Structural and Functional Characterization of Clp Chaperones and Proteases in the Human Malaria Parasite *Plasmodium falciparum*

By

Andre Pow

A thesis submitted in conformity with the requirements for the degree of Master of Science

Department of Biochemistry

University of Toronto

© Copyright by Andre Pow (2012)
Structural and Functional Characterization of Clp Chaperones and Proteases in the Human Malaria Parasite \textit{Plasmodium falciparum}

Master of Science

2012

Andre Pow

Department of Biochemistry

University of Toronto

Abstract

The Clp chaperones and proteases play a pivotal role in maintaining cellular homeostasis. They are highly conserved across prokaryotes and can also be found in the mitochondria of eukaryotes and chloroplast of plants. For my thesis, I provide an analysis of the Clp chaperones and protease in the human malaria parasite \textit{Plasmodium falciparum}. The parasite contains four Clp ATPases, which I term \textit{Pf}ClpB1, \textit{Pf}ClpB2, \textit{Pf}ClpC, and \textit{Pf}ClpM. One \textit{Pf}ClpP, the proteolytic protomer, and one \textit{Pf}ClpR, an inactive isoform, were also identified. All proteins, with the exception of \textit{Pf}ClpB2, were found to be localized to the apicoplast, a non-photosynthetic relic plastid in \textit{P. falciparum}. Both \textit{Pf}ClpP and \textit{Pf}ClpR form mostly homoheptameric rings as observed by various techniques. Through X-ray crystallography, \textit{Pf}ClpP assumed a compacted tetradecamer structure similar to that observed for other ClpPs. My data suggest the presence of a ClpCRP complex in the apicoplast of \textit{P. falciparum}. 
# Table of Contents

**Abstract** ii  
**Table of Contents** iii  
**List of Tables** v  
**List of Figures** vi  
**List of Abbreviation** vii  
1. Introduction 1  
  1.1. Multi-subunit chaperone-protease complexes 2  
  1.2. The AAA+ Superfamily of Proteins 2  
  1.3. The Clp Chaperone Family 5  
    1.3.1 ClpX 9  
    1.3.2 ClpA 10  
    1.3.3 ClpC 11  
    1.3.4 ClpB 12  
  1.4. The Protease ClpP 13  
  1.5 Proteasome 23  
  1.6 Malaria 25  
    1.6.1 Tissue schizogony (pre-erythrocytic schizogony) 27  
    1.6.2 Erythrocytic schizogony 30  
    1.6.3 Sporogony 31  
    1.6.4 Apicoplast 34  
  1.7. Thesis Rationale 42  
2. Materials and Methods 43  
  2.1 Plasmid constructs 44  
  2.2 Expression and Purification 44  
  2.3 Antibodies for ClpP and ClpR 46  
  2.4 Preparation of *P. falciparum* parasitized erythrocyte at mature stages 46  
  2.5 Western Blot and Saponin Treatment 47  
  2.6 Peptidase assays 48
2.7 Protease Assay 48
2.8 GFP-SsrA Degradation Assay 49
2.9 Size-exclusion chromatography 49
2.10 Analytical Ultracentrifugation – Sedimentation Equilibrium 50
2.11 Transmission Electron Microscopy 50
2.12 STEM – Scanning Transmission Electron Microscopy 50
2.13 Homology Modeling 51
2.14 Tagging of PfClpC, PfClpB1, and PfClpB2 gene products by 3’ replacement 52
2.15 Immunofluorescence assays 52
2.16 Surface Plasmon Resonance experiments 53
2.17 X-ray crystallography 53

3. Results and Discussion 56
3.1. Identification of putative clp genes in Plasmodium falciparum 57
3.2. Apicoplast targeting predictions 64
3.3. Experimental localization of the Clp proteins in P. falciparum. 67
3.4. Sequence boundaries of the mature forms of PfClpP and PfClpR 72
3.5. Oligomeric state of N-terminally truncated PfClpP and PfClpR 77
3.6. The interaction between PfClpP(179-370) and PfClpR(49-244) 85
3.7. Peptidase activity of PfClpP(179-370) 89
3.8. Structure of PfClpP(179-370) 93
3.9. Modeling of the ClpR Monomer 98

4. Discussion and Future Directions 102
4. Discussion and Future Directions 103

5. References 108
5. References 109
List of Tables

Table 3.1. The Clp proteins of *Plasmodium falciparum*. 59
Table 3.2. Apicoplast targeting prediction of PfClp proteases and ATPases. 66
Table 3.3. Data collection and structure refinement statistics for H6-PfClpP (179-370). 97
List of Figures

Figure 1.1. The domain structures of the Clp ATPases. 8
Figure 1.2. *E.coli* ClpP structure. 14
Figure 1.3. Clp ATPases and Protease 17
Figure 1.4. Life cycle of *Plasmodium falciparum*. 29
Figure 1.5. Anatomy of a RBC infected by *Plasmodium falciparum*. 33
Figure 1.6. Protein targeted to the apicoplast. 36
Figure 3.1. Sequence properties of the *P. falciparum* Clp proteases. 62
Figure 3.2. Domain arrangement of the *P. falciparum* Clp ATPases. 63
Figure 3.3. Expression and localization of the *Pf*Clp proteins. 70
Figure 3.4. Mature *Pf*ClpP and *Pf*ClpR in *Plasmodium falciparum*. 75
Figure 3.5. Oligomeric state of *Pf*ClpP(179-370) and *Pf*ClpR(49-244). 81
Figure 3.6. Electron microscopy analysis of *Pf*ClpP and *Pf*ClpR. 83
Figure 3.7. SPR analysis of the interaction between *Pf*ClpP(179-370) and *Pf*ClpR(49-244). 87
Figure 3.8. Peptidase activity of *Pf*ClpP(179-370). 91
Figure 3.9. The X-ray structure of H6-*Pf*ClpP(179-370). 95
Figure 3.10. Modeling of the ClpR subunit. 100
List of Abbreviation

aa Amino acid
AAA+ ATPases associated with a diverse cellular activities plus
ACP Acyl carrier protein
ADP Adenosine 5'-diphosphate
ASCE Additional strand catalytic E
ATP Adenosine 5'-triphosphate
AUC-SE Analytical Ultracentrifugation - Sedimentation Equilibrium
BSA Bovine serum albumin
CC Coiled coil
Clp Caseinolytic protease
COG Clusters of Orthologous Groups
D1 First AAA+ domain
D2 Second AAA+ domain
DG Dense granules
DNA Deoxyribonucleic acid
DTT Dithiothreitol
EDTA Ethylenedinitrilotetraacetic acid
ELISA Enzyme-Linked ImmunoSorbent Assay
EM Electron microscopy
ER Endoplasmic reticulum
FPLC Fast protein liquid chromatography
FV Food vacuoles
GFP Green fluorescent protein
HA Hemagglutinin
Hepes 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP Horseradish peroxidase
Hsl Heat shock locus
Hsp Heat shock protein
HV 6x His sequence followed by a tobacco etch virus recognition sequence
IPTG Isopropyl -D-1-thiogalactopyranoside
MD Middle Domain
Mic Micronemes
NCBI National Centre for Biotechnology Information
Ni-NTA Nickel-nitrilotriacetic acid
NMR Nuclear magnetic resonance spectroscopy
O/N overnight
PATS predict apicoplast-targeted sequences
PBS Phosphate-buffered saline
PCR Polymerase chain reaction
PDB Protein Data Bank
PSORT Prediction of Protein Sorting Signals and Localization Sites in
Amino Acid Sequences

RBC  Red blood cells (erythrocyte)
RNA  Ribonucleic acid
RPMI Roswell Park Memorial Institute
SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SP  Signal peptide
SPP  Stromal Processing Peptidase
SPR  Surface Plasmon Resonance
SsrA Small stable RNA A
STEM Scanning transmission electron microscope
TEV  Tobacco etch virus
TIC  Translocon of outer envelope of chloroplast
TOC  Translocon of inner envelope of chloroplast
TP  Transit peptide
WT  Wildtype
ZBD  Zinc binding domain
Chapter 1. INTRODUCTION
1.1. Multi-subunit chaperone-protease complexes

The presence of proteases is vital for cellular homoeostasis in almost all organisms. Short-lived regulatory and misfolded proteins must be degraded for proper cellular function. Much of this proteolytic activity is carried out by energy-dependent proteases (Wickner, Maurizi et al. 1999). They share a multi-subunit, compartmentalized and two-component architecture in which the proteases associate with molecular chaperones. Molecular chaperones in these systems are proteins that recognize and typically assist in the unfolding of proteins. These unfoldase chaperones that associate with proteases are part of the ATPases Associated with diverse cellular Activities (AAA+) superfamily of proteins. Together the chaperones and proteases form a coordinated proteolytic system in which substrates are recognized, unfolded, and translocated by the chaperone into the protease’s chamber for degradation. In eukaryotic cells, the predominant protease is the proteasome which is located in both the nucleus and the cytosol. In prokaryotic cells, the list of proteases includes ClpXP, ClpAP, and ClpYQ (also known as HslUV) proteases. These systems consist of an unfoldase chaperone (ClpX, ClpA, and ClpY) and a cylindrical protease (ClpP and ClpQ).

1.2. The AAA+ Superfamily of Proteins

The AAA+ superfamily of proteins is a unique lineage of the much larger class of P-loop NTPases. Members of the AAA+ family are found across all organisms where they are essential for various cellular functions such as protein degradation, vesicular fusion, control of cell division and gene expression, and biogenesis of peroxisome and endosome (Erdmann et al., 1991).
The AAA+ superfamily is a major structural group of the P-loop NTPases referred to as additional strand catalytic E (ASCE) (Erdmann et al., 1991). Members of this group contain a core αβα nucleotide-binding domain which contains two major motifs referred to as Walker A and Walker B, involved in ATP binding and hydrolysis respectively (Walker, Saraste et al., 1982). Similar to other members of ASCE, AAA+ proteins typically function as hexameric rings (Iyer, Leipe et al., 2004). AAA+ proteins contain a highly conserved domain of 200-250 amino acids referred to as the “AAA+ module”. AAA+ proteins can have multiple AAA+ modules. Each AAA+ module is comprised of two discrete sub domains: a core αβα nucleotide-binding subdomain consisting of parallel β-sheets sandwiched between two sets of α-helices and a smaller α-helical subdomain consisting of two helical hairpins arranged in a left-handed, superhelical structure (Ogura and Wilkinson 2001).

The hallmarks of AAA+ proteins are the highly conserved Walker A (GX4GKT) and Walker B (RX4Φ4DE) motifs in each ATPase domain, where X represents any amino acid and Φ represents hydrophobic amino acids (Walker, Saraste et al., 1982). The Walker A motif lies between the first strand of the core β-sheet β1 and the subsequent helix α1. It facilitates nucleotide binding and metal-ion coordination (Erzberger and Berger 2006). The Walker A motif binds the β and γ-phosphate of bound ATP and Mg2+ ion through the conserved Lys and Thr/Ser residues (Neuwald, Aravind et al., 1999). The Walker B motif is associated with the third strand of the core β-sheet. It facilitates ATP hydrolysis and metal-ion coordination through the carboxylate of Asp and Glu respectively (Leipe, Koonin et al., 2003).
In addition to Walker A and Walker B motifs, the AAA+ module also contains other motifs that are not characteristic of P-loop NTPases. These motifs include sensor 1, sensor 2, and “box” sequences. Sensor 1 is found on the fourth core β-strand after the Walker B motif and is characterized by a conserved polar residue, which can be Asn, Asp, Thr, or Ser, that senses nucleotide binding and hydrolysis. The sensor 1 motif has been postulated to interact directly with the γ-phosphate of ATP for nucleotide binding and hydrolysis (Story and Steitz 1992). It has also been hypothesized to act in concert with the Glu residue of the Walker B motif to properly orient a water molecule for subsequent nucleophilic attack on the γ-phosphate of ATP (Story and Steitz 1992). The Sensor 2 is found in the third helix of the small α-helical domain and contains a conserved Arg residue. This residue is proposed to interact with the γ-phosphate of bound ATP substrate (Neuwald, Aravind et al., 1999).

Lastly, the ‘Box’ motifs include Box II and Box VII. Box II motif is located on the first helix before the core β-sheet and is hypothesized to be involved in adenine recognition. The Box VII motif is located at the amino terminus of the fifth β-strand. It contains an arginine finger that is proposed to be necessary for ATP hydrolysis and plays an important role in inter AAA+ subunit communication (Davey, Indiani et al., 2003).

In terms of functionality, the AAA+ superfamily exhibits a tremendous mechanistic variety within its ATP machinery. ATP binding, hydrolysis and sensing are all mediated by different motifs and sequence elements within the AAA+ module. The current mechanistic model for AAA+ is that ATP hydrolysis is hypothesized to initiate at the Walker B motif through proton abstraction from a molecule of water by the catalytic Glu residue. The deprotonated water subsequently performs a nucleophilic attack on the
γ-phosphate of bound ATP. The conserved Lys and Ser/Thr residues of the Walker A motif serve to bind the β- and γ-phosphates of the bound nucleotide and the Mg$^{2+}$ ion, respectively (Walker, Saraste et al., 1982). During the entire hydrolysis process, the N-terminal and C-terminal domains of the AAA+ module are postulated to move relative to one another, generating a mechanical force to affect remodeling events in associated molecules (Ogura and Wilkinson 2001).

1.3. The Clp Chaperone Family

Clp stands for Caseinolytic Protease as initially identified Clp proteins were found to degrade the protein substrate casein commonly found in milk (Katayama, Gottesman et al., 1988). However, Clp also stands for chaperone-like proteins since some Clp proteins were found to be ATPases that have chaperone-like activity. Subsequently, it was realized that Clp proteases act together in complex with Clp ATPases. Clp proteins are ubiquitous in most living organisms. In bacteria, they are found in the cytoplasm. In eukaryotes, they can be found in the cytoplasm or residing in organelles such as the mitochondria and chloroplasts.

The Clp chaperones can be categorized into two classes (Figure 1.1). Members of the first class, Class I, which include the AAA+ ATPases ClpA, ClpB, ClpC, ClpD, ClpE, and ClpL, are characterized by having two ATP nucleotide-binding (AAA+) domains, which are termed D1 and D2. Members of Class I, except for ClpA and ClpL, also contain a middle coiled coil domain which separates the D1 from D2. Members of the Class II Clp ATPase contain only one AAA+ domain which is similar to the second AAA+ domain (D2) of Class I. Members of the Class II include both ClpX and ClpY.
(Figure 1.1). Only ClpA, ClpB, ClpX and ClpY are present in *E. coli*. ClpC can be found in cyanobacteria, plants, and most Gram positive bacteria. ClpD has only been found in chloroplast of plants thus far, and both ClpE and ClpL are found in certain Gram positive bacteria.

Except for ClpB and ClpL, all Clp ATPases are discovered to have associative cognate protease partners; ClpA, ClpC, ClpE, and ClpX all form complexes with the serine protease ClpP; ClpY forms complexes with the threonine protease ClpQ. Clp ATPases serve as molecular chaperones and as regulatory guardian components of the Clp proteolytic complex. Clp ATPases are involved in unfolding globular protein substrates for subsequent degradation and concurrently regulate access of substrates to its associated proteolytic machinery (Wawrzynow, Banecki *et al.*, 1996).

Briefly, the structure of the associated Clp proteases is that of double stacked rings of identical subunits that can have either seven fold or six fold symmetries. ClpP exhibits seven fold symmetry with 2 heptameric rings stacked coaxially forming a tetradecamer. ClpQ has a similar structure as ClpP except that each ring exhibits six fold symmetry. Much like their proteasome cousins, the Clp proteases have an internal chamber where the catalytic resides for proteolytic activities are located. Access to this chamber is regulated by the narrow axial pores of the protease, as well as, the associated Clp ATPases. On their own, the proteases are capable of degrading small peptides that are presumably diffused into the Clp protease chamber through its axial pores. The association of protease with its Clp ATPase partner increases the rate of proteolysis of small peptides. Clp chaperones bind to one or both ends of the symmetric protease, acting as gate keepers to the proteolytic chamber.
The clpP is found upstream of clpX on the same operon in bacteria, but not in
eukaryotes; Human mitochondrial clpP and clpX genes are located on different
chromosomes (Corydon, Bross et al., 1998). Clp ATPases form homo-hexamers in a
nucleotide dependent manner. Given that the Clp proteases have seven fold symmetry,
there is a symmetry mismatch between the double heptameric ClpP and the hexameric
Clp ATPase ring. The reason for such asymmetry remains enigmatic, but it has been
proposed that the mismatch is an evolutionary result in which the rings reciprocate or
rotate about each other enhancing the rate of translocation of unfolded substrates from the
ATPase to the protease (Buron, Maurizi et al., 1998), although this has not been
experimentally observed. The same symmetry mismatch is also observed in the 20S-19S
proteasome complex.

