Spring 2012

Determination of the atomic resolution structure of a DNA Polymerase I isolated from Rhodothermus marinus

Natalie S. Omattage
University of Richmond

Determination of the Atomic Resolution Structure of a DNA Polymerase I Isolated from *Rhodothermus marinus*

by

Natalie S. Omattage

Honors Thesis
in
Program in Biochemistry and Molecular Biology
University of Richmond
Richmond, VA

Spring 2012

Advisor: Dr. Eugene Wu
Secondary Reader: Dr. Michelle Hamm

This thesis has been accepted as part of the honors requirements

in the Program in Biochemistry and Molecular Biology.

_____ _____

(advisor signature) *(date)*

_____ _____

(reader signature) *(date)*

Dedication Page

This thesis is dedicated to my parents, Dr. O.P. & Thishya Perera, for igniting in me a passion for the sciences, and for always supporting my academic and scientific endeavors.

Table of Contents

Acknowledgements

Foremost, I would like to express my sincere gratitude to my mentor, Dr. Eugene Wu, for his continuous support of my undergraduate studies and research, and for his patience, motivation, enthusiasm, and immense knowledge.

I would like to thank the Advanced Photon Source, Argonne National Laboratory for granting synchrotron beamtime and the Southeast Region Collaborative Access Team for assistance with data collection. I would also like to thank the University of Richmond School of Arts & Sciences for Undergraduate Summer Fellowship grants in 2010 and 2011. Finally, I would like to thank Emily Kornberg, Miguel Angel Rubio Gomez, Robert Dow, and Lori Spicer for their contributions to this project.

Abstract

DNA polymerase I employs a multistep mechanism for sorting correctly paired nucleotides from mismatches. We aim to characterize reaction intermediates during nucleotide selection to better understand how this class of enzymes achieves high DNA replication fidelity. DNA polymerase I from *R. marinus* contains an unusual and disruptive proline in the mobile O helix near the active site. To characterize this enzyme, the structure of the large (5′-to-3′ exo-deficient) fragment of the *R. marinus* DNA polymerase I (RF) was solved to 2.95 Å (R = 0.234) using multi-wavelength anomalous dispersion. Alignment with homologous *Escherichia coli* Klenow Fragment (KF) DNA polymerase I confirmed that the active sites of each structural domain were conserved. In order to study the polymerase activity in isolation, 3'-to-5' exonuclease activity was eliminated using site-directed mutagenesis of an aspartic acid involved in binding DNA (D497A). Unexpectedly, mutation of an aspartic acid involved in binding a catalytic magnesium ion (D421A in RF) failed to abolish exonuclease activity despite its ability to do so in the KF (D424A). Future structural studies will include crystallization of the polymerase in both its binary and ternary complex to study the conformational changes associated with the process of nucleotide selection as well as examination of the exonuclease domain to characterize its unusual activity in this enzyme.

I. Background

DNA polymerases are responsible for DNA synthesis, the reaction that copies genetic material. These enzymes have been classified into seven distinct families: A, B, C, D, X, Y and RT [1]. Within the Family A of polymerases is DNA polymerase I, the first polymerase to ever be identified, and the most abundant of DNA polymerases within *Escherichia coli*[2]. Within bacteria, DNA polymerase I actively participates in processing Okasaki fragments generated during lagging strand synthesis by removing RNA primers to fill in the gaps, and maintains the high fidelity of replication by performing nucleotide excision repair[1].

DNA Polymerases are characterized by a multi-domain structure consisting of polymerase activity as well as 3'-5' and/or 5'-3' exonuclease activity (Figure 1)[3].

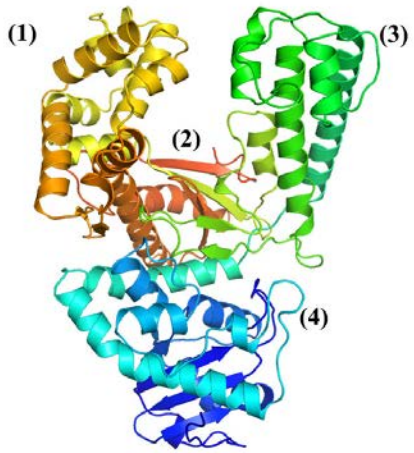

Figure 1. Bacillus Fragment DNA Polymerase I (apo form). Numbered regions are (1) Fingers region, (2) Palm region, and (3) Thumb region of the polymerase domain, and (4) 3'-5' exonuclease domain [PDB ID: 1XWL].

The polymerase domain is responsible for performing DNA synthesis, the 3'-5' exonuclease domain has a proofreading function and performs nucleotide excision repair, and the 5'-3' exonuclease domain removes RNA primers for the filling in of Okasaki fragments[1]. The polymerase domain is often likened to a right hand in that it contains palm, fingers, and thumb motifs that aid in the binding of DNA[4] (Figure 1). The "palm" domain grasps the nascent template DNA[5,6]. The finger motif consists of many helices that serve as key contributors to the catalytic cleft in this domain[7]. Mutation of key residues in one helix, the O-helix, demonstrated the importance of this secondary structure as a deoxyribonucleotide phosphate (dNTP) binding cleft7. For high-fidelity bacterial DNA polymerase I enzymes, the template base is flipped out of the DNA base stack and resides in a "pre-insertion site" prior to encountering a free dNTP. Preliminary models suggested that binding of a nucleotide substrate induces a change in the structure of the enzyme from an open state to a closed state[8]. As seen in multiple experimental systems, the active site O-helix rotates greater than 40° towards the catalytic site upon binding a nucleotide complementary to the template base[9]. Below, Figure 2 depicts DNA polymerase I from *Bacillus Stearothermophilus* (Bacillus Fragment, BF) adopting both the open and closed conformations in the presence of a DNA substrate. The blue helix represents the mobile O-helix in the active site that serves as a gating structure during nucleotide selection. Enclosure of the nucleotide opposite the template base forms an "insertion site" in which the dNTP undergoes in-line nucleophilic attack by the 3'-hydroxyl of the primer, which results in the formation of a phosphodiester bond concomitant with pyrophosphate cleavage from the

nucleotide[5',9,10,11,12,13] (Figure 3). Once the pyrophosphate is released, the fingers domain opens, and polymerase transitions to the next nucleotide for another round of nucleotide incorporation.

