BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

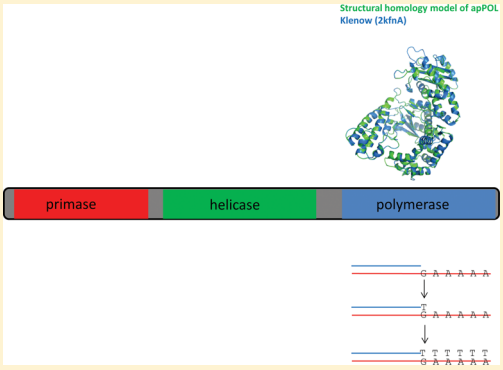
Fidelity, Mismatch Extension, and Proofreading Activity of the *Plasmodium falciparum* Apicoplast DNA Polymerase

Bentley M. Wingert,[†,‡] Eric E. Parrott,[†,§] and Scott W. Nelson*[,†]

[†]Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, Iowa 50011, United States
[‡]Biochemistry Undergraduate Program at Iowa State University, Ames, Iowa 50011, United States
[§]Biochemistry B.S./M.S. Program at Iowa State University, Ames, Iowa 50011, United States

Supporting Information

ABSTRACT: *Plasmodium falciparum*, a parasitic organism and one of the causative agents of malaria, contains an unusual organelle called the apicoplast. The apicoplast is a nonphotosynthetic plastid responsible for supplying the parasite with isoprenoid units and is therefore indispensable. Like mitochondria and the chloroplast, the apicoplast contains its own genome and harbors the enzymes responsible for its replication. In this report, we determine the relative probabilities of nucleotide misincorporation by the apicoplast polymerase (apPOL), examine the kinetics and sequence dependence of mismatch extension, and determine the rates of mismatch removal by the 3′ to 5′ proofreading activity of the DNA polymerase. While the intrinsic polymerase fidelity varies by >50-fold for the 12 possible nucleotide misincorporations, the most dominant selection step for overall polymerase fidelity is conducted at the level of mismatch extension, which varies by >350-fold. The efficiency of mismatch extension depends on both the nature of the DNA mismatch and the templating base. The proofreading activity of the 12 possible mismatches varies <3-fold. The data for these three determinants of polymerase-induced mutations indicate that the overall mutation frequency of apPOL is highly dependent on both the intrinsic fidelity of the polymerase and the identity of the template surrounding the potential mismatch.



Malaria kills nearly 800000 people each year, with the majority of those deaths occurring in children under 5 years of age.[1] More than 40% of the world's population live in areas where malaria is a serious health risk, and there are approximately 225 million new diagnosed cases each year.[1] Resistance to commonly used malaria drugs is spreading, such that chloroquine and sulfadoxin-pyrimethamine are largely ineffective in parts of Brazil, central Africa, India, and southeast Asia.[2] It is imperative that new drugs for the treatment of malaria be developed, preferably those that target novel aspects of the parasite's biology for which resistance has not already developed.[3]

The human parasite *Plasmodium falciparum* is a member of the phylum Apicomplexa and the most common causative agent of malaria.[4] Apicomplexa also contains the causative agents of several other economically important animal diseases such as toxoplasmosis (human and feline), babesiosis (cattle and human), and coccidiosis (poultry).[5] Nearly all members of Apicomplexa contain an unusual organelle called the apicoplast. The apicoplast is evolutionarily related to the chloroplast and is thought to have arisen through a symbiotic event with red algae.[4] The function of the apicoplast has been enigmatic until recently. The apicoplast participates in several metabolic pathways, including the biosynthesis of fatty acids, heme, iron−sulfur clusters, and isoprenoids.[6] The parasite is completely dependent on the apicoplast for the synthesis of isoprenoids and defects in apicoplast metabolism, and its failure to replicate and divide leads to the death of the organism.[7] Because of the essential nature of the apicoplast and its uniqueness among eukaryotes, the apicoplast has been identified as a highly promising drug target.[8] It has been estimated that approximately 550 proteins reside within the apicoplast, with a significant number of those being dedicated to translation (8%), transcription (1%), and DNA replication (3%).[9] The proteins that conduct these fundamental processes are absolutely required for apicoplast function and are therefore promising drug targets. Indeed, several antibiotics with confirmed antimalarial activity target the following apicoplast enzymes: the 70S ribosome (clindamycin and tetracycline), isoleucine tRNA synthetase (mupirocin), and DNA gyrase (ciprofloxacin).[10]

The *P. falciparum* apicoplast genome is replicated by an A-family DNA polymerase (apPOL) that is clearly of prokaryotic origin. A recent bioinformatic analysis has shown that there are four lineages that are related to, but separate from, the prototypical A-family DNA polymerases.[11] These four lineages

Received: June 5, 2013
Revised: September 25, 2013
Published: October 22, 2013

consist of DNA polymerases that belong to thermophilic viruses, Aquificaceae/Hydrogenothermaceae, Apicomplexa (which includes *Plasmodium*), and various unrelated bacteria.[11] While many A-family DNA polymerases from the "typical" lineage have been extensively characterized at the biochemical and structural level (e.g., *Escherichia coli* Pol I, *Taq* polymerase, T7 phage DNA polymerase, and mitochondrial DNA polymerase γ), a thorough characterization of polymerases from the other four distinct A-family lineages is lacking. In addition to being a member of one of the atypical A-family lineages, apPOL is an attractive drug target because of its role in apicoplast genome replication. Although humans contain three A-family polymerases (pol θ, pol ν, and pol γ), the level of homology between apPOL and these polymerases is very low (<24%). On the other hand, the level of homology between the apPOLs from the two primary causative agents of human malaria is quite high (*P. falciparum* and *Plasmodium vivax* are 84% identical), suggesting that drugs targeted against the *P. falciparum* apPOL would be effective in treating malaria caused by *P. vivax*, as well. The *P. falciparum* apicoplast DNA polymerase is expressed as part of a polyprotein consisting of a DNA primase, a helicase, and a polymerase.[12] Following import into the apicoplast, a protease(s) of unknown identity separates the three proteins.[13]

