mitochondrial genome of the parasite (4). Around the same time, an intracellular organelle of unknown function was detected in electron micrographs of different apicomplexans (5). Evidence began to build in favor of a plastidic origin of both this “spherical body” and the 35 kb DNA (6). Nearly 15 years after its misidentification, the true mitochondrial genome of *P. falciparum* was identified (7) and subsequent work showed that the unidentified 35 kb genome belonged to what is now known as the apicoplast (8). Full sequencing of the apicoplast genome established that its organization and gene content are reminiscent of an algae plastid (9).

To account for the presence of a plastidic-like organelle which is surrounded by an average of four membranes (as appose to the two membranes found in typical plastids), a precursor to apicomplexans must have acquired the plastid through a secondary endosymbiosis event. The identity of the players involved in the creation of the apicoplast were debated for some time, but more recent phylogenetic analysis has provided definitive evidence that the apicoplast originated from a photosynthetic red algae (10–12) that was engulfed and retained by an ancestral dinoflagellate through a secondary endosymbiosis event (13). The apicoplast has since lost its photosynthetic abilities, but is still maintained within the cell presumably due to its role in isoprenoid biosynthesis (14).
**Apicoplast Genome Structure**

The apicoplast genome has low complexity and encodes genes involved in its own expression. The apicoplast has one of the most A/T-rich genomes known to date with 86.9% A/T (12). It contains 68 genes coding for the large and small subunit rRNAs, a minimal but complete set of tRNAs, ribosomal proteins, a bacterial-like RNA polymerase, and several protein chaperones (15). A series of inverted repeats contain duplicate genes for rRNAs and tRNAs (12). It is believed that the genome has been under high selective pressure, consequently reducing the genome from the 150 kb of a typical plastid to 35 kb (15). This has resulted in many deletions and rearrangements to conserve the minimal autonomy of the apicoplast (15). This review will focus on how the apicoplast genome is replicated and maintained. We will look at each of the known proteins involved in replication as well as the proteins that are possibly involved in repair. There are still many unknowns in regards to replication and repair of the apicoplast genome. Sufficient groundwork has been laid to enable us to begin exploiting this system for new antimalarial drugs and we will discuss the potential of these proteins as drug targets.

**Method of Replication**

Understanding the means by which the *P. falciparum* apicoplast genome is replicated could lead to the development of more specific antimalarial compounds with novel mechanisms of action. Replication initiates within a large inverted repeat region through twin single-stranded displacement loops (D-loops) (16). This method of replication has been shown to be sensitive to ciprofloxacin (CIP), a topoisomerase inhibitor (17). At other locations on the genome, replication takes place through a rolling-circle mechanism, which is less sensitive to CIP (16). Replication of the apicoplast genome initiates in the late trophozoite stage of the parasite life cycle within the
red blood cell. Replication begins slightly before that of the nuclear genome and no more than 3 copies of plastid DNA are observed per cell (16).

The relationship between D-loop and rolling circle is unclear. When parasites are subjected to non-lethal doses of CIP, the amount of 35 kb linear DNA increases but the ratio of replication bubbles to replication forks decreases. This suggests the two mechanisms are truly theta and sigma varieties and are independent form each other as they are in the plastids of other species (16).

**Proteins Involved in DNA Replication**

The genes encoding for the apicoplast DNA proteins reside within the nuclear genome. Genome replication requires a helicase to unwind the double-stranded DNA, a primase to synthesize RNA primers for Okazaki fragment initiation, and a polymerase to replicate the leading and lagging stand templates. Single-stranded binding (SSB) proteins are also essential and aid in replication by protecting exposed single stranded DNA, melting DNA secondary structure, and interacting with proteins in the replisome. Topoisomerases are required to relieve positive supercoils that are generated ahead of the replication fork and are used to separate interlocked strands of replicated circular DNA. To date, all of these proteins have been identified and to some extend biochemically characterized. This review will examine at what is known about each of these proteins and their involvement in the replication and maintenance of the *P. falciparum* apicoplast genome.
Figure 1. *P. falciparum* Prex. Schematic of Prex depicting key motifs as determined by BLAST. A 20 residues signaling sequence (yellow) is followed by an apicoplast targeting sequence of unknown length. The fading of colors between subdomains represents the unknown cleavage sites between the primase, helicase, and polymerase regions. Black bars below the schematic represent the different constructs that have been studied for each region.

**Prex (transcription, transport, and processing)**

The *P. falciparum* gene PF14_0112 is 6051 bp long, contains no introns, and resides on the 14th chromosome of the nuclear genome. It was originally annotated as *pom1*, for “polymerase of malaria,” when a portion of the open reading frame was identified in *P. chabaudi* during a screen for antigenic variation in the murine parasite (18). The gene has since been renamed Prex for **P**lastidic DNA **R**eplication/repair **E**nzyme Complex (19). Prex is a single open reading frame encoding DNA primase, helicase, and polymerase (Fig. 1). A 20 amino acid long targeting sequence ensures that the gene is co-translated as a single polyprotein into the ER lumen, where it is then transported as a polyprotein into the apicoplast (20–22). Prex is thought to contain the only DNA polymerase targeted to the apicoplast, and thus is responsible for both DNA replication and repair (23).

The parasite maintains low levels of Prex expression with the maximum amount occurring during the late trophozoite stage, with a subsequent decline in the schizonts stages (19). Expression of Prex is correlated with an increase in apicoplast DNA (16), consistent with
its proposed role in the replication of the apicoplast genome (Fig. 2). Lindner et al. were unable to generate a Prex knockout in *P. yoelli* (a mouse model for malaria), strongly suggesting Prex is an essential protein in the blood stage of parasite infection (24).

Several groups have shown that the targeting sequence of Prex is specific to the apicoplast. Fusing GFP to the Prex targeting sequence causes it to be transported into the apicoplast without any transport into other subcellular compartments (19). Subcellular localization of Prex was further confirmed when the primase domain was fused to GFP and observed throughout the parasite’s life cycle in the red blood cell (24). Using immunofluorescence, antibodies to the Prex helicase and apicoplast SSB co-localize to the apicoplast in both early and late stages of the parasite’s growth within a red blood cell (25).

Upon entering the apicoplast, Prex is post-translationally modified by unknown protease(s) (20). On the basis of sequence analysis, there are 21 proteases containing predicted apicoplast targeting sequences, but their targets and recognition sequences are unknown (26). Through Western blot analysis, Seow et al. provided evidence that the polymerase domain is separated from the primase and helicase domains (19). Subsequently, Lindner et al. demonstrated that the primase domain may sometimes be cleaved from the helicase domain, with cleavage products of approximately 55 kDa, 80 kDa, and 160 kDa (24). Antibodies towards the helicase domain...
domain resulted in several bands of approximately 70 kDa and 55 kDa (25). Our understanding of the Prex post-translational modification is still greatly lacking and is possible that there may be several forms of the mature Prex proteins, as multiple isoforms of replicative proteins have been observed in the T7 bacteriophage primase/helicase protein (27). It is likely that proteolysis is used for regulation of Prex. This regulation would be necessary to obtain the optimal stoichiometry of each component, as helicases generally form hexamers, while polymerases act as a single subunit (28). Identification of the protease(s) responsible for separating the Prex primase, helicase, and polymerase domains would lead to a better understanding of post-translational modifications of apicoplast proteins, and identify potential targets for drug designed.

**Primase**

The Prex primase domain clusters phylogenetically with T7 phage-like primases (19). T7 bacteriophages contain a bi-functional primase-helicase protein which forms a hexamer with a distinct primase and helicase functional domain (29, 30). It is not clear if the Prex primase is active when covalently bound helicase like its T7 ancestry, or if primase and helicase domains are proteolytically separated. Bacterial primases generally have a three domain organization: zinc-binding domain (ZBD), topoisomerase-primase domain (TOPRIM), and a helicase-binding domain (HBD). Bacteriophages can have a slight variation in this arrangement, since the helicase is expressed with the primase, and thus would have no need for an HBD. Annotation of the Prex primase region reveals a ZBD and TOPRIM but no HBD is detected (24). It is possible that the Prex primase and helicase are maintained as one protein product after post-translational modification or the primase interacts with the helicase in a novel fashion.
Seow et al. expressed the primase and helicase domains of Prex together as one polypeptide (Fig. 1) and observed Mg$^{2+}$ dependent primase and helicase activity (19). Lindner et al. expressed a protein containing just the primase ZBD and TOPRIM (residues 115 to 465, Fig. 1) (24). It was found that the PfPrex primase binds zinc in a 1:1 molar ratio and catalyzed DNA dependent RNA synthesis, creating > 20 nt long RNA primers in the absence of other replisomal components (24).

Although it is homologous to the T7 helicase-primase, the absence of a HBD, and functional studies with the fused protein suggest that the native primase and helicase domains of Prex remain as a single unit in the apicoplast, there is also evidence suggesting this may not always be the case. Western blot analysis of parasite extract using antibodies to the TOPRIM domain revealed a native primase that runs slightly higher than the Lindner recombinant construct, suggesting that the primase is cleaved from the helicase. There were other predominates band at ~80 kDa and ~160 kDa which could be uncleaved primase-helicase domains and larger fragment Prex, respectively (24).

Due to its essential nature within the replisome, and structural features that set it apart from other primases, the Prex primase could serve as a lead target in the efforts towards identifying new anti-malarial compounds. Successful high-throughput assays have been developed to screen for primase inhibitors in other organisms that could be easily adapted for screening the Prex primase (31).

**Helicase**

Like the Prex primase, the helicase clusters with T7 phage-like helicase (19). All helicases bind and hydrolyze nucleotides. The Prex helicase has the Walker A and B motifs needed for ATPase activity and, due to its homology to the T7 helicase, likely belongs to
helicase superfamily 4. Its homology to T7 and the mitochondrial helicase Twinkle, suggests that the Prex helicase will form a hexamer ring with single stranded DNA in threaded through the center and will translocate along the single-stranded DNA in the 5’ to 3’ direction (28).

The T7 phage-like lineage supports a conjoined primase-helicase domain (32). As mentioned above, Western blot analysis has yet to result in a definitive answer regarding the post-translational modification of the primase and helicase domains or the native protein constructs. Regardless, recombinant expression of separate and functional primase and helicase domains has been successful (24, 25).

Bhowmick et al. expressed the Prex helicase (residues 669-1000, Fig. 1) fused to a maltose-binding protein (MBP) tag (25). Even with the MBP-fusion, the helicase was able to unwind DNA and the activity was enhanced upon addition of SSB. Using several truncation mutants (discussed more below in the SSB section), it was concluded that the C-terminus of the \textit{P. falciparum} SSB (residues 237-256) stimulates Prex helicase unwinding activity (25). The potential protein-protein interaction was further characterized in pulldown assays where the MBP-tagged Prex helicase pulled-down recombinant and endogenous SSB. To identify the domain that was primarily responsible for the interaction, SSB was separated into two domains and it was found that an N-terminal construct of SSB (residues 77-200), but not a C-terminal construct (residues 200-284) interacted with the helicase (25). It was concluded that the N-terminus of SSB is likely involved in diverse interactions with the apicoplast helicase and that helicase activity is stimulated by the SSB C-terminus (25). It is not known how removal of the N-terminal domain of SSB impacts helicase unwinding activity. The Prex helicase was not stimulated by non-\textit{P. falciparum} SSBs, although only \textit{Helicobacter pylori} SSB, which shares 25% sequence identity, was tested (25). It should be noted that it is not clear what impact the
presence of the MBP has on helicase unwinding activity or protein-protein interactions. While Bhowmick et al. carefully showed that the MBP domain itself was not directly responsible for their observations, maintaining a large tag (larger than the helicase itself) could affect helicase functionality.

It is common to observe helicases interacting with their SSB counterparts (33, 34). PriA, a protein responsible for restarting a stalled replication fork, has been shown to be loaded onto the DNA by SSB (35, 36). It is conceivable that the interaction between the Prex helicase and SSB could be reminiscent of the PriA-SSB interaction, resulting in the apicoplast SSB acting as a helicase loader for the Prex helicase. It is likely that the Prex helicase is a homo-hexameric ring that encloses around the single stranded DNA and therefore may require a helicase loading protein (29, 37). The T7 helicase and mitochondrial Twinkle helicase are loaded onto the DNA without a specific helicase loader, but it is likely that the adjoining primase domain aids in loading (37, 38). Thus far, no homolog to a helicase loader has been identified for the apicoplast. As seen in T7, it is possible that no helicase loader is present in the apicoplast and that SSB or an attached primase domain facilitates the loading for the Prex helicase onto the single-stranded DNA.

Western blots employing polyclonal antibodies raised to the helicase identified two prominent bands at 70 kDa and 55 kDa, which were attributed to different helicase isoforms similar to what is seen in T7 bacteriophage (25, 39). The 70 kDa band is present throughout the blood stage of parasite infection; the 55 kDa band appears during and after the trophozoite stage (25). Neither of these bands correspond in size to the Bhowmick et al. recombinant protein, which has a theoretical molecular weight of 38 kDa, suggesting that the starting and/or ending points chosen for the Prex helicase construct are not those found in the native protein. This
highlights the importance of identifying the Prex proteases and their cleavage sites, as the functional properties of the native proteins could differ from the truncated counterparts.

Compounds targeting the Prex helicase would be highly toxic to replication, but due to its similarities to the mitochondrial Twinkle helicase, drugs targeted to the apicoplast Prex helicase could inhibit mitochondria replication in humans. Drugs designed against the Prex helicase would need to be designed to be highly specific to avoid cross reactivity. Assays have been successfully developed for high-throughput screening of viral helicase inhibitors (40). These assays could possibly be used to identify compounds that inhibit the Prex helicase.

**Polymerase**

The apicoplast DNA polymerase is the best studied of the Prex domains. Genetic analysis classifies the polymerase as a prokaryotic A-family DNA polymerase, which are divided into five clades: prototypical bacterial PolAs, thermophilic viruses, Aquificaceae and Hydrogenothermaceae, Apicomplexa, and other viral-like bacterial polymerases that are found in bacteria that also contain typical PolAs (41). There is evidence that the atypical A-family polymerase originated through lateral gene transfer when a thermophilic virus infected a bacterium, transferring its polA gene to the host. It is likely that the transfer occurred prior to the second endosymbiosis event that gave rise to the apicoplast (41).

Typical PolAs contain a polymerase domain, 3’-5’ exonuclease domain for proofreading, and a 5’-3’ exonuclease domain from removing RNA primers from Okazaki fragments (42). Members of Apicomplexa have lost the 5’-3’ exonuclease domain, and instead are fused to the N-terminal primase and helicase domains.

DNA polymerases often contain accessory proteins that act as processivity factors. Currently there is no known processivity factor associated with the apicoplast DNA polymerase.
If a processivity factor does exist, it likely will not be similar to known processivity factors since no homologs to known processivity factors have been identified. However, based on the relatively small size of the apicoplast genome and the time required for its replication, it may be possible for apPOL to fully replicate the genome without the assistance of a processivity factor.