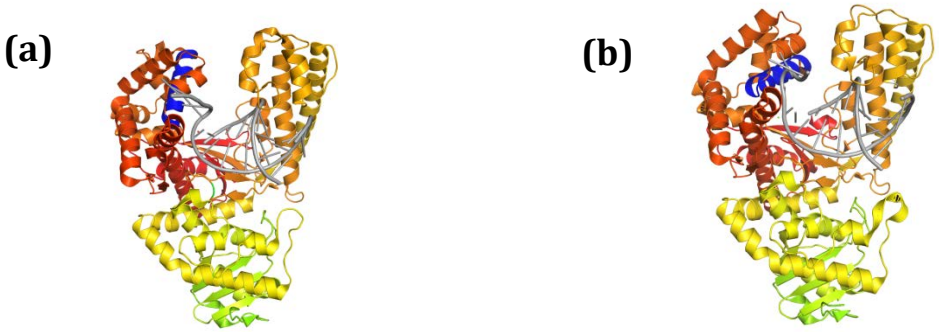

Figure 2. Bacillus Fragment DNA Polymerase I bound to DNA in (a) Open and (b) Closed Conformations. Bacillus Fragment bound to DNA substrate (gray) transitions from (a) open conformation [PDB ID: 1L3S] to (b) closed conformation [PDB ID: 2HVI] through movement of mobile O-helix (blue).

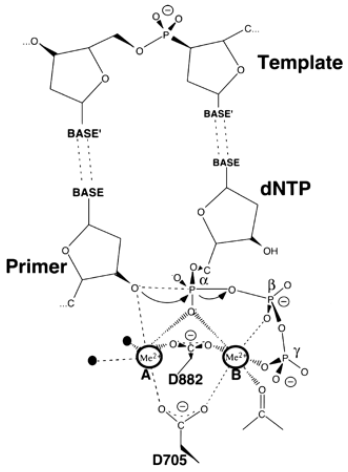

Figure 3. The two-metal ion mechanism of DNA polymerase. The active site features two metal ions that stabilize the resulting pentacoordinated transition state. Metal ion A activates the primer's 3'-OH for attack on the α-phosphate of the dNTP. Metal ion B plays the dual role of stabilizing the negative charge that builds up on the leaving oxygen and chelating the β- and γ-phosphates. Figure and caption reproduced from Figure 3, Reference 13.

While matching complementary nucleotides with templates bases, the mismatch frequency of these enzymes varies from one in five thousand to one in one million nucleotides[14]. DNA polymerases are thought to exhibit extremely high fidelity by utilizing hydrogen bonding, active site water exclusion, geometric selection, and conformational checkpoints for sorting correctly paired base pairs[14]. Although much is known about the open and closed conformational changes of the enzyme in the reaction pathway, the steps in between are not well characterized. Determining these reaction intermediates or "kinetic checkpoints" allows us to understand how these enzymes achieve such a high level of fidelity throughout the process of nucleotide selection. While studying the enzyme's reaction pathway, the rate-limiting step of dNTP insertion was found to be a non-covalent process that precedes chemical bond formation (see references from review)[15]. However, when and how this process occurs has remained unclear. Many models have been proposed describing this non-covalent process. One model posits that the closed conformation is the rate-limiting step, and that one or more conformational changes allow the dNTP to "preview" the template base in the open conformation[15]. A complementary dNTP would stabilize the binding interaction, and would favor the formation of the closed complex, while a noncomplementary dNTP would favor dissocation[15]. Below is a theoretical schematic of the polymerase reaction pathway.

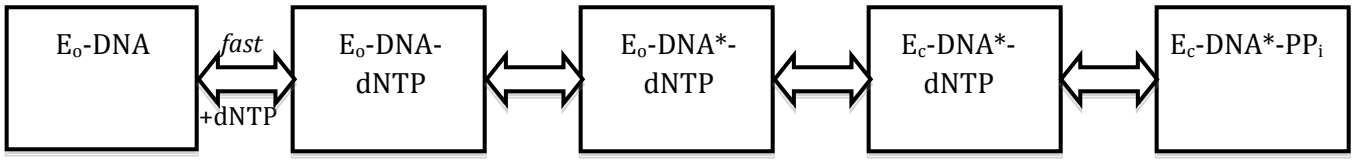

Figure 4. Schematic of Nucleotide Selection. Polymerase (E_o) bound to DNA binds dNTP in fast equilibrium step, undergoes conformational rearrangement (E_o-DNA*-dNTP), and then goes from open to closed conformation (E_c)in a slow, rate-limiting step. Pyrophosphate release (PP_i) from the complex marks the formation of a phosphodiester bond and the end of one round of nucleotide incorporation. Reproduced from Figure 2, Reference 15.

This model has been supported by many experimental studies. For example, a study in DNA polymerase I from *Thermus aquaticus* indicated that there is a step prior to dNTP binding that is dependent on the incoming nucleotide and precedes the closed conformation[16]. Another study in the Klenow Fragment (KF) isolated from *E.coli* demonstrated that mismatched nucleotides were discriminated against in the reaction pathway prior to the fingers closing, suggesting that selection against mismatched nucleotides occurs at an uncharacterized, earlier step[17].

