BIOCHEMICAL CHARACTERIZATION OF THE MALARIA PARASITE PLASMODIUM FALCIPARUM CLPB HOMOLOGUE PFCLPB1 LOCALIZED TO THE APICOPLAST

by

FABRICE NGANSOP

B.S., Wichita State University, Kansas, USA, 2008

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Michal Zolkiewski

Abstract

ClpB is a molecular chaperone that is essential for infectivity and pathogen survival in a host. It belongs to the AAA+ protein family, which cooperates with the DnaK chaperone system to reactivate aggregated proteins. In this study, we purified and then studied the biochemical properties of the apicoplast targeted ClpB isoform from the malaria parasite *Plasmodium falciparum*: PfClpB1. *Plasmodium falciparum* is the parasite responsible for the most severe form of malaria. In contrast to the parasitophorous vacuole targeted PfClpB2 from *Plasmodium falciparum* which contains all characteristic AAA+ sequence motifs, PfClpB1 also includes a 52-residue long non-conserved insert in the middle domain. The ATPase activity study shows that PfClpB1 hydrolyzes ATP in presence of Poly-lysine and α-casein. Similar to most AAA+ ATPases, addition of ATP induces hexamer formation in PfClpB1. Lastly, PfClpB1 reactivates aggregated firefly luciferase. However, PfClpB1 is unable to efficiently reactivated luciferase in the presence of the E. coli DnaK chaperone system or human Hsp70 and Hsp40 (Hdj1). This can be explained by the extra middle domain sequence of PfClpB1. Our data may suggest that PfClpB1 activity is essential for *Plasmodium falciparum* survival by preserving the activity of apicoplast proteins.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAA+: ATPases Associated with a diverse cellular Activities

ATP: Adenosine Triphosphate

ClpB: Caseinolytic protease B

D1: First AAA+ domain

D2: Second AAA+ domain

DNA: Deoxyribonucleic acid

DTT: Dithiothreitol

EDTA: Ethylenediamine-tetra-acetic acid

HIS-tagged: Histine-tagged

HPLC: High Performance Liquid Chromatography

Hsp: Heat shock protein

IPTG: Isopropyl β-D-1 Thiogalactopyranoside

KCL: Potassium chloride

KDa: Kilo-Dalton

MD: Middle Domain

$MgCl_2$: Magnesium chloride

MS: mass spectrometry

$(NH_4)_2S_2O_8$: Ammonium persulfate

Ni-Nta: Nickel-Nitrilotriacetic acid

PCR: Polymerase chain reaction

PDB: Protein Data Bank

PfClpB1: Plasmodium Falciparum ClpB1

PfClpB2: Plasmodium Falciparum ClpB2

Pi: Inorganic phosphate

RBC: Red blood cells

RT: Room temperature

SDS: Sodium dodecylsulphate

SDS-page: Sodium Dodecylsulfate Poly-Acrylamide Gel Electrophoresis

SP: Signal peptide

TCEP: Tris (2-carboxyethyl) phosphine

TP: Transit peptide

Tris: Tris (hydroxymethyl) aminomethane

Acknowledgements

I would like to express my deep gratitude to Professor Michal Zolkiewski my research advisor, for his patient guidance, enthusiastic encouragement and useful critiques of my thesis project. I would also like to thank my close friends for their advice and assistance in keeping my progress on schedule. My grateful thanks are also extended to Dr. Hui Li for her help in the cloning step, and to my fellow graduate students and particularly Karanbir Aulakh for his guidance and support. I would also like to extend my thanks to Dr. Anna Zolkiewska and the technicians of the plant pathology department for their help in offering me the resources to complete my research.

Finally, I wish to thank my parents for their support and encouragement throughout my study.

Dedication

This thesis is dedicated to my dear and beautiful daughter, Abigail Ngansop, my parents and my family who support me toward my dreams.

Chapter 1 - Introduction

1.1 Malaria

The history of malaria predates humanity. This conclusion can be inferred from the discovery of mosquitos in Baltic amber stones dating from 40 to 60 million years. Malaria comes from the Italian word “malaria” which means bad air. The etymology of the word is related to the fact that the ancient Romans though the disease was caused by the foul air coming from the swamps [1]. Reference to the ravage caused by malaria has been recorded through history [2]. During World War II and the Vietnam War most casualties in the US army was mainly caused by malaria rather than bullets.

According to the Centers for Disease Control and Prevention (CDC) [3], malaria is caused by the genus *Plasmodium* which spends parts of its life cycle between two hosts: humans and mosquitos. *Plasmodium falciparum* is responsible for most deaths in Africa. *Plasmodium vivax* is more geographically widespread and the symptoms are less severe. *Plasmodium malariae* persists in the human blood for a long period, possibly a decade [4]. From the genus of plasmodium, *Plasmodium falciparum* is the most virulent [5]. Malaria is an endemic disease in Asia, Africa, India, South and Latin America (Figure 1.1). People in non-endemic area often get the disease during travel.

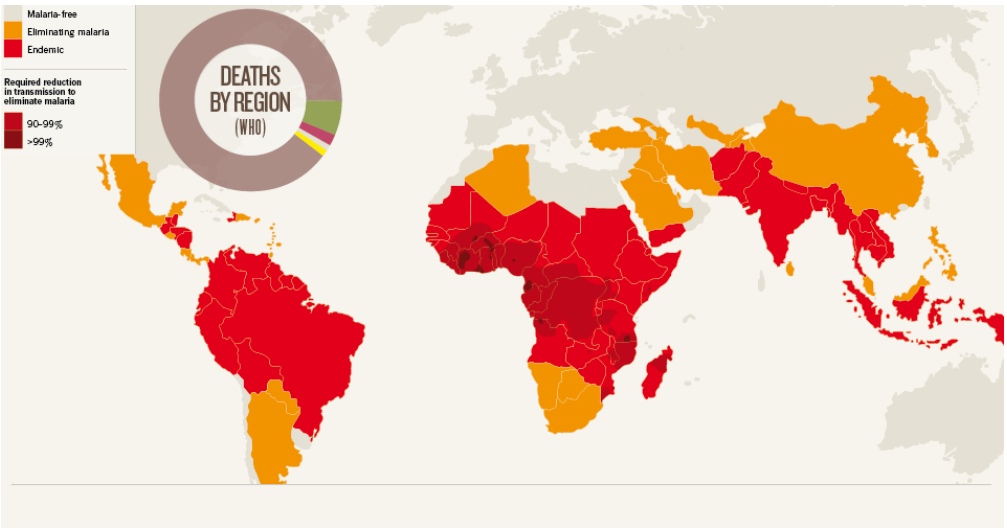


Figure 1.1 Malaria endemic countries

The map represents the spread of malaria infection in the world. The more predominant endemic areas are Africa, South America, Asia and India. (http://www.nature.com/nature/journal/v484/n7395_supp/interactive/malaria.html)

1.1.1 Malaria disease

The most vulnerable groups to malaria are children under 5 year's old, pregnant women, and people living with HIV [6]. According to the World Health Organization in 2008 [7], 89 % of the deaths are from Africa, 5% from Southeast Asia, 6% from Eastern Mediterranean. The main symptoms are: fever, sweating, headache, convulsion, chills, and muscle pain. Complication can arise such as hemolytic anemia: destruction of blood cells, liver and kidney failure. Sometimes rupture of the spleen can occur.

For thousands of years, malaria has been treated with herbal remedies. Quinine was the first effective treatment, which was quickly replaced by chloroquinine which has fewer side effects [8]. Today, the choice of medication depends on where you live, because of the development of antimalarial drug resistance in *Plasmodium* [9]. Novel approaches need to be developed to fight the disease, since antimalarial drug resistance is becoming a serious issue. The genome of *P. falciparum* clone 3D7 was the first to be sequenced and annotated in 2002. The nuclear genome is A+T rich, with an overall composition of 81% of A+T [10]. The components

of some anabolic pathways for the synthesis of lipid and iron sulfur complexes seem to be localized to the apicoplast [11, 12]. The function of apicoplast will be discussed later.

1.1.2 Life cycle of *Plasmodium falciparum*

The complexity of the *Plasmodium* life cycle resides in the fact that it requires two hosts: female mosquito and human host (Figure 1.2). During the **liver stages**, anopheles has a blood meal and injects sporozoites into the human bloodstream until they reach the liver cell. In the infected liver cells, sporozoites mature and develop into schizonts, which will burst out of the cells as merozoites. At the next stage: the **human blood stages**, the released *merozoites* infect red blood cells. In this stage, we have two scenarios at play. After infection of red blood cells, the mature merozoites can divide and infect some new red blood cells or they can differentiate into gametocytes. The anophele will ingest the gametocytes during the **mosquito stages**. Female and male gametocytes will then form an oocyst, which will mature to release new sporozoites that will infect the liver and the plasmodium life cycle will repeat indefinitely. The human blood stage is responsible for malaria. The internment and synchronous lysis of the red blood cells is responsible for the fever during malaria [13]. In addition, *Plasmodium falciparum* is a deadly disease because of its capability to cause severe anemia. Infected red bloods undergo morphological changes and can block blood flow to vital organs such as the brain: leading to cerebral malaria [86].

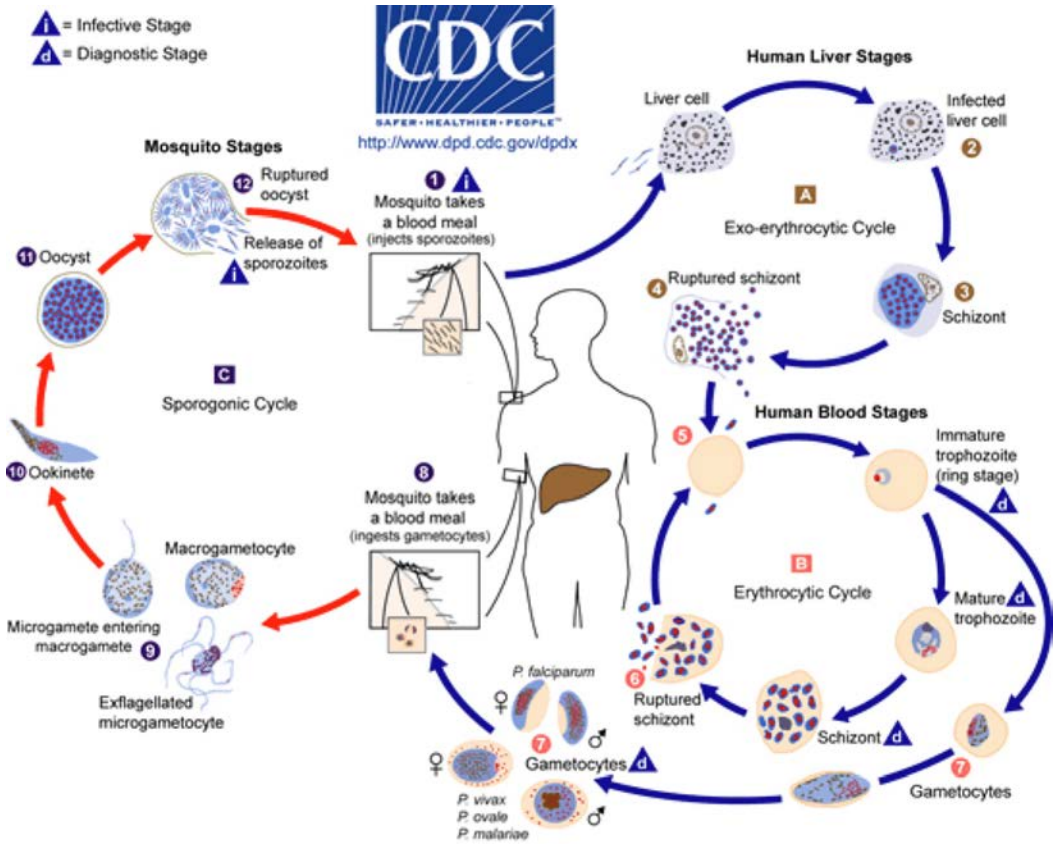


Figure 1.2 Life cycle of *Plasmodium falciparum*

The life cycle is divided into 3 stages: Mosquito stages, human liver stages, and human blood stages. *Plasmodium falciparum* requires human and mosquisto host to complete its full life-cycle. (1): Human host infection, (2): Invasion liver cells, (4): Rupture of infected liver cells, (5): Infection of red blood cells, (6): Division and multiplication, (7): Sexual forms cycles, (8): Transfer gametocytes into mosquito, (11): Formation oocyst, (12): release of sporozoites. (http://www.dpd.cdc.gov/dpdx/HTML/malaria.htm)

1.2 Caseinolytic peptidase B (ClpB)

Caseinolytic peptitase B (ClpB) is heat shock protein that is involved in the disaggregation of mis-folded protein in several organisms [14, 15, 16]. ClpB is an ATP dependent molecular chaperone that belongs to the AAA+ superfamily of ATPase [17]. It is found in eukaryotic and prokaryotic cells and cooperates with the KJE chaperone system (DnaK, DnaJ, and GrpE) to disaggregate and refold proteins [18]. Many ATPases such as ClpA, ClpX,

ClpE, ClpP and ClpL form a complex with a peptidase unit, but ClpB does not associate with a peptidase unit and does not participate in protein degradation [19].

1.2.1 AAA+ superfamily protein ATPase

The hallmark of the AAA+ family of protein (ATPases Associated with diverse cellular Activities) is a 200-250 amino-acid ATP binding domain that contains different conserved motifs: Walker A, Walker B, N-linker, the second region of homology, the pore region, Sensor 1 and 2 [20]. AAA+ protein are involved in different cellular processes which range from protein disaggregation, degradation, DNA replication and thermo-tolerance. AAA+ proteins function as oligomers, but the extent of the mechanism of nucleotide binding and hydrolysis are still ill defined. It is usually assumed that Walker A and B are involved in ATP binding and hydrolysis [21]. AAA+ domain consists of an N-terminal nucleotide-binding subdomain and a smaller C-terminal helical subdomain. The αβα fold with five parallel β sheets characterizes the N-terminal subdomain. The α-helix of the C-terminal subdomain is one of main features of AAA + proteins that differentiate it from other nucleotide-binding proteins [22]. All AAA+ proteins are classified depending of the number of conserved AAA+ domains (nucleotide binding domains). A Class I has two AAA+ domains, when a class II has one AAA+ domain (Figure 1.3) [23].

N-terminal	AAA+ domain	Middle-domain	AAA+ domain

Class I: ClpB

N-Terminal	AAA+ domain

Class II: ClpX

Figure 1.3 The domain structures of the Clp ATPases

1.2.2 Structure and function of ClpB domains

ClpB is a molecular chaperone with multiple domains. Like other AAA-type ATPases, ClpB contains two nucleotide binding domains (NBDs), including a Walker A and B consensus sequences, an arginine finger motif, the pore loop and the sensor 1 and 2 [24]. To achieve full disaggregation of stress-damaged protein, ClpB cooperates with the KJE chaperone system (DnaK, DnaJ, GrpE). Walker A and B coordinate ATP bonding and hydrolysis. The arginine finger is necessary for nucleotide hydrolysis and oligomerization. Sensor 1 and 2 also assist with

nucleotide binding and hydrolysis. Finally, the pore loop is required for substrate binding, translocation and disaggregation activity [25]. The Middle-domain is speculated to facilitate protein disaggregation [26, 27]. Despite major advance in our understanding of the middle-domain, some of its functions still remain elusive

Protein aggregates binds to the ClpB hexamers in the presence of ATP. Before translocation is initiated, the KJE (DnaK, DnaJ and GrpE) system must interact with the Middle domain of ClpB to commit to the interaction between the pore loop of NBD1 and aggregates [28]. Then ATP hydrolysis is followed by translocation of the aggregates through the central channel of ClpB. The mechanism was discovered by constructing a ClpB mutant: BAP that could bind to ClpP [29]. If aggregates are hard to unfold, ClpB hexamers will try a different position to pull the aggregates through the central channel.

The N-terminal domain of ClpB is more mobile that the rest of the protein (Figure 1.5). The N-terminal is required for protein disaggregation of some substrates *in vitro* and *in vivo* [30, 31]. Furthermore mutation or deletion of the N-terminal domain blocked activation of ATPase of ClpB by α-casein [32, 33]. It has been speculated that the mobility of ClpB allows it to interact efficiently with the substrate and Dnak. The N-terminal domain of ClpB may play a role in substrate specificity. The increase mobility of the N-terminal domain can result in increased chaperone activity. In short, the N-terminal domain is required for aggregate binding and is also important for protein disaggregation.

The middle domain is less mobile than the N-terminal domain. It is located between two nucleotide binding domains (Figure 1.4). The presence of the middle domains helps to stabilize the oligomers [34]. The function of the middle domain is still unknown. One model proposes that the middle domain facilitates the interaction between KJE and ClpB [35, 36]. It was observed that lack of cooperation between the Middle domain and the KJE system stop the reactivation of aggregates [37]. The most recent studies confirm that the middle domain is responsible for the species specific cooperation with the Hsp 70 system [38, 39].

The C-terminal domain of ClpB is made principally of α-helices and linked to the second nucleotide binding domain. The C-terminal is also required for oligomerization [40, 41, and 42]. It is being proposed that the C-terminal domain binds to some substrates and it is therefore called the sensor domains [43].

The two nucleotide binding domains NBD1 & NBD2 are the essential part of the ClpB machine. NBD1 and NBD2 undergo conformational change during translocation to allow the passage of unfolded polypeptides through the central channel [44]. In a recent study, wild type ClpB was mixed with an inactive mutant, ClpB chaperone activity was blocked in the presence of co-chaperone with some substrates, indicating the requirement of cooperation between subunits of ClpB hexamer [45].

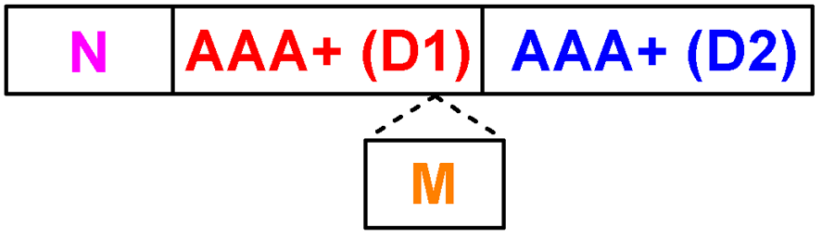


Figure 1.4 ClpB domains

The figure shows the main domain of ClpB. The N-terminus, the Middle domain and the two AAA+ domains.

