Structural and functional insights into apicomplexan gliding and its regulation

Dissertation

to obtain the degree of

Doctor of Natural Sciences

University of Hamburg

Faculty of Mathematics, Informatics and Natural Sciences

at the Department of Biology

by

Samuel Pažický

from Bratislava, Slovakia

Hamburg 2020

Examination commission

Examination commission chair

Prof. Dr. Jörg Ganzhorn (University of Hamburg)

Examination commission members

Prof. Jonas Schmidt-Chanasit (Bernhard Nocht Institute for Tropical Medicine and University of Hamburg)

Prof. Tim Gilberger (Bernhard Nocht Institute for Tropical Medicine, Centre for Structural Systems Biology and University of Hamburg)

Dr. Maria Garcia-Alai (European Molecular Biology Laboratory and Centre for Structural Systems Biology)

Dr. Christian Löw (European Molecular Biology Laboratory and Centre for Structural Systems Biology)

Date of defence: 29.01.2021

This work was performed at European Molecular Biology Laboratory, Hamburg Unit under the supervision of Dr. Christian Löw and Prof. Tim-Wolf Gilberger.

The work was supported by the Joachim Herz Foundation.

Evaluation

Prof. Dr. rer. nat. Tim-Wolf Gilberger
Bernhard Nocht Institute for Tropical Medicine (BNITM)
Department of Cellular Parasitology
Hamburg

Dr. Christian Löw
European Molecular Biology Laboratory
Hamburg unit
Hamburg

Prof. Dr. vet. med. Thomas Krey
Hannover Medical School
Institute of Virology

Declaration of academic honesty

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, 22.9.2020 Samuel Pažický

List of contents

Acknowledgements

First, I would like to thank Dr. Christian Löw for giving me the opportunity to work in his laboratory. As a mentor, he always found a time to consult my results, provide me with a valuable feedback and steer the work in the right direction. With his scientific excellence and great organizational skills, he helped me to improve as a scientist and as a person, too.

My thanks also go to my thesis advisory committee members: Dr. Orsolya Barabas, Dr. Jan Kosinski and Prof Tim Gilberger, whose advice always turned my project in the right direction. Special thanks go to the scientists that directly contributed to my project: Dr. Ulrich Weininger, Dr. Arne Alder, Dr. Karol Kaszuba and Dr. Haydyn Mertens.

I would also like to thank (Dr.-to-be) Karthikeyan Dhamotharan who I had a chance to mentor for one year. His dedication, precision and fast learning crucially helped with the progress of the project and I am very thankful for all his hard work.

I would also like to acknowledge all the current and past members of the Löw group, who have not only given me infinite amount of scientific advice, but also have become my good friends.

At the same time, I am very thankful for the community at EMBL Hamburg, the wet lab team, SPC team, IT group, beamline scientists and all fellow predocs, postdocs and scientists, without whom I would not be able push the project so far. Thank you for making EMBL a special place.

Not least, I want to thank Samantha Seah for her tremendous support and for the amazing journey that we went through over the last four years. My thanks also go to my closest friends who have always been there for me in spite of the long distance.

Finally, I would like to thank my family. Without their unprecedented love and support, I would not have become the person that I am now.

Summary

Apicomplexans are a group of single cell parasitic organisms that are infectious agents of many human and animal diseases. For example, *Plasmodium* species cause malaria, which infects over 200 million and kills almost 400 thousands people every year. *Toxoplasma* species have infected about 30-50% of the entire population. While infection is benign for most, it is a threat for immunocompromised patients and pregnant women. Several other apicomplexan species are causative agents of diseases in domesticated animals in Africa, representing further financial and humanitarian burdens.

A typical apicomplexan parasite completes a sophisticated life cycle, usually switching between two host organisms to engage in sexual reproduction in one and the asexual reproduction in the other. During their life cycle, the parasites differentiate into distinct life stages that possess unique shapes and morphologies. Motile life stages, in which the parasites move through the tissues and invade their host cells, are of specific interest because the parasite cells are exposed to the extracellular environment outside the host cells. Consequently, they are more accessible to the host immune system and drugs, making them and their proteins major drug targets.

To move across tissues and invade host cells, apicomplexans use a specific type of motility that does not require a change in their cell shape, called gliding. Many organelles and subcellular structures of the motile apicomplexan stages are designed to mediate efficient gliding and host cell invasion. These parasitic cells are elongated and polarized, forming an apical pole in the anterior and basal end in the posterior of the parasite. The parasites glide and invade the host cell in the apical direction. The apical pole contains specialized secretory organelles, such as micronemes and rhoptries, that discharge important virulence factors essential for gliding and for host cell invasion. Other specialized organelles, alveoli, also called inner membrane complex (IMC), form a network of large flat vesicles underneath the parasite`s plasma membrane to accommodate the molecular motor that powers the motility and invasion process.

This thesis cumulates three separate bodies of work that represent advances in understanding of the processes that are important in gliding motility and host cell invasion by apicomplexan parasites.

The first work (manuscript 1) reveals the structure and the function of the essential light chains (ELCs) of *Plasmodium falciparum* and *Toxoplasma gondii*. In apicomplexan parasites, ELCs are part of the glideosome, a larger complex residing between the parasite`s plasma membrane and the IMC. The glideosome consists of a myosin motor protein MyoA, two light chains – ELC and myosin light chain 1 (MLC1) and three glideosome-associated proteins (GAPs). It has been shown that ELCs bind the neck

domain of MyoA and facilitate an enhanced motor efficiency. However, the molecular mechanism behind this phenomenon remains unclear.

We have shown that the two *T. gondii* ELCs (TgELC1 and TgELC2) as well as *P. falciparum* ELC (PfELC) bind a conserved sequence of the MyoA neck domain. We have determined the structures of all ELCs bound in trimeric complexes with their respective MyoA neck domains and MLC1s, representing the first glideosome trimeric complexes shown so far. These structures reveal that the C-terminus of PfELC, which is partially unfolded in an unbound state, maintains some degree of flexibility even in the complex, resulting in a different interaction angle compared to TgELCs and explaining lower affinity of PfELC to MyoA compared to its *T. gondii* analogues. Additionally, we have shed a light on the regulatory function of ELCs, showing that two phosphorylation sites can regulate their binding to MyoA. On the other hand, we proved that the binding of calcium does not regulate the ELC binding and only increases the stability of the trimeric complexes. Finally, we have shown that upon binding, the ELCs induce a formation of secondary structure in MyoA neck, leading to a stiffer MyoA neck that consequently serves more efficiently as a lever arm and increases the motor speed.

In the second project (manuscript 2), we have tackled the role of a rhoptry protein Armadillo Repeat-Only (ARO) and its putative interaction partner, ARO interacting protein (AIP). ARO is anchored to the membrane of the rhoptries from their cytoplasmic side by palmitoylation and myristoylation. In *T. gondii*, ARO has been shown to be essential for the correct positioning of rhoptries in the apical pole and to interact with AIP1, myosin F (MyoF) and adenylate cyclase beta (ACβ). However, it was not clear whether the same arrangement holds true in *P. falciparum*.

We solved the crystal structure of PfARO that, with five armadillo repeat motifs, assumes a bean-shaped conformation. We identified the *P. falciparum* homologue of *T. gondii* AIP1 and showed that AIP1 proteins are only present in apicomplexan organisms, lacking any structural homologues outside of apicomplexan species. Further localisation studies, using super-resolution microscopy, revealed that in *P. falciparum*, PfARO does not co- localise with PfAIP. While PfARO localises to rhoptry bulbs, PfAIP1 is present in rhoptry necks. On the other hand, deletion of the conserved loop of PfARO, that was identified thanks to the solved structure, leads to mis localisation of PfAIP1, suggesting that the two proteins do associate. Finally, proximity-dependent biotin identification approach showed that PfAIP associates with PfACβ and MyoF as in *T. gondii*, but does not interact with PfARO. Based on these conflicting results, we have proposed two models of PfARO-PfAIP interaction.

The third work (manuscript 3) investigates the structure and function of *P. falciparum* glycogen synthase kinase (PfGSK3). Human GSK3 regulates diverse processes, such as glycogen metabolism,

cell cycle and grow, translation, embryonic development or differentiation of neurons. On the other hand, the role of PfGSK3 is rather elusive. PfGSK3 has been shown to have an impact on parasite`s invasion. During host cell invasion, the parasite forms the moving junction, an interface between the parasitic cell and host cell, through which the parasites glide inside the host cell. The dominant protein forming the moving junction is apical membrane antigen 1 (AMA1) that interacts with rhoptry neck (RON) protein complex on the surface of the red blood cells. AMA1 is a transmembrane protein residing on the surface of *Plasmodium* species with an N-terminal extracellular domain and a short C-terminal cytoplasmic tail. The phosphorylation of this cytoplasmic tail regulates the function of AMA1 and is crucial in the invasion process. The C-terminal tail of AMA1 is first phosphorylated by protein kinase A (PKA) and subsequently also by PfGSK3. Consequently, the inhibitors of GSK3 have been shown to impair host cell invasion of the parasite. Although PfGSK3 is an important drug target, its structure and means of regulation are unknown.

In our investigation, we have discovered two factors that regulate the activity of PfGSK3. I have found that PfGSK3 exhibits autophosphorylation, phosphorylating its residues in the activation loop and at the N-terminus. We have shown that the N-terminus of PfGSK3 is indispensable for its function and that the amount of N-terminal phosphorylation positively correlates with PfGSK3 activity. Additionally, we found that bivalent heavy metal ions, such as those of zinc and copper, induce reversible formation of high-molecular-weight species of PfGSK3 that are heterogeneous and enzymatically inactive. Thus, our work provide the first insights into processes that possibly regulate the function of PfGSK3 *in vivo*.

Zusammenfassung

Apicomplexa sind einzellige parasitäre Organismen, die Infektionserreger vieler menschlicher und tierischer Krankheiten sind. Zum Beispiel, die Arten der *Plasmodien* infizieren jedes Jahr über 200 Millionen und töten fast 400 Tausend Menschen mit Malaria. *Toxoplasma* Arten infizieren etwa 30-50% der Gesamtbevolkerung und obwohl sie für die meisten Menschen nicht gefährlich sind, stellen sie eine Bedrohung für immungeschwächte Patienten und schwangere Frauen dar. Andere *Apicomplexa* Arten sind Krankheitserreger bei domestizierten Tieren in Afrika und stellen dadurch eine finanzielle und humanitäre Belastungen dar.

Ein typischer *Apicomplexa* Parasit weißt einen anspruchsvollen Lebenszyklus auf, typischerweise wechselnd zwischen zwei Wirtsorganismen, um in einem die asexuelle Vermehrung und in dem anderen die sexuelle Reproduktion durchzuführen. Während des Lebenszyklus differenzieren sich die Parasiten in verschiedene Lebensstadien, welche immer eine einzigartige Form und Morphologie besitzen. Motile Lebensstadien, die das Gewebe penetrieren können, sind von besonderem Interesse, da die Parasit-Zellen außerhalb der Wirtszellen exponiert sind. Die Parasiten sind folglich für die Interaktion mit dem Immunsystem des Wirts und für Arzneimitteln leichter zugänglich, wodurch die motile Lebensstadien und ihre Proteine wichtige Wirkstoffziele sind.

Um das Gewebe zu penetrieren und in die Wirtszellen einzudringen, verwenden *Apicomplexa* eine bestimmte Art der Motilität, die keine Änderung ihrer Form erfordert: das Gleiten. Viele Organellen und subzelluläre Strukturen der motilen *Apicomplexa* Stadien vermitteln ein effizientes Gleiten und Eindringen in die Wirtszellen. Diese Parasiten sind länglich und ihre Zellen sind polarisiert. Durch die Polarität formen sie einen apikalen Pol im vorderen Teil und ein basales Ende im hinteren Teil. Die Parasiten gleiten und dringen in apikaler Richtung in die Wirtszelle ein. Der apikale Pol enthält spezialisierte sekretorische Organellen, wie Mikronemen und Rhoptrien, die wichtige Virulenzfaktoren sekretieren, und für das Gleiten und die Invasion der Wirtszellen wichtig sind. Andere spezielle Organellen, Alveolen (auch als Innerer Membran Komplex, IMC, bezeichnet), bauen ein Netzwerk von großen flachen Vesikel unter der Plasmamembran der Parasiten auf, um den molekularen Motor zu beherbergen, der die Motilität und den Invasionsprozess antreibt.

Diese Arbeit fasst drei Manuskripte zusammen, die Fortschritte beim Verständnis der Prozesse darstellen, die für die Gleitmotilität und die Invasion von Wirtszellen durch *Apicomplexa* Parasiten wichtig sind.

In der ersten Arbeit (Manuskript 1) wurde die Struktur und Funktion Essentieller Leichtketten (*essential light chains*, ELCs) von *Plasmodium falciparum* und *Toxoplasma gondii* entdeckt. In *Apicomplexa* sind

ELCs ein Teil eines größeren Proteinkomplexes zwischen der Plasmamembran des Parasiten und dem IMC, der als Glideosom bezeichnet wird. Das Glideosom besteht aus einem Myosin Motorprotein MyoA, zwei Leichtketten – ELC und Myosin-Leichkette 1 (MLC1) und drei Glideosom-assoziierten Proteinen (GAPs). Es war bereits bekannt, dass ELCs die Halsdomäne von MyoA binden und dadurch die Motoreffizienz verbessern. Der molekulare Mechanismus hinter diesem Effekt war jedoch unklar.

Wir haben gezeigt, dass die beiden *T. gondii* ELCs (TgELC1 und TgELC2), sowie auch *P. falciparum* ELC (PfELC) eine konservierte Sequenz der MyoA-Halsdomäne binden. Wir konnten die Strukturen aller ELCs gebunden in Trimerkomplexen mit den jeweiligen MyoA-Halsdomänen und MLC1s bestimmen. Diese sind die ersten atomoranen Strukturen der Trimerkomplexe des Glideosoms. Die Strukturen zeigen, dass der C-Terminus von PfELC, der im ungebundenen Zustand teilweise ungeordnet ist, auch im Komplex ein gewisses Maß an Flexibilität beibehält, was zu einem unterschiedlichen Interaktionswinkel im Vergleich zu TgELCs führt und die geringere Affinität von PfELC zu MyoA im Vergleich zu *T. gondii* erklärt. Zusätzlich haben wir Aufschluss über die Rolle von ELCs in der Regulation des Glideosoms gegeben. Wir haben gezeigt, dass die Bindung der ELCs zu MyoA durch zwei Phosphorylierungsstellen reguliert werden kann. Darüber hinaus haben wir bewiesen, dass die Bindung von Kalzium die ELC-Bindung nicht reguliert, sondern nur die Stabilität der Trimerkomplexe erhöht. Zuletzt haben wir gezeigt, dass die ELCs bei der Bindung eine Sekundärstruktur im MyoA-Hals induzieren, was zu einem steiferen MyoA-Hals führt, der folglich effizienter als Hebelarm dient und den Motorumsatz erhöht.

In der zweiten Arbeit (Manuskript 2) haben wir uns mit der Rolle eines Rhoptrienproteins *Armadillo Repeat-Only* (ARO) und mit seinem mutmaßlichen Interaktionspartner *ARO Interacting Protein* (AIP) befasst. ARO wird auf der zytoplasmatischen Membranseite der Rhoptrien durch Palmitoylierung und Myristoylierung verankert. Für *T. gondii* wurde schon gezeigt, dass ARO für die korrekte Positionierung von Rhoptrien im apikalen Pol essentiell ist und dass es mit AIP1, Myosin F (MyoF) und Adenylatzyklase Beta (ACβ) interagiert. Wir sind deshalb der Frage nachgegangen, ob ein ähnliches Arrangement auch in *P. falciparum* existiert.

Dazu haben wir die Kristallstruktur von PfARO gelöst, die mit fünf sogenannten „Armadillo repeat“ Motiven eine bohnenförmige Konformation annimmt. Wir identifizierten das *Plasmodium*-Homolog von *T. gondii* AIP1 und zeigten dass die AIP1 Proteine nur in *Apicomplexa* vorhanden sind und ihre strukturelle Homologen in anderen *Taxa* fehlen. Weitere Untersuchungen der Proteinlokalisation, auch unter Verwendung von Hochauflösender Mikroskopie, zeigten, dass PfARO in *P. falciparum* nicht zusammen mit PfAIP lokalisiert ist, da PfARO in Rhoptrien-Bauch lokalisiert ist, während AIP1 in Rhoptrien-Hals vorhanden ist. Andererseits führt die Deletion der konservierten Schleife von PfARO,

die dank der gelösten Struktur identifiziert wurde, zu einer Fehllokalisierung von PfAIP. Dies deutet darauf hin, dass die beiden Proteine trotz allem assoziieren. Schließlich zeigte der bioID Pulldown-Assay, dass PfAIP, wie auch bei *T. gondii*, mit PfACβ und MyoF interagiert, jedoch nicht mit PfARO. Basierend auf diesen widersprüchlichen Ergebnissen schlagen wir zwei Modelle der PfARO-PfAIP Interaktion vor, die mit unseren Daten übereinstimmen.

In der dritten Arbeit (Manuskript 3) wird die Struktur und Funktion der Glykogensynthase-Kinase 3 von *P. falciparum* (PfGSK3) untersucht. Humanes GSK3 reguliert verschiedene Prozesse wie den Glykogenstoffwechsel, den Zellzyklus und das Wachstum, die Translation, die Embryonalentwicklung oder die Differenzierung von Neuronen. Andererseits ist die Rolle von PfGSK3 schwer fassbar. Es wurde gezeigt, dass PfGSK3 einen Einfluss auf die Invasion von Parasiten hat. Während des Eindringens in den Wirtszellen bildet der Parasit sogenannte *moving junction*, eine Grenzfläche zwischen der parasitären Zelle und der Wirtszelle, durch die der Parasit in der Wirtszelle gleitet. Das Protein, das einen Großteil der *moving junction* bildet, ist das apikale Membranantigen 1 (AMA1), das mit dem Rhoptrien-Hals (RON) auf der Oberfläche der roten Blutkörperchen interagiert. AMA1 ist ein Transmembranprotein, das sich auf der Oberfläche von Plasmodium-Spezies befindet. Es besitzt eine N-terminale extrazelluläre Domäne und eine kurze C-terminale zytoplasmatische Sequenz. Die Phosphorylierung dieser zytoplasmatischen Sequenz reguliert die Funktion von AMA1 und ist für den Invasionsprozess von entscheidender Bedeutung. Die C-terminale Sequenz von AMA1 wird zuerst durch Proteinkinase A (PKA) und anschließend auch durch PfGSK3 phosphoryliert. Folglich wurde gezeigt, dass die Inhibitoren von GSK3 die Invasion der Wirtszellen durch den Parasiten beeinträchtigen. Obwohl PfGSK3 ein wichtiges Wirkstoffziel ist, sind Struktur und Regulationsmittel unbekannt.

In unserer Untersuchung wurden zwei Faktoren entdeckt, die die Aktivität von PfGSK3 regulieren. Ich habe festgestellt, dass PfGSK3 durch Autophosphorylierung im Protein selbst Aminosäurereste in der sogenannten Aktivierungsschleife und am N-Terminus phosphoryliert. Wir haben gezeigt, dass der N-Terminus von PfGSK3 für die Funktion unverzichtbar ist und dass die Menge der N-terminalen Phosphorylierung positiv mit der PfGSK3-Aktivität korreliert. Zusätzlich fanden wir, dass zweiwertige Schwermetalle-Ionen, wie z. B. Zink und Kupfer Ionen, die reversible Bildung von PfGSK3-Spezies mit einem hohen Molekulargewicht induzieren. Diese Spezies sind heterogen und enzymatisch inaktiv. Unsere Arbeit liefert daher erste Einblicke in Prozesse, die möglicherweise die Funktion von PfGSK3 *in vivo* regulieren.

List of publications

S. Pazicky, K. Dhamotharan, K. Kaszuba, H. Mertens, T. Gilberger, D. Svergun, J. Kosinski, U. Weininger, C. Löw (2020): Structural role of essential light chains in the apicomplexan glideosome. Posted on BioRxiv and accepted in Communications Biology. doi: 10.1101/867499

M. Geiger, C. Brown, J. S. Wichers, J. Strauss, A. Lill, R. Thuenauer, B. Liffner, L. Wilcke, S. Lemcke, D. Heincke, **S. Pazicky**, A. Bachmann, C. Löw, D. Wilson, M. Filarsky, P. C. Burda, K. Zhang, M. Junop, T. W. Gilberger (2020): Structural insights into PfARO and characterization of its interaction with PfAIP. Journal of Molecular Biology 432(4), 878-896. doi: 10.1016/j.jmb.2019.12.024

J. Pieprzyk., **S. Pazicky**, C. Löw (2018): Transient Expression of Recombinant Membrane-eGFP Fusion Proteins in HEK293 Cells. Methods in Molecular Biology, Mol Biol. 1850, 17-31. doi: 10.1007/978-1-4939-8730-6_2

T. F. Custódio, H. Das, D. J Sheward, L. Hanke, **S. Pazicky**, J. Pieprzyk, M. Sorgenfrei, M. Schroer, A. Gruzinov, C Jeffries, M. Graewert, D. Svergun, N. Dobrev, K. Remans, M. A. Seeger, G. M. McInerney, B. Murrell, B. M. Hällberg and C. Löw (2020): Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. Posted on BioRxiv and accepted in Nature Structural and Molecular Biology. doi: 10.1101/2020.06.23.165415

Scientific contribution to the manuscript

Contribution to the manuscripts that are part of this thesis

1. **S. Pazicky**, K. Dhamotharan, K. Kaszuba, H. Mertens, T. Gilberger, D. Svergun, J. Kosinski, U. Weininger, C. Löw (2020): Structural role of essential light chains in the apicomplexan glideosome. Posted on BioRxiv and accepted in Communications Biology.

As the first author of this publication, I designed, performed and evaluated the vast majority of the experiments, I communicated the results with the co-authors and in co-operation with them, designed the experiments that were performed by them. At the same time, I analyzed and interpreted the majority of the results. In the end, I was responsible for the conceptualization and writing of the manuscript and creation of the figures.

2. M. Geiger, C. Brown, J. S. Wichers, J. Strauss, A. Lill, R. Thuenauer, B. Liffner, L. Wilcke, S. Lemcke, D. Heincke, **S. Pazicky**, A. Bachmann, C. Löw, D. Wilson, M. Filarsky, P. C. Burda, K. Zhang, M. Junop, T. W. Gilberger (2020): Structural insights into PfARO and characterization of its interaction with PfAIP. Journal of Molecular Biology 432(4), 878-896.

For this publication, I have adapted the purification strategy of biotinylated proteins in *P. falciparum* samples, including the optimization of the washing procedure, sample tubes and optimization of the negative control. I was responsible for the communication of the results with the Proteomics Core Facility at European Molecular Biology Laboratory and delivery of the result to the main authors.

3. **S. Pazicky**, A. Alder, M. Killer, E. Round, L. Hauke, T. Gilberger and C. Löw (2020): N terminal autophosphorylation regulates the activity of Plasmodium falciparum GSK3. Unpublished manuscript.

In this manuscript, as the first author, I have designed, performed and evaluated the majority of the experiments. I organized the communication with the co-authors and scientific facilities, and co-designed the experiments performed by them. I have conceptualised and written the manuscript and assembled the figures.

Contribution to the manuscripts that are not part of this thesis

4. J. Pieprzyk., **S. Pazicky**, C. Löw (2018): Transient Expression of Recombinant Membrane-eGFP Fusion Proteins in HEK293 Cells. Methods in Molecular Biology, Mol Biol. 1850, 17-31.

As the second author of this publication, I have written the "Introduction" and "Methods" sections of the paper and consulted the presented results with the other authors, focusing on the clarity of the text in context of my previous experience with the topic.

5. T. F. Custódio, H. Das, D. J Sheward, L. Hanke, **S. Pazicky**, J. Pieprzyk, M. Sorgenfrei, M. Schroer, A. Gruzinov, C Jeffries, M. Graewert, D. Svergun, N. Dobrev, K. Remans, M. A. Seeger, G. M. McInerney, B. Murrell, B. M. Hällberg and C. Löw (2020): Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. Posted on BioRxiv and accepted in Nature Structural and Molecular Biology.

In this work, I helped with the expression of the recombinant virus proteins, selection and the purification of nanobodies, performed their characterization by circular dichroism and was responsible for the preparation of the samples for small-angle X-ray scattering measurements. In the discussion with the co-authors, I have contributed to the overall workflow. Additionally, I have written the "Introduction" section of the manuscript.

Abbreviations

ACβ	Adenylate cyclase beta
ACT	Artemisinin-based combination therapy
AIP	ARO interacting protein
AMA1	Apical membrane antigen 1
AP1	Adaptor protein 1
APH	Apical plekstrin homology protein
ARO	Armadillo Repeat-Only
ATP	Adenosine triphosphate
bioID	Proximity-dependent biotin identification
CaM	Calmodulin
cAMP	Cyclic adenosine monophosphate
CAP	Cyclase-associated protein
CDPK	Calcium-dependent protein kinase
cGMP	Cyclic guanosine monophosphate
CSP	Circumsporozoite surface protein
CTRP	Circumsporozoite- and TRAP-related protein
CyRPA	Cysteine-rich protective antigen
DAG	diacylglycerol
DGK	diacylglycerol kinase
DrpB	Dynamin-related protein B
EBA-175	Erythrocyte binding antigen 175
ELC	Essential light chain
IMC	Inner membrane complex
IAP1	IMC-interacting protein 1
IP_3	Inositol-3-phosphate
FRM2	Formin 2
GAP	Glideosome-Associated Protein
GAPM	Glideosome-Associated Protein with multiple membrane spans
GSK3	Glycogen synthase kinase 3
GTP	Guanosine triphosphate
high-MW species	Species of high molecular weight
HSP90	Heat shock protein 90
MIC	Microneme protein
MLC	Myosin light chain
MSP	Merozoite surface protein
MTIP	Myosin tail interacting protein
MTRAP	Merozoite thrombospondin-related anonymous protein
Myo	Myosin
PIP_2	Phosphatidylinositol 4,5-bisphosphate
P.	*Plasmodium*
PA	Phosphatidic acid
Pf	*Plasmodium falciparum*
PhiL1	Photosensitized INA-labelled protein 1
PI	Phosphatidylinositol
PKA	Protein kinase A
PKA-C	Catalytic domain of PKA
PKA-R	Regulatory domain of PKA
PKB	Protein kinase B
PKG	Protein kinase G
PLC	Phospholipase C
PLP	Perforin-like protein
PNP	Purine nucleoside phosphorylase

PTRAMP	*Plasmodium* thrombospondin-related apical merozoite protein
RAP	Rhoptry-associated protein
RBC	Red blood cell
Rh5	Reticulocyte-binding protein homologue
RIPr	Rh5 interacting protein
RON	Rhoptry neck protein
SAG	Surface antigen
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SIM	Super-resolution microscopy
T.	*Toxoplasma*
Tg	*Toxoplasma gondii*
TRAP	Thrombospondin-related anonymous protein

1. Introduction

1.1. Apicomplexa

Apicomplexa are a diverse phylum of unicellular organisms that are highly relevant as causative agents of fatal human and animal diseases [1,2]. All apicomplexans, with the exception of symbiotic *Nephromyces*, are obligate intracellular parasites, meaning that they need a host cell environment in order to complete their life cycle [3]. The life cycle of apicomplexans is typically very sophisticated because the parasites exist as distinct life stages with variable cellular morphology and host cell requirements. Moreover, the apicomplexans usually switch between two host organisms, undergoing sexual reproduction in one and the asexual reproduction in the other [1]. As eukaryotic organisms, they contain typical eukaryotic organelles, including mitochondria and a type of plastid called apicoplast. The presence of apicoplast, together with the apical complex, are typical characteristics that define the phylum *Apicomplexa* [1].

1.1.1. Classification

Taxonomically, *Apicomplexa* fall within the kingdom *Chromista* and infrakingdom *Alveolata* [4,5]. *Alveolata* are single cell organisms with characteristic membrane vesicles, termed alveoli, that are located just underneath their outer plasma membrane. Besides apicomplexans, the phylum *Alveolata* accommodates many ubiquitous protists such as ciliates and dinoflagellates. *Apicomplexa*, besides their typical organelle structures, are characterised by their unique proteome [6]. More than 60% of their gene sequences are unique to *Apicomplexa*, while most of them are further specific to smaller taxa within Apicomplexa (Figure 1). Apicomplexans can be further subdivided into two classes: *Aconoidasa* and C*onoidasa* (Figure **1**) [7]. As the name suggests, the typical characteristics of *Aconoidasa* is a lack of conoid, which is a rigid tubulin-derived structure located at the tip of apicomplexans [8]. The most notorious representatives of this class are the *Plasmodium* species, pathogens that cause malaria [9]. Other members of *Aconoidasa*, the *Babesia* species, represent the second most common parasites that infect animals [10]. Although rarely infecting humans, they cause babesiosis in cattle and other mammals. Similarly, the *Theileria* species are common cattle parasites in Africa that cause East Coast fever [11]. On the other hand, *Conoidasa*, apicomplexans that do contain a conoid, are further divided into two subclasses: *Coccidia* and *Gregarinia*. Gregarines typically infect gut epithelium of segmented worms, arthropods and molluscs [12], while *Coccidia* are able to infect, besides gut epithelium, any nucleated cells. The main difference between the two subclasses lies in the morphology of their gamonts

(gamete producing cells): while coccidian gamonts are small and intracellular, gamonts in *Gregarinia* are large and extracellular, and contain additional organelles, such as epimerites (hook-like anchoring structures) and mucrons (similar adhesive organelles) [13,14]. *Toxoplasma* species are the most studied coccidians as they infect up to 50% of humans and pose a health risk to immunocompromised individuals or pregnant women. Moreover, *Toxoplasma gondii* was adopted as a model apicomplexan because of the relative ease of genetic manipulation and propagation in the cell cultures [15]. Other *Coccidia*, *Cryptosporidium* species, are of bigger interest as they are the leading cause of enteric diseases in the developing world [16]. The *Eimeria*, *Neospora* and *Sarcocystis* species are also relevant for humans because they infect cattle and poultry [17].

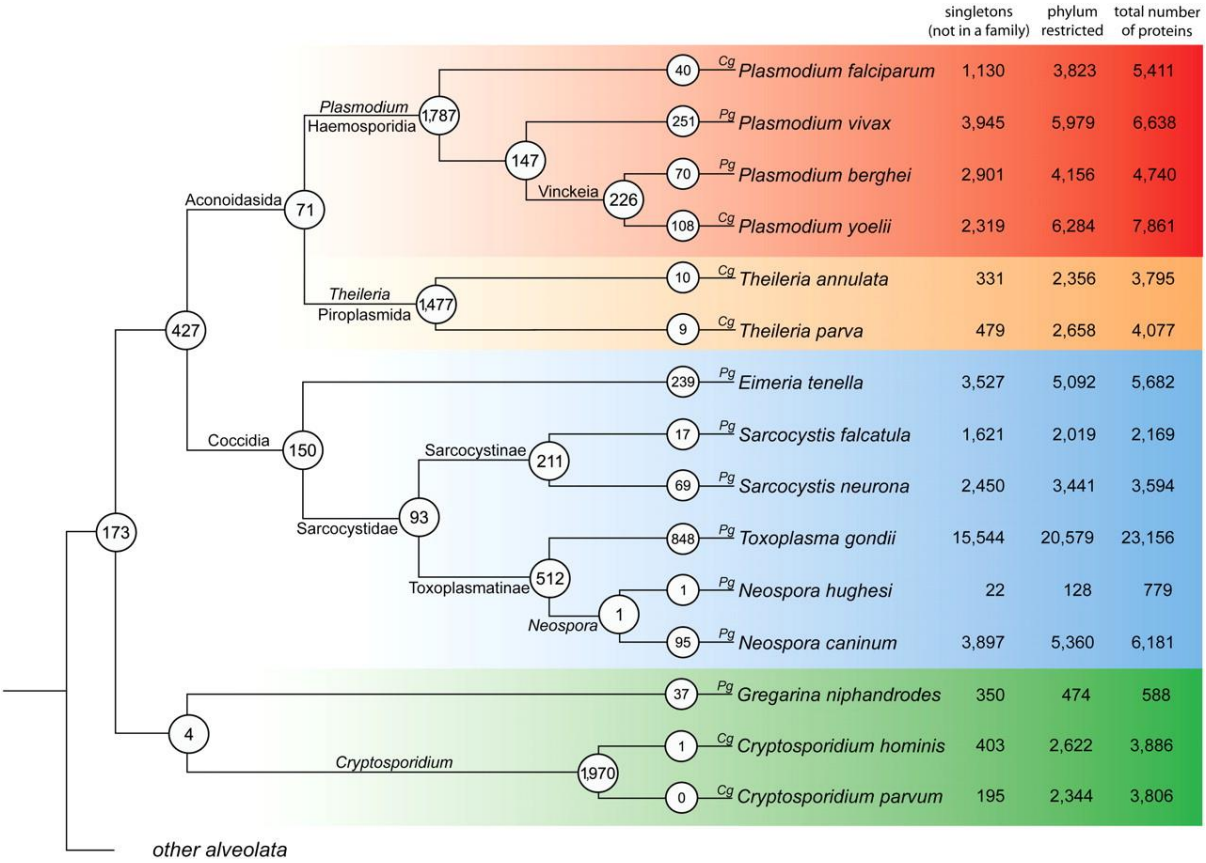


Figure 1. Conservation of protein families across apicomplexan phyla. The numbers in the nodes represent the number of conserved protein families that are shared between the two daughter phyla. Number of singletons (proteins that do not belong to any family), phylum-restricted proteins and total number of proteins are given in the columns. Abbreviations Cg and Pg refer to completely or partially sequences genome, respectively. Figure from Wasmuth et al, 2009.

1.1.2. *Plasmodium falciparum* life cycle

Apicomplexans go through a complex life cycle, typically switching between two hosts (Figure 2). *Plasmodium* species that infect humans are typically transmitted by the mosquitoes from the genus *Anopheles*, whereas other mosquito genera, including *Anopheles*, can also transmit the parasite to

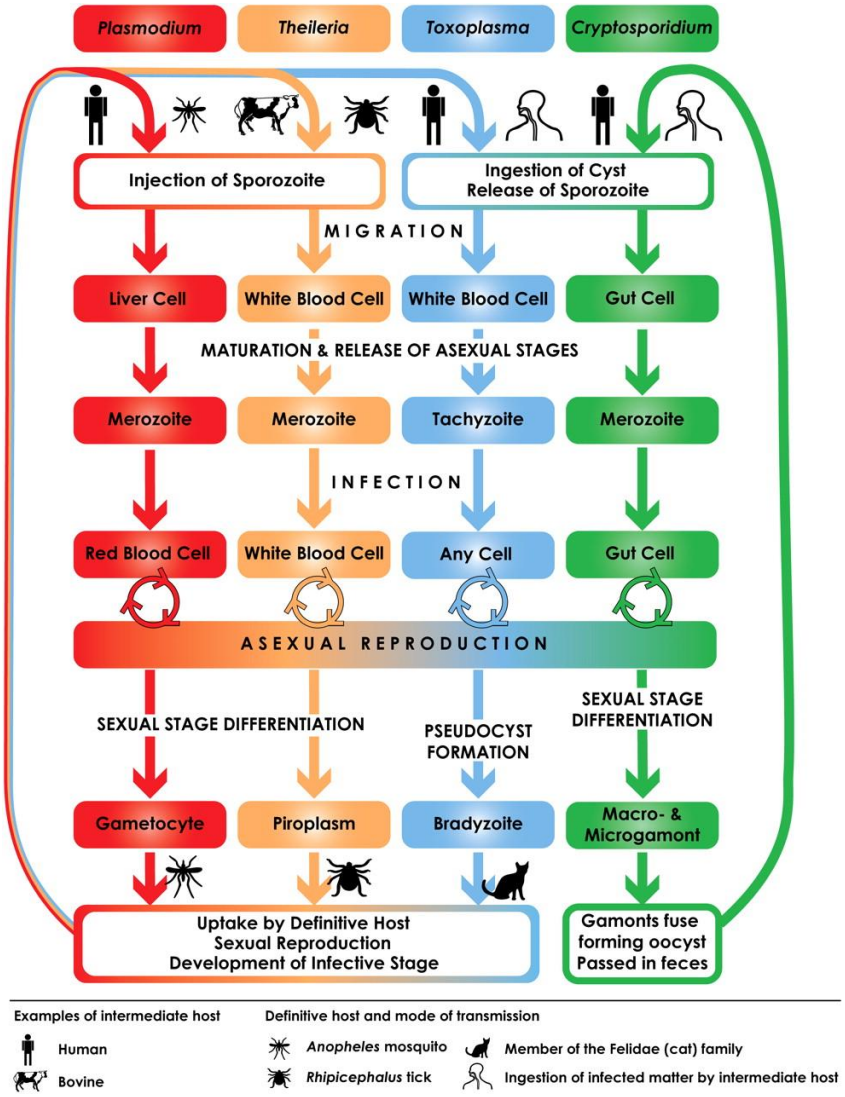


Figure 2: Life cycle of selected apicomplexan organisms. Most apicomplexans require two hosts to undergo asexual and sexual reproduction, but *Cryptosporidium* suffices with a single host. In general, upon infection, the parasites maturate in one host cell and release motile asexual stages. Those undergo rapid reproduction by schizogony in other host cells. Upon differentiation, the asexual forms develop into gametes that need another host organism to undergo fertilization and development of infective stages. Figure from Wasmuth et al, 2009.

animals [18]. The mosquitoes usually feed on nectar, but as oogenesis takes place in the female mosquitoes, they require additional nutrition in the form of human blood. During their blood meal, the female mosquitoes use saliva to prevent blood coagulation. If the mosquitoes are infected with *Plasmodium*, they release about a hundred *Plasmodium* sporozoites with their saliva into the subcutaneous tissues of humans. To move across tissues to reach the blood stream, the sporozoites use a type of motility that is specific for apicomplexans, called gliding [19,20]. In the blood stream, the sporozoites are transported to the liver, where they pass through Kupfer and endothelial cells to reach hepatocytes, in which they undergo merogony (commonly referred to as schizogony) [21]. During schizogony, the parasitic nucleus divides first, which is followed by cell segmentation. One parasite

divides into more than hundred schizonts that fill up the entire hepatocyte before they egress from the host cell and escape back into the blood stream in form of another motile stage, merozoites. Alternatively, some *Plasmodium* species can enter another stage, called hypnozoites, that can remain dormant for up to 30 years from the initial infection [18]. In the blood stream, merozoites actively invade red blood cells, where they develop to ring stages and then to trophozoites, large metabolically active cells that undergo schizogony to produce 8-16 merozoites. Consequently, the merozoites egress from their host cell to invade more red blood cells. The evasion from the host cells thus occurs cyclically and causes a periodic occurrence of symptoms in patients, such as fever, headache and muscle pain [18]. After host cell re-invasion, a small portion of merozoites develops into gametocytes that further undergo five stages of development inside the red blood cell. With another mosquito blood meal, the gametocytes can get into the mosquito midgut, where they differentiate upon decrease of temperature and increase of pH into female and male gametes [22]. The gametes subsequently go through a relatively long process of sporogony. First, the gametes undergo fertilization to form a diploid zygote. Then, the zygote develops into a motile ookinete that penetrates the midgut wall and takes up to 16 days to mature into oocysts. In the end, the oocyst releases infectious sporozoites that glide to the salivary glands, where they are ready to be passed to another human.

1.1.3. *Toxoplasma gondii* life cycle

The life cycle of *T. gondii* is simpler because the parasite is able to infect any nucleated cell [2,23]. The definitive hosts of *T. gondii* are cats that shed the oocysts in their faeces. These usually get in the gastrointestinal tract of other mammalian or avian animals, typically mice. In the stomach or intestines, the oocysts release motile sporozoites that invade the gastrointestinal epithelium and undergo endodyogony to differentiate into tachyzoites that escape the host cells, causing acute toxoplasmosis. Tachyzoites are able to differentiate into larger and less immunogenic bradyzoites that slowly grow in the cysts. Bradyzoites can be either ingested by another non-feline host, where they are able to differentiate back into tachyzoites and continue asexual reproduction; or they can be ingested by a feline host. In such cases, the cyst wall is destroyed by gastric enzymes and the bradyzoites are released to undergo self-limiting number of asexual reproduction cycles to finally undergo gametogony – the formation of gametes. Compared to *P. falciparum*, *T. gondii* fertilization and formation of the oocysts take place in the host cells. The oocysts then disrupt the host cell and are released with feline faeces to continue asexual reproduction in another host [2,23].

1.2. Morphology of motile apicomplexan stages

Because the shape and subcellular structures of the apicomplexans serve stage-specific roles, all motile stages of the apicomplexan life cycle (also called zoites) share a large set of specific features that are associated with host cell egress, motility and host cell invasion (Figure 3) [8,24–27]. The motile sporozoites, merozoites and ookinetes in *Plasmodium*, and sporozoites, tachyzoites and bradyzoites in *Toxoplasma*, have polarized cells, meaning that their cellular components are asymmetrically distributed [17]. Along their long axis (see Figure 3), the organelles follow a specific distribution from

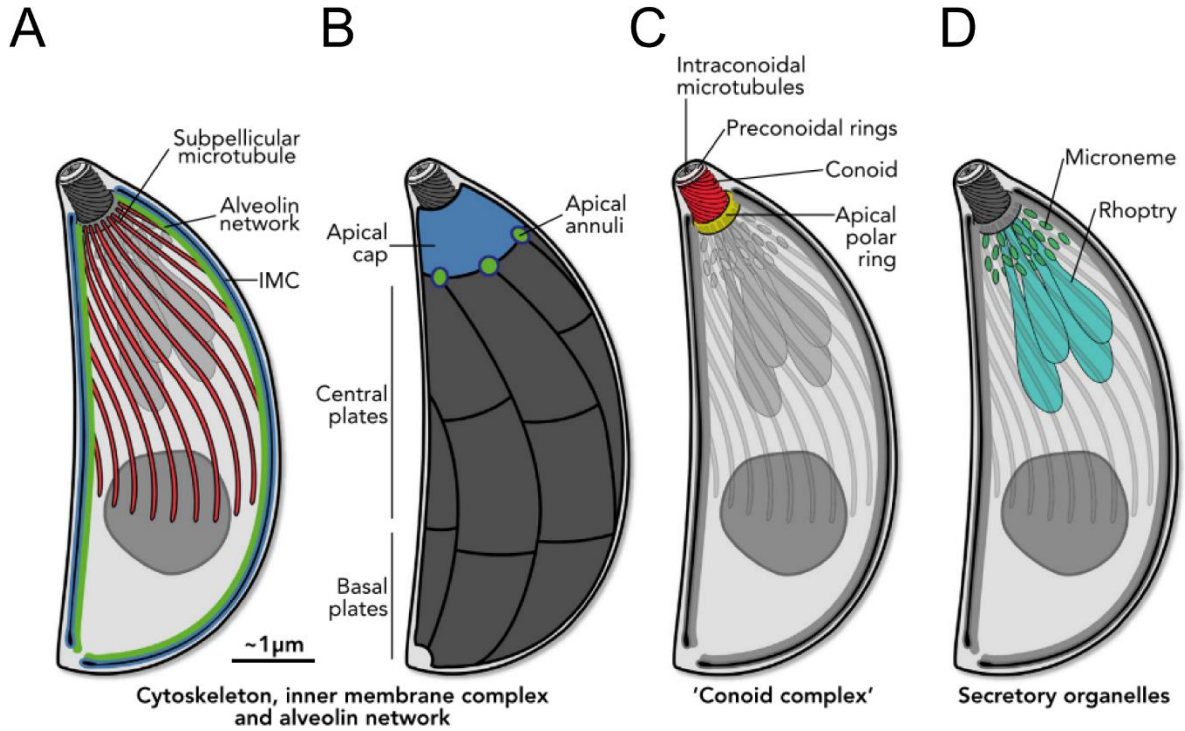


Figure 3. Subcellular structures typical for motile apicomplexan cells. The motile apicomplexan stages are here exemplified by *T. gondii* tachyzoites. (A)two thirds of the parasite length are supported by subpellicular microtubules that underlay the inner membrane complex (IMC). The IMC is composed of flattened vesicles that underlay the plasma membrane and are attached to the subpellicular microtubules through an alveolin network.(B)the IMC vesicles can be divided to apical cap, central plates and basal plate that are sutured together by a mesh of proteins. Special sutures called apical annuli connect the apical cap with the rest of the IMC.(C) The apical tip of the coccidian parasites possesses a typical microtubule structure called the conoid. The conoid is usually associated with further microtubule forms called apical polar rings and preconoidal rings. (D) Secretory organelles that discharge important virulence factors upon parasite egress, gliding and host cell invasion, locate to the anterior pole of apicomplexan parasite. These are long pear-shaped rhoptries and small rod-like micronemes that associate with the subpellicular microtubules. Figure from Pacheco at al, 2020.

apical tip (anterior pole) to basal end (posterior pole). The apical tip is defined by the location of the apical complex, a phylum-defining feature that consists of the following organelles: (a) a peripheral membrane vesicle called the apical cap, (b) conoid, a tubulin-rich structure, present only in *Conoidasa*, often associated with additional tubulin rings, and (c) specialized secretory vesicles (rhoptries, micronemes and dense granules), located inside and behind the conoid [8]. The conoid is a rigid structure composed of tightly packed microtubule filaments that form a shape of a hollow cone (Figure

4) [8]. The conoid sits below the apical polar ring, but during cell egress and invasion, it rapidly protrudes through the apical polar ring [28]. At the same time, the apical polar ring serves as a unique microtubules organizing centre [27]. Several microtubules, their number depending on the organism and tubulin expression level, radiate from the apical polar ring towards the posterior end of the parasite, reaching about two third of the length of the apicomplexan cell [29,30]. These microtubules have been termed subpellicular microtubules, because they are located close to the periphery of the parasite cell and they delineate pellicle [27]. Pellicle is an umbrella title for all subcellular structures located between the subpellicular microtubules and parasite plasma membrane. It is dominated by alveoli, large flattened vesicles located just between the subpellicular microtubules and plasma membrane (Figure 3) [31]. The alveoli and their residing proteins are collectively known as the inner membrane complex (IMC) . The inner membrane of IMC is tethered to subpellicular microtubules by a meshwork of proteins of the alveolin family [32]. These associate with IMC either *via* lipidation motifs that directly anchor them in the inner IMC membrane, or presumably further interact with other proteins, such as photosensitized INA-labelled protein 1 (PhiL1) or glideosome-associated proteins with multiple membrane spans (GAPMs [33,34]. Interestingly, many IMC proteins are not homogeneously scattered in the IMC membranes, but form a pattern of regularly interspaced detergent-resistant foci along the long axis of the parasite [35,36]. Although the individual IMC vesicles are sutured together, three distinct populations can be distinguished based on their location and specific proteins that they accommodate: basal plates, central plates and the above-mentioned apical cap [37] (Figure 3). On the other side of the IMC, the space between the outer IMC membrane and plasma membrane is accommodated by an actomyosin motor that is crucial for the host cell egress, motility and cell invasion [38]. The actomyosin motor relies on a network of IMC proteins that anchor myosin to the microtubules, while actin is immobilized by plasma membrane receptors that bind the host cell surface proteins. The majority of the plasma membrane receptors and even some of the host cell surface proteins are trafficked and secreted through secretory organelles located in the apical complex [24,25,39].

1.2.1. Secretory organelles of zoites

The main apicomplexan secretory organelles are rhoptries, micronemes and dense granules (Figure 4). Rhoptries and micronemes empty their contents at the very apical tip of the parasite prior to and during host cell invasion, while dense granules only play a role in host cell modification after invasion [40]. In general, the secretory organelles arise from the Golgi apparatus (similarly to the IMC vesicles); however, the process of their biogenesis still remains to be elucidated. Both the pear-shaped rhoptries and the rod-like micronemes localise near the apical tip: the rhoptries point towards the apical tip with

their narrow rhoptry neck inside the conoid, followed by lipid-rich rhoptry bulb, whereas the micronemes associate with subpellicular microtubules just below the conoid.

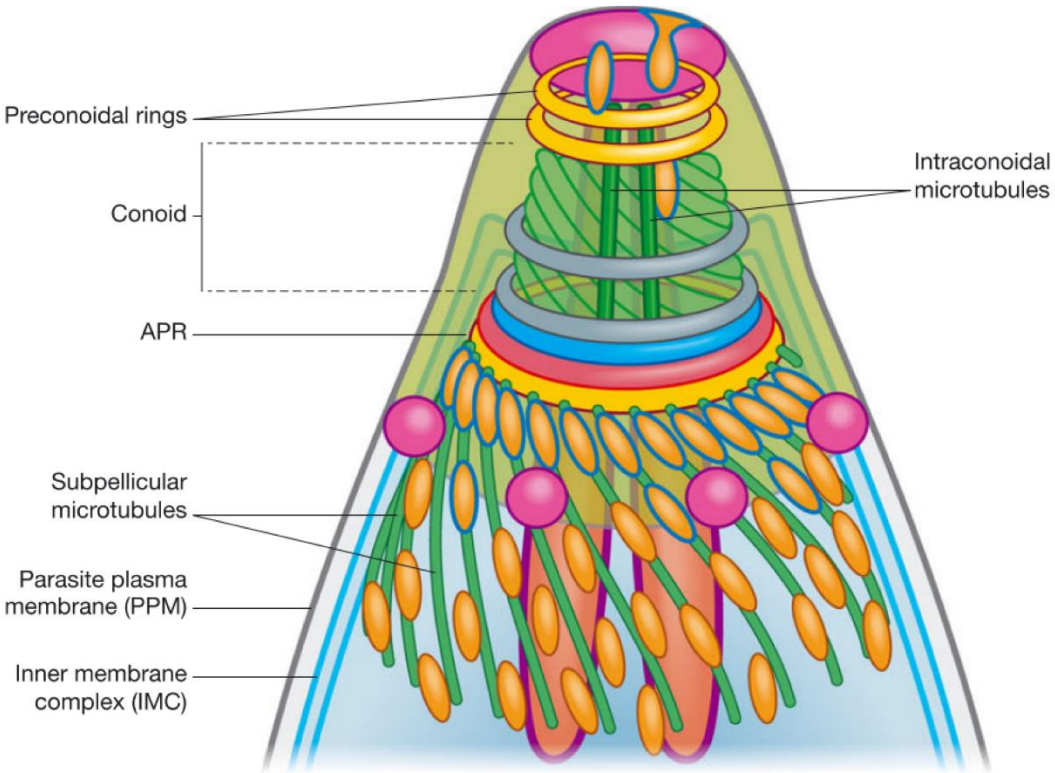


Figure 4. Anterior pole of *T. gondii* parasite. The conoid is placed in the apical tip between the preconoidal rings and the apical polar ring (APR). The apical polar ring also serves as a microtubule organizing centre, radiating several subpellicular microtubules towards the parasite posterior which delineate peripheral inner membrane complex (IMC) and parasite plasma membrane (PPM). Two intraconoidal microtubules located inside the conoid are speculated to play a role in the secretion of micronemes (depicted in yellow). Rhoptries, (depicted in pink) are divided to neck (inside conoid) and bulb (below conoid), and discharge sequentially in a currentlyunknown manner. Dense granules (depicted in magenta), only play a role after host cell invasion. Figure from Dubois et al, 2019.

A common apicomplexan zoite has approximately 10 rhoptries that locate inside the conoid in *Conoidasa* [8]. The distinction of rhoptry neck and bulb is important because they accommodate different proteins. Moreover, they are secreted sequentially: the rhoptry neck proteins are discharged prior to the invasion, whereas the less conserved bulb proteins are only secreted after the invasion and play a role in host cell modification. Micronemes, on the other hand, secrete mostly plasma membrane proteins that serve as adhesive molecules during host cell binding [25]. The adhesive molecules bind both proteins on the host cell surface, and sialic acid or collagen in the extracellular matrix [41]. In *Plasmodium*, a model has been suggested, in which the rhoptry necks fuse with the micronemes upon their discharge [42], but this is not plausible in *Toxoplasma* because the events are temporarily separated [43].

The protein trafficking into specific parts of rhoptries could be dependent on the time of gene expression during schizont development, but it remains unclear, what guides the trafficking of the proteins into rhoptries and micronemes. In many cases, the proteins form larger complexes, where the trafficking of one protein guides the trafficking of the other proteins, such as the rhoptry neck protein complex (RON complex) [44], rhoptry-associated protein complex (RAP complex) [45] and microneme proteins (MICs) [46]. Some proteins use acylation to achieve the correct localisation, opening up further possibilities for their regulation [47]. For example, armadillo repeat only protein 1 (ARO1) uses both myristoylation and palmitoylation to anchor themselves in the outer leaflet of rhoptry bulbs [48].

ARO1 is of a specific interest because its knock-out in *T. gondii* results in a faulty localisation of nascent rhoptries, suggesting that it plays a role in rhoptry biogenesis [49,50]. Moreover, several interaction partners of TgARO1 were identified: ARO interacting protein (TgAIP1), adenylate cyclase beta (TgACβ) and myosin F (TgMyoF) [49,50]. However, it is not clear if these interactions hold true in *P. falciparum*. The role of ARO and AIP1 in *P. falciparum* and their comparison to *T. gondii* analogs are investigated in the second part of this work (Manuscript 2).

1.3. Gliding motility and host cell invasion

The subcellular structures and features described above assist the unique gliding motility of the zoites, which enables them to pass through soft tissues, invade cells and egress from host cells [20]. When zoites are placed on coated glass slides, gliding can be visualized and three types of movement can be discerned: helical gliding, circular gliding and upright twirling, whilst different mutations in the parasite can lead to impairments in one, two or all gliding types [51]. Gliding is a tightly orchestrated process that is dependent on three major factors: (a) surface molecules that mediate attachment to the host cell or tissue substrate [52], (b) a motor that empowers the movement of the parasite [52–55] and (c) the secretory organelles that produce factors important for gliding and invasion [24,25,39].

1.3.1. Surface molecules in gliding and invasion

A palette of surface proteins and secreted factors is necessary for successful gliding and host cell invasion [39]. Prior to the invasion, the parasite cell first attaches to the host cell surface using a plethora of surface proteins. The most important surface proteins are P25 and P28 in ookinetes [56], merozoite surface proteins (MSPs) in merozoites [57], the highly abundant circumsporozoite surface protein (CSP) in sporozoites [58] and surface antigens (SAGs) in *Toxoplasma* tachyzoites [59]. Because of their

abundancy and exposure to immune system, many of these proteins are also targets for drug and vaccine development [60]. After initial host cell attachment, the contents of secretory organelles are discharged

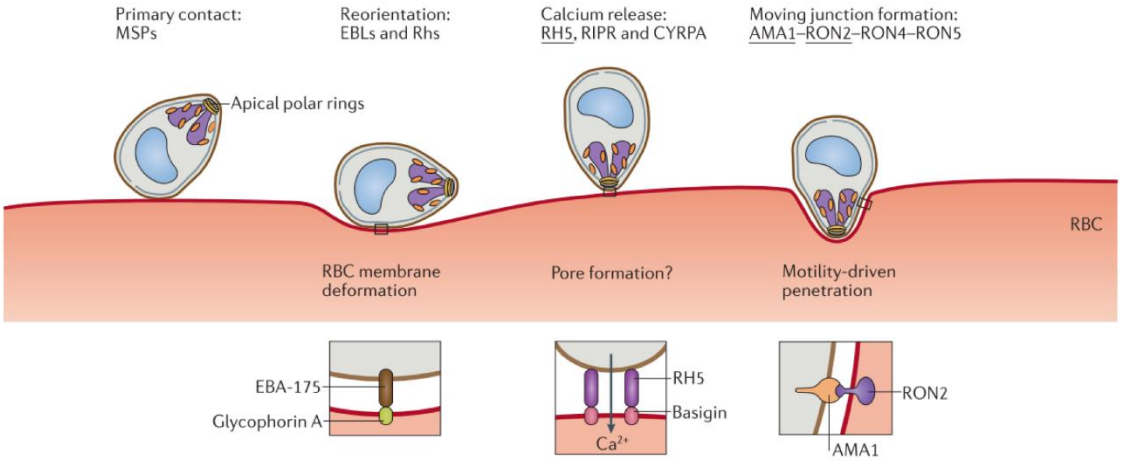


Figure 5. Simplified model of parasite reorientation. The parasites first reversibly attach to the host cell surface using less specific abundant surface receptors. Upon initial binding, the secretory organelles at the parasite`s apical tip deposit additional membrane proteins on the cell surface that mediate reorientation of the parasite (such as EBA-175 binding glycophorin A on the red blood cell surface) and juxtaposition of the apical tip at the parasite`s membrane (RH5 complex binding basigin). In the end, the moving junction is established by the interaction of AMA1 with the RON complex. mediating host cell invasion. Figure from Frénal et al, 2017.

at the apical tip, which in turn guides the reorientation of the parasite, moving the apical tip adjacent to the host cell (Figure 5) [61,62]. The secretory organelles release membrane proteins that, upon discharge, re- localise to the parasite`s plasma membrane and mediate further contacts with the host cell [39]. For example, micronemes of merozoites discharge a membrane receptor, erythrocyte binding antigen 175 (EBA-175), that binds glycophorin on the surface of the red blood cells (RBCs) [63]. As the concentration of the discharged proteins is higher at the apical tip, the parasite changes its orientation. At the apical tip, additional important cell-cell contacts are established. For example, reticulocyte-binding protein homologue 5 (Rh5, released from rhoptries) in complex with Rh5 Interacting Protein (RIPr) and Cysteine-Rich Protective Antigen (CyRPA) bind the RBC receptor “basigin” during merozoite invasion [64]. Finally, apical membrane antigen 1 (AMA1), released from micronemes to the parasite`s surface, binds the complex of rhoptry neck proteins (RON2, RON4, RON5), that is released from the rhoptries and localises to the host cell surface [65,66]. This interaction forms the basis of the so-called moving junction: as the parasite glides into the host cell, a ring forms around the entry point and at the site of cell-cell contacts, which is delineated by the proteins that form the moving junction [67]. As the parasite moves into the cell, the ring moves back relative to the parasite and thus, the proteins of the moving junction are translocated to the basal end of the parasite. Importantly, the surface proteins of the moving junction become undesirable inside the host cells,

therefore, they are subjected to shedding, a proteolytic cleavage that releases their extracellular domains, just prior to host cell entry. While gliding into the host cell, the parasite pushes the host cell plasma membrane inwards, forming a so called parasitophorous vacuole that protects the parasite from the defence mechanisms of the host cell.

1.3.1.1. AMA1 and GSK3

On their way from the ribosome to the plasma membrane, the surface molecules of the parasite undergo multiple steps of posttranslational modification that regulate their function [24,25]. The surface molecules are typically expressed in their inactive forms, presumably to prevent their cross-interaction inside the parasite [39]. In the secretory organelles and upon their deposition on the plasma membrane, the surface proteins are usually cleaved by co-localised proteases, making their binding sites accessible to their binding partners. Additionally, phosphorylation of their cytoplasmic C-termini can further impact their binding properties [68].

This is also the case for AMA1, which possesses multiple phosphorylated residues at its C-terminus [69–71]. The protein kinase A (PKA) is the primary kinase that phosphorylates the AMA1 C-terminus [69]. PfGSK3 only recognises the phosphorylated sequence of AMA1 and catalyses phosphorylation of an additional residue, thereby further increases the invasion efficiency [72]. Hence, the function of PfGSK3 in host cell invasion is known. However, the means of PfGSK3 regulation have not been investigated and PfGSK3 structure has not been determined so far. In the last part of this work (Manuscript 3), I provide a detailed analysis of the possible means of PfGSK3 regulation by N-terminal phosphorylation, autophosphorylation and binding of heavy metal ions.

1.3.2. The invasion motor

While the parasite surface proteins guide attachment to the extracellular matrix or the host cell and ensure the correct orientation of the parasite towards the host cell, parasite motility and invasion are powered *via* actomyosin motor that is located in the narrow space between the parasite`s plasma membrane and the inner membrane complex (Figure 6) [73–76]. Myosin A (MyoA) is attached to the IMC, while short actin filaments are immobilized to the plasma membrane [77]. MyoA and several other proteins form the glideosome, a protein complex that mediates the attachment of MyoA to the outer leaflet of IMC [53]. The glideosome is presumably further interconnected with other IMC proteins that in turn anchor MyoA to the microtubules that lie beneath the IMC membranes (Figure 3). The second component of the actomyosin motor, the short actin filaments, are attached to the C-terminal tails of plasma membrane receptors through glideosome associated connector (GAC) [78]. These

plasma membrane receptors bind to the surface of the host cells or molecules of extracellular matrix, which in turn mediates the immobilization of actin. As myosin is anchored to the intracellular structures of the parasite, but actin is attached to the extracellular proteins, the movement of myosin along the actin filaments generates relative movement of the parasite against the extracellular substrate.

1.3.2.1. Actin and its attachment to the host cell surface

The role of actin in gliding and invasion was recognised early because molecules that interfere with the actin polymerization, such as cytochalasin D, inhibit invasion and cause aberrations in the gliding [79]. Actin is a globular protein (G-actin) that has the ability to form filaments (F-actin) by polymerization. It is encoded by one gene in coccidians (e.g. *Toxoplasma*), but by two genes (*Act1* and *Act2*) in *Aconoidasa* (e.g. *Plasmodium*) [80]. Presumably because of the low sequence similarity to other actin proteins, apicomplexan actin filaments tend to be unstable and short, and therefore are hard to visualize in the living parasites. Only recently, using fluorescent antibodies, it was shown that actin filaments co- localise with the moving junction and that during the invasion, they re-localise from the anterior to the posterior pole [81,82]. Formin 2 (FRM2), a protein that mediates actin nucleation and is essential for actin polymerization, also co- localises with actin and the moving junction [82].

Actin does not directly interact with the plasma membrane proteins of the parasite, but requires a connecting protein to mediate this interaction. Aldolase, the enzyme involved in glycolysis, was first incorrectly identified as the mediator of actin immobilization [83,84]. Only later, after the role of aldolase in gliding was disputed, glideosome-associated connector (GAC) was discovered both in *T. gondii* and *P. falciparum*. The N-terminus of TgGAC was shown to bind to actin, whereas the C-terminal domain binds the cytosolic tail of microneme protein 2 (MIC2) [78].

MIC2 is a membrane protein from the family of thrombospondin-related anonymous proteins (TRAPs) that is crucial for apicomplexan invasion [85]. MIC2 possesses an extracellular N-terminus bearing several adhesion domains, one transmembrane helix and a short C-terminus that mediates the interaction with GAC. Binding partners of most TRAP and TRAP-like proteins are not known, with the exception of the abundant RBC surface protein semaphorin A that binds *Plasmodium* thrombospondin-related apical merozoite protein (PTRAMP) [86]. Other proteins of TRAP protein family are found in other *Plasmodium* stages, such as merozoite TRAP (MTRAP) in *Plasmodium* merozoites [87] , TRAP protein in *Plasmodium* sporozoites [88,89] and circumsporozoite- and TRAP-related protein (CTRP) in *Plasmodium* ookinetes [90]. Moreover, other proteins of the moving junction, such as AMA1, are assumed to use their C-terminal tails to bind GAC or other mediators of actin attachment [83,91]. As the parasite moves forward during invasion, actin remains attached to the surface proteins that form the

moving junction, which relocates towards the posterior of the parasite relative to the rest of the parasite cell.

1.3.2.2. Myosin A and the glideosome

The immobilization of actin on the host cell surface or extracellular matrix substrate enables myosins to drive the parasite along this substrate. In general, myosins are molecular motors that are able to "walk" on actin filaments, using the energy from ATP hydrolysis [92]. Structurally, they consist of an N-terminal head domain (or motor domain), that binds actin filaments and ATP. The head domain is followed by the neck domain bearing IQ domains that enable binding and regulation of myosins by myosin light chains. Finally, the myosin C-terminal tail domains are typically long coiled-coils that usually form dimers or higher oligomers and bind a substrate that is transported along the actin filaments [92]. In apicomplexans, the myosin involved in invasion, MyoA, is classified in the class XIV myosin family that lacks the entire tail domain [73,93]. Consequently, the role of the MyoA in the invasion is different from that of the classical myosins. Instead of transporting a substrate along actin filaments, its C-terminus is fixed and the N-terminal head domain shifts the short actin filaments rather than "walking" on them. The MyoA C-terminal neck domain bears two degenerate IQ motifs and therefore binds two atypical light chains: a distal myosin light chain 1 (MLC1, also termed myosin tail-interacting protein or MTIP in *Plasmodium*) [94], and an essential light chain (ELC) [95–97]. MyoAs of both *P. falciparum* and *T. gondii* have been exhaustively studied. Recently, the motor domain structure of both PfMyoA and TgMyoA have been determined, revealing unique inter-domain interactions that maintain the chemomechanical coupling [98,99]. *In vitro* functional assays have also shown that the speed, at which TgMyoA or PfMyoA shift the actin filaments, correlates with the speed of apicomplexan gliding, supporting the hypothesis that they are the main motors empowering the gliding motility [100,101].