Recently, an X ray crystal structure of a mismatched base pair in the active site of a DNA polymerase I from Bacillus Fragment revealed that the enzyme did not proceed directly to the closed conformation[18]. Instead, the polymerase established an intermediate conformation, called the “ajar” conformation, in which the template base had moved into the insertion site but misaligned the incorrect nucleotide. This ajar conformation was characterized as adopting elements characteristic of both the open and closed conformations; the mobile O, O1, and N helices in the active site all retained structural similarity to the open conformation while the C-terminal helical turn of the O-helix resembled the closed conformation. This portion of the O-helix was different in that a

conserved active site tyrosine was displaced, thereby unraveling the end of the O-helix and moving the O-helix in a position intermediate to that of the open and closed conformations. The study used these structural findings to hypothesize that the ajar conformation is the missing "preview" checkpoint during which the template probes incoming nucleotides for complementarity before proceeding to the closed conformation[18]. The schematic below represents the theorized process of nucleotide binding and nucleotide selection as the enzyme proceeds from the open to the ajar conformation, and finally to the closed conformation if the incoming nucleotide is complementary (Figure 5, reproduced from Reference 18, with permission).

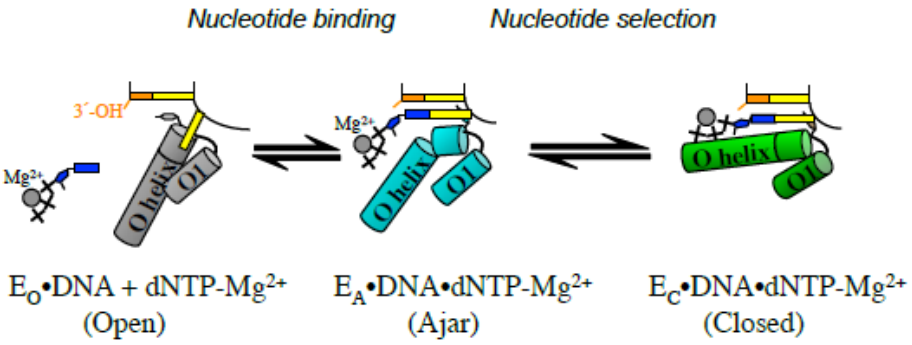

Figure 5. Schematic representation of Open, Ajar, and Closed states, respectively. The ajar conformation resembles the E_o-DNA*-dNTP complex described in Figure 3.

Although experimental data corroborates the existence of this preview checkpoint in BF, it would be useful to observe the ajar conformation in other polymerases of the same class. By catching them adopting this transitional ajar conformation, biochemists would be able to ascertain valuable information about the process of nucleotide selection, and possibly observe other unseen intermediates in the DNA synthesis reaction.

II. Introduction

Rhodothermus marinus is a gram-negative, thermophilic eubacterium found in the alkaline environment of submarine hot springs[19]. It is an obligate aerobe that grows optimally at 65 °C and pH 7.0 in 2% sodium chloride. The *R. marinus* DNA polymerase I has 43% sequence identity with the Pol I sequences of *E. coli* and *B. stearothermophilus*[19]. This polymerase consists of 5'-3' and 3'-5' exonuclease domains as well as a polymerase domain. In addition, it was observed that a highly conserved phenylalanine residue in the O-helix of *E. coli* Pol I has been naturally substituted with a tyrosine at position 757 in the *R.marinus* Pol I. This polymerase and others with this substitution are able to incorporate dideoxynucleotides (ddNTPs) with high efficiency, making this polymerase useful for applications in biotechnology, such as DNA sequencing.

This enzyme is particularly interesting because it contains a proline residue at the beginning of mobile O-helix, a natural substitution from isoleucine and valine seen in KF and BF, respectively (Figure 6)[19].

```
RF    747    EQRRRAKMVNYGIPYGISAWGLAQRLRC
KF    752    EQRRSAKAINFGLIYGMSAFGLARQLNI
BF    700    NMRRQAKAVNFGIVYGISDYGLAQNLNI
             : ** ** :*:*: **:* :***:.*.
```

Figure 6. Polymerase domain active site sequence alignment of RF, KF, and BF.

When the study that characterized the ajar conformation mutated the valine residue in BF to a proline residue (V713P), they found that they were able to stabilize the ajar conformation and examine it in the presence of a complementary nucleotide[18]. Moreover, it was discovered that the proline mutation blocked the formation of a hydrogen bond between the carbonyl residue 709 and the backbone nitrogen of residue 713, and that it sterically hindered the formation of a contiguous O-helix. The V713P mutant was found to be slower than the wild-type BF when incorporating either the correct or incorrect nucleotide. Additionally, the rate of catalysis for V713P was slower in the presence of a mismatch than when in the presence of a correct nucleotide, indicating that both reaction pathways adopt the ajar conformation before transitioning to either the open state to release the mismatched dNTP or to the closed state to continue with nucleotide incorporation. Based on the information garnered from the BF study, and due to the fact that RF contains a natural proline substitution in place of valine in the polymerase active site, *R. marinus* DNA polymerase I appears to be a candidate that may preferentially adopt the ajar conformation.

In order to observe the conformational changes of the Rhodothermus Fragment (RF) in the presence of a DNA substrate, and correct or incorrect nucleotides, we aim to utilize X-ray crystallography to obtain atomic resolution structures of the enzyme in various stages of the reaction pathway. The aims of this study were three-fold: solving the structure of the apo form of RF, of RF in its binary complex with a DNA substrate, and of RF in its ternary complex with DNA and in the presence of correct/incorrect nucleotides. Studying these three forms of the polymerase would allow us to visualize the

conformational changes that it undergoes at an atomic level, and would identify whether it adopts the ajar conformation during nucleotide selection. Understanding this process would not only contribute to a greater understanding of the DNA synthesis reaction, but it would also allow for the application of this newly garnered information to improving biosynthesis reactions in the field of biotechnology.