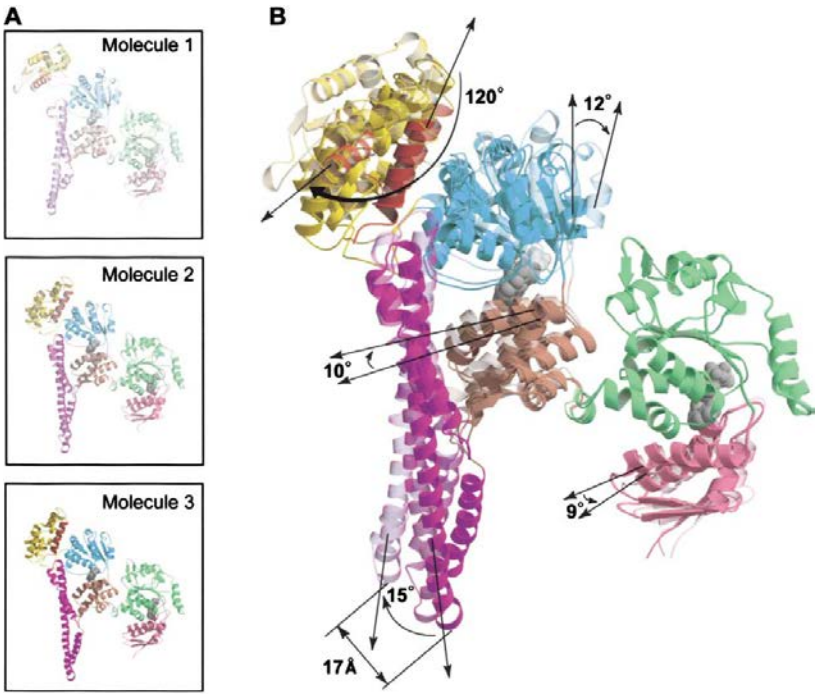


Figure 1.5 The structure and mobility of *Thermus thermopholis* ClpB domains

The N-terminal domain is shown in yellow color. The Middle domain is in purple color. The blue and green color represents the nucleotide binding domains. (A) The different conformation of TClpB. (B) Degree of mobility of each domain with TClpB molecule 1, 2, 3 superimposed through the atoms of NBD2. The higher mobility of the N-terminal domain allows it to react with substrate in different conformation during binding. (Cell, Vol. 115, 229–240, October 17, 2003, Copyright 2003 by Cell Press The Structure of ClpB: A Molecular Chaperone that Rescues Proteins from an Aggregated State)

1.2.3 Cell stress and protein aggregation

Classical “molecular chaperones” such as Hsp40, Hsp60, Hsp70, Hsp90 and Hsp100 assist in promoting protein degradation after stress or injury and facilitate protein folding (Table 1.1), but they cannot mediate protein disaggregation. Heat-shock response can be caused by circumstances as diverse as viral and bacterial infection or exposure to transition heavy metals and oxidants [46]. The transcription of heat-shock genes is regulated by an array of stress conditions. First of all, there are environmental stresses such as toxic chemical, drugs, heat shock. Next, there are cell growth and cycle, differentiation and activation of cells by some oncogenes. Finally, there are pathophysiological stresses such as fever, inflammation and infection [46]. However, ClpB “has the remarkable capacity to rescue proteins from aggregated

state by mediating disaggregation of stress-damaged proteins" [47]. The DnaK/Hsp 70 chaperone system is required for full recovery of active protein [48] ClpB uses ATP-driven conformational change to mediate the disaggregation of protein (Figure 1.6). Stress-damaged proteins arising from partially folded intermediates are inclusion bodies, thermal aggregates and refolding aggregates. In vitro, protein aggregation is responsible for protein instability, in vivo protein aggregation promote formation of disordered aggregates [49].

Table 1.1 Heat shock proteins in eukaryotes.

(Whitley, D., Goldberg, S., and Jordan, N. (1998). "Heat shock proteins: A review of the molecular chaperone." J. Vasc Surg. 29:748-751)

	Molecular size (kDa)	Location	Major functions
Hsp40	40	Cytosol/nucleus/ Mitochondria	Stabilization of mis-folded proteins, co-chaperone for Hsp70
Hsp60	60	Cytosol/nucleus	Protein folding (limited substrates in eukaryotic cytoplasm)
Hsp70	70	Cytosol/nucleus	Protein folding, membrane transport of proteins.
Hsp90	90	Cytosol/nucleus	Regulatory interaction with signaling proteins, stabilization of misfolded proteins
Hsp100	100	Cytosol/nucleus	Protein disaggregation, thermotolerance

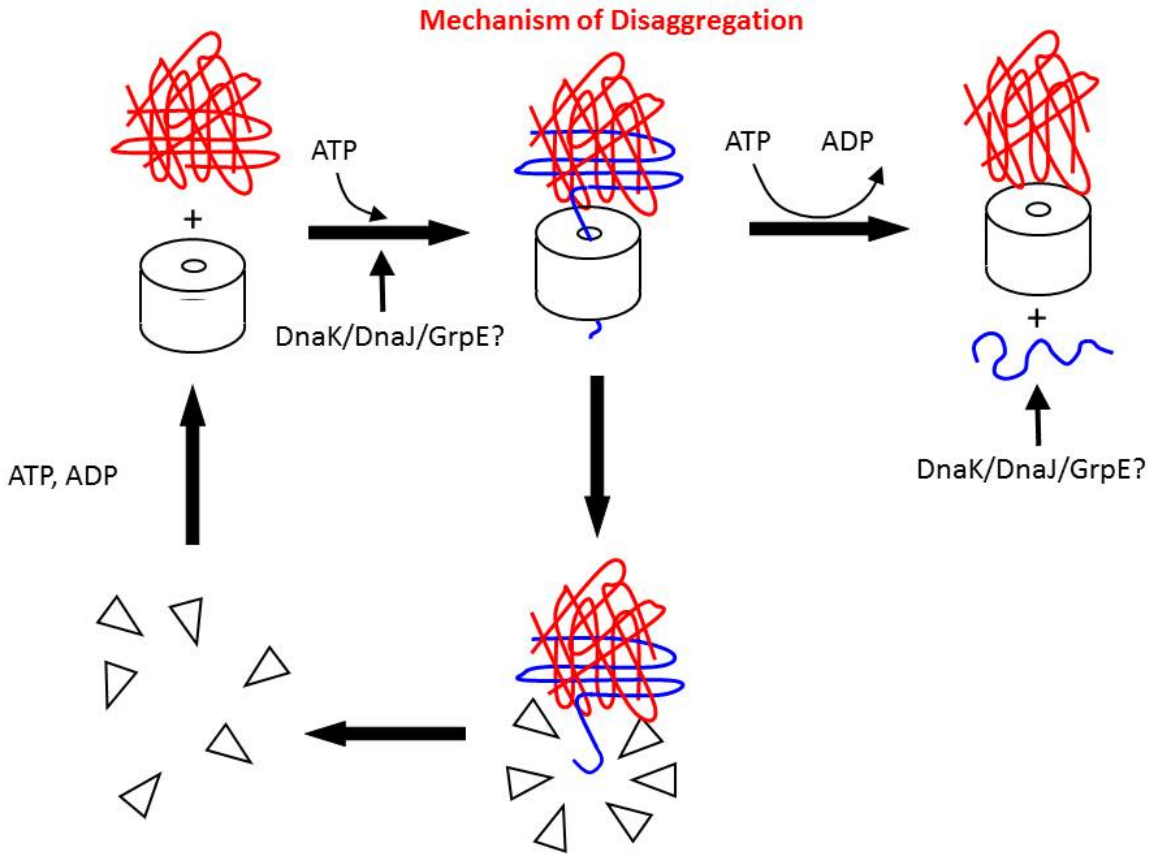


Figure 1.6 Mechanism of protein disaggregation mediated by ClpB

The Cylinders with a central channel represent the ClpB hexamer. ATP hydrolysis is required to pull the aggregates through the channel. The ClpB hexamer and the KJE chaperone system facilitate the insertion of aggregates into the central channel.

1.2.4 Biochemical properties of ClpB

The crystal structure of *Thermus thermophilus* ClpB has been solved and provides valuable tools to link structure to function (Figure 1.5) [50]. The ClpB monomer is about 95 KDa with multiple domains. ClpB forms hexamers that are stabilized by ATP binding and hydrolysis [51]. Full reactivation of insoluble aggregated proteins by ClpB requires the assistance of the chaperone helpers: DnaK, DnaJ and GrpE [85]. Substantial evidence of ClpB involvement in pathogenic microorganism virulence has been reported.

1.3 ClpB protein needed for infectivity

Like many other pathogenic organisms, the malaria parasite *Plasmodium falciparum* is subjected to hostile environment during infection in the human host. The environmental stressors are temperature, pH, reactive nitrogen, oxidative stress, salt concentration. Stressful condition acts as a stimulus to induce change in gene expression and transcription. Although the molecular mechanism for survival in the host cell is not clearly understood, some of the genes that are necessary for multiplication and expansion of the disease are well known. During infection, most bacteria up-regulate the expression of heat shock protein to cope with challenging cellular stress in their host [52]. One of those proteins is ClpB which helps to disaggregate mis-folded intracellular proteins because of stressful condition. ClpB chaperone is highly conserved across prokaryotes and eukaryotes. ClpB is involved in stress response in various bacteria. For instance in *E. coli*, ClpB is required for thermo-tolerance with the assistance of DnaK, DnaJ and GrpE. The heat shock protein or Hsp can vary from organism to organism. In *Drosophila melanogaster*, Hsp70 provides stress protection. Previous studies show that classical heat shock proteins are necessary for the pathogenicity of *mycobaxterium tuberculosis* [53, 54, 55]. Heat shock is known to confer thermo-tolerance to bacteria. Hsp 100 (ClpB) in *Leishmania donovani* is responsible for the differentiation from promastigotes (insect stages) to amastigotes (mammalian stages). Its absence shifts the balance toward promastigotes. So, The ClpB gene is required for full amastigote development and during first stages of a mammalian infection [56]. Loss of Hsp100 in *leishmania major* retards lesion development in infected mice, because Hsp 100 is important for thermotolerance during the mammalian stage, when cellular stress is intense. Absence of Hsp100 affects infectivity and impaired asmatigote development [57]. In *Staphylococcus aureus* and *P.gingivalis*, ClpB is associated with intracellular survival and multiplication [58, 59]. The lack of ClpB in *L. monocytogenes* and *P. gingivalis* attenuated virulence in murine model of infection [60, 61]. Finally, Inactivation of ClpB in *Leptospira interrogans* is involved in general stress response and directly or indirectly reduces virulence [62]. Thus, the data clearly demonstrate that ClpB is required for the infectivity and virulence of the disease in different organisms. Since ClpB does not exist in humans, it might be a good target for the development of therapeutic drug. For instance, the development of antibodies specific for *Flavobacterium psychrophilum* ClpB may be useful for protective immunity from bacterial disease in coldwater [63].

1.4 Apicoplast organelle

1.4.1 Origin and function of the apicoplast

In 1975, Araaxie Kilejian found the first evidences that ultimately confirm the existence of a plastid in *Plasmodium*. Under the microscope, she saw an extra-chromosomal DNA molecule in *Plasmodium* [64]. The apicoplast is derived from two serial endosymbiotic events (Figure 1.7). The primary endosymbiotic event occurred between a nucleated biciliate phagotroph and a photosynthetic cyanobacterium bounded by two membranes. This process gave rise to several organisms: red algae, green algae and land plants. The second endosymbiosis involved the engulfment and retention of a red alga by a second phagotroph giving rise to *P. falciparum* [65, 66]. The process of a phagotrophic host cell engulfing, retaining and ultimately enslaving another cell is called endosymbiont. Thus, apicoplast is an endosymbiont: an organism living in another living organism (*Plasmodium falciparum*). The function of the apicoplast was discovered by identifying genes necessary for fatty acid biosynthesis [67]. Soon after, it was also discover that apicoplast has a pathway to synthesize isopentenyl diphosphate, a precursor of isoprenoids [68]. Finally, the iron sulphur complexes were discovered through data mining [69]. Consequently, the apicoplast supplies carbon, energy and reducing power in a similar way to algal plastid. An understanding of the machinery involved in translocating proteins across the four membranes surrounding the apicoplast is coming to light.

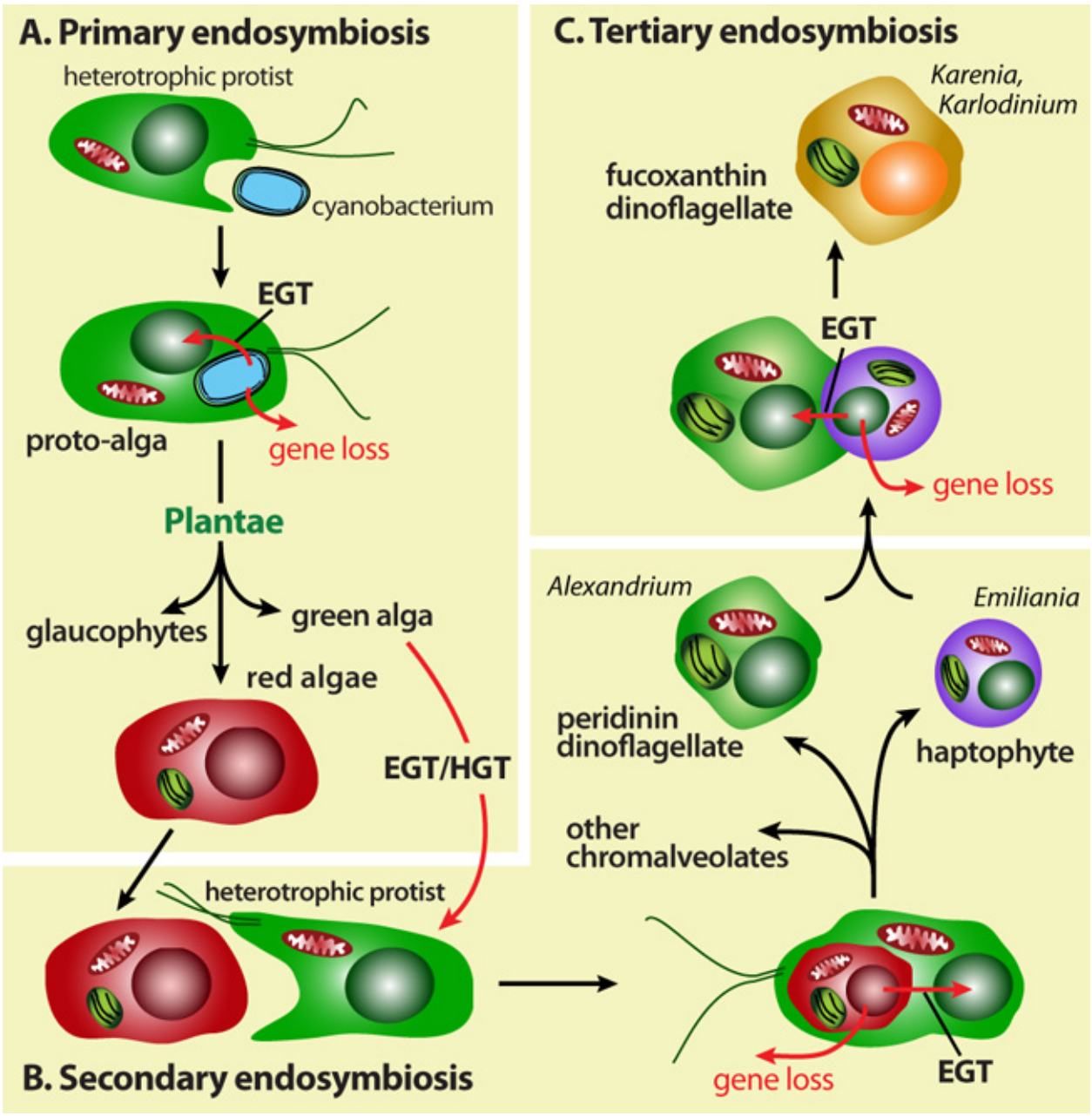


Figure 1.7 The concept of endosymbiosis

The chromalveolate hypothesis can explain some endosymbiotic events in dinoflagellates. (Chan, C. X. & Bhattacharya, D. (2010) The Origin of Plastids. Nature Education 3(9):84)

1.4.2 Apicoplast targeting sequence

Bipartite apicoplast-targeting leaders start with a signal peptide [70]. Transit peptides are adjacent to the signal peptides and exhibit a surplus of basic over acidic residues, and they tend to be enriched for Asparagines and/or Lysines (Figure 1.8). Apicoplast proteins use the co-

translational translocation pathway. They are imported into the ER, and their import is directed by an N-terminal signal sequence or an internal signal sequence. The transit peptide domain of luminal proteins is bound at the ER by a receptor protein that directs these proteins into vesicles. These vesicles are directly exported to the apicoplast where they fuse with its outer membrane. The specific mechanisms that direct the distribution of the proteins to the organelle are still unknown. The apicoplast targeting signal is cleaved during transport to the apicoplast.

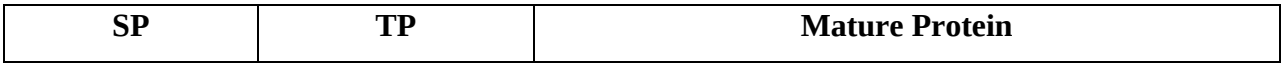


Figure 1.8 Schematic representation of a typical apicoplast-targeted protein
The mature protein remains after cleavage of the signal and transit peptide.
SP: signal peptide, TP: transit peptide.

1.4.3 Apicoplast needed for infectivity

The *P. falciparum* apicoplast is required for survival of the parasite in both its erythrocytic [71] and liver stages [72, 73]. The apicoplast is a non-photosynthetic organelle that is involved in several biochemical pathways, such as biosynthesis of fatty acids, iron-sulfur cluster and isoprenoids [74]. Two different lines of evidence demonstrated that apicoplasts is required for parasite survival. First, any chemical affecting apicoplast metabolisms resulted in the death of parasite. Second, the inability to replicate the apicoplast also resulted in death. Surprisingly, the parasite is able to survive with a damaged apicoplast, but only die in the next generation. This is called “delayed death”. One probable hypothesis is that apicoplast secretes a molecule that is needed for the infection. Plant scientists are trying to develop non-toxic herbicides that may act upon the apicoplast, by using tools such as bioinformatics and experimental approaches. Because apicoplasts share similitude with chloroplasts and prokaryotes, they strike as an attractive target for known antibiotics. The non-mammalian characteristics of the apicoplast make it a perfect target for anti-malarial drugs [75].

1.5 *Plasmodium Falciparum* ClpB homologue (PfClpB1)

1.5.1 Up regulation of PfClpB1 during malaria

The ClpB protein plays an important role in the cell homeostasis. During infection, *Plasmodium falciparum* navigate in hostile environment and it is subjected to different stressors,

which can stress-damage its proteins. According to the PlasmoDB database (plasmodb.org), PfClpB1 and PfClpB2 are ClpB homologues in *Plasmodium falciparum* that are up-regulated during infection.

1.5.2 Structure and biochemical properties

The parasite is known to have two ClpB homologues: PfClpB1 and PfClpB2. PfClpB2 (hsp101) is localized to the parasitophorous vacuole and pfClpB1 can be found in the apicoplast [76], a non-photosynthetic organelle that accommodates several important metabolic pathways [77] and is necessary for plasmodium survival [78]. All PfClpB ATPase have the predicted two ATP nucleotide-binding domains with the conserved Walker A and Walker B. Different sensor motif can be recognized in the AAA+ domains of PfClpB ATPases. PfClpB1 (PF08_0063) is a 123 KDa protein whereas pfClpB2 (PF11_0175) is a 103 KDa protein.

1.5.3 PfClpB1 and PfClpB2 localization

PfClpB1 is localized to the apicoplast, however PfClpB2 also known as Hsp101 is found in the parasitophous vacuole and it is part of the PTEX complex [79]. During infection, to gain access to the host cell cytosol, plasmodium must export its proteins through the parasitophous vacuole membrane. Trafficking of malaria protein across the parasitophorous vacuole membrane, requires a Plasmodium export element known as PEXEL [80]. Identification of the PEXEL motif has improved our understanding of plasmodium protein export. Some Algorithms predict that 5 to 8% of the plasmodium genome should be exported [81]. PEXEL motif recognition occurs in the ER, where the fate of the proteins destined for trafficking across the parasitophous vacuole membrane is decided. The plasmodium translocon of exported proteins (PTEX) was discovered by an array of strategy combining proteomic, bioinformatics, biochemical and genetic methodologies. PTEX is made of five components: PfClpB2 (HSP101), PTEX150, EXP2, PTEX88 and TRX2. The complex interacts with native exported protein to move them across the parasitophous vacuole membrane. HSP101 is predicted to oligomerize as a hexamer and hydrolyse ATP, which helps to unfold *Plasmodium* proteins [82]. EXP2 is predicted to a pore-forming integral membrane [83]. Finally, TRX2 is believed to play a regulatory role, while PTEX150 and PTEX88 must probably help to recognize the malaria protein arriving in the parasitophous vacuole or feeding the protein into HSP101 [9]. PfClpB1 is an apicoplast protein with a longer N-terminal domain compared to PfClpB2 (Table 1.2). It has a mass of 123 kDa.

PfClpB1 belongs to Class I AAA+ ATPase with an extra sequence in the middle domain. This characteristic makes PfClpB1 a good candidate for further study.

