Title

Structural mapping of the ClpB ATPases of Plasmodium falciparum: Targeting protein folding and secretion for antimalarial drug design

Journal

Protein Science, 24(9)

ISSN

0961-8368

Authors

AhYoung, Andrew P
Koehl, Antoine
Cascio, Duilio
et al.

Publication Date

2015-09-01

DOI

10.1002/pro.2739

Peer reviewed

THE PROTEIN SOCIETY

Structural mapping of the ClpB ATPases of *Plasmodium falciparum*: Targeting protein folding and secretion for antimalarial drug design

Andrew P. AhYoung,[1] Antoine Koehl,[1] Duilio Cascio,[2] and Pascal F. Egea[1,3]*

[1]Department of Biological Chemistry, David Geffen School of Medicine, UCLA, Los Angeles, California
[2]Department of Energy Institute for Genomics and Proteomics, UCLA, Los Angeles, California
[3]Molecular Biology Institute, UCLA, Los Angeles, California

Received 19 May 2015; Accepted 26 June 2015
DOI: 10.1002/pro.2739
Published online 00 Month 2015 proteinscience.org

Abstract: Caseinolytic chaperones and proteases (Clp) belong to the AAA+ protein superfamily and are part of the protein quality control machinery in cells. The eukaryotic parasite *Plasmodium falciparum*, the causative agent of malaria, has evolved an elaborate network of Clp proteins including two distinct ClpB ATPases. ClpB1 and ClpB2 are involved in different aspects of parasitic proteostasis. ClpB1 is present in the apicoplast, a parasite-specific and plastid-like organelle hosting various metabolic pathways necessary for parasite growth. ClpB2 localizes to the parasitophorous vacuole membrane where it drives protein export as core subunit of a parasite-derived protein secretion complex, the *Plasmodium* Translocon of Exported proteins (PTEX); this process is central to parasite virulence and survival in the human host. The functional associations of these two chaperones with parasite-specific metabolism and protein secretion make them prime drug targets. ClpB proteins function as unfoldases and disaggregases and share a common architecture consisting of four domains—a variable N-terminal domain that binds different protein substrates, followed by two highly conserved catalytic ATPase domains, and a C-terminal domain. Here, we report and compare the first crystal structures of the N terminal domains of ClpB1 and ClpB2 from *Plasmodium* and analyze their molecular surfaces. Solution scattering analysis of the N domain of ClpB2 shows that the average solution conformation is similar to the crystalline structure. These

Abbreviations: Pfal, *Plasmodium falciparum*; AAA+, ATPase associated with diverse activities; Clp, caseinolytic protease; PTEX, Plasmodium translocon of exported proteins; PV, parasitophorous vacuole; PVM, parasitophorous vacuole membrane; GFP, green fluorescent protein; SAXS, small angle X-ray scattering; RMSD, root-mean-square deviation; ADP, atomic displacement parameters; SEC, size exclusion chromatography; Ec, Escherichia coli; Tth, Thermus thermophilus; Bsu, Bacillus subtilis; Vch, Vibrio cholera.

Additional Supporting Information may be found in the online version of this article.

Andrew P. AhYoung and Antoine Koehl contributed equally to this work.

Grant sponsor: National Center for Research Resources; Grant number: 5P41RR015301-10; Grant sponsor: National Institute of General Medical Sciences; Grant number: 8P41GM103403-10; Grant sponsor: DOE Integrated Diffraction Analysis (IDAT); Grant number: DE-AC02-05CH11231; Grant sponsor: National Institutes of Health; Grant sponsors: UCLA Geffen School of Medicine, the UCLA Scholars in Translational Medicine Program Award, the Alexander and Renée Kolin Endowed Professorship in Molecular Biology and Biophysics and the National Center for Advancing Translational Sciences UCLA CTSI; Grant number: UL1TR000124 (to P.F.E.); Grant sponsor: Gates Millennium and Dissertation of the Year fellowships (to A.P.A.Y.).

*Correspondence to: Pascal F. Egea, University of California Los Angeles, David Geffen School of Medicine, Department of Biological Chemistry, Boyer Hall 356, 611 Charles E. Young Drive East, Los Angeles, CA 90095. E-mail: pegea@mednet.ucla.edu

structures represent the first step towards the characterization of these two malarial chaperones and the reconstitution of the entire PTEX to aid structure-based design of novel anti-malarial drugs.

Keywords: malaria; ClpB AAA+ proteins; *Plasmodium* translocon of exported proteins; protein export and trafficking; chaperone; apicoplast; drug target; antimalarial design; SAXS; hybrid methods

Introduction

Malaria is a mosquito-borne parasitosis caused by the obligate intracellular protozoan *Plasmodium*. Efforts to eradicate this devastating disease include drug-based therapies, mosquito control, and development of a vaccine. Emergence of drug resistance to existing anti-malarial therapies among parasites constitutes a serious threat to the current disease control efforts. *Plasmodium falciparum* (*Pfal*) is responsible for its most severe and fatal form. Modern research on malaria seeks to identify and understand biological processes that are unique to the parasite. Once characterized, such processes can be targeted for drug design.

All living cells have evolved sophisticated mechanisms of protein homeostasis. Proteostasis includes the elimination of misfolded proteins or otherwise abnormal proteins, maintenance of aminoacid pools and the refolding/reactivation of salvageable proteins in cells subjected to stresses such as heat shock or starvation. Caseinolytic proteases and chaperones (Clp) constitute one large group of proteins involved in protein quality control; they belong to the superfamily of AAA+ proteins (ATPases associated with diverse cellular activities) and are found in bacteria, fungi, but also in the mitochondria of eukaryotes or the chloroplast of plants.[1]

They are several subgroups of Clp chaperones. ClpA, ClpC and ClpX belong to the Clp/HSP100 family of AAA+ proteins[2] that also includes the molecular chaperones ClpB/HSP104.[3] The hexameric chaperones ClpA, ClpC and ClpX bind and unfold specific substrate proteins before translocating the unfolded polypeptide to the protease ClpP for its final degradation.[4,5] In contrast, ClpB/HSP104 proteins are essential for the survival of bacteria, fungi and plants where they act as 'disaggregases' assisting resolubilization and reactivation of aggregated proteins[6] in response to stress. These chaperones support infectivity and host survival in a number of pathogens.[7–9] Bacterial Clp proteases are *bona fide* targets for the screening and development of novel antibacterials.[10]

Sequence analyses indicate that the genome of *Plasmodium falciparum* encodes six Clp proteins.[11] ClpP and ClpR[12] are two proteases while ClpB1, ClpB2/HSP101, ClpC and ClpM are ATPases. Interestingly, despite the plethora of Clp AAA+ proteins identified in *Pfal*, no ClpA has been detected. *Plasmodium* belongs to the phylum of Apicomplexa, characterized by the presence of a peculiar organelle, the apicoplast. Apicomplexans are thought to have originated from Dinoflaggellates, a large group of photosynthetic protozoa. As such, the apicoplast is distantly related to the chloroplast of higher plants. Some 500 proteins of *Plasmodium* have been predicted to localize to this organelle where several prokaryotic biochemical pathways have been annotated.[13] Chemical rescue experiments performed on malaria parasites lacking an apicoplast have helped define the function of this organelle and suggest that isoprenoid precursor biosynthesis is the main essential function of the apicoplast during blood stage growth.[14] Other crucial biosynthetic pathways include fatty acid, heme and aminoacid synthesis. Because the apicoplast is vital to parasite survival[15] and is an organelle unique to apicomplexans, it provides an enticing target for the design of novel antimalarial drugs aimed at disrupting biological pathways essential to the parasite.

In *Plasmodium*, all Clp proteins, except ClpB2/HSP101, are localized in the apicoplast [Fig. 1(A)]. ClpB2/HSP101 instead localizes exclusively to the encasing parasitophorous vacuole (PV) where it functions as a core subunit of a protein secretion complex, the *Plasmodium* Translocon of Exported proteins (PTEX) [Fig. 1(B)].[16] The intracellular survival of *Plasmodium* within the infected red blood cells is dependent on its ability to export several hundreds of its proteins (~7% of its proteome) that hijack and remodel the infected host cells to support its virulence and parasitic lifestyle.[17–20] Parasitic proteins destined for export through the PTEX harbor a signal sequence to reach the vacuole by entering the endoplasmic reticulum (the regular secretory pathway) *and* a specific *Plasmodium* Export Element sequence (PEXEL)[21–23] for their vacuolar translocation. Thus, to access the host erythrocyte cytosol, exported proteins must cross two membranes, first the parasite plasma membrane, and then the PV membrane (PVM) [Fig. 1(A,B)].

PTEX is a large membrane-associated complex composed of five subunits: EXP2, ClpB2/HSP101, PTEX150, TRX2 and PTEX88 [Fig. 1(B)]. A stable detergent-resistant core composed of subunits EXP2, PTEX150 and ClpB2/HSP101 has been characterized.[24] The PVM-associated Exported protein-2 (EXP2)[25,26] is a possible candidate for the transmembrane protein-conducting conduit and may structurally resemble the bacterial pore-forming