Several labs have biochemically characterized the apicoplast DNA polymerase. One complication has been determining the native start site of the polymerase domain as isolation of the native polymerase has been unsuccessful. Thus far, three different apicoplast DNA polymerase constructs have been studied. Seow et al. first expressed a protein construct encompassing residues 1107 to 2016 (Fig. 1), which was named PfPREXpol (19). Kennedy et al. employed construct expressing the polymerase domain beginning at residue 1426 (Fig. 1), which was referred to as KPom1, based on sequence similarity to the Klenow construct of *E. coli* Pol I (23). Most recently, the Nelson laboratory designed a construct (referred to as apPOL) based on the sequence alignment of members of the *Plasmodium* genus (43, 44) and determined that a likely polymerase boundary conserved across *Plasmodium* spanning residues 1389 to 2016 (Fig. 1). Below, we will summarize and compare the findings of each group’s apicoplast DNA polymerase construct.

Protein expression levels vary greatly between constructs. *PfPREXpol* was reported to yield 4 mg of purified recombinant protein per liter of culture (19), whereas, the yield of apPOL is 50 mg of protein per liter (45). The total yield of KPom1 was not reported and the protein was expressed with a maltose-binding-protein (MBP) fusion tag as opposed to the hexa-histidine tags used for *PfPREXpol* and apPOL (23). The MBP fusion tag may serve to increase the solubility of KPom1, as we have found that the expression and stability of the apicoplast polymerase is very sensitive to modifications at the N-terminal region. Upon solving the crystal structure of
apPOL (PDB 5DKU), it became clear why removal of the N-terminal region impacts stability. The polymerase N-terminal region forms a long β-hairpin that spans the bottom of the exonuclease domain. The β-hairpin and flanking loops are involved in numerous hydrophobic and hydrogen bonding interactions with the rest of the exonuclease domain. The start site of KPom1 disrupts the β-hairpin and excludes a majority of these interactions, which could result in a large unstructured loop at the N-terminus of the construct. It is possible that the differences in construct start sites are responsible for the differences in polymerase activity discussed below.

Biochemical assays examining the polymerase and exonuclease activities, as well as the optimal pH, Mg$^{2+}$ concentrations, and pH have been carried out. PfPREXpol shows maximal DNA polymerase activity at a Mg$^{2+}$ concentration of 4 mM, whereas Kpom1 requires 10 mM (reference). KPom1 has pH and temperature optimums of 9.0 and 40°C, respectively, whereas PfPREXpol optimum values are 7.0 and 75°C, respectively (19, 23). It is unclear if differences in construct alone can account for these discrepancies. The diparity in temperature tolerated by the polymerase is particularly interesting. The temperature where PfPREXpol shows the highest activity (75°C) is consistent with the thermophilic virus ancestry of the polymerase. On the other hand, the temperature optimum of 40°C for KPom1 is more physiologically relevant, as the replication of the apicoplast genome takes place in infected red blood cells at 37°C (16). Our findings also suggest that apPOL operates at an optimal temperature that is close to 37°C. We find that apPOL has a melting temperature of about 45°C and at temperatures higher than 45°C the activity significantly decreases and cannot be recovered upon returning to lower temperatures. Since the native apicoplast DNA polymerase boundaries are unknown, it is difficult to tell what structural features may be causing the differences in temperature tolerance observed between PfPREXpol and KPom1/apPOL.
Kennedy et al. determined the fidelity of KPom1 and found that it was surprisingly error-prone, but the exonuclease domain corrected a majority of its mistakes, which is consistent with a replicative polymerase (23). They concluded that without an active exonuclease domain, KPom1 has a unique error signature error with a mutation spectrum more closely resembling to lesion bypass polymerases than replicative polymerases. KPom1 was found to have a strong bias towards misincorporating a dGTP across from a dTMP (23), which is in conflict with the A/T-rich aspect of the apicoplast genome.

On the other hand, apPOL behaves more like Klenow and its fidelity is more reminiscent of other replicative DNA polymerases (43). Differences in the reported fidelity of KPom1 and apPOL may not be solely due the structural variances between the constructs as the methods of analysis differed between the two proteins (43). Mutation rates are not exclusively determined by polymerase fidelity, as the ability of a polymerase to extend a mismatch and removal of misincorporated nucleotides also play a role. Mismatch extension by apPOL is highly sequence dependent. The incoming nucleotide must compensate for alternative structures in the polymerase active site caused by upstream mismatches. By combining misincorporation, exonuclease, and mismatch extension rate data, the most likely mutations caused by apPOL were ranked. apPOL is most likely to mutate a G to an A if the downstream base is a T, followed by C replacing a T if a T, C, or G is downstream (43). Of course, these observations do not take into account potential DNA repair enzymes which could impact mutation probabilities in the apicoplast. Also, the high fidelity of apPOL is not reminiscent of a lesion bypass polymerase, in fact it is likely that apPOL stalls at common DNA lesions (43). This suggests that DNA repair pathways are in place in the apicoplast.
Using a complementation assay with heat sensitive Pol I, Kennedy et al. has shown that KPom1 can substitute for the *E. coli* Pol I *in vivo*. Over expressed KPom1 restored wild-type growth by 60%. Induction with IPTG appears to be required, as the no IPTG growth condition only restored wild-type growth by 29% (23). Growths were normalized to the growth of Pol I containing vectors in the absence of IPTG, suggesting that KPom1 cannot directly replace Pol I even though they share nearly identical active site residues. This is potentially due to the extremely low $k_{cat}$ of KPom1 but other structural factors could be responsible.

Polymerases appear to be well-suited for inhibition (46). Drugs developed towards the Prex polymerase would likely be potent and specific (47). Seow et al. demonstrated that their recombinantly expressed polymerase (*Pf*PREXpol) was sensitive to two drugs known to inhibit DNA polymerase activity, chloroquine and suramin, but was unaffected by aphidicolin, an inhibitor of the eukaryotic DNA polymerase α (19). A promising high-throughput assay has been developed for screening for inhibitors of apPOL (45). The assay is able to detect compounds that inhibit polymerase activity through four main mechanisms: DNA mimics/inhibitors that compete for the DNA binding site, nucleotide mimics that compete for nucleotide binding site, DNA intercalators, and small molecules that inhibit by binding somewhere else on the polymerase. Nonspecific inhibitors can easily be counter screened using other DNA polymerases such as Pol I.

Although there is some contradictory data surrounding the details, the results of these three groups are in strong favor of the Prex polymerase domain being the polymerase responsible for replication of the apicoplast genome. Determination of the native construct would aid in understanding the discrepancies in polymerase activities between *Pf*PREXpol, KPom1, and
apPOL. Until then, the structure of apPOL will act as a starting point for further understanding of this atypical A-family polymerase and aid in the development of new antimalarial compounds.

**SSB**

Single stranded binding proteins (SSB) are responsible for protecting and stabilizing ssDNA during replication and repair and melting secondary structures in the DNA. SSBs from other organisms have been shown to interact with numerous proteins involved with replication and repair (34). The gene encoding the apicoplast SSB resides on chromosome five of the nuclear genome and is of bacterial origin (48). Although *E. coli* SSB shares only a 39% sequence identity and 66% homology with *P. falciparum* SSB, the two proteins share a high degree of structural similarity. The N-terminus is composed of an oligonucleotide-binding domain (OBD) that facilitates the binding of ssDNA using three tryptophan residues that base stack with the DNA (49). *P. falciparum* SSB maintains the positioning of these residues (48, 50). The OBD also contains a conserved histidine that facilitates oligomerization (48, 51). Prokaryotic SSBs have an acidic C-terminal tail that is involved in mediating protein-protein interactions in DNA replication, repair, and recombination (34). *P. falciparum* SSB lacks this acid sequence, suggesting that its binding partners are different and likely specific to apicoplast function (48, 52). In addition to having a distinct tail sequence, *P. falciparum* SSB has a 28 residue extension on its C-terminus that has no sequence homology to its *E. coli* counterpart (48). Removal of these 28 residues results in an increase in helicase unwinding activity in the presence of SSB, suggesting that the extreme C-terminus of SSB could be involved in helicase regulatory functions (25). *P. falciparum* SSB also contains an N-terminal apicoplast targeting sequence believed to be 76 amino acids long (48).
Prusty et al. demonstrated that SSB localizes to the apicoplast and that it is expressed during all three stages of the parasite’s life cycle in the red blood cell (Fig. 2) (48). Native and recombinantly expressed protein form a stable tetramer in solution and in the crystal (48, 49, 52). Gel shift assays along with a co-crystal with DNA, confirm that SSB binds ssDNA (48, 49). Similar assays with dsDNA show that SSB specifically interacts with ssDNA and that the cause of the gel shift is due to SSB binding (48).

Despite being nearly structurally identical to the *E. coli* SSB (49), *P. falciparum* SSB is unable to complement *E. coli* SSB *in vivo* and removal of the 28 residue C-terminal extension does not induce complementation (48). The inability to complement is likely due to the different ssDNA binding properties observed in the *P. falciparum* SSB crystal structure. It is also probably that does not bind to the same proteins. While *P. falciparum* SSB wraps ssDNA with the same topology as *E. coli*, the DNA backbone polarity is reversed (49). There are other notable differences between the two proteins: protein-DNA contacts, symmetry of inter-subunit contact sites, and protein-protein interactions between adjacent tetramers. *P. falciparum* SSB only binds ssDNA in a tight binding, fully wrapped mode, while *E. coli* SSB is able to bind in full and partially wrapped modes (52). Even in a fully wrapped mode, *P. falciparum* SSB is able to fully diffuse along the ssDNA and melt DNA secondary structures (52). A different interface is observed for tetramer-tetramer interactions for *P. falciparum* SSB than those seen in *E. coli*. In *E. coli* a loop from neighboring tetramers pack against each other, but this interface is not seen in *P. falciparum*. All of these differences combined could result in an SSB that is functionally different from that of *E. coli*. The findings of Prusty et al. and Antony et al. suggest that the *P. falciparum* SSB may interact with other proteins in a unique manner (48, 49, 52).
In the crystal structure of PfSSB, only the residues in OBD (residues 77-194) were observed, the C-terminal tail and additional 28 residues had no visible electron density. Analysis of the crystals showed that the C-terminal region was at least partially cleaved (49). Unfortunately, the crystal structure does not shed light on the helicase binding interaction. Antony et al. showed that removal of the extreme C-terminal 28 residues from P. falciparum SSB to create a more E. coli-like construct, shifts the binding equilibrium to favor DNA binding in a partially wrapped mode (52). An alternative to Bhowmick et al.’s hypothesis that the C-terminal stimulates helicase activity is that the difference in DNA binding modes of the two SSB constructs could impact helicase unwinding activity. Clearly, additional work is necessary to determine the precise role of the PfSSB C-terminal region. It appears that this region is responsible for novel protein-protein interactions not seen in any other SSB. It is also likely that this region is involved in regulation of the ssDNA binding mode and regulation of replication partners.

**5ʹ-3ʹ exonuclease**

5ʹ to 3ʹ exonucleases are responsible for removing the RNA primers synthesized by primase so that the space can be filled in with DNA and ligated. In typical A-family polymerases, an N-terminal 5ʹ to 3ʹ exonuclease domain preforms this activity, but the Prex polymerase lacks such a domain and instead has an N-terminal primase and helicase domain (41). The 5ʹ to 3ʹ exonuclease domain can function independently of the 3ʹ to 5ʹ exonuclease domain and polymerase domain (53). In fact, thermophilic virus PolAs frequently have a separate gene which encodes a 5ʹ to 3ʹ exonuclease (41). We have identified the putative apicoplast targeted 5ʹ to 3ʹ exonuclease on chromosome 7 of the nuclear genome of P. falciparum (gene ID Pf3D7_0203900). BLASTP analysis reveals a DNA Pol I 5ʹ to 3ʹ
exonuclease domain with metal binding sites. We propose that this is the 5’ to 3’ exonuclease involved in DNA replication within the apicoplast and will act in conjunction with the Prex polymerase.

**Potential for Anti-malarial Drug Target**

Since its discovery as a unique and essential organelle, the apicoplast has been investigated for its potential as a target for anti-malarial compounds (17). It was observed that antibiotics that affect protein or nucleic acid synthesis have anti-malarial activity. Interestingly, these compounds induced a “delayed-death” phenotype (17, 47). Delayed-death refers to the phenomenon that upon administration of the drug, the first generation of parasites remain viable, continuing to divide and infect new red blood cells but the second generation is unable to form functional merozoites (54). Parasites treated with doxycycline, which inhibits the 70 S ribosome, cannot replicate their apicoplast genomes and were thus nonfunctional. Interestingly, the mitochondria is left unaffected (54). This suggests that drugs are able to penetrate the four membranes encompassing the apicoplast and impact proteins involved in apicoplast replication and transcription.

**Current drugs that impact replicative proteins**

Many antibiotics inhibit proteins involved in transcription and translation of the apicoplast genome: tetracycline, clindamycin, azithromycin, and rifampin (17, 55). Ciprofloxacin and coumermycin inhibit the DNA gyrase of the apicoplast (56, 57) and the Prex polymerase is inhibited by chloroquine and suramin (19). Unfortunately, these drugs are not specific to the apicoplast and frequently have high IC_{50}s. More potent and specific compounds need to be identified.
Medicine for Malaria Venture assembled a library of 400 compounds identified through phenotypic screening that inhibit malaria. This library, known as the Malaria Box, is refined from 20,000 compounds identified from four million compounds screened by St. Jude’s Children’s Research Hospital, Novartis, and GlaxoSmithKline (58). All compounds inhibit malaria in the blood stage and are nontoxic to humans. Bowman et al. identified roughly 40 small molecules from the Malaria Box that have an apicoplast-targeting phenotype (59). We have found that one of these 40 compounds likely inhibits the Prex polymerase. The small molecules within the Malaria Box will provide a solid starting point for developing new antimalarial therapies, but a phenotypical screen sometimes misses the potential of some potent targets.

The replication proteins mentioned in this review could be potent and highly specific drug targets. There may be the concern of cross-reactivity between DNA ligase, SSB, and Prex helicase and human cellular functions but it is likely that apicoplast-specific features could be exploited. Efforts to identify inhibitors of the Prex polymerase are underway (45). As we gain more biochemical and structural knowledge concerning the proteins involved with replication and maintenance of the apicoplast genome, we begin to elucidate key features that can be employed in our efforts in anti-malarial drug discovery.

Dissertation Organization

This work focuses on the structure and biochemical mechanisms of the *Plasmodium falciparum* apicoplast DNA polymerase (apPOL), centering on its “atypical” nature and potential use as an antimalarial drug target. Since apPOL has no direct orthologs to mammalian polymerases, the *P. falciparum* apPOL is an attractive antimalarial drug target that provides
structural and biochemical insight into a poorly characterized subgroup of the A-family polymerases.

Chapter two concentrates on the initial crystallization process, data collection, and solving of the apPOL x-ray crystallography structure. Due to the small dimensions of the crystals, tendency towards disordered crystal packing, and lack of a close structural homolog for molecular replacement, data collection and processing posed a significant challenge. Advances in synchrotron beamline technology through the use of a mini-beam allowed for the collection of a full data set on a single apPOL microcrystal. Due to complications with traditional molecular replacement, MR-Rosetta was employed to solve the structure of apPOL. MR-Rosetta couples traditional x-ray crystallographic software with structure prediction algorithms, resulting in energy-optimized models that can improve electron density maps. The work published in this paper focuses on the technical aspects of data collection and processing and lays the groundwork for the structural characterization of apPOL.

