doi:10.1016/j.jmb.2010.09.051 J. Mol. Biol. (2010) 404, 456–477

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

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

Majida El Bakkouri[1]†, Andre Pow[1]†, Anne Mulichak[2]†, Kevin L. Y. Cheung[3]†, Jennifer D. Artz[4], Mehrnaz Amani[4], Stuart Fell[5], Tania F. de Koning-Ward[6], C. Dean Goodman[5], Geoffrey I. McFadden[5], Joaquin Ortega[3], Raymond Hui[4] and Walid A. Houry[1]*

[1]*Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8*
[2]*IMCA-CAT, Argonne National Lab, Argonne, IL 60439, USA*
[3]*Department of Biochemistry and Biomedical Sciences and MG DeGroote Institute for Infectious Diseases Research, McMaster University, Hamilton, Ontario, Canada L8N 3Z5*
[4]*Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada M5G 1L5*
[5]*School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia*
[6]*School of Medicine, Deakin University, Pigdons 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 *Pf*ClpB1, *Pf*ClpB2, *Pf*ClpC and *Pf*ClpM. One *Pf*ClpP, the proteolytic subunit, and one *Pf*ClpR, 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 *Pf*ClpB2 (also known as Hsp101), which was found in the parasitophorous vacuole. Both *Pf*ClpP and *Pf*ClpR form mostly homoheptameric rings as observed by size-exclusion chromatography, analytical ultracentrifugation and electron microscopy. The X-ray structure of *Pf*ClpP 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*.

*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; OD_{600}, optical density at 600 nm; OM, outer membrane; PBP, periplasmic binding protein; SBP, substrate-binding protein; SeMet, selenomethionine.

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 protease oligomer.[5] Such proteases include the proteasome, ClpP (*c*aseino*l*ytic *p*roteases) and HslV proteases.

The ClpPs are highly conserved across prokaryotes[6] 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 (HslU).

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 *Se*ClpP$_1$P$_2$ protease complex that interacts with *Se*ClpX, a *Se*ClpP$_3$R complex that interacts with *Se*ClpC and a *Se*ClpP$_1$R 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 pathways[28] 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 protein[31] and Clp ATPases[32] 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] Dysregulation[34] or selective inhibition[35] 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 (AAL03913) were retrieved from the NCBI database[36] and used in BLASTP queries of the *P. falciparum* nuclear, plastid and mitochondrial protein databases of PlasmoDB.[37] Putative hits were further confirmed by reciprocal BLASTP against the UniProt protein database.[38]

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[37] 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 COGNITOR program.[39] Both sequences were identified as belonging to COG0740, a COG group corresponding to Clp proteases. We subsequently aligned *Ec*ClpP 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 *Pf*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.[19] PF14_0348 is referred to here as *Pf*ClpR (Table 1).

Four putative proteins displayed very significant sequence homology to *S. elongatus* ClpC. One putative *clp* ATPase gene (PFC10_API0060) is found on the plastid DNA,[30] 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 *Pf*ClpB1, *Pf*ClpB2, *Pf*ClpC and *Pf*ClpM, 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 *Pf*HslU (*Pf*ClpY, PFI0355c) and *Pf*HslV (*Pf*ClpQ, PFL1465c) proteins, which are not discussed further here but have been described elsewhere.[40]

All the *Pf*Clp 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 *Pf*Clp ATPases (Fig. 2), except for D1 of apicoplast-encoded PFC10_API0060. D1 of PFC10_API0060 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.*,[12] we have named PFC10_API0060 as *Pf*ClpM (Table 1).

Unlike *Ec*ClpA, *Pf*Clp ATPases have an insertion of 18-174 residues at a site in the small subdomain of the D1 AAA+ module (Fig. 2) with *Pf*ClpM 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.[41] Subsequently, the presence of certain motifs was used for such a classification.[12] However, since the X-ray structure of *Thermus thermophilus* ClpB (*Tt*ClpB) is solved and shows that the insertion region forms a coiled coil,[42] 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[43] and the Paircoil,[44] the insertion region of both PF08_0063 and PF11_0175 are predicted to form coiled coils; hence, we name them *Pf*ClpB1 and *Pf*ClpB2, respectively. Close inspection of the insertion region based on sequence alignment of *Pf*ClpB1, *Pf*ClpB2, *Tt*ClpB, *Se*ClpB and *Ec*ClpB reveals that residues that are part of the leucine-

Table 1. The Clp proteins of *P. falciparum*

Name	Mass (kDa)[a]	Gene ID[b]	Gene location	Protein type	Localization[c]
*Pf*ClpP	43	PFC0310c	Nucleus	Clp protease	Apicoplast
*Pf*ClpR	28	PF14_0348	Nucleus	Inactive Clp protease	Apicoplast[d]
*Pf*ClpB1	123	PF08_0063	Nucleus	Clp ATPase	Apicoplast
*Pf*ClpB2	103	PF11_0175	Nucleus	Clp ATPase	Parasitophorous vacuole[e]
*Pf*ClpC	156	PF14_0063	Nucleus	Clp ATPase	Apicoplast
*Pf*ClpM	91	PFC10_API0060	Apicoplast	Clp ATPase	Apicoplast

[a] This is the theoretical molecular mass of the unprocessed protein based on its gene size.
[b] This is the PlasmoDB gene ID.[37]
[c] As determined in this study.
[d] Prediction only.
[e] Refer also to de Koning-Ward.[53]

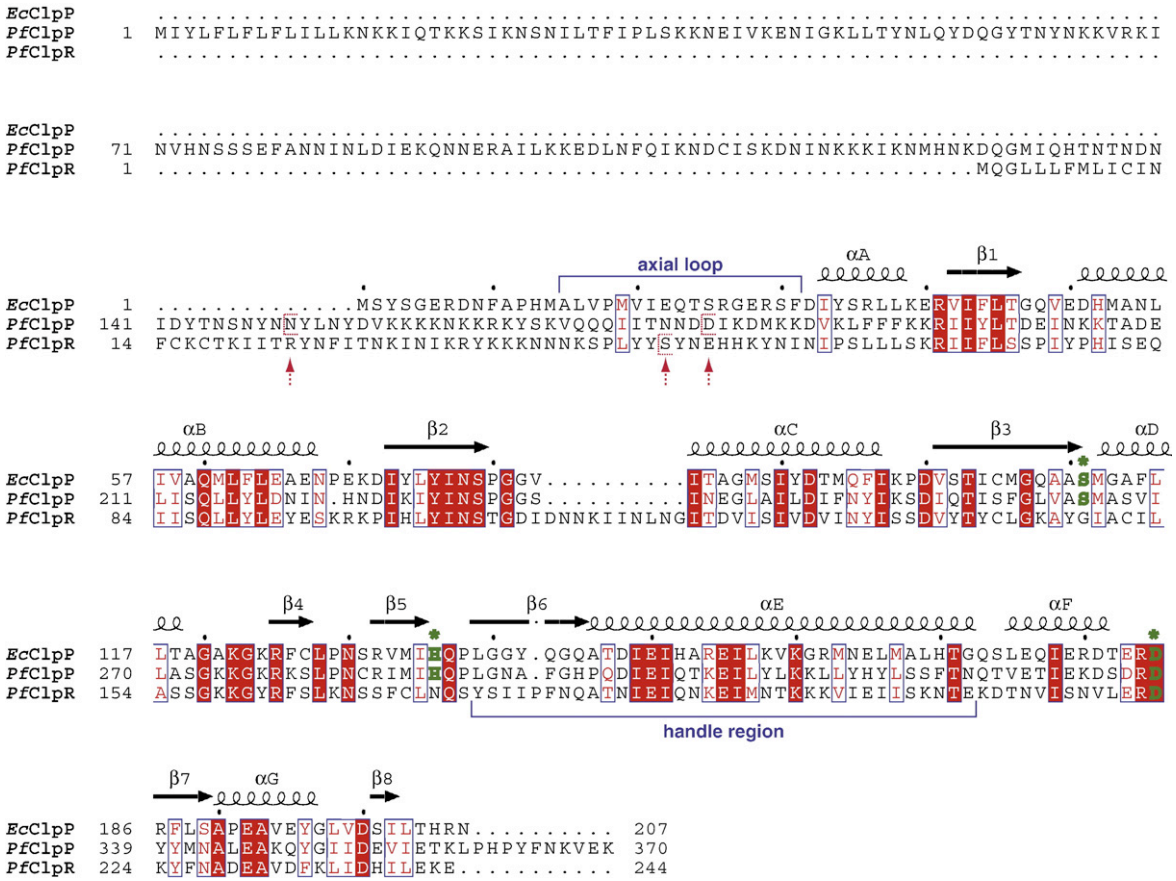

Fig. 1. Sequence properties of the *P. falciparum* Clp proteases. Sequence alignment of *Ec*ClpP, *Pf*ClpP and *Pf*ClpR using ClustalW2[90] and drawn with ESPript.[91] Residues that are 100% identical in the three sequences are highlighted in red, while those in red font are highly similar. Residues of the Ser-His-Asp catalytic triad are in bold green and are indicated by an asterisk. N150 and D179 of *Pf*ClpP and S49 of *Pf*ClpR are indicated with brackets and red arrows. Secondary structure elements shown on top of the sequence alignment are based on the *Ec*ClpP structure (PDB code 1yg6).[64] The axial loop present at the N-terminus of *Ec*ClpP and the handle region formed by β6 and αE are indicated.