Various structural loops are important in mediating the interaction between ClpP
and its ATPases. There are ClpP interaction loop in the ATPases which has the consensus
motif [LIVMP]G[FL]. There are also N-terminal axial loops on ClpP which surround the
axial pores of the protease protruding from the ClpP apical surface to interact with its
ATPases partner. Our group has previously discovered that both the ClpP interaction loop
in the ATPases and the axial loops in ClpP are essential for the formation of ATPase-
ClpP complex (Figure 1.3) (Gribun, Kimber et al., 2005). It has also been demonstrated
that two different ATPases may associate simultaneously with each side of the ClpP
cylinder in vitro (Ortega, Lee et al., 2004). I will briefly summarize the characteristics of
the following Clp ATPases: ClpX, ClpA, ClpC, and ClpB.
Figure 1.1. The domain structures of the Clp ATPases. Cartoon representation of the domain organization of Class I and Class II Clp ATPases. The domains are labeled and colored accordingly. Orange represents an insertion sequence. CC refers to the presence of coiled coil in the middle domain of ClpB. The N-terminal domain of ClpX is a zinc binding domain (ZBD).
1.3.1 ClpX

ClpX associates with its protease partner ClpP, and the ClpXP proteolytic complex degrades the substrate α-casein, a protein abundant in milk. It was first discovered to promote the degradation of λO, a bacteriophage λ replication initiation protein in an ATP-dependent manner (Gottesman, Clark et al., 1993). Protein substrates targeted for degradation by ClpXP include λO, MuA, Mu repressor, as well as SsrA-tagged proteins (Flynn, Neher et al., 2003). SsrA is an 11 amino acids long peptide which is added cotranslationally to the C-terminus of nascent polypeptides on stalled ribosomes to target the polypeptides by degradation by ClpXP. Despite the sheer amount of substrates, ClpX is not an essential protease.

In *E. coli*, clpX is located downstream of clpP on the same operon. As mentioned previously, ClpX is a member of the Class II Clp family which has one AAA+ domain (D1) with an N-terminal zinc binding domain (ZBD) (Figure 1.1) (Donaldson, Wojtyra et al., 2003). ClpX forms a hexameric ring in an ATP-dependent manner and associates with tetradecameric ClpP as expected of Class II Clp ATPases with a ClpP partner. Both ClpP and ClpX are ubiquitous in most living organisms except in archaebacteria, mollicutes, and certain fungi. Interestingly, a ClpX paralogue is present while a ClpP paralogue is absent in *S. cerevisiae*.

Only EM images of the ClpXP complex are currently available. EM images show that the hexameric ClpX binds to one or both sides of tetradecameric ClpP (Ortega, Lee et al., 2002). It was also reported that ClpXP is more active when a 2:1 ratio of ClpX:ClpP is used instead of a 1:1 ratio (Ortega, Lee et al., 2002). As mentioned previously, it was observed that ClpA and ClpX can bind simultaneously to opposite
sides of ClpP in vitro implying that the hetero-complex ClpAXP might be a bifunctional protease which can target different classes of substrates (Ortega, Lee et al., 2004).

The structure of the first 60 N-terminal residues of E. coli ClpX corresponding to the Zinc Binding Domain (ZBD) was solved by NMR and it was found by our group to be a constitutive dimer (Donaldson, Wojtyra et al., 2003). In ZBD, four cysteine residues are involved in coordinating a Zn$^{2+}$ essential for dimerization. Given that ZBD assumes dimer formation and ClpX is a hexamer, our group postulated that ZBD forms a trimer-of-dimers on the surface of the AAA+ ring of ClpX.

1.3.2 ClpA

ClpA is the first Clp chaperone identified as part of a two component ATP-dependent protease system where it was shown to degrade $\alpha$-casein and was originally named Ti (Hwang, Park et al., 1987). ClpA belongs to Class I of the Clp ATPases family characterized by two ATPase domains, D1 and D2 (Figure 1.1).

ClpA is able to function as a molecular chaperone in the absence of ClpP. In vitro, ClpA alone converts the phage P1 origin-binding protein, RepA, by dissociating an inactive dimer into active monomers (Wickner, Gottesman et al., 1994). ClpA also has the ability to protect RepA and firefly luciferase from irreversible heat inactivation in vitro. ClpA is only found in Gram negative bacteria and is most extensively characterized in the model bacteria E.coli.

ClpA hexamer, in complex with ClpP tetradecamer, carries out ATP-dependent protein degradation. As expected of the Clp proteolytic system, when in complex, ClpA
translocates substrate proteins into the ClpP chamber for degradation with small peptides as byproducts (Reid, Fenton et al., 2001).

In general, ClpAP recognize protein substrates that contain signal tags of ~10 residues either at the N- or C-terminus. Upon recognition, ClpAP unfolds and translocates the substrates into the chamber of ClpP for degradation. ClpAP targets RepA for degradation through an N-terminal recognition tag (Hoskins, Kim et al., 2000).

It was shown by EM that ClpA hexamer can bind on one or both sides of ClpP tetradecamer (Kessel, Maurizi et al. 1995). Moreover, it was observed in vitro that ClpA and ClpX can bind simultaneously to opposite sides of ClpP suggesting that ClpAXP might be a bifunctional protease which can target different classes of substrates (Ortega, Lee et al., 2004).

1.3.3 ClpC

ClpC was first discovered in a genetic screen for repressors of competence development in Bacillus subtilis (Dubnau and Roggiani 1990). In addition to general proteolysis, ClpC is also known to be responsible for controlling key steps of developmental processes such as competence, sporulation, and regulation of stress response in B. subtilis (Kruger, Zuhlke et al., 2001).

ClpC is a member of the Class I Clp ATPases with 2 AAA+ domains. ClpC has an N-domain homologous to the N-domains of ClpA and ClpB (Figure 1.1). The first AAA+ domains (D1) of ClpC is interrupted by an insertion sequence similar to the middle domain of ClpB. ClpC is present in all low GC Gram-positive bacteria. Interestingly, these bacteria are also coincidentally devoid of ClpA, suggesting that ClpC
may serve to replace ClpA’s functionality in these organisms (Frees, Savijoki et al., 2007). It should be noted that there exist major functional differences between the ClpC from B. subtilis and the Clp ATPases present in E. coli. In contrast to other characterized Clp ATPases, ClpC has very low ATPase activity and is dependent on the presence of cofactors. MecA, a ClpC cofactor, is required for ClpC oligomerization, substrate recognition and interaction with ClpP (Kirstein, Schlothauer et al., 2006).

1.3.4 ClpB

ClpB are essential chaperones for cellular stress responses, especially in disaggregation. As a member of Class I Clp ATPases, ClpB consists of an N-terminal domain, two nucleotide binding domain (D1 and D2), a middle domain (MD) which is inserted in D1, and a small C-terminal region (Figure 1.1). It is interesting to note that the amino acid sequence of the N-terminal and the middle domains of ClpB are not highly conserved. It has been postulated that this may confer different substrate and functional specificities in different organisms. The crystal structure of the full length protein ClpB has been solved from Thermus thermophilus organism (Li and Sha 2003). The ClpB hexamer is characterized by a double stacked ring structure with a height of 90 Å and a diameter of 140 Å (D2) and 150 Å (D1). As expected, the structure of ClpB is similar to other members of class 1 Clps such as E. coli ClpA, and the overall topology is similar to other AAA+ proteins (Guo, Esser et al., 2002).

The MD of ClpB is involved in Hsp 70 interaction. It is a long coiled-coil of 85 Å in length located on the outside of the hexamer and is inserted into the α-helical domain of the D1 (Li and Sha 2003). The MD is postulated to facilitate a concerted ATP-driven
conformational change in D1 important for protein disaggregation (Lee, Sowa et al., 2003). The close proximity of MD to D1 suggests that the coiled-coil repeats will undergo significant motion in response to nucleotide hydrolysis occurring at D1.

ClpB was initially found in E.coli. Orthologs of ClpB have now been found in eukaryotic yeast and in plants. They are called Hsp104 and Hsp101 respectively. In yeast, it was found that deletion of the gene hsp104 has no consequence on the growth of cells under normal condition but lethal if under thermal stress (Sanchez and Lindquist 1990). Mutagenesis study revealed that Hsp104 mediates the resolubilization of heat-inactivated luciferase from insoluble aggregates (Parsell, Kowal et al., 1994). Furthermore, in vitro remodeling activity of Hsp104 on protein aggregates revealed that the assistance of Hsp70 and Hsp40 (yeast Ssa1 and Ydj1) is needed for Hsp104 disaggregate activity in vitro (Glover and Lindquist 1998). Hsp104 along with Hsp70 chaperone system orchestrates an effective protein system capable of disaggregating stress-damaged protein aggregates in vivo.

1.4. The Protease ClpP

ClpP is ubiquitous in most living organisms except for archaea, mollicutes, and several fungi. ClpP is a serine protease that was first identified along with ClpA in a strain of E. coli lacking Lon protease (Hwang, Park et al., 1987). ClpP in E. coli consists of 207 amino acids in which the first 14 residues constitute the regulatory prosequence. The prosequence must be cleaved off autocatalytically to form the active mature protease of 193 residues. Initially, through electron microscopy (EM), ClpP was discovered to have a similar structure to the eukaryotic proteasome in which there are stacked rings of
Figure 1.2. *E.coli* ClpP structure. The crystal structure of *E. coli* ClpP (ITYF). (A) Side view of ClpP monomer. Each repeats of $\alpha/\beta$ units is highlighted with a different color. The catalytic residues are labeled. (B) Side view of ClpP tetradecamer. Each protomer is highlighted with a different color. (C) Top view of ClpP tetradecamer.
multiple subunits with a central cavity (Maurizi, Clark et al., 1990). ClpP was later shown by EM again to be a tetradecamer arranged from two heptameric rings that form a large chamber containing the active sites (Flanagan, Wall et al., 1995).

The ClpP proteolytic complex is made of 14 ClpP subunits (Figure 1.2). Each ClpP subunit contains an active site, which is composed of three highly conserved residues; Serine, Histidine, and Aspartic acid, that are located at positions 111, 136, and 185, respectively in *E. coli*. Thus ClpP contains 14 active sites within its proteolytic chamber given its tetradecamer structure. ClpP typically associates with the regulatory Clp ATPases to degrade protein substrates. ATPases unfold targeted substrates that are translocated through the axial pores into the ClpP proteolytic chamber for degradation. ClpP can degrade substrates into peptides of 7-8 residues which are then released from the chamber (Choi and Licht 2005). Without its ATPase partner, ClpP on its own can only degrade small peptides (Thompson and Maurizi 1994).

The release mechanism of peptide byproducts from the ClpP chamber remains controversial. One group suggests that degraded products exit the ClpP proteolytic chamber by the same axial pores that allow polypeptide entry by passive diffusion (Wang, Hartling et al., 1997). Given that the ATPase chaperones associate with Clp on its axial ends, this model of diffusion from the axial pores implies that the bound chaperones must first dissociate from the ClpP cylinder in order to facilitate peptide release from the chamber. Support for this model has been reported in the case of proteasome in which ATPase dissociates when cleaved peptides are to be released (Babbitt, Kiss et al., 2005). Our group favors an alternative model in which the disruptive dissociation of the Clp chaperones is not necessary and the peptides are released through the equatorial side
Figure 1.3. Clp ATPases and Protease. Cartoon representation of Clp ATPases (Red) and Clp Protease (Blue). Left side: Two hexameric Clp ATPases stacked on top and bottom of the tetradecameric Clp protease. Orange arrows denote the release of degraded peptide products from equatorial region of the Clp protease. Right side: Clp ATPases with ClpP interaction loop with consensus motif [LIVMP]G[FL] and Clp Protease with N-terminal axial loops.
pores that are transiently formed at the interface between the two heptameric ClpP rings (Figure 1.3). NMR studies combined with biochemical data suggest that the equatorial regions of ClpP barrel are highly dynamic and plastic, thus providing putative exit sites of cleaved peptides (Gribun, Kimber et al., 2005).

The first crystal structure was solved for *E. coli* ClpP at 2.3 Å resolution (Wang, Hartling et al., 1997). The crystal structure agreed with the EM images in which ClpP forms a tetradecamer. The protease complex is a cylinder of ~90 Å in height by ~51 Å in diameter with two axial pores that are ~10 Å wide (Figure 1.2). The narrow pore suggests that only peptides and unfolded proteins can be threaded in, while big globular proteins are sequestered outside, as expected of a self-compartmentalized protease. The crystal structure for ClpP has also been solved for various organisms such as *Steptococcus pneumonia* (Gribun, Kimber et al., 2005), the malaria agent *Plasmodium falciparum* (Vedadi, Lew et al., 2007), *Mycobacterium tuberculosis* (Ingvarsson, Mate et al., 2007), and *Homo sapiens* (Kang, Maurizi et al., 2004). Overall, ClpP structures across various organisms are very similar.

The *E. coli* ClpP protomer is composed of six repeats of α/β units with an additional protruding α/β unit (Figure 1.2 A). The β-strands form two layers of β-sheets that pack against a layer of α-helices creating a unique fold around the serine proteolytic active site. The contact between the monomers within a ring is mainly reinforced by hydrophobic interactions. The ClpP protomer can be divided into a handle region (β-strand 9 and α-helix E) and a head domain comprised of residues 42-134 and 174-202 in *E. coli* (Wang, Hartling et al., 1997). Based on the crystal structures, the interaction between the two ClpP hexameric rings is mediated mainly by the handle region (Figure
1.2 A). Our group has previously discovered that the handle region exhibits a high degree of plasticity as truncations in the handle region do not result in the dissociation of the two rings. (Gribun, Kimber et al., 2005). It was proposed that the electrostatic interactions between networks of residues leading to the head domains stabilize the interactions between the two ClpP rings (Gribun, Kimber et al., 2005).

In humans, ClpP is encoded on chromosome 19 (Corydon et al., 2000). Through localization studies by immunofluorescence, human ClpP was found to reside within the mitochondrial matrix (Casta et al., 1999), although its role within the organelle remains enigmatic. Human ClpP contains an N-terminal prosequence and mitochondrial targeting sequence, which combine to be approximately 56 residues long. The targeting sequence is essential for delivering the protein into the mitochondrion (Herrmann et al., 1989). Similar to its bacterial cousin, the N-terminal sequence must be cleaved to yield the mature active protease. Mature human ClpP shares 56% sequence identity and 71% sequence similarity with *E. coli* ClpP. Unlike its *E. coli* cousin, human ClpP has an additional 28 residues present at its C-terminus with unknown function. With regards to associated chaperones, human ClpX is encoded on chromosome 15 and it is the only ATPase that has been demonstrated to associate with human ClpP (Corydon et al., 2000). Human ClpX also contains a mitochondrial targeting sequence, and it is also found to reside within the mitochondrial matrix as expected. Mature human ClpX shares 44% identity and 62% similarity with *E. coli* ClpX.

*E. coli* ClpP exists as double-ring tetradecamers *in vivo* and *in vitro*. When assembled as a full tetradecamer of two stacked heptameric rings, the catalytic sites in *E. coli* ClpP are compartmentalized within the chamber inside two heptameric rings.
However, human ClpP was initially demonstrated to only exist as a single heptameric ring under physiological conditions (Kang, Ortega et al., 2002). Without the double ring assembly, the active sites in human ClpP are predicted and subsequently shown to be exposed to the environment (Kang, Dimitrova et al., 2005). The exposed and openly accessible catalytic sites are hypothesized to be detrimental for the cell because its proteolytic activity will be uncontrolled without the restrictions of compartmentalization. However, it was demonstrated that the single heptameric human ClpP does not exhibit such rampant protease activity and has a very low peptidase activity compared to *E. coli* ClpP (Kang, Dimitrova et al., 2005). The reason behind the lack of proteolytic activity is that the catalytic triads are found to be in an inactive configuration. When the protease assumes a single heptameric structure, there exists an increased plasticity at the handle region compared to the protease when it assumes a tetradecameric structure in which the handle regions are interacting with one another. It was hypothesized that such increase in the plasticity of the handle region disrupts the proper orientation of the catalytic triad that is located at the interface between the handle and head regions (Kang, Dimitrova et al., 2005).

Tetradecameric human ClpP does exist under physiological conditions but only in the presence of human ClpX and ATP (Kang, Dimitrova et al., 2005). Single heptameric human ClpP will assume its native active tetradecameric form of double rings upon binding to human ClpX in the presence of ATP. The binding of the ATPases is hypothesized to induce conformational changes in the ClpP heptameric rings favoring the formation of double rings. The human ClpXP tetradecameric complex that is formed has the expected proteolytic activity, as well as, increased peptidase activity compared to
heptameric human ClpP (Kang, Dimitrova et al., 2005). This further supports the hypothesis that catalytic residues in human ClpP single heptamers are in an inactive configuration through disruption by the increased plasticity of the handle region. The catalytic residues revert back to their active configuration when proper compartmentalized ClpP tetradecamer is formed upon binding of ClpX. It should be noted that such induction of proper tetradecamer formation by the presence of ATPase in ClpPs is also observed in Bacillus subtilis (Kirstein, J., Schlothauer et al. 2006).

Unexpectedly, human ClpP is observed to form a hybrid complex with E. coli ClpX in vitro, although the reverse in which human ClpX forming a complex with E.coli ClpP did not occur (Kang, Dimitrova et al., 2005). This hybrid complex has a similar structure as E. coli ClpXP complex as observed by EM. Such hybrid complex occurs with ClpX and ClpA thus far. Interestingly, the human ClpP and E. coli ClpX hybrid complex is able to degrade E. coli ClpXP substrates, thus supporting the hypothesis that substrate recognition is solely dependent on the ATPase in the Clp proteolytic system.

