2024

Isolation, purification, crystallization and preliminary X-ray data analysis of the B family DNA polymerase from Thermococcus thioreducens

Emika Miyamoto

ISOLATION, PURIFICATION, CRYSTALLIZATION AND PRELIMINARY X-RAY DATA ANALYSIS OF THE B FAMILY DNA POLYMERASE FROM *THERMOCOCCUS THIOREDUCENS*

Emika Miyamoto

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science
in
Biological Sciences
to
The Graduate School
of
The University of Alabama in Huntsville
August 2024

Approved by:

Dr. Joseph Ng, Research Advisor & Committee Chair
Dr. Jerome Baudry, Committee Member
Dr. Luis Cruz-Vera, Committee Member
Dr. Paul Wolf, Department Chair
Dr. Rainer Steinwandt, College Dean
Dr. Jon Hakkila, Graduate Dean

Abstract

ISOLATION, PURIFICATION, CRYSTALLIZATION AND PRELIMINARY X-RAY DATA ANALYSIS OF THE B FAMILY DNA POLYMERASE FROM *THERMOCOCCUS THIOREDUCENS*

Emika Miyamoto

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

Biological Sciences

The University of Alabama in Huntsville
August 2024

Family B DNA polymerases are the core polymerizing enzymes that carry out the essential process of DNA replication in many Archaeal organisms. Archaea depend on the unique properties of their DNA polymerases to faithfully replicate DNA within extreme environments. Variation in the catalytic properties of DNA polymerases from different organisms can be elucidated through detailed structural analysis by means of macromolecular X-ray crystallography. Crystallization is the crucial and bottleneck step in the process of determining the protein's three-dimensional structure by X-ray diffraction—essentially no crystal, no structure. This study optimizes the isolation and purification of recombinant DNA polymerase B from *Thermococcus thioreducens* (TthiPolB) for crystallization. Purified TthiPolB is screened using a sparse-matrix approach by sitting droplet vapor diffusion. Crystals were obtained and finally optimized with Polyethylene glycol (PEG) as the precipitant to produce crystals suitable for X-ray diffraction. TthiPolB crystals diffracted X-ray to 4.5 Å and they were determined to have orthorhombic space group $P2_12_12_1$ with unit cell dimensions a=70.70 Å, b=

109.56 Å, c=111.75 Å, α=β=γ=90^{o}. X-ray data was collected to about 85% completeness and a preliminary TthiPolB structure was determined by molecular replacement revealing the molecular arrangement of the crystal lattice.

Acknowledgements

I am immensely grateful to my research advisor, Dr. Ng, and committee members, Dr. Baudry and Dr. Cruz-Vera, for their sincere support and guidance. I could not have had more patient and insightful mentors, and it has made all the difference.

A very special thank you to Noriko whose unparalleled enthusiasm and thoughtfulness has been both an absolute pleasure to work with and instrumental to my work.

I would also like to express my gratitude to Dr. Adcock for her encouragement and wholehearted support of my efforts throughout this journey.

And of course, I greatly appreciate the continuous support of my parents and family who always believed I could accomplish anything.

Table of Contents

List of Figures

List of Tables

The limits of the investigator's patience may prove to be a formidable constraint.

– Alex McPherson & Jose A. Gavira

Chapter 1. Introduction

DNA polymerases are the key components of the many proteins that carry out DNA replication and repair. The proliferation of cells and maintenance of genetic material depends heavily on the polymerase's ability to accurately synthesize a new complementary strand of DNA by the correct sequential polymerization of dNTPs.

The DNA replication process is conserved in all domains of life. However, the replisome of different organisms can vary in their components. Most organisms utilize multiple DNA polymerases, unlike some Archaea that achieve DNA replication and repair with only one DNA polymerase. Polymerases differ in their fidelity, extension rate, processivity and other catalytic properties. The catalytic subunit of all DNA polymerases possesses a common architecture that resembles a right hand, with a thumb, finger and palm subdomain. The thumb domain is responsible for positioning the template DNA by binding to the double stranded part as it moves through the polymerase. The fingers interact with the single-stranded template and align the nucleotide for incorporation at the active site in the palm domain, at which DNA is bound in its active conformation (Hopfner, 1999; Kuznetsova *et al*., 2022).

Currently, DNA polymerases are categorized into seven different families based primarily on their amino acid sequence similarity in addition to phylogenetic analysis. These families are A, B, C, D, X, Y and RT. Families A, B, and C were originally established as *E. coli* Pol I, Pol II, and Pol III homologs, respectively. Family X was defined by eukaryotic polymerase β, followed by the discovery of a heterodimeric polymerase in the Euryarchaeota

subdomain, introducing family D (Marsic *et al*., 2008; Cann & Ishino, 1999). The DNA polymerase families known to-date are listed in Table 1.1 with their names and functions.

Table 1.1 Overview of DNA polymerase families.

Family	Taxon	DNA Polymerases	Functions
A	Eukaryota Bacteria Viruses	Pol γ, Pol θ, and Pol ν Pol I T7 DNA Pol	Replication, repair
B	Eukaryota Bacteria Archea Viruses	Pol ζ, Pol α, Pol δ, Pol ε Pol II DNA pol B T4 DNA Pol	Replication, repair
C	Bacteria	Pol III	Replication
D	Archea	Pol D	Replication
X	Eukaryota Bacteria Archea Viruses	Pol β, Pol σ, Pol λ, Pol μ, TdT Pol X Pol X ASFV DNA Pol	Repair
Y	Eukaryota Bacteria Archea	Rev1, Pol ι, Pol κ, and Pol η Dbh, Pol IV and Pol V Dpo4 DNA Pol	Translesion synthesis
RT	Eukaryota Viruses	Telomerase Reverse transcriptase	RNA-dependent DNA synthesis

From Kuznetsova *et al*., 2022

Archaeal DNA replication is relatively not well understood, particularly the process of lagging strand and Okazaki fragment synthesis and maturation. So far, most species characterized in Archaea, not including Crenarchaeota, possess both Family B DNA polymerases (PolB) along with the Family D DNA polymerase (PolD) unique to Archaea. Some species only require PolD for viability while others require both PolB and PolD (Greenough *et al*., 2015). Considering the complex evolution of Eukaryotic polymerases and their relationship to Archaeal polymerases in their homology, the distribution of Archaeal DNA polymerases across phyla is relatively

unresolved. So far, PolB1, PolB2, PolB3, and PolD have been recognized. However, no type seems to be ubiquitous within archaea (Lyu & Whitman, 2017; Makarova *et al*., 2014).

Archaeal family B polymerases possess the unique ability to sense uracil ahead of the replication fork as a way of preventing G-C to A-T mutations, thus increasing its fidelity (Barry & Bell, 2006). *Thermococcus thioreducens* DNA PolB (TthiPolB) kinetic parameters, including its relative fidelity, have been reported by Marsic *et al*. summarized in Table 1.2.

Table 1.2 A comparison of TthiPolB to other thermostable DNA polymerase parameters.

Source species	Pol family	Half-life at 95°C (min)	Relative fidelity	K_m^{DNA} (nM)	K_m^{dNTP} (μM each)	Extension rate (nt/s)
Thermococcus thioreducens	B	127[a]	1.7[a]	0.66 ± 0.02[a]	10.5 ± 1.5[a]	≥96[a]
Thermococcus sp. 9°N-7	B	–	–	0.05 ± 0.03[b]	75 ± 36[b]	–
Thermococcus fumicolans	B	200[c]	1.5[c]	–	–	–
Thermococcus kodakarensis	B	720[d]	3.7[d]	–	–	106–138[d]
Thermococcus litoralis	B	300[f], 480[e]	2.9[g]	0.07–0.12[e]	41–57[e]	17[e]
Pyrococcus furiosus	B	1,140[f]	6.2 [g]	0.7[f]	16 ± 2[f]	9.3 ± 1.7[f]
Thermus aquaticus	A	45[a]	1	4.0 ± 0.7[f]	24 ± 2[f]	47 ± 16[f]

From Marsic *et al*., 2008

The fidelity of TthiPolB was relatively lower than expected when compared to other *thermococcus* species. Overall, TthiPolB outperformed Taq family A polymerase with respect to its thermal stability, relative fidelity, and extension rate, leading to its interest in using TthiPolB for biotechnology applications in place of TaqPolA. However, other Archaeal polymerases such as that of *Thermococcus kodakarensis,* outperform even TthiPolB in addition to TaqPolA (Marsic *et al*., 2008). The question of what structural features in TthiPolB are responsible for its unique biochemical features still remains to be answered.

There are currently no reported efforts made in crystallizing TthiPolB. Macromolecular structure determination by x-ray crystallographic analysis is faced with the major challenge of the crystallization bottleneck in which producing a crystalline sample is an essential step in structure determination. The protein crystallization process is complicated by the large conformationally complex nature of macromolecules, which form relatively few interactions with neighboring protein molecules in a crystal lattice assuming its high molecular mass relative to small molecule compounds (McPherson & Gavira, 2014).