rich heptad repeat in the *Tt*ClpB structure[42] are generally conserved on all these ClpBs, but to a lesser degree in *Pf*ClpB2 (Supplementary Data Fig. 1B). Note also that the insertion region of *Pf*ClpB1 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

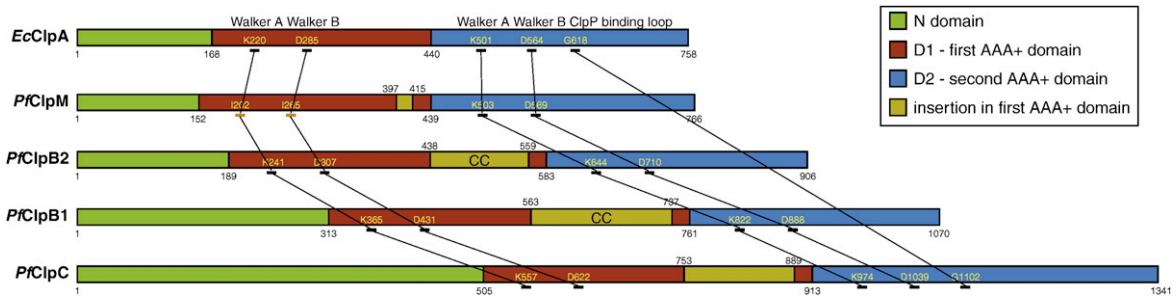

Fig. 2. Domain arrangement of the *P. falciparum* Clp ATPases. Cartoon representation of the domain organization of the *P. falciparum* Clp ATPases and *Ec*ClpA shown for comparison. The residue numbers of the lysine in the Walker A motif G*K*T, the aspartic acid in the Walker B motif ΦΦ*D*E, and the 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 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].[45] Clp ATPases that lack the tripeptide motif are not known to form a complex with the ClpP protease. Based on multiple sequence alignment, only *Pf*ClpC 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 *Pf*ClpC (Supplementary Data Fig. 1C). Hence, we propose that *Pf*ClpC associates with *Pf*ClpP and *Pf*ClpR 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 bipartite leader sequence.[47] Initial targeting begins via the secretory pathway into the endoplasmic reticulum with the classical signal peptide at the very N-terminal region. Subsequent targeting across the inner membranes of the apicoplast is achieved by a downstream transit peptide that has properties similar to those of the transit peptides directing proteins into plastids of plants and algae.[48,49] The sequence composition of the transit peptide in apicoplast-targeted proteins is not well defined, but it is generally basic. As a result, it is often challenging to identify the start position of the functional domain of these proteins. Both the signal and transit peptides are absent from mature apicoplast proteins because they are cleaved during apicoplast import.[50]

*Pf*ClpM does not require an apicoplast targeting sequence because it is encoded and presumably translated in the apicoplast. The targeting sequence predictor algorithms PlasmoAP[51] and PATS[52] were used to screen for the presence of an apicoplast targeting sequence in the other *P. falciparum* Clp proteins (Supplementary Data Table 1). Both algorithms predict that *Pf*ClpP, *Pf*ClpR and *Pf*ClpB1 contain apicoplast targeting sequences (Supplementary Data Table 1) and, hence, should be imported into the apicoplast. *Pf*ClpB2 is predicted by PlasmoAP but not by PATS to have a targeting sequence. However, recently published experimental data show that *Pf*ClpB2 (also known as Hsp101) is located in the parasitophorous vacuole membrane and is part of a protein export machine.[53] We re-confirm parasitophorous vacuole localization below and, hence, the PlasmoAP prediction is wrong. Finally, PATS but not PlasmoAP predicts an apicoplast targeting sequence in the *Pf*ClpC protein (Supplementary Data Table 1).

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-3×hemagluttinin (Strp-3×HA) tag at the 3′ end.

A 120 kDa protein was detected by 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 (Fig. 3a). The apparent mass of *Pf*ClpB1-Strp-3×HA is close to that predicted, while the apparent mass of *Pf*ClpC-Strp-3×HA is slightly greater than predicted (Table 1). Western blots thus confirm that both fusion proteins are expressed in asexual blood stages. No antibody reaction was observed in the parental, untagged line of parasites. Colocalization using antibodies detecting the stromal apicoplast marker acyl carrier protein (ACP),[48] and anti-HA to detect either *Pf*ClpB1-Strp-3×HA or *Pf*ClpC-Strp-3×HA produce immunofluorescence that overlaps (Fig. 3b1 and b2), demonstrating that *Pf*ClpB1-Strp-3×HA and *Pf*ClpC-Strp-3×HA are in the apicoplast. We confirmed the report that *Pf*ClpB2 (Hsp101) is expressed as a 100 kDa protein (Fig. 3a), and it is localized in the parasitophorous vacuole[53] by colocalizing *Pf*ClpB2-Strp-3xHA with the parasitophorous vacuole marker *Pf*PTEX150 (Fig. 3b3).

A Western blot analysis of ring-stage parasite strain 3D7 lysate probed with anti-*Pf*ClpM antibodies revealed a single band of ~95 kDa (Fig. 3a), which is congruent with the predicted mass from the apicoplast gene (Table 1). Immunofluorescence assays showed that *Pf*ClpM colocalized with the apicoplast ACP (Fig. 3b4).

Fig. 3. Expression and localization of the *Pf*Clp proteins. (a) Expression of *Pf*ClpB1, *Pf*ClpC, *Pf*ClpB2, *Pf*ClpM, *Pf*ClpP and *Pf*ClpR was checked in blood stage parasites. Lanes 1–3, Western blots of lysates from parasites having *Pf*ClpB1-Strp-3xHA, *Pf*ClpC-Strp-3xHA and *Pf*ClpB2-Strp-3xHA probed with anti-HA antisera. 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 *Pf*ClpB1, *Pf*ClpC, *Pf*ClpB2, *Pf*ClpM and *Pf*ClpP of *P. falciparum* was tested by immunofluorescent colocalization. (1) *Pf*ClpB1-Strp-3× HA and (2) *Pf*ClpC-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). (3) *Pf*ClpB2-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). (4) and (5) wild type 3D7 strain was probed with anti-ACP antisera (red), with anti-*Pf*ClpM antisera or anti-ClpP antisera (green), and with Hoescht 33342 (blue). (1)–(5) The rightmost image represents the merged fluorescence images with the DIC or transmission image. The scale bars represent 5 μm.

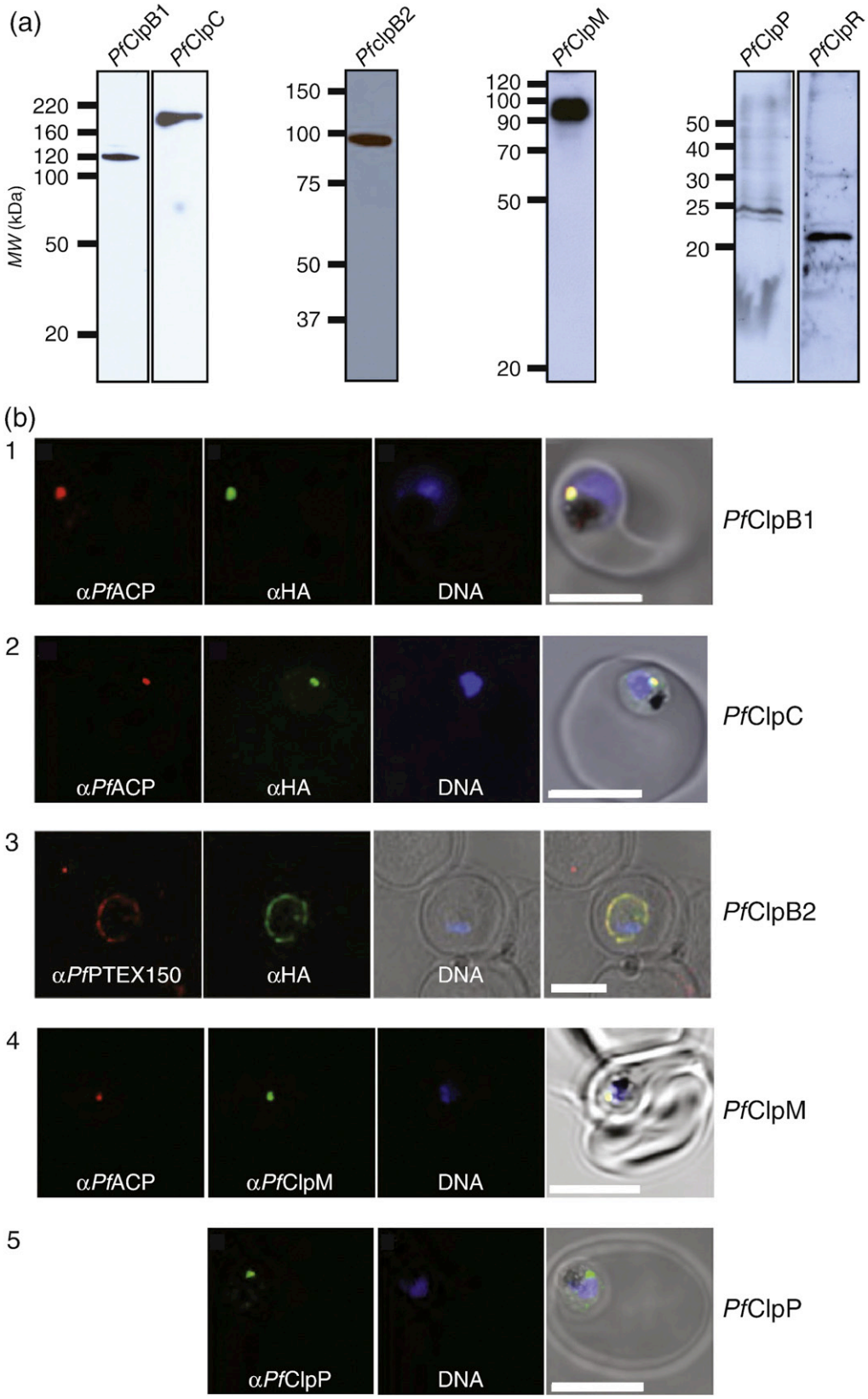

Fig. 3 (*legend on previous page*)