In the present study, we describe the 2.95 Å resolution structure of the large (5'-3' exonuclease deficient) Rhodothermus Fragment (RF) containing a mutation (D421A) in active site of the 3'-5' exonuclease domain, solved using the Multiwavelength Anomalous Dispersion (MAD) method. MAD is a method used in X-ray crystallography to aid in the determination of a three-dimensional atomic resolution structure. X-ray crystallography involves bombarding a crystallized molecule with X-rays, and then detecting the scattered radiation from the electron clouds of atoms in the crystal lattice to produce a diffraction pattern[20]. This diffraction pattern yields information regarding the amplitude of the diffracted rays, but not their relative phases. In order to generate an electron density map of a molecule, the phases of the diffracted rays must be determined. The MAD method involves the incorporation of heavy atoms, such as selenomethionine, into the molecule of interest so that when bombarded with varying wavelengths of X-rays, the resonance centers of the heavy atoms produce unique scattering patterns that yield the phases of the diffracted rays[20]. Once the phases have been determined, an electron density map is generated for use in model building and structure determination.

III. Results

A. Overall structure

The D421A RF protein was crystallized in space group $P2_12_12_1$, an orthorhombic unit cell with α, β, γ dimensions of 90 Å (Table 1).

Table 1. Data collection, phasing, and refining statistics.

	MAD peak	MAD edge	MAD remote
Data collection statistics			
Wavelength	0.97926	0.97943	0.97481
Space Group	*P212121*		
Unit Cell (A)	a = 59.13 Å, b = 107.46 Å c =124.05 Å, $\alpha=\beta=\gamma$ = 90 Å		
Resolution Limit	50.00 to 2.80 (3.03 to 2.95)	50.00 to 2.80 (3.03 to 2.95)	50.00 to 2.80 (3.03 to 2.95)
Unique reflections	32,050 (2,358)	32,081 (2,358)	32,115 (2,357)
Completeness	99.7 (99.9)	99.8 (99.9)	99.9 (99.9)
Rmerge	10.3 (71.5)	9.2 (67.2)	10.7 (92.9)
$I/\sigma I$	11.4 (2.03)	13.33 (1.96)	11.36 (1.44)
Redundancy	3.8	3.8	3.9
MAD phasing			
Heavy atom sites	16	16	16
Refinement statistics			
Resolution limit			47.80 to 2.95
R-factor			23.45%
R-free			29.19%
r.m.s.d. bond length			0.0080 Å
r.m.s.d. bond angles			1.21
Average of B-Factors			42.5 Å
Ramachandran plot			
Most favored			92.1%
Additional allowed			6.7%

The resulting structure was solved to 2.95 Å using MAD, and had R-free and R-factor values of 0.292 and 0.234, respectively (Table 1, Figure 7). The root mean square deviations (r.m.s.d) bond length and bond angles were 0.0080 Å and 1.21, respectively. The Ramachandran plot showed that 92.1% of residues were in the favored region while 98.8% of residues were in the allowed regions of the plot. Residues 569-607, which correspond to the tip of the thumb motif within the polymerase domain, were omitted from the structure due to poor electron density.

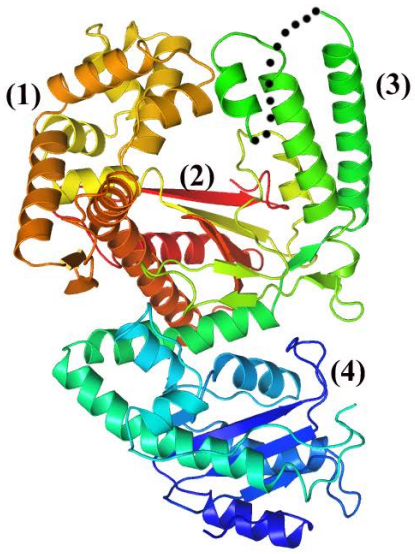

Figure 7. Ribbon representation of apo D421A RF solved at 2.95 Å. (1) Fingers Domain, (2) Palm Domain, (3) Thumb Domain, and (4) 3'→5' exonuclease domain. Dotted line depicts missing region of thumb motif.

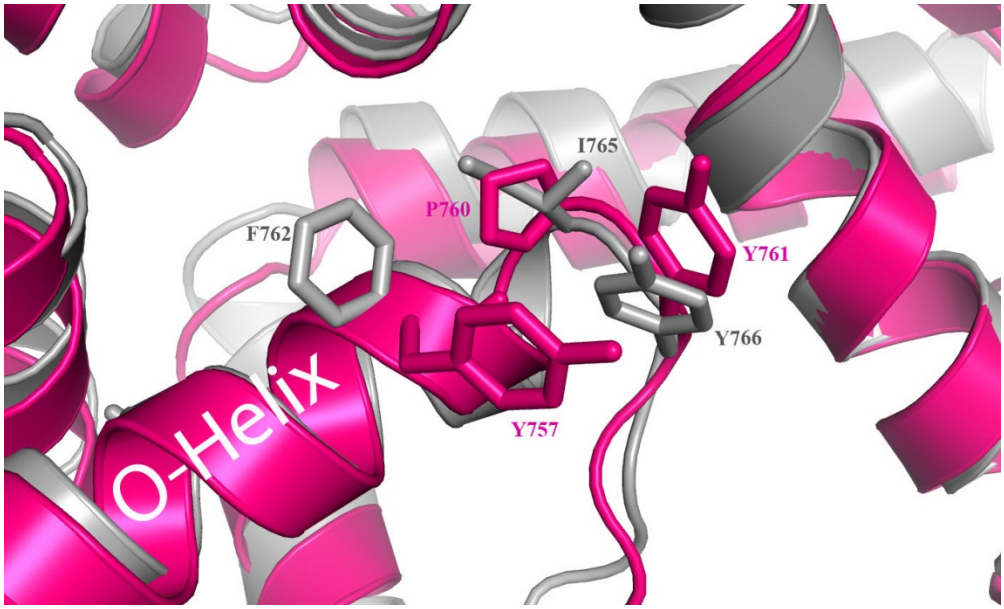

Figure 8. Polymerase active site structural alignment of apo RF (pink) and apo KF (gray). Residues P760 in RF and I765 in KF mark the C-terminal end of the mobile O-helix. Residues Y757 in RF and F762 in KF are shown to demonstrate natural substitution of conserved phenylalanine residue to tyrosine in RF.