Replication-induced mutations are governed by a combination of the fidelity of the DNA polymerase (i.e., the probability of inserting an incorrect nucleotide opposite the templating base), its ability to extend the mismatch, and its counteracting exonuclease activity (i.e., proofreading).[14,15] The relative contribution of each of these processes is often overlooked, and DNA polymerase fidelity is the focus of most studies.[16−18] The fidelity of a fragment of *P. falciparum* apicoplast DNA polymerase has been recently reported.[18] To obtain a more complete picture of the mutagenic profile of apPOL, we have extended these studies to include mismatch extension and proofreading activity. On the basis of bioinformatic analysis, a different polymerase fragment that contains 38 additional amino acids at the N-terminus was constructed, and our steady-state kinetic analysis indicates that it is significantly more active than the previous version, with a fidelity more reminiscent of that of a replicative DNA polymerase.[18] It was found that the efficiency of mismatch extension is more important to the overall mutation frequency than the intrinsic fidelity of the polymerase. While the proofreading activity of each mismatch is relatively similar, the ability of the polymerase to extend a particular mismatch varies by >350-fold.

MATERIALS AND METHODS

Materials. Oligodeoxynucleotides used for mutagenesis were purchased from either Integrated DNA Technologies or the Iowa State University DNA Facility. DNA sequencing was performed at the Iowa State University DNA Facility. Nickel-agarose was purchased from Sigma-Aldrich Chemical Co. Deoxyribonucleotides were purchased from Invitrogen or Sigma-Aldrich.

Cloning of the *P. falciparum* Apicoplast DNA Polymerase and Creation of the Exonuclease Negative Mutant (apPOL^{exo-}). The open reading frame containing the apicoplast DNA polymerase encodes a polyprotein consisting of a DNA primase, a helicase, and a polymerase.[12] A linker region between the primase and helicase is proteolytically cleaved, and it is thought that the protein is cleaved between the helicase and polymerase, as well.[13] On the basis of protein sequence alignments of POM1 from *Plasmodium*, a likely boundary for the polymerase protein spanning amino acid residues 1389−2016 was identified. This protein sequence was then converted to DNA sequence using optimal *E. coli* codons and synthesized (Genescript). The synthesized gene was subcloned from the puc18 vector into the pet28b expression vector using the NdeI and BamHI restriction sites. The Quickchange method of mutagenesis was employed to produce the exonuclease negative mutant (apPOL^{exo-}). The sequence of the forward primer used for mutagenesis was 5′-gatattaaatatgcggcctg**aa**t**a**tc**caa**accacgggtctggaagtg-3′ (codons for the D1470N and E1472Q mutations in bold). The reverse primer was the reverse complement of the forward.

Protein Expression and Purification. The purification protocols for the wild-type (apPOL) and exonuclease deficient (apPOL^{exo-}) polymerases were identical. Either the pet28-apPOL or pet28-apPOL^{exo-} vector was transformed into *E. coli* BL21(DE3) cells, and a single colony was used to inoculate 100 mL flasks of LB-kanamycin that were shaken for 16 h at 37 °C. A 10 mL portion of starter culture was used to inoculate two 1 L flasks of LB-kanamycin per protein, which were shaken at 225 rpm at 37 °C until the A_{600} reached 0.8. The flasks were then cooled to 18 °C, and expression was induced by the addition of 0.2 mM (final) isopropyl 1-thio-β-D-galactopyranoside. After 16 h, the cells were collected by centrifugation at 4000*g* for 20 min, and pellets were frozen at −20 °C.

Target protein purification relied on a hexahistidine tag provided by the pet28 vector. Cell pellets containing expressed apPOL or apPOL^{exo-} (2 L) were resuspended in 100 mL of loading buffer [20 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole, and 20% glycerol (pH 8.0, 4 °C)] and stored frozen at −20 °C. After the pellets had thawed, lysis was accomplished by passage through an EmulsiFlex-C5 column (Avestin, Inc.) at ∼16K psi. The lysate was clarified by centrifugation at ∼32500*g* and the supernatant loaded onto ∼3 mL of Ni-agarose resin. The column was washed with 100 mL of loading buffer, followed by 100 mL of high-salt buffer [5 mM imidazole, 1 M NaCl, 20 mM Tris-HCl, and 20% glycerol (pH 8.0, 4 °C)] and then 30 mL (10 column volumes) of Ni wash buffer [20 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, and 20% glycerol (pH 8.0, 4 °C)]. The protein was eluted with elution buffer [20 mM Tris-HCl, 200 mM NaCl, 150 mM imidazole, and 20% glycerol (pH 8.0, 4 °C)]. The fractions containing protein were pooled and loaded onto a 320 mL HiLoad 26/60 Superdex 200 prep grade column equilibrated with 20 mM Tris-HCl, 400 mM NaCl, and 20% glycerol (pH 8.0, 4 °C). The column was washed with 400 mL of 20 mM Tris-HCl, 400 mM NaCl, and 20% glycerol (pH 8.0, 4 °C). The fractions containing protein were pooled, and the concentration was determined spectrophotometrically using an extinction coefficient (ε_{280} = 56750 M^{-1} cm^{-1}) calculated from the deduced protein composition.