Each protein possesses unique chemical properties that define its potential to crystallize under given conditions, conditions that consists of multiple physical, chemical, and biochemical parameters, from posttranslational modifications to the effects of gravity (McPherson & Cudney, 2014; Krauss *et al*., 2013). Some of the primary factors considered in a crystallization experiment include pH, protein concentration, precipitant type and precipitant concentration (Bergfors, 1999; McPherson *et al*., 1995; McPherson & Cudney, 2014). Crystallization of a protein requires that these parameters are conducive to creating the precise solubility environment in which the protein molecules have a greater tendency to interact with each other in an orderly, repetitive manner, than with that of the surrounding solvent. However, this process must be fine-tuned to assure molecules come out of a supersaturated protein solution in a crystalline form instead of an amorphous aggregate. Favorable protein crystallization can be thought of in two stages; 1) nucleation occurs when molecules associate in a three-dimensional, stable aggregate that sets the foundation for 2) crystal growth as more molecules adsorb to each surface in an ordered lattice (Krauss *et al*., 2013). For either of these events to take place, the concentration of protein in solution must reach above its solubility limit, a state that's referred to as supersaturation. There are three principal regions to consider in the context of the

supersaturation state. First, the "precipitation" region is where the supersaturation state is so high that disordered aggregates (*i.e.*, amorphous precipitation) form out of solution. Secondly, a moderate "labile" region is where crystal nucleation and growth can occur. Finally, a lower supersaturation metastable region is where protein crystal growth occurs from an already pre-nucleated state. (McPherson & Gavira, 2014; Krauss *et al*., 2013). These regions are illustrated in the form of a phase diagram such as that shown in Figure 1.1 (Krauss *et al*., 2013). The phase diagram shows barriers of solubility and the principal areas of supersaturation at selective protein concentration as a function of the concentration of a precipitating agent.

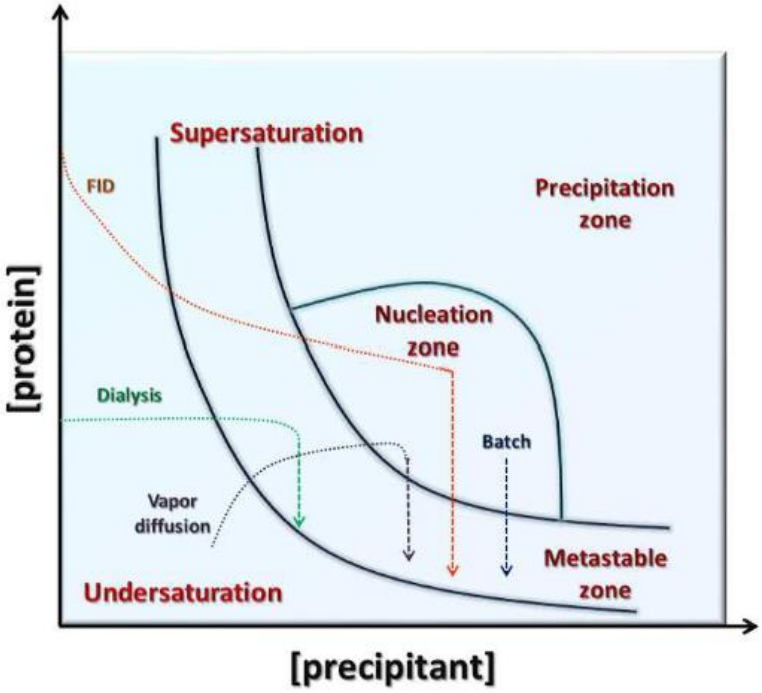

Figure 1.1 Protein phase diagram. Solubility of a protein is described at different protein concentration as a function of precipitants. The principal areas of supersaturation are shown as the metastable, nucleation (labile) and precipitation zones. Different paths used to reach nucleation and metastable zones are shown for different crystallization methods. Only vapor diffusion is discussed in this thesis. Diagram is taken from Krauss *et al*. (2013).

Vapor diffusion is one of the most common methods used to carefully control the protein phase changes during an equilibration process against precipitating reagents. This can be easily carried out by the sitting drop technique where a droplet containing protein and precipitant sits in a raised well exposed to the same chamber in which a reservoir, containing a much greater volume of precipitant, lies below. Within this sealed chamber, water leaves the droplet of lower precipitant concentration in the form of evaporation to equilibrate with the reservoir. This results in a gradual increase in precipitant and protein concentration of the droplet, bringing the solution towards supersaturation. This process can be conceptualized on the protein phase diagram (Figure 1.1). (Krauss *et al*., 2013). Since each protein is unique, there are no preexisting diagrams such as these to guide the initial crystallization process. Often, a sparse-matrix screen consisting of a wide range of precipitants is used to discover initial "hits" to be further optimized (Jancarik & Kim, 1991). In most cases, these screens result in crystals unsuitable for x-ray diffraction or no crystals at all. However, formation of various levels of precipitation can serve as useful information for rationalizing some form of systematic optimization (Cudney *et al*., 1994).

The size and the quality of the protein crystal is correlated with its ability to diffract X-ray and thus the quality of the final protein structure determination. Therefore, finding suitable chemical reagents that can drive a protein solubility condition to an optimal supersaturation state can determine downstream success/failure in obtaining diffracting quality crystals. Stable protein-solvent and protein-protein interactions are required for the periodic assembly in a crystal lattice. Protein heterogeneity and the incorporation of impurities may compromise proper intermolecular interactions that can lead to crystal defects rendering high mosaicity and low-resolution x-ray diffraction. Raw diffraction x-ray data used to resolve a structure are a direct

and literal reflection of the crystal lattice (McPherson & Cudney, 2014; McPherson & Gavira, 2014). Having quality crystals is essential for x-ray diffraction analysis as is refining protein purification procedures for crystallization. Slight differences in purification methodology have the potential to affect protein crystallization to an even greater extent than manipulations of crystallization conditions since these purification procedures bear the risk of altering the molecule's structure or chemical properties (McPherson & Gavira, 2014).

Efforts towards resolving the structure of TthiPolB would add to the growing repertoire of structural data that would permit the comparison of structure-function relationships across phyla. In addition to addressing such questions about *T. thioreducen's* replisome, structural studies of thermal stable proteins in general enable efforts to better understand the chemical properties of protein stability and its evolution. Determining the three-dimensional structure of TthiPolB would contribute to important medical and biotechnological applications such as extending enzyme shelf life, stability and PCR usage among other things. A thorough understanding of archaeal functioning could even allow us to use archaea as a model for human disease by "humanizing" archaeal genes (Lyu & Whitman, 2017). This study aims to provide the necessary, preliminary data that allow for the eventual structure determination of the *Thermococcus thioreducens* B family DNA polymerase.

Chapter 2. Methodology

2.1 Purification

The family B DNA polymerase from *T. thioreducens* (TthiPolB) was cloned and expressed as described by Marsic *et al*. (2008). TthiPolB recombinant *E. coli* cells, harvested from 10 l expression, were resuspended in SPB for lysis (10 mM Sodium Phosphate, pH 7.0, 100 mM NaCl, 0.1 mM EDTA, 0.3 mg/ml Lysozyme) at a volume to cell paste ratio of 10 ml per 1 g cell paste. PMSF and DTT were added to all buffers right before use to a concentration of 0.5 mM for each. Once thawed, the suspended cells were kept on ice and further lysed by sonication for 20 seconds 30 times, with 30 second intervals between sonication, using a Fisher Scientific™ model 550 sonic dismembrator. The cell lysate was centrifuged at 60,000 g for 30 min at 4 ℃. The supernatant was decanted into a 200 ml erlenmeyer flask and kept in a heat bath at 75 ℃ for 1 hour, then centrifuged again at 60,000 g for 30 min at 4 ℃ to remove precipitation. The supernatant was filtered through a 0.45 µm, Millex® - HV, PVDF syringe filter (Merck Millipore Ltd. Tullagreen, Carrigtwohill, Co. Cork, IRL), then loaded at a flow rate of 4 ml/min onto a HiTrap™ Heparin HP Affinity column (GE Healthcare, Uppsala, Sweden), pre-washed with a higher salt, elution buffer (10 mM Sodium Phosphate, pH 7.0, 1 M NaCl, 0.1 mM EDTA), and pre-equilibrated with a lower salt, running buffer (The same as lysis buffer, except without lysozyme). Following application of the supernatant, the column was washed with 10 x column volume (CV) of running buffer and the protein was eluted with a linear salt gradient of 0.1 M to 1 M NaCl, SPB at a flow rate of 3 ml/min. Elution Fractions corresponding to a relative

absorbance peak of the TthiPolB protein on an elution graph (Figure 3.1) were pooled and concentrated to approximately 1.5 ml by centrifugal filtration using a 50kDa cutoff Amicon® centrifugal filter (Merck Millipore Ltd. Tullagreen, Carrigtwohill, Co. Cork, IRL). The protein concentrate was loaded onto a Sephacryl™ S-200 HR size exclusion column (GE Healthcare, Uppsala, Sweden), pre-equilibrated with the same running buffer, at a flow rate of 0.7 ml/min. Fractions corresponding to the elution peak (Figure 3.2) were pooled and concentrated to 12 mg/ml by centrifugal filtration using a 50kDa cutoff Amicon® centrifugal filter (Merck Millipore Ltd. Tullagreen, Carrigtwohill, Co. Cork, IRL). Purity of the final protein solution was evaluated by SDS-PAGE and A280/A260 absorbance. Gel images were analyzed using Image Lab™ (Bio-Rad Laboratories, inc.) which evaluated purity at 85 % homogeneity. Enzyme activity was confirmed by using the pure TthiPolB as the polymerase in a PCR experiment (Figure 3.4).