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 *Pf*ClpP or *Pf*ClpR, respectively (Fig. 3a). Immunofluorescence using the antibodies against *Pf*ClpP 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 *Pf*ClpP with ACP antibody because both sera are from the rabbit. The antibodies against *Pf*ClpR 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 *Pf*ClpB1, *Pf*ClpC, *Pf*ClpM and likely *Pf*ClpP are localized to the apicoplast. Secretion of *Pf*ClpB2 to the parasitophorous vacuole, where it participates in export of virulence proteins to the red cell,[53] is confirmed. At this stage, *Pf*ClpR is predicted to be targeted to the apicoplast (Supplementary Data Table 1).

Sequence boundaries of the mature forms of *Pf*ClpP and *Pf*ClpR

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 *Pf*ClpP and *Pf*ClpR 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 *Pf*ClpP and *Pf*ClpR, 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-*Pf*ClpP and anti-*Pf*ClpR antibodies, respectively (Figs. 3a and 4). The polyclonal antibodies generated are against the C-terminal peptides of *Pf*ClpP and *Pf*ClpR 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 *Pf*ClpP and *Pf*ClpR constructs (see below) to be used as references in the Western blot analysis. It was found that truncated *Pf*ClpP beginning at residue N150 and truncated *Pf*ClpR beginning at residue S49 are closest in molecular mass to the observed processed *Pf*ClpP and *Pf*ClpR as found by SDS-PAGE (Fig. 4). Hence, we propose that *Pf*ClpP(150-370) with a theoretical mass of 25.7 kDa and *Pf*ClpR(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 *Pf*ClpP and *Pf*ClpR

To investigate the oligomeric state of the predicted mature forms of the *Pf*Clp protease subunits, *Pf*ClpP (150-370) and *Pf*ClpR(49-244) were expressed in *E. coli* and purified as described in Materials and Methods. Both *Pf*ClpP(150-370) and *Pf*ClpR(49-244) are 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

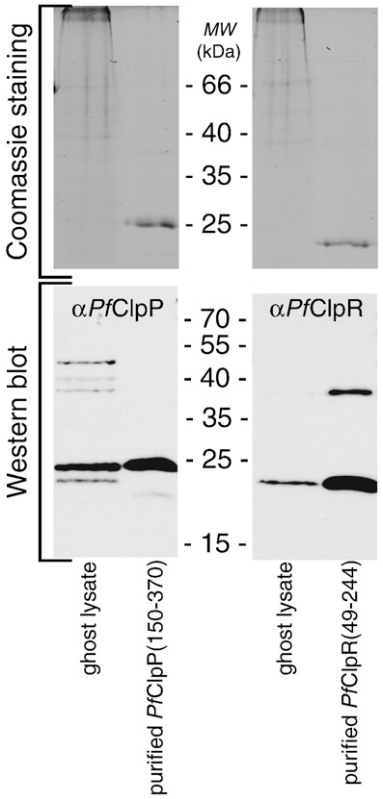

Fig. 4. Size of the mature *Pf*ClpP and *Pf*ClpR in *P. falciparum*. Ghosts' lysates and recombinant N-terminally truncated *Pf*ClpP or *Pf*ClpR were loaded onto the same gel. Upper panels, proteins stained with Coomassie brilliant blue; lower panels, Western blot analysis of the same samples using anti-*Pf*ClpP and anti-*Pf*ClpR.

10/30 column. Consequently, 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 (Fig. 1). Each of these constructs was expressed and purified. *Pf*ClpP(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.

*Pf*ClpP(179-370) migrated at about 146 kDa and *Pf*ClpR(49-244) migrated at about 134 kDa when subjected to SEC (Fig. 5a). Both proteins seemed to form heptameric complexes as compared to tetradecameric *Ec*ClpP (theoretical mass of 303.8 kDa;

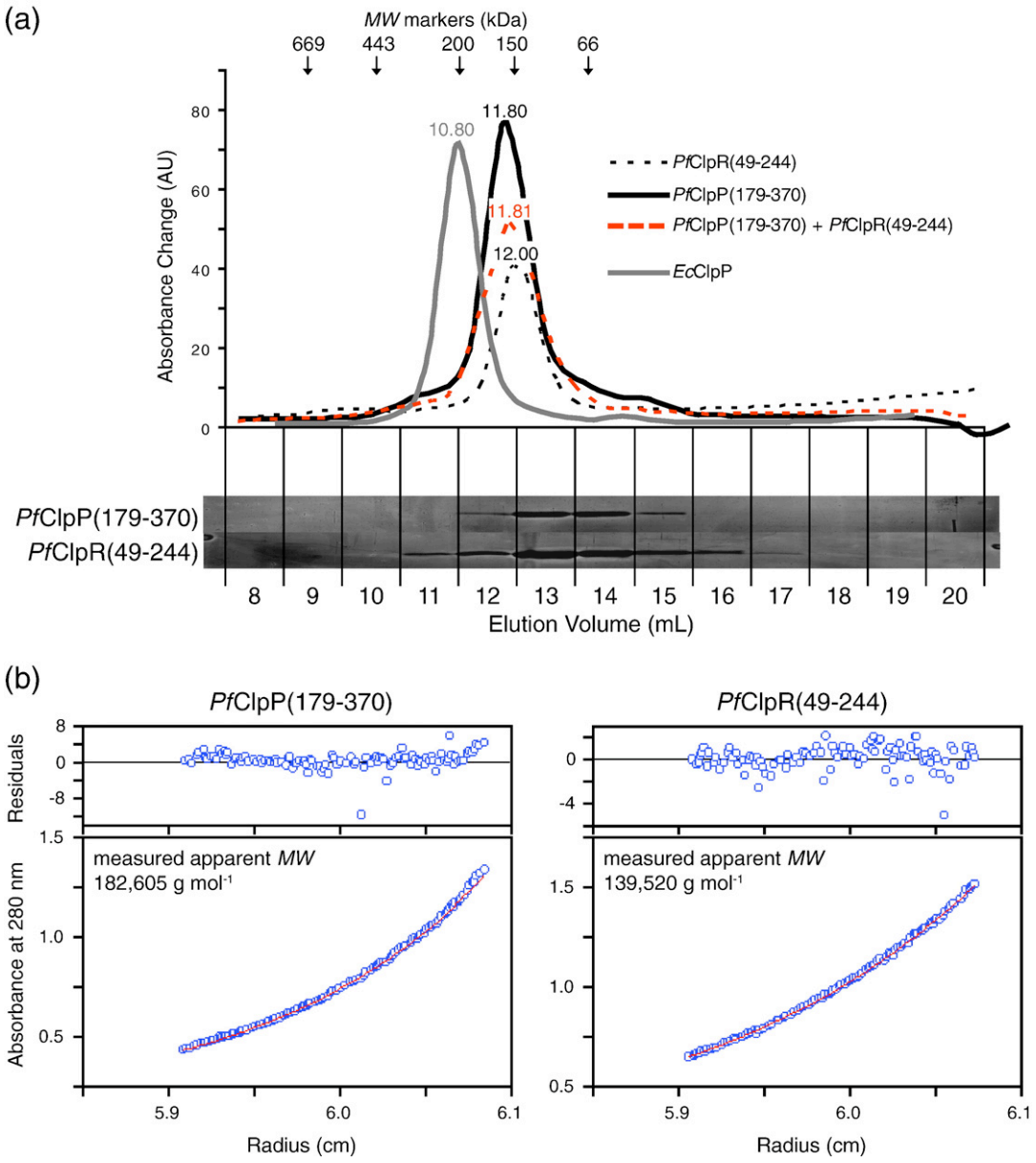

Fig. 5. Oligomeric state of *Pf*ClpP(179-370) and *Pf*ClpR(49-244). (a) SEC of the proteins on a Superdex 200 HR 10/30 column in buffer P. The position of the molecular mass markers are shown at the top. The amounts loaded were: *Pf*ClpP (179-370), 2.50 mg; *Pf*ClpR(49-244), 2.50 mg; *Pf*ClpP(179-370)+*Pf*ClpR(49-244), 1.25 mg+1.25 mg; *Ec*ClpP, 1.00 mg. The silver stained SDS-PAGE gels shown below the chromatograms are of *Pf*ClpP(179-370) and *Pf*ClpR(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 8000 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 molecular mass as given. The residual deviations from the theoretical fits are given in the upper panels.

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 *Pf*ClpP(179-370) or *Pf*ClpR(49-244).