Table 1.2 The Clp proteins of P. falciparum

(The Clp Chaperones and Proteases of the Human Malaria Parasite Plasmodium falciparum, Journal of Molecular Biology, Volume 404, Issue 3, 3 December 2010, Pages 456-477)

Name	Mass (kDa)[a]	Gene ID[b]	Gene location	Protein type	Localization[c]
PfClpP	43	PFC0310c	Nucleus	Clp protease	Apicoplast
PfClpR	28	PF14_0348	Nucleus	Inactive Clp protease	Apicoplast[d]
PfClpB1	123	PF08_0063	Nucleus	Clp ATPase	Apicoplast
PfClpB2	103	PF11_0175	Nucleus	Clp ATPase	Parasitophorous vacuole[e]
PfClpC	156	PF14_0063	Nucleus	Clp ATPase	Apicoplast
PfClpM	91	PFC10_API0060	Apicoplast	Clp ATPase	Apicoplast

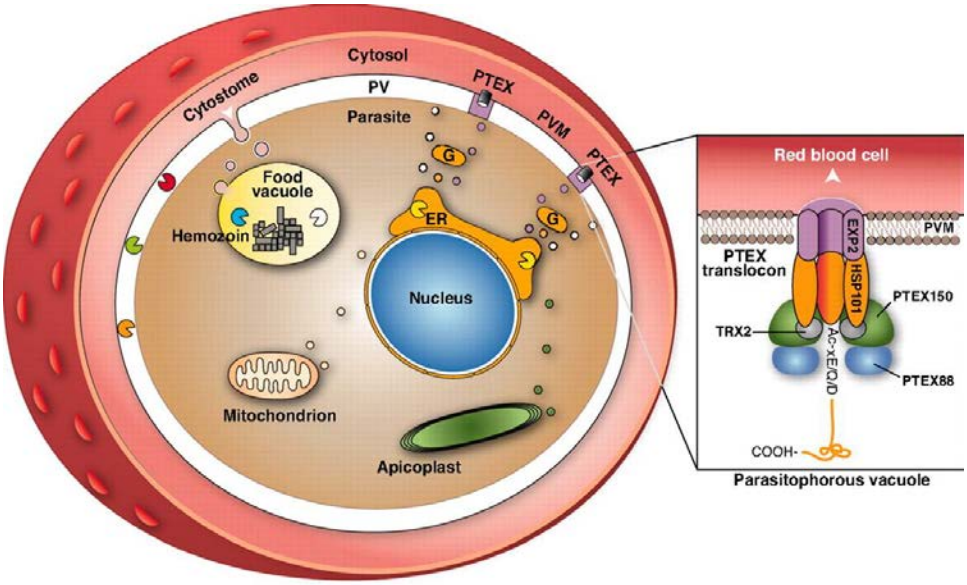


Figure 1.9 Putative model of PTEX

This putative model of PTEX shows that the complex is formed by HSP101 (PfClpB2), PTEX 150, PTEX88, EXP2. (That was then but this is now: malaria research in the time of an eradication agenda: Science 14 May 2010: Vol. 328 no. 5980 pp. 862-866)

1.6 Hypothesis and objectives

Previous studies localize PfClpB1 in the apicoplast of *Plasmodium falciparum*. The characterization of biochemical and structural properties of pfClpB1 have yet not been fully established. Most of what is known about PfClpB1 is the fact that it is a 123 kDa protein with an apicoplast targeting domain that directs the newly synthesized protein toward the apicoplast. We hypothesize that PfClpB1 is involved in the folding and refolding of newly synthesized and denatured malaria parasite proteins through its action as a ClpB homologue. During heat stress conditions, its chaperone activity is increased resulting in the survival of the malarial parasite. The main goal of this study is to produce enough soluble and fully active pfClpB1 protein in *E. coli* for thorough biochemical characterization. First, the similarity between pfClpB1 and *Thermus thermophilus* ClpB is achieved using bioinformatics tools. Then, we study the biochemical properties of PfClpB1 as a homologue of *E. coli* ClpB.

Chapter 2 - Bioinformatics analysis

2.1 Introduction

Bioinformatics is a field of science that merge biology, statistics, mathematics and computer science to generate useful biological knowledge. Bioinformatics is crucial in the field of genomics, proteomics and homology modeling. Homology modeling is a practical tool to understand the relationship between protein function and structure when an experimental model is not accessible. The prediction of functional residues and secondary structure is based on multiple sequence alignment and sequence similarity search. The function and structure prediction is maximized when the number of matches in the sequence alignment is increased.

For the bioinformatics analysis, the degree of conservation of the middle domain, C terminal and N terminal domains between *E.coli* ClpB and PfClpB1 was measured with Clustal omega tool (http://www.ebi.ac.uk/Tools/msa/clustalo/). Homology modeling tool I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used with *Thermus thermophilus* ClpB structure (PDB 1qvrA) to build the molecular model of PfClpB1. Pymol, the molecular visualization software, was used to view the predicted model. Codon analysis was carried out with the graphical codon usage analyser (http://gcua.schoedl.de/). Finally, SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/) predicted the location of a signal peptide cleavage site in the amino acid sequences of PfClpB1 and apicoplast targeting sequence analysis tool, PlasmoAP (http://v4-4.plasmodb.org/restricted/PlasmoAPcgi.shtml), helped to identify potential amino acid sequences that will be part of the apicoplast targeting signal.

The broad objective of this study was to apply bioinformatics tools, such as primary sequence alignments and homology modelling to identify structural features potentially of functional importance. Functional motifs and domains were identified by searching for conserved blocks within multiple sequence alignment of ClpB from plasmodium falciparum. The predicted three-dimensional structure was generated by Itasser computer program and visualized using pymol, allowing identification of domains and positions of important residues.

2.2 Prediction of apicoplast targeting sequence

The signal peptide cleavage sites in amino acid sequences were detected by SignalP V4.1 (http://www.cbs.dtu.dk/services/SignalP). The co-translational pathway is initiated after the signal peptide is recognized by the signal-recognition particle (SRP). That long stretch of amino acid that is recognized and will be cleaved by the signal peptidase. The apicoplast targeting sequence was detected with PlasmoAP (http://v4-4.plasmodb.org/restricted/PlasmoAPcgi.shtml) following these rule sets: apicoplast targeting sequence tend to be enriched for asparagines and/or lysines, exhibit a surplus of basic over acidic residues, and finally start with a signal peptide.

2.2.1 Prediction of signal peptide cleavage site

Prediction of the cleavage site was done by SignalP V4.1. The signal peptide is cleaved during post-translational processing by signal-peptidase. It was important that the signal peptide did not be expressed in *E.coli*, since it was not part of PfClpB1 protein after cleavage. The predicted protein sequence (Figure 2.1) was analyzed after specifying the exclusion of trans-membrane regions to SignalP. The first 70 amino acids were analyzed and the cleavage of signal peptide site was predicted to be between position 23 and 24(Figure 2.2).

*MVNSFFFCFVIIGLIYVWDITYSK**KAKIFFNKNDIFSIKNTHWDIYDKKKYFFIGNNHLK*
NEESFLPEVRKDYKSQIKEYKNSTNGIIYHNNKNRLSYTINDQVNYDNNMTSGINKKRKV
*KDSSIHMNNSYEKNRNKNKFALF*MSDEEYTINSDDYTEKAWEAISSLNKIGEKYDSAYVE
AEMLLLALLNDSPDGLAERILKESGIDTQLLVQEIDDYLKKQPKMPSGFGEQKILGRTLQ
TVLSTSKRLKKEFNDEYISIEHLLLSIISEDSKFTRPWLLKYNVNYEKVKKAVEKIRGKK
KVTSKTPEMTYQALEKYSRDLTALARAGKLDPVIGRDNEIRRAIQILSRRTKNNPILLGD
PGVGKTAIVEGLAIKIVQGDVPDSLKGRKLVSLDMSSLIAGAKYRGDFEERLKSILKEVQ
DAEGQVVMFIDEIHTVVGAGAVAEGALDAGNILKPMLARGELRCIGATTVSEYRQFIEKD
KALERRFQQILVEQPSVDETISILRGLKERYEVHHGVRILDSALVQAAVLSDRYISYRFL
PDKAIDLIDEAASNLKIQLSSKPIQLENIEKQLIQLEMEKISILGDKQKNLFNYSSVANT
HNNNNNSSISSNNSSSYGNAEETEATVDYTKSPNFLKKRINEKEIDRLKMIDRIMSELRK
EQRKILDSWSTEKSYVDNIRAIKERIDVVKIEIEKAERYFDLNRAAELRFETLPDLEKQL
KKAEENYLNDIPEKSRILKDEVTSEDIVNIVSMSTGIRLNKLLKSEKEKILNLENELHKQ
IIGQDDAVKVVTKAVQRSRVGMNNPKRPIASLMFLGPTGVGKTELSKVLADVLFDTPEAV
IHFDMSEYMEKHSISKLIGAAPGYVGYEQGGLLTDAVRKKPYSIILFDEIEKAHPDVYNL
LLRVIDEGKLSDTKGNVANFRNTIIIFTSNLGSQSILDLANDPNKKEKIKEQVMKSVRET
FRPEFYNRIDDHVIFDSLSKKELKEIANIEIRKVANRLFDKNFKITIDDAVFSYIVDKAY
DPSFGARPLKRVIQSEIETEIAVRILDETFVENDTINISLKDQKLHFSKS*
Sequence Length: 1070 aa

Figure 2.1 Predicted Protein Sequence of PfClpB1

The yellow highlight represents the putative mature protein of PfClpB1 after the signal peptide and apicoplast targeting sequence are cleaved by post-translational processing. The green highlight represents the final putative signal peptide sequence. The red highlight represents the final putative apicoplast targeting sequence.

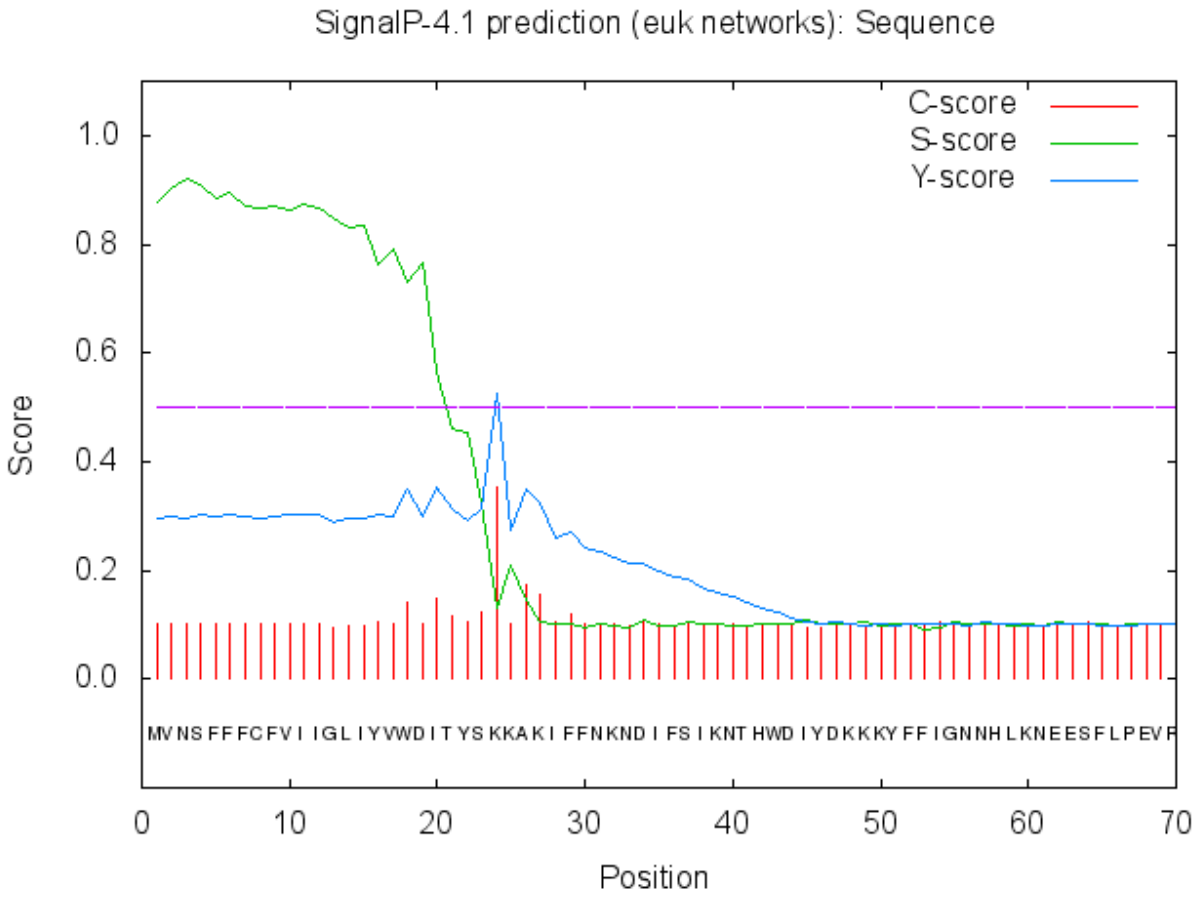


Figure 2.2 Prediction signal peptide cleavage site

SignalP detected the presence of signal peptide cleavage site in the protein sequence. The cleavage site was predicted to be between position 24 and 25.

C-score: Raw cleavage site score is trained to be high at the position immediately after the cleavage site (mature protein).

S-score: Signal peptide score is trained to distinguish positions within signal peptides from positions in the mature part of the proteins and from proteins without signal peptides.

Y-score: Combined cleavage site score is trained to distinguish between C-score peaks by choosing the one where the slope of the S-score is steep

2.2.2 Apicoplast targeting sequence prediction

The apicoplast targeting sequence was predicted by PlasmoAP (http://v4-4.plasmodb.org/cgi-bin/plasmoap.cgi). After the prediction of the signal peptide cleavage site,

the next step was to find the apicoplast targeting sequence. PfClpB1 protein is localized to the apicoplast, but the PfClpB1 gene is localized to the nucleus. PlasmoAP outputs predicted the presence of an apicoplast targeting sequence in the first 150 amino acids following the signal peptide (Table 2.1). Comparison of acidic and basic amino acids content (Figure 2.3) of PfClpB1 revealed the transit peptide switch from basic to acidic at Phe 143. Change of acidic/basic residues ratio is one of the characteristic of bipartite apicoplast. So, we decided the mature protein start at Ser 145.

Table 2.1 Complete PlasmoAP output for query

Only the first 150 AA have been taken into account for the analysis. The final decision is indicated by "++, +, 0 or -", where apicoplast-localization for a given sequence is considered. ++ very likely, + likely, 0 undecided, - unlikely. There is a great probability for the presence of a signal-peptide and apicoplast targeting sequence in the first 150 amino acids.

Criterion	Value	Decision
Signal-peptide	4 of 4 tests positive	++
apicoplast-targeting peptide	5 of 5 tests positive	++
Ruleset 1		
Ratio acidic/basic residues in first 22 amino acids <=0.7	0.333	Yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) with a ratio acidic/basic <=0.9	0.455	Yes
Ruleset 2		
number of acidic residues in first 15 amino acids (<=2)	1	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N)? Ratio acidic/basic residues in this region <0.6	0.455	yes
Is the first charged amino acid basic ?		yes

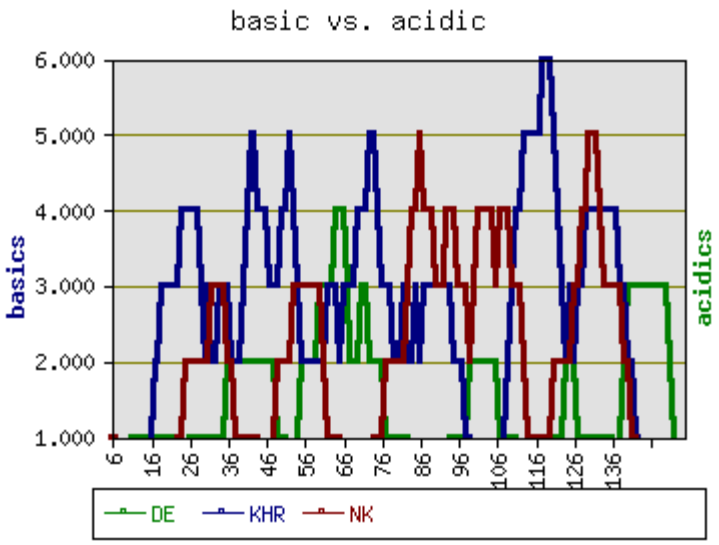


Figure 2.3 Comparison of contents of acidic and basic amino acids

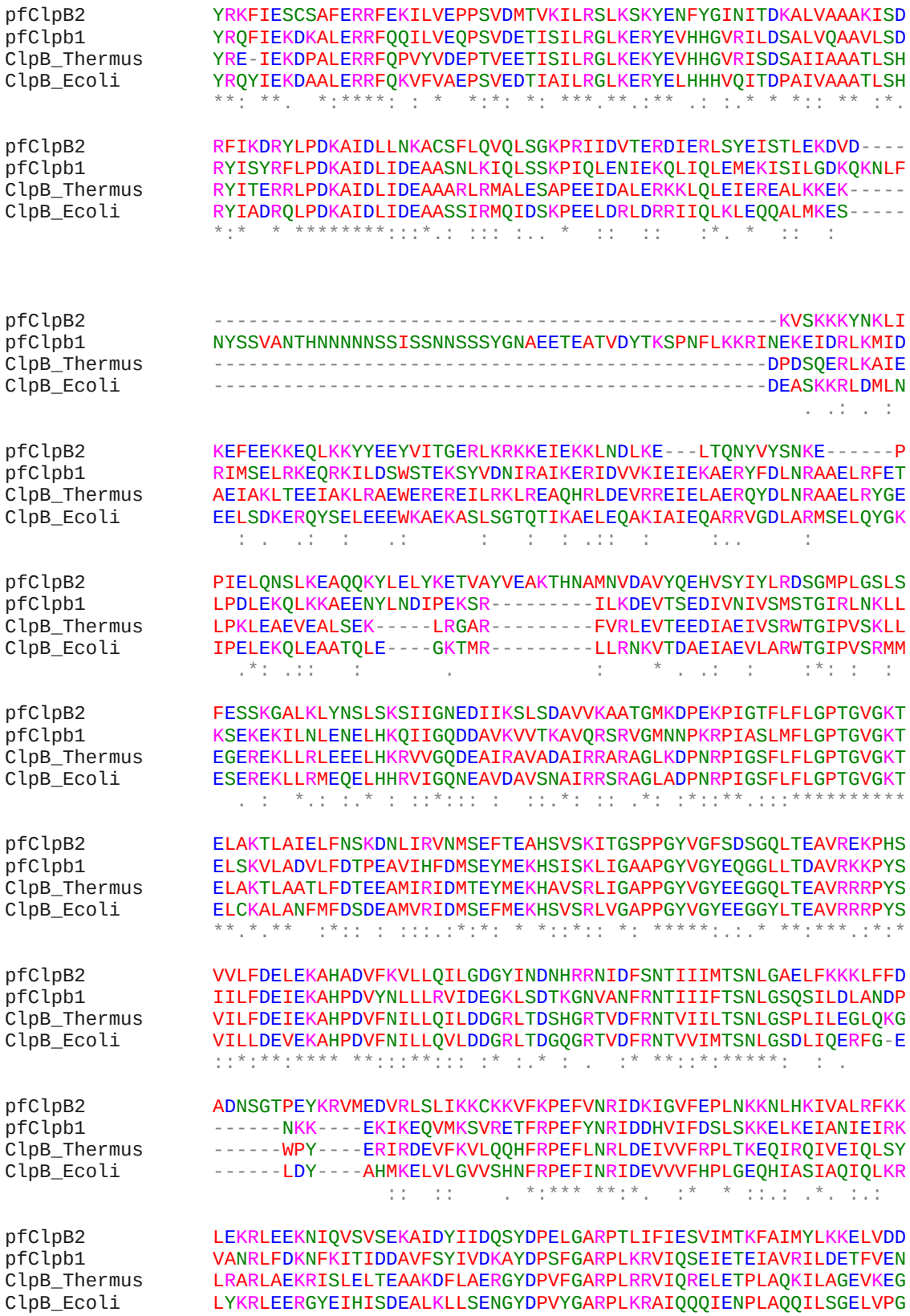
The figure shows that PfClpB1 sequence shifts from basic to acidic close to Phe143. The transit peptide (adjacent to the signal peptide) exhibits a surplus of basic over acidic. This is a characteristic of bipartite apicoplast.