Interestingly, the efficiency of apicomplexan MyoAs double when ELCs are bound [97,102]. In conventional myosins, ELCs are thought to regulate the myosin motor *via* interaction with the converter domain of myosin head and stabilization of its interaction with the motor domain. At the same time, the binding of ELCs is supposed to stiffen the myosin lever arm, although no mechanism of such stiffening on molecular level has been proposed so far [103]. Structurally, the conventional ELCs are part of the calmodulin family, typically consisting of an N-terminal and a C-terminal lobe. Each lobe consists of two pairs of α-helices, while each helix pair has a so-called EF hand placed between them, which is a loop with calcium binding properties. This enables further regulation of classical myosins by calcium ions [104,105]. Interestingly, in apicomplexan atypical ELCs, the ability to bind calcium ions is partially

or completely lost [102]. In *P. falciparum*, only one ELC (PfELC) has been identified [97], and all of its EF hands bear a degenerate sequence that prevents them from binding calcium. In *T. gondii*, two ELCs have been identified (TgELC1 and TgELC2), and both of them only maintain a single functional EF hand [106]. Moreover, different MyoA binding sites have been identified for TgELCs and PfELC, although the sequence of MyoA tail domains are conserved between *T. gondii* and *P. falciparum* [96,106,107]. These intriguing differences between the apicomplexan ELCs are tackled in the largest part of this work (Manuscript 1). Importantly, recent data shows that PfELC is essential in the host cell

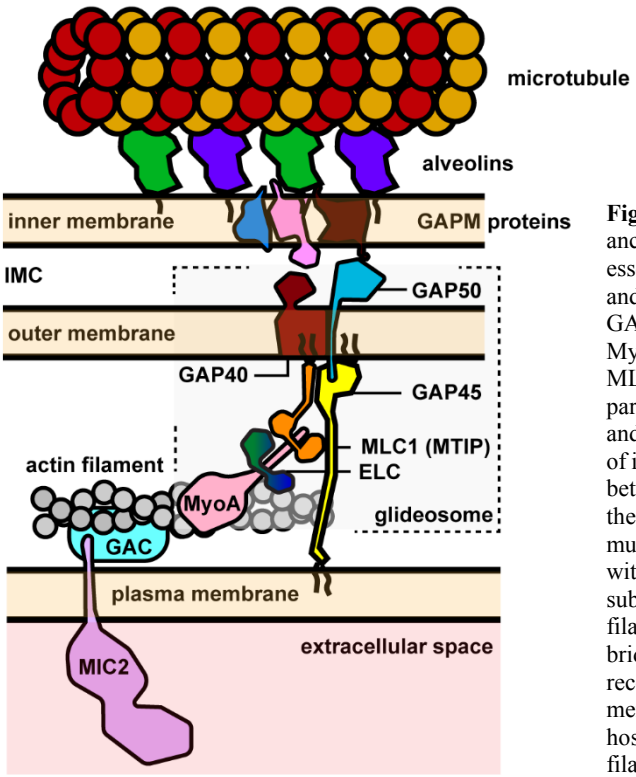


Figure 6: Glideosome in the pellicle. Glideosome complex anchors myosin MyoA in the IMC. It consists of MyoA, essential light chain (ELC), myosin light chain 1 (MLC1) and three glideosome associated proteins (GAP40, GAP45, GAP50). In linear glideosome model, the C-terminus of MyoA interacts with ELC and MLC1 and N-terminus of MLC1 interacts with C-terminus of GAP45. GAP45 bridges parasite plasma membrane with the outer membrane of IMC and further interacts with GAP40 and GAP50. The topology of interactions between the GAPs as well as of the interaction between MLC1 and GAP45 remains unclear. Presumably, the membrane GAPs further interact with GAPs with multiple membrane spans (GAPMs), that further associate with a network of alveolins. Those bridge the IMC onto subpellicular microtubules. On the other side, short actin filaments use glideosome associated connector (GAC) to bridge onto C-terminal tails of parasite plasma membrane receptors, in this case microneme protein MIC2. The plasma membrane receptors subsequently bind the receptors of the host cell and thus, mediate immobilization of the actin filaments along the parasite surface.

invasion process, where mutant parasites, although able to attach to the cell membrane and undergo apical re-orientation, are not able to penetrate the host cells [108].

The structure of typical MLCs is similar to that of ELCs, with an N-terminal and C-terminal lobe and four calcium-binding EF hands [109]. The function of classical MLCs is also partially overlapping with the function of ELCs: they stiffen the myosin lever arms to increase the motor efficiency and regulate the motor by calcium binding. Additionally, MLCs can be phosphorylated by light chain kinases that provide another means of motor regulation [110–113]. Apicomplexan MLC1s are structurally unusual because, besides the typical N- and C-terminal lobes, they possess a 60-residues-long N-terminal extension that is predicted to be disordered [94,114]. Indeed, although the structures of both *T. gondii* MLC1 as well as *P. falciparum* MTIP have been determined bound to the respective MyoA C-termini,

they lack the 60 N-terminal residues [107,115]. The N-termini of MLC1/MTIP are of a specific interest because they anchor the myosin to the outer IMC membrane through N-terminal myristoylation and, presumably, mediate the interaction with the GAPs [116].

1.3.2.3. Glideosome-associated proteins

Glideosome-associated proteins 40, 45 and 50 (GAP40, GAP45, GAP50) were identified as interacting partners of MyoA and MLC1 and named after the size at which they migrate on SDS-PAGE gels [87,116,117]. These proteins are indispensable for apicomplexan parasites, as besides their role they play in the glideosome, they are also crucial for maintaining the cellular morphology [118].

GAP45, in fact a 27 kDa protein, is myristoylated and palmitoylated at the N-terminus, whereas another lipidation site is predicted near its C-terminus [116,119]. Whereas the N-terminal half of the protein is assumed to form a coiled-coil structure, the C-terminus, assumedly responsible for the interaction with MLC1/MTIP, is probably disordered. The exact localisation of GAP45 in the pellicle remains a subject of discussion because previous evidence was ambiguously suggestive of an association with the plasma membrane or with the IMC. In the current model, GAP45 is anchored in both membranes: in the plasma membrane *via* N-terminal palmitoylation and myristoylation and in the outer membrane of the IMC *via* C-terminal lipidation [116]. However, this model raises some questions because the gap between the plasma membrane and IMC seems to be too large to be bridged by a single small protein [27]. Similarly, the function of GAP45 in the glideosome remains unclear. It is known that, together with MLC1/MTIP, MyoA and presumably also ELC, GAP45 forms a proto-glideosome and associates with GAP40 and GAP50 later during schizogony [117]. The conditional knock-out of GAP45 does not influence the morphology of *P. falciparum* merozoites, but inhibits the gliding, host cell invasion and egress from the host cells [120]. On the other hand, the conditional knock-out of GAP45 in *T. gondii* tachyzoites also slightly alters their shape [116,121].

GAP40 and GAP50 are integral membrane proteins with 9 and 1 predicted membrane helices, respectively, with each having either an N-terminal (GAP50) or C-terminal (GAP40) soluble domain. Weak evidence suggests that the short C-terminal tail of GAP50 is responsible for the interaction with the other glideosome members [117], but the orientation of GAP40 in the IMC membrane and the role of both GAP40 and GAP50 in the glideosome remain elusive. Conditional knock-outs of GAP40 and GAP50, unlike in case of GAP45, additionally cause abnormal parasite cell morphology during schizogony [118]. The daughter cells are larger and round compared to the wild type parasites and sheets of IMC membranes scatter throughout the entire parasite, consequently leading to mis localisation of MyoA and MLC1/MTIP. Therefore, GAP40 and GAP50 not only anchor the other

glideosome members in the IMC membranes, but also are essential for the IMC integrity [118]. Interestingly, GAP40 is not continuously distributed along the IMC, but localises in the distinct regularly spaced foci, similar to those observed under the electron microscope [118]. This suggests that GAP40 interacts with other IMC proteins that in turn bind the glideosome to the microtubules that underlay the IMC. Through pull-down experiments, several candidate proteins were identified that could play this role, such as glideosome-associated proteins with multiple membrane spans (GAPM1-3) or photosensitized INA-labelled protein 1 (Phil-1) [33,34,96,118]. These presumably bind the alveolin network which interacts directly with the subpellicular microtubules, finally ensuring the immobilization of MyoA [32,122].

The structure of the glideosome remains undetermined so far. Structures of several individual components have been published, namely of soluble domain of PfGAP50 [123] and the motor domains of PfMyoA and TgMyoA [98,99]. Additionally, structures of MLC1/MTIP bound to their respective MyoA termini have been solved [107,115]. However, these structures do not reveal the role of the individual proteins in the glideosome, nor do they shed light on the role of the glideosome *per se*. Extensive research is therefore necessary to uncover the molecular details of parasite gliding and the mechanism by which the glideosome functions.

2. Discussion

Apicomplexans are a diverse group of unicellular organisms with a complex life cycle and even more complicated subcellular morphology. The motile stages of apicomplexans have unique subcellular structures and organelles that enable efficient host cell egress and invasion, and mediate the parasite motility. The specific motility, called gliding, does not require changes in the cell shape, as is usual in other unicellular organisms, but rather relies on an actomyosin molecular motor coupled to surface receptors that bind extracellular matrix and the host cell surface. While the understanding of the gliding mechanism has advanced in recent years, the molecular details of the motor empowering gliding have not been elucidated.

2.1. Structure of the glideosome

The glideosome, a complex embedded in the outer membrane of the IMC, is responsible for the anchoring of MyoA within the internal structures of the parasite [19,54]. Besides MyoA, the members of the glideosome are MLC1/MTIP, ELC, GAP40, GAP45 and GAP50 [73,74,87,94–96,114,116,117]. MLC1/MTIP and ELC serve as a myosin light chain, while the N-terminal MLC1/MTIP extension mediates interaction with GAP45. These further interact with GAP40 and GAP50 in an unknown manner. The only determined structures of interacting glideosome members are those of MTIP/MLC1 bound to the C-terminal portions of MyoAs [107,115]. In this study, we extended the knowledge about the structure and interactions within the glideosome and have presented the first structure of glideosome subcomplex consisting of three proteins: MLC1/MTIP, ELC and MyoA.

2.1.1. Structure of MyoA with bound light chains

The structures confirm the previously suggested MyoA binding site in *T. gondii* and clearly show that PfELC binds a homologous site of PfMyoA. The PfELC has previously been incorrectly shown to bind two different sequences of PfMyoA, with one proposed binding site even overlapping with the MTIP binding site [96]. In our hands, mixing peptides containing these binding sites leads to a visible precipitation. The affinities of PfELC to the previously identified binding sites were measured by bilayer interferometry, which directly measures the changes in the thickness of the molecules bound to the surface of a chip. I suppose that the authors of this study actually observed precipitation, which would also display as a change in the thickness on the sensor chip and could be easily misinterpreted as a binding event.

At the same time, I realized that PfELC can only bind to PfMyoA once MTIP is bound beforehand, otherwise the proteins precipitate. In contrast, an opposite effect was observed with *T. gondii* proteins: MLC1 only bound TgMyoA after TgELCs, otherwise the proteins would precipitate. The results with *P. falciparum* light chains are consistent with the previous study, in which the authors could not co-express PfELC in insect cells with PfMyoA in the absence of MTIP [97]. On the other hand, no precipitation of *T. gondii* MLC1 have been previously observed [106]. This suggests that these results cannot be interpreted within the *in vivo* context. It is possible that other factors, such as buffer pH or salt concentration, could have an impact on the precipitation behaviour of the light chains *in vitro*.

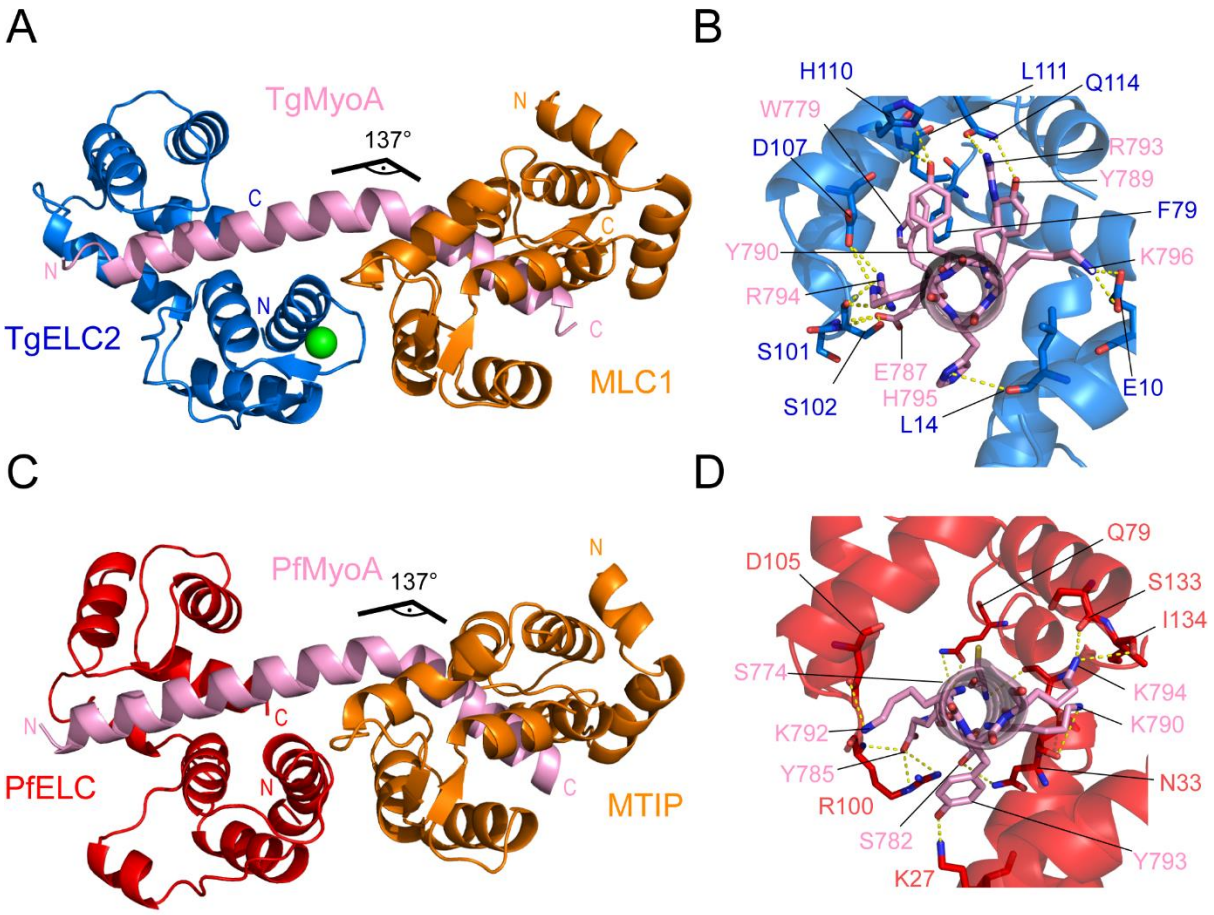


Figure 7. The solved structure of the glideosome subcomplexes. (A) The structure of the trimeric complex consisting of TgELC2, TgMyoA C-terminus and MLC1. (B) The interaction interface between TgELC2 and TgMyoA C-terminus. (C) The structure of the trimeric complex consisting of PfELC, PfMyoA C-terminus and MTIP. (D) The interaction interface between PfELC and PfMyoA C-terminus.

The structures of the trimeric complexes are topologically similar, but also differ between *P. falciparum* and *T. gondii* (Figure 7). In particular, the last five helices of PfELC deviate from those of TgELCs, resulting in a different orientation of the C-terminal lobe of PfELC and thus, a distinct binding mode. Indeed, the C-terminal lobe is much less conserved between PfELC and TgELCs compared to the N-terminal lobe. This further correlates with our measurements that showed that in an unbound state,

the PfELC C-terminus is unfolded. I assume that even upon binding to PfMyoA, the C-terminus of PfELC remains flexible as the electron density for the PfELC C-terminal lobe is less well defined than the density for the rest of the structure. Such less well-defined density could also be explained by the exposure of this part of the protein complex to the solution in the crystal packing, however, the C-terminal lobe is also solvent-exposed in the *T. gondii* structures, where it remains well defined. Our data show that despite the similarities in their binding sites and their sequences, *T. gondii* and *P. falciparum* assume distinct conformation upon binding to MyoA.

2.1.2. Impact of phosphorylation on ELC binding

The differences between the orientation and binding of ELC C-terminal lobes prompted us to investigate if these differences expand to their regulation. The glideosome proteins are known to be heavily phosphorylated, but the exact role of the phosphorylation in the regulation of their function is mostly unknown [124]. We used phosphomimetic mutations based on the available *T. gondii* and *P. falciparum* phosphoproteomes in order to investigate the potential role of phosphorylation on ELC binding. No residue, that would be amenable to modification by phosphorylation, is conserved across apicomplexan ELCs. In TgELC2, the residue S102 has previously been shown to be phosphorylated [125] and at the same time, it is directly involved in the formation of hydrogen bonds with the TgMyoA residue E787, making it an interesting regulation target. On the other hand, the residue S127 of PfELC is located in the intriguingly flexible C-terminal lobe and forms a polar interaction with PfELC N75, helping to maintain the compact conformation of the PfELC C-terminus [125]. Interestingly, phosphomimetic mutations of both aforementioned residues decrease the binding affinity of their respective proteins to MyoA twofold. Thus, the phosphorylation of ELCs might negatively regulate the glideosome assembly. However, even after the twofold decrease, the affinity of ELCs is in the nanomolar range, which makes it unlikely that a single phosphorylation even would be sufficient for the regulation of glideosome assembly. However, the published proteomics data show that these residues are phosphorylated which alone suggests that they play a regulatory role. It might be that only the sum of the individual phosphorylation events contributes to the glideosome regulation and it cannot be excluded that the phosphorylation of ELCs may be one of them. Finally, the examination of the phenotypes of these mutants in parasite cultures would be required to evaluate their role *in vivo*.

2.1.3. Role of ELCs in the regulation of the glideosome

After identification of the sequence of MyoAs that binds ELCs, and further elucidation of their interaction interface and binding regulation, we investigated how their binding improves the efficiency of the MyoA motor. The data measured on other essential light chains suggest that they stabilize the

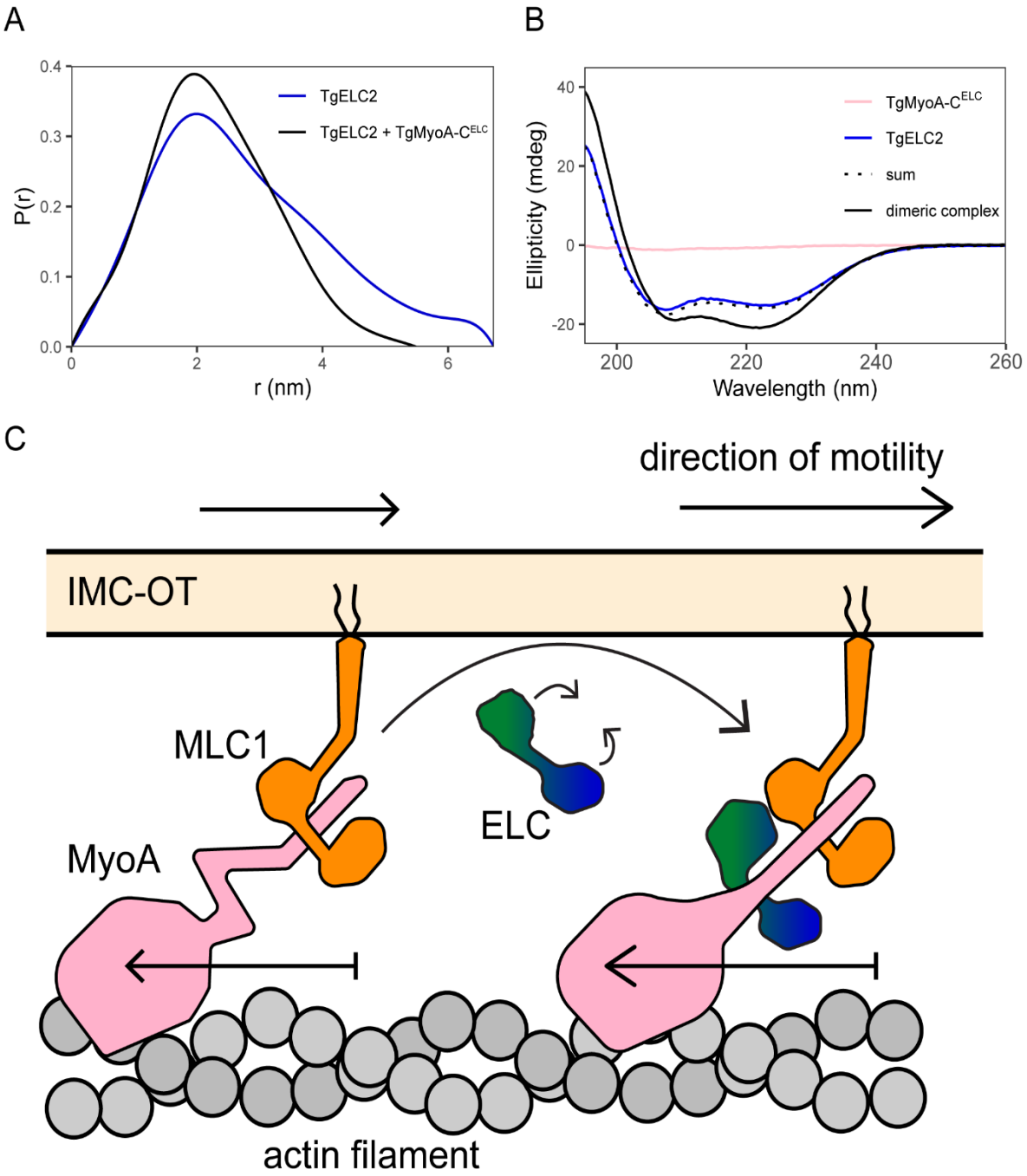


Figure 8. Role of ELCs in the glideosome. (A) The distance distributions derived from SAXS measurements show that upon binding to TgMyoA C-terminus, TgELC2 assumes a more compact conformation. (B) Circular dichroism data reveal that upon binding of TgELC2 to TgMyoA C-terminus, the dimeric complex has a higher amount MyoA (pink) lever arm is bound to MLC1 (orange), but the absence of ELC (green/blue) allows certain level of disorder in the lever arm (left). Upon binding, ELC compresses (centre) to induce the secondary structure in the MyoA lever arm (right). As a result, the MyoA undergoes a larger step, increasing the efficiency of gliding.

myosin neck regions, thereby increasing the neck efficiency [103,126]. However, the molecular mechanism behind this stabilization were unknown. We have shown that the ELCs tightly wrap around the disordered MyoA neck region, which in turn induces secondary structure in MyoA and stiffens the entire myosin lever arm (Figure 8). This is in agreement with MyoA functional assays that showed that the presence of ELCs results in an almost twofold increase in the motor speed by increasing the size of the myosin step.

2.1.4. Role of calcium in the glideosome regulation

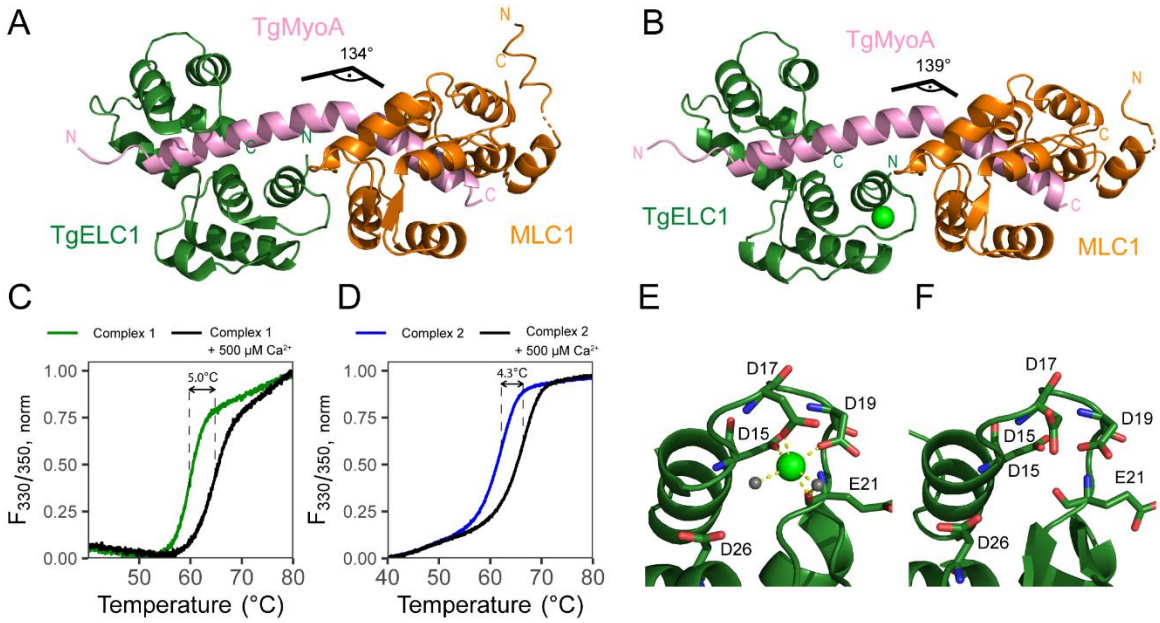


Figure 9. Role of calcium in the glideosome. (A) Structure of the calcium-free trimeric complex of TgELC1, TgMyoA C-terminus and MLC1. (B) Structure of the calcium-bound trimeric complex of TgELC1, TgMyoA C-terminus and MLC1. The structure display a conserved topology and TgELC1-TgMyoA interaction interface, showing that the calcium does not change the properties of ELC binding. (C,D) Thermal unfolding curves point out that the trimeric complexes of MLC1, TgMyoA C-termini and TgELC1 (complex 1) or TgELC2 (complex 2) are dramatically stabilized by the addition of calcium. (E,F) The comparison between the EF hand loops of TgELC1 with and without a bound calcium stress the importance of the residue D17 in calcium binding but also show that their loop structure is independent of the bound calcium.

As *T. gondii* ELCs bind calcium ions in their first EF hand, it has been suggested that calcium could regulate binding of ELCs to MyoA [106]. Indeed, the light chains of myosin generally do use calcium to positively regulate myosin binding [104]. Moreover, the regulation of gliding and host cell invasion is initiated by an increase in the parasite intracellular calcium concentration [127]. However, conflicting data were published about the role of calcium in ELCs. Several studies showed that calcium increases the binding affinity of TgELCs to TgMyoA, presumably by mediating tighter interactions between MLC1 and TgELCs [106,107]. On the other hand, functional studies of TgMyoA *in vitro* showed that calcium does not impact the MyoA motor speed [97]. We could demonstrate that calcium strengthens

the stability of the trimeric complexes (Figure 9). However, when we compared the binding affinity of TgELC1 and TgELC2 to TgMyoA in the absence and in the presence of calcium, we showed that the role of calcium is limited, increasing the affinity only twofold. Further structural insights into the calcium binding were revealed with the calcium-free structure of trimeric complex of TgELC2, MLC1 and the MyoA C-terminus. Comparing the calcium-bound and calcium-free structures, it became clear that the calcium ion does not impact the overall structure of the complex and does not alter the interactions between the two light chains. Two additional facts also suggest a rather limited role of calcium in ELCs: (a) PfELC does not possess the ability to bind calcium, rendering such regulation impossible in *P. falciparum*; and (b) recent studies have shown that ELCs are crucial in parasite egress, but the increase in intracellular calcium concentration only takes place after the egress [128]. Thus, our biophysical and structural data agree with previous functional evidence, suggesting that calcium does not regulate the myosin motor *via* interaction with ELCs.

2.1.5. Glideosome associated proteins

Our study has revealed the structures of the MyoA neck domains with both light chains bound. These structures, together with the structures of MyoA motor domains, provide a thorough understanding of MyoA motor and its regulation. However, the process by which MyoA and the light chains interact with glideosome associated proteins, remains unclear. It has been shown that MyoA and the light chains first form a “proto-glideosome” that only later binds to GAP40 and GAP50 [117]. Elegant *in vivo* experiments have suggested that the unusual MTIP/MLC1 N-terminal extension interacts with the C-terminus of GAP45 [116]. In the current model, GAP45 uses N-terminal and C-terminal lipidation to stretch between the plasma membrane and the outer membrane of IMC, respectively. However, the small size of GAP45 could only hardly bridge a 20 nm wide gap between these two membranes [27], calling the current glideosome model into question. More questions arise about how the “proto-glideosome” binds GAP40 and GAP50. GAP40 is distantly related to the major facilitator superfamily of membrane proteins that usually transport small molecules across membrane [116]. It is therefore imaginable that GAP40 serves as a specific anchor point for the MLC1/MTIP or GAP45 lipidation moieties. On the other hand, experiments using fluorescence recovery after photobleaching (FRAP) and super-resolution microscopy data show that GAP40 distributes along the IMC in distinct foci, whereas MLC1/MTIP and GAP45 display a continuous peripheral distribution [118,129]. These data rather point towards GAP50 as the anchoring point of the proto-glideosome, but the distinct distribution of GAP40 calls its role in the glideosome into question. The structure of the large GAP50 soluble domain has previously been determined [123], however, its glycosylation in *T. gondii* suggests that it is hidden inside the alveoli [130] and only the short GAP50 C-terminus is exposed to the rest of

the glideosome. Indeed, the conserved C-terminus of GAP50 was proposed as the interaction partner of GAP45, however, any direct evidence for that is lacking thus far [117]. Therefore, the structure of the glideosome and the mechanism of gliding remains a mystery to this date.

2.1.6. Unidentified glideosome members

It cannot be excluded that additional proteins are also members of the glideosome. Even the current glideosome members were discovered step by step, with the last identified member being PfELC of *P. falciparum* in 2017 [96,97]. Indeed, all glideosome members were discovered by pull-down assays followed by mass spectrometry. If a protein only interacts transiently or weakly, or does not interact with the glideosome at the parasite stage under investigation, it is not pulled down. Moreover, the mass spectrometry techniques are limited by the availability of the proteins upon extraction. For example, the intramembrane parts of membrane proteins are hydrophobic and therefore precipitate easily or unfold if not handled with mild detergents. Indeed, the IMC is packed with a cassette of membrane and membrane-associated proteins, the functions of which is yet to be deciphered [31,32,37]. For example, the involvement of PfIMCl in *P. falciparum* glideosome has been suggested, but this still remains to be verified [131].

2.1.7. Towards the full glideosome structure

Even if all the glideosome members are known, it would still require a large effort to solve the structure of the entire complex. First, the recombinant expression of apicomplexan proteins in traditional expression hosts can be inefficient or impossible, or they can only be expressed as unfolded proteins due to the lack of their interacting partners or of the apicomplexan chaperons. The expression and purification of the IMC proteins might be specifically challenging because of the cholesterol-rich composition of the IMC membranes [31]. I assume that for the *in vitro* assembly of the glideosome, the proteins must be co-expressed in order to form a complex directly in a eukaryotic host, while the co-expression of apicomplexan chaperons might be necessary, as it is in the case of MyoA expression [97,102].

As an alternative strategy, the glideosome complex could be pulled down from apicomplexan cultures. This strategy would ensure that the proteins are correctly folded and could directly provide interesting information about the phosphorylation state of the individual glideosome members. On the other hand, it might be even more difficult to solubilize the membrane protein residing in the IMC. In both cases, the purified complexes could be subjected to structural studies by X-ray crystallography or single-particle electron microscopy to obtain a high-resolution structure.

Nevertheless, in case of the glideosome, even obtaining a low-resolution structure would be a success because it would uncover the overall topology of the complex and the existing crystal structures could be docked into the low-resolution glideosome envelope, which would enable localisation of the remaining glideosome members. To obtain a low-resolution structure, *in situ* cryo-electron tomography could be used [132]. In this approach, the protein complex does not need to be purified and the low-resolution structure of the complex would be reconstructed from a tilt series of images of frozen parasite.

In any case, the determination of the glideosome structure will require a huge effort and will represent a large step towards understanding of apicomplexan motility.

2.2. Drug against malaria and the glideosome as a drug target

Apicomplexans remain the deadliest pathogens on the Earth, with the *Plasmodium* species being responsible for approximately 400 thousands deaths with over 200 million cases in 2018 [133]. At least five *Plasmodium* species have been shown to cause malaria in humans, broadening the complexity of the efforts to control the disease [134]. *Plasmodium falciparum* is the dominant causative malaria agent in Africa and can cause the most severe complications, such as cerebral malaria. Patients with cerebral malaria present with impaired consciousness that is caused by infected erythrocytes that bind the endothelium in the brain and induce immunological and systemic inflammatory responses as well as dysfunction of coagulation in the brain [135]. *Plasmodium falciparum* accounts for over 99% malaria deaths [133]. *Plasmodium vivax*, a less deadly malaria agent, is most prevalent in Southeast Asia and in the Americas [133]. In a manner similar to *Plasmodium ovale*, they possess the ability to enter a dormant stage after the invasion of hepatocyte, called hypnozoites, that can relapse as late as several years after the initial infection [18,134]. *Plasmodium ovale* and another species, *Plasmodium malariae*, are rare compared to *P. falciparum* and *P. vivax* and because they only cause relatively mild influenza-like symptoms, they are said to cause „benign“ malaria. *Plasmodium knowlesi* infect humans as a zoonotic species, meaning that the humans are only their secondary host [134]. The primary hosts of *P. knowlesi* are macaques and it was thought to only cause malaria in humans very rarely, but in the last decade, the surveys confirmed that in some places in Southeast Asia, most malaria infections were caused by *P. knowlesi* [136].

2.2.1. Antimalarial compounds

Despite the complexity of malaria, the global fight against the disease in the last decades has mostly been a success. The number of malaria cases dramatically dropped after WWII due to increased

mosquito control by insecticides and increased usage of insecticide-treated mosquito nets in Africa [133]. Moreover, the death rate decreased due to the introduction of novel treatment options against malaria [133]. Historically, the first discovered active antimalarial compound was quinine, which was isolated from the bark of South American of cinchona tree [137,138]. Quinine is toxic to *Plasmodium* parasites because it interferes with the parasite's haem detoxification mechanism. Although it is predicted that quinine simultaneously targets multiple parasite proteins, only *P. falciparum* purine nucleoside phosphorylase (PfPNP) has been recently shown to bind and be inhibited by quinine [139]. Although no large-scale resistance to quinine has been observed so far, WHO does not recommend its usage unless the artemisinin resistance is indicated or in some cases of cerebral malaria [140]. Chloroquine, developed by the chemical modification of quinine before the WWII, has only been used in large scale afterwards [137]. Many structurally similar drugs were developed since then, that are usually used in a combination with artemisinin-derived drugs, such as piperaquine or meflaquine. Another class of compounds developed during WWII are derivatives of pyrimidine, such as proguanil and pyrimethamine. These inhibit folate dehydrogenase, an enzyme that is essential in DNA synthesis of bacteria and proteazon organisms [141]. Another class of antimalarial compounds, sulfones and sulfamides, inhibit an enzyme called dihydropteroate synthetase, that is also essential in DNA synthesis and folate metabolism [142]. The most important antimalarial drug of the last decades is artemisinin, originally isolated from sweet wormwood (*Artemisia annua*) that has been used for a long time by Chinese herbalists [137]. Artemisinin, in a yet not fully understood process, exposes the parasite cells to an oxidative stress by inducing the generation of free radicals. Due to the emergence of artemisinin resistance in Southeast Asia, artemisinin and its derivates should only be used in a combination with another antimalarial compound, together constituting artemisinin-based combination therapies (ACTs) [143].

In contrast to drug development, the development of vaccines against *Plasmodium* species has been unsuccessful. In particular, the sporozoites have been an attractive target for the vaccine development because only about a hundred individual parasites enter the human body. The most successful vaccine candidate, called "RTS,S", is directed against the abundant PfCSP present on the surface of *P. falciparum* sporozoites. However, it only confers immunity in 51% of the patients one year after the last dose and 26% of the patients two years after the last dose [144]. Thus, the current antimalarial drugs remain the main means of protection against fatal consequences of malaria.

Nevertheless, a resistance to all known antimalarial compounds has emerged and therefore, there is a need for the identification of new drugs and drug targets. The identification of the drug targets cannot be achieved without an improved understanding of the essential processes that the parasites undergo during their life cycle. Cell egress, motility and host cell invasion represent processes that are essential

for completion of the parasite`s life cycle and for their reproduction in the host organisms. The molecular mechanism that underlies gliding is therefore an important goal in apicomplexan research.

2.2.2. Drugs directed against the glideosome

Currently, the design of compounds inhibiting the interactions within the glideosome complex is only hardly possible due to the lack of functional and structural data. MTIP is an obvious drug target because MTIP anchors MyoA in the IMC membrane and it has been shown that both proteins are crucial for the parasite`s motility and the host cell invasion. A small compound that inhibits MTIP lipidation was first shown to impair the *P. falciparum* invasion process. However, a recent study has shown that palmitoylation of MLC1 in *T. gondii* is not required for the parasite motility [145]. Blocking of MTIP-MyoA interaction therefore seems as a more plausible option in MTIP targeting. The structure of MTIP-MyoA complex has been previously determined and their interaction is thus well defined. Indeed, *in vitro* studies showed that peptides and compounds mimicking the very C-terminus of MyoA are able to reduce the parasite motility and the host cell invasion [146–150]. However, these experiment were only performed *in vitro* and the effect of these compounds *in vivo* might not be sufficient or the compounds might be toxic.

2.2.2.1. PfELC as a potential drug target

Due to the lack of functional and structural data, it has not been possible to design inhibitory peptides directed against ELC. Intriguingly, different ELC binding sites have originally been identified on PfMyoA compared to TgMyoA [96]. With our binding data and subcomplex structures, we could finally resolve this discrepancy and show the interaction mode between PfELC and PfMyoA C-terminus. At the same time, as the recently published data show that the merozoites with knocked-out PfELC are not able to invade the red blood cells, PfELC represents a new potential drug target [108]. Thanks to our study, the protein is now well characterised and its structure and binding mode are described.

Additionally, PfELC might represent a more interesting drug target than MTIP, because its interaction with PfMyoA requires half the number of salt bridges (8 vs 4) and hydrogen bonds (16 vs 7). On the other hand, the interaction interface between PfELC and PfMyoA is considerably large (over 1400 $Å^2$) and a complete abolition of their interaction would require a rather large molecule, similar to the modified peptides used in the case of the MTIP-MyoA interaction. According to Lipinski`s rule of five, the drug-like molecules usually contains a limited number of hydrogen bond donors and acceptors (5 and 10, respectively), have a molecular mass less than 500 Da and are rather hydrophobic (the octanol-water partition coefficient typically does not exceed 5) [151]. Taking into account the fact that the

smallest PfMyoA peptide that would still maintain all interacting residues with PfELC would have a molecular mass of about 2000 Da, one of Lipinski's rules would have to be broken. Drugs usually cannot be too large because their bioavailability, which defines how accessible the drug is to its target, decreases with the size [152]. This is especially important for drugs targeted against apicomplexans, as the parasite drug targets reside under the RBC plasma membrane, membrane of parasitophorous vacuole and their own plasma membrane. It might be still possible that a compound could be designed such that it would only interfere with a part of the MyoA-ELC interface and still block the interaction. For example, our mutational analysis of TgELC2-TgMyoA binding showed that the mutation of two residues (E10A, H110A) led to an almost 30-fold decrease in the binding affinity. As these and other residues of TgELC2, which mediate polar interactions with TgMyoA, are located only within a limited section of TgMyoA sequence, it might be possible to design smaller compounds inhibiting TgELC2-TgMyoA interaction. Similarly, PfELC residues that mediate polar interactions with PfMyoA are only localised to a confined site that could be regulated by the binding of small molecules. Further research will be required to validate PfELC as a drug target and to identify compounds that could hinder its function.

2.2.2.2. TgELCs as a potential drug target

PfELC is a good drug target because the knockout of PfELC in the parasite leads to a fatal phenotype (a lack of host cell invasion) [108]. This is, however, not the case in *T. gondii*, which possesses two ELCs – TgELC1 and TgELC2 [106]. Both of these proteins are expressed in the parasite blood stages and, consequently, a knockout of only one of them does not impede the processes of egress or invasion. Both egress and invasion are, however, completely inhibited in the double knockout of both TgELCs [106]. TgELCs are, therefore, essential in *T. gondii* similarly to *P. falciparum*, but they can compensate each other's role in the glideosome. Such compensation makes it difficult to use them as a drug target such as in the case of *P. falciparum*.

Interestingly, other pairs or groups of proteins that can compensate each other's function have been described in *T. gondii* glideosome, such as MyoA and MyoC [153,154], or TgGAP45 with TgGAP70 and TgGAP80 [154,155]. However, in these cases, the proteins have distinct roles and localisation in the wild-type context and only compensate for the loss of the other in the laboratory-produced knockouts. In contrast, TgELCs seem to have an overlapping function in the wild-type parasites and therefore, it remains unclear why two ELC genes have evolved in *T. gondii*. The transcriptomic data shows that *tgelc1* is expressed 1.5-2x more than *tgelc2* in sporozoites and bradyzoites, but *tgelc2* exceeds *tgelc1* expression in oocysts by a factor of more than 60 [156]. Because the oocysts are non-motile parasite forms, the gene expression data suggest that the primary role of TgELC2 lies in a

different process, whereas the involvement in the glideosome is its secondary role. Indeed, some apicomplexan light chains have been found to bind more than one myosin, thus, it is plausible that TgELC2 binds a different myosin than MyoA in another process that requires myosin motors, e. g. cell division.

2.2.2.3. Apical and basal glideosomes as drug targets

At least 6 different myosins exist in *P. falciparum*, whereas 12 myosins were described in *T. gondii*, raising the question of whether some of them are also involved in gliding, host cell invasion and egress [93,157]. Indeed, in the TgMyoA knockout strain, the wild-type phenotype can be partially rescued by TgMyoC that was observed to be redistributed along the periphery of the parasite. The same has been also observed *vice versa*: missing TgMyoC can be substituted by TgMyoA [153,158].

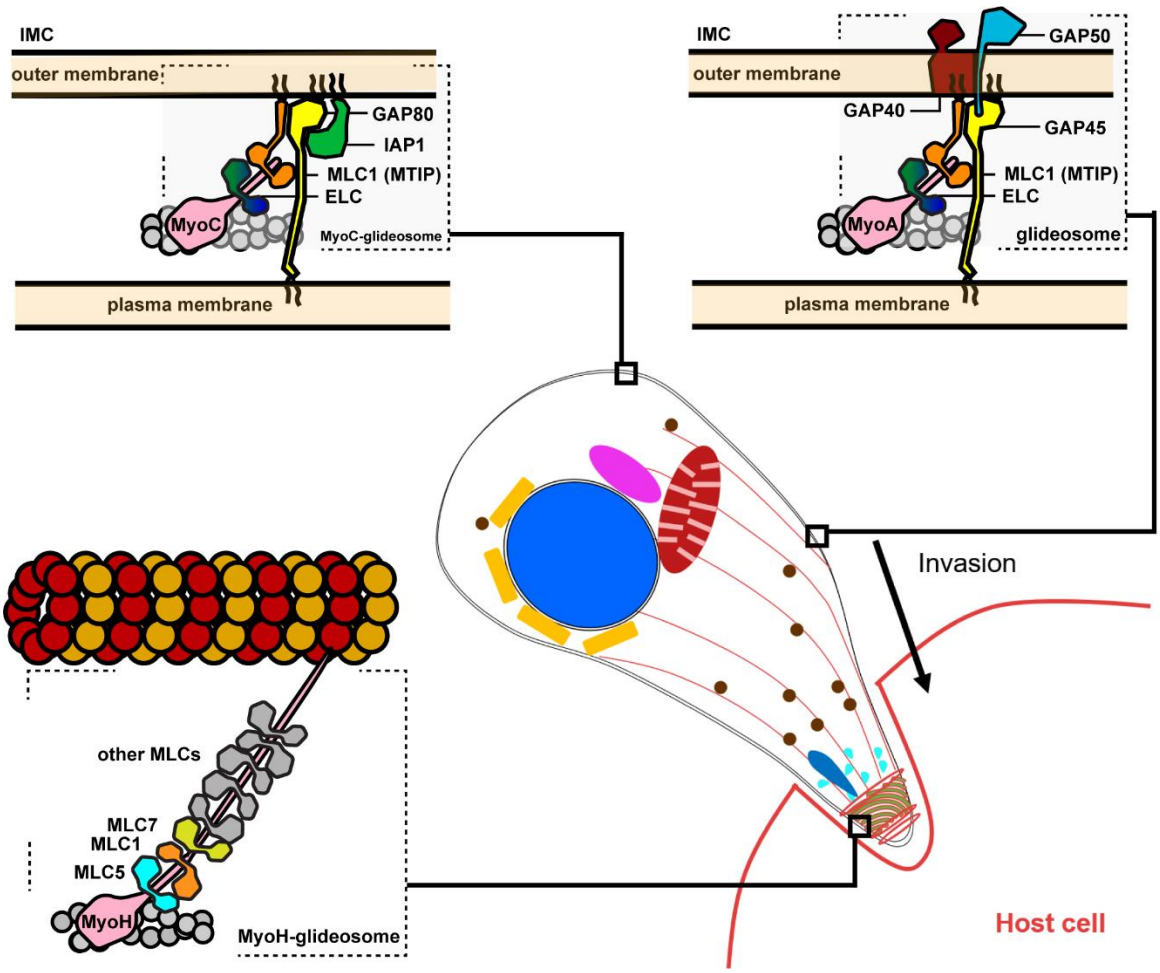


Figure 10. Alternative glideosomes in *T. gondii*. The glideosome was first identified as a complex of MyoA with ELC, MLC1/MTIP, GAP40, GAP45 and GAP50. In *T. gondii*, another complex located in the posterior of the parasite was identified, composed of MyoC, TgELC1, MLC1, GAP80 and IAP1. Another MyoH motor is essential for the invasion, binding eight light chains (among them MLC1, MLC5 and MLC7), and anchors in the microtubules in the apical pole.

Indeed, TgMyoC localises to the basal end of *T. gondii* and has been found to be associated with the MyoA light chains MLC1 and TgELC1. These further interact with a GAP45 homologue, GAP80 and anchor to the basal plate of the IMC presumably *via* the IMC-interacting protein (IAP1, Figure 10) [153]. In a manner similar to MyoA and MyoC, the loss of GAP80 could be partially compensated by the re- localisation of GAP45 and *vice versa*. These data thus show that an alternative glideosome (MyoC-glideosome) with a different localisation (posterior pole) forms that is important for gliding (Figure 10). The fact that most proteins of the MyoA-glideosome are identical or exchangeable with the proteins of MyoC-glideosome also suggests functional and structural conservation [73,153,158,159].

Another *T. gondii* myosin that has been shown to play a crucial role in gliding and host cell invasion is TgMyoH (Figure 10) [160]. TgMyoH localises to the apical tip where it binds to microtubules of apical polar ring *via* the C-terminal tail domain and eight myosin light chains, among them MLC1, MLC5 and MLC7 [160]. As TgMyoH directly attaches to apical polar ring, presumably no additional proteins, that would resemble glideosome complex, are necessary. TgMyoH is crucial for the initiation of the parasite motility and host cell invasion because it is responsible for retrograde translocation of the apical polar ring components, including actin and formin 2, towards the IMC, where the retrograde translocation is taken over by the MyoA-glideosome.

Myosins with the same localisation patterns, and thus, identical function, were also identified in *P. falciparum*. PfMyoE is present in the basal end of the *P. falciparum* motile stages and its absence results in reduced parasite motility. PfMyoB, on the other hand, localises to the parasite`s anterior, where it presumably plays the same role as TgMyoH. Unlike TgMyoH, however, PfMyoB resembles MyoA because it requires a light chain with an unusual N-terminus, in this case MLC-B, to maintain the apical localisation, suggesting that a glideosome-like complex could be associated with both PfMyoB and PfMyoE [73,157,159,161].

The presence of additional glideosomes only increases the complexity of the gliding mechanism, but, on the other hand, offers more targets for antimalarial drug development.

2.2.3. Drug targets upstream of the glideosome and regulation of the gliding motility

The proteins of the glideosome are good drug targets because many of them are unique to *Apicomplexa* or *Alveolata*, therefore such drugs would less likely interfere with human metabolism. At the same time, targeting a single protein or protein complex might result in fast emergence of resistant parasite species because they would need to circumvent a loss of only one phenotype. Therefore, the processes upstream of the glideosome, such as gliding regulating pathways, offer themselves as attractive drug targets.

Indeed, parasite egress, gliding and the host cell invasion are initiated and regulated by a complex cascade of phosphorylation events (Figure 11) [127,162]. The egress is first initiated by an unknown stimulus, upon which diacylglycerol kinase 2 (DGK2), residing in the parasitophorous vacuole, converts diacylglycerol (DAG) into phosphatidic acid (PA) in the outer leaflet of the plasma membrane [163]. PA then activates plasma membrane associated guanylate cyclase [163,164]. Guanylate cyclase then converts GTP to cGMP that further activates protein kinase G (PKG) [165]. PKG, through the activation of phosphatidylinositol (PI) kinase and phosphatidylinositol 4-phosphate kinase, catalyses the conversion of PI into PI-4-phosphate and in the end, into PI-4,5-bisphospate (PIP_2) [127]. PIP_2 is further cleaved by phospholipase C (PLC) into soluble inositol triphosphate (IP_3) and membrane-located DAG that resides in the inner leaflet of the plasma membrane. IP_3 stimulates the release of calcium from yet unidentified internal sources (probably from endoplasmic reticulum) through unknown calcium transporters [166]. The released intracellular calcium regulates the activity of calcium dependent protein kinases (CDPKs) which supports the exocytosis of micronemes [25]. On the other hand, membrane soluble DAG is converted by DAG kinase to phosphatidic acid (PA), which is further sensed by apical plekstrin homology domain proteins (APH) that are embedded in the microneme membrane by acylation and which regulates microneme exocytosis [167].

This cascade of processes leads to the discharge of micronemes that deposit perforin like protein 1 (PLP1) in the parasitophorous vacuole [168]. This leads to perforation of the PV, granting access to the host cell calcium that is transported into the parasite, causing a second spike of calcium in the parasite and finally initiating the egress [128].

The kinase cascade downstream of PKG and calcium release also regulate the proteins of the glideosome and the plasma membrane receptors of the parasite [55,68,95,127,166,169]. PKG decreases the activity of CDPK1, whereas CDPK1 directly phosphorylates MLC1/MTIP and GAP45 [170–174]. At the same time, CDPK1 mediates the dissociation of the protein kinase A regulatory domain (PKA-R) from its catalytic domain (PKA-C) [174] and PKA further phosphorylates MyoA and GAP45 [175]. Moreover, activated PKA can downregulate CDPK1 in a negative feedback loop [174]. CDPK3 has also been suggested to downregulate PKA, whilst it also phosphorylates MyoA and the cyclase associated protein (CAP) that stabilizes actin filaments [176]. Another glideosome protein, GAP40, is also phosphorylated by an unknown kinase in a PKG dependent manner [177]. Besides motor proteins, these kinases also phosphorylate membrane receptors. For example, phosphorylation of AMA1 by PKA [178], which is supported by glycogen kinase 3 (GSK3) [72], is crucial for the formation of the moving junction. A PKG-independent cascade is regulated by an increase in calcium concentration that mediates binding of calmodulin (CaM) to protein kinase B (PKB), promoting its autophosphorylation and consequent activation [179,180]. PKB in turn phosphorylates GAP45 [173].

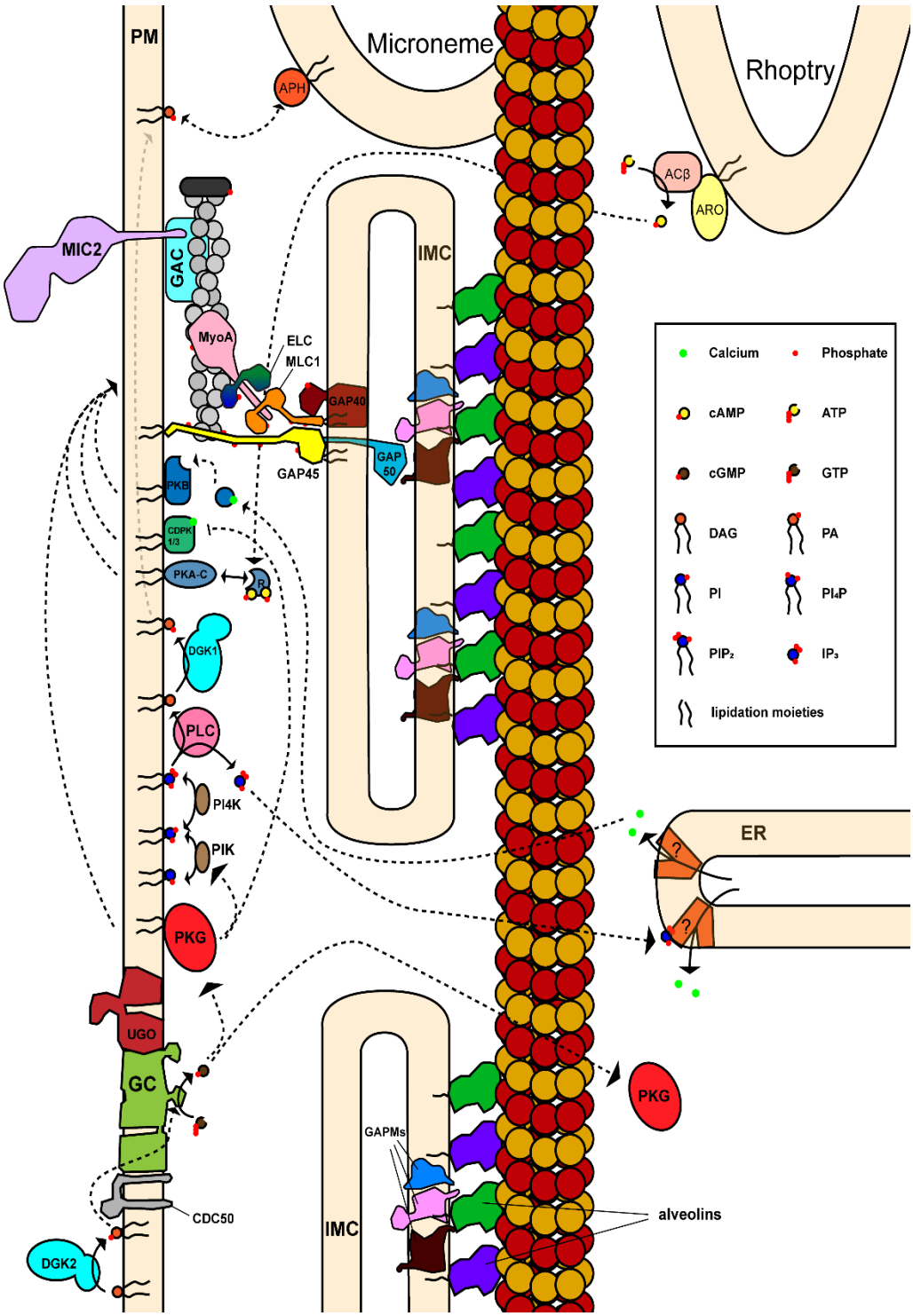


Figure 11 Simplified scheme of glideosome regulation. Diacylglycerol (DAG) is converted to phosphatidic acid (PA) upon stimulation of DAG kinase 2 (DGK2). PA activates guanylate cyclase (GC) that resides in the plasma membrane (PM) in a complex with unique GC organiser (UGO) and CDC50. GC converts guanosine triphosphate (GTP) to cyclic guanosine monophosphate (cGMP) that activates protein kinase G (PKG). PKG directly phosphorylates some glideosome members, regulates some glideosome-phosphorylating kinases (such as calcium dependent protein kinases CDPK1 and CDPK3) and regulates the activity of phosphatidylinositol (PI) kinase (PIK). PIK with phosphatidylinositol-4-phosphate (PI_4P) kinase (PI4K) catalyse conversion of PI to PI_4 and finally to phosphatidylinositol-4,5-bisphosphate (PIP_2). PIP_2 is then cleaved by phospholipase C (PLC) to DAG and inositol triphosphate (IP_3). DAG is converted to PA in the inner leaflet of the plasma membrane by DGK1 and PA is sensed by apical plekstrin homology protein (APH) that regulates microneme discharge. IP_3 regulates gating of calcium channels, probably in ER and calcium subsequently regulates protein kinases either through direct binding (CDPK1, CDPK3) or through interacting partners (calmodulin, C, binding to PKB). The regulatory domain (R) of another kinase – PKA, is regulated by cyclic adenosine monophosphate (cAMP) that is produced from adenosine triphosphate (AT) by adenylate cyclase beta (ACβ) that is tethered to the rhoptry membrane by Armadillo Repeat-Only (ARO). PKA, PKB, CDPK1, CDPK3 and PKG all phosphorylate glideosome members.

An active involvement of the above-mentioned kinases in the regulation of glideosome is also evident from their localisation. CDPK1, CDPK3 as well as PKA, PKB and PKG have all been observed to localise to the periphery of the parasites and to co-localise with some of the glideosome proteins [170,180–182]. PKB, for example, can be pulled down by GAP45 and MTIP antibodies. CDPK1 and CDPK3 possess a myristoylation and palmitoylation pattern similar to that of GAP45. Such close association of these kinases with the myosin motor enables temporarily tight phosphorylation that would even further increase the complexity of glideosome regulation.

As the kinases involved in glideosome regulation also regulate secretory pathways and activation of plasma membrane receptors, they are suitable drug target candidates [183]. For example, CDPKs are conserved across plants and protists but their genes are not present in animals, making the unintended drug interaction with human proteins less of a concern [184].

Multiple inhibitors with sub-micromolar IC_{50} values have been designed based on the TgCDPK1 structure [185]. Some derivatives of these compounds were also found to inhibit *P. falciparum* PfCDPK1, however, they were also found to target cGMP-dependent kinases and heat shock protein 90 (HSP90), questioning PfCDPK1 as a drug target [186]. On the other hand, potent and selective inhibitors have been identified that target PKG with picomolar IC_{50} values [187,188]. Selective inhibitors of GSK3, a regulator of AMA1 function, have also been found to have antimalarial properties [189,190].

This exhaustive but incomplete list of the protein kinases and their functions in glideosome regulation testifies to the sheer amount of potential drug targets involved in gliding and it will be an enormous effort to experimentally validate them. Further basic research into the function and structure of these proteins is therefore necessary to narrow down the list of potential drug targets.

At the same time, the number of proteins involved in glideosome regulation raises various questions, such as (1) which phosphorylation sites are important for proper function of the glideosome? and (2) how does the phosphorylation impact the glideosome structure and function? A majority of the phosphorylation sites are located at the N-terminus of MLC1/MTIP and on GAP45, two proteins that have been hypothesised to interact but this interaction has never been shown *in vitro*. As both the N-termini of MLC1/MTIP and GAP45 are predicted to form coiled coils, I hypothesise that their phosphorylation could lead to the induction of secondary structure and either direct interaction or, as would be typical for coiled-coil domains, dimerization, which could further promote interaction. Indeed, both the N-termini of MTIP/MLC1 and GAP45 are partially or completely unfolded when expressed in *E. coli*, supporting the hypothesis that the phosphorylation could induce their secondary structure. Although an extensive study of GAP45 showed that phosphomimetic mutations do not change its

structure, it is impossible to test all possible permutations of possible phosphorylation sites. Just 16 serine residues of GAP45, potentially serving as phosphorylation sites, give 65536 possible permutations. Moreover, the phosphomimetic mutations might not sufficiently mimic an actual phosphate group, resulting in a different phenotype or structure.

In summary, not only glideosome proteins, but also the upstream processes in glideosome regulation, such as protein kinases, can be suitable protein targets and their future investigation will be important in anti-apicomplexan drug development.

2.3. Regulation of PfGSK3

PfGSK3 is a kinase that is well conserved across taxa. Although human GSK3 plays a role in multiple cellular processes, PfGSK3 has proven to be an important drug target: a class of organic compounds built on a thieno[2,3-b]pyridine scaffold has been found to specifically inhibit PfGSK3 and decrease the invasion efficiency [72,190]. The protein is thus of special interest, although its structure and means of its regulation are not known.

In our study, we provided a thorough purification protocol with yields of up to 1.5 mg of pure protein per 1 L of *E. coli* culture, representing a solid starting point for further PfGSK3 research. During the optimization of the purification protocol, I noticed that the protein precipitates upon contact with nickel beads, but only in the absence of the N-terminal purification tag. The investigation of this behaviour led us to discover that other bivalent heavy atom metals besides Ni^{2+}, such as Co^{2+}, Cu^{2+} and Zn^{2+} modify the thermal unfolding of the protein, suggesting that these atoms trigger changes in its structure. However, data from circular dichroism have shown that PfGSK3 remains folded independent of the presence of these heavy metals. Finally, we found that the heavy metals induce the formation of high-MW species of over 1 MDa that are heterogeneous in size and shape. The fact that this behaviour is only observed after the cleavage of the N-terminal tag and only with metal ions of similar size (1.09-1.21 Å) lets us hypothesise that the effect is mediated by a specific binding site at the N-terminus of PfGSK3. Interestingly, the metal-induced high-MW species do not possess enzymatic activity, indicating that PfGSK3 could be negatively regulated by the changes in the concentration of the metal ions in the parasite. Importantly, we could show that the high-MW species can be reverted back into the enzymatically active monomeric species, further supporting the hypothesis that such regulation could take place *in vivo*. Indeed, high-MW species of another serine/threonine kinase, CK2, have been shown to appear *in vivo* in a process termed autoinhibitory polymerisation [191]. On the other hand, in *P. falciparum*, the N-terminus of PfGSK3 could be bound by another protein, preventing the induction of high-MW species, or the induction of high-MW species could be regulated by the phosphorylation

of the N-terminus. Thus, the presence and the role of PfGSK3 high-MW species remains to be validated *in vivo*.

The restricted conservation of the PfGSK3 N-terminus in *Plasmodium* species led us to investigate it in more detail. First, we have confirmed that, similarly to human GSK3, PfGSK3 exhibits autophosphorylation, phosphorylating important residues in its activation loop but also N-terminal residues that are unique for *Plasmodium* species. To investigate the function of the N-terminus, otherwise predicted to be disordered, we cloned several constructs missing the first 23, 46 and 64 N-terminal residues. However, all these constructs were expressed as insoluble proteins. The analysis of the Δ64-PfGSK3 revealed that it is devoid of any phosphorylation, linking the function of the N-terminus with the enzymatic activity of the protein: PfGSK3 that lacks the N-terminus does not maintain the enzymatic activity and because of the lack of autophosphorylation, is also not soluble.

To shed more light on the role of the N-terminal phosphorylation of PfGSK3, we used the differently phosphorylated protein fractions, which were separated by ion exchange chromatography, to assess the relationship between the phosphorylation state and the enzymatic activity of PfGSK3 (Figure 12). The results show that the activity of PfGSK3 increases with the amount of phosphorylation at the N-terminus. Interestingly, this suggests that the N-terminus in PfGSK3 fulfils an opposite role in comparison with human GSK3, which is inhibited by the phosphorylation of the S9 residue [192]. Although the positions of the serine residues in the respective GKS3 termini of different *Plasmodium* species are not conserved, all of them contain many serine residues that could be amenable to phosphorylation. Thus, I assume that the means of GSK3 regulation by N-terminal phosphorylation are conserved across the species.

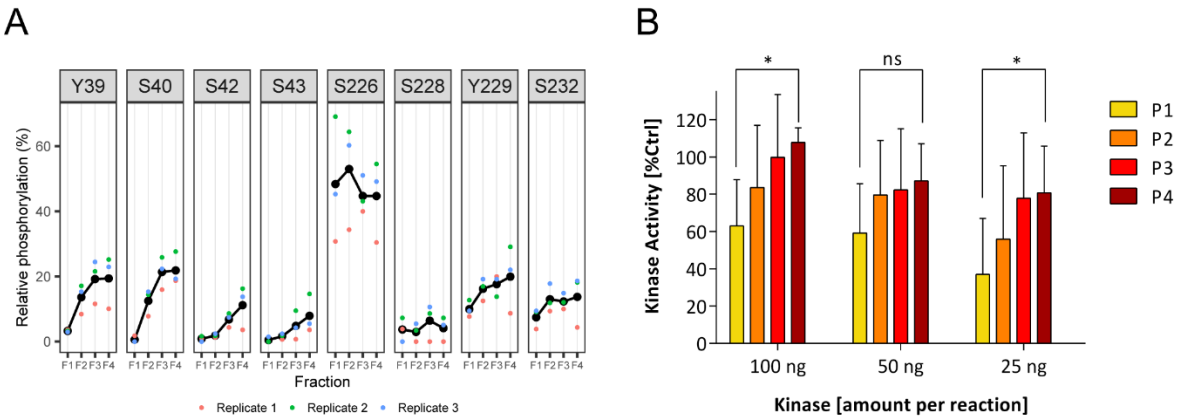


Figure 12. N-terminal phosphorylation of PfGSK3 increases its enzymatic activity. (A) The relative phosphorylation of the four PfGSK3 fractions separated by ion exchange chromatography. The phosphorylation of the N-terminal residues Y39, S40, S42 and S43 increases from fraction 1 to fraction 4. (B) PfGSK3 kinase activity measured for four separated fractions. The fractions with increased phosphorylation exhibit higher enzymatic activity, suggesting that the N-terminal phosphorylation increases the kinase activity of PfGSK3.

Our analysis does not provide the mechanism by which N-terminal phosphorylation increases PfGSK3 activity. However, a phosphorylation-induced dimerization and subsequent structural changes in the active site could explain this effect. The N-terminus of PfGSK3 is predicted to be disordered, however, replacing the phosphorylated residues by phosphomimetic glutamates dramatically increases the probability of coiled-coil formation and a consequent protein dimerization *in silico*. Indeed, modulation of coiled-coil formation by phosphorylation has already been shown in several cases [193–195].

Determination of the structure of PfGSK3 could provide us with the molecular details of all the above-described phenomena. Unfortunately, our attempts to crystallize the protein have been unsuccessful, presumably due to the protein's heterogeneity in phosphorylation, disordered N-terminus and possibly the presence of dimer or oligomer fractions. Our work has also shown that removing the presumably disordered N-terminus does not pave the way towards the GSK3 structure because such protein constructs are insoluble. We have also attempted to insert a cleavage site between the GSK3 N-terminus and its folded domains, however, such protein constructs also ended up in the insoluble fraction, pinpointing the importance of an intact N-terminus. Thus, PfGSK3 alone seems to be a difficult crystallization target. Selection of PfGSK3-nanobodies could ease the crystallization by covering the phosphorylated sequences and mediating crystal contacts. Additionally, if nanobodies specific for distinct phosphorylation species were selected, they would enable preparation of more homogenous protein samples. Cryo-electron microscopy (cryo-EM) could be an alternative approach towards the determination of the PfGSK3 structure. Although the protein itself is too small to determine its monomeric structure by cryo-EM, the high-MW species could be subject to such analysis. Further experiments need to be performed to figure out whether a homogeneous high-MW PfGSK3 sample can be prepared, possibly with the help of cross-linking.

Nevertheless, PfGSK3 remains an important drug target and further investigation of its function, regulation and structure will have a high impact in the search for antimalarial compounds.

2.4. Rhoptries and their biogenesis

The glideosome and its regulation, as described above, are tightly connected to the parasitic secretory organelles – rhoptries and micronemes [24,25]. Both rhoptries and micronemes discharge their content upon parasite egress and host cell invasion, depositing proteins that are essential for the attachment of the host cell and extracellular matrix attachment to the parasite plasma membrane. Functional rhoptries and micronemes therefore play a crucial role in the apicomplexan life cycle. However, their biogenesis is poorly understood. In *T. gondii*, clathrin adaptor protein 1 (AP1) and dynamin-related protein B (DrpB), proteins related to typical eukaryotic factors for organelle biogenesis, have been shown to be

involved in the biogenesis of rhoptries and micronemes [185,196]. Small GTPases, such as Rab11A, Rab5a and Rab5b were also demonstrated to be essential for rhoptry development [197–199].