Steady-State Polymerase Extension Kinetics. DNA templates were made by annealing a 20-nucleotide primer (P1 in Table S1 of the Supporting Information) that had been separately labeled with ^{32}P using T4-polynucleotide kinase to a 26-nucleotide fragment (T2, T5, T10, and T14). Reactions were performed at 25 °C in 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitiol, and 0.1 mg/mL bovine serum albumin. Polymerase was diluted in 20 mM Tris-HCl, 200 mM NaCl, and 0.1 mg/mL BSA (standard buffer) prior to being added to the reaction mixture. For correct nucleotide incorporations, steady-state assays were

Table 1. Steady-State Kinetic Parameters of *P. falciparum* Apicoplast DNA Polymerase[a] for Correct and Incorrect Nucleotide Incorporations

nucleotide	template	k_{cat} (s^{-1})	error[b]	K_m (μM)	error[b]	k_{cat}/K_m (s^{-1} μM^{-1})	fidelity[c]
dTTP	dAMP	0.33	0.01	5.1	0.7	0.065	1
dATP	dAMP	0.0026	0.0001	148	31	1.76×10^{-5}	2.71×10^{-4}
dGTP	dAMP	0.0012	0.0005	262	66	4.58×10^{-6}	7.08×10^{-5}
dCTP	dAMP	0.0013	0.0001	345	134	3.77×10^{-6}	5.82×10^{-5}
dATP	dTMP	0.34	0.02	1.7	0.4	0.2	1
dGTP	dTMP	0.0034	0.0001	73	11	4.66×10^{-5}	2.33×10^{-4}
dCTP	dTMP	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]
dTTP	dTMP	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]
dCTP	dGMP	0.24	0.009	7.3	1.7	0.033	1
dTTP	dGMP	0.0018	0.0005	142	16	1.27×10^{-5}	3.86×10^{-4}
dGTP	dGMP	0.0006	0.00005	562	127	1.07×10^{-6}	3.25×10^{-5}
dATP	dGMP	0.0033	0.0002	903	136	3.65×10^{-6}	1.11×10^{-4}
dGTP	dCMP	0.33	0.01	0.7	0.2	0.47	1
dATP	dCMP	0.0061	0.0002	241	34	2.53×10^{-5}	5.37×10^{-5}
dTTP	dCMP	0.0008	0.00007	159	42	5.03×10^{-6}	1.07×10^{-5}
dCTP	dCMP	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]

[a]The polymerase used in these experiments lacked 3′ to 5′ proofreading activity. [b]Standard error of the data fit to the Michaelis–Menten equation. [c]Fidelity calculated as $(k_{cat}/K_m)_{incorrect}/(k_{cat}/K_m)_{correct}$. [d]Not detectable.

performed with 2 μM DNA substrate, 8 nM apPOL^{exo-}, and varying amounts of the correct dNTP. For the misincorporation reactions, the apPOL^{exo-} concentration was increased to either 80 or 200 nM, depending on the observed rates in preliminary assays. The enzyme concentration was increased so that approximately ~10–20% of the substrate was converted to product at the longest time point used. The correct nucleotide following the misincorporation position (to allow for runoff synthesis) was held at a concentration of 20 μM. The reactions were initiated at time zero by mixing equal volumes of the DNA polymerase and the DNA substrate and/or nucleotides. Reactions were quenched after various times with equivalent volumes of a mixture containing 0.1 M EDTA and 80% (v/v) formamide. Reaction products were resolved on a 20% denaturing polyacrylamide gel electrophoresis gel containing 7.5 M urea in 1× Tris-Borate-EDTA (TBE) buffer. Gels were run for 3–3.5 h at a constant power of 60 W, visualized using a Typhoon Phosphorimager, and analyzed using ImageJ (National Institutes of Health, Bethesda, MD). The initial rate velocities were averaged (each assay was performed two to five times) and fit to the Michaelis–Menten equation. The errors given for each parameter (k_{cat} and K_m) are standard errors of the fit.

Steady-State Mismatch Extension Kinetics. The following 48 different substrates were used to investigate the mismatch extension kinetics of apPOL^{exo-}: P2 with T1–T4 and T9–T16, P3 with T5–T16, P4 with T1–T12, and P5 with T1–T8 and T13–T16 (Table S1 of the Supporting Information). The reactions were performed in standard assay buffer at 25 °C; mixtures contained 2 μM DNA, 80 nM apPOL^{exo-}, and 2 mM dNTP, and reactions were quenched at times points ranging from 1.25 to 60 min, depending on the reaction. Analysis of the gel products was the same as that described above, and the rates are averages of three independent reactions.

Steady-State Exonuclease Kinetics. The exonuclease experiments were performed using 8 nM apPOL and 2 μM DNA, and reactions were quenched at time points of 1 and 5 min. Substrates used for these experiments were P2–P5 annealed to each of T2, T6, T10, and T14. Analysis of the gel products was the same as that described for the mismatch extension kinetics, and the rates are averages of two independent reactions.

RESULTS AND DISCUSSION

***P. falciparum* apPOL^{exo-} Misincorporation Kinetics.** The mutagenesis frequency of a DNA polymerase depends on several factors. The first is the intrinsic fidelity of the polymerase active site (i.e., the probability of incorrect nucleotide incorporation), which is a commonly examined property of DNA polymerases, and several different methods have been used to determine it, including pre-steady-state kinetics, steady-state kinetics, and direct competition assays.[19] Apparently, all three methods yield similar results, although there are theoretical concerns about using single-nucleotide incorporation steady-state kinetics when the rate of dissociation of the polymerase from the DNA product is very rate-limiting.[20] Slow product dissociation may occur when a highly processive polymerase is stalled at a correctly matched $n + 1$ DNA product and cannot continue polymerizing because of the omission of the next correct nucleotide.[21] To weaken the effect of this potential concern, the DNA templates were designed to contain a five-nucleotide homopolymeric extension, and a low concentration (20 μM) of the correct nucleotide for incorporation across the repeat is included in the reaction mixtures. Control reactions indicate that at this low concentration the nucleotide does not incorporate across from the position of interest, yet the nucleotide tail is very rapidly filled in. A substrate of this type allows a processive polymerase to quickly "run off" the DNA substrate via five rounds of very rapid nucleotide incorporation.