2.3 Multiple sequence alignment of PfClpB1

2.3.1 N-terminus

Amino acid alignment of PfClpB1 and its homologues was produced to deduce conserved residues of the N-terminal domain. A blast search was performed in order to identify PfclpB1 homologues. PfClpB1 and PfClpB2 sequences were used to find identical and similar residue among residue. *Thermus thermophilus* ClpB and *E. coli* ClpB were used for amino acid sequence alignment. Alignment of the N-terminal shows highly conserved regions after the signal peptide and the apicoplast targeting sequence have been removed due to post-translational processing.

```
CLUSTAL O(1.2.0) multiple sequence alignment


pfClpB2           ------------------------------------------------------------
pfClpb1           MVNSFFFCFVIIGLIYVWDITYSKKAKIFFNKNDIFSIKNTHWDIYDKKKYFFIGNNHLK
ClpB_Thermus      ------------------------------------------------------------
ClpB_Ecoli        ------------------------------------------------------------


pfClpB2           ------------------------------------------------------------
pfClpb1           NEESFLPEVRKDYKSQIKEYKNSTNGIIYHNNKNRLSYTINDQVNYDNNMTSGINKKRKV
ClpB_Thermus      ------------------------------------------------------------
ClpB_Ecoli        ------------------------------------------------------------


pfClpB2           ------MTRRYLKYYIFVTLLFFVQVINNVLCAPDNKQEQGKYLNRTINILNAGKNIAKS
pfClpb1           KDSSIHMNNSYEKNRNKNKFALFM-------SDEEYTINSDDYTEKAWEAISSLNKIGEK
ClpB_Thermus      -----------------------------------MNLERWTQAAREALAQAQVLAQR
ClpB_Ecoli        -----------------------------------MRLDRLTNKFQLALADAQSLALG
                                                      .     :     :   : :.


pfClpB2           YGHNKLKPIHILSALAKSDYG---STLFKENNVNAANLKEYIDIALEQTRAGAPLDNKSK
pfClpb1           YDSAYVEAEMLLLALLNDSPDGLAERILKESGIDTQLLVQEIDDYLKKQPKMPSGFGE-Q
ClpB_Thermus      MKHQAIDLPHLWAVLLK-DERSLAWRLLEKAGADPKALKELQERELARLPKVEGAEV--G
ClpB_Ecoli        HDNQFIEPLHLMSALLN-QEGGSVSPLLTSAGINAGQLRTDINQALNRLPQVEGTGG--D
                       :.   :  .* : .       :: .   :   *    :  * :


pfClpB2           IVNSAEVKETLALAEAAANKYKSPKVDVEHLLSGLSND--ELVNEIFNEVYLTDEAIKAI
pfClpb1           KILGRTLQTVLSTSKRLKKEFNDEYISIEHLLLSIISEDSKFTRPWLLKYNVNYEKVKKA
ClpB_Thermus      QYLTSRLSGALNRAEALMEELKDRYVAVDTLVLALAEATPGLPG---------LEALKGA
ClpB_Ecoli        VQPSQDLVRVLNLCDKLAQKRGDNFISSELFVLAALESRGTLAD-ILKAAGATTANITQA
                        :  .*  ..   ::  .  :  : :: .  .    :              :.


pfClpB2           LK-RKFEKTKKDKDGKTGTLYIEQFGSNMNEKVRNGKLQGIYGRDEEIRAIIESLLRYNK
pfClpb1           VEKIRGKKKVTSKTPEMTYQALEKYSRDLTALARAGKLDPVIGRDNEIRRAIQILSRRTK
ClpB_Thermus      LKELRGGRTVQTEHAESTYNALEQYGIDLTRLAAEGKLDPVIGRDEEIRRVIQILLRRTK
ClpB_Ecoli        IEQMRGGESVNDQGAEDQRQALKKYTIDLTERAEQGKLDPVIGRDEEIRRTIQVLQRRTK
                  ::  :  ..   :  :    ::::  ::.  .  ***: : ***:***  *: * *  .*


pfClpB2           NSPVLVGNPGTGKTTIVEGLVYRIEKGDVPKELQGYTVISLNFRKFTSGTSYRGEFETRM
pfClpb1           NNPILLGDPGVGKTAIVEGLAIKIVQGDVPDSLKGRKLVSLDMSSLIAGAKYRGDFEERL
ClpB_Thermus      NNPVLIGEPGVGKTAIVEGLAQRIVKGDVPEGLKGKRIVSLQMGSLLAGAKYRGEFEERL
ClpB_Ecoli        NNPVLIGEPGVGKTAIVEGLAQRIINGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERL
                  *.*:*:*:**.***:*****. :* :*:**. *:*  :::*::  : :*:.***:** *:


pfClpB2           KNIIKELKNKKNKIILFVDEIHLLLGAGK-AEGGTDAANLLKPVLSKGEIKLIGATTIAE
pfClpb1           KSILKEVQDAEGQVVMFIDEIHTVVGAGAVAEGALDAGNILKPMLARGELRCIGATTVSE
ClpB_Thermus      KAVIQEVVQSQGEVILFIDELHTVVGAGK-AEGAVDAGNMLKPALARGELRLIGATTLDE
ClpB_Ecoli        KGVLNDLAKQEGNVILFIDELHTMVGAGK-ADGAMDAGNMLKPALARGELHCVGATTLDE
                  * ::::: . : ::::*:**:* ::***  *:*. **.*:*** *::**:: :****: *
```


pfClpB2 YRKFIESCSAFERRFEKILVEPPSVDMTVKILRSLKSKYENFYGINITDKALVAAAKISD
pfClpb1 YRQFIEKDKALERRFQQILVEQPSVDETISILRGLKERYEVHHGVRILDSALVQAAVLSD
ClpB_Thermus YRE-IEKDPALERRFQPVYVDEPTVEETISILRGLKEKYEVHHGVRISDSAIIAAATLSH
ClpB_Ecoli YRQYIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAATLSH
pfClpB2 RFIKDRYLPDKAIDLLNKACSFLQVQLSGKPRIIDVTERDIERLSYEISTLEKDVD----
pfClpb1 RYISYRFLPDKAIDLIDEAASNLKIQLSSKPIQLENIEKQLIQLEMEKISILGDKQKNLF
ClpB_Thermus RYITERRLPDKAIDLIDEAAARLRMALESAPEEIDALERKKLQLEIEREALKKEK-----
ClpB_Ecoli RYIADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRIIQLKLEQQALMKES-----
pfClpB2 ----------KVSKKKYNKLI
pfClpb1 NYSSVANTHNNNNNSSISSNNSSSYGNAEETEATVDYTKSPNFLKKRINEKEIDRLKMID
ClpB_Thermus ----------DPDSQERLKAIE
ClpB_Ecoli ----------DEASKKRLDMLN
pfClpB2 KEFEEKKEQLKKYYEEYVITGERLKRKKEIEKKLNDLKE---LTQNYVYSNKE------P
pfClpb1 RIMSELRKEQRKILDSWSTEKSYVDNIRAIKERIDVVKIEIEKAERYFDLNRAAELRFET
ClpB_Thermus AEIAKLTEEIAKLRAEWERE REILRKLREAQHRLDEVRREIELAERQYDLNRAAELRYGE
ClpB_Ecoli EELSDKERQYSELEEEWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGK
pfClpB2 PIELQNSLKEAQQKYLELYKETVAYVEAKTHNAMNVDAVYQEHVSYIYLRDSGMPLGSLS
pfClpb1 LPDLEKQLKKAEENYLNDIPEKSR---------ILKDEVTSEDIVNIVSMSTGIRLNKLL
ClpB_Thermus LPKLEAEVEALSEK-----LRGAR---------FVRLEVTEEDIAEIVSRWTGIPVSKLL
ClpB_Ecoli IPELEKQLEAATQLE----GKTMR---------LLRNKVTDAEIAEVLARWTGIPVSRMM
pfClpB2 FESSKGALKLYNSLSKSIIGNEDIIKSLSDAVVKAATGMKDPEKPIGTFLFLGPTGVGKT
pfClpb1 KSEKEKILNLENELHKQIIGQDDAVKVVTKAVQRSRVGMNNPKRPIASLMFLGPTGVGKT
ClpB_Thermus EGEREKLLRLEEELHKRVVGQDEAIRAVADAIRRARAGLKDPNRPIGSFLFLGPTGVGKT
ClpB_Ecoli ESEREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAGLADPNRPIGSFLFLGPTGVGKT
pfClpB2 ELAKTLAIELFNSKDNLIRVNMSEFTEAHSVSKITGSPPGYVGFSDSGQLTEAVREKPHS
pfClpb1 ELSKVLADVLFDTPEAVIHFDMSEYMEKHSISKLIGAAPGYVGYEQGGLLTDAVRKKPYS
ClpB_Thermus ELAKTLAATLFDTEEAMIRIDMTEYMEKHAVSRLIGAPPGYVGYEEGGQLTEAVRRRPYS
ClpB_Ecoli ELCKALANFMFDSDEAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGYLTEAVRRRPYS
pfClpB2 VVLFDELEKAHADVFKVLLQILGDGYINDNHRRNIDFSNTIIIMTSNLGAELFKKKLFFD
pfClpb1 IILFDEIEKAHPDVYNLLLRVIDEGKLSDTKGNVANFRNTIIIFTSNLGSQSILDLANDP
ClpB_Thermus VILFDEIEKAHPDVFNILLQILDDGRLTDSHGRTVDFRNTVIILTSNLGSPLILEGLQKG
ClpB_Ecoli VILLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSDLIQERFG-E
pfClpB2 ADNSGTPEYKRVMEDVRLSLIKKCKKVFKPEFVNRIDKIGVFEPLNKKNLHKIVALRFKK
pfClpb1 ------NKK----EKIKEQVMKSVRETFRPEFYNRIDDHVIFDSLSKKELKEIANIEIRK
ClpB_Thermus ------WPY----ERIRDEVFKVLQQHFRPEFLNRLDEIVVFRPLTKEQIRQIVEIQLSY
ClpB_Ecoli ------LDY----AHMKELVLGVVSHNFRPEFINRIDEVVVFHPLGEQHIASIAQIQLKR
pfClpB2 LEKRLEEKNIQVSVSEKAIDYIIDQSYDPELGARPTLIFIESVIMTKFAIMYLKKELVDD
pfClpb1 VANRLFDKNFKITIDDAVFSYIVDKAYDPSFGARPLKRVIQSEIETEIAVRILDETFVEN
ClpB_Thermus LRARLAEKRISLELTEAAKDFLAERGYDPVFGARPLRRVIQRELETPLAQKILAGEVKEG
ClpB_Ecoli LYKRLEERGYEIHISDEALKLLSENGYDPVYGARPLKRAIQQQIENPLAQQILSGELVPG

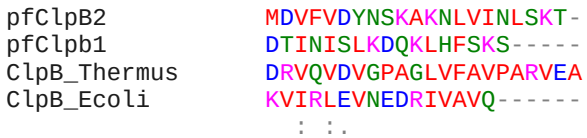


Figure 2.4 Multiple sequence alignment of the N-terminal domains, Middle domain and C-termini of PfClpB1 homologues.

The Walker A and the Walker B are highly conserved in the AAA domains of The PfClpB1. The main difference is found in the N-terminal and the Middle domain. Sequence alignment was performed using Clustal Omega. Identical amino acid residues are marked with "*" symbol and "." and ":" represent the degree of similarity.

2.3.2 Middle domain

The middle domain is responsible for the species specific cooperation with the HSP70 system. The PfClpB1 middle domain has 52 extra amino acid residues (Figure 2.4) that are not present in *Thermus thermophilus ClpB* and *E. coli ClpB*. The function of that specific sequence is still unknown.

2.3.3 C-terminus

The C-terminal domain is required for oligomerization and it is called the sensor domain. The C-terminal is highly conserved in PfclpB1, which confirm its importance during protein disaggregation. The sequence alignment was performed with *Thermus thermophilus ClpB* and *E. coli ClpB.* PfClpB1 and PfClpB2 show some dissimilarity in the C-terminal domain.

2.4 Modeling and DNA analysis of PfClpB1

2.4.1 Homology modeling of PfClpB1

I-TASSER was used to predict PfClpB1 structure using *Thermus Thermophilus* ClpB (PDB 1qvrA) as the main template. PfClpB1 showed conserved structural similarity in the N-terminal, C-terminal, Middle domain and the Nucleotide binding domains. *Thermus*

Thermophilus ClpB was used for homology modeling because its structure was known and it was a close homologue of PfClpB1. Pymol was used to model the multiple domains of PfClpB1. The homology modeling was carried on after removing the putative signal peptide signal and the apicoplast targeting sequence since those must be remove during post-translational processing. Comparison of the two models (PfClpB1 and TClpB) showed that the domains were significantly similar (Figure 2.5) except mainly for the middle domain where some extra amino acid residues were really specific for PfClpB1 and could not be properly aligned.

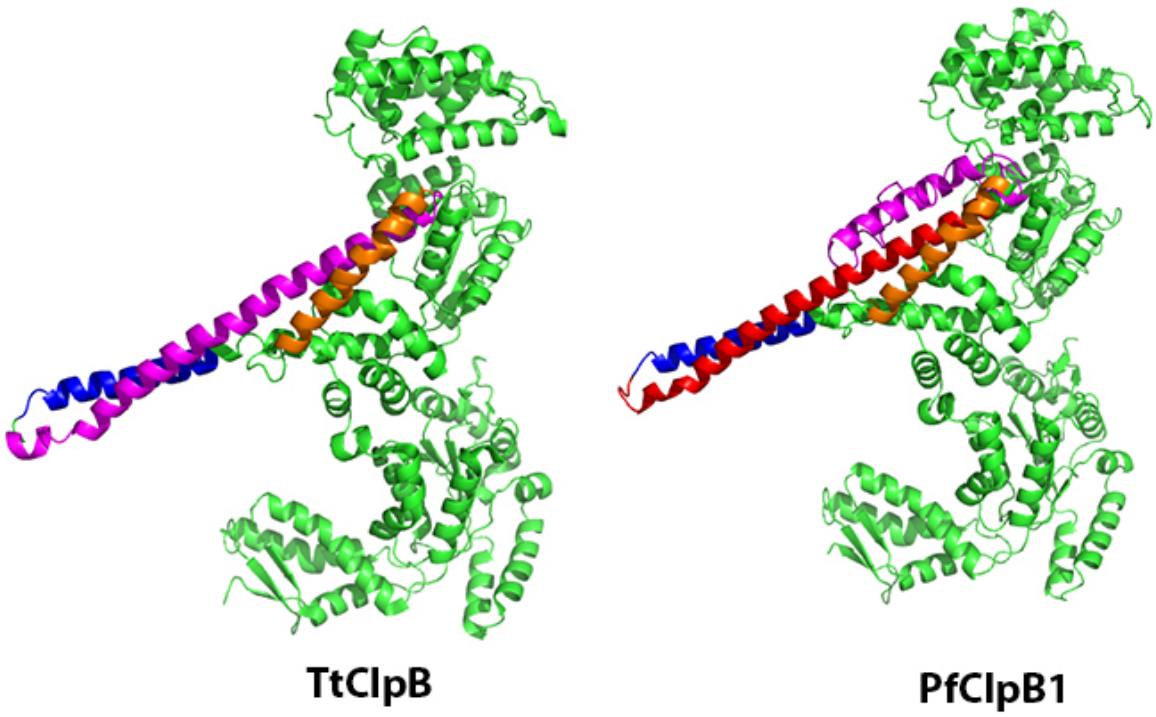


Figure 2.5 Homology modeling of the predicted Tertiary structure of PfClpB1.
Thermus Thermophilus ClpB (1qvrA) was used as a template for the modeling of PfClpB1. The N-terminal α/β subdomain (on the bottom) and the C-terminal α-helical subdomain (at the top) are shown in the figure respectively. The N-terminal domain, C-terminal domain, the NBD1, and NBD2 are conserved. The middle domain was represented with different colors. The red helix represents the extra amino acid in the middle domain whose function is still unknown in pfClpB1. The model was generated with I-TASSER and visualized using pymol.

2.4.2 Codon analysis

The relative adaptiveness of PfClpB1 in *E. coli* was analyzed with the graphical codon usage analyzer (http://gcua.schoedl.de/). The Genscript tools (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) were used to visualize the codon usage and GC content of PfClpB1 gene. Finally, the occurrence of the rare codon in PfClpb1 was mapped with Rare Codon Calculator (http://nihserver.mbi.ucla.edu/RACC/). The distribution of codon usage frequency along the length of PfClpB1 (Figure 2.7) showed that the CG content was low and would be problematic during protein expression. The presence of rare codon (Figure 2.9) confirmed the need for a special strain of *E.coli.*

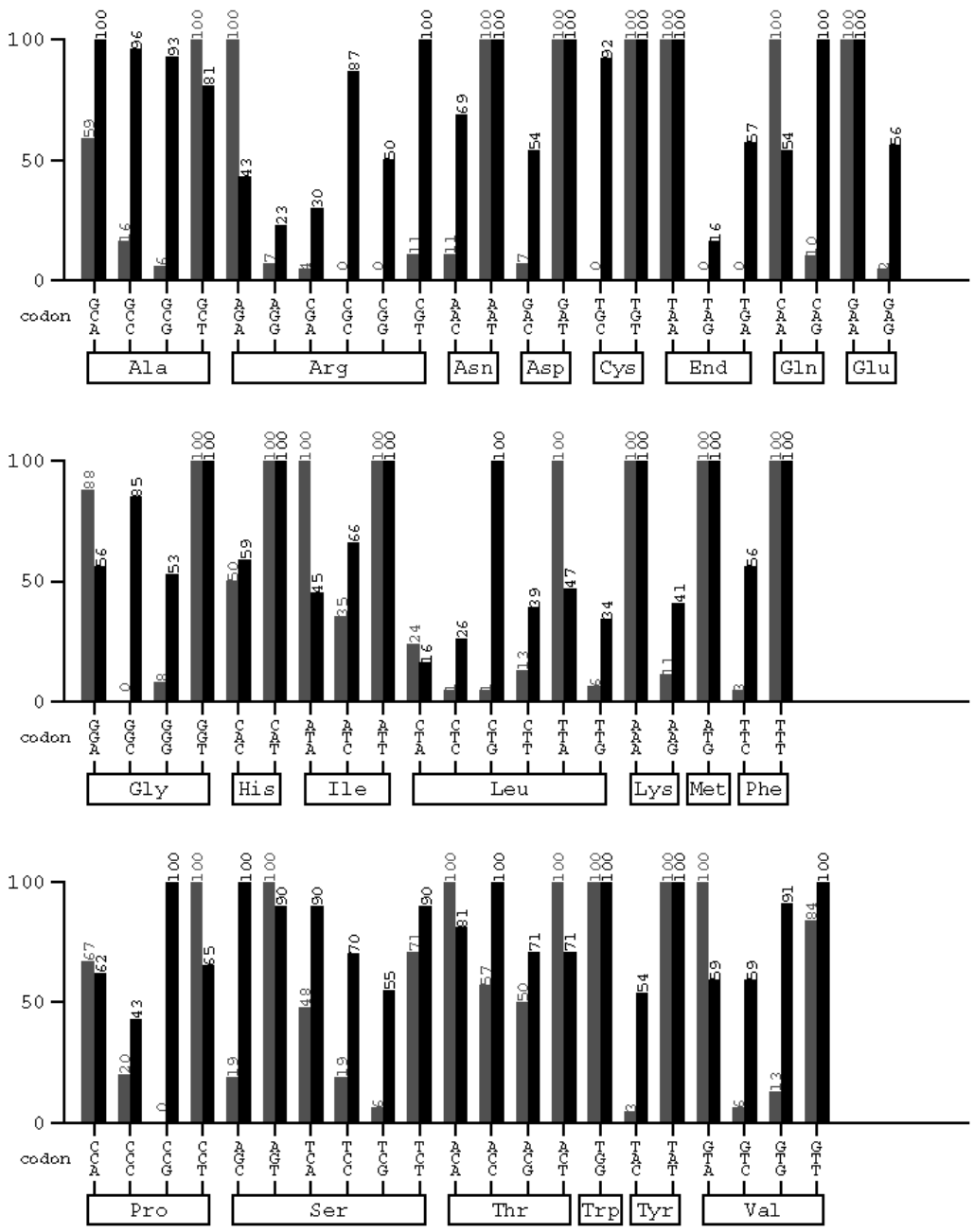


Figure 2.6 Graph shows the relative adaptiveness of PfClpB1.