2.2 Crystallization

TthiPolB crystals were produced by sitting drop vapor diffusion at 23 °C for all crystallization condition screening experiments. An initial screen was performed in a 3-drop format 96-well plate, using Crystal Screen HT ™ - HR2-130 crystal screen; conditions 1 - 68 (Hampton Research, Aliso Viejo, California, USA). A 10 mg/ml TthiPolB protein solution was used for each condition in a 1:1 ratio of protein to crystallization solution; 1µl of protein to 1µl of crystallization (reservoir) solution. The 2 µl droplet for each condition was equilibrated against a 35 µl reservoir. Crystals appeared as soon as 5 days after the droplet set in two of the 68 conditions tested: a 0.1 M sodium acetate trihydrate pH 4.6 / 8% PEG 4000 solution (D1), and a 10% PEG 1000 / 10% PEG 8000 solution (E7). The crystals were let to equilibrate for 3 weeks, after which, condition E7 was selected for further optimization. The E7 condition was re-tested under 6 different protein-to-reservoir drop volumes, using a 12.5 mg/ml protein solution obtained from a different purification batch as the initial screen (Figure 3.6).

A Chryschem 24 well format sitting drop plate (Hampton Research, Aliso Viejo, California, USA) was used for a scaled-up crystallization experiment using 6 µl and 7 µl total droplet volumes and a reservoir of 600 µl of the E7 precipitating reagent. Keeping the protein concentration at 12 mg/ml, various protein-to-reservoir volume ratios were tested; 1:1, 1:2 and 1:5 in 6 µl, and 1:1, 3:4, 2:5 and 1:6 in 7 µl. Set drops were let to equilibrate for 2 weeks. Crystals formed in the 1:1 protein-to-reservoir drops for both total volumes, 3 µl and 3.5 µl protein to 3 µl and 3.5 µl reservoir respectively (Figure 3.7). After 1 month these crystals were stored at 4 °C to be used for x-ray diffraction.

2.3 X-Ray Diffraction and Data Collection

X-ray diffraction data were collected by an in-house Rigaku diffractometer. The crystal was transferred to a cryoprotectant solution consisting of 20% PEG 8000, 20% PEG 1000, 25 mM SPB pH 7.2, and 50 mM NaCl. The crystal was soaked for no more than 10 s, then flash-cooled in a 110.3 K liquid nitrogen stream upon mounting. Data were collected from a single crystal. 360 frames were collected, using an oscillation angle of 0.5°, a crystal-to-detector distance of 100 mm and exposure time of 180 s. X-ray diffraction data were indexed and scaled using HKL-3000R (Minor *et al*., 2006).

Chapter 3. Results and Discussion

3.1 Purification and Crystallization

TthiPolB can be purified using a single heat-cut selection and two chromatography steps. Protein concentration and purity are evaluated at each step and summarized in Table 3.1. A total of ~3 mg can be reliably isolated and purified from 10 g of *E. coli* cell paste. After the second size exclusion chromatography step, the purified protein was further purified and concentrated by centrifugation using a 50 kDa filter until a protein concentration of ~10 mg/ml was achieved. The purity at this level increased by 10% as seen by the SDS-PAGE analysis (Figure 3.3), where the lowest molecular weight band is diminished between the gel filtration and centrifugal filtration step. After the heparin column step, a distinct band can be seen near the 100 kDa marker representing the 89.97 kDa TthiPolB protein. Higher molecular weight protein impurities, greater than 55 kDa and a small fragment corresponding to 14 kDa were observed throughout the purification process and persisted even after the final filtration step. These higher molecular weight contaminants may be limited degradative fragments of TthiPolB or minor contaminating proteins that resisted heat denaturation and co-purified with TthiPolB. It is also possible that the small 14kDa contaminating band may be the lysozyme used for cell lysis. Since these fragments are of different sizes, a narrower selection and collection of the primary elution peak in the size-exclusion chromatography may help in removing these fragments. Nonetheless, it is unclear whether these contaminating fragments facilitate or hinder downstream crystallization.

A total of 68 different crystallization conditions were screened. Conditions containing 0.1 M sodium acetate trihydrate pH 4.6 / 8% w/v PEG 4000 (D1) mix, and a 10% PEG 1000 / 10% PEG 8000 mix (E7) produced crystals which are shown in Figure 3.5. The latter was selected for subsequent optimization. The volume ratio optimization suggests that a 1:1 protein-to-precipitant (reservoir) volume ratio was optimal for the crystallization kinetics of TthiPolB (Figure 3.6), as the resulting crystals in these volumes were larger and more consistent in their external character which is marked by a lack of defects such as ridges, cracks, and clusters (McPherson & Cudney, 2014). In the two conditions of higher protein-to-reservoir volumes, significantly less aggregation is observed. The crystals formed under these volume ratios are smaller, more numerous and inconsistent in their character. Most crystals appeared in as soon as 5 days. In fewer cases, crystals appeared much later, such as the case in the "1P : 2R" drop well shown in Figure 3.6, where a crystal eventually forms out of high aggregate precipitation after as much as 12 days. However, these crystals are of relatively poor quality based on the ridges, branching, and other imperfections that can be seen of the crystal. Crystals were also obtained under the same 1µl : 1µl, protein-to-E7 conditions using protein solutions at 14 mg/ml and 15 mg/ml. However, fewer crystals resulted from these conditions and were less optimal compared to when a 12 mg/ml protein solution was used in this volume and are not shown here.

One of the attractive features of using PEG as precipitants is its cryoprotectant properties for subsequent X-ray diffraction under cryogenic conditions. Only slight adjustments to the concentration of the same PEGs for cryoprotection were made for soaking the crystals, which only had to occur for mere seconds. This likely reduced the shock that would otherwise occur when cryoprotecting crystals taken from other crystallization reagents. For these reasons, more broad screening for conditions did not seem necessary. In general, PEG is well known to be a

successful and reliable crystallization reagent. It is unsurprising that conditions that grew crystals are PEG based, as PEG precipitants have shown to be one of the most successful in protein crystallization (McPherson, 1999, p. 209). Although polymers like PEG are comparable to salts and organic solvents in their ability to influence solvent competition, and therefore protein solubility, PEGs provide a different macromolecule - solvent relationship by an "excluded volume" mechanism where other proteins and nucleic acids are "crowded out" of solvent, making the solvent more available for the predominant protein of interest (Atha & Ingham, 1981; McPherson, 1999). This property could allow for diffraction quality crystals in cases where less than ideal protein homogeneity is obtained. This may also partially explain the relatively poor quality of the crystals obtained by condition D1 of the initial crystallization screen, which utilized a precipitant with a lower PEG composition of only 8%, compared to the total 20% w/v PEG in E7, and the PEG that it does contain is of a different MW (4000).

However, 25 other PEG containing conditions failed to provide any crystals at all. Crystals that form in PEG of a certain molecular weight are likely to also form in other molecular weight PEGs within that range. These ranges are thought to be 200-600, 600-1500, 3350-8000, and 10000-20000 (McPherson & Cudney, 2014). Assuming this as a rule, the crystallizing ability of both conditions with this protein may be more attributable to the higher MW PEG component and is something that could be explored further in attempts to obtain high quality crystals. It can be difficult to assert the exact reason for the failure or success of crystal growth under certain conditions, especially in a unique mixture of precipitants. If future crystallization attempts were without knowledge of preliminary diffraction data, trying combinations of a lower MW PEG or sodium acetate trihydrate and a higher MW PEG would be a recommended starting point. For example, it may be interesting to test a 0.1 M sodium acetate

trihydrate pH 4.6 and 8% PEG 8000 mixture, and a 10% PEG 1000 and 10% PEG 4000 mixture to better understand the relative roles of these components, such as if the sodium acetate is providing a similar contribution to protein-solvent interactions as the lower molecular weight PEG in the mixture.