Analytical ultracentrifugation sedimentation equilibrium analysis at 4 °C (Fig. 5b) was used to further establish the oligomeric state of *Pf*ClpP(179-370) and *Pf*ClpR(49-244). The plots of ln(A) *versus* r^2 (where A is absorbance at 280 nm and r is radius) for both *Pf*ClpP(179-370) and *Pf*ClpR(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 *Pf*ClpP(179-370) and 139.5 kDa for *Pf*ClpR(49-244) (Fig. 5b). The ratio of the oligomer mass to the theoretical monomer mass for *Pf*ClpP(179-370) is 8.3:1 and for *Pf*ClpR(49-244) it is 6.2:1. The slightly elevated ratio observed for *Pf*ClpP(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

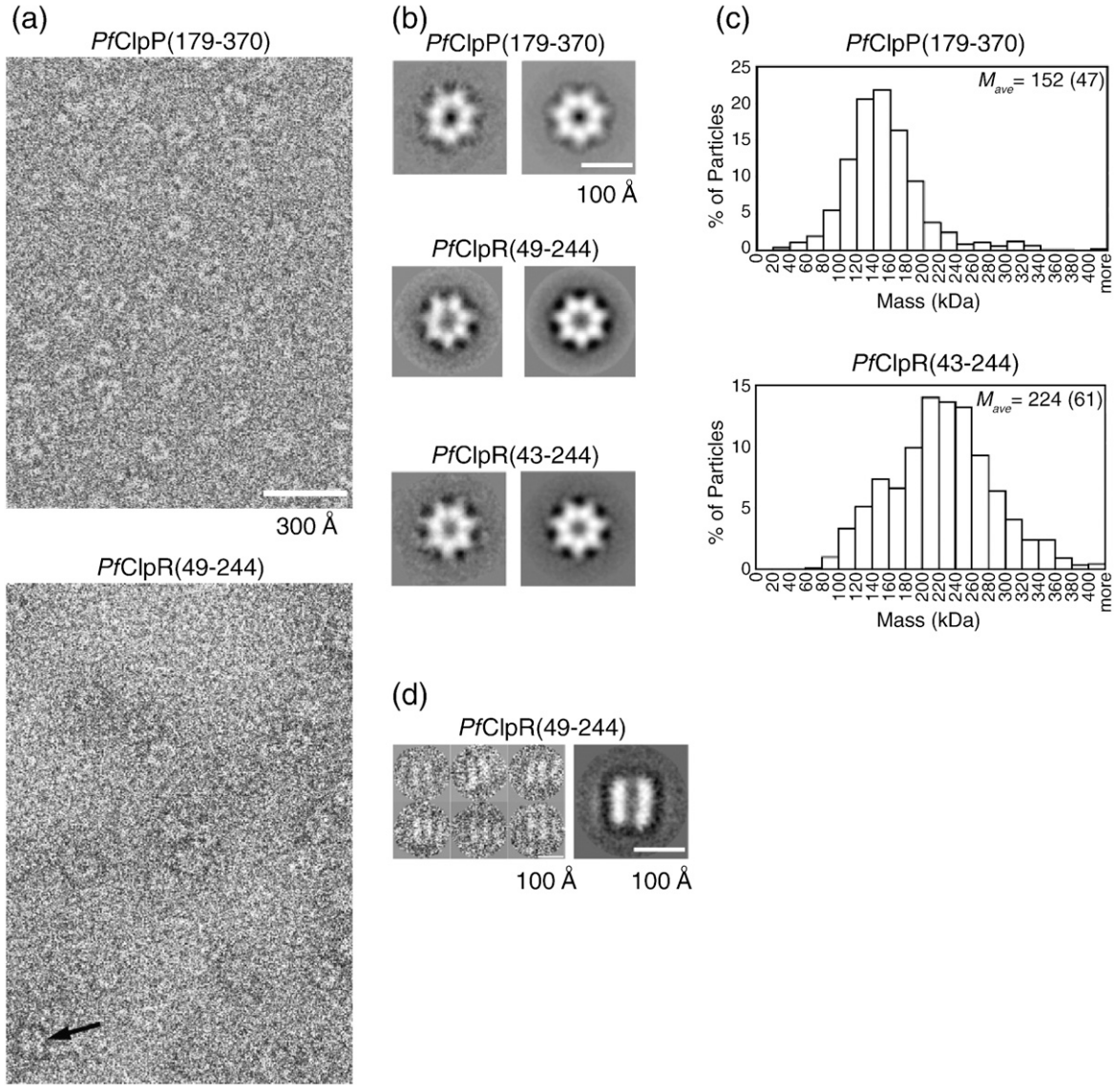

Fig. 6. Electron microscope analysis of *Pf*ClpP and *Pf*ClpR. (a) Negatively stained electron micrographs showing ring-shaped particles of *Pf*ClpP(179-370) and *Pf*ClpR(49-244). For *Pf*ClpR(49-244), particles made of double parallel striations are present in low abundance (black arrow). (b) The calculated averages for the top views of *Pf*ClpP(179-370), *Pf*ClpR(49-244), and *Pf*ClpR(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 (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 *Pf*ClpR(49-244) heptameric rings (left-hand panels). The calculated average for this type of particles is shown in the right-hand panel.

ratio observed for *Pf*ClpR(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 *Pf*ClpP and *Pf*ClpR. When *Pf*ClpP(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[54] 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 *Ec*ClpP.[55] In order to discriminate whether these top-view projections of *Pf*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 (Fig. 6c). 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. These results are consistent with those obtained by SEC and analytical ultracentrifugation sedimentation equilibrium experiments (Fig. 5) and all these together indicate that *Pf*ClpP (179-370) assembles predominantly into a single heptameric ring. Interestingly, this is unlike other ClpPs, such as that of *E. coli*[56] or *Helicobacter pylori*[57] that readily form tetradecamers, but similar to human ClpP that assembles mainly into heptamers and forms tetradecamers in the presence of its cognate ClpX chaperone.[58]

Similar experiments were done with *Pf*ClpR(49-244) and *Pf*ClpR(43-244), two constructs of *Pf*ClpR. 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 *Pf*ClpP(179-370) (Fig. 6b). However, in addition to the ring-shaped particles observed in the electron micrographs for *Pf*ClpR, 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 *Pf*ClpR heptameric rings (Fig. 6d). These results, together with our earlier SEC and analytical centrifugation experiments, suggest that *Pf*ClpR(49-244) and *Pf*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 *Pf*ClpR(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 *Pf*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. In addition, some particles were observed with a mass corresponding to a tetradecamer. These results suggest that *Pf*ClpR constructs form a heterogeneous oligomeric mixture containing heptamers and tetradecamers and other oligomeric forms of variable mass.

The interaction between *Pf*ClpP(179-370) and *Pf*ClpR(49-244)

Several experiments were done to determine whether *Pf*ClpP(179-370) and *Pf*ClpR(49-244) are able to form the typical tetradecameric Clp protease complex. Because *Pf*ClpP(179-370) and *Pf*ClpR(49-244) can readily form heptamers, we speculated that a *Pf*ClpRP complex would be formed by a *Pf*ClpP homoheptameric ring interacting with a *Pf*ClpR homoheptameric ring. Pre-incubation of *Pf*ClpP(179-370) with *Pf*ClpR(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 *Pf*ClpP(179-370) and *Pf*ClpR(49-244) might form heteroheptamers, the two proteins were co-expressed in *E. coli* from a polycistronic pST39 plasmid[59] with one of the proteins being tagged. The tagged protein was pulled down; however, there was no evidence of significant interaction between *Pf*ClpP(179-370) and *Pf*ClpR(49-244). The isolated proteins formed heptamers, as detected by SEC, indicating that the complexes formed are likely homoheptamers 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 *Pf*ClpP(179-370) or *Pf*ClpR(49-244) heptamer was immobilized on the sensorchip and sensograms were

Table 2. Analysis of the rotational symmetry of *Pf*ClpP and *Pf*ClpR

Sample	Particles analyzed	Symmetry detected	*p* (*t*-test)	Spectral ratio product
*Pf*ClpP(179-370)	174	7	<0.000001	4.28×10^{10}
*Pf*ClpR(49-244)	150	7	<0.000001	6.68×10^{26}
*Pf*ClpR(43-244)	980	7	<0.000001	4.85×10^{30}