The steady-state kinetic parameters of the exonuclease deficient apPOL^{exo-} for the four correct nucleotide incorporations and 9 of the 12 possible mismatches are given in Table 1. The values determined here for apPOL^{exo-} do not correspond well to a previous investigation using a shorter version of the *P. falciparum* apicoplast DNA polymerase.[18] The most striking difference between the two reports is that our k_{cat} values are significantly greater (~750-fold) than those of Kennedy et al. [average $k_{cat(correct)}$ of 0.3 s^{-1} for our construct vs

dx.doi.org/10.1021/bi400708m | *Biochemistry* 2013, 52, 7723–7730

0.00044 s^{-1} for theirs]. The reason for this very large discrepancy is not clear. The polymerase construct used in our studies contains an extra 38 residues at the N-terminus, which are partially conserved only within the genus *Plasmodium* and are not found in other DNA polymerases (Figure S1 of the Supporting Information). The absence of these residues may be responsible for the greatly reduced activity. It is also possible that the reduced k_{cat} values reported by Kennedy et al. are due to the unusual conditions used for their steady-state polymerase assays. The DNA substrate concentration used was low (10 nM) and possibly not saturating. Additionally, the polymerase concentration that was used (20 nM) exceeded that of the substrate concentration by 2-fold, indicating that the reported values should not be considered true steady-state kinetic parameters.[22] These types of conditions (low DNA concentration and excess protein) also have the potential to obscure potential problems caused by dNTP contamination. For example, kinetic time courses revealed that our original source of dTTP was contaminated with dATP at a level of approximately 0.02%. Contamination can be identified by nonlinear initial velocity kinetic time courses that display a faster phase with an amplitude greater than the enzyme concentration. To observe this phase, the DNA concentration must be greater than the concentration of the contaminating nucleotide. If not, then the resulting kinetic data may produce k_{cat} values similar to those of correct incorporation and elevated K_m values that are linked to the percent contamination and the k_{cat}/K_m for the intended dNTP. In our unpublished assays using polymerase concentrations in excess of DNA substrate concentrations (i.e., pre-steady-state kinetic assays), the substrate is completely converted to product in <2 s when the correct dNTP is used (an average k_{pol} of ~20 s^{-1}) (E. E. Parrott, B. M. Wingert, and S. W. Nelson, unpublished results). The relatively slow rates previously reported[18] (5 min for ~10% substrate conversion with an excess of polymerase) are consistent with a polymerase defect, or the polymerase has a K_d-DNA that is significantly >10 nM.

In addition to significantly greater k_{cat} values for correct nucleotide incorporation, a much wider variation in k_{cat} values between correct and incorrect incorporations was observed (Table 1). In the previous study, the k_{cat} values ranged from a 13.7-fold decrease to a 1.5-fold increase for incorrect versus correct incorporation.[18] The decreases in k_{cat} values reported here for incorrect incorporation range from 412- to 54-fold. This magnitude of the decrease is within a range similar to those of most other well-characterized high-fidelity DNA polymerases.[19,23] Again, the differences between the previously published data and those reported here may be the result of using different polymerase constructs or different methods of kinetic analysis.

There is also a lack of correspondence between the K_m-dNTP data reported here and the values reported by Kennedy et al., although not as great as with the k_{cat} values.[18] As expected, we find the purine nucleotides to have a smaller K_m than the pyrimidine nucleotides (1.7 and 0.7 μM for dATP and dGTP, respectively, vs 5.1 and 7.3 μM for dTTP and dCTP, respectively). There are no clear trends in the misincorporation data, other than the fact that pyrimidine-pyrimidine misincorporations tend to be strongly disfavored.

The fidelity of a polymerase can be determined by the $(k_{cat}/K_m)_{incorrect}/(k_{cat}/K_m)_{correct}$ ratio.[24,25] On the basis of this formula, the fidelity of apPOL^{exo-} ranges from 3.86×10^{-4} for the most likely misincorporation (dTTP:dGMP) to $<1.07 \times 10^{-5}$ for the least likely misincorporations (dCTP:dCMP, dCTP:dTMP, and dTTP:dTMP), where the first position represents the incoming nucleotide and the second is the template (i.e., primer:template) (Table 1). The data indicate that our limit of detection, while still maintaining true steady-state conditions ([DNA] ≫ [apPOL^{exo-}]), is around a k_{cat} value of 0.0006 s^{-1}. The fidelity values we have determined are approximately 5-fold smaller than previous measurements, which is likely a reflection of the larger differences in k_{cat} between correct and incorrect incorporations that is observed here. However, the correspondence between our relative ranking of misincorporations based on fidelity measurements and those measured previously is, for the most part, quite good (Table 1 and ref 18). Two of the three most likely misincorporations are identical between this report and the earlier report (dTTP:dGMP and dGTP:dTMP), and two of the three least likely misincorporations are the same, as well (dCTP:dCMP and dTTP:dTMP). The only significant differences between this report and the previous report in terms of the fidelity calculation are the dATP:dAMP (second and seventh most likely, respectively) and dCTP:dTMP (10th and fifth most likely, respectively) misincorporations. The source of these discrepancies is unclear, but it appears that our polymerase construct behaves more like Klenow polymerase, where dATP:dAMP is also the second most likely mutation and dCTP:dTMP is the 11th most likely using steady-state kinetic methods.

