The Clp Chaperones and Proteases of the Human Malaria Parasite *Plasmodium falciparum*

Majida El Bakkouri1†, Andre Pow1†, Anne Mulichak2†, Kevin L. Y. Cheung3†, Jennifer D. Artz4, Mehrnaz Amani4, Stuart Felli5, Tania F. de Koning-Ward6, C. Dean Goodman5, Geoffrey I. McFadden5, Joaquin Ortega3, Raymond Hui4 and Walid A. Houry1⁎

1Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8
2IMCA-CAT, Argonne National Lab, Argonne, IL 60439, USA
3Department of Biochemistry and Biomedical Sciences and MG DeGroote Institute for Infectious Diseases Research, McMaster University, Hamilton, Ontario, Canada L8N 3Z5
4Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada M5G 1L5
5School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia
6School of Medicine, Deakin University, Piddons Rd, Waurn Ponds, Victoria 3217, Australia

Received 15 July 2010; received in revised form 8 September 2010; accepted 22 September 2010
Available online 29 September 2010

Edited by R. Huber

Keywords:
*Plasmodium falciparum*; apicoplast; PfClp ATPases; PfClp proteases; protein homeostasis

The Clp chaperones and proteases play an important role in protein homeostasis in the cell. They are highly conserved across prokaryotes and found also in the mitochondria of eukaryotes and the chloroplasts of plants. They function mainly in the disaggregation, unfolding and degradation of native as well as misfolded proteins. Here, we provide a comprehensive analysis of the Clp chaperones and proteases in the human malaria parasite *Plasmodium falciparum*. The parasite contains four Clp ATPases, which we term PfClpB1, PfClpB2, PfClpC and PfClpM. One PfClpP, the proteolytic subunit, and one PfClpR, which is an inactive version of the protease, were also identified. Expression of all Clp chaperones and proteases was confirmed in blood-stage parasites. The proteins were localized to the apicoplast, a non-photosynthetic organelle that accommodates several important metabolic pathways in *P. falciparum*, with the exception of PfClpB2 (also known as Hsp101), which was found in the parasitophorous vacuole. Both PfClpP and PfClpR form mostly homoheptameric rings as observed by size-exclusion chromatography, analytical ultracentrifugation and electron microscopy. The X-ray structure of PfClpP showed the protein as a compacted tetradecamer similar to that observed for *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* ClpPs. Our data suggest the presence of a ClpCRP complex in the apicoplast of *P. falciparum*.

© 2010 Elsevier Ltd. All rights reserved.

*Corresponding author. E-mail address: walid.houry@utoronto.ca.
† M.E.B., A.P., A.M. and K.L.Y.C. contributed equally to this work.

Abbreviations used: RBC, red blood cell; SEC, size-exclusion chromatography; STEM, scanning transmission electron microscope; ABC, ATP-binding cassette; hmu, heme uptake system; DMSO, dimethyl sulfoxide; ITC, isothermal titration calorimetry; OD600, optical density at 600 nm; OM, outer membrane; PBP, periplasmic binding protein; SBP, substrate-binding protein; SeMet, selenomethionine.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
Introduction

Protein homeostasis in the cell is regulated by a wide array of molecular systems consisting of molecular chaperones and proteases.1–4 These systems ensure that proteins fold to their native state and maintain their proper conformation throughout their lifetime in the cell. Surveillance systems also ensure that misfolded proteins and short-lived regulatory proteins are targeted for degradation. Many proteases that clear the cells of proteins are selective in their target substrates. One highly conserved class of proteases is the self-compartmentalizing proteases in which the proteolytic active sites are sequestered within a chamber inside the body of the proteolytic oligomer.5 Such proteases include the proteasome, ClpP (caseinolytic proteases) and HslV proteases.

The ClpPs are highly conserved across prokaryotes6 and are found in the mitochondria of eukaryotes and the plastids of plants and algae, which are derived from endosymbiotic bacteria.7 ClpP is a cylinder-shaped multi-subunit oligomeric complex that associates with chaperones that dock on either end of the ClpP cylinder. Although the majority of bacteria contain only one copy of ClpP, some such as cyanobacteria and actinobacteria have multiple isoforms of the protease.6 Interestingly, cyanobacteria, Arabidopsis thaliana and other eukaryotes also express a non-catalytic ClpP paralog that lacks residues of the Ser-His-Asp catalytic triad. This non-catalytic paralog is termed ClpR, and its role within the Clp protease complex remains enigmatic.

The chaperones that dock on either end of the ClpP cylinder target substrate proteins and then thread them through into the ClpP proteolytic chamber for degradation. These chaperones are members of the ATPases associated with diverse cellular activities (AAA+) superfamily.8–11 They are structurally diverse and are divided into two main classes.12 Members of class I, which includes ClpA, ClpB, ClpC, ClpD, ClpE and ClpL, contain two ATP nucleotide-binding domains (AAA+ domains) that have characteristic Walker A and Walker B nucleotide binding and recognition motifs.8,10,13–15 Most members of class I contain a middle region of varying length inserted within AAA+1. Proteins of class II are smaller and contain only one AAA+ domain, which is homologous to the second AAA+ domain of proteins of class I. Members of class II include ClpX and ClpY (HisU).

The basic organization of the Clp chaperone–protease complex is expected to be the same across all species. However, complexes with variable composition and subunit layout have been found in organisms containing several paralogs of ClpP and multiple ATPases. For instance, the model cyanobacterium, Synechococcus elongatus PCC7942, has three ClpP paralogs and one ClpR and possesses ClpX and ClpC ATPases. Biochemical and genetic studies of the Clp chaperone–protease complexes in this organism suggested the presence of a SeClpP1P2 protease complex that interacts with SeClpX, a SeClpP3R complex that interacts with SeClpC and a SeClpP1R complex that might interact with membranes.16,17 A dramatic example of complex diversity is found in higher plants, where the number of ClpP proteins is greatly increased.7,18,19 For example, A. thaliana has six ClpPs, four ClpRs, three ClpXs, two ClpCs, one ClpD and three ClpBs.19–22 While all of the clp genes are nuclear encoded except for clpP1, which is a plastid-encoded gene, all the Clp proteins probably reside in the chloroplast stroma; indeed, a 325–350 kDa Clp chaperone–protease complex has been identified in this compartment.20 Based on proteomic studies, the Clp chaperone–protease complex seems to consist of five ClpP paralogs (ClpP1 and ClpP3 – ClpP6), four non-catalytic ClpR subunits (ClpR1 – ClpR4), three Clp AAA+ chaperones (ClpC1, ClpC2 and ClpD) and several additional members of unknown function. The organization of the subunits in the complex is not known. Deletion or disruption of the clp genes in A. thaliana results in many severe phenotypes,19,23–25 highlighting the importance of the Clp system in plant growth and development and protein homeostasis.