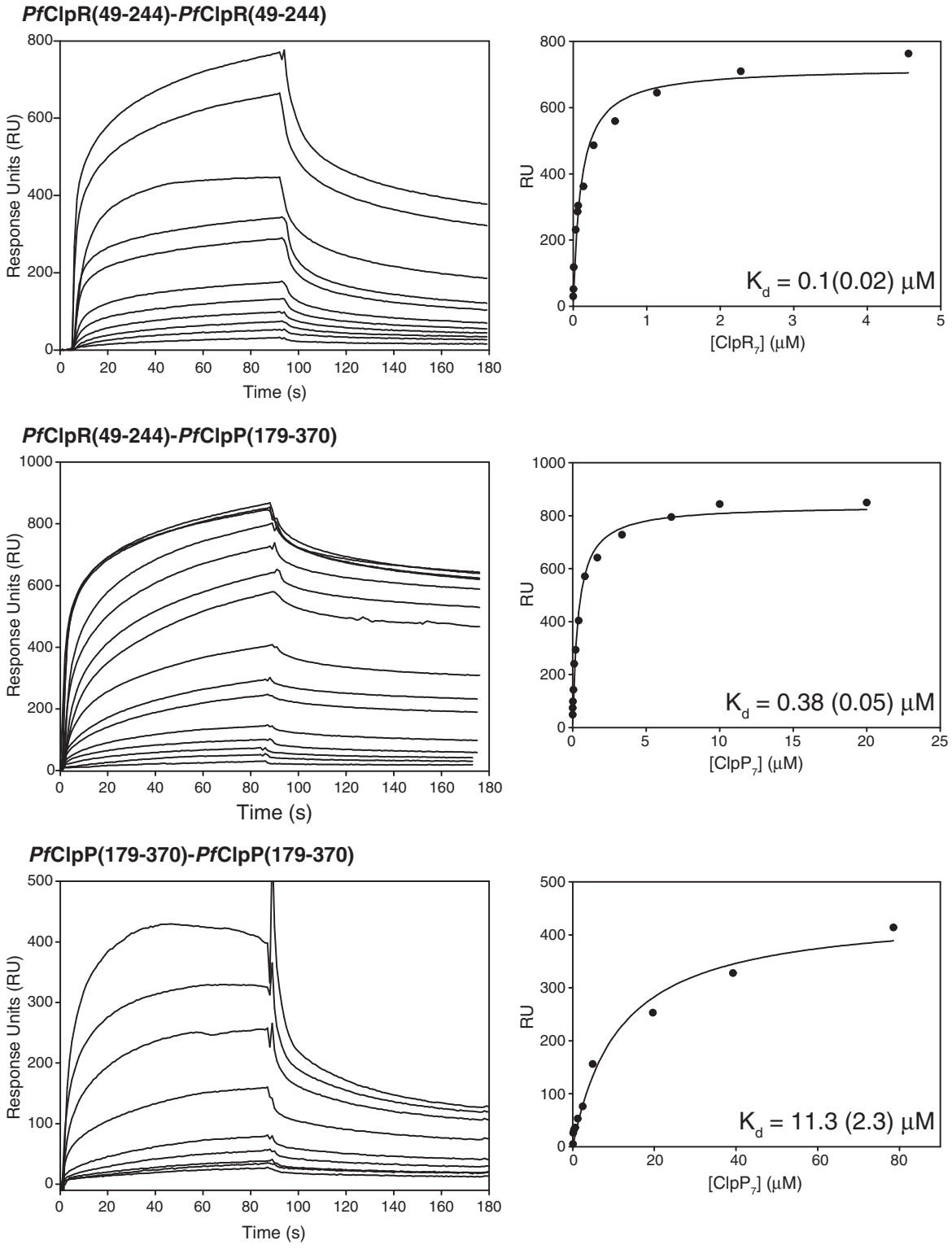

Fig. 7. Surface plasmon resonance analysis of the interaction between *Pf*ClpP(179-370) and *Pf*ClpR(49-244). BIAcore sensorgrams and the derived equilibrium binding curves for the interaction between *Pf*ClpR(49-244)-*Pf*ClpR(49-244), *Pf*ClpP(179-370)-*Pf*ClpR(49-244) and *Pf*ClpP(179-370)-*Pf*ClpP(179-370) 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.

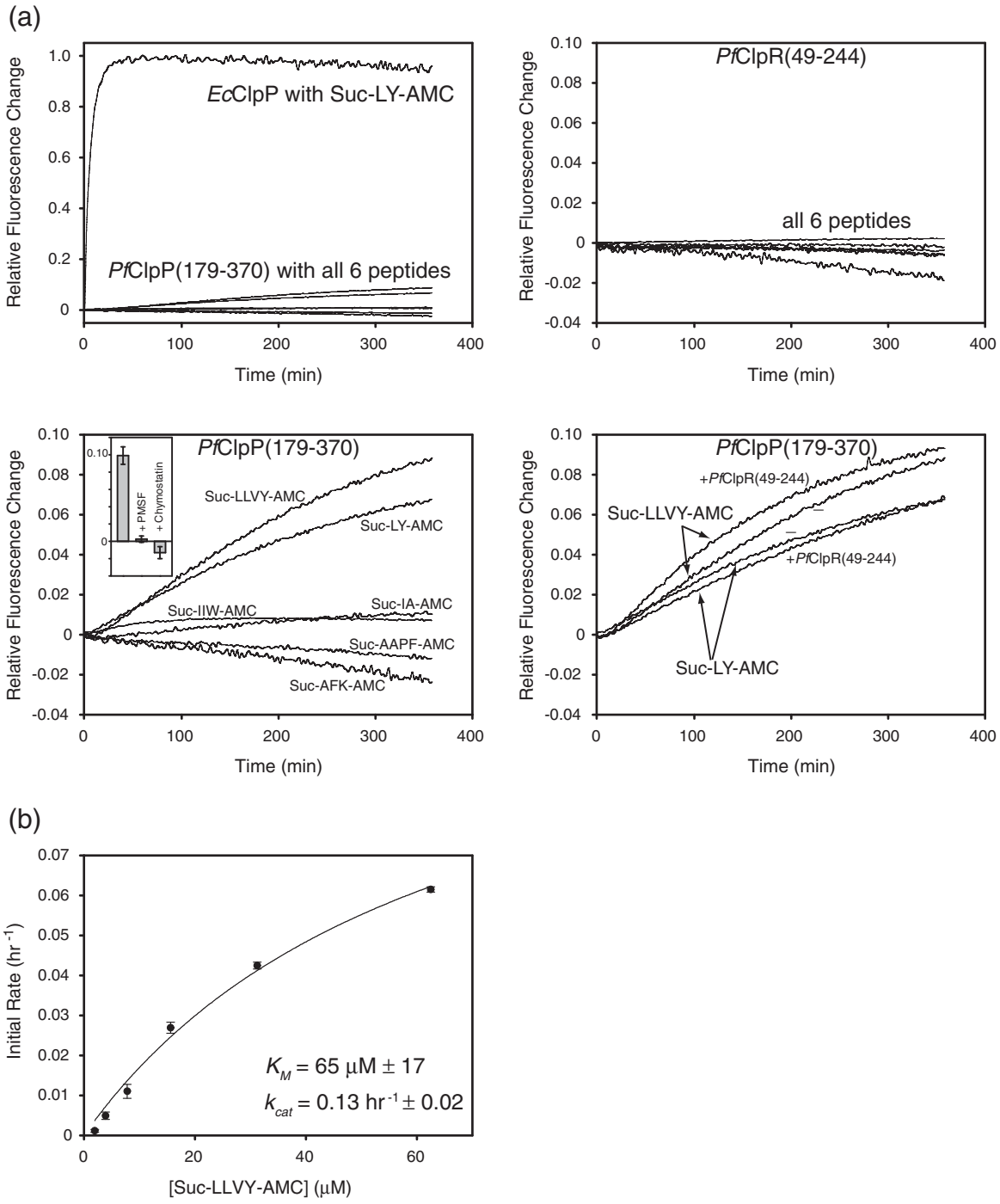

Fig. 8. Peptidase activity of *Pf*ClpP(179-370). (a) Peptidase activity of 20 μM *Pf*ClpP(179-370), 20 μM *Pf*ClpR(49-244) or 20 μM *Pf*ClpP(179-370) + 20 μM *Pf*ClpR(49-244) were measured at 30 °C by their ability to cleave 50 μM Suc-LLVY-AMC, Suc-LY-AMC, Suc-IIW-AMC, Suc-IA-AMC, Suc-AAPF-AMC or Suc-AFK-AMC. 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 fluoresence intensity change after incubation for 6 h. (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 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: *Pf*ClpR(49-244) – *Pf*ClpR(49-244) > *Pf*ClpP(179-370) – *Pf*ClpR(49-244) > *Pf*ClpP(179-370) – *Pf*ClpP(179-370). This is consistent with the observation of double ring complexes by electron microscopy for *Pf*ClpR(49-244) and *Pf*ClpR(43-244) (Fig. 6a, c and d). Hence, if *Pf*ClpP and *Pf*ClpR are at equimolar concentrations in the apicoplast, then the major tetradecameric complexes observed would be that of the *Pf*ClpR homo-oligomer and a *Pf*ClpP-*Pf*ClpR double ring formed from a heptameric *Pf*ClpP and heptameric *Pf*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 *Pf*ClpC (Fig. 2), would drive tetradecamer formation and might dictate whether the complex formed is a $PfClpP_7$-$PfClpR_7$, $PfClpP_{14}$,or $PfClpR_{14}$ oligomer.

Peptidase activity of *Pf*ClpP(179-370)

The peptidase activity of *Pf*ClpP(179-370), *Pf*ClpR (49-244) and a mixture of *Pf*ClpP(179-370) with *Pf*ClpR(49-244) was assessed using model peptides such as the Suc-LY-AMC typically used for measuring *Ec*ClpP peptidase,[60] as well as Suc-AFK-AMC, Suc-AAPF-AMC, Suc-IA-AMC, Suc-IIW-AMC and Suc-LLVY-AMC (Fig. 8a). The peptidase activity of *Ec*ClpP against Suc-LY-AMC was used as a reference. No activity was observed for *Pf*ClpR(49-244), while very weak activity was detected for *Pf*ClpP (179-370) against Suc-LLVY-AMC and Suc-LY-AMC. The peptidase activity of *Pf*ClpP(179-370) was inhibited by two serine protease inhibitors, chymostatin and 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 (Fig. 8a, inset in lower left panel). The activity of *Pf*ClpP(179-370) was not significantly affected by the presence of *Pf*ClpR(49-244) (Fig. 8a). The *Pf*Clp proteases were unable to degrade longer peptides such as:

MCA-YEVHHQKLVFK(DNP)-NH_2

and

MCA-SEVNLDAEFR-Ednp-KRR-NH2·3TFA

or casein, which was reported to be degraded by *Ec*ClpP in the absence of the ATPase chaperone.[61,62]

K_M and k_{cat} for Suc-LLVY-AMC hydrolysis by *Pf*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 *Ec*ClpP against Suc-LY-AMC[60] are 1 mM and 150 min^{-1}, respectively. Hence, *Pf*ClpP (179-370) is a very weak peptidase on its own.

