UBIQUITIN RELATED PROTEINS FACILITATE IMPORT INTO THE APICOPLAST OF TOXOPLASMA GONDII

By

JUSTIN FELLOWS

(Under the Direction of Boris Striepen)

ABSTRACT

Apicomplexans are single celled eukaryotes and parasites that cause veterinarian and human diseases. Important human diseases caused by these parasites include malaria, cryptosporidiosis, toxoplasmosis and babesiosis. Most members of the phylum (with the exception of Cryptosporidium) are characterized by the presence of a non-photosynthetic plastid called the apicoplast. The apicoplast is derived from a secondary endosymbiosis event where a red algal endosymbiont was engulfed by a heterotrophic organism that resulted in the four membranes that now surround the apicoplast. Although the organelle has lost its photosynthetic properties, the apicoplast is the location of important biochemical processes including isoprenoid, fatty acid and heme biosynthesis. The apicoplast's divergent origin and its essential biochemical processes suggest the organelle as an ideal therapeutic target. It is estimated that almost 500 endosymbiont genes were transferred to the host's nucleus. Consequently, these now nuclear encoded apicoplast proteins must be transported back across the four membranes of the apicoplast to reach their site of action. Previous studies have demonstrated a series of distinct translocons that are dedicated to transport across each of the organelle's multiple membranes. Here

we focus on a system that is derived from the endoplasmic reticulum associated degradation (ERAD) pathway and has been repurposed for import across the second outermost membrane or periplastid membrane of the apicoplast. While previous research has demonstrated several components of the apicoplast ERAD machinery as critical for import, the crucial ubiquitin modifier typically associated with the ERAD machinery remained elusive. My dissertation provides the first evidence of an apicoplast specific ubiquitin-like protein and defines its function by generating a conditional mutant and establishing complementation which allowed definitive analysis to characterize the protein. My findings revealed that the apicoplast specific ubiquitin-like protein is significantly different from any known ubiquitin-like proteins, critical for parasite viability and essential for protein import across the periplastid membrane. I demonstrate that $CDC48_{AP}$, another component of the ERAD machinery that may interact with ubiquitin, is crucial for apicoplast protein import across the periplastid membrane and apicoplast biogenesis. We identify additional components of the apicoplast ubiquitin-like machinery including candidate ubiquitin ligases and deubiquitinases.

INDEX WORDS: Apicomplexans, *Toxoplasma gondii*, apicoplast, ubiquitin, ERAD

UBIQUITIN RELATED PROTEINS FACILITATE IMPORT INTO THE APICOPLAST OF TOXOPLASMA GONDII

By

JUSTIN FELLOWS

B.Sc, Southern Methodist University 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

UBIQUITIN RELATED PROTEINS FACILITATE IMPORT INTO THE APICOPLAST OF TOXOPLASMA GONDII

by

JUSTIN FELLOWS

Major Professor: Boris Striepen

Committee: Vasant Muralidharan
Silvia Moreno
Robert Sabatini

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

DEDICATION

This work is dedicated to my parents Joan Fellows and Richard Toupal, my brother Jason Fellows and to the rest of my family for their support during my time in graduate school.

ACKNOWLEDGEMENTS

I would like express my deepest gratitude to Dr. Boris Striepen who has been supportive and patient through every step of my scientific career. He has been a fantastic advisor who has helped shaped me into the scientist I am today. This work would not have been possible without his guidance.

I would also like to acknowledge my committee members Dr. Silvia Moreno, Dr. Robert Sabatini and Dr. Vasant Muralidharan. Their suggestions and advice over the years have been very helpful in the advancement of this project.

I would also like to thank all current and previous lab members. While the members are too numerous to name them all, I thank all of them for the encouragement, advice, friendship and memories. You all have made my time in the Striepen Lab more enjoyable. Lilach Sheiner deserves a special acknowledgment for the time she spent mentoring me which undeniably made me a better scientist. I would be remised if I didn't mention Carrie Brooks for all her hard work and time spent helping me from scientific to non-scientific problems. None of this work (or any of the Striepen Lab projects) would be possible without her help. I would also like to thank Dr. Julie Nelson and Dr. Muthugapatti Kandasamy for their help and training.