The structure of the apicoplast DNA polymerase, the first of an atypical A-family DNA polymerase to be solved, is presented in chapter three. apPOL contains the canonical right handed finger, thumb, and palm subdomains seen in all polymerases. Although apPOL resembles a typical A-family polymerase, a close inspection reveals several distinct structural features. This paper compares structural and active site elements of apPOL with other structurally characterized A-family members. Several of the structural differences between apPOL and typical A-family members reside in signature motifs determined through sequence alignment. This work focuses on the composition of these signature motifs and other significant structurally differences we identified, as well as how these regions may impact polymerase activity and fidelity. The atypical elements described in this paper will assist in investigating the biochemistry of apPOL,
and will aid in the development of antimalarial drug compounds that are potent and specific to apPOL.

Chapter four details with the development of a fluorescence-based assay that can be used to screen for inhibitors of apPOL in a high-throughput manner. The assay is based on a molecular beacon and is able to indirectly detect polymerase activity in a continuous manner. The assay is robust and has been validated for the National Institutes of Health’s requirements for a high-throughput screening assay. A pilot screen demonstrated that the assay can be used to screen large libraries of small molecules and is able to identify apPOL inhibitors with a wide array of inhibitory mechanisms. The assay was further tested for its ability to be used to perform steady state kinetics in a low throughput manner. A known inhibitor of nucleic acid binding proteins was selected, and its mechanism for inhibiting apPOL was established. This work provides a solid platform for the high-throughput identification of apPOL inhibitors.

Altogether, this work establishes the means for identifying compounds which target apPOL. Determining the structure of apPOL allows for further characterization of inhibitors, and the means for understanding the atypical nature of the apicoplast DNA polymerase. This work will allow for the optimization of lead antimalarial compounds through the coupling for target based drug discovery and structure-guided drug design.

References


CHAPTER 2
CRYSTALLIZATION AND PRELIMINARY X-RAY ANALYSIS OF THE PLASMODIUM FALCIPARUM APICOPLAST DNA POLYMERASE

Morgan E Milton, Jun-Yong Choe, Richard B Honzatko, and Scott W Nelson


Abstract

Infection by the parasite Plasmodium falciparum is the leading cause of malaria in humans. The parasite has a unique and essential plastid-like organelle called the apicoplast. The apicoplast contains a genome that undergoes replication and repair through the action of a replicative polymerase (apPOL). apPOL has no direct orthologs to mammalian polymerases and is therefore an attractive anti-malarial drug target. No structural information exists for apPOL and the Klenow fragment of E. coli DNA polymerase I, which is its closest structural homolog, shares only 28% sequence identity. Here, we report conditions of crystallization and preliminary X-ray diffraction data from crystals of P. falciparum apPOL. Data complete to 3.5 Å resolution were collected from a single crystal (2 x 2 x 5 µm) using a 5 µm beam. The space group P6522 (unit cell \( a = b = 141.8 \, \text{Å}, \, c = 149.7 \, \text{Å}, \, \alpha = \beta = 90^\circ \) and \( \gamma =120^\circ \)) was confirmed by molecular replacement. Refinement is in progress.

Introduction

Almost half of the world’s population is at risk of malaria. In 2012, there were approximately 207,000,000 cases worldwide, resulting in 627,000 deaths (1). The cause of malaria is a group of parasites from the genus Plasmodium, with the most common being P. falciparum and P. vivax. Drug resistance has been a constant problem and the spread of drug-
resistant parasites in central Africa, India, southeast Asia, and northern South America has left common antimalarial drugs such as chloroquine, artemisinin, and sulfadoxin–pyrimethamine ineffective (2).

*Plasmodium* spp. are members of the phylum Apicomplexa, which is composed exclusively of unicellular protozoan parasites and contains several other important human and animal pathogens such as *Toxoplasma gondii* (toxoplasmosis), *Babesia bovis* (babesiosis), *Cyclospora cayetanensis* (cyclosporiasis), and *Eimeria falciformis* (coccidiosis). All of these organisms contain a unique organelle called the apicoplast that has been shown to be essential for the survival of the parasites within their host. The apicoplast arose through a secondary endosymbiosis event with red algae and can be viewed as an ancient chloroplast (3). While it has not maintained the photosynthetic abilities of its parent organelle, the apicoplast is responsible for the execution of several biochemical processes such as the synthesis of isoprenoids, fatty acids, and hemes, along with iron–sulfur cluster maturation (4).

Like the mitochondria and chloroplast, the apicoplast houses its own genome that must undergo replication and repair, but its genome lacks genes coding for enzymes involved in DNA replication (5). In 2004, the *prex* (plastidic DNA replication/repair enzyme complex) gene was identified. *Prex* encodes the putative apicoplast DNA primase, helicase and polymerase in the form of a polyprotein that has an N-terminal sequence targeting it to the apicoplast. Following import, Prex is cleaved by an unknown protease into three separate proteins (6). The *P. falciparum* apicoplast DNA polymerase (apPOL) is the only known DNA polymerase targeted to the apicoplast, and we have previously demonstrated that the activity and fidelity of apPOL is consistent with that of a replicative DNA polymerase (7).
Owing to the essential nature of the apicoplast in malaria parasites, apPOL is an especially attractive target for the development of antimalarial drugs. Bioinformatics studies indicate that apPOL is of prokaryotic origin and is a member of the A-type family of DNA polymerases (8). A-family DNA polymerases fall into five main groups: typical bacterial PolAs, thermophilic viruses, Aquificaceae/Hydrogenothermaceae, Apicomplexa, and other unrelated bacteria. While many members of the typical A-family DNA polymerases (for example, *Escherichia coli* PolII, *Taq* polymerase, and mitochondrial DNA polymerase γ) have been extensively studied at a biochemical and structural level, very little is known regarding the polymerases from the other four distinct A-family clades. *P. falciparum* apPOL shares only 28% sequence identity with the Klenow fragment of *E. coli* PolII (PDB 2KFN, 9), which is its closest structurally characterized homolog (10). On the other hand, the identity between different apPOLs is relatively high (84% identity between *P. falciparum* and *P. vivax*, the two primary agents of human malaria). This suggests that drugs designed to target apPOL from *P. falciparum* would be effective in treating malaria caused by other *Plasmodium* species. Here, we report the expression, purification, crystallization, and preliminary X-ray analysis of apPOL. The apicoplast DNA polymerase sequence is highly conserved among members of the *Plasmodium* genus and the structural information gathered through this study will aid in the discovery and development of antimalarial drugs.

**Materials and Methods**

**Protein preparation**

The pET-28b vector (Novagen) containing apPOL (*P. falciparum* Prex residues 1389-2016) was used to generate a 3’-to-5’ exonuclease-deficient mutant apPOL\(^{\text{exo}^-}\) (D1470N and E1470Q) as described previously (7, Table 1). The resulting vector was transformed into *E. coli*
BL21(DE3) cells (Novagen), which were grown in LB medium at 310 K to an optical density of 0.7 at 600 nm. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a concentration of 0.2 mM to induce translation.

The induced cultures were grown overnight at 291 K. The cells were harvested by centrifugation for 20 min at 3000 g and 277 K, suspended in a minimal volume of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole (buffer 1), and stored at 193 K. The polymerase was purified at 277 K as described by Miller et al. (11). The lysate from homogenized cells (EmulsiFlex-C5) was centrifuged for 1 h at 30,000 g and 277 K. The supernatant was loaded onto a 5 ml Ni-agarose column and washed with buffer 1 and then with 20 mM Tris-HCl pH 8.0, 1 M NaCl, 25 mM imidazole. 1 M NaCl removes nucleic acids from the immobilized polymerase. 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 150 mM imidazole was used to elute the polymerase. The eluted apPOLexo− was concentrated to 10 mg ml⁻¹ as determined spectrophotometrically using an extinction coefficient ε₂₈₀ of 56750 M⁻¹ cm⁻¹. The sample was then subjected to size-exclusion chromatography using a 320 ml HiLoad 26/200 Superdex 200 column (GE Healthcare Biosciences) equilibrated with 20 mM Tris-HCl pH 8.0,

<table>
<thead>
<tr>
<th>Source organism</th>
<th>P. falciparum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA source</td>
<td>GenScript</td>
</tr>
<tr>
<td>Forward primer†</td>
<td>5'-GGTGGTCATATGGAATCACCACAAATTACCTCAGAG-3'</td>
</tr>
<tr>
<td>Reverse primer‡</td>
<td>5'-GGTTGGGATATTGATCCCAAAAAATATCAGGC-3'</td>
</tr>
<tr>
<td>Exonuclease mutation</td>
<td>5'-GATATAAATTGTCGGCTGAAATATCAGAAACTGAG-3'</td>
</tr>
<tr>
<td>Expression vector</td>
<td>pET-28b</td>
</tr>
<tr>
<td>Expression host</td>
<td>E. coli BL21(DE3)</td>
</tr>
<tr>
<td>Complete amino-acid sequence of the construct produced</td>
<td>MGSSHSHHHHSSGSLFYVRGSHDEITKYYIKDNIINVDDNI IKKDKAFLKNENRTECAFAEYFEGKKFDDDIESRF F I INDNYYNEINLYIYDIYCIYGCLDRISTTGELEVDENIRL IQIAVENVPIYIDMFNIKDDILGLRKVLARKNIIKI IQNGKFADKFLHNNFKIENFDTYAISSLKLDKNKNMYG FKLNNIEYKLNVLILDQQNSVWNNSLNNQFLYAAAR DSCSCLLKLYKKLEEIKKENLHVIDREINCIILPICDME LNGKIVDENLQKSTNEILNENLEKUINLKLKGDENIN VSNSQQVQKLQRKRRVDSNKLIENTDSNKNFLAHE EIIISLNRPRKILDAIAFILKLLHLINTKNIHTTPFNQ LKTFSGRFSSEKPLQQIQTPQKNKEREIPFNNDNFIFIIDFKFQELKIAEITNEIMLAIKNIIIDHLHLSAITIKK KNIPEIKKEDIHAIKAINFGLIGMYVGNLDKNKYTYG LNMSSLDQCLFYFSFSEYKAIYKWNQYVQKRALGQYST LSRKVTIFPYFSFKALKYFVQGTCDDIKLALVYDYN LKDINGKIIILCVHDEIIIEEVKFKQEEALKILVQSMENS ASYFLKKVKEVSUIIAENWGSKD</td>
</tr>
</tbody>
</table>

† The NdeI site is underlined. ‡ The BamH1 site is underlined.
400 mM NaCl. Fractions containing apPOL\textsuperscript{exo−} were pooled and concentrated to approximately 15 mg ml\textsuperscript{-1} using 50,000 NMWL Amicon Ultra-15 Centrifugal Filter Devices (Merck Millipore). High salt (400 mM) was used throughout the purification and crystallization process to ensure the stability of apPOL, which precipitates after prolonged periods (approximately 2 d) at lower salt concentrations. Protein purity was evaluated at each step by SDS–PAGE using Coomassie Blue stain. Control reactions indicated that our preparations were free of nuclease and phosphatase activity which would possibly interfere with polymerization assays. Polymerase activity was confirmed using methods described previously (11). Aliquots of purified apPOL\textsuperscript{exo−} were flash-frozen in liquid nitrogen and stored at 193 K. The freezing process does not impact on the polymerase activity or crystallization. Cells for the preparation of SeMet- apPOL\textsuperscript{exo−} were grown in M9 minimal medium supplemented with 50 mg SeMet 15 min prior to induction with IPTG. Labeled polymerase was purified in an identical manner to unlabeled polymerase and resulted in a slightly depressed protein yield, with unaltered catalytic activity.

Protein for crystallization was evaluated by multi-angle light scattering (MALS). The polymerase sample was passed through a 0.22 mm cellulose acetate Spin-X centrifuge tube filter (Electron Microscopy Sciences) prior to data collection. MALS data were collected by passing the elutant from a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8.0, 400 mM NaCl into a DAWN HELEOS II light-scattering system and Optilab T-rEX refractive-index detector (Wyatt Technology). The data were analyzed with the ASTRA software package.
Crystallization and X-ray data collection

Initial attempts employed Crystal Screen, Crystal Screen 2 and Index (Hampton Research) and vapor diffusion in 96-well sitting-drop plates. Promising outcomes were replicated and optimized by hanging-drop vapor-diffusion experiments. Polymerase (4–18 mg ml\(^{-1}\)) in 20mM Tris-HCl pH 8.0, 400 mM NaCl was combined with the well solutions in a 1:1 ratio to give a final droplet size of 4 ml (Table 2). Variation of the ratio of solutions and the drop size did not appear to affect the quality and the size of the crystals produced. The trays were equilibrated at room temperature. Prior to data collection, crystals were looped and immediately flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K using a MAR300 CCD detector on Advanced Photon Source beamline 23-ID-B (GM/CA-CAT). At a crystal-to-detector distance of 350 mm, 50 frames were collected using an oscillation range of 1° and a 5 mm mini-beam. Data were processed using *HKL*-3000 (12). Additional analysis and molecular replacement was performed using *PHENIX* (13).

<table>
<thead>
<tr>
<th><strong>Table 2</strong></th>
<th><strong>Crystallization</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td>Vapor diffusion</td>
</tr>
<tr>
<td><strong>Plate type</strong></td>
<td>Hanging drop</td>
</tr>
<tr>
<td><strong>Temperature (K)</strong></td>
<td>298</td>
</tr>
<tr>
<td><strong>Protein concentration (mg ml(^{-1}))</strong></td>
<td>4–18</td>
</tr>
<tr>
<td><strong>Buffer composition of protein solution</strong></td>
<td>20 mM Tris pH 8.0, 400 mM NaCl</td>
</tr>
<tr>
<td><strong>Composition of reservoir solution</strong></td>
<td>0.2 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 30% (w/v) PEG monomethyl ether</td>
</tr>
<tr>
<td><strong>Volume and ratio of drop</strong></td>
<td>4 μl, 1:1 protein:well solution</td>
</tr>
<tr>
<td><strong>Volume of reservoir (μl)</strong></td>
<td>500</td>
</tr>
</tbody>
</table>

Results and Discussion

The yield of apPOL\(^{exo^{-}}\), which was 98% pure according to ImageQ (GE Healthcare) analysis of a Coomassie Blue stained SDS-PAGE, was approximately 50 mg per liter of LB (Figure 1a). MALS data revealed a single peak corresponding to a molecular mass of 68.9 kDa (Figure 1b), which is in agreement with the gene-calculated mass of 73.9 kDa. Microcrystals
grew within two weeks in droplets consisting of a 1:1 ratio of protein solution and well solution consisting of 0.2 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 30% PEG monomethyl ether 5000 (Crystal Screen 2 condition No. 26). Crystals produced under this condition resembled short-grain rice, with approximate dimensions of 2 x 2 x 5 µm (Figure 2). Although crystallization was highly reproducible using the premade solution from Hampton Research (regardless of the production lot), in-house solutions failed to consistently generate crystals and SeMet-labeled protein did not produce crystals under this, seeding or any other conditions.