Little is known of the Clp system in the parasitic protozoan genus Plasmodium, which is the causative agent of malaria. Plasmodium falciparum contains a plastid organelle called the apicoplast,26 which is derived from an ancient red algal endosymbiont.27 The apicoplast is a non-photosynthetic organelle that does, however, accommodate several important metabolic pathways28 and is essential for parasite survival.29 The prokaryote-like processes of the apicoplast present many potential drug targets that are absent from the human host and, thus, the apicoplast appears to be a promising target for new anti-malarial drugs. The identification in P. falciparum of a putative Clp ATPase gene in the apicoplast DNA,30 a nuclear-encoded ClpP-type protein31 and Clp ATPases32 that might be targeted to the apicoplast, led us to speculate that a Clp-related complex might play an important role in protein quality control in the apicoplast of P. falciparum. Hence, targeting the activity of this complex might be a promising approach for the development of anti-malarial drugs as was recently suggested.33 Dysregulation34 or selective inhibition35 of ClpP function has already proven to be a promising avenue for the development of antibiotics and such approaches could potentially be adapted to malaria.

Here, we describe the identification of four Clp ATPases, one ClpP and one ClpR in P. falciparum. Experiments are presented showing the localization of these proteins as well as the biochemical and structural characterization of ClpP and ClpR. Our data suggest the presence of a ClpCRP complex in the apicoplast of P. falciparum.
Results

Identification of putative clp genes in Plasmodium falciparum

Protein sequences from Escherichia coli K12 ClpP (accession no. P0A6G7), Synechococcus elongatus PCC7942 ClpC (BAD79443), and S. elongatus PCC7942 ClpX (AAI03913) were retrieved from the NCBI database and used in BLASTP queries of the P. falciparum nuclear, plastid and mitochondrial protein databases of PlasmoDB. Putative hits were further confirmed by reciprocal BLASTP against the UniProt protein database.

Six putative clp genes were identified in P. falciparum (Table 1). Two P. falciparum sequences (PFC0310c and PF14_0348) were found to have a high level of sequence homology with E. coli ClpP. Sequence data from PlasmoDB showed that both are nuclear-encoded proteins that are absent from the apicoplast genomes. PFC0310c is on chromosome 3 and PF14_0348 is on chromosome 14. Each sequence was compared against the COG database using the COGITOR program. Both sequences were identified as belonging to the COG0740, a COG group corresponding to Clp proteases. We subsequently aligned E.coliP with PFC0310c and PF14_0348 using ClustalW with default parameters (Fig. 1). From the alignment, we observed that PFC0310c possesses the three catalytic residues (Ser264, His289 and Asp338) and, therefore, we refer to it as ClpP (Table 1). PF14_0348 lacks the catalytic serine and histidine residues (Fig. 1) and, hence, seems to be an inactive paralog of ClpP that is found also in plants and cyanobacteria. PF14_0348 is referred to here as PfClpR (Table 1).

Four putative proteins displayed very significant sequence homology to S. elongatus ClpC. One putative clp ATPase gene (PFC10_API01060) is found on the plastid DNA, while three other genes (PF11_0175, PF14_0063 and PF08_0063) are found on the nuclear chromosomes. The proteins corresponding to these genes are referred to here as PfClpB1, PfClpB2, PfClpC and PfClpM, as shown in Table 1, for the reasons given below. No proteins were found to bear high sequence identity to S. elongatus ClpX. It should be noted that P. falciparum also has P/HslU (PFCpY, PFH0355c) and PfHslV (PFCpQ, PFH1465c) proteins, which are not discussed further here but have been described elsewhere.

All the PfClp ATPases have a predicted N-domain and two AAA+ domains (Fig. 2), D1 and D2, which comprise a large and a small subdomain. The conserved Walker A, Walker B and different sensor motifs can be identified easily in all the AAA+ domains of PfClp ATPases (Fig. 2), except for D1 of apicoplast-encoded PFC10_API01060. D1 of PFC10_API01060 does not have the canonical GKT or (I/V)DEI sequences of the Walker A and Walker B motifs (Supplementary Data Fig. 1A), respectively; hence, this domain seems to have degenerated and is not expected to be an active ATPase. Following the nomenclature for Clp ATPases recommended by Schirmer et al., we have named PFC10_API01060 as PfClpM (Table 1).

Unlike EcClpA, PfClp ATPases have an insertion of 18-174 residues at a site in the small subdomain of the D1 AAA+ module (Fig. 2) with PfClpM having the smallest insertion region. The size of this insertion was originally used as a basis for categorizing the Clp ATPases into different subfamilies, such as ClpB, ClpC and ClpD. Subsequently, the presence of certain motifs was used for such a classification. However, since the X-ray structure of Thermus thermophilus ClpB (TtClpB) is solved and shows that the insertion region forms a coiled coil, we classify Clp ATPases whose insertion region is predicted to form a coiled coil as belonging to the ClpB subfamily. Based on the programs COILS and the Paircoil, the insertion region of both TtClpB and the Paircoil, the insertion region of both PFC10_API01060 and PFC10_API01060 is solved and shows that the insertion region forms a coiled coil; hence, we name them PfClpB1 and PfClpB2, respectively. Close inspection of the insertion region based on sequence alignment of PfClpB1, PfClpB2, TtClpB, ScClpB 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 (Supplementary Data Fig. 1B). Note also that the insertion region of PfClpB1 is interrupted by a 53 amino acid residue segment (Supplementary Data Fig. 1B).

It is known that the interaction between the ATPase and proteolytic subunits of a Clp chaperone–protease complex is mediated partly by a surface loop in the Clp ATPase proteins termed the ClpP binding loop.\(^{45,46}\) This loop is located after the Walker B motif in the AAA+ module between the

---

**Plasmodium falciparum Clp Chaperones and Proteases**

![Fig. 1. Sequence properties of the *P. falciparum* Clp proteases. Sequence alignment of EcClpP, PfClpP and PfClpR using ClustalW\(^{29}\) and drawn with ESPript.\(^{91}\) 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).\(^{64}\) The axial loop present at the N-terminus of EcClpP and the handle region formed by \(\beta_6\) and \(\alpha E\) are indicated.

![Fig. 2. Domain arrangement of the *P. falciparum* Clp ATPases. Cartoon representation of the domain organization of the *P. falciparum* Clp ATPases and EcClpA shown for comparison. The residue numbers of the lysine in the Walker A motif GKT, the aspartic acid in the Walker B motif ΦΦDE, and the glycine in the ClpP binding loop motif [L/V]G[F/L] are indicated. CC refers to the presence of coiled coils as predicted by the programs COILS\(^{43}\) and Paircoil.\(^{44}\)
Sensor I and Box VII motifs and has a conserved tripeptide consensus sequence of [L/I/V]-G-[F/L].

Clp ATPases that lack the tripeptide motif are not known to form a complex with the ClpP protease. Based on multiple sequence alignment, only PfClpC is found to have a putative LGF ClpP binding loop in the correct position (Supplementary Data Fig. 1C). There is also an Asn-rich insertion of 94 amino acids in this putative ClpP-binding loop of PfClpC (Supplementary Data Fig. 1C). Hence, we propose that PfClpC associates with PfClpP and PfClpR to form the *P. falciparum* chaperone–protease complex that, as discussed below, is proposed to be localized to the apicoplast.

**Apicoplast targeting predictions**

The apicoplast is the relic plastid found in most apicomplexan parasites except *Cryptosporidium*. The apicoplast is different from typical plant plastids in that it has four bounding membranes thought to have resulted from two consecutive endosymbiotic events; thus, targeting to this organelle requires a different targeting mechanism.