I would like to express my gratitude to all participants of the 2014 Biology of Parasitisim course for the memorable experience and for their friendship.

I would like to give a special thanks to all my friends in the graduate school who have made the past five years worthwhile and special. I am very grateful for your friendship

and value all the experiences we have shared. Thank you: Snehal Chaudhari, Zach Detweiler, Catherine Sullenberger, Riju Balachandran, Madhumati Mukherjee, Jared Wilmoth, Jackie Uribe, Sri Ramakrishnan, Jayesh Tendel, Robert Ng, Zhibo Ma, Panayiota Louka and Mayukh Guha.

Most importantly I would like to thank my family for their support and love. Thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Introduction

Numerous parasitic diseases including toxoplasmosis, malaria, cryptosporidiosis, cyclosporidiosis, babesiosis and east coast fever are caused by obligate intracellular parasites that belong to the phylum Apicomplexa. These parasites share unique characteristics and organelles that define the phylum including an apical complex used in host cell invasion, the inner membrane complex involved in motility and the apicoplast organelle (excluding *Cryptosporidium*). The apicoplast is an organelle related to plastids that originated in a secondary endosymbiosis event where a red algal endosymbiont was engulfed by a heterotrophic eukaryote resulting in the four membrane organelle [1]. The apicoplast has lost photosynthetic properties, but it is still necessary for parasite survival as it is the site of important pathways including fatty acid, isoprenoid and heme biosynthesis [1-3]. Therefore, the apicoplast is a prime candidate as a therapeutic target. It has been demonstrated that interfering with apicoplast transcription, translation and metabolic function results in parasite death which emphasizes the importance of studying the organelle [4-6]. Most apicoplast proteins are nuclear encoded and must be imported across the four membranes of the apicoplast. The molecular machinery that imports proteins across the second outermost membrane (periplastid membrane) is particularly fascinating and the system that facilitates transport has only been partially characterized.

Most of the insight into the molecular machinery that controls import into the apicoplast has been obtained by work on *Plasmodium falciparum* and *Toxoplasma gondii*. The life cycle of *T. gondii* relies on infection and transmission between intermediate and definitive hosts. The definitive host was an enigma until the identification of oocysts from cat feces and the discovery that sexual stages were occurring in the intestines of the family Felidae [7, 8]. Initially, parasites are released from tissue cysts and undergo asexual replication by endodyogeny in the small intestines of felines. This is followed by multiple replication cycles by endopolygeny and schizogony in the epithelial cells of the small intestine of felines to form schizonts [9]. Schizonts release the asexual stage merozoites which eventually form male and female gametes. The male gamete has two flagella which are used to fertilize the female gamete and initiate the formation of oocysts in the intestines. Oocysts are shed from the feline with feces and subsequently sporogony occurs resulting in oocysts containing two sporocysts with four sporozoites each [10]. This entire cycle in felines can be completed as quickly as 66 hours after feeding cats tissue cysts [11]. Oocysts are quite hardy and can survive up to 18 months in the environment [12]. Almost all warm blooded animals can be infected through the intake of oocysts and the bradyzoite stages from the environment and undercooked meat respectively [13]. These infected animals act as an intermediate host for the asexual stages of the parasite and transmission of the disease. Parasites can persist as two asexual stages during infection of intermediate hosts. The tachyzoite stage is the asexual proliferative form that rapidly invades cells, forms a parasitophorous vacuole inside the host cell, divides by endodyogeny and egresses from the host cell. The tachyzoites invade neighboring cells and continue to lyse host cells in

the process. This stage results in clinical manifestation of the disease and prompts a strong immune response. The immune response is often accompanied by the restriction of the parasite to the bradyzoite stage [14]. Bradyzoites are much slower growing stages that can persistent throughout an individual's life and form tissue cysts surrounded by a cyst wall that is resistant to pepsin and trypsin digestion [15].