A beam with a 20 µm cross-section provided diffraction from microcrystals to a resolution of only 8Å; however, a beam with a 5 µm cross-section decreased the background radiation without loss of signal, revealing intensities to a resolution of 2.8 Å in a single image resulting from an exposure time of 10 s (14). The crystals were radiation-sensitive, so the exposure time was limited.
to 2 s in order to collect complete data from a single crystal (at least 30 images). The crystal received an average radiation dose of 134.61 MGy as calculated by RADDOSE-3D (15). This well exceeds the maximum dose of 30 MGy recommended for macromolecular crystallography (16). Radiation damage was visually apparent over the 50 frames of data collected as the resolution decreased from just beyond 3.5 Å to approximately 4.5 Å. Also influencing the strategy for data collection was a frequency of one in 60 flash-cooled crystals that exhibited high-resolution ordered diffraction. The aforementioned circumstances resulted in complete data, but only to a resolution of 3.5 Å largely owing to the short exposure time of 2 s and radiation damage (Figure 3a, Table 3). Data reduction initially assumed a trigonal lattice, but further analysis indicated space group P6122 (or its enantiomorph), with unit-cell parameters \(a = b = 141.8, c = 149.7\), \(\alpha = \beta = 90, \gamma = 120^\circ\). The Matthews coefficient of 2.94 Å\(^3\) Da\(^{-1}\) is consistent with one polymerase molecule per asymmetric unit and a solvent content of 58.2%.

Data were originally scaled with a resolution cutoff at 3.5 Å (Table 3). The inclusion of all available diffraction data resulted in qualitative improvement of the electron-density map, but the \(I/\sigma(I)\) values for each shell suggest that the nominal resolution of the data is 3.5 Å.

Initial molecular replacement was performed using Phaser (17) and resulted in a translation-function Z-score (TFZ) of 4.4, a rotation-function Z-score (RFZ) of 3.1, and a log-

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection and processing.</td>
</tr>
<tr>
<td>Values in parentheses are for the outer shell.</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>Total No. of reflections</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Multiplicity</td>
</tr>
<tr>
<td>(\langle I/\sigma(I)\rangle)</td>
</tr>
<tr>
<td>R(_{merge})</td>
</tr>
</tbody>
</table>
likelihood gain (LLG) of 49.555 when an unmodified Klenow fragment (PDB 2KFN, 9) was used as the molecular-replacement model. Use of only the polymerase domain of the Klenow fragment (residues 519–928), which has 31% sequence identity to apPOL\textsuperscript{exo-} but alone only accounts for 63% of the total apPOL\textsuperscript{exo-} molecule, resulted in a molecular-replacement solution (electron density revealing right-handed \(\alpha\)-helices as shown in Figure 3b) only in space group P6\(_{5}\)22, with a TFZ of 6.2, an RFZ of 3.9, and an LLG of 63.248 (13). Subsequent use of MR-Rosetta (18) using the unmodified PDB 2KFN (complete with DNA and ions) as the model independently indicated the same space group and a partial model with approximately 450 of the 628

\textbf{Figure 3.} (a) X-ray diffraction pattern of an apPOL\textsuperscript{exo-} microcrystal produced from a 2 s exposure. The first image collected from a data set of 50 frames is shown. Weak diffraction extends to 2.8 Å resolution, with stronger diffraction extending to approximately 3.5 Å resolution as shown by the black circle. (b) Density map and model from early refinement after molecular replacement. The density shows clear right-handed helices and crystal lattice contacts. Coordinates are shown in yellow and symmetry atoms are shown in pink.
expected residues placed in electron density. MR-Rosetta failed to model 67 residues at the N-terminus which make up part of the apPOL$^{\text{exo-}}$ exonuclease domain. Over half of these residues are found exclusively in members of the *Plasmodium* genus and appear to be important in producing an active polymerase (7). Density is clearly present for these residues and a model is currently being manually built.

Several circumstances were unfavorable for structure determination: radiation-sensitive microcrystals, most of which failed to diffract in an ordered manner, low sequence identity to known structures and failure to grow SeMet-derivative crystals. Nonetheless, the advanced capabilities of synchrotron facilities resulted in complete data and successful structure determination. Further development of the exonuclease domain will require manual building in regions of low or absent sequence identity. Preliminary work, however, clearly supports the thumb-palm-finger structure observed in other A-type polymerases.

References


CHAPTER 3

CRYSTAL STRUCTURE OF THE APICOPLAST DNA POLYMERASE FROM PLASMODIUM FALCIPARUM: THE FIRST LOOK AT AN “ATYPICAL” A-FAMILY DNA POLYMERASE

Morgan E Miller, Jun-Yong Choe, Richard B Honzatko, and Scott W Nelson

Modified from a paper submitted for publication in PNAS Plus.

Abstract

Plasmodium falciparum, the primary cause of malaria, contains a non-photosynthetic plastid called the apicoplast. The apicoplast exists in most members of the phylum Apicomplexa and has its own genome along with organelle-specific enzymes for its replication. The only DNA polymerase in the apicoplast (apPOL) is an atypical A-family polymerase and here we present its crystal structure at a resolution of 2.9 Å. apPOL, the first structural representative of the atypical A-family, diverges from typical polymerases in two of three previously identified motifs and a region not implicated by sequence. Moreover, apPOL has an additional N-terminal subdomain, the absence of which severely diminishes 3’ to 5’ exonuclease activity. A compound known to be toxic to Plasmodium is a potent inhibitor of apPOL, suggesting apPOL is a viable drug target. The structure of apPOL will facilitate the mapping of inhibitors to binding loci and provide a basis for structurally-guided antimalarial drug design.

Introduction

The phylum Apicomplexa includes parasites responsible for malaria, toxoplasmosis, cyclosporiasis, and babesiosis. Malaria, caused by Plasmodium, is a mosquito-borne disease with 200 million cases annually, resulting in almost 450,000 deaths (1). Half of the world’s
population is at risk of infection by *Plasmodium falciparum*, the most common and deadly cause of malaria. Malarial parasites in central Africa, India, southeast Asia, and northern South America are becoming resistant to common antimalarial drugs, such as chloroquine, sulfadoxin-pyrimethamine, and artemisinin (2). Intense efforts to develop a malaria vaccine have provided only a single candidate that has progressed past Phase 3 clinical trials (3). Hence, the development of new and potent antimalarial treatments will likely play a central role in the eradication of malaria (1).

Nearly all members of the phylum Apicomplexa have an essential organelle called the apicoplast. Genetic analysis supports the evolution of the apicoplast from the chloroplast of a red algae (4), presumably captured in an endosymbiotic event. The apicoplast has lost its photosynthetic capabilities, but maintains several biochemical processes within the parasite. Proteins involved in fatty acid biosynthesis, heme synthesis, Fe-S cluster maturation, and isoprenoid synthesis contain an apicoplast targeting sequence (5). In *P. falciparum*, the synthesis of isoprenoid precursors is an essential role of the apicoplast during the liver and blood stages of infection in humans (6). Drugs that disable the apicoplast may be potent treatments and chemoprotectants against malaria (7).

As a consequence of gene transfer to the host cell nucleus, the genomes of endosymbiotic organelles are of reduced size relative to those of related free-living organisms (8). The apicoplast of *P. falciparum* contains a 35 kilobase (kb) circular genome with homology in genes and organization to plastid genomes of algae. Nearly all of the apicoplast genome is dedicated to gene expression, with no genes encoding for DNA replication proteins (9). Instead, the nuclear genome houses the prex gene (10), which is translated into a single polyprotein targeted to the apicoplast. Following translocation into the apicoplast, an unknown protease cleaves Prex into
separate and functional helicase, primase, and polymerase proteins (10). We have previously demonstrated that the polymerase derived from Prex (apPOL), the only known DNA polymerase targeted to the apicoplast, has activity and fidelity consistent with a replicative DNA polymerase (11).

A-Family polymerases predominantly come from bacteria, mitochondria, and bacteriophage (12). Bacterial DNA polymerases often elongate Okazaki fragments, remove RNA primers during replication, and repair DNA. Such polymerases employ a 5ʹ to 3ʹ polymerase domain, a 3ʹ to 5ʹ exonuclease proofreading domain, and a 5ʹ to 3ʹ exonuclease domain for primer removal and DNA repair (11, 12). The Klenow fragment from *Escherichia coli* Polymerase I (Klenow), the large fragment of DNA polymerase from *Bacillus stearothermophilus* (BF), the DNA polymerase from bacteriophage T7 (T7), the DNA polymerase from *Thermus aquaticus* (Taq), and the mitochondrial DNA polymerase (polγ) comprise the known structures of the A-family. Each resembles a right hand with finger, palm, and thumb subdomains. The palm subdomain is well-conserved, whereas the finger and thumb subdomains are relatively divergent across species (13). The 3ʹ to 5ʹ exonuclease domains are poorly conserved except for three amino acid motifs at or near the exonuclease active site (14). Outside of Apicomplexa, the closest homolog to apPOL is from *Cyanothece* sp. PCC 8802 (35% sequence identity), putatively acquired through lateral gene transfer from a lysogenic phage (15). Klenow is the closest structural homolog of apPOL, sharing 28% sequence identity.

A-family DNA polymerases divide into five clades: typical PolA bacteria, thermophilic viruses, Apicomplexa, *Aquificaceae/Hydrogenothermaceae*, and other unrelated bacteria (15). Typical A-family DNA polymerases have been well studied biochemically and structurally. Little is known, however, about the other four A-family clades, which have been deemed
atypical on the basis of three sequence motifs (14, 15). These signature motifs could represent distinct and unknown biochemical or physiological attributes, and indeed one lies in a region that could impact polymerase fidelity (15, 16). Bioinformatic studies classify apPOL as prokaryotic in origin and an atypical member of the A-family DNA polymerases (10).

Detailed mechanisms of DNA polymerization for typical A-family members come from extensive studies of Klenow and BF (17–21). In contrast, atypical attributes of apPOL are largely hypothetical. Presented here is the crystal structure of \textit{P. falciparum} apicoplast DNA polymerase, the first structure of an atypical A-family polymerase. The crystal structure defines atypical elements and suggests mutations that probe function. Moreover, several distinguishing features of apPOL may afford selective inhibition and facilitate the development of new antimalarial therapies.

Materials and Methods

\textbf{Protein Preparation}

All protein constructs were purified as described (22). Mutagenesis was performed using QuickChange mutagenesis to generate exonuclease deficient mutants (D82N and E84Q), catalytic base mutants (H578Q), and finger tyrosine mutants (Y481A/485A/Y486A).

\textbf{Crystallization and data collection}

Crystallization experiments employed the method of hanging drops. N-apPOL\textsuperscript{exo−} produced microcrystals (22) by the combination of equal parts protein solution (4–16 mg ml\textsuperscript{−1} N-apPOL\textsuperscript{exo−}, 20 mM Tris-HCl pH 8.0, and 400 mM NaCl) and precipitant solution (0.2 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5, and 35\% w/v polyethylene glycol monomethyl ether 5,000). We flash froze crystals in liquid nitrogen without additional cryoprotectant. We collected X-ray diffraction data at the Advanced Photon Source beamline 23-
ID-B (GM/CA) at 100 K and wavelength 1.0332 Å. We recorded 50 images from a single N-apol{exo} crystal (physical dimensions of 2×2×5 μm) using a MAR300 CCD detector, crystal-to-detector distance of 350 mm, an exposure time of 2.0 s, oscillation of 1°, and beam size of 5 μm.

Crystals of C-apol{exo} appeared after the structure determination of the N-apol{exo} crystal form. C-apol{exo} crystallized from equal parts protein solution (10–15 mg/ml C-apol{exo}, 0.2 M magnesium acetate tetrahydrate, and 0.1 M sodium cacodylate trihydrate pH 6.5) and precipitant solution (20% w/v polyethylene glycol 8,000). We collected X-ray diffraction data at the Advanced Photon Source beamline 23-ID-B (GM/CA) at 100 K and wavelength 1.0332 Å. We recorded 360 images from a single crystal (10×10×100 μm) using a MAR300 CCD detector crystal-to-detector distance of 350 mm, an exposure time of 2.0 s, oscillation of 1°, and beam size of 10 μm.

Data processing

We processed data from the N-apol{exo} crystal using HKL-2000 (23) by two approaches: a sigma cutoff of zero, which excluded all negative intensities, and a sigma cutoff of −1.0, which eliminated large-negative but retained weak-negative intensities. In subsequent refinements, the two data sets provided virtually identical values for R-work, R-free, average B, and parameters of stereochemistry. However, the completeness of the data set in the last shell (2.9–3.0 Å) was 96.4% and 58.2%, for sigma cutoffs of −1.0 and zero, respectively. The model and associated data deposited in the PDB results from refinement using the more complete data set (overall Rsym of 0.345) instead of the negative-purged data set (overall Rsym of 0.178). Statistical analysis of intensities indicated a 100% probability for the space groups P6_22 or P6_122 for crystals of N-apol{exo}. 
Data reduction included only the first 60 images (overall Rsym of 0.094) from the crystal of C-apPOL\(_{exo}\) in order to eliminate the impact of radiation damage. Statistical analysis of intensities indicated a 99% probability for the space group P6\(_5\) or P6\(_1\) for crystals of C-apPOL\(_{exo}\).

**Structure determination and refinement**

The first crystals to appear were those of N-apPOL\(_{exo}\). These presented a challenge related to crystal size (< 5 μm), limited lifetime in the X-ray beam (< 120 s), modest resolution of diffraction, and low sequence identity (28%) to the closest available structure, Klenow (PDB 2KFN). Regardless, the polymerase domain of Klenow led to an initial molecular replacement solution of the crystal structure of N-apPOL\(_{exo}\). Electron density recognizable as α-helices appeared only in the space group P6\(_5\)22. Subsequent molecular replacement and automated model building employed MR-Rosetta (24) and Klenow (PDB 2KFN). Operational parameters of MR-Rosetta were those of Terwilliger et al. (25), except 3 models rather than 100 were built during each round of MR-Rosetta to reduce computational time. Two cycles of MR-Rosetta resulted in a model with approximately 450 of 628 expected residues. We built an additional 125 residues manually using Coot (26) in regions between elements of secondary structure and in regions for which no structural analog existed (first 67 residues of apPOL). Refinement with Phenix.refine (27) along with user-specified donor-acceptor restraints aided in the correct registration of the amino-acid sequence to the electron density by enforcing canonical hydrogen-bonding patterns. Donor-acceptor restraints involved main-chain atoms initially, then all combinations of main-chain and side-chain atoms were imposed with target distances of 2.86 ± 0.1 Å. We did not use donor-acceptor restraints for residues preceding and following prolines in α-helices. We assigned significant difference density in the exonuclease and active site to two
molecules of ammonium sulfate (concentration of approximately 200 mM in crystallization droplets) and a partially hydrated Na$^+$ (concentration of approximately 400 mM in droplets). We restrained coordination distances to Na$^+$ involving the oxygen atoms of water molecules and a protein side-chain to 2.5 ± 0.01 Å. We placed seven ammonium sulfate molecules in the polymerase active site and at a lattice contact. In the late stages of refinement, we removed donor-acceptor restraints if refined distances exceeded 3.26 Å (4-sigma threshold), or if difference density clearly indicated that restraints restricted optimal placement of main- and/or side-chain atoms. Restraints on donor-acceptor distances maintained canonical secondary structure, favorable rotamer angles (0.2% unfavorable) and allowed conformations (99.3%) by Ramachandran analysis. The inclusion of hydrogen atoms using the Phenix routine ReadySet!, reduced clashscores from ~10 to 1.23 and resulted in an overall score out of MolProbity (28) of 1.38.