There are two apicoplast targeting mechanisms: the transit peptide targeting and the Clp-protease targeting. The transit peptide targeting is characterized by a tripeptide consensus sequence of [L/I/V]-G-[F/L].

Based on multiple sequence alignment, only PfClpP, PfClpR, and PfClpM are known to form a complex with the ClpP protease. PfClpP, PfClpR, and PfClpM contain a conserved tripeptide consensus sequence of [L/I/V]-G-[F/L].

There are two apicoplast targeting mechanisms: the transit peptide targeting and the Clp-protease targeting. The transit peptide targeting is characterized by a tripeptide consensus sequence of [L/I/V]-G-[F/L].

Based on multiple sequence alignment, only PfClpP, PfClpR, and PfClpM are known to form a complex with the ClpP protease. PfClpP, PfClpR, and PfClpM contain a conserved tripeptide consensus sequence of [L/I/V]-G-[F/L].

**Experimental localization of the Clp proteins in *P. falciparum***

The detection of the expression and localization of the different *P. falciparum* Clp proteins was done by using either antibodies to these proteins or by tagging the respective parasite genes with a streptavidin-

**Figure 3.** Expression and localization of the PfClp proteins. (a) Expression of PfClpB1, PfClpC, PfClpB2, PfClpP, PfClpM, and PfClpR was checked in blood stage parasites. Lanes 1–3, Western blots of lysates from parasites having PfClpB1-3xHA, PfClpC-3xHA, and PfClpB2-3xHA probed with anti-HA antibodies. Lanes 4–6, lysates from 3D7 wild type parasites probed with anti-ClpM, anti-ClpP, and anti-ClpR peptide antisera. Molecular mass markers are indicated to the left of the lanes. (b) Localization of PfClpB1, PfClpC, PfClpB2, PfClpM, and PfClpP of *P. falciparum* was tested by immunofluorescent colocalization. (1) PfClpB1-3xHA and (2) PfClpC-3xHA parasites were probed with the apicoplast marker anti-ACP antisera (red), with anti-HA antisera (green), and with Hoechst 33342 to stain for DNA (blue). (3) PfClpB2-3xHA parasites were probed with the parasitophorous vacuole marker anti-PTEX150 antisera (red), with anti-HA antisera (green), and with Hoechst 33342 (blue). (4) and (5) Wild type 3D7 strain was probed with anti-ACP antisera (red), with anti-ClpM antisera or anti-ClpP antisera (green), and with Hoechst 33342 (blue). (1)–(5) The rightmost image represents the merged fluorescence images with the DIC or transmission image. The scale bars represent 5 μm.

The rightmost image represents the merged fluorescence images with the DIC or transmission image. The scale bars represent 5 μm.
Fig. 3 (legend on previous page)
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 $PfClpP$ or $PfClpR$, respectively (Fig. 3a). Immunofluorescence using the antibodies against $PfClpP$ in an early trophozoite stage parasite identified a small globular structure (Fig. 3b5) that is distinct from the nucleus and the food vacuole that we believe to be the apicoplast. We were unable to colocalize $PfClpP$ with ACP antibody because both sera are from the rabbit. The antibodies against $PfClpP$ gave no signal in immunofluorescence. Unfortunately, both $PfClpP$ and $PfClpR$ were refractory to 3′end tagging.

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 $PfClpB1$, $PfClpC$, $PfClpM$ and likely $PfClpP$ are localized to the apicoplast. Secreton of $PfClpB2$ to the parasitophorous vacuole, where it participates in export of virulence proteins to the red cell,53 is confirmed. At this stage, $PfClpR$ is predicted to be targeted to the apicoplast (Supplementary Data Table 1).

Sequence boundaries of the mature forms of $PfClpP$ and $PfClpR$

Since expressing and purifying the full-length Clp ATPases was not successful, we concentrated our efforts on the characterization of the ClpRP protease system. Existing prediction algorithms are unable to identify the precise cleavage site of the apicoplast targeting sequence. Thus, Western blot analysis was used to estimate the sequence boundaries of the mature $PfClpP$ and $PfClpR$ polypeptides. As described in Materials and Methods, anti-sera were raised against peptides:

ETKLPHPYFNKVEK

and

ADEAVDFKLIDHILEKE

which correspond to the C-terminal amino acid sequences of $PfClpP$ and $PfClpR$, respectively. Western blot analysis of ghost lysates of infected red blood cells (RBCs) revealed prominent bands of approximately 25 kDa and 22 kDa using anti-$PfClpP$ and anti-$PfClpR$ antibodies, respectively (Figs. 3a and 4). The polyclonal antibodies generated are against the C-terminal peptides of $PfClpP$ and $PfClpR$ so the mature proteins must be N-terminally truncated forms. To estimate the size of these processed proteins, we generated and purified a series of N-terminally truncated $PfClpP$ and $PfClpR$ constructs (see below) to be used as references in the Western blot analysis. It was found that truncated $PfClpP$ beginning at residue N150 and truncated $PfClpR$ beginning at residue S49 are closest in molecular mass to the observed processed $PfClpP$ and $PfClpR$ as found by SDS-PAGE (Fig. 4). Hence, we propose that $PfClpP(150-370)$ with a theoretical mass of 25.7 kDa and $PfClpR(49-244)$ with a theoretical mass of 22.5 kDa are likely to correspond or are very close to the mature proteins residing in the apicoplast.

Oligomeric state of N-terminally truncated $PfClpP$ and $PfClpR$

To investigate the oligomeric state of the predicted mature forms of the $PfClp$ protease subunits, $PfClpP$(150-370) and $PfClpR$(49-244) were expressed in E. coli and purified as described in Materials and Methods. Both $PfClpP$(150-370) and $PfClpR$(49-244) are expressed as soluble proteins. However, while $PfClpR$(49-244) was stable, $PfClpP$(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. Consequently, a series of N-terminally truncated PfClpP constructs was generated, starting at: S146, Y148, L152, Y154, D155, N161, Y166, V169, Q171, N177 and D179 (Fig. 1). Each of these constructs was expressed and purified. PfClpP(179-370) (theoretical mass 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 biochemical, biophysical and structural studies.

PfClpP(179-370) migrated at about 146 kDa and PfClpR(49-244) migrated at about 134 kDa when subjected to SEC (Fig. 5a). Both proteins seemed to form heptameric complexes as compared to tetradecameric EcClpP (theoretical mass of 303.8 kDa; Fig. 5).
Fig. 5a). Changing the concentration of salt between 150 mM and 1 M NaCl or changing the pH between 4 and 9 did not change the migration of PfClpP(179-370) or PfClpR(49-244).

Analytical ultracentrifugation sedimentation equilibrium analysis at 4 °C (Fig. 5b) was used to further establish the oligomeric state of PfClpP(179-370) and PfClpR(49-244). The plots of ln(A) versus r² (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. 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) (Fig. 5b). 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. 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 (see below). The slightly lower

![Fig. 6. Electron microscope 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 present 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 7-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 (Mave) 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-hand panels). The calculated average for this type of particles is shown in the right-hand panel.](image-url)
ratio observed for PfClpR(49-244) might be due to the presence of monomeric or lower order oligomers.