The relative adaptiveness was analyzed by comparing the nucleotide sequence of PfClpB1 with the codon table of E.coli. The graph shows the percentage of adaptiveness of every codon of PfClpB1 expressed in *E.Coli.* The gray and black bars indicate PfClpB1 and E.coli codon table respectively.

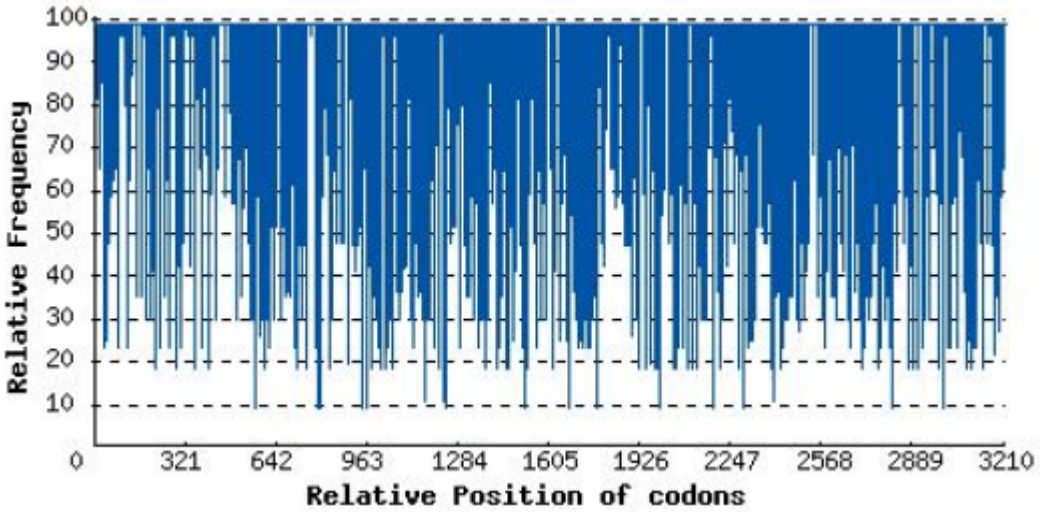


Figure 2.7 The distribution of codon usage frequency along the length of PfClpB1 expressed in E.coli.

The distribution of codon usage frequency was made using the genscript rare codon analysis tools. The Codon Adaptation Index (CAI) was 0.62 which was below the level required for good protein expression.

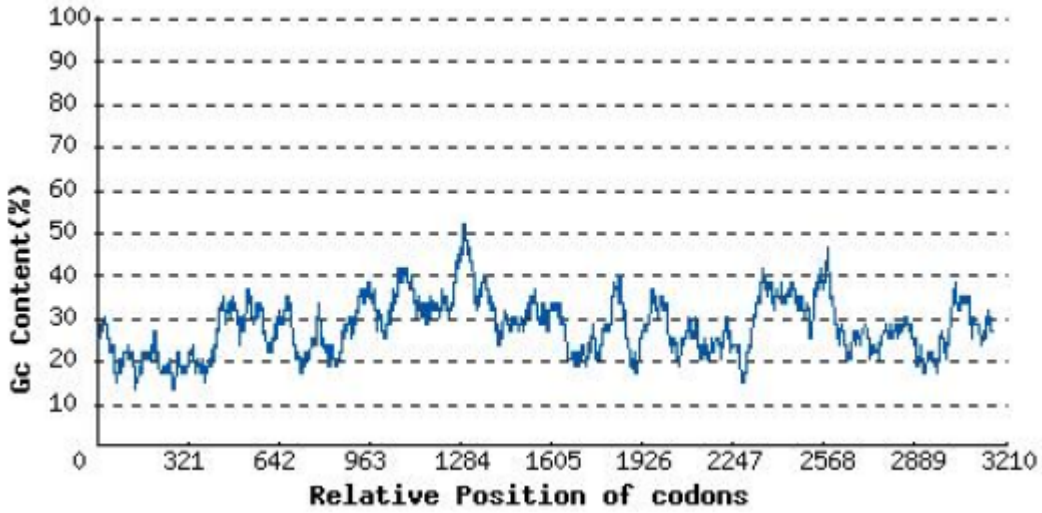


Figure 2.8 The GC content of PfClpB1.

The Average GC content was 27.71%. Any peaks outside 30% to 70% will adversely affect transcriptional and translational efficiency.

atg gtt aat agt ttt ttt ttt tgt ttt gtg ATA att ggt ctt att tat gta tgg gac
atc acg tac agt aaa aaa gct aaa ATA ttt ttt aat aaa aat gat atc ttt
agt ATA aaa aat acg cac tgg gat att tat gat aag aag aaa tat ttt ttt att gga
aat aat cat tta aaa aat gaa gaa agt ttt tta cca gaa gta AGA aag gat tat aaa
tca caa ATA aaa gaa tat aag aat tca acg aat ggt att ATA tat cat aat aat aaa
aac AGA tta agt tat aca ATA aat gat caa gta aat tat gat aat aat atg aca agt
ggt att aat aaa aaa AGA aaa gtt aaa gat agt agt ATA cac atg aat aat tct tat
gaa aaa aat AGA aac aaa aat aaa ttt gct tta ttt atg agt gat gaa gaa tat acc
att aat tca gat gat tat acc gaa aaa gct tgg gaa gct att agc tcc tta aat aaa
att gga gaa aaa tat gat tcg gca tat gta gaa gct gaa atg tta tta tta
gct CTA CTA aat gat tca CCC gat ggt tta gct gaa AGA ATA tta aaa gaa agt
ggtATA gat acc caa tta tta gtt caa gaa att gat gat tat tta aaa aaa caa cct
aag atg cct agt ggt ttt gga gaa cag aaa ATA tta ggt AGA act tta caa act gta
tta agt act agt aaa AGA tta aaa aaa gaa ttt aat gat gaa tat att tcc ATA gaa
cac CTA tta CTA agt atc att tca gaa gat tct aaa ttt act AGA CCC tgg tta tta
aaa tat aat gta aat tat gaa aaa gta aaa aaa gct gta gaa aaa att CGA gga aaa
aaa aaa gtt act tct aaa aca cca gaa atg act tat caa gct CTA gaa aaa tat
agt AGA gat CTA aca gct ttg gca AGA gca gga aaa tta gat cct
gtt ATA ggt AGA gat aat gaa att AGA AGA gcc ATA caa att tta tcc AGA AGA act
aaa aat aat cct atc tta tta gga gat cct ggt gtt ggg aaa aca gct att gtt gaa
ggg tta gcc ATA aaa atc gta caa gga gat gta cct gac tca tta aaa gga AGG aaa
tta gta tct tta gat atg tct tct ctt ATA gct ggt gca aaa tat AGA ggt gat ttt
gaa gaaAGG CTA aaa tca att ctg aaa gaa gta caa gat gct gaa ggt caa gtt gtt
atg ttt ATA gat gaa atc cat act gtt gtg gga gct gga gcg gtc gca gaa ggt gca
tta gat gct ggt aatATA tta aaa cct atg tta gct AGA ggt gaa tta cgt tgt att
ggt gct acg acg gtt agt gaa tat AGA caa ttt ATA gaa aag gat aaa gca tta
gaa AGA AGA ttt caa caa att ctt gtt gaa caa cca agt gtt gat gaa act att
agt ATA tta AGA ggt CTA aaa gaa AGA tat gaa gtt cat cat ggt gta cgt ATA tta
gat tct gca tta gta caa gct gct gtt tta tca gat cgt tat att agt tat AGA ttc
tta cca gat aaa gcg att gat ctt att gac gaa gct gca tct aat ctt
aaa ATA caa CTA tct agt aaa cct att caa tta gaa aat ATA gaa aaa caa
ctt ATAcaa tta gaa atg gaa aaa ATA tcc ATA tta gga gat aaa caa aag
aat CTA ttt aat tat tct agt gta gct aac aca cac aat aat aat aat aat agt agt
att agt agc aat aac tcg tca tca tat ggt aac gct gaa gaa act gaa gca act gtt
gat tat act aaa agc CCC aat ttt tta aaa aaa AGA att aat gaa aaa gaa att
gat AGA tta aaa atg atc gat CGA atc atg agc gaa tta AGA aaa gaa caa AGA aaa
atc CTA gat tct tgg tcc acc gaa aaa agc tat gta gat aat atc AGA gct att aaa

gaa AGA ATA gat gtt gtt aaa ATA gaa att gaa aaa gct gaaAGA tat ttt gat tta
aat AGA gca gct gaa ttg AGA ttt gaa aca tta cct gat tta gaa aaa caa tta aaa
aaa gca gaa gaa aat tat CTA aat gat atc cct gaa aaa agt AGA ATA tta aaa gat
gaa gtt aca agt gaa gat att gtt aat att gta agt atg tct acc ggt atc AGA tta
aat aaa tta CTA aaa tct gaa aaa gaa aaa ATA ctt aat ctt gaa aat gaa tta cat
aaa caa att atc ggt caa gat gat gcc gta aaa gtt gta acc aaa gct gtt
caa AGA tct AGG gtt gga atg aat aac cct aaa AGA cca ATA gca tct tta atg ttt
tta gga cca aca gga gta gga aaa acg gaa tta tct aag gta ttg gca gat gta tta
ttt gac aca cca gaa gca gta att cat ttt gat atg tct gaa tat atg gag aag cat
tca att agt aaa tta ATA ggt gcc gca cca ggt tat gtg gga tat gaa caa gga gga
tta tta aca gat gca gta cgt aaa aaa cca tat tct atc att tta ttt gat
gaa ATA gaa aaa gca cat cct gat gta tat aat tta tta ttaAGA gtt ATA gat gag
gga aaa tta tct gat acc aaa gga aat gta gct aat ttt AGA aat aca att
att ATA ttt aca tcc aat tta gga agt caa agt ATA CTA gat CTA gct aat gat cca
aat aaa aaa gaa aaa atc aaa gaa cag gta atg aaa tca gtg AGA gaa aca
ttt AGA cct gaa ttt tat aac AGA att gat gat cat gtt ATA ttt gat agc tta tca
aaa aaa gaa tta aaa gaa att gca aat att gaa att AGA aaa gta gct aat
cgt CTA ttt gat aaa aat ttt aaa ATA act ATA gac gat gct gtc ttt tca
tat ATA gta gat aaa gcc tat gat cct tct ttt ggt gct AGA cct ctt
aaa AGA gtt ATA caa tct gaa ATA gaa acg gaa att gct gta AGA ATA tta gat gaa
acc ttt gta gaa aat gat act att aat ATA tct ctc aag gat cag aag ttg cac ttt
tca aaa agt taa

Figure 2.9 The occurrence of rare codons in PfClpB1

Red = rare Arg codons AGG, AGA, CGA./ Green = rare Leu codon CTA./ Blue = rare Ile codon ATA / Orange = rare Pro codon CCC. The occurrence of rare codons was mapped with Rare Codon Calcμl ator (http://nihserver.mbi.ucla.edu/RACC/).

Chapter 3 - Expression and purification of PfClpB1

3.1 Cloning

Subcloning was performed by moving the PfClpB1 gene from the parental vector P15TV-L, a gift from Dr. Raymond Hui of the University of Toronto, to the destination vector pET28a(+) (Figure 3.1), using PfuUltra II DNA polymerase (Agilent Technologies) for amplification. Codons for arginine (AGG or AGA), glycine (GGA) and isoleucine (AUA), are found to be rarely used and have a damaging effect on protein expression in E.Coli (Table 3.1). *P. falciparum* contains an A-T rich genome, consequently causing rare codon usage in *E. coli* as confirmed by codon analysis. A special strain of *E.coli* such as Rosetta strain DE3 (Novagen) was used to provide rare tRNAs for Leucine, Isoleucine, Arginine, Glycine, Proline to enhance protein expression. The setup for the PCR reaction was conducted with the following protocol for PfuUltra II DNA Polymerase (Agilent Technologies). 2.0 µl PfClpB1 template (100 ng/µl) was used with 0.8 µl dNTP mix (25 mM each dNTP), 10.0 µl of 10× PfuUltra II reaction buffer, 200 ng of each designed forward and reverse primer, 2.0 µl of PfuUltra HF DNA polymerase (2.5 U/µl) , finally the volume was brought to 100µl with distilled water (dH_2O) . The PCR cycling parameters for PfuUltra II DNA Polymerase were performed as follows: 95°C 2 mins, followed by 30 cycles: 30 seconds at 95°C, 30 seconds at 55°C, and 4 mins at 72°C in each cycle. Then, the amplified products were elongated for an additional 15 mins at 72°C. PCR products were gel purified with QIAEX II Gel Extraction Kit (Qiagen). UV absorbance at 260 nm was used to find the DNA concentration. Then PCR products and pET28a(+) were digested with restriction enzymes XhoI and NheI from New England Biolabs at 37°C for 4 h, and purified from agarose gels using QIAEX II Gel Extraction Kit (Qiagen). The digested vector and PCR products were ligated with Rapid DNA Ligation Kit (Roche) at 16°C overnight. Rosetta2 (DE3) cells were transformed with the ligated vector. The transformed colonies were selected against chloramphenicol (Rosetta 2 (DE3)) and kanamycin (pET28a(+)). DNA sequencing was performed at the Plant Pathology sequencing facility to check the accuracy of our transformation.

Primers used for amplification of the PfClpB1 with the restriction site for NheI and XhoI.

Forward primer:

5’ CGCGCTAGCAGTGATGAAGAATATACCATT 3’ (NheI)

Reverse primer:

5’ CGCGCACTCGAGTTAACTTTTTGAAAAGTGCAA 3’ (XhoI)

Table 3.1 Rarely used codons in E.coli.

Lee, S.F., Li, Y.J., and Halperin, S.A. (2009). “Overcoming codon-usage bias in heterologous protein expression in Streptococcus gordonii.” Microbiology 155:3581-3588.

Amino acid	Rare codon(s)
Arginine	AGG, AGA, CGG, CGA
Leucine	CUA, CUC
Isoleucine	AUA
Serine	UCG, UCA, AGU, UCC
Glycine	GGA, GGG
Proline	CCC, CCU, CCA
Threonine	ACA

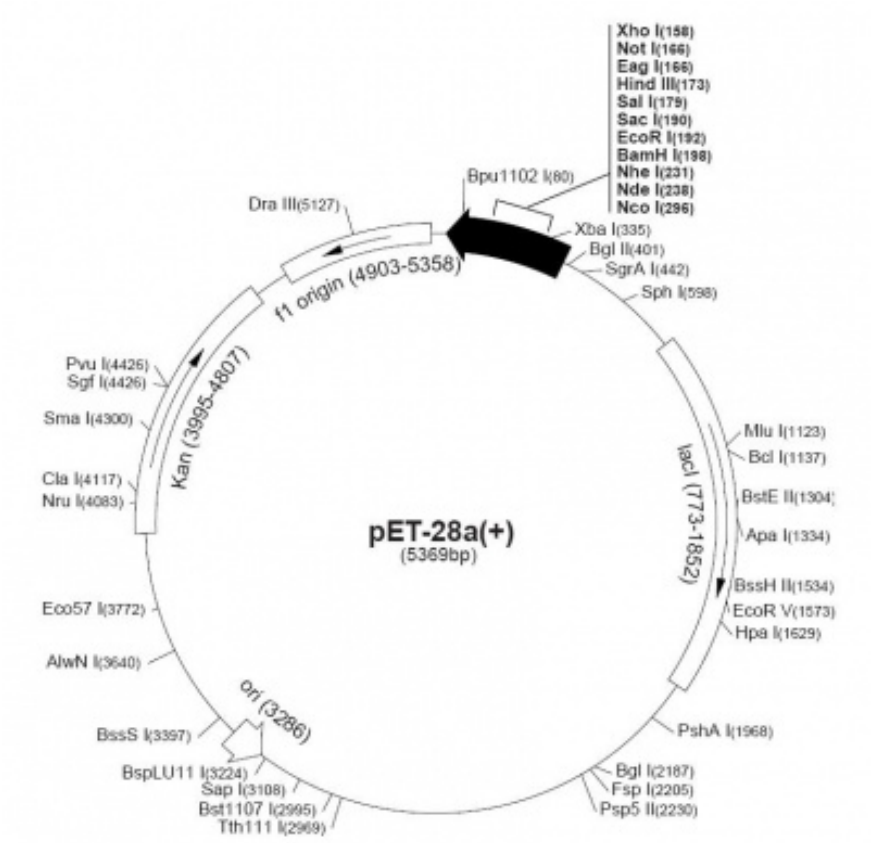


Figure 3.1 Novagen pET-28a(+)

Plasmid map pET-28a(+) showing restriction sites of XhoI and NheI

3.2 Protein expression

The pfClpB1 coding sequence (amino acids Ser145-Ser1070) was cloned into pET28 vector with a His-tag on the N-terminus. The protein expression was started by inoculating a single colony holding the pET28vector with pfClpB1 gene into 50 ml LB containing 50ug/mL chloramphenicol and 50ug/mL kanamycin. The cultures were incubated overnight at 37°C with shaking at 225 rpm. Afterward, a 1:20 dilution of the overnight cultures was poured into 1000 ml fresh LB and was grown at the same conditions with similar concentration of chloramphenicol and kanamycin until an OD600 reading of 0.5. Finally, protein expression was induced at 37°C for 2 h with 0.4 mM IPTG. SDS-page analysis was conducted (Figure 3.3) to detect the expression of a 107 KDA protein. MS analysis would be done along the way to confirm that our protein is PfClpB1.