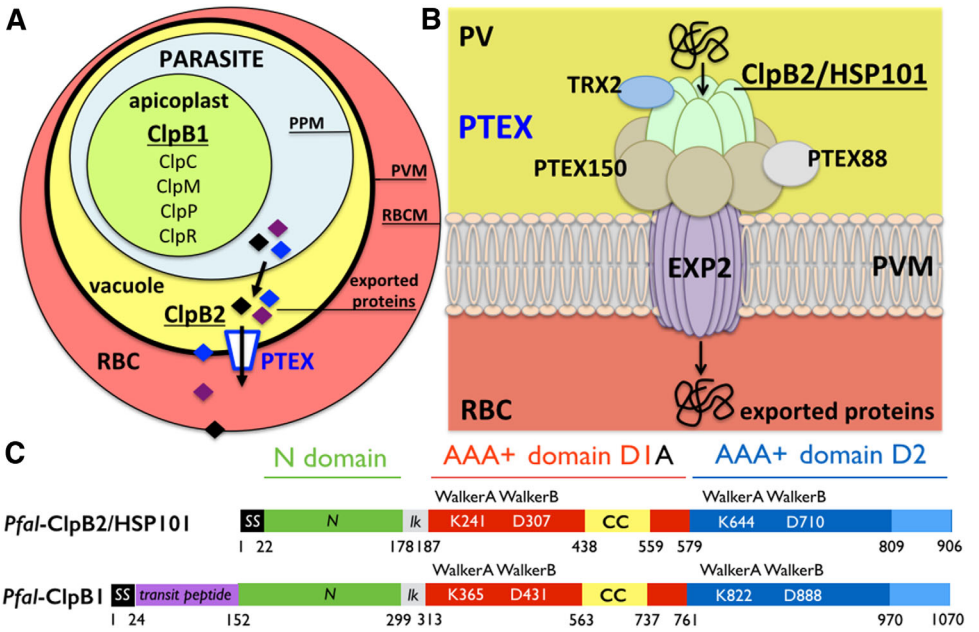

Figure 1. The Clp proteases and chaperones in *Plasmodium* and the *Plasmodium* translocon of exported proteins. (A) Simplified diagram showing a red blood cell (RBC) infected by a parasite encased in the parasitophorous vacuole. The nomenclature and localization of all Plasmodium Clp chaperones and proteases are show. There are two ClpB proteins, ClpB1 and ClpB2/HSP101. All Clp proteins are localized in the apicoplast except the vacuolar ClpB2/HSP101. (B) ClpB2/HSP101 is a subunit of the parasite-derived Plasmodium translocon of exported proteins (PTEX). *Plasmodium* proteins are translocated across the parasitophorous vacuole membrane (PVM) into the red blood cell cytoplasm through a translocation pore composed of Exported protein-2 (EXP2). Prior to their transport, exported proteins are unfolded by the hexameric AAA+ protein ClpB2/HSP101. For some cargos, unfolding also necessitates reduction of disulfide bonds by the accessory subunit thioredoxin-2 (TRX2). The other subunits, PTEX150 and PTEX88, have unknown functions. The arrows indicate the directionality of transport from the parasitophorous vacuole (yellow) towards the erythrocyte cytoplasm (red). The relative positions of the different subunits within the complex are unknown. (C) Domain organization of the two ClpB proteins present in *Plasmodium falciparum*: the apicoplastic ClpB1 and the vacuolar ClpB2/HSP101. The signal sequence (*SS* in black) targets proteins to the general secretory pathway and the transit peptide (in magenta) is a specific signal that specifies subcellular localization to the apicoplast. The N-terminal domain (in green) is linked to the two catalytic nucleotide-binding domains (NBD) D1 (in red) and D2 (in blue) by a short basic linker (*lk* in grey). The D1 domains of ClpB and ClpC chaperones are characterized by the presence of a coiled-coil insertion, the middle domain also named "arm" (CC in yellow). The D2 domains usually contain a C-terminal module (in light blue). Each NBD contains two signature Walker (A and B) motifs.

cytolytic α-helical toxin hemolysins.[27–29] In *Plasmodium*, ClpB2/HSP101 actively drives the export process by harnessing the energy from ATP hydrolysis to unfold and thread the diverse cargos through the trans-membrane conduit EXP2. While ClpB2/HSP101 seems to be the primary 'engine' driving protein export, the malarial thioredoxin-2 (TRX2) auxiliary protein assists protein unfolding by reducing the disulfide bonds present in some cargo proteins, thus facilitating the ATP-driven export process. Protein unfolding is required for export across the PVM as it was shown that proteins must pass into the erythrocyte cytosol in an unfolded state.[30] Molecular chaperones appear to play important roles in keeping parasite proteins in a 'translocation competent' state prior to crossing the PVM.

Attempts to generate gene complete knockouts of any of the three PTEX components EXP2, PTEX150 and ClpB2/HSP101, initially failed[16] suggesting their essential functions. Mutant parasites lacking TRX2 or PTEX88 are severely impaired with considerably slower rates of development during the blood stage.[31] These studies have led to the conclusions that while EXP2, HSP101, and PTEX150 play central roles during the blood infection in *Plasmodium*, PTEX88 and TRX2 are auxiliary yet important regulators of PTEX-mediated protein export and required for maintaining normal blood-stage growth.[32] The integrity of the PTEX complex is required for efficient protein export. Ablation or inactivation of HSP101/ClpB2 in parasites results in the nearly complete block in export with substrates accumulating in the vacuole;[33] furthermore this blockade affects all classes of exported proteins, whether they harbor a PEXEL motif or not.[34] As a central and parasite-specific portal, PTEX appears as a weak point in the parasitic life cycle and hence an attractive drug target.

Like all class 1 Clp AAA+ proteins, *Pfal*- ClpB1 and ClpB2/HSP101 are modular proteins composed of a N-terminal domain involved in the binding of substrate proteins, followed by the two ATPase catalytic domains (D1 and D2) responsible for the

ATP-driven unfolding of polypeptides and a short C-terminal domain [Fig. 1(C)]. They likely function as homo-hexamers where the N-domains are connected to the D1D2 ATPase catalytic core via flexible linkers and are structurally mobile, which may facilitate binding to target proteins[35–37] and enhance chaperone activity.[38] Switching N-domains between different HSP100 proteins alters substrate specificity accordingly, indicating that N-domains function as independent units in substrate selection.[39,40] N-domains can contact substrates directly[41,42] or via adaptor proteins.[43] The hexameric association state and ATPase-dependent chaperone activity of the apicoplastic ClpB1 have been demonstrated *in vitro*; however nothing is known about its roles *in vivo*; it might be involved in maintaining a functional apicoplastic proteome during red blood cell infection.[44]

To aid design novel anti-malarials using computational tools and combinatorial screening of libraries, knowledge of the three-dimensional structure of the targeted component is a prerequisite. For ClpB1 and ClpB2, identification of the protein surfaces and structural features involved in the essential steps of cargo recognition/delivery, unfolding and/or translocation can help define and identify novel pharmacophores. PTEX appears a prime drug target and there is considerable interest in targeting the components involved in cargo unfolding such as ClpB2/HSP101 and the thioredoxin TRX2 to block protein export. The structures of malarial TRX2[45,46] provided valuable templates for rational structure-based drug design. Our current efforts focus on the characterization of ClpB2/HSP101 and EXP2 proteins at the core of PTEX. To this end, we describe here the first crystal structures of the N-terminal domain of ClpB1 and ClpB2/HSP101 proteins from *Plasmodium* and also characterize the conformation of its vacuolar ATPase N domain using solution small-angle X-ray scattering (SAXS). We analyze and compare the surface properties of the N domains from the two parasitic ClpB chaperones in terms of substrate and partner protein binding sites.

Results

Structures of the N-terminal domains of the ClpB1 and ClpB2/hsp101 ATPases from the eukaryotic pathogen plasmodium falciparum

We solved the crystal structures of the N domains from ClpB1 and ClpB2/HSP101 of *Pfal* at resolutions ranging from 1.7 to 2.0 Å by molecular replacement using the structures of the N domain from the ClpB proteins from *Thermus thermophilus*[47] and *Escherichia coli*[48] as best search models (Table I and Material and Methods). Despite relatively weak sequence identities ranging from ~12 to 24%, the corresponding N-domains of ClpA, ClpB, ClpC and ClpV from bacteria such as *Escherichia*, *Thermus*, *Bacillus* and *Vibrio* can be superimposed with a root mean square deviation (RMSD) ranging from ~1.5 to 2.4 Å for corresponding Cα atoms [Fig. 2(A) and Supporting Information Table S1]. The fold of the N domain is composed of eight α-helices (α1-α8) folding into a compact globular domain characteristic of all Clp chaperones belonging to the A, B and C subgroups [Fig. 2(B)]. It can be divided into two four α-helix bundles (α1-α4 and α5-α8) connected by a ~15 residue-long loop. The excellent quality of the electron density (Supporting Information Figure S1A) enabled us to trace the entire domain of N-ClpB2/HSP101 in each of the two crystal structures reported here; the loop of N-ClpB1 showed some disorder in one of the two chains constituting the asymmetric unit.

We initially tried to solve the structure of N-ClpB1 by itself (Material and Methods). Bacterial expression and purification yielded homogenous protein and crystals of N-ClpB1 appeared in a single condition. However they were twinned and although a molecular replacement solution could be identified, we sought to find a way to prevent twinning. To overcome this problem, we applied carrier-driven crystallization and selected Green Fluorescent Protein (GFP)[49] over Maltose Binding Protein (MBP)[50] as fusion partner. We were able to crystallize a fusion protein between N-ClpB1 and a superfolder variant of GFP[51] (Material and Methods). The resulting crystals were not twinned and diffracted to 2.0 Å resolution with two copies of the fusion protein in the asymmetric unit [Fig. 2(C)].

Superposition of the two copies using the GFP as reference structure reveals that the two ClpB1 adopt strikingly different orientations within the asymmetric unit [Supporting Information Fig. S2(A)]. In one copy (chain B), the C-terminal helix α8 of ClpB1 and the N-terminal α-helix of GFP align to form a single continuous helix including the $G_{149}G_{150}$ sequence linking the C-terminus of N-ClpB1 to the N-terminus of GFP. In the second copy (chain A), this linker forms a sharp break/turn resulting in a 90° angle between the two helices. This results in a local breakage of the two-fold non-crystallographic symmetry that relates the two GFP units [Fig. 2(C)]. Although the linker displays remarkable conformational plasticity, it is well resolved in the experimental electron density maps [Supporting Information Fig. S1(B)]. Most importantly, the two ClpB1 N domain copies are superimposable, thus showing that fusion to the GFP using a very short linker did not distort the structure of the N-terminal domain [Supporting Information Figure S2(B)]. These are the first structures of N domains from eukaryotic ClpBs and in particular of a fragment of the vacuolar ATPase ClpB2/HSP101, an essential core subunit of the PTEX.