Next, we used electron microscopy to observe the oligomers formed by PfClpP and PfClpR. When PfClpP(179-370) was observed under negative staining conditions, the electron micrographs showed abundant ring-shaped particles of ~100 Å in diameter (Fig. 6a). Particle images selected from the electron micrographs were analyzed with algorithms that detect rotational symmetry and a strong 7-fold symmetry component was found (Table 2). Averaging these images showed a top view of a seven-membered ring (Fig. 6b) similar to those observed for EcClpP. In order to discriminate whether these top-view projections of PfClpP(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 (Fig. 6c). The histogram showed 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. These results are consistent with those obtained by SEC and analytical ultracentrifugation sedimentation equilibrium experiments (Fig. 5) and all these together indicate that PfClpP (179-370) assembles predominantly into a single heptameric ring. Interestingly, this is unlike other ClpPs, such as that of E. coli or Helicobacter pylori that readily form tetrads, but 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 PfClpR(49-244) and PfClpR(43-244), two constructs of PfClpR. In both cases, negatively stained electron micrographs showed ring-shaped particles (Fig. 6a; data not shown) that showed a strong 7-fold symmetry component when analyzed for rotational symmetry (Table 2). The top-view averages calculated for these two constructs were remarkably similar to those obtained for PfClpP(179-370) (Fig. 6b). However, in addition to the ring-shaped particles observed in the electron micrographs for PfClpR, we found particles made of two parallel striations (Fig. 6a black arrow in lower panel, and Fig. 6d), albeit at low abundance. The particles most likely represent side views of two stacked PfClpR heptameric rings (Fig. 6d). These results, together with our earlier SEC and analytical centrifugation experiments, suggest that PfClpR(49-244) and PfClpR(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 PfClpR(43-244) construct to estimate the proportion of single and double heptameric particles in solution. Interestingly, the histogram obtained (Fig. 6c) showed that the mass of the PfClpR(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. In addition, some particles were observed with a mass corresponding to a tetradecamer. These results suggest that PfClpR constructs form a heterogeneous oligomeric mixture containing heptamers and tetradecamers and other oligomeric forms of variable mass.

The interaction between PfClpP(179-370) and PfClpR(49-244)

Several experiments were done to determine whether PfClpP(179-370) and PfClpR(49-244) are able to form the typical tetradecameric Clp protease complex. Because PfClpP(179-370) and PfClpR(49-244) can readily form heptamers, we speculated that a PfClpRP complex would be formed by a PfClpP homohexameric ring interacting with a PfClpR homohexameric ring. Pre-incubation of PfClpP(179-370) with PfClpR(49-244) and analyzing the mixture by SEC did not show the presence of a tetradecameric complex (chromatogram in Fig. 5a). Analysis of the mixture by analytical ultracentrifugation sedimentation equilibrium gave similar results; only a heptameric complex was found. To test the possibility that PfClpP(179-370) and PfClpR(49-244) might form heteroheptamers, the two proteins were co-expressed in E. coli from a polycistronic pST39 plasmid with one of the proteins being tagged. The tagged protein was pulled down, however, there was no evidence of significant interaction between PfClpP(179-370) and PfClpR(49-244). The isolated proteins formed heptamers, as detected by SEC, indicating that the complexes formed are likely homohexamers rather than heteroheptamers (not shown).

Subsequently, surface plasmon resonance experiments using the BIAcore system were used to determine direct binding between the two proteins. The PfClpP(179-370) or PfClpR(49-244) heptamer was immobilized on the sensorchip and sensograms were

Table 2. Analysis of the rotational symmetry of PfClpP and PfClpR

<table>
<thead>
<tr>
<th>Sample</th>
<th>Particles analyzed</th>
<th>Symmetry detected</th>
<th>t (t-test)</th>
<th>Spectral ratio product</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfClpP(179-370)</td>
<td>174</td>
<td>7</td>
<td>&lt;0.0000001</td>
<td>4.28 × 10^{10}</td>
</tr>
<tr>
<td>PfClpR(49-244)</td>
<td>150</td>
<td>7</td>
<td>&lt;0.0000001</td>
<td>6.68 × 10^{26}</td>
</tr>
<tr>
<td>PfClpR(43-244)</td>
<td>980</td>
<td>7</td>
<td>&lt;0.0000001</td>
<td>4.85 × 10^{26}</td>
</tr>
</tbody>
</table>
Fig. 7. Surface plasmon resonance analysis of the interaction between PfClpP(179-370) and PfClpR(49-244). 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) are shown. The $K_d$ values given refer to the apparent dissociation constants between the different protein heptamers. The numbers in parentheses are the standard deviations on the $K_d$ values.
Fig. 8. Peptidase activity of PfClpP(179-370). (a) Peptidase activity of 20 µM PfClpP(179-370), 20 µM PfClpR(49-244) or 20 µM PfClpP(179-370) + 20 µM PfClpR(49-244) were measured at 30 °C by their ability to cleave 50 µM Suc-LLVY-AMC, Suc-LY-AMC, Suc-IIW-AMC, Suc-IAMC, Suc-AAPF-AMC or Suc-AFK-AMC. The peptidase activity of EcClpP using Suc-LY-AMC as a substrate is shown as a reference. The inset in the lower left panel shows the inhibition of PfClpP(179-370) peptidase activity against Suc-LLVY-AMC by PMSF and chymostatin measured by the fluorescence intensity change after incubation for 6 h. (b) The change in the initial rate of Suc-LLVY-AMC hydrolysis by 20 µM PfClpP(179-370) as a function of peptide concentration is shown. The continuous line is the fit to the data using the Michaelis–Menten kinetic model. The values of $K_M$ and $k_{cat}$ derived from the fit are given.
recorded by injecting the other protein (Fig. 7); the apparent dissociation constants were then derived as described in Materials and Methods. The binding constants obtained were in the range 0.1–10 μM (Fig. 7). The strength of the binding interactions was: \( P/Pc \text{ClpP}(49-244) > P/Pc \text{ClpP}(49-244) > P/Pc \text{ClpP}(179-370) > P/Pc \text{ClpP}(179-370) > P/Pc \text{ClpP}(179-370) > P/Pc \text{ClpP}(179-370). \) This is consistent with the observation of double ring complexes by electron microscopy for \( P/Pc \text{ClpP}(49-244) \) and \( P/Pc \text{ClpP}(43-244) \) (Fig. 6a, c and d). Hence, if \( P/Pc \text{ClpP} \) and \( P/Pc \text{ClpR} \) are at equimolar concentrations in the apicoplast, then the major tetradsamic complexes observed would be that of the \( P/Pc \text{ClpR} \) homo-oligomer and a \( P/Pc \text{ClpP-P} \) \( P/Pc \text{ClpR} \) double ring formed from a heptameric \( P/Pc \text{ClpP} \) and heptameric \( P/Pc \text{ClpR} \). However, as observed with human ClpP,\(^{58}\) it is reasonable to speculate that the cognate ATPase, which we propose, on the basis of the bioinformatic analysis presented above, is \( P/Pc \text{ClpC} \) (Fig. 2), would drive tetradsamic formation and might dictate whether the complex formed is a \( P/Pc \text{ClpP}_7-P/Pc \text{ClpR}_7 \), \( P/Pc \text{ClpP}_{14} \) or \( P/Pc \text{ClpR}_{14} \) oligomer.