Lastly, ARO was shown to localise in the membrane of rhoptries, where it utilises its lipidation moiety to anchor at the cytosolic side of the rhoptry membrane [48,50]. The presence of ARO in the rhoptries is a prerequisite for correct positioning of rhoptries in *T. gondii* and its absence from rhoptries leads to defects in rhoptry biogenesis and in parasite invasion. At the same time, several proteins interacting with ARO were identified in *T. gondii*: AIP, ACβ and MyoF. Knock-out studies suggest that ARO, embedded in the outer rhoptry membrane, recruits AIP that in turn ensures correct localisation of ACβ [49,200].

Our results suggest that a different organization of these proteins developed in *P. falciparum*. First, while PfARO localises to the rhoptry bulbs, PfAIP is distinctively located in the rhoptry neck with only a minimal co- localisation with PfARO. Second, using the bioID approach that specifically enriches proteins in the vicinity of PfAIP, we only pulled down PfACβ and PfMyoF but not PfARO, in agreement with their distinct localisation. On the other hand, deletion of the conserved loop 1 of PfARO or point mutations within this loop cause mislocalisation of PfAIP, indicating that the proteins are functionally connected in *P. falciparum*.

These apparently contradictory results could be explained by the fact that PfARO and PfAIP interact only transiently. In such a model, PfARO, using the conserved loop 1, would guide the PfAIP to the rhoptries, but PfAIP would finally assume a distinct location in the rhoptry neck, as observed in *P. falciparum* schizonts. Indeed, PfAIP, with a size of 49 kDa, is much smaller than TgAIP (89 kDa), and does not contain the flanking disordered regions present in TgAIP. These additional residues of TgAIP could explain a functional difference in comparison to PfAIP.

In the second model, the interaction between PfARO and PfAIP could be mediated by another protein. In this case, the absence of PfARO in the list of proteins identified by bioID could be a false negative. In bioID, a target protein (in this case PfAIP) is tagged with BirA, an enzyme that catalyses attachment of a biotin label to primary amines of neighbouring amino acid residues. Therefore, the absence of such residues in proximity of BirA or absence of trypsin cleavage sites in the later sample processing step can cause false negative results.

Having determined the N-terminal loop 1 of PfARO as a functional unit responsible for the recruitment of PfAIP, we hypothesise that the rest of the protein binds PfMyoF. Indeed, the chaperone UNC-45, with a structure similar to the five armadillo repeats of ARO, binds myosin in *Drosophila melanogaster*, where it is essential for muscle myosin stability [201]. However, the function of MyoF in the rhoptries

remains elusive because the biogenesis and positioning of the rhoptries in schizonts are independent of the presence of actin in *T. gondii*. It has been hypothesised that while a microtubule-associated motor is responsible for the transport of rhoptries to the apical pole, MyoF might be important for their apical tethering in the developing daughter cells. In melanocytes, myosin from the same myosin family (myosin Va) have been found to have a similar role in tethering of melanosomes [202].

The last protein associating with PfARO is PfACβ. Interestingly, through production of cyclic adenosine monophosphate (cAMP), PfACβ can directly contribute to regulation of egress and invasion [127]. cAMP binds the regulatory domain of PKA and thereby mediate its dissociation from the PKA catalytic domain. In turn, PKA directly phosphorylates its targets, such as some glideosome proteins, and regulates the activity of other kinases, thereby modulating parasite motility and invasion [181].

Further research is necessary to reveal the function of the PfARO-interacting proteins. The investigation of AIP1 is of particular interest, because its function cannot be predicted from its sequence. In both *T. gondii* and *P. falciparum*, only the core region of AIP (160 amino acid residues) is predicted to fold into a Pfam domain, while the rest of the protein sequences are only restricted to apicomplexans, potentially representing new drug targets. Therefore, AIP is also an interesting target for structural studies.

3. Outlook and concluding remarks

The glideosome, the protein complex responsible for the gliding movement of the motile apicomplexan parasites, requires extensive investigation for the elucidation of its structure and of the function of its components. However, conflicting or lacking data in the literature about the role of lipidation in MLC1/MTIP, and about the localisation and the role of GAP45 and the means of their anchoring by GAP40 and GAP50 call the concept of the glideosome into question. To tackle the various issues with the model that have been raised, an interplay between the methods of structural and molecular biology, and the modern approaches in cellular biology and imaging will play a crucial role.

On one hand, a wider range of strategies for the recombinant expression of the glideosome proteins, including diverse eukaryotic expression systems, cell-free expression system, protein co-expression, co-expression with apicomplexan chaperones or specific kinases or phosphatases, can be the key towards successful determination of the glideosome structure by X-ray crystallography or single particle cryoEM. On the other hand, modern pull-down strategies, using carefully selected detergents that support the integrity of the glideosome membrane proteins, can potentially mediate the identification of new glideosome components, a step that would undoubtedly change our view of the glideosome and have an impact on the application of all other approaches. Finally, a clean pulled-down sample could represent an expensive but interesting approach for the investigation of the glideosome structure by electron microscopy.

Similarly to the structure and the function of the glideosome, biogenesis of the rhoptries is only poorly understood. Protein ARO, depletion of which leads to aberrant function and positioning of the rhoptries, is therefore of a special interest. Although our study provided a thorough look into its function of *P. falciparum*, more questions were raised than answered. Specifically, the role and even the existence of the interaction between ARO and AIP in *P. falciparum* appears elusive. Why do ARO and AIP co- localise in *T. gondii*, but clear distinct localisation to the rhoptry bulb and neck, respectively, supported by bioID, is observed in *P. falciparum*? And if they are localised into distinct rhoptry subcompartments, why and how is PfARO essential for the correct localisation of PfAIP? Given the larger molecular weight of TgAIP compared to PfAIP, investigation of the localisation of AIP proteins of other apicomplexan species could reveal, whether the observed distinct localisations in *T. gondii* and *P. falciparum* can be attributed to additional protein-terminal sequences. At the same time, recombinant expression of ARO and AIP, and characterization of their potential interaction interface by means of structural biology, or other methods of molecular biology such as cross-linking or

hydrogen-deuterium exchange, can help the functional dissection of these proteins on a molecular level. Finally, such results could be confirmed by specific genetic modifications of ARO and AIP *in vivo*.

As drug development is one of the main motivations behind the investigation of apicomplexan organisms, the study of PfGSK3, which is a relevant drug target, is of high importance. We have provided the first insights into the regulation of PfGSK3 and dissected the function of its unique N-terminus. Investigation of PfGSK3 phosphorylation throughout the different *P. falciparum* life stages, for example using pull-down approaches, would be beneficial in order to understand the contribution of N-terminal phosphorylation in particular biological processes. Nevertheless, we could not determine the structure of PfGSK3 as the protein did not crystalize, presumably due to its high heterogeneity in phosphorylation and potentially also oligomerization. I propose two strategies that can facilitate the determination of PfGSK3 structure. The first strategy is based on the selection of nanobodies that would serve as crystallization chaperons of PfGSK3. The selection of the nanobodies can stochastically lead to the identification of binders that specifically interact with a protein with certain phosphorylation patterns, additionally providing a potential platform for the selective purification of homogeneously phosphorylated GSK3. Additionally, the nanobodies can be genetically linked with larger proteins or antibody domains that would enable the determination of PfGSK3 structure *via* single particle cryoEM. The second strategy embraces the utilization of metal-induced high-MW species that can be directly used for single particle cryoEM. This approach would require extensive optimization of the sample preparation to achieve sample homogeneity, but should be feasible as the high-MW species of PfGSK3 maintain their secondary structure.

In summary, it is evident that the amalgamation of biophysical methods, including structural biology and modern *in vivo* approaches of cell biology, will be essential for further advances in the field of apicomplexan research.

4. Bibliography

1. Votýpka J, Modrý D, Obornik M, Šlapeta J, Lukeš J. Apicomplexa. Handbook of the Protists: Second Edition. 2017. doi:10.1007/978-3-319-28149-0_20

2. Seeber F, Steinfelder S. Recent advances in understanding apicomplexan parasites. F1000Research. 2016;5: 1–14. doi:10.12688/f1000research.7924.1

3. Saffo MB, McCoy AM, Rieken C, Slamovits CH. Nephromyces, a beneficial apicomplexan symbiont in marine animals. Proc Natl Acad Sci U S A. 2010;107: 16190–16195. doi:10.1073/pnas.1002335107

4. Cavalier-Smith T. Kingdom protozoa and its 18 phyla. Microbiological Reviews. 1993. doi:10.1128/mmbr.57.4.953-994.1993

5. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. J Eukaryot Microbiol. 2012. doi:10.1111/j.1550-7408.2012.00644.x

6. Wasmuth J, Daub J, Peregrín-Alvarez JM, Finney CAM, Parkinson J. The origins of apicomplexan sequence innovation. Genome Res. 2009;19: 1202–1213. doi:10.1101/gr.083386.108

7. Mehlhorn H, Peters W. The formation of kinetes and oocyst in Plasmodium gallinaceum (Haemosporidia) and considerations on phylogenetic relationships between Haemosporidia, Piroplasmida and other Coccidia. Protistologica. 1980;16: 135–154.

8. Dos Santos Pacheco N, Tosetti N, Koreny L, Waller RF, Soldati-Favre D. Evolution, Composition, Assembly, and Function of the Conoid in Apicomplexa. Trends Parasitol. 2020;36: 688–704. doi:10.1016/j.pt.2020.05.001

9. Wirth D. Malaria: biology in the era of eradication. Malaria: biology in the era of eradication. 2017. doi:10.1111/j.1467-8330.1974.tb00606.x

10. Homer MJ, Aguilar-delfin I, Iii SAMRT, Krause PJ, Persing DH, Ev CLINMIR. Babesiosis. 2000;13: 451–469.

11. Mans BJ, Pienaar R, Latif AA. A review of Theileria diagnostics and epidemiology. Int J Parasitol Parasites Wildl. 2015;4: 104–118. doi:10.1016/j.ijppaw.2014.12.006

12. Leander BS. Marine gregarines: evolutionary prelude to the apicomplexan radiation? Trends Parasitol. 2008;24: 60–67. doi:10.1016/j.pt.2007.11.005

13. Valigurová A, Michalková V, Koudela B. Eugregarine trophozoite detachment from the host epithelium via epimerite retraction: Fiction or fact? Int J Parasitol. 2009;39: 1235–1242. doi:10.1016/j.ijpara.2009.04.009

14. Valigurová A, Hofmannová L, Koudela B, Vávra J. An ultrastructural comparison of the attachment sites between Gregarina steini and Cryptosporidium muris. J Eukaryot Microbiol. 2007;54: 495–510. doi:10.1111/j.1550-7408.2007.00291.x

15. Robert-Gangneux F, Dardé ML. Epidemiology of and diagnostic strategies for toxoplasmosis. Clin Microbiol Rev. 2012;25: 264–296. doi:10.1128/CMR.05013-11

16. Thompson RCA, Olson ME, Zhu G, Enomoto S, Abrahamsen MS, Hijjawi NS. Cryptosporidium and cryptosporidiosis. Adv Parasitol. 2005;59: 77–158. doi:10.1016/S0065-308X(05)59002-X

17. Votýpka J, Modrý D, Oborník M, Jan Š, Luke J. Handbook of the Protists. 2017. doi:10.1007/978-3-319-32669-6

18. Cowman AF, Healer J, Marapana D, Marsh K. Malaria: Biology and Disease. Cell. 2016;167: 610–624. doi:10.1016/j.cell.2016.07.055

19. Frénal K, Dubremetz JF, Lebrun M, Soldati-Favre D. Gliding motility powers invasion and egress in Apicomplexa. Nat Rev Microbiol. 2017;15: 645–660. doi:10.1038/nrmicro.2017.86

20. Heintzelman MB. Cellular and Molecular Mechanics of Gliding Locomotion in Eukaryotes. Int Rev Cytol. 2006;251: 79–129. doi:10.1016/S0074-7696(06)51003-4

21. Frevert U. Sneaking in through the back entrance: The biology of malaria liver stages. Trends Parasitol. 2004;20: 417–424. doi:10.1016/j.pt.2004.07.007

22. Guttery DS, Holder AA, Tewari R. Sexual development in plasmodium: Lessons from functional analyses. PLoS Pathog. 2012;8: 8–10. doi:10.1371/journal.ppat.1002404

23. Dubey JP. Advances in the life cycle of Toxoplasma gondii. Int J Parasitol. 1998;28: 1019–1024. doi:10.1016/S0020-7519(98)00023-X

24. Counihan NA, Kalanon M, Coppel RL, Koning-ward TF De. Plasmodium rhoptry proteins : why order is important. Trends Parasitol. 2013;29: 228–236. doi:10.1016/j.pt.2013.03.003

25. Dubois DJ, Favre DS. Biogenesis and secretion of micronemes in Toxoplasma gondii. 2019; 1–14. doi:10.1111/cmi.13018

26. Santos JM, Lebrun M, Daher W, Soldati D, Dubremetz JF. Apicomplexan cytoskeleton and motors: Key regulators in morphogenesis, cell division, transport and motility. Int J Parasitol. 2009;39: 153–162. doi:10.1016/j.ijpara.2008.10.007

27. Morrissette NS, Sibley LD. Cytoskeleton of Apicomplexan Parasites. Microbiol Mol Biol Rev. 2002;66: 21–38. doi:10.1128/mmbr.66.1.21-38.2002

28. Monteiro VG, De Melo EJT, Attias M, De Souza W. Morphological changes during conoid extrusion in Toxoplasma gondii tachyzoites treated with calcium ionophore. J Struct Biol. 2001;136: 181–189. doi:10.1006/jsbi.2002.4444

29. Cyrklaff M, Kudryashev M, Leis A, Leonard K, Baumeister W, Menard R, et al. Cryoelectron tomography reveals periodic material at the inner side of subpellicular microtubules in apicomplexan parasites. J Exp Med. 2007;204: 1281–1287. doi:10.1084/jem.20062405

30. Spreng B, Fleckenstein H, Kübler P, Di Biagio C, Benz M, Patra P, et al. Microtubule number and length determine cellular shape and function in Plasmodium . EMBO J. 2019;38: 1–22. doi:10.15252/embj.2018100984

31. Harding CR, Meissner M. The inner membrane complex through development of Toxoplasma gondii and Plasmodium. Cell Microbiol. 2014;16: 632–641. doi:10.1111/cmi.12285

32. Gould SB, Tham WH, Cowman AF, McFadden GI, Waller RF. Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. Mol Biol Evol. 2008;25: 1219–1230. doi:10.1093/molbev/msn070

33. Saini E, Zeeshan M, Brady D, Pandey R, Kaiser G, Koreny L, et al. Photosensitized INA-Labelled protein 1 (PhIL1) is novel component of the inner membrane complex and is required for Plasmodium parasite development. Sci Rep. 2017. doi:10.1038/s41598-017-15781-z

34. Bullen HE, Tonkin CJ, O'Donnell RA, Tham WH, Papenfuss AT, Gould S, et al. A novel family of apicomplexan glideosome-associated proteins with an inner membrane-anchoring role. J Biol Chem. 2009;284: 25353–25363. doi:10.1074/jbc.M109.036772

35. Sanders PR, Cantin GT, Greenbaum DC, Gilson PR, Nebl T, Moritz RL, et al. Identification of protein complexes in detergent-resistant membranes of Plasmodium falciparum schizonts. Mol Biochem Parasitol. 2007;154: 148–157. doi:10.1016/j.molbiopara.2007.04.013

36. Wang L, Mohandas N, Thomas A, Coppel RL. Detection of detergent-resistant membranes in asexual blood-stage parasites of Plasmodium falciparum. Mol Biochem Parasitol. 2003;130: 149–153. doi:10.1016/S0166-6851(03)00165-8

37. Chen AL, Kim EW, Toh JY, Vashisht AA, Rashoff AQ, Van C, et al. Novel components of the toxoplasma inner membrane complex revealed by BioID. MBio. 2015;6: 1–12. doi:10.1128/mBio.02357-14

38. Keeley A, Soldati D. The glideosome : a molecular machine powering motility and host-cell invasion by Apicomplexa. 2004;14: 528–532. doi:10.1016/j.tcb.2004.08.001

39. Beeson JG, Drew DR, Boyle MJ, Feng G, Fowkes FJI, Richards JS. Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. FEMS Microbiol Rev. 2016;40: 343–372. doi:10.1093/femsre/fuw001

40. Mercier C, Adjogble KDZ, Däubener W, Delauw MFC. Dense granules: Are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites? Int J Parasitol. 2005;35: 829–849. doi:10.1016/j.ijpara.2005.03.011

41. Carruthers VB, Tomley FM. Receptor-ligand interaction and invasion: Microneme proteins in apicomplexans. Subcell Biochem. 2008;47: 33–45. doi:10.1007/978-0-387-78267-6_2

42. Bannister LH, Hopkins JM, Dluzewski AR, Margos G, Williams IT, Blackman MJ, et al. Plasmodium falciparum apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development. J Cell Sci. 2003. doi:10.1242/jcs.00665

43. Carruthers VB, Sibley LD. Sequential protein secretion front three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. Eur J Cell Biol. 1997.

44. Besteiro S, Michelin A, Poncet J, Dubremetz JF, Lebrun M. Export of a Toxoplasma gondii rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. PLoS Pathog. 2009. doi:10.1371/journal.ppat.1000309

45. Ghosh S, Kennedy K, Sanders P, Matthews K, Ralph SA, Counihan NA, et al. The Plasmodium rhoptry associated protein complex is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth. Cell Microbiol. 2017;19: 1–16. doi:10.1111/cmi.12733

46. Wang Y, Yin H. Research advances in microneme protein 3 of Toxoplasma gondii. Parasites and Vectors. 2015;8: 1–12. doi:10.1186/s13071-015-1001-4

47. Santos JM, Hedberg C, Soldati-Favre D. Protein Acylation: New Potential Targets for Intervention against the Apicomplexa. Apicomplexan Parasites Mol Approaches Towar Target Drug Dev. 2011; 335–357. doi:10.1002/9783527633883.ch18

48. Cabrera A, Herrmann S, Warszta D, Santos JM, John Peter AT, Kono M, et al. Dissection of minimal sequence requirements for rhoptry membrane targeting in the malaria parasite. Traffic. 2012;13: 1335–1350. doi:10.1111/j.1600-0854.2012.01394.x

49. Mueller C, Klages N, Jacot D, Santos JM, Cabrera A, Gilberger TW, et al. The toxoplasma protein ARO mediates the apical positioning of rhoptry organelles, a prerequisite for host cell invasion. Cell Host Microbe. 2013;13: 289–301. doi:10.1016/j.chom.2013.02.001

50. Beck JR, Fung C, Straub KW, Coppens I, Vashisht AA, Wohlschlegel JA, et al. A Toxoplasma Palmitoyl Acyl Transferase and the Palmitoylated Armadillo Repeat Protein TgARO Govern

Apical Rhoptry Tethering and Reveal a Critical Role for the Rhoptries in Host Cell Invasion but Not Egress. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003162

51. Håkansson S, Morisaki H, Heuser J, Sibley LD. Time-lapse video microscopy of gliding motility in Toxoplasma gondii reveals a novel, biphasic mechanism of cell locomotion. Mol Biol Cell. 1999;10: 3539–3547. doi:10.1091/mbc.10.11.3539

52. Boucher LE, Bosch J. The apicomplexan glideosome and adhesins -- structures and function. J Struct Biol. 2016;190: 93–114. doi:10.1016/j.jsb.2015.02.008.The

53. Opitz C, Soldati D. "The glideosome": A dynamic complex powering gliding motion and host cell invasion by Toxoplasma gondii. Mol Microbiol. 2002;45: 597–604. doi:10.1046/j.1365-2958.2002.03056.x

54. Keeley A, Soldati D. The glideosome: A molecular machine powering motility and host-cell invasion by Apicomplexa. Trends in Cell Biology. 2004. doi:10.1016/j.tcb.2004.08.002

55. Daher W, Soldati-Favre D. Mechanisms controlling glideosome function in apicomplexans. Curr Opin Microbiol. 2009;12: 408–414. doi:10.1016/j.mib.2009.06.008

56. Baton LA, Ranford-Cartwright LC. Do malaria ookinete surface proteins P25 and P28 mediate parasite entry into mosquito midgut epithelial cells? Malar J. 2005;4: 1–8. doi:10.1186/1475-2875-4-15

57. Lin CS, Uboldi AD, Epp C, Bujard H, Tsuboi T, Czabotar PE, et al. Multiple plasmodium falciparum merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes. J Biol Chem. 2016;291: 7703–7715. doi:10.1074/jbc.M115.698282

58. Swearingen KE, Lindner SE, Shi L, Shears MJ, Harupa A, Hopp CS, et al. Interrogating the Plasmodium Sporozoite Surface: Identification of Surface-Exposed Proteins and Demonstration of Glycosylation on CSP and TRAP by Mass Spectrometry-Based Proteomics. PLoS Pathog. 2016;12: 1–32. doi:10.1371/journal.ppat.1005606

59. Manger ID, Hehl AB, Boothroyd JC. The surface of Toxoplasma tachyzoites is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. Infect Immun. 1998;66: 2237–2244. doi:10.1128/iai.66.5.2237-2244.1998

60. Bejon P, White MT, Olotu A, Bojang K, Lusingu JPA, Salim N, et al. Efficacy of RTS,S malaria vaccines: Individual-participant pooled analysis of phase 2 data. Lancet Infect Dis. 2013. doi:10.1016/S1473-3099(13)70005-7

61. Dvorak JA, Miller LH, Whitehouse WC, Shiroishi T. Invasion of erythrocytes by malaria merozoites. Science (80-). 1975. doi:10.1126/science.803712

62. Gilson PR, Crabb BS. Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. Int J Parasitol. 2009;39: 91–96. doi:10.1016/j.ijpara.2008.09.007

63. Sim BKL, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum. Science (80-). 1994. doi:10.1126/science.8009226

64. Wong W, Huang R, Menant S, Hong C, Sandow JJ, Birkinshaw RW, et al. Structure of Plasmodium falciparum Rh5–CyRPA–Ripr invasion complex. Nature. 2019;565: 118–121. doi:10.1038/s41586-018-0779-6

65. Alexander DL, Arastu-Kapur S, Dubremetz JF, Boothroyd JC. Plasmodium falciparum AMA1 binds a rhoptry neck protein homologous to TgRON4, a component of the moving junction in Toxoplasma gondii. Eukaryot Cell. 2006. doi:10.1128/EC.00040-06

66. Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC. Identification of the moving

junction complex of Toxoplasma gondii: A collaboration between distinct secretory organelles. PLoS Pathog. 2005. doi:10.1371/journal.ppat.0010017

67. Besteiro S, Dubremetz JF, Lebrun M. The moving junction of apicomplexan parasites: A key structure for invasion. Cell Microbiol. 2011. doi:10.1111/j.1462-5822.2011.01597.x

68. Santos JM, Soldati-Favre D. Invasion factors are coupled to key signalling events leading to the establishment of infection in apicomplexan parasites. Cell Microbiol. 2011;13: 787–796. doi:10.1111/j.1462-5822.2011.01585.x

69. Leykauf K, Treeck M, Gilson PR, Nebl T, Braulke T, Cowman AF, et al. Protein kinase a dependent phosphorylation of apical membrane antigen 1 plays an important role in erythrocyte invasion by the malaria parasite. PLoS Pathog. 2010. doi:10.1371/journal.ppat.1000941

70. Narum DL, Thomas AW. Differential localisation of full-length and processed forms of PF83/AMA-1 an apical membrane antigen of Plasmodium falciparum merozoites. Mol Biochem Parasitol. 1994. doi:10.1016/0166-6851(94)90096-5

71. Hodder AN, Crewther PE, Matthewll MLSM, Reid GE, Moritz RL, Simpson RJ, et al. The disulfide bond structure of Plasmodium apical membrane antigen-1. J Biol Chem. 1996. doi:10.1074/jbc.271.46.29446

72. Prinz B, Harvey KL, Wilcke L, Ruch U, Engelberg K, Biller L, et al. Hierarchical phosphorylation of apical membrane antigen 1 is required for efficient red blood cell invasion by malaria parasites. Sci Rep. 2016;6: 34479. doi:10.1038/srep34479

73. Heintzelman MB, Schwartzman JD. A novel class of unconventional myosins from Toxoplasma gondii. J Mol Biol. 1997;271: 139–146. doi:10.1006/jmbi.1997.1167

74. Pinder JC, Fowler RE, Dluzewski AR, Bannister LH, Lavin FM, Mitchell GH, et al. Actomyosin motor in the merozoite of the malaria parasite, Plasmodium falciparum: Implications for red cell invasion. J Cell Sci. 1998;111: 1831–1839.

75. Meissner M, Schlüter D, Soldati D. Role of Toxoplasma gondii myosin a in powering parasite gliding and host cell invasion. Science (80-). 2002;298: 837–840. doi:10.1126/science.1074553

76. Dobrowolski JM, Carruthers VB, Sibley LD. Participation of myosin in gliding motility and host cell invasion by Toxoplasma gondii . Mol Microbiol. 1997. doi:10.1046/j.1365-2958.1997.5671913.x

77. Schatten H, David Sibley L, Ris H. Structural evidence for actin-like filaments in Toxoplasma gondii using high-resolution low-voltage field emission scanning electron microscopy. Microsc Microanal. 2003. doi:10.1017/S1431927603030095

78. Jacot D, Tosetti N, Pires I, Stock J, Graindorge A, Hung YF, et al. An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion. Cell Host Microbe. 2016. doi:10.1016/j.chom.2016.10.020

79. Dobrowolski JM, Sibley LD. Toxoplasma invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. Cell. 1996. doi:10.1016/S0092-8674(00)81071-5

80. Dobrowolski JM, Niesman IR, Sibley LD. Actin in the parasite toxoplasma gondii is encoded by a single copy gene, act1 and exists primarily in a globular form. Cell Motil Cytoskeleton. 1997. doi:10.1002/(SICI)1097-0169(1997)37:3<253::AID-CM7>3.0.CO;2-7

81. Periz J, Whitelaw J, Harding C, Gras S, Minina MIDR, Latorre-Barragan F, et al. Toxoplasma gondii F-actin forms an extensive filamentous network required for material exchange and parasite maturation. Elife. 2017;6: 1–29. doi:10.7554/eLife.24119

82. Dos N, Pacheco S, Soldati-favre D, Jacot D. Three F-actin assembly centers regulate organelle inheritance , cell-cell communication and motility in Toxoplasma gondii. 2019; 1–32.

83. Buscaglia CA, Krumm B, Ingason BP, Lucas R, Roach C. Aldolase provides an unusual binding site for thrombospondin-related anonymous protein in the invasion machinery of the malaria parasite. 2007.

84. Nemetski SM, Cardozo TJ, Bosch G, Weltzer R, Malley KO, Ejigiri I, et al. Inhibition by stabilization : targeting the Plasmodium falciparum aldolase – TRAP complex. Malar J. 2015; 1–18. doi:10.1186/s12936-015-0834-9

85. Kappe S, Bruderer T, Gantt S, Fujioka H, Nussenzweig V, Menard R. Conservation of a Gliding Motility and Cell Invasion Machinery in Apicomplexan Parasites. J Cell Biol. 1999;147: 937–943. doi:10.1002/9783527633883

86. Bartholdson SJ, Bustamante LY, Crosnier C, Johnson S, Lea S, Rayner JC, et al. Semaphorin-7A Is an Erythrocyte Receptor for P. falciparum Merozoite-Specific TRAP Homolog, MTRAP. PLoS Pathog. 2012;8. doi:10.1371/journal.ppat.1003031

87. Baum J, Richard D, Healer J, Rug M, Krnajski Z, Gilberger TW, et al. A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. J Biol Chem. 2005;281: 5197–5208. doi:10.1074/jbc.M509807200

88. Robson KJH, Hall JRS, Jennings MW, Harris TJR, Marsh K, Newbold CI, et al. A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. Nature. 1988. doi:10.1038/335079a0

89. Sultan AA, Thathy V, Frevert U, Robson KJH, Crisanti A, Nussenzweig V, et al. TRAP is necessary for gliding motility and infectivity of Plasmodium sporozoites. Cell. 1997;90: 511–522. doi:10.1016/S0092-8674(00)80511-5

90. Trottein F, Triglia T, Cowman AF. Molecular cloning of a gene from Plasmodium falciparum that codes for a protein sharing motifs found in adhesive molecules from mammals and Plasmodia. Mol Biochem Parasitol. 1995. doi:10.1016/0166-6851(95)02489-1

91. Buscaglia CA, Coppens I, Hol WGJ, Nussenzweig V. Sites of Interaction between Aldolase and Thrombospondin-related Anonymous Protein in Plasmodium. Mol Biol Cell. 2003. doi:10.1091/mbc.E03-06-0355

92. Edition S. Myosins.

93. FrÉnal K, Foth BJ, Soldati D. Myosin Class XIV And Other Myosins In Protists. Myosins. 2007; 421–440. doi:10.1007/978-1-4020-6519-4_15

94. Herm-Götz A, Weiss S, Stratmann R, Fujita-Becker S, Ruff C, Meyhöfer E, et al. Toxoplasma gondii myosin A and its light chain: A fast, single-headed, plus-end-directed motor. EMBO J. 2002;21: 2149–2158. doi:10.1093/emboj/21.9.2149

95. Nebl T, Prieto JH, Kapp E, Smith BJ, Williams MJ, Yates JR, et al. Quantitative in vivo analyses reveal calcium-dependent phosphorylation sites and identifies a novel component of the toxoplasma invasion motor complex. PLoS Pathog. 2011;7. doi:10.1371/journal.ppat.1002222

96. Green JL, Wall RJ, Vahokoski J, Yusuf NA, Mohd Ridzuan MA, Stanway RR, et al. Compositional and expression analyses of the glideosome during the Plasmodium life cycle reveal an additional myosin light chain required for maximum motility. J Biol Chem. 2017;292: 17857–17875. doi:10.1074/jbc.M117.802769

97. Bookwalter CS, Tay CL, McCrorie R, Previs MJ, Lu H, Krementsova EB, et al. Reconstitution of the core of the malaria parasite glideosome with recombinant Plasmodium class XIV myosin A and Plasmodium actin. J Biol Chem. 2017;292: 19290–19303. doi:10.1074/jbc.M117.813972

98. Powell CJ, Ramaswamy R, Kelsen A, Hamelin DJ, Warshaw DM, Bosch J, et al. Structural and mechanistic insights into the function of the unconventional class XIV myosin MyoA from

Toxoplasma gondii. Proc Natl Acad Sci U S A. 2018. doi:10.1073/pnas.1811167115

99. Robert-Paganin J, Robblee JP, Auguin D, Blake TCA, Bookwalter CS, Krementsova EB, et al. Plasmodium myosin A drives parasite invasion by an atypical force generating mechanism. Nat Commun. 2019. doi:10.1038/s41467-019-11120-0

100. Green JL, Martin SR, Fielden J, Ksagoni A, Grainger M, Yim Lim BYS, et al. The MTIP-myosin A complex in blood stage malaria parasites. J Mol Biol. 2006;355: 933–941. doi:10.1016/j.jmb.2005.11.027

101. Siden-Kiamos I, Pinder JC, Louis C. Involvement of actin and myosins in Plasmodium berghei ookinete motility. Mol Biochem Parasitol. 2006;150: 308–317. doi:10.1016/j.molbiopara.2006.09.003

102. Bookwalter CS, Kelsen A, Leung JM, Ward GE, Trybus KM. A toxoplasma gondii class XIV myosin, expressed in Sf9 cells with a parasite co-chaperone, requires two light chains for fast motility. J Biol Chem. 2014;289: 30832–30841. doi:10.1074/jbc.M114.572453

103. Logvinova DS, Levitsky DI. Essential Light Chains of Myosin and Their Role in Functioning of the Myosin Motor. Biochemistry (Moscow). 2018. doi:10.1134/S0006297918080060

104. Houdusse A, Silver M, Cohen C. A model of Ca2+-free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch. Structure. 1996. doi:10.1016/S0969-2126(96)00154-2

105. Zhang Y, Kawamichi H, Kohama K, Nakamura A. Calcium-mediated regulation of recombinant hybrids of full-length Physarum myosin heavy chain with Physarum/scallop myosin light chains. Acta Biochim Biophys Sin (Shanghai). 2016. doi:10.1093/abbs/gmw031

106. Williams MJ, Alonso H, Enciso M, Egarter S, Sheiner L, Meissner M, et al. Two Essential Light Chains Regulate the MyoA Lever Arm To. 2015;6: 1–16. doi:10.1128/mBio.00845-15.Editor

107. Powell CJ, Jenkins ML, Parker ML, Ramaswamy R, Kelsen A, Warshaw DM, et al. Dissecting the molecular assembly of the Toxoplasma gondii MyoA motility complex. J Biol Chem. 2017. doi:10.1074/jbc.M117.809632

108. Blake TCA, Haase S, Baum J. Actomyosin forces and the energetics of red blood cell invasion by the malaria parasite Plasmodium falciparum. bioRxiv. 2020. doi:10.1101/2020.06.25.171900

109. Heissler SM, Sellers JR. Myosin light chains: Teaching old dogs new tricks. Bioarchitecture. 2014. doi:10.1080/19490992.2015.1054092

110. Ni S, Hong F, Haldeman BD, Baker JE, Facemyer KC, Cremo CR. Modification of interface between regulatory and essential light chains hampers phosphorylation-dependent activation of smooth muscle myosin. J Biol Chem. 2012. doi:10.1074/jbc.M112.343491

111. Colson BA, Gruber SJ, Thomas DD. Structural dynamics of muscle protein phosphorylation. Journal of Muscle Research and Cell Motility. 2012. doi:10.1007/s10974-012-9317-6

112. Espinoza-Fonseca LM, Colson BA, Thomas DD. Effects of pseudophosphorylation mutants on the structural dynamics of smooth muscle myosin regulatory light chain. Mol Biosyst. 2014. doi:10.1039/c4mb00364k

113. Kast D, Espinoza-Fonseca LM, Yi C, Thomas DD. Phosphorylation-induced structural changes in smooth muscle myosin regulatory light chain. Proc Natl Acad Sci U S A. 2010. doi:10.1073/pnas.1001941107

114. Bergman LW. Myosin A tail domain interacting protein (MTIP) localises to the inner membrane complex of Plasmodium sporozoites. J Cell Sci. 2003;116: 39–49. doi:10.1242/jcs.00194

115. Douse CH, Green JL, Salgado PS, Simpson PJ, Thomas JC, Langsley G, et al. Regulation of the

plasmodium motor complex: Phosphorylation of myosin a tail-interacting protein (MTIP) loosens its grip on MyoA. J Biol Chem. 2012;287: 36968–36977. doi:10.1074/jbc.M112.379842

116. Frénal K, Polonais V, Marq JB, Stratmann R, Limenitakis J, Soldati-Favre D. Functional dissection of the apicomplexan glideosome molecular architecture. Cell Host Microbe. 2010. doi:10.1016/j.chom.2010.09.002

117. Gaskins E, Gilk S, DeVore N, Mann T, Ward G, Beckers C. Identification of the membrane receptor of a class XIV myosin in Toxoplasma gondii. J Cell Biol. 2004;165: 383–393. doi:10.1083/jcb.200311137

118. Harding CR, Egarter S, Gow M, Jiménez-Ruiz E, Ferguson DJP, Meissner M. Gliding Associated Proteins Play Essential Roles during the Formation of the Inner Membrane Complex of Toxoplasma gondii. PLoS Pathog. 2016. doi:10.1371/journal.ppat.1005403

119. Rees-Channer RR, Martin SR, Green JL, Bowyer PW, Grainger M, Molloy JE, et al. Dual acylation of the 45 kDa gliding-associated protein (GAP45) in Plasmodium falciparum merozoites. Mol Biochem Parasitol. 2006;149: 113–116. doi:10.1016/j.molbiopara.2006.04.008

120. Perrin AJ, Collins CR, Russell MRG, Collinson LM, Baker DA, Blackman MJ. The actinomyosin motor drives malaria parasite red blood cell invasion but not egress. MBio. 2018. doi:10.1128/mBio.00905-18

121. Egarter S, Andenmatten N, Jackson AJ, Whitelaw JA, Pall G, Black JA, et al. The toxoplasma acto-myoA motor complex is important but not essential for gliding motility and host cell invasion. PLoS One. 2014;9. doi:10.1371/journal.pone.0091819

122. Tosetti N, Pacheco N dos S, Bertiaux E, Maco B, Bournonville L, Hamel V, et al. Essential function of the alveolin network in the subpellicular microtubules and conoid assembly in Toxoplasma gondii. Elife. 2020;9: 1–22. doi:10.7554/eLife.56635

123. Bosch J, Paige MH, Vaidya AB, Bergman LW, Hol WGJ. Crystal structure of GAP50, the anchor of the invasion machinery in the inner membrane complex of Plasmodium falciparum. J Struct Biol. 2012;178: 61–73. doi:10.1016/j.jsb.2012.02.009

124. Jacot D, Frénal K, Marq JB, Sharma P, Soldati-Favre D. Assessment of phosphorylation in Toxoplasma glideosome assembly and function. Cell Microbiol. 2014;16: 1518–1532. doi:10.1111/cmi.12307

125. Treeck M, Sanders JL, Elias JE, Boothroyd JC. The phosphoproteomes of plasmodium falciparum and toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. Cell Host Microbe. 2011;10: 410–419. doi:10.1016/j.chom.2011.09.004

126. Xie X, Harrison DH, Schlichting I, Sweet RM, Kalabokis VN, Szent-Györgyi AG, et al. Structure of the regulatory domain of scallop myosin at 2.8 Å resolution. Nature. 1994. doi:10.1038/368306a0

127. Bisio H, Soldati-favre D. Signaling Cascades Governing Entry into and Exit from Host Cells by Toxoplasma gondii. Annu Rev. 2019.

128. Vella SA, Moore CA, Li Z-H, Hortua Triana MA, Potapenko E, Moreno SN. The Role of Potassium and Host Calcium Signaling in Toxoplasma gondii egress. bioRxiv Microbiol. 2020; 2020.03.06.980508. doi:10.1101/2020.03.06.980508

129. Yeoman JA, Hanssen E, Maier AG, Klonis N, Maco B, Baum J, et al. Tracking glideosome-associated protein 50 reveals the development and organization of the inner membrane complex of Plasmodium falciparum. Eukaryot Cell. 2011;10: 556–564. doi:10.1128/EC.00244-10

130. Fauquenoy S, Hovasse A, Sloves P-J, Morelle W, Dilezitoko Alayi T, Dilezitoko Ayali T, et al. Unusual N-glycan structures required for trafficking Toxoplasma gondii GAP50 to the inner

membrane complex regulate host cell entry through parasite motility. Mol Cell Proteomics. 2011;10: M111.008953. doi:10.1074/mcp.M111.008953

131. Kumar V, Behl A, Kapoor P, Nayak B, Singh G, Singh AP, et al. Inner membrane complex 1l protein of Plasmodium falciparum links membrane lipids with cytoskeletal element "actin" and its associated motor "myosin". Int J Biol Macromol. 2019;126: 673–684. doi:10.1016/j.ijbiomac.2018.12.239

132. Wan W, Briggs JAG. Cryo-Electron Tomography and Subtomogram Averaging. Methods in Enzymology. 2016. doi:10.1016/bs.mie.2016.04.014

133. World Health Organization. World Malaria Report. https://www.who.int/malaria. 2019. Available: https://www.who.int/publications-detail/world-malaria-report-2019

134. Antinori S, Galimberti L, Milazzo L, Corbellino M. Biology of human malaria plasmodia including Plasmodium knowlesi. Mediterr J Hematol Infect Dis. 2012;4. doi:10.4084/MJHID.2012.013

135. Bruneel F. Human cerebral malaria: 2019 mini review. Rev Neurol (Paris). 2019;175: 445–450. doi:10.1016/j.neurol.2019.07.008

136. Singh B. Human Infections and Detection of Plasmodium knowlesi. 2013;26: 165–184. doi:10.1128/CMR.00079-12

137. Tse EG, Korsik M, Todd MH. The past, present and future of anti-malarial medicines. Malar J. 2019;18: 1–21. doi:10.1186/s12936-019-2724-z

138. Achan J, Talisuna AO, Erhart A, Yeka A, Tibenderana JK, Baliraine FN, et al. Quinine, an old anti-malarial drug in a modern world: Role in the treatment of malaria. Malar J. 2011;10: 144. Available: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L51444620%5Cnhttp://dx.doi.org/10.1186/1475-2875-10-144%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=14752875&id=doi:10.1186%2F1475-2875-10-144&atitle=Quinine%2C+an+old+anti-m

139. Dziekan JM, Yu H, Chen D, Dai L, Wirjanata G, Larsson A, et al. Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay. Sci Transl Med. 2019. doi:10.1126/scitranslmed.aau3174

140. Organización Mundial de la Salud (OMS). World health organization model list of essential medicines. Ment Holist Heal Some Int Perspect. 2019;21: 119–134.

141. Fidock DA, Wellems TE. Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Proc Natl Acad Sci U S A. 1997;94: 10931–10936. doi:10.1073/pnas.94.20.10931

142. Bzik DJ, Li WB, Horii T, Inselburg J. Molecular cloning and sequence analysis of the Plasmodium falciparum dihydrofolate reductase-thymidylate synthase gene. Proc Natl Acad Sci U S A. 1987. doi:10.1073/pnas.84.23.8360

143. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, et al. Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. N Engl J Med. 2014;371: 411–423. doi:10.1056/NEJMoa1314981

144. Keating C. The history of the RTS,S/AS01 malaria vaccine trial. Lancet. 2020;395: 1336–1337. doi:10.1016/S0140-6736(20)30815-1

145. Rompikuntal PK, Foe IT, Deng B, Bogyo M, Ward GE. Blocking palmitoylation of Toxoplasma gondii myosin light chain 1 disrupts glideosome composition but has little impact on parasite motility. bioRxiv. 2020; 2020.08.13.250399. doi:10.1101/2020.08.13.250399

146. Douse CH, Maas SJ, Thomas JC, Garnett JA, Sun Y, Cota E, et al. Crystal structures of stapled and hydrogen bond surrogate peptides targeting a fully buried protein-helix interaction. ACS Chem Biol. 2014;9: 2204–2209. doi:10.1021/cb500271c

147. Douse CH, Vrielink N, Wenlin Z, Cota E, Tate EW. Targeting a dynamic protein-protein interaction: Fragment screening against the malaria myosin a motor complex. ChemMedChem. 2015;10: 134–143. doi:10.1002/cmdc.201402357

148. Kortagere S, Welsh WJ, Morrisey JM, Daly T, Ejigiri I, Sinnis P, et al. Structure-based Design of Novel Small-Molecule Inhibitors of Plasmodium falciparum. J Chem Inf Model. 2010;50: 840–849. doi:10.1161/CIRCULATIONAHA.110.956839

149. Anam Z, Joshi N, Gupta S, Yadav P, Chaurasiya A. A De novo Peptide from a High Throughput Peptide Library Blocks Myosin A -MTIP Complex Formation in Plasmodium falciparum. Int J Mol Sci. 2020;21: 6158. doi:10.3390/ijms21176158

150. Welsh WJ, Kortagere S, Bergman LW. MECHANISM-BASED SMALL-MOLECULE PARASITE INHIBITOR. WO 2009/065096 Al, 2009.

151. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. Drug Discovery Today: Technologies. 2004. doi:10.1016/j.ddtec.2004.11.007

152. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002. doi:10.1021/jm020017n

153. Frénal K, Marq JB, Jacot D, Polonais V, Soldati-Favre D. Plasticity between MyoC- and MyoA-glideosomes: an example of functional compensation in Toxoplasma gondii invasion. PLoS Pathog. 2014;10: e1004504. doi:10.1371/journal.ppat.1004504

154. Frénal K, Soldati-Favre D. Plasticity and Redundancy in Proteins Important for Toxoplasma Invasion. PLoS Pathog. 2015;11: 1–6. doi:10.1371/journal.ppat.1005069

155. Frénal K, Polonais V, Marq JB, Stratmann R, Limenitakis J, Soldati-Favre D. Functional dissection of the apicomplexan glideosome molecular architecture. Cell Host Microbe. 2010;8: 343–357. doi:10.1016/j.chom.2010.09.002

156. Fritz HM, Buchholz KR, Chen X, Durbin-Johnson B, Rocke DM, Conrad PA, et al. Transcriptomic analysis of toxoplasma development reveals many novel functions and structures specific to sporozoites and oocysts. PLoS One. 2012. doi:10.1371/journal.pone.0029998

157. Wall RJ, Zeeshan M, Katris NJ, Limenitakis R, Rea E, Stock J, et al. Systematic analysis of Plasmodium myosins reveals differential expression, localisation, and function in invasive and proliferative parasite stages . Cell Microbiol. 2019; 1–12. doi:10.1111/cmi.13082

158. Whitelaw JA, Latorre-Barragan F, Gras S, Pall GS, Leung JM, Heaslip A, et al. Surface attachment, promoted by the actomyosin system of Toxoplasma gondii is important for efficient gliding motility and invasion. BMC Biol. 2017;15. doi:10.1186/s12915-016-0343-5

159. Delbac F, Sänger A, Neuhaus EM, Stratmann R, Ajioka JW, Toursel C, et al. Toxoplasma gondii myosins B/C: One gene, two tails, two localisations, and a role in parasite division. J Cell Biol. 2001;155: 613–623. doi:10.1083/jcb.200012116

160. Graindorge A, Frénal K, Jacot D, Salamun J. The Conoid Associated Motor MyoH Is Indispensable for Toxoplasma gondii Entry and Exit from Host Cells. 2016; 1–26. doi:10.1371/journal.ppat.1005388

161. Heintzelman MB, Schwartzman JD. Characterization of Myosin-A and Myosin-C: Two class XIV unconventional myosins from Toxoplasma gondii. Cell Motil Cytoskeleton. 1999;44: 58–

67. doi:10.1002/(SICI)1097-0169(199909)44:1<58::AID-CM5>3.0.CO;2-R

162. Kumpula EP, Kursula I. Towards a molecular understanding of the apicomplexan actin motor: On a road to novel targets for malaria remedies? Acta Crystallogr Sect FStructural Biol Commun. 2015;71: 500–513. doi:10.1107/S2053230X1500391X

163. Bisio H, Lunghi M, Brochet M, Soldati-favre D. Phosphatidic acid governs natural egress in Toxoplasma gondii via a guanylate cyclase receptor platform. Nat Microbiol. 2019;4: 420–428. doi:10.1038/s41564-018-0339-8

164. Bullen HE, Soldati-Favre D. A central role for phosphatidic acid as a lipid mediator of regulated exocytosis in apicomplexa. FEBS Lett. 2016;590: 2469–2481. doi:10.1002/1873-3468.12296

165. et al. C. Essential cGMP signaling in Toxoplasma is initiated by a hybrid P-type ATPase / guanylate cyclase Kevin. Physiol Behav. 2018;24: 804–516. doi:10.1016/j.chom.2018.10.015

166. Moreno SNJ, Ayong L, Pace DA. Calcium storage and function in apicomplexan parasites. Essays Biochem. 2011;51: 97–110. doi:10.1042/BSE0510097

167. Bullen HE, Jia Y, Yamaryo-Botté Y, Bisio H, Zhang O, Jemelin NK, et al. Phosphatidic Acid-Mediated Signaling Regulates Microneme Secretion in Toxoplasma. Cell Host Microbe. 2016;19: 349–360. doi:10.1016/j.chom.2016.02.006

168. Guerra AJ, Zhang O, Bahr CME, Huynh MH, DelProposto J, Brown WC, et al. Structural basis of Toxoplasma gondii perforin-like protein 1 membrane interaction and activity during egress. PLoS Pathog. 2018;14: 1–21. doi:10.1371/journal.ppat.1007476

169. Brochet M, Billker O. Calcium signalling in malaria parasites. Mol Microbiol. 2016;100: 397–408. doi:10.1111/mmi.13324

170. Green JL, Rees-Channer RR, Howell SA, Martin SR, Knuepfer E, Taylor HM, et al. The motor complex of Plasmodium falciparum: Phosphorylation by a calcium-dependent protein kinase. J Biol Chem. 2008;283: 30980–30989. doi:10.1074/jbc.M803129200

171. Winter D, Kugelstadt D, Seidler J, Kappes B, Lehmann WD. Protein phosphorylation influences proteolytic cleavage and kinase substrate properties exemplified by analysis of in vitro phosphorylated Plasmodium falciparum glideosome-associated protein 45 by nano-ultra performance liquid chromatography-tandem mass s. Anal Biochem. 2009;393: 41–47. doi:10.1016/j.ab.2009.06.022

172. Ridzuan MAM, Moon RW, Knuepfer E, Black S, Holder AA, Green JL. Subcellular location, phosphorylation and assembly into the motor complex of GAP45 during Plasmodium falciparum schizont development. PLoS One. 2012;7. doi:10.1371/journal.pone.0033845

173. Thomas DC, Ahmed A, Gilberger TW, Sharma P. Regulation of Plasmodium falciparum glideosome associated protein 45 (PfGAP45) phosphorylation. PLoS One. 2012;7: 15–17. doi:10.1371/journal.pone.0035855

174. Kumar S, Kumar M, Ekka R, Dvorin JD, Paul AS, Madugundu AK, et al. PfCDPK1 mediated signaling in erythrocytic stages of Plasmodium falciparum. Nat Commun. 2017;8: 1–12. doi:10.1038/s41467-017-00053-1

175. Lasonder E, Green JL, Camarda G, Talabani H, Holder AA, Langsley G, et al. The plasmodium falciparum schizont phosphoproteome reveals extensive phosphatidylinositol and cAMP-protein kinase A signaling. J Proteome Res. 2012;11: 5323–5337. doi:10.1021/pr300557m

176. Treeck M, Sanders JL, Gaji RY, LaFavers KA, Child MA, Arrizabalaga G, et al. The Calcium-Dependent Protein Kinase 3 of Toxoplasma Influences Basal Calcium Levels and Functions beyond Egress as Revealed by Quantitative Phosphoproteome Analysis. PLoS Pathog. 2014;10. doi:10.1371/journal.ppat.1004197

177. Govindasamy K, Khan R, Snyder M, Lou HJ, Du P, Kudyba HM, et al. Plasmodium falciparum cyclic GMP-dependent protein kinase interacts with a subunit of the parasite proteasome. Infect Immun. 2019;87: 1–13. doi:10.1128/IAI.00523-18

178. Harvey KL, Yap A, Gilson PR, Cowman AF, Crabb BS. Insights and controversies into the role of the key apicomplexan invasion ligand, Apical Membrane Antigen 1. Int J Parasitol. 2014. doi:10.1016/j.ijpara.2014.08.001

179. Kumar A, Vaid A, Syin C, Sharma P. PfPKB, a novel protein kinase B-like enzyme from Plasmodium falciparum. I. Identification, characterization, and possible role in parasite development. J Biol Chem. 2004;279: 24255–24264. doi:10.1074/jbc.M312855200

180. Vaid A, Thomas DC, Sharma P. Role of Ca2+/calmodulin-PfPKB signaling pathway in erythrocyte invasion by Plasmodium falciparum. J Biol Chem. 2008;283: 5589–5597. doi:10.1074/jbc.M708465200

181. Uboldi AD, Wilde ML, McRae EA, Stewart RJ, Dagley LF, Yang L, et al. Protein kinase A negatively regulates Ca 2+ signalling in Toxoplasma gondii. PLoS Biology. 2018. doi:10.1371/journal.pbio.2005642

182. Brown KM, Long S, Sibley LD. Plasma membrane association by N-acylation governs PKG function in Toxoplasma gondii. MBio. 2017;8: 1–14. doi:10.1128/mBio.00375-17

183. Cabrera DG, Horatscheck A, Wilson CR, Basarab G, Eyermann CJ, Chibale K. Plasmodial Kinase Inhibitors: License to Cure? J Med Chem. 2018;61: 8061–8077. doi:10.1021/acs.jmedchem.8b00329

184. Harmon AC, Gribskov M, Gubrium E, Harper JF. The CDPK superfamily of protein kinases. New Phytologist. 2001. doi:10.1046/j.1469-8137.2001.00171.x

185. Jonathan Posner, James A. Russell and BSP. Development of Toxoplasma gondii Calcium-Dependent Protein Kinase 1 (TgCDPK1) Inhibitors with Potent Anti-Toxoplasma Activity. J Med Virol. 2012;55: 2416–2426. doi:10.1021/jm201713h

186. Lemercier G, Fernandez-Montalvan A, Shaw JP, Kugelstadt D, Bomke J, Domostoj M, et al. Identification and characterization of novel small molecules as potent inhibitors of the plasmodial calcium-dependent protein kinase. Biochemistry. 2009. doi:10.1021/bi9005122

187. Vanaerschot M, Murithi JM, Pasaje CFA, Ghidelli-Disse S, Dwomoh L, Bird M, et al. Inhibition of Resistance-Refractory P. falciparum Kinase PKG Delivers Prophylactic, Blood Stage, and Transmission-Blocking Antiplasmodial Activity. Cell Chem Biol. 2020. doi:10.1016/j.chembiol.2020.04.001

188. Diaz CA, Allocco J, Powles MA, Yeung L, Donald RGK, Anderson JW, et al. Characterization of Plasmodium falciparum cGMP-dependent protein kinase (PfPKG): Antiparasitic activity of a PKG inhibitor. Mol Biochem Parasitol. 2006. doi:10.1016/j.molbiopara.2005.10.020

189. Masch A, Nasereddin A, Alder A, Bird MJ, Schweda SI, Preu L, et al. Structure-activity relationships in a series of antiplasmodial thieno[2,3-b]pyridines. Malar J. 2019;18: 1–10. doi:10.1186/s12936-019-2725-y

190. Schweda SI, Alder A, Gilberger T. 4-Arylthieno[2,3-b]pyridine-2-carboxamides Are a New Class of Antiplasmodial Agents Sandra. Molecules. : 1–39.

191. Lolli G, Pinna LA, Battistutta R. Structural determinants of protein kinase CK2 regulation by autoinhibitory polymerization. ACS Chem Biol. 2012;7: 1158–1163. doi:10.1021/cb300054n

192. Sutherland C, Leighton IA, Cohen P. Inactivation of glycogen synthase kinase-3β by phosphorylation: New kinase connections in insulin and growth-factor signalling. Biochem J. 1993. doi:10.1042/bj2960015

193. Szilák L, Moitra J, Krylov D, Vinson C. Phosphorylation destabilizes α-helices. Nat Struc. 1997;4: 112–114. doi:10.1038/nsb0297-112

194. Derewenda U, Tarricone C, Choi WC, Cooper DR, Lukasik S, Perrina F, et al. The Structure of the Coiled-Coil Domain of Ndel1 and the Basis of Its Interaction with Lis1, the Causal Protein of Miller-Dieker Lissencephaly. Structure. 2007;15: 1467–1481. doi:10.1016/j.str.2007.09.015

195. Szilák L, Moitra J, Vinson C. Design of a leucine zipper coiled coil stabilized 1.4 kcal mol-1 by phosphorylation of a serine in the e position. Protein Sci. 1997;6: 1273–1283. doi:10.1002/pro.5560060615

196. Venugopal K, Werkmeister E, Barois N, Saliou JM, Poncet A, Huot L, et al. Dual role of the Toxoplasma gondii clathrin adaptor AP1 in the sorting of rhoptry and microneme proteins and in parasite division. PLoS Pathogens. 2017. doi:10.1371/journal.ppat.1006331

197. Kremer K, Kamin D, Rittweger E, Wilkes J, Flammer H, Mahler S, et al. An Overexpression Screen of Toxoplasma gondii Rab-GTPases Reveals Distinct Transport Routes to the Micronemes. PLoS Pathog. 2013. doi:10.1371/journal.ppat.1003213

198. Bradley PJ, Ward C, Cheng SJ, Alexander DL, Coller S, Coombs GH, et al. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii. J Biol Chem. 2005. doi:10.1074/jbc.M504158200

199. Agop-Nersesian C, Naissant B, Rached F Ben, Rauch M, Kretzschmar A, Thiberge S, et al. Rab11A-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis. PLoS Pathog. 2009. doi:10.1371/journal.ppat.1000270

200. Mueller C, Samoo A, Hammoudi P, Klages N. Structural and functional dissection of Toxoplasma gondii armadillo repeats only protein (TgARO). 2016.

201. Lee CF, Melkani GC, Yu Q, Suggs JA, Kronert WA, Suzuki Y, et al. Drosophila UNC-45 accumulates in embryonic blastoderm and in muscles, and is essential for muscle myosin stability. J Cell Sci. 2011. doi:10.1242/jcs.078964

202. Weisman LS. Organelles on the move: Insights from yeast vacuole inheritance. Nature Reviews Molecular Cell Biology. 2006. doi:10.1038/nrm1892

5. Manuscript 1

Structural role of essential light chains in the apicomplexan glideosome

Samuel Pazicky[1,2], Karthikeyan Dhamotharan[1,2], Karol Kaszuba[1,2], Haydyn DT Mertens[2], Tim Gilberger[1,3,4], Dmitri Svergun[2], Jan Kosinski[1,2,5], Ulrich Weininger[6] and Christian Löw[1,2*]

[1] Centre for Structural Systems Biology (CSSB), Notkestrasse 85, D-22607 Hamburg, Germany.

[2] Molecular Biology Laboratory (EMBL), Hamburg Unit c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, D-22607 Hamburg, Germany.

[3] Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Strasse 74, D-20359 Hamburg, Germany.

[4] Department of Biology, University of Hamburg, Hamburg, Germany

[5] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[6] Martin-Luther-University Halle-Wittenberg, Institute of Physics, Biophysics, D-06120 Halle (Saale), Germany

Abstract

Gliding, a type of motility based on an actin-myosin motor, is specific to apicomplexan parasites. Myosin A binds two light chains which further interact with glideosome associated proteins and assemble into the glideosome. The role of individual glideosome proteins is unclear due to the lack of structures of larger glideosome assemblies. Here, we investigate the role of essential light chains (ELCs) in *Toxoplasma gondii* and *Plasmodium falciparum* and present their crystal structures as part of trimeric sub-complexes. We show that although ELCs bind a conserved MyoA sequence, *P. falciparum* ELC adopts a distinct structure in the free and MyoA-bound state. We suggest that ELCs enhance MyoA performance by inducing secondary structure in MyoA and thus stiffen its lever arm. Structural and biophysical analysis reveals that calcium binding has no influence on the structure of ELCs. Our work represents a further step towards understanding the mechanism of gliding in *Apicomplexa*.

Introduction

Apicomplexa are a phylum of intracellular, parasitic, single cell eukaryotes with high medical and agricultural relevance. For instance, *Plasmodium species* are the causative agents of malaria, that lead to 414.000 deaths per year[1]. Another apicomplexan parasite, *Toxoplasma gondii*, infects more than 30% of the population worldwide with no clinical symptoms but can cause severe damage in immunocompromised patients and in pregnant women[2]. Proliferation and transmission of these obligate endoparasites in their host organisms rely on efficient cell invasion[3]. This active process is based on the motility of the parasite, referred to as gliding, and is empowered by an actin/myosin motor[4,5]. This motor is localized within the intermembrane space between the parasite's plasma membrane and inner membrane complex (IMC), an additional double-layer of membranes that is unique for these single cell organisms[6]. The IMC provides stability to invasion competent stages of the parasite and functions as an anchor for the actin/myosin motor. While motility is achieved by the interaction of the myosin motor with actin filaments, myosin is linked to the IMC by a membrane-embedded multi-protein complex referred to as the glideosome[7–9] (Fig. 1).

According to the current model, the apicomplexan glideosome is composed of six proteins: myosin MyoA, essential light chain ELC, myosin light chain MLC1, and the glideosome-

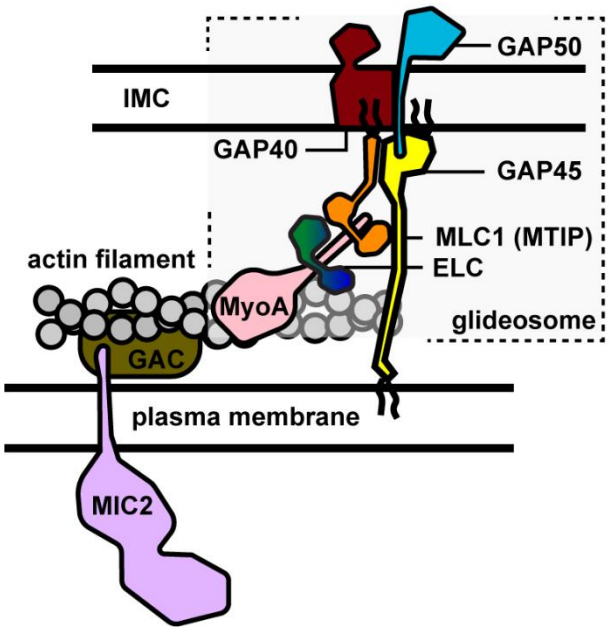


Fig. 1. Scheme of the glideosome. Schematic representation of the current model of the glideosome and its localization in the *T. gondii* intermembrane space. Actin polymerization occurs between the plasma membrane (PM) and the inner membrane complex (IMC) whereas myosin A is part of the glideosome, which binds the essential light chains ELC and myosin light chain MLC1 (called myosin tail interacting protein, MTIP, in *Plasmodium spp.*). Myosin A and its light chains further interact with glideosome associated proteins GAP40, GAP45 and GAP50, which anchor the glideosome in the outer membrane of the inner membrane complex. On the other side, glideosome associated connector (GAC) facilitates the association of actin filaments with surface transmembrane proteins such as MIC2.

associated proteins GAP40, GAP45 and GAP50[7,8,10]. MyoA is an unusually small myosin protein of the unconventional myosin class XIV[11,12], which lacks the typical myosin tail domain and binds the two light chains at the C-terminal myosin neck region[13,14]. MLC1 (in *P. falciparum*: myosin A tail-interacting protein, MTIP) binds at the very C-terminus of MyoA, while ELC is expected to interact with the C-terminus of MyoA upstream of MLC1[15]. Two ELC homologs recognizing the same MyoA region, termed TgELC1 and TgELC2, were identified in *T. gondii*[16], whereas only one PfELC homolog is known in *P. falciparum*[14,17]. Both light chains have been shown to stabilize MyoA *in vivo* and to be essential for parasite egress or invasion[16,18,19]. Myosin A and the light chains interact with the C-terminus of the glideosome associated protein 45 (GAP45) to form a pre-complex in the earlier stages of intracellular parasite development[7], which subsequently assembles with the remaining glideosome members (GAP40 and GAP50). N-terminal palmitoylation modification at its N-terminus anchors MLC1 (MTIP) to the IMC[20], whereas N-terminal myristoylation and palmitoylation sites tie GAP45 to the plasma membrane[21–23]. GAP45 is essential for the correct localization of MyoA with its light chains and GAP45 depletion leads to impairment of host cell invasion[10]. Depletion of GAP40 or GAP50 changes the morphology of the parasites and the integrity of the IMC and thereby also alters the localization of MyoA and the light chains[24].

Structural information on individual members and sub-complexes of the glideosome are limited and the architecture of the entire glideosome is elusive. So far, only structures of *P. falciparum* PfGAP50 soluble domain[25], a *T. gondii* dimeric complex between the TgMyoA C-terminus and MLC1[15], a homologous dimeric complex in *P. falciparum* between PfMyoA C-terminus and MTIP[26], and the motor domains of the *T. gondii* TgMyoA[27] and *P. falciparum* PfMyoA[28] are available (Supplementary Table 1).

Here, we present crystal structures of *T. gondii* and *P. falciparum* light chains bound to the respective MyoA C-termini in the presence of calcium, an additional calcium-free structure as well as the X-ray and NMR solution structures of the N-terminal domain of *P. falciparum* PfELC. We provide a thorough characterization of all identified interaction surfaces and discuss the differences between both species. We demonstrate that ELCs bind to a conserved

binding site on MyoA to induce its α-helical secondary structure and stiffen the MyoA neck. Our work deepens the mechanistic understanding of the gliding motility in *Apicomplexa*.

Results

Structures of isolated ELCs

Crystal structures of *T. gondii* and *P. falciparum* MyoA and of their distal light chains MLC1 (MTIP)[15,26] have already been determined. To shed light on the role of proximal essential light chains (ELCs), we studied their structure in isolation and in the context of their interaction partners. TgELC1 and TgELC2 share a high degree of sequence similarity (65.2%), whereas PfELC has only 40.6% similarity to TgELC1 (Supplementary Fig. 1a), pointing towards structural differences. Likewise, the disorder probability differs between *T. gondii* and *P. falciparum* ELCs (Supplementary Fig. 2a). We recombinantly expressed N-terminally His-tagged ELCs in *E. coli* (Supplementary Fig. 1b) and purified them to homogeneity. In spite of similar molecular weights, PfELC elutes earlier than TgELC2 when subjected to size exclusion chromatography (Supplementary Fig. 2b), indicative of a larger hydrodynamic radius for PfELC. Small angle X-ray scattering (SAXS) measurements further confirm that PfELC has a larger overall size in solution compared to TgELC2, with respective radii of gyration (R_g) of 2.71 ± 0.05 nm and 2.14 ± 0.05 nm (Supplementary Fig. 2d-e, Table 1 and Supplementary Table 2,3). The SAXS data also provide evidence that the increased R_g of PfELC likely results from conformational flexibility (Supplementary Fig. 2f, Supplementary Table 3). This is also apparent from circular dichroism data which show that PfELC has lower α-helical and higher random coil content compared to TgELC2 (Supplementary Fig. 2c, Supplementary Table 2). To map the structured elements and disordered regions of PfELC, we performed triple-resonance NMR experiments that facilitated the near complete assignment of the amide backbone resonances (Supplementary Fig. 2j). Heteronuclear NOEs ({^{1}H}-^{15}N NOE) and chemical shift analysis revealed that the protein consists of an α-helical N-terminal domain, while the C-terminal part is disordered (Supplementary Fig. 2g). Based on this finding, we were able to determine the structure of the N-terminal PfELC fragment (amino acids 1-74, PfELC-N; see Supplementary Fig. 1b) by both X-ray crystallography to 1.5 Å resolution (Fig. 2a, Table 2) and by NMR spectroscopy (Fig. 2b, Table 3). The lowest energy NMR

conformers are very similar to the crystal structure, with an average backbone RMSD of 1.4 Å over residues 1-68. The N-terminal domain of PfELC has a typical calmodulin fold with two EF-hands formed by two helix-loop-helix motifs. Both EF-hands lack the canonical residues that usually bind calcium in calmodulins[29] and in agreement with that, we did not observe any electron density corresponding to a bound ion. PfELC-N crystallized as a dimer covalently linked *via* disulfide bridge, but both NMR and non-reducing SDS-PAGE indicate that the protein exists as a monomer in solution (Supplementary Fig. 1i) and the scattering data calculated from a protein monomer structure fit the measured X-ray scattering profile with X^2=1.37 (Supplementary Fig. 2h). A comparison between the crystal and the NMR structure highlights that the loop of the first EF hand (residues 16-22) and the third helix (residues 40-47) display the highest degree of flexibility, in agreement with the heteronuclear NOE experiment (Supplementary Fig. 2g,j). In general, the assigned backbone resonances in the NMR spectra superimpose for both full-length protein PfELC and the N-terminal domain, proving that the N-terminal domain maintains the same structure as in the full-length context (Supplementary Fig. 2j). These results show that isolated PfELC is monomeric in solution and adopts a calmodulin-like N-terminal fold and differs from TgELCs with a disordered C-terminal region.