Table I. *X-Ray Data Collection and Structure Refinement Statistics*

Data set	APS 060612 24-ID-E	APS 120912 24-ID-C	APS 10162014 24-ID-C
Protein	*Pfal* N ClpB2/HSP101	*Pfal* N ClpB2/HSP101	*Pfal* N ClpB1-sfGFP
Data collection statistics	PDB ID 4IOD	PDB ID 4IRF	PDB ID 4XBI
Wavelength	0.97918 Å	0.97918 Å	0.97918 Å
Space group	$C222_1$	$P22_12_1$	$P6_5$
	$a = 89.1$ Å $b = 141.88$ Å $c = 91.8$ Å	$a = 31.8$ Å $b = 92.1$ Å $c = 97.0$ Å	$a = 127.5$ Å, $c = 92.6$ Å
AU content	Three molecules	Two molecules	Two molecules
Solvent content	58%	45%	52%
Resolution (last shell)	44.6–1.80 Å (1.85–1.80 Å)	66–1.65 Å (1.67–1.65 Å)	50–2.01 Å (2.07–2.01 Å)
Unique reflections	53,488 (3,809)	35,522 (2,439)	55,948 (3,785)
Completeness	99.7% (96.7%)	99.1% (94.9%)	98.6% (90.4%)
$I/\sigma(I)$	21.6 (3.2)	15.0 (2.7)	10.9 (2.7)
Redundancy	9.9 (8.9)	8.2 (8.1)	5.0 (4.3)
R_{sym}	7.1% (74.3%)	6.3% (47.8%)	7.9% (48.8%)
R_{meas}	7.5% (78.6%)	6.8% (50.9%)	8.9% (55.4%)
CC(1/2)	99.9% (82.6%)	99.8% (96.0%)	99.6% (85.6%)
Refinement statistics			
Resolution	44.6–1.8 Å	66.78–1.65 Å	70.95–2.01 Å
Reflections	53,476	35,144	55,908
Work set/test set	48,142/5,334	31,631/3,513	53,104/2,804
R_{free}/R_{cryst}	21.2%/17.0%	23.4%/21.4%	20.1%/16.4%
B_{wilson}	21.2 $Å^2$	28.2 $Å^2$	36.5 $Å^2$
Protein atoms, ADP	3,524, 29.2 $Å^2$	2,249, 35.6 $Å^2$	5,832, 41.1 $Å^2$
Solvent atoms, ADP	371, 34.6 $Å^2$	223, 43.0 $Å^2$	383, 44.9 $Å^2$
Others atoms, ADP	3 sulfates, 53.6 $Å^2$	–	5 sulfates, 79.0 $Å^2$
RMSD bonds	0.014 Å	0.018 Å	0.007 Å
RMSD angles	1.209°	1.904°	1.035°
Ramachandran analysis			
Allowed regions	99.6%	99.0%	98.2%
Generously allowed	0.4%	0.7%	1.6%
Outliers	0%	0.3%	0.2%

$R_{sym} = \sum_{hkl}\sum_i |I_{hkl,I} - \langle I_{hkl,i}\rangle| / \sum_{hkl}\sum_i |I_{hkl,i}|$ where $\langle I_{hkl,i}\rangle$ is the average intensity of the multiple hkl, i observations for symmetry-related reflections.
R_{meas} is the redundancy independent R-factor.[70]
CC(1/2) percentage of correlation between intensities from random half-datasets.[71]
$R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$. F_{obs} and F_{calc} are observed and calculated structure factors, R_{free} is calculated from a set of randomly chosen reflections, and R_{cryst} is calculated over the remaining reflections.
ADP is the atomic displacement parameter.

Surface properties of the two ClpB N domain proteins in plasmodium

Within the N domains, the two four-helix bundles are related to each other by a pseudo two-fold symmetry axis [Fig. 3(A)]. This topology results in the most salient feature of these structures: a solvent-accessible grove that is delineated by helices α1/α2 and their pseudo-symmetry related counterparts α5/α6. This surface is composed of hydrophobic residues [Fig. 3(B)] and can be referred to as a "hydrophobic patch." The ~15 residue-long loops connecting the two halves are slightly basic in contrast with the so-called "acidic loop" present in all N domains of ClpA ATPases;[52] it is well ordered and adopts the same conformation in the two different crystal forms of N-ClpB2/HSP101 we obtained (Material and Methods).

In *Plasmodium* ClpB2/HSP101, the corresponding surface is wider and flatter and seems more suited to the promiscuous binding of unfolded and/or hydrophobic polypeptides. In ClpA, the same hydrophobic patch was shown to bind to peptides with low affinity and loose specificity.[52] This particular surface also overlaps partially with the cargo-specific binding site revealed in the structure of the ClpV-VipB peptide[53] (Supporting Information Fig. S3). However, the ClpV protein N domain contains a short extra N-terminal helix appended to the eight α-helical bundle common to all N domains of ClpA, ClpB and ClpC HSP100 AAA+ proteins; this architecture results in an hydrophobic groove that specifically accommodates the α-helical VipB peptide (Supporting Information Fig. S3).

The specificity of individual AAA+ proteins is achieved by their direct or adaptor-mediated association with their proteinaceous substrates. ClpA and ClpC ATPases have been shown to recognize some of their substrates through adaptor proteins that dock to their N-terminal domains and deliver their cargo for subsequent unfolding by the D1 and D2 catalytic domains. Only a few AAA+ Clp proteins have been crystallized with their adaptor or their specific cargo proteins. These include the N-ClpA/ClpS adaptor

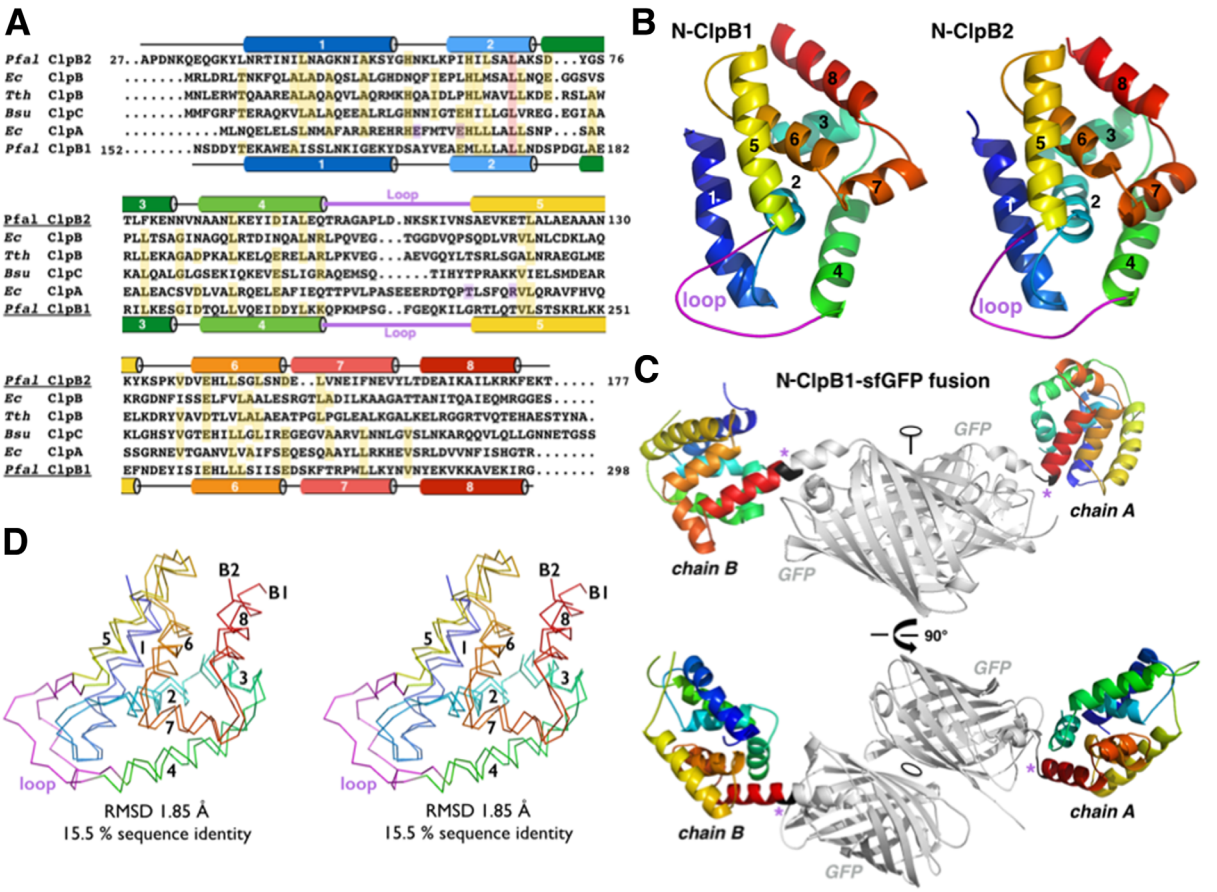

Figure 2. General architecture of the N-terminal cargo-binding domain of the *Plasmodium* AAA+ proteins ClpB1 and ClpB2/HSP101. (A) Structure-based sequence alignments of N domains from ClpA, ClpB and ClpC AAA+ proteins of known structures; the position of the secondary structure elements correspond to our ClpB2/HSP101 and ClpB1 N domain structures. The N domain structures from *E. coli* ClpA (PDB 1K6K), *E. coli* ClpB (PDB 1KHY), *T. thermophilus* ClpB (PDB 1QVR) and *B. subtilis* ClpC (PDB 2Y1Q) were used. (B) Crystal structures of the two ClpB N terminal domains from *Plasmodium falciparum*. Secondary structure elements are labeled and colored using a rainbow pattern for the N domain. (C) Two views showing the two chains (A and B) of the N domain of ClpB1 fused to superfolder GFP (in grey) present in the asymmetric unit. The magenta star highlights the different conformations adopted at the Gly-Gly linkage (in black) connecting the N domain to the sfGFP used to assist crystallization. The two-fold non-crystallographic symmetry axis that relates the two GFP chains but not the two N domains is indicated. (D) Stereo view of the two malarial ClpB N domains superposed.

complex from *Escherichia*,[54,55] the ClpC/MecA adaptor complex[56] from *Bacillus* (Fig. 4) and the N-ClpV/VipB cargo peptide complex from *Vibrio*.[53] ClpA, ClpB, ClpC, and ClpV share structurally similar N-terminal domains. Although structurally unrelated, the bacterial adaptors ClpS and MecA primarily use a single α-helix to interact with a common binding site at the surface of the N domains of ClpA and ClpC. Strikingly, in both cases a glutamate residue from the adaptor (E64 in ClpS and E184 in MecA) interacts with this pocket of the Clp protein. This "hot-spot" for adaptor binding is delineated by the connecting loop and helices α2, α4, α5, and α7 [Fig. 4(A)]. We compared the electrostatic features of the N domains from Clp proteins representative of the three subgroups (*i.e.,* ClpA/B/C): the bacterial ClpA and ClpC and our two malarial ClpB structures. For each of these proteins representing three subclasses of Clp proteins, a shallow pocket is present. However, the electrostatic potential of malarial ClpB2/HSP101 shows a remarkable charge distribution asymmetry on this specific face of the protein [Fig. 4(B)] that sets it apart from its homologues.