Molecular replacement determination of the C-apPOL$^{\text{exo}}$ crystal form employed the model for N-apPOL$^{\text{exo}}$. Unit cell parameters for the C-apPOL$^{\text{exo}}$ crystal form indicated two polymerase molecules in the asymmetric unit. We used the program Phaser (29) in the Phenix suite in placing the first polymerase molecule (TFZ > 8), identifying the space group as P6$_5$. A second run of Phaser with fixed coordinates for the first molecule resulted in the complete molecular replacement solution (TFZ > 16). We refined donor-acceptor restraints, but only between main-chain atoms. We placed an octahedrally coordinated Mg$^{2+}$ (approximate concentration of 200 mM in the crystallization droplet) at the exonuclease active site, with the oxygen atoms of five water molecules and a protein side chain restrained to a distance of 2.10 ± 0.01 Å. We added hydrogen atoms with the program ReadySet!. The refined model has 99.5% of residues in allowed/favorable Ramachandran conformations, 0.9% residues with unfavorable
rotamer angles, a clashscore of 0.77, and a score out of MolProbity of 1.07. We generated figures using PyMOL (http://www.pymol.org/) and VMD (http://www.ks.uiuc.edu/Research/vmd/) and calculated the electrostatic surface with PBEQ Solver (30).

**Pre-steady state polymerase activity assay**

Oligonucleotide template (5’–TTTTTTTCAGTAGCTGACTGACACCTG) and primer (5’–HEX-CAGGTGTCAGTCAGCTAGTG) were obtained from Integrated DNA Technologies. Reactions occurred at 25°C in 20 mM Tris-HCl pH 8.0, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT, and 0.1 mg ml\(^{-1}\) BSA with 200 µM dATP, 4 µM apPOL, and 400 nM annealed duplex DNA. Reactions were quenched with 100 mM EDTA. Time points less than 5 s were performed on a Biologic QFM-400. Data was visualized on a 16% denaturing acrylamide gel containing 7.5 M urea in Tris-borate-EDTA (TBE) buffer. Gels were scanned on a Typhoon FLA9500 (GE Healthcare). Data was analyzed using ImageQuant TL (GE Healthcare) and apparent \(k_{pol}\) values were determined by data fitting to a single exponential equation, \(P=P_0+P_\infty(1-e^{-k_{pol}t})\), in SigmaPlot. Assays were performed in triplicate.

**DNA-binding determination**

Oligonucleotide template (5’–TTTTTTTCAGTAGCTGACTGACACCTG) and primer (5’–CAGGTGTCAGTCAGCTAGTG) were obtained from the Iowa State University DNA Facility. The primer was radioactively labeled with \(^{32}\)P using T4-polynucleotide kinase (New England Biolabs). Reactions were carried out in the same buffer as above with 20 nM final duplex DNA and 0–2 µM apPOL. A 96-well Bio-Dot (Bio-Rad) was assembled with 0.45 µm nitrocellulose membrane (Bio-Rad), 0.45 µm positively charged nylon transfer membrane (GE Water + Process Technologies), and Whatman filter paper (Fisher Scientific). Membranes were washed...
three times with 100 µL of reaction buffer prior to loading 100 µL of sample and three more washes with 100 µL each of reaction buffer. Membranes were visualized using a Typhoon FLA9500 and analyzed using ImageJ (National Institutes of Health, Bethesda, MD). The $k_d$ was determined using DYANFIT (31) and experiments were performed three times with two replicates each.

**Apparent-$K_m$ and $V_{max}$ determination**

Parameters of steady state kinetics (apparent-$K_m$ for nucleoside triphosphates and $V_{max}$) for N-apPOL $\text{exo}$- and N-apPOL $\text{exo}$--Y481A/Y485A/Y486A were determined using the DNA hairpin substrate containing the Cy3 and BQ1 dye/quencher pair as described previously (32). Reactions were carried out at 25 °C in 20 mM Tris-HCl pH 8.0, 10 mM MgAc, 50 mM KAc, 1 mM DTT, 0.1 mg/ml BSA, and 2% DMSO with 50 nM hairpin DNA, 10 nM polymerase, and nucleotide concentration varying from 0 to 4 µM. All data were collected on a Cary Eclipse Fluorescence Spectrophotometer using the Cary Kinetics software at an excitation wavelength of 545 nm and emission wavelength of 570 nm. Initial rates were analyzed using SigmaPlot software and the Michaelis-Menten equation. Assays were performed in triplicate.

**Exonuclease activity assay**

Oligonucleotide template (5’–TTTTTTCACTAGCTGACTGACACCTG) and primer (5’–HEX-CAGGTGTCAGTCAGCTAGTG) were obtained from Integrated DNA Technologies. Reactions occurred at 37°C in 20 mM Tris-HCl pH 8.0, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT, and 0.1 mg ml$^{-1}$ BSA with 250 nM apPOL and 100 nM annealed duplex DNA. Reactions were quenched with 100 mM EDTA. Data was visualized on a 16% denaturing acrylamide gel containing 7.5 M urea in Tris-borate-EDTA (TBE) buffer. Gels were scanned on a Typhoon FLA9500 (GE Healthcare). Data was analyzed using ImageQuant TL (GE
Healthcare) and apparent $k_{exo}$ values were determined by data fitting in SigmaPlot fit to
$P = P_0 + P_\infty (1 - e^{-k_{exo}})$. Assays were performed in triplicate.

Results and Discussion

Overview of apPOL structure

Structural work employed an exonuclease-deficient mutant of apPOL (D82N/E84Q) with
either an N-terminal hexahistidine tag (N-apPOL\textsubscript{exo−}) or a C-terminal tag (C-apPOL\textsubscript{exo−}). Crystals
of N-apPOL\textsubscript{exo−} provided data to a nominal resolution ($I/\sigma I = 2$) of 3.7 Å (22). MR-Rosetta (24)
with Klenow [PDB 2KFN (33)] as a starting model
provided a molecular
replacement solution. The
model output by MR-Rosetta
had 450 of the expected 628
residues of N-apPOL\textsubscript{exo−}. We
manually built an additional
125 residues between
elements of secondary
structure and in a region for
which no structural analog
existed (the first 67 residues
of apPOL). Inclusion of data
from 2.9-3.7 Å improved the
interpretability of electron

<table>
<thead>
<tr>
<th>Data collection</th>
<th>N-apPOL\textsubscript{exo−} (5DKT)</th>
<th>C-apPOL\textsubscript{exo−} (5DKU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6\textsubscript{3}22</td>
<td>P6\textsubscript{5}</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>141.78, 141.78, 148.76</td>
<td>146.38, 146.38, 165.67</td>
</tr>
<tr>
<td>$a$, $\beta$, $\gamma$ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>47.36-2.90 (3.00-2.90)\textsuperscript{a}</td>
<td>47.90-2.90 (3.00-2.90)</td>
</tr>
<tr>
<td>$R_{sym}$</td>
<td>0.345 (over 1)*</td>
<td>0.094 (0.526)</td>
</tr>
<tr>
<td>$I / \sigma I$</td>
<td>3.4 (0.8)</td>
<td>14.9 (2.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.8 (96.4)</td>
<td>99.3 (98.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.2 (3.4)</td>
<td>3.5 (3.2)</td>
</tr>
</tbody>
</table>

| Refinement |                                          |                                          |
| Resolution (Å) | 47.34-2.90                             | 47.91-2.90                             |
| No. reflections | 19696                                   | 44,271                                 |
| $R_{work} / R_{free}$ | 0.27/0.30                            | 0.19 / 0.23                           |
| No. atoms |                                          |                                          |
| Protein  | 4,800                                    | 9,665                                  |
| Ligand/ion | 46                                      | 2                                      |
| Water    | 63                                       | 96                                     |
| $B$-factors (Å\textsuperscript{2}) |                                          |                                          |
| Protein  | 48.6                                     | 65.5                                   |
| Ligand/ion | 38.1                                    | 63.2                                   |
| Water    | 20                                       | 33                                     |
| R.m.s. deviations |                                          |                                          |
| Bond lengths (Å) | 0.003                                   | 0.002                                  |
| Bond angles (°) | 0.548                                   | 0.477                                  |

\textsuperscript{a}Values in parenthesis correspond to the highest-resolution shell.

\*See methods for comments on $R_{sym}$. 
density maps. Auxiliary donor-acceptor restraints maintained the canonical secondary structure and increased the number of residues in allowed/favorable regions of the Ramachandran plot.

The inclusion of hydrogen atoms in refinement reduced unfavorable interatomic contacts by one order of magnitude (clashscore decreased from 10 to 1). The polymerase model developed from the N-apPOL\textsuperscript{exo−} crystal form (PDB 5DKT) was the basis for a structure determination by molecular replacement of the C-apPOL\textsuperscript{exo−} crystal form (PDB 5DKU), resulting in a nominal

![Figure 1. Structure of the *P. falciparum* apicoplast DNA polymerase. The polymerase domain of apPOL\textsuperscript{exo−} with finger (orange), thumb (purple), and palm (blue) subdomains. An N-terminal region (NTR, gray) extends the exonuclease domain (green). The polymerase and exonuclease active sites are marked by green spheres. Selected elements of secondary structure are labeled with all elements defined in Figure 8.](image-url)
resolution of 2.9 Å. Auxiliary donor-acceptor restraints maintained canonical secondary structure, and the inclusion of hydrogen atoms reduced unfavorable interatomic contacts. Models of the polymerase, a total of three independent polypeptide chains over two crystal forms, exhibit only small conformational differences primarily in regions of lattice contact. Statistics of data collection and refinement are in Table 1.

Much of apPOL (Figure 1 and Figure 3) structurally resembles typical members of the A-family. The polymerase domain of apPOL shares finger, thumb, and palm subdomains of A-family polymerases. The palm subdomain is structurally conserved (Figure 3) and includes residues responsible for the coordination of active-site metals involved in DNA primer extension (34). The finger subdomain plays a role in binding template DNA and incoming nucleoside triphosphate (35). Helix O of the finger subdomain adopts an “open” conformation as defined by the corresponding helix of BF in its open [PDB 1L3S (18)] and closed conformation [PDB 2HVI (36)] (Figure 2). The thumb subdomain positions and translocates the growing duplex DNA and

**Figure 2.** Polymerase “open” and “closed” conformation. Helix N and Helix O of apPOL (blue) align with those of the open BF structure (cyan, PDB 1L3S) rather than those of the closed BF structure (pink, PDB 2HVI).
may play a role in processivity. Approximately 50 residues between Helix H and Helix I of the thumb subdomain of apPOL are not modelled due to weak electron density. In other A-family polymerases, the corresponding region has either high thermal parameters or weak electron density not represented by a model. However, in primer/template complexes, conserved residues in this region interact with the DNA backbone (13).

Even though exonuclease domains of A-family DNA polymerases vary broadly in sequence and secondary structure (14), active site residues are well-conserved in proofreading polymerases (19). apPOL has residues corresponding in position and type to polymerases with proofreading activity (details below). Nonetheless, a contiguous region of the apPOL exonuclease domain structurally diverges from the typical A-family polymerase...
(Figure 3). The N-terminal region (NTR), elements of the exonuclease domain juxtaposed to the NTR, and residues 170–206 spanning between the NTR and palm subdomain exhibit structural variations. Residues 170–192 include lysine residues which extend the electropositive surface charge from the polymerase active site to the exonuclease active site (see section on DNA-polymerase models), residues 193–199 cover the exonuclease active site, and residues 204–206 interact with the NTR.

**Polymerase and exonuclease active sites**

DNA polymerization putatively employs a two-metal-ion mechanism of nucleotidyl transfer (37) (Figure 4a, b). Asp410, Asp579, and Glu580 in apPOL correspond to conserved residues that bind Mg$^{2+}$ at sites A and B. The site B magnesium ion accompanies the binding of nucleoside triphosphate (16, 38). The magnesium ion at site A coordinates the 3’-hydroxyl group of the DNA primer (39). No metals are present in the polymerase active sites of either N-apPOL$^{exo-}$ (PDB 5DKT) or C-apPOL$^{exo-}$ (PDB 5DKU). The side chains of Asp579 and Glu580 do not adopt χ1 rotamers consistent with those in metal-bound complexes, instead hydrogen bonding with other active site residues (Figure 4c). Asp579 interacts with a conserved Arg377, whereas Glu580 interacts with Trp624 and/or Arg395. Metal-coordinating side chains of human Polβ, an X-family DNA polymerase (40), and A-family structures which lack DNA, exhibit similar salt links with arginyl side chains. In computational simulations of Pol β, the breaking of the salt link corresponding to apPOL Asp579-Arg377 is a significant energy barrier between the open and closed states (41). Consistent with this analysis, mutation of the arginine residue in Pol β results in a 2.5-fold enhancement in open-to-closed transitions (42). The rate-limiting step of
the polymerase reaction may be the change that brings about the closed conformation required for catalysis. This conformational change may be a checkpoint for correct nucleotide insertion.

A conserved Helix O lysine (Lys463 in apPOL Figure 4a,b) likely serves as a general acid for all classes of polymerases, protonating the pyrophosphate leaving group (43). A catalytic base is not firmly established, as several functional groups are close to the proton of the 3′-hydroxyl group of the DNA primer. Mounting evidence, however, weighs in favor of a highly conserved histidine (His578 in apPOL) as the proton acceptor. Mutation of this histidine (His881) to alanine in Klenow causes a 10-fold decrease in activity (44). In molecular dynamic simulations, the corresponding histidine in BF, which crystallizes in a constrained conformation, undergoes a rapid conformational change to hydrogen bond with the 3′-hydroxyl group of the primer (45). Mutation of His578 in apPOL to glutamine reduces catalytic activity by 100-fold (Figure 4d). apPOL\textsuperscript{exo} has a pre-steady state nucleotide incorporation rate (apparent-\(k_{pol}\)) of 22 ± 3 s\(^{-1}\), whereas apPOL\textsuperscript{exo/His578Q} has an apparent-\(k_{pol}\) of 0.21 ± 0.02 s\(^{-1}\). Sequence conservation, structural proximity, computational modeling, and biochemical studies together implicate His578 as the probable proton acceptor from the 3′-hydroxyl group in the polymerase reaction of apPOL.