The Clp family becomes increasingly diverse and complex in higher plant organisms. In the model plant Arabidopsis thaliana, there are 6 ClpP paralogs (ClpP1–6) and 4 ClpR paralogs (ClpR1–4), a total of 10 different proteolytic Clp proteins. ClpR is an inactive ortholog of ClpP lacking two or three of the catalytic residues necessary for proteolysis. Unlike E. coli in which the Clp protease exists as a homocomplex of solely ClpP subunits, both ClpP and ClpR protomers are found in the Clp proteolytic complex in A. thaliana, forming a ClpPR heterocomplex (Peltier et al., 2004). In terms of chaperones, in A. thaliana, there are 7 Class I (ClpB1–4, ClpC1–2, ClpD) and 3 Class II (ClpX1-3), a total of 10 Clp/Hsp100 AAA+ ATPase. These Clp chaperones have been
found to associate with the ClpPR complexes (Peltier et al., 2004). The same variety of Clp proteases and chaperones are also found in the plastids of non-green plants Brassica rapa roots and Brassica oleracea petals (Peltier et al. 2001). ClpT is found in A. thaliana and is an ortholog of the bacterial ClpS. ClpS is a cofactor that binds to ClpA and is essential to the N-end rule pathway in E. coli (Erbse et al. 2006). Two previously unknown plant Clp chaperones called ClpS1 and ClpS2 were also discovered. It should be noted that they have no similarity with bacterial ClpS and should not be confused with ClpT. In terms of localization, a majority of Clp ATPases and proteases of Arabidopsis are found in the stroma of chloroplasts including ClpB3, ClpC1–2, ClpD, ClpP1, CpP3–6, ClpR1–4, ClpS1–2, and ClpT. They are all encoded in the nuclear genome with the exception of ClpP1 which is plastid-encoded (Peltier et al., 2004, Sjogren et al., 2006).

In terms of composition, in the chloroplast of A. thaliana, the Clp proteolytic complex has been found to consist of a variety of subunits including ClpP1, ClpP3–6, ClpR1–4 and ClpS1–2 (Peltier et al., 2001). Modeling studies indicate that ClpP and ClpR subunits contribute to the formation of the central core ring structure. Interestingly, in the mitochondria of A. thaliana, the ClpP complex is a composed of only ClpP2 subunits (Peltier et al., 2004, Sjogren et al., 2006). The ClpS1–2 subunit is hypothesized to associate with the hydrophobic pockets on the axial sites of the ClpPR core complex (Peltier et al., 2004). The role of ClpS1-2 is hypothesized to regulate association of ClpPR with its cognate ATPase such as ClpC or ClpD. It was further postulated by another group that ClpS1–2 might also regulate degradation by competitive binding with the ATPase chaperones (Majeran et al., 2005).
Eventhough the function of numerous ClpP and ClpR subunits in plants remains enigmatic, they are essential for chloroplast development. In *A. thaliana*, ClpR2 is essential for Clp core complex assembly since reducing the levels of ClpR2 resulted in reduction of the ClpPRS protease complex levels. Reduction in ClpPRS results in pale-green phenotype, delayed shoot development, reduced chloroplast size, decreased thylakoid accumulation, and increased plastoglobule levels (Rudella *et al.*, 2006).

As mentioned previously, in plants, the cognate Clp proteolytic complex consist of both ClpP and ClpR subunits. The function of the inactive ClpR protomers remains unclear, but it has been postulated to have a regulatory role in proteolysis. ClpRs in *A. thaliana* seem to have extended C-termini compared to *E. coli* ClpP (Peltier *et al.*, 2004). The C-terminal extension in ClpR might fold on top of the proteolytic core and serve to regulate interaction between the proteolytic core and its chaperones. Furthermore, alignments of ClpRs with ClpPs in *A. thaliana* displays the presence of insertion loops 9–10 residues long in ClpR1, ClpR3, and ClpR4. Based on homology remodeling, the loop is situated as part of the substrate-binding cleft in the active site. The insertion loop of non-catalytic ClpR is now hypothesized to serve in presenting substrates to its neighboring catalytic ClpP (Peltier *et al.*, 2004).

1.5 Proteasome

Perhaps the protease that garnered the most attention is that of the eukaryotic 26S proteasome. It is the main compartmentalized protease found in archaea, bacteria, eukaryotic cytosol and nucleus. Named after the Svedberg sedimentation coefficient of its proteolytic (20S) and regulatory (19S) complexes, the 26S proteasome is very well
conserved with a cylindrical structure consisting of two distinct types of subunits, α and β. The proteasome exhibits hydrolytic activities similar to chymotrypsin, trypsin and caspase. Protein degradation mediated by 20S proteasomes is an energy-dependent process requiring ATP hydrolysis by associating with hexameric AAA+ ATPase complexes such as PAN in archaea and the 26S proteasome in eukaryotes (Lupas, Zwickl et al., 1994). Proteasomes can catalyze the degradation of ubiquitinated and non-ubiquitinated polypeptide substrates.

The eukaryotic 26S proteasome is large with an approximate size of 2.6 MDa and is approximately 150 by 115 Å in dimension (Whitby, Masters et al., 2000), while the entrance axial pores are approximately 13Å, and the interior chamber is 53Å wide. The proteasome is made of 62 subunits encoded by 31 genes. The 26S proteasome can be divided into two subparticles, the 20S core and the 19S regulatory cap. The 20S core particle is composed of four heptameric rings with α and β subunits arranged alternatively; α\textsubscript{7}β\textsubscript{7}β\textsubscript{7}α\textsubscript{7}. The catalytic sites of the proteasome are formed by β-subunits within the inner core of the cylinder, however only three eukaryotic β subunits out of seven contain an active catalytic site, namely β1, β2, and β5 (Baumeister, Walz et al., 1998). Similar to ClpP, access into the interior chamber of the cylinder is regulated. In the case of the proteasome, access is regulated by the outer two α rings, which concurrently serve as docking sites for regulatory particles.

Similar to ClpP, the 20S proteasome associates with various regulatory cap complexes. The interaction of the cap complex with the 20S proteasome induces a conformational change that allows access into the core complex (Whitby, Masters et al., 2000). In eukaryotes, the 19S regulatory cap complex is approximately 700 kDa and is
aptly named PA700. It is composed of 20 subunits and has 2 main regions; the base and the lid regions. The base contains six ATPases and two non-ATPase subunits. The lid region is composed of eight ATPases and one non-ATPase subunit. The lid region is thought to bind and disassemble ubiquitin-conjugated substrates (Voges, Zwickl et al., 1999). The 19S cap can recognize and unfold ubiquitinated proteins, which eventually remove ubiquitin chains and provide a passageway for threading unfolded proteins into the proteasome core complex (Thrower, Hoffman et al., 2000).

In terms of substrate, a highly conserved 76 residues long protein called ubiquitin is attached to the substrate repeatedly in an ATP-dependent fashion by ubiquitin ligases creating a polyubiquitin chain that will subsequently target the protein for degradation by the proteasome. Ubiquitinated proteins are recognized by the 19S cap similar to that of Clp ATPases for Clp proteases (Liu, Li et al., 2006).

1.6 Malaria

Malaria is first associated as a disease from marshy regions, hence the term “mal’aria”, which is an Italian word meaning “bad air”. Since then, it was discovered that a parasite that resides within red blood cells is the causative agent of malaria, and this parasite can only be transmitted by Anopheles mosquitoes. Over 300 million cases of humans developing clinical malarial symptoms and at least 2 million deaths (Snow et al., 2004) can be attributed to malaria each year. Malaria remains a major pediatric killer in many parts of sub-Saharan Africa.

Malaria is transmitted through the bite of an infected female Anopheles mosquito. The causative agent of malaria is the eukaryotic single-celled microorganisms that belong
to the genus *Plasmodium*, a member of the *Apicomplexa*. The name *Apicomplexa* is derived from the specialized apical complex that is central to the invasion process that all parasites of this group share. More than 100 species of *Plasmodium* have been discovered that are capable of infecting various animal species such as reptiles, birds and mammals. Yet, only four species of parasite can infect humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. These four species differ morphologically, immunologically, in their geographical distribution, in their relapse patterns and in their drug responses. Relapse is the appearance of symptoms followed by a latent period after the primary infection. *P. vivax* is the most widespread parasite but infections with this species are usually not fatal (Miller and Greenwood 2002). The most severe form of malaria is caused by *P. falciparum* and it is the principal cause of malaria deaths in young children in Africa (Snow et al., 2004). The reason behind *P. falciparum* severity lies in its ability to cause severe anemia. Also, red blood cells infected with *P. falciparum* undergo major morphological changes and can obstruct small blood vessels and capillaries, thus reducing blood flow to major organs (Tuteja 2007). Obstruction of blood flow to the brain leads to cerebral malaria, a complication often fatal in African infants.

In terms of genomics, the genome of *P. falciparum* clone 3D7 was the first to be sequenced and annotated in 2002. The *P. falciparum* genome is 23 mega base pairs of DNA long with 14 chromosomes. It also has a circular plastid-like genome within the apicoplast and a linear mitochondrial genome (Bozdech et al., 2003). The apicoplast is a relic non-photosynthetic plastid within the parasite. The nuclear genome is very A+T rich, with an overall A+T composition of 81% (Gardner et al., 2002). About 5300 genes have
been predicted from the genomic sequence, of which only a few have been identified to date as encoding enzymes.

Homologs of genes involved in basic pathways such as translation initiation, DNA replication, repair and recombination are expectedly present in the nuclear genome of the parasite (Gardner et al., 2002). However, it appears to lack genes for enzymes involved in several key metabolic pathways. The synthesis of essential amino acids, synthesis of purines and the salvage of pyrimidines, as well as two protein components of ATP synthase and NADH dehydrogenase complex are missing in the nuclear genome (Gardner et al., 2002). However, components of some anabolic pathways for the synthesis of fatty acids, isoprenoid precursors, heme and iron sulfur complexes seem to be localized in the apicoplast (Gardner et al., 2002; Gornicki 2003). The apicoplast will be discussed further in section 1.6.4.

The life cycle of malaria parasites is extremely complex and involves both the invertebrate and vertebrate hosts (Figure 1.4). The life cycle can be divided into several stage; 1) Tissue schizogony, 2) Erythrocytic schizogony and 3) Sporogony. The cycle begins with sporozoites entry into the bloodstream of the mammalian host.

1.6.1 Tissue schizogony (pre-erythrocytic schizogony)

Infective sporozoites reside within the salivary gland of the *Anopheles* mosquito (Figure 1.4). During feeding, the sporozoites are injected into the human host and enter the bloodstream. As they are delivered to the liver, the *P. falciparum* sporozoites penetrate the hepatocytes where they will remain for 9–16 days and undergo asexual replication known as exo-erythrocytic schizogony (Miller and Greenwood 2002). While
residing within the hepatocyte, each sporozoite can potentially give rise to tens of thousands merozoites. Each merozoite can then invade an erythrocyte when release from the liver. The entire tissue schizogony stage takes about 8-25 days for *P. falciparum* (Tuteja 2007).
Figure 1.4. Life cycle of *Plasmodium falciparum*. The life cycle is divided into 3 stages; tissue schizogony, erythrocytic schizogony, and sporogony. The cycle begins when an female *Anopheles* mosquito injects sporozoites into the human host during feeding.

Tissue schizogony describes the asexual stage within the liver. Erythrocytic schizogony describes the asexual repetitive stage when merozoites lyse and re-infect addition RBC (red blood cells). This is also the stage when clinical symptoms are manifested.

Sporogony describes the sexual stage within the gut of the mosquito.
1.6.2 Erythrocytic schizogony

Upon release from the hepatocytes through lysis, the merozoites are free to invade erythrocytes within the host’s circulation. Once inside the erythrocytes, the merozoites undergo multiple nuclear divisions forming schizonts (Figure 1.4). A mature schizont contains approximately 20 merozoites that will be released upon lysis of the erythrocyte (Gornicki 2003). These merozoites are then free to invade other uninfected erythrocytes; thus repeating the cycle again. This repetitive intraerythrocytic cycle of invasion, multiplication, release and invasion is called erythrocytic schizogony and continues for approximately 48 h in *P. falciparum* (Tuteja 2007). Erythrocytic schizogony is also the stage in which clinical symptoms are manifested on the mammalian host. This is due to repeated lysis of infected erythrocytes which stimulate the production of tumor necrosis factor and other cytokines (Gornicki 2003).

The invasion commences when merozoites engage binding receptors on the erythrocyte membrane reversibly. The parasite subsequently induces a vacuole derived from the erythrocyte’s plasma membrane and enters and reorientates within the vacuole by a moving junction. The moving junction is an irreversible attachment between the erythrocyte membrane and the apical (invading) end of the merozoite (Tuteja 2007). Various enzymes and chemicals are released from the rhoptry and microneme organelles, leading to formation of the parasitophorous vacuole (Figure 1.5). It should be noted that both rhoptries and micronemes are apical organelles that define the phylum Apicomplexa. The receptors that mediate invasion of erythrocytes by merozoites and invasion of liver by sporozoites are found in these organelles (Tuteja 2007). The next stage of the invasion process involves subsequent movements of the junction and invagination of the
erythrocyte membrane around the merozoite followed by removal of the merozoite’s surface coat. Lastly, the parasitophorous vacuole and erythrocyte membranes are sealed and the invasion by the merozoites is completed (Gornicki 2003).

Once inside its erythrocytic host, the parasite undergoes different stages of its asexual division. The early trophozoite is referred to as the ‘ring form’, named after its characteristic ring-like morphology. A black pigment is often visible when viewing ring stage trophozoites. As the parasite degrades the host’s hemoglobin, heme is released as a byproduct. The parasite is unable to degrade heme and heme is toxic to the parasite. As such, free heme is polymerized into hemozoin, a crystalline pigmented substance that is subsequently stored within the food vacuoles (Gornicki 2003).

1.6.3 Sporogony

Sporogony describes the sexual reproductive stage of the parasite’s life cycle that takes place within the female *Anopheles* mosquito (Figure 1.4). Briefly, besides the repeating asexual cycle within the erythrocytes, a small proportion of the merozoites differentiate to produce male and female gametocytes. These gametocytes do not infect additional erythrocytes but remain in circulation until a female *Anopheles* mosquito ingests these gametocytes into its midgut upon taking a blood meal on an infected individual (Miller and Greenwood 2002). These gametes fuse, undergo fertilization and form a zygote. The zygote transforms into an ookinete, which penetrates the wall of a cell in the midgut and develops into an oocyst (Eksi et al., 2006). The oocyst produces many sporozoites. Upon the rupture of the oocyst, the sporozoites are released and they migrate to the salivary glands in preparation for transmission into another host (Tuteja 2007).
When an infected mosquito bites a susceptible host, the sporozoites are injected into the host’s circulation and *Plasmodium* life cycle repeats again.
Figure 1.5. Anatomy of a RBC infected by *Plasmodium falciparum*. Cartoon depiction of the major cellular organelles of a red blood cell infected by *Plasmodium falciparum* during early erythrocytic schizogony. RBC = red blood cell; DG = dense granules; Mic = micronemes; FV = food vacuoles; ER = endoplastic reticulum.
1.6.4 Apicoplast

The apicoplast is a vestigial plastid present in almost every parasite of the Phylum Apicomplexa. In terms of origin, the apicoplast is postulated to be a result of secondary endosymbiosis (Cavalier-Smith, 1982). Briefly, an eukaryote initially obtained a plastid by engulfing a cyanobacteria through primary endosymbiosis. The new photosynthetic eukaryote was subsequently engulfed by another heterotrophic eukaryote. Thus the plastid had undergone 2 separate endosymbiotic events – secondary endosymbiosis (Cavalier-Smith, 1982). As a result, secondary endosymbiotic plastids typically have three or four membranes, whereas primary plastids invariably have only two membranes. The apicoplast has been reported by a majority of groups to exhibit 4 membranes (Dubremetz 1995; Köhler et al., 1997; McFadden and Roos, 1999).

The outermost membrane of the apicoplast is derived from the phagosomal membrane of the host cell. The second outermost membrane of the apicoplast is derived from the plasma membrane of the original engulfed alga cell. The inner pair of membranes is equivalent to the outer and inner envelopes of the plastid, which evolved from the cell surface envelope of the engulfed cyanobacterium by the primary host (McFadden and Roos, 1999).

One striking feature of the apicoplast is its close proximity to the single mitochondrion (van Dooren et al., 2006). The closeness of the two organelles suggests their metabolic dependences on each other (Bender et al., 2003). Throughout the various asexual stages of the parasite, the apicoplast is always in close contact with the mitochondrion (van Dooren et al., 2006).
The apicoplast genome is circular and is approximately 35 kb in length, smaller than the usual photosynthetic plastid genomes. As a result of its endosymbiotic origin, many genes of the apicoplast have transferred to the host cell nucleus (Martin & Herrmann 1998). Such endosymbiotic gene transfer likely took place to minimize the effects of Muller’s ratchet, whereby non-recombining genomes accumulate deleterious mutations (Martin & Herrmann 1998). As a result, the apicoplast genome itself encodes less than 50 proteins and a majority of proteins are encoded by nuclear genes and targeted into the apicoplast instead (Wilson et al., 1996).