Unlike ClpA and ClpC, ClpB proteins do not seem to require adaptors to bind their substrates. The two ClpB chaperones from *Plasmodium* are localized in two distinct cellular compartments and thus interact with different subsets of substrates, in different subcellular contexts. ClpB1 most likely functions as a 'house-keeping' chaperone in the apicoplast. On the other hand, the vacuolar ClpB2/HSP101 is a quite specialized chaperone involved in the specific export across the vacuolar membrane of a subset of malarial proteins; the molecular mechanisms governing cargo selection and triage are not presently understood and it is likely that some other PTEX subunit, namely the core subunit PTEX150 that tightly associates with ClpB2/HSP101 and EXP2,

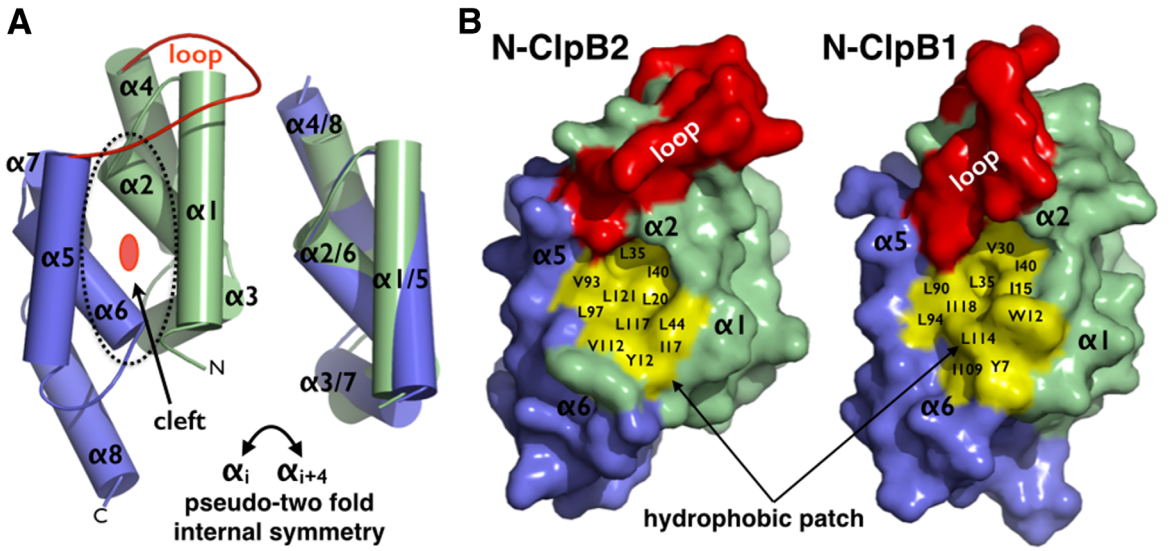

Figure 3. Pseudo two-fold internal symmetry creates a hydrophobic patch at the surface of the N terminal domains of the two ClpB chaperones from *Plasmodium*. (A) Two four-α-helical bundles related by a pseudo two-fold axis perpendicular to the plane of the figure (red symbol) constitute the N domain (left). The first bundle (α1– α4) is colored in green; the second (α5–α8) is colored in blue. This arrangement, shown here for N-ClpB2 creates a cleft. The two bundles are shown superposed (right). (B) Solvent accessible surface representation highlighting the hydrophobic nature of the cleft present in N-ClpB2 (left) and N-ClpB1 (right). Residues delineating these putative binding sites are labeled. The same orientation is used in A and B.

might play this role of delivering the diverse cargos to the chaperone for their translocation.

Conformation of the N-terminal domain of the plasmodium ClpB2/hsp101 ATPase in solution

ClpA/B/C proteins function as homo-hexamers. Purified N-terminal domains of ClpB1 and ClpB2, together with the N-ClpB1-GFP fusion protein [Fig. 5(A)] were analyzed by size exclusion chromatography (SEC). The three proteins behaved as monomers [Fig. 5(B)]. We further characterized the association state and conformation of the N domain of ClpB2/HSP101 in solution using SAXS[57] (Material and Methods). Kratky plots indicate that the protein is folded in solution and is overall compact [Fig. 5(C)]. We derived the experimental values of radius of gyration (R_G) values by Guinier analysis of scattered intensities at very small angles [Table II and Fig. 5(D)]. Experimental scattering curves were fitted against theoretical curves calculated using our crystal structures [Fig. 5(E)]. The pair distance distribution function ($P(r)$) estimated by Fourier inversion of the experimental intensities was compared to those calculated from our crystal structures model and used to estimate D_{max} (the longest intra-molecular distance) and R_G values independent of the Guinier analysis [Table II and Fig. 5(F)]. The experimental R_G value of 16.3 Å measured for N-ClpB2/HSP101 is similar to the value calculated from the X-ray structure. The excellent agreement between theoretical and experimental scattering curves (as quantified by a discrepancy factor of $\chi^2 = 1.008$) and the overlap of the associated $P(r)$ distributions indicate that the average conformation sampled by the N domain of ClpB2/HSP101 in solution is similar to our crystal structures and the monomeric state observed by SEC. It also suggests that the N-terminal domain of ClpB2/HSP101 is a rather rigid structure where the loop adopts the conformation observed in the crystal structures and does not sample a wide range of conformations.

Conclusion

This study describes the first high-resolution crystal structures of the N-terminal domain from the two ClpB chaperones from the eukaryotic pathogen causing malaria, the parasite *Plasmodium falciparum*. The two *Plasmodium* ClpB chaperones are prime drug targets. Because of their ubiquitous distribution and the conserved architecture of their catalytic core, AAA+ proteins are challenging targets for the design of selective inhibitors. Nevertheless, such small molecules inhibitors have been described. A di-benzyl-quinazoline-diamine was the first reported reversible inhibitor for the p97/VCP ATPase[58] offering a promising new therapeutic avenue in the treatment of cancer. More recently, small compounds named ciliobrevins were characterized as specific inhibitors of the motor cytoplasmic dynein.[59] Disrupting cargo recognition, either through direct binding or via adaptors, might prove another viable strategy. Using these crystal structures as templates, future analyses of their protein surfaces might reveal potential binding sites that can be targeted for inhibitor design using fragment-based and computational screening of chemical libraries. The characterization of the other

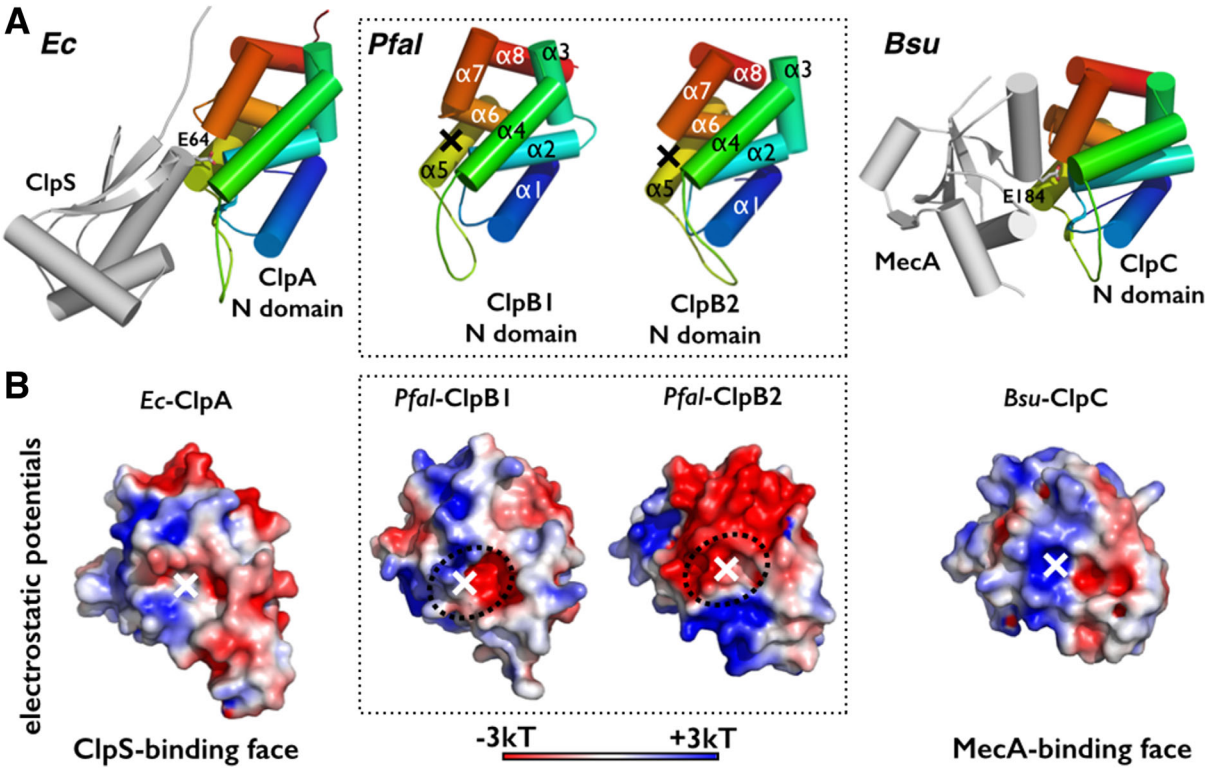

Figure 4. Modalities of interaction between N domains from different AAA+ proteins of the Clp subfamily and their know adaptors. (A) The structures of the *E. coli* ClpS/ClpA N domain (PDB 1LZW) and the *B. subtilis* MecA/ClpC N domain (PDB 2Y1R) complexes are displayed next to our *Plasmodium* ClpB1 and ClpB2/HSP101 N domain structures. In both complexes, the structurally unrelated adaptors use a conserved glutamate residue to interact with the N-domain of their cognate Clp ATPase. (B) Electrostatic potential properties as calculated using *APBS*[72] for the binding interfaces of ClpA and ClpC with their respective adaptors; the structurally equivalent surface in the ClpB1 and ClpB2/HSP101 N domains are shown in the same orientation. Red, blue and white surfaces correspond to negative, positive and neutral surface potential, respectively. The cross indicates the binding pockets accommodating residues E64 (in ClpS bound to ClpA) and E184 (in MecA bound to ClpC) contributed by the two adaptor proteins. The 'hot-spot' for adaptor binding is marked with an × symbol.

Plasmodium Clp ATPases and their yet to be discovered effectors or adaptors will also identify novel targets for anti-malarial design. For ClpB2/HSP101, this also represents a step towards the determination of the structure of the full-length ATPase through a hybrid approach combining SAXS, electron microscopy and X-ray crystallography. Upon reconstitution of the entire PTEX, our structures will facilitate future modeling, biochemical and, biophysical studies aimed at deciphering the mechanisms of cargo recognition, unfolding, and delivery by ClpB2/HSP101 to the predicted protein-translocating channel in the vacuolar membrane of *Plasmodium*.