The exonuclease reaction putatively involves a pair of divalent metal ions which coordinate the phosphodiester backbone of DNA (46) (Figure 4e). In apPOL, three aspartate residues (82, 143, and 215) and a glutamate residue 84 probably coordinate metal ions. The mutation of Asp82 and Glu84 to corresponding amides likely reduces affinity for metal ions; nonetheless, Mg\(^{2+}\) and Na\(^{+}\) are present in the exonuclease active sites of C-apPOL\textsuperscript{exo} and N-apPOL\textsuperscript{exo}, respectively. The N-apPOL\textsuperscript{exo} structure contains two sulfate anions roughly 8.5 Å apart that map to phosphoryl groups of DNA in the model of a DNA editing complex (described below). Conserved Tyr211 in apPOL may orient a metal-ligated hydroxide anion in the
Figure 4. Active Site Residues. (a) Alignment of active site residues of apPOL (blue), Klenow [gray, PDB 1KLN (17)], BF (cyan, PDB 1L3S), and Taq [green, PDB 1TAQ (29)] modeled with the nucleoside triphosphate and site B magnesium ion from Taq [PDB 2KTQ (26)]. (b) Schematic of polymerase active site with the 3′-end of the DNA primer (green), nucleoside triphosphate (blue), and divalent cations (red). Dashed lines represent donor-acceptor interactions and coordinate bonds. (c) Conserved residues Asp579 and Glu580 in apPOL (blue) hydrogen bond with side-chains of arginine residues, whereas in the open DNA-BF complex (cyan, PDB 1L3S) the corresponding interactions do not occur. (d) Pre-steady state polymerase kinetics of apPOLexo− (blue) and apPOLexo−H578Q (red). (e) The exonuclease active site of apPOL from PDB 5DKT with magnesium ion (green), and yellow spheres mark the positions of sulfate anions from PDB 5DKU. (f) Schematic of the exonuclease active site showing the 3′-end of the DNA primer (blue) and divalent cations (red).
exonuclease reaction (46). However, its side chain is rotated away from the active site in N-apPOL\textsuperscript{exo−}. In C-apPOL\textsuperscript{exo−} (sulfate anions absent), Tyr211 is directed into the active site, consistent with a role in catalysis. In addition to the aforementioned residues, two leucine residues (128 and 222, buried in separate locations of the domain core) and an aspartate residue 159 are conserved (Figure 3). Asp159 hydrogen bonds with the backbone amides of residues 139 and 140, orienting the main-chain carbonyls toward the metal binding sites of the exonuclease active site. This residue is present even in polymerases Taq and BF, which lack signature residues for exonuclease activity.

**DNA binding in the polymerase and exonuclease active sites**

Polymerase active sites have electropositive pockets which balance the negative charge of the phosphodiester backbone of DNA (47). apPOL exhibits the characteristic electropositive surface at the polymerase active site (Figure 5). For Klenow, BF, Taq, and T7, the polymerase active site electropositive region closely follows the template DNA strand, tapers off near the finger subdomain, and vanishes short of the exonuclease site (Figure 5b). In contrast, apPOL has a continuous lysine-rich electropositive surface running from the polymerase active site to the exonuclease site, along with an electropositive band that wraps around the finger subdomain and extends down the side of the protein (Figure 5a).

The DNA complex of BF polymerase (PDB 1L3S) is perhaps the most definitive representation of catalytically productive polymerase-DNA interactions, as it can extend a DNA primer within the crystal. The polymerase domain of BF shares only 25% sequence identity with apPOL, yet superimposes with a root mean squared deviation (rmsd) of 1.9 Å based on Ca atoms, placing the DNA primer/template into a region of electropositive charge with favorable contacts between DNA and the polymerase domain (Figure 5a). A nucleotide base from the
Figure 5. Surface charge of A-family polymerases. (a) The electropositive surface (blue) of apPOL extends well beyond the polymerase active site. (b) The electropositive surface (blue) of the DNA-BF complex (PDB 1L3S) maps to bound DNA (green), but is electronegative (red) in the region corresponding to the extended electropositive surface of apPOL.
template occupies the template pre-insertion site created from a slot between Helix O and Helix O1 (18) (Figure 6a, blowout). Residues that bind directly to DNA in the BF-DNA complex (Lys582, Ser585, Lys593, Arg615, Asn622, and Asn625) correspond to residues in apPOL of identical or similar type (Lys344, Ser347, Lys357, Arg377, Asn384, and Gln387, respectively). Lys372 and Lys382 of apPOL are proximal to the phosphodiester backbone of the DNA duplex and contribute to an electropositive surface that extends toward the exonuclease site. Tyr714 in open (PDB 1L3S) and in closed (PDB 2HVI) BF-DNA complexes exhibits a substantial displacement. In the open complex, Tyr714 blocks the template from access to its site for base insertion opposite that of the incoming nucleoside triphosphate, whereas in the closed complex the insertion site for the template base is accessible. Tyr714 is part of Helix O, the movement of which defines the open and closed conformations of BF. Other helices of the finger subdomain of BF with or without DNA superimpose well, as do the helices of Klenow (DNA absent). In the “open” apPOL structure, however, Tyr471 coincides with Tyr714 in its closed and unblocked BF-DNA complex. Moreover, helices of the finger subdomain of BF and Klenow exhibit substantial displacements relative to those of apPOL (Figure 6a). As these displacements remain in the superposition of apPOL onto Klenow and BF lacking DNA, the conformational differences are not necessarily due to the binding of DNA. Instead the helix displacements could be a consequence of sequence and structural differences in Helix Q and Helix P (described below).

The crystal structure of Klenow (PDB 1KLN) putatively reveals DNA bound in the proofreading mode. Superposition of that complex onto apPOL (rmsd of 2.4 Å based on Cα atoms) docks the DNA primer strand into the apPOL exonuclease site (Figure 6b). The phosphodiester backbone of the DNA primer tracks along the extended electropositive surface
Figure 6. Modeled DNA-apPOL complexes. (a) DNA (template strand yellow, primer strand orange) from BF (PDB 1L3S) modeled by superposition onto PDB 5DKT. Blow out shows interactions of the 5′-template overhang (yellow) in two orientations with BF (cyan) aligned to apPOL (blue). (b) DNA (template strand yellow, primer strand orange) from Klenow (PDB 1KLN) transformed onto apPOL (blue) by the superposition onto apPOL with active site magnesium ion (green sphere).
that leads from the polymerase active site into the exonuclease active site (Figure 4a). Residues in Klenow that interact directly with the primer strand correspond to apPOL Asp/Asn82, Glu/Gln84, Leu88, Asp143, Tyr211, and Asp215. The 3'-hydroxyl group of the primer strand is within hydrogen bonding distance of Glu/Gln84, corresponding to Glu357 of Klenow, which

Figure 7. Structure corresponding to atypical regions of the polymerase domain. Klenow (gray, PDB 1KLN) is superimposed onto apPOL (blue) in Panels a–c. (a) Six amino acids (dark blue) extend Helix O₁ of apPOL relative to that of Klenow. (b) apPOL (blue) lacks six amino acids (light gray) found in Klenow. (c) The connection from Strand 11 to Helix Q is shortened by 11 residues in apPOL relative to Klenow. (d) Superposition of apPOL (blue) onto T7 [purple, PDB 1ZYQ (40)] shows a short Helix Q in T7, and also an additional loop before Strand 10. (e) Superposition of apPOL (blue) onto polγ [yellow, PDB 3IKM (41)] reveals a vastly different structure between Strand 11 and Helix Q.
hydrogen bonds with the 3’-hydroxyl group and a metal ion (46). Leu361 in Klenow is in direct contact with the base at the 3’-end of the primer strand and maps to Leu88 in apPOL. However, Phe473 of Klenow does not align with a corresponding residue in the sequence of apPOL, but comes close (~4 Å) to Trp199. The two active site sulfates of the N-apPOL\(^\text{exo}^-\) (PDB 5DKT) are within 2 Å of the phosphoryl groups of the docked DNA primer strand.

In the polymerase editing complex, attention has historically focused on the 3’-end of the DNA primer. The DNA template strand, however, binds to a site distinct from that in the DNA complex at the polymerase active site (Figure 6). The template DNA docks onto the electropositive surface that extends around apPOL between the finger subdomain and the exonuclease domain (Figure 5a), suggesting the possibility of extended protein-template interactions. This additional electropositive surface, coupled with the electropositive surface along the phosphodiester backbone of the primer strand, implies a stable editing complex. Kinetics experiments revealed an unusually stable complex in apPOL due to the misincorporation of a triphosphate nucleoside (11).

**Atypical regions**

Three sequence motifs define the atypical A-family polymerase (15); however, only two motifs map to divergent structure. The first motif (\(^\text{481YANTYYG}^{486}\) in apPOL) extends Helix O\(_1\) (Figure 7a). Tyrosines 481, 485, and 486, conserved within the atypical A-family, point toward the base moieties of template DNA and nucleoside triphosphate. Mutation of Tyr481, Tyr485, and Tyr486 to alanine (N-apPOL\(^\text{exo}^-/\text{Y481A/Y485A/Y486A}\)) results in a 3-fold decrease in

<table>
<thead>
<tr>
<th></th>
<th>(K_{\text{dDNA}}) (nM)</th>
<th>Apparent-(K_{\text{m,dNTPs}}) (nM)</th>
<th>(V_{\text{max}}) (nM*min(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-apPOL(^\text{exo}^-)</td>
<td>38 ± 3</td>
<td>100 ± 30</td>
<td>6.6 ± 0.5</td>
</tr>
<tr>
<td>N-apPOL(^\text{exo}^-/\text{Y481A/Y485A/Y486A})</td>
<td>160 ± 30</td>
<td>25 ± 10</td>
<td>1.4 ± 0.1</td>
</tr>
</tbody>
</table>
polymerase activity, a 4.5-fold increase in $k_d\text{-DNA}$, and a 4-fold decrease in apparent-$K_{m,dNTP}$ (Table 2). The opposite effect on $k_d\text{-DNA}$ and apparent-$K_{m,dNTP}$ due to mutation suggest binding antagonism (direct or indirect) between the dNTP and primer/template DNA. Evidently, these helix-extending tyrosines stabilize the binding of DNA primer/template at the expense of dNTP binding affinity. Interestingly, human polγ has a five-residue extension of Helix O₁. However, the residue types do not match those of apPOL.

The second motif shortens the loop from Helix J to Helix K in the palm subdomain. Typical A-family polymerases have a six-residue loop, whereas apPOL connects helical elements with two residues 390–391 (Figure 7 and Figure 3). The six-residue loop includes a conserved arginine and glycine (positions 682 and 686, respectively in Klenow). Conceivably Arg390 of apPOL corresponds to Arg682 of Klenow which is involved in contacting duplex DNA. The third atypical region, an insertion of two amino acids ($^{523}{QY}^{524}$), results in no secondary structural differences between apPOL and the typical A-family polymerases. A two-residue deletion at the N-terminal end of Helix P compensates for the insertion located on Strand 10 of the finger subdomain.

The structure of apPOL reveals another difference relative to Klenow, BF, and Taq that was not previously recognized on the basis of sequence analysis. An eleven-residue deletion in apPOL eliminates an extended loop between Strand 11 and Helix Q, as well as residues from the N-terminal end of Helix Q (Figure 7c). The sequence deletion, which is conserved in atypical A-family polymerases, may be responsible for a shortened Helix P and the aforementioned conformational displacements of helices of the finger subdomain. These structural differences are in proximity to the 5'-overhang of the template DNA strand (Figure 6a). Arg784 in BF (836 in Klenow) is absent in apPOL, and Ser717, Tyr719, and Arg789 in BF correspond respectively
to Asn474, Val476, and Thr540 in apPOL. These residues in BF and Klenow interact with the 5′-overhang of the DNA template strand in its primer-extension binding mode (20), (43). T7 DNA and poly polymerases also have shortened Q helices. However, other elements of secondary structure fill the void (Figure 7d, e). In spite of the altered structure, N- and C-apPOLexo binds primer/template DNA with a $k_{d-DNA}$ of 38 ± 3 and 23 ± 2 nM respectively, comparable to other stable protein/DNA complexes. The potential loss of DNA-interacting residues in this loop may be compensated for by the triple tyrosine insertion loop.

A 50-residue extension to the N-terminal side of the exonuclease domain is the most apparent difference between apPOL and other A-family members (Figure 3 and Figure 8). Residues in the NTR are exclusive to and highly conserved within the *Plasmodium* genus and may be necessary for a properly functioning polymerase (11, 22). The NTR has a short helical region followed by two anti-parallel β-strands (Strand 1A and Strand 1B). The antiparallel strands

*Figure 8. Plasmodium N-terminal region. A network of hydrogen bonds and hydrophobic interactions connects the NTR of apPOL to the exonuclease domain. The residues involved in these interactions are highly conserved across Plasmodium spp.*
extend the central β-sheet of the exonuclease domain. However, Strand 1B connects to Strand 1 through only two main-chain hydrogen bonds. Additional hydrogen bonds are between Strand 1B and residues 202 to 206 of a long loop (residues 170 to 206) that covers the exonuclease active site and extends to the palm subdomain. Hydrophobic interactions between the NTR and the rest of the exonuclease domain suggest a stable interface that is unlikely to separate. Moreover, the interface between the NTR and exonuclease domain proper is invariant over three independent polypeptide chains subject to different lattice contacts and conditions of crystallization. The 12-residue β-strands of the NTR coil into a right-handed helix similar to the β-hairpin structures of the Arc and MetJ repressors. The hairpin structures of Arc and MetJ repressors bind to the major groove of duplex DNA (49). NTR-DNA interactions could increase the intrinsic processivity of apPOL, as an extrinsic processivity factor for apPOL has not been identified.

Deletion of the NTR (C-apPOL52–628) significantly decreases expression levels (20-fold less relative to wild-type), purity, and protein stability. Identification of C-apPOL52–628 on gels employed Western blot analysis (Figure 9a) and densitometry analysis of Coomassie-stained gels determined the percent purity of the enriched protein (Figure 9b). Low enzyme purity confounds assays for DNA binding and pre-steady state kinetics. However, we captured reliable results from steady-state kinetics of enriched proteins. Preparations of exonuclease deficient C-apPOL 52–628/exo− and exonuclease and polymerase deficient C-apPOL 52–628/exo−/pol− determined levels of exonuclease and polymerase contamination in C-apPOL 52–628. C-apPOL 52–628/exo−/pol− had no detectable polymerase activity, indicating no measurable polymerase contamination in preparations of C-apPOL 52–628. Therefore, at saturating DNA (400 nM) and dNTP (16 μM), the observed specific activities of 0.04 s⁻¹ and 0.03 s⁻¹ for C-apPOLexo− and C-apPOL 52–628/exo−, respectively, suggest little or no effect on polymerase activity due to NTR deletion. No
exonuclease activity is observed for N-apPOL<sup>exo−</sup>. Exonuclease activities of C-apPOL<sup>52−628</sup> and C-apPOL<sup>52−628/exo−</sup> are similar and less than that of C-apPOL<sup>52−628/exo−/pol−</sup> (Figure 9c), suggesting exonuclease activity is severely affected by the NTR deletion and that observed exonuclease activity comes from contaminants. Hydrophobic interactions and a network of hydrogen bonds connect the NTR to the exonuclease domain (Figure 8). Removal of the NTR disrupts these interactions and likely results in an unstable or misfolded exonuclease domain.
Potential for Drug Discovery

DNA polymerases of HIV, herpes simplex virus, hepatitis B virus, and cytomegalovirus are targets of therapeutic drugs, the majority of which are nucleoside-based (50). Such drugs, although effective inhibitors, often lack selectivity due to the conservation of polymerase active sites. The optimization of lead compounds, guided by structure, however, has yielded potent non-nucleoside and uncompetitive inhibitors against HIV reverse transcriptase (13). The structure of apPOL reveals attributes that may facilitate structure-guided drug design. MetaPocket (51), a software predictor of potential ligand binding sites, correctly identifies the polymerase and exonuclease active sites, and six other potential ligand binding sites. Four of the six sites are in the exonuclease domain and two are in the polymerase domain (Figure 10). A similar analysis of poly reveals no pockets corresponding to those of apPOL. Moreover, Malaria Box (52) compound MMV666123 inhibits polymerase activity of apPOL (IC_{50} of ~700 nM), and preferentially goes (on the basis of in silico analysis) to one of the potential ligand binding sites on apPOL. Hence, the structure of apPOL offers a point of departure in the design and discovery of specific inhibitors through computer docking and co-crystallization experiments.