The general pathway by which most nuclear encoded proteins are targeted into the apicoplast is mediated by a bipartite leader sequence at the N-terminus of a polypeptide chain (Figure 1.6) (Waller et al., 1998, 2000). The bipartite leader sequence comprises an N-terminal signal peptide (SP) followed immediately by a transit peptide (TP). The SP commits the nascent polypeptide chain into the endomembrane system in which the apicoplast is contiguous with; The TP subsequently delivers the protein into the plastid (Waller et al., 2000). The endomembrane system is thought to derive from the food vacuole of the host eukaryote during the secondary endosymbiotic event (Waller et al., 1998). The TP is laced with positively charged residues and these positive charges at the N-terminus of the TP are essential for apicoplast targeting (Foth et al., 2003).

Elegant deletion studies have been done to elucidate the mechanism of action of both SP and TP using the acyl carrier protein (ACP), which is a known nuclear-encoded
Figure 1.6. Protein targeted to the apicoplast. Cartoon depiction of newly synthesized nuclear encoded proteins targeted to the apicoplast. Cyan represents signal peptide; Green represents transit peptide; Blue represents mature protein. ER = endoplasmic reticulum.
protein that is targeted towards the apicoplast. If the TP domain of ACP is removed, leaving only the SP domain, ACP is secreted outside of the cell into the parasitophorous vacuolar space (Figure 1.5) for both *Plasmodium falciparum* (Waller *et al.*, 2000) and *Toxoplasma gondii* (McFadden and Roos 1999). This indicates that the SP is responsible for trafficking protein into the endomembrane system. If the SP domain is removed leaving only the transit peptide domain, ACP accumulates in the cytosol (McFadden and Roos 1999; Waller *et al.*, 2000). Together these experiments suggest that once within the endomembrane system, the TP domain is required to sort the protein towards the apicoplast. Without this sorting mediated by TP, the ACP simply follows a default route of secretion out of the cell (Figure 1.6). Also, the apicoplast sorting machinery that recognizes the TP is within the endomembrane system and cannot facilitate apicoplast sorting from the cytosol. Protein targeting to apicoplasts is therefore at least a two-step process, with co-translational import into the endomembrane system as the first step, and subsequent sorting to the apicoplast within that compartment as the second step. A signal peptide followed by a transit peptide is therefore both necessary and sufficient for apicoplast import in *Plasmodium falciparum* and *Toxoplasma gondii* (Waller *et al.*, 2000). It should be noted that the leader sequence is interchangeable between parasites, as the bipartite sequence of *Plasmodium falciparum* can be used to target proteins towards the apicoplast of *Toxoplasma gondii*, and vice versa (Waller *et al.*, 2000).

Putative sequence motifs of the transit peptide have remained enigmatic. The TP sequence varies greatly in size, has no primary sequence conservation, and lacks any stable secondary structure. The median length of TP is approximately 78 residues and there exists enormous sequence redundancies (DeRocher *et al.*, 2000). In algae and plants,
the sequence of TP seems to exhibit enrichment in hydroxylated residues, particularly serine and threonine, and a net positive charge (Cline and Henry, 1996). The TP of *P. falciparum* also demonstrated a net positive charge and interestingly this bias is predominantly at the N-terminal region of the transit peptide (Waller *et al.*, 1998). A net positive charge in the first 17 residues has been adopted as a useful defining character for predictive software for *P. falciparum* transit peptides. *P. falciparum* also demonstrates a strong preference for lysine and asparagine residues in its transit peptides (Zuegge *et al.*, 2001), perhaps a consequence of the *P. falciparum* AT bias (82%) given that both residues can be encoded without a C or G. Given that *P. falciparum* TP have been demonstrated to work in *T. gondii* and vice versa (Jomaa *et al.*, 1999), it seems more likely that the differences between them reflect genome pressures rather than modified trafficking mechanisms.

It is postulated that a minimal length rather than specific residues is important for apicoplast targeting. It was found that 55 residues could be deleted from the C-terminal end of S9 protein without loss of apicoplast targeting. Upon the deletion of a further six residues, TP function was lost and the reporter protein was secreted from the cell. Yet, there is no effect on apicoplast trafficking if just one or all six of these residues are removed (Yung *et al.*, 2001).

In terms of mechanism of action, during protein trafficking SP and TP are removed sequentially. The SP is presumed to be removed during co-translational import into the endoplasmic reticulum (ER) (Waller *et al.*, 1998). In plants, the removal of the TP from chloroplast-targeted proteins is achieved by the Stromal Processing Peptidase (SPP) once the proteins arrived at the chloroplast (Richter and Lamppa, 1999). A nuclear-
encoded homologue of plant SPP has recently been identified from *P. falciparum*. The *P. falciparum* SPP homologue bears a bipartite presequence for apicoplast-targeting suggesting it is apicoplast located (van Dooren *et al.*, 2002). This presents a very likely candidate for the role of transit peptide processing in the apicoplast. However, the *P. falciparum* cleavage sites within the bipartite sequence remain enigmatic.

Bioinformatic tools have been derived to predict the presence or absence of the bipartite leader sequence. Signal peptide predictors such as neural network-based SignalP, and “knowledge-based” PSORT have been used for identifying the signal peptides on the bipartite presequences (Zuegge *et al.*, 2001). In terms of transit sequence, a predictive system called PATS (predict apicoplast-targeted sequences) has been developed. It is a neural network using a training set of 84 sequences most probably targeted to the apicoplast, and a negative set of 102 non-apicoplast proteins (Zuegge *et al.*, 2001) and can be used in conjunction with SignalP for increased accuracy. When PATS was applied to the *P. falciparum* chromosome 2 and 3 data, it predicted that 69 genes (15% of total genes on these chromosomes) encoded apicoplast-targeted proteins (Zuegge *et al.*, 2001).

A second system called PlasmoAP is a “knowledge-based” system where observed features of *Plasmodium falciparum* presequences have been used to model a set of criteria for discriminating apicoplast-targeting presequences (Foth *et al.*, 2003). The first criterion is to identify the presence of signal peptide using SignalP. The second criterion is that within the first 22 residues of the transit peptide the ratio of acid to basic residue must not exceed 0.7. Next criterion is that within the first 80 amino acids there must be a stretch of 40 residues with a total of at least 9 asparagines and/or lysines residues. The final criterion is the 40 residue window must have a ratio of acid to basic residues of no
greater than 0.9. Using these sets of criteria PlasmoAP was able to obtain similar results to that of PATS (Foth et al., 2003).

Despite understanding the function of the bipartite leader sequence as well as predicting several patterns within its sequence, the precise mechanism of action of how the protein is delivered across 4 membranes is still not fully understood. As mentioned previously, transport across the outermost membrane is mediated by the SP. The SP is a short hydrophobic peptide that facilitates entry into the endomembrane system, a derived food vacuole. The membrane immediately beneath the outermost membrane is called the periplastid membrane. Passage through the periplastid membrane is postulated to be mediated by an extra set of endosymbiont-derived endoplasmic reticulum-associated degradation complex (Spork et al., 2009). The next pair of apicoplast membranes beneath the periplastid membrane is derived from the original pair of membranes belonging to the engulfed cyanobacteria. The translocon of outer envelope of chloroplast (TOC) and translocon of inner envelope of chloroplast (TIC) complexes are postulated to facilitate protein import across these 2 membranes (van Dooren et al., 2001). However, only two TIC components—Tic20 and Tic22—have been found to associate with the inner apicoplast membrane (van Dooren et al., 2008).

One of the interesting issues relating to the apicoplast is the unique ‘delayed death’ phenomenon. It characterizes the death of parasites only in the generation following drug intervention. Treatment of the malaria parasites with ciprofloxacin, an inhibitor of the bacterial DNA gyrase, and other antibiotics including chloramphenicol and tetracycline resulted in the arrest of growth in the second asexual cycle, while the parasites in the current cell cycle appeared relatively unaffected (Surolia et al., 2004). As expected of its
eubacterial ancestry, the machinery that maintains the apicoplast is prokaryotic in origin. Interestingly, drugs that disrupted the biosynthesis of fatty acids, isoprenoids and heme resulted in relatively rapid death of the parasites (Ramya et al., 2007). Such rapid elimination of parasites with drugs suggests that one or more anabolic products of the plastid being essential for the parasites to establish new infections (Ramya et al., 2007). It has been postulated that the apicoplast probably plays a crucial role in the successful establishment of parasite–host interaction through the formation of a functional parasitophorous vacuole (Ralph et al., 2004). Despite recent efforts in metabolic mapping, the precise identity, quantity and functionality of metabolic products by the apicoplast remain enigmatic.
1.7. Thesis Rationale

The Clp chaperone and proteolytic system together represent complex disaggregation, unfolding and degradation machinery that play a pivotal role in protein homeostasis. While ClpP in most organisms is encoded by a single gene, ClpP becomes increasingly diverse in plants, algae, and cyanobacteria in which multiple paralogs of ClpP are present along with an inactive variant called ClpR. The Clp protease complex in these organisms contains both ClpP and ClpR. Both ClpP and ClpR have been found in *Plasmodium falciparum*. Here we attempt to characterize the Clp chaperone and proteolytic system in *Plasmodium falciparum*, the blood-borne obligatory protozoa parasite responsible for malaria in humans. Given that only one isoform of each nuclear-encoded ClpP and ClpR have been identified, we chose *P. falciparum* as an ideal organism for characterization of the ClpPR hetero-proteolytic complex. The Clp system is predicted to be located within the apicoplast, the relic plastid organelle in *Plasmodium*. The apicoplast is indispensable and its prokaryotic origin has been viewed as the parasite’s Achilles’ heel. The current greatest challenge in malaria management is resistance to conventional drugs. It is imperative to discover new drug targets to manage the disease. Knowledge of the apicoplast’s enzymatic machinery will be beneficial towards that goal.
Chapter 2. MATERIALS AND METHODS

The *Plasmodium falciparum* 3D7 cell lysate was prepared by Kodjo Ayi from the McLaughlin Rotman Center for Global Health at MaRS, Toronto. Transmission Electron Microscopy and Scanning Transmission Electron Microscopy were performed by Kevin L.Y. Cheung from Dr. Joaquin Ortega’s lab at the McMaster University, Hamilton, Canada. Parasite gene constructs tagging experiments and Immunofluorescence experiments were performed by Dr. Geoffrey McFadden from the University of Melbourne, Australia. Surface Plasmon Resonance experiments were performed by Majida El Bakkouri from Dr. Walid Houry’s lab at the University of Toronto, Toronto. X-ray crystallography. The X-ray structure was reported by Dr. Raymond Hui’s group from the Structural Genomic Consortium, Toronto. The X-ray structure analyzed and described here by Majida El Bakkouri from Dr. Walid Houry’s lab at the University of Toronto, Toronto.
Unless indicated otherwise, from here forth, all Clp proteins are of *Plasmodium falciparum* origin. All nucleotides and reagents were purchased from Sigma. The AKTA FPLC system and all chromatography columns were from Amersham Biosciences.

### 2.1 Plasmid constructs

The pET21aClpP and pET15ClpR plasmids were a generous gift from Dr. Raymond Hui from the Structural Genomic Consortium, Toronto. The pET21aClpP plasmid has a single methionine deletion (Δ-1M) that removes an internal translation initiation site, which we refer to simply as ClpP. All N-terminal deletion constructs were generated by amplification (PCR) using primers that introduce an *Nde I* site in-frame into the first codon of each construct and a *BamHI* site after the termination codon. The amplified DNA was cloned into both expression plasmids, p11 (a kind gift from Dr. Alexei Savchenko, Clinical Genomic Centre, Toronto) and pET-9a. p11 vector is a modified pET vector with a N-terminal His<sub>6</sub> tag followed by a TEV cleavage site to facilitate Ni<sup>2+</sup>-NTA agarose beads affinity binding protein purification. pET-9a does not have an N-terminal His<sub>6</sub> tag, thus protein purification will be carried out using progressive chromatography columns. All constructs were verified by DNA sequencing.

### 2.2 Expression and Purification

The expression of the recombinant ClpP and ClpR were performed in BL21(DE3) CodonPlus pRIL *E. coli* by inducing the transformed strain at an absorbance (600 nm) of 0.6 for 18°C O/N in Terrific Broth with 1 mM IPTG. Cells were harvested by centrifugation and lysed by sonication for 5 mins at 10 s intervals on ice with the addition
of 5 mg/mL of lysozyme. For His$_6$ tagged ClpP and ClpR, protein purification was carried out with Ni$^{2+}$-NTA agarose beads (Qiagen) using standard manufacture’s protocol. Both purified ClpP and ClpR were dialyzed into Buffer P (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10% glycerol and 1 mM DTT). For untagged ClpP constructs, cells were resuspended in Buffer A (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10% glycerol, 1 mM DTT) and lysed by sonication. Cell lysate was filtered and subjected to crude ion exchange chromatography purification using Q-sepharose column attached to an AKTA FPLC system. ClpP eluted at approximately 15-20% Buffer B (50 mM Tris-HCl, pH 7.5, 1 M NaCl, 10% glycerol, and 1 mM DTT). Elution fractions containing ClpP are pooled, dialyzed back in Buffer A, and subjected to more resolved ion exchange chromatography using MonoQ 5/5 HR with similar procedures. Fractions containing ClpP were pooled, concentrated, and subjected to size exclusion chromatography purification using calibrated Superdex 200 HR 10/30 column in Buffer P (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10% glycerol and 1 mM DTT). ClpP elutes at fractions corresponding to the molecular weight of its heptameric form. These fractions were pooled, concentrated, flash frozen in liquid nitrogen and stored in -80°C freezer. For untagged ClpR constructs, cells were resuspended in Buffer A with 5 mg/ml of lysozyme and lysed by sonication. Crude lysate was filtered and subjected to both Q-sepharose and SP-sepharose ion exchange chromatography columns. Contaminants were bound to columns and removed, while the unbound ClpR proteins were recovered in flow-through. Flow-through fractions containing ClpR were pooled, concentrated, dialyzed in Buffer B and subjected to hydrophobic interaction chromatography using phenyl sepharose 6 FF high-sub column. ClpR eluted at approximately 75-85% Buffer C (50 mM Tris-HCl, pH 7.5, 10% glycerol
and 1 mM DTT). The elution fractions were pooled, concentrated, and subjected to size exclusion chromatography purification using calibrated Superdex 200 HR 10/30 column in Buffer P. ClpR elutes at fractions corresponding to the molecular weight of its heptameric form. These fractions were pooled, concentrated, and flash frozen in liquid nitrogen and stored in -80°C freezer. All proteins were analyzed on SDS-PAGE to be >95% pure. Protein concentrations were determined using Bradford assay (Bio-Rad). All reported concentrations are calculated as monomers unless indicated otherwise.

2.3 Antibodies for ClpP and ClpR

Short peptides, ETKLPHPYFNKVEK and ADEAVDFKLIDHILEKE, based on the C-terminal amino acid sequences of ClpP and ClpR respectively, were synthesized at the Advanced Protein Technology Center at the Hospital for Sick Children. These peptides were conjugated to Keyhole Limpet Hemocyanin and were used to generate polyclonal antisera in rabbit by four subcutaneous injections (500 mg injection followed by 250 mg boosts), using Freund’s adjuvant and standard protocols.

2.4 Preparation of *P. falciparum* parasitized erythrocyte at mature stages

*Plasmodium falciparum* 3D7 cells were a generous gift from Dr. Kain’s laboratory, the McLaughlin Rotman Center for Global Health at MaRS, Toronto. Briefly, red blood cells (RBC) were isolated from plasma and white blood cells by percoll-mannitol gradient (80% percoll, 6% mannitol, 10 mM glucose and PBS) centrifuged and washed three time in wash medium (RPMI 1640 medium containing 25 mmol/L Hepes, 20 mmol/L glucose and 1.25 ml/L gentamicin, pH 6.8). Infection of culture was
performed using *P. falciparum* 3D7 strain, mycoplasma-free. 3D7 parasite was cultivated at 2% hematocrit and 4% parasitemia in culture medium. Culture medium (RPMI 1640 medium with Hapes and glutamine) was supplemented with 10% fresh autologous serum, 20 mmol/L glucose, 1.35 mg/L hypoxanthine and 1.25 mg/L gentamycin, and changed daily to ensure optimal parasite growth. Mature stages (trophozoites and schizonts) culture was separated on percoll-mannitol gradients. After purification, the parasitemia was usually >95% of mature stages. Mature stage layer was harvested, washed, and resuspended in PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na2HPO4, 0.24 g KH2PO4 in 1 mL of sterilized H2O adjusted the pH to 7.4 with HCl) and subjected to freeze/thaw cycles and sonication for 10 mins at 10 s intervals for lysis.