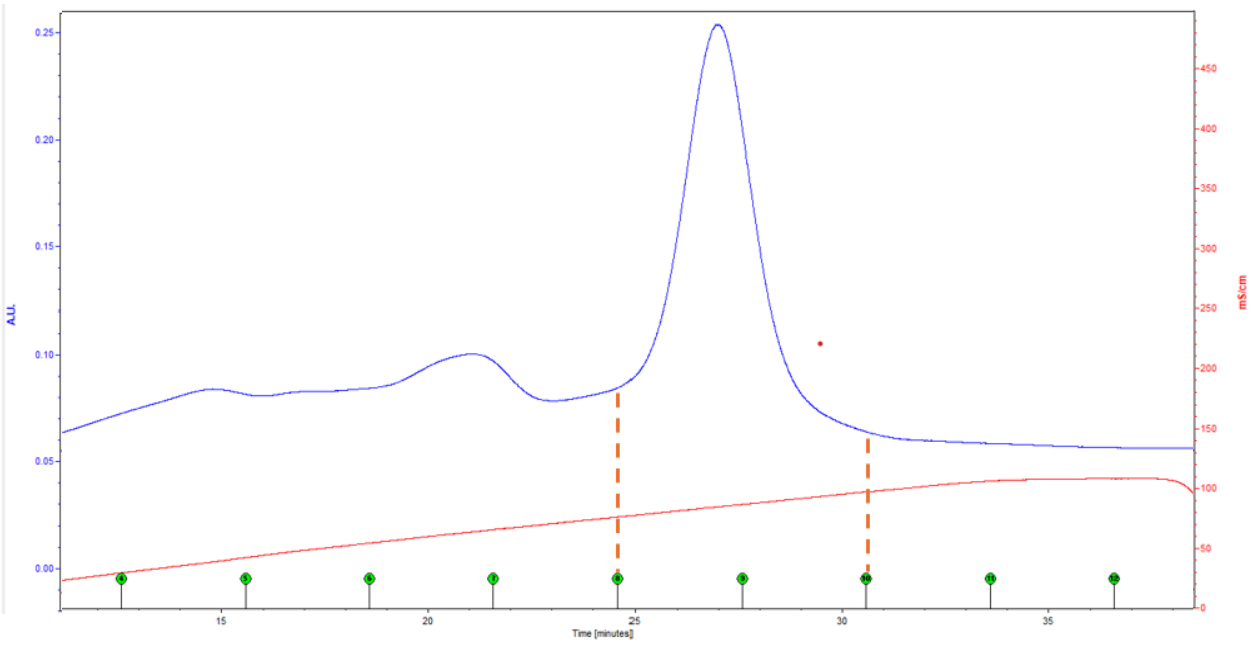

Figure 3.1 Heparin column chromatography elution graph. The major peak corresponds to TthiPolB elution brought about by a gradient increase in salt concentration of the elution buffer, indicated by the red line representing conductivity. Elution fractions indicated within the orange dotted lines, were collected and pooled for further purification.

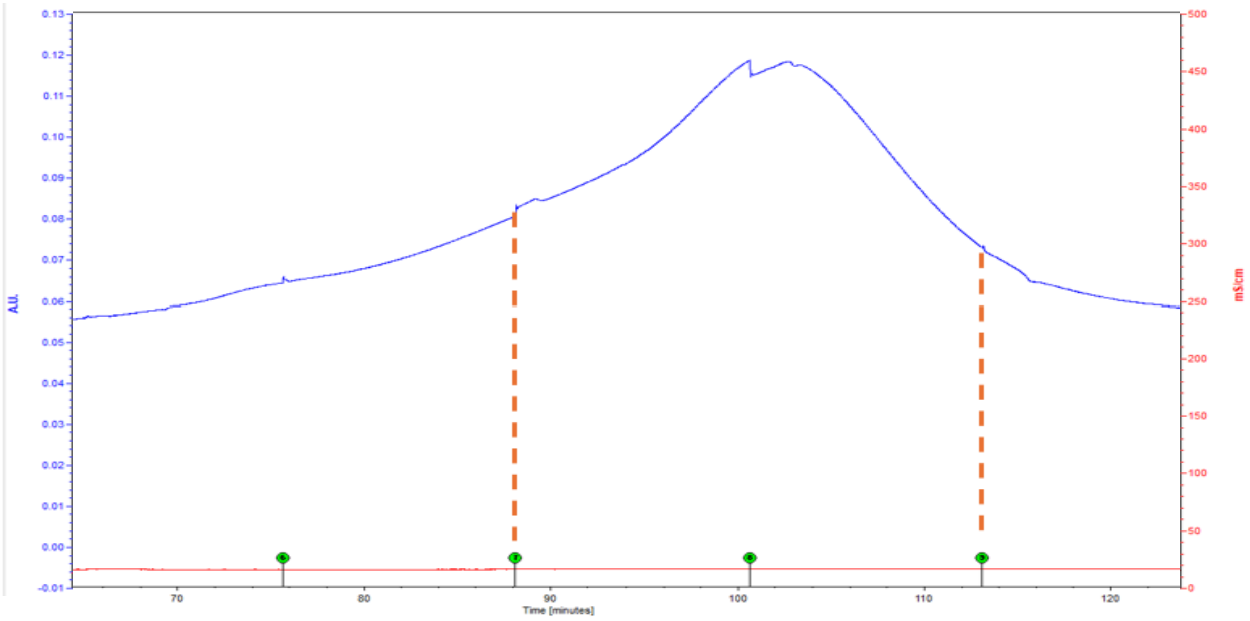

Figure 3.2 Size-exclusion S-200 gel column chromatography elution graph showing a primary peak corresponding to TthiPolB elution. Elution fractions indicated within the orange dotted lines, were collected and pooled to be concentrated for crystallization.

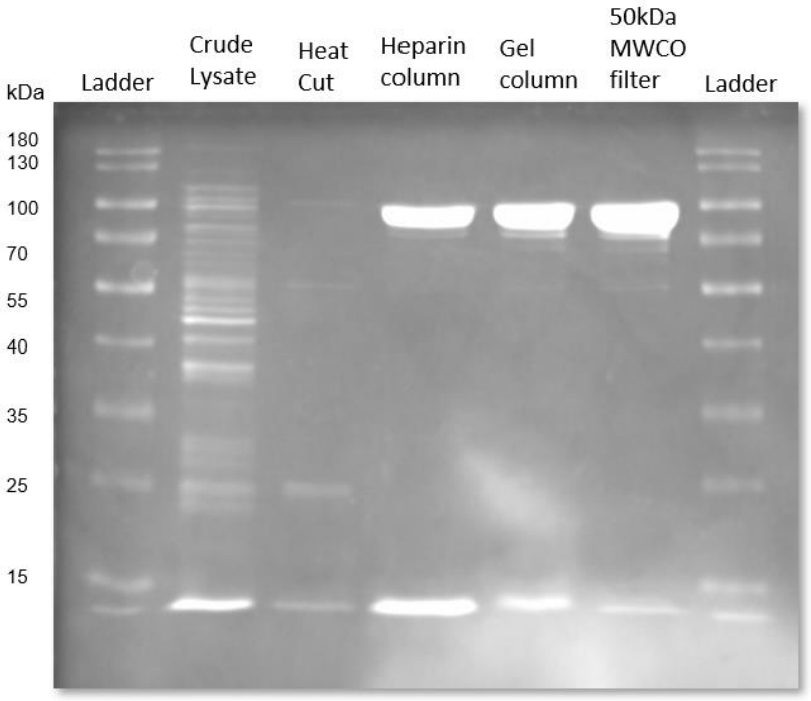

Figure 3.3 SDS-PAGE analysis of TthiPolB for each purification step. The last "50kDa MWCO filter" column shows a band around the 100 kDa MW mark representing TthiPolB and was measured at 85% homogeneity.

Table 3.1 TthiPolB purification table listing the volume, concentration and total protein in solution at each step in the purification process, starting from 11 g of recombinant *E. coli* cell paste. Concentrations are calculated using A280 absorbance and the theoretical extinction coefficient for TthiPolB (Abs 0.1% = 1.3).

Step		**Fraction Volume (ml)**	**Conc. (mg/ml)**	**Total Protein (mg)**	**A280 / A260**
1	Crude cell free extract	130 ml	40.92 mg/ml	5319.6 mg	0.59
2	Heat cut	150 ml	30.59 mg/ml	4588.5 mg	0.48
3	Heparin Column	14 ml	0.57 mg/ml	7.98 mg	1.56
4	Size exclusion (gel) column	15 ml	0.246 mg/ml	3.69 mg	1.95

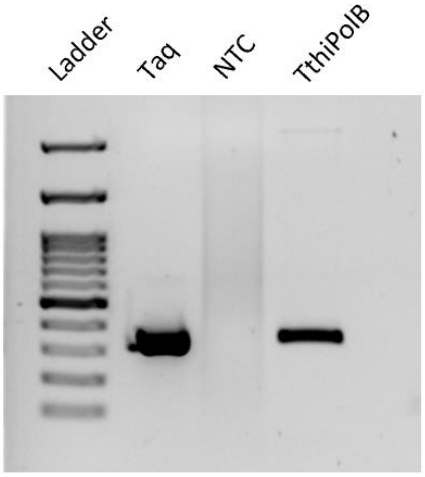

Figure 3.4 PCR agarose gel amplification showing polymerizing activity of purified TthiPolB.