The benefits of a new drug targeting the apicoplast of P. falciparum and related organisms could transcend those of the typical rapid elimination of an invading organism. Inhibition of apicoplast genome replication, transcription, and/or translation results in the “delayed death” of the organism (53). In the presence of inhibitors of apicoplast topoisomerase, parasites grow, divide, and produce daughter cells that die instead of establishing productive infections (54). The delayed death phenotype occurs in both the red blood cell and liver stages of
infection, and timely treatment of liver infection results in daughter parasites that cannot establish a productive blood cell infection (55). Mice exposed to live sporozoites simultaneously with apicoplast inhibitors exhibit no infection of red blood cells and acquire immunity to subsequent infections with live sporozoites. The delayed death phenotype may extend to the sporozoite, which would block transmission of the disease from human to mosquito (56). Hence, effective drugs targeting the apicoplast may facilitate a mechanism of natural immunization toward the most prevalent infectious disease known to human civilization.

References


CHAPTER 4
A HIGH-THROUGHPUT ASSAY TO IDENTIFY INHIBITORS OF THE APICPLAST DNA POLYMERASE FROM PLASMODIUM FALCIPARUM

Morgan E Miller, Eric E Parrott, Risham Singh, and Scott W Nelson


Abstract

Infection by Plasmodium falciparum is the leading cause of malaria in humans. The parasite contains a unique and essential plastid-like organelle called the apicoplast that, similar to the mitochondria and chloroplast, houses its own genome that must undergo replication and repair. The putative apicoplast replicative DNA polymerase, POM1, has no direct orthologs in mammals, making the P. falciparum POM1 an attractive antimalarial drug target. Here, we report on a fluorescent high-throughput DNA polymerase assay that relies on the ability of POM1 to perform strand-displacement synthesis through the stem of a DNA hairpin substrate, thereby separating a Cy3 dye from a quencher. Assay-validation experiments were performed using 384-well plates and resulted in a signal window of 7.90 and a Z' factor of 0.71. A pilot screen of a 2880-compound library identified 62 possible inhibitors that cause more than 50% inhibition of polymerase activity. The simplicity and statistical robustness of the assay suggest it is well suited for the screening of novel apicoplast polymerase inhibitors that may serve as lead compounds in antimalarial drug-discovery efforts.

Introduction

More than 3 billion people live in areas at risk of malaria, and the disease causes an estimated 650,000 deaths per year worldwide (1). The majority of these deaths occur in African
children younger than age five. Malaria is caused by parasites of the genus *Plasmodium*, mainly *P. falciparum* and *P. vivax*, with the former being the most common and most deadly (2).

*Plasmodium* is in the phylum Apicomplexa, the members of which are exclusively animal parasites and are responsible for several human and agricultural diseases, such as toxoplasmosis (*Toxoplasma gondii*), babesiosis (*Babesia bovis*), cyclosporiasis (*Cyclospora cayetanensis*), and coccidiosis (*Eimeria falciformis*). All of these organisms contain an unusual organelle called the apicoplast (3). The apicoplast is evolutionarily related to a chloroplast and was obtained through a secondary endosymbiotic event with red algae (4). The apicoplast has lost its photosynthetic ability, but it has retained the biochemical pathways for the synthesis of isoprenoids, fatty acids, and heme, along with iron–sulfur cluster assembly (5). The parasite has come to completely rely on the apicoplast for production of isoprenoids (6). This dependent relationship has brought a great deal of attention to the apicoplast and its potential as a drug target (7).

Because the apicoplast is derived from a chloroplast, it contains its own genome and harbors the proteins necessary for DNA replication, transcription, and translation (8). At 35 kb, the genome is greatly reduced compared to the chloroplast. Although many of the proteins required for transcription and translation are encoded by the apicoplast genome, all of the DNA replication proteins are encoded in the nuclear genome and are imported into the apicoplast following translation. Inhibition of apicoplast replication, transcription, and translation with antibiotics such as tetracycline and ciprofloxin causes a phenomenon referred to as the “delayed-death” phenotype (8). When treated with these antibiotics, parasites continue to grow, divide, and produce daughter cells that die instead of establishing productive infections. Although the biological basis of this delayed-death phenotype is poorly understood, it clearly establishes the apicoplast as a target of drug action.
The apicoplast DNA polymerase (POM1; hereafter referred to as apPOL, or apPOL^{exo−} for the exonuclease deficient mutant) represents an especially attractive target for antimalarial drug development. apPOL is clearly of prokaryotic origin, with its nearest homolog outside of Apicomplexa being the replicative polymerase from the cyanobacteria *Cyanothece* sp. PCC 8802 (35% identity) (9). The most similar human DNA polymerases are the lesion bypass polymerases theta and nu (23 and 22%, respectively), with the other human DNA polymerases displaying < 20% identity. In contrast, there is 84% identity between the apPOLs from the two primary causative agents of human malaria, *P. falciparum* and *P. vivax*, suggesting that drugs targeted against the *P. falciparum* apPOL would be effective in treating *P. vivax* infections as well. In addition, polymerases appear to be well suited to therapeutic inhibition because there are commercial drugs targeting the polymerases of HIV, herpes simplex virus, hepatitis B virus, and cytomegalovirus (9).

Many assays for DNA polymerase activity have been reported, and several have been used in a high-throughput fashion. To facilitate the high-throughput screening for apPOL inhibitors, we have modified an established assay based on a molecular-beacon DNA substrate for use in high-throughput screening (10). We have used this assay in a continuous mode to determine kinetic parameters for apPOL and in a quenched-time mode for high-throughput screening of a small library of 2880 compounds. The assay proved to be statistically robust with a Z’ factor of 0.71, a signal window of 7.9, and a signal-to-background ratio of 41. Many potential small-molecule inhibitors were identified in the pilot screen.
Materials and Methods

Materials

The 33-mer DNA hairpin substrate (5' CCTCTCCGTGTTTGTACTTCCCGTCAG-AGAGG) containing either Cy3 or hexachlorofluorescein (HEX) dyes at the 5’ end, and either Iowa Black® FQ (IBQ) or Black Hole Quencher®-1 (BQ1) quenchers at the 3’ end were purchased from Integrated DNA Technologies. The hairpin contains a 6-nucleotide-long double-stranded stem and 21-nucleotide-long single-stranded loop. A 9-nucleotide primer (5’- GACGGGAAG) that is complementary to a portion of the loop region was obtained from the Iowa State University DNA Facility. Nickel-agarose was bought from the Sigma- Aldrich Chemical Company, deoxyribonucleotides were acquired from Invitrogen or Sigma-Aldrich, chloroquine from Santa Cruz Biotechnology, and aurintricarboxylic acid (ATA) from Acros Organics.

Purification of apPOL$_{\text{exo}^-}$

The exonuclease-deficient apPOL (D1470N and E1472Q) was expressed and purified as previously described with a few exceptions (11). Cells were lysed in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, and 5 mM imidazole. Clarified lysate was loaded on ~5 mL of Ni-agarose resin and washed with 100 mL of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, and 5 mM imidazole followed by a high-salt wash of 100 mL of 20 mM Tris-HCl pH 8.0, 1 M NaCl, and 25 mM imidazole. Protein was then eluted in a minimal volume of elution buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, and 150 mM imidazole. The resulting protein was then further purified through size exclusion chromatography using a 320 mL HiLoad 26/60 Superdex 200 equilibrated in 20 mM Tris-HCl pH 8.0 and 400 mM NaCl.
Kinetic Assays

The enzymatic activity of apPOL<sup>exo−</sup> was determined using the DNA hairpin substrate containing the Cy3 and BQ1 dye-quencher pair (Figure 1A). The reactions were carried out at 25 °C in 20 mM Tris-HCl pH 8.0, 10 mM MgAc, 50 mM KAc, 1 mM DTT, 0.1 mg/ml BSA, and

![Diagram A](image1)

**Figure 1.** Polymerase activity with the high-throughput DNA substrate. (A) A schematic of the DNA substrate and the steps leading to an increase in fluorescence. Additional details can be found in the text. (B) The increase in fluorescence signal is dependent on protein concentration. A polymerase concentration of 3 nM was determined to be optimal for high-throughput screening because approximately half of the DNA substrate would be consumed after 7.5 minutes. The dotted line represents 1 nM apPOL<sup>exo−</sup>, the short dashed line 2 nM apPOL<sup>exo−</sup>, the large dashed line 4 nM apPOL<sup>exo−</sup>, and the solid line 8 nM apPOL<sup>exo−</sup>. The inset is a plot of the initial velocities versus enzyme concentration. The slope of the line represents the apparent-<i>k</i><sup>catalytic</sup>, which is 4.95 min<sup>−1</sup>. (C) Denaturing urea–polyacrylamide gel analysis of product formation using 32P-labeled primer. The unextended primer and fully extended product are marked by arrows at the bottom and top of the gel, respectively. (D) Overlay of polymerase activity based on polyacrylamide gel analysis (♦) and fluorescence assay using the hairpin substrate (gray line).
2% DMSO, and all data were collected on a Cary Eclipse Fluorescence Spectrophotometer using the Cary Kinetics software at an excitation wavelength of 545 nm and an emission wavelength of 570 nm.

To validate that the Cy3 dye was not affecting the rate of apPOL\textsuperscript{exo−} polymerization, two separate reactions were carried out with 3 nM apPOL\textsuperscript{exo−}, 16 µM deoxynucleoside triphosphates (dNTPs) (~5× the K\textsubscript{M}), and 50 nM DNA substrate (which provided a strong fluorescent signal). In the first reaction, polymerase activity was determined on a Cary Eclipse, as stated above. For the second reaction, the primer that is annealed to the hairpin DNA was radioactively labeled with \textsuperscript{32}P using T4-poly nucleotide kinase (New England Biolabs). The reaction was initiated by mixing equal volumes of apPOL\textsuperscript{exo−} and a mixture containing dNTPs and DNA. Time points were then removed and quenched with 0.1 M ethylenediaminetetraacetic acid (EDTA) and 80% (v/v) formamide. Each time point was analyzed using a 16% denaturing polyacrylamide gel containing 7.5 M urea in a Tris-borate-EDTA buffer. The gel was visualized using a FujiFilm FLA-5100 Fluorescent Image Analyzer and analyzed using ImageJ (National Institutes of Health, Bethesda, MD).

To determine the optimum enzyme concentration and time point for the stopped-time high-throughput assay, the concentration of apPOL\textsuperscript{exo−} was varied from 0 to 8 nM. It was found that for a reaction containing a final concentration of 16 µM dNTPs and 50 nM DNA substrate, 3 nM apPOL\textsuperscript{exo−} led to 50–90% of substrate use over a time course of 7.5 minutes (a convenient time for the HT robotic platform). The steady-state kinetics values (K\textsubscript{M} for nucleotide and the DNA substrate) were established using 20 nM apPOL\textsuperscript{exo−}. To determine the K\textsubscript{M}-DNA, the nucleotide concentration was held at 16 µM while the DNA concentration was varied from 0 to 800 nM. To determine the K\textsubscript{M}-dNTPs, the DNA concentration was held at 50 nM and the
nucleotide concentration was varied from 0 to 16 µM. Initial rates were analyzed using the Matlab R2011b software (Mathworks) using the standard Michaelis-Menten equation.

ATA was identified through high-throughput screening as an inhibitor of apPOL\textsuperscript{exo−}. Although ATA has previously been classified as an inhibitor by mimicking DNA, recent studies have shown that this may not always be the case (12). For this reason, ATA was selected for further kinetic studies. The IC\textsubscript{50} of ATA was determined using 16 µM dNTPs, 50 nM DNA substrate, 3 nM apPOL\textsuperscript{exo−}, and varying concentrations of ATA from 0 to 12 µM. ATA was incubated with apPOL\textsuperscript{exo−} in the absence of DNA and nucleotide for several minutes prior to the start of the reaction. The reaction was initiated by simultaneous addition of DNA and dNTPs, and time courses lasted approximately 15 minutes. The slope of the linear portion of the time course was used to determine the initial velocities. The resulting initial velocities were fit with the Matlab R2011b software (Mathworks) using the following equation:

\[
V_0 = \frac{V_{max}}{1 + \frac{I}{IC_{50}}}
\]

where \(V_0\) represents initial velocity, \(V_{max}\) represents maximum velocity, \(I\) is the ATA concentration, and \(IC_{50}\) is the concentration of ATA that produces 50% inhibition.

Determination of the ATA inhibition mechanism was carried out by incubating 20 nM apPOL\textsuperscript{exo−} with varying concentrations of ATA for several minutes. To establish the inhibition mechanism versus nucleotides, the DNA substrate was held constant at 50 nM, and the dNTPs were varied from 0 to 16 µM. To establish the inhibition mechanism versus DNA, the nucleotide was held constant at 16 µM nucleotides, and the DNA substrate was varied from 0 to 800 nM. The resulting data were fit against inhibition models describing competitive, noncompetitive,
mixed, and uncompetitive inhibition using the Dynafit (13) software. The best-fitting model was determined using the model discrimination analysis function in the Dynafit software.

**Validation Assay and Pilot Screen**

The validation assay was performed at the University of Iowa High Throughput Screening Facility. The reaction was carried out with a final concentration of 16 µM dNTPs, 50 nM DNA substrate, and 3 nM of apPOL\textsuperscript{exo−} in 20 mM Tris-HCl pH 8.0, 10 mM MgAc, 50 mM KAc, 1 mM DTT, 0.1 mg/mL BSA, and 2% DMSO in the presence and absence of 8 mM chloroquine. Solutions were loaded into a Nunc 384-well plate (Thermo Scientific) using a MicroLab Pipettor (Hamilton). A 2× protein solution was incubated with 16 mM chloroquine (8 mM final) when applicable, and then mixed with an equal volume of 2× DNA solution to initiate the reaction. After 7.5 minutes at room temperature, the reaction was quenched with 100 mM EDTA. Fluorescence was detected at 545 nm using an Envision High Throughput Plate Reader (Perkin Elmer). Resulting data were analyzed and evaluated for robustness through determination of the signal window and Z' (14,15). Maximum signal values are those resulting from reactions containing no inhibitors, and minimum signal values are background readings with no fluorescent substrate controls.