Active site sequence and structural alignments of the polymerase domains in RF and KF demonstrated that mobile O helices were conserved in both sequence and structure, with the exception of P760 and Y757 (Figure 6, 8). Active site sequence and structural alignment of the 3'-5' exonuclease domains in RF and KF demonstrated that catalytically important amino acid residues were conserved between the two species (Figure 9). Mutation of 3'-5' exonuclease active site acidic residues in KF have been shown to decrease exonuclease activity by several orders of magnitude[21].

(a)

```
RF  336  -------VRNRQQLDELVARLDGLERLAIDTETTSTEAMWASLVGIAFSW
KF  326  SYDNYVTILDEETLKAWIAKLEKAPVFAFDTETDSLDNISANLVGLSFAI
BF  297  -------AKMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIAVVN
                      *   ::.       *: .*. . :   * :**::.

RF  379  EKGQGYYVPTP-----LPDGTPTETVLERLAPILR-RAQRKVGQNLKYDL
KF  376  EPGVAAYIPVAHDYLDAPDQISRERALELLKPLLEDEKALKVGQNLKYDR
BF  340  EHGRFFLRPET------------ALADPQFVAWLGDETKKKSMFDSKRAA
         * *     * .              .   : . *  .   *   : *

RF  423  VVLAQHGVEVPPPYFDTMVAHYLIAPEE-PHNLDVLARQYLRYQMVAITE
KF  426  GILANYGIELRGIAFDTMLESYILNSVAGRHDMDSLAERWLKHKTITFEE
BF  378  VALKWKGIELCGVSFDLLLAAYLLDPAQGVDDVAAAAKMKQYEAVRPDEA
           *   *:*:    ** ::  *:: .    .::   *.        .

RF  472  LIGSGRDQKSMRDVSIDEVGPYACEDTDIALQLADVLAAELDRH
KF  476  IAGKGKNQLTFNQIALEEAGRYAAEDADVTLQLHLKMWPDLQKH
BF  428  VYGKGAKRAVPDEPVLAEH---LVRKAAAIWELERPFLDELRRN
         : *.* .:    :  : *      ..:    :*   :  :* ::
```

(b)

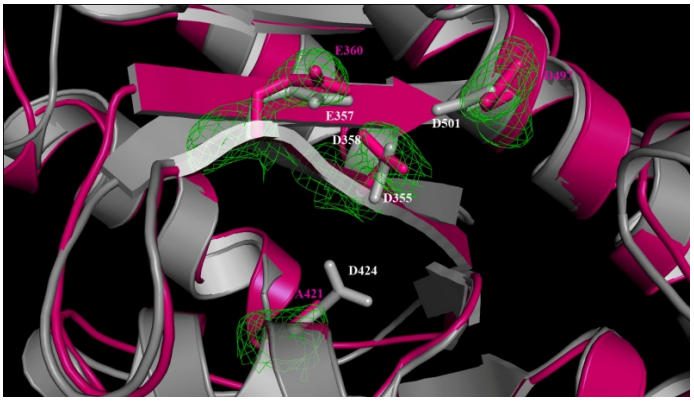

Figure 9a. 3'-5' exonuclease domain active site sequence alignment and 9b. 3'-5' exonuclease domain active site structural alignment in RF (pink) and KF (gray). Highlighted residues in red (9a.) represent conservation of amino acids essential for exonuclease activity in RF and KF. Same residues are depicted in structural alignment (9b.), with the exception of D421A mutation in RF. Electron density (green) is shown for residues in RF.

B. 3' to 5' Exonuclease Domain

DNA polymerase I is able to retain polymerase activity in the absence of a 5'-3' exonuclease domain. Therefore, the large fragment (5'-3' exo deficient) was used for crystallization, and site-directed mutagenesis (D421A) was utilized to inactivate the 3'-5' exonuclease domain of RF for future kinetic studies and co-crystallization trials. To assay exonuclease inactivation, a double-stranded DNA substrate containing 2-aminopurine (2-AP), an environmentally sensitive fluorophore, was bound to RF and its mutants. When base paired to thymine, 2-AP fluorescence is quenched. When the primer is in position "X-1" fluorescence is low, in position "X" fluorescence is high and in position "X+1" fluorescence decreases to an intermediate level (Figure 10)[22].

Figure 10. Double-stranded DNA Template used for 2-Aminopurine assay. Labeled in red are the positions that 2-AP will adopt as a result of exonuclease activity. When quenched at position (X-1), (X), and (X+1), 2-AP emits low, high, and intermediate levels of fluorescence, respectively.

The 2-Aminopurine fluorescence assay revealed that both wild-type RF and D421A had similar 3'-5' exonuclease activity as demonstrated by the sharp increase to a high level of fluorescence followed by a decrease to an intermediate level of fluorescence (Figure 11). Site-directed mutagenesis was utilized to make another mutant by changing a

separate single amino acid residue (D497A) in RF, and the 2-AP fluorescence assay revealed an initial slight change in fluorescence followed by a constant level of fluorescence for the remainder of the time period (Figure 11).