Structure of *Pf*ClpP(179-370)

We were able to crystallize and solve the X-ray structure of the *Pf*ClpP construct that had the N-terminal tag:

MGSSHHHHHHSSGRENLYFQGHM

followed by the protein sequence from D179–K370, H_6-*Pf*ClpP(179-370) (Table 3; Fig. 9a). 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.

The ClpP protomer structure can be divided into three parts (Fig. 1). The first part is the N-terminal

Table 3. Data collection and structure refinement statistics for H_6-*Pf*ClpP(179-370)

A. *Data collection*	
Space group	$C222_1$
Cell parameters	
a (Å)	158.3
b (Å)	196.5
c (Å)	139.2
α (°)	90.0
β (°)	90.0
γ (°)	90.0
Wavelength (Å)	1.10
Resolution (Å)	50-2.45
Measured reflections	1,525,816
Unique reflections	78,895
R_{sym}[a](last shell)[b]	0.075 (0.316)
$\langle I/\sigma(I)\rangle$	15.0 (4.7)
Completeness	0.992 (0.985)
Redundancy	7.5 (4.5)
B. *Structure refinement*	
Resolution range (Å)	30–2.45
Reflections, working set	72,825
Reflections, test set	3790
R_{work}[c]	0.210
R_{free}[d]	0.238
Protein atoms	9941
Water atoms	248
Mean *B*-factor (Å^2)	50.6
Ramachandran	
Favored (%)	95.2
Outliers (%)	0.4
r.m.s.d. from ideal	
Bond lengths (Å)	0.005
Bond angles (°)	1.00

Values in parentheses are for the last shell.
[a] $R_{sym} = \Sigma |I - \langle I\rangle| / \Sigma I$, where I is the observed intensity of a measured reflection and $\langle I\rangle$ is the mean intensity of that reflection.
[b] The last shell includes all reflections between 2.45 and 2.54 Å.
[c] R_{work} is $\Sigma |F_o - F_c| / \Sigma F_o$.
[d] R_{free} is the cross-validation *R*-factor computed for a test set of reflections (5% of total).

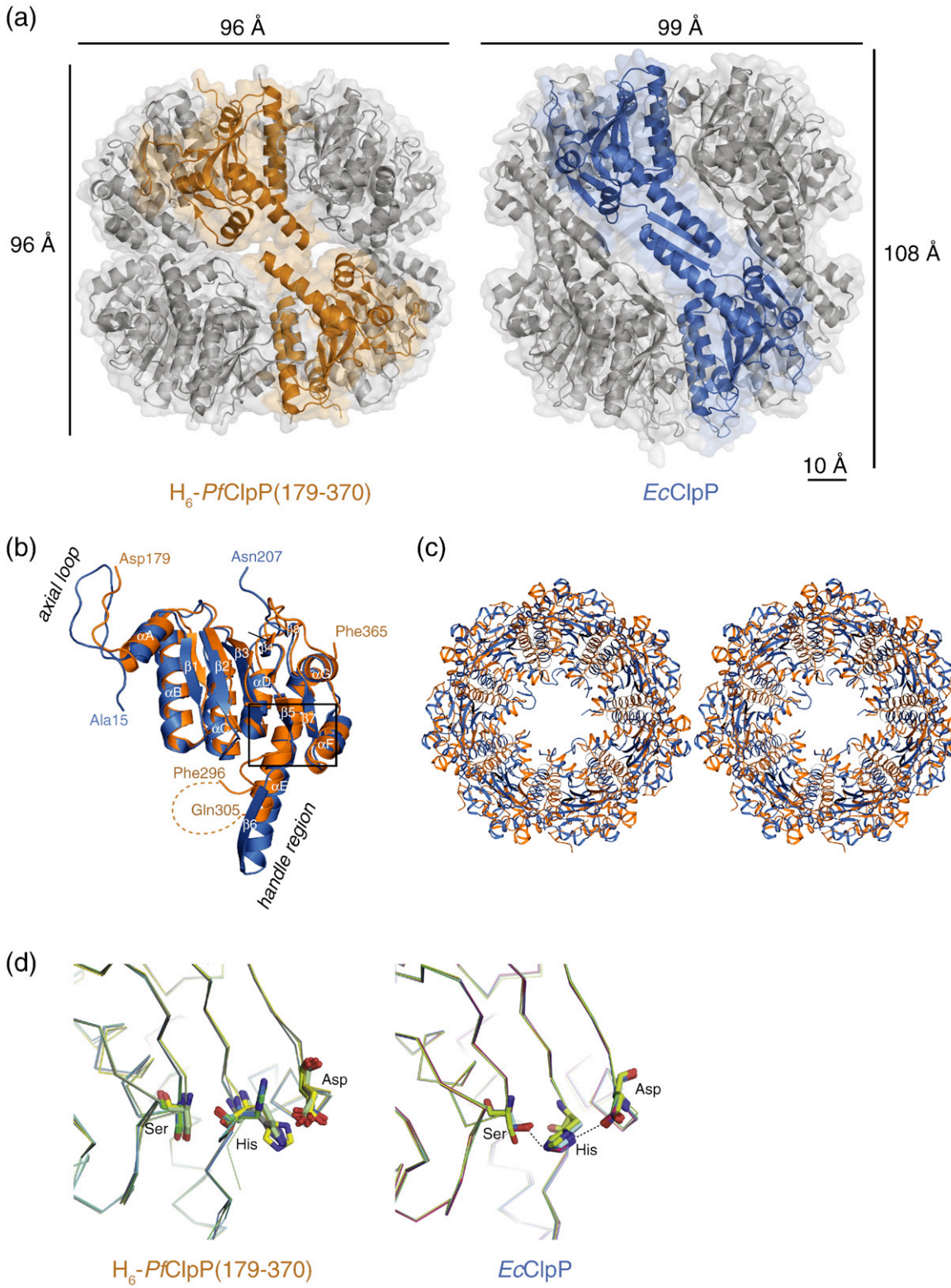

Fig. 9 (*legend on next page*)

axial loop consisting of residues N-terminal to αA, which is required for the interaction of ClpP with its cognate chaperones.[63–65] In *Ec*ClpP, 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 (Fig. 9b). While this complete N-terminal loop arrangement is not observed in *Pf*ClpP, resolved residues N-terminal to αA appear to follow a trajectory similar to that observed for *Ec*ClpP axial loops (Fig. 9b).

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 *Pf*ClpP superimposes well with that of *Ec*ClpP with minimal perturbations (rmsd 1.15 Å) consistent with the high level of conservation in the protein sequence.

The third part of ClpP is the handle region, which is formed by β6 and αE (Fig. 1). Handle regions from two ClpP single rings interdigitate in the double ring structure. Residues F296–Q305 are unstructured in the handle region of H_6-*Pf*ClpP(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 closer to each other in *Pf*ClpP than in *Ec*ClpP resulting in a more compact tetradecamer. Indeed, the opposing apical surfaces are ~10 Å closer in H_6-*Pf*ClpP(179-370) than in *Ec*ClpP, as measured from the 2D projections of the proteins. Furthermore, if the subunits of one ring of *Ec*ClpP are superposed on the subunits of one ring of H_6-*Pf*ClpP(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 H_6-*Pf*ClpP(179-370) have multiple configurations, especially His289, when compared to the respective residues of *Ec*ClpP, which form a hydrogen bonded network. This observation and the disordered β6 guide strand suggest that the *Pf*ClpP structure corresponds to a catalytically inactive state of the protease.

Discussion

We have provided a comprehensive overview of the Clp system in *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 *Pf*ClpCRP complex in the apicoplast of the parasite. The apicoplast also harbors a nuclear-encoded (*Pf*ClpB1) and an apicoplast-encoded (*Pf*ClpM) Clp ATPases.

An earlier report concluded that *Pf*ClpP is localized to the nucleus based on the expression of the parasite gene in human cells,[66] but our localization of *Pf*ClpP to the apicoplast with antisera to the native protein contradicts the proxy localization in a mammalian surrogate system. The high lysine content of *Pf*ClpP[67] 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 *Pf*ClpP was published that also demonstrated that the protease is localized to the apicoplast as we have found.[33] Consistent with our findings, those authors showed that purified *Pf*ClpP 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 *PfclpP* gene was found to be maximally transcribed in late trophozoite and early schizont stage parasites and the inactivation of *Pf*ClpP using β lactone-based inhibitors, albeit at relatively high concentrations, disrupted apicoplast development.

The molecular mass detected for both *Pf*ClpP and *Pf*ClpR 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 *Pf*ClpP and 48 amino acids are removed from the N-terminus of *Pf*ClpR. In contrast, the *Ec*ClpP 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 *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 *E. coli*, *Ec*ClpS has been found to recognize substrates that are degraded following the N-end rule pathway.[68,69] In *S. elongatus*

Fig. 9. The X-ray structure of H_6-*Pf*ClpP(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*, chain F, (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 H_6-*Pf*ClpP(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 the 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 (www.pymol.org).