A 2880-small-molecule pilot screen composed of the Spectrum Collection (composed of 2320 compounds) and University of Iowa Legacy Collection (composed of 560 compounds) was carried out at the University of Iowa High Throughput Screening Facility. The screen took place under the same conditions as the validation assay with a 2× protein solution being incubated with 20 µM (10 µM final) small molecules. Small molecules were added using a Star Robotic Liquid Handler System (Hamilton). Controls containing no protein, 8 mM chloroquine, and no small
molecules were performed on each 384-well plate. Fluorescence was detected at 545 nm using an Envision High Throughput Plate Reader. Resulting data were analyzed using Microsoft Excel.

Results and Discussion

**Purification of apPOL$^{\text{exo}^-}$**

The expression of apPOL$^{\text{exo}^-}$ is extremely robust with yields approaching 50 mg of pure protein per liter of Luria broth. The purified protein is stable at room temperature for up to 8 hours and for at least 1 year frozen at −80 °C. The enzyme is also highly active (11), enabling a very small amount of protein to be used in each assay (3 nM). In the 384-well format, a single preparation of protein from 1 L of media is adequate to screen more than 13,000 plates.

**Fluorescence Assays**

The hairpin DNA substrate used here was previously developed by others but contained a carboxytetramethylrhodamine (TAMRA) fluorophore and a Dabcyl quencher (10). The substrate was used in a continuous DNA polymerase assay in a low-throughput fashion (Figure 1A) (10). Although the assay, as previously reported, is suitable for continuous assays, because of incomplete quenching of the TAMRA dye fluorescence by the Dabcyl quencher, the signal-to-background ratio is relatively low (~fivefold) and is borderline for high-throughput screening. For this reason, we attempted to find a better fluorophore-quencher pair. We initially replaced the pair with hexachlorofluorescein and the Iowa Black® RQ quencher, but saw only modest improvements in the statistical parameters. Using 96-well plates, this dye-quencher pair resulted in a signal window of ~6.6, a $Z'$ factor of 0.60, and a signal-to-background ratio of 7.1. We then tried the Cy3 fluorophore and BQ1 quencher pair, which resulted in better statistical parameters. Using 96-well plates, the Cy3-BQ1 dye-quencher pair resulted in a signal window of ~8.9, a $Z'$
factor of 0.73, and a signal-to-background ratio of 9.7. Because some proteins have a strong affinity for dye molecules, we confirmed that the observed rate was linearly dependent on the concentration of apPOL<sup>exo−</sup> concentration. As seen in Figure 1B, the increase in fluorescent signal is proportional to the amount of apPOL<sup>exo−</sup> in the reaction, with an apparent-<i>k<sub>catalytic</sub></i> of 4.95 min<sup>−1</sup>. To directly compare the increase in fluorescence signal to extension of the primer, we <sup>32</sup>P labeled the primer strand of the substrate and analyzed the polymerase time course using denaturing urea-polyacrylamide gel electrophoresis (Figure 1C). Comparison of the fluorescence signal with the quantification of the gel indicates that the increase in fluorescence is directly proportional to the amount of primer extended (Figure 1D). The gel analysis also indicates that there is some degree of pausing when apPOL<sup>exo−</sup> reaches the stem portion of the substrate, where it must perform strand-displacement synthesis to separate the fluorophore from the quencher. We then used the assay in continuous mode to determine the steady-state kinetic parameters of apPOL<sup>exo−</sup>, which resulted in apparent-<i>K<sub>M</sub></i> values of 3.4 ± 1.0 µM and 305 ± 99 nM for dNTPs and DNA, respectively (Figure 2). These values are in close agreement with those determined using the standard <sup>32</sup>P-labeled primer-extension assay (11).
Following the tests using 96-well plates, the assay was validated for high-throughput screening using 384-well plates. Half of the plate was used as a minimum signal reference (8 mM chloroquine and 3 nM polymerase), and the other half was used as the maximum signal control (3 nM polymerase added). The reactions were initiated by the addition of DNA and nucleotide, and all wells were quenched with EDTA after 7.5 minutes. The plates were analyzed, and resulting values for the signal window, $Z'$ factor, and signal-to-background ratio were determined to be 9.27, 0.71, and 17.96, respectively.

We next performed a pilot screen using the Spectrum Collection and the University of Iowa Legacy Collection (Figure 3). The screen consisted of nine 384-well plates with the first two columns of each plate reserved for reference wells (no polymerase addition), the upper half of the last two columns for the fully inhibited reaction (8 mM chloroquine), and the lower half of the last two columns for the uninhibited reaction (polymerase addition only). The average fluorescent signals from the fully active and inhibited reactions were $1638 \pm 143$ (signal-to-noise ratio).

**Figure 3.** High-throughput pilot screen. The small-molecule pilot screen of the Spectrum Collection (2320 compounds) and the University of Iowa Legacy Collection (560 compounds) resulted in 62 hits that inhibit polymerase activity by at least 50%. No polymerase controls (gray X’s that occur periodically near the bottom of the graph) were performed on each of the nine 384-well plates. Each symbol type represents the results from an individual 384-well plate.
of 11) and $39 \pm 14$ (signal-to-noise of 2.8), respectively. These data were used to calculate values for the signal window, $Z'$ factor, and signal-to-background ratio of 7.9, 0.71, and 42, respectively. The increase in the signal-to-background ratio compared to the initial characterization is due to an optimization of the plate-reader photomultiplier voltage prior to fluorescence acquisition.

The pilot screen identified 62 compounds that resulted in a fluorescence signal of less than 700 (i.e., causing more than 50% inhibition), giving a hit rate of approximately 2.1%. Among these 62 compounds were several nonspecific inhibitors of nucleic acid enzymes, such as DNA intercalators (e.g., ethidium bromide, epirubicin, and acriflavine) and known or suspected DNA mimics (16) (e.g., suramin and ATA). It is expected that nonspecific inhibitors can be removed by using this assay to perform counterscreens against a panel of selected DNA polymerases (e.g., *Escherichia coli* Pol I and human pol γ). In addition to removing compounds that act as DNA intercalators and mimics, the counterscreens would also identify compounds that may be interfering with the fluorescent signal of the DNA substrate and/or product. Alternatively, an orthogonal assay similar to the one shown in Figure 1C could be used to rule out dye interference.

ATA was selected for determination of its inhibition mechanism because it appears to be capable of multiple inhibition mechanisms depending on the enzyme it is acting on. Although it is generally assumed that the ATA polymerizes and acts as a nonspecific DNA mimic (i.e., competitive with DNA), recent studies using HCV helicase found that ATA was competitive with adenosine triphosphate and noncompetitive with DNA (12). To determine the inhibition of ATA for apPOL$^{exo-}$, we first measured the dose–response of ATA under our standard assay conditions and determined the $IC_{50}$ to be $0.9 \pm 0.2$ µM (Figure 4A). We then determined the
inhibition mechanism of ATA against dNTPs. The concentrations of ATA and dNTPs were varied above and below their IC$_{50}$ and K$_M$, respectively, and the resulting data were fit to a variety of inhibition mechanisms using the Dynafit software (13). The best-fitting mechanism for ATA versus dNTPs was true noncompetitive with K$_I$-ATA and K$_M$-dNTPs values of 0.99 ± 0.08 µM and 3.6 ± 0.5 µM, respectively.

We also determined the inhibition mechanism of ATA against the DNA substrate. Again, the concentrations of ATA and DNA were varied, and the data were fit to several possible inhibition mechanisms. The best-fitting mechanism was true noncompetitive with a K$_I$-ATA and K$_M$-DNA of 1.47 ± 0.09 µM and 261 ± 15 µM, respectively. Because the inhibition mechanism of ATA is true noncompetitive against both dNTPs and DNA (i.e., the binding of ATA does not affect the apparent-K$_M$’s for the substrates), the two determined K$_I$’s and the IC$_{50}$ values should
be the same. We have determined values of 0.90, 0.99, and 1.47 µM for the IC50, K_i (against dNTPs), and K_i (against DNA), respectively, which are all in reasonable agreement with each other. In addition, the apparent-K_M values determined using the data shown in Figure 2 correspond very well to the K_M values determined using the data in Figure 4B and Figure 4C (3.4 and 3.6 µM for K_M-dNTPs and 305 and 261 nM for K_M-DNA, respectively). Based on this analysis, it appears that in the case of apPOL_exo−, ATA is not acting simply as a DNA mimic, but binds to the enzyme at a unique site that does not directly overlap with either the nucleotide- or DNA-binding site.

In conclusion, we have improved on a previously developed DNA polymerase assay by altering the fluorophore and quencher pair, thereby increasing its dynamic range. We found that the assay provides steady-state kinetic parameters that are comparable to the traditional 32P-based assay and is both highly sensitive and reproducible using 96- and 384-well plates. A pilot screen resulted in excellent statistical parameters (Z’ factor and signal window) and identified a number of potential inhibitors of the apPOL_exo− DNA polymerase. We used the assay to determine the kinetic inhibition mechanism of ATA and somewhat unexpectedly found that it was not behaving as a DNA mimic but was noncompetitive against both dNTPs and DNA. The robustness of the modified assay strongly suggests that it is well suited to the screening of much larger small-molecule libraries in an effort to identify inhibitors of the P. falciparum apicoplast DNA polymerase that can serve as lead compounds in antimalarial drug-discovery efforts.

References


CHAPTER 5
CONCLUSIONS

Conclusions

We have only just begun to understand the *P. falciparum* apicoplast DNA polymerase, apPOL, and its role in replication and repair of the apicoplast genome. In this dissertation, I presented the crystal structure of apPOL, the first atypical A-family DNA polymerase to be structurally characterized, and provided insight into a poorly characterized subfamily. Before we can develop drugs to inhibit the apicoplast polymerase in a highly potent and specific manner, a deeper understanding of the structure and mechanisms of apPOL is required. This crystal structure provides a starting point for structure-aided anti-malarial drug design, and will assist in the biochemical characterization of apPOL.

As discussed in chapter one, the apicoplast DNA polymerase is expressed as a polyprotein, Prex, containing primase, helicase, and polymerase subdomains. It has been clearly shown that Prex is targeted to the apicoplast where it undergoes post translational modification which may separate these three subdomains, though the degree of separation is unclear. (1–3). Despite having been identified in 2005, it is still uncertain where Prex is cleaved and what proteins are involved in its processing.

Based on sequence alignment of *Plasmodium* Prex genes, it is likely that our apPOL construct is representative of functional replicative polymerase within the apicoplast (4). apPOL surpasses other Prex polymerase constructs in both activity and recombinant expression levels (1,5). While the protease responsible for cleaving the polymerase domain of Prex from the primase and helicase domains is still unknown, apPOL folds into a well ordered protein. The crystal structure shows ordered density for all N-terminal residues in our construct, and that the
extreme N-terminus residues form an ordered region involved in many stabilizing interactions. While we have not investigated the effect what additional upstream residues could have on polymerase activity, removal of the N-terminal region has a negative impact. As described at the end of chapter three, deletion of the N-terminal region results in a 20-fold decrease in polymerase expression and affects exonuclease activity, suggesting that this region is important for proper protein function.

The structure of apPOL provides substantial information regarding the structural significances of the signature motifs that distinguish atypical A-family members (6). Two of the three signature motifs manifest as structural differences when apPOL is compared to typical A-family polymerases from *E. coli* Pol I (Klenow), *B. stearothermophilus*, and *T. aquaticus* (Taq). One motif involves the deletion of a loop at the base of the thumb subdomain. This loop has been shown to be involved in the binding of the double stranded primer DNA and polymerase fidelity (7). It is possible that this absence of this loop in apPOL accounts for some of the differences in its fidelity as compared to Klenow. Further investigation of this motif could provide insight into the fine-tuning of DNA binding and polymerase fidelity.

The second motif to show significant structural differences is a six amino acid addition which extends a helix at the tip of the finger subdomain. The most notable feature of this motif is three tyrosine residues at the end of the helix, pointed inwards towards the polymerase active site. Mutation of these tyrosines to alanines affects apPOL’s DNA binding affinity, nucleotide binding affinity, and rate of polymerization. Due to its close proximity to the single stranded template DNA binding site, it is possible that this motif is involved in positioning the template DNA. Removal of the tyrosines may cause a conformational change in the DNA which prevents
proper Watson-Crick base pairing with the incoming nucleotide, or positioning the 3′-OH in a suboptimal location for chemistry.

A key structural difference between apPOL and typical A-family members that was not identified through previous sequence alignment is the deletion of eleven residues at the base of the finger subdomain. These residues comprise the end of a large helix and a loop which contain many residues involved in binding template DNA (8,9). In typical A-family members, this helix and loop form a groove at the base of the finger subdomain where the template DNA is threaded through (10). Removal of this region provides a much more open interface for DNA to bind on apPOL. Even without this loop, apPOL is still able to bind DNA with an affinity comparable to other replicative DNA polymerases. It is possible that the tyrosine signature motif mentioned above compensates for this missing region, providing apPOL with the necessary interactions to position DNA into the active site.

Information about the structure and biochemistry of these atypical regions of apPOL will aid in the development of potent inhibitors which are specific to apPOL. To identify compounds that inhibit apPOL, we developed a fluorescence-based assay that can be used in a high-throughput manner to screen large libraries of small molecules. The assay is simple and statistically robust, making it well suited for high-throughput screening. We have shown that the assay behaves as expected in solution, and can be used in a continuous and stop-time manner to determine polymerase activity. The assay has been effectively used to perform a pilot screen, and can be used to determine the mechanism of inhibition through steady-state kinetics methods.

Future work will involve further exploration of the biochemical relevance of the apPOL atypical regions. We will express a polymerase construct which incorporates the palm deletion from Klenow (described in Chapter three), and test if polymerase fidelity has shifted in favor of a
more Klenow-like profile. We will also test the impact of this signature motif on DNA binding, nucleotide affinity, and exonuclease activity. A construct that adds the missing loop from the base of the finger subdomain (described in Chapter three) will also be created. If apPOL binds DNA in the same region as Klenow, adding in the missing eleven residues at the tip of the Q-helix should significantly impact DNA binding in apPOL. Point mutations within the N-terminal region will help map the specific residues and interactions which are necessary for proper folding and exonuclease activity. Understanding the role these regions play in regulating polymerase activity will provide further knowledge regarding what separates atypical from typical A-family polymerases, and how these areas can be exploited for drug targeting. We also hope to obtain a crystal structure of apPOL with DNA bound into the active site. This could provide information about how the polymerase interacts with the DNA substrate, and if those interactions vary from typical A-family members.

Future work will also include screening large libraries of small molecules to identify compounds that inhibit apPOL. Those compounds will then be counter screened for their specificity to apPOL, and a mechanism of inhibition will be determined. Cocrystallization with apPOL screens and testing the effects of point mutations that disrupt binding pockets on polymerase activity will be performed in order to map out the inhibitor’s binding site. This information will then be used by a collaborator to develop more potent and specific variations of the compound. Subsequent compounds will then be tested for effectiveness and kill rates within \textit{P. falciparum} in blood culture.

Overall, this work has and will continue to lead toward better understanding of some of the unusual features of the apicoplast DNA polymerase at a basic scientific level. It will also
provide the means for identifying and optimizing small molecules that inhibit the function of apPOL, which in turn will result in the development of potent antimalarial compounds.

References