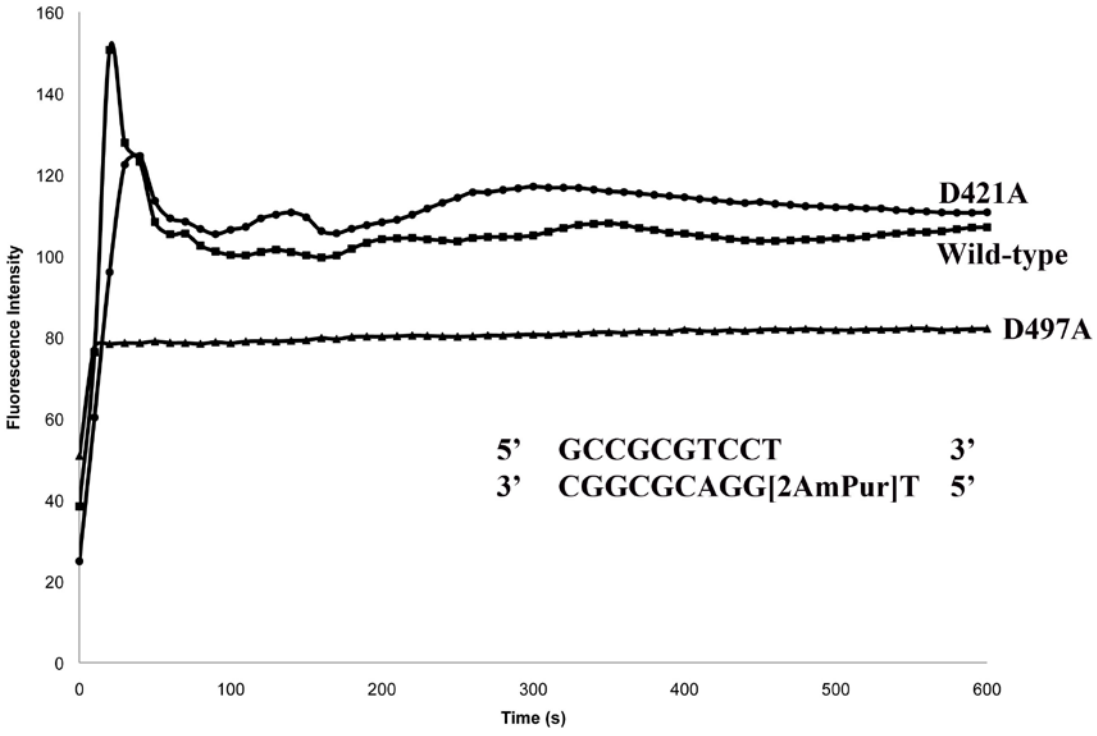

Figure 11. 2-AP Fluorescence Assay of RF, D421A RF and D497A RF. 3'-5' exonuclease activity in RF, D421A RF, and D497A RF was measured using 10ntPrimer+2-AP-11ntTemplate DNA. Fluorescence was measured over a 10-minute time course at 10 second intervals. The increase in fluorescence to an intensity of 80 most likely represents the binding event of protein to DNA.

IV. Discussion

A. *Overall Structure*

The overall structure of the Rhodothermus Fragment (RF) is similar to previously reported polymerases such as the Klenow Fragment (KF) and Bacillus Fragment (BF). DNA polymerase I consists of three structural domains: one that catalyzes DNA synthesis (polymerase domain), one that catalyzes proofreading and nucleotide excision repair (3'-5' exonuclease domain) of the synthesized DNA strand, and one that removes the RNA primers on Okasaki fragments (5'-3' exonuclease domain). Because the structure of the 5'-3' exonuclease deficient RF was solved, only the 3'-5' exonuclease domain and polymerase domain are depicted in the ribbon representation of RF (Figure 7). The top portion "claw-like" portion of the structure represents the polymerase domain (labeled 1-3) while the bottom portion is the exonuclease domain (labeled 4). As seen with other DNA polymerase I structures without DNA, the thumb tip is flexible and difficult to visualize in the crystal structure. Co-crystallization of RF with a DNA substrate should allow us to build in the missing region, as a substrate would stabilize this region that is solvent-exposed in the open conformation. The thumb has been shown to interact with DNA in other DNA polymerases and change its conformation when bound.

Active site structural alignment of RF and KF confirmed the natural substitution of the phenylalanine residue to tyrosine (F762 in KF and Y757 in RF) and isoleucine residue to proline (I765 in KF and P760 in KF) in the active site of RF (Figure 8). Both exonuclease active site sequence and structural alignment demonstrated that amino acid residues

deemed essential for the exonuclease reaction in KF (D355, E357, D501, and D424) are conserved in RF (Figures 9)[21]. Based on these structural observations, we hypothesize that both the polymerase and exonuclease activity of this enzyme should operate by a reaction mechanism similar to KF and BF.

B. *3' to 5' Exonuclease Domain*

Crystallographic studies of the exonuclease domain in KF have demonstrated that the product and inhibitor of the exonuclease reaction, deoxynucleoside monosphosphate (dNMP), binds to the active site of the 3'-5' domain[23]. Studies have also demonstrated that the presence of two divalent metal ions is crucial for binding the substrate and catalysis of the exonuclease reaction. A structural model of the KF exonuclease active site with a bound dNMP and metal ions has indicated that the phosphate oxygen atoms of dNMP interact with one metal ion bound to the protein by three amino acids—two aspartic acids, Asp355 and Asp501, and one glutamic acid, Glu357 (Figure 12, reproduced from Reference 23).

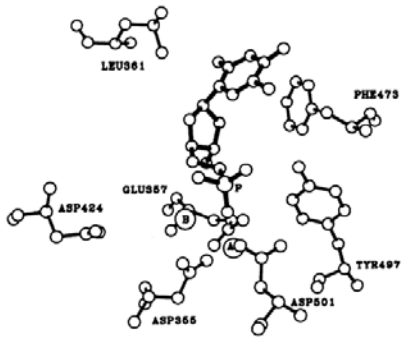

Figure 12. KF 3'-5' exonuclease active site. Deoxycytidine monophosphate (dCMP) is shown with thickened bonds, and side chains of active site residues are shown interacting with metal ions (labeled A and B) and dCMP. Reproduced from Figure 1, Reference 23.