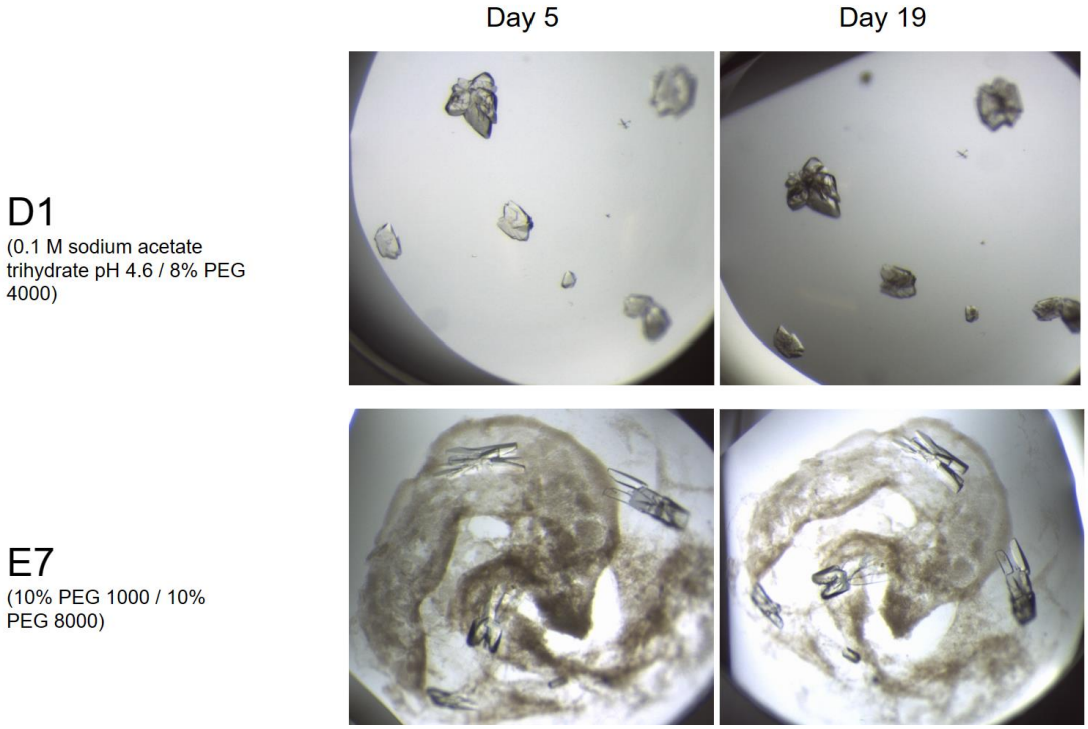

Figure 3.5 Crystals formed in the D1 and E7 precipitant conditions from the initial crystallization screen (Hampton Crystal Screen HT ™ - HR2-130) using a 10 mg/ml TthiPolB solution. The same drop well for each condition 5 days and 19 days after droplet set are shown side by side. Crystals in the D1 condition appear to degrade between day 5 and day 19.

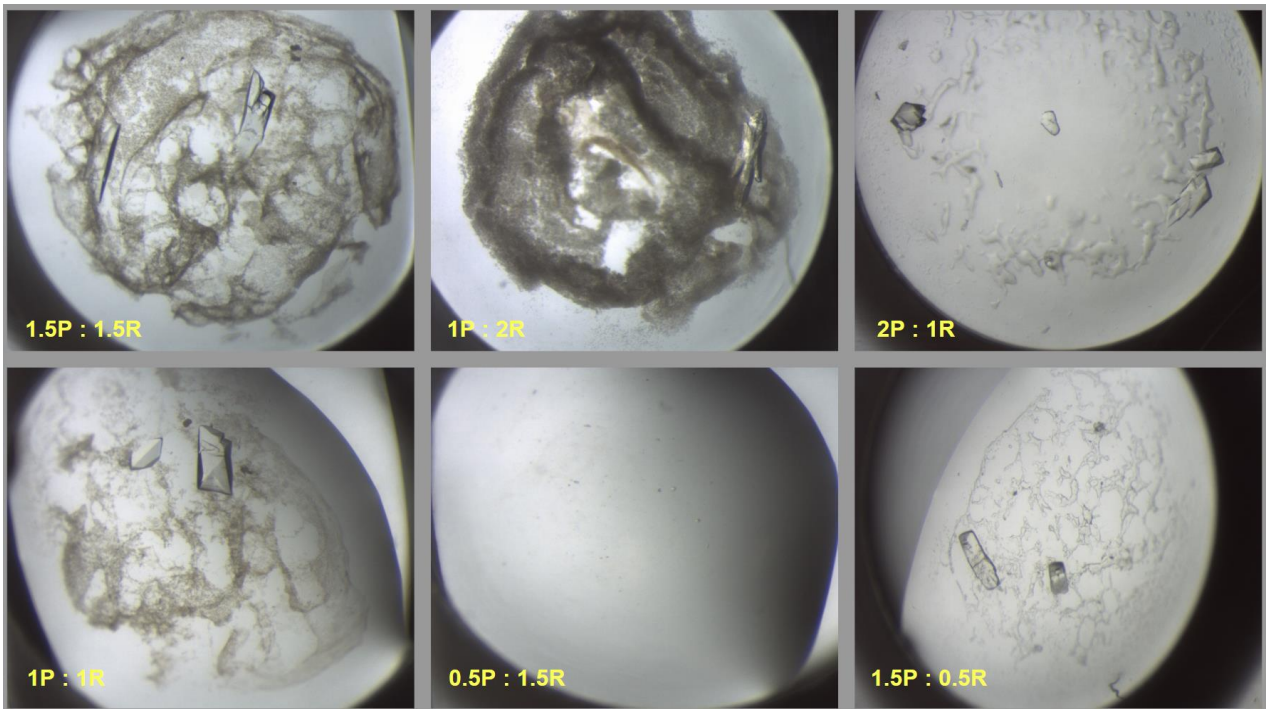

Figure 3.6 Crystals grown from 6 different protein to reservoir (P : R) drop volume ratios (shown at the bottom left of each image in yellow) using a 12 mg/ml TthiPolB protein solution and Hampton HR2-130 crystallization reagent E7. The 1.5P:1.5R and 1P:1R ratios show the largest crystals with the fewest imperfections.

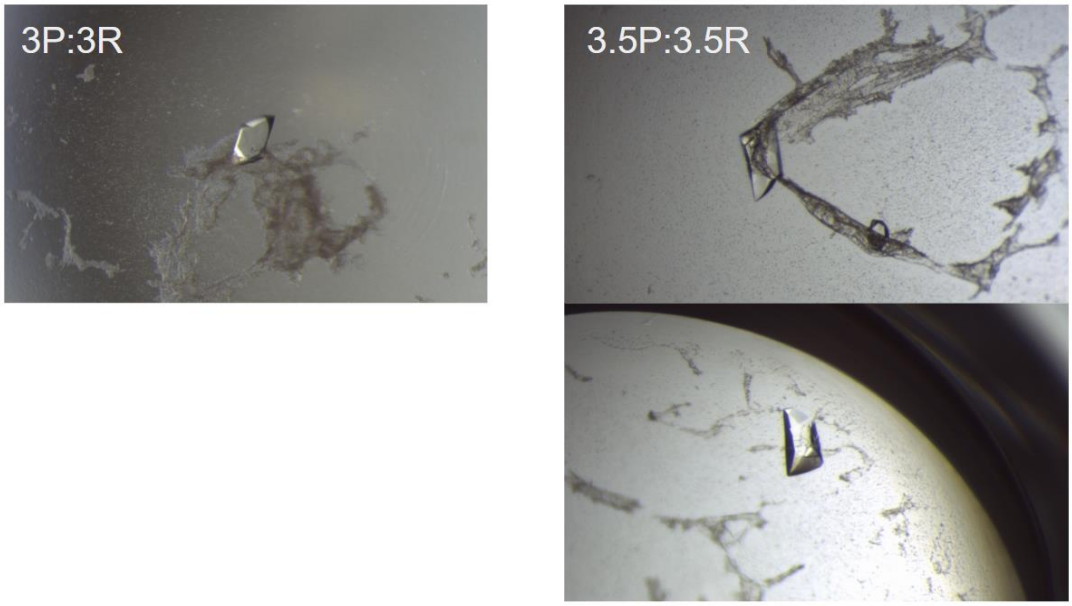

Figure 3.7 Crystals grown in 6 µl and 7 µl total drop volumes in a 1:1 ratio of protein to reservoir solutions using a 12 mg/ml TthiPolB protein solution.

3.2 X-ray Data Statistics and Structure Modeling

X-ray diffraction data were indexed and scaled with HKL3000 (Minor *et al*., 2006). The TthiPolB crystals diffracted X-ray to 4.5 Å and were determined to have orthorhombic space group $P2_12_12_1$ with unit cell dimensions a=70.70 Å, b= 109.56 Å, c=111.75 Å, α=β=γ=90°. X-ray data were collected to about 85% completeness. The x-ray data collection statistics are shown in Table 3.2. A preliminary TthiPolB structure was determined by molecular replacement using the PHENIX software package (Liebschner, 2019) revealing the molecular arrangement of the crystal lattice. The PolB structure from *Thermococcus kodakarensis* (PDB ID 4K8Z) was selected as the template structure which has a 90% sequence identity. This selection was primarily based on the highest sequence identity and sequence coverage out of known Protein Data Bank structures using SWISS-MODEL (Waterhouse *et al*., 2018).

Table 3.2 X-ray data collection statistics for TthiPolB.

Space group	$P2_12_12_1$
Unit cell parameters	
a	70.704 Å
b	109.564 Å
c	111.751 Å
Wavelength	1.54 Å
Resolution range	49.78 Å - 4.462 Å (4.51 Å -4.46 Å)
Total reflections	283092
Unique reflections	5443 (40)
Multiplicity	6.4 (6.2)
Completeness (%)	85.11 (13.33)
Mean $I/\sigma(I)$	8.5 (7.0)
Wilson B-factor	50.8
R-merge	0.217

$R_{merge} = \sum|I_{obs} - \langle I\rangle| \,/\, \sum I_{obs}$, where I_{obs} is the observed intensity and $\langle I\rangle$ is the mean intensity of the reflection.

3.3 Crystal Packing

A single TthiPolB molecule makes up the asymmetric unit and is shown colored by domain in Figure 3.8. The TthiPolB crystals are of a $P2_12_12_1$ space group. In a primitive orthorhombic space group, molecules assemble a crystal lattice by a half unit cell translation and 180° rotation, forming two-fold rotational symmetry or screw axis (Figure 3.9). TthiPolB molecules in a $P2_12_12_1$ space group form 3 unique interfaces that repeat throughout the lattice. The outer edge of the palm/thumb surface interfaces with that of a thumb/n-terminal surface, as does a palm/n-terminal and n-terminal/exonuclease surface. Additionally, the two major polymerase surfaces create the largest interface, with the finger domain contacting the thumb domain on the opposite face.