**Peptidase activity of \( P/Pc \text{ClpP}(179-370) \)**

The peptidase activity of \( P/Pc \text{ClpP}(179-370) \), \( P/Pc \text{ClpP}(49-244) \) and a mixture of \( P/Pc \text{ClpP}(179-370) \) with \( P/Pc \text{ClpP}(49-244) \) was assessed using model peptides such as the Suc-LY-AMC typically used for measuring EcClpP peptidase,\(^{60}\) as well as Suc-AFK-AMC, Suc-AAPF-AMC, Suc-IIW-AMC and Suc-LLVY-AMC. The peptidase activity of \( P/Pc \text{ClpP}(179-370) \) against Suc-LLVY-AMC and Suc-LY-AMC was used as a reference. No activity was observed for \( P/Pc \text{ClpP}(49-244) \), while very weak activity was detected for \( P/Pc \text{ClpP}(179-370) \) (Fig. 8a, inset in lower left panel). The activity of \( P/Pc \text{ClpP}(179-370) \) was not significantly affected by the presence of \( P/Pc \text{ClpR}(49-244) \) (Fig. 8a). The \( P/Pc \text{ClpP} \) proteases were unable to degrade longer peptides such as:

- MCA-YEVHHDQLVFK(DNP)-NH\(_2\)

and

- MCA-SEVNLDAEFR-Ednp-KRR-NH\(_2\)2TFA

or casein, which was reported to be degraded by EcClpP in the absence of the ATPase chaperone.\(^{51,62}\) \( K_M \) and \( k_{cat} \) for Suc-LLVY-AMC hydrolysis by \( P/Pc \text{ClpP}(179-370) \) were 65 μM and 0.13 h\(^{-1} \) (expressed as mol peptide bonds cleaved/mol ClpP protomers), respectively (Fig. 8b). As a comparison, \( K_M \) and \( k_{cat} \) reported for EcClpP against Suc-LY-AMC\(^{60}\) are 1 mM and 150 min\(^{-1} \), respectively. Hence, \( P/Pc \text{ClpP}(179-370) \) is a very weak peptidase on its own.

**Structure of \( P/Pc \text{ClpP}(179-370) \)**

We were able to crystallize and solve the X-ray structure of the \( P/Pc \text{ClpP} \) construct that had the N-terminal tag:

\[
\text{MGSSHHHHHSGRENLYFQGHM}
\]

followed by the protein sequence from D179–K370, H\(_{179-370} \)PClpP(179-370) (Table 3; Fig. 9a). The protein crystallized as a tetradsamic with one heptamer in the asymmetric unit. The high concentration in the crystal could have promoted double ring formation.

The \( P/Pc \text{ClpP} \) protomer structure can be divided into three parts (Fig. 1). The first part is the N-terminal

<table>
<thead>
<tr>
<th>Table 3. Data collection and structure refinement statistics for H(_{179-370} )PClpP(179-370)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Data collection</strong></td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell parameters</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Measured reflections</td>
</tr>
<tr>
<td>Unique reflections</td>
</tr>
<tr>
<td>( R_{sym}(\text{last shell}) )</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td><strong>B. Structure refinement</strong></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>Reflections, working set</td>
</tr>
<tr>
<td>Reflections, test set</td>
</tr>
<tr>
<td>( R_{work})</td>
</tr>
<tr>
<td>( R_{free})</td>
</tr>
<tr>
<td>Protein atoms</td>
</tr>
<tr>
<td>Water atoms</td>
</tr>
<tr>
<td>Mean B-factor (Å(^2))</td>
</tr>
<tr>
<td>Ramachandran</td>
</tr>
<tr>
<td>Favored (%)</td>
</tr>
<tr>
<td>outliers (%)</td>
</tr>
<tr>
<td>r.m.s.d. from ideal</td>
</tr>
</tbody>
</table>

Values in parentheses are for the last shell.

- \( R_{sym} = \left| \Sigma I(I-\langle I \rangle) / \Sigma I \right| \), where \( I \) is the observed intensity of a measured reflection and \( \langle I \rangle \) is the mean intensity of that reflection.
- \( R_{work} \) is \( \left| \Sigma I_{true} - I_{calc} / \Sigma I_{true} \right| \), where \( I_{true} \) is the observed intensity and \( I_{calc} \) is the calculated intensity.
- \( R_{free} \) is the cross-validation R-factor computed for a test set of reflections (5% of total).
Fig. 9 (legend on next page)
axial loop consisting of residues N-terminal to \( \alpha A \), which is required for the interaction of ClpP with its cognate chaperones.\(^{63-65}\) In \( \text{EcClpP} \), these residues start within the heptameric ring complex then rise above the surface of the ring and, subsequently, trace backwards towards the ring to connect with \( \alpha A \) (Fig. 9b). While this complete N-terminal loop movement is not observed in \( \text{PfClpP} \), resolved residues N-terminal to \( \alpha A \) appear to follow a trajectory similar to that observed for \( \text{EcClpP} \) axial loops (Fig. 9b).

The second part of ClpP is the head domain, consisting of most of the protease sequence except for \( \beta 6 \) and \( \alpha E \). The head domain of \( \text{PfClpP} \) superimposes well with that of \( \text{EcClpP} \) with minimal perturbations (rmsd 1.15 \( \AA \)) consistent with the high level of conservation in the protein sequence.

The third part of ClpP is the handle region, which is formed by \( \beta 6 \) and \( \alpha E \) (Fig. 1). Handle regions from two \( \text{ClpP} \) single rings interdigitate in the double ring structure. Residues F296–Q305 are unstructured in the handle region of \( \text{H}_4\text{PfClpP}(179-370) \), although the exact boundaries vary between protomers (Fig. 9a and b). This appears to be a consequence of the two heptameric rings being close to each other in \( \text{PfClpP} \) than in \( \text{EcClpP} \) resulting in a more compact tetradecamer. Indeed, the opposing apical surfaces are ~10 \( \AA \) closer in \( \text{H}_4\text{PfClpP}(179-370) \) than in \( \text{EcClpP} \), as measured from the 2D projections of the proteins. Furthermore, if the subunits of one ring of \( \text{EcClpP} \) are superposed on the subunits of one ring of \( \text{H}_4\text{PfClpP}(179-370) \), the subunits of the remaining rings are rotated with respect to each other by ~10° (Fig. 9c). Residues of the Ser264-His289-Asp338 catalytic triad in \( \text{H}_4\text{PfClpP}(179-370) \) have multiple configurations, especially His289, when compared to the respective residues of \( \text{EcClpP} \), which form a hydrogen bonded network. This observation and the disordered \( \beta 6 \) guide strand suggest that the \( \text{PfClpP} \) structure corresponds to a catalytically inactive state of the protease.

**Discussion**

We have provided a comprehensive overview of the Clp system in \( \text{P. falciparum} \). Our bioinformatic analysis (Figs. 1 and 2; Supplementary Data Fig. 1 and Table 1) and localization studies (Fig. 3) suggest the presence of a chaperone–protease \( \text{PfClpCRP} \) complex in the apicoplast of the parasite. The apicoplast also harbors a nuclear-encoded (\( \text{PClpB1} \)) and an apicoplast-encoded (\( \text{PClpM} \)) Clp ATPases.