Material and Methods

Protein expression and purification

A codon optimized synthetic gene (*DNA2.0 Inc.*) encoding full-length *Plasmodium falciparum* ClpB2/HSP101 (*Pfal*-ClpB2/HSP101: PF3D7_1116800) was used as starting material to construct several expression vectors. Two expression constructs encompassing residues A27-L188 (long-N101) or Q32-T177 (short-N101) were cloned into a pJexp401 vector (*DNA2.0 Inc.*) (for long-N101) or a pCDF vector (*Novagen*) (for short-N101) to express the corresponding proteins as fusions harboring a cleavable deca-histidine tag at their C-terminus. Both constructs lacked residues M1 through C26 corresponding to the signal sequence of *Pfal*-ClpB2/HSP101.

A codon optimized synthetic gene (*Invitrogen Life Technologies*) encoding residues G152-G298 corresponding to the N-terminal domain of *Plasmodium falciparum* ClpB1 (*Pfal*-ClpB1: PF3D7_0816600) was used as starting material to construct two expression vectors. Residues M1-I151 correspond to the signal sequence and the apicoplast targeting transit peptide. A first construct encoded the N-domain of ClpB1 cloned in the pET28b vector (Novagen) to be expressed as an N-terminal hexa-histidine-tagged cleavable fusion protein. A second construct encoding the N-ClpB1 gene fused to the superfolder GFP gene with a single G residue linker connecting residues G253 of ClpB1 and S2 of sfGFP; the N-ClpB1-sfGFP fusion protein contained a cleavable C-terminal octa-histidine tag.

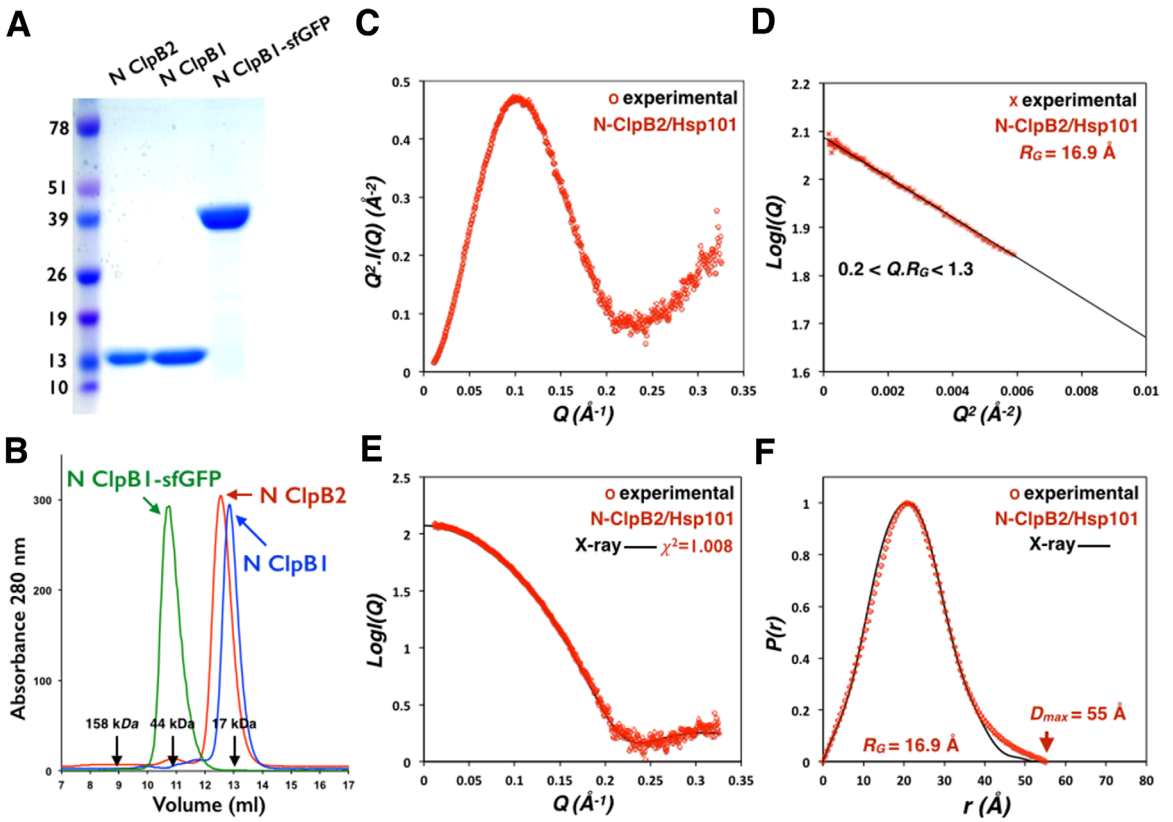

Figure 5. Homogeneity of the N domains of malarial ClpB1 and ClpB2 proteins purified from *E. coli* and solution analysis by SAXS of the *Plasmodium* ClpB2/HSP101 N domain. (A) Coomassie-stained SDS-PAGE analysis of N-ClpB1 (17.2 kDa), N-ClpB2 (16.8 kDa) and the N-ClpB1sfGFP fusion (43.2 kDa) proteins. (B) Size exclusion chromatography analysis of the three purified proteins used for structure determination on a Superdex S75 HR 10/30 column. The three proteins elute as monomers. Elution volumes for three molecular mass standards are indicated with arrows. (C) Kratky plot showing that *Pfal*-ClpB2/HSP101 N domain is folded in solution. (D) Guinier analysis. Guinier plot shown with the corresponding experimental value for the radius of gyration. The angular range (expressed in QR_G values) used for the regression is indicated. Guinier's law is applied in its most conservative form within a QR_G range not exceeding 1.3. (E) Scattering curve fit. The experimental curve (in color) is fitted against the curves calculated from the X-ray structures (in black) with *CRYSOL*, assuming a monomeric state of the N-domain in solution. The discrepancy factor for the fit is indicated. (F) Pair distance distribution functions. Experimental curve (in color) and curve calculated from the X-ray structures (in black) with *GNOM*. D_{max} and R_G experimental values inferred from the analysis are indicated.

All proteins were expressed using C43(DE3) *E. coli* cells grown in LB media at 37°C until they reached OD600 = 0.6 and protein expression was induced with 0.8 m*M* IPTG. Cells were harvested, washed in 150 m*M* KCl and 20 m*M* Tris pH = 7.8, and stored at −80°C until processing. Thawed cell pellets were resuspended in lysis buffer (500 m*M* NaCl, 20 m*M* Tris pH = 7.8, 15% glycerol) supplemented with 2.8 m*M* β-mercaptoethanol (β-ME), 0.2 m*M* of phenylmethylsulfonyl fluoride (PMSF), and 1 tablet of EDTA-free protease inhibitor cocktail (Roche). All subsequent steps were carried out at 4°C or on ice. Cells were disrupted by three passes through a C-3 Emulsiflex (*Avestin*) pressurized at 15,000 psi. The lysate was clarified by centrifugation at 25,000 g for 1 hr, after which the resulting supernatant corresponding to the total soluble extract was applied onto a gravity-flow column (*BioRad*) packed with 5-10 mL of Cobalt-NTA IMAC resin (*Qiagen*). Non-specifically bound bacterial proteins were subsequently washed from the column using Cobalt-A wash buffer (12.5 m*M* imidazole, 500 m*M* NaCl, 20 m*M* Tris pH = 7.8, 10% glycerol, 2.8 m*M* β-ME and 0.2 m*M* PMSF). The protein was eluted from the column with Cobalt-B buffer (125 m*M* imidazole, 500 m*M* NaCl, 20 m*M* Tris pH = 7.8, 10% glycerol, 2.8 m*M* β-ME, and 0.2 m*M* PMSF). The IMAC eluate was concentrated to 0.5-1 mL using a 10 kDa cutoff centricon, and desalted using a PD-10 desalting column (*BioRad*) equilibrated in 100 m*M* NaCl, 20 m*M* Tris pH = 7.8, 2% glycerol, 2.8 m*M* β-ME and 0.2 m*M* PMSF, to remove imidazole. Following desalting, the protein was purified by ion exchange chromatography (IEX) on a CaptoS (ClpB2) or a CaptoQ (for ClpB1-sfGFP) HiTrap column (*GE Healthcare*). The sample was then treated with thrombin for 24 hr at 4°C (0.25 units enzyme/mg of protein) to remove the histidine purification tag. Following thrombin treatment, the sample was further purified on a mixture of Nickel-NTA IMAC (*Qiagen*)

Table II. *Small Angle X-ray Scattering Analysis*

Protein	Sequence Mass[a]	X-ray structures: Models R_G	X-ray structures: Models D_{max}	SAXS: Guinier analysis R_G	SAXS: Fourier analysis R_G	SAXS: Fourier analysis D_{max}
N ClpB2[b]	16,782 Da	16.3 Å	53 Å	16.9 ± 0.1Å	16.9 ± 0.1Å	55 ± 3 Å

[a] Theoretical masses calculated for the monomer of the protein constructs described in this study.
[b] The theoretical R_G and D_{max} values correspond to the monomer of the X-ray structure reported in this study.

and Benzamidine Sepharose (*GE Healthcare*) resins to simultaneously subtract unprocessed tagged-proteins, tags, and thrombin. The final purification step consisted in a size exclusion chromatography (SEC) on a Superdex 200 HR16/60 or a Superdex 75 HR10/30 gel-filtration column (*GE Healthcare*) equilibrated in 150 m*M* NaCl, 20 m*M* Tris pH = 7.8, 2% glycerol, and 7 m*M* β-ME to condition samples for crystallization trials.

Crystallization

Crystals of the *Pfal*-N-ClpB1-sfGFP fusion protein grew at room temperature in 2.0*M* Ammonium sulfate and 100 m*M* Hepes at pH = 7.0 at a protein concentration of 13–17 mg/mL and belong to hexagonal space group $P6_5$ with unit cells parameters a = 127.5 Å, c = 92.6 Å and a solvent content of 52% corresponding to two molecules of fusion protein in the asymmetric unit. The N-terminal domain of *Pfal*-ClpB2/HSP101 was concentrated at 30 mg/mL and yielded multiple crystallization conditions at 4°C. A first crystal form grew at 4°C in 20% PEG 3350, 10% glycerol and 50 m*M* CAPSO buffer at pH = 9.4 and belong to orthorhombic space group $P22_12_1$ with unit cell parameters $a = 31.8$ Å $b = 92.1$ Å and $c = 97.0$ Å for two copies in the asymmetric unit and a solvent content of 45%. A second form of crystals belonging to orthorhombic space group $C222_1$ was obtained in 2.2*M* ammonium sulfate with unit cell parameters $a = 89.1$ Å $b = 141.9$ Å and $c = 91.8$ Å for three copies in the asymmetric unit and a solvent content of 58%.