DNA Mismatch Removal and Extension. A qualitative examination of the misincorporation gels indicated that some mismatches were easily extended to the end of the DNA substrate, whereas other mismatches appeared to extend only to the $n + 1$ and $n + 2$ positions. This observation prompted an investigation of the second factor that governs mutation frequency, which is the ratio between the rate of mismatch removal (i.e., proofreading) and the ability of the polymerase to extend the mismatch.[15] The mismatch removal rate is relatively constant for all 12 possible DNA mismatches [<3-fold variation (Table 2)] but is 2−14-fold faster than removal of the four correctly matched base pairs. In contrast to mismatch removal,

Table 2. Steady-State Exonuclease Rate of *P. falciparum* Apicoplast DNA Polymerase[a]

substrate	apparent k_{cat} (s^{-1})	error[b] (s^{-1})
A:T	0.07	0.01
A:C	0.36	0.11
A:G	0.15	0.03
A:A	0.14	0.02
T:A	0.06	0.03
T:C	0.37	0.07
T:T	0.29	0.03
T:G	0.30	0.05
G:C	0.06	0.03
G:T	0.32	0.03
G:G	0.37	0.06
G:A	0.27	0.05
C:G	0.03	0.01
C:T	0.28	0.10
C:C	0.41	0.029
C:A	0.26	0.08

[a]Wild-type polymerase was used. [b]Standard deviation of at least three independent replicates.

dx.doi.org/10.1021/bi400708m | *Biochemistry* 2013, 52, 7723−7730

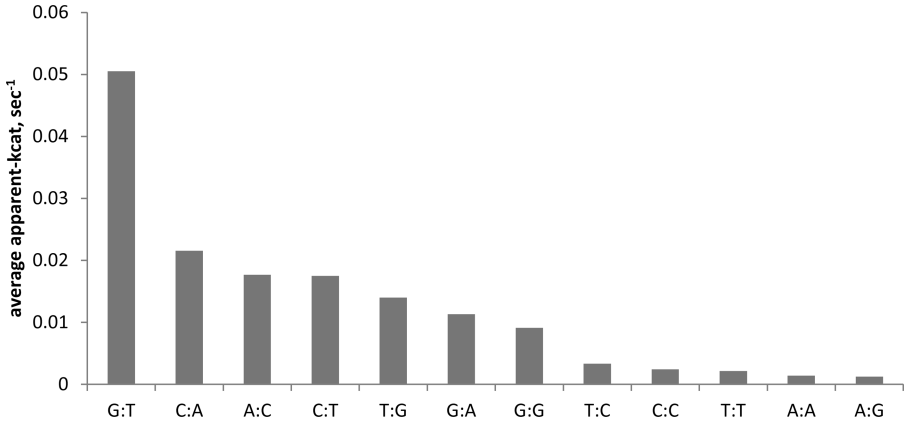


Figure 1. Average apparent k_{cat} for mismatch extension. The first letter of the DNA substrate represents the base at the 3′ end of the primer strand, and the letter following the colon represents the template base that is mismatched with the primer (primer:template). Each value is an average of the apparent k_{cat} values for all four templating nucleotides (e.g., G:TT, G:TA, G:TC, and G:TG).

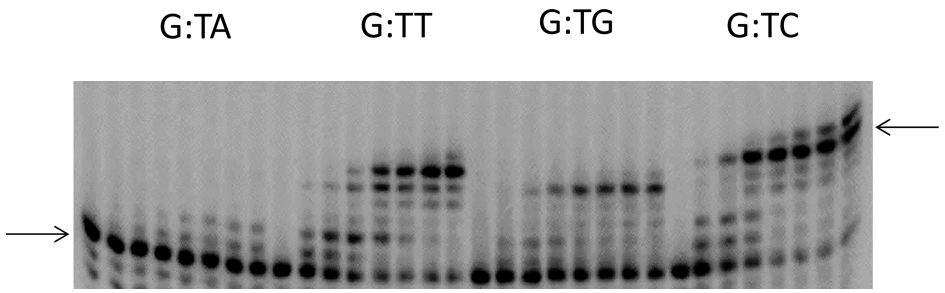


Figure 2. Polyacryamide gel analysis of the G:T mismatch extension reaction. The fully extended and unextended primers are indicated by the arrows on the right and left sides of the gel, respectively. In reactions performed with dATP and dGTP, the addition of a single 3′ overhanging nucleotide is often observed. The time points for each reaction are 0, 1.25, 2.5, 5, 10.75, 15, 20, and 30 min.

the rate of mismatch extension varies much more widely (>350-fold), depending on both the nature of the mismatch and the templating base that follows the mismatch (Figures 1−3). Because the downstream templating base was initially observed to strongly affect the mismatch extension frequency (Figure 2), the 12 mismatches with all four possible templating bases were examined (i.e., 48 total DNA substrates). Because of the large number of DNA substrates, a determination of K_m and k_{cat} for each substrate was beyond the scope of this investigation, and therefore, only the apparent k_{cat} for each mismatch extension reaction was determined using a nucleotide concentration that was assumed to be saturating or near-saturating (2.0 mM dNTP).

There are several general trends that are observed in the mismatch extension data (Figures 1 and 3). The most efficiently extended mismatches tend to be purine-pyrimidine or pyrimidine-purine mismatches (G:T, C:A, A:C, and C:T), and the least efficiently extended mismatches are purine-purine and pyrimidine-pyrimidine mismatches (A:A, T:T, A:G, and T:C). Although useful, the average apparent k_{cat} values for mismatch extension given in Figure 1 obscure the significant dependency of mismatch extension efficiency on the templating base that follows the mismatch. For example, the C:A mismatch is extended with a relatively high rate when T or C is the templating base (C:AT and C:AC) but is much less efficient when A is the templating base (C:AA). In the case of the A:C mismatch, only the DNA substrate with a templating T supports a high rate of mismatch extension (A:CT). This pattern could possibly be the result of primer slippage (i.e., the mismatched template nucleotide looping out and allowing the primer to anneal to the downstream base); however, a close analysis of the gel indicated that the product of this reaction was of the expected length, indicating that primer slippage did not occur.