Figure 3.8 Model of the TthiPolB protein based on the homologous template structure from *T. kodakarensis* (4K8Z) colored by domain. Red-thumb, yellow-exonuclease, blue-N-terminal, green-palm, orange-fingers.

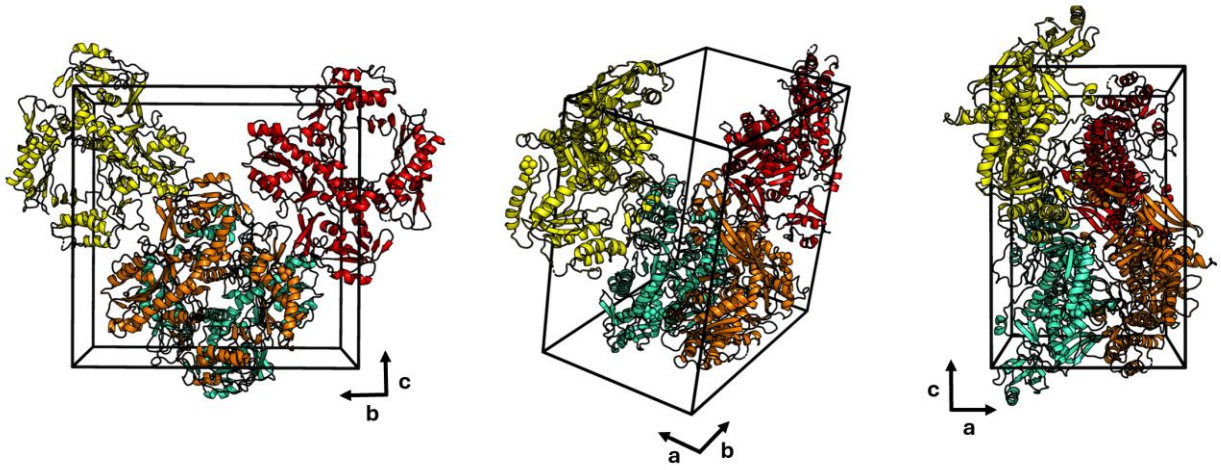

Figure 3.9 Crystal packing within the P212121 unit cell. a = 70.70 Å, b = 109.56 Å and c = 111.75 Å and α, β, γ = 90°. Molecules assume space within the unit cell by a half unit cell translation and 180° rotation along each axis.

3.4 Residue Interactions

In addition to its ability to form powerful salt bridges, arginine can provide strong polar interactions due to its charged side chain creating a dipole environment. This seems to be the case in Figure 3.10 where polar interactions are indicated between Arg-739 and Ser-301 and Gly-302. The neighboring Glu-303 residue may further stabilize these bonds by forming polar interactions between the same two residues interacting with Arg-739. All residue interactions discussed here are summarized in Table 3.3.

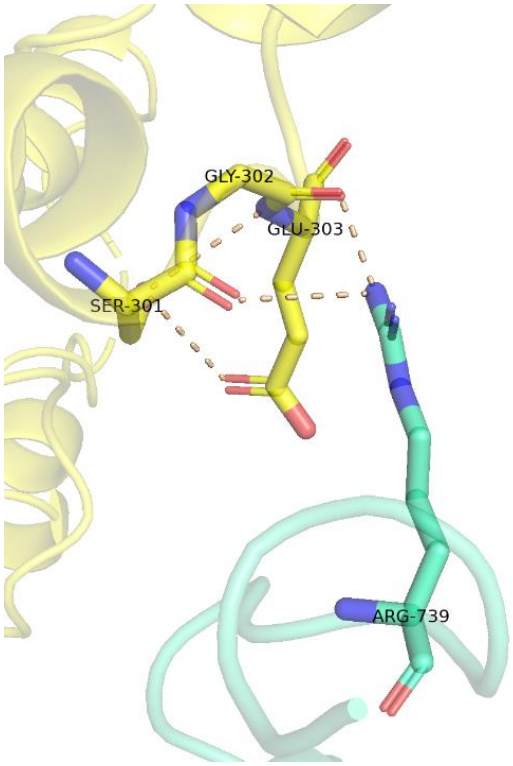

Figure 3.10 Polar interactions indicated by the dotted lines between Arg-739 and Ser-301 and Gly-302. The neighboring Glu-303 residue partakes in polar interactions between the same two residues interacting with Arg-739.

Figure 3.11 shows the next closest residues on this interface. However, these residues are further apart with the closest potentially interacting atoms at least 5Å apart. In addition, the side chain orientations suggest relatively poor interactions if any. Mutating Trp-397 and Pro-537 to a polar

molecule such as Serine or Glutamine may reduce steric hinderance and hydrophobic effect of the large ring structures on these amino acid side chains, while also increasing electrostatic and polar interactions with the adjacent Thr-711 residue and surrounding water molecules. A glutamine residue may be the more appropriate replacement for Trp-397 since the point from which the tryptophan side chain originates in the peptide backbone is approximately 10 Å from the nearest point on Thr-711 of the adjacent molecule. This may also help the neighboring Asp-398 partake in polar interactions with residues on the exposed region of the adjacent molecule that bears a positively charged Lys-712, and polar Thr-711 and Asp-709. Suggested point mutations are listed in Table 3.4.

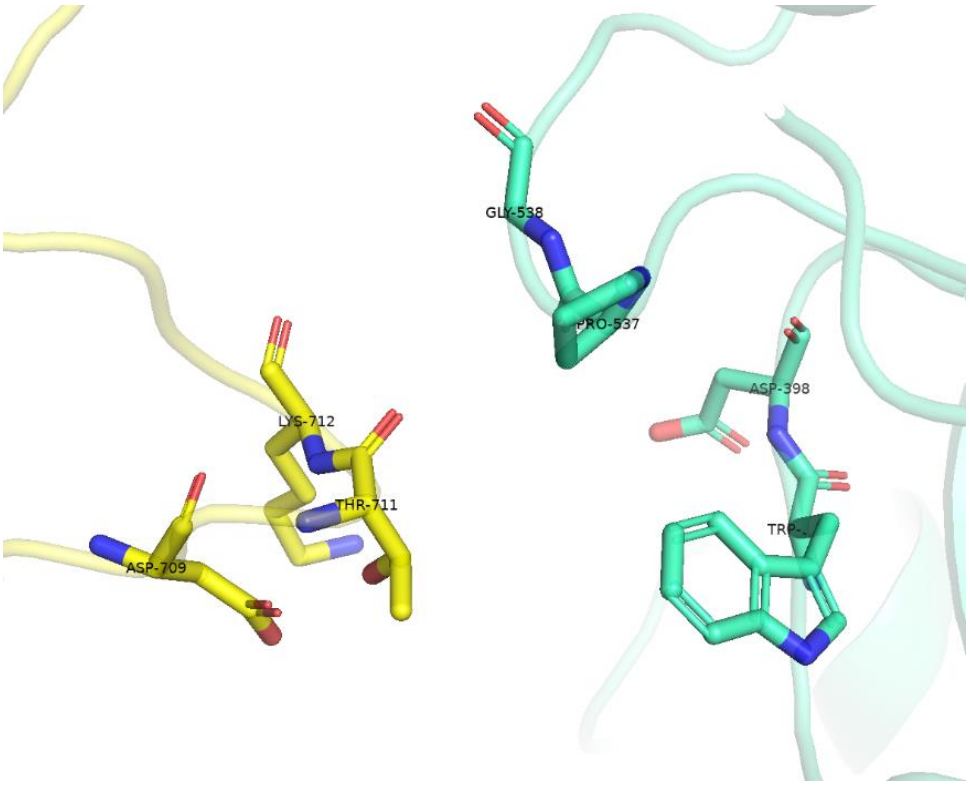

Figure 3.11 Asp-709, Lys-712, and Thr-711 (yellow molecule), and Gly-538, Pro-537, Asp-398, and Trp-397 (blue molecule) interfacing.

The residues shown in Figure 3.12 indicate two of the closest, and likely most important, areas of contact at this interface. Glu-10 and Arg-433 form a salt bridge at their charged side chains that is suggested to be 3.5 Å in length. At the same time, the backbone nitrogen on Arg-433 is partaking in a polar interaction with an oxygen on Asp-11. Glu-81 and Lys-188 are also shown on this interface. Similarly, these residues sidechains are oppositely charged and about 3.3 Å apart at their closest point. It can be assumed that they would provide an important point of contact in stabilizing the crystal lattice. Although lysine salt bridge energies are lower than that of Arginine due to Lysine's lower pKa value, charged amino acids can provide a strong electrostatic interaction even 5 Å apart (Xie *et al*., 2015; Zhou & Pang, 2018). Besides the possibility of an electrostatic interaction happening here, these residues would likely form hydrogen bonds with water molecules that could mediate longer range interactions (Levy & Onuchic, 2004). Relative to their size, protein crystals in general exhibit fewer bonds between molecules in a lattice compared to those that form within small molecule crystals, making it even more important that these few electrostatic interactions be preserved at this interface (McPherson & Gavira, 2014).