It is estimated that ~30% of the human population is chronically infected with *T. gondii* [9]. Most infections are asymptomatic and toxoplasmosis in immunocompetent individuals only results in mild symptoms including fever and muscle aches. However, severe disease may occur in immunocompromised individuals which include patients undergoing immunosuppressive treatment during organ transplants, patients with acquired immunodeficiency syndrome (AIDS) and developing fetuses [16]. A weakened immune system results in the reactivation of dormant bradyzoites into the rapidly dividing tachyzoite stage. This is frequent in AIDS patients or organ transplant recipients which results in severe diseases including encephalitis and myocarditis [13, 17]. At one point 18% to 25% of AIDS patients in the US suffered from toxoplasmosis [18]. Another major concern is congenital toxoplasmosis attributed to the vertical transmission of the parasite from the mother to the unborn fetus during the primary infection of a pregnant woman. Congenital toxoplasmosis is often marked by the clinical symptoms of hydrocephalus, chorioretinitis, intracerebral calcification, mental retardation and in severe cases abortion [19]. Congenital toxoplasmosis can be highly prevalent with studies showing that 1 per 3000 births and 3.3 per 10,000 births in Brazil and France suffer from congenital toxoplasmosis respectively [20, 21].

The main drugs used to treat *T. gondii* infection are pyrimethamine and sulfonamides [22]. It was first described that sulfonamides had therapeutic effects on *T. gondii* in 1941, but the combined and synergistic treatment using sulfonamides and pyrimethamine first established in 1953 is still a mainstay of *T. gondii* treatment [23]. It was later shown that spiramycin is also an effective agent in treating the disease. Although spiramycin is considered less effective than the combination of sulfonamindes and pyrimethamine, spiramycin is often used to treat toxoplasmosis during pregnancy as spiramycin is safe for fetuses and newborns [24, 25]. Other drugs used to treat toxoplasmosis include azithromycin and clindamycin used in combination with pyrimethamine [22]. A commercial vaccine that uses a live non cyst forming strain of *T. gondii* is available for sheep [26]. However, there is no effective vaccine available to prevent toxoplasmosis for humans and there is a need for new drug development. Current drugs manage acute infection, but are unable to eradicate chronic infection. Drugs also suffer from adverse effects particularly when multiple or prolonged courses of treatment are required [27]. *T. gondii* is an ideal model organism to study the apicoplast due to its strong genetic tools and well established functional assays. Studies of the apicoplast in *T. gondii* may ultimately lead us to identifying novel targets suitable for the development of better treatment. In this work we experimentally demonstrate the importance of two endosymbiont derived proteins and provide evidence of their role in apicoplast protein import.

1.2 Structure of the Dissertation

In this dissertation I focus on a molecular machine involved in apicoplast protein import across the periplastid membrane that is related to the endoplasmic reticulum associated

degradation (ERAD) system. The ERAD pathway typically recognizes misfolded proteins in the secretory pathway, extracts them from the ER and marks the misfolded proteins with ubiquitin for subsequent degradation by the proteasome. I seek to understand how the ERAD derived proteins and ubiquitination machinery in the apicoplast have been retooled for import across the periplastid membrane of the apicoplast of *T. gondii*. This dissertation is presented in five chapters. The second chapter is a literary review of the present knowledge of the molecular machinery and mechanisms involved in protein import into the apicoplast and related complex plastid organelles. Chapter 3 presents the work to identify and characterize a novel plastid ubiquitin-like protein (PUBL) in the apicoplast. I demonstrate through genetic studies that PUBL and $CDC48_{AP}$ the molecular motor of the ERAD derived machinery are required for apicoplast protein import and apicoplast biogenesis. This work was submitted for publication as a manuscript to *mBio* and is currently going through a revision. In Chapter 4 I investigate the localizations of multiple candidates of potential apicoplast specific ubiquitin ligases and deubiquitinases. Chapter 5 describes future work needed to gain a better understanding of the biology of protein import into the apicoplast and presents an overall conclusion to the dissertation. In addition, there is an appendix chapter of a published article from *Traffic* which showcases the finding of the Toc75 translocon of the second innermost membrane of the apicoplast in *T. gondii* and presents the first experimental evidence that Toc75 mediates apicoplast protein import [28]. I contributed by performing several apicoplast import assays with the Toc75 conditional mutant. Overall, the dissertation provides a deeper understanding of a process that is essential for *T. gondii* growth and pathogenesis.