An earlier report concluded that \( \text{PfClpP} \) is localized to the nucleus based on the expression of the parasite gene in human cells,\(^{66}\) but our localization of \( \text{PfClpP} \) to the apicoplast with antisera to the native protein contradicts the proxy localization in a mammalian surrogate system. The high lysine content of \( \text{PfClpP} \) likely confounded the mammalian cells and led to a false nuclear localization signal being read. Indeed, more recently, during the preparation of this manuscript, a study of \( \text{PfClpP} \) was published that also demonstrated that the protease is localized to the apicoplast as we have found.\(^{63}\) Consistent with our findings, those authors showed that purified \( \text{PfClpP} \) forms a heptameric complex and has very low peptidase activity with \( K_M \) close to what we report here using the Suc-LLVY-AMC peptide (Fig. 8). The \( \text{PfClpP} \) gene was found to be maximally transcribed in late trophozoite and early schizont stage parasites and the inactivation of \( \text{PfClpP} \) using \( \beta \) lactone-based inhibitors, albeit at relatively high concentrations, disrupted apicoplast development.

The molecular mass detected for both \( \text{PfClpP} \) and \( \text{PfClpR} \) is much smaller than the predicted value (Fig. 4), indicating that a large part of both proteins is removed upon translocation into the apicoplast. About 149 amino acids are removed from the N-terminus of \( \text{PfClpP} \) and 48 amino acids are removed from the N-terminus of \( \text{PfClpR} \). In contrast, the \( \text{EcClpP} \) pro-sequence, which is autocatalytically cleaved, is only about 13 amino acids (Fig. 1). It is not clear at this stage why such a large pre/pro-sequence is required for the \( \text{P. falciparum} \) proteases but one possibility is that a large pre/pro-sequence might actively prevent the proteases from folding and assembling during translocation into the apicoplast. Another possibility is that the N-terminal residues, upon release from the rest of the polypeptide, might fold into an active stable protein with a specific function in the apicoplast.

Several adaptor proteins typically modulate the activity of the Clp chaperone–protease system by interacting directly with the chaperone. One such adaptor protein is ClpS. In \( \text{E. coli} \), \( \text{EcClpS} \) has been found to recognize substrates that are degraded following the N-end rule pathway.\(^{68,69}\) In \( \text{S. elongatus} \),...
A ClpS ortholog also exists in *P. falciparum* (gene ID MAL13P1.111) and it is predicted to be localized to the apicoplast. Hence, this suggests that the N-end rule pathway plays an important role in regulating protein degradation in the apicoplast of the parasite.

The crystal structure of the *Pf*ClpP tetr decamer revealed a compact structure relative to that observed for wild type *Ec*ClpP, with the catalytic triads in an inactive conformation. Such inactive compact ClpP structures have been observed for *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* ClpPs. The active extended ClpP structure has so far been observed for WT *Ec*ClpP, *H. pylori* ClpP, *Bacillus subtilis* ClpP. In a recent analysis, we have argued, on the basis of experimental and theoretical data, that the compact structures of *S. pneumoniae, M. tuberculosis* and *P. falciparum* ClpPs represent a naturally sampled compact state of the ClpP cylinder. Based on our data, we have proposed a model whereby the ClpP cylinder switches dynamically between an active extended state required for substrate degradation and an inactive compact state allowing peptide product release.

The region N-terminal of αA in mature *Pf*ClpP is made of 36 residues whereas in *Ec*ClpP there are only 16 residues (Fig. 1). These 16 residues in *Ec*ClpP form axial loops that are located at the entrance of the tetradecameric cylinder and, hence, control the access of substrates to the degradation chamber and mediate the communication with the bound chaperone. The axial loops are rather flexible but can form β-hairpins that delimitate a narrow pore of 10–12 Å diameter. In the structure of *Pf*ClpP reported here, only a few residues of the N-terminal region proximal to αA are visible and they follow a trajectory similar to that observed for *Ec*ClpP. This finding suggests that the N-terminal region of *Pf*ClpP might adopt a conformation similar to that of *Ec*ClpP, at least for the first few residues proximal to αA. However, the conformation and role of the additional residues at the N-terminal region of *Pf*ClpP remain to be determined. One possibility is that these residues adopt the β-hairpin conformation and trace into the digestion chamber. Alternatively, they might fold back and protrude on the apical surfaces of *Pf*ClpP. In any case, these additional residues might have important roles in the translocation of substrate proteins or in the transmission of allosteric signals between *Pf*ClpP and *Pf*ClpC.

In summary, our study provides the first comprehensive overview of the Clp chaperones and proteases of *P. falciparum* and highlights a possible important role of these proteins in maintaining protein homeostasis in the apicoplast. Disregulation of this homeostasis by developing drugs that target the Clp chaperones and proteases might provide a novel approach in the fight against malaria.

**Materials and Methods**

**Plasmid construction**

Unless indicated otherwise, all *Pf*ClpP and *Pf*ClpR proteins are of *P. falciparum* 3D7 origin. All nucleotides and reagents were purchased from Sigma. The *Pf*clpP (gene ID in PlasmoDB database is *PFC0310c*) and *Pf*clpR (*PF14_0348*) genes were cloned from genomic DNA into the pET28a-LIC (Genbank accession number EF442785.1) and p15TV-L (Genbank accession number EF456736.1) vectors, respectively. All N-terminal deletion constructs were generated by PCR amplification using primers that introduce an NdeI cut site in-frame with the first codon of each construct and a BamHI cut site after the termination codon. The resulting pET28a-*Pf*ClpP plasmid has a single methionine deletion (Δ-1 M) that removes an internal translation initiation site and was used as a template for constructs expressed for crystallization experiments. For biochemical and biophysical studies, the amplified DNA was cloned into p11 (a kind gift from Dr Alexei Savchenko, Clinical Genomic Centre, Toronto) and pET9a (New England Biolabs) plasmids. The p11 vector is a modified pET vector with an N-terminal His6 tag followed by a tobacco etch virus (TEV) cleavage site. pET9a does not introduce any tag to the protein.

**Tagging of *Pf*ClpC, *Pf*ClpB1 and *Pf*ClpB2 gene products by 3′ replacement**

Streptavidin and three tandemly linked hemagglutinin epitopes were inserted into the 3′ end of the *Pf*clpC, *Pf*clpB1 and *Pf*clpB2 genes of the 3D7 strain using the Gateway multisite systems vectors as described. *P. falciparum* parasites were transfected by electroporation of ring stage cultures using 150 μg of plasmid DNA as described. Following recovery of drug-resistant parasites, cultures were grown without the selective drug, 5 mM 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 weeks before reapplication of the drug. This cycling between growth with drug and without drug was continued until it was determined that the integration at the 3′ end of the endogenous gene had occurred, resulting in the incorporation of an epitope tag. This was verified by PCR screening.