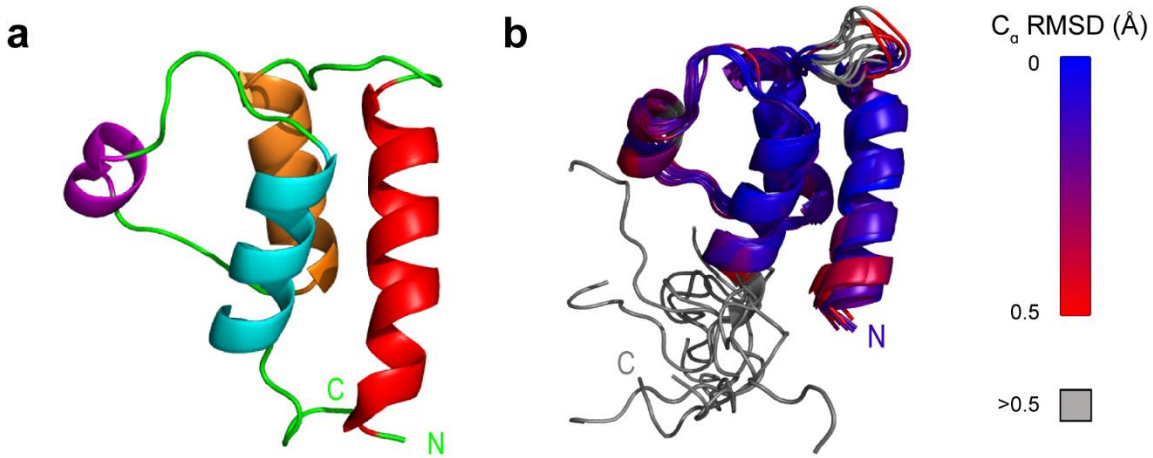


Fig. 2. Crystal structure and NMR structures of PfELC N-terminal domain. (A) Crystal structure of the N-terminal domain of PfELC, residues 1-68. PfELC displays a typical calmodulin fold with two helix-loop-helix motifs. The degenerated EF hand loops do not bind any ion. In agreement with the NMR data of full length PfELC, the protein consists of four α-helices (from N terminus red, orange, violet, cyan, loops and disordered regions in green). (B) Ten lowest-energy NMR structures of PfELC (residues 1-74, all atom RMSD of 1.23 Å) colored from lowest (blue) to highest (red) backbone RMSD compared to the crystal structure show that the loop of the PfELC first EF hand (residues 16-22) and the third helix (residues 40-47) display a certain degree of flexibility.

Essential light chains bind conserved sequence of MyoA

Based on the available literature, *T. gondii* TgELCs and *P. falciparum* PfELC bind to different sites of the MyoA C-terminus[13,14,17]. For PfELC, two binding sites at the PfMyoA C-terminus (PfMyoA residues 786-803 and 801-818) were identified[14], while only one distinct binding site was experimentally confirmed for TgELCs (TgMyoA 775-795; see Fig. 3a and Supplementary Fig. 1c)[15,16]. To resolve this discrepancy, we measured the binding affinity of TgELC1, TgELC2 and PfELC to peptides that correspond to the proposed MyoA binding sites.

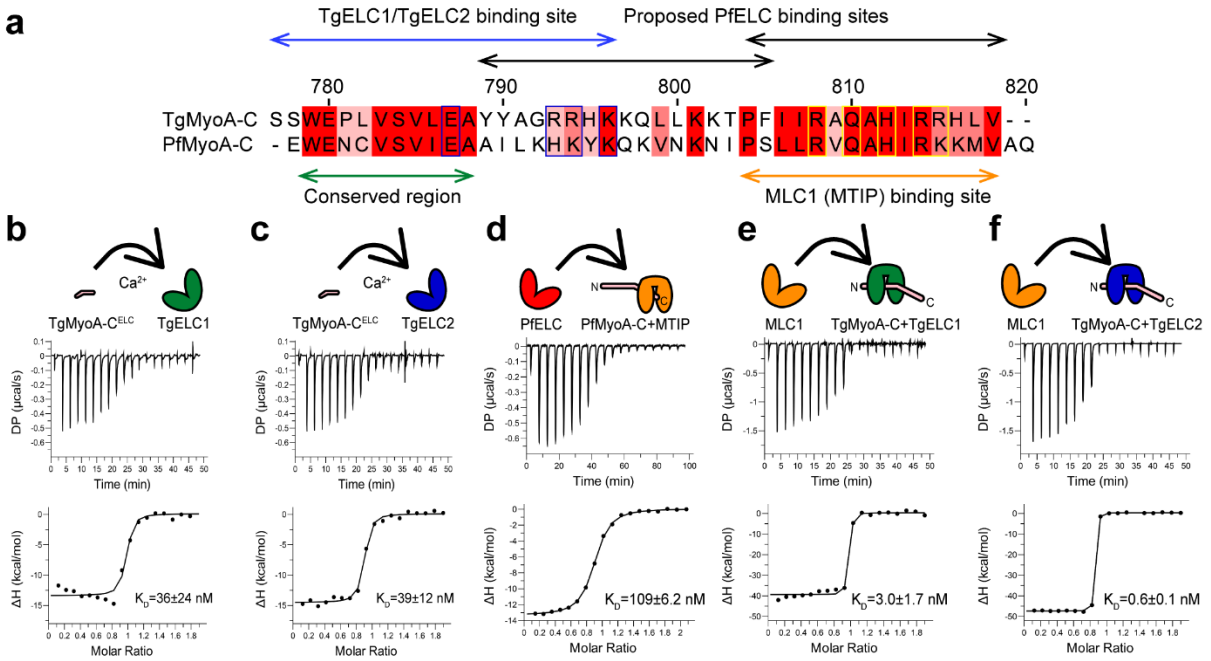


Fig. 3. Assembly of glideosome sub-complexes in *T. gondii* and *P. falciparum*. (A) Sequence comparison of TgMyoA and PfMyoA C-termini shows a conserved region (green arrow) upstream of the MLC1 (MTIP) binding site. Whereas two binding sites of PfELC at the very C-terminus of PfMyoA were proposed (black arrows)[14], our data show that the actual binding site of PfELC encompasses the MyoA conserved region and is similar to the TgELC/TgMyoA binding site (blue arrows). The blue boxed residues indicate residues involved in polar interactions with TgELC1 and TgELC2, while yellow boxed residues form polar interactions with MLC1 (see Fig. 3c-d and Supplementary Table 4). (B,C) Isothermal titration of TgMyoA-C^{ELC} with TgELC1 and TgELC2 show that both dimeric complexes form with nanomolar affinity. The upper panel shows the signal recorded directly after each injection of TgELC1 and TgELC2 and represents the thermal power that has to be applied to maintain a constant temperature in the sample cell during recurring injections. In the lower panel, the integrated heats are plotted against the peptide/protein concentration ratio. The thermodynamic binding parameters were obtained by nonlinear regression of the experimental data using a one-site binding model. (D) Binding isotherm of PfELC titrated to the preformed MTIP/PfMyoA-C complex proves that the conserved hydrophobic region of MyoA is indispensable for ELC binding. (E,F) Binding isotherms of MLC1 titrated into the pre-complex of TgMyoA-C with TgELC1 and TgELC2. MLC1 binds the pre-complex with high nanomolar affinity. All thermodynamic parameters derived from ITC measurements are summarized in Table 1

Both TgELC1 and TgELC2 bound TgMyoA-C^{ELC} (residues 777-799, see Supplementary Fig. 1c) with high affinity (36 ± 24 nM and 39 ± 12 nM, respectively), in agreement with the

Table 1. Overview of thermodynamic constants measured by ITC.

Dimeric interactions

Protein (cell)	MyoA peptide (syringe)	Molar ratio	K_d (nM)	ΔH (kcal/mol)	-TΔS (kcal/mol)
MTIP	PfMyoA-C^{ELC}	0.74 ± 0.01	303 ± 43	-14.4 ± 0.4	5.5
TgELC1	TgMyoA-C^{ELC}	1.05 ± 0.01	36 ± 24	-13.0 ± 0.2	3.2
TgELC1 (EDTA)	TgMyoA-C^{ELC}	0.81 ± 0.01	57 ± 18	-13.0 ± 0.6	3.4
TgELC2	TgMyoA-C^{ELC}	0.85 ± 0.01	39 ± 12	-15.0 ± 0.3	4.5
TgELC2 (EDTA)	TgMyoA-C^{ELC}	0.77 ± 0.01	82 ± 7	-18.0 ± 0.1	8.2
TgELC2^{E10A}	TgMyoA-C^{ELC}	0.79 ± 0.01	190 ± 25	-17.0 ± 0.3	8.2
TgELC2^{F79A}	TgMyoA-C^{ELC}	0.84 ± 0.01	280 ± 34	-18.0 ± 0.3	9.5
TgELC2^{S101A}	TgMyoA-C^{ELC}	0.88 ± 0.02	280 ± 85	-18.0 ± 0.8	9.3
TgELC2^{S102A}	TgMyoA-C^{ELC}	0.79 ± 0.01	76 ± 26	-16.0 ± 0.5	6.6
TgELC2^{S102E}	TgMyoA-C^{ELC}	0.77 ± 0.01	140 ± 26	-18.0 ± 0.3	8.9
TgELC2$^{E10A+H110A}$	TgMyoA-C^{ELC}	0.75 ± 0.02	1100 ± 220	-21.0 ± 0.9	12.0

Trimeric interactions

Pre-complex with MyoA-C (cell)	Protein (syringe)	Molar ratio	K_d (nM)	ΔH (kcal/mol)	-TΔS (kcal/mol)
MTIP	PfELC	0.86 ± 0.01	109 ± 6.2	-13.4 ± 0.1	4
MTIP	PfELC^{S127D}	0.81 ± 0.01	260 ± 26	-12.6 ± 0.2	4
TgELC1	MLC1	0.92 ± 0.01	4.7 ± 2.5	-39.1 ± 0.8	28
TgELC2	MLC1	0.81 ± 0.01	0.6 ± 0.1	-47.6 ± 0.1	35
TgELC2^{R17A}	MLC1	0.92 ± 0.01	4.6 ± 0.4	-49.7 ± 0.2	38
TgELC2^{E22A}	MLC1	0.92 ± 0.01	5.2 ± 1.9	-45.9 ± 0.7	35
TgELC2	MLC1^{K168A}	0.79 ± 0.01	1.2 ± 0.8	-47.7 ± 0.2	36
TgELC2	MLC1^{Q169A}	0.89 ± 0.01	2.3 ± 1.9	-48.8 ± 0.5	37
TgELC2	MLC1^{N172A}	0.84 ± 0.01	4.3 ± 4.3	-41.6 ± 0.9	30

The thermodynamic parameters were fitted by a one site binding model with the MicroCal PEAQ-ITC Analysis Software.

previously published data (Fig. 3b-d, Table 1). Strikingly, we could not monitor any binding of PfELC to the previously described binding sites but observed precipitation upon mixing PfELC with the respective peptides. Therefore, we hypothesized that the ELC binding sites are conserved between *T. gondii* and *P. falciparum* (see conserved MyoA region in Supplementary Fig. 2a) and extended the PfMyoA peptide based on homology with the binding site of TgMyoA. However, precipitation occurred again and we speculated that in *P. falciparum*, the presence of MTIP bound to PfMyoA is a prerequisite for PfELC binding. Thus, we first formed a dimeric complex between MTIP and the PfMyoA neck region peptide (PfMyoA-C, residues 775-816; Fig. 3a and Supplementary Fig. 3c) and then titrated this

pre-complex to PfELC. This time, PfELC bound to the dimeric pre-complex with an affinity of 109 ± 6.2 nM (Fig. 3d and Table 1). These results indicate a particular order in which the *P. falciparum* light chains bind to PfMyoA: MTIP has to interact first and only then PfELC can bind. This is in agreement with previous reports, highlighting that PfELC co-expressed with full-length PfMyoA in insect cells can only be co-purified when MTIP is co-expressed as well[17]. On the other hand, *T. gondii* light chains showed an inverse behavior. We observed no precipitation upon binding of TgMyoA peptides to TgELCs and we were able to further titrate in MLC1 to form the trimeric complexes with high affinity (4.7 ± 2.5 nM and 0.6 ± 0.1 nM, respectively) (Fig. 3e-f and Table 1). However, the addition of MLC1 to TgMyoA peptides caused precipitation. It remains to be investigated whether the different order of binding events required for the formation of the trimeric complexes in *T. gondii and P. falciparum in vitro* play any role *in vivo*. We have demonstrated that the MyoA binding sites are conserved and topologically identical trimeric complexes form in both apicomplexan species.

TgELCs form similar complexes with TgMyoA and MLC1

The successful formation of trimeric assemblies of MyoA with its light chain proteins allowed us to crystallize and determine the structures of the following complexes: (i) *T. gondii* MLC1/TgMyoA-C/TgELC1 complex at 2.4 Å resolution (hereafter named complex 1) and (ii) *T. gondii* MLC1/TgMyoA-C/TgELC2 complex at 2.3 Å resolution (hereafter named complex 2) (Fig. 4a-b, d-e, Table 2). Both complexes constitute a similar architecture. TgMyoA folds into an extended α helix with a characteristic kink between residues 801-803 (angle of 139° in complex 1 and 137° in complex 2). Both TgELCs display a typical calmodulin fold with one N-terminal and one C-terminal lobe, with each lobe comprising two EF hands. Clear additional electron density was visible only in the first EF hand of each complex and assigned to a calcium ion coordinated in a tetragonal bipyramidal geometry. Both TgELCs form conserved polar interactions with TgMyoA, involving TgMyoA residues E787, R793, R794 and K796, a π-π stacking interaction between the conserved residue pair W779-F79 and a group of

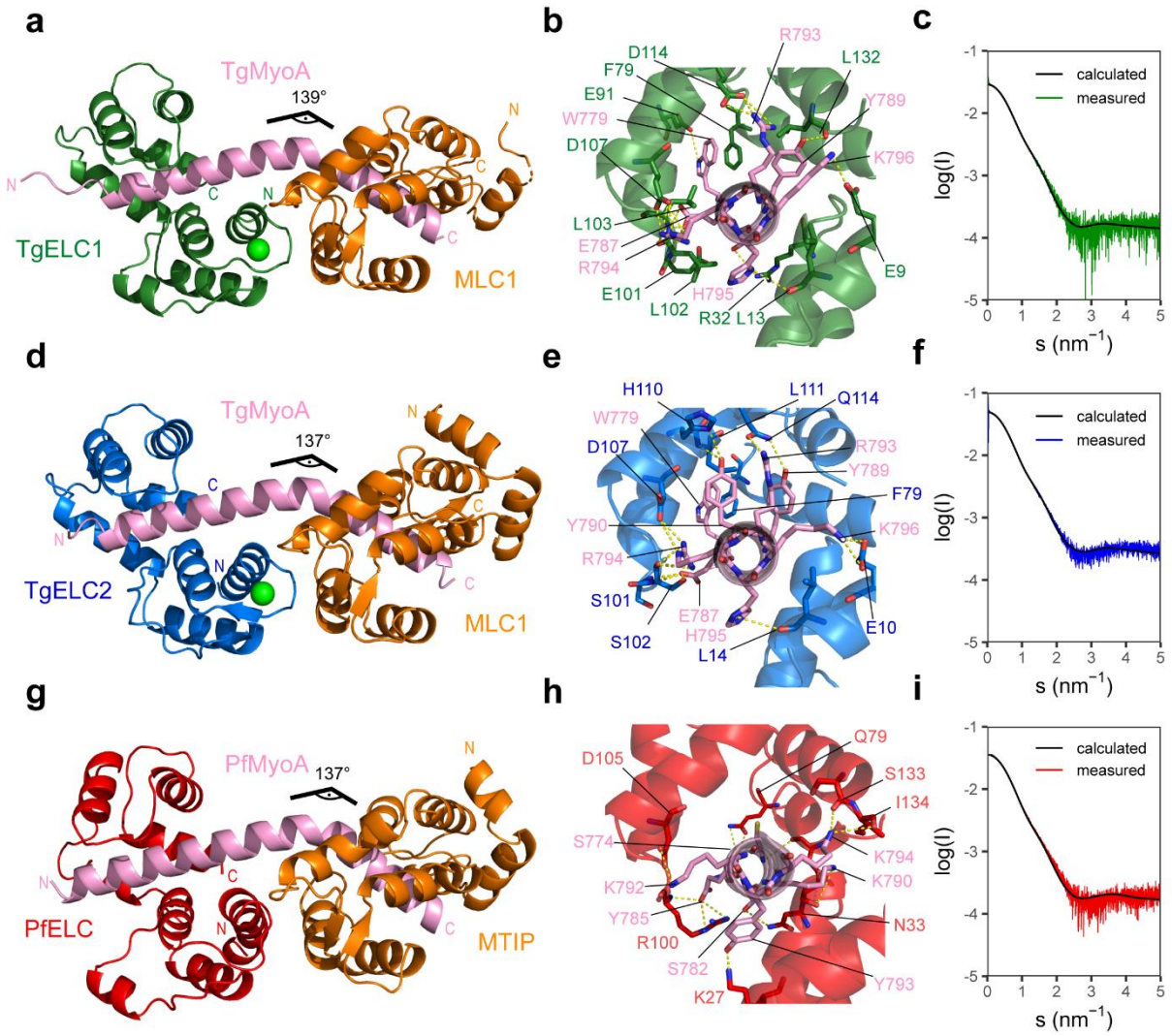


Fig. 4. X-ray structures of trimeric glideosome sub-complexes. (A,D,G) Crystal structures of trimeric complex of TgELC1 (green) or TgELC2 (blue) or PfELC (red) with MyoA-C (pink) and MLC1/MTIP (orange). The complexes are topologically similar and the MyoA helix displays a characteristic kink between residues 801-803. ELCs bind upstream of the MLC1/MTIP binding site. (B,E,H) Binding interface between MyoA-C (pink) and TgELC1 (green) in complex 1, TgELC2 (blue) in complex 2 or PfELC (red) in *P. falciparum* complex. Residues involved in polar interactions are labelled with the corresponding colour and shown in stick representation. Most polar interactions are mediated by the C-terminal lobes of ELCs and the hydrophobic interactions between ELCs and the conserved hydrophobic MyoA residues play a crucial role in complex formation as evident from ITC measurements. (C,F,I) SAXS analysis of the trimeric complexes. Calculated scattering curves of complex 1, complex 2 and *P. falciparum* complex fit the respective experimental data with χ^2 equal to 1.26, 2.41 and 3.9, respectively.

hydrophobic residues clustered around the conserved TgMyoA region P801-Y810 (Fig. 4b,e, Supplementary Table 4). Mutational analysis on TgELC2 (Table 1, Supplementary Fig. 3a) showed that disrupting one of the polar interactions of the conserved π-π stacking interaction W779-F79 has only a moderate effect on the binding affinity of TgMyoA to TgELC2 and suggests that the hydrophobic residues in the conserved MyoA region play a crucial role for

complex formation. In agreement, the phosphomimetic mutation of residue S102, previously shown to be phosphorylated[30], only had a moderate effect on the affinity of TgELC2 to the MyoA peptide, indicating that a single phosphorylation event is likely not sufficient to regulate complex formation (Table 1 and Supplementary Fig. 3b). Complexes 1 and 2 are monomeric in solution, but while the calculated scattering data of complex 1 fit the experimental scattering data with a X^2 = 1.26, the structure of complex 2 displays a higher X^2 = 2.41, indicative of small structural differences in solution (Fig. 4c,f, Supplementary Table 3). Taken together, *T. gondii* TgELCs form tight complexes with MyoA and MLC1 and the corresponding binding interfaces are dominated by hydrophilic and hydrophobic interactions.

Table 2. X-ray data collection and refinement statistics.

	PfELC-N	Complex 1	Complex 1f	Complex 2	*P. falciparum* complex[1]
Data collection					
Space group	P 21 21 21	P 41	P 41	I 21 21 21	P 43
Cell dimensions					
a, b, c (Å)	30.24, 57.51, 86.34	87.32, 87.32, 56.75	86.13, 86.13, 53.7	84.63, 93.48, 108.15	211.88 211.88 75.46
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.86 - 1.50 (1.55 - 1.50)	47.58 - 2.39 (2.48 - 2.39)	40.94 - 2.00 (2.07 - 2.00)	40.96 - 2.30 (2.38 - 2.30)	47.42 – 2.51 (2.58-2.51)
R_{merge}	0.03382 (0.495)	0.106 (1.599)	0.0431 (1.35)	0.08044 (1.007)	0.0874 (3.79)
$I / \sigma I$	17.68 (2.06)	19.06 (1.40)	31.02 (1.74)	13.84 (1.79)	12.69 (0.55)
Completeness (%)	99.0 (98.0)	99.9 (99.7)	99.9 (99.4)	99.9 (99.9)	83.5 (8.0)
Total no. reflections	104329 (9981)	226789 (23610)	373831 (34891)	124597 (12474)	768342 (75189)
Redundancy	4.2 (4.2)	13.3 (13.7)	13.5 (12.9)	6.4 (6.6)	6.7 (6.6)
Refinement					
Resolution (Å)	1.5	2.4	2.0	2.3	2.5
No. reflections	104329	226789	373831	124597	768342
R_{work} / R_{free}	0.167/0.193	0.189/0.231	0.190/0.225	0.186/0.219	0.200/0.238
No. atoms	1319	2523	2610	2687	12968
Protein	1126	2457	2458	2578	12965
Ligands	n.a.	2	5	33	n.a.
Solvent	193	64	147	76	3
B-factors	36.5	78.2	65.4	65.3	99.69
Proteins	34.8	78.3	65.4	65.0	99.69
Ligands	n.a.	96.6	111	94.2	n.a.
Solvent	46.5	72.8	63.2	62.6	70.51
R.m.s. deviations					
Bond lengths (Å)	0.013	0.008	0.003	0.007	0.015
Angles (°)	1.16	0.97	0.60	0.87	2.04

[1] The native data of the P. falciparum complex was subjected to anisotropic scaling and truncation. Without truncation, I/σI of the native data set used for refinement falls below 2.0 between a maximum resolution of 2.75 and 2.70 Å at an overall completeness of over 99%

Table 3. NMR and refinement statistics for PfELC (residues 1-74).

	Protein
NMR distance and dihedral constraints	
Distance constraints	2320
Total NOE	2320
Intra-residue	900
Inter-residue	1420
Sequential ($\|i-j\| = 1$)	324
Medium-range ($\|i-j\| < 4$)	373
Long-range ($\|i-j\| > 5$)	723
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	118
ϕ	59
ψ	59
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.03 ± 0.02
Dihedral angle constraints (º)	1.2 ± 0.5
Max. dihedral angle violation (º)	16.5
Max. distance constraint violation (Å)	1.53
Deviations from idealized geometry	
Bond lengths (Å)	0.0023 ± 0.0004
Bond angles (º)	0.38 ± 0.06
Impropers (º)	0.3 ± 0.1
Average pairwise r.m.s. deviation** (Å)	
Heavy	1.23 ± 0.14
Backbone	0.87 ± 0.11

** "Pairwise r.m.s. deviation was calculated among 10 refined structures."

PfELC binds PfMyoA in a structurally distinct manner

To investigate whether the homologous complexes from *T. gondii* and *P. falciparum* are structurally similar, we determined the crystal structure of the *P. falciparum* trimeric complex (PfMyoA, MTIP, PfELC) at 2.6 Å resolution (Fig. 4g, Table 2). Overall, this structure resembles a similar fold and conformation compared to the *T. gondii* trimeric complexes, with the typical MyoA helix kink of 131° between the MTIP and PfELC binding sites. While the secondary structure elements are maintained, the position of the PfELC helices differ. The N-terminal lobe of PfELC aligns well to TgELCs structures (backbone RMSD of 2.5 Å to TgELC1), but the C-terminal lobe adopts a different orientation with respect to the MyoA helix (backbone RMSD of 3.5 Å to TgELC1), resulting in a reduced number of polar interactions between PfELC and PfMyoA (Supplementary Fig. 3e). This explains the lower binding affinity of trimeric

complex formation in *P. falciparum* (Fig. 4h, Table 1). Of note, the electron density of the PfELC C-terminal lobe is less well defined compared to the remaining structure, which is likely caused by increased flexibility of the C-terminal loop of PfELC. This is also reflected in the comparison of the calculated scattering data from the *P. falciparum* complex structure and the recorded SAXS data with $X^2 = 3.9$ (Fig. 4i, Supplementary Table 3).

Due to the reduced number of interacting residues at the PfELC C-terminus (Supplementary Table 4), it seems plausible that C-terminal phosphorylation could play a regulatory role in binding of PfELC to PfMyoA. To test this hypothesis *in vitro*, we mutated residue S127, that has previously been shown to be phosphorylated *in vivo*[30], to a phosphomimetic aspartate residue and observed that the affinity for this variant to PfMyoA C-terminal peptide dropped twofold (Table 1 and Supplementary Fig. 3d). S127 does not directly interact with PfMyoA, but forms a polar interaction with PfELC residue N75, maintaining the tertiary structure of the C-terminal lobe. Based on available data, it is likely that phosphorylation of S127 has a direct impact on the interaction of PfELC with PfMyoA, however, *in vivo* experiments are necessary to study the impact of this phosphorylation on the glideosome assembly and function.

ELCs induce α-helical structure in MyoA

Previous reports have shown that the presence of *P. falciparum* and *T. gondii* essential light chains increase the speed of the myosin A motor twofold[14,16,17]. To understand the function of ELCs on a molecular level, we characterized TgELC2 in a free and bound state with TgMyoA-C^{ELC} (see Supplementary Fig. 1c). On size exclusion chromatography, the dimeric complex of TgELC2 and TgMyoA-C^{ELC} elutes later than TgELC2 alone, indicating that the hydrodynamic radius of TgELC2 decreases upon binding of TgMyoA-C^{ELC} (Fig. 5a). To quantify the structural changes upon binding, we compared the parameters calculated from the SAXS data of TgELC2 alone and in complex with TgMyoA-C^{ELC} (Fig. 5b-c, Supplementary Fig. 4a-b, Supplementary Table 3). Changes in the dimensionless Kratky plot (Fig. 5b) as well as the drop of the radius of gyration (2.15 nm to 1.73 nm) and maximum particle size (6.7 nm to 5.5 nm, Fig. 5c) highlight that the dynamic TgELC2 protein undergoes compression upon

interaction with the TgMyoA C-terminus. This rigid conformation allows the neck region to act as the lever arm of myosin and its stiffness directly correlates with the myosin step size and speed[31–33]. Although our crystal structures show that TgMyoA-C forms a continuous α helix, we noticed that both TgMyoA-C as well as PfMyoA-C are unfolded or partially unfolded in the absence of binding partners (Supplementary Fig. 4c). Indeed, the C-terminal amino acid residues of the recently published TgMyoA[27] and PfMyoA[28] motor domain structures could not

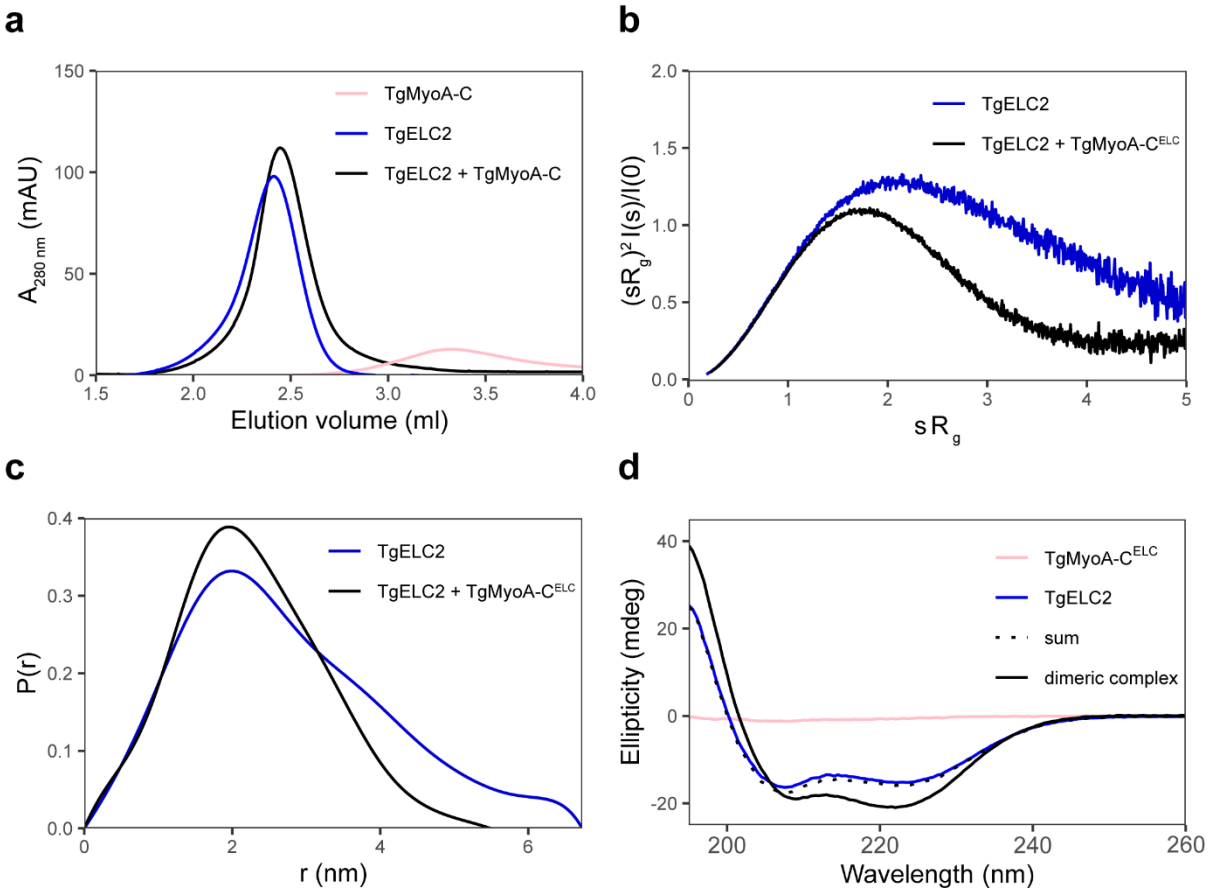


Fig. 5. TgELCs and TgMyoA undergo large conformational changes upon binding. (A) Dimeric complex of TgELC2 and TgMyoA-C elutes at shorter retention times than isolated TgELC2 on Superdex 200 5/150 column, suggesting that the hydrodynamic radius of TgELC2 decreases upon TgMyoA-C binding. (B) Dimensionless Kratky plots of isolated TgELC2 and in complex with TgMyoA-C. The plot of TgELC2 in complex with TgMyoA-C^{ELC} (black) has a maximum close to $sR_g = \sqrt{3}$ and converges to zero, unlike isolated TgELC2 (blue), suggesting that TgELC2 in isolation is rather extended and compacts upon binding to TgMyoA. (C) The distance distribution calculated by Guinier analysis from the SAXS data further confirms that TgELC2 undergoes compaction upon TgMyoA binding. TgELC2 displays wider distance distribution with d_{max} = 6.7 nm, whereas the distance distribution of the dimeric complex is narrower with d_{max} = 5.5 nm. (D) The far-UV CD data indicate that TgELC2 induces a α-helical structure in TgMyoA upon binding. The individual spectra of TgELC2 and TgMyoA-C^{ELC} do not sum up to the CD spectrum of their dimeric complex and the CD spectrum of the dimeric complex displays more pronounced features of α-helical secondary structure with lower ellipticity at 222 nm and higher ellipticity at 195 nm compared to the sum of individual components. CD spectra were recorded in a 1 mm cuvette at a concentration of 5 μM of each component in 10 mM NaP (pH 7.5), 150 mM NaF and 0.25 mM TCEP at 20°C.

be resolved, likely due to their intrinsically disordered nature. We hypothesized that the essential light chains can induce the formation of an α-helical structure in MyoA upon binding. Therefore, we measured far-UV CD spectra of TgMyoA-C^{ELC} and TgELC2 in isolation and in complex (Fig. 5d). The data revealed that TgMyoA-C^{ELC} is predominantly unstructured while TgELC2 has an α-helical fold. However, the CD spectrum of the dimeric complex displays a markedly higher α-helical content than the sum of the spectra of the two individual components, suggesting that the content of the α-helical secondary structure increased upon formation of the complex. We also observed a similar, albeit less pronounced effect for the TgELC1-TgMyoA-C^{ELC} and *P. falciparum* trimeric complex assembly (Supplementary Fig. 4d,e). We anticipate that the increase in α-helical secondary structure content corresponds to the induction of the structure of the TgMyoA C-terminus, which in turn stiffens the TgMyoA lever arm. As a result, the myosins are capable of undergoing a larger step size and thus increase their speed, in agreement with the published functional measurements for both *T. gondii* and *P. falciparum* myosin A motors[14,16,17].

Calcium stabilizes but has no impact on complex assembly

The myosin light chains together with the myosin heavy chain neck region constitute a regulatory domain that influences the biochemical and mechanical properties of myosins either upon phosphorylation[34–37] or by direct binding of calcium[38,39]. Apicomplexan invasion is a tightly regulated process, which involves an increase in intracellular calcium concentration[40]. To investigate the role of calcium bound in the first EF hand of both TgELCs, we determined an additional crystal structure of the calcium-free complex TgELC1/MLC1/MyoA-C at 2.0 Å resolution (complex 1f, Fig. 6a, Table 2). Complex 1f generally adopts the same conformation as complex 1. The MyoA-C helix is kinked at a similar angle (134°), and the binding interfaces between MLC1 and TgMyoA as well as between TgELC1 and MyoA are identical to complex 1 (Supplementary Table 4). The first EF hand loop and the calcium binding residues remain in the same conformation as in complex 1 except for the side chain of aspartate 17 which is flipped by 120° and thereby enables the release of calcium from the binding pocket (Fig. 6b).

In complex 1, calcium is coordinated in a tetragonal bipyramidal geometry by the carboxyl groups of side chains D15, D17, D19, the carbonyl group of E21 and two water molecules. In complex 2, calcium is similarly coordinated by the homologous side chain residues of D16, N18, D20, the carbonyl group of E22 and two water molecules. Additionally, in complex 2, these water molecules are further stabilized by interactions with the side chains of E27 and Q49. Contrary to *T. gondii* TgELCs, the homologous EF hand loop of PfELC (in isolation or in complex) is bent to the other side and does not possess the residues needed for coordination of calcium (Fig. 6b). In agreement with the presented crystal structures, calcium has no major influence on the secondary structure of individual TgELCs or PfELC (Supplementary Fig. 5a). Powell *et al.* recently showed that the absence of calcium notably reduces the affinity of TgELC1 for the MyoA C-terminus[15]. To investigate this effect in both *T. gondii* essential light chains, we measured the affinity of TgELC1 and TgELC2 to the TgMyoA peptide with wild type proteins either in the presence of 5 mM calcium or 5 mM EDTA. Strikingly, the difference in affinity is only minor in both cases, with an observed twofold decrease in affinity in the presence of 5 mM EDTA compared to 5 mM calcium (Supplementary Fig. 5b). This is rather surprising, considering the fact that the regulatory role of calcium has been proposed for other myosin light chains[38,41]. Our binding data are supported by the available crystal structures, where a clear role for calcium regulation is not directly evident. While the presence of calcium affects the affinity of ELCs only to a minor extend, we observed a pronounced effect of calcium ions on the thermal stability of the trimeric complex in a concentration dependent manner (Fig. 6c-d, Supplementary Fig. 5c). This reveals that calcium ions bind TgELCs and mediate substantial stabilization of their sub-complexes, although they do not markedly change their structure or affinity. This is in agreement with previously published functional data, reporting that the absence of calcium does not alter the function of the myosin A motor in both *P. falciparum*[17] and *T. gondii*[13]. It is likely that the presence of calcium could have a rather indirect effect, for example by modulating the activity of kinases which in return change the

phosphorylation status of members of the glideosome[42–44]. In conclusion, calcium binding by the first EF hand of TgELCs does not structurally impact the formation of the complex but increases the stability of the complexes *per se*.

Light chain interactions do not trigger structural changes

Based on our structural work, we have shown that the formation of the TgMyoA-TgELCs dimeric complexes leads to large structural changes and folding of the MyoA C-terminus. In the trimeric complexes, interactions between the light chains have been proposed to mediate the transmission of regulatory signals from distal (MLC) to proximal light chain (ELC) light chains[34]. To assess the structural changes that could result from the interaction between the two light chains, we recorded SAXS data of the TgELC2-TgMyoA-C^{ELC} dimeric complex and compared them to the scattering profile calculated from complex 2 without MLC1

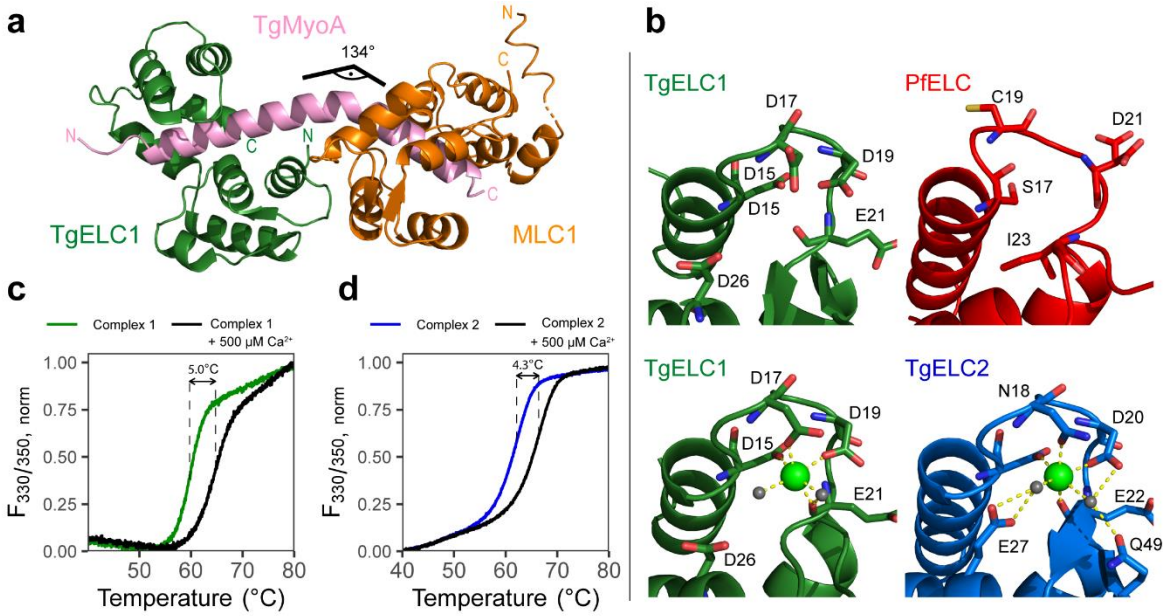


Fig. 6. Role of calcium in TgELCs. (A) Crystal structure of the glideosome trimeric complex composed of TgELC1 (green), MLC1 (orange) and TgMyoA-C (pink) in the absence of calcium (complex 1f). The absence of calcium does not cause a major structural rearrangement (see Fig. 5A). (B) Structural comparison of the first EF hand in ELCs and calcium coordination between complex 1f, complex 1, complex 2 and PfELC-N. Whereas PfELC does not bind any ion due to a degenerated sequence in its EF hand, both TgELCs in complex 1 and complex 2 bind calcium in a tetragonal bipyramid coordination, including two water molecules. These water molecules are further stabilized in complex 2 by additional residues (E27, Q49, D20). In complex 1f, the side chain of residue D17 is flipped by 120°, enabling the release of calcium. (C-D) Thermal stability change of trimeric complex 1 and complex 2 upon addition of calcium measured by nanoDSF. The stability of both complexes strongly increases upon calcium binding.

(Supplementary Fig. 6a). Based on a resulting X^2 of 1.16 Å, it is unlikely that TgELC2 undergoes structural changes upon trimeric complex formation. Similarly, MLC1 and MTIP adopt the same conformation as in already described structures of their dimeric complexes with MyoA (PDB IDs 5vt9 and 4aom, respectively) and the key interactions remain unperturbed in the presence of ELCs (Supplementary Fig. 6e-g, Supplementary Table 4). Thus, light chains do not exhibit any major structural rearrangements upon trimeric complex formation, although the crystal structures revealed a small interaction surface between both

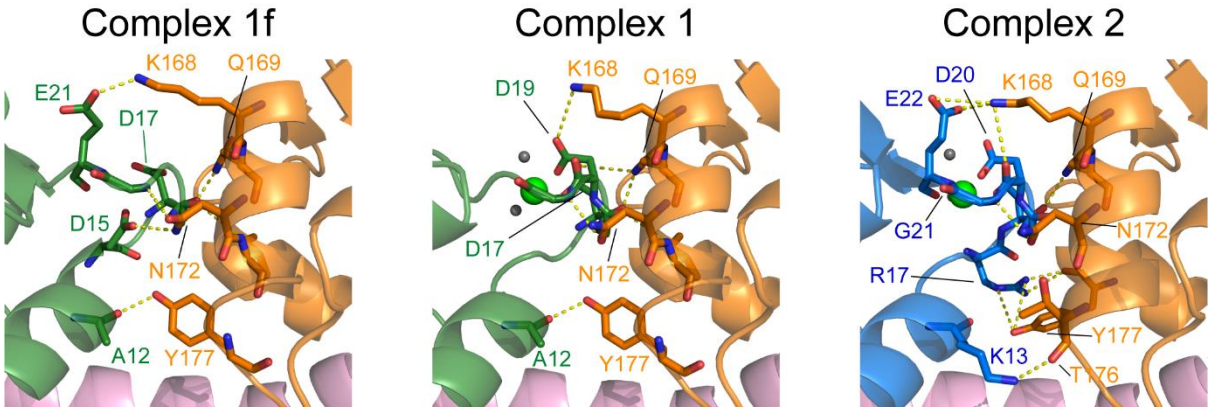


Fig. 7. Light chain interactions upon trimeric complex formation. Binding interfaces between TgELCs and MLC1 in the trimeric complex structures of (from left) complex 1f, complex 1 and complex 2. Corresponding residues are labelled with the respective colour. The same set of residues (K168, Q169, N172, Y177) is involved in polar interactions (indicated by yellow dashes) on the MLC1 site, but various residues are utilized by TgELCs.

light chains formed by several polar interactions near the ELC calcium binding site. These interactions were previously proposed to be only present when calcium is bound[16]. However, the calcium-free crystal structure shows that these interactions are rather independent of the presence of calcium and conserved between complex 1 and complex 2 (Fig. 7). We additionally performed mutational analysis of the interacting residues at the interface of MLC1 and TgELC2. We observed only a minor decrease in affinity upon mutation, but the measured affinities reached the limitations of reliable high affinity ITC measurements (Table 1 and Supplementary Fig. 6d). This leaves open the possibility of cross-talk between the two light chains, however, we do not expect these to have a large impact on the overall structure and myosin motor function because the effect of the mutations at the light chain interface is only minor.

To complete our analysis, we examined whether the formation of the trimeric complexes impacts the structure of the MLC1 N-terminus. The disordered N-termini of MLC1 and MTIP are of particular interest because they are expected to anchor myosin A to the IMC *via* interaction with GAP45[22]. Our SAXS data reveal that the trimeric complex containing full-length MLC1 displays a notably larger maximum particle size (D_{max}=14 nm) and radius of gyration (3.50±0.02 nm) in comparison to the complex used for crystallization (Fig. 4d, Supplementary Fig. 6b and Supplementary Table 3), indicating that the MLC1 N-terminus remains disordered even in the trimeric complex with TgMyoA and TgELC1.

Next, we explored the stretch of MLC1 residues 66-77, which are on the border of the disordered N-terminus and the structured domains. The electron density in complex 2 reveals an additional α helix for this area, which is absent in complex 1, suggesting that residues 66-77 are disordered. We hypothesized that in solution, this helix is in equilibrium with a disordered state. To investigate this possibility, we compared the distance distributions calculated from SAXS data measured on complex 1 using MLC1 constructs spanning residues 66-210 or 77-210. In case residues 66-76 form exclusively an α helix in solution, we expect them to fold back towards the center of the molecule and the maximum particle distance D_{max} should stay identical. However, D_{max} in the trimeric complex with $MLC1^{77-210}$ (8.2 nm) is markedly lower compared to the construct containing residues 66-76 (9.5 nm with $MLC1^{66-210}$, see Supplementary Fig. 6b). Moreover, SAXS data of the trimeric complex with $MLC1^{66-210}$ agree less with the corresponding crystal structure than with the shorter MLC^{77-210} construct (X^2 equals 1.26 vs 1.04, Supplementary Fig. 6c). The flexibility within residues 66-77 is additionally apparent from the normal mode analysis (Supplementary Fig. 7a, see below). We assume that residues 66-77 of MLC1 exist in equilibrium between α-helical and disordered conformation in solution and believe that this feature may have further implications on the function of the protein, namely anchoring MyoA to the membranes of the IMC or interacting with other members of glideosome, such as GAP45. Knowing that the stiffening of the MyoA lever arm by ELCs increases the motor activity[14,16,17], we find it unlikely that the MLC1/MTIP N-termini are disordered when assembled within the glideosome. We propose that, similarly

to the MLC1 helix 67-77, the secondary structure can be induced in the entire MLC1/MTIP N-terminus upon binding to presumably GAP45, as described here for the ELC-MyoA interaction.

TgMyoA complexes follow the dynamics of traditional myosins

Previously reported structures of myosins in complex with their light chains suggest that the converter domains interact with the essential light chain to further stabilize the rigid lever arm and possibly transmit the structural changes from the myosin motor domain to the lever arm[45,46]. Similarly, it has been proposed that TgELC1 might constitute a small binding interface with the TgMyoA converter domain[15]. To investigate whether the crystal structures of *T. gondii* complexes are compatible with these observations and to ensure that they do not clash with the TgMyoA core, we built structural models of the TgMyoA motor and neck domain bound to MLC1 and TgELC1 or TgELC2 (Fig. 8).

In both cases, the energy-minimized models did not contain any clashes, indicating that our structures are compatible within the full-length context of TgMyoA (Fig. 8a-b). TgMyoA residues 762-818, which constitute the lever arm, maintained a continuous α helix after energy minimization, with both TgELC1 and TgELC2 forming a small number of contacts with the TgMyoA converter domain. These contacts mainly involve the side chain of arginine 81 of TgELC1 or TgELC2 and residues 720-724 of TgMyoA, which is in agreement with the previously published HDX data[15]. To further explore the dynamics of full-length TgMyoA with its light chains, we performed normal mode analysis in an all-atom representation on five energy-minimized models from complex 1 and complex 2, and subsequent deformation analysis which allowed us to identify potential hinge regions within these structures. In both cases, all five reconstructed models displayed nearly identical pattern of motions (see Supplementary Fig. 7a for complex 2): the structures undergo bending in the hinge region of TgMyoA residues 773-777 in two perpendicular directions (mode 7 and 8) as well as twisting in the same region (mode 9). In the remaining modes (modes 10 and higher), the movement

further propagates throughout the lever arm helix up to TgMyoA residue 799. As a result, the deformation analysis of the 20 lowest energy modes predicts the hinge region of the TgMyoA lever arm between TgELCs and the converter domain, and an additional hinge between TgELCs and MLC1 (complex 2 in Fig. 8c and complex 1 in Supplementary Fig. 7b). Such dynamics of myosin light chains is similar to what has been previously described in conventional myosins[46,47] and the flexibility in the first TgMyoA hinge could contribute to the efficient rebinding of the myosin motor domain to actin in the pre-power stroke state (Supplementary Fig. 7c)[41]. In conclusion, the structures of the trimeric complexes composed of the TgMyoA light chains and TgMyoA C-terminus are compatible with full-length TgMyoA and exhibit dynamics that are similar to the dynamics of conventional myosins.

Finally, ELCs generally interact with the myosin converter domain and likely stabilize the hinge region of the myosin neck between the ELC and the converter domain (TgMyoA residues 775-777)[41,47]. A small interaction interface between the converter domain and TgELC1 has

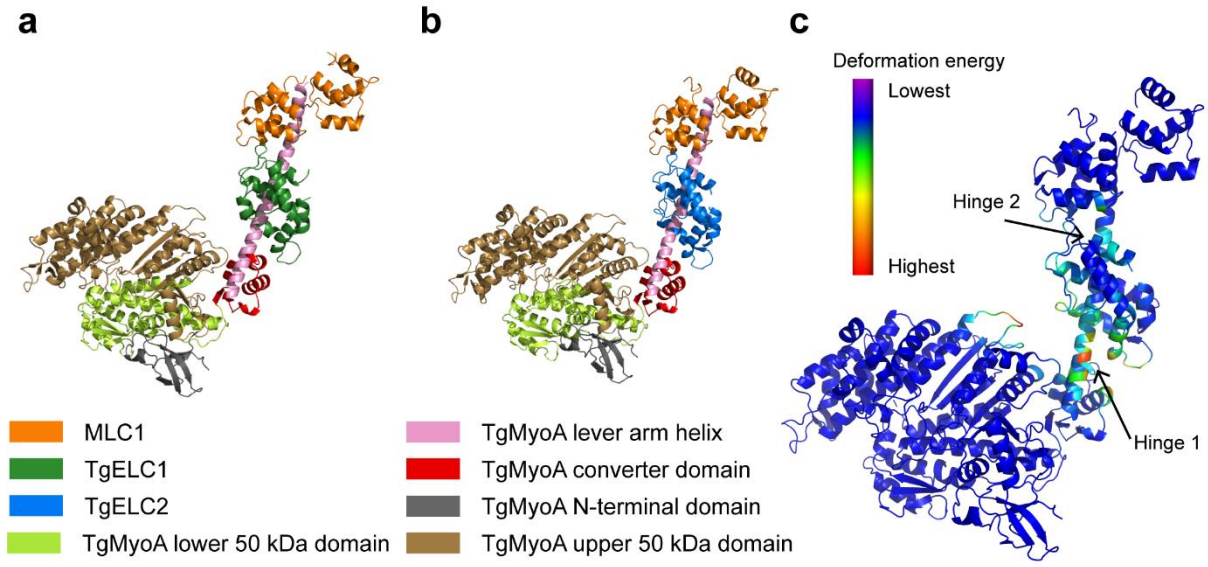


Fig. 8. Trimeric complexes modelled in the full-length MyoA context. (A) Energy-minimized model of complex 1 as a part of TgMyoA. (B) Energy-minimized model of complex 2 as a part of TgMyoA. The models show that the crystal structures of the trimeric complexes are compatible with the structure of TgMyoA and maintain the α-helical structure of the TgMyoA lever arm. No clashes between TgMyoA and TgELCs were observed. (C) Deformation analysis of complex 2 identified two hinge regions in the lever arm of myosin A, which contribute to most of the observed dynamics of the protein complex within the 20 lowest-energy modes. The model is coloured by deformation energy from lowest (violet) to highest (red). The hinges localize to the TgMyoA lever arm between the converter domain and the TgELC2 binding site (hinge 1, residues 773-777) as well as between the TgELC2 and MLC1 binding sites (hinge 2, residues 799-801). These deformations agree with the role of TgMyoA in the pre-power stroke state in the context of a power stroke cycle, where the myosin is probing the conformational space to bind to actin.

also been suggested previously[15]. Our models now highlight that both TgELC1 and TgELC2 form polar interactions with the converter domain, however, these are not sufficient to maintain the rigid structure, and the TgMyoA hinge between ELC and the converter domain contributes to most of the movement of the myosin complex. Nevertheless, the normal mode analysis was performed in the absence of a bound nucleotide or actin and the interface between TgELCs and the converter domain might become more rigid once TgMyoA binds actin, as has been previously described for other myosins[48].

Conclusion

Although, both gliding and invasion of apicomplexan parasites have been intensively studied in the past, the lack of structural data inhibits the broader understanding of these processes on a molecular level. Our work represents a further step towards grasping glideosome function and the mechanism of apicomplexan gliding and invasion. We have determined crystal structures of the glideosome trimeric sub complexes of two main apicomplexan representatives, *P. falciparum* and *T. gondii*. Our structures together with binding data show that ELCs bind a conserved sequence of MyoAs. The C-terminus of PfELC is disordered in isolation compared to TgELCs and also adopts a distinct position when bound to PfMyoA, compared to *T. gondii* complexes. The structures also reveal potential regulatory phosphorylation sites on ELCs and our mutational analysis indicates that phosphorylation events can decrease the ELC binding affinity. We have further investigated the role of ELCs in glideosome assembly as well as the impact of calcium ions that we have observed to be bound in the first EF hands of TgELCs. An additional calcium-free structure of a *T. gondii* trimeric sub complex shows that no major structural changes occur upon calcium binding. Indeed, we observe that calcium ions have no impact on the assembly of the complexes but rather stabilizes the trimeric complexes *per se*. Finally, our biophysical analysis demonstrates that ELCs undergo compression upon binding to MyoA, which induces an α helical structure and thereby stiffens the MyoA lever arms. Our functional observations explain previously published data showing that ELCs can double the speed of a myosin A motor whereas calcium

has no effect. In conclusion, our study complements and rationalizes the role of glideosome components that have been previously observed while providing new structural and functional data that will be important in the future elucidation of glideosome structure and mechanism of apicomplexan gliding.

Methods

Cloning

Open reading frames encoding TgELC2 (*TGME49_305050*) and TgMLC1 (*TGME49_257680*) sub cloned *via* NdeI/XhoI restriction enzymes into pET28a(+)-TEV vector were purchased from *GenScript*. The TgELC1 gene was cloned, by extending the *TGME49_269442* open reading frame (*GenScript*) into a pNIC28_Bsa4[49] vector *via* BsaI restriction sites. DNA sequences of PfELC (*PF3D7_1017500*), PfELC-N (residues 1-74), PfMTIP (*PF3D7_1246400*), PfMTIP-S (residues 60-204) and PfMTIP$^{77-204}$ were amplified from *P. falciparum* 3D7 cDNA and cloned into a pNIC28_Bsa4 vector *via* BsaI restriction sites. These constructs have an N-terminal TEV-cleavable His_6-tag. TgMLC1-S (residues 66-146) was sub cloned into a pNIC_CTHF[49] vector *via* the BfuI restriction site. The vector has a C-terminal TEV-cleavable His_6-tag and FLAG-tag. The sequence encoding TgMyoA-C was amplified by two complementary primers and cloned *via* NcoI/KpnI restriction enzymes into a pET_GB1 vector. This construct contains an N-terminal TEV-cleavable His-GB1 domain. Expression cassettes of His-TgELC1 and His-GB1-TgMyoA-C were then sub cloned *via* NdeI/XbaI restriction enzymes into a pPYC[50] vector. The His-GB1-TgMyoA-C gene was then cut by SpeI/XbaI restriction enzymes and inserted into SpeI-cut pPYC-His_TgELC1 to construct the co-expression vector pPYC with TgELC1 and TgMyoA-C.

Mutagenesis

Site directed mutants were generated by blunt-end PCR. Briefly, the plasmids were amplified by primers which contain the alternative bases on their 5' ends and anneal upstream and downstream of the target triplet. The PCR products were digested by DpnI (NEB) overnight at 37°C and purified by a PCR purification kit (Qiagen). Subsequently, the 5' ends of the PCR products were phosphorylated by T4 polynucleotide kinase (NEB), the products were purified and the free ends of the plasmid re-ligated by T4 DNA ligase (NEB). The positive clones were subsequently selected and their sequence was verified by sequencing.

Protein expression and purification

The proteins were overexpressed in *E. coli* BL21(DE3) (MLC1, MTIP, MTIP-S, co-expressed TgELC1-TgMyoA-C + MLC1-S) or *E. coli* BL21-CodonPlus(DE3)-RIL (TgELC1, TgELC2, PfELC, PfELC-N, MLC1-S), in TB medium. The bacterial cultures were induced at OD_{600nm} of 0.6 with 1 mM IPTG and harvested after 4 hours at 37°C (TgELC1, TgELC2, PfMTIP) or induced at OD_{600nm} of 0.6 by 0.2 mM IPTG and harvested after 16 hours at 18°C (PfELC, PfELC-N, MLC1). The expression of PfELC and PfELC-N for NMR measurements was performed in minimal expression medium as described elsewhere[51].

The cell pellets were resuspended in lysis buffer (20 mM NaP (pH 7.5), 300 mM NaCl, 5% glycerol, 15 mM imidazole, 5 units/ml DNase I, 1 tablet of protease inhibitors (*Roche*) per 100 mL buffer, 1 mg/mL lysozyme, 0.5 mM TCEP) and the bacteria were lysed by three passages through an emulsifier (EmulsiFlex-C3, Avestin) with a maximum pressure of 10 000 psi. The lysate was centrifuged (20 min, 19 000*g*) and incubated with 2 ml of Ni-IMAC beads (ThermoFisher) per 1 l of culture on a rotatory wheel (1 h, 4 RPM). The lysate was then transferred into a gravity column and washed twice with 10 ml wash buffer (20 mM NaP (pH 7.5), 300 mM NaCl, 5% glycerol, 15 mM imidazole, 0.5 mM TCEP). The bound protein was eluted with 10 ml and subsequently with 5 ml of elution buffer (20 mM NaP (pH 7.5), 150 mM NaCl, 5% glycerol, 250 mM imidazole, 0.5 mM TCEP). The elution fractions were pooled and 0.5 mg of TEV protease per liter of bacterial culture was added. The samples were dialyzed (2 kDa cut-off) against 500 ml wash buffer or, in case of PfELC and PfELC-N, against 50 mM Tris (pH 8.0), 20 mM NaCl, 0.5 mM TCEP overnight. Next day, the samples were incubated on a gravity column with 1 ml Ni-beads per 1 l of culture. The flow-through was concentrated (10 kDa cut-off) to maximum of 10 mg/ml and further purified by size exclusion chromatography on a Superdex 200 HiLoad column (GE Healthcare; PfELC, MTIP, MTIP-S, MLC1, MLC1-S) or on a Superdex 75 HiLoad column (GE Healthcare; TgELC1, TgELC2, PfELC-N, co-expressed TgELC1-TgMyoA-C), using gel filtration buffer (20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5 mM TCEP). Finally, the samples were concentrated

(10 kDa cut-off) up to 15 mg/ml and either directly used or flash-frozen for later use. Due to instability, PfELC was always directly used within 3 days of the purification without freezing. All steps were performed at 4°C.

SDS-PAGE analysis

The concentrated samples of PfELC were dialyzed against 50 mM Tris (pH 8.0), 20 mM NaCl, and 0, 0.25, 0.5 or 1 mM TCEP overnight at 4°C. Subsequently, the protein concentration was adjusted to 1 mg/ml and 50 µl of each sample was mixed with a fivefold excess of 2-iodoacetamide. The samples were incubated for 1 h at 37°C and afterwards, 10 µl of each sample was mixed with 5 µl of non-reducing loading dye. The gel was run at 180 V for 40 min and stained by Direct Blue.

Analytical gel filtration

The proteins and protein complexes were analyzed by analytical gel filtration using a Superdex 200 5/150 column (GE Healthcare) and the 1260 Infinity Bio-inert high-performance liquid chromatography system (Agilent Technologies) at 10°C. The system and column were equilibrated in 20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5mM TCEP and 30 µl of each sample was injected by an auto sampler. The system was run at 0.2 ml/min for 20 minutes and the elution profile was recorded by a UV detector.

Thermal shift assay

The stability of the different proteins was measured by nanoDSF (Prometheus NT.48, NanoTemper Technologies, GmbH). The proteins were first dialyzed against 1 l of gel filtration buffer supplemented with 5 mM EDTA overnight at 4°C and subsequently 2x against 1 l of gel filtration buffer without EDTA overnight at 4°C. The protein concentration was then adjusted to 100 µM (individually or 100 µM each component of a complex) in gel filtration buffer and varying concentrations of calcium chloride (0 – 500 µM). 10 µl of sample was loaded in the glass capillaries and heated from 20°C to 95°C with a heating rate of 1°C/min. The fluorescence signals with an excitation wavelength of 280 nm and emission wavelengths of

330 and 350 nm were recorded and the melting temperature was calculated as either the maximum of the derivative of the ratio of fluorescence at 330 and 350 nm, or as maximum of the derivative of the fluorescence recorded at 330 nm.

Circular dichroism

To estimate the secondary structure content of the proteins and peptides, we measured circular dichroism on a Chirascan CD spectrometer (Applied Photophysics). For spectrum measurements, the protein or peptide concentration was adjusted to 100 µM and diluted tenfold by 10 mM NaP (pH 7.5), 20 mM NaCl, 0.25 mM TCEP just prior to the measurement. To measure the difference in secondary structure content in presence or absence of calcium, the proteins were first dialyzed against 1 l of gel filtration buffer supplemented with 5 mM EDTA overnight at 4°C and subsequently 2x against 1 l of gel filtration buffer supplemented with ±1 mM $CaCl_2$ overnight at 4°C. The proteins were then diluted to 5 µM or 10 µM with 10 mM NaP (pH 7.5), 20 mM NaCl, 0.25 mM TCEP and ±1 mM $CaCl_2$ just prior to the measurement. The CD spectrum was measured between 200 nm and 260 nm with 1 nm steps in triplicates using a 2 mm quartz cuvette. To assess the induction of structure in the dimeric protein complexes, each component was diluted by 10 mM NaP (pH 7.5), 150 mM NaF and 0.25 mM TCEP to a final concentration of 5 µM. The circular dichroism was measured 10x between 195 nm and 260 nm with 0.5 nm step in 1 mm quartz cuvette. The data were averaged, background subtracted and analyzed by K2D algorithm[52] using DichroWeb[53].

Isothermal titration calorimetry

To measure the interaction of TgELC1 or TgELC2 with the TgMyoA-C^{ELC} peptide (S777-Q798), the peptides were dissolved and the proteins were dialyzed in gel filtration buffer supplemented with either 5mM $CaCl_2$ or EDTA overnight at 4°C and 2 µl of a 200 µM peptide solution were injected 19 times into 20 µM protein. To measure the interaction of the trimeric complex, first, the peptides were dissolved and the proteins dialyzed against gel filtration buffer supplemented with 1 mM $CaCl_2$. The complex of TgELC1, TgELC2 or MTIP-S with the MyoA

peptide (S777-V818 in *T. gondii*, V775-V816 in *P. falciparum*) was first formed in 1:1.1 molar ratio, respectively, and incubated for 1 h at 4°C. For measurement, 2 µl of 200 µM TgMLC-S or PfELC was injected 19 times into 20 µM of the pre-formed complex. The measurements were performed with a MicroCal PEAQ-ITC (Malvern) at 25°C. The data were processed using the MicroCal PEAQ-ITC Analysis Software and fitted with a one-site binding model.

Bioinformatics methods

The homologous protein sequences were aligned with the program MAFFT[54]. The protein disorder probability was calculated using the disEMBL[55] server with loops and coils defined by dictionary of secondary structure of proteins [56]. The secondary structure prediction of PfELC, TgELC1 and TgELC2 was calculated in JPred[57].

Small angle X-ray scattering

The SAXS data were collected at the P12 BioSAXS beamline[58] at the PETRA III storage ring (DESY, Hamburg, Germany). The concentrated samples of TgELC2 and PfELC (10 mg/ml) were dialyzed against the buffer (20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5 mM TCEP for TgELC2; 20 mM Tris (pH 8.0), 150 mM NaCl, 0.5 mM TCEP for PfELC-N) overnight at 4°C. Further, the samples were centrifuged (5 min, 15 000*g*, 4°C) and a dilution series of each sample (typically in a range of 0.5 – 10 mg/ml) and their corresponding solvent were measured at room temperature under continuous flow with a total exposure of 1 s (20 x 50 ms frames). The dimeric complex TgELC2/TgMyoA-C, as well as the trimeric complexes using different constructs, were mixed in 1:1 or 1:1:1 molar ratio, purified by SEC and concentrated to 10 mg/ml prior to the measurement. The X-ray scattering data were measured in an on-line SEC-SAXS mode, using a SD200 Increase column (GE Healthcare) at 0.5 ml/min with 1 frame recorded per second. The sample of PfELC was concentrated to 10 mg/ml and the X-ray scattering was measured in the on-line SEC-SAXS mode, using a SD200 5/150 column at 0.4 ml/min. The automatically processed data were further analyzed using the ATSAS suite[59] programs CHROMIXS[60] and PRIMUS[61] to determine the overall parameters and distance distribution, CRYSOL[62] to compute the scattering from the crystal structures and CORAL[63] to

compute the scattering from the crystal structures with dummy residues mimicking the missing flexible parts. The results of all SAXS measurements are summarized in Supplementary Table 3. All SAXS data and models have been deposited in the SASBDB (www.sasbdb.org) with accession codes: SASDH64, SASDH74, SASDH84, SASDH94, SASDHA4, SASDHB4, SASDHC4, SASDHD4 and SASDHE4.

NMR

All NMR experiments were conducted on a Bruker Avance II 800 NMR spectrometer equipped with a cryoprobe at 288 K in 50 mM HEPES, 20 mM NaCl, 0.5 mM TCEP and 10% (v/v) D_2O at pH 7.0, except for H(CCO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments that were performed on a Bruker Avance III 600 NMR spectrometer equipped with a room temperature probe. Full-length PfELC (residues 1-134) was ^{15}N and $^{15}N^{13}C$ labeled and concentrated to 500 µM. PfELC-N was also ^{15}N and $^{15}N^{13}C$ labeled and in addition site-selectively ^{13}C labeled [64–66] by using 1-$^{13}C_1$ and 2-$^{13}C_1$ glucose. Samples were concentrated to about 1 mM. All spectra were processed suing NMRPipe[67] and analyzed using NMRView[68].

Backbone resonances of $^{15}N^{13}C$ labeled samples (1-74 and 1-134) were assigned using HNCACB[69] and HN(CO)CACB[70] experiments. Aliphatic side chains (1-74) were assigned using H(CCO)NH-TOCSY[71] (H)C(CO)NH-TOCSY and H(C)CH-TOCSY[72] experiments. Aromatic side chains (1-74) were assigned by (HB)CB(CGCD)HD[73] and aromatic H(C)CH-TOCSY experiments and verified by the site-selective ^{13}C labeling.