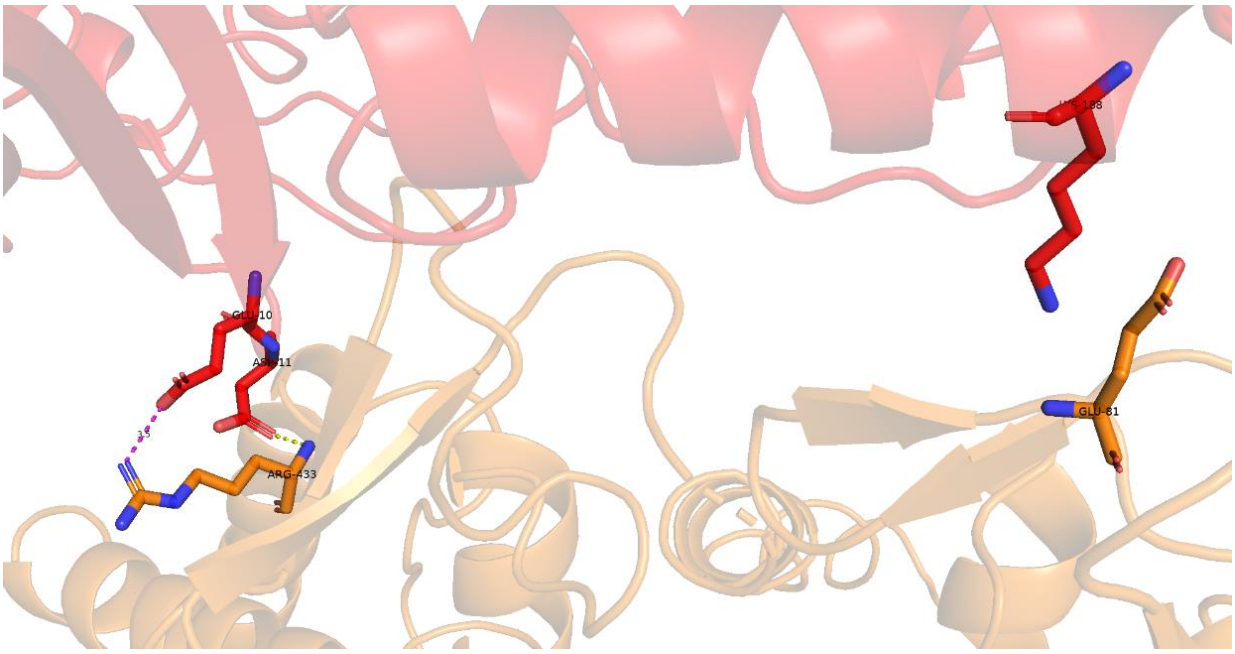

Figure 3.12 Arg-433 and Glu-81 (orange molecule), and Asp-11, Glu-10 and Lys-188 (red molecule) interfacing. The salt bridge between Glu-10 and Arg-433 is shown as a pink dotted line. The Polar contact is shown in yellow. This interface seems to depend on the suggested electrostatic interactions.

The interface of the two largest polymerase surfaces provides a diversity of interactions at the fingers domain, which can be seen imbedding itself into the thumb domain of the neighboring molecule (Figure 3.13). Lys-466 can be seen in proximity to Tyr-719, Gln-724 and Trp-603, allowing it to form polar interactions with the oxygen of the Tyr and Gln, and a pi-cation with the Trp. Creating a salt bridge here by mutating Tyr, Gln or Trp may be possible to strengthen the interaction. However, the large branched or ring structure of these residues may be important in folding and maintaining a hydrophobic environment in the interior, meaning a mutation to these residues could have harmful effects on the main chain folding (Levy & Onuchic, 2004). In any case, this region is also supported by Asp-472 forming a relatively close salt bridge (2.4-2.8 Å) with two other residues, Arg-600 and Lys-581. In general, this interface is well supported by numerous other polar interactions due to its proximity and large surface area of contact which are not shown.

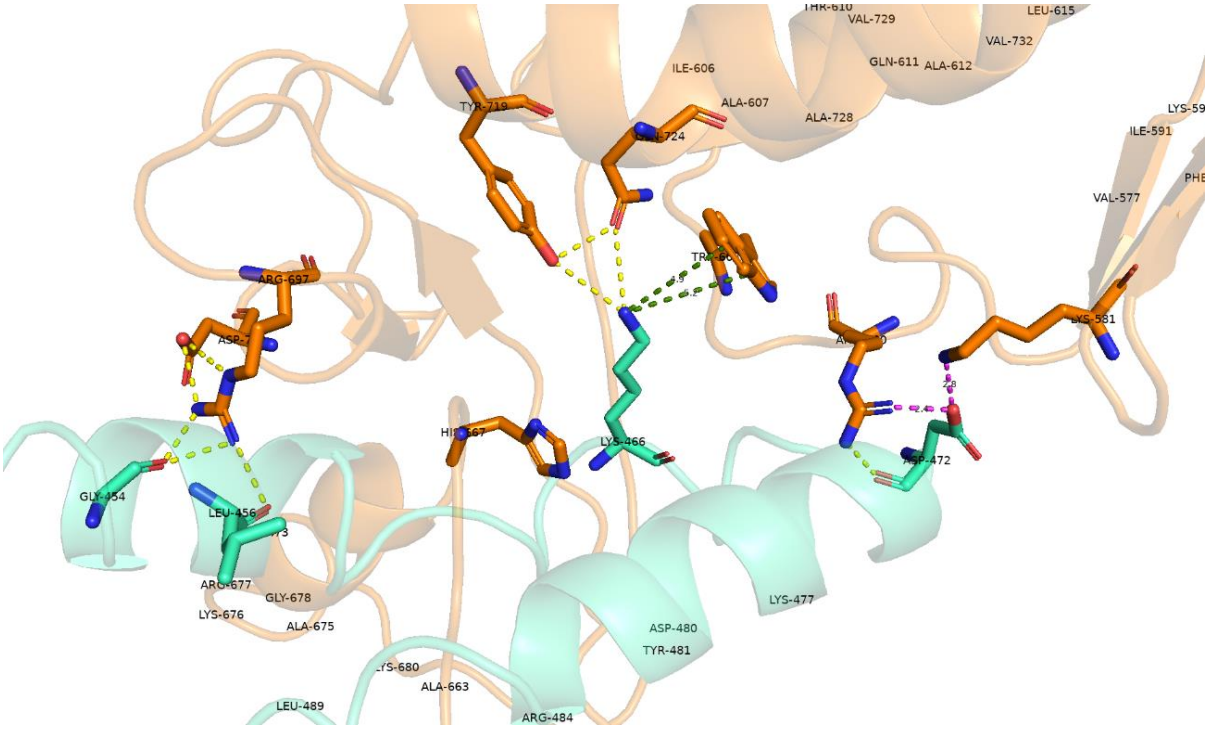

Figure 3.13 The finger domain (blue) interacting with residues within the thumb domain (orange) of a neighboring molecule. Asp-472 marks the innermost point of this interface and can be seen forming two salt bridges with the inner thumb domain residues, Arg-600 and Lys-581. Interior hydrophobic residues are shown interacting with Lys-466.

Table 3.3 Likely residue interactions at each of the three interfaces found in the P212121 crystal lattice. Two interacting residues ("Residue A" and "Residue B") are listed with the involved atom types, listed respectively for multiple interactions between two residues.

Residue A ID / Atom(s)	Residue B ID / Atom(s)	Interaction(s)	Distance (Å)
GLU-10 / O	ARG-433 / N	Salt bridge	3.5
ASP-11 / O	ARG-433 / N	Polar	2.5
LYS-188 / N	GLU-81 / O	Salt bridge	4.2
SER-301 / O	ARG-739 / N	Polar	3.4
GLY-302 / O	ARG-739 / N	Polar	2.4
GLY-454 / O	ARG-697 / N ; N	Polar	2.5 ; 2.7
LEU-456 / O	ARG-697 / N	Polar	2.7
LYS-466 / N	TYR-719 / O	Polar	3.4
LYS-466 / N	GLN-724 / O	Polar	3.4
LYS-466 / N	TRP-603 / -	Pi-cation	4.9 ; 5.2
ASP-472 / O ; O	ARG-600 / N ; N	Polar ; Salt bridge	2.2 ; 2.4
ASP-472 / O	LYS-581 / N	Salt bridge	2.8

Table 3.4 Suggested residue mutations for strengthening crystal packing.