**Antibody generation**

The peptides: ETKLHPYFNKVEK ADEAVDFKLDHILEKE corresponding to the C-terminal amino acid sequences of *Pf*ClpP and *Pf*ClpR, respectively, were synthesized at the
Advanced Protein Technology Center at the Hospital for Sick Children (Toronto, Canada). These peptides were conjugated to keyhole limpet hemocyanin and were used to generate polyclonal antiserum in rabbit by four subcutaneous injections (500 mg/injection followed by 250 mg boosts), using Freund’s adjuvant and standard protocols. Antisera against PfClpM were generated using the synthetic peptide KKNFLSKDLKEIC fused to diphtheria toxin ( Mimotopes, Clayton, Victoria, Australia). Serum recovered after four subcutaneous injections into goats was affinity-purified using unconjugated peptide bound to CNBr-activated Sepharose 4B resin (GE Healthcare) according to the manufacturer’s instructions.  

Western blot analysis

Parasites were grown to 5–15% parasitaemia in 2% hematocrit using standard techniques. Red blood cells were lysed in 0.15% saponin, 0.075% (w/v) BSA in PBS to release the parasite, then the parasite pellet was washed three times in PBS. SDS-PAGE and Western blotting were done as described. For the experiments shown in Fig. 3a, the primary antibodies and dilutions used were as follows: Roche rat anti-hemaglutinin (1:500), affinity-purified goat anti-Roche rat anti-hemaglutinin (1:500), rabbit anti-PfClpM (1:500), rabbit anti-PfClpP (1:4000) and rabbit anti-PfClpR (1:4000). Secondary antibodies were used were DAKO anti-rat IgG horseradish peroxidase (1:750), Sigma anti-goat IgG horseradish peroxidase and Sigma anti-rabbit IgG (1:2000).

For the experiments shown in Fig. 4, P. falciparum 3D7 cell lysate was centrifuged at 15,000 × g for 15 min at 4 °C then 20 μL of supernatant were subjected to SDS-PAGE (12% polyacrylamide (w/v) gel) and blotted onto nitrocellulose or poly(vinylidene fluoride) membranes for Western blot analysis using the anti-PfClpP or anti-PfClpR serum at dilutions of 1:1500 and 1:1000, respectively. To isolate whole parasite cells from RBCs, fresh parasite culture was incubated with 0.02% (w/v) saponin (Sigma) in PBS for 10 min on ice. The mixture was then centrifuged at 15,000g. The pellets, which contained whole parasite cells surrounded by empty RBC membranes, also known as ghosts, were collected by centrifugation while the supernatants, which contained RBC ghosts, were collected by centrifugation and lysed by sonication for 5 min at 10 s intervals. All fractions were subjected to SDS-PAGE analysis. 

Immunofluorescence assays

Cultures of P. falciparum were grown to at least 5% parasitaemia then fixed and labeled as described. Antibodies were used as follows: primary antibodies were rabbit anti-ACP (1:1000), Roche rat anti-hemaglutinin (1:200), affinity-purified anti-PfClpM (1:100) and anti-PfClpP (1:1000). Secondary antibodies were Alexa Fluor 546 red anti-rabbit IgG (1:750), Alexa Fluor 488 green anti-rat IgG (1:750) and AlexaFlour 488 green anti-goat IgG (1:1000). Images were captured on a Leica TCS 4D confocal microscope.

Protein expression and purification

The expression of recombinant PfClpP and PfClpR was done in BL21(DE3) CodonPlus-RIL E. coli (Stratagene, La Jolla, CA) by inducing the transformed strain with 1 mM IPTG at an absorbance of 0.6 and grown overnight in Terrific Broth at 18 °C. Cells were harvested by centrifugation and lysed by sonication for 5 min at 10 s intervals on ice with the addition of 5 mg/mL of lysozyme. For His6-tagged PfClpP and PfClpR, protein purification was done with Ni2+–NTA agarose beads (Qiagen) following the manufacturer’s protocol. Purified PfClpP and PfClpR were dialyzed into buffer P (50 mM Tris–HCl, pH 7.5, 300 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). They were then further purified as described for the untagged proteins.

For untagged PfClpP constructs, cells were suspended in buffer A (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 10% glycerol, 1 mM DTT) and lysed. Cell lysate was filtered and subjected to crude ion-exchange chromatography using a Q-Sepharose column attached to an AKTA FPLC system (GE Healthcare). PfClpP eluted at 15–20% buffer B (50 mM Tris–HCl, pH 7.5, 1 M NaCl, 10% glycerol, 1 mM DTT). Elution fractions containing PfClpP were pooled, dialyzed in buffer A, and subjected to ion-exchange chromatography using a MonoQ 5/5 HR column. Fractions containing PfClpP were pooled, concentrated, and subjected to SEC using a calibrated Superdex 200 HR 10/30 column in buffer P. PfClpP elutes at fractions corresponding to the molecular mass of its heptameric form. These fractions were pooled, concentrated, flash-frozen in liquid nitrogen and stored at −80 °C. For untagged PfClpR constructs, cells were suspended in buffer A with 5 mg/mL of lysozyme and lysed by sonication. Crude lysate was filtered and subjected to ion-exchange chromatography in both Q-Sepharose and SP-Sepharose columns. Contaminants were bound to the columns and the unbound PfClpR proteins were recovered in the flow-through. Flow-through fractions containing PfClpR were pooled, concentrated, dialyzed in buffer B and subjected to hydrophobic interaction chromatography using phenyl Sepharose 6 FF high-sub column. PfClpR eluted at 75–85% buffer C (50 mM Tris–HCl, pH 7.5, 10% glycerol, 1 mM DTT). The elution fractions were pooled, concentrated, and subjected to SEC purification using a calibrated Superdex 200 HR 10/30 column in buffer P. PfClpR elutes at fractions corresponding to the molecular mass of its heptameric form. These fractions were pooled, concentrated, flash-frozen in liquid nitrogen and stored at −80 °C. All proteins were analyzed by SDS-PAGE and found to be >95% pure. Protein concentrations were determined using the Bradford assay (Bio-Rad). All reported concentrations are calculated as monomers unless indicated otherwise.

Analytical ultracentrifugation

Sedimentation equilibrium analytical ultracentrifugation experiments were done at the Ultracentrifugation Service Facility at the University of Toronto, Department of Biochemistry. PfClpP and PfClpR proteins at different
concentrations in buffer P were spun at 6000 rpm, 8000 rpm and 10000 rpm at 4 °C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti rotor. Absorbance at 230 nm and 280 nm was monitored. Data analysis was done with the Origin MicroCal XL-A/CL-I Data Analysis Software Package version 4.0.

**Surface plasmon resonance experiments**

Surface plasmon resonance experiments were done at 25 °C using a BiaCore X instrument (GE Healthcare). PfClpP (179-370) or PfClpR(49-244) at a concentration of 100 μg/mL were covalently linked to an activated Biacore CM5 sensorchip 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) P2O surfactant). The surface was regenerated between injections with a 1 min pulse of 2 M NaCl. The steady-state responses were plotted against the corresponding analyte concentrations and the dissociation constants were derived by fitting the data to a Langmuir binding isotherm using BioEvaluation 4.1 software (GE Healthcare).