### 2.5 Western Blot and Saponin Treatment

*Plasmodium falciparum* 3D7 cell lysate was centrifuged at 15,000 g (15 min at 4°C). 20 mL of supernatant was separated on 12% SDS-PAGE gels and blotted onto nitrocellulose or PVDF for Western blotting analysis using the anti-ClpP or anti-ClpR serum at 1:1500 and 1:1000 dilutions, respectively. To isolate whole parasite cells from RBC, fresh parasite culture was incubated with 0.02% saponin (Sigma) in PBS for 10 min on ice. The mixture is then centrifuged at 15,000 g. The pellets, which contain whole parasite cells surrounded by empty RBC membranes, also known as ghost, were collected by centrifugation while the supernatant, which contains RBC cytosol, was discarded. The parasite pellets were washed twice with PBS and lysed by sonication for 10 mins at 10 s intervals. All fractions were subjected to analysis by SDS polyacrylamide gel.
2.6 Peptidase assays

Peptidase assays for ClpP(179-370) and ClpR(49-244) were carried out in a 200 µL reaction volume in a 96 well plate. The reaction mix contained 20 µM of recombinant protein in assay buffer D (0.1 M sodium acetate, pH 7.0, and 1 mM DTT) in presence or absence of an inhibitor (chymostatin, phenylmethyl sulphonyl fluoride, pepstatin, leupeptin, or aprotinin at 100 µM). Six different fluorogenic peptide substrates were used: Suc-LLVY-AMC, Suc-LY-AMC, Suc-IIW-AMC, Suc-IA-AMC, Suc-AAPF-AMC, and Suc-AFK-AMC added at 50 µM final concentration. AMC release was continuously monitored by the increase in fluorescence intensity (excitation 355 nm, emission 460 nm) for 6 hours at 30°C using the EnSpire 2300 Multilabel Plate Reader (PerkinElmer). For comparison, *Ec*ClpP protease activity was tested under the same conditions but in buffer E (50 mM TrisHCl, pH 8.0, 0.2 M KCl, and 1 mM DTT). $K_M$ and $k_{cat}$ for the hydrolysis of Suc-LLVY-AMC substrate (1.9 – 62.5 µM) at constant enzyme concentration (20 µM) were obtained by fitting the enzyme kinetic data to the Michaelis-Menten model using SigmaPlot.

2.7 Protease Assay

Protease activity at 37 °C of ClpP and ClpR was quantified by measuring the rate of processive degradation of the α-casein. 4 µM ClpP and ClpR were preincubated in Buffer D for 3 min at 37 °C. 50 µM of α-casein was then added, and degradation was allowed to proceed for 24 h at 37 °C. Aliquots were taken at indicated time points and analyzed on SDS-PAGE gels.
2.8 GFP-SsrA Degradation Assay

This assay exploits the 11 amino acid long sequence (AANDENYALAA) called SsrA. It is a model target of the ClpXP system in *E.coli* as part of the strategy for the organism to remove stalled nascent peptide chains in the ribosomes. Thus, substrates with a C-terminal SsrA tag will be degraded processively by ClpXP *in vitro* in the presence of ATP. Typical reaction mixture includes 16 mM of creatine phosphate, 4 mM ATP, 1.3 units of creatine kinase, 3.9 µM substrate protein, 1.2 µM ClpP or ClpR in Buffer E (50 mM Tris, 5 mM MgCl₂, 5 mM KCl, 0.03% Tween-20, and 10% glycerol). Mixture is preincubated at 37°C for 3 mins prior to the addition of 1.0 µM of *E.coli* ClpX, which initiates the degradation. Reaction is allowed to continue at 37°C. Samples were taken at the indicated time points and the degradation was stopped by addition of 4x Laemmlli buffer and boiling. Samples were then analyzed on SDS-PAGE. Degradation of GFP-SsrA was also monitored by fluorescence using a Fluorolog Spectrofluorometer (Jobin Yvon) with the excitation wavelength set at 395 nm and the emission wavelength set at 509 nm.

2.9 Size-exclusion chromatography

Size-exclusion chromatography was performed using a calibrated Superdex 200 HR 10/30 column attached to an AKTA FPLC system. The column was equilibrated with Buffer P for ClpP and ClpR. Experiments were performed at 4°C, and absorbance was monitored at 280 nm. Equal molar concentrations of ClpP and ClpR proteins were added.
2.10 Analytical Ultracentrifugation – Sedimentation Equilibrium

Sedimentation equilibrium experiments and analyses were performed at the Ultracentrifugation Service Facility at the University of Toronto Department Of Biochemistry. ClpP and ClpR proteins at different concentrations in buffer P were spun at 6000, 8000, and 10,000 rpm at 4°C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti rotor. Absorbance was monitored at 230 and 280 nm. Data analysis was performed using the Origin MicroCal XL-A/CL-I Data Analysis Software Package version 4.0.

2.11 Transmission Electron Microscopy

ClpP and ClpR protein samples were applied by floating a 10 μL drop to carbon-coated grids previously glow discharged and negatively stained with 2% uranyl acetate. Specimens were observed in a JEOL 2010F electron microscope operated at 200 kV. Images were collected at 50,000x with a dose of ~10 electrons/Å². All images were recorded on Kodak SO-163 films, scanned on a Nikon Super COOLSCAN 9000 ED.

2.12 STEM – Scanning Transmission Electron Microscopy

STEM was performed in collaboration with Dr. Joaquin Ortega at the McMaster University, Hamilton, Canada. Specimens were prepared according to the standard method of the Brookhaven National Laboratory STEM facility. Briefly, a 2-μL drop of a 100-μg/mL solution of tobacco mosaic virus (TMV), used as a calibration standard, was absorbed for 1 min onto freshly prepared carbon films supported on a holey membrane on a titanium grid. The grid was washed four times in water, and a 3-μL drop of a 1/10
dilution of previously assembled ring-shaped particles was applied to the grid for 1 min and thoroughly washed five times with 100 mM ammonium acetate and five more times with 20 mM ammonium acetate. The excess liquid was blotted with filter paper, and the grid was plunged into liquid nitrogen slush for quick freezing. The ∼1-μm-thick ice layer was freeze dried by transferring the grid into an ion-pumped chamber and gradually increasing the temperature (overnight) to −80 °C. Grids were then transferred under vacuum to the STEM. Images were collected on the Brookhaven National Laboratory 40-kV STEM with a 0.25-nm probe. Low-dose techniques were used to collect the images with an average electron dose <1000 e−/nm², and the grid was kept at −150 °C during data collection, which keeps the mass loss to 2% due to radiation damage. Images were collected digitally from the large detector (40–200-mrad acceptance angles) and used for mass determination. Large-angle signals are proportional to the mass in the irradiated pixel for thin specimens. Obtained images had a scan width of 0.512 μm.

2.13 Homology Modeling

The monomeric structure of ClpR was modeled based on Myobacterium tuberculosis ClpP1 (PDB code 2ce3C) using SWISS-MODEL followed by alignment computed by hhsearch. The MtClpP1 subunit has the highest sequence identity of 28.4% with ClpR. Other homologous templates with structures available in PDB have been found using the psi-BLAST analysis on the NCBI database; however, none of them have suitable alignments with ClpR for homology modeling. Structural correctness on the generated ClpR model, such as phi/psi angles, peptide bond planarity, bond lengths, bond
angles, and side chain conformations has been evaluated using PROCHECK. Images were prepared using SPDBV and PyMol.

2.14 Tagging of PfcClpC, PfcClpB1, and PfcClpB2 gene products by 3' replacement

Tagging experiments were done in collaboration with Dr. Geoffrey McFadden from the University of Melbourne, Australia. Streptavidin and three tandemly linked haemagglutinin (HA) epitopes were inserted into the 3' end of the clpC, clpB1, clpB2 genes of the 3D7 strain using the Gateway multisite systems vectors (Tonkin et al., 2004) as previously described (Kalanon et al., 2009). P. falciparum parasites were transfected by electroporation of ring stage cultures using 150 µg of plasmid DNA as previously described (Tonkin et al., 2004). Following recovery of drug resistant parasites, cultures were grown without the selective drug, 5 nM WR99210, for 14 days before drug selection was reapplied. When the parasite cultures had recovered from reapplication of the selective drug, they were grown without the drug for another two-week cycle, before reapplication of drug. This cycling between growth on drug and no drug was continued until it was determined that the wild type 3' loci had been replaced by the insertion tag and this was verified by PCR screening.

2.15 Immunofluorescence assays

Immunofluorescence experiments were done in collaboration with Dr. Geoffrey McFadden from the University of Melbourne, Australia. Cultures of P. falciparum were grown to at least 5% parasitaemia then fixed and labeled as described (Tonkin et al., 2004). Antibodies were used as follows: Primary antibodies were rabbit anti-ACP
(1:1000), Roche rat anti-HA (1:200), affinity purified anti-ClpM (1:100), and anti-ClpP (1:1000). Secondary antibodies were Alexa Fluor 546 red anti-rabbit IgG (1:750), Alexa Fluor 488 green anti-rat IgG (1:750), and AlexaFluor 488 green anti-goat IgG (1:1000). Images were captured on a Leica TCS 4D confocal microscope.

2.16 Surface Plasmon Resonance experiments

Surface Plasmon Resonance experiments were done at 25 °C using a Biacore X instrument (GE Healthcare). ClpP (179-370) or ClpR(49-244) at a concentration of 100 μg/mL were covalently linked to an activated Biacore CM5 sensor chip by amine coupling following the manufacturer's protocols. For binding experiments, sensograms were recorded at 20 μL/min flow in running buffer R (10 mM Hepes, pH 7.5, 150 mM NaCl, 3 mM EDTA, 0.005% (w/v) P20 surfactant). The surface was regenerated between injections with a 1 min pulse of 2 M NaCl. The steadystate responses were plotted against the corresponding analyte concentrations and the dissociation constants were derived by fitting the data to a Langmuir binding isotherm using BiaEvaluation 4.1 software (GE Healthcare).

2.17 X-ray crystallography

The structure was reported by our collaborators earlier as part of a structural genomics project (accession number 2F6I) (Vedadi et al., 2007), however, the structure was not analyzed or described in that study. The structure was further refined here. Many different ClpP truncation constructs were generated for crystallization trials. All constructs had an N-terminal His6-tag followed by a tobacco etch virus (TEV) cleavage
site added prior to the ClpP sequence being tested. The truncated protein constructs were expressed and purified according to protocols previously described (Vedadi et al., 2007) using the Lex bioreactor system (Harbinger Biotechnology and Engineering, Markham, Ontario, Canada).

The protein that crystallized had the N-terminal tag MGSSHHHHHHSSGRENLYFQGHM followed by the protein sequence from D179 – K370. Crystals of the purified ClpP protein were grown using the hanging drop vapor diffusion method. The drop was formed by mixing equal parts protein solution and a reservoir solution of 23% PEG MME 550, 200 mM ammonium sulfate in 100 mM cacodylate buffer, pH 7.0. The rod shaped crystals belong to orthorhombic space group C222₁, with unit cell dimensions a=158.3 Å, b=196.5 Å, c=139.2 Å. For data collection, the crystals were first transferred to a cryo-protectant solution consisting of 50 µL of reservoir solution supplemented with 15 mg sucrose and were then flash frozen in liquid nitrogen. Data at 3.0 Å resolution used for initial phasing were collected at the IMCA-CAT beamline 17-ID at the Advanced Photon Source, Argonne National Laboratory. Subsequently, higher resolution data at 2.45 Å used for structure refinement were collected at beamline X25, National Synchrotron Light Source, Brookhaven National Laboratory. All data were processed and scaled using HKL-2000 (Otwinowski and Minor, 1997).

The positions of seven molecules in the asymmetric unit were determined by AMoRe (Navaza 1994) molecular replacement programs from the CCP4 crystallographic program suite using as a search model the highly conserved portions of the E. coli ClpP heptamer, PDB entry 1TYF (Wang et al., 1997). The model was completed using Coot
interactive graphics (Emsley and Cowtan, 2004), alternated by cycles of refinement with CNS Solve version 1.1 software (Brunger et al., 1998). In early stages of refinements, simulated annealing protocols were used, followed by simple positional refinement and individual $B$ factor refinement. In later stages of the refinement, water molecules were added to the model where $|F_o-F_c|$ electron density maps showed peaks at least $3\sigma$ above background and in positions appropriate for hydrogen bonding. Non-crystallographic symmetry restraints between the seven protein molecules were imposed initially during refinement, but restraints were released in later stages. The final ClpP model includes a seven-fold heptamer, with the full tetradecamer biological unit formed by crystallographic symmetry. The seven independent molecules each include residues 179-365 or 366, but contain a 6-15 residue break in the peptide chain where electron density is not observed, typically occurring prior to residue 304 or 305. Statistics for data collection and structure refinement are summarized in Table 3. Molprobity (Davis et al., 2004) was used to evaluate the structure.
Chapter 3. Results and Discussion

Expression and localization of the PfClp proteins were performed by Dr. Geoffrey McFadden’s lab from the University of Melbourne, Australia. Surface Plasmon Resonance analysis of the interaction between PfClpP(179-370) and PfClpR(49-244) was performed by Majida El Bakkouri from Dr. Walid Houry’s lab at the University of Toronto, Toronto. Transmission Electron Microscopy and Scanning Transmission Electron Microscopy were performed by Kevin L.Y. Cheung from Dr. Joaquin Ortega’s lab at the McMaster University, Hamilton, Canada. The X-ray structure of PfClpP was analyzed and described here by Majida El Bakkouri from Dr. Walid Houry’s lab at the University of Toronto, Toronto.
3.1. Identification of putative clp genes in *Plasmodium falciparum*

From PlasmoDB (Aurrecoechea *et al.*, 2009), an online *Plasmodium* genome database, we identified putative *Plasmodium* Clp proteins using BlastP analysis. Protein sequences from *Escherichia coli* K12 ClpP (accession no. P0A6G7), *Synechococcus elongatus* PCC7942 ClpC (BAD79443), and *S. elongatus* PCC7942 ClpX (AAL03913) from the NCBI database were used as input protein sequences using default parameters. Two sequences were found producing high-scoring segment pairs; PFC0310c with a high score of 380 and smallest sum probability of 5.9e-37, and PF14_0348 with a high score of 296 and smallest sum probability of 4.7e-28. Sequence data from GeneDB showed that both are nuclear-encoded proteins that are absent from the apicoplast genomes; PFC0310c on chromosome 3 and PF14_0348 on chromosome 14. PFC0310c has 43% sequence identity and 65% sequence similarity with *E.coli* ClpP, and PF14_0348 has 35% and 56% respectively. Each sequence was also compared against the Clusters of Orthologous Groups (COG) database using the COGNITOR program. Both sequences were identified as belonging to COG0740, a COG group corresponding to subunit of ATP-dependent Clp proteases. The COG groups were developed by comparing protein sequences in complete genomes consisting of orthologous and/or paralogous protein sequences. We subsequently aligned *E.coli* ClpP with PFC0310c and PF14_0348 using ClustalW with default parameters (Figure 3.1). From the alignment, we observed that PFC0310c possesses three catalytic residues (Ser264, His289, Asp338). The unique arrangement of the catalytic triad residues (Ser, His, Asp) in that order further suggests that PFC0310c belongs specifically to the MEROPS peptidase family S14 of the ClpP endopeptidase family from Clan SK. PF14_0348 lacks two of the three catalytic residues
(Gly148, Asn173, Asp223); the serine and histidine were replaced by glycine and asparagine (Figure 3.1). Catalytic serine and histidine are critical residues in the enzyme’s active site whose side chains are responsible for initiating nucleophilic attacks on the trigonal carbonyl carbon of scissile peptides. Given its high sequence similarity to ClpP along with its inability to cleave substrates, we believe PF14_0348 belongs to an inactive isoform of ClpP of unknown function called ClpR. Unless indicated otherwise, we will refer to PFC0310c as ClpP, and to PF14_0348 as ClpR.

In addition to ClpP and ClpR, four putative proteins from the PlasmoDB database displayed significant sequence homology to *S. elongatus* ClpC (Table 1). One putative clp ATPase gene (PFC10_API0060) is located within the plastid DNA of the apicoplast, while three other genes (PF11_0175, PF14_0063 and PF08_0063) are found on the nuclear chromosomes. We will refer to these four ATPases as *Pf*ClpM, *Pf*ClpB1, *Pf*ClpB2, *Pf*ClpC respectively (Table 1). No proteins were found to bear high sequence identity to *S. elongatus* ClpX. *P. falciparum* also possess *Pf*HslU (*Pf*ClpY, PFI0355c) and *Pf*HslV (*Pf*ClpQ, PFL1465c) proteins localized to the mitochondria, but they will not be discussed in my thesis.
Table 3.1. The Clp proteins of *Plasmodium falciparum*. List of *Plasmodium falciparum* Clp proteins identified using BlastP analysis from the NCBI and PlasmoDB database. Protein sequences from *Escherichia coli* K12 ClpP (accession no. P0A6G7), *Synechococcus elongatus* PCC7942 ClpC (BAD79443), and *S. elongatus* PCC7942 ClpX (AAL03913) were used as input protein sequences. MW represents the theoretical molecular weight of the unprocessed protein based on its gene size. Gene ID is based on PlasmoDB gene ID.