Structure determination and refinement

All diffraction data were collected on beamlines 24-ID-C or 24-ID-E at the Advanced Photon Source in the Argonne National Laboratory using crystals cryoprotected in mother liquor supplemented with 20–25% glycerol. Data were indexed, scaled and reduced using *XDS*.[60] For N-ClpB2, the $C222_1$ 1.8 Å resolution structure was solved by molecular replacement using the structures of *E. coli* CplB (PDB ID 1KHY) and *T. thermophilus* ClpB (PDB ID 1QVR) N terminal domains as search probes. Molecular replacement using *Phaser*[61] combined with partial automatic rebuilding in place with *Phenix*[62] yielded a preliminary model that was manually improved in *Coot*.[63] ADP refinement was carried in *Phenix* using TLS restraints combined with individual isotropic ADP refinement. This structure was then used to determine the $P22_12_1$ 1.65 Å resolution structure by molecular replacement using *Phaser* and refined with *Phenix* using TLS restraints combined with individual isotropic ADP refinement. For N-ClpB1-sfGFP, the structure was solved at 2.0 Å resolution by molecular replacement in *Phaser* using the structures of the N domains from *E. coli*, *T. thermophilus* ClpBs and our *P. falciparum* ClpB2 N domain structure (PDB 4IRF and 4IOD, this work) and the superfolder GFP structure (PDB 2B3P) as search probes. The resulting maps enabled partial automatic rebuilding of the model with further rebuilding in *Coot*. Refinement was carried out in *Phenix* using a strategy combining simulated annealing and individual ADP refinement. At the last stage of refinement, TLS refinement was performed where each chain was split into two TLS groups corresponding to the ClpB1 N domain and the GFP. For each structure, all molecules observed in the asymmetric unit were refined independently without applying non-crystallographic symmetry restraints.

Small angle X-ray scattering

Solution scattering data were collected at the SIBYLS high-throughput SAXS beamline 12.3.1[64] of the Advanced Light Source at the Lawrence Berkeley National Laboratory. Proteins and complexes were first prepared as described in the above section. *S*amples were further purified by gel filtration (Superdex 75 HR 10/30) equilibrated in the SAXS experimental buffer 200 m*M* NaCl, 20 m*M* Tris pH = 8.0, 2% glycerol, 0.5 m*M* EDTA and 5 m*M* fresh DTT with concentrations ranging from 1 to 25 mg/mL as estimated by UV absorption spectroscopy using the extinction coefficient values calculated from the amino-acid compositions. All intensity curves were collected with a sample-to-detector distance of 1.5m and an X-ray wavelength of $\lambda = 1$ Å and each contained about 500 data points. This corresponds to a Q ranging from 0.001 to 0.32 $Å^{-1}$, with Q the scattering vector defined as $Q = 4\pi\sin\theta/\lambda$ where 2θ and λ are the scattering angle and X-ray wavelength, respectively. For each sample several exposures were sequentially recorded (0.5, 1, 2, and

4 sec). Corresponding to each protein sample, data were collected on a buffer sample under identical experimental conditions. A beamline-specific software was used for radial averaging of images and for subtracting buffer signal from protein signal.[65] Each scattering profile was visually inspected to check for aggregation, repulsive or attractive interactions, X-ray induced damage, detector saturation or anomalies in buffer subtraction. Following data reduction, analysis of experimental scattering data was performed using *Scatter* (http://bl1231.als.lbl.gov/saxs_protocols/index.php and http://www.bioisis.net/) to determine the radius of gyration (R_G) by Guinier analysis, *P(r)* and D_{max}. Kratky plots [$Q^2.I(Q)$ vs. Q] were used to assess the particle flexibility and folded nature.[66] Scattering curves corresponding to the X-ray structures reported in this study and fitted against the experimental scattering curves were calculated using *CRYSOL*[67] or *FoXS*[68] with default parameters; the discrepancy factor χ^2 was used to estimate the quality of the fits between experimental and calculated scattering curves. Pair distance distribution functions (*P(r)*) were derived by Fourier transform using *GNOM*[69] to estimate D_{max}, the longest distance occurring within the particle, and R_G, independently of Guinier analysis.

Acknowledgments

The authors are grateful to Kevin Dyer from the Advanced Light Source SYBILS beamline 12.3.1 for his help with SAXS data collection. They acknowledge the DOE Integrated Diffraction Analysis (IDAT) grant DE-AC02-05CH11231 that supports the development and implementation of the SIBYLS beamline at the Advanced Light Source of the Lawrence Berkeley National Laboratory. They thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov, and F. Murphy at NECAT beamlines 24-ID at the Advanced Photon Source at the Argonne National Laboratory

Accession Numbers

Atomic coordinates and structure factors have been deposited at the RCSB Protein Data Bank (http://www.rcsb.org/pdb) under accession numbers 4IOD and 4IRF for the ClpB2/HSP101 N domain and 4XBI for the ClpB1 N domain GFP fusion.

Author Contributions

Performed the experiments: APAY, AK, DC, and PFE. Designed the experiments: APAY, AK, DC, and PFE. Contributed reagents/materials/analysis: APAY, AK, DC, and PFE. Wrote the paper: APAY, AK, and PFE. Contributed equally to this work: APAY and AK.

References

1. Nishimura K, van Wijk KJ (2014) Organization, function and substrates of the essential Clp protease system in plastids. Biochim Biophys Acta 1847:915–930.
2. Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 9:27–43.
3. Doyle SM, Wickner S (2009) Hsp104 and ClpB: protein disaggregating machines. Trend Biochem Sci 34:40–48.
4. Baker TA, Sauer RT (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. Biochim Biophys Acta 1823:15–28.
5. Yu AY, Houry WA (2007) ClpP: a distinctive family of cylindrical energy-dependent serine proteases. FEBS Lett 581:3749–3757.
6. Parsell DA, Kowal AS, Singer MA, Lindquist S (1994) Protein disaggregation mediated by heat-shock protein Hsp104. Nature 372:475–478.
7. Krobitsch S, Clos J (1999) A novel role for 100 kD heat shock proteins in the parasite Leishmania donovani. Cell Stress Chaperones 4:191–198.
8. Capestany CA, Tribble GD, Maeda K, Demuth DR, Lamont RJ (2008) Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in Porphyromonas gingivalis. J Bacteriol 190:1436–1446.
9. Meibom KL, Dubail I, Dupuis M, Barel M, Lenco J, Stulik J, Golovliov I, Sjostedt A, Charbit A (2008) The heat-shock protein ClpB of Francisella tularensis is involved in stress tolerance and is required for multiplication in target organs of infected mice. Mol Microbiol 67:1384–1401.
10. Brotz-Oesterhelt H, Sass P (2014) Bacterial caseinolytic proteases as novel targets for antibacterial treatment. Int J Med Microbiol 304:23–30.
11. El Bakkouri M, Pow A, Mulichak A, Cheung KL, Artz JD, Amani M, Fell S, de Koning-Ward TF, Goodman CD, McFadden GI, Ortega J, Hui R, Houry WA (2010) The Clp chaperones and proteases of the human malaria parasite *Plasmodium falciparum*. J Mol Biol 404:456–477.
12. El Bakkouri M, Rathore S, Calmettes C, Wernimont AK, Liu K, Sinha D, Asad M, Jung P, Hui R, Mohmmed A, Houry WA (2013) Structural insights into the inactive subunit of the apicoplast-localized caseinolytic protease complex of *Plasmodium falciparum*. J Biol Chem 288:1022–1031.
13. Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS, McFadden GI (2004) Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2:203–216.
14. Yeh E, DeRisi JL (2011) Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. PLoS Biol 9:e1001138
15. Goodman CD, Su V, McFadden GI (2007) The effects of anti-bacterials on the malaria parasite *Plasmodium falciparum*. Mol Biochem Parasitol 152:181–191.
16. de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF, Crabb BS (2009) A newly discovered protein export machine in malaria parasites. Nature 459:945–949.
17. Crabb BS, Cooke BM, Reeder JC, Waller RF, Caruana SR, Davern KM, Wickham ME, Brown GV, Coppel RL, Cowman AF (1997) Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell 89:287–296.