A

MGSSHHHHHHSSGRENLYFQGASDEEYTINSDDYTEKAWEAISSLNKIGEKYDSAYVEAEMLLLA
LLNDSPDGLAERILKESGIDTQLLVQEIDDYLKKQPKMPSGFGEQKILGRTLQTVLSTSKRLKKEFNDEYISI
EHLLLSIISEDSKFTRPWLLKYNVNYEKVKKAVEKIRGKKKVTSKTPEMTYQALEKYSRDLTALARAGKLD
PVIGRDNEIRRAIQILSRRTKNNPILLGDPGVGKTAIVEGLAIKIVQGDVPDSLKGRKLVSLDMSSLIAGAKY
RGDFEERLKSILKEVQDAEGQVVMFIDEIHTVVGAGAVAEGALDAGNILKPMLARGELRCIGATTVSEYRQ
FIEKDKALERRFQQILVEQPSVDETISILRGLKERYEVHHGVRILDSALVQAAVLSDRYISYRFLPDKAIDLID
EAASNLKIQLSSKPIQLENIEKQLIQLEMEKISILGDKQKNLFNYSSVANTHNNNNNSSISSNNSSSYGNAEET
EATVDYTKSPNFLKKRINEKEIDRLKMIDRIMSELRKEQRKILDSWSTEKSYVDNIRAIKERIDVVKIEIEKAE
RYFDLNRAAELRFETLPDLEKQLKKAEENYLNDIPEKSRILKDEVTSEDIVNIVSMSTGIRLNKLLKSEKEKI
LNLENELHKQIIGQDDAVKVVTKAVQRSRVGMNNPKRPIASLMFLGPTGVGKTELSKVLADVLFDTPEAVI
HFDMSEYMEKHSISKLIGAAPGYVGYEQGGLLTDAVRKKPYSIILFDEIEKAHPDVYNLLLRVIDEGKLSDT
KGNVANFRNTIIIFTSNLGSQSILDLANDPNKKEKIKEQVMKSVRETFRPEFYNRIDDHVIFDSLSKKELKEIA
NIEIRKVANRLFDKNFKITIDDAVFSYIVDKAYDPSFGARPLKRVIQSEIETEIAVRILDETFVENDTINISLKD
QKLHFSKS

Figure 3.2 The amino acid sequence of 6×His- tag pfClpB1 after cleavage of the signal peptide and apicoplast targeting sequence

The polyhistidine region is underlined. The target protein is marked as red

E. coli lysate 1	E. coli lysate 2	E. coli lysate 3	E. coli lysate 4

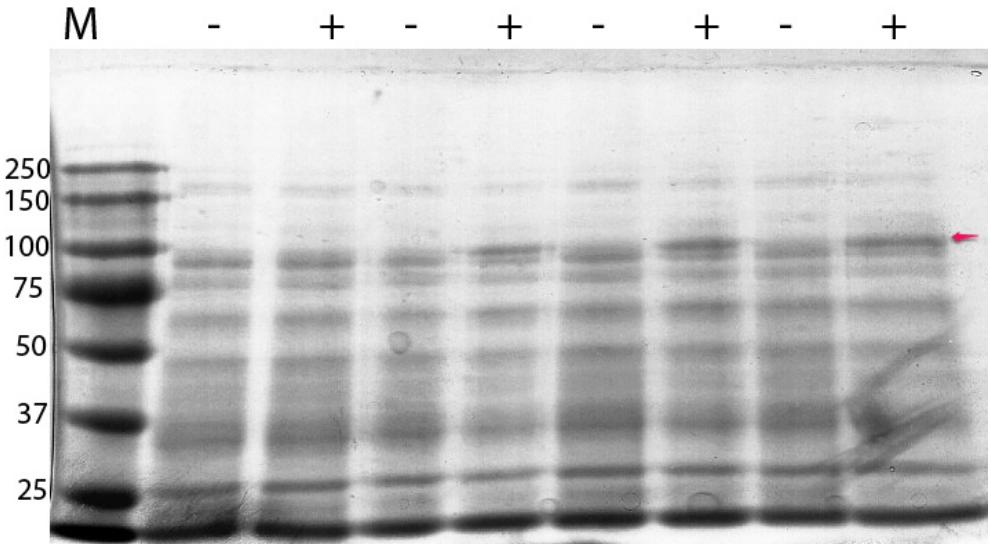


Figure 3.3 SDS gel electrophoresis analysis of PfClpB1 after protein expression.
M: Molecular mass ladder. +: IPTG induction. -: No induction. The red arrow indicates the band of the putative PfClpB1protein with a molecular weight of 107 kDa. Shown is the protein expression of 4 independent E. coli cultures either non-induced (-) or induced with IPTG (+) after 2 hours.

3.3 Solubility test

Isolation of the cell pellets for solubility test was initiated by centrifuging at 4000 x g for 20 min at 4°C, and the mass of the cells was weighted. Next, 4 ml/g of ice-cold lysis buffer (300 mM NaCl, 10 mM imidazole in 50 mM sodium phosphate buffer, pH 8.0) was added to the pellets. Sonication on ice for 25 ×15 seconds with a 10 sec cooling between each burst was used to disrupt the cells. 100 µL lysates was collected after sonication for centrifugation at 14,000 × g for 5 min at 4°C. Then, the insoluble fraction was taken out by solubilizing the pellets in 100 µL cold lysis buffers (300 mM NaCl, 10 mM imidazole in 50 mM sodium phosphate buffer, pH 8.0). The remaining lysates were centrifuged at 4000 × g for 20 min at 4°C. For the solubility test, 100 µL supernatant as soluble fraction was collected. The rest of soluble fractions were saved for purification. Finally, SDS-Page followed by coomassie blue staining was performed on 15 µL aliquots of each cellular fraction.

3.4 Purification of His3tagged PfClpB1

The soluble fraction of pfClpB1 was purified under native conditions by nickel- affinity Chromatography (Ni2+-NTA). Pre-equilibrated Ni-NTA resin (Invitrogen) with the supernatant was mixed overnight at 4°C. The mixture was then poured into a 1.5 cm diameter column and the flow-through was saved. After washing, the pfClpB1 proteins were eluted with 250 mM imidazole. Purified proteins fractions were analyzed with SDS-PAGE (Figure 3.4), followed by coomassie blue staining. Finally, the concentration of purified pfClpB1 proteins was measured spectrophotometrically and reported in monomer units.

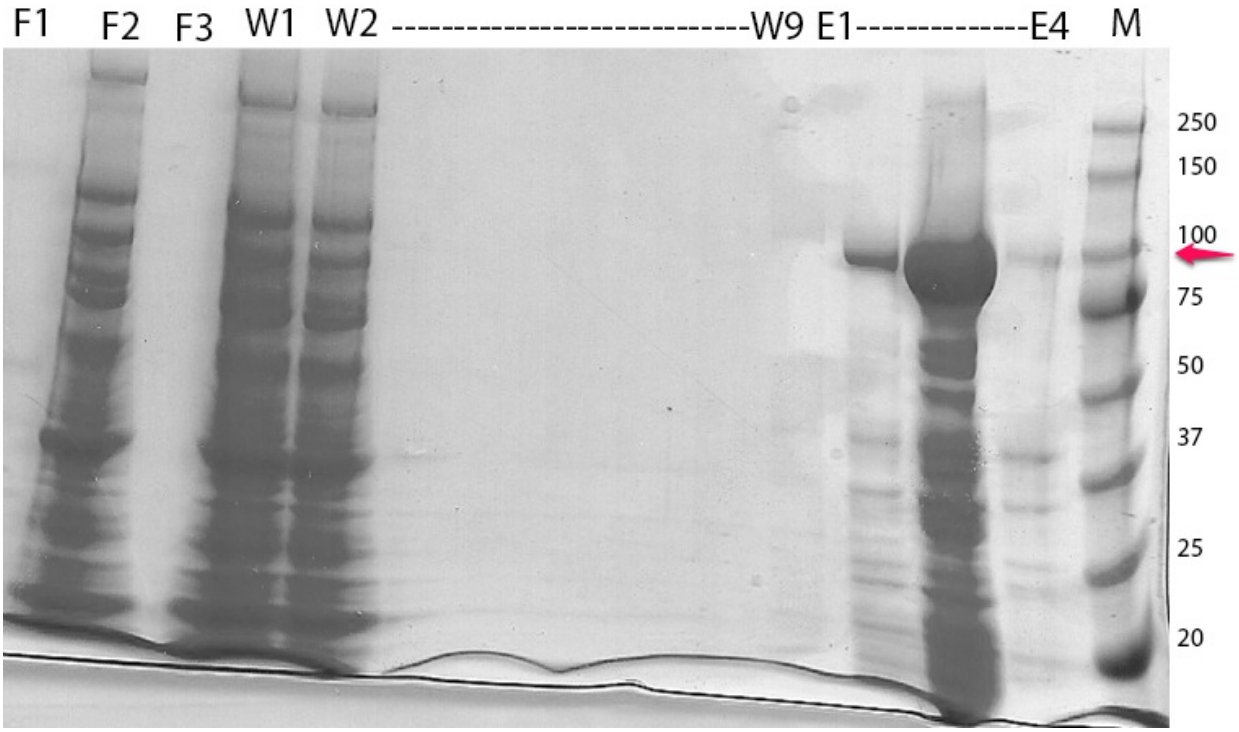


Figure 3.4 SDS gel electrophoresis analysis of purification of 6×His-tagged PfClpB1 by using Ni-NTA affinity chromatography under native conditions.

Each lane represents different protein fractions collected overtime. The flow-through was saved for analysis. The elution fraction was collected with an elution buffer from a purification kit. M: molecular mass ladders. The red arrow: The band of PfClpB1. F1-F3: Flow-through. W1-W9: wash. E1-E4: Elution.

3.5 Gel filtration

Gel filtration was used for further purification of the PfClpB1 proteins. A Superdex 200® (GE LifeSciences) was equilibrated with 50 mM Tris-HCl PH 7.5, 1mM EDTA, 1mM DTT, 20 mM $MgCl_2$, 0.1 M KCl, at a flow rate of 2 mL/min. Elution was carried with 250 mM imidazole. The partially purified pfClpB1 from Ni-NTA filtration was loaded into the system at a rate of 0.1 mL/min. All collected samples were analyzed with 10% SDS-PAGE to figure out the pfClpB1 fractions with the least amount of contaminants. The resulting gel (Figure 3.5) revealed a protein with a molecular weight of 107Kda.

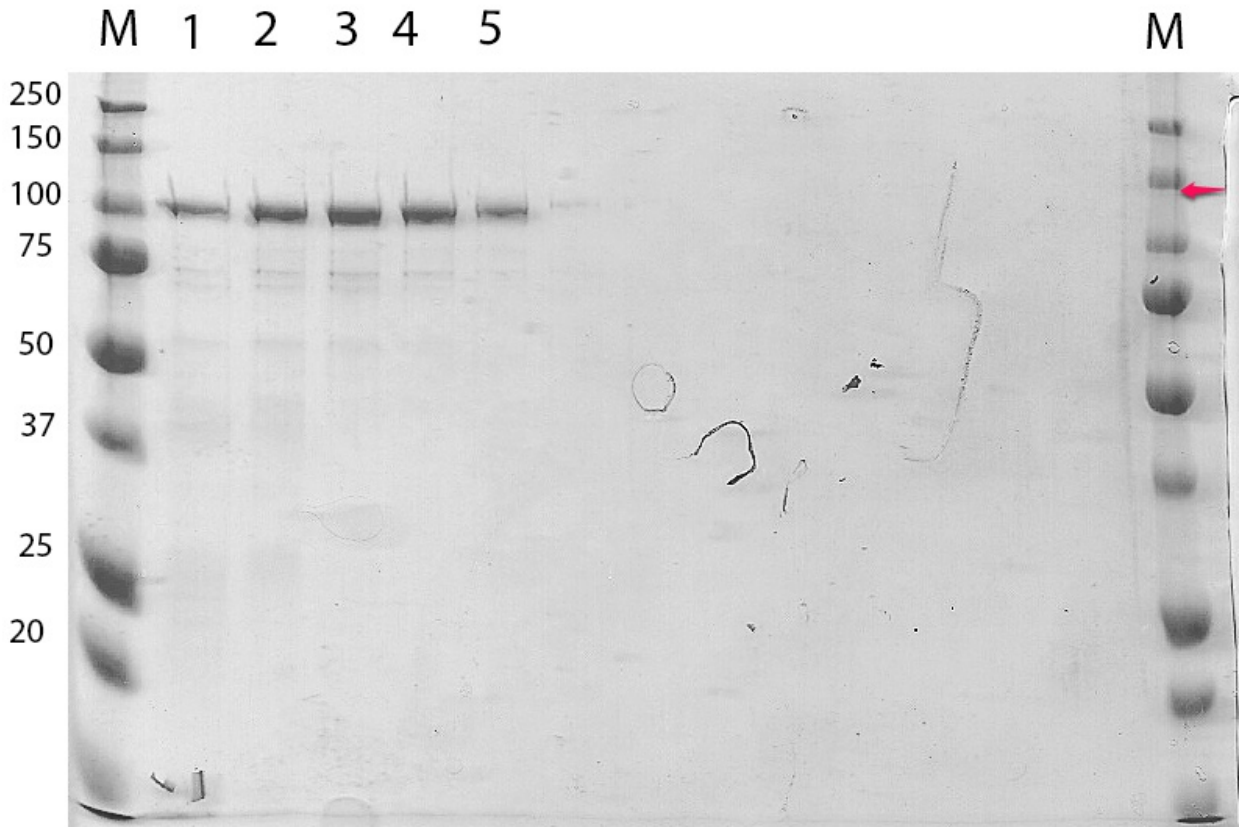


Figure 3.5 SDS gel electrophoresis analysis of Gel filtration

M: Molecular mass ladder.

Lanes 1, 2, 3, 4, 5 represents fractions containing PfClpB1 protein after gel filtration purification. The red arrow indicates the band of PfClpB1.

3.6 Mass spectrometry

The purified proteins were heated at 95 °C for 6min and treated with 2 x SDS loading buffer made of SDS and β-mercaptoethanol. The sample was loaded with 10% SDS-PAGE, and run on constant voltage of 200 V for approximately 30 min. The Electrophoresis buffer was made of Tris-glycine-SDS buffer. The gels were then stained with coomassie blue R-250 for 60 min and destained with 40% (v/v) methanol and 10% (v/v) acetic acid. Finally, MS analysis of tryptic peptides (Figure 3.6) was performed by KSU Biotechnology/Proteomics facility to confirm that the purified protein was a match to PfClpB1 peptide after gel purification.

1 MGSSHHHHHH SSGLVPR**GSH MASDEEYTIN SDDYTEKAWE AISSLNK**IGE
51 KYDSAYVEAE MLLLALLNDS PDGLAER**ILK ESGIDTQLLV QEIDDYLKK**Q
101 PKMPSGFGEQ KILGR**TLQTV LSTSKR**LK**KE FNDEYISIEH LLLSIISEDS**
151 **KFTRPWLLK**Y NVNYEKVKKA VEKIRGKKKV TSK**TPEMTYQ ALEK**YSRDLT
201 ALARAGK**LDP VIGRDNEIR**R AIQILSRR**TK NNPILLGDPG VGK**TAIVEGL
251 AIK**IVQGDVP DSLKGRKLVS LDMSSLIAGA K**YRGDFEERL KSILK**EVQDA**
301 **EGQVVMFIDE IHTVVGAGAV AEGALDAGNI LKPMLAR**GEL R**CIGATTVSE**
351 **YR**QFIEKDKA LER**RFQQILV EQPSVDETIS ILR**GLKERYE VHHGVR**ILDS**
401 **ALVQAAVLSD RYISYRFLPD KAIDLIDEAA SNLKIQLSSK PIQLENIEKQ**
451 **LIQLEMEK**IS ILGDKQKNLF NYSSVANTHN NNNNSSISSN NSSSYGNAEE
501 TEATVDYTKS PNFLKKRINE KEIDRLKMID RIMSELRKEQ R**KILDSWSTE**
551 **K**SYVDNIRAI KERIDVVKIE IEKAER**YFDL NR**AAELR**FET LPDLEK**QLK**K**
601 **AEENYLNDIP EK**SR**ILKDEV TSEDIVNIVS MSTGIR**LNKL LKSEKEK**ILN**
651 **LENELHKQII GQDDAVKVVT K**AVQRSRVGM NNPK**RPIASL MFLGPTGVGK**
701 TELSK**VLADV LFDTPEAVIH FDMSEYMEK**H SISK**LIGAAP GYVGYEQGGL**
751 **LTDAVRKKPY SIILFDEIEK AHPDVYNLLL R**VIDEGKLSD TKGNVANFR**N**
801 **TIIIFTSNLG SQSILDLAND PNKKEKIKEQ VMKSVRETFR PEFYNRIDDH**
851 **VIFDSLSK**KE LK**EIANIEIR** KVANRLFDKN FK**ITIDDAVF SYIVDKAYDP**
901 **SFGAR**PLKR**V IQSEIETEIA VRILDETFVE NDTINISLKD QK**LHFSKS

Figure 3.6 Mass spectrometry of purified PfClpB1

The peptides detected by MS analysis are shown in red.

3.7 Conclusion

A purified form of mature recombinant PfClpB1was produced in Rosetta BL21(DE3) strain of E. coli. Rosetta BL21(DE3) provides tRNAs for the rare codons found in the PfClpB1. The mature PfClpB1 is without the bipartite N-terminal targeting sequence.

Chapter 4 - In vitro Characterization of PfClpB1: Methods

E. coli ClpB has been the object of extensive research regarding its biochemical properties. ClpB is up-regulated and can reactivate aggregated proteins during cellular stress. Better knowledge about the mechanism of aggregate related diseases has been established because of our understanding of ClpB function. During infection, ClpB is up-regulated in many pathogenic bacteria and helps to overcome stress from the immune system of the host. Previous studies confirm the need of ClpB during the pathogenic stage. Since, ClpB exists in bacteria, but not in humans, it is a good target for therapeutic drugs. Sequence alignment of PfClpB1 shows two conserved nucleotide binding domains in PfClpB1, containing the Walker A and Walker B motifs. Sequence identity between PfClpB1 and *E. coli* ClpB is also quite high, but the middle domain carries extra amino acid residues. In this chapter, we studied the oligomeric structure of PfClpB1, ATPase activity and chaperone activity.

4.1 ATPase activity

Analysis of the ATPase activity of PfClpB1 was performed with a Malachite green phosphate assay. The assay dye was prepared in a 3:1 mixture of 0.045% malachite green and 4.2% ammonium molydate in 4N HCL. 6 ml of malachite green (0.045%) was added to 2ml ammonium molydate in 4N HCL (4.2%) and finally 150 µl of 10% tween solution was used to complete the mixture. The assay dye mixture was equilibrated for 30 minutes before usage. 34% sodium citrate was also prepared. From 10mM of phosphate stock solution, a 2.5 mM phosphate solution was prepared by adding 20 µl of the stock solution to 60 µl of distilled water. The 2.5 mM phosphate solution was used to build the standard curve. PfClpB1 and ClpB1 were incubated in assay buffer (100 mM Tris/HCl pH 8.0, 5 mM ATP, 1 mM DTT, 10 mM MgCl2, and 1 mM EDTA) at 37 °C for 30 min without or with 0.1 mg/ml α-casein or 0.04 mg/ml poly-lysine. The concentration of ClpB and pfClpB1 were 0.1 mg/ml in the presence of α-casein and poly-lysine, 1 mg/ml for the basal activity. For the blank 45 µl buffer assay was mixed with 5 µl buffer B (Buffer B: 50 mM Tris pH 7.5, 0.1 M KCL, 0.1 M EDTA and 1 M DTT was used during gel filtration for pfClpB1 extraction) and also incubated at 37 °C for 30 min. After incubation, 800 µl of dye solution was added and this was followed one minute later with 100 µl

of 34% sodium citrate. The ATPase activity was measured with a spectrophotometer at 660 nm as described before [15].

4.2 Sedimentation velocity

A Beckman Xl-I analytical ultracentrifuge with 2-channel analytical cells was used for the sedimentation experiment. The data were analyzed with the software included with the instrument using a time-derivative approach. The sedimentation velocity analysis of pfClpB1 was performed at 49,000 rpm and 20 °C with absorption profiles measured at 242 nm after 0.7-mg/ml protein sample was dialyzed in 50 mM Tris/HCL pH 7.5, 0.2 M KCl, 20 mM $MgCl_2$, 1 mM EDTA, 2 mM β-mercaptoethanol. Before sedimentation velocity, protein dialysis was performed to exchange buffer B with buffer for sedimentation velocity. Finally, 2 mM ATPγS [adenosine-50-(γ-thio)-triphosphate] was added to the sample and ultracentrifugation was performed at 42,000 rpm with absorption profiles measured at 291 nm.