**Peptidase assays**

Peptidase assays for PfClpP(179-370) and PfClpR(49-244) were done in a 200 μL reaction volume in a 96-well plate. The reaction mix contained 20 μM recombinant protein in assay buffer D (0.1 M sodium acetate, pH 7.0, 1 mM DTT) in the presence or in the absence of an inhibitor (100 μM chymostatin, PMSF, pepstatin, leupeptin or aprotinin). Six different fluorogenic peptide substrates were used (purchased from Sigma): Suc-LLVY-AMC, Suc-LLVY-AMC, Suc-IWV-AMC, Suc-IAV-AMC, Suc-AAPF-AMC and Suc-AFk-AMC added to a final concentration of 50 μM. AMC release was monitored continuously by the increase in fluorescence intensity (excitation 355 nm, emission 460 nm) for 6 h at 30 °C using an EnSpire 2300 Multilabel Plate Reader (PerkinElmer). For comparison, EctClpP protease activity was tested under the same conditions but in buffer E (50 mM Tris-HCl, pH 8.0, 0.2 M KCl, 1 mM DTT). K_m and k_cat for the hydrolysis of Suc-LLVY-AMC substrate (1.9–6.25 μM) at constant enzyme concentration (20 μM) were obtained by fitting the enzyme kinetic data to the Michaelis–Menten model using SigmaPlot.

**Electron microscopy and image processing**

PfClpP and PfClpR protein samples were visualized using transmission electron microscopy (TEM) by floating a previously glow-discharged carbon-coated grid on a 10 μL drop of sample with a protein concentration of ∼50 μg/mL. The samples were negatively stained with 1% (w/v) uranyl acetate and observed in a JEOL 2010F electron microscope operated at 200 kV. Images were collected at a magnification of 50,000× with a dose of ∼10 e⁻/Å². All images were recorded on Kodak SO-163 films, scanned on a Nikon Super COOLSCAN 9000 ED at 6.35 μm/pixel and averaged twice to produce data at 2.54 Å/pixel. Particles were extracted from the electron micrographs using the Boxer program (EMAN). Rotational symmetry analysis was done with two statistical tests, the spectral ratio product and Student's t-test as described. Cross-correlation methods as implemented in the Xmipp program were used to align the particles and to calculate the two-dimensional averages.

Specimens were prepared for analysis with scanning transmission electron microscopy (STEM) according to the standard method of the Brookhaven National Laboratory STEM facility as described. 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⁻/Å² and the grid was kept at <150 °C during data collection. Images were collected digitally from the large detector (40–200 mrad acceptance angles) and used for mass determination. The images were collected with a scan width of 0.512 μm. Processing the STEM images and mass determination of the particle images was done using the program PCMass as described.

**X-ray crystallography**

The structure was reported by us earlier as part of a structural genomics project (accession number 2F6I), however, the structure was not analyzed or described in that study. Many different PfClpP truncation constructs were generated for crystallization trials. All constructs had an N-terminal His tag followed by a TEV cleavage site added before the PfClpP sequence was tested. The truncated protein constructs were expressed and purified as described using the Lex bioreactor system (Harbinger Biotechnology and Engineering, Markham, Ontario, Canada). The protein that crystallized had the N-terminal tag: 5GSSHHHHHHHSGRENLYFQGHM followed by the protein sequence from D179–K370. Crystals of the purified PfClpP protein were grown using the hanging-drop, vapor-diffusion method. The drop was formed by mixing equal parts of protein solution and a reservoir solution of 23% polyethylene glycol monomethyl ether 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 parameters as given in Table 3. For data collection, the crystals were first transferred to a cryoprotectant solution consisting of 50 μL of reservoir solution supplemented with 15 μg of 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, data at 2.45 Å resolution 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.

The positions of seven molecules in the asymmetric unit were determined by AMoRe, a 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. The model was completed using Coot interactive graphics.
biological unit formed by crystallographic symmetry. The includes a 7-fold heptamer, with the full tetradecamer Table 3. Molprobity was used to evaluate the structure.89

Acknowledgements

W.A.H. and A.P. thank Dr Kevin C. Kain and his student Kodjo Ayi at the University of Toronto for their help with parasite work. J.O. and K.L.Y.C. acknowledge Joel Kooistra for early electron microscopy work on this project. M.E.B. is a postdoctoral fellow of the Canadian Institutes of Health Research Training Program in Protein Folding and Interaction Dynamics: Principles and Diseases. A.P. is a graduate student of the Canadian Institutes of Health Research Training Program in Protein Folding: Principles and Diseases. This work was supported by grants from the National Institutes of Health Research to W.A.H. (MOP-67210) and J.O. (MOP-82930). J.O. is a Canadian Institutes of Health Research New Investigator. G.I.M. is an Australian Research Council Federation Fellow and a Howard Hughes International Research Scholar. Program Grant support from the National Health and Medical Research Council of Australia is gratefully acknowledged. The Australian Red Cross generously supplied human red blood cells. The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co. Inc., the Novartis Research Foundation, the Petrus and Augusta Hedlund's Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust. Use of the IMCA-CAT beamline 17ID at the Advanced Photon Source was supported by the member companies of the Industrial Macromolecular Crystallography Association through a contract with the University of Chicago. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number W-31-109-Eng-38. Use of beamline X25 of the National Synchrotron Light Source was supported principally from the Office of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy, and the National Center for Research Resources of the National Institutes of Health.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.09.051

References


**Supplementary Material**

**SUPPLEMENTARY TABLE 1**

Apicoplast targeting sequence prediction

<table>
<thead>
<tr>
<th>Name</th>
<th>PlasmoAP (signal peptide, apicoplast targeting peptide)</th>
<th>PATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfClpP</td>
<td>Yes (+,+ )</td>
<td>Yes</td>
</tr>
<tr>
<td>PfClpR</td>
<td>Yes (+,+ )</td>
<td>Yes</td>
</tr>
<tr>
<td>PfClpB1</td>
<td>Yes (+,+ )</td>
<td>Yes</td>
</tr>
<tr>
<td>PfClpB2</td>
<td>Yes (+,+ )</td>
<td>No</td>
</tr>
<tr>
<td>PfClpC</td>
<td>No (-,+ )</td>
<td>Yes</td>
</tr>
<tr>
<td>PfClpM</td>
<td>not applicable</td>
<td>not applicable</td>
</tr>
</tbody>
</table>
Legend for Supplementary Figure

Supplementary Fig. 1. Sequence properties of the *P. falciparum* Clp ATPases. (A) Sequence alignment of the Walker A and Walker B motifs in the *P. falciparum* Clp ATPases and *Ec*ClpA. (B) Sequence alignment of the insertion domains of *Pf*ClpB1 (residues 563 – 737), *Pf*ClpB2 (residues 438 – 559), *Tt*ClpB (SwissProt CLPB_THET2), *Se*ClpB (SwissProt CLPB1_SYNE7), and *Ec*ClpB (SwissProt CLPB_ECOLI). Residues that form the leucine-rich heptad repeat according to *Tt*ClpB structure\(^1\) are highlighted in blue boxes and indicated by blue arrows. (C) Sequence alignment of the region spanning the ClpP binding loop between Sensor 1 and Box VII in the *P. falciparum* Clp ATPases and *Ec*ClpA. All alignments were done using ClustalW2\(^2\). Residues that are 100% identical in the five sequences are highlighted in red, while those in red font are highly similar.
References