PCC7942, *Se*ClpS1 was found to inhibit the degradation of generic model substrates such as casein and to promote the degradation of N-end rule substrates.[17] A ClpS ortholog also exists in *P. falciparum* (gene ID MAL13P1.111) and it is predicted to be localized to the apicoplast.[28] 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 tetradecamer 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*[63] and *Mycobacterium tuberculosis* ClpPs.[70] The active extended ClpP structure has so far been observed for WT *Ec*ClpP,[56,64,65] *Homo sapiens* mitochondria ClpP,[71] *H. pylori* ClpP[57] and *Bacillus subtilis* ClpP.[72] 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.[73] Based on our data,[73] 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 axial pore on each end of the tetradecameric cylinder and, hence, control the access of substrates to the degradation chamber and mediate the communication with the bound chaperone.[63,64,74] 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 *PfclpP* (gene ID in PlasmoDB database[37] is ***PFC0310c***) and *PfclpR* (***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 His_6 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 hemagluttinin epitopes were inserted into the 3′ end of the *PfclpC*, *PfclpB1* and *Pf*clpB2 genes of the 3D7 strain using the Gateway multisite systems vectors[75] as described.[76] *P. falciparum* parasites were transfected by electroporation of ring stage cultures using 150 μg of plasmid DNA as described.[77] 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 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:

ETKLPHPYFNKVEK
ADEAVDFKLIDHILEKE

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 antisera in rabbit by four subcutaneous injections (500 mg/injection followed by 250 mg boosts), using Freund's adjuvant and standard protocols. Antisera against *Pf*ClpM were generated using the synthetic peptide

KNKNFLSKSDLKEIC

fused to diptheria 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.[78] 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.[79] For the experiments shown in Fig. 3a, the primary antibodies and dilutions used were as follows: Roche rat anti-hemagluttinin (1:500), affinity-purified goat anti-*Pf*ClpM (1:500), rabbit anti-*Pf*ClpP (1:4000) and rabbit anti-*Pf*ClpR (1:4000). Secondary antibodies 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-*Pf*ClpP or anti-*Pf*ClpR 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 15,000*g*. 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 cytosol, were discarded. The parasite pellets were washed twice in PBS and lysed by sonication for 10 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.[77] Antibodies were used as follows: primary antibodies were rabbit anti-ACP (1:1000), Roche rat anti-hemagluttinin (1:200), affinity-purified anti-*Pf*ClpM (1:100) and anti-*Pf*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.

Protein expression and purification

The expression of recombinant *Pf*ClpP and *Pf*ClpR 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 at 600 nm 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 His_6-tagged *Pf*ClpP and *Pf*ClpR, protein purification was done with Ni^{2+}-NTA agarose beads (Qiagen) following the manufacturer's protocol. Purified *Pf*ClpP and *Pf*ClpR 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 *Pf*ClpP 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). *Pf*ClpP eluted at 15~20% buffer B (50 mM Tris–HCl, pH 7.5, 1 M NaCl, 10% glycerol, 1 mM DTT). Elution fractions containing *Pf*ClpP were pooled, dialyzed in buffer A, and subjected to ion-exchange chromatography using a MonoQ 5/5 HR column. Fractions containing *Pf*ClpP were pooled, concentrated, and subjected to SEC using a calibrated Superdex 200 HR 10/30 column in buffer P. *Pf*ClpP 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 *Pf*ClpR 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 *Pf*ClpR proteins were recovered in the flow-through. Flow-through fractions containing *Pf*ClpR were pooled, concentrated, dialyzed in buffer B and subjected to hydrophobic interaction chromatography using phenyl Sepharose 6 FF high-sub column. *Pf*ClpR 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. *Pf*ClpR 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. *Pf*ClpP and *Pf*ClpR 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). *Pf*ClpP (179-370) or *Pf*ClpR(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) P20 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 BiaEvaluation 4.1 software (GE Healthcare).

Peptidase assays

Peptidase assays for *Pf*ClpP(179-370) and *Pf*ClpR(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-LY-AMC, Suc-IIW-AMC, Suc-IA-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, *Ec*ClpP 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 (1.9–62.5 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

*Pf*ClpP and *Pf*ClpR 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 $e^-/Å^2$. 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).[80] Rotational symmetry analysis was done with two statistical tests, the spectral ratio product and Student's *t*-test as described.[54] Cross-correlation methods as implemented in the Xmipp program[81] 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.[82] 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^-/nm^2 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 PCMass29[83] as described.[82]

X-ray crystallography

The structure was reported by us earlier as part of a structural genomics project (accession number 2F6I),[31] however, the structure was not analyzed or described in that study. Many different *Pf*ClpP truncation constructs were generated for crystallization trials. All constructs had an N-terminal His_6 tag followed by a TEV cleavage site added before the *Pf*ClpP sequence was tested. The truncated protein constructs were expressed and purified as described[31] 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 *Pf*ClpP 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 $C222_1$, 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 mg 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.[84]

The positions of seven molecules in the asymmetric unit were determined by AMoRe,[85] a molecular replacement programs from the CCP4 crystallographic program suite,[86] using as a search model the highly conserved portions of the *E. coli* ClpP heptamer, PDB entry 1TYF.[56] The model was completed using Coot interactive graphics,[87]

alternated by cycles of refinement with CNS Solve version 1.1 software.[88] In early stages of refinement, simulated annealing protocols were used, followed by simple positional refinement and individual *B*-factor refinement. In later stages of refinement, water molecules were added to the model where $|F_o–F_c|$ electron density maps showed peaks at least 3σ 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 *Pf*ClpP model includes a 7-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 before residue 304 or 305. The statistics for data collection and structure refinement are summarized in 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 Canadian 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

1. Baker, T. A. & Sauer, R. T. (2006). ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem. Sci.* **31**, 647–653.
2. Mogk, A., Haslberger, T., Tessarz, P. & Bukau, B. (2008). Common and specific mechanisms of AAA+ proteins involved in protein quality control. *Biochem. Soc. Trans.* **36**, 120–125.
3. Hartl, F. U. & Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581.
4. Kress, W., Maglica, Z. & Weber-Ban, E. (2009). Clp chaperone-proteases: structure and function. *Res. Microbiol.* **160**, 618–628.
5. Lupas, A., Flanagan, J. M., Tamura, T. & Baumeister, W. (1997). Self-compartmentalizing proteases. *Trends Biochem. Sci.* **22**, 399–404.
6. Yu, A. Y. & Houry, W. A. (2007). ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett.* **581**, 3749–3757.
7. Sakamoto, W. (2006). Protein degradation machineries in plastids. *Annu. Rev. Plant Biol.* **57**, 599–621.
8. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43.
9. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A. *et al.* (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell*, **119**, 9–18.
10. Ammelburg, M., Frickey, T. & Lupas, A. N. (2006). Classification of AAA+ proteins. *J. Struct. Biol.* **156**, 2–11.
11. Snider, J. & Houry, W. A. (2008). AAA+ proteins: diversity in function, similarity in structure. *Biochem. Soc. Trans.* **36**, 72–77.
12. Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. (1996). HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.* **21**, 289–296.