NOEs for the structure determination were derived from 3D-NOESY-HSQC experiments for ^{15}N, ^{13}C aliphatic nuclei and ^{13}C aromatic nuclei (on 1-$^{13}C_1$ and 2-$^{13}C_1$ glucose labeled samples). Phi-Psi dihedral angle constraints were derived using TALOS[74]. Structure calculations were performed using ARIA 2.3[75] and standard parameters. The lowest-energy models have been deposited in the PDB with accession number 6tj3. Secondary structure elements were determined from chemical shifts and the dynamics of the PfELC backbone was probed using heteronuclear NOEs ({^1H}-^{15}N NOE). This ^{15}N based dynamics experiment allowed us to distinguish between rigid ({^1H}-^{15}N NOE > 0.7, secondary structure elements),

somewhat flexible ({^{1}H}-^{15}N NOE ~ 0.5-0.7, loops and turns) and extremely flexible ({1H}-^{15}N NOE < 0.5, unfolded/ random coil) regions of the protein. Ramachandran analysis was performed by PROCHECK[76].

{^{1}H}-^{15}N NOE saturation was performed using a train of shaped 180° pulses in a symmetric fashion[77–79] for 3 s and a total inter-scan relaxation period of 10 s. Data collection, processing and analysis details are summarized in Table 3.

Crystallization

PfELC-N was concentrated (5kDa cut-off) to 26 mg/ml and 200 nl of the sample was mixed with 100 nl of reservoir solution (0.1M Tris-HCl (pH 8.5), 0.2M Li_2SO_4, 30% PEG 4000). The crystals grew in sitting drop plates at 19°C for 7 days.

The trimeric complex of MLC1-S, TgELC2 and TgMyoA-C (S777-V818) was mixed in a molar ratio of 1.1: 1.1: 1, respectively. After 1 h of incubation, the trimeric complex was separated by gel filtration in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP using a Superdex 75 16/600 column (GE Healthcare). The fractions containing the peak of the trimeric complex were concentrated (5 kDa cut-off) to 10 mg/ml. The crystals grew for 7 days at 19°C in sitting drop plates prepared by mixing 200 nl of the sample with 100 nl of reservoir solution (0.1 M imidazole, 0.1 M MES monohydrate pH 6.5, 20% v/v PEG 500 MME, 10% w/v PEG 20 000, 0.12 M 1,6-hexadiol, 0.12 M 1-butanol, 0.12 M 1,2-propanediol, 0.12 M 2-propanol, 0.12 M 1,4-butanediol, 0.12 M 1,3-propanediol).

The recombinantly expressed dimeric complex of TgELC1 and TgMyoA-C (S777-V818) was mixed with MLC1 in 1:1.1 molar ratio, incubated for 1 h and the trimeric complex was separated by gel filtration in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP using a Superdex 75 16/600 column (GE Healthcare). The fractions containing the peak of the trimeric complex were concentrated (5 kDa cut-off) to 10 mg/ml. The crystals of the calcium-bound complex grew within 7 days at 19°C in a sitting drop prepared by mixing 200 nl of the sample with 100 nl of reservoir solution (20% w/v ethylene glycol, 10% w/v PEG 8000, 0.1M Tris (base), 0.1M bicine pH 8.5, 0.09 M sodium nitrate, 0.09 M sodium phosphate dibasic, 0.09 M

ammonium sulfate). The crystals of the calcium-free complex grew within 7 days at 19°C in a sitting drop plate prepared by mixing 200 nl of the sample with 100 nl of reservoir solution (32% w/v PEG 8000, 0.1M Tris pH 7.0, 0.2M LiCl).

The recombinantly expressed dimeric complex of MTIP (residues 77-204) and PfMyoA-C (775-816) were mixed with excess of His-tagged PfELC, the complex was purified by NiNTA IMAC, dialyzed and TEV-cleaved overnight at 4°C and further purified by negative NiNTA IMAC and size exclusion chromatography using Superdex 200 column with 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The crystals grew within 3 days at 4°C in a sitting drop prepared by mixing 150 nl of the sample with 150 nl of reservoir solution (0.1M imidazole/MES pH 6.5, 20% w/v ethylene glycol, 10% PEG 8000, 0.03M of each di-ethylene glycol, tri-ethylene glycol, tetra-ethylene glycol and penta-ethylene glycol).

Data collection and structure determination

The diffraction data of the trimeric complexes were collected at the P13 EMBL beamline of the PETRA III storage ring (c/o DESY, Hamburg, Germany) at 0.9762 Å wavelength and 100 K temperature using a Pilatus 6 M detector (DECTRIS). The diffraction data of PfELC-N were collected at the P14 EMBL beamline of the PETRA III storage ring (c/o DESY, Hamburg, Germany) at 1.0332 Å and 100 K temperature using an EIGER 16 M detector (DECTRIS). The diffraction data were processed using XDS[80], merged with Aimless[81] or (STARANISO[82] in case of the *P. falciparum* trimeric complex) and phase information were obtained by molecular replacement with Phaser[83], using the structure of peptide-bound TgMLC1 (PDB ID 5vt9) as a search model in case of the trimeric complexes and the NMR structure as search model in case of PfELC-N. In all cases, the models were further built and refined in several cycles using PHENIX[84], Refmac[85] and *Coot*[86]. Data collection and refinement statistics are summarized in Table 2. In all structures, over 98% residues are in the favored region of the Ramachandran plot and each structure contains no more than one Ramachandran outlier. PyMOL was used to generate figures, measure the angle of the helical kink, inter-molecular angles, distances and RMSDs. PDBePISA[87] was used to characterize the intermolecular

interfaces. The atomic coordinates and the structure factors have been deposited in the PDB with accession numbers 6tj4, 6tj5, 6tj6, 6tj7 and 6zn3.

Modelling

The modelling procedure was performed in Modeller version 9.18[88]. We built 50 models for the TgMyoA residues 772-791. These 50 models were fused to the structure of TgMyoA (PDB ID 6due; residues 33-771). All 50 models were tilting along the bond/dihedral angle between residue 771 and the first modelled residue, that is 772; at the same time, the residues 33-771 of the 6due structure remained fixed. Thus, each of the produced models consisted of an intact crystal structure 6due (till residue 771) and *de novo* modelled fragment of 772-791. Restraints in a form of i-i+4 h-bonding pattern were imposed in order to ensure that all 50 models have an α-helical conformation along the whole length of the *de novo* modelled fragment, and also at the junction between residues 771 and 772. The crystal structure of complex 1 (PDB ID 6tj5) or complex 2 (PDB ID 6tj7) were superposed on the 50 models using the TgMyoA residues 780-791. After superposition, the modelled conformation of this fragment was removed from the merged structures, which produced models consisting of an intact crystal structure of TgMyoA (PDB 6due), the modelled helix of TgMyoA (residues 772-779) and the intact crystal structure of the complex 1 (50 models) or complex 2 (50 models), starting from the TgMyoA residue 780 of these structures. Next, all reconstructed complexes were screened against the existence of atomic clashes using the Chimera software[89] and the best five models (both complex 1 and complex 2) were energy minimized by executing 1000 steps of conjugate gradient energy minimization in the NAMD program[90]. All energy minimizations were performed in a water box with ions.

Normal mode analysis

Normal mode analysis (NMA)[91] was used to probe essential dynamics of the reconstructed trimeric models. The NMA was performed in an all-atom representation on the best five energy-minimized models using the BIO3D software[92] . The deformation analysis was

performed, using the first 20, 50 and 100 modes, and also on the first 10 modes separately. This allowed us to not only identify possible hinge points within the studied structures of trimeric complexes, but also to determine which hinges correspond to which modes.

Statistics and reproducibility

In all reported experiments, the protein samples were expressed and purified under identical experimental conditions. The figures represent the results from one experiment, unless stated otherwise. The CD experimental curves were recorded 10 times, averaged and buffer-subtracted. The SAXS data recorded in batch mode represent a buffer-subtracted average of 20 measurements of the same sample measured under continuous flow.

Acknowledgments

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support with nanoDSF, ITC measurements and with protein crystallization. We acknowledge all group members for continuous support and feedback on the project and during manuscript preparation. We would like to thank the group of Thomas R. Schneider at EMBL Hamburg for access to the EMBL beamlines P13 and P14 and Guillaume Pompidor and Grzegorz Chojnowski for help with data processing and initial model building.

Author contributions

Designed research: S.P., K.K., J.K., U.W. and C.L.; Performed research: S.P., K.D., K.K., H.M., U.W., C.L.; Analyzed data: S.P., K.D., K.K., H.M., T.G., D.S., J.K., U.W., C.L.; Prepared figures: S.P.; Wrote the paper: S.P. and C.L. and all other authors contributed to writing of the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on request. The data source data underlying the charts in the main and supplementary figures is deposited in Figshare repository[93]. Coordinates and structure factors as well as NMR structures were deposited in the PDB at the Research Collaboratory for Structural Bioinformatics (RCSB) with the following identifying codes: 6tj3, 6tj4, 6j5, 6tj6, 6tj7, 6zn3. The averaged and subtracted SAXS data were deposited in SASBDB with the following identifying codes: SASDH64, SASDH74, SASDH84, SASDH94, SASDHA4, SASDHB4, SASDHC4, SASDHD4 and SASDHE4. The structural models of full lengths MyoA-MLC1-ELCs have been uploaded to Zenodo (https://zenodo.org).

References

1. *World Malaria Report 2018*. (2018).
2. Dubey, J. P. *Toxoplasmosis of animals and humans*. (CRC Press, 2010). doi:10.1201/9781420092370
3. Cowman, A. F., Tonkin, C. J., Tham, W. H. & Duraisingh, M. T. The Molecular Basis of Erythrocyte Invasion by Malaria Parasites. *Cell Host Microbe* **22**, 232–245 (2017).
4. Frénal, K., Dubremetz, J. F., Lebrun, M. & Soldati-Favre, D. Gliding motility powers invasion and egress in Apicomplexa. *Nature Reviews Microbiology* (2017). doi:10.1038/nrmicro.2017.86
5. Heintzelman, M. B. Gliding motility in apicomplexan parasites. *Seminars in Cell and Developmental Biology* (2015). doi:10.1016/j.semcdb.2015.09.020
6. Kono, M., Prusty, D., Parkinson, J. & Gilberger, T. W. The apicomplexan inner membrane complex. *Front. Biosci.* (2013). doi:10.2741/4157
7. Gaskins, E. *et al.* Identification of the membrane receptor of a class XIV myosin in Toxoplasma gondii. *J. Cell Biol.* **165**, 383–393 (2004).
8. Jones, M. L., Kitson, E. L. & Rayner, J. C. Plasmodium falciparum erythrocyte invasion : A conserved myosin associated complex. **147**, 74–84 (2006).
9. Soldati-Favre, D. Molecular dissection of host cell invasion by the Apicomplexans: the glideosome. *Parasite* (2008). doi:10.1051/parasite/2008153197
10. Frénal, K. *et al.* Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host Microbe* (2010). doi:10.1016/j.chom.2010.09.002
11. Dobrowolski, J. M., Carruthers, V. B. & Sibley, L. D. Participation of myosin in gliding motility and host cell invasion by Toxoplasma gondii . *Mol. Microbiol.* (1997). doi:10.1046/j.1365-2958.1997.5671913.x
12. Meissner, M., Schlüter, D. & Soldati, D. Role of Toxoplasma gondii myosin a in powering parasite gliding and host cell invasion. *Science (80-.).* **298**, 837–840 (2002).
13. Bookwalter, C. S., Kelsen, A., Leung, J. M., Ward, G. E. & Trybus, K. M. A toxoplasma gondii class XIV myosin, expressed in Sf9 cells with a parasite co-chaperone, requires two light chains for fast motility. *J. Biol. Chem.* **289**, 30832–30841 (2014).
14. Green, J. L. *et al.* Compositional and expression analyses of the glideosome during the Plasmodium life cycle reveal an additional myosin light chain required for maximum motility. *J. Biol. Chem.* **292**, 17857–17875 (2017).
15. Powell, C. J. *et al.* Dissecting the molecular assembly of the Toxoplasma gondii MyoA motility complex. *J. Biol. Chem.* (2017). doi:10.1074/jbc.M117.809632
16. Williams, M. J. *et al.* Two essential light chains regulate the MyoA lever arm to promote toxoplasma gliding motility. *MBio* **6**, 1–16 (2015).
17. Bookwalter, C. S. *et al.* Reconstitution of the core of the malaria parasite glideosome with recombinant Plasmodium class XIV myosin A and Plasmodium actin. *J. Biol. Chem.* **292**, 19290–19303 (2017).
18. Sebastian, S. *et al.* A Plasmodium calcium-dependent protein kinase controls zygote development and transmission by translationally activating repressed mRNAs. *Cell Host Microbe* **12**, 9–19 (2012).
19. Egarter, S. *et al.* The toxoplasma acto-myoA motor complex is important but not essential for gliding motility and host cell invasion. *PLoS One* **9**, (2014).
20. Jones, M. L., Collins, M. O., Goulding, D., Choudhary, J. S. & Rayner, J. C. Analysis of Protein Palmitoylation Reveals a Pervasive Role in Plasmodium Development and Pathogenesis. 246–258 (2012). doi:10.1016/j.chom.2012.06.005
21. Rees-Channer, R. R. *et al.* Dual acylation of the 45 kDa gliding-associated protein (GAP45) in Plasmodium falciparum merozoites. *Mol. Biochem. Parasitol.* **149**, 113–

116 (2006).

22. Frénal, K. *et al.* Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host Microbe* **8**, 343–357 (2010).
23. Ridzuan, M. A. M. *et al.* Subcellular location, phosphorylation and assembly into the motor complex of GAP45 during Plasmodium falciparum schizont development. *PLoS One* **7**, (2012).
24. Harding, C. R. *et al.* Gliding Associated Proteins Play Essential Roles during the Formation of the Inner Membrane Complex of Toxoplasma gondii. *PLoS Pathog.* (2016). doi:10.1371/journal.ppat.1005403
25. Bosch, J., Paige, M. H., Vaidya, A. B., Bergman, L. W. & Hol, W. G. J. Crystal structure of GAP50, the anchor of the invasion machinery in the inner membrane complex of Plasmodium falciparum. *J. Struct. Biol.* **178**, 61–73 (2012).
26. Douse, C. H. *et al.* Regulation of the plasmodium motor complex: Phosphorylation of myosin a tail-interacting protein (MTIP) loosens its grip on MyoA. *J. Biol. Chem.* **287**, 36968–36977 (2012).
27. Powell, C. J. *et al.* Structural and mechanistic insights into the function of the unconventional class XIV myosin MyoA from Toxoplasma gondii. *Proc. Natl. Acad. Sci. U. S. A.* (2018). doi:10.1073/pnas.1811167115
28. Robert-Paganin, J. *et al.* Plasmodium myosin A drives parasite invasion by an atypical force generating mechanism. *Nat. Commun.* (2019). doi:10.1038/s41467-019-11120-0
29. Dovega, R. *et al.* Structural and biochemical characterization of human PR70 in isolation and in complex with the scaffolding subunit of protein phosphatase 2A. *PLoS One* (2014). doi:10.1371/journal.pone.0101846
30. Treeck, M., Sanders, J. L., Elias, J. E. & Boothroyd, J. C. The phosphoproteomes of plasmodium falciparum and toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* **10**, 410–419 (2011).
31. Batters, C. *et al.* Myo1c is designed for the adaptation response in the inner ear. *EMBO J.* (2004). doi:10.1038/sj.emboj.7600169
32. Sun, Y. & Goldman, Y. E. Lever-arm mechanics of processive myosins. *Biophysical Journal* (2011). doi:10.1016/j.bpj.2011.05.026
33. Sakamoto, T., Yildez, A., Selvin, P. R. & Sellers, J. R. Step-size is determined by neck length in myosin V. *Biochemistry* (2005). doi:10.1021/bi0512086
34. Ni, S. *et al.* Modification of interface between regulatory and essential light chains hampers phosphorylation-dependent activation of smooth muscle myosin. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M112.343491
35. Colson, B. A., Gruber, S. J. & Thomas, D. D. Structural dynamics of muscle protein phosphorylation. *Journal of Muscle Research and Cell Motility* (2012). doi:10.1007/s10974-012-9317-6
36. Espinoza-Fonseca, L. M., Colson, B. A. & Thomas, D. D. Effects of pseudophosphorylation mutants on the structural dynamics of smooth muscle myosin regulatory light chain. *Mol. Biosyst.* (2014). doi:10.1039/c4mb00364k
37. Kast, D., Espinoza-Fonseca, L. M., Yi, C. & Thomas, D. D. Phosphorylation-induced structural changes in smooth muscle myosin regulatory light chain. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.1001941107
38. Houdusse, A., Silver, M. & Cohen, C. A model of Ca2+-free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch. *Structure* (1996). doi:10.1016/S0969-2126(96)00154-2
39. Zhang, Y., Kawamichi, H., Kohama, K. & Nakamura, A. Calcium-mediated regulation of recombinant hybrids of full-length Physarum myosin heavy chain with Physarum/scallop myosin light chains. *Acta Biochim. Biophys. Sin. (Shanghai).* (2016). doi:10.1093/abbs/gmw031

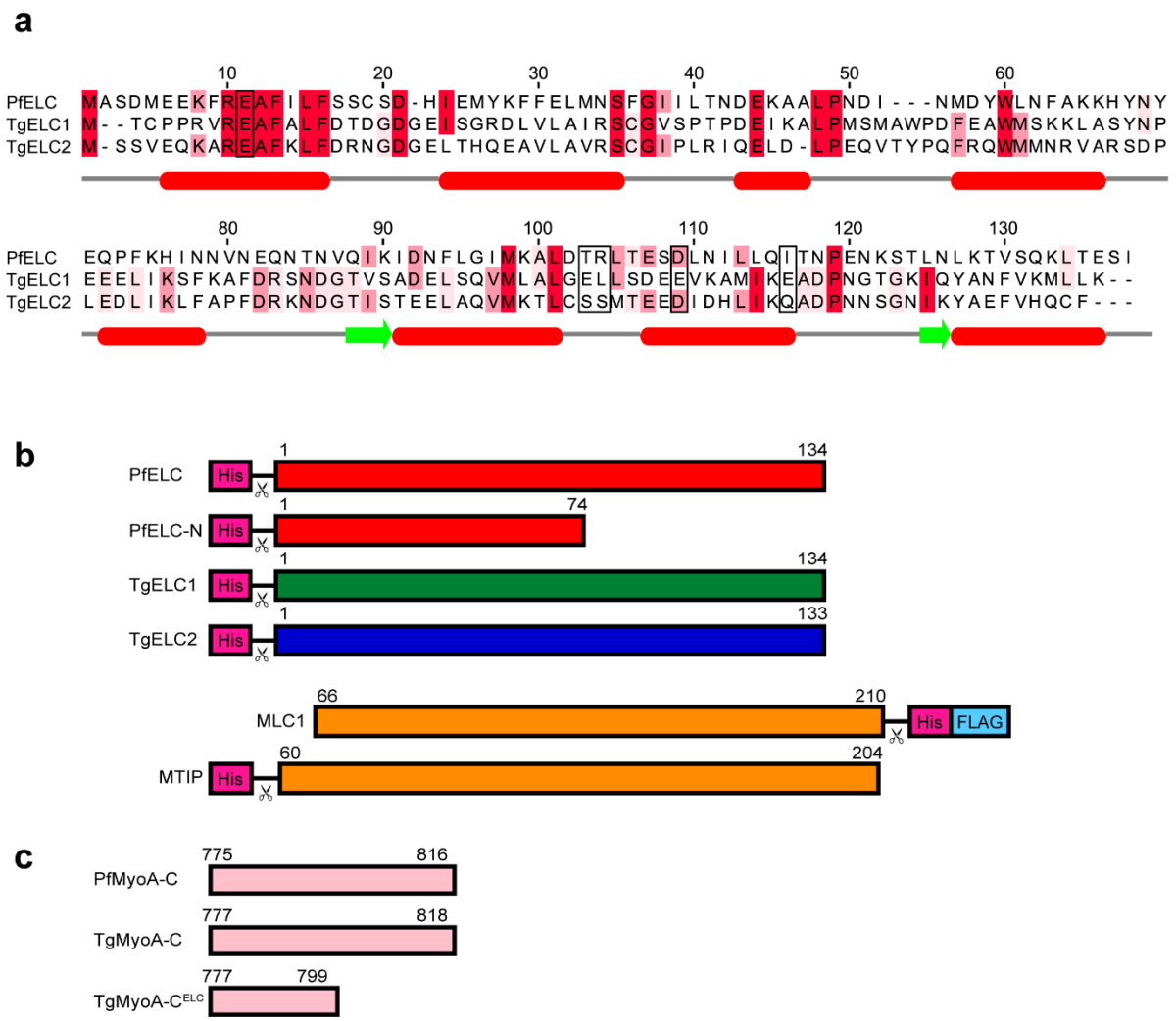
40. Brochet, M. & Billker, O. Calcium signalling in malaria parasites. *Mol. Microbiol.* **100**, 397–408 (2016).
41. Pylypenko, O. & Houdusse, A. M. Essential 'ankle' in the myosin lever arm. *Proceedings of the National Academy of Sciences of the United States of America* (2011). doi:10.1073/pnas.1017676108
42. Lovett, J. L. & Sibley, L. D. Intracellular calcium stores in Toxoplasma gondii govern invasion of host cells. *J. Cell Sci.* (2003). doi:10.1242/jcs.00596
43. Green, J. L. *et al.* The motor complex of Plasmodium falciparum: Phosphorylation by a calcium-dependent protein kinase. *J. Biol. Chem.* **283**, 30980–30989 (2008).
44. Nebl, T. *et al.* Quantitative in vivo analyses reveal calcium-dependent phosphorylation sites and identifies a novel component of the toxoplasma invasion motor complex. *PLoS Pathog.* **7**, (2011).
45. Dominguez, R., Freyzon, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: Visualization of the pre-power stroke state. *Cell* (1998). doi:10.1016/S0092-8674(00)81598-6
46. Houdusse, A., Szent-Györgyi, A. G. & Cohen, C. Three conformational states of scallop myosin S1. *Proc. Natl. Acad. Sci. U. S. A.* (2000). doi:10.1073/pnas.200376897
47. Thomas, D. D., Kast, D. & Korman, V. L. Site-Directed Spectroscopic Probes of Actomyosin Structural Dynamics. *Annu. Rev. Biophys.* (2009). doi:10.1146/annurev.biophys.35.040405.102118
48. Borejdo, J. *et al.* The power stroke causes changes in the orientation and mobility of the termini of essential light chain 1 of myosin. *Biochemistry* (2001). doi:10.1021/bi002527u
49. Savitsky, P. *et al.* High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* (2010). doi:10.1016/j.jsb.2010.06.008
50. Diebold, M. L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. *J. Struct. Biol.* (2011). doi:10.1016/j.jsb.2011.02.001
51. Löw, C. *et al.* Crystal structure determination and functional characterization of the metallochaperone SlyD from thermus thermophilus. *J. Mol. Biol.* (2010). doi:10.1016/j.jmb.2010.03.014
52. Andrade, M. A., Chacón, P., Merelo, J. J. & Morán, F. Evaluation of secondary structure of proteins from uv circular dichroism spectra using an unsupervised learning neural network. *Protein Eng. Des. Sel.* (1993). doi:10.1093/protein/6.4.383
53. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* (2008). doi:10.1002/bip.20853
54. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz342
55. Linding, R. *et al.* Protein disorder prediction: Implications for structural proteomics. *Structure* (2003). doi:10.1016/j.str.2003.10.002
56. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen- bonded and geometrical features. *Biopolymers* (1983). doi:10.1002/bip.360221211
57. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn238
58. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J.*

Appl. Crystallogr. (2015). doi:10.1107/S160057671500254X

59. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* (2017). doi:10.1107/S1600576717007786
60. Panjkovich, A. & Svergun, D. I. CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* (2018). doi:10.1093/bioinformatics/btx846
61. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* (2003). doi:10.1107/S0021889803012779
62. Svergun, D., Barberato, C. & Koch, M. H. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* (1995). doi:10.1107/S0021889895007047
63. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* (2012). doi:10.1107/S0021889812007662
64. Lundström, P. *et al.* Fractional 13C enrichment of isolated carbons using [1-13C]- or [2-13C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins. *J. Biomol. NMR* (2007). doi:10.1007/s10858-007-9158-6
65. Teilum, K., Brath, U., Lundström, P. & Akke, M. Biosynthetic 13C labeling of aromatic side chains in proteins for NMR relaxation measurements. *J. Am. Chem. Soc.* (2006). doi:10.1021/ja055660o
66. Weininger, U. Optimal Isotope Labeling of Aromatic Amino Acid Side Chains for NMR Studies of Protein Dynamics. in *Methods in Enzymology* (2019). doi:10.1016/bs.mie.2018.08.028
67. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* (1995). doi:10.1007/BF00197809
68. Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* (2004).
69. Wittekind, M. & Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B* (1993). doi:10.1006/jmrb.1993.1033
70. Grzesiekt, S. & Bax, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* (1992). doi:10.1021/ja00042a003
71. Bax, A., Clore, G. M. & Gronenborn, A. M. 1H1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *J. Magn. Reson.* (1990). doi:10.1016/0022-2364(90)90202-K
72. Grzesiek, S., Anglister, J. & Bax, A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. *Journal of Magnetic Resonance, Series B* (1993). doi:10.1006/jmrb.1993.1019
73. Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. Two-Dimensional NMR Experiments for Correlating 13Cβ and 1Hδ/ϵ Chemical Shifts of Aromatic Residues in 13C-Labeled Proteins via Scalar Couplings. *Journal of the American Chemical Society* (1993). doi:10.1021/ja00076a099
74. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* (1999). doi:10.1023/A:1008392405740

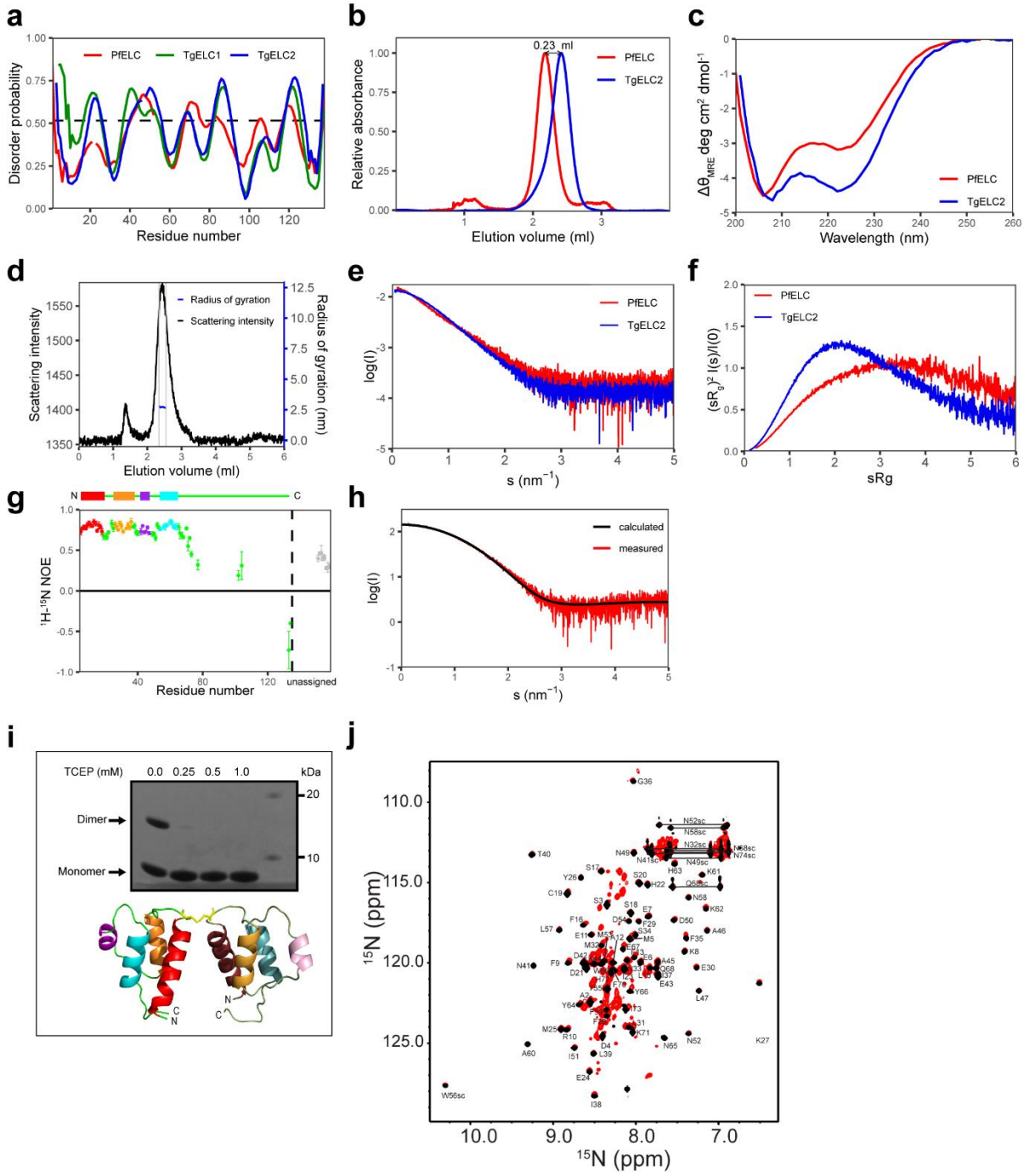
75. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics* (2003). doi:10.1093/bioinformatics/19.2.315
76. Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. PROCHECK : validation of protein-structure coordinates . in (2012). doi:10.1107/97809553602060000882
77. Ferrage, F., Piserchio, A., Cowburn, D. & Ghose, R. On the measurement of 15N-{1H} nuclear Overhauser effects. *J. Magn. Reson.* (2008). doi:10.1016/j.jmr.2008.03.011
78. Ferrage, F., Cowburn, D. & Ghose, R. Accurate sampling of high-frequency motions in proteins by steady-state 15N-{ 1H} nuclear overhauser effect measurements in the presence of cross-correlated relaxation. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja809526q
79. Weininger, U., Diehl, C. & Akke, M. 13C relaxation experiments for aromatic side chains employing longitudinal-and transverse-relaxationoptimized NMR spectroscopy. *J. Biomol. NMR* (2012). doi:10.1007/s10858-012-9650-5
80. Kabsch, W. 1 XDS. *Acta Crystallogr. Sect. D* (2010). doi:10.1107/s0907444909047337
81. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2013). doi:10.1107/S0907444913000061
82. Tickle, I. J. *et al.* STARANISO. *Cambridge, United Kingdom: Global Phasing Ltd.* (2018).
83. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* (2007). doi:10.1107/S0021889807021206
84. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909052925
85. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* (1997). doi:10.1107/S0907444996012255
86. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444910007493
87. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* (2007). doi:10.1016/j.jmb.2007.05.022
88. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* (2016). doi:10.1002/cpbi.3
89. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* (2018). doi:10.1002/pro.3235
90. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* (2005). doi:10.1002/jcc.20289
91. Bahar, I., Lezon, T. R., Bakan, A. & Shrivastava, I. H. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chem. Rev.* (2010). doi:10.1021/cr900095e
92. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* (2006). doi:10.1093/bioinformatics/btl461
93. Pazicky, S., Dhamotharan, K. Kaszuba, K., Mertens, H. D. T., Gilberger, T., Svergun, D., Kosinski, J., Weininger, U., Löw, C.: Structural role of essential light chains in the apicomplexan glideosome data sets. figshare http://dx.doi.org/10.6084/m9.figshare.12769841 (2020)

Supplementary Fig. 1



(A) Sequence alignment of *P. falciparum* PfELC and *T. gondii* TgELC1 and TgELC2. Identical residues between these proteins are highlighted in red. The boxed residues indicate the residues involved in the polar interactions with TgMyoA (see Fig. 3c-d and Supplementary Table 6). Secondary structure elements of PfELC as predicted by JPred are graphically shown under the sequence alignment. (B) Schematic representation of the myosin light chain constructs used in this study. The numbers indicate the sequence residues of the particular protein; the scissor symbol represents a TEV cleavage site. (C) The peptide constructs representing the C-terminal regions of MyoAs with indicated domain boarders used in this study. The constructs PfMyoA-C and TgMyoA-C encompass both MLC1/MTIP binding sites as well as the upstream conserved region which binds the essential light chains. The construct TgMyoA-C^{ELC} only consists of the TgELC binding site.

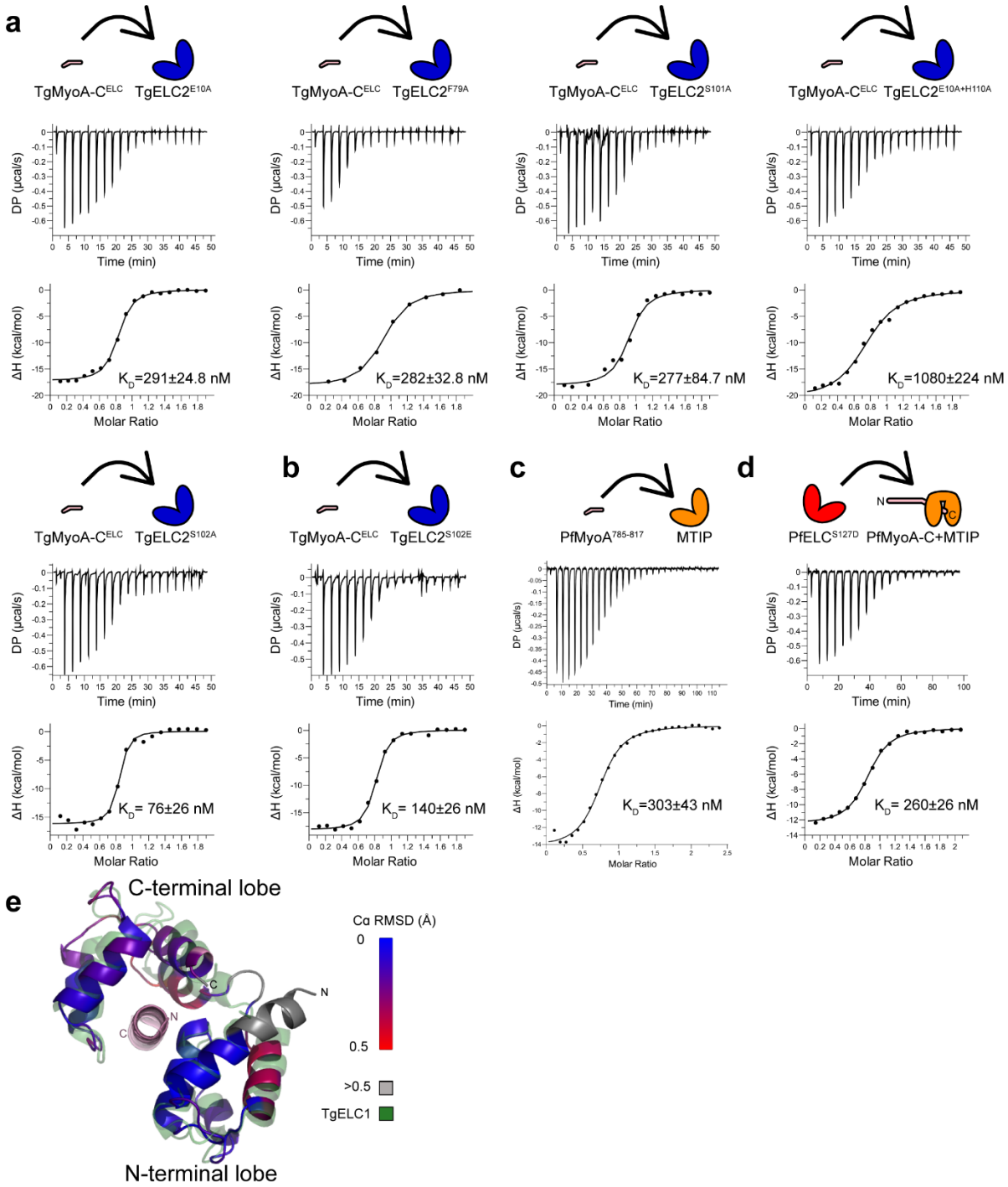
Supplementary Fig. 2



(A) Disorder probability prediction calculated by the disEMBL server shows differences between PfELC and TgELCs, predominantly in the C-terminal region of the sequence. The disorder for the prediction was defined by dictionary of secondary structure probabilities. The amino acid residues with disorder probability above the threshold (dashed line) are predicted to be disordered. (B) Gel filtration profile of PfELC (red) and TgELC2 (blue) on a Superdex 200 5/150 column. PfELC elutes at a smaller elution volume, suggesting that it has a larger hydrodynamic radius compared to TgELC2. (C) Far-UV circular dichroism spectrum of PfELC (red) and TgELC2 (blue) shows that PfELC has a lower α-helical and higher random coil content compared to TgELC2. (D) Elution profile of on-line SEC-SAXS measurement of PfELC using a Superdex 200 5/150 column with the region used for the analysis highlighted in grey. (E) Recorded SAXS curves of isolated PfELC and TgELC2 indicate conformational differences of these homologous proteins. (F) The dimensionless Kratky plot shows that TgELC2 is more compact than

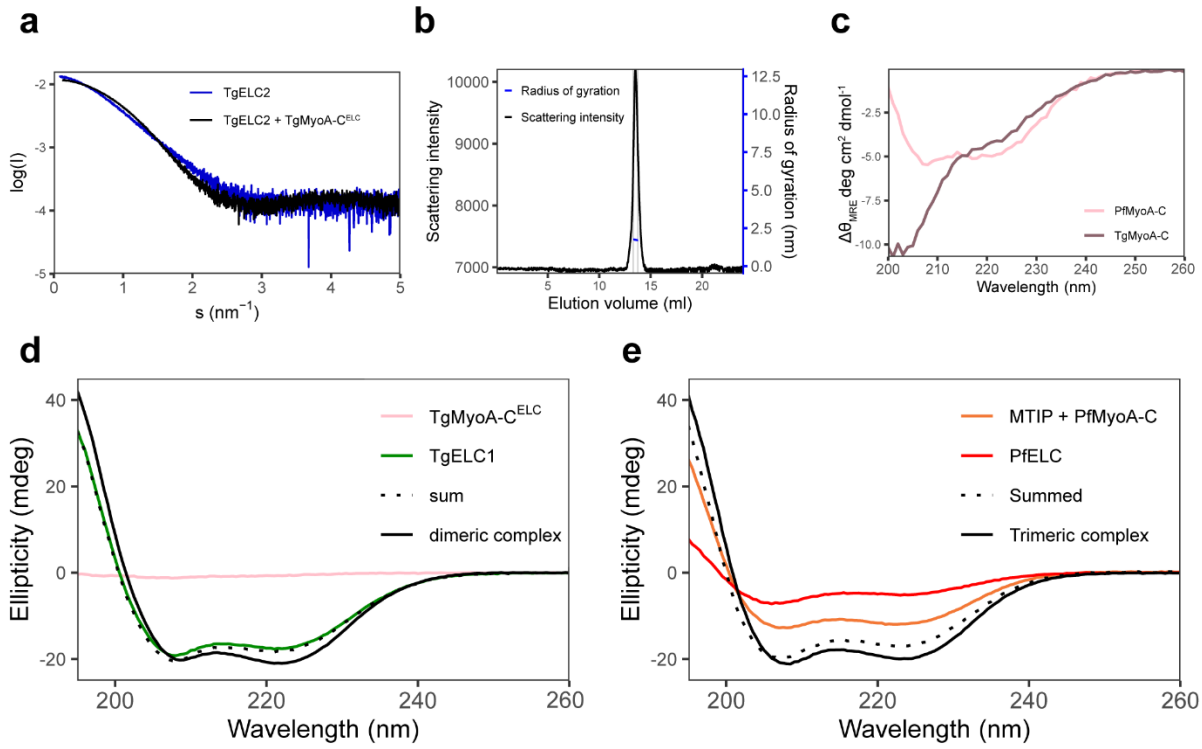
PfELC. This is evident because the maximum is considerably closer to $sR_g = \sqrt{3}$ for TgELC2 than PfELC and the plot converges closer to zero in case of TgELC2. (G) Backbone dynamics of PfELC on a picosecond to nanosecond time scale. Heteronuclear NOE ({^{1}H}-^{15}N NOE) of PfELC on a residue basis. The C-terminus of PfELC is disordered as indicated by the low heteronuclear NOEs for this region. Residues are colored according to secondary structure elements (four α-helices: from N terminus red, orange, violet, cyan, random coil/loop residues are green, unassigned C-terminal residues in grey). (H) Experimental small angle X-ray scattering curve of PfELC (red) and calculated scattering (black line) from the crystal structure of the PfELC monomer fit with a X^2 value of 1.37, confirming that the protein is a structurally rigid globular monomer in solution. (I) SDS-PAGE gel with PfELC-N samples dialyzed against buffers with varying concentration of TCEP and subsequently alkylated by 2-iodoacetamide and dimer of PfELC-N formed by a cysteine bond between two symmetry related molecules. The results show that the protein is monomeric at the concentration of TCEP used for its biophysical characterization. (J) Overlay of ^{15}N HSQC spectra of full-length PfELC (red) and its N-terminal construct PfELC-N (black), indicating that the construct PfELC-N is identical to the N-terminal domain of full-length PfELC. Assigned resonances are labeled.

Supplementary Fig. 3



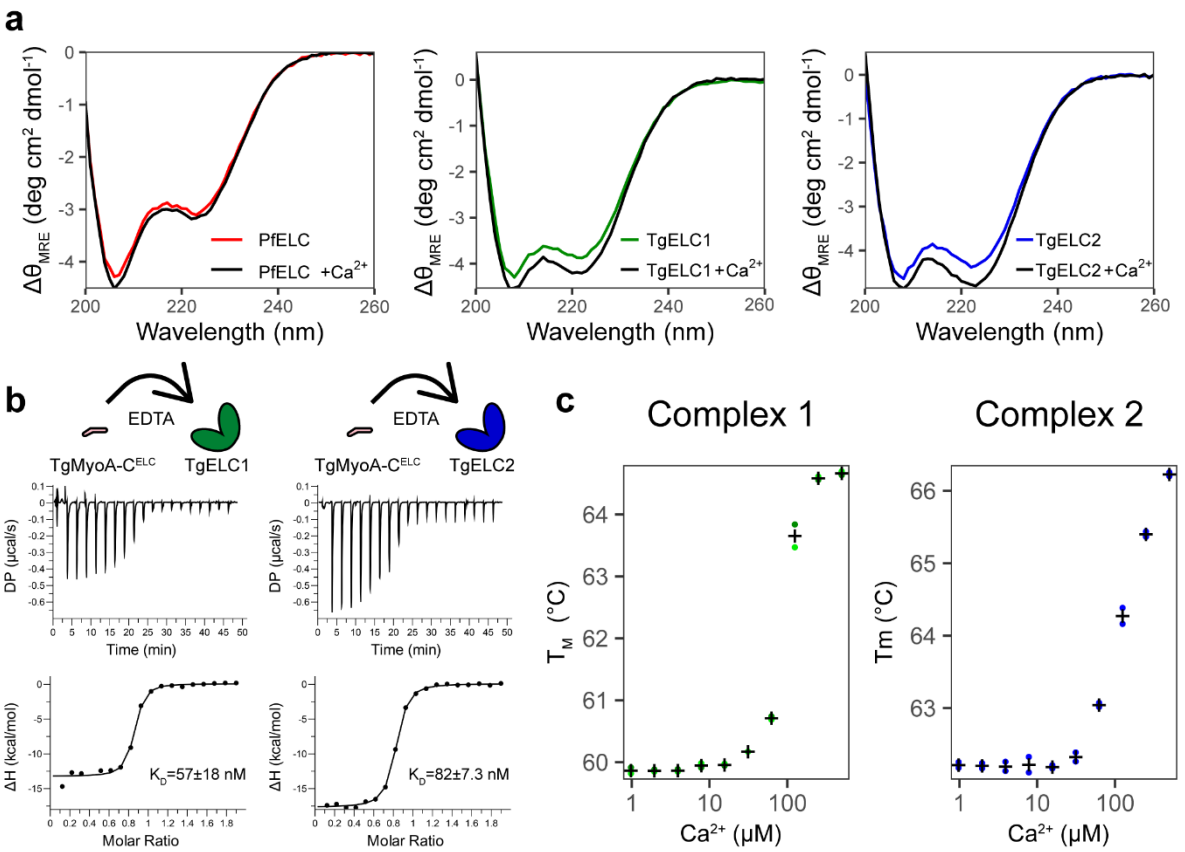
(A) Isothermal titration calorimetry of TgELC2 mutants binding to TgMyoA-C^{ELC}. Individual mutations of polar residues (E10A, F79A, S101A, S102A) of TgELC2 interacting with TgMyoA-C^{ELC} do not cause major changes in the affinity of the two components, but the double mutant TgELC2E10A+H110A shows a substantially lower affinity. (B) Isothermal titration calorimetry of TgELC2 phosphomimetic mutant S102E leads to a small increase of the resulting K_D which suggests that phosphorylation of S102 is unlikely to regulate complex formation. (C) Isothermal titration calorimetry of PfMyoA$^{785-815}$ titrated into MTIP. (D) Isothermal titration calorimetry of PfELC phosphomimetic mutant S127E causes a twofold decrease in K_D. (E) Structural alignment of TgELC1 and PfELC from the crystal structure of their complexes shows that the C-terminal lobe and helix 4 of PfELC display different conformation than TgELC1.

Supplementary Fig. 4



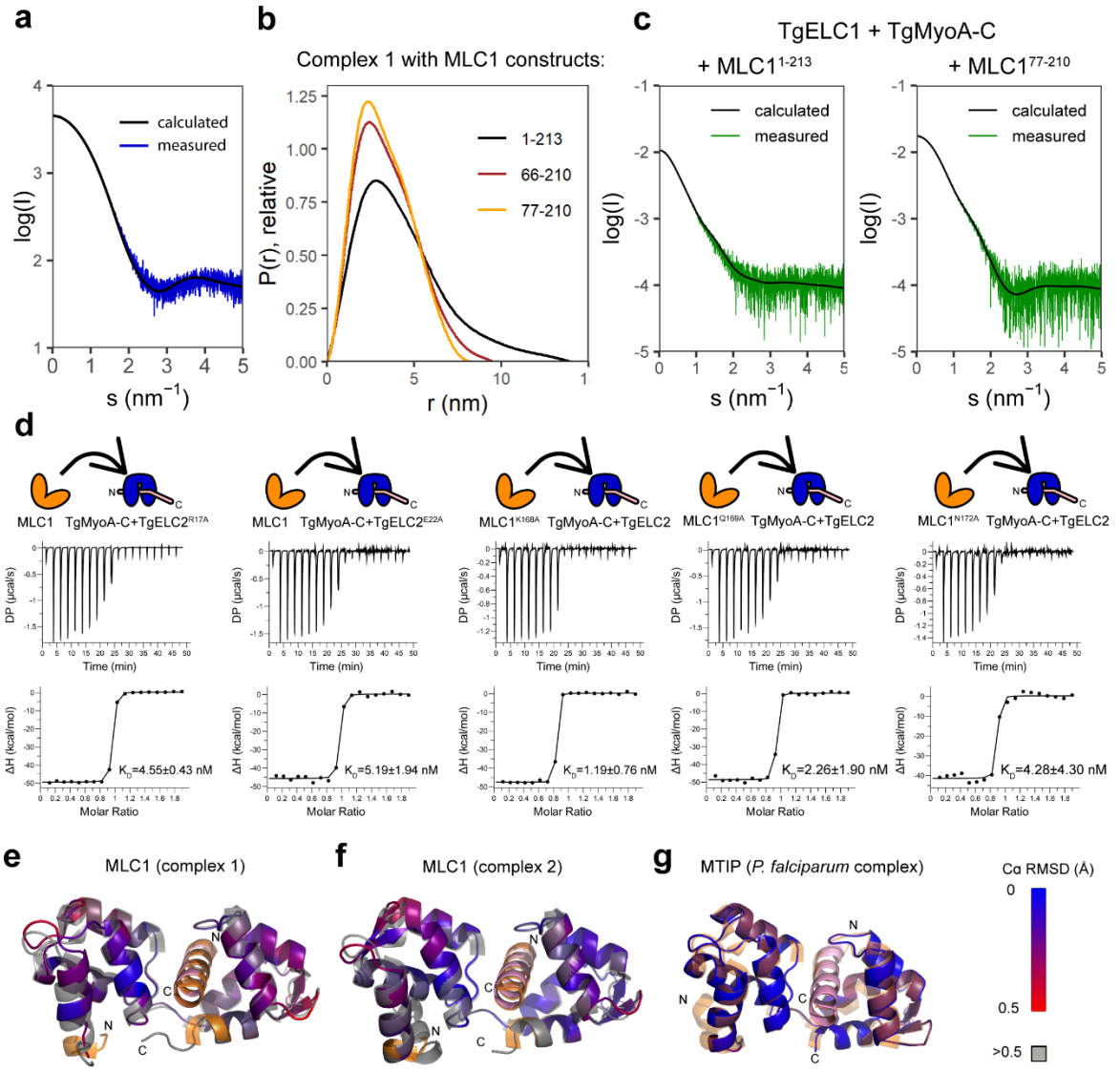
(A) Small Angle X-ray scattering profiles of TgELC2 (blue) and in complex with TgMyoA-C (black) show conformational changes upon interaction. (B) Elution profile of on-line SEC-SAXS measurement of TgELC2 using a Superdex 200 10/300 column with the region used for the analysis highlighted in grey. (C) Far-UV CD spectra of both PfMyoA-C (pink) and TgMyoA-C (violet) indicate that the unbound C-terminus of MyoA is disordered (TgMyoA) or partially disordered (PfMyoA). (D) Far-UV CD data indicate that TgELC1 induces α-helical structure in TgMyoA upon binding. The individual spectra of TgELC1 (green) and TgMyoA^{ELC} (pink) do not sum up (dotted black line) to the spectrum of their dimeric complex (black continuous line), which has more pronounced features of α-helical secondary structure with lower ellipticity at 222 nm and higher ellipticity at 195 nm. (E) The far-UV CD data show that, similarly to TgELCs, the amount of α-helical structure increases upon binding of PfELC to MyoA C-terminus. The data were collected in a 1 mm cuvette at a concentration of 5 µM of each component in 10 mM NaP (pH 7.5), 150 mM NaF and 0.25 mM TCEP at 20°C.

Supplementary Fig. 5



(A) Comparison of far-UV circular dichroism spectra of individual ELC proteins in presence and absence of calcium ions show that calcium does not significantly alter the secondary structure of the ELCs. (B) Binding isotherms of MyoA-C^{ELC} titrated to TgELC1 or TgELC2 in the presence of 5mM EDTA shows that calcium does not have a major influence on the affinity of TgELC2 to the myosin A neck. (C) Stability dependence of the trimeric complex upon addition of increasing concentrations of calcium illustrated by the increase in T_M (°C). Stability data for complex one are shown on the left and for complex two on the right. The colored points are individual measurements and "**+**" represents the average. The experiment shows that the stability of the trimeric complex is greatly enhanced by the addition of calcium in a concentration-dependent manner.

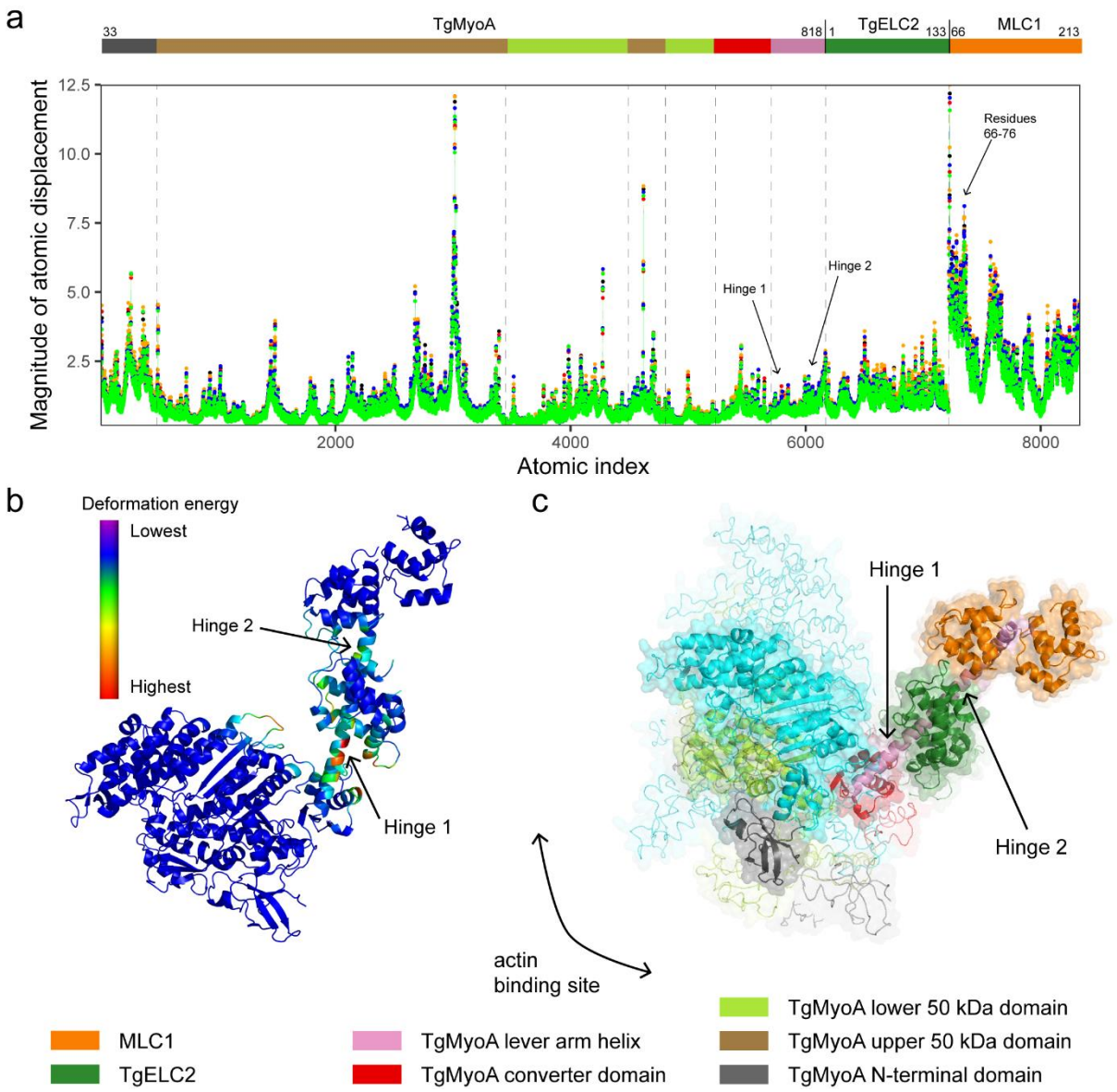
Supplementary Fig. 6



(A) Experimental small angle X-ray scattering curve of TgELC2 bound to TgMyoA-C^{ELC} (blue). The fit to the scattering pattern computed from the crystal structure of complex 2 omitting MLC1 (black line, χ^2 = 1.16) shows that TgELC2 does not undergo major conformational changes upon binding of MLC1 and the formation of the trimeric complex. (B) The distance distribution plots of complex 1 calculated from experimental small angle X-ray scattering data change by shortening the MLC1 N-terminal domain, indicating flexibility of MLC1 upstream of residue 77. The distance distribution is narrower upon N-terminal truncation of MLC1, with d_{max} decreasing from 14 nm (complex 1 with full-length MLC1$^{1-213}$) to 9.5 nm (MLC1$^{66-210}$) and further to 8.2 nm (MLC1$^{77-210}$). (C) Experimental small angle X ray scattering curves of complex 2 with the short MLC1 construct (MLC1$^{77-210}$) and full-length MLC1 (MLC1$^{1-210}$). The calculated scattering curve computed from the crystal structure of complex 1 fits the scattering data of complex 1 with construct MLC1$^{77-210}$ with χ^2 = 1.04, suggesting that MLC1 residues 77-210 form a folded and rigid entity in the complex. The experimental data of complex 1 with full length MLC1$^{1-213}$ fit the calculated scattering data from the crystal structure of complex 1 and N-terminal MLC1 residues modelled by CORAL with χ^2 = 1.15. (D) ITC binding isotherms of MLC1 titrated to the TgELC2/TgMyoA-C pre-complex. TgELC2 or MLC1 residues forming the binding interface were mutated individually and their affinity was measured to assess the contribution to the binding interface within the trimeric complex. The measured mutants were TgELC2 mutants R17A and E22A, and MLC1 mutants K168A, Q169A and N172A. The binding affinities were in the low nanomolar range. (E-G) Overlays of MLC1 or MTIP derived from the published dimeric complex structure in grey (PDB ID 5vt9

and 4aom, orange) with the protein chains of the trimeric complex structures show that MLC1 and MTIP do not undergo any major structural changes upon ELC binding. The dimeric complex structure agrees with a backbone RMSD of 0.96 Å and 0.75 Å with complex 1 and complex 2, respectively and 1.41 Å with *P. falciparum* complex. The key interactions of TgMyoA with MLC1 are also conserved between the structures of the trimeric and dimeric complexes (R808, H812, R814) with the exception of few weak polar interactions (see Supplementary Table 6). Color code of MLC1 derived from the trimeric complexes according to RMSD deviation of Cα is indicated.

Supplementary Fig. 7



(A) Summary plot of the atomic displacements predicted by NMA based on the five lowest-energy models for complex 2 selected by the lowest clash score. The relative atomic displacement of the individual amino acid residues follows the same pattern in all five models, confirming that the results of the normal mode analysis are independent of the chosen starting conformation or the energy-minimized model. The results for complex 1 are similar (data not shown). (B) The deformation analysis of complex 1 averaged through the 20 lowest-energy modes predicts two main hinge regions, with the hinge 1 (residues 773-777), having the largest contribution to the observed motions. The model is coloured by the deformation energy from low (violet) to high (red). (C) The ensemble of the structures of complex 2 based on the two lowest-energy modes, which contribute most to the large-scale dynamics of proteins. The original model is drawn in cartoon representation with shown semitransparent surface, whereas the deformed structures are partially transparent and drawn in ribbon representation with faded surface. The structures were aligned on MLC1 to reflect the immobilization of MLC1 in the IMC membrane as in the current model of the glideosome (see Fig. 1a).

Supplementary Table 1. A list of published *P. falciparum* and *T. gondii* glideosome protein structures. So far, only structures of individual proteins of the glideosome and two homologous sub-complexes (MTIP/PfMyoA and MLC1/TgMyoA) have been determined.

Organism	**Protein**	**Residues**	**PDB ID**	**Year**	**Ref.**
P. falciparum	GAP50	24-365	3tgh	2012	25
P. falciparum	MyoA	2-768	6i7d, 6i7e	2019	28
T. gondii	MyoA	33-778	6due	2018	27
P. falciparum	MTIP+MyoA	60-204, 799-816	4aom	2012	26
T. gondii	MLC1+MyoA	66-210, 801-831	5vt9	2017	15

Supplementary Table 2. Biophysical characterization and comparison of PfELC and TgELC2.

	SEC data	**CD data**			**SAXS data**		
	Elution volume (ml)	**α helix**	**β sheet**	**random**	***Rg* (nm)**	**MW (kDa)**	**D_{max} (nm)**
TgELC2	2.41	41%	15%	45%	2.15	17.0	6.73
PfELC	2.18	34%	16%	51%	2.83	16.3	9.50

Supplementary Table 3. SAXS sample details, data acquisition parameters, structural parameters and atomistic modelling.

Sample details									
Sample	**PfELC-N**	**PfELC**	**TgELC2**	**TgELC2 + TgMyoA-C^{ELC}**	**Complex 1 (MLC1$^{1-213}$)**	**Complex 1 (MLC1$^{66-210}$)**	**Complex 1 (MLC1$^{77-210}$)**	**Complex 2 (MLC1$^{66-210}$)**	**Complex of *P. falciparum***
Organism	*P. falciparum*	*P. falciparum*	*T. gondii*	*T. gondii*	*T. gondii*	*T. gondii*	*T. gondii*	*T. gondii*	*P. falciparum*
Source	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21	*E. coli* BL21
UniProt ID	Q8IJM4	Q8IJM4	B9PZ33	B9PZ33 + S8G527	XYZ* + S8G527 + Q95UJ7	XYZ* + S8G527 + Q95UJ7	XYZ* + S8G527 + Q95UJ7	B9PZ33 + S8G527 + Q95UJ7	Q8IJM4 + Q8IDR3 + Q8I4Q8
Extinction coefficient ε (at 280 nm, M^{-1} cm^{-1})	11460	8480	16760	25240	49850	35870	35870	30370	35870
Molecular weight from chemical composition (Da)	9011.2	15763.9	15471.5	20681.34	44183.1	37181.1	36619.5	37785.6	37459.5
Concentration (analysis or injection, mg/ml)	3.75	10	5	10	3.9	5.6	4.8	15.5	8.2
Solvent composition	20mM HEPES pH 7.5, 150mM NaCl, 0.5mM TCEP								
SAS data collection parameters									
Beamline	P12, DESY/EMBL, Hamburg (Germany)								
Detector	Pilatus 6M								
Energy (kEV)	10.0								
Sample-to-detector distance (mm)	3000								
q-measurement range (Å-^{10})	0.003 – 0.732								
Absolute scaling method	Relative to the scattering of pure water								
Method for monitoring radiation damage	Frame comparison								
Exposed time for frame	1 s (20 x 0.05 s)	900 s (900 x 1 s)	1 s (20 x 0.05 s)	1500 s (1500 x 1 s)	1500 s (1500 x 1 s)	1500 s (1500 x 1 s)	1500 s (1500 x 1 s)	1500 s (1500 x 1 s)	1500 s (1500 x 1 s)
Mode	Batch	SEC-SAXS	Batch	SEC-SAXS	SEC-SAXS	SEC-SAXS	SEC-SAXS	SEC-SAXS	SEC-SAXS
Sample temperature (°C)	20	20	20	20	20	20	20	20	20
Structural parameters									
Guinier Analysis									
I(0) (cm^{-1})	0.007±0.001	0.007±0.001	0.013±0.001	0.021±0.001	0.011±0.001	0.029±0.001	0.017±0.001	0.050±0.001	0.037±0.001
R_g (Å)	13.6±1.9	27.2±1.7	21.4±0.5	17.3±0.1	32.4±0.2	26.7±3.0	25.2±4.0	26.7±3.2	27.4±1.8
q-range ($Å^{-1}$)	0.01-0.09	0.01-0.05	0.01-0.06	0.01-0.07	0.01-0.04	0.01-0.05	0.01-0.05	0.01-0.05	0.01-0.05
Fidelity (Quality of fit parameter, *AutoRg*)	0.96	0.93	0.97	0.97	0.93	0.96	0.92	0.98	0.96
Molecular weight (Da)**									
Rel. to standard (BSA, 66 kDa)	9762	n.d.	16824	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
From V_c	8486	16299	16991	16223	38396	34547	32210	35435	35029
From *MoW*	7311	15636	16700	14536	27615	35069	32610	33378	36809
P(r) analysis									

I(0) (cm^{-1})	0.007±0.001	0.007±0.001	0.013±0.001	0.021±0.001	0.011±0.001	0.029±0.001	0.017±0.001	0.050±0.001	0.038±0.001
R_g (Å)	13.7±0.01	28.3±0.42	21.5±0.05	17.3±0.01	35.0±0.24	27.2±0.05	25.5±0.05	27.4±0.04	28.4±0.05
d_{max} (Å)	43.0	95.0	67.3	54.9	140	95.0	82.0	100	107
q-range ($Å^{-1}$)	0.01-0.59	0.01-0.30	0.01-0.37	0.01-0.46	0.01-0.25	0.01-0.30	0.01-0.32	0.01-0.30	0.01-0.29
Total quality estimate	0.85	0.52	0.80	0.86	0.71	0.84	0.91	0.79	0.78
Porod volume x 10^3 ($Å^3$)	11.80	22.81	26.72	27.59	63.74	49.68	47.03	49.92	50.76
Atomistic modelling	*CRYSOL* with constant subtraction and maximum order of harmonics equal to 50								
Crystal structures	6tj4, chain A, residues 1-68			6tj7, chains A+C (res. 777-798)		6tj6, chains A+B+C	6tj6, chains A+B$^{(res. 81-214)}$+C	6tj7, chains A+B+C	XXX, chains A+B+C
q-range for modelling	0.01-0.50			0.01-0.50		0.01-0.50	0.01-0.50	0.01-0.50	0.01-0.50
χ^2, P-value	1.37, 0.00			1.16, 0.00		1.26, 0.00	1.04, 0.20	2.41, 0.00	3.94, 0.00
Predicted R_g (Å)	13.19			17.31		6.42	25.67	26.26	25.80
Vol (Å), Ra (Å), Dro (e $Å^{-3}$)	10436, 1.760, 0.030			22222, 1.780, 0.025		42336, 1.800, 0.065	46310, 1.400, 0.022	46306, 1.800, 0.050	45540, 1.800, 0.055
	CORAL hybrid rigid body modeling								
Starting crystal structures					6tj6, chains A+B$^{81-210}$+C				
Flexible residues					1-80 of chain B				
q-range for modelling					0.01-0.58				
χ^2, P-value					1.15, 0.00				
SASBDB IDs for data and models									
	SASDH64	SASDH74	SASDH84	SASDH94	SASDHA4	SASDHB4	SASDHC4	SASDHD4	SASDHE4

* the protein with accession number *TGME49_069440* in ToxoDB database has been, as our data show, incorrectly split to two different genes and is not available in the current versions of database.