An analysis of the data shown in Figure 3 allows some general conclusions to be drawn from the templating nucleotide data. Mismatches are most efficiently extended when the templating base is a pyrimidine rather than a purine, with average apparent k_{cat} values of 0.02 and 0.005 s^{-1} for a pyrimidine and a purine, respectively. The strong sequence dependence that is observed in the mismatch extension reactions suggests that the incoming nucleotide plays a significant role in compensating for alterations in the structure of the polymerase active site because of upstream mismatched DNA. As purines are capable of providing more binding energy than pyrimidines, it is not unexpected that these nucleotides are able to produce a stronger compensatory effect.

On the basis of our misincorporation, exonuclease, and mismatch extension data, we have ranked possible mutations from the most to least likely to occur (Figure 4). These values were obtained by multiplying the fidelity by the ratio of mismatch extension to mismatch removal rate [ranking = fidelity × (extension/removal)]. While these values do not represent a kinetic constant and have no true physical meaning, they are useful for comparisons. The intent of this calculation is to take into account all three selection steps that govern mutagenesis frequency. The higher the value, the more likely it is that a particular mutation will occur. On the basis of these values, the four most likely mutations are the G → A transition mutation when a T is the downstream base and the T → C

dx.doi.org/10.1021/bi400708m | *Biochemistry* 2013, 52, 7723−7730

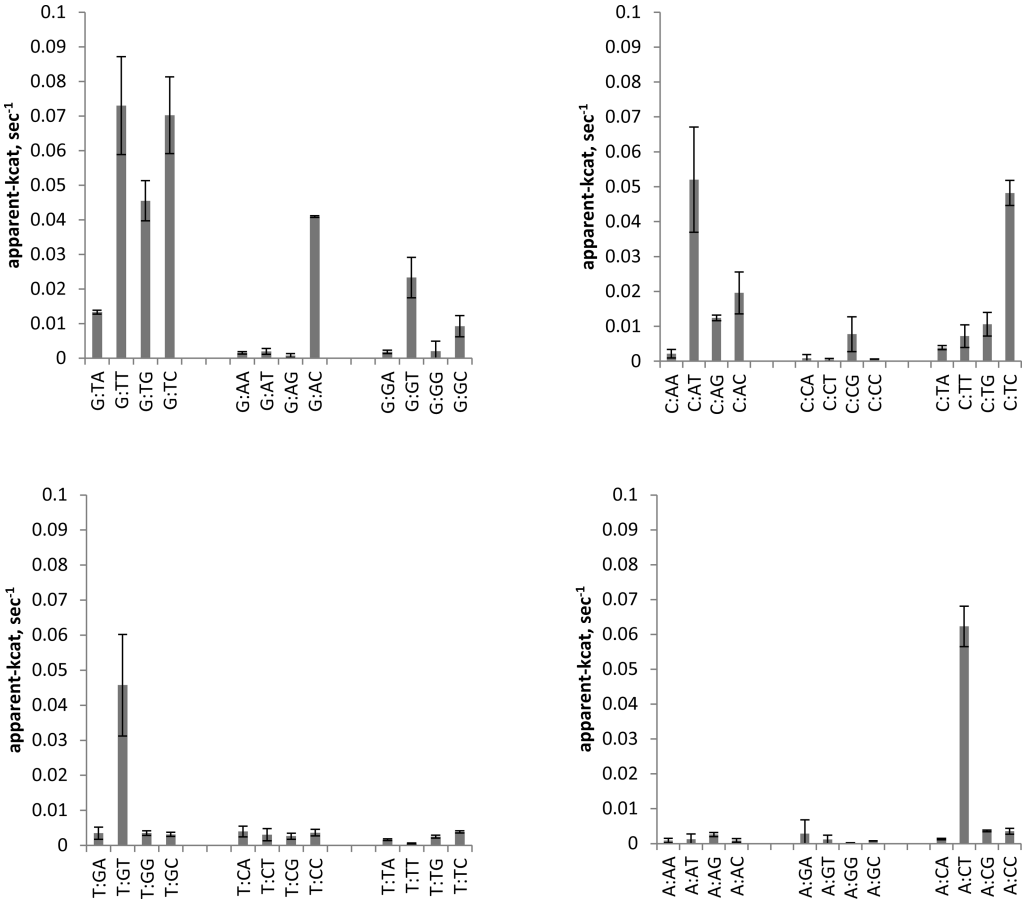


Figure 3. Individual extension rates for the 12 possible mismatches with each templating base. The first letter of the DNA substrate represents the base at the 3′ end of the primer strand; the letter immediately after the colon represents the base that is incorrectly paired with the primer, and the last letter represents the templating base for the correct incoming nucleotide.The values are considered to be apparent k_{cat} values as they were determined at saturating or near-saturating dNTP concentrations. The rates were derived through analysis of polyacrylamide gels such as the one shown in Figure 2. Reactions that showed no detectable product over the entire time course of the reaction were considered to be zero.

transition mutation when T, C, or G is the downstream base. There are only a few transversions that rank in the upper third of the most likely mutations (Figure 4A). Those are the A → C mutation when C is the downstream base, the A → T mutation when G or T is the downstream base, and the G → T mutation when A is the downstream base.

There is a lack of correspondence between the relative rankings for mismatch extension and misincorporation probability (Figure 1 and Table 1). For example, A:A and G:A are ranked second and fifth for misincorporation probability but 10th and ninth for mismatch extension, respectively (using the averaged data from Figure 1). On the other hand, the C:A and A:C mismatches are seventh and eighth for misincorporation probability but second and third for mismatch extension, respectively. The opposing rankings for these mutations weaken the effects of either low fidelity or high mismatch extension rate, reducing the probability that they will occur. The most notable exception to this pattern is G:T, which is ranked near the top in both misincorporation probability and mismatch extension.