The second metal ion lies adjacent to an aspartic acid, Asp424, only when the dNMP is bound, and shares Asp355 only when the first metal is bound[21]. Mutations to each of these amino acid residues have shown that exonuclease activity drastically decreased by at least four orders of magnitude, indicating that they are crucial to this reaction[21]. A mutation of Asp424 to alanine (D424A) in KF reported the largest decrease in exonuclease activity, followed by a mutation to Asp501 (D501A)[21]. In this study, we made a mutation in RF (D421A) that corresponded to the D424A mutation in KF in order to eliminate 3'-5' exonuclease activity. In spite of this mutation, a 2-AP fluorescence assay demonstrated that the D421A RF exonuclease activity was just as active as the wild-type exonuclease activity (Figure 11). This is peculiar in that this aspartic acid was demonstrated to chelate the second metal ion required for exonuclease catalysis; therefore, mutation of this residue should have prevented catalysis of this reaction. After making a separate second mutation (D497A) in RF that corresponded to the D501A mutation in KF, a 2-AP fluorescence assay demonstrated that exonuclease activity was abolished in this mutant. Although we have not been able to determine why the D421A RF mutant is still active, we hope to explore this question in future structural studies. The D497A mutant is currently being used for kinetic analysis as well as for co-crystallization trials.

In conclusion, the atomic resolution structure of the D421A, 5'→3' exonuclease deficient DNA polymerase I isolated from *Rhodothermus marinus* has been solved to 2.95 Å. Structural analysis has demonstrated that key amino acids at the active sites of both the polymerase and exonuclease domains of the polymerase are conserved with

other well-characterized polymerases such as the Bacillus and Klenow Fragments. Upon closer study of two RF mutants, one with D421A and another with D497A, which contain single amino acid mutations to acidic amino acids in the exonuclease active site, it was found that D421A RF 3'-5' exonuclease activity was similar to the wild-type RF, and that D497A RF lacked any inherent exonuclease activity. Future studies include crystallization of the polymerase in its binary and ternary complexes to study the conformational changes associated with nucleotide selection, and additional structural analysis of the exonuclease domains of wild-type RF and its mutants. Information garnered from these studies will allow us to characterize previously unseen reaction intermediates in the DNA replication pathway, and will provide useful information on how to improve these enzymes for applications in biotechnology.

V. Experimental

A. Cloning

DNA fragment encoding the 5'-'3 exonuclease deficient Rhodothermus Fragment (RF) was amplified by PCR, and inserted into pCR2.1-TOPO vector (Invitrogen) as EcoRI restriction-enzyme fragment. Oligonucleotides used for amplification were (5'-3') CATATGGAAAAGGCGGACTACCGGATCGTC and TCAGTGGGCATCCAGCCAGTTG-TC. Gene was sequenced using Sanger sequencing and inserted into bacterial expression vector, pET-21a (Novagen) as BamHI-NdeI restriction-enzyme fragments.

B. Site-Directed Mutagenesis: D421A

The D421A mutation was generated using the QuikChange site-directed mutagenesis kit (Stratagene) with Phusion High-Fidelity DNA Polymerase (NEB). The oligonucletoides used for the mutagenesis reaction were (5'-3')
AGAACCTGAAGTACGCTCTGGTGGTGCTGGC and
GCCAGCACCACCAGAGCGTACTTCAGGTTCT.

C. Selenomethionine protein labeling, expression and purification

The pET-21a plasmid containing the gene of interest was transformed into *Escherichia coli* B834(DE3) cells (Novagen, methionine auxotroph). Cells were cultured overnight with shaking at 37 °C in 60 mL Starter Media (Minimal Media + 5% LB) containing

ammonium sulfate (1 g/L), potassium dihydrogen phosphate (4.5 g/L), potassium phosphate (10.5 g/L), sodium citrate (0.5 g/L), amino acids (19 L-amino acids, no methionine; 0.8 g/L), nucleotide mix, glucose (0.5% w/v), magnesium sulfate (1 mM), thiamine (4 mg/mL), d-biotin (4 mg/mL), L-selenomethionine (30 mg/mL), ampicillin (100 mg/mL), and Luria-Bertani broth (5%). Inoculated 500 mL Minimal Media with 20 mL overnight culture and grew at 37 °C to OD_{600}=0.400. Expression of selenomethionine labeled protein was induced with IPTG (5 mM) and the cells were incubated overnight at 37°C. Cells were harvested by centrifugation at 4000 RPM for 15 minutes, and stored at 4 °C.

A bacterial pellet from cell culture was resuspended and lysed in B-PER (Bacterial Protein Extraction Reagent 20 mM Tris Buffer, pH 7.5; Pierce). Lysate was centrifuged at 7500 RPM for 30 minutes and supernatant was heated in 65 °C water bath for 20 minutes. Lysates were centrifuged again at 7500 RPM for 30 minutes, and supernatant was dialyzed in 1x Buffer A (50 mM Tris HCl pH 7.5, 1 mM EDTA, 6.5 mM BME). Supernatant was passed through heparin sepharose column equilibrated in 1x Buffer A using a linear gradient containing 1x Buffer A and 50% 1x Buffer A & 1.5 M sodium chloride (Buffer B). Fractions containing the protein were collected and exchanged back into 1x Buffer A through dialysis. The fractions were then passed through S1 ion exchange column (BioRad) pre-equlibrated with 1x Buffer A using a linear gradient containing 1x Buffer A and 1x Buffer B. Fractions containing protein were combined and concentrated using 10 kDa Amicron centrifugal concentrators (Millipore). Protein sample was run through S1 ion-exchange column again to remove remaining impurities,

again using linear gradient of 1x Buffer A and 1x Buffer B. Fractions containing RF were exchanged into 1x Buffer A through three cycles of concentration and dilution, and concentrated to 1 mL in 50 kDa Amicron centrifugal concentrator (Millipore).