18. Waterkeyn JG, Wickham ME, Davern KM, Cooke BM, Coppel RL, Reeder JC, Culvenor JG, Waller RF, Cowman AF (2000) Targeted mutagenesis of *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) disrupts cytoadherence of malaria-infected red blood cells. EMBO J 19:2813–2823.
19. van Ooij C, Tamez P, Bhattacharjee S, Hiller NL, Harrison T, Liolios K, Kooij T, Ramesar J, Balu B, Adams J, Waters AP, Janse CJ, Haldar K (2008) The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog 4:e1000084
20. Maier AG, Rug M, O'Neill MT, Brown M, Chakravorty S, Szestak T, Chesson J, Wu Y, Hughes K, Coppel RL, Newbold C, Beeson JG, Craig A, Crabb BS, Cowman AF (2008) Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. Cell 134:48–61.
21. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306:1930–1933.
22. Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306:1934–1937.
23. Gruring C, Heiber A, Kruse F, Flemming S, Franci G, Colombo SF, Fasana E, Schoeler H, Borgese N, Stunnenberg HG, Przyborski JM, Gilberger TW, Spielmann T (2012) Uncovering common principles in protein export of malaria parasites. Cell Host Microbe 12:717–729.
24. Bullen HE, Charnaud SC, Kalanon M, Riglar DT, Dekiwadia C, Kangwanrangsan N, Torii M, Tsuboi T, Baum J, Ralph SA, Cowman AF, de Koning-Ward TF, Crabb BS, Gilson PR (2012) Biosynthesis, localisation and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEX). J Biol Chem 287:7871–7884.
25. Fischer K, Marti T, Rick B, Johnson D, Benting J, Baumeister S, Helmbrecht C, Lanzer M, Lingelbach K (1998) Characterization and cloning of the gene encoding the vacuolar membrane protein EXP-2 from *Plasmodium falciparum*. Mol Biochem Parasit 92:47–57.
26. Riglar DT, Rogers KL, Hanssen E, Turnbull L, Bullen HE, Charnaud SC, Przyborski J, Gilson PR, Whitchurch CB, Crabb BS, Baum J, Cowman AF (2013) Spatial association with PTEX complexes defines regions for effector export into *Plasmodium falciparum*-infected erythrocytes. Nat Commun 4:1415
27. Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) E-coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. Cell 100:265–276.
28. Eifler N, Vetsch M, Gregorini M, Ringler P, Chami M, Philippsen A, Fritz A, Muller SA, Glockshuber R, Engel A, Grauschopf U (2006) Cytotoxin ClyA from Escherichia coli assembles to a 13-meric pore independent of its redox-state. EMBO J 25:2652–2661.
29. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. Nature 459:726–730.
30. Gehde N, Hinrichs C, Montilla I, Charpian S, Lingelbach K, Przyborski JM (2009) Protein unfolding is an essential requirement for transport across the parasitophorous vacuolar membrane of *Plasmodium falciparum*. Mol Microbiol 71:613–628.
31. Matthews K, Kalanon M, Chisholm SA, Sturm A, Goodman CD, Dixon MW, Sanders PR, Nebl T, Fraser F, Haase S, McFadden GI, Gilson PR, Crabb BS, de Koning-Ward TF (2013) The Plasmodium translocon of exported proteins (PTEX) component thioredoxin-2 is important for maintaining normal blood-stage growth. Mol Microbiol 89:1167–1186.
32. Matz JM, Matuschewski K, Kooij TW (2013) Two putative protein export regulators promote Plasmodium blood stage development in vivo. Mol Biochem Parasitol 191:44–52.
33. Beck JR, Muralidharan V, Oksman A, Goldberg DE (2014) PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature 511:592–595.
34. Elsworth B, Matthews K, Nie CQ, Kalanon M, Charnaud SC, Sanders PR, Chisholm SA, Counihan NA, Shaw PJ, Pino P, Chan JA, Azevedo MF, Rogerson SJ, Beeson JG, Crabb BS, Gilson PR, de Koning-Ward TF (2014) PTEX is an essential nexus for protein export in malaria parasites. Nature 511:587–591.
35. Lee S, Choi JM, Tsai FT (2007) Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB. Mol Cell 25: 261–271.
36. Effantin G, Ishikawa T, De Donatis GM, Maurizi MR, Steven AC (2010) Local and global mobility in the ClpA AAA+ chaperone detected by cryo-electron microscopy: functional connotations. Structure 18: 553–562.
37. Zhang T, Ploetz EA, Nagy M, Doyle SM, Wickner S, Smith PE, Zolkiewski M (2012) Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency. Proteins 80:2758–2768.
38. Chow IT, Barnett ME, Zolkiewski M, Baneyx F (2005) The N-terminal domain of Escherichia coli ClpB enhances chaperone function. FEBS Lett 579:4242–4248.
39. Kirstein J, Schlothauer T, Dougan DA, Lilie H, Tischendorf G, Mogk A, Bukau B, Turgay K (2006) Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. EMBO J 25:1481–1491.
40. Djuranovic S, Hartmann MD, Habeck M, Ursinus A, Zwickl P, Martin J, Lupas AN, Zeth K (2009) Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. Mol Cell 34:580–590.
41. Wojtyra UA, Thibault G, Tuite A, Houry WA (2003) The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. J Biol Chem 278:48981–48990.
42. Bonemann G, Pietrosiuk A, Diemand A, Zentgraf H, Mogk A (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. EMBO J 28:315–325.
43. Maurizi MR, Xia D (2004) Protein binding and disruption by Clp/Hsp100 chaperones. Structure 12:175–183.
44. Ngansop F, Li H, Zolkiewska A, Zolkiewski M (2013) Biochemical characterization of the apicoplast-targeted AAA+ ATPase ClpB from *Plasmodium falciparum*. Biochem Biophys Res Commun 439:191–195.
45. Sharma A, Dixit S (2011) Structural insights into thioredoxin-2: a component of malaria parasite protein secretion machinery. Sci Rep 1:179
46. Peng M, Cascio D, Egea PF (2015) Crystal structure and solution characterization of the thioredoxin-2 from *Plasmodium falciparum*, a constituent of an essential parasitic protein export complex. Biochem Biophys Res Commun 456:403–409.

47. Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, Tsai FT (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell 115:229–240.
48. Li J, Sha B (2003) Crystal structure of the E. coli Hsp100 ClpB N-terminal domain. Structure 11:323–328.
49. Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, Dikic I, Drew D, Iwata S, Wakatsuki S, Kawasaki M (2010) Crystallization of small proteins assisted by green fluorescent protein. Acta Cryst D66: 1059–1066.
50. Moon AF, Mueller GA, Zhong X, Pedersen LC (2010) A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci 19:901–913.
51. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 24:79–88.
52. Xia D, Esser L, Singh SK, Guo F, Maurizi MR (2004) Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. J Struct Biol 146:166–179.
53. Pietrosiuk A, Lenherr ED, Falk S, Bonemann G, Kopp J, Zentgraf H, Sinning I, Mogk A (2011) Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. J Biol Chem 286:30010–30021.
54. Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B, Dougan DA (2002) Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. Nat Struct Biol 9:906–911.
55. Guo F, Esser L, Singh SK, Maurizi MR, Xia D (2002) Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. J Biol Chem 277:46753–46762.
56. Wang F, Mei Z, Qi Y, Yan C, Hu Q, Wang J, Shi Y (2011) Structure and mechanism of the hexameric MecA-ClpC molecular machine. Nature 471:331–335.
57. Blanchet CE, Svergun DI (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. Ann Ref Phys Chem 64:37–54.
58. Chou TF, Brown SJ, Minond D, Nordin BE, Li K, Jones AC, Chase P, Porubsky PR, Stoltz BM, Schoenen FJ, Patricelli MP, Hodder P, Rosen H, Deshaies RJ (2011) Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. Proc Natl Acad Sci U S A 108:4834–4839.
59. Firestone AJ, Weinger JS, Maldonado M, Barlan K, Langston LD, O'Donnell M, Gelfand VI, Kapoor TM, Chen JK (2012) Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein. Nature 484:125–129.
60. Kabsch W (2010) Xds. Acta Cryst D66:125–132.
61. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
62. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
63. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst D60:2126–2132.
64. Classen S, Hura GL, Holton JM, Rambo RP, Rodic I, McGuire PJ, Dyer K, Hammel M, Meigs G, Frankel KA, Tainer JA (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the advanced light source. J Appl Cryst 46:1–13.
65. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL, 2nd Tsutakawa SE, Jenney FE, Jr., Classen S, Frankel KA, Hopkins RC, Yang SJ, Scott JW, Dillard BD, Adams MW Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature Meth 6:606–612.
66. Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers 95:559–571.
67. Svergun DI, Barberato C, Koch MH (1995) CRYSOL: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 28:768–773.
68. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J 105:962–974.
69. Svergun DI (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria. J Appl Cryst 25:495–503.
70. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4:269–275.
71. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. Science 336:1030–1033.
72. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98:10037–10041.

Supplementary Information for

Structural mapping of the two ClpB ATPases of *Plasmodium falciparum*: targeting protein folding and secretion for antimalarial drug design

by

Andrew P. AhYoung, Antoine Koehl, Duilio Cascio and Pascal F. Egea

Table S1. RMSD deviation and sequence identity between the N domain structures of malarial ClpB proteins and other Clp proteins of known structure.

N domains	**RMSD (Å)**	**sequence identity (%)**
Clp proteins (*PDB code*)	**N-ClpB1 / N-ClpB2**	**N-ClpB1 / N-ClpB2**
ClpA *Escherichia (1K6K)*	1.90 / 1.98	17.7 / 13.8
ClpB *Escherichia (1KHY)*	1.45 / 1.54	22.7 / 19.2
ClpB *Thermus (1QVR)*	1.78 / 2.44	24.3 / 20.1
ClpC *Bacillus (2Y1Q)*	2.15 / 2.23	18.0 / 17.3
ClpV *Vibrio (3ZRI)*	2.26 / 2.25	16.4 / 12.0

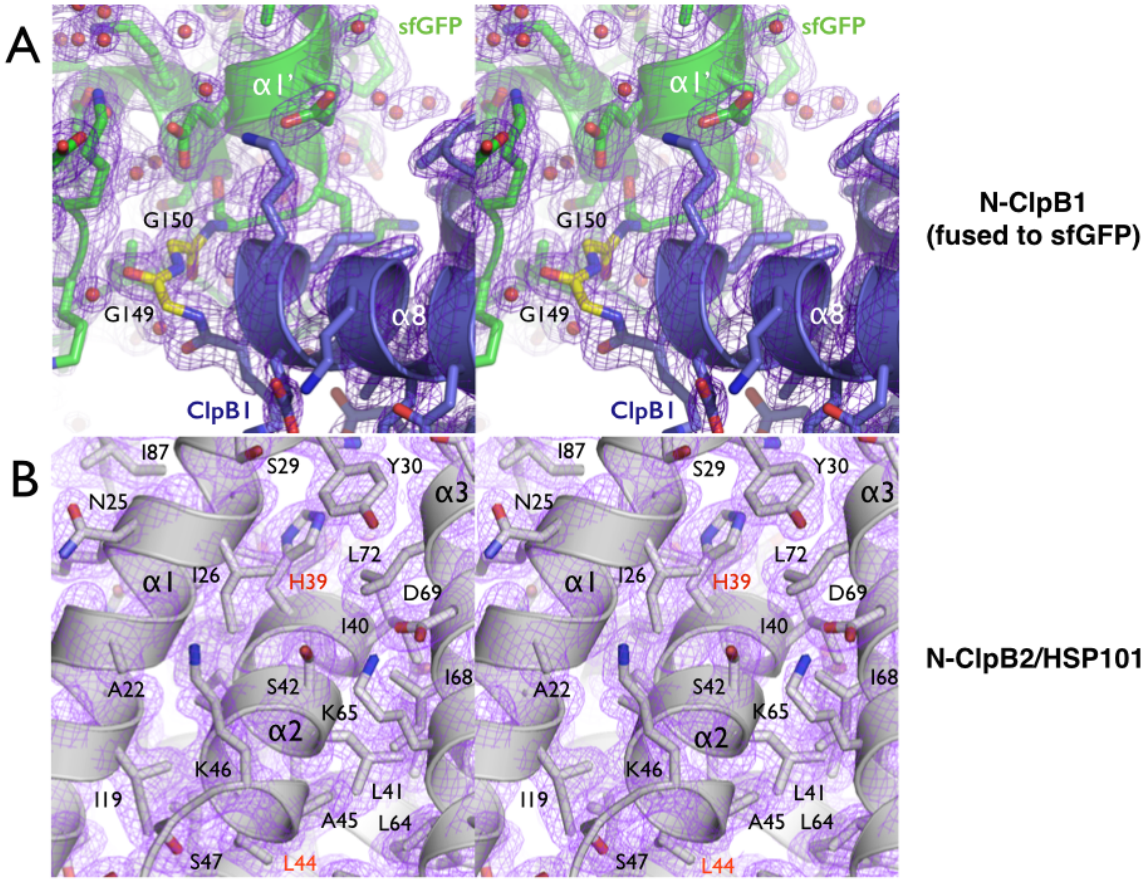

Figure S1. Electron density maps. Stereo views of representative 2m*Fo*-D*Fc* Fourier difference electron density maps contoured at 1.5σ for the N-ClpB1-sfGFP (**A**) and the N-ClpB2/HSP101 (**B**) crystal structures. In (**A**) the Gly-Gly linker connecting the C-terminus of the ClpB1 N domain (blue) to the N-terminus of the sfGFP protein (green) is highlighted in yellow.