Of course, the rank order of the mutations shown in Figure 4 may not entirely reflect the order of polymerase-induced mutation probability in the apicoplast of *P. falciparum*. Very little is known regarding mismatch repair within the apicoplast. Inspection of the *P. falciparum* genome revealed several genes encoding putative DNA repair proteins that appear to contain apicoplast targeting signals.[26] One of these proteins contains a MutS domain and the typical ABC motifs of MutS homologues; therefore, some form of mismatch repair may occur within the apicoplast. Human MutS appears to show no strong preference for any particular mismatch (<2-fold differences),[27] whereas prokaryotic MutS displays large differences in efficiencies.[28] Notably, the mismatch least efficiently repaired (pyrimidine-pyrimidine) is also the least likely to occur.

The fidelity and mutagenic potential are of considerable importance to the biology of *Plasmodium* because its apicoplast apparently contains only a single DNA polymerase. High fidelity is necessary for preserving genetic information and avoiding harmful mutations that may lead to organismal dysfunction, whereas mutagenesis is necessary for generating genetic diversity and the long-term evolution of a species. The fidelity measurements reported here clearly place apPOL in the group of high-fidelity DNA polymerases with error rates ranging from 10^{-4} to $<10^{-6}$. Unlike the mitochondria and apicoplast, which contain only a single DNA polymerase, most organisms harbor several low-fidelity polymerases that are responsible for replicating past DNA lesions (i.e., translesion synthesis polymerases). The high fidelity of apPOL suggests that DNA lesions are not easily bypassed, and it is likely that the apPOL stalls at most or all types of DNA lesions. This suggests the existence of DNA repair pathways like those found

dx.doi.org/10.1021/bi400708m | *Biochemistry* 2013, 52, 7723−7730

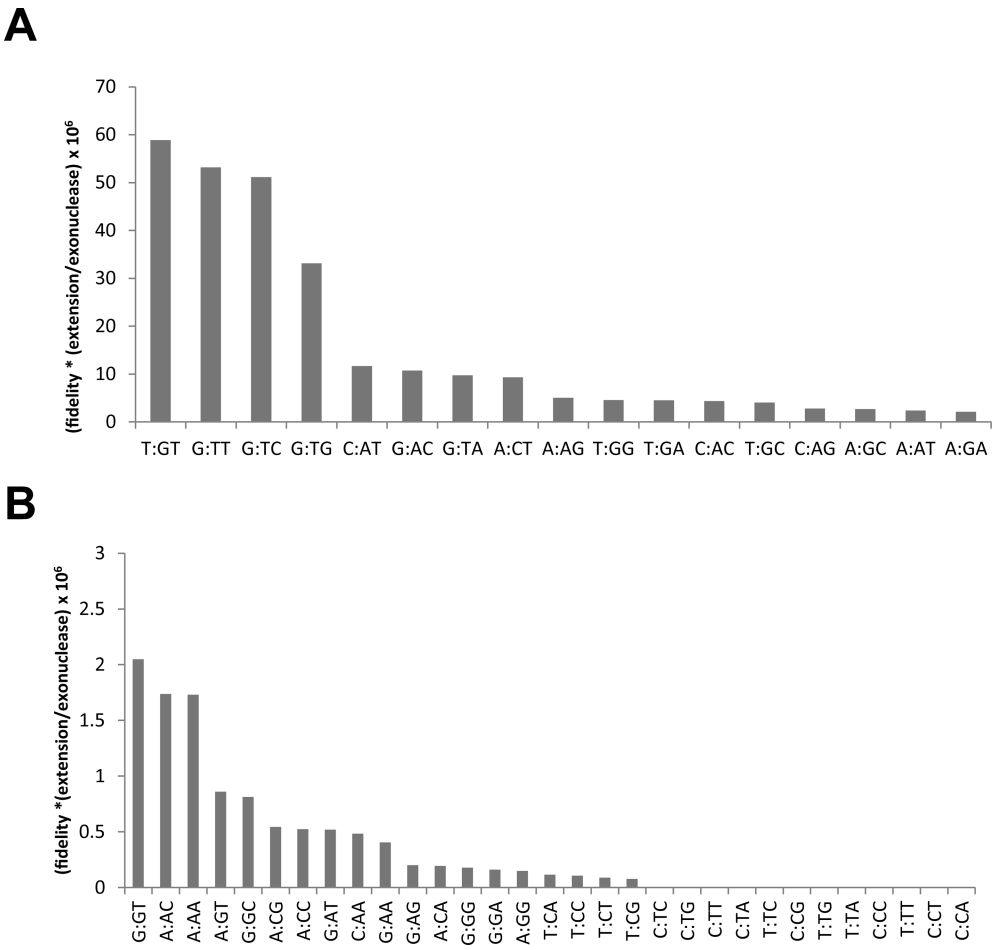


Figure 4. Rank order of polymerase-induced mutation probability based on intrinsic fidelity, mismatch extension activity, and exonuclease rate. As described in the text, the calculated values are for the purpose of comparison only and do not represent a true kinetic constant. (A) The 17 DNA substrates with the highest calculated values. (B) The 31 DNA substrates with the lowest calculated values. Note the 100-fold difference in the *y*-axis scale between panels A and B.

in the mitochondria (e.g., base excision repair).[29] As mentioned above, there are several proteins containing predicted apicoplast targeting signal sequences that are annotated as DNA repair proteins, and it appears that a base excision repair system may be present.[26] The replication and repair of the *P. falciparum* apicoplast genome is poorly understood, and with the ever-increasing drug resistance of *P. falciparum*, combined with the potential of the apicoplast as a drug target, determining the fundamental mechanisms of apicoplast genome replication and repair is an important step toward finding new treatments for malaria that target novel aspects of the biology of the parasites.

■ ASSOCIATED CONTENT

Supporting Information

A table listing the oligonucleotides that were used in this study (Table S1) and a figure showing a protein sequence alignment of selected apicoplast DNA polymerases (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: swn@iastate.edu. Phone: (515) 294-3434. Fax: (515) 294-0453.