References

1. van Dooren GG, Striepen B: The Algal Past and Parasite Present of the Apicoplast. *Annual review of microbiology* 2013.
2. Mazumdar J, E HW, Masek K, C AH, Striepen B: Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in Toxoplasma gondii. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103(35):13192-13197.
3. Nair SC, Brooks CF, Goodman CD, Sturm A, McFadden GI, Sundriyal S, Anglin JL, Song Y, Moreno SN, Striepen B: Apicoplast isoprenoid precursor synthesis and the molecular basis of fosmidomycin resistance in Toxoplasma gondii. *The Journal of experimental medicine* 2011, 208(7):1547-1559.
4. Fichera ME, Roos DS: A plastid organelle as a drug target in apicomplexan parasites. *Nature* 1997, 390(6658):407-409.
5. Fleige T, Soldati-Favre D: Targeting the Transcriptional and Translational Machinery of the Endosymbiotic Organelle in Apicomplexans. *Current drug targets* 2008, 9(11):948-956.
6. Wiesner J, Reichenberg A, Heinrich S, Schlitzer M, Jomaa H: The plastid-like organelle of apicomplexan parasites as drug target. *Current pharmaceutical design* 2008, 14(9):855-871.
7. Ferguson DJ, Hutchison WM, Dunachie JF, Siim JC: Ultrastructural study of early stages of asexual multiplication and microgametogony of Toxoplasma

gondii in the small intestine of the cat. *Acta pathologica et microbiologica Scandinavica Section B: Microbiology and immunology* 1974, 82(2):167-181.

8. Hutchison WM: Experimental transmission of Toxoplasma gondii. *Nature* 1965, 206(987):961-962.

9. Tenter AM, Heckeroth AR, Weiss LM: Toxoplasma gondii: from animals to humans. *International journal for parasitology* 2000, 30(12-13):1217-1258.

10. Dubey JP: Toxoplasmosis - a waterborne zoonosis. *Veterinary parasitology* 2004, 126(1-2):57-72.

11. Dubey JP, Frenkel JK: Cyst-induced toxoplasmosis in cats. *The Journal of protozoology* 1972, 19(1):155-177.

12. Frenkel JK, Ruiz A, Chinchilla M: Soil survival of toxoplasma oocysts in Kansas and Costa Rica. *The American journal of tropical medicine and hygiene* 1975, 24(3):439-443.

13. Dubey JP: History of the discovery of the life cycle of Toxoplasma gondii. *International journal for parasitology* 2009, 39(8):877-882.

14. Lyons RE, McLeod R, Roberts CW: Toxoplasma gondii tachyzoite-bradyzoite interconversion. *Trends in parasitology* 2002, 18(5):198-201.

15. Jacobs L, Remington JS, Melton ML: The resistance of the encysted form of Toxoplasma gondii. *The Journal of parasitology* 1960, 46:11-21.

16. Weiss LM, Dubey JP: Toxoplasmosis: A history of clinical observations. *International journal for parasitology* 2009, 39(8):895-901.

17. Mastrobuoni S, Dell'aquila AM, Herreros J: Fatal Toxoplasma gondii Dissemination in a Heart Transplant Recipient: Description of a Case. *Case reports in transplantation* 2012, 2012:524279.

18. Kasper LH, Buzoni-Gatel D: Some opportunistic parasitic infections in AIDS: Candidiasis, pneumocystosis, cryptosporidiosis, toxoplasmosis. *Parasitology today* 1998, 14(4):150-156.

19. Jones JL, Lopez A, Wilson M, Schulkin J, Gibbs R: Congenital toxoplasmosis: a review. *Obstetrical & gynecological survey* 2001, 56(5):296-305.

20. Neto EC, Anele E, Rubim R, Brites A, Schulte J, Becker D, Tuuminen T: High prevalence of congenital toxoplasmosis in Brazil estimated in a 3-year prospective neonatal screening study. *International journal of epidemiology* 2000, 29(5):941-947.