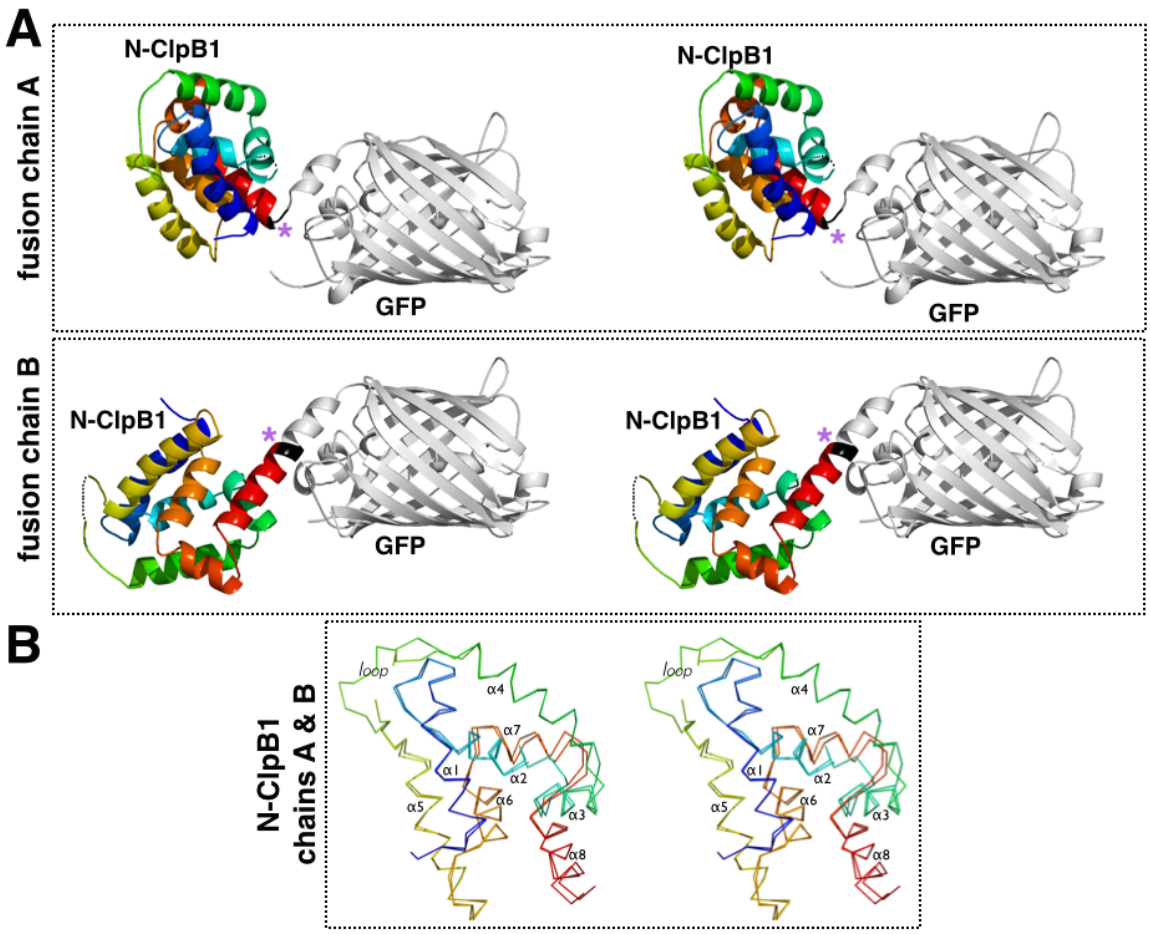

Figure S2. Conformational plasticity of the linker in the N-ClpB1-GFP fusion. (A) Stereo views showing the two copies of N-ClpB1-GFP fusion protein present in the asymmetric unit. The two chains are displayed using the GFP carrier protein as reference point for superposition. The linker is colored in black and indicated with a magenta star. (**B**) The backbone of the two copies of N-ClpB1 are superposed showing that despite the different relative orientations adopted by each N-domain relative to the GFP carrier, this does not affect the N domain structures. The N-domain chains are colored using a rainbow pattern.

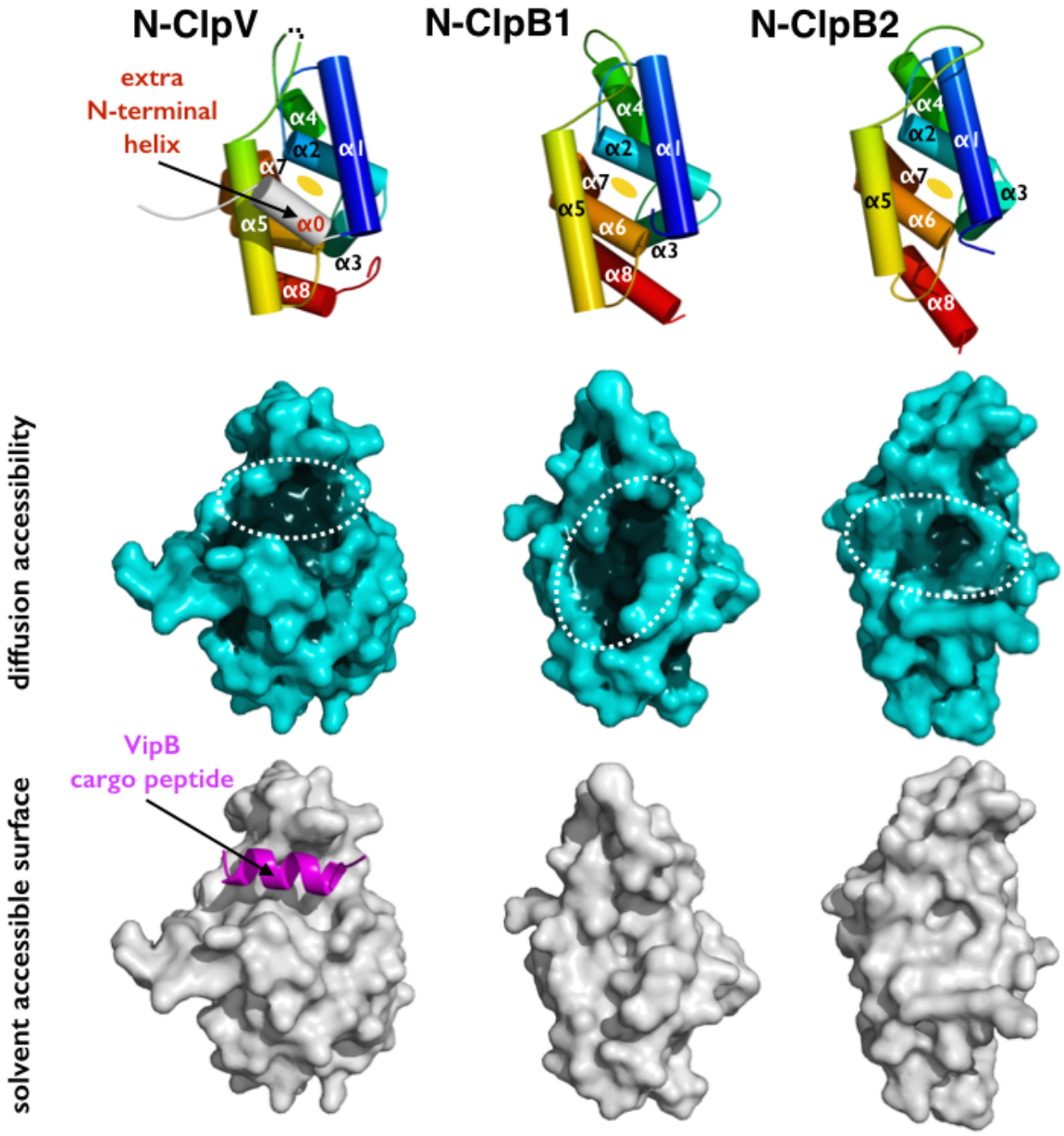

Figure S3. Comparative analysis of the protein surfaces in the N terminal domains of *Plasmodium falciparum* ClpB1 and ClpB2/HSP101 and in the *Vibrio cholerae* N-ClpV/VipB cargo peptide complex. The putative binding sites of *Plasmodium* ClpB1 and ClpB2 N domains corresponding to the 'hydrophobic patches' resulting from the internal pseudo twofold symmetry axis characteristic of the N domain fold are compared to the peptide binding site observed in the *Vibrio* N domain ClpV/VipB peptide complex (PDB 3ZRJ) [1]. The cylindrical representation of helices highlights the presence of the extra N-terminal helix (α0) in ClpV. Surfaces in blue are shaded according to depth, based on a calculation of diffusion accessibility [2]. Solvent accessible surfaces are colored in grey. The VipB cargo peptide bound to ClpV is colored in pink. The cylindrical representations of the N-domains are colored using a rainbow pattern. All views correspond to the same orientation.

Supplementary references

1. Pietrosiuk A, Lenherr ED, Falk S, Bonemann G, Kopp J, Zentgraf H, Sinning I, Mogk A. Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. J Biol Chem 2011;286(34):30010-30021.
2. Yeates TO. Algorithms for evaluating the long-range accessibility of protein surfaces. Journal of molecular biology 1995;249(4):804-815.