13. Dougan, D. A., Mogk, A., Zeth, K., Turgay, K. & Bukau, B. (2002). AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.* **529**, 6–10.
14. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31.
15. Snider, J., Thibault, G. & Houry, W. A. (2008). The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216.
16. Stanne, T. M., Pojidaeva, E., Andersson, F. I. & Clarke, A. K. (2007). Distinctive types of ATP-dependent Clp proteases in cyanobacteria. *J. Biol. Chem.* **282**, 14394–14402.
17. Andersson, F. I., Tryggvesson, A., Sharon, M., Diemand, A. V., Classen, M., Best, C. *et al.* (2009). Structure and function of a novel type of ATP-dependent Clp protease. *J. Biol. Chem.* **284**, 13519–13532.
18. Shanklin, J., DeWitt, N. D. & Flanagan, J. M. (1995). The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease. *Plant Cell*, **7**, 1713–1722.
19. Adam, Z., Rudella, A. & van Wijk, K. J. (2006). Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. *Curr. Opin. Plant Biol.* **9**, 234–240.
20. Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P. & van Wijk, K. J. (2001). Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J. Biol. Chem.* **276**, 16318–16327.
21. Adam, Z. & Clarke, A. K. (2002). Cutting edge of chloroplast proteolysis. *Trends Plant Sci.* **7**, 451–456.
22. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J. *et al.* (2004). Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J. Biol. Chem.* **279**, 4768–4781.
23. Stanne, T. M., Sjogren, L. L., Koussevitzky, S. & Clarke, A. K. (2009). Identification of new protein substrates for the chloroplast ATP-dependent Clp protease supports its constitutive role in Arabidopsis. *Biochem. J.* **417**, 257–268.
24. Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybailov, B., Olinares, P. D. & van Wijk, K. J. (2009). Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in Arabidopsis. *Plant Cell*, **21**, 1669–1692.
25. Zybailov, B., Friso, G., Kim, J., Rudella, A., Rodriguez, V. R., Asakura, Y. *et al.* (2009). Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol. Cell Proteomics*, **8**, 1789–1810.
26. McFadden, G. I., Reith, M. E., Munholland, J. & Lang-Unnasch, N. (1996). Plastid in human parasites. *Nature*, **381**, 482.
27. Gould, S. B., Waller, R. F., McFadden, G. I. *et al.* (2008). Plastid evolution. *Annu. Rev. Plant Biol.* **59**, 491–517.
28. Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J. *et al.* (2004). Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. *Nat. Rev. Microbiol.* **2**, 203–216.
29. Goodman, C. D., Su, V. & McFadden, G. I. (2007). The effects of anti-bacterials on the malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **152**, 181–191.
30. Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A. *et al.* (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155–172.
31. Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A. *et al.* (2007). Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Mol. Biochem. Parasitol.* **151**, 100–110.
32. Gangwar, D., Kalita, M. K., Gupta, D., Chauhan, V. S. & Mohmmed, A. (2009). A systematic classification of *Plasmodium falciparum* P-loop NTPases: structural and functional correlation. *Malar. J.* **8**, 69.
33. Rathore, S., Sinha, D., Asad, M., Bottcher, T., Afrin, F., Chauhan, V. S. *et al.* (2010). A cyanobacterial serine protease of *Plasmodium falciparum* is targeted to the apicoplast and plays an important role in its growth and development. *Mol. Microbiol*, **77**, 873–890.
34. Brotz-Oesterhelt, H., Beyer, D., Kroll, H. P., Endermann, R., Ladel, C., Schroeder, W. *et al.* (2005). Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat. Med.* **11**, 1082–1087.
35. Bottcher, T. & Sieber, S. A. (2008). Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *J. Am. Chem. Soc.* **130**, 14400–14401.
36. Jenuth, J. P. (2000). The NCBI. Publicly available tools and resources on the Web. *Methods Mol. Biol.* **132**, 301–312.
37. Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B. *et al.* (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543.
38. The UniProt Consortium (2010). The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.* **38**, D142–D148.
39. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, **278**, 631–637.
40. Ramasamy, G., Gupta, D., Mohmmed, A. & Chauhan, V. S. (2007). Characterization and localization of *Plasmodium falciparum* homolog of prokaryotic ClpQ/HslV protease. *Mol. Biochem. Parasitol.* **152**, 139–148.
41. Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J. S. *et al.* (1990). Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. *Proc. Natl Acad. Sci. USA*, **87**, 3513–3517.
42. Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M. & Tsai, F. T. (2003). The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell*, **115**, 229–240.
43. Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.

44. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995). Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci. USA*, **92**, 8259–8263.
45. Kim, Y. I., Levchenko, I., Fraczkowska, K., Woodruff, R. V., Sauer, R. T. & Baker, T. A. (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat. Struct. Biol.* **8**, 230–233.
46. Kim, D. Y. & Kim, K. K. (2003). Crystal structure of ClpX molecular chaperone from Helicobacter pylori. *J. Biol. Chem.* **278**, 50664–50670.
47. Tonkin, C. J., Kalanon, M. & McFadden, G. I. (2008). Protein targeting to the malaria parasite plastid. *Traffic*, **9**, 166–175.
48. Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N. *et al.* (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **95**, 12352–12357.
49. Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* **19**, 1794–1802.
50. van Dooren, G. G., Su, V., D'Ombrain, M. C. & McFadden, G. I. (2002). Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J. Biol. Chem.* **277**, 23612–23619.
51. Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S. *et al.* (2003). Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. *Science*, **299**, 705–708.
52. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I. & Schneider, G. (2001). Deciphering apicoplast targeting signals–feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene*, **280**, 19–26.
53. de Koning-Ward, T. F., Gilson, P. R., Boddey, J. A., Rug, M., Smith, B. J., Papenfuss, A. T. *et al.* (2009). A newly discovered protein export machine in malaria parasites. *Nature*, **459**, 945–949.
54. Kocsis, E., Cerritelli, M. E., Trus, B. L., Cheng, N. & Steven, A. C. (1995). Improved methods for determination of rotational symmetries in macromolecules. *Ultramicroscopy*, **60**, 219–228.
55. Ortega, J., Singh, S. K., Ishikawa, T., Maurizi, M. R. & Steven, A. C. (2000). Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol. Cell*, **6**, 1515–1521.
56. Wang, J., Hartling, J. A. & Flanagan, J. M. (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell*, **91**, 447–456.
57. Kim, D. Y. & Kim, K. K. (2008). The structural basis for the activation and peptide recognition of bacterial ClpP. *J. Mol. Biol.* **379**, 760–771.
58. Kang, S. G., Dimitrova, M. N., Ortega, J., Ginsburg, A. & Maurizi, M. R. (2005). Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX. *J. Biol. Chem.* **280**, 35424–35432.
59. Tan, S. (2001). A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. *Protein Expr. Purif.* **21**, 224–234.
60. Maurizi, M. R., Thompson, M. W., Singh, S. K. & Kim, S. H. (1994). Endopeptidase Clp: ATP-dependent Clp protease from *Escherichia coli*. *Methods Enzymol.* **244**, 314–331.
61. Jennings, L. D., Lun, D. S., Medard, M. & Licht, S. (2008). ClpP hydrolyzes a protein substrate processively in the absence of the ClpA ATPase: mechanistic studies of ATP-independent proteolysis. *Biochemistry*, **47**, 11536–11546.
62. Bewley, M. C., Graziano, V., Griffin, K. & Flanagan, J. M. (2009). Turned on for degradation: ATPase-independent degradation by ClpP. *J. Struct. Biol.* **165**, 118–125.
63. Gribun, A., Kimber, M. S., Ching, R., Sprangers, R., Fiebig, K. M. & Houry, W. A. (2005). The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J. Biol. Chem.* **280**, 16185–16196.
64. Bewley, M. C., Graziano, V., Griffin, K. & Flanagan, J. M. (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J. Struct. Biol.* **153**, 113–128.
65. Szyk, A. & Maurizi, M. R. (2006). Crystal structure at 1.9 Å of *E. coli* ClpP with a peptide covalently bound at the active site. *J. Struct. Biol.* **156**, 165–174.
66. Lin, W., Chan, M. & Sim, T. S. (2009). Atypical caseinolytic protease homolog from *Plasmodium falciparum* possesses unusual substrate preference and a functional nuclear localization signal. *Parasitol. Res.* **105**, 1715–1722.
67. Singer, G. A. & Hickey, D. A. (2000). Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol. Biol. Evol.* **17**, 1581–1588.
68. Mogk, A., Schmidt, R. & Bukau, B. (2007). The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol.* **17**, 165–172.
69. Dougan, D. A., Truscott, K. N. & Zeth, K. (2010). The bacterial N-end rule pathway: expect the unexpected. *Mol Microbiol.* **75**, 545–558.
70. Ingvarsson, H., Mate, M. J., Hogbom, M., Portnoi, D., Benaroudj, N., Alzari, P. M. *et al.* (2007). Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of *Mycobacterium tuberculosis* ClpP1. *Acta Crystallogr. D*, **63**, 249–259.
71. Kang, S. G., Maurizi, M. R., Thompson, M., Mueser, T. & Ahvazi, B. (2004). Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. *J. Struct. Biol.* **148**, 338–352.
72. Lee, B. G., Park, E. Y., Lee, K. E., Jeon, H., Sung, K. H., Paulsen, H. *et al.* (2010). Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat. Struct. Mol. Biol.* **17**, 471–478.
73. Kimber, M. S., Yu, A. Y. H., Borg, M., Leung, E., Chan, H. S. & Houry, W. A. (2010). Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. *Structure*, **18**, 798–808.
74. Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. (2005). Quantitative NMR

spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl Acad. Sci. USA*, **102**, 16678–16683.
75. van Dooren, G. G., Marti, M., Tonkin, C. J., Stimmler, L. M., Cowman, A. F. & McFadden, G. I. (2005). Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Mol. Microbiol.* **57**, 405–419.
76. Kalanon, M., Tonkin, C. J. & McFadden, G. I. (2009). Characterization of two putative protein translocation components in the apicoplast of *Plasmodium falciparum*. *Eukaryot Cell*, **8**, 1146–1154.
77. Tonkin, C. J., van Dooren, G. G., Spurck, T. P., Struck, N. S., Good, R. T., Handman, E. *et al.* (2004). Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol.* **137**, 13–21.
78. Trager, W. & Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science*, **193**, 673–675.
79. Mullin, K. A., Lim, L., Ralph, S. A., Spurck, T. P., Handman, E. & McFadden, G. I. (2006). Membrane transporters in the relict plastid of malaria parasites. *Proc. Natl Acad. Sci. USA*, **103**, 9572–9577.
80. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97.
81. Scheres, S. H., Nunez-Ramirez, R., Sorzano, C. O., Carazo, J. M. & Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* **3**, 977–990.
82. Gribun, A., Cheung, K. L., Huen, J., Ortega, J. & Houry, W. A. (2008). Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. *J. Mol. Biol.* **376**, 1320–1333.
83. Hainfeld, J. F., Wall, J. S. & Desmond, E. J. (1982). A small computer system for micrograph analysis. *Ultramicroscopy*, **8**, 263–270.
84. Otwinowski, Z. & Minor, W. (1997). Macromolecular crystallography. In *Methods in Enzymology* (Carter, C. W., Jr & Sweet, R. M., eds), *Methods in Enzymology*, vol. 276, pp. 307–326, Academic Press, New York.
85. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
86. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
87. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
88. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
89. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619.
90. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
91. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.