<table>
<thead>
<tr>
<th>Name</th>
<th>MW (kDa)</th>
<th>Gene ID</th>
<th>Gene Location</th>
<th>Protein Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfClpP</td>
<td>43</td>
<td>PFC0310c</td>
<td>nucleus</td>
<td>Clp protease</td>
</tr>
<tr>
<td>PfClpR</td>
<td>28</td>
<td>PF14_0348</td>
<td>nucleus</td>
<td>inactive Clp protease</td>
</tr>
<tr>
<td>PfClpB1</td>
<td>123</td>
<td>PF08_0063</td>
<td>nucleus</td>
<td>Clp ATPase</td>
</tr>
<tr>
<td>PfClpB2</td>
<td>103</td>
<td>PF11_0175</td>
<td>nucleus</td>
<td>Clp ATPase</td>
</tr>
<tr>
<td>PfClpC</td>
<td>156</td>
<td>PF14_0063</td>
<td>nucleus</td>
<td>Clp ATPase</td>
</tr>
<tr>
<td>PfClpM</td>
<td>91</td>
<td>PFC10_API0060</td>
<td>apicoplast</td>
<td>Clp ATPase</td>
</tr>
</tbody>
</table>
All *Plasmodium falciparum* Clp ATPases that we have identified have a predicted characteristic N-domain and two AAA+ domains, D1 and D2, which comprise a large and a small subdomain each (Figure 1.1). The conserved Walker A, Walker B and different sensor motifs can be identified in all the AAA+ domains of *P. falciparum* Clp ATPases. The only exception is the apicoplast-encoded PFC10_API0060 which does not possess the canonical GKT or (I/V)DEI sequences of the Walker A and Walker B motifs in its AAA+1 domain (Figure 3.2). Thus we postulated that PFC10_API0060 AAA+1 is not an active ATPase. In accord with nomenclature for Clp ATPases as per recommendation by Schirmer *et al.*, we have named the inactive ATPase PFC10_API0060 as *PfClpM* (Table 1).

Unlike *EcClpA*, *PfClp* ATPases have an insertion of 18-174 residues within the small subdomain of the D1 module (Figure 3.2). Compared to other *PfClp* ATPases, *PfClpM* has the smallest of such insertion regions. We decided to use the size of this insertion, as well as, the presence of certain motifs as basis for categorizing the Clp ATPases into different subfamilies, such as ClpB, ClpC and ClpD (Figure 3.2). In the case of ClpB, we classify Clp ATPases whose insertion region is predicted to form a coiled coil as belonging to the ClpB subfamily. The reason behind such classification is the said insertion region forms a coiled coil in *Thermus thermophilus* ClpB (*TtClpB*) based on its X-ray structure. Using programs COILS (Lupas *et al.*, 1991) and Paircoil (Berger *et al.*, 1995), the insertion region of both PF08_0063 and PF11_0175 are predicted to form coiled coils; hence, we name them *PfClpB1* and *PfClpB2*, respectively (Table 1). Close inspection of the insertion region based on sequence alignment of *PfClpB1*, *PfClpB2*, *TtClpB*, *SeClpB* and *EcClpB* reveals that residues that are part of the
leucine rich heptad repeat in the TtClpB structure are generally conserved on all these ClpBs, but to a lesser degree in PfClpB2 (Figure 3.2). Note also that the insertion region of PfClpB1 is interrupted by a 53 amino acid residue segment.

In terms of putative association with ClpP or ClpR, it is known that the interaction between the ATPase and the Clp protease complex is mediated in part by a surface loop in the Clp ATPase proteins termed the ClpP binding loop (Figure 1.3). This loop is located after the Walker B motif in the AAA+ module between the Sensor I and Box VII motifs and has a conserved tripeptide consensus sequence of [L/I/V]-G-[F/L]. Clp ATPases that lack this tripeptide motif are not known to form a complex with the ClpP protease (Figure 1.3). Based on multiple sequence alignment, only PfClpC is found to have the LGF ClpP binding loop in the correct position (Figure 3.2). There is also an Asn-rich insertion of 94 amino acids in this putative ClpP-binding loop of PfClpC. Hence, we propose that PfClpC associates with ClpP and ClpR to form the P. falciparum chaperone–protease complex that, as discussed below, is proposed to be localized to the apicoplast.
**Figure 3.1. Sequence properties of the *P. falciparum* Clp proteases.** Sequence alignment of *EcClpP*, *PfClpP*, and *PfClpR* using ClustalW2 (Larkin et al., 2007). The alignment is drawn using ESPript (Gouet et al., 1999). Residues that are 100% identical in the three sequences are highlighted in red, while those in red font are highly similar. Residues of the Ser-His-Asp catalytic triad are in bold green and are indicated by an asterisk. N150 and D179 of *PfClpP* and S49 of *PfClpR* are indicated with brackets and red arrows. Secondary structure elements shown on top of the sequence alignment are based on the *EcClpP* structure (PDB code 1yg6). The axial loop present at the N-terminus of *EcClpP* and the handle region formed by β6 and αE are indicated.
Figure 3.2. Domain arrangement of the *P. falciparum* Clp ATPases. Cartoon representation of the domain organization of the *P. falciparum* Clp ATPases. *EcClpA* is shown for comparison. The residue numbers of the lysine in the Walker A motif ‘GKT’, the aspartic acid in the Walker B motif ‘ΦDE’, and the glycine in the ClpP binding loop motif ‘[L/I/V]G[F/L]’ are indicated. ‘CC’ refers to the presence of coiled coils as predicted by the COILS (Lupas *et al.*, 1991) and Paircoil (Berger *et al.*, 1995) programs.
3.2. Apicoplast targeting predictions

Through sequence alignment, we observed that both ClpP and ClpR have an N-terminal extension compared to *E.coli* ClpP, albeit ClpR has a shorter one in comparison (Figure 3.1). The sequences seem to be rich in lysine and asparagine with an overall positive charge and variable in size, suggesting a very plastic composition. In plants, numerous proteins contain similar N-terminal extensions that, upon translation, target nuclear-encoded proteins to the chloroplast, a relict organelle involved in photosynthesis (Figure 1.6). A similar non-photosynthetic version exists in several *Plasmodium* spp., called apicoplast, the plastid of apicomplexan parasites. We believe both ClpP and ClpR possess N-terminal leader sequences that target them towards the apicoplast upon translation. The apicoplast is slightly different than its plant cousin in that it has 4 membranes which are postulated to be a result of secondary endosymbiosis (Figure 1.6). Therefore, entry of protein into the apicoplast necessitates a bipartite leader sequence that has 2 functional domains; an N-terminal signal peptide (SP) followed immediately by a transit peptide (TP). The SP commits the nascent polypeptide chain into the endomembrane system which is contiguous with the apicoplast; The TP subsequently delivers the protein into the plastid (Waller *et al.*, 1998, 2000).

The targeting sequence predictor algorithms PlasmoAP and PATS were used to screen for the presence of an apicoplast targeting sequence in the *P. falciparum* Clp proteins. Both algorithms predict that *Pf*ClpP, *Pf*ClpR and *Pf*ClpB1 contain apicoplast targeting sequences, and, hence, should be imported into the apicoplast (Table 3.2). *Pf*ClpB2 is predicted by PlasmoAP but not by PATS to have a targeting sequence. However, the PlasmoAP prediction is wrong as we discovered through
immunofluorescence studies on parasite cells in section 3.4 that PfClpB2 is localized in the parasitophorous vacuole. Finally, PATS but not PlasmoAP predicts an apicoplast targeting sequence in the PfClpC protein (Table 3.2).
Table 3.2. Apicoplast targeting prediction of *Pf*Clp proteases and ATPases. The N-terminal bipartite leader sequence of *Plasmodium falciparum* ClpP, ClpR, ClpB1, ClpC and ClpM are analyzed using PlasmoAP (Foth *et al.*, 2003) and PATS (Zuegge *et al.*, 2001) to predict if it is an apicoplast targeting sequence or not.

<table>
<thead>
<tr>
<th>Name</th>
<th>PlasmoAP (SP, apicoplast TP)</th>
<th>PATS</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pf</em>ClpP</td>
<td>Yes (+,+)</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Pf</em>ClpR</td>
<td>Yes (+,+)</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Pf</em>ClpB1</td>
<td>Yes (+,+)</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Pf</em>ClpB2</td>
<td>Yes (+,+)</td>
<td>No</td>
</tr>
<tr>
<td><em>Pf</em>ClpC</td>
<td>No (-,+)</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Pf</em>ClpM</td>
<td>not applicable</td>
<td>not applicable</td>
</tr>
</tbody>
</table>
3.3. Experimental localization of the Clp proteins in *P. falciparum*.

In addition to prediction using online algorithms, we attempted to detect the expression and localization of the different *P. falciparum* Clp ATPases and proteases using immunofluorescence technique on fixed parasite cells (Figure 3.3). This is achieved using antibodies to these proteins or by tagging the respective parasite genes with a streptavidin-3×hemagglutinin (Strp-3×HA) tag at the 3′ terminal. Immunofluorescence analysis was done in collaboration with Dr. McFadden’s group at the University of Melbourne, Australia.

A 120 kDa protein was detected using anti-HA antibodies in Western blots from the *PfclpB1*-Strp-3×HA line and a 200 kDa protein was observed in the *PfclpC*-Strp-3×HA (Figure 3.3 A). The apparent mass of *PfClpB1*-Strp-3×HA is close to that predicted, while the apparent mass of *PfClpC*-Strp-3×HA is slightly greater than predicted (Figure 3.3 A). Western blots confirm that both fusion proteins are expressed in the asexual blood stage parasite cells. No antibody reaction was observed in the parental, untagged line of parasites, suggesting the absence of any cross-interactions with the antibodies.

Co-localization analysis was done using antibodies detecting the stromal apicoplast marker acyl carrier protein (ACP) and antibodies against HA. We have discovered that both *PfClpB1*-Strp-3×HA and *PfClpC*-Strp-3×HA produce immunofluorescence that overlaps with ACP, demonstrating that *PfClpB1*-Strp-3×HA and *PfClpC*-Strp-3×HA are indeed localized in the apicoplast (Figure 3.3 B). Western blot analysis demonstrated that *PfClpB2* (Hsp101) is expressed as a 100 kDa protein (Figure 3.3 A). It was found that *PfClpB2*-Strp-3xHA produces immunofluorescence
that overlaps with the parasitophorous vacuole marker \( P/J \text{TEx150} \), suggesting that 
\( P/J \text{ClpB2-Strp- 3xHA} \) is localized within the parasitophorous vacuole indicating that the 
prediction from PlasmoAP on \( P/J \text{ClpB2} \) is wrong (Figure 3.3 B, Table 3.2). A Western 
blot analysis on ring-stage parasite strain 3D7 lysate probed with anti-\( P/J \text{ClpM} \) antibodies 
revealed a single band of \( \sim 95 \) kDa close to its theoretical molecular weight from the 
apicoplast gene (Table 3.2). Co-localization immunofluorescence assays demonstrated 
overlapping signals between \( P/J \text{ClpM} \) and ACP, suggesting that the apicoplast-encoded 
\( P/J \text{ClpM} \) resides within the apicoplast (Figure 3.3 B).

Either a 25 kDa band or a 22 kDa band was identified in Western blots of wild 
type parasites using the antisera against a peptide from the C-terminus of \( P/J \text{ClpP} \) or 
\( P/J \text{ClpR} \), respectively. Immunofluorescence using the antibodies against \( P/J \text{ClpP} \) in early 
trophozoite stage parasite identified a small globular structure that is distinct from the 
nucleus and the food vacuole within the vicinity of the apicoplast. We believe this small 
globular structure is the apicoplast. We were unable to co-localize \( P/J \text{ClpP} \) with ACP 
anticluster because both sera are derived from rabbit. The antibodies against \( P/J \text{ClpR} \) gave 
no signal in immunofluorescence. Unfortunately, both \( P/jclpP \) and \( P/jclpR \) were refractory 
to 3’ end tagging, rendering anti-HA localization studies unavailable at this point.

In summary, the Western blots and immunolocalizations demonstrate that all the 
Clp proteins are expressed in the blood-stage parasites and localization studies confirm 
that \( P/J \text{ClpB1, P/JClpC, P/JClpM} \) are localized to the apicoplast (Figure 3.3). \( P/J \text{ClpP} \) is also 
very likely to be localized to the apicoplast. \( P/J \text{ClpB2} \) is confirmed to be secreted to the 
parasitophorous vacuole, where it may participate in export of virulence proteins to the
red blood cell. At this stage, \textit{PfClpR} is predicted to be targeted to the apicoplast using the prediction algorithms.
**Figure 3.3. Expression and localization of the PfClp proteins.** (A) Expression of PfClpB1, PfClpC, PfClpB2, PfClpM, PfClpP, and PfClpR was checked in blood stage parasites. Shown in the first three lanes are Western blots of lysates from parasites having PfClpB1-Strp-3xHA, PfClpC-Strp-3xHA, and PfClpB2-Strp-3xHA probed with anti-HA antisera. Shown on the right three lanes are lysates from 3D7 wild type parasites probed with anti-ClpM, anti-ClpP, and anti-ClpR peptide antisera. Molecular weight markers are indicated on the left of the lanes. (B) Localization of PfClpB1, PfClpC, PfClpB2, PfClpM, and PfClpP of *P. falciparum* was tested by immunofluorescent colocalization. In panels 1 and 2, PfClpB1-Strp-3×HA or PfClpC-Strp-3×HA parasites were probed with the apicoplast marker anti-ACP antisera (red), with anti-HA antisera (green), and with Hoescht 33342 to stain for DNA (blue). In panel 3, PfClpB2-Strp-3×HA parasites were probed with the parasitophorous vacuole marker anti-PTEX150 antisera (red), with anti-HA antisera (green), and with Hoescht 33342 (blue). In panels 4 and 5, wild type 3D7 strain was probed with anti-ACP antisera (red), with anti-PfClpM antisera or anti-ClpP antisera (green), and with Hoescht 33342 (blue). For all panels, the right most image represents the merged fluorescence images with the DIC or transmission image. The scale bars correspond to 5 μm.
3.4. Sequence boundaries of the mature forms of PfClpP and PfClpR

We attempt to express both PfClpP and PfClpR as described in methods. The full length constructs do not express well in E.coli (data not shown). The presence of the full apicoplast targeting sequence at the N-terminus is not conducive to expression, as reported by various groups. We attempted to selectively truncate the signal peptide (SP) portion of the targeting sequence, but the protein also does not express well (data not shown). Thus the removal of the full targeting sequence is required for proper expression of the protein. Existing prediction algorithms are unable to identify the precise cleavage site of the transit peptide after it has been processed and has entered the apicoplast. Thus, we attempt to examine the processing by analyzing its molecular weight using Western blot analysis.

We generated anti-sera against ClpP and ClpR using short peptides of 15-17 residues long as antigens. These peptides correspond to the C-terminal sequences of PfClpP and PfClpR respectively (ETKLPHPYFNKVEK and ADEAVDFKLIDHILEKE). We chose these as antigens because based on the X-ray structure of EcClpP (Figure 1.2), the C-terminus is relatively unstructured and is projected up and outward near the equatorial region of the cylindrical. Thus the C-termini should be readily accessible in solution by antibodies as epitopes. The two sequences also share very little similarity, therefore, peptides corresponding to these C-terminal sequences were chosen as antigens to confer better antibody-antigen interaction and higher antibody-antigen specificity compared to native proteins, which have similar cylindrical structures. Indeed, initial Western blot tests revealed no cross reactivity between PfClpP and PfClpR as well as among PfClpP, PfClpR and EcClpP (data not shown).
We performed Western blot analyses on native PfClpP and PfClpR (Figure 3.4). Complete genome-wide transcriptome study of *Plasmodium falciparum* performed by the DeRisi lab showed that both *PfclpP* and *PfclpR* were transcribed throughout the erythrocytic schizogony, also known as the asexual intraerythrocytic developmental cycle (IDC), the cycle in which the parasite propagate within the human erythrocytes (Figure 1.4). Thus, mature stages (trophozoites and schizonts) of *Plasmodium falciparum* 3D7 culture were used for our Western blot experiments.