Table 4.1 Buffer for sedimentation velocity

50 mM Tris/HCL pH 7.5	0.1 L of 1 M Tris
0.2 M KCl	0.2 L of 1 M KCL
20 mM $MgCl_2$	0.02 L of 1M MgCl2
1 mM EDTA	0.02 L of 0.1 M EDTA
2 mM β-mercaptoethanol	0.31 mL of 13M β-mercaptoethanol
dH_2O	Water to 2 L

4.3 Protein reactivation assay

Next we tested the chaperone activity of pfClpB1. 226 μM Luciferase stock from Promega was used. The recombinant firefly luciferase was diluted 300-fold into buffer 1 (50 mM Tris pH 7.5, 120 mM KCL, 10 mM MgCl2, 1 mM EDTA, 6 mM ATP and 10 mM DTT) and incubated at 45 degree Celsius for 12 minutes to denature the protein. Then, the denatured firefly luciferase (0.75 μM) was diluted 20-fold with buffer 1 containing 2 μM *E. coli* ClpB or pfClpB1, 6 mM ATP, 1 μM DnaK or human Hsp70, 1 μM DnaJ or human Hdj1 , and 0.5 μM

GrpE. Denatured firefly luciferase diluted with buffer 1 without the chaperones was used as a negative control. To measure firefly luciferase reactivation, aliquots from each reaction was pipetted into a 96 well plates followed by measurement with a luminometer.

Chapter 5 - In vitro Characterization of PfClpB1: Results and Discussion

PfClpB1 is a good target for the development of therapeutic drug because of its presence in the apicoplast, which is an organelle that is absent in mammalian. Adequate expression of PfClpB1 protein was the initial step for the biochemical characterization of pfClpB1.

Bioinformatics analysis

The tRNAs of each organism determine the codon bias of that population. Codon bias varies among different organisms. To determine how codon usage affects protein expression in PfClpB1, Graphical Codon Usage Analyzer was used. Some of pfClpB1 codons have low relative adaptiveness compared to E.Coli codons. Bioinformatics analysis reveals a low GC content in PfClpB1 and confirms the need for a special host system such as the Rosetta (DE3) Competent Cells to enhance protein production during expression.

The prediction of a signal peptide and an apicoplast targeting sequence suggest that the mature protein will be shorter. Since we do not fully understand the apicoplast cleavage mode of action, we propose that the N-terminus of the mature protein is close to Phe143. PfClpB1 sequence shifts from basic to acidic close to Phe143: a characteristic of transit peptide, and then sequence similarity between E. coli ClpB and PfClpB1 is visible in the region following that residue.

Multiple sequence alignment of PfClpB1 to Thermus thermophilus ClpB and E. coli ClpB confirms that PfClpB1 shares a lot of similarities with the member of the ClpB families. The N-terminal shows highly conserved regions after the signal peptide and the apicoplast targeting sequence. The C-terminal and the nucleotide binding domains are also well conserved. Multiple sequence analysis predicts that PfClpB1 would form an oligomer which is a characteristic of the AAA+ superfamily protein ATPase.

These first analyses are used as a premise to predict the final form of PfClpB1 after protein expression. This is important because we want to express the protein in its final form in E. coli. Homology modeling and multiple sequence alignment suggest that the mature PfClpB1 protein structure is similar to Thermus thermophilus ClpB. However, the 52 extra middle domain amino acids of PfClpB1 seem to be a hard fit in the predicted structure. The middle domain is

responsible for the specificity and binding with the KJE chaperone system of E.coli ClpB. Consequently, the 52 extra amino acid residues inserted in the middle domain could interfere with KJE chaperone cooperation.

Expression and purification of pfClpB1 in Rosetta strain of E.Coli

Expression of plasmodium genes is complex in *E.Coli* because of the codon bias and the presence of A-T rich genome. There are 2 ways in which full-length protein may be expressed: (i) by changing the rare codons of the Plasmodium gene to match *E. coli* through mutagenesis, or (ii) To engineer an *E. coli* strain with a plasmid that produces tRNAs that spots the rare codons. *E.coli* Rosetta (DE3) strain was used because it provides the tRNAs for the rare codons. The original Plasmid p15TV-L was replaced by pet28a (+) because we needed to introduce a 6xHis-tag codons on the N-terminal. DNA sequencing revealed that the insert in p15TV-L was a copy of pfClpB1 gene. Next, we purified the protein by Ni2+-affinity chromatography under native condition. The results show that the target protein efficiently binds to the column with a few non-specific bindings. The identity of pfClpB1 was confirmed by MS analysis. Then we used gel filtration for further purification. Overall, the results reveal that it is inherently hard to purify PfClpB1 protein in large quantity in its native form. However the protein expression level was enough for our biochemical studies.

ATP Hydrolysis of PfClpB1

Basal ATP hydrolysis of PfClpB1 was found to be similar to that of *E.Coli* ClpB. Poly-lysine and α-casein was found to stimulate PfClpB1, but at a lower rate than ClpB. These results indicate that PfClpB1 is involved in disaggregation activity. Work on the DnaJ proteins of Plasmodium falciparum (Pfj 1-4) would have aided to reveal to what extent Pfjs stimulates the ATPase activity of PfClpB1.

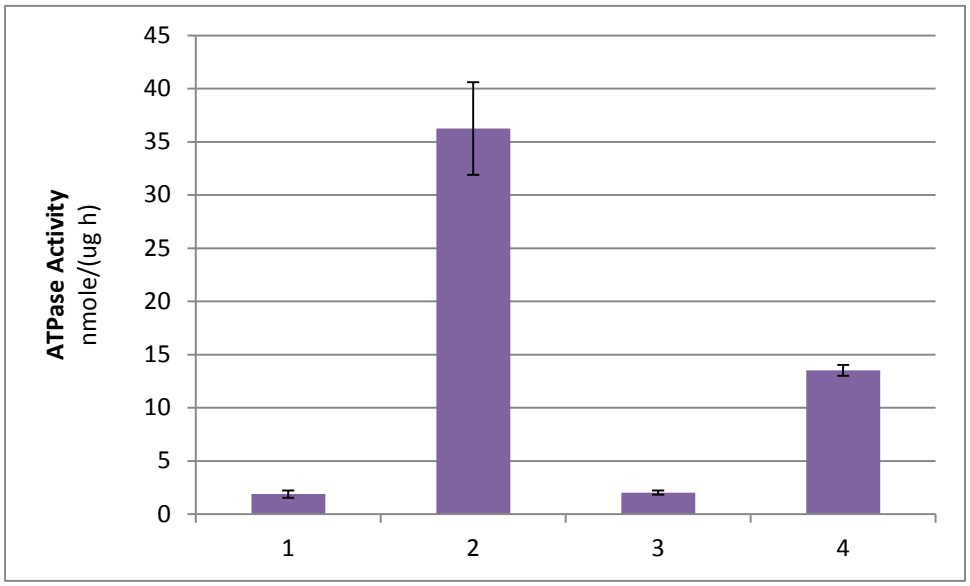


Figure 5.1 ATPase activity of PfClpB1 and *E. coli* ClpB in the presence of polylysine

The initial rate of hydrolysis of ATP catalyzed by PfClpB or EcClpB was determined at 37 degree Celsuis for basal activity and with 0.04 mg/ml poly-lysine. The average values from four separate experiments are shown with the standard deviations.
1: E.Coli ClpB. 2: E.Coli ClpB + polylysine. 3: PfClpB1. 4: PfClpB1 + + poly-lysine.

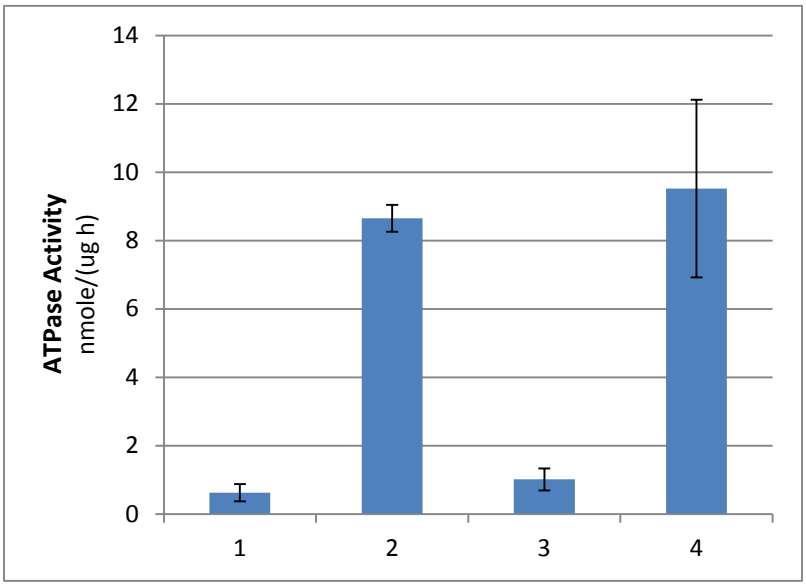


Figure 5. 2 ATPase activity of PfClpB1 and *E. coli* ClpB in the presence of α-casein

The initial rate of hydrolysis of ATP catalyzed by PfClpB or EcClpB was determined at 37 degree Celsuis for basal activity and with 0.2 mg/ml α-casein. The average values from four separate experiments are shown with the standard deviations.

1: E.Coli ClpB. 2: E.Coli ClpB + α-casein. 3: PfClpB1. 4: PfClpB1 + α-casein.

Sedimentation Velocity

The oligomerization of PfClpB1 was tested by sedimentation velocity. In the absence of nucleotide PfClpB1 stayed mainly monomeric. When ATPγS was provided, distributions of the sedimentation coefficients g(s*) shifted from 4.5S (monomeric) to 10S and 16.5S (oligomeric), indicating oligomers were the main component in the solution (Figure 4.1 and figure 4.2). Thus, it can be deducted that the oligomerization of PfClpB1occurs, but it is very dynamic process because the addiction of nucleotide pull it toward the oligomeric form. The data also suggests that there a two oligomeric form of PfClpB1. All those findings are consistent with the characteristic of members in the AAA+ family.

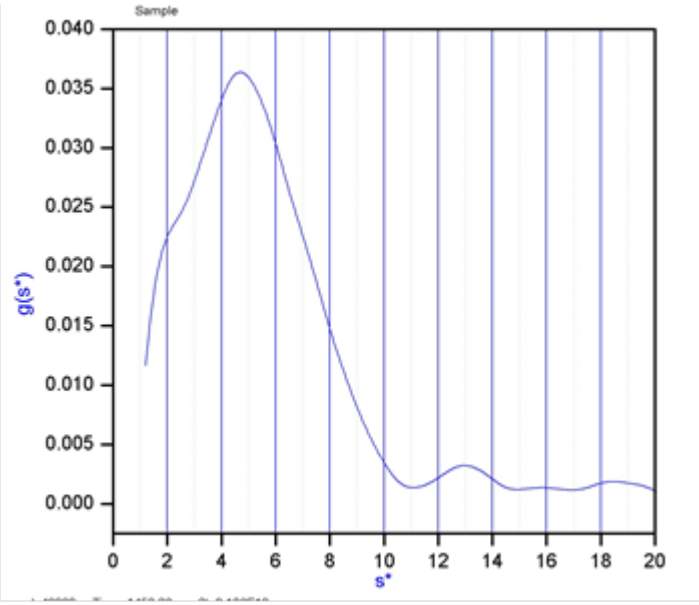


Figure 5.3 Sedimentation velocity analysis of pfClpB1.

Ultracentrifugation was performed at 49,000 rpm and 20 °C for the 0.7 mg/ml protein samples in 50 mM Tris/HCL pH 7.5, 0.2 M KCl, 20 mM MgCl2, 1 mM EDTA, 2 mM β-mercaptoethanol. The apparent sedimentation coefficient distributions g(s*) are shown.

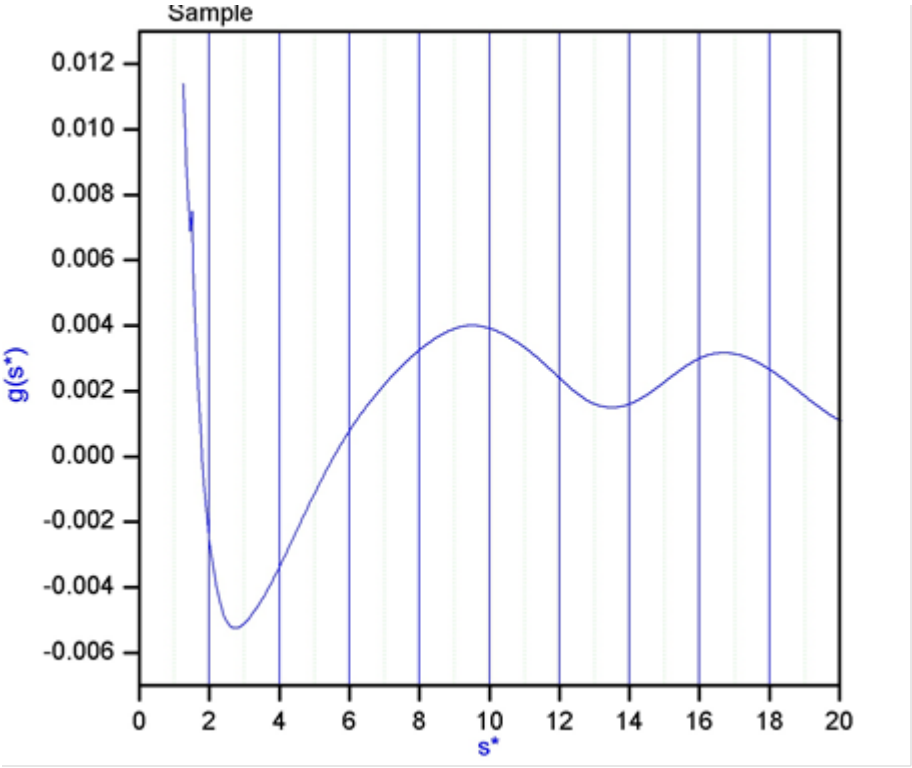


Figure 5.4 Sedimentation velocity analysis of pfClpB1 with ATPγS.
Ultracentrifugation was performed at 49,000 rpm and 20 °C for the 0.7 mg/ml protein samples in 50 mM Tris/HCL pH 7.5, 0.2 M KCl, 20 mM MgCl2, 1 mM EDTA, 2 mM β-mercaptoethanol with ATPγS. The apparent sedimentation coefficient distributions g(s*) are shown.

Protein reactivation assay

The mechanism of aggregate reactivation by pfClpB1 machinery in the presence of luciferase was studied. *E. coli* ClpB shows an increase in aggregate reactivation activity in the presence of the KJE chaperone system as expected. However pfClpB1 shows a decrease in aggregate reactivation activity in the presence of the KJE chaperone system. This can be explained by the extra sequence of amino acids in the middle domain of pfClpB1 which affect the specificity and binding with the KJE chaperone system. The results are not surprising, but support the important role played by the middle domain. Furthermore, the experiment also suggests that PfClpB1 binds with the KJE chaperone system, but it seems like there are

inhibiting each other because of the lack of specificity. Aggregate reactivation with pfClpB1 or KJE chaperone system alone was very successful. Further studies are required to understand the mechanism of aggregate reactivation in PfClpB1.

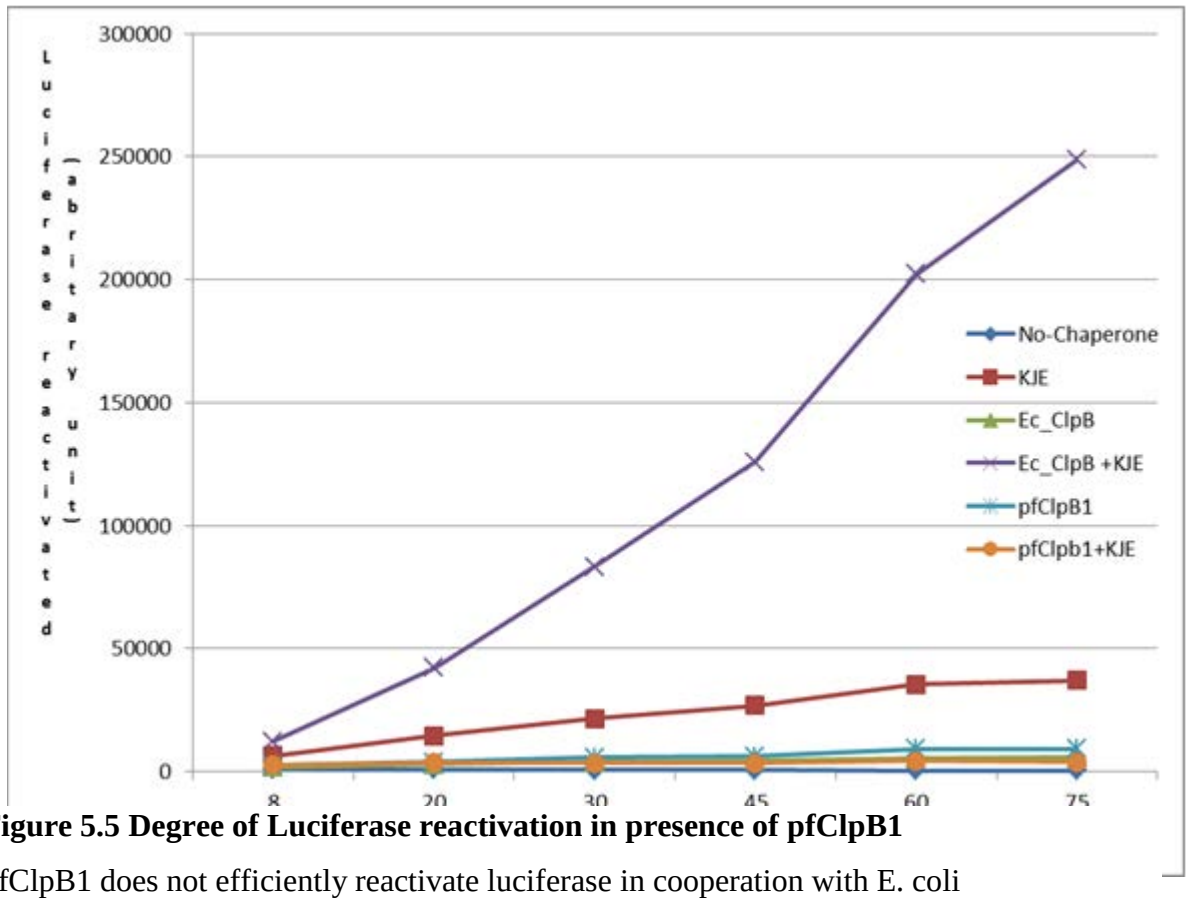


Figure 5.5 Degree of Luciferase reactivation in presence of pfClpB1

pfClpB1 does not efficiently reactivate luciferase in cooperation with E. coli DnaK/DnaJ/GrpE in presence of luciferase. Luciferase was denatured by heat at 45 ℃. Then pfClpB1 and co-chaperones were added. The reactivation of luciferase was monitored by luminometer.