** calculated from BSA standard for batch mode samples, and V_c and *MoW* volume estimates for SEC-SAXS samples, using the appropriate functions in ATSAS

Supplementary methods

Oligos used for cloning

Plasmodium falciparum MTIP primers for LIC cloning into pNIC28_Bsa4 vector

MTIP_FL_N28_Fw	TACTTCCAATCCATGAAACAAGAATGCAATGTATGTTATTTT
MTIP_FL_N28_Rv	TATCCACCTTTACTGTTATTGTAATATATCTTCACAGAATAATTTGT

Plasmodium falciparum MTIP-S primers for Slice cloning into pNIC28_Bsa4 vector

MTIPTr_Nt_SliceFW	ggtgtagatctgggtaccgagaacctgtacttccaatccatgGAATCAGTTGCTGACATA
MTIPTr_NtSliceRv	tgtcgacggagctcgaattcggatccgtatccacctttactgTTATTGTAATATATCTTCACAGAA

Plasmodium falciparum ELC primers for Slice cloning into pNIC28_Bsa4 vector

10175_FL_Nt_Fw	ggtgtagatctgggtaccgagaacctgtacttccaatccatgATGGCATCTGATATGG
10175_FL_Nt_Rv	tgtcgacggagctcgaattcggatccgtatccacctttactgTTATATCGATTCCGTTAA

Plasmodium falciparum ELC-N primers for LIC cloning into pNIC28_Bsa4 vector

pNIC28_ELC_LIC_Fw	TACTTCCAATCCATGGCATCTGATATGGAAGAAAAATTTAGAGA
pNIC28_ELC_LIC_N74_Rv	TATCCACCTTTACTGTTAATTAATATGTTTAAATGGTTGTTCATAGTTGTAGTG

Plasmodium falciparum PfMyoA-C primers for restriction cloning into pETM11_SUMO3 vector

PfMyoA_SUMO_Slice_FW	ttccagcaacagaccggtggatccGTTGAATGGGAAAATTGTGTGAGT
PfMyoA_SUMO_Slice_RV	gtgctcgagtgcggccgcaagcttTTATACCATTTTTTTTCTTATATGAGC

Toxoplasma gondii TgMyoA-C primers for restriction cloning into pET_GB-1a vector

TgMyoA_pep1_FW	AAAACCATGGCTTCTTCTTGGGAGCCTCTCGTCTCAGTGCTCGAGGCGTACTACGCTGGCAGACGCCACAAGAAGCAGCTGC
TgMyoA_pep1_RV	AAAAGGATCCTTACACCAGGTGTCTGCGGATGTGAGCCTGGGCGCGAATGATGAAGGGGGTCTTTTTCAGCAGCTGCTTCTTG

Toxoplasma gondii TgMLC1 short constructs primers for_LIC cloning into pNIC_CTHF vector

pNIC_CTHF_TgMLC1_66_FW	TTAAGAAGGAGATATACTATGGCAGACGAAGACATGCAG
pNIC_CTHF_TgMLC1_77_FW	TTAAGAAGGAGATATACTATGGTGGAGGCCGACGAAATG
pNIC_CTHF_TgMLC1_210_RV	GATTGGAAGTAGAGGTTCTCTGCCTCGAGCATTGCCTTGC

Plasmodium falciparum PfELC primers to create mutation S127D by blunt end PCR

PfELC_S127D_BEfw	TCAAAAATTAACGGAATCGATATAACagta
PfELC_S127D_BErv	TCTACTGTCTTAAGGTTTAAGGTTGATTTAT

Toxoplasma gondii TgELC1 primers for extending the purchased gene

ELC1-part1_FW	TACTTCCAATCCATGACCTGCCCGCCGCGTGTTCGTGAAGCGTTCGCGCTGTTCGACACCGAC
ELC1-part2_RV	CCGCAAGAACGGATCGCCAGAACCAGGTCACGACCAGAGATTTCACCGTCACCGTCGGTGTCGAACAGCG
ELC1-part3_FW	GCGATCCGTTCTTGCGGTGTTTCTCCGACCCCGGACGAAATCAAAGCGCTGCCGatgTCAATGGCGTGGCC

Toxoplasma gondii TgELC1 primers for Slice cloning into pNIC28_Bsa4

TgELC1_Slice_FW	ggtgtagatctgggtaccgagaacctgtacttccaatccATGACCTGCCCGCCGCG
TgELC1_Slice_RV	tgtcgacggagctcgaattcggatccgtatccacctttactgTTATTTCAGCAGCATCTTGACAAAG

Toxoplasma gondii TgELC2 primers for single amino acid mutations by blunt-end PCR

TgELC2_E10A_BErv	GCGCGCGCTTTTTGTTCGAC
TgELC2_E10A_BEfw	GGCATTCAAGCTTTTCGATCGC
TgELC2_R17A_BErv	GCATCGAAAAGCTTGAATGCCTCG
TgELC2_R17A_BEfw	GAATGGTGATGGCGAGTTAACG
TgELC2_E22A_BErv	GCGCCATCACCATTGCGA
TgELC2_E22A_BEfw	GTTAACGCATCAAGAAGCTGTCC
TgELC2_F79A_BErv	GCAGGCGCAAACAGTTTGATCAG
TgELC2_F79A_BEfw	GGATCGCAAAAATGATGGCACGA
TgELC2_D80A_BErv	GCGAAAGGCGCAAACAGTTTGATCAG
TgELC2_D80A_BEfw	GCGCAAAAATGATGGCACGATC
TgELC2_G101A_BErv	GCGCAGAGAGTCTTCATGACTTGAGC
TgELC2_G101A_BEfw	GTCTATGACGGAGGAGGACATCG
TgELC2_H110A_BErv	GCATCGATGTCCTCCTCCGTCA
TgELC2_H110A_BEfw	GCTCATTAAACAAGCGGATCCAAACAAC

Toxoplasma gondii MLC1 primers for single amino acid mutations by blunt-end PCR

TgMLC1_K168A_BEfw	GCAGATGGGGAACATCCTCA
TgMLC1_K168A_BErv	GCGCGCGTCAGGTAACCG
TgMLC1_Q169A_BEfw	GATGGGGAACATCCTCATGACC
TgMLC1_Q169A_BErv	GCCTTGCGCGTCAGGTAACCG
TgMLC1_N172A_BEfw	GATCCTCATGACCTACGGAGAGC
TgMLC1_N172A_BErv	GCCCCCATCTGCTTGCGCG

Primers for re-cloning gene cassettes into pPYC vector

NdeI_to_pPYC_FW	AAAAcatatgcaccatcatcatcatcattc
XbaI_to_pPYC_RV	AAAAtctagacgacggagctcgaattcgga

6. Manuscript 2

Structural Insights Into PfARO and Characterization of its Interaction With PfAIP

Michael Geiger[1,2,5], Chris Brown[3], Jan Stephan Wichers[1,2,5], Jan Strauss[1,2,5], Andrés Lill[1,5], Roland Thuenauer[1,5], Benjamin Liffner[4], Louisa Wilcke[1,2], Sarah Lemcke[1,2,5], Dorothee Heincke[1,2,5], Samuel Pazicky[1,7], Anna Bachmann[1,2,5], Christian Löw[1,7], Danny William Wilson[4,6], Michael Filarsky[1,5], Paul-Christian Burda[1,2,5], Kun Zhang[3], Murray Junop[3], Tim Wolf Gilberger[1,2,5]

[1] Centre for Structural Systems Biology, Notkestraße 85, 22607, Hamburg, Germany

[2] Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Strasse 74, 20359, Hamburg, Germany

[3] Western University, Department of Biochemistry, London, ON, Canada

[4] Research Centre for Infectious Diseases, School of Biological Sciences, University of Adelaide, Adelaide, Australia

[5] Department of Biology, University of Hamburg, Hamburg, Germany

[6] Burnet Institute, 85 Commercial Road, Melbourne, 3004, Victoria, Australia

[7] Molecular Biology Laboratory (EMBL), Hamburg Unit c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, 22607, Hamburg, Germany

Structural Insights Into *Pf*ARO and Characterization of its Interaction With *Pf*AIP

Michael Geiger[1,2,5,†], Chris Brown[3,†], Jan Stephan Wichers[1,2,5], Jan Strauss[1,2,5], Andrés Lill[1,5], Roland Thuenauer[1,5], Benjamin Liffner[4], Louisa Wilcke[1,2], Sarah Lemcke[1,2,5], Dorothee Heincke[1,2,5], Samuel Pazicky[1,7], Anna Bachmann[1,2,5], Christian Löw[1,7], Danny William Wilson[4,6], Michael Filarsky[1,5], Paul-Christian Burda[1,2,5], Kun Zhang[3], Murray Junop[3,] and Tim Wolf Gilberger[1,2,5]**

1 - Centre for Structural Systems Biology, *Notkestraße 85, 22607, Hamburg, Germany*
2 - Bernhard Nocht Institute for Tropical Medicine, *Bernhard-Nocht-Strasse 74, 20359, Hamburg, Germany*
3 - Western University, *Department of Biochemistry, London, ON, Canada*
4 - Research Centre for Infectious Diseases, *School of Biological Sciences, University of Adelaide, Adelaide, Australia*
5 - Department of Biology, *University of Hamburg, Hamburg, Germany*
6 - Burnet Institute, *85 Commercial Road, Melbourne, 3004, Victoria, Australia*
7 - Molecular Biology Laboratory (EMBL), *Hamburg Unit c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, 22607, Hamburg, Germany*

Correspondence to Tim Wolf Gilberger and Murray Junop: *Centre for Structural Systems Biology, Hamburg, Germany. Fax: +49 40 42818 600; Western University, Department of Biochemistry, London, ON, Canada.*
mjunop@uwo.ca, gilberger@bnitm.de
https://doi.org/10.1016/j.jmb.2019.12.024
Edited by Shao Feng

Abstract

Apicomplexan parasites contain rhoptries, which are specialized secretory organelles that coordinate host cell invasion. During the process of invasion, rhoptries secrete their contents to facilitate interaction with, and entry into, the host cell. Here, we report the crystal structure of the rhoptry protein Armadillo Repeats-Only (ARO) from the human malaria parasite, *Plasmodium falciparum* (*PfARO*). The structure of *Pf*ARO comprises five tandem Armadillo-like (ARM) repeats, with adjacent ARM repeats stacked in a head-to-tail orientation resulting in *Pf*ARO adopting an elongated curved shape. Interestingly, the concave face of *Pf*ARO contains two distinct patches of highly conserved residues that appear to play an important role in protein-protein interaction. We functionally characterized the *P. falciparum* homolog of ARO interacting protein (*Pf*AIP) and demonstrate that it localizes to the rhoptries. We show that conditional mislocalization of *Pf*AIP leads to deficient red blood cell invasion. Guided by the structure, we identified mutations of *Pf*ARO that lead to mislocalization of *Pf*AIP. Using proximity-based biotinylation we probe into *Pf*AIP interacting proteins.

Introduction

Apicomplexa are a phylum of single-cell parasitic organisms that include the major human pathogens *Plasmodium*, *Toxoplasma*, and *Cryptosporidium*. Malaria, caused by a number of *Plasmodium* spp., infects ~200 million people and causes over 400,000 deaths annually [1]. *Plasmodium falciparum* is responsible for the majority of malaria-related mortality. Clinical symptoms of malaria are caused by 48 h cycles of invasion, growth, and replication of the parasite within human red blood cells (RBCs), and subsequent RBC rupture [2]. RBC invasion, by a parasite stage called the merozoite, represents an attractive therapeutic target as the merozoite is free from its host cell and directly exposed to the host immune system. Invasion is coordinated by specialized merozoite secretory organelles including the

micronemes, rhoptries, and dense granules that secrete their contents to facilitate interaction with and invasion into RBCs [3]. Rhoptries are the largest of these invasive organelles and are Golgi-derived dual-club-shaped organelles located at the apical pole of merozoites [4]. Despite the important role they play in merozoite invasion, rhoptry biogenesis is poorly understood. Previous studies implicate the adapter protein 1 complex (AP1) [5,6] and the dynamin-related protein B (DrpB) in the biogenesis of secretory organelles in the related apicomplexan parasite *T. gondii* [7]. Studies in both *Plasmodium* and *Toxoplasma* localized Rab11A to the rhoptry membranes [8,9]. For the small GTPases Rab5a and Rab5c, it was shown that depletion of these proteins leads to development of parasites that lack rhoptries [10]. In addition to these classical, molecular membrane trafficking switches, an Apicomplexa-specific armadillo (ARM) repeats containing protein was identified in *P. falciparum* [11] that localizes to the cytosolic side of the rhoptries and is named Armadillo Repeats-Only (*Pf*ARO, PF3D7_0414900) [12]. This 31 kDa protein relies on both myristoylation and palmitoylation motifs for membrane attachment [12]. Subsequent studies in *T. gondii* showed that *Tg*ARO is essential for the correct positioning of the nascent rhoptries at the apical pole of parasites [13,14]. Additionally, these studies demonstrated that *Tg*ARO interacts with an 82 kDa protein termed ARO interacting protein (*Tg*AIP); the key regulatory enzyme adenylate cyclase beta (*Tg*ACβ); and the motor protein myosin F (*Tg*MyoF). Knockout studies of *Tg*AIP showed that in the absence of *Tg*AIP, *Tg*ACβ is unstable and no longer recruited to the rhoptries, whereas *Tg*ARO localization is unchanged. The authors concluded that *Tg*ARO likely interacts directly with *Tg*AIP and that *Tg*AIP recruits *Tg*ACβ to the rhoptry neck [15]. Whether these interactions and their functional consequences also hold true for the *P. falciparum* homologs of these proteins is currently unknown.

Studies in *T. gondii* have analyzed the structure of ARO using small-angle X-ray scattering (SAXS). Homology modeling of *Tg*ARO suggested it exists as a globular monomeric protein with a maximum intramolecular distance of ~9 nm, containing at least five ARM repeats and demonstrating similarity to a *Caenorhabditis elegans* myosin chaperone [13]. However, a crystal structure for *Tg*ARO or *Pf*ARO has not been determined.

Here, we report the crystal structure of *Pf*ARO to 1.8 Å resolution as well as identify and localize the *Pf*ARO interacting protein (*Pf*AIP) in *P. falciparum.* We show that mutations within the loop 1 region of *Pf*ARO lead to a mislocalization of *Pf*AIP and that conditional loss of *Pf*AIP from the rhoptries results in an inhibition of merozoite invasion. Using proximity-based biotinylation, we probe into *Pf*AIP interacting proteins.

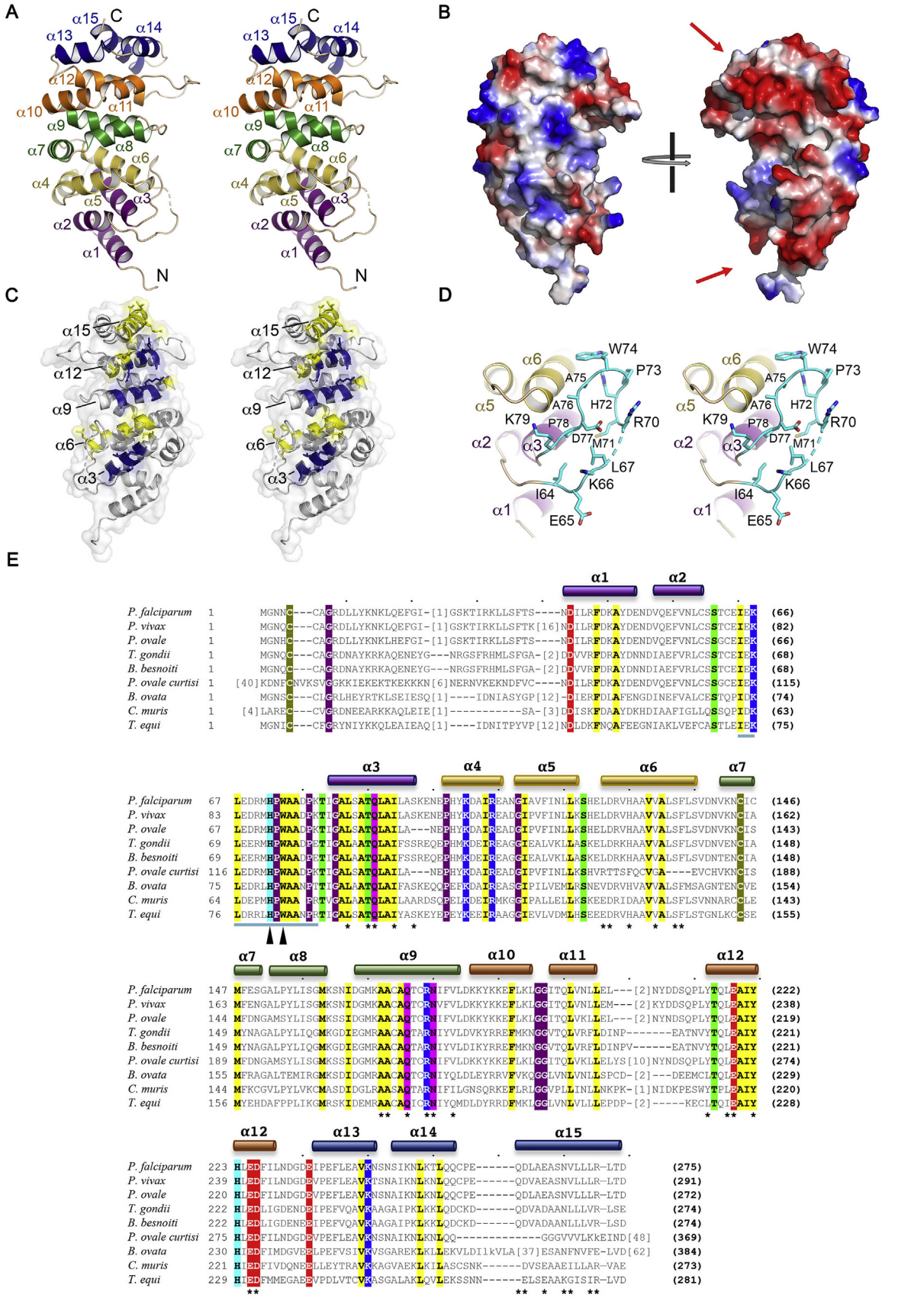
Results

The crystal structure of PfARO

*Pf*ARO is predicted to contain up to 6 ARM repeats spanning residues 9–274 [14]. However, based on partial proteolysis studies the first ~20 residues of *Pf*ARO appeared unstructured (Sup. Fig. S1). We therefore cloned and expressed *Pf*ARO$_{23\text{-}275}$ for structural studies. The structure of *Pf*ARO$_{23\text{-}275}$ was solved by single-wavelength anomalous dispersion phasing (using a Selenomethione-derivatized protein) and refined against a native data set to 1.8 Å. *Pf*ARO crystallized in space group P1 and contained a dimer (RMSD 0.19 Å) within the unit cell. ARO exists as a monomer in solution [13], the observed dimer can therefore be attributed to crystallographic packing. The final structure of *Pf*ARO spanned amino acid residues 32–274. Two residues (E68 and D69) within an extended loop (loop 1), as well as residues at the N- (23–31) and C-termini (D275), were disordered and could not be modeled into electron density maps. The final model was refined to R and R_{free} values of 16.9% and 19.2%, respectively. A complete list of X-ray diffraction data and model refinement statistics are provided in Table S1.

The *Pf*ARO$_{23\text{-}275}$ monomer (Fig. 1A) comprises 5 ARM-like repeats, each containing three helices. Although the first ARM repeat (Fig. 1A, α1-α3, purple) is similar to the other four ARM repeats in terms of fold, it differs in the relative positioning of its first helix (α1) and is therefore a somewhat atypical or "degenerate" ARM repeat. It is possible that removal of residues 1–22 caused a small change in the position of helix 1, but the first 31 amino acids of the N-terminus are not required for formation of α1 *per se*.

Similar to what has been observed in other ARM-containing proteins such as β-Catenin (PDB 2I22), helices from adjacent ARM repeats of *Pf*ARO stack in a head-to-tail fashion resulting in an elongated right-handed superhelix. As shown in Fig. 1B, *Pf*ARO adopts an overall shape that resembles a kidney bean with the concave surface formed by the last helix from each ARM repeat (α3, α6, α9, α12, α15). Similar to ARM repeat proteins such as importin α7 (PDB 4UAD, 6N88), the concave surface of *Pf*ARO has been suggested to potentially function as an interaction surface for *Pf*ARO binding partners [15]. In addition to its compelling shape, the surface has a significant negative charge (Fig. 1B) that might help mediate interaction. Electrostatic surface potential of *Pf*ARO is not evenly distributed. While the front face (Fig. 1B, left) of *Pf*ARO is slightly positive, the opposing back face is almost entirely covered with negative charge, suggesting the back face may be well suited for interaction with a positively charged


A
B
C
D
E
α1
α2
α3
α4
α5
α6
α7
α8
α9
α10
α11
α12
α13
α14
α15
C
N
W74
P73
A75
A76
H72
R70
K79
P78
D77
M71
L67
K66
I64
E65
P. falciparum
P. vivax
P. ovale
T. gondii
B. besnoiti
P. ovale curtisi
B. ovata
C. muris
T. equi

protein or helping to orient *Pf*ARO relative to a negatively charged surface. Despite these favorable features, the concave surface is composed of residues that are generally not highly conserved (Fig. 1E). Exceptions occur for several surface exposed residues found on α3, α9, and α12 of the concave surface (labeled with asterisks in Fig. 1E). As shown in Fig. 1C, residues from α9 and α12 form a continuous surface suggesting important functions potentially involving protein interaction.

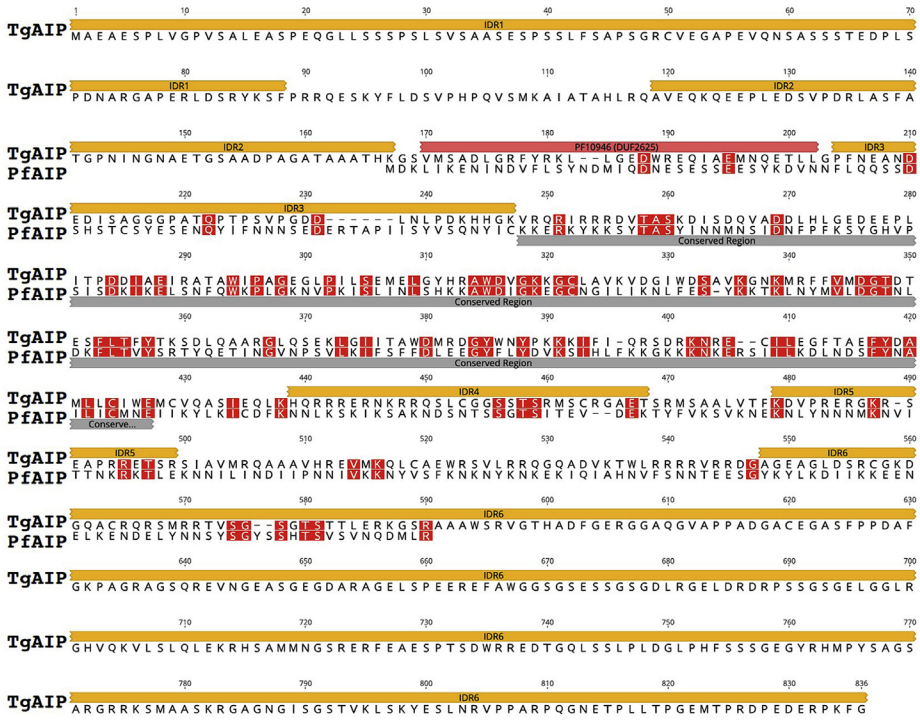
Perhaps the most notable feature of the *Pf*ARO structure is the presence of two loops inserted between α2-α3 of ARM1 and α11-α12 of ARM4 (Fig. 1A). Both loops protrude from the same surface; however, loop 1 (residue 59−80) is considerably larger than loop 2 (residue 203−214) and contains a surprisingly large number of highly conserved residues (Fig. 1E). In fact, loop 1 and the adjacent α3 represent the most highly conserved region of *Pf*ARO (Fig. 1E). Residues from the apex of loop 1 (residues 71−78) extend toward α3 forming a continuous surface of highly conserved residues between these elements in three-dimensional space (Fig. 1D). As such, this region is expected to be important for *Pf*ARO function by mediating interaction with binding partners.

Identification and phylogenetic analysis of PfARO interacting protein homolog in *P. falciparum*

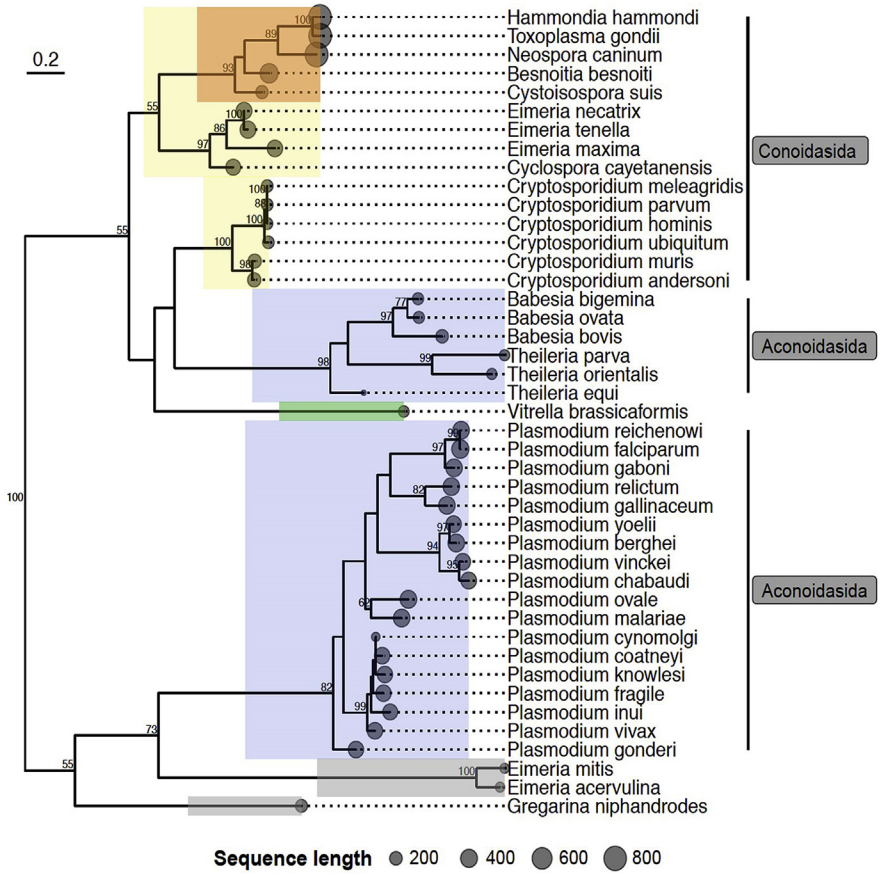
The *T. gondii* homolog of ARO was shown to interact with *Tg*AIP (TGME49_309190) [15]. Using the blastp function of the PlasmoDB database (www.plasmoDB.org) [16] we retrieved PF3D7_1136700 as a putative AIP homolog in the *P. falciparum* genome. The *Pfaip* gene contains 7 exons and a coding sequence of 1266 bp that is translated into 421 amino acids. The predicted molecular mass of *Pf*AIP is 49.1 kDa and is therefore significantly smaller than *Tg*AIP, which contains 822 amino acids and has a molecular mass of 89.3 kDa. AIP protein sequence alignments identified a conserved core region that is free of intrinsically disordered regions that are predicted for *Tg*AIP (Fig. 2A). Additionally, except for a Pfam domain of unknown function (PF10946) that is predicted for *Tg*AIP, no conserved domains could be identified for both *Tg*AIP and *Pf*AIP sequences (Fig. 2A). Because the initially identified homolog AIP sequences appeared to be restricted to the phylogenetic group of apicomplexan parasites, we investigated their phylogenetic relationship in more detail. An additional blastp search was performed against the nr database [17] retrieving 89 unique hits exclusively within the phylogenetic clade of alveolates including apicomplexan parasites and the nonparasitic photosynthetic chromerid *Vitrella brassicaformis*. The number of putative homologous AIP sequences could be extended to 99 unique hits based on batch blastp searches of both the *Tg*AIP (TGME_309190) and *Pf*AIP (PF3D7_1136700) sequences after combining results. Our phylogenetic analysis recovered AIP sequences in all sequenced apicomplexan protist genomes to date except for *Theileria annulata*. The reconstructed phylogenetic tree based on 43 representative sequences (Supp. Table 3) revealed well-supported separate clades for most apicomplexan genera (Fig. 2B), except likely misplaced apicomplexan taxa *Eimeria mitis*, *Eimeria acervulina* and *Gregarina niphandrodes* and the chromerid *Vitrella brassicaformis* (Fig. 2B; highlighted in gray and green, respectively), for which only short sequences (108−181 aa) were identified. In contrast, the longest AIP sequences were found in the family of Sarcostidae (*Hammondia hammondi*, *Toxoplasma gondii*, *Neospora caninum*, *Besnoita besnoiti*, *Cystoisospora suis*) with the longest sequence length of up to 829 amino acids (*N. caninum*). In comparison, the length of AIP sequences recovered from *Plasmodium* spp. was around half the length (~400 amino acids). Interestingly, not considering the 4 likely misplaced single taxa with short sequences and low bootstrap support, the overall

Fig. 1. Structure of *Pf*ARO. (**A**) Stereo view of ARO structure. Each ARM-like domain is colored separately (ARM1, purple; ARM2, gold; ARM3, green; ARM4, orange; ARM5, blue). Individual helices are labeled in sequential order starting with α1. N and C indicate residues 32 and 274, respectively. (**B**) Surface electrostatic map of *Pf*ARO. Front and back views of *Pf*ARO are presented. The front view (left) is oriented identical to structure shown in (A). Red, blue, and gray colors represent negative, positive and neutral electrostatic potential, respectively. (**C**) *Pf*ARO concave surface. Helices 3, 6, 9, 12, and 15 are shown lining the ARO concave surface. Conserved surface exposed residues, dark blue; nonconserved surface exposed residues, yellow. (**D**) Close-up view of the loop 1 region of *Pf*ARO. ARM domains are colored as in (A). The portion of loop 1 deleted (residues 64−79) for functional studies (cyan) have individual residues displayed. (**E**) Sequence alignment of *Pf*ARO proteins. ARO sequences (*Plasmodium falciparium* 3D7*, Plasmodium vivax, Plasmodium ovale, Toxoplasma gondii* ME49*, Besnoitia besnoiti, Plasmodium ovale curtisi, Babesia ovata, Cryptosporidium muris* RN66*, Theileria equi* WA) were aligned using COBALT [63]. Individual helices are numbered according to the structure of ARO. Helices forming ARM-like domains are colored as in Fig. 1. Residues that are absolutely conserved are highlighted (hydrophobic, yellow; positive charge, blue; negative charge, red; G/P, purple; S/T, green; C, beige). The portion of the loop between α1 and α2 that was deleted for functional studies is underlined in cyan. Black arrows mark residues altered during functional studies. Surface exposed residues lining the concave surface composed of α3, α6, α9, α12, and α15 are marked with asterisks.

A


TgAIP
PfAIP
IDR1
IDR2
IDR3
IDR4
IDR5
IDR6
PF10946 (DUF2625)
Conserved Region

B


0.2
Hammondia hammondi
Toxoplasma gondii
Neospora caninum
Besnoitia besnoiti
Cystoisospora suis
Eimeria necatrix
Eimeria tenella
Eimeria maxima
Cyclospora cayetanensis
Cryptosporidium meleagridis
Cryptosporidium parvum
Cryptosporidium hominis
Cryptosporidium ubiquitum
Cryptosporidium muris
Cryptosporidium andersoni
Babesia bigemina
Babesia ovata
Babesia bovis
Theileria parva
Theileria orientalis
Theileria equi
Vitrella brassicaformis
Plasmodium reichenowi
Plasmodium falciparum
Plasmodium gaboni
Plasmodium relictum
Plasmodium gallinaceum
Plasmodium yoelii
Plasmodium berghei
Plasmodium vinckei
Plasmodium chabaudi
Plasmodium ovale
Plasmodium malariae
Plasmodium cynomolgi
Plasmodium coatneyi
Plasmodium knowlesi
Plasmodium fragile
Plasmodium inui
Plasmodium vivax
Plasmodium gonderi
Eimeria mitis
Eimeria acervulina
Gregarina niphandrodes
Conoidasida
Aconoidasida
Aconoidasida
Sequence length
200
400
600
800

pattern of the tree revealed a clear distinction between sequences of the two apicomplexan classes Conoidasida [18] and Aconoidasida [19] (Fig. 2B).

Localization and functional characterization of PfAIP in *P. falciparum*

*Pf*AIP shares only a 17.9% identity and a 60.1% similarity (Blosum42 with threshold 0) with its *T. gondii* counterpart. To validate PF3D7_1136700 as a putative AIP homolog and to get functional insights in its physiological role, we first generated the transgenic parasite line *Pf*AIP-FKBP-GFP, by tagging the endogenous protein with the FK506 binding protein (FKBP) and green fluorescent protein (GFP) using the selection-linked integration (SLI) system [20] (Fig. 3A). This approach not only allows for the localization of the targeted protein, but also allows the conditional mislocalization of the protein using the knock-sideways approach [20]. Correct insertion of the plasmid in the *Pfaip* gene locus was verified by PCR (Fig. 3B). Expression of *Pf*AIP-FKBP-GFP was then analyzed by Western blot in late schizont stage parasites, as this is the stage where mature rhoptries are present (Fig. 3C). Probing parasite lysates with an anti-GFP antibody identified a single *Pf*AIP-FKBP-GFP protein band at ~110 kDa (calculated MW = 106 kDa). To assess if, like *Tg*AIP, *Pf*AIP localizes to the rhoptries, we performed live-cell widefield fluorescence microscopy on *Pf*AIP-FKBP-GFP parasites (Fig. 3D). In both schizonts and free merozoites, *Pf*AIP localized to the apical tip, consistent with its localization in *T. gondii*. To determine if *Pf*AIP colocalized with the rhoptry protein *Pf*ARO, *Pf*AIP-FKBP-GFP parasites were cotransfected with a *Pf*ARO-mCherry overexpressing plasmid. Confocal fluorescence microscopy of these cotransfected parasites showed that *Pf*AIP localizes in close proximity to *Pf*ARO with minimal overlap (Fig. 3E–G). This was confirmed using superresolution microscopy (Fig. 3H, Sup. Fig. 2A) that allowed a clear visualization of the individual bulbs of the rhoptries by *Pf*ARO (colocalized with the inner membrane marker GAPM2). Rhoptry localization of *Pf*AIP [12] was further confirmed by colocalization with the rhoptry neck resident protein Rhoptry-associated leucine zipper-like protein (RALP1) [21,22] (Sup. Fig. 2B).

To provide functional insights on *Pf*AIP, we created a transgenic parasite line that allows the rapalog-inducible conditional mislocalization of *Pf*AIP. We cotransfected the parasite line *Pf*AIP-FKBP-GFP with a plasmid coding for a nuclear localized FRB-domain (mislocalizerN) [20] (2xNLS-FRB-mCherry, Fig. 4A) and subsequently analyzed the phenotypic consequences of *Pf*AIP mislocalization in these *Pf*AIP$_{condKS}$ schizonts. Using flow cytometry, we assessed the effect of *Pf*AIP loss on parasite growth, by treating synchronized trophozoite stage parasites (30 hpi) with rapalog for 24 h and comparing against untreated controls (Fig. 4B). Relative to untreated controls, *Pf*AIP mislocalization led to a 55.3% (2.2-fold) reduction in parasitemia in the following cycle suggesting that the function of *Pf*AIP at the rhoptries is crucial for parasite growth.

As *Pf*AIP is most highly expressed in schizonts and its homolog in *T. gondii* interacts with the rhoptry protein ARO [13,15], we hypothesized that this growth inhibition was likely due to defects in merozoite formation, schizont rupture or RBC invasion. To differentiate these possibilities, we first used Giemsa-stained, methanol-fixed, slides to count the number of segmented merozoites inside *Pf*AIP$_{condKS}$ mature schizonts; either in the presence or absence of rapalog (Fig. 4C). Mislocalization of *Pf*AIP did not result in any changes in the number of merozoites per schizont, with both rapalog treated and untreated schizonts containing an average of 25 fully formed merozoites with no apparent morphological differences. Subsequently, parasite egress (percentage of ruptured schizonts, Fig. 4D) and parasite invasion (number of rings per ruptured schizonts, Fig. 4E) after rapalog induced *Pf*AIP mislocalization was quantified and compared with the untreated control. Mislocalization of *Pf*AIP resulted in a significant reduction (53.4%) in the number of newly invaded rings per ruptured schizont (mean 4.8) compared with untreated parasites (mean 10.3), but there was no significant difference in the number of ruptured schizonts. This suggested that mislocalization of *Pf*AIP interferes directly with merozoite invasion.

Fig. 2. Sequence homology and phylogenetic analysis. (**A**) Sequence alignment of AIP from *T. gondii* and *P. falciparum*. Sequence alignment of *Tg*AIP (TGME49_309190) and *Pf*AIP (PF3D7_1136700) highlighting conserved residues (red). Additionally, annotations for the conserved AIP core region (*Pf*AIP 82–260; gray) as well as a Pfam domain with unknown function (PF10946; DUF2625) predicted by MotifFinder (red) and 6 intrinsically disordered regions (IDR1-6) predicted by MobiDB (yellow) for *Tg*AIP are highlighted. (**B**) Phylogenetic tree of AIP proteins across the alveolates. This rooted maximum-likelihood tree is based on 43 protein sequences and was estimated from an alignment of AIP sequences (60–830 amino acids). Numbers at branches indicate statistical support (bootstrap of 100 replicates) of >50% in the corresponding consensus tree. The tip points are scaled according to sequence length. Selected taxa (as referred to in the text) are highlighted in yellow (Conoidasida class), blue (Aconoidasida class), red (Sarcostidae family), green (*V. brassicaformis*), and gray (*E. mitis*, *E. acervulina*, *G. niphandrodes*).

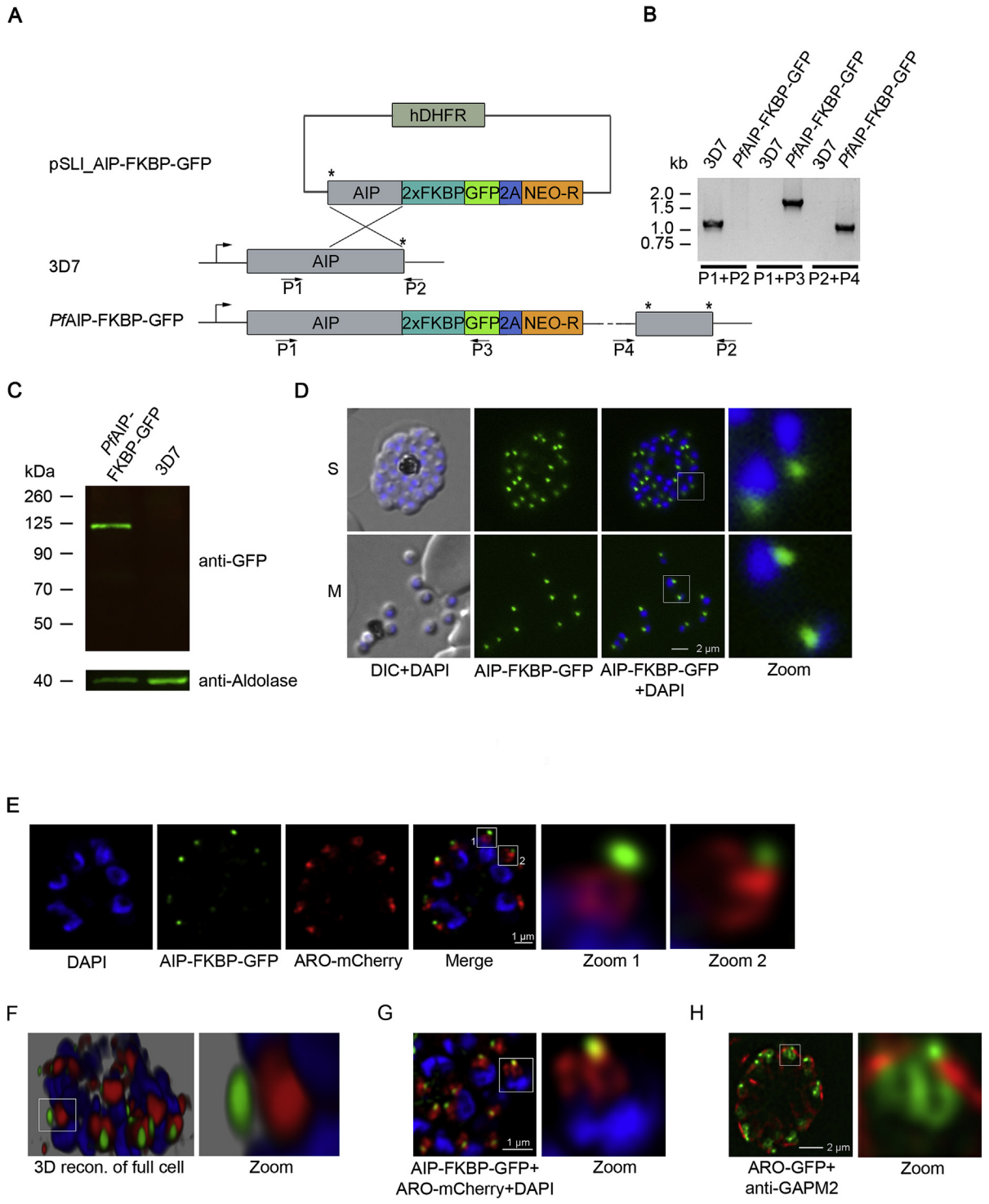


Fig. 3. *Pf*AIP localizes to the rhoptry neck. (**A**) Schematic representation of pSLI-*Pf*AIP-FKBP-GFP vector integration in the genomic *Pf*AIP locus via homolog recombination resulting in endogenous tagging of *Pf*AIP with FKBP-GFP (*Pf*AIP-FKBP-GFP). Gray: *Pf*AIP coding region, turquoise: 2xFKBP, green: GFP, blue: T2A skip peptide [64,65] orange: neomycin resistance gene. Arrows (P1−P4) are indicating position of oligonucleotides used for B. (**B**) PCR analysis of the rendered genomic locus using primers as shown in the schema using gDNA of 3D7 wild type and *Pf*AIP-FKBP-GFP parasites. (**C**) Western blot analysis using anti-GFP antibody detects an approx. 110 kDa protein in *Pf*AIP-FKBP-GFP (expected size MW 106kD) and no protein in the parental cell line. Anti-Aldolase antibody was used as loading control. (**D**) *Pf*AIP-FKBP-GFP (green) can be localized in unfixed cells at the apical pole of schizont and merozoite stage parasites. DAPI (blue) was

Mutation of PfARO leads to mislocalization of PfAIP

To probe into the putative interaction of *Pf*ARO with *Pf*AIP, we used the structural information of *Pf*ARO to create *Pf*ARO variants that are likely to interfere with protein-protein interactions. As discussed above, loop 1 and α3 are closely positioned in 3D space, forming the most highly conserved surface of *Pf*ARO. Loops that only function to join secondary structure elements do not typically contain large numbers of highly conserved residues, suggesting that the highly conserved residues within loop 1 are functionally important for mediating *Pf*ARO interaction. Therefore, we targeted loop 1 (residue 64−79), which is highly conserved between *Tg*ARO and *Pf*ARO. To test if this region is essential for interaction with *Pf*AIP we created a mutant (PfARO$_{\Delta 64\text{-}79}$) and coexpressed this in a parasite with *Pf*AIP using a bicistronic overexpression plasmid (Fig. 5A). While the coexpression of wild type *Pf*ARO-GFP with *Pf*AIP-mCherry leads to normal rhoptry association of both proteins (Fig. 5B, first row), the deletion of the loop region (AA 64−79) leads to redistribution of *Pf*AIP resulting in mostly cytosolic localization whereas the mutant *Pf*ARO is still targeted to the rhoptries (Fig. 5B, second row). Of note, given the presence of endogenous proteins in this bicistronic-overexpression parasite line, some of the *Pf*AIP-mCherry appears to still be recruited to the rhoptries. The same mislocalization can be achieved by point mutations within the loop region (H72 and W74) suggesting that the function of this conserved *Pf*ARO region is necessary to recruit *Pf*AIP either directly or indirectly to the neck of the rhoptries.

Identification of Pf*AIP* interacting proteins in late schizonts

Although bicistronic overexpression of *Pf*ARO mutants led to mislocalization of *Pf*AIP, the localization analysis of *Pf*AIP and *Pf*ARO showed minimal overlap between these two rhoptry proteins. To further probe into their interaction we first applied coimmunoprecipitation approaches using antibodies directed either against the GFP or mCherry tag and late schizont parasite material. We were unable to detect *Pf*ARO/*Pf*AIP complex in subsequent Western Blot analysis (data not shown). Next we applied proximity based biotinylation (BioID [23]) that exploits the activity of the BirA* ligase for protein biotinylation and allows determination of putative protein-protein interaction by subsequent mass spectrometry. To minimize the pool of false positives we applied a recently developed assay that is termed dimerization induced quantitative proximity-dependent biotin identification method [24]. DiQ-BioID relies on the rapalog inducible recruitment of the BirA* ligase to the site of interest i.e. the rhoptry neck. To achieve this, the *Pf*AIP-FKBP-GFP cell line was cotransfected with a BirA*-C^{L} (mcherry-FRB-BirA) plasmid [24]. Efficient relocalization of the BirA-FRB-mCherry construct to the rhoptries was confirmed by live microscopy (Fig. 6A). Additionally, using fluorescently labeled streptavidin we visualized *Pf*AIP-based biotinylation at the rhoptries in the presence of rapalog (Fig. 6B). Proteins found to be enriched by mass are depicted in Fig. 6C and D and Supp. Tab. 3. Enriched proteins (log2FC > 0.75) include *Pf*AIP, *Pf*ACβ (PF3D7_0802600), dihydrofolate synthase/folylpolyglutamate synthase (PF3D7_1324800), peptidyl-prolyl cis-trans isomerase FKBP35 (FKBP35) (PF3D7_1247400), heat shock protein 90 (PF3D7_0708400), and vacuolar protein sorting-associated protein 9 (PF3D7_0815800, PfVPS9), but not *Pf*ARO. A complete list of enriched proteins is provided in Supp. Tab. 4.

Discussion

The crystal structure of *Pf*ARO is comprised 5 tandem ARM repeats spanning residues 32−274. These repeats are oriented in a head to tail fashion generating a larger right-handed superhelical structure. The fact that residues 23−31 were disordered in the structure suggests that the N-terminal region of *Pf*ARO is highly flexible. This finding is consistent with results from partial proteolysis studies that showed the first ~ 20 residues are readily removed by treatment with different proteases (Sup. Fig. S1). It was previously suggested that residues 9−37 might form an additional ARM repeat; however, based on the evidence presented here, this seems unlikely. Rather, this region appears to form a flexible linker that tethers *Pf*ARO to the cytosolic face of rhoptry membranes through acylation at

used to stain the nuclei. Scale bar, 2 μm; DIC, differential interference contrast; 5x zoom is indicated by a white square. (**E-G**) Colocalization of *Pf*AIP-FKBP-GFP with *Pf*ARO-mCherry using confocal microscopy reveals a more restricted distribution of *Pf*AIP (green) on the rhoptries with only minimal overlap with *Pf*ARO-mCherry (red). DAPI (blue) was used to stain the nuclei. In **(E)** a single confocal image section acquired with a Leica SP8 microscope is displayed, whereas in **(F)** a full three-dimensional reconstruction of a cell is shown. For **(G)** a Zeiss Airyscan microscope was used. (**H**) SIM image from parasites expressing *Pf*ARO-GFP [12] (green) and stained with anti-GAPM2 antibodies that visualize the inner membrane complex (red).

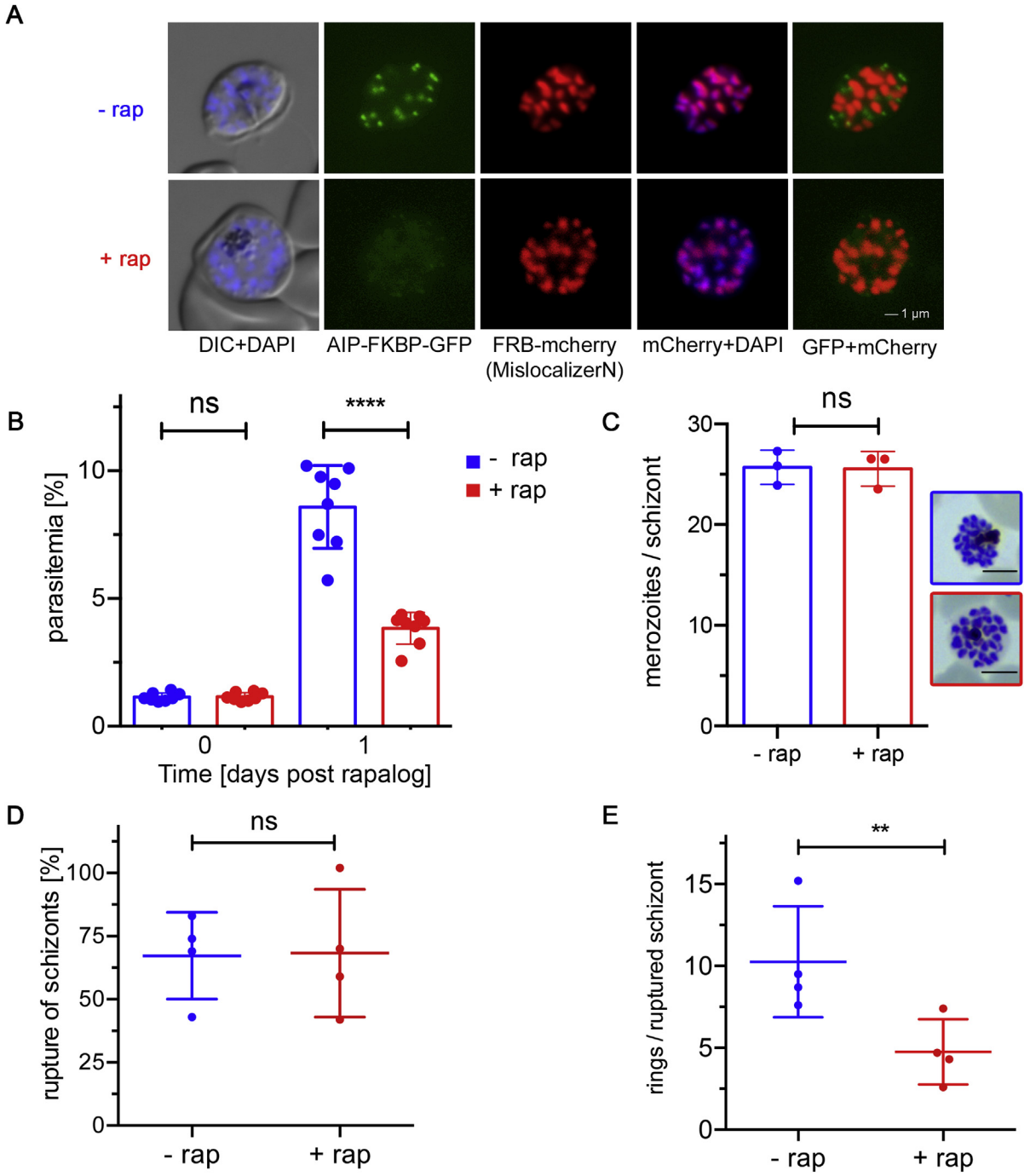


Fig. 4. Knock sideways of *Pf*AIP leads to decrease in parasitemia due to impaired invasion. (**A**) Colocalization of *Pf*AIP-FKBP-GFP (green) and the nuclear targeted mislocalizer (mislocalizerN, NLS-FRB-mCherry, red) in late schizonts in the absence (-rap) and presence (+rap) of rapalog (added at 30 hpi). *Pf*AIP-FKBP-GFP is depleted from rhoptries. Scale bar, 1 µm. Nuclei stained with DAPI. DIC, differential interference contrast. (**B**) Parasite proliferation determined by FACS analysis in the absence (-rap) or presence (+rap) of *Pf*AIP at the rhoptries. The mislocalization leads to 53,4% decreased parasitemia in the subsequent parasite cycle after rapalog addition at 30 hpi. Mean parasitemia values were determined from n = 8 independent experiments performed in duplicate. (**C**) The number of merozoites per schizont was assessed by Giemsa-stained thin blood smears taken before rupture. 10–12 schizonts per condition in biological triplicates were analyzed. Size bars, 5 µm. (**D**) Egress (percentage of ruptured schizonts) and (**E**) invasion (number of rings per ruptured schizont) after rapalog induced *Pf*AIP mislocalization was quantified. Approximately 6000 cells were analyzed for each experiment. Experiments were performed in quadruplicates on different days using biologically independent samples. Statistical significances (** = p-value < 0.01, (**** = p-value < 0.0001) were determined by performing a ratio-paired *t*-test. Error bars show standard deviation.

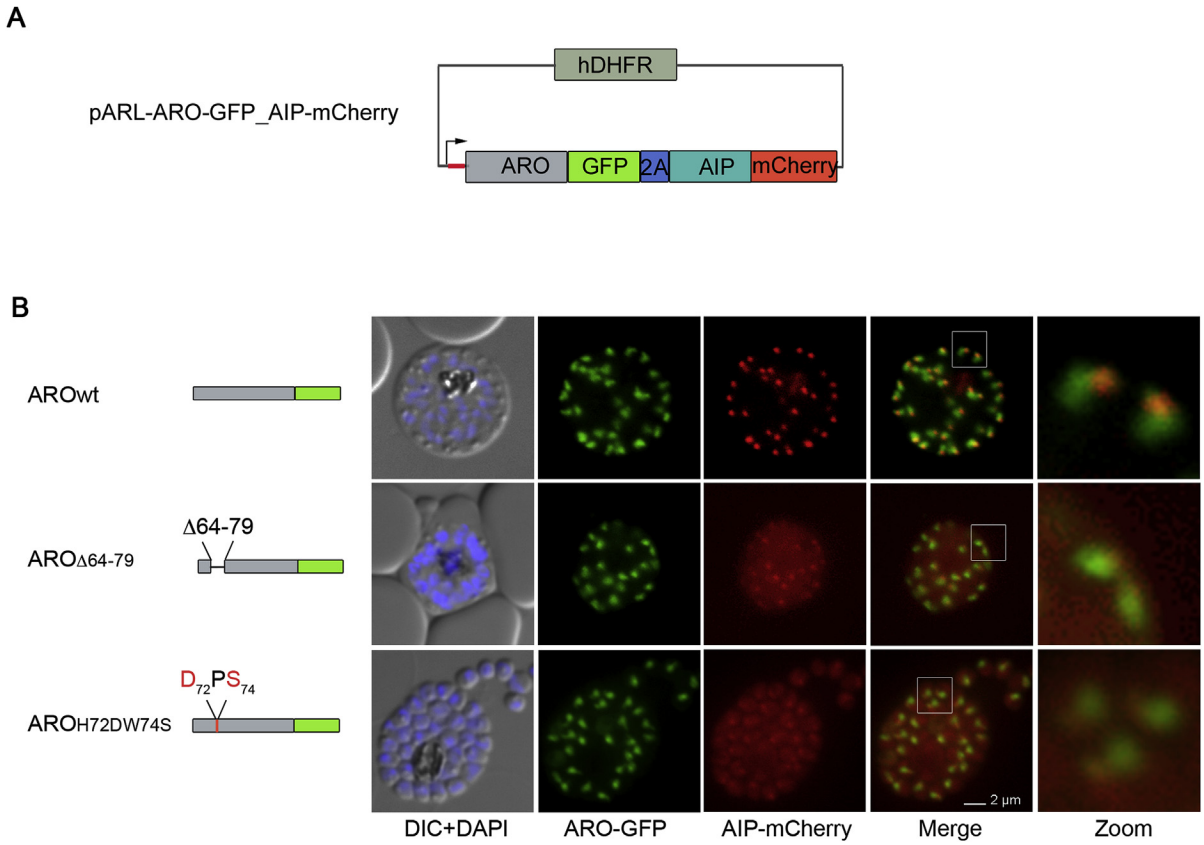
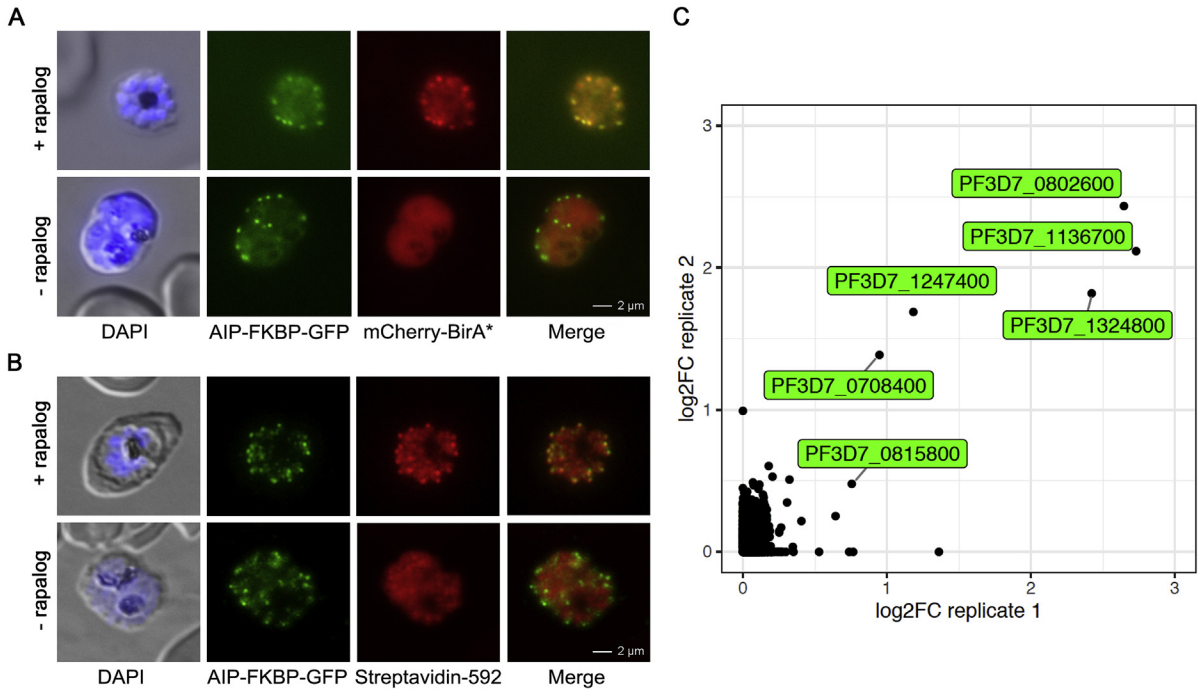


Fig. 5. Mutations in putative *Pf*ARO interaction domain cause cytosolic distribution of *Pf*AIP. (**A**) Schematics of the biscistronic expression plasmid pARL_*Pf*ARO-GFP_*Pf*AIP-mCherry under the control of the AMA1 promoter. Gray: *Pf*ARO coding region, green: GFP, blue: T2A skip peptide [64,65] turquoise: *Pf*AIP coding sequence, red: mcherry (**B**) Schematic representation of the *Pf*ARO-GFP variants (*Pf*ARO$_{WT}$, *Pf*ARO$_{\Delta 64\text{-}79}$ *Pf*ARO$_{H72DW74S}$) and their colocalization with *Pf*AIP-mCherry. Scale bars, 2 µm. Nuclei stained with DAPI. DIC, differential interference contrast; 5x zoom indicated by white square.

glycine and cysteine residues [12] located within the linker region. Tethering *Pf*ARO to the membrane by a flexible linker would allow *Pf*ARO to readily accommodate simultaneous interactions with multiple binding partners. Indeed, the *T. gondii* ARO interaction protein *Tg*AIP has been shown to interact with an ARO homolog in *T. gondii* [15], and we identified *Pf*AIP (PF3D7_1136700) as a putative binding partner of *Pf*ARO using blastp searches. Notably, the retrieved AIP sequences appear to be largely restricted to apicomplexan parasites. The one exception to this was a single homologous AIP sequence found in the phylogenetically related nonparasitic photosynthetic chromerid *Vitrella brassicaformis* (Fig. 2A), suggesting *aip* gene gain in a photosynthetic free-living proto-apicomplexan ancestor and an ancestral function unrelated to invasion. Interestingly, *Pf*AIP is also highly expressed in ookinete [25] and oocyst [26] stage of the parasite.

Overall, the crystal structure of *Pf*ARO agrees well with previously reported SAXS analysis of full-length *Tg*ARO [15]. In agreement with the crystal structure, the SAXS-based solution model (in combination with a homology model) suggested the presence of only 5 ARM repeats. Although the homology model predicted a folded N-terminal region (residues 1–37), it should be noted that this portion of the model was poorly structured and did not fit well into the *ab initio* determined SAXS envelope, implying the N-terminal region may be poorly structured. The largest difference, however, between the crystal structure and homology model relates to loop placement. While the crystal structure revealed insertion of large loops between the second and third helices of ARM 1 and 4, the homology model predicted these regions to be extensions of helices, which resulted in misplacement of key surface-exposed conserved residues. This is particularly significant for loop 1 (Fig. 1D), which forms an



D

Accession number	PlasmoDB Annotation	log2 FC replicate 1	log2 FC replicate 2
PF3D7_1136700	conserved Plasmodium protein, unknown function (aro-interacting protein (AIP))	2.73	2.12
PF3D7_0802600	adenylyl cyclase beta (ACβ)	2.65	2.44
PF3D7_1324800	dihydrofolate synthase/folylpolyglutamate synthase	2.42	1.82
PF3D7_0612200	leucine-rich repeat protein	1.36	NA
PF3D7_0610400	histone H3	1.30	-0.03
PF3D7_1247400	peptidyl-prolyl cis-trans isomerase FKBP35 (FKBP35)	1.18	1.69
PF3D7_0708400	heat shock protein 90	0.95	1.39
PF3D7_0617900	histone H3 variant (H3.3)	0.77	NA
PF3D7_0815800	vacuolar protein sorting-associated protein 9, putative (VPS9)	0.76	0.48
PF3D7_1029900	conserved Plasmodium protein, unknown function	0.74	NA
PF3D7_0907200	GTPase-activating protein, putative	NA	0.99

Fig. 6. DiQ-BioID reveals proteins in close proximity of *Pf*AIP. (**A**) Relocalization of BirA-FRB-mcherry (red) to the AIP (green) localization on addition of rapalog Nuclei stained with DAPI. Scale bar 1 µm (**B**) *Pf*AIP based biotinylation at the rhoptries. MeOH fixed schizonts expressing *Pf*AIP-FKBP-GFP (mouse anti-GFP, green). Biotinylation is visualized by streptavidin-594 (red). DAPI was used for DNA staining. (**C**) Scatterplot showing proteins with a log2FC $\geq$ 0 in both DiQ-BioID replicates. Proteins with a log2FC $\geq$ 0.4 in both replicates are labeled in green (corresponding to D) (**D**) Proteins with a log2FC > 0.75 in at least one of the replicates are listed. Proteins found to be enriched in both replicates are depicted in green, whereas proteins only found in one replicate are shown in red. FC, fold change.

extended surface with α3 of ARM1. Together, these elements comprise the most highly conserved surface on *Pf*ARO and, as shown through mutational analysis, likely mediate interaction with its binding partners.

As noted previously [15], *Tg*ARO shares structural similarity with the five C-terminal ARM repeats of the myosin chaperone UNC-45 from *C. elegans*. The concave surface formed by the third helix from each ARM repeat of UNC-45 has been implicated in myosin interaction [27,28]. Similarly, *Pf*ARO may be expected to use this surface for related interactions; however, at this point there is no clear evidence to support such a claim. In addition, prior functional analysis of individual ARM deletions indicated that *Tg*ARO requires each repeat to properly bring the rhoptries to the apical pole; however, it does not require ARM 5 for rhoptry clustering or *Tg*ACβ association [15]. Although this analysis is of interest, it is likely that the individual deletion of ARM 1–4 repeats destabilized the overall structure of ARO, limiting the interpretation of structure-function relationships. In our analysis, mapping of highly conserved residues to the concave surface of the *Pf*ARO crystal structure indicated an importance of this surface for interaction with one or more binding

partners. As shown in Fig. 1C, *Pf*ARO has two patches of conserved sequences lining the concave surface. While the first patch is composed of sequence from α3 of ARM 1 and loop 1, the second is formed by residues contributed from α9 and α12 of ARM 3 and 4, respectively. These patches are separated by a nonconserved surface from ARM 2, suggesting that each patch may interact with a separate binding partner. Deletion of loop 1 disrupts the first patch and as shown here, disrupts the ability to recruit *Pf*AIP to the rhoptries; however, it remains to be determined what the underlying mechanics of the *Pf*ARO/*Pf*AIP are, and what interactions are facilitated by second conserved amino acid patch formed by ARM 3/4. Obvious candidates for binding to the conserved ARM 3/4 residues include *Pf*MyoF and *Pf*ACβ, as they were shown as interacting partners in *T. gondii* [15,29].

To establish PF3D7_1136700 as a true *Tg*AIP homolog, we first analyzed its localization and subsequently its putative role in erythrocyte invasion. *Tg*AIP has been described previously as a rhoptry neck protein [13]. Consistent with this, endogenous GFP-tagged *Pf*AIP shows distinct apical localization that can be colocalized with the rhoptry neck marker RALP1 [21,22] but shows basically no colocalization with the rhoptry bulb associated *Pf*ARO as visualized by confocal and superresolution microscopy (Fig. 3E–H). Using the knock-sideways approach [20] we showed that the loss of *Pf*AIP at the rhoptries leads to about 55% decreased parasitaemia in the following parasite cycle due to impaired invasion (Fig. 4B, E). This finding is consistent with a genome-wide saturation mutagenesis screen in *P. falciparum* [30], in which *Pf*AIP was considered essential and indispensable for parasite proliferation. As *Pf*AIP is essential, and parasite growth was not completely inhibited, it can be inferred that the mislocalization was not 100% effective and that some cells invaded with the remaining rhoptry-localized *Pf*AIP. In contrast, in *T. gondii* the *AIP* gene could be successfully knocked out and no growth perturbation for *Tg*AIP-KO parasites was reported [15], indicating that *Tg*AIP is not essential for *T. gondii*. Given the significant differences between *Pf*AIP and *Tg*AIP (49.1 kDa vs 89.3 kDa), we can only speculate that in *T. gondii* other proteins might functionally compensate for *Tg*AIP or that *Pf*AIP has additional functions in malaria parasites.

The apparent distinct localization of *Pf*AIP and *Pf*ARO in late schizont stage of the parasite (Fig. 3E–H) is also reflected in DiQ-BioID results, which did not retrieve *Pf*ARO (Fig. 6C and D). Instead DiQ-BioID provided a list of six enriched proteins (including *Pf*AIP) that were identified in both biological replicates. Two of these (PF3D7_1247400, PF3D7_0708400) can be most likely assigned as false positives. Peptidyl-prolyl cis-trans isomerase FKBP35 (FKBP35) (PF3D7_1247400) the immunophilin of the FKBP family in *P. falciparum* [31] has been shown to be sensitive to rapamycin and binds *Pf*Hsp90 (PF3D7_0708400) [31] indicating that both are false positive hits in our DiQ-BioID caused by the dimerization of the FRB-BirA* construct to FKBP35. Furthermore for FKBP35 a nuclear and cytoplasmic [31] and for *Pf*Hsp90 a cytoplasmic localization [32] has been observed.