Cell lysates were prepared as described in methods. Initial Western blot analyses on whole cell lysates revealed a prominent band of approximately 25 kDa for both *PfClpP* and *PfClpR* antibodies (Figure 3.4). At first, we believe that to be the mature processed version of *PfClpP* and *PfClpR*. However, to eliminate the possibility of a common cross reacting contaminant, we treated whole cell parasite culture with 0.02% saponin, which selectively punctures the red blood cell membrane without damaging the parasite within. This allows the removal of the cytosolic contents of the red blood cell, leaving behind a ghost, which consists of the intact parasite only surrounded by remaining red blood cell membrane. Western analysis on the concentrated ghost lysate demonstrated that the prominent band observed initially might have been a cross-reacting contaminant of erythrocyte origin. Saponin treatment successfully removed this contaminant, and revealed what we believed to be the mature, processed *PfClpP* and *PfClpR* (Figure 3.4). To estimate the size of these processed proteins, we generated and purified a series of truncated *PfClpP* and *PfClpR* constructs to be used as molecular weight references on SDS-PAGE gels. We found that *PfClpP* beginning at residue N150 and *PfClpR* beginning at residue S49 to be the closest in molecular weight to the
observed processed PfClpP and PfClpR (Figure 3.4). We believe these are the correct processed mature sequence of PfClpP and PfClpR in which their pre-sequences have been cleaved *in vivo*. It should be noted that the presumed auto-catalytic processing of the presequences of PfClpP and PfClpR must have happened at the N-terminus and not the C-terminus, because the antibodies to PfClpP and PfClpR used in the Western blot analysis were raised using small peptides with sequences identical to the C-termini of the corresponding PfClp proteins.
Figure 3.4. Mature \textit{PfClpP} and \textit{PfClpR} in \textit{Plasmodium falciparum}. Erythrocytes (RBC) infected with parasites were lysed with saponin, leaving behind whole parasite cells surrounded by RBC membrane (ghost). Ghosts were sonicated in PBS, and the pellet fractions were washed and subjected to (A) Western blot analysis and (B) simple Coomassie staining.
gel electrophoresis. 100 µg of pure RBC lysate was loaded on the first lane from the left, followed by equal amount of infected RBC lysate and ghost lysate. Purified truncated proteins: PfClpP:N150-K370 and PfClpR:S49-E244 were included on the last lane as reference.
3.5. Oligomeric state of N-terminally truncated PfClpP and PfClpR

Given that PfClpP (150-370) and PfClpR(49-244) represent the predicted mature processed subunits of *P. falciparum* ClpP and ClpR (Figure 3.4), we attempt to investigate their oligomeric state as a complex as well. *Pf*ClpP (150-370) and *Pf*ClpR(49-244) were expressed in *E. coli* and purified as described in Methods. Both *Pf*ClpP(150-370) and *Pf*ClpR(49-244) expressed as soluble proteins. However, while *Pf*ClpR(49-244) was stable, *Pf*ClpP(150-370) seemed to form soluble aggregates (>1 MDa) that appeared in the void volume when subjected to size-exclusion chromatography (SEC) using a Superdex 200 HR 10/30 column. To search for a construct similar in molecular weight that will not form soluble aggregates, a series of N-terminally truncated *Pf*ClpP constructs was generated, starting at: S146, Y148, L152, Y154, D155, N161, Y166, V169, Q171, I173, T175, N177 and D179. Each of these constructs was expressed and purified. *Pf*ClpP(179-370) with a theoretical mass of 22.1 kDa was the only construct to express a soluble and stable protein with minimal soluble aggregates appearing in the void volume. We used this construct for our biochemical, biophysical and structural studies. It should be noted that we have also used *Pf*ClpR(43-244), another stable and soluble construct, for several of our experiments as well.

*Pf*ClpP(179-370) migrated at about 146 kDa and *Pf*ClpR(49-244) migrated at about 134 kDa when subjected to SEC (Figure 3.5 A). The molecular weight obtained from SEC suggested that both proteins form heptameric complexes as compared to tetradecameric EcClpP with a theoretical mass of 303.8 kDa. Changing the concentration of salt between 150 mM and 1 M NaCl or changing the pH between 4 and 9 did not seem
to affect the oligomeric state of PfClpP(179-370) or PfClpR(49-244) as inferred from the lack of change in migration when subjected to SEC (data not shown).

Analytical ultracentrifugation sedimentation equilibrium analysis at 4 °C was performed to further establish the oligomeric state of PfClpP(179-370) and PfClpR(49-244). The plots of ln(A) versus r^2 (where A is absorbance at 280 nm and r is radius) for both PfClpP(179-370) and PfClpR(49-244) are very close to linear, suggesting that the majority of the sample for both proteins is composed of one species (Figure 3.5 B). The global self-association model was used to analyze the data giving an average molecular mass of 182.6 kDa for PfClpP(179-370) and 139.5 kDa for PfClpR(49-244). The ratio of the oligomer mass to the theoretical monomer mass for PfClpP(179-370) is 8.3:1 and for PfClpR(49-244) it is 6.2:1 (Figure 3.5 B). The slightly elevated ratio observed for PfClpP(179-370) might be explained by the possible presence of a small amount of higher oligomeric population, possibly a tetradecameric complex. The slightly lower ratio observed for PfClpR(49-244) might be due to the presence of monomeric or higher order oligomers.

We also employed electron microscopy (EM) to observe the oligomers formed by PfClpP and PfClpR (Figure 3.6). EM was done in collaboration with Dr. Joaquin Ortega from McMaster University, Hamilton. When PfClpP(179-370) was observed under negative staining conditions, the electron micrographs showed abundant ring-shaped particles of ~100 Å in diameter (Figure 3.6 A). Particle images selected from the electron micrographs were analyzed with algorithms that detect rotational symmetry and a strong 7-fold symmetry component was found (Figure 3.6 B). Averaging these images showed a top view of a seven membered ring similar to those observed for EcClpP (Figure 3.6 B).
In order to discriminate whether these top view projections of \( Pf\text{ClpP}(179-370) \) were made of a single or a double heptameric ring, scanning transmission electron microscope (STEM) images were recorded after the specimen was freeze-dried on carbon films and the mass of the round particles in the images was calculated and plotted (Figure 3.6 C). The histogram shows that the distribution of particle mass follows a unimodal distribution centered at the mass expected for a single heptameric ring and only a few particles were found with the mass expected for a double heptameric ring (Figure 3.6 C). These results are consistent with those obtained by SEC and analytical ultracentrifugation sedimentation equilibrium experiments (Figure 3.5). Together these indicate that \( Pf\text{ClpP}(179-370) \) oligomerizes predominantly into a single heptameric ring. It should be noted that single heptameric complex is unlike other ClpPs, such as that of \textit{E. coli} or \textit{Helicobacter pylori} that readily form tetradecamers. However, it is similar to human ClpP that assembles mainly into heptamers and forms tetradecamers in the presence of its cognate ClpX chaperone.

Similar experiments were done with \( Pf\text{ClpR}(49-244) \) and \( Pf\text{ClpR}(43-244) \), two constructs of \( Pf\text{ClpR} \). In both cases, negatively stained electron micrographs showed ring-shaped particles that had a strong 7-fold symmetry component when analyzed for rotational symmetry (Figure 3.6 A). The top-view averages calculated for these two constructs were remarkably similar to those obtained for \( Pf\text{ClpP}(179-370) \) (Figure 3.6 B). However, in addition to the ring-shaped particles observed in the electron micrographs for \( Pf\text{ClpR} \), we found particles made of two parallel striations, albeit at low abundance (Figure 3.6 D). The particles most likely represent side views of two stacked \( Pf\text{ClpR} \) heptameric rings. These results, together with our earlier SEC and analytical
ultracentrifugation experiments, suggest that \( P/\text{ClpR}(49-244) \) and \( P/\text{ClpR}(43-244) \) form single heptameric rings in solution but coexist with a small proportion of double heptameric ring oligomers. We used STEM imaging of the \( P/\text{ClpR}(43-244) \) construct to estimate the proportion of single and double heptameric particles in solution (Figure 3.6 C). Interestingly, the histogram obtained showed that the mass of the \( P/\text{ClpR}(43-244) \) particles followed a bimodal distribution with one peak centered around 150 kDa, the mass expected for a single-heptameric ring, and a second broader peak centered around 230 kDa, which does not correspond to the mass expected for either a single or a double heptameric ring structure (Figure 3.6 C). In addition, some particles were observed with a mass corresponding to a tetradecamer. These results suggest that \( P/\text{ClpR} \) constructs form a heterogeneous oligomeric mixture containing heptamers and tetradecamers and other oligomeric forms of variable mass.
**Figure 3.5. Oligomeric state of PfClpP(179-370) and PfClpR(49-244).** (A) Size exclusion chromatography of the proteins on a Superdex 200 HR 10/30 column in buffer P. The position of the MW markers is shown on top. Amounts loaded are as follows: PfClpP(179-370), 2.50 mg; PfClpR(49-244), 2.50 mg; PfClpP(179-370) + PfClpR(49-244), 1.25 mg + 1.25 mg; EcClpP, 1.00 mg. The silver stained SDS-PAGE gels shown below the chromatograms are of PfClpP(179-370) and PfClpR(49-244) run separately. (B) Data from analytical ultracentrifugation sedimentation equilibrium experiments are shown for the two proteins. The data points depicted were collected at 8,000 rpm and 4°C for both proteins. The red lines in the lower panels correspond to the fit of the data to a monodisperse heptameric model with apparent MWs as given. The residual deviations from the theoretical fits are given in the upper panels.
Figure 3.6. Electron microscopy analysis of PfClpP and PfClpR. (A) Negatively stained electron micrographs showing ring-shaped particles of PfClpP(179-370) and PfClpR(49-244). For PfClpR(49-244), particles made of double parallel striations are also observed in low abundance (black arrow). (B) The calculated averages for the top views of PfClpP(179-370), PfClpR(49-244), and PfClpR(43-244) show pronounced seven-fold symmetry (left) and differ very little from the symmetrized version of the average (right). (C) Histograms show the mass distribution of the particles obtained by STEM. The average mass ($M_{ave}$) and standard deviation (in parentheses) are given. (D) Gallery of selected raw particle images showing double parallel striations of density representing
side views of two stacked PfClpR(49-244) heptameric rings (left panels). The calculated average for this type of particles is shown in the right panel.
3.6. The interaction between *PfClpP(179-370)* and *PfClpR(49-244)*

We attempted to determine if *PfClpP(179-370)* and *PfClpR(49-244)* are able to form the typical hetero-tetradecameric Clp protease complex as observed in plants. Since *PfClpP(179-370)* and *PfClpR(49-244)* can readily form single heptameric rings, we speculated that a *PfClpRP* complex would be formed by a *PfClpP* homoheptameric ring interacting with a *PfClpR* homoheptameric ring at the handle region. Pre-incubation of *PfClpP(179-370)* with *PfClpR(49-244)* and analyzing the mixture by size exclusion chromatography did not show the presence of a tetradecameric complex. Analysis of the mixture by AUC-SE gave similar results; only heptameric complexes were found.

To further test the possibility that *PfClpP(179-370)* and *PfClpR(49-244)* might form heteroheptamers *in vivo*, the two proteins were co-expressed in *E. coli* from a polycistronic pST39 plasmid (Tan 2001) with one of the proteins being tagged. The tagged protein was pulled down, however, there was no evidence of significant interactions between *PfClpP(179-370)* and *PfClpR(49-244)* (data not shown). The isolated proteins formed heptamers as detected by size exclusion chromatography indicating that the complexes formed are likely homoheptamers rather than heteroheptamers (not shown).

Surface Plasmon Resonance (SPR) experiments using the BIAcore system were carried out to determine direct binding between the two proteins. *PfClpP(179-370)* or *PfClpR(49-244)* heptamer was immobilized on the sensorchip and sensograms were recorded by injecting the other protein; the apparent dissociation constants were then derived as described in the Methods (Figure 3.7). The binding constants obtained are in the range of 0.1 to 10 μM. The strength of the binding interactions were as follows:

---

85
\( Pf\text{ClpR}(49-244) - Pf\text{ClpR}(49-244) > Pf\text{ClpP}(179-370) - Pf\text{ClpR}(49-244) > Pf\text{ClpP}(179-370) - Pf\text{ClpP}(179-370) \) (Figure 3.7). This is consistent with the observation of double ring complexes by EM for \( Pf\text{ClpR}(49-244) \) and \( Pf\text{ClpR}(43-244) \).

Based on this observation, we speculate that if \( Pf\text{ClpP} \) and \( Pf\text{ClpR} \) were at equimolar concentrations in the apicoplast, then the major tetradecameric complexes observed \textit{in vivo} would be that of \( Pf\text{ClpR} \) homo-oligomers and a \( Pf\text{ClpP}-Pf\text{ClpR} \) double ring formed from a heptameric \( Pf\text{ClpP} \) and heptameric \( Pf\text{ClpR} \). While the reason for a \text{ClpR} homo-complex remains enigmatic, it should be noted that similar heptameric \text{ClpP} rings exist for human \text{ClpP}, and the human ATPase \text{ClpX} is required to facilitate the oligomerization of human \text{ClpP} (Kang \textit{et al.}, 2005). Thus it is reasonable to speculate that the cognate ATPase, which we propose based on the bioinformatics analysis presented above is \( Pf\text{ClpC} \), would be necessary to drive tetradecamer formation and might dictate whether the complex formed is a \( Pf\text{ClpP}_7-Pf\text{ClpR}_7 \), \( Pf\text{ClpP}_{14} \), or \( Pf\text{ClpR}_{14} \) oligomer.
Figure 3.7. SPR analysis of the interaction between PfClpP(179-370) and PfClpR(49-244). Shown are BIAcore sensorgrams and the derived equilibrium binding curves for the interaction between PfClpR(49-244)-PfClpR(49-244), PfClpP(179-370)-PfClpR(49-244), and PfClpP(179-370)-PfClpP(179-370). The $K_d$'s given refer to the apparent dissociation constants between the different protein heptamers. The numbers in parenthesis are the standard deviations on the $K_d$'s.
3.7. Peptidase activity of PfClpP(179-370)

We attempted to assess the peptidase activity of PfClpP and PfClpR. The constructs PfClpP(179-370), PfClpR(49-244), and a mixture of PfClpP(179-370) with PfClpR(49-244) were assessed for peptidase activity (Figure 3.8). We used the model substrate Suc-LY-AMC typically used for measuring EcClpP peptidase (Maurizi et al., 1994), as well as: Suc-AFK-AMC, Suc-AAPF-AMC, Suc-IA-AMC, Suc-IIW-AMC, and Suc-LLVY-AMC. The peptidase activity of EcClpP against Suc-LY-AMC was used as a reference. No activity was observed for PfClpR(49-244) as expected, while very weak activity was detected for PfClpP(179-370) against Suc-LLVY-AMC and Suc-LY-AMC (Figure 3.8 A). The peptidase activity of PfClpP(179-370) was inhibited by two serine protease inhibitors, chymostatin and phenyl methyl sulphonyl fluoride (PMSF). Other inhibitors were tested including leupeptin (serine and cysteine protease inhibitor), pepstatin (aspartic protease inhibitor), and aprotinin (serine protease inhibitor), but no inhibition was observed (Figure 3.8 A). The activity of PfClpP(179-370) was not significantly affected by the presence of PfClpR(49-244). The PfClp proteases were unable to degrade longer peptides such as MCA-YEVHHQKLVFK(DNP)-NH₂ and MCA-YEVHHQKLVFK(DNP)-NH₂ or casein, which was reported to be degraded by EcClpP in the absence of the ATPase chaperone (Jennings et al., 2008; Bewley et al., 2009).

The K_M and k_cat for Suc-LLVY-AMC hydrolysis by PfClpP(179-370) were 65 µM and 0.13 hr⁻¹ (expressed as moles of peptide bonds cleaved per mole of ClpP protomer), respectively (Figure 3.8 B). As a comparison, the K_M and k_cat reported for EcClpP against Suc-LY-AMC (Maurizi et al., 1994) is 1 mM and 150 min⁻¹, respectively.
Hence, $Pj\text{ClpP}(179-370)$ is a very weak peptidase on its own.
Figure 3.8. Peptidase activity of *Pf*ClpP(179-370). (A) Peptidase activity of 20 μM *Pf*ClpP(179-370), 20 μM *Pf*ClpR(49-244), or 20 μM *Pf*ClpP(179-370) + 20 μM *Pf*ClpR(49-244) were measured at 30°C by their ability to cleave 50 μM of: Suc-LLVY-AMC, Suc-LY-AMC, Suc-IIW-AMC, Suc-IA-AMC, Suc-AAPF-AMC, or Suc-AFK-AMC. The peptidase activity of *Ec*ClpP using Suc-LY-AMC as a substrate is shown as a reference. The inset in the lower left panel shows the inhibition of *Pf*ClpP(179-370) peptidase activity against Suc-LLVY-AMC by PMSF and chymostatin measured by the fluorescence intensity change after 6 hrs. of incubation. (B) The change in the initial rate of Suc-LLVY-AMC hydrolysis by 20 μM *Pf*ClpP(179-370) as a function of peptide concentration is shown. The solid line is the fit to the data using the Michaelis-Menten kinetic model. The $K_M$ and $k_{cat}$ derived from the fit are given.
3.8. Structure of PfClpP(179-370)

We were able to crystallize and solve the X-ray structure of the PfClpP construct (Figure 3.9, Table 3.3). This was done in collaboration with Dr. Raymond Hui from the Structural Genomic Consortium, Toronto. The PfClpP construct has an N-terminal tag MGSSHHHHHHHSGRENLYFQGHM followed by the protein sequence from D179 – K370, H₆-PfClpP(179-370). The protein crystallized as a tetradecamer with one heptamer in the asymmetric unit. The high concentration in the crystal could have promoted double ring formation (Figure 3.9).

The ClpP protomer structure can be divided into three parts. The first part is the N-terminal axial loop, consisting of residues N-terminal to αA (Figure 3.9 B), which is required for the interaction of ClpP with its cognate chaperones (Gribun et al., 2005; Bewley et al., 2006; Szyk and Maurizi 2006). In EcClpP, these residues start within the heptameric ring complex then rise above the surface of the ring and, subsequently, trace back towards the ring to connect with αA (Figure 1.2). While such N-terminal loop arrangement is not completely observed in PfClpP, resolved residues N-terminal to αA seem to follow a similar trajectory observed for EcClpP axial loops (Figure 3.9 B).