No-chaperone: only buffer 1

KJE: DnaK, DnaJ, GrpE

Ec_ClpB: E.coli ClpB

Ec_ClpB+KJE: E.coli ClpB + KJE chaperone system

pfClpB1: plasmodium falciparum ClpB located to the apicoplast

pfClpB1+KJE: pfClpB1 + KJE chaperone system

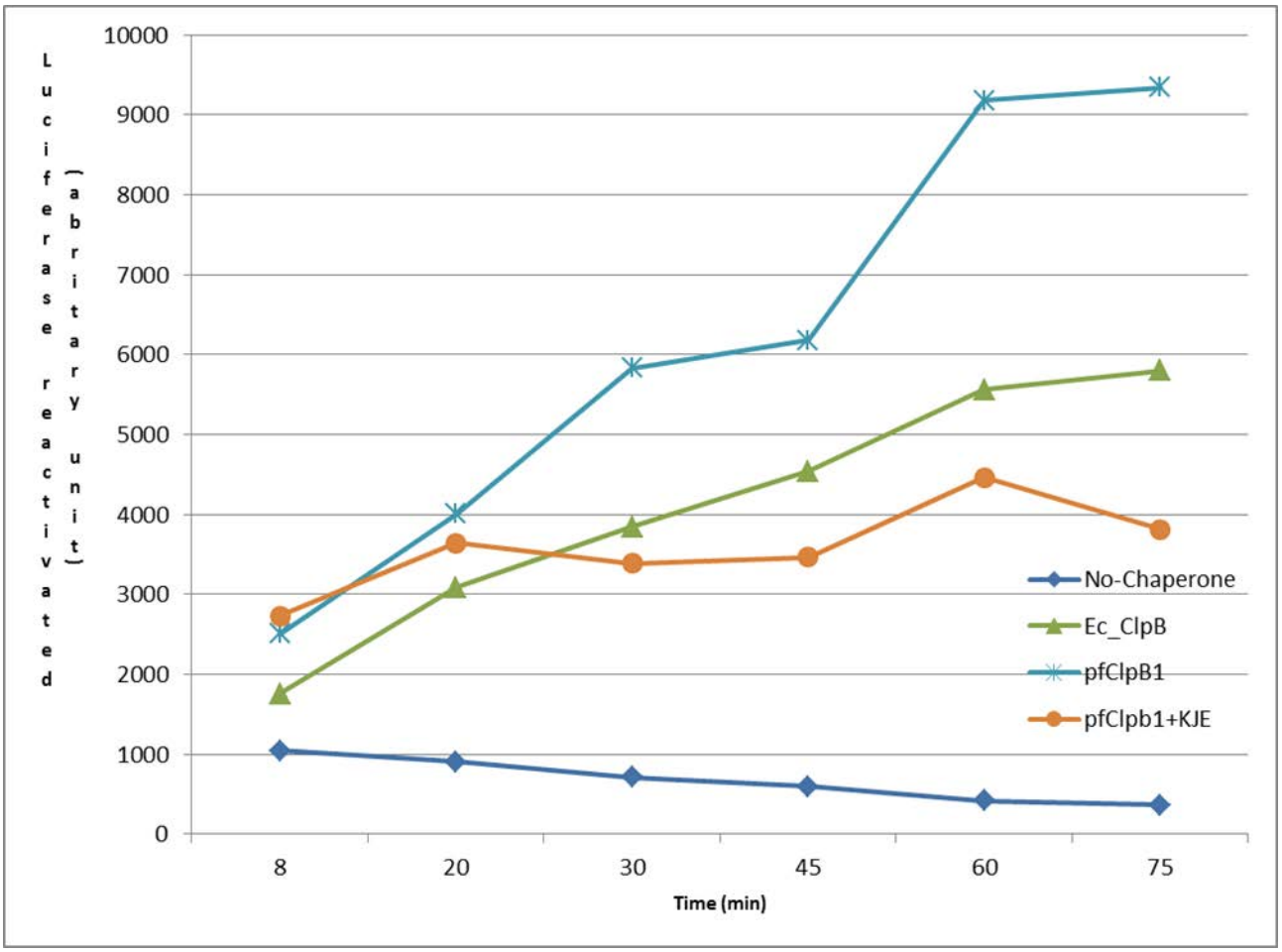


Figure 5.6 Degree of Luciferase reactivation in presence of pfClpB1 (Without E.coli ClpB + KJE chaperone system for a better scale)

pfClpB1 does not efficiently reactivate luciferase in cooperation with E. coli DnaK/DnaJ/GrpE in presence of luciferase. Luciferase was denatured by heat at 45 ℃. Then pfClpB1 and co-chaperones were added. The reactivation of luciferase was monitored by luminometer.

No-chaperone: only buffer 1

KJE: DnaK, DnaJ, GrpE

Ec_ClpB: E.coli ClpB

pfClpB1: plasmodium falciparum ClpB located to the apicoplast

pfClpB1+KJE: pfClpB1 + KJE chaperone system

Further works

This study suggests that PfClpB1 is able to reactivate disaggregated protein in presence of ATP. However, more work is required to figure out the ATP hydrolysis of PfClpB1 and protein reactivation in presence of the unknown apicoplast-targeted *Plasmodium falciparum* co-chaperones such as the Hsp70 and Hsp40 families. Our data may imply that PfClpB1 chaperone activity may help to preserve the activity of the apicoplast targeted protein under environmental stress, which may support the survival of Plasmodium. Lastly, a multiple sequence alignment reveals the presence of extra amino acid sequences in the middle domain, whose function is still uncharacterized. A better understanding of the 52 amino acid inserts in the middle domain of PfClpB1 can help to determine its function. In short, PfClpB1and other apicoplast chaperones may become good targets for the development of pharmaceutical drug therapy helping to eliminate Apicomplexa-borne diseases.

APPENDIX A: SUPPLEMENTARY RESULTS

Table: Preparation phosphate standard

2.5 mM Phosphate (µl)	**DiH_2O (µl)**	**concentration (nmole)**
0	50	0
1	49	2.5
2	48	5
3	47	7.5
4	46	10
5	45	12.5
6	44	15

References

1. Hempelmann E, Krafts K (2013). "Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria". Malar J. 12 (1): 213.

2. Lalremruata A, Ball M, Bianucci R, Welte B, Nerlich AG, Kun JF, Pusch CM. "Molecular identification of falciparum malaria and human tuberculosis co-infections in mummies from the Fayum depression (lower Egypt)". PLoS One 8(4):e60307.

3. Centers for Disease Control and Prevention (2013). “Malaria facts”. Retrieved from http://www.cdc.gov/malaria/about/facts.html

4. National Institute of Allergy and Infectious Diseases (2007). “Understanding Malaria: Fighting an Ancient Scourge”. Pp 6.

5. Rhian E. Hayward, Bela Tiwari, Karen P. Piper, Dror I. Baruch, and Karen P. Day (1999). “Virulence and transmission success of the malarial parasite Plasmodium falciparum” PNAS 96 (8): 4563–4568.

6. Centers for Disease Control and Prevention (2013). “Impact of Malaria”. Retrieved from http://www.cdc.gov/malaria/malaria_worldwide/impact.html

7. World Health Organization (2013). “Malaria cases of and deaths”. Retrieved from http://www.who.int/gho/mdg/diseases/malaria/situation_trends_mortality/en/index.html

8. National Institute of Allergy and Infectious Diseases (2007). “Understanding Malaria: Fighting an Ancient Scourge”. Pp 20.

9. Naman K Shah, Gajender P S Dhillon, Adtiya P Dash et al (2011). “Antimalarial drug resistance of Plasmodium falciparum in India: changes over time and space”. Lancet Infect Dis. 11(1): 57–64.

10. Malcolm J. Gardner, et al (2002). “Genome sequence of the human malaria parasite Plasmodium falciparum”. Nature 419, 498-511

11. Stuart A. Ralph, et al (2004). “Metabolic maps and functions of the Plasmodium falciparum Apicoplast”. Nature reviews 2: 205-216

12. Liting Lim, et al (2010). “The evolution, metabolism and functions of the apicoplast”. Phil. Trans. R. Soc. B 365:749–763

13. National Institute of Allergy and Infectious Diseases (2007). “Understanding Malaria: Fighting an Ancient Scourge”. Pages 8-11.

14. K. Motohashi, Y. Watanabe, M. Yohda, M. Yoshida (1999). “Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones” Proc. Natl. Acad. Sci. U. S. A. 96: 7184–7189

15. M. Zolkiewski, (1999). “ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli, J”. Biol. Chem. 274: 28083–28086.

16. A. Mogk, T. Tomoyasu, P. Goloubinoff, S. Rudiger, D. Roder, H. Langen, B. Bukau (1999) “Identification of thermolabile E. coli proteins: prevention and reversion of aggregation by DnaK and ClpB” EMBO J. 18: 6934–6949.

17. Schirmer, E.C., Glover, J.R., Singer, M.A., and Lindquist, S. (1996). “Hsp100/Clp proteins: a common mechanism explains diverse functions”. Trends Biochem. Sci. 21: 289–296.

18. Zolkiewski, M. (2006). “A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases”. Molecular microbiology 61: 1094-1100.

19. Woo KM, Kim KI, Goldberg AL, Ha DB, Chung CH (1992) “The heat-shock protein ClpB in Escherichia coli is a protein-activated ATPase”. J Biol Chem 267: 20429–20434.

20. Wendler P., Ciniawsky S., Kock M., Kube S. (2012). “Structure and function of the AAA+ nucleotide binding pocket”. Biochem. Biophys. Acta. 1823:2–14

21. Walker J.E., Saraste M., Runswick M.J., Gay N.J. (1982) “Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold”. EMBO J. 1:945–951.

22. Phyllis I. Hanson1 & Sidney W. Whiteheart (2005) “AAA+ proteins: have engine, will work” Nature Reviews Molecular Cell Biology 6: 519-529

23. Neuwald, A. F., L. Aravind, et al. (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res 9(1): 27-43.

24. Tucker, P.A., Sallai, L. (2007). “The AAA+ superfamily – a myriad of motions”. Curr. Opin. Struct. Biol. 17: 641–652.

25. Skye Hodson, Jacqueline J.T. Marshall, Steven G. Burston (2012). “Mapping the road to recovery: The ClpB/Hsp104 molecular chaperone”. J Struct Biol. 179(2):161-7

26. DeSantis, M.E., Shorter, J. (2012). “The elusive middle domain of Hsp104 and ClpB: location and function”. Biochem. Biophys. Acta 1823: 29–39.

27. Haslberger, T., Weibezahn, J., Zahn, R., Lee, S., Tsai, F.T.F., et al. (2007). “M domains couple the ClpB threading motor with the DnaK chaperone activity”. Mol. Cell 25: 247–260.

28. Haslberger, T., J. Weibezahn, et al. (2007). "M domains couple the ClpB threading motor with the DnaK chaperone activity." Mol Cell 25(2): 247-260.

29. Weibezahn, J., P. Tessarz, et al. (2004). "Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB." Cell 119(5): 653-665.

30. Hung, G. C. and D. C. Masison (2006). "N-terminal domain of yeast Hsp104 chaperone is dispensable for thermotolerance and prion propagation but necessary for curing prions by Hsp104 overexpression." Genetics 173(2): 611-620.

31. Sielaff, B. and F. T. Tsai (2010). "The M-domain controls Hsp104 protein remodeling activity in an Hsp70/Hsp40-dependent manner." J Mol Biol 402(1): 30-37.

32. Park, S. K., K. I. Kim, et al. (1993). "Site-directed mutagenesis of the dual translational initiation sites of the clpB gene of Escherichia coli and characterization of its gene products." J Biol Chem 268(27): 20170-20174.

33. Liu, Z., V. Tek, et al. (2002). "Conserved amino acid residues within the amino-terminal domain of Clp are essential for the chaperone activity.” J Mol Biol 321 (1): 111-120.

34. del Castillo, U., C. Alfonso, et al. (2011). "A quantitative analysis of the effect of nucleotides and the M domain on the association equilibrium of ClpB." Biochemistry 50(12): 1991-2003.

35. Weibezahn, J., P. Tessarz, et al. (2004). "Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB." Cell 119(5): 653-665.

36. Zietkiewicz, S., A. Lewandowska, et al. (2006). "Hsp70 chaperone machine remodels protein aggregates at the initial step of Hsp70-Hsp100-dependent disaggregation." J Biol Chem 281(11): 7022-7029.

37. Haslberger, T., J. Weibezahn, et al. (2007). "M domains couple the ClpB threading motor with the DnaK chaperone activity." Mol Cell 25(2): 247-260.

38. Sielaff, B. and F. T. Tsai (2010). "The M-domain controls Hsp104 protein remodeling activity in an Hsp70/Hsp40-dependent manner." J Mol Biol 402(1): 30-37.

39. Miot, M., M. Reidy, et al. (2011). "Species-specific collaboration of heat shock proteins (Hsp) 70 and 100 in thermotolerance and protein disaggregation." Proc Natl Acad Sci USA 108(17): 6915-6920.

40. Barnett, M. E., A. Zolkiewska, et al. (2000). "Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains." J Biol Chem 275(48): 37565-37571.

41. Mogk, A., C. Schlieker, et al. (2003). "Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity." J Biol Chem 278(20): 17615-17624.

42. Mackay, R. G., C. W. Helsen, et al. (2008). "The C-terminal extension of Saccharomyces cerevisiae Hsp104 plays a role in oligomer assembly." Biochemistry 47(7): 1918-1927.

43. Lo, J. H., T. A. Baker, et al. (2001). "Characterization of the N-terminal repeat domain of Escherichia coli ClpA-A class I Clp/HSP100 ATPase." Protein Sci 10(3): 551-559.

44. Schlieker, C., I. Tews, et al. (2004). "Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides." FEBS Lett 578(3): 351-356.

45. Hoskins, J. R., S. M. Doyle, et al. (2009). "Coupling ATP utilization to protein remodeling by ClpB, a hexameric AAA+ protein." Proc Natl Acad Sci USA 106(52): 22233-22238.

46. Morimoto, R. I., Kline, M. P., Bimston, D, N., and Cotto, J. J. (1997). “The heat-shock response: regulation and function of heat-shock proteins and molecular chaperones.” In Essays in Biochemistry volume 32 (Bowles, D. J., ed) pp.17-27, Portland and Press, London

47. Hsp104. Parsell DA, Kowal AS, Singer MA, Lindquist S. (1996). “Protein disaggregation mediated by heat-shock protein Source” Nature. 372(6505):475-478.

48. Zolkiewski M. (1999). “ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli.” J Biol Chem. 274(40):28083-28086.

49. Fink, A. L. (1998). "Protein aggregation: folding aggregates, inclusion bodies and amyloid." Fold Des 3(1): R9-23.

50. Lee, S., M. E. Sowa, et al. (2003). "The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state." Cell 115(2): 229-240.

51. Akoev, V., E. P. Gogol, et al. (2004). "Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB." Protein Sci 13(3): 567-574.

52. Hartl FU. (1996). “Molecular chaperones in cellular protein folding.” Nature 381: 571

53. Garbe TR, Hibler NS, Deretic V. (1999). “Response to reactive nitrogen intermediates in Mycobacterium tuberculosis: induction of the 16 kDa α-crystallin homolog by exposure to nitric oxide donors.” Infect Immun 67: 460.

54. Kohler S, Ekaza E, Paquet JY, et al. (2002). “Induction of DnaK through its native heat shock promoter is necessary for intramacrophagic replication of Brucella suis.” Infect Immun 70: 1631

55. Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK. (2001). “Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding α-crystallin.” Proc Natl Acad Sci USA 98: 7534

56. Krobitsch S, Clos J. (1999). “A novel role for 100 kD heat shock proteins in the parasite Leishmania donovani.” Cell Stress Chaperones 4(3):191-198.

57. Hübel A, Krobitsch S, Hörauf A, and Clos J. (1997). “Leishmania major Hsp100 is required chiefly in the mammalian stage of the parasite.” Mol Cell Biol. 17(10): 5987–5995.

58. Frees D, Chastanet A, Qazi S, Sørensen K, Hill P, Msadek T, Ingmer H. (2004). “Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in Staphylococcus aureus.” Mol Microbiol 54: 1445–1462.

59. Capestany CA, Tribble GD, Maeda K, Demuth DR, Lamont RJ. (2008). “Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in Porphyromonas gingivalis.” J Bacteriol. 190(4): 1436-46.

60. Chastanet A, Derre I, Nair S. & Msadek T. (2004). “ClpB, a novel member of the Listeria monocytogenes CtsR regulon, is involved in virulence but not in general stress tolerance.” J Bacteriol 186: 1165–1174.

61. Yuan L, Rodrigues PH, Bélanger M, Dunn W Jr, Progμl ske-Fox A. (2007). “The Porphyromonas gingivalis clpB gene is involved in cellular invasion in vitro and virulence in vivo.” FEMS Immunol Med Microbiol 51: 388–398.

62. Lourdaμl t K, Cerqueira GM, Wunder EA Jr, Picardeau M. (2011). “Inactivation of clpB in the pathogen Leptospira interrogans reduces virulence and resistance to stress conditions.” Infect Immun. 79(9):3711-3717.

63. LaFrentz, B. R., S. E. LaPatra, et al. (2011). "Identification of immunogenic proteins within distinct molecular mass fractions of Flavobacterium psychrophilum." J Fish Dis 34(11): 823-830.

64. Kilejian A. (1975). “Circular mitochondrial DNA from the avian malarial parasite Plasmodium lophurae.” Biochim Biophys Acta 390:276–284.

65. Moore RB, Obornik M. et al. (2008). “A photosynthetic alveolate closely related to apicomplexan parasites.” Nature 451:959–963.

66. Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ. (2010). “A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids.” Proc Natl Acad Sci USA 107:10949–10954.

67. Waller RF, Keeling PJ. et al.(1998). “Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum.” Proc Natl Acad Sci. USA 95:12352–12357

68. Ralph SA, van Dooren GG. et al. (2004). “Metabolic maps and functions of the Plasmodium falciparum apicoplast.” Nat Rev Microbiol 2(3):203–216.

69. Seeber F. (2003). “Biosynthetic pathways of plastid-derived organelles as potential drug targets against parasitic apicomplexa.” Curr Drug Targets Immune Endocr Metabol Disord 3:99–109

70. Marilyn Parsons, Anuradha Karnataki, Jean E. Feagin and Amy DeRocher (2007). “Protein Trafficking to the Apicoplast: Deciphering the Apicomplexan Solution to Secondary Endosymbiosis.” Eukaryot Cell. 6(7): 1081–1088.

71. Goodman, C.D., Su, V. and McFadden, G.I. (2007). “The effects of anti-bacterials on the malaria parasite Plasmodium falciparum.” Mol. Biochem. Parasitol. 152: 181–191.

72. Vaughan A, O’Neill M, Tarun A, Camargo N, Phuong T, Aly A, Cowman A and Kappe S. (2009). “Type II fatty acid synthesis is essential only for malaria parasite late liver stage development.” Cell. Microbiol. 11: 506–520.

73. Yu M, Kumar TR, Nkrumah LJ, Coppi A. et al. (2008). “The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites.” Cell Host Microbe 4:567–578.

74. Ralph SA, van Dooren, GG. et al. (2004). “Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast.” Nature Reviews Microbiology 2(3):203–216.

75. Ralph SA, van Dooren, GG. et al. (2004). “Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast.” Nature Reviews Microbiology 2(3):203–216.

76. El Bakkouri M, Pow A. et al. (2010). “The Clp chaperones and proteases of the human malaria parasite Plasmodium falciparum.” J Mol Biol. 404(3):456-477.

77. Ralph SA, van Dooren GG. et al. (2004). “Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast.” Nat Rev Microbiol. 2(3):203-216.

78. Goodman CD, Su V, McFadden GI. (2007). “The effects of anti-bacterials on the malaria parasite Plasmodium falciparum.” Mol Biochem Parasitol. 152(2):181-191.

79. El Bakkouri M, Pow A. et al. (2010). “The Clp chaperones and proteases of the human malaria parasite Plasmodium falciparum.” J Mol Biol. 404(3):456-477.

80. Marti M, Good RT, Rug M. et al. (2004). “Targeting malaria virulence and remodeling proteins to the host erythrocyte.” Science. 306(5703):1930-1933.

81. Hiller NL, Bhattacharjee S, van Ooij C. et al. (2004). “A host-targeting signal in virulence proteins reveals a secretome in malarial infection.” Science. 306(5703):1934-1937.

82. De Koning-Ward TF, Gilson PR, Boddey JA. et al. (2009). “A newly discovered protein export machine in malaria parasites.” Nature. 459(7249):945-949.

83. Bullen HE, Charnaud SC. et al. (2012). “Biosynthesis, localization, and macromolecular arrangement of the Plasmodium falciparum translocon of exported proteins (PTEX).” J Biol Chem. 287(11):7871-7884.

84. De Koning-Ward TF, Gilson PR, Boddey JA. et al. (2009). “A newly discovered protein export machine in malaria parasites.” Nature. 459(7249):945-949.

85. Doyle SM, Hoskins JR, Wickner S. (2007). “Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system.” Proc Natl Acad Sci U S A. 104(27):11138-11144.

86. Tuteja R. (2007). “Malaria: an overview.” FEBS J. 274(18):4670-4679.