The *T. gondii* homolog of the adenyl cyclase beta (*Pf*ACβ, PF3D7_0802600) was previously identified as a *Tg*ARO interacting protein and a recent study localized *Pf*ACβ to the rhoptries [33]. DiCre-mediated gene excision of the *Pfacβ* gene resulted in parasites that developed normally to mature schizonts that ruptured and released merozoites. However, no new ring stage parasites were observed in the *Pf*ACβ-null parasite line. Given the congruent phenotype of *Pf*AIP knock-sideways and *Pf*ACβ-null parasites, a functional interaction between these two proteins appears likely. The homolog of the vacuolar protein sorting-associated protein 9 (PF3D7_0815800, *Pf*VPS9) and other members of the VPS protein family have been shown to be required for the transport of rhoptry and dense granule proteins in *Toxoplasma gondii* [34]. Lastly, the bifunctional dihydrofolate synthase/folyl-polyglutamate synthase (PF3D7_1324800) plays an important role in the folate biosynthesis [35]. No localization data is available and future studies have to validate its putative rhoptry neck localization.

Combining the localization, BioID, structural, and *Pf*ARO mislocalization/mutation data generated in this study along with similar analysis of *Tg*ARO/AIP interactions, we can hypothesize two models for how *Pf*ARO and *Pf*AIP may interact during the course of merozoite development and invasion. Firstly, *Pf*ARO and *Pf*AIP might interact directly with each other as evidenced by mislocalization of *Pf*ARO by deletion of select residues in loop 1 leading to increasing localization of *Pf*AIP to the cytosol rather than the rhoptry neck. Because disruption of loop 1 does not contribute to the overall fold of *Pf*ARO, it is likely that loop 1 and the juxtaposed α3 of ARM1 could the key binding surface required for *Pf*AIP rhoptry association and apparent translocation to the rhoptry neck evident in fluorescence localization studies. Given that *Pf*ARO localizes to the rhoptry bulb and *Pf*AIP to the rhoptry neck in mature schizonts and that we were unable to co-IP *Pf*ARO with *Pf*AIP, it seems plausible that the interaction between *Pf*ARO and *Pf*AIP is transient. An ARO/AIP complex might be present during nascent rhoptry development and missed by our BioID approach. A second model could see the essential interaction between *Pf*ARO and *Pf*AIP mediated by another protein. Because the proximity-based biotinylation approach identified *Pf*ACβ but not *Pf*ARO as a putative interaction

partner, it may be that the *Pf*AIP/*Pf*ACβ complex interacts with *Pf*ARO through *Pf*ACβ or another part of the complex. In addition, the finding that a second patch of highly conserved surface exposed residues is formed by α9 and α12 of ARM 3 and 4 suggests that *Pf*ARO may use this region to interact with another protein such as *Pf*MyoF or *Pf*ACβ. Because *Pf*ARO appears to be a prerequisite for targeting *Pf*AIP, likely in complex with *Pf*ACβ, to the rhoptry, it seems most likely that the ARM3/4 region would be required for *Pf*MyoF interaction. It is also possible that this region could be used to stabilize interactions between the *Pf*AIP/*Pf*ACβ complex with *Pf*ARO. Further structure-function studies using targeted single amino acid substitutions in the ARM3/4 region will be required to test these possibilities. Given that *Tg*ARO and *Tg*AIP were identified in complex, it is likely that there are differences in how these rhoptry associated proteins interact between these different parasites that need further characterization going forward.

In summary, the mislocalization of *Pf*AIP on the mutation of the putative protein-protein interacting domain of *Pf*ARO suggests that *Pf*ARO is essential for the correct trafficking of *Pf*AIP. This recapitulates the *Tg*ARO knockdown phenotype, which revealed that, in the absence of *Tg*ARO, *Tg*AIP is not localized to the rhoptries. DiQ-BioID revealed *Pf*ACβ as well as *Pf*VPS9 as likely interacting partner of *Pf*AIP, but not *Pf*ARO, that is also reflected in their different localization. Future studies elucidating *Pf*AIP structure and *Pf*AIP binding partners will be an important step towards understanding the precise function of *Pf*AIP and its interplay with *Pf*ARO and *Pf*ACβ during erythrocyte invasion.

Material and Methods

Crystallization

*Pf*ARO$_{23\text{-}275}$ was expressed in *Escherichia coli* BL21(DE3) cells as an N-terminal hexahistidine tagged protein. Cells were grown at 37 °C until reaching an OD_{600} of ~0.5, after which expression was induced for 3 h with 1 mM isopropyl β-D-thiogalactopyranoside. *Pf*ARO labeled with selenomethionine was expressed in a methionine auxotrophic strain of *E. coli* (B834) using SeMet M9 media from Shanghai Medicilon Inc. Cells were lysed by sonication in buffer containing 20 mM Tris pH 8.0, 500 mM KCl and 5 mM imidazole. Soluble lysate was loaded onto a 5 mL Ni-IMAC column (GE Healthcare). The column was washed with 10 column volumes of lysis buffer containing 100 mM imidazole, and protein eluted with the same buffer supplemented with 250 mM imidazole. Protein was exchanged into cleavage buffer (20 mM Tris pH 8.0 and 150 mM KCl) and allowed to react with TEV protease for 2 h at room temperature before passing the sample over a 5 mL Ni-IMAC column. Cleaved *Pf*ARO was collected in the flow through fraction and buffer exchanged (into 20 mM Tris pH 7.5, 150 mM KCl, 1 mM EDTA). Purified *Pf*ARO was then concentrated to 5 mg/ml by ultrafiltration (Corning).

Concentrated protein was mixed in a 1:1 ratio with the crystallization condition (0.2 M potassium fluoride, 20% w/v PEG 3350) and dehydrated over a reservoir of 1.5 M ammonium sulfate using hanging drop vapor diffusion. Crystals grew after approximately 96 h at 20 °C.

Structure determination

No additional cryoprotection was required for data collection. Crystals were flash frozen in liquid nitrogen and maintained at 100 K for data collection. Diffraction data were collected using the 08ID-1 beamline at the Canadian Light Source (CLS). The diffraction data were indexed and integrated with iMOSFLM [36] and then scaled, merged and converted to structure factors using SCALA in CCP4 [37,38]. Data sets for both native and seleno-methionine labeled *Pf*ARO were processed using the XTRIAGE module in PHENIX to assess cell content. Following this, experimental phasing was performed using the AUTOSOL module of PHENIX [39] to determine the position of the selenium atoms. The hybrid-substructure search determined the location of 8 Selenium atoms. The resulting density modified map was used for molecular replacement with a search model based on a homology model of β-Catenin (PDB 2I22). Phenix-Autobuild was then used with native data to generate an initial model, which was rebuilt-in-place into a simulated annealing OMIT map. Missing residues and sidechains were manually added and refined using iterative cycles of COOT [40] and Phenix-Refine until R_{work} and R_{free} values converged and geometry statistics reached suitable ranges (Table S1). Ramachandran analysis indicated that 98% of residues fell within the most favored positions and no outliers were present. The final structure was deposited to the PDB with the accession code 5EWP.

Sequence alignments and phylogenetic methods

A blastp search of the *Tg*AIP (TGME_309190) was performed against the nr database (June 17, 2019) using Geneious 10.2.3 (https://www.geneious.com) and an E-value of 10e-0 (BLOSUM62 substitution matrix) to identify 89 hits for *Tg*AIP and 92 hits for *Pf*AIP. We performed initial sequence alignments to detect AIP core sequences using MUSCLE [41] using the R package msa v1.16.0 [33].

For phylogenetic tree construction, we merged BLAST results for *Tg*AIP and *Pf*AIP sequences and manually filtered the list for redundancies and representative sequences. We kept the longest sequence in case multiple sequence copies with similar sequence similarity from a single species were identified resulting in a final list of 43 homologous AIP sequences (Supplementary Table 3). We then used MUSCLE v3.26.0 to align sequences and the R package phangorn v2.5.3 [42] for phylogenetic tree building. We selected the JTT + G + I amino acid substitution model based on its smallest Bayesian Information Criterion (BIC) after testing all amino acid substitution models on our data, and computed a

maximum likelihood tree with optimized topology and branch length including bootstrap analysis based on 100 samples. The exported Newick tree was visualized and annotated using the R package ggtree v1.16.1 [43,44].

Nucleic acids and constructs

All oligonucleotides used for plasmid construction are listed in Sup. Table S2.

For generation of the transgenic cell line *Pf*AIP-FKBP-GFP the 3′end of the gene (634bp) was amplified from 3D7 gDNA and cloned into pSLI-2xFKBP-GFP vector [20] in frame with *2xfkbp-gfp.* Integration of the plasmid into the *aip*-locus was verified by PCR using the oligonucleotides listed as in Sup. Table S2 using genomic DNA (gDNA) prepared from the *Pf*AIP-FKBP-GFP parasite line.

For colocalization of *P*fAIP with *Pf*ARO the parasite line *Pf*AIP-FKBP-GFP was transfected with an overexpression plasmid expressing *Pf*ARO-mCherry under the control of the late stage specific *ama-1* promoter using a blasticidin resistance cassette for selection [12]. For generation of the conditional *Pf*AIP knock side-ways cell line the *Pf*AIP-FKBP-GFP cell line was transfected with a mislocalizer plasmid 2xNLS-FRB-mCherry (mislocalizerN) [20].

For generation of transgenic parasites overexpressing *Pf*ARO-GFP variants in conjunction with *Pf*AIP-mCherry, full length coding regions were obtained using either cDNA library (*Pf*AIP) or plasmid (*Pf*ARO, [12]). *Pf*ARO was PCR amplified with *Kpn*I/*Avr*II and *Pf*AIP with *Mlu*I/*Sal*I restriction sites restriction sites (Sup. Table S2) that allows the insertion into the corresponding cloning sites of a skip vector that enable bicistronic expression under the control of the late stage specific *ama1* promoter [45]. Mutations of *Pf*ARO were generated by overlap PCR [46] using the oligonucleotides listed Sup. Table S2. PCR amplified sequences were verified by Sanger Sequencing.

For structural studies, the coding region of *Pf*ARO corresponding to residues 23–275, along with an N-terminal TEV protease recognition sequence, was cloned into the *Eco*RI/*Xho*I sites of a pET-28a vector (EMD Biosciences). The final vector expresses $PfARO_{23\text{-}275}$ with a TEV protease-cleavable N-terminal His_{6x} fusion (MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRSEF ENLYFQG).

P. falciparum culture and transfection

The *P. falciparum* clone 3D7 [47] was cultivated at a hematocrit of 4% in human O^+ erythrocytes according to standard procedures [48]. For transfection, late schizont-stage parasites were transfected with 50 μg of plasmid DNA using Amaxa Nucleofactor 2b (Lonza, Switzerland) as previously described [49]. All transfectants were selected with 4 nM WR99210 (Jacobus Pharmaceuticals). Integration of the pSLI constructs was selected with geneticin (G418) with a final concentration of 400 μg/ml (ThermoFisher, USA) initially added to a 10% parasitemia culture [20]. *Pf*AIP-FKBP-GFP parasites overexpressing *Pf*ARO-mCherry, BirA*-C^L or p2xNLS-FRB-mCherry were selected with 4 μg/ml blasticidin S. To obtain tightly synchronized parasites, ring stage parasites were treated twice 6 h apart with 5% w/v sorbitol [50].

Wide-field fluorescence microscopy

All fluorescence images were observed and captured using a Zeiss AxioImager M1 equipped with a Hamamatsu Orca C4742-95 camera and the Zeiss Axiovision software (v 4.7). A 100 × /1.4–NA lens was used. The images were processed in ImageJ v 1.5 [51]. Microscopy of unfixed IEs was performed as previously described [52]. Briefly, parasites were incubated in RPMI1640 culture medium with 1 mg/ml 4′,6′-diamidine-2′-phenylindole dihydrochloride (DAPI) (Roche) for 15 min at 37 °C before imaging. 7 μl of IEs were added on a glass slide and covered with a cover slip. Immunofluorescence assay (IFA) was performed as previously described [12]. Briefly, IEs were smeared on slides and air-dried. Cells were fixed in 100% ice cold methanol for 3 min at −20 °C. Afterward, cells were blocked with 5% bovine serum albumin (BSA) in PBS for 30 min. Next, primary antibodies (mouse anti-GFP (Roche) (1:1000), rabbit anti-RALP1-C (1:500) [21]) was diluted in PBS/3% BSA and incubated for 2 h, followed by three washing steps in PBS. Secondary antibodies (1:1000) and streptavidin-594 (1:4000) (Licor) were applied for 2 h in PBS/3% BSA containing DAPI (Roche), followed by 3 washes with PBS. One drop of mounting medium (Dako S3023) was added and the slide sealed with a cover slip for imaging.

Confocal and SIM imaging

For Fig. 3E–G and Supp. Fig. 2A, *Pf*AIP-FKBP-GFP parasites overexpressing *Pf*ARO-mCherry were fixed with 4% formaldehyde, stained with 1 μg/ml of DAPI (Roche) to visualize nuclei, and washed 3 times with DPBS. For imaging, 2 μl of a suspension of fixed cells was applied on a microscope slide and covered with a 22 mm × 22 mm high precision cover glass (Marienfeld, No. 1.5H). For Fig. 3H, immunofluorescence assays were performed using parasites overexpressing *Pf*ARO-GFP. They were fixed with 4% formaldehyde and 0.0075% glutaraldehyde, permeabilized with 0.1% Triton X-100 and blocked with 3% bovine serum albumin. Afterward parasites were incubated for 1 h with polyclonal mouse anti-GAPM2 (1:1000) [53]. Subsequently, cells were washed three times with PBS and the secondary antibody Alexa-fluor 594 anti-mouse IgG (1:2000, Molecular Probes) incubated for 1 h. After removal of unbound antibodies (three times washing with PBS) 2 μl of the cell suspension was applied on a slide and covered with a 22 mm × 22 mm high precision cover glass (Marienfeld, No. 1.5h). Confocal images (Fig. 3E and F) were acquired with a Leica SP8 microscope with laser excitation at 405 nm, 490 nm, and 550 nm for DAPI, GFP, and mCherry excitation, respectively. An HC PL APO 63x NA 1.4 oil immersion objective was used and images were acquired with the HyVolution mode of the LASX microscopy software. After recording, images were deconvolved using Huygens (express deconvolution, setting 'Standard'). Airyscan confocal imaging (Fig. 3G) was carried out with a Zeiss Airyscan LSM 880 microscope set to Airyscan superresolution (SR) mode. The microscope was equipped with laser lines at 405 nm, 488 nm, and 561 nm, and a 63x Plan APO NA 1.4 oil immersion objective. SIM imaging (Fig. 3H and Supp. Fig. 2A) was carried out with a Nikon N-SIM E system, equipped with laser lines at

488 nm and 561 nm, and a HP APO 100x NA 1.49 oil immersion objective. SIM reconstruction was conducted with the Nikon NIS Elements software (slice reconstruction mode). All images were processed using ImageJ v 1.52p [51].

Western blot analysis

Immunoblots were performed using saponin lysed infected erythrocytes. Parasite proteins were separated on a 10% SDS-PAGE gel using standard procedures [12] and transferred to a nitrocellulose membrane (Li-COR Odyssey Nitrocellulose Membrane) using a transblot device (Bio-Rad) according to manufacturer's instructions. The antibodies used for detection were mouse anti-GFP (Roche) antibody (1:1000), and rabbit antialdolase [54] antibody (1:2000), IRDye 680RD goat antimouse (Licor) (1:5000), IRDye 800CW goat anti rabbit (Licor) (1:10,000). Infrared fluorescent dye signal was visualized using an Odyssey© Fc imaging system.

Flow cytometry assessment of parasite growth and stage quantification

Tightly synchronized parasites were adjusted to 1% parasitemia at 30 hpi before culture was split evenly into two dishes. To one dish, Rapalog was added in a final concentration of 250 nM, and the other served as an untreated control. Parasitemia was measured after 24 h via flow cytometry using a previously established protocol, with minor modifications [55]. Briefly, *P. falciparum* cultures were resuspended and 20 µl of packed RBCs were transferred to an Eppendorf tube. 80 µl RPMI containing SYBRGreen (Sigma-Aldrich) and dihydroethidium (DHE) (ThermoFischer) was then added to obtain final concentrations of 0.25x and 5 µg/ml, respectively. Samples were incubated for 20 min (protected from UV light) at room temperature and parasitemia was determined using a NovoCyte® 1000 (ACEA Biosciences Inc.). For every sample, 100,000 events were recorded, and parasitemia was determined with NovoExpress® software. Assay was performed 8 times in duplicates.

For parasite stage quantification tightly synchronized *Pf*AIP-FKBP-GFP parasites overexpressing 2xNLS-FRB-mCherry were split evenly into two dishes at 30 hpi, with one dish left untreated and the other treated with rapalog at a final concentration of 250 nM. The number of merozoites per schizont was assessed by methanol-fixed, Giemsa-stained, thin blood smears taken before rupture (33 schizonts for -rap, 32 schizonts + rap (in total); biological triplicate (each 10−12 schizonts)). To determine if *Pf*AIP mislocalization produced either an egress (percentage of ruptured schizonts) or invasion (number of rings per ruptured schizont) defect, parasite stages during the process of invasion were quantified as described previously [56]. Giemsa-stained thin blood smears of synchronous cultures were taken before invasion at 38 hpi (t_0) and after invasion at 6 hpi the following cycle (t_1) either in the presence or absence of rapalog. For each time point a series of 30 images and the number of RBCs, schizonts and rings was determined manually for each image. Approximately 6000 cells were analyzed for each culture. The percentage of schizonts and rings within each biological replicate was determined.

Dimerization induced quantitative proximity-dependent biotin identification (DiQ-BioID)

The protocol for BioID in *P. falciparum* was adapted from previous published assays [24,57]. For DiQ-BioID [24] 100 ml of highly synchronous (Percoll-purified [58] schizonts were added to erythrocytes and grown for 5 h, followed by a sorbitol synchronization [50]). PfAIP-2xFKBP-GFP + BirA*-C^L (mCherry-FRB-BirA) parasite culture with a parasitemia >10% were grown to 38 h post invasion. Next the culture was divided into two identical cultures. Both cultures were grown for 8 h in RPMI supplemented with 50 µM biotin (Sigma-Aldrich) and 2 µg/ml blasticidin S (Invitrogen) and to one of the cultures rapalog was added to a final concentration of 250 nM Compound2, a protein kinase G inhibitor was added to a final concentration of 1 µM for 4 h to arrest schizont before egress [59]. Parasites were harvested and washed 2x with D-PBS containing Compound2. Infected erythrocytes were purified using MACS-column [60], lysed in 2 ml lysis buffer (50 mM Tris-HCL pH 7.5, 500 mM NaCl, 1% Triton-X-100, 1 mM DTT 1 mM PMSF and 1x protein inhibitor cocktail (Roche) and frozen at −80 °C. Three freeze-thaw cycles were performed, the sample was centrifuged at 25'000 g for 30 min and the supernatant was stored at −80 °C. For purification of biotinylated proteins, 50 µL streptavidin sepharose (GE Healthcare) was added to the lysate and incubated by rotating over night at 4 °C. The beads were washed twice in lysis buffer, once in dH_2O, twice in Tris-HCl (pH 7.5) and three times in 100 mM Triethylammonium bicarbonate buffer (TEAB) pH 7.5 (Sigma-Aldrich). The washed beads were resuspended in 200 µl ammonium bicarbonate (AmBic) (pH 8.3) and on-bead trypsin digestion (rolling with 1 µg of trypsin (Roche) for 16 h at 37 °C followed by a second trypsin digest with 0.5 µg trypsin for 2 h) was performed. Next the sample was centrifuged at 2000 g for 5 min, resuspended in 2 × 150 µl ammonium bicarbonate (pH 8.3), transferred and collected in a spin column (Pierce Spin Columns with Snap Cap, Thermo Scientific) placed in a low binding tube (Low Protein Binding Microcentrifuge tubes, Thermo Scientific). Subsequently, the left-over biotinylated peptides were eluted from the beads by 2 × 150 µl 80% ACN and 20% TFA. Next, the samples were dried using SpeedVac (Thermo Scientific) and stored at −20 °C.

Dried peptides were sent to Proteomics Core Facility at EMBL Heidelberg. Peptides were dissolved in 1% Formic acid/4% acetonitrile, sonicated in the ultrasonic bath for 5 min and desalted using an OASIS® HLB µElution Plate (Waters). Cleaned peptides were dissolved in 50 mM HEPES pH8.5 and labeled with TMT6plex Isobaric Label Reagent (ThermoFisher) according the manufacturer's instructions. After labeling, samples were pooled and purified from unreacted TMT label using OASIS® HLB µElution Plate (Waters).

An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100,

5 μm, 300 μm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 μm × 250 mm C18, 1.8 μm, 100 Å, Waters). Trapping was carried out with a constant flow of solvent A (0.1% formic acid in water) at 30 μL/min onto the trapping column for 6 min. Subsequently, peptides were eluted via the analytical column with a constant flow of 0.3 μL/min with increasing percentage of solvent B (0.1% formic acid in acetonitrile) from 2% to 4% in 4 min, from 4% to 8% in 2 min, then 8%–28% for a further 37 min, and finally from 28% to 40% in another 9 min. The outlet of the analytical column was coupled directly to a QExactive plus (Thermo) mass spectrometer using the proxeon nanoflow source in positive ion mode.

The peptides were introduced into the QExactive plus via a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (New Objective) and an applied spray voltage of 2.1 kV. The capillary temperature was set at 275 °C. Full mass scan was acquired with mass range 375–1200 *m/z* in profile mode with resolution of 70000. The filling time was set at maximum of 10 ms with a limitation of 3 × 106 ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 17500, with a fill time of 50 ms and a limitation of 2 × 105 ions. A normalized collision energy of 32 was applied. Dynamic exclusion time of 30 s was used. The peptide match algorithm was set to "preferred" and charge exclusion "unassigned," charge states 1, 5–8 were excluded. MS2 data was acquired in profile mode.

IsobarQuant [61] and Mascot (v2.2.07) were used to process the acquired data, which was searched against a Uniprot *Plasmodium falciparum* (UP000001450) proteome database containing common contaminants and reversed sequences. The following modifications were included into the search parameters: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (Protein N-term), Oxidation (M) and TMT10 (N-term) (variable modifications). For the full scan (MS1) a mass error tolerance of 10 ppm and for MS/MS (MS2) spectra of 0.02 Da was set. Further parameters were set: Trypsin as protease with an allowance of maximum two missed cleavages: a minimum peptide length of seven amino acids; at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01.

Statistical analysis

Flow cytometry were tested for normal distribution with D'Agostino Pearson test. Statistical significances were determined with ratio paired *t*-test. Statistical analysis was performed using GraphPad Prism 6 or 8 (GraphPad Software).

Accession Numbers

The structure of *Pf*ARO was deposited to the PDB with the accession code 5EWP. The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE [62] repository with the data set identifier PXD016687.

Funding

This work was supported by an operating grant from the Canadian Institutes of Health Research (MOP-89903) (MJ), DAAD/Universities Australia joint research co-operation scheme (TWG, DW, BL), Joachim Herz Foundation Project grant (TWG), NHMRC (Project Grant APP1143974) (DW), University of Adelaide Beacon Fellowships (DW), ARC PhD Scholarship (BL), German Research Foundation (DFG) grant BA 5213/3-1 (JSW, AB), EMBL Interdisciplinary Postdoc (EIPOD) program under Marie Sklodowska-Curie Actions COFUND program (grant number 664726) (JS), Joachim Herz Stiftung Add-on Fellowship for Interdisciplinary Sciences (project number 850016) (JS).

Acknowledgments

The authors are grateful to Tobias Spielmann for the pSLI-2xFKBP-GFP, p2xNLS-FRB-mCherry, BirA*-C^L plasmids and anti-aldolase antibody. They thank the Proteomics Core Facility at EMBL Heidelberg for support with mass spectrometry sample preparation, measurements and data analysis.

Author Contributions

Conceptualization: TWG, MJ; Methodology: TWG, MJ, JS, DW, SP, CL; Validation: MG, CB, JSW, LW, SL, MJ, PB; Formal Analysis: TWG, MG, DW, KZ, JS, MJ; PB, Investigation: MG, CB, JSW, LW, SL, PB, BL; Writing –Original Draft: MG, CB, JSW, BL, MF, MJ and TWG; Writing –Review & Editing: MG, CB, JSW, JS, MF, MJ, PB, AB, TWG, DW; Visualization: MG, RT, AL, DH, CB, JSW, JS, BL, MJ and TWG; Funding Acquisition: BL, DW, AB, MF, MJ and TWG; Resources: MJ, TWG; Project Administration: MJ, TWG; Supervision: MW, TWG, DW. *All authors read and approved the manuscript*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.12.024.

Received 5 August 2019;
Received in revised form 11 December 2019;
Accepted 12 December 2019
Available online 23 December 2019

Keywords:
malaria;
plasmodium;
host cell invasion;
armadillo proteins;
BioID

Abbreviations used:
ARM, armadillo; ARO, armadillo repeat only protein; AIP, ARO interacting protein; *Tg*, *Toxoplasma gondii*; *Pf*, *Plasmodium falciparum*; DAPI, 4′,6-diamidino-2-phenylindole; DiCRE, dimerizable CRE recombinase; FKBP, FK506 binding protein; FRB, FKBP rapamycin binding protein; rapa, rapalog; RBC, red blood cell; condKS, conditional knock-sideways; iEs, infected erythrocytes; SIM, structured illumination microscopy; SLI, selection-linked integration; NLS, nuclear localization signal; RALP, rhoptry-associated leucine zipper-like protein; WT, wild type.

References

[1] World Health Organization (WHO), WHO | World Malaria Report 2017, WHO, 2017.
[2] A.F. Cowman, J. Healer, D. Marapana, K. Marsh, Malaria: biology and disease, Cell 167 (2016) 610–624, https://doi.org/10.1016/j.cell.2016.07.055.
[3] A.F. Cowman, B.S. Crabb, Invasion of red blood cells by malaria parasites, Cell 124 (2006) 755–766, https://doi.org/10.1016/j.cell.2006.02.006.
[4] N.A. Counihan, M. Kalanon, R.L. Coppel, T.F. de Koning-Ward, Plasmodium rhoptry proteins: why order is important, Trends Parasitol. 29 (2013) 228–236, https://doi.org/10.1016/j.pt.2013.03.003.
[5] K. Venugopal, E. Werkmeister, N. Barois, J.-M. Saliou, A. Poncet, L. Huot, F. Sindikubwabo, M.A. Hakimi, G. Langsley, F. Lafont, S. Marion, Dual role of the Toxoplasma gondii clathrin adaptor AP1 in the sorting of rhoptry and microneme proteins and in parasite division, PLoS Pathog. 13 (2017), e1006331, https://doi.org/10.1371/journal.ppat.1006331.
[6] H.M. Ngô, M. Yang, K. Paprotka, M. Pypaert, H. Hoppe, K.A. Joiner, AP-1 in toxoplasma gondii mediates biogenesis of the rhoptry secretory organelle from a post-golgi compartment, J. Biol. Chem. 278 (2003) 5343–5352, https://doi.org/10.1074/jbc.M208291200.
[7] M.S. Breinich, D.J.P. Ferguson, B.J. Foth, G.G. van Dooren, M. Lebrun, D.V. Quon, B. Striepen, P.J. Bradley, F. Frischknecht, V.B. Carruthers, M. Meissner, A dynamin is required for the biogenesis of secretory organelles in toxoplasma gondii, Curr. Biol. 19 (2009) 277–286, https://doi.org/10.1016/j.cub.2009.01.039.
[8] C. Agop-Nersesian, B. Naissant, F. Ben Rached, M. Rauch, A. Kretzschmar, S. Thiberge, R. Menard, D.J.P. Ferguson, M. Meissner, G. Langsley, Rab11A-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis, PLoS Pathog. 5 (2009), e1000270, https://doi.org/10.1371/journal.ppat.1000270.
[9] P.J. Bradley, C. Ward, S.J. Cheng, D.L. Alexander, S. Coller, G.H. Coombs, J.D. Dunn, D.J. Ferguson, S.J. Sanderson, J.M. Wastling, J.C. Boothroyd, Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii, J. Biol. Chem. 280 (2005) 34245–34258, https://doi.org/10.1074/jbc.M504158200.
[10] K. Kremer, D. Kamin, E. Rittweger, J. Wilkes, H. Flammer, S. Mahler, J. Heng, C.J. Tonkin, G. Langsley, S.W. Hell, V.B. Carruthers, D.J.P. Ferguson, M. Meissner, An overexpression screen of toxoplasma gondii rab-GTPases reveals distinct transport routes to the micronemes, PLoS Pathog. 9 (2013), e1003213, https://doi.org/10.1371/journal.ppat.1003213.
[11] G. Hu, A. Cabrera, M. Kono, S. Mok, B.K. Chaal, S. Haase, K. Engelberg, S. Cheemadan, T. Spielmann, P.R. Preiser, T.-W. Gilberger, Z. Bozdech, Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum, Nat. Biotechnol. 28 (2010) 91–98, https://doi.org/10.1038/nbt.1597.
[12] A. Cabrera, S. Herrmann, D. Warszta, J.M. Santos, A.T. John Peter, M. Kono, S. Debrouver, T. Jacobs, T. Spielmann, C. Ungermann, D. Soldati-Favre, T.W. Gilberger, Dissection of minimal sequence requirements for rhoptry membrane targeting in the malaria parasite, Traffic 13 (2012) 1335–1350, https://doi.org/10.1111/j.1600-0854.2012.01394.x.
[13] C. Mueller, N. Klages, D. Jacot, J.M. Santos, A. Cabrera, T.W. Gilberger, J.F. Dubremetz, D. Soldati-Favre, The toxoplasma protein ARO mediates the apical positioning of rhoptry organelles, a prerequisite for host cell invasion, Cell Host Microbe 13 (2013) 289–301, https://doi.org/10.1016/j.chom.2013.02.001.
[14] J.R. Beck, C. Fung, K.W. Straub, I. Coppens, A.A. Vashisht, J.A. Wohlschlegel, P.J. Bradley, A toxoplasma palmitoyl acyl transferase and the palmitoylated armadillo repeat protein TgARO govern apical rhoptry tethering and reveal a critical role for the rhoptries in host cell invasion but not egress, PLoS Pathog. 9 (2013), e1003162, https://doi.org/10.1371/journal.ppat.1003162.
[15] C. Mueller, A. Samoo, P.-M. Hammoudi, N. Klages, J.P. Kallio, I. Kursula, D. Soldati-Favre, Structural and functional dissection of Toxoplasma gondii armadillo repeats only protein, J. Cell Sci. 129 (2016) 1031–1045, https://doi.org/10.1242/jcs.177386.
[16] C. Aurrecoechea, J. Brestelli, B.P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O.S. Harb, M. Heiges, F. Innamorato, J. Iodice, J.C. Kissinger, E. Kraemer, W. Li, J.A. Miller, V. Nayak, C. Pennington, D.F. Pinney, D.S. Roos, C. Ross, C.J. Stoeckert, C. Treatman, H. Wang, PlasmoDB: a functional genomic database for malaria parasites, Nucleic Acids Res. 37 (2009) D539–D543, https://doi.org/10.1093/nar/gkn814.
[17] E.W. Sayers, R. Agarwala, E.E. Bolton, J.R. Brister, K. Canese, K. Clark, R. Connor, N. Fiorini, K. Funk, T. Hefferon, J.B. Holmes, S. Kim, A. Kimchi, P.A. Kitts, S. Lathrop, Z. Lu, T.L. Madden, A. Marchler-Bauer, L. Phan, V.A. Schneider, C.L. Schoch, K.D. Pruitt, J. Ostell, Database resources of the national center for biotechnology information, Nucleic Acids Res. 47 (2019) D23–D28, https://doi.org/10.1093/nar/gky1069.
[18] N.D. Levine, The Protozoan Phylum Apicomplexa, CRC Press, 2018, https://doi.org/10.1201/9781351076111.
[19] H. Mehlhorn, W. Peters, The formation of kinetes and oocyst in Plasmodium gallinaceum (Haemosporidia) and considera-

tions on phylogenetic relationships between Haemosporidia, Piroplasmida and other Coccidia, Protistologica 16 (1980) 135–154.

[20] J. Birnbaum, S. Flemming, N. Reichard, A.B. Soares, P. Mesén-Ramírez, E. Jonscher, B. Bergmann, T. Spielmann, A genetic system to study Plasmodium falciparum protein function, Nat. Methods 14 (2017) 450–456, https://doi.org/10.1038/nmeth.4223.

[21] S. Haase, A. Cabrera, C. Langer, M. Treeck, N. Struck, S. Herrmann, P.W. Jansen, I. Bruchhaus, A. Bachmann, S. Dias, A.F. Cowman, H.G. Stunnenberg, T. Spielmann, T.-W. Gilberger, Characterization of a conserved rhoptry-associated leucine zipper-like protein in the malaria parasite Plasmodium falciparum, Infect. Immun. 76 (2008) 879–887, https://doi.org/10.1128/IAI.00144-07.

[22] D. Ito, T. Hasegawa, K. Miura, T. Yamasaki, T.U. Arumugam, A. Thongkukiatkul, S. Takeo, E. Takashima, J. Sattabongkot, E.-T. Han, C.A. Long, M. Torii, T. Tsuboi, RALP1 is a rhoptry neck erythrocyte-binding protein of Plasmodium falciparum merozoites and a potential blood-stage vaccine candidate antigen, Infect. Immun. 81 (2013) 4290–4298, https://doi.org/10.1128/IAI.00690-13.

[23] K.J. Roux, D.I. Kim, M. Raida, B. Burke, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells, J. Cell Biol. 196 (2012) 801–810, https://doi.org/10.1083/jcb.201112098.

[24] J. Birnbaum, S. Scharf, S. Schmidt, E. Jonscher, W.A.M. Hoeijmakers, S. Flemming, C.G. Toenhake, M. Schmitt, R. Sabitzki, B. Bergmann, U. Fröhlke, P. Mesén-Ramírez, A.B. Soares, H. Herrmann, R. Bártfai, T. Spielmann, A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites, Science 367 (2020) 51–59, https://doi.org/10.1126/science.aax4735.

[25] M.J. López-Barragán, J. Lemieux, M. Quiñones, K.C. Williamson, A. Molina-Cruz, K. Cui, C. Barillas-Mury, K. Zhao, X. Su, Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum, BMC Genomics 12 (2011) 587, https://doi.org/10.1186/1471-2164-12-587.

[26] G. Zanghì, S.S. Vembar, S. Baumgarten, S. Ding, J. Guizetti, J.M. Bryant, D. Mattei, A.T.R. Jensen, L. Rénia, Y.S. Goh, R. Sauerwein, C.C. Hermsen, J.-F. Franetich, M. Bordessoulles, O. Silvie, V. Soulard, O. Scatton, P. Chen, S. Mecheri, D. Mazier, A. Scherf, A specific PfEMP1 is expressed in P. Falciparum sporozoites and plays a role in hepatocyte infection, Cell Rep. 22 (2018) 2951–2963, https://doi.org/10.1016/j.celrep.2018.02.075.

[27] L. Gazda, W. Pokrzywa, D. Hellerschmied, T. Löwe, I. Forné, F. Mueller-Planitz, T. Hoppe, T. Clausen, The myosin chaperone UNC-45 is organized in tandem modules to support myofilament formation in C. elegans, Cell 152 (2013) 183–195, https://doi.org/10.1016/j.cell.2012.12.025.

[28] C.F. Lee, G.C. Melkani, Q. Yu, J.A. Suggs, W.A. Kronert, Y. Suzuki, L. Hipolito, M.G. Price, H.F. Epstein, S.I. Bernstein, Drosophila UNC-45 accumulates in embryonic blastoderm and in muscles, and is essential for muscle myosin stability, J. Cell Sci. 124 (2011) 699–705, https://doi.org/10.1242/jcs.078964.

[29] D. Jacot, W. Daher, D. Soldati-Favre, Toxoplasma gondii myosin F, an essential motor for centrosomes positioning and apicoplast inheritance, EMBO J. 32 (2013) 1702–1716, https://doi.org/10.1038/emboj.2013.113.

[30] M. Zhang, C. Wang, T.D. Otto, J. Oberstaller, X. Liao, S.R. Adapa, K. Udenze, I.F. Bronner, D. Casandra, M. Mayho, J. Brown, S. Li, J. Swanson, J.C. Rayner, R.H.Y. Jiang, J.H. Adams, Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis, Science 360 (2018), eaap7847, https://doi.org/10.1126/science.aap7847.

[31] R. Kumar, B. Adams, A. Musiyenko, O. Shulyayeva, S. Barik, The FK506-binding protein of the malaria parasite, Plasmodium falciparum, is a FK506-sensitive chaperone with FK506-independent calcineurin-inhibitory activity, Mol. Biochem. Parasitol. 141 (2005) 163–173, https://doi.org/10.1016/j.molbiopara.2005.02.007.

[32] G.W. Gitau, P. Mandal, G.L. Blatch, J. Przyborski, A. Shonhai, Characterisation of the Plasmodium falciparum Hsp70-Hsp90 organising protein (PfHop), Cell Stress Chaperones 17 (2012) 191–202, https://doi.org/10.1007/s12192-011-0299-x.

[33] A. Patel, A.J. Perrin, H.R. Flynn, C. Bisson, C. Withers-Martinez, M. Treeck, C. Flueck, G. Nicastro, S.R. Martin, A. Ramos, T.W. Gilberger, A.P. Snijders, M.J. Blackman, D.A. Baker, Cyclic AMP signalling controls key components of malaria parasite host cell invasion machinery, PLoS Biol. 17 (2019), e3000264, https://doi.org/10.1371/journal.pbio.3000264.

[34] J. Morlon-Guyot, S. Pastore, L. Berry, M. Lebrun, W. Daher, Toxoplasma gondii Vps11, a subunit of HOPS and CORVET tethering complexes, is essential for the biogenesis of secretory organelles, Cell Microbiol. 17 (2015) 1157–1178, https://doi.org/10.1111/cmi.12426.

[35] P. Wang, Q. Wang, Y. Yang, J.K. Coward, A. Nzila, P.F.G. Sims, J.E. Hyde, Characterisation of the bifunctional dihydrofolate synthase-folylpolyglutamate synthase from Plasmodium falciparum; a potential novel target for antimalarial antifolate inhibition, Mol. Biochem. Parasitol. 172 (2010) 41–51, https://doi.org/10.1016/j.molbiopara.2010.03.012.

[36] A.G.W. Leslie, H.R. Powell, Processing diffraction data with mosflm, in: Evol. Methods Macromol. Crystallogr. NATO Sci. Ser. II Math. Phys. Chem, 245, Springer Netherlands, 2007, ISBN 978-1-4020-6314-5, pp. 41–51, https://doi.org/10.1007/978-1-4020-6316-9_4, 41, 2007.

[37] S. French, K. Wilson, On the treatment of negative intensity observations, Acta Crystallogr. A 34 (1978) 517–525, https://doi.org/10.1107/S0567739478001114.

[38] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr. Sect. D Biol. Crystallogr. 67 (2011) 235–242, https://doi.org/10.1107/S0907444910045749.

[39] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX : a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 213–221, https://doi.org/10.1107/S0907444909052925.

[40] P. Emsley, K. Cowtan, Coot : model-building tools for molecular graphics, Acta Crystallogr. Sect. D Biol. Crystal-

logr. 60 (2004) 2126–2132, https://doi.org/10.1107/S0907444904019158.
[41] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[42] K.P. Schliep, , phangorn: phylogenetic analysis in R, Bioinformatics 27 (2011) 592–593, https://doi.org/10.1093/bioinformatics/btq706.
[43] G. Yu, D.K. Smith, H. Zhu, Y. Guan, T.T.-Y. Lam, ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data, Methods Ecol. Evol. 8 (2017) 28–36, https://doi.org/10.1111/2041-210X.12628.
[44] G. Yu, T.T.-Y. Lam, H. Zhu, Y. Guan, Two methods for mapping and visualizing associated data on phylogeny using ggtree, Mol. Biol. Evol. 35 (2018) 3041–3043, https://doi.org/10.1093/molbev/msy194.
[45] M. Kono, D. Heincke, L. Wilcke, T.W.Y. Wong, C. Bruns, S. Herrmann, T. Spielmann, T.W. Gilberger, Pellicle formation in the malaria parasite, J. Cell Sci. 129 (2016) 673–680, https://doi.org/10.1242/jcs.181230.
[46] S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, L.R. Pease, Site-directed mutagenesis by overlap extension using the polymerase chain reaction, Gene 77 (1989) 51–59, https://doi.org/10.1016/0378-1119(89)90358-2.
[47] D. Walliker, I.A. Quakyi, T.E. Wellems, T.F. McCutchan, A. Szarfman, W.T. London, L.M. Corcoran, T.R. Burkot, R. Carter, Genetic analysis of the human malaria parasite Plasmodium falciparum, Science 236 (1987) 1661–1666, 80-.
[48] W. Trager, J.B. Jensen, Continuous culture of Plasmodium falciparum: its impact on malaria research, Int. J. Parasitol. 27 (1997) 989–1006, https://doi.org/10.1016/S0020-7519(97)00080-5.
[49] R.W. Moon, J. Hall, F. Rangkuti, Y.S. Ho, N. Almond, G.H. Mitchell, A. Pain, A.A. Holder, M.J. Blackman, Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes, Proc. Natl. Acad. Sci. 110 (2013) 531–536, https://doi.org/10.1073/pnas.1216457110.
[50] C. Lambros, J.P. Vanderberg, Synchronization of Plasmodium falciparum erythrocytic stages in culture, J. Parasitol. 65 (1979) 418, https://doi.org/10.2307/3280287.
[51] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675.
[52] C. Grüring, T. Spielmann, Imaging of live malaria blood stage parasites, Methods Enzymol. 506 (2012) 81–92, https://doi.org/10.1016/B978-0-12-391856-7.00029-9.
[53] M. Kono, S. Herrmann, N.B. Loughran, A. Cabrera, K. Engelberg, C. Lehmann, D. Sinha, B. Prinz, U. Ruch, V. Heussler, T. Spielmann, J. Parkinson, T.W. Gilberger, Evolution and architecture of the inner membrane complex in asexual and sexual stages of the malaria parasite, Mol. Biol. Evol. 29 (2012) 2113–2132, https://doi.org/10.1093/molbev/mss081.
[54] P. Mesén-Ramírez, F. Reinsch, A. Blancke Soares, B. Bergmann, A.K. Ullrich, S. Tenzer, T. Spielmann, Stable translocation intermediates jam global protein export in Plasmodium falciparum parasites and link the PTEX component EXP2 with translocation activity, PLoS Pathog. 12 (2016), e1005618, https://doi.org/10.1371/journal.ppat.1005618.
[55] B. Malleret, C. Claser, A.S.M. Ong, R. Suwanarusk, K. Sriprawat, S.W. Howland, B. Russell, F. Nosten, L. Rénia, A rapid and robust tri-color flow cytometry assay for monitoring malaria parasite development, Sci. Rep. 1 (2011) 118, https://doi.org/10.1038/srep00118.
[56] J.S. Wichers, J.A.M. Scholz, J. Strauss, S. Witt, A. Lill, L.-I. Ehnold, N. Neupert, B. Liffner, R. Lühken, M. Petter, S. Lorenzen, D.W. Wilson, C. Löw, C. Lavazec, I. Bruchhaus, E. Tannich, T.W. Gilberger, A. Bachmann, Dissecting the gene expression, localization, membrane topology, and function of the Plasmodium falciparum STEVOR protein family, MBio 10 (2019) e01500–e01519, https://doi.org/10.1128/mBio.01500-19.
[57] M. Khosh-Naucke, J. Becker, P. Mesén-Ramírez, P. Kiani, J. Birnbaum, U. Fröhlke, E. Jonscher, H. Schlüter, T. Spielmann, Identification of novel parasitophorous vacuole proteins in P. falciparum parasites using BioID, Int. J. Med. Microbiol. 308 (2018) 13–24, https://doi.org/10.1016/J.IJMM.2017.07.007.
[58] E.M. Rivadeneira, M. Wasserman, C.T. Espinal, Separation and concentration of schizonts of Plasmodium falciparum by Percoll gradients, J. Protozool. 30 (1983) 367–370.
[59] C.R. Collins, F. Hackett, M. Strath, M. Penzo, C. Withers-Martinez, D.A. Baker, M.J. Blackman, Malaria parasite cGMP-dependent protein kinase regulates blood stage merozoite secretory organelle discharge and egress, PLoS Pathog. 9 (2013), e1003344, https://doi.org/10.1371/journal.ppat.1003344.
[60] C. Ribaut, A. Berry, S. Chevalley, K. Reybier, I. Morlais, D. Parzy, F. Nepveu, F. Benoit-Vical, A. Valentin, Concentration and purification by magnetic separation of the erythrocytic stages of all human Plasmodium species, Malar. J. 7 (2008) 45, https://doi.org/10.1186/1475-2875-7-45.
[61] H. Franken, T. Mathieson, D. Childs, G.M.A. Sweetman, T. Werner, I. Tögel, C. Doce, S. Gade, M. Bantscheff, G. Drewes, F.B.M. Reinhard, W. Huber, M.M. Savitski, Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry, Nat. Protoc. 10 (2015) 1567–1593, https://doi.org/10.1038/nprot.2015.101.
[62] J.A. Vizcaíno, A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, Erratum, Update of the PRIDE database and its related tools, 2016, Nucleic Acids Res. 44 (D1) (2016) D447–D456, https://doi.org/10.1093/nar/gkv1145. Nucleic Acids Res. 44 (2016) 11033. doi:10.1093/nar/gkw880.
[63] J.S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics 23 (2007) 1073–1079, https://doi.org/10.1093/bioinformatics/btm076.
[64] A.L. Szymczak, C.J. Workman, Y. Wang, K.M. Vignali, S. Dilioglou, E.F. Vanin, D.A.A. Vignali, Correction of multigene deficiency in vivo using a single "self-cleaving" 2A peptide–based retroviral vector, Nat. Biotechnol. 22 (2004) 589–594, https://doi.org/10.1038/nbt957.
[65] J. Straimer, M.C.S. Lee, A.H. Lee, B. Zeitler, A.E. Williams, J.R. Pearl, L. Zhang, E.J. Rebar, P.D. Gregory, M. Llinás, F.D. Urnov, D.A. Fidock, Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases, Nat. Methods 9 (2012) 993–998, https://doi.org/10.1038/nmeth.2143.

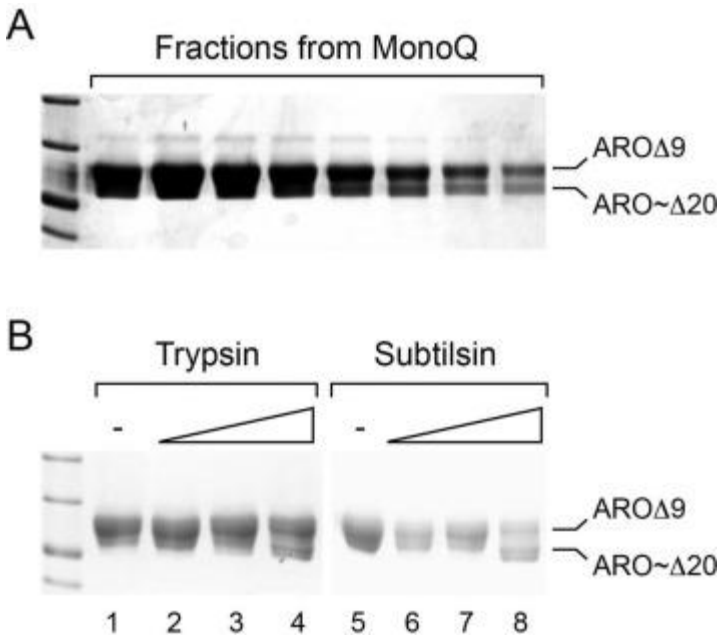


Fig. S1. Recombinant PfARO N-terminus is sensitive to proteolytic cleavage. (A) rPfARO fractions obtained from FPLC chromatography using MonoQ column. Following rPfARO purification by IMAC and removal of the N-terminal HIS6 fusion by TEV protease cleavage, rPfARO10-275 protein was further purified using a MonoQ ion exchange column. Elution fractions containing rPfARO were observed to have a lower species with a molecular weight consistent with rPfARO truncation at ~ residue 20. (B) rPfARO10-275 partial proteolysis with trypsin or subtilisin. Lane 1 and 5, PfARO10-275 only; lanes 2-4, rPfARO in presence of increasing concentration of trypsin; lanes 6-7, PfARO in presence of increasing concentration of subtilisin. Concentration of protease in lanes 2 and 6, 3 and 7, 4 and 8 were 0.1 µg/mL, 1 µg /mL and 10 µg/mL, respectively. Both proteases generated similar cleavage patterns suggesting the N-terminal ~20 residues of rPfARO are flexible and susceptible to proteolytic degradation. Experimental procedure: rPfARO protein was observed to undergo proteolytic cleavage following purification. To estimate stable domain boundaries, purified rPfARO was subjected to partial proteolysis with trypsin and subtilisin. Briefly, rPfARO (5 µg) was incubated for 30 min at 25°C in the presence of 10-fold increasing dilutions of either trypsin or subtilisin, with a starting concentration of protease set at 10 u µg/mL. Reactions were carried out in buffer containing 10 mM HEPES pH 7.5, 5 mM CaCl2 and 500 mM NaCl and stopped by addition of 2x load dye.

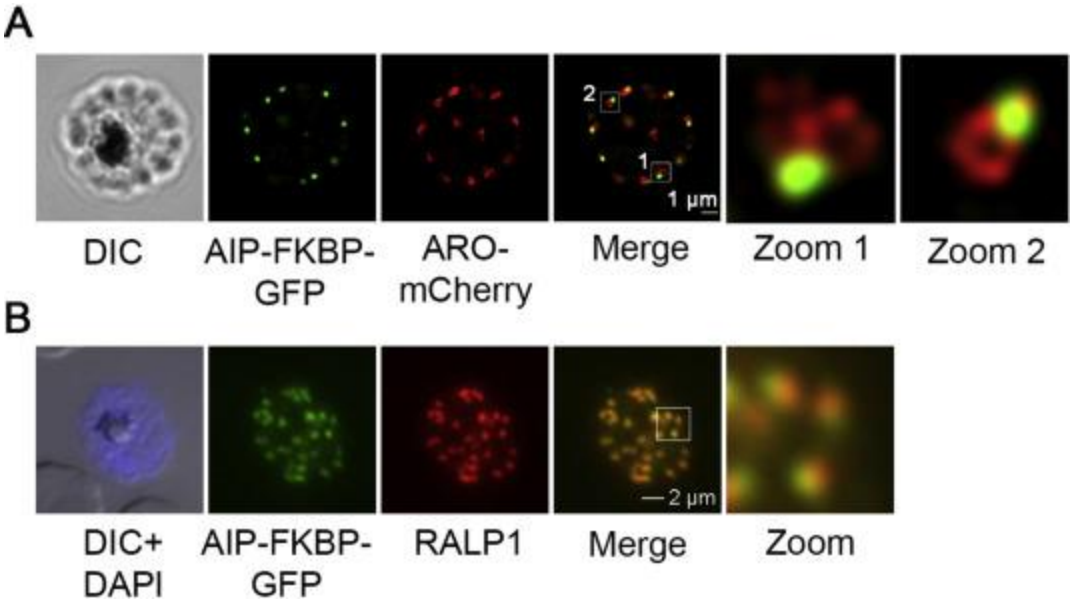


Fig. S2. (A) Additional SIM image from parasites overexpressing PfAIP-FKBP-GFP (green) and PfARO-mCherry (red), showing the localization of PfAIP and PfARO in super-resolution. (B) Co-localization of PfAIP with PfRALP1- Representative image of a fixed schizont expressing PfAIP-FKBP-GFP (mouse anti-GFP, green) co-localizing with RALP1 (rabbit anti-RALP1-C, red). DAPI was used for DNA staining. Scale bar, 2 µm. Zoomed region of interest indicated by white square.

Supp Table 1

Data Collection and Model Refinement Statistics

Data Collection	
Space group	P1
Unit cell parameters	
a,b,c (Å)	46.8, 54.9, 64.7
α,β,γ (°)	106.7, 100.6, 99.6
Molecules in ASU	2
Resolution range (Å)[a]	47.36 – 1.80 (1.84 – 1.80)
Observed reflections[a]	96,059 (5647)
Unique reflections[a]	53,118 (3132)
Redundancy[a]	1.8 (1.8)
Completeness (%)[a]	97.1 (95.7)
I/σ(I)[a]	7.9 (2.3)
R_{merge} (%)[a]	4.6 (34.6)
Wilson B Factor (Å^2)	22.1

Model and refinement	
Resolution (Å)[a]	33.9 – 1.80
R_{work}/R_{free} (%)	16.9/19.2
$\text{Reflections}_{observed}$	53,065
$\text{Reflections}_{Rfree}$	2,623
No. atoms	
Protein	3,677
Water	638
R.m.s.d. bond	
Lengths (Å)	0.007
Angles (°)	1.03
Average B Factor (Å^2)	30.0
PDB Accession Code	5EWP

[a] Statistics for the highest resolution shell are shown in parentheses.

Supp Table 2

Primer name	Sequence
pSLI-PfAIP_fwd _NotI	GGCGCGGCCGCTAACTTTTTTGATTTGGAAGAAGGC
pSLI-PfAIP_rev_pARL_AvrII	GGCCCTAGGTCTTAACATATCTTGATTAACAC
pSLI-PfAIP_int_check _fwd	GGTAATGTCTTACACAAAGAATAATATTTTAC
GFP_as_272	CCTTCGGGCATGGCACTC
pSLI-PfAIP_int_check_re	GATACATACATACGTCTTTTTTATATGG
pARL-PfARO_wt_fwd_KpnI	GCGCGGTACCATGGGAAATAATTGCTGTGC
pARL-PfARO_rev_AvrII	GCGCCCTAGGATCCGTTAGTCTCAATAAGAGAACATTG
pARL-PfAIP_fwd_MluI	GCGCACGCGTATGGATAAATTAATAAAAGAAAATATTAATG
pARL-PfAIP_rev_SalI	GCGCGTCGACTCTTAACATATCTTGATTAACACTAAC
$PfARO_{H72DW74S}$ PCRprod1_rev	GGGGTCAGCAGCACTAGGATCCATTCTATCTTC
$PfARO_{H72DW74S}$ PCRprod2_fwd	GAAGATAGAATGGATCCTAGTGCTGCTGACCCC
$PfARO_{\square 64\text{-}79}$PCRprod1_rev	GAGTTGCGGATAATGCACCAATAGTTTCGCATGTTGAAGAACATA AGTTTAC
$PfARO_{\square 64\text{-}79}$PCRprod2_fwd	GTAAACTTATGTTCTTCAACATGCGAAACTATTGGTGCATTATCCG CAACTC

Restriction site marked in blue

Supp Table 3

TAXA	Accession	Organism	Sequence_length
POM84195_hypothetical_protein_CmeUKMEL1_11180_	POM84195	Cryptosporidium meleagridis	169
XP_628227_hypothetical_protein_	XP_628227	Cryptosporidium parvum	169
XP_668365_hypothetical_protein_	XP_668365	Cryptosporidium hominis	154
XP_028875127_uncharacterized_protein_cubi_02709_	XP_028875127	Cryptosporidium ubiquitum	172
XP_002142081_hypothetical_protein_	XP_002142081	Cryptosporidium muris	226
OII74411_hypothetical_protein_cand_005110_	OII74411	Cryptosporidium andersoni	226
XP_012766862_hypothetical_protein-_conserved_	XP_012766862	Babesia bigemina	155
XP_028866417_armadillo_interacting_protein-_putative_	XP_028866417	Babesia ovata	155
XP_001611071_hypothetical_protein_	XP_001611071	Babesia bovis	198
XP_766674_hypothetical_protein_	XP_766674	Theileria parva	131
PVC57447_hypothetical_protein_MACJ_00001339_	PVC57447	Theileria orientalis	131
XP_004833724_hypothetical_protein_BEWA_043130_	XP_004833724	Theileria equi	60
CEM01199_unnamed_protein_product_	CEM01199	Vitrella brassicaformis	143
KEP62791_UNVERIFIED_CONTAM-_hypothetical_protein_HHA_309190_	KEP62791	Hammondia hammondi	823
TgAIP_-TGME49_309190	TgAIP_-TG	Toxoplasma gondii	822
XP_003884963_conserved_hypothetical_protein_	XP_003884963	Neospora caninum	829
PFH33909_hypothetical_protein_BESB_070610_	PFH33909	Besnoitia besnoiti	474
PHJ19779_armadillo_interacting_protein_	PHJ19779	Cystoisospora suis	193
XP_013438625_hypothetical_protein-_conserved_	XP_013438625	Eimeria necatrix	350
XP_013235948_hypothetical_protein-_conserved_	XP_013235948	Eimeria tenella	350
XP_013338132_hypothetical_protein-_conserved_	XP_013338132	Eimeria maxima	326
XP_026191212_uncharacterized_protein_LOC34617608_	XP_026191212	Cyclospora cayetanensis	299
XP_012763768_hypothetical_protein_PRSY57_1135100_	XP_012763768	Plasmodium reichenowi	381
PfAIP_-PF3D7_1136700	PfAIP_-PF	Plasmodium falciparum	421
XP_018641281_hypothetical_protein_PGSY75_1136700_	XP_018641281	Plasmodium gaboni	381
XP_028533244_conserved_Plasmodium_protein-_unknown_function_	XP_028533244	Plasmodium relictum	382
XP_028529224_conserved_Plasmodium_protein-_unknown_function_	XP_028529224	Plasmodium gallinaceum	381

XP_022812157_conserved_Plasmodium_protein-_unknown_function_	XP_022812157	Plasmodium yoelii	329
XP_676854_conserved_Plasmodium_protein-_unknown_function_	XP_676854	Plasmodium berghei	386
XP_008624911_hypothetical_protein_YYE_03037_	XP_008624911	Plasmodium vinckei	329
XP_016653586_conserved_Plasmodium_protein-_unknown_function_	XP_016653586	Plasmodium chabaudi	338
SBT77248_conserved_Plasmodium_protein-_unknown_function_	SBT77248	Plasmodium ovale	390
XP_028861782_conserved_Plasmodium_protein-_unknown_function_	XP_028861782	Plasmodium malariae	382
XP_004222589_hypothetical_protein_PCYB_094260_	XP_004222589	Plasmodium cynomolgi	92
XP_019915130_Uncharacterized_protein_PCOAH_00027060_	XP_019915130	Plasmodium coatneyi	327
XP_002259374_hypothetical_protein-_conserved_in_Plasmodium_species_	XP_002259374	Plasmodium knowlesi	324
XP_012333180_hypothetical_protein_AK88_00106_	XP_012333180	Plasmodium fragile	323
XP_008816184_hypothetical_protein_C922_02363_	XP_008816184	Plasmodium inui	298
XP_001615482_hypothetical_protein-_conserved_	XP_001615482	Plasmodium vivax	323
XP_028543723_hypothetical_protein-_conserved_	XP_028543723	Plasmodium gonderi	326
XP_013354856_hypothetical_protein-_conserved_	XP_013354856	Eimeria mitis	108
XP_013249560_hypothetical_protein-_conserved_	XP_013249560	Eimeria acervulina	109
XP_011131164_hypothetical_protein_GNI_101370_	XP_011131164	Gregarina niphandrodes	181

7. Manuscript 3

N-terminal autophosphorylation regulates the activity of *Plasmodium falciparum* GSK3

Samuel Pazicky[1,2], Arne Alder[1,3], Maxime Killer[1,2], Ekaterina Round[2], Lilie Hauke[4], Tim Gilberger[1,3,5] and Christian Löw[1,2]

[1] Centre for Structural Systems Biology (CSSB), Notkestrasse 85, D-22607 Hamburg, Germany.

[2] Molecular Biology Laboratory (EMBL), Hamburg Unit c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, D-22607 Hamburg, Germany.

[3] Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Strasse 74, D-20359 Hamburg, Germany.

[4] Charles Tanford Protein Centre, Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg , Kurt-Mothes-Str. 3a, 06120 Halle, Germany

[5] Department of Biology, University of Hamburg, Hamburg, Germany

Abstract

The number of malaria cases has been growing over the last five years and therefore, investigations of novel targets are necessary for the development of new drugs. A crucial step in the *Plasmodium falciparum* life cycle is its attachment to the host cell and a formation of the tight junction at their interface. A protein involved in the formation of the tight junction, Apical Membrane Antigen 1 (AMA1), is activated upon phosphorylation by the Glycogen Synthase Kinase 3 (PfGSK3). PfGSK3 is a promising drug target because several inhibitors have been shown to specifically impact the life cycle of *Plasmodium falciparum* but molecular mechanisms of PfGSK3 function have not been well characterized so far. Here, we present a detailed PfGSK3 expression and purification protocol. We characterized the activity of the purified protein and investigated the role of the unique PfGSK3 N-terminus. Our data show that PfGSK3 exhibits autophosphorylation, which is important for protein solubility and activity. Residues S226 and Y229 as well as the N-terminal domain are crucial for PfGSK3 activity and we propose that phosphorylation of N-terminal residues Y39, S40, S42 and S43 further fine-tunes the activity of PfGSK3.

Introduction

Plasmodium falciparum is an intracellular eukaryotic parasite that causes the most severe form of malaria in humans, infecting over 200 million people every year [1]. The emergence of malaria parasites resistant to all currently used antimalarial drugs imposes a serious threat to public health [2]. Identification of new drug targets and drugs is therefore crucial for future handling of the malaria pandemics. Kinases comprise up to 10% of all drug targets [3] and a number of potent inhibitors were discovered that are directed against *P. falciparum* kinases, such as glycogen synthase kinase 3 (GSK3) [4]. In humans, GSK3 plays a crucial role in a number of signalling processes, regulating glycogen metabolism [5–7], cell cycle and growth [8–10], translation [7], embryonic development [9,11] or differentiation of neurons [12–16]. Consequently, the inhibitors of GSK3 are investigated as potential treatment options for neurodegenerative and psychiatric diseases [17–23]. In *P. falciparum*, PfGSK3 likely regulates the red blood cell invasion by phosphorylation of the apical membrane antigen 1 (PfAMA1) [24–29]. During host cell invasion, the parasites squeeze inside the cell through tight junction, a circular interface between the parasite and the host cell at the site of the host cell entry [30]. The N-terminal ectodomain of PfAMA1 forms a complex with rhoptry neck (RON) proteins, forming a basis of tight junction in *P. falciparum* merezoites [31–35]. On the other hand, the short C-terminus of PfAMA1 locates inside the parasite and thus, is amenable to regulation by phosphorylation [25,36,37]. In a two-step phosphorylation event, the protein kinase A (PfPKA) first phosphorylates S610 of PfAMA1 [36,38,39], which in turn enables subsequent phosphorylation of PfAMA1 residue T613 by PfGSK3 [24]. Thus, PfGSK3 serves as a secondary kinase, although both phosphorylation sites are equally important for the PfAMA1 function and *P. falciparum* invasion [24]. Indeed, a set of PfGSK3 inhibitors built on a thieno[2,3-b]pyridine scaffold has been discovered that display antiplasmodial activity in the low micromolar range [40,41]. Although these inhibitors bind to the conserved ATP binding pocket, they selectively interact with PfGSK3 compared to human GSK3 [41]. Further development of these inhibitors could be guided by the structure of PfGSK3 that has, however, not been determined so far. Based on sequence homology, PfGSK3 consists of two conserved

structural domains: a β-sheet domain with an ATP binding pocket and an α-helical domain with the activation loop and a substrate binding site. These domains are typically preceded by a disordered N-terminal domain, phosphorylation of which generally leads to inhibition of GSK3 activity [42]. In PfGSK3, the important residues in the ATP binding pocket and in the activation loop remain conserved, however, its N-terminal domain is unique to *Plasmodium* species and its structure and function are unknown [43].

To address these matters, we developed a robust expression and purification protocol of PfGSK3 that enabled us to perform thorough structural and functional characterization of the protein *in vitro*. We show that PfGSK3 is sensitive to bivalent heavy metal ions that induce PfGSK3 oligomerization, which in turn obstructs the kinase activity. Furthermore, we show evidence that PfGSK3 exhibits autophosphorylation *in vitro* and its kinase activity is dependent on phosphorylation of the unique N terminus. Phosphorylation of the N-terminal residues of PfGSK3 increases the activity of PfGSK3, providing an additional regulatory mechanism to the protein.

Results

Expression and purification of PfGSK3 from *E. coli* yields a pure protein

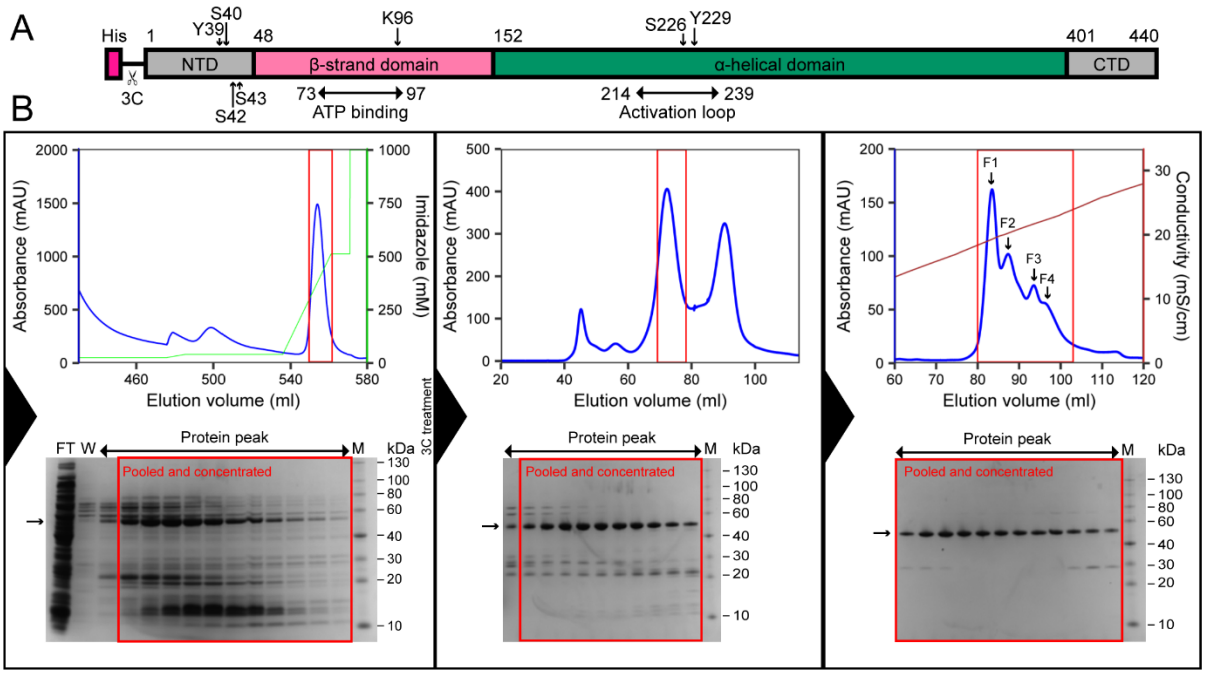


Fig 1. Expression and purification of PfGSK3. (A) Construct of PfGSK3 used for the expression. The construct consists of the full-length sequence of PfGSK3 with N-terminal His-tag and 3C cleavage site. The domain organization and phosphorylation sites are marked. (B) Purification of PfGSK3. From left to right: imidazole gradient elution profile from His-Trap column; elution profile from Superdex 200 size exclusion column and NaCl gradient elution profile from Resource Q ion exchange column. The peaks of the ion exchange elution represent fractions that are phosphorylated to different extent, which is increasing from F1 to F4.