The second part of ClpP is the head domain, consisting of most of the protease sequence except for β6 and αE. The head domain of PfClpP superimposes well with that of EcClpP with minimal perturbations (rmsd 1.15 Å) consistent with the high conservation in the protein sequence.

The third part of ClpP is the handle region, which is formed by β6 and αE (Figure 3.9 A). Handle regions from two ClpP single rings interdigitate in the double ring structure. Residues F296 – Q305 are unstructured in the handle region of H₆-PfClpP(179-
370), although the exact boundaries vary between protomers. This seems to be a consequence of the two heptameric rings being closer to each other in PfClpP than in EcClpP resulting in a more compact tetradecamer (Figure 3.9 A). Indeed, the opposing apical surfaces are approximately 10 Å closer in H6-PfClpP(179-370) than in EcClpP, measured from the 2D projections of the proteins (Figure 3.9 A). Furthermore, if the subunits of one ring of EcClpP are superposed on the subunits of one ring of H6-PfClpP(179-370), then the subunits of the remaining rings are rotated with respect to each other by about 10° (Figure 3.9 C). Residues of the Ser264-His289-Asp338 catalytic triad in H6-PfClpP(179-370) have multiple configurations, especially His289, when compared to the respective residues of EcClpP, which form a hydrogen bonded network (Figure 3.9 D). This observation in addition to the disordered β6 guide strand, suggests that the PfClpP structure corresponds to a catalytically inactive state of the protease.
Figure 3.9. The X-ray structure of H₆-PfClpP(179-370). (A) Space-filling models of *P. falciparum* (left) and *E. coli* (right, 1yg6) ClpP tetradecamers are shown. In each complex, two monomers from two different rings are colored. The dimensions were measured from the 2D projections of the proteins. (B) Superposition of *P. falciparum* (orange) and *E. coli* (blue) ClpP protomers. The first and last residues resolved in the X-ray structures are indicated for each protomer. The unstructured region in H₆-PfClpP(179-370), F296-Q305, is indicated by a dotted line. The box marks the location of the catalytic triad. (C) Stereoviews of the *P. falciparum* (orange) and *E. coli* (blue) ClpP heptameric rings viewed from top when the subunits of the bottom heptameric rings (not shown) are superposed. The subunits in the upper rings are rotated with respect to each other by about 10°. (D) Overlay of the seven Ser-His-Asp catalytic triads from the protomers within one heptameric ring of *P. falciparum* or *E. coli* ClpP. The structures were generated using PyMOL.
<table>
<thead>
<tr>
<th><strong>Data Collection</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>C222₁</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>158.3</td>
</tr>
<tr>
<td>b (Å)</td>
<td>196.5</td>
</tr>
<tr>
<td>c (Å)</td>
<td>139.2</td>
</tr>
<tr>
<td>α (°)</td>
<td>90.0</td>
</tr>
<tr>
<td>β (°)</td>
<td>90.0</td>
</tr>
<tr>
<td>γ (°)</td>
<td>90.0</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>1.10</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>50-2.45</td>
</tr>
<tr>
<td><strong>Measured reflections</strong></td>
<td>1525816</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>78895</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (last shell)</td>
<td>0.075 (0.316)</td>
</tr>
<tr>
<td>&lt;I&gt;/σ(I)&gt; (last shell)</td>
<td>15.0 (4.7)</td>
</tr>
<tr>
<td>Completeness (last shell)</td>
<td>0.992 (0.985)</td>
</tr>
<tr>
<td>Redundancy (last shell)</td>
<td>7.5 (4.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Structure Refinement</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>30-2.45</td>
</tr>
<tr>
<td><strong>Reflections, Working set</strong></td>
<td>72825</td>
</tr>
<tr>
<td><strong>Reflections, Test set</strong></td>
<td>3790</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; (R&lt;sub&gt;free&lt;/sub&gt;)</td>
<td>0.210 (0.238)</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>9941</td>
</tr>
<tr>
<td>Number of water atoms</td>
<td>248</td>
</tr>
<tr>
<td>Mean B factor (Å²)</td>
<td>50.6</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>95.2</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.4</td>
</tr>
<tr>
<td>r.m.s.d. bond lengths (Å)</td>
<td>0.005</td>
</tr>
<tr>
<td>r.m.s.d. bond angles (°)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Values in parentheses are for the last shell.

<sup>a</sup>R<sub>sym</sub> = Σ[I| – <I>|]/ΣI, where I is the observed intensity of a measured reflection and <I>| is the mean intensity of that reflection.

<sup>c</sup>The last shell includes all reflections between 2.45 and 2.54 Å.

<sup>d</sup>R<sub>work</sub> = Σ |F<sub>o</sub> – F<sub>c</sub>/ Σ F<sub>o</sub>.

<sup>d</sup>R<sub>free</sub> is the cross-validation R factor computed for a test set of reflections (5% of total).
3.9. Modeling of the ClpR Monomer

The existence of a ClpR, an unique isoform of ClpP which lacks the catalytic residues necessary for proteolytic activity, has always been an interesting question. The purpose of incorporating these inactive subunits into the proteolytic core forming a ClpP/ClpR heterocomplex is puzzling. The structure of ClpP has been solved in numerous organisms, and it is becoming important to identify the structure of ClpR as well. We have attempted to solve the crystal structure of ClpR in collaboration with Dr. Raymond Hui from the Structural Genomic Consortium in Toronto, Canada. However, the crystals we have obtained so far did not yield significant data for diffraction.

In the absence of experimental data, we decided to take advantage of the vast number of ClpP structures available as basis for homology modeling of ClpR to gain structural information. The three-dimensional crystal structure of *Mycobacterium tuberculosis* ClpP1 (2ce3 – 2.6Å) was chosen as template for homology modeling of ClpR (Figure 3.10). Both proteins share a sequence similarity of 28.4% (Figure 3.10 A). Other homologous templates with structures available in PDB have been found using the psi-BLAST analysis in the NCBI database; however, none of them gave suitable alignments with ClpR for homology modeling.

A portion is truncated on the *Mt*ClpP1 structure from Pro-125 to Ile-136, which is an unstructured region N-terminal to the E-helix. Evaluation of the structural correctness of the model did not show any residues in the disallowed region of the Ramachandran plot. All spatial restraints were satisfied with violations of distance and dihedral angle restraints from the template structure (2ce3) minimized. The ClpR model revealed a high degree of structural conservation with ClpP, except for a 9- to 10-residue insertion
(Figure 3.10 B, C). This insertion is seen in a majority of ClpRs. Based on the model, the insertion can be seen as protruding into the inner chamber of a putative ClpRP core complex. When the ClpR model is aligned with the crystal structure of an *E. coli* ClpP monomer with a substrate bound (2FZS) (Figure 3.10 B), the insertion seems to be a lid covering the active sites, which would potentially prevent access of substrates to the catalytic domain. Along with the lack of catalytic residues in ClpR, this suggests that ClpR may play a role as a regulator of proteolysis in a putative ClpRP core complex.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClpP_Ec</td>
<td>GERSFDIYSRLLKERVFILTXQVMDHANLIVAQLMFLAENPEKDIYLYINSFGV--- 83</td>
<td>113</td>
</tr>
<tr>
<td>ClpP_Mt</td>
<td>LSLTDLGVEKLLERIIIFLGEVDMVIEANRLCAQLLLAAEADASKDCIILYINSFGS--- 70</td>
<td>102</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>QMRRKQVKEFPRKIIYLTDEIMRRTDEQILYYLNLNHRDIKIVYINSFGS--- 236</td>
<td>250</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>KRRNYNIFQHKLLERIIIFLGMIDPIYQEGQVQILYLYLEVESKLPEKVIYINSQK--- 113</td>
<td>126</td>
</tr>
<tr>
<td>ClpP_Ec</td>
<td>-------ITGMSIVDQMQIYKFDVTTICMQQAALSAAFLTTAQGKVARFCLPSFVMH--- 136</td>
<td>150</td>
</tr>
<tr>
<td>ClpP_Mt</td>
<td>-------ISAMQAIYMTVDAPCDIAYAFIAAMAGKMEFLAAAGQGKRYALPHARILL--- 123</td>
<td>137</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>-------INIGALIQIFYKQQDIQTSICFLVAGMAYILASSGKGYKQKLPNCIRMK--- 289</td>
<td>300</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>KLIHANGYTVVIVVIVYVNTLESQVTYTCGLKQNYAGCAILASQGKRYLFQKLNSIFL--- 173</td>
<td>187</td>
</tr>
<tr>
<td>ClpP_Ec</td>
<td>QQLRGVQC--QTDEIEHARELKLKGGWILMAFOTQKELQIQFDFYTOPEAEVE--- 195</td>
<td>210</td>
</tr>
<tr>
<td>ClpP_Mt</td>
<td>QQLGQVVS--SAADIAIQQPOFAQIKKMMFVRQEQPTQPLIEEADPIRDRWFTAAKE--- 182</td>
<td>200</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>QQLGQVVS--SAADIAIQQPOFAQIKKMMFVRQEQPTQPLIEEADPIRDRWFTAAK--- 348</td>
<td>360</td>
</tr>
<tr>
<td>ClpP_PF</td>
<td>QQLGQVVS--SAADIAIQQPOFAQIKKMMFVRQEQPTQPLIEEADPIRDRWFTAAK--- 233</td>
<td>244</td>
</tr>
</tbody>
</table>

**Diagram**

- **A**
- **B**
- **C**
Figure 3.10. Modeling of the ClpR subunit. (A) Sequence alignment of *E.coli* ClpP, *Mycobacterium tuberculosis* ClpP1, with *Plasmodium falciparum* ClpP and ClpR was computed using ClustalW with default annotations. The modeling of ClpR is performed using the crystal structure of *Mt*ClpP1 (2ce3) as template. The N-terminal sequences are removed. The catalytic residues in ClpPs are colored in red, and they are missing in ClpR, except for the last residue – Asp. The unique insertion in ClpR is underlined. (B) The homology model of ClpR (cyan) is superimposed onto the crystal structure of *E.coli* ClpP (red) with tripeptide substrate (yellow) bound to its active site (2FZS). (C) The homology model of ClpR superimposed onto its template structure, *Mt*ClpP1 (blue).
Chapter 4. DISCUSSION AND FUTURE DIRECTIONS
4. Discussion and Future Directions

In this work, we have provided a comprehensive overview of the Clp system in *P. falciparum*. We have identified the Clp proteolytic components, *Pf*ClpP and *Pf*ClpR, as well as, the Clp chaperone components, *Pf*ClpB1, *Pf*ClpB2, *Pf*ClpC and *Pf*ClpM, in *P. falciparum*. All of these proteins are expressed in blood stage parasites, and are localized within the apicoplast, with the exception of *Pf*ClpB2, which is localized within the parasitophorous vacuole. Based on our bioinformatics analysis (Figures 3.1 and 3.2) and localization studies (Figure 3.3), we suggest the presence of a chaperone-protease *Pf*ClpCRP complex that is expected to reside within the apicoplast of the parasite.

Immunofluorescence using antibodies identified the location of *Pf*ClpP within a non-nucleus, non-vacuole organelle within the vicinity of the apicoplast, thus we believed that organelle is in fact the apicoplast. Even though we did not obtain a signal for *Pf*ClpR in our immunofluorescence studies, online algorithms such as PATS and PlasmAP have successfully predicted the localization of *Pf*ClpR within the apicoplast. Furthermore, our analysis using Surface Plasmon Resonance suggested an interaction between *Pf*ClpP and *Pf*ClpR. We chose *Pf*ClpC as our candidate cognate Clp ATPase partner to the *Plasmodium* Clp protease because it is the only Clp ATPase we identified in the parasite to have a putative ClpP-binding loop necessary for interaction with ClpP. In addition, co-localization immunofluorescence studies using stromal apicoplast marker acyl carrier protein (ACP) demonstrated that *Pf*ClpC resides within the same apicoplast as *Pf*ClpP and *Pf*ClpR.

Our localization studies contradicted with a previous report from Lin *et al.* which concluded that *Pf*ClpP is localized to the nucleus. Lin *et al.* employed expression of the
parasite gene in human cells for their localization studies, whereas we used antisera to the native protein in live parasite cells. Furthermore, a study on PfClpP by Rathore et al. also demonstrated that the protease is localized to the apicoplast which is consistent with our findings. Rathore et al. demonstrated that purified PfClpP forms the same heptameric complex as we did, and that it possesses very low peptidase activity with $K_M$ close to our results using the Suc-LLVY-AMC peptide (Figure 3.8). We believe the high lysine content of PfClpP (Singer and Hickey, 2000) likely lead to a false nuclear localization signal being read in mammalian cells.

Based on our biochemical analyses, the molecular weights that we detected for both PfClpP and PfClpR are much smaller than the predicted molecular weight (Figure 3.5 and 3.6), suggesting that a significant number of residues of both proteins are removed upon translocation into the apicoplast. Approximately 149 amino acids are removed from the N-terminus of PfClpP, while 48 amino acids are removed from the N-terminus of PfClpR. This is in stark contrast with EcClpP in which the pro-sequence that is auto-catalytically cleaved is only approximately 13 residues (Figure 3.1). Aside from the fact that the N-terminal leader sequences are required for proper translocation into the 4 membrane thick apicoplast, it remains uncertain as to why such a large pre/pro-sequence is required for the P. falciparum proteases. We believe it is possible that perhaps a large pre/pro-sequence might actively prevent premature folding and assembly of the protease prior to its translocation into the apicoplast through the TiC and ToC complexes.

In our X-ray crystallography studies, the PfClpP tetradecamer revealed a relatively compacted structure compared to wild type EcClpP, with the catalytic triads in
an inactive conformation. Such inactive compact ClpP structures have been observed for
*Streptococcus pneumonia* (Gribun *et al.* 2005) and *Mycobacterium tuberculosis* (Ingvarsson *et al.* 2007) ClpPs. The complementary active extended ClpP structure have thus far been observed for WT *Ec*ClpP (Wang *et al.* 1997; Bewley *et al.* 2006; Szyk and Maurizi 2006), *H. sapiens* mitochondria ClpP (Kang *et al.* 2004), *H. pylori* ClpP (Kim and Kim, 2008), and *B. subtilis* ClpP (Lee *et al.* 2010). Based on experimental and theoretical data, our group has recently argued that the compact structures of *S. pneumoniae*, *M. tuberculosis*, and *P. falciparum* ClpPs represent a naturally sampled compact state of the ClpP cylinder (Kimber *et al.* 2010). Together with our earlier studies which suggested plastic ring-ring interactions between ClpP double rings, we have proposed a model whereby the *Pf*ClpP cylinder switches dynamically between an active extended state required for substrate degradation and an inactive compact state allowing peptide product release.

In mature *Ec*ClpP, the 16 N-terminal residues form the axial loops that are situated at the entrance of the axial pore on each end of the ClpP tetradecamer. The axial loops are rather flexible but can form β-hairpins that confer a narrow entrance pore of approximately 10-12 Å diameter to the inner core of the proteolytic cylinder. It have been previously discovered that these axial loops control the access of substrates to the degradation chamber and also mediate the communication with the bound chaperone (Gribun *et al.* 2005; Sprangers *et al.* 2005; Bewley *et al.* 2006). In the case of mature *Pf*ClpP the N-terminal region of αA is made of 36 residues instead (Figure 3.1). In our X-ray crystal structure of *Pf*ClpP (Figure 3.9), only a few residues of the N-terminal region proximal to αA are visible but they follow a trajectory similar to the one observed for
EcClpP. We believe the N-terminal region of PfClpP may adopt a similar conformation to EcClpP, at least for the first few residues proximal to αA. Yet, the conformation and role of the additional residues at the N-terminal region of PfClpP remains enigmatic. It is possible that these residues adopt the β-hairpin conformation and trace into the digestion chamber. Another possibility is that these extended N-terminal residues might fold back and protrude on the apical surfaces of PfClpP. In any case, these additional residues might have important roles in the translocation of substrate proteins or in the transmission of allosteric signals between PfClpP and PfClpC.

Given our hypothesis that the chaperone PfClpC associates with the protease PfClpPR, we hope to reproduce the PfClpCPR complex in vitro in the future. The crystal structure of PfClpC, PfClpR as well as the overall PfClpCPR complex will need to be solved. A member of our lab is currently solving the crystal structure of PfClpR as we speak. Once the PfClpCPR complex can be isolated, we hope to assess its catalytic activity as well as its potential substrates. It will be interesting to observe the role of the inactive ClpR subunits in the overall ClpPR protease. Continual collaboration with Dr. McFadden’s group can be done to perform further in vivo analyses on the PfClpCPR complex within the parasite cells.

In summary, we have provided the first comprehensive overview of the Clp chaperones and proteases of *P. falciparum*. We have identified the location of the Clp proteins within the apicoplast, an indispensable organelle of prokaryotic origin believed to be the parasite’s Achilles’ heel. We have highlighted a possible important role of Clp proteins in maintaining protein homeostasis in this cellular compartment. With the increase in malarial drug resistance, we believe our study has shed light towards a novel
approach in malarial management though disregulation of homeostasis by developing drugs that target the Clp chaperones and proteases.
CHAPTER 5. REFERENCES
5. References


Kimber, M. S., et al. (2010). Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. Structure, 18, 798–808.