Funding

This work was financially supported by The Roy J. Carver Charitable Trust and Iowa State University.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

apPOL, *P. falciparum* apicoplast DNA polymerase; apPOL^{exo-}, exonuclease deficient *P. falciparum* apicoplast DNA polymerase.

■ REFERENCES

(1) World Malaria Report (2010) World Health Organization, Geneva.

(2) Petersen, I., Eastman, R., and Lanzer, M. (2011) Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Lett. 585*, 1551−1562.

(3) Burrows, J. N., Leroy, D., Lotharius, J., and Waterson, D. (2011) Challenges in antimalarial drug discovery. *Future Med. Chem. 3*, 1401−1412.

(4) Kalanon, M., and McFadden, G. I. (2010) Malaria, *Plasmodium falciparum* and its apicoplast. *Biochem. Soc. Trans. 38*, 775−782.

(5) Escalante, A. A., and Ayala, F. J. (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc. Natl. Acad. Sci. U.S.A. 92*, 5793.

(6) Seeber, F., and Soldati-Favre, D. (2010) Metabolic pathways in the apicoplast of apicomplexa. *Int. Rev. Cell Mol. Biol. 281*, 161−228.

(7) Yeh, E., and Derisi, J. L. (2011) Chemical Rescue of Malaria Parasites Lacking an Apicoplast Defines Organelle Function in Blood-Stage *Plasmodium falciparum*. *PLoS Biol. 9*, e1001138.
(8) Dahl, E. L., and Rosenthal, P. J. (2008) Apicoplast translation, transcription and genome replication: Targets for antimalarial antibiotics. *Trends Parasitol. 24*, 279−284.
(9) Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S., and McFadden, G. I. (2004) Tropical infectious diseases: Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat. Rev. Microbiol. 2*, 203−216.
(10) Pradel, G., and Schlitzer, M. (2010) Antibiotics in malaria therapy and their effect on the parasite apicoplast. *Curr. Mol. Med. 10*, 335−349.
(11) Schoenfeld, T. W., Murugapiran, S. K., Dodsworth, J. A., Floyd, S., Lodes, M., Mead, D. A., and Hedlund, B. P. (2013) Lateral Gene Transfer of Family A DNA Polymerases between Thermophilic Viruses, Aquificae, and Apicomplexa. *Mol. Biol. Evol. 30*, 1653−1664.
(12) Seow, F., Sato, S., Janssen, C. S., Riehle, M. O., Mukhopadhyay, A., Phillips, R. S., Wilson, R. J. M., and Barrett, M. P. (2005) The plastidic DNA replication enzyme complex of *Plasmodium falciparum*. *Mol. Biochem. Parasitol. 141*, 145−153.
(13) Lindner, S. E., Llinás, M., Keck, J. L., and Kappe, S. H. I. (2011) The primase domain of PfPrex is a proteolytically matured, essential enzyme of the apicoplast. *Mol. Biochem. Parasitol. 180*, 69−75.
(14) Echols, H., and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem. 60*, 477−511.
(15) Goodman, M. F., Creighton, S., Bloom, L. B., and Petruska, J. (1993) Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol. 28*, 83−126.
(16) Fiala, K. A., and Suo, Z. (2004) Pre-steady-state kinetic studies of the fidelity of *Sulfolobus solfataricus* P2 DNA polymerase IV. *Biochemistry 43*, 2106−2115.
(17) Dieckman, L. M., Johnson, R. E., Prakash, S., and Washington, M. T. (2010) Pre-steady state kinetic studies of the fidelity of nucleotide incorporation by yeast DNA polymerase *δ*. *Biochemistry 49*, 7344−7350.
(18) Kennedy, S. R., Chen, C. Y., Schmitt, M. W., Bower, C. N., and Loeb, L. A. (2011) The biochemistry and fidelity of synthesis by the apicoplast genome replication DNA polymerase Pfprex from the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol. 410*, 27−38.
(19) Bertram, J. G., Oertell, K., Petruska, J., and Goodman, M. F. (2010) DNA polymerase fidelity: Comparing direct competition of right and wrong dNTP substrates with steady state and pre-steady state kinetics. *Biochemistry 49*, 20−28.
(20) Johnson, K. A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim. Biophys. Acta 1804*, 1041−1048.
(21) Hacker, K. J., and Alberts, B. M. (1994) The slow dissociation of the T4 DNA polymerase holoenzyme when stalled by nucleotide omission. An indication of a highly processive enzyme. *J. Biol. Chem. 269*, 24209−24220.
(22) Cook, P. F., and Cleland, W. W. (2007) *Enzyme kinetics and mechanism*, Springer, New York.
(23) Kuchta, R. D., Benkovic, P., and Benkovic, S. J. (1988) Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry 27*, 6716−6725.
(24) Fersht, A. R., Shi, J. P., and Tsui, W. C. (1983) Kinetics of base misinsertion by DNA polymerase I of *Escherichia coli*. *J. Mol. Biol. 165*, 655−667.
(25) Boosalis, M. S., Petruska, J., and Goodman, M. F. (1987) DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *J. Biol. Chem. 262*, 14689−14696.
(26) Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J., Jr., and Whetzel, P. (2003) PlasmoDB: The *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res. 31*, 212−215.
(27) Hays, J. B., Hoffman, P. D., and Wang, H. (2005) Discrimination and versatility in mismatch repair. *DNA Repair 4*, 1463−1474.
(28) Kramer, B., Kramer, W., and Fritz, H. J. (1984) Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell 38*, 879−887.
(29) Kazak, L., Reyes, A., and Holt, I. J. (2012) Minimizing the damage: Repair pathways keep mitochondrial DNA intact. *Nat. Rev. Mol. Cell Biol. 13*, 659−671.