D. Protein crystallization

Crystallization trials were set up in 96-well trays using PEG/Ion HT and Crystal Screen HT commercial screens (Hampton Research). D421A RF crystals were obtained in 20% w/v Polyethylene glycol 3350, 0.2 M Potassium thiocyanate at 17 °C using a protein concentration of 100 µM. Optimization of crystallization conditions was carried out by hanging drop vapor diffusion using a 2% gradient of 10-20% PEG 3350, 0.2 M Potassium thiocyanate. Crystals with average dimensions of 150 x 150 x 500 µM were grown by mixing 1 µl protein solution with 1 µl reservoir solution (18% PEG 3350, 0.2 M Potassium thiocyanate) and incubating at 17 °C for 5-7 days. Crystal was harvested with nylon cryoloop and transferred into 10 µl drop of 18% PEG 3350, 0.2 M Potassium thiocyanate. Crystal was cryoprotected by adding 18% PEG 3350, 0.2 M Potassium thiocyanate and 50% Ethylene glycol four times to the drop until final concentration of ethylene glycol was 33%, and was flash frozen in liquid nitrogen.

E. Data collection, structure determination and refinement

Multiwavelength anomalous dispersion data for a SeMet derivative crystal was collected at radiation wavelengths of 0.97926, 0.97943, and 0.97481 Å at the Advanced Photon Source at U.S. Department of Energy's Argonne National Laboratory, using beamline 22-ID-D at -173 °C with a MAR 300-mm CCD. All diffraction data were indexed, integrated and scaled with XDS.

The crystal structure of RF was determined using the MAD phasing method from a SeMet-substituted protein crystal to a maximum resolution of 2.95 Å. The PHENIX program suite was used to locate 16 selenium atoms and for phasing. Multiple rounds of iterative model building and refinement were performed using Coot and REFMAC5; 5% of the data were excluded from modification and used for cross-validation. Geometry of the structure was analyzed using MolProbity.

F. Site-Directed Mutagenesis: D497A

The D497A mutation was generated using the QuikChange site-directed mutagenesis kit (Stratagene) with Phusion High-Fidelity DNA Polymerase (NEB). The oligonucletoides used for the mutagenesis reaction were (5'-3') CCCTATGCCTGT GAA GCC ACGGACATT GCACTG and CAGTGCAATGTCCGTGGCTTCACAGGCATAGGG.

G. 2-Aminopurine (2-AP) Fluorescence Assay

Fluorescence assay using a 2-AP tagged DNA substrate was utilized for detecting RF exonuclease activity. The oligonucleotides used for the assay were (5'-3') GCCGCGTCCT and T[2AmPur]GGACGCGGC. Protein mixed with DNA was excited at 313 nm and an emission spectrum was collected at 360 nm over a 10 minute time course at 10 second intervals using the JASCO FP-6200 spectrofluorometer. This assay was conducted for the wild-type protein as well as the D421A and D497A mutants.

---

[1] Rothwell, P.J. and Waksman, G. (2010) *Nature* **465**, 1044-1048.
[2] Fijalkowska, I.J., Schaaper, R.M. and Jonczyk, P. (2012) *FEMS Microbiol Rev. Accepted Online.*
[3] Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990) *Protein Engineering* **3**, 461-467.
[4] Ollis.D.L , Brick.P., Hamlin.R., Xuong.N.G. and Steitz.T.A. (1985a) *Nature* **313**, 762-766.
[5] Johnson, S. J., Taylor, J. S., and Beese, L. S. (2003) *Proc Natl Acad Sci U S A* **100**, 3895-3900.
[6] Li, Y., Kong, Y., Korolev, S., and Waksman, G. (1998) *Protein Sci* **7**, 1116-1123.
[7] Kaushik, N., Pandey, V.N., Modak, M.J. (1996) *Biochemistry* **35**, 7526-7566.
[8] Johnson, K.A. (2010) *Biochim Biophys Acta.* ***1804***, 1041-1048.
[9] Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* **391**, 251-258.
[10] Franklin, M. C., Wang, J., and Steitz, T. A. (2001) *Cell* **105**, 657-667.
[11] Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999) *Structure* **7**, R31-35.
[12] Johnson, S. J., Taylor, J. S., and Beese, L. S. (2003) *Proc Natl Acad Sci U S A* **100**, 3895-3900.
[13] Steitz, T.A. (1999) *J Biol Chem* **274**, 17395-17398.
[14] Kunkel, T. A. (2004) *J Biol Chem* **279**, 16895-16898.
[15] Joyce, C.A. and Benkovic, S.J. (2004) *Biochemistry* **43**, 14317-14324.
[16] Rothwell, P.J. and Waksman, G. (2007) *J Biol Chem* **282**, 28884-28892.
[17] Joyce, C.A., Potapova, O., DeLucia, A.M., Hang, W., Basu, V.P., and Grindley, N.D.F. (2008) *Biochemistry* **47**, 6103-6116.
[18] Wu, E.Y. and L.S. Beese. (2011) *J Biol Chem* **286**, 19758-67.
[19] Alfredsson, G.A., Kristjánsson, J.K., Hjörleifsdóttir, S. and Stetter, K.O. (1988) *J Gen. Microbiol.* **134**, 299-306.
[20] Lattman, E.E. and Joll, P.J. (2008) *Protein Crystallography: A Concise Guide*. The Johns Hopkins University Press, Baltimore, MD, 136 p.
[21] Derbyshire, V., Grindley, N.D.F, and Joyce., C.M. (1991) *The EMBO Journal* **10**, 17-24.
[22] Purohit, V., Grindley, N.D.F., and Joyce, C.M. (2003) *Biochemistry* **42**, 10200-10211.
[23] Derbyshire, V., et al. (1988) *Science* **240**, 199-201.