A major challenge in obtaining high-quality functional and structural data on PfGSK3 *in vitro* is sample purity and homogeneity. For PfGSK3 expression, we used a construct consisting of a N-terminal Histidine tag, a 3C-protease cleavage site and the open reading frame of full-length PfGSK3 (Fig 1A) [41]. The protein was expressed in the *E. coli* C41(DE3) strain and further purified in a 1-day multi-step process (see *Methods* and Fig 1B). In short, after cell lysis, the supernatant is first subjected to immobilized metal affinity chromatography (IMAC) with gradient imidazole elution. The resulting protein peak is concentrated and at the same time cleaved with the 3C protease. At this point, the protocol deliberately excludes a negative IMAC step because after the removal of the histidine tag, PfGSK3 becomes sensitive to NiNTA beads and precipitates. Therefore, the concentrated protein is directly separated *via* size exclusion chromatography. As the fractions containing PfGSK3 still comprise a significant amount of contaminants, we opted to include ion exchange chromatography as a last purification step (Fig 1C). This results in the desired level of purity of PfGSK3, while the protein

elutes in several partially separated peaks with an increasing concentration of sodium chloride, indicating that the heterogeneity of PfGSK3 is potentially caused by different posttranslational modifications . Following this protocol, we were routinely able to obtain around 1.5 mg of pure protein from 1l of bacterial culture.

PfGSK3 purified from *E. coli* retains its structure and function

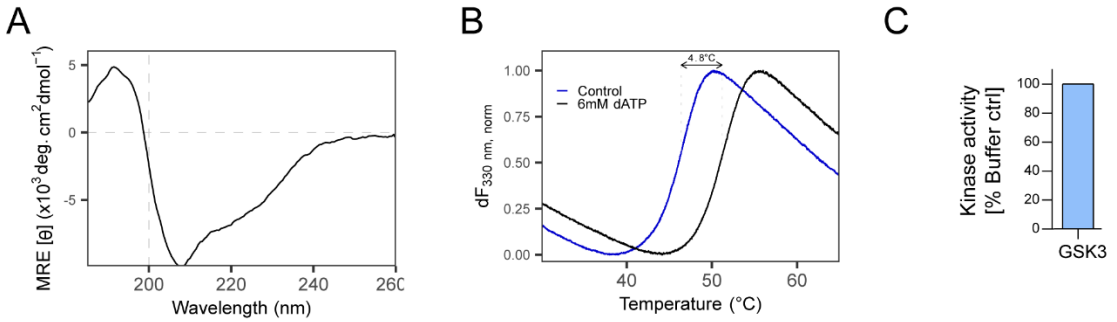


Fig 2. Characterization of PfGSK3. (A) Circular dichroism shows secondary structure composition similar to what is expected from GSK3 protein, with 25% α helix, 19% β sheet and 56% random coil. The circular dichroism was measured 10x and the data were averaged, buffer subtracted, and analyzed by DichroWeb. (B) Thermal unfolding profiles of PfGSK3 in presence or absence of 6mM dATP. dATP dramatically stabilizes PfGSK3, suggesting that it binds in the ATP binding pocket. (C) Activity of PfGSK3 measured with a peptide substrate. The calculated catalyst rate constant is k_{cat} = 9.5 s^{-1}, which is in the range typical to homologous proteins.

To assess the quality of the purified protein and to characterize its structure and function, we combined several biophysical methods. Dynamic light scattering data confirm high homogeneity of the sample with a calculated molecular weight of 52 kDa, in agreement with the expected value for a monomer of GSK3 (SFig 2A). The secondary structure content derived from circular dichroism data is 22% α helix, 26% β sheet, 24% turns and 29% disordered. This corresponds to a typical domain organisation of homologous GSK3 proteins with an N-terminal β-sheet-rich domain and α helical C-terminal domain that are flanked by disordered N-terminal and C-terminal regions (SFig 1A, Fig 2A). To verify that the protein retains its function, we tested weather its thermal stability changes in the presence of ATP and ATP analogs. Indeed, the stability of PfGSK3 markedly increased in a concentration dependent manner from 45°C to over 50°C with increasing concentration of dATP or ATP (Fig 2B and SFig 2B) and the protein could also be stabilized by several non-hydrolyzable ATP analogs (SFig 2C). Additionally, we also observed a concentration-dependent increase

of PfGSK3 thermal stability in the presence of sulphate ions (SFig 2D). Sulphate ions are bound in the activation loop of several human GSK3 structures, indicating that the structural elements in this region are conserved. Finally, we measured the PfGSK3 activity with a resulting catalyst rate constant k_{cat} = 9.5 s^{-1}, which is in the range that is typical for protein kinases and comparable to human GSK3 (Fig 2C) [44]. Thus, the purified PfGSK3 is active and retains its structure that is similar to that of human GSK3.

Bivalent ions of heavy metals induce formation of PfGSK3 oligomers

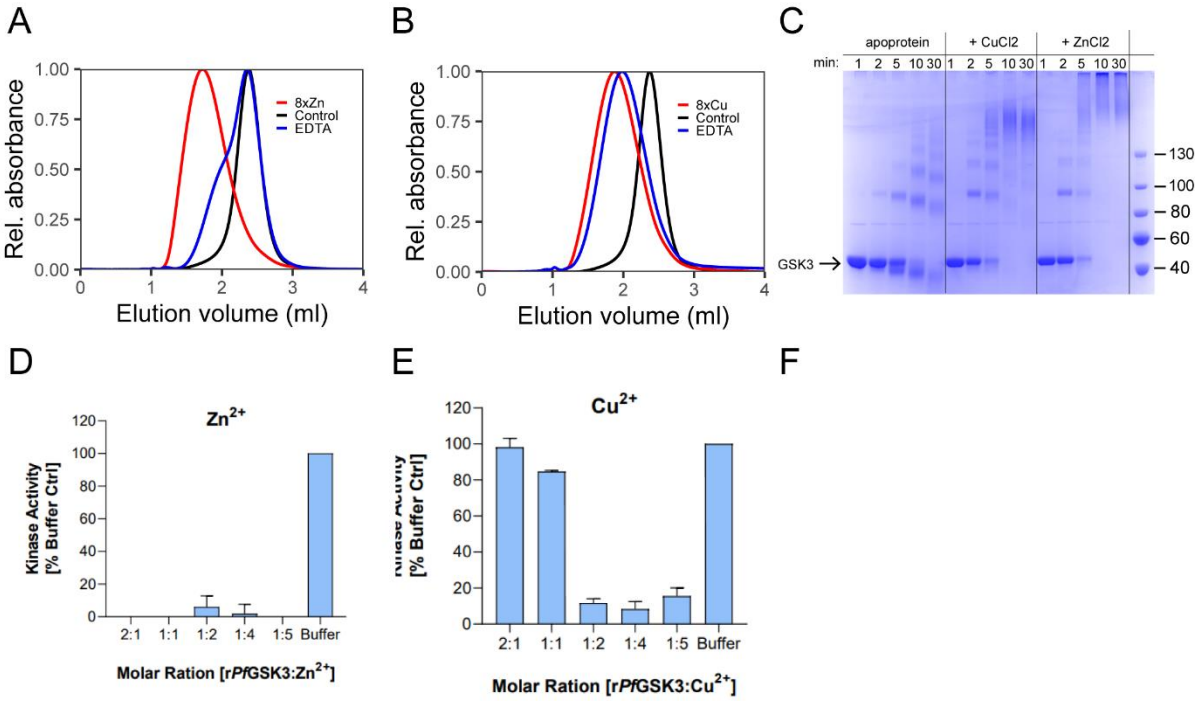


Fig 3. PfGSK3 autophosphorylation. (A) Analytical size exclusion chromatography profiles on home-packed Superose 6 column of PfGSK3 apoprotein (black), after addition of 8x molar excess of zinc chloride (red) and after subsequent addition of EDTA (blue). (B) Analytical size exclusion chromatography profiles on home-packed Superose 6 column of PfGSK3 apoprotein (black), after addition of 8x molar excess of copper chloride (red) and after subsequent addition of EDTA (blue). Both analyses show that zinc and copper induce formation of high-MW PfGSK3 species. The formation is reversible because addition of EDTA shifts the elution profiles towards higher elution volumes. (C) Cross-linking experiment confirms that compared to apoprotein, PfGSK3 forms high-MW species that are induced by addition of copper or zinc ions. (D-E) The activity of PfGSK3 is strongly reduced in the presence of zinc or copper ions.

The observation that PfGSK3 with a cleaved histidine tag has a tendency to precipitate upon interaction with NiNTA beads prompted us to investigate the impact of various heavy metals on PfGSK3 structure. To this end, we measured thermal unfolding profiles of the protein in the presence of different metal ions. Under standard conditions, PfGSK3 displays a transition

midpoint at 45°C (fluorescence recorded at 330 nm). However, in the presence of bivalent ions, such as zinc, cobalt, nickel and copper cations, the typical unfolding transition completely disappears, suggesting structural changes in the protein (SFig 3A). We selected zinc and copper to further investigate the effect of bivalent heavy metal ions on PfGSK3 because they have the most pronounced effect on PfGSK3 unfolding. First, the secondary structure content measured by circular dichroism did not change upon addition of zinc or copper, demonstrating that the protein maintains its secondary structure (SFig 3D). However, analytical size exclusion chromatography revealed that both zinc and copper induce formation of high molecular weight (high-MW) protein species that elute earlier than the apoprotein in a concentration-dependent manner (Fig 3A-B, SFig 3B-C). Indeed, in a cross-linking experiment, the high-MW species appear only in the presence of zinc or copper (Fig 3C). Cross-linking experiments further show that the copper-induced species are of lower molecular weight than the zinc-induced species, which is in agreement with the observed size exclusion elution times (1.7 vs 1.9 ml, respectively). The high-MW species can be reversed into monomers by addition of EDTA, showing that the effect of heavy metal ions is reversible (Fig 3A-B). To study the impact of the heavy metals on the PfGSK3 function, we measured its activity in presence of zinc or copper in different concentration. Heavy metals cause a dramatic decrease in the PfGSK3 activity that is concentration-dependent and stronger with zinc compared to copper (Fig 3D-E). Thus, although PfGSK3 retains the secondary structure in the metal-induced high-MW species, its enzymatic activity is completely inhibited. Importantly, the reversed PfGSK3 monomers fully regain their kinase activity (Fig 3F), showing that the process is reversible. To quantify the size of the high-MW PfGSK3 species, we recorded small angle X-ray scattering (SAXS, SFig 3E-F) and analytical ultracentrifugation (AUC, Fig3F) data. Both methods show that heavy metals the induce formation of large particles (>1 MDa). Moreover, the particles display a broad sedimentation coefficient distribution derived from AUC (around 20 S), as well as broad distance distribution derived from SAXS (D_{max}=89.23 nm, but the distance distribution peaks at 20 nm), which indicates that they are highly heterogeneous. These data are in good agreement with the observed broadening of peaks on

size exclusion chromatography (from 0.27 ml peak width of PfGSK3 monomer to 0.74 ml with Cu and 0.69 ml with Zn), which further confirms that the metal-induced high-MW species are rather heterogeneous. To visualize them, we separated the copper-induced large-MW particles on size exclusion chromatography and imaged them by negative EM (SFig 3G). Particles of various shapes and sizes underline the heterogeneity of the high-MW species, although repetitive particles of similar size and shape were observed in the sample, too. In summary, the heavy metal ions, in particular zinc and copper, induce a reversible formation of large PfGSK3 particles while the protein, although being folded, completely loses its enzymatic activity.

PfGSK3 exhibits autophosphorylation

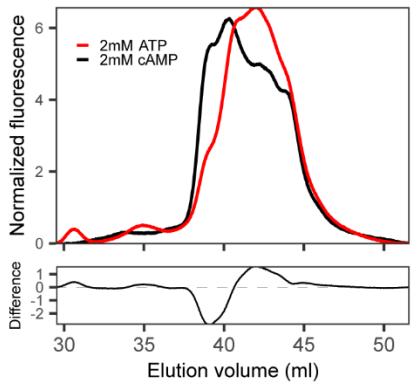


Fig 4. PfGSK3 autophosphorylation. The IEX elution chromatograms (upper plot) of PfGSK3 after incubation with (red) or without (black) ATP suggest that PfGSK3 exhibits autophosphorylation. The difference between the two chromatograms (bottom plot) shows access of later-eluting species after ATP treatment, suggesting higher amount of phosphorylation.

In the last step of PfGSK3 purification protocol, ion exchange chromatography (Fig 1B) reveals the heterogeneous nature of the recombinant protein. PfGSK3 elutes in at least four different species of comparable purity and the difference in the elution volume can only be explained by the heterogeneous phosphorylation patterns. To investigate this possibility, we separately analyzed the four top peak fractions (F1-F4) from the IEX elution profile of the wild type PfGSK3 preparation (Fig 1B). SDS-PAGE stained with the Pro-Q Diamond dye, which discriminates between unphosphorylated and phosphorylated proteins, shows that the amount of total phosphorylation increases towards later elution volumes (SFig 4A). Our mass spectrometry analysis then showed that the protein is partially phosphorylated at several

residues (SFig 4B). Multiple phosphorylation sites were identified at the disordered N-terminus, predominantly Y39, S40, S42 and S43. Additionally, several residues in the activation loop, mainly S226, but also S228, Y229 and S232, were found to be phosphorylated. As the *E. coli* expression system only possesses a limited capacity to phosphorylate proteins [45–47], we tested the ability of PfGSK3 to exhibit autophosphorylation by incubation with ATP/$MgCl_2$ and subsequently analyzed of the ion exchange chromatography profiles. Due to the usage of an analytical setup compared to traditional purification systems, the elution profile displays a more complicated chromatogram compared to the purification setup (Fig. 4 vs Fig. 1B). Nevertheless, a clear drop in the fluorescence signal from the earlier eluting peaks and gain in the fluorescence signal from later eluting peaks suggests that PfGSK3 is indeed able to exhibit autophosphorylation *in vitro*. As autophosphorylation of human PfGSK3β plays a role in its regulation, it is plausible that the function is conserved in PfGSK3. However, the N-terminus of PfGSK3 is distinct from the N-terminus of HsGSK3, prompting us to investigate the phosphorylation and the N-terminus of PfGSK3 in detail.

N-terminal domain is indispensable for the PfGSK3 function

To investigate the role of phosphorylation in PfGSK3, we attempted to express inactive mutants. We mutated lysine K96 in the ATP binding pocket and residues from the activating loop S226 and Y229 to alanine. The small scale expression screen showed that the inactive mutants did not express in a soluble form but only formed inclusion bodies (Fig 5A). We then isolated the inclusion bodies of the $PfGSK3^{K96A}$ and $PfGSK3^{S226A/Y229A}$ mutants and examined their phosphorylation states by mass spectrometry. The results show that these mutants display the full length sequence but are devoid of any phosphorylation (K96A) or the amount of phosphorylation is markedly lower compared to the wild-type protein (S226A/Y229A) (Fig 5C). This indicates that the mutations of these residues impair PfGSK3 folding and function. At the same time, the experiment suggests that PfGSK3 phosphorylation is crucial for its solubility. This is further confirmed by co-expression of wild-type PfGSK3 with the de-

phosphorylating lambda phosphatase, resulting in insoluble protein expression only (data not shown).

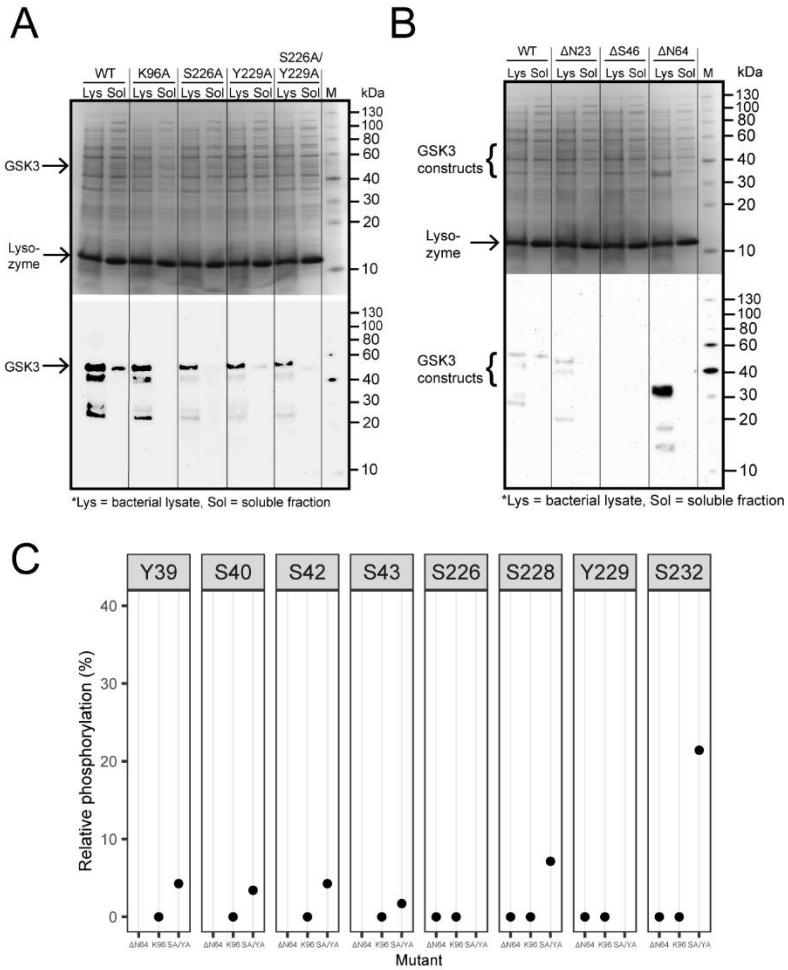


Fig. 5. Role of PfGSK3 N-terminus. (A) Small-scale expression test of PfGSK3 mutants with inactivating mutation in ATP binding site (K96A) and activation loop (Y226A and S229A). Whereas all protein constructs are expressed, only the wild type PfGSK3 is soluble, which indicates that the autophosphorylation of PfGSK3 is important for its solubility. (B) Small-scale expression test of PfGSK3 N-truncated constructs that start with residues N23, S46 or N64. In spite of a very strong expression of PfGSK3-N64, none of the proteins is soluble, indicating that the PfGSK3 N-terminus is crucial for the autophosphorylation process. The bacterial lysates (Lys) and their soluble fractions (Sol) were analyzed on SDS-PAGE and by Western blot with anti-His antibodies. (C) Analysis of phosphorylation of PfGSK3 mutants shows that the phosphorylation is completely lost (K96A, ΔN64) or reduced (S226A/Y229A) compared to the wild-type protein.

As we found the N-terminal extension of PfGSK3 to carry several phosphorylation sites, we cloned a series of N-terminally truncated protein constructs (SFig 1A). None of these constructs expressed in a soluble form (Fig 5B) and mass spectrometry confirmed again a complete lack of phosphorylation in the PfGSK3 activation loop (Fig 5C). Expression tests at different expression temperatures indicate that not only the above-mentioned mutants and constructs, but also wild type PfGSK3 are insoluble and less expressed at lower temperatures. In fact, the constructs with truncated N-termini do not express at 18C at all. (SFig 5A-B). Thus, these data further stress the importance of both PfGSK3 phosphorylation as well as of its N-terminus. Because the N-terminally truncated PfGSK3 constructs completely lack the phosphorylation, we propose that the N-terminal PfGSK3 domain, unique for *Plasmodium* species, is essential for PfGSK3 activity.

N-terminal phosphorylation fine-tunes the activity of PfGSK3

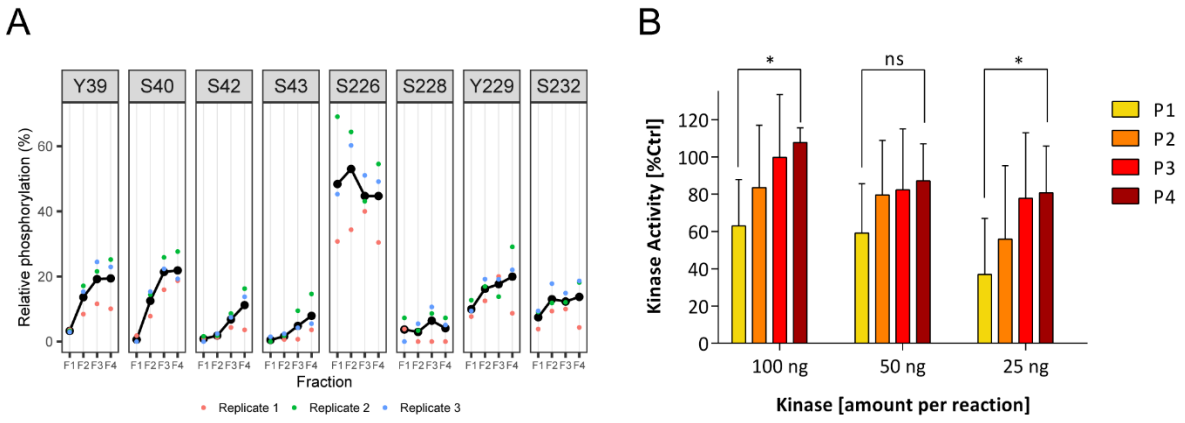


Fig. 6. Impact of PfGSK3 N-terminal phosphorylation. (A) Relative phosphorylation of selected residues in the individual PfGSK3 fractions separated by ion exchange chromatography. The relative phosphorylation represents fraction of all identified residues, mass of which corresponds to phosphorylation as measured by LC-MS/MS with Mascot score > 32 and MD score >= 5. The data show that the residues in the N-terminal domain are gradually more phosphorylated with increasing IEX elution volume, whereas the phosphorylation at the activating loop remains constant. The different colors represent three biological triplicates. (B) Activity of individual PfGSK3 fractions separated by ion exchange chromatography. The data show a general trend towards more active protein with increasing phosphorylation in the N-terminal domain. The activity was measured in biological triplicates and at different PfGSK3 concentration.

Provided that the PfGSK3 N-terminus is indispensable for its function, we suspected that this region could also be a potential site of PfGSK3 regulation. Therefore, we analyzed the individual separated fractions from IEX chromatography (Fig 1C) by LC/MS-MS analysis (Fig 6A). Here we observed that the extent of phosphorylation of residues in the activating loop of PfGSK3 is comparable between the four fractions, while significant differences in the N-terminal phosphorylation pattern are obvious. Specifically, all N-terminal residues are significantly less phosphorylated in the first fraction (F1), while in the second fraction (F2), the extent of phosphorylation of residues Y39 and S40 is strongly increased and the third and fourth fractions (F3 and F4) have generally higher level of N-terminal phosphorylation, which is most pronounced on residues Y39, S40, S42 and S43. Off note, although only S40 is well conserved across *Plasmodium* species, all of them contain multiple N-terminal residues that are potential sites of modification by phosphorylation (SFig 6). To assess the relationship between the activity of PfGSK3 and its phosphorylation pattern, we made use of the differences in N-terminal phosphorylation between the individual IEX fractions and measured their kinase activity (Fig 6B). PfGSK3 activity rises with an increase in N-terminal

phosphorylation, with a significant difference comparing the first and fourth fraction. Therefore, our data support that the phosphorylation of the PfGSK3 N-terminus enhances its activity. Taken together, the presence of the PfGSK3 N-terminal domain is crucial for the protein to exhibit autophosphorylation and thus its solubility, whereas phosphorylation of the N-terminal residues, mainly Y39, S40, S42 and S43 increases the activity of PfGSK3.

Discussion

Identification and characterization of novel malaria drug targets remain an important goal as resistant malaria species emerge. PfGSK3 has been identified as a drug target because its inhibitors have been shown to possess a potent antiplasmodial activity. Understanding its function is therefore important for further development of its inhibitors.

Here, we have shown that PfGSK3 can be expressed and purified from *E. coli* as a folded and functional protein. So far, the protein resisted our crystallization efforts, presumably due to its heterogeneity: we have shown that PfGSK3 is heterogeneously phosphorylated and is able to form high-MW particles. Moreover, the PfGSK3 N-terminus is predicted to be disordered, which represents another obstacle in protein crystallization. Removal of the N-terminal residues leads to an insoluble protein that lacks phosphorylation. This result alone indicates the importance of the PfGSK3 phosphorylation, and functional and potential regulatory role of its N-terminus.

The phosphorylation status of PfGSK3 specifically drew our attention because it is unlikely that endogenous bacterial kinases would phosphorylate a recombinantly expressed protein to such extent. Indeed, we have shown that PfGSK3 exhibits autophosphorylation, confirming the data in a recent publication, where the authors suspect that a high background in their phosphorylation assay could be attributed to autophosphorylation [48]. Human GSK3 has also been shown to exhibit autophosphorylation directly after the protein synthesis while being associated with chaperons, such as HSP90 [49]. Interestingly, a major contaminant during our PfGSK3 purification process was *E. coli* DnaK that shares 60% identity with *P. falciparum*

Hsp70-3, suggesting that chaperon-assisted autophosphorylation might be conserved across these taxa.

In general, the locations of PfGSK3 phosphorylation sites are similar to the pattern found in human GSK3, where the phosphorylation of the activation loop increases its activity and a phosphorylation at the N-terminal serine 9 inhibits its activity. However, the N-terminus of PfGSK3 is distinctively unique within the *Plasmodium* species and serine 9 of human GSK3 is not conserved across the *taxa*. Across *Plasmodium* species, serine 40 is conserved, while the phosphorylation of the same residue is most upregulated by PfGSK3 autophosphorylation, which indicates that autophosphorylation is also conserved across *Plasmodium* and therefore, is likely relevant *in vivo*. We hypothesize that during the invasion process, upstream kinases first phosphorylate PfGSK3, which boosts its activity by autophosphorylation while it further phosphorylates AMA1 C-terminus. Thus, autophosphorylation might play a role in a timely efficient regulation of red blood cell invasion. To confirm this hypothesis, future work analyzing the phosphorylation status of PfGSK3 in blood stages *in vivo* would be necessary.

Unfortunately, the mechanism of the activity upregulation upon PfGSK3 phosphorylation remains elusive. However, an *in silico* secondary structure prediction offers a possibility of large structural changes upon N-terminal phosphorylation. The N-terminus of PfGSK3 is predicted to be unfolded, but replacing the phosphorylated N-terminal residues with phosphomimetic glutamates leads to high increase in the calculated probability of coiled-coil formation. Indeed, the phosphorylation was previously suggested to destabilize the coiled-coil association [50,51] and stabilization of coiled-coil interaction was described in genetically engineered proteins [52].

It is worth mentioning that the N-termini of GSK3 homologues are not well conserved even among the representatives of *Apicomplexa*, however, a number of residues that can be phosphorylated could still enable similar mode of regulation as we described here (SFig 4). In *Plasmodium*, serine 40 is conserved across all main species and residues that can be phosphorylated are also present on positions 38, 39 and 42 in most *Plasmodium species*,

suggesting that the mechanism of regulation via N-terminal phosphorylation is preserved in *Plasmodium* GSK3 proteins.

Additionally, we have revealed another means of PfGSK3 regulation: the bivalent heavy metal ions, such as zinc and copper, inhibit the activity of PfGSK3 by inducing a formation of heterogeneous high-MW particles. Interestingly, the ionic radii of these cations are in a small range of 1.09-1.21 Å, indicating that the effect is mediated through a specific binding site. The fact that the protein becomes more sensitive to the metal ions only after the cleavage of the N-terminal tag suggests, that this potential binding site could be located to the PfGSK3 N-terminus. Interestingly, partially overlapping set of bivalent cations inhibiting human GSK3 can be found in the literature, including lithium, beryllium, zinc, copper and mercury [17,53,54]. The lithium and beryllium ions inhibit GSK3 by competing with magnesium ions that mediate ATP binding [54–56], but they have not been shown to induce a formation of high-MW particles. The inhibition mechanism of heavy metal cations thus remains unknown and to our knowledge, there is no structural information of any GSK3 homolog with a bound heavy metal ion in the PDB database. Importantly, as the high-MW particles can dissociate back into active PfGSK3 monomers, it cannot be excluded that this regulation is applied *in vivo*. The concentration of zinc in *Plasmodium* cytosol peak at the late blood stages and should be sufficient to induce the high-MW PfGSK3 species [57]. Moreover, the concept of “autoinhibitory polymerization” has already been described in detail for another serine/threonine kinase, CK2 [58]. On the other hand, this effect might be dependent on the phosphorylation status of PfGSK3 and PfGSK3 could also be stabilized by unknown interacting proteins that could prevent the formation of high-MW particles. It therefore remains to be evaluated experimentally *in vivo* whether the PfGSK3 high-MW particles play a role in native environment.

In conclusion, our work represents the first insights into the function of PfGSK3 and its regulation by phosphorylation and heavy metals, showing that N-terminal phosphorylation

boosts the kinase activity, whereas the heavy metals induce formation of high-MW species that are inactive.

Materials and methods

Cloning and mutagenesis

The vector with N-terminally His-tagged PfGSK3 was generated by PCR amplification of the GSK3 coding sequence from *P. falciparum* cDNA followed by Ligation Independent Cloning into HindIII/KpnI-cleaved plasmid pOPIN F [59] using In-Fusion HD EcoDry Cloning Kit (Takara Clontech) according to the manufacturer's instructions. The mutants S226A, Y229A and S226A/Y229A were generated by overlap extension PCR amplification from the original vector and Ligation Independent Cloning as described above. The wild-type protein and the mutant K96A cloned in pET28a vector were ordered from GenScript. The N-terminally truncated constructs were cloned by amplifying the sequence from the original vector and subcloning into BsaI-cleaved plasmid pNIC28_Bsa4 by SLiCE cloning.

PfGSK3 expression and purification

E. coli C41(DE3) culture transformed with PfGSK3 vector was grown in TB supplemented with ampicillin at 37°C, induced at O.D.=0.7 with 0.5 mM IPTG and harvested after 4 hours. The pellets were resuspended in 5 mL of lysis buffer (20mM NaP pH 7.5, 300mM NaCl, 15mM imidazole, 5% glycerol, 0.5mM TCEP, 1 mg/ml lysozyme, 5 U/ml DNase, 1 Roche protease inhibitor tablet/100mL) per 1 g of culture. The suspension was passed 3x through emulsifier at 15 000 psi, centrifuged at 19 000 x g and the supernatant was loaded on a His-Trap column. The column was washed with buffer containing 40 mM imidazole and the protein was eluted at increasing imidazole concentration gradient. The protein peak was concentrated (10 kDa concentrator) with addition of 1 mg of 3C protease per 3 L of culture volume and further separated on Superdex 200 HiLoad column (GE Healthcare) with 50mM Tris pH 8.0, 20mM NaCl, 0.5mM TCEP. The protein peak was concentrated again, loaded on ResourceQ anion exchange column (GE Healthcare) and eluted with increasing concentration of NaCl. The peaks of PfGSK3 were pooled together, concentrated, dialyzed against the final (typically 20mM Tris pH 8.0, 100mM NaCl, 0.5mM TCEP) overnight and flash-frozen. All purification steps were performed at 4°C.

Expression and purification of insoluble PfGSK3 constructs

The insoluble PfGSK3 constructs (K96A, S226A/Y229A, ΔN64) were expressed as described for the wild-type PfGSK3. The pellets of harvested bacteria were resuspended in 5 mL of lysis buffer (20mM NaP pH 7.5, 300mM NaCl, 15mM imidazole, 5% glycerol, 0.5mM TCEP, 1 mg/ml lysozyme, 5 U/ml DNase, 1 Roche protease inhibitor tablet/100mL) per 1 g of culture. The suspension was passed 3x through emulsifier at 15 000 psi, centrifuged at 19 000 x g and the pellet was dissolved in buffer (20mM NaP (pH 7.5), 300mM NaCl, 25mM imidazole, 5% glycerol, 6M urea, 0.5mM TCEP) and incubated with 1 mL NiNTA bead slurry (Sigma) for 1h, 4°C. The beads were subsequently washed twice with 10 ml of the same buffer and subsequently eluted with 5 ml of the same buffer containing 250 mM imidazole. The protein was concentrated (10 kDa c/o) and the sample for mass spectrometry was separated on SDS-PAGE gel.

Small-scale expression tests

4 mL of TB supplemented with ampicillin was inoculated with a pre-cultured *E. coli* C41(DE3) that had been transformed with the tested plasmids. The cultures were grown at 37°C until the O.D.=0.6 and further either induced with 0.5mM IPTG and grown for 4 hours at 37°C or 25°C, or induced with 0.1mM IPTG and grown for 16h at 18°C. O.D. was measured and culture volume corresponding to O.D. of 2.0/mL was harvested. The pellets were resuspended in 400 uL of lysis buffer (lysate fractions), then lysed by 10 min of vortexing with 100 µl glass beads and the cell debris were removed by centrifugation (soluble fractions). 5 uL of each fraction was mixed with 10 uL of loading dye and run on SDS-PAGE in duplicates. One gel of each set of samples was blotted on Western blot membranes. The membranes were blocked by 1% BSA, washed, incubated for 1h with anti-His antibodies coupled to horseradish peroxidase (ThermoFisher), washed and the protein bands were visualised by SuperSignal™ Western Blot Substrates Femto and Pico (ThermoFisher) in 1:10 ratio.

Thermal shift assay

The stability of PfGSK3 under different buffer conditions and additives was measured by nanoDSF (Prometheus NT.48, NanoTemper Technologies, GmbH). The protein concentration was adjusted to 1 mg/ml. 10 µl of samples was loaded in the glass capillaries and heated from 20 °C to 90 °C with a heating rate of 1 °C/min. The fluorescence signals with excitation wavelength of 280 nm and emission wavelengths of 330 and 350 nm were recorded and the melting temperature was calculated as maximum of the derivative of the fluorescence at 330 nm.

Circular dichroism

Circular dichroism was measured on a Chirascan CD spectrometer (Applied Photophysics). The protein concentration was adjusted to 1 µM in 10 µM by 2 mM NaP (pH 7.5), 4 mM NaCl, 0.05 mM TCEP prior to the measurement. The circular dichroism was measured 10x between 185 nm and 260 nm with 1 nm step in 1 mm quartz cuvette and analyzed by CDSSTR algorithm [60,61] using DichroWeb [62].

Analytical size exclusion chromatography

The analytical size exclusion chromatography was performed on Agilent Bio-LC system using a home-packed Superose 6 column (25 µl sample) or a Superose 6 Increase column (100 µl sample). PBS supplemented with 0.5mM TCEP was used as a mobile phase and the resulting chromatograms were recording as fluorescence signal with excitation wavelength of 280 nm and emission wavelength of 350 nm.

Autophosphorylation IEX assay

0.5 mg/ml of PfGSK3 was incubated either at 25°C with 200 µM ATP and 5mM $MgCl_2$ or at 37°C with 2mM ATP and 10mM $MgCl_2$ overnight. The samples were subsequently dialyzed at 4°C overnight in 50mM Tris pH 8.0, 20mM NaCl, 0.5mM TCEP. The samples were analyzed by analytical ion exchange chromatography using a Resource Q column (GE Healthcare) and the 1260 Infinity Bio-inert high-performance liquid chromatography system (Agilent Technologies) at 10 °C. The system and column were equilibrated in 50mM Tris pH 8.0,

20mM NaCl, 0.5mM TCEP. 100 uL of sample was injected and eluted with an increasing concentration of NaCl. The system was run at 0.2 ml/min ad the elution profile was analyzed by UV fluorescence detector with absorbance at 280 nm and emission at 350 nm.

Small angle X-ray scattering

The SAXS data were collected at the P12 BioSAXS Beamline at the PETRAIII storage ring (c/o DESY, Hamburg, Germany). The concentrated sample was dialyzed against (20 mM Tris pH 8.0, 100 mM NaCl, 0.5 mM TCEP overnight at 4 °C. The X-ray scattering data were measured in an on-line SEC-SAXS mode, using a SD200 Increase column (GE Healthcare) at 0.5 ml/min with 1 frame recorded per second. The automatically processed data were further analyzed using the ATSAS suite [63] programs CHROMIXS and PRIMUS to determine the molar mass, radius of gyration and distance distribution, and CRYSOL and CORAL to compare the data with the structural models and calculate X^2 values.

Kinase assay

To investigate enzymatic activity of recombinant PfGSK3 a commercial luminescence-based kinase assay (KinaseGlo Plus, Promega) was used as previously described (Schweda et al., 2020). Briefly, 20 ng PfGSK3, 12 µM GS-1 peptide substrate (residues 636-661 of the human glycogen synthase 1; sequence YRRAAVPPSPSLSRHSSPHQ(pS)EDEEE; pS = prephosphorylated serine, Promega) and 6 µM ATP (UltraPure, Promega) in kinase reaction buffer (40 mM Tris-HCl pH 7.5; 20 mM $MgCl_2$; 0.1 mg/mL BSA) were used as standard reaction mix (final reaction volume 5 µL). The kinase reaction was performed for 30 min at 30°C followed by incubation with KinaseGlo reagent according to the manufacturer's instructions. The reaction was transferred to a solid white 384-well plate (NUNC, ThermoFisher) and the luminescence signal was measured in an EnVision Multilable Plate Reader (PerkinElmer, Integration time 0.5 sec/well). To investigate reaction kinetics kinase and substrate concentrations were varried between 0 and 100 ng or 0 and 15 µM respectively.

Mass spectrometry

PfGSK3 samples were isolated from SDS-PAGE gels, cleaved by trypsin protease and the resulting peptides were purified and analyzed by LC-MS/MS approach. Mascot (v2.2.07) were used to process the acquired data, whereas the following modifications were included into the search parameters: Carbamidomethylm acetyl (protein N-term), oxidation and phosphorylation. For the full scan (MS1) a mass error tolerance of 10 ppm and for MS/MS (MS2) spectra of 0.02 Da was set. Further parameters were set: Trypsin as protease with an allowance of maximum two missed cleavages: a minimum peptide length of seven amino acids; at least two unique peptides were required for a protein identific9ation. Relative phosphorylation was calculated as the number of peptides identified to be phosphorylated with Mascot score >= 32 and delta.mod number >=5 divided by a total number of times the site was identified.

Negative stain electron microscopy

4 ul of 0.01 mg/ml protein solution were applied to carbon-coated Cu/Pd grids and incubated for 1 min. Excess liquid was removed with the filter paper (Whatman), and the grid was washed with twice with water for 5 s and 30 s. 4 ul of 2% uranyl acetate were applied twice for 5 s and 30 s. The micrographs were collected on alos L120C (CSSB Cryo-EM mulituser facility) with 70 000x magnification.

Bioinformatics

The homologous protein sequences were aligned with the program MAFFT [64] and depicted in Jalview. The secondary structure was predicted by Jpred [65].

Acknowledgments

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support with nanoDSF and DLS measurements. We would like to thank Dr. Haydyn Mertens for support operating the BioSAXS beamline P12 (EMBL, Hamburg) and help with the data analysis. This work was funded by a grant from the Joachim Herz foundation (Nr. 800026).

References

1. World Health Organization. World Malaria Report. https://www.who.int/malaria [Internet]. 2019. (accessed 15 Dec. 2019). Available from: https://www.who.int/publications-detail/world-malaria-report-2019
2. Haldar K, Bhattacharjee S, Safeukui I. Drug resistance in Plasmodium. Nat Rev Microbiol [Internet]. 2018;16(3):156–70. Available from: http://dx.doi.org/10.1038/nrmicro.2017.161
3. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa G, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017;16(1):19–34.
4. Cabrera DG, Horatscheck A, Wilson CR, Basarab G, Eyermann CJ, Chibale K. Plasmodial Kinase Inhibitors: License to Cure? J Med Chem. 2018;61(18):8061–77.
5. PARKER PJ, CAUDWELL FB, COHEN P. Glycogen Synthase from Rabbit Skeletal Muscle; Effect of Insulin on the State of phosphorylation of the Seven Phosphoserine Residues in vivo. Eur J Biochem. 1983;
6. Cross DAE, Alessi DR, Cohen P, Andjelkovich M, Hemmings BA. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. Nature. 1995;
7. Welsh GI, Proud CG. Glycogen synthase kinase-3 is rapidly inactivated in response to insulin and phosphorylates eukaryotic initiation factor eIF-2B. Biochem J. 1993;
8. Diehl JA, Cheng M, Roussel MF, Sherr CJ. Glycogen synthase kinase-3β regulates cyclin D1 proteolysis and subcellular localization. Genes Dev. 1998;
9. Yamamoto H, Kishida S, Kishida M, Ikeda S, Takada S, Kikuchi A. Phosphorylation of axin, a Wnt signal negative regulator, by glycogen synthase kinase-3β regulates its stability. J Biol Chem. 1999;
10. Turenne GA, Price BD. Glycogen synthase kinase3 beta phosphorylates serine 33 of p53 and activates p53's transcriptional activity. BMC Cell Biol. 2001;
11. Ikeda S, Kishida S, Yamamoto H, Murai H, Koyama S, Kikuchi A. Axin, a negative regulator of the Wnt signaling pathway, forms a complex with GSK-3β and β-catenin and promotes GSK-3β-dependent phosphorylation of β-catenin. EMBO J. 1998;
12. Hanger DP, Hughes K, Woodgett JR, Brion JP, Anderton BH. Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: Generation of paired helical filament epitopes and neuronal localisation of the kinase. Neurosci Lett. 1992;
13. Yang S -D, Song J -S, Yu J -S, Shiah S -G. Protein Kinase FA/GSK-3 Phosphorylates on Ser235-Pro and Ser404-Pro that Are Abnormally Phosphorylated in Alzheimer's Disease Brain. J Neurochem. 1993;
14. Aplin AE, Gibb GM, Jacobsen JS, Gallo JM, Anderton BH. In vitro phosphorylation of the cytoplasmic domain of the amyloid precursor protein by glycogen synthase kinase-3β. J Neurochem. 1996;
15. Ryder J, Su Y, Liu F, Li B, Zhou Y, Ni B. Divergent roles of GSK3 and CDK5 in APP processing. Biochem Biophys Res Commun. 2003;
16. Ferrao Santos S, Tasiaux B, Sindic C, Octave JN. Inhibition of neuronal calcium oscillations by cell surface APP phosphorylated on T668. Neurobiol Aging. 2011;
17. Eldar-Finkelman H, Martinez A. GSK-3 Inhibitors: Preclinical and Clinical Focus on CNS. Front Mol Neurosci. 2011;4(October):1–18.
18. Bourel E, Grieco SF, Jope RS. Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. 2016;114–31.
19. Jope RS. Glycogen Synthase Kinase-3 in the Etiology and Treatment of Mood Disorders. Front Mol Neurosci. 2011;
20. Singh KK. An emerging role for Wnt and GSK3 signaling pathways in schizophrenia. Clin Genet. 2013;
21. De Sarno P, Axtell RC, Raman C, Roth KA, Alessi DR, Jope RS. Lithium Prevents and

Ameliorates Experimental Autoimmune Encephalomyelitis. J Immunol. 2008;

22. Del Ser T, Steinwachs KC, Gertz HJ, Andrés M V., Gómez-Carrillo B, Medina M, et al. Treatment of Alzheimer's disease with the GSK-3 inhibitor tideglusib: A pilot study. J Alzheimer's Dis. 2013;
23. Lovestone S, Boada M, Dubois B, Hüll M, Rinne JO, Huppertz HJ, et al. A phase II trial of tideglusib in alzheimer's disease. J Alzheimer's Dis. 2015;
24. Prinz B, Harvey KL, Wilcke L, Ruch U, Engelberg K, Biller L, et al. Hierarchical phosphorylation of apical membrane antigen 1 is required for efficient red blood cell invasion by malaria parasites. Sci Rep [Internet]. 2016;6(June):34479. Available from: http://www.nature.com/articles/srep34479
25. Narum DL, Thomas AW. Differential localization of full-length and processed forms of PF83/AMA-1 an apical membrane antigen of Plasmodium falciparum merozoites. Mol Biochem Parasitol. 1994;
26. Solyakov L, Halbert J, Alam MM, Semblat J-P, Dorin-Semblat D, Reininger L, et al. Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. Nat Commun [Internet]. 2011;2:565. Available from: http://www.nature.com/doifinder/10.1038/ncomms1558
27. Lasonder E, Treeck M, Alam M, Tobin AB. Insights into the Plasmodium falciparum schizont phospho-proteome. Microbes Infect. 2012;
28. Treeck M, Sanders JL, Elias JE, Boothroyd JC. The phosphoproteomes of plasmodium falciparum and toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. Cell Host Microbe. 2011;10(4):410–9.
29. Treeck M, Zacherl S, Herrmann S, Cabrera A, Kono M, Struck NS, et al. Functional analysis of the leading malaria vaccine candidate AMA-1 reveals an essential role for the cytoplasmic domain in the invasion process. PLoS Pathog. 2009;
30. Mordue DG, Desai N, Dustin M, Sibley LD. Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring. J Exp Med. 1999;
31. Vulliez-Le Normand B, Tonkin ML, Lamarque MH, Langer S, Hoos S, Roques M, et al. Structural and functional insights into the malaria parasite moving junction complex. PLoS Pathog. 2012;
32. Cao J, Kaneko O, Thongkukiatkul A, Tachibana M, Otsuki H, Gao Q, et al. Rhoptry neck protein RON2 forms a complex with microneme protein AMA1 in Plasmodium falciparum merozoites. Parasitol Int. 2009;
33. Besteiro S, Michelin A, Poncet J, Dubremetz JF, Lebrun M. Export of a Toxoplasma gondii rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. PLoS Pathog. 2009;
34. Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC. Identification of the moving junction complex of Toxoplasma gondii: A collaboration between distinct secretory organelles. PLoS Pathog. 2005;
35. Alexander DL, Arastu-Kapur S, Dubremetz JF, Boothroyd JC. Plasmodium falciparum AMA1 binds a rhoptry neck protein homologous to TgRON4, a component of the moving junction in Toxoplasma gondii. Eukaryot Cell. 2006;
36. Leykauf K, Treeck M, Gilson PR, Nebl T, Braulke T, Cowman AF, et al. Protein kinase a dependent phosphorylation of apical membrane antigen 1 plays an important role in erythrocyte invasion by the malaria parasite. PLoS Pathog. 2010;
37. Hodder AN, Crewther PE, Matthewll MLSM, Reid GE, Moritz RL, Simpson RJ, et al. The disulfide bond structure of Plasmodium apical membrane antigen-1. J Biol Chem. 1996;
38. Patel A, Perrin AJ, Flynn HR, Bisson C, Withers-Martinez C, Treeck M, et al. Cyclic AMP signalling controls key components of malaria parasite host cell invasion machinery. PLoS Biol. 2019;

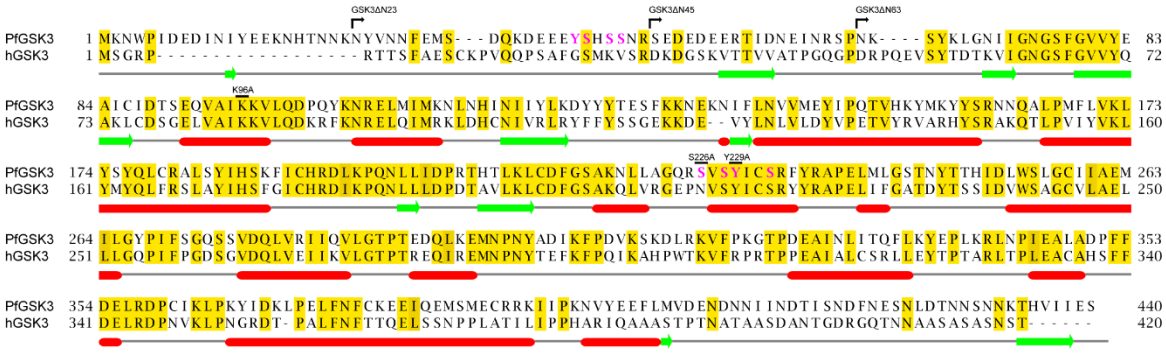
39. Wilde ML, Triglia T, Marapana D, Thompson JK, Kouzmitchev AA, Bullen HE, et al. Protein kinase A is essential for invasion of Plasmodium falciparum into human erythrocytes. MBio. 2019;10(5):1–15.
40. Fugel W, Oberholzer AE, Gschloessl B, Dzikowski R, Pressburger N, Preu L, et al. 3,6-Diamino-4-(2-halophenyl)-2-benzoylthieno[2,3-b]pyridine-5-carbonitriles are selective inhibitors of Plasmodium falciparum glycogen synthase kinase-3. J Med Chem [Internet]. 2013;56(1):264–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23214499
41. Masch A, Nasereddin A, Alder A, Bird MJ, Schweda SI, Preu L, et al. Structure-activity relationships in a series of antiplasmodial thieno[2,3-b]pyridines. Malar J [Internet]. 2019;18(1):1–10. Available from: https://doi.org/10.1186/s12936-019-2725-y
42. Sutherland C, Leighton IA, Cohen P. Inactivation of glycogen synthase kinase-3β by phosphorylation: New kinase connections in insulin and growth-factor signalling. Biochem J. 1993;
43. OSOLODKIN DI, ZAKHAREVICH N V., PALYULIN VA, DANILENKO VN, ZEFIROV NS. Bioinformatic analysis of glycogen synthase kinase 3: human versus parasite kinases. Parasitology [Internet]. 2011;138(06):725–35. Available from: http://www.journals.cambridge.org/abstract_S0031182011000151
44. Dajani R, Fraser E, Roe SM, Yeo M, Good VM, Thompson V, et al. Structural basis for recruitment of glycogen synthase kinase 3β to the axin-APC scaffold complex. EMBO J. 2003;
45. Macek B, Gnad F, Soufi B, Kumar C, Olsen J V., Mijakovic I, et al. Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics. 2008;7(2):299–307.
46. Deutscher J, Saier MH. Ser/Thr/Tyr protein phosphorylation in bacteria - For long time neglected, now well established. J Mol Microbiol Biotechnol. 2006;9(3–4):125–31.
47. Mann NH. Protein Phosphorylation in Escherichia coli and Purification of a Protein Kinase. J Biol Chem. 1984;259(1):526–33.
48. Bui HTN, Niederwieser I, Bird MJ, Dai W, Brancucci NMB, Moes S, et al. Mapping and functional analysis of heterochromatin protein 1 phosphorylation in the malaria parasite Plasmodium falciparum. Sci Rep. 2019;
49. Lochhead PA, Kinstrie R, Sibbet G, Rawjee T, Morrice N, Cleghon V. A chaperone-dependent GSK3β transitional intermediate mediates activation-loop autophosphorylation. Mol Cell. 2006;24(4):627–33.
50. Szilák L, Moitra J, Krylov D, Vinson C. Phosphorylation destabilizes α-helices. Nat Struc. 1997;4(2):112–4.
51. Derewenda U, Tarricone C, Choi WC, Cooper DR, Lukasik S, Perrina F, et al. The Structure of the Coiled-Coil Domain of Ndel1 and the Basis of Its Interaction with Lis1, the Causal Protein of Miller-Dieker Lissencephaly. Structure. 2007;15(11):1467–81.
52. Szilák L, Moitra J, Vinson C. Design of a leucine zipper coiled coil stabilized 1.4 kcal mol-1 by phosphorylation of a serine in the e position. Protein Sci. 1997;6(6):1273–83.
53. Ilouz R, Kaidanovich O, Gurwitz D, Eldar-Finkelman H. Inhibition of glycogen synthase kinase-3β by bivalent zinc ions: Insight into the insulin-mimetic action of zinc. Biochem Biophys Res Commun. 2002;
54. Ryves WJ, Harwood AJ, Dajani R, Pearl L. Glycogen synthase kinase-3 inhibition by lithium and beryllium suggests the presence of two magnesium binding sites. Biochem Biophys Res Commun. 2002;
55. Ryves WJ, Harwood AJ. Lithium inhibits glycogen synthase kinase-3 by competition for magnesium. Biochem Biophys Res Commun. 2001;
56. Mudireddy SR, Abdul ARM, Gorjala P, Gary RK. Beryllium is an inhibitor of cellular GSK-3β that is 1,000-fold more potent than lithium. BioMetals. 2014;
57. Marvin RG, Wolford JL, Kidd MJ, Murphy S, Ward J, Que EL, et al. Fluxes in 'Free' and

Total Zinc are Essential for Progression of Intraerythrocytic Stages of Plasmodium falciparum. Bone. 2012;19(6):731–41.

58. Lolli G, Pinna LA, Battistutta R. Structural determinants of protein kinase CK2 regulation by autoinhibitory polymerization. ACS Chem Biol. 2012;7(7):1158–63.
59. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. 2007;
60. Manavalan P, Johnson WC. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem. 1987;
61. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem. 2000;
62. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers. 2008;
63. Franke D, Petoukhov M V., Konarev P V., Panjkovich A, Tuukkanen A, Mertens HDT, et al. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr. 2017;
64. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;
65. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 2008;

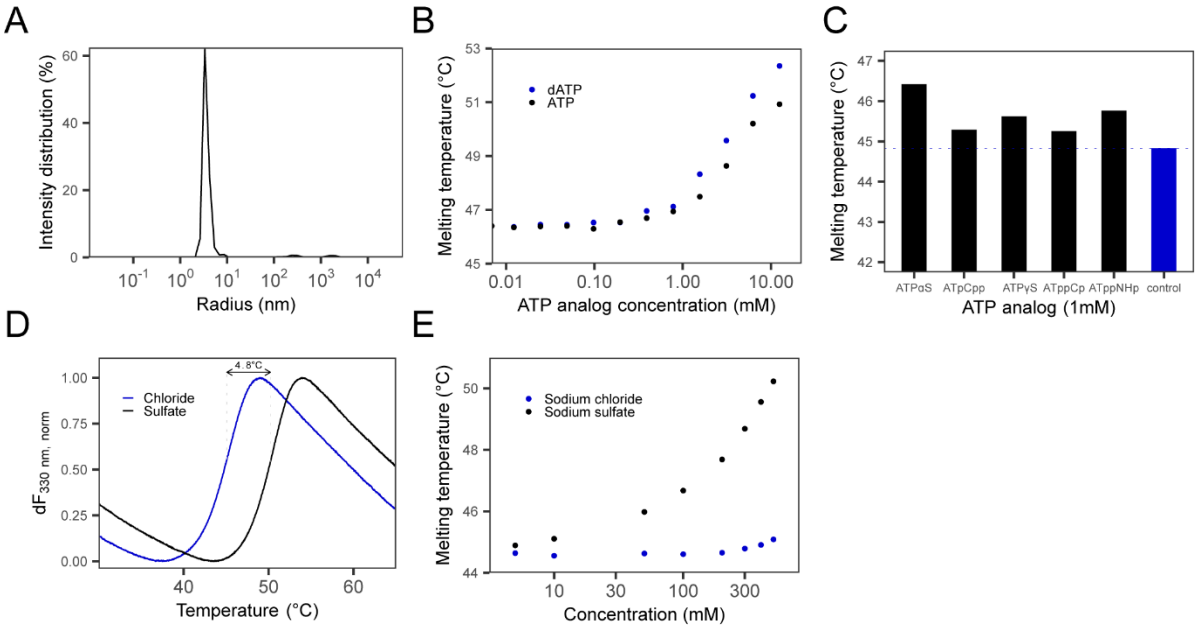
Supplementary figures

Supplementary figure 1



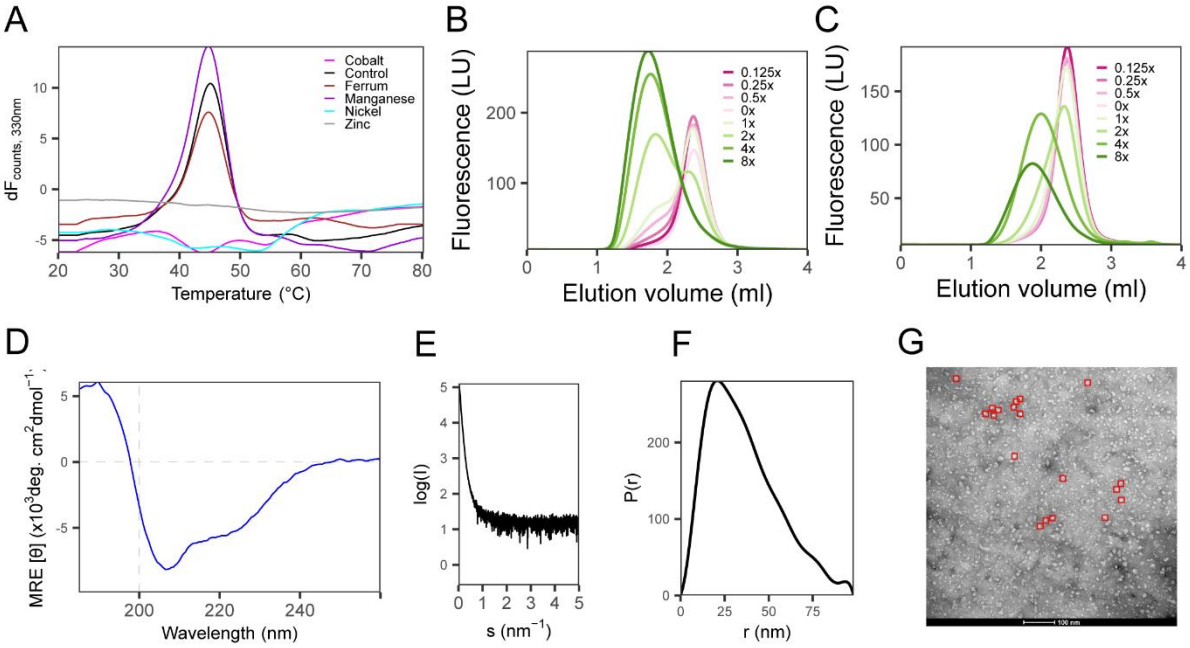
Sequence alignment of PfGSK3 with human GSK3. The conserved residues are highlighted in yellow. Residues mutated in this study and the N-truncated constructs are indicated. Residues that are phosphorylated are shown in magenta. The PfGSK3 secondary structure prediction by JPred is indicated below the sequence. The sequences were aligned using MAFFT and plotted in Jalview.

Supplementary figure 2



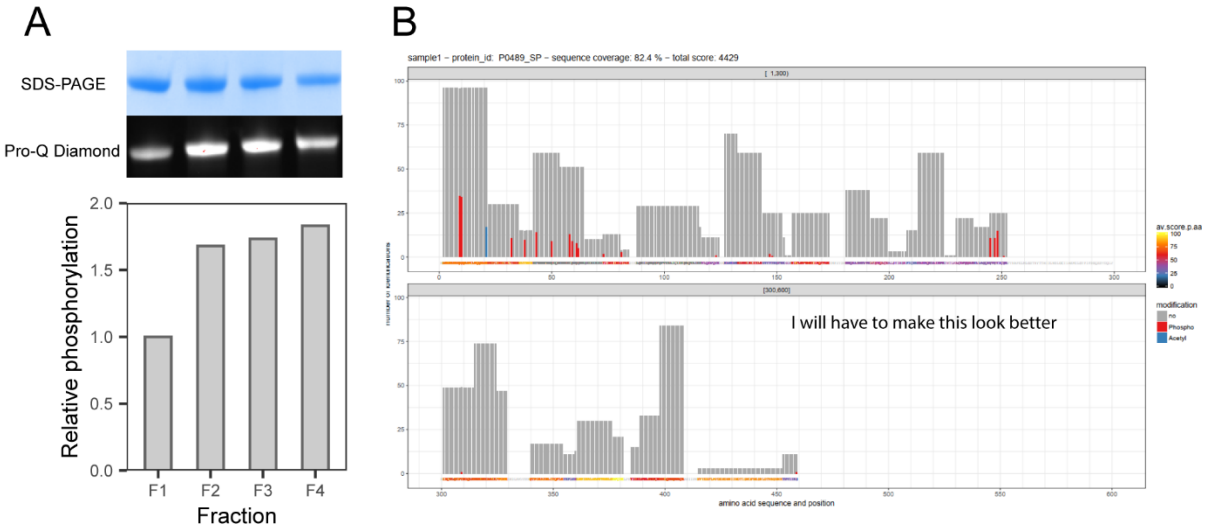
(A) Analysis of Ddnamic light scattering data of PfGSK3 reveals a high degree of homogeneity with a calculated molecular weight equal to the expected value of 52 kDa. (B) Stability dependence of the PfGSK3 melting temperature measured by nano differential scanning fluorimetry (nDSF) in the presence of different concentrations of ATP or dATP c. Melting temperature increases with the concentration of ATP or dATP, indicative for binding ATP and its analogs. (C) Melting temperature of PfGSK3 measured by nDSF with or without 1mM non-hydrolyzable ATP analogs. (D) Thermal unfolding profiles of PfGSK3 in presence of 500mM sodium chloride or sodium sulfate show that PfGSK3 is preferentially stabilized by sulfate ions. (E) Stabilization of PfGSK3 by sulfate ions is concentration-dependent, indicating that stabilization effect is caused by specific sulfate ion binding.

Supplementary figure 3



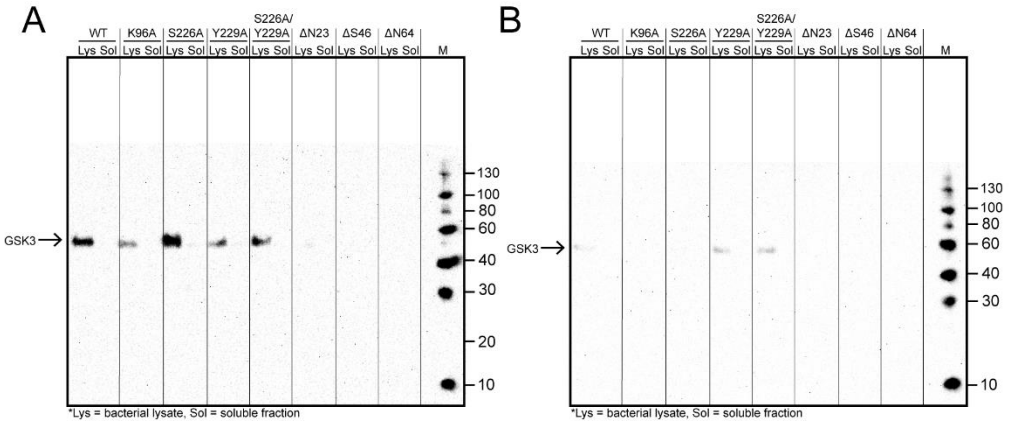
(A) Effect of heavy metals on thermal unfolding of PfGSK3. The unfolding trace was measured by nano differential scanning fluorimetry at a protein concentration of 1 mg/ml and a metal concentration of 1mM. Whereas a clear unfolding transition is observed in PfGSK3, some heavy metal ions suppress the thermal unfolding. (B-C) Effect of zinc chloride (B) or copper chloride (C) on the elution profile of PfGSK3 on a Superose 6 column. The experiment shows that these metal ions induce the formation of high-MW particles in a concentration-dependent manner. The numbers indicate the molar access of metal over the protein. (D) Circular dichroism spectrum of PfGSK3 in the presence of zinc shows that the secondary structure is retained even in the induced high-MW particles. (E) SAXS curve of PfGSK3 in the presence of zinc shows that the high-MW particles are large (>1MDa) and heterogeneous. (F) The distance distribution profile calculated from the PfGSK3/Zn SAXS data indicate the heterogeneity of the protein. (G) Negative stain electron microscopy image of PfGSK3 high-MW particles induced by copper further stress the large heterogeneity.

Supplementary figure 4



(A) The intensity ratio of bands between SDS-PAGE and Pro-Q diamond staining for individual IEX fraction shows that the later eluting species of PfGSK3 are more phosphorylated. (B) LC MS-MS analysis of the PfGSK3 phosphorylation sites reveals heterogeneous phosphorylation at the N-terminus and in the activation loop of PfGSK3.

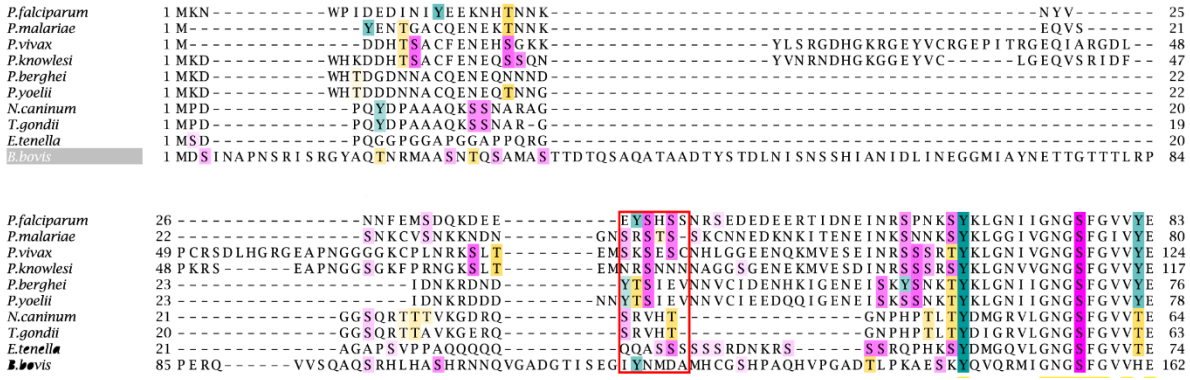
Supplementary figure 5



(A-B) Western blot analysis of expression test of PfGSK3 mutants in its ATP binding site (K96) or activating loop (S226 and Y229) and of PfGSK3 N-terminally truncated constructs at 25°C (A) and 18°C (B). The results show that the lower temperatures decrease the expression level and solubility of the proteins.

Supplementary figure 6

A



Sequence alignment of PfGSK3 with homologous proteins from other *Plasmodia* and *Apicomplexa*. Only N-terminal residues are shown. Serine, threonine and tyrosine residues are depicted with magenta, yellow and cyan background, respectively. The region around regulatory N-terminal residues is in a red box. The alignment shows that residue S40 and neighboring residues are conserved in *Plasmodium* but not within *Apicomplexa*.