Residue	Mutation
GLY-427	THR
TRP-397	GLN
PRO-537	SER

Chapter 4. Conclusion

DNA polymerases reliably synthesize and maintain the genetic blueprint for the very building blocks of life, and do so to varying extents as a result of the organism's unique phylogeny and environmental demands. As these functional differences are often reflected in the unique structural properties of a protein, a structural analysis is required for understanding this functioning. The isolation, purification, crystallization and x-ray diffraction analysis of the B family DNA polymerase from the hyper-thermophilic Archaeon, *T. thioreducens* is described. The thermal stability, DNA binding property and 89.97 kDa size of TthiPolB allows for the utilization of a major purification step in the form of heat denaturation, followed by an affinity and size exclusion chromatography. A promising indication of this protein's potential in crystallization for x-ray diffraction analysis is its ability to crystallize in two conditions of an initial sparse matrix screen. Crystals of ~0.5 mm in the largest dimension can be produced from an 85% homogenous TthiPolB solution using a 10% PEG 1000 and 10% PEG 8000 mixture as the crystallization reagent in a protein-to-precipitant volume ratio of 1:1 for a total volume of 6-7 µl in a sitting drop vapor diffusion set-up. The resulting crystals provide x-ray diffraction data to 4.5 Å resolution. Conveniently, crystals can be obtained without alterations to the pH 7 buffer at any point, including during crystallization optimizations. Under these conditions, TthiPolB molecules pack in the $P2_12_12_1$ orthorhombic space group, resulting in three unique molecule interfaces that repeat throughout the crystal lattice. Important electrostatic interactions can be seen at all interfaces, with one interface seeming to depend on this type of interaction. Efforts should be made to preserve these relatively strong non-covalent bonds, which can be done by ensuring that the protein buffer solution is kept at a pH of 7, especially at the crystallization step to keep charged residues in their protonated state. The largest interface may be an important area

of overall lattice stability as it provides a large surface area of contact and numerous potential residue interactions. Although strengthening attractive forces and/or decreasing repulsive forces of the large, hydrophobic amino acids at this internal interface may be tempting, introducing any point mutation here should be approached with caution as these residues may be essential for the main chain folding. This crystal lattice may benefit more from other specific residue mutations, especially a Trp-397 to Glutamine point mutation, which could provide an electrostatic attraction at an interface that is seemingly distant and lacking in interactions. Increasing protein solution homogeneity in attempts to obtain higher quality crystals should be explored first, as this can be done by resolving the size exclusion chromatography elution, something that does not require changes to amino acid sequence or pH, factors that could disrupt protein folding and complicate crystallization. Nonetheless, this work sets a solid basis for optimizing TthiPolB crystallization. With this insight, achieving high resolution structure determination of TthiPolB is a very feasible prospect.

References

Atha, D. H., & Ingham, K. C. (1981). Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *The Journal of biological chemistry*, *256*(23), 12108–12117.

Barry, E. R., & Bell, S. D. (2006). DNA replication in the archaea. *Microbiology and molecular biology reviews : MMBR*, *70*(4), 876–887. https://doi.org/10.1128/MMBR.00029-06

Bergfors, T.M. (1999). *Protein Crystallization: Techniques, Strategies and Tips*. International University Line.

Cann, I. K., & Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics*, *152*(4), 1249–1267. https://doi.org/10.1093/genetics/152.4.1249

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., & McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. *Acta crystallographica. Section D, Biological crystallography*, *50*(Pt 4), 414–423. https://doi.org/10.1107/S0907444994002660

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. https://doi.org/10.1385/1-59259-890-0:571.

Greenough, L., Kelman, Z., & Gardner, A. F. (2015). The roles of family B and D DNA polymerases in Thermococcus species 9°N Okazaki fragment maturation. *The Journal of biological chemistry*, *290*(20), 12514–12522. https://doi.org/10.1074/jbc.M115.638130

Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from Thermococcus

gorgonarius. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7), 3600–3605. https://doi.org/10.1073/pnas.96.7.3600

Jancarik, J. A. K. S., & Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of applied crystallography*, *24*(4), 409-411.

Krauss, I. R., Merlino, A., Vergara, A., & Sica, F. (2013). An overview of biological macromolecule crystallization. *International journal of molecular sciences*, *14*(6), 11643–11691. https://doi.org/10.3390/ijms140611643

Kuznetsova, A. A., Fedorova, O. S., & Kuznetsov, N. A. (2022). Structural and Molecular Kinetic Features of Activities of DNA Polymerases. *International journal of molecular sciences*, *23*(12), 6373. https://doi.org/10.3390/ijms23126373

Levy, Y., & Onuchic, J. N. (2004). Water and proteins: a love-hate relationship. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3325–3326. https://doi.org/10.1073/pnas.0400157101

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta crystallographica. Section D, Structural biology*, *75*(Pt 10), 861–877. https://doi.org/10.1107/S2059798319011471

Lyu, Z., & Whitman, W. B. (2017). Evolution of the archaeal and mammalian information processing systems: towards an archaeal model for human disease. *Cellular and*

molecular life sciences : CMLS, *74*(2), 183–212. https://doi.org/10.1007/s00018-016-2286-y

Makarova, K. S., Krupovic, M., & Koonin, E. V. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Frontiers in microbiology*, *5*, 354. https://doi.org/10.3389/fmicb.2014.00354

Marsic, D., Flaman, J. M., & Ng, J. D. (2008). New DNA polymerase from the hyperthermophilic marine archaeon Thermococcus thioreducens. *Extremophiles : life under extreme conditions*, *12*(6), 775–788. https://doi.org/10.1007/s00792-008-0181-7

McPherson, A. (1999). *Crystallization of Biological Macromolecules* (p. 209). Cold Spring Harbor Laboratory Press.

McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta crystallographica. Section F, Structural biology communications*, *70*(Pt 11), 1445–1467. https://doi.org/10.1107/S2053230X14019670

McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications*, *70*(Pt 1), 2–20. https://doi.org/10.1107/S2053230X13033141

McPherson, A., Malkin, A. J., & Kuznetsov, Y. G. (1995). The science of macromolecular crystallization. *Structure (London, England : 1993)*, *3*(8), 759–768. https://doi.org/10.1016/s0969-2126(01)00211-8

Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta crystallographica. Section D, Biological crystallography*, *62*(Pt 8), 859–866. https://doi.org/10.1107/S0907444906019949

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 307–326. https://doi.org/10.1016/s0076-6879(97)76066-x

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, 46(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Xie, N. Z., Du, Q. S., Li, J. X., & Huang, R. B. (2015). Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design. *PloS one*, *10*(9), e0137113. https://doi.org/10.1371/journal.pone.0137113

Zhou, H. X., & Pang, X. (2018). Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical reviews*, *118*(4), 1691–1741. https://doi.org/10.1021/acs.chemrev.7b00305

Appendix A.

The Amino Acid sequence for TthiPolB:

```
        10         20         30         40         50         60
MILDADYITE DGKPVVRIFK KENGEFKIEY DREFEPYIYA LLRDDSAIEE IKKITADRHG

        70         80         90        100        110        120
KVVKVKRAEK VQKKFLGRPI EVWKLYFTHP QDVPAIRDEI RKHPAVVDIY EYDIPFAKRY

       130        140        150        160        170        180
LIDKGLIPME GDEELKMLAF DIETLYHEGE EFGTGPILMI SYADEDGARV ITWKKIDLPY

       190        200        210        220        230        240
VDVVSTEKEM IKRFLKVVKE KDPDVLITYN GDNFDFAYLK KRCEKLGIKF TLGRDGSEPK

       250        260        270        280        290        300
IQRMGDRFAV EVKGRIHFDL YPLIRRTINL PTYTLEAVYE AVFGKPKEKV YAEEIALAWE

       310        320        330        340        350        360
SGEGLERVAR YSMEDAKVTF ELGREFFPME AQLSRLIGQS LWDVSRSSTG NLVEWFLLRK

       370        380        390        400        410        420
AYERNELAPN KPNERELARR RGGYAGGYVK EPERGLWDNI VYLDFRSLYP SIIITHNVSP

       430        440        450        460        470        480
DTLNREGCKE YDRAPQVGHK FCKDVPGFIP SLLGSLLDER QKIKRKMKAT IDPIEKKLLD

       490        500        510        520        530        540
YRQRAIKILA NSYYGYYGYA RARWYCRECA ESVTAWGREY IEMAIRELEE KFGFKVLYAD

       550        560        570        580        590        600
TDGLHATIPG ADAETVKKKA MEFLKYINPK LPGLLELEYE GFYARGFFVT KKKYAVIDEE

       610        620        630        640        650        660
GKITTRGLEI VRRDWSEIAK ETQARVLEAI LRHGDVEEAV RIVKEVTEKL SKYEVPPEKL

       670        680        690        700        710        720
VIHEQITREL KDYRATGPHV AIAKRLAKRG IKIRPGTVIS YIVLKGSGRI GDRAIPFDEF

       730        740        750        760        770
DPTKHKYDAE YYIENQVLPA VERVLKAFGY RKDDLRYQKT RQVGLGAWLK VKKR
```

List of parameters and amino acid composition based on the TthiPolB sequence computed by the ProtParam Expasy server tool (Gasteiger *et al*., 2005):

Molecular weight:	89.97 kDa
Theoretical pI:	8.79
Extinction Coefficient:	122300
Abs 0.1% (=1 g/l):	1.359
Total negatively charged residues (Asp + Glu):	127
Total positively charged residues (Arg + Lys):	137
Aliphatic index:	88.44
Grand average of hydropathicity:	-0.501
Instability index:	36.13

Ala (A)	54	7.0 %
Arg (R)	61	7.9%
Asn (N)	14	1.8%
Asp (D)	48	6.2%
Cys (C)	5	0.6%
Gln (Q)	13	1.7%
Glu (E)	79	10.2%
Gly (G)	52	6.7%
His (H)	12	1.6%
Ile (I)	59	7.6%
Leu (L)	64	8.3%
Lys (K)	76	9.8%
Met (M)	11	1.4%
Phe (F)	31	4.0%
Pro (P)	35	4.5%
Ser (S)	22	2.8%
Thr (T)	31	4.0%
Trp (W)	10	1.3%
Tyr (Y)	45	5.8%
Val (V)	52	6.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%