21. Villena I, Ancelle T, Delmas C, Garcia P, Brezin AP, Thulliez P, Wallon M, King L, Goulet V, Toxosurv n *et al*: Congenital toxoplasmosis in France in 2007: first results from a national surveillance system. *Euro surveillance : bulletin europeen sur les maladies transmissibles = European communicable disease bulletin* 2010, 15(25).

22. Wei HX, Wei SS, Lindsay DS, Peng HJ: A Systematic Review and Meta-Analysis of the Efficacy of Anti-Toxoplasma gondii Medicines in Humans. *PloS one* 2015, 10(9).

23. Eyles DE, Coleman N: Synergistic effect of sulfadiazine and daraprim against experimental toxoplasmosis in the mouse. *Antibiotics & chemotherapy* 1953, 3(5):483-490.

24. Couvreur J, Thulliez P, Daffos F, Aufrant C, Bompard Y, Gesquiere A, Desmonts G: In utero treatment of toxoplasmic fetopathy with the combination pyrimethamine-sulfadiazine. *Fetal diagnosis and therapy* 1993, 8(1):45-50.

25. Stray-Pedersen B: Treatment of toxoplasmosis in the pregnant mother and newborn child. *Scandinavian journal of infectious diseases Supplementum* 1992, 84:23-31.

26. Buxton D, Innes EA: A commercial vaccine for ovine toxoplasmosis. *Parasitology* 1995, 110 Suppl:S11-16.

27. Soheilian M, Sadoughi MM, Ghajarnia M, Dehghan MH, Yazdani S, Behboudi H, Anisian A, Peyman GA: Prospective randomized trial of trimethoprim/sulfamethoxazole versus pyrimethamine and sulfadiazine in the treatment of ocular toxoplasmosis. *Ophthalmology* 2005, 112(11):1876-1882.

28. Sheiner L, Fellows JD, Ovciarikova J, Brooks CF, Agrawal S, Holmes ZC, Bietz I, Flinner N, Heiny S, Mirus O *et al*: Toxoplasma gondii Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins. *Traffic* 2015.

CHAPTER 2

COMPLEX PLASTID LITERATURE REVIEW

2.1 Introduction

Endosymbiosis is an important biological event that is believed to have contributed to the origin of the mitochondria and eukaryotic cell by the uptake of an α-proteobacterium. Most evolutionary biologist hypothesize that primary plastids originated from a similar endosymbiotic process where a phototrophic cyanobacteria was engulfed by a heterotrophic eukaryote and eventually integrated as an organelle resulting in plastids found in glaucophytes, red and green algal [1]. This endosymbiotic event allowed eukaryotes to harness the power of photosynthesis which has drastically shaped the world. Plastids have become widespread due to their ability to move laterally from secondary and tertiary endosymbiosis events. The descendants of the secondary endosymbiosis events have resulted in complex plastids with multiple membranes surrounding the plastid. Organisms containing complex plastids belong to a wide array of ecological niches, but they all share the need for a mechanism to import nuclear encoded plastid proteins across these multiple membranes.

This literary review will focus on protein import into complex plastids with an emphasis on the biology of the complex plastid of *Toxoplasma gondii* named the apicoplast. I will initially discuss the discovery of the apicoplast and the current hypothesis on the evolutionary origin of the apicoplast. Throughout the review I will refer to organisms and phyla with complex plastids that have been informative in constructing

a model for *T. gondii* apicoplast protein import (reviewed in Fig. 2.1). I will then examine signals encoded in the proteins that determine whether a protein is imported into a complex plastid. The bulk of the review is spent analyzing the current literature of the molecular machinery that governs protein import across the four membranes that surround the apicoplast and complex plastids. I will conclude by focusing on protein import across the second outermost membrane of complex plastids as the majority of my research is done on this membrane. I have added a figure of the current model for apicoplast protein import for easier comprehension (Fig. 2.2).

2.2 History and identification of the apicoplast

The apicoplast was first identified through electron microscopy studies as a multi-membrane organelle with unknown functions. In 1975 the first key step to characterizing the apicoplast organelle came from the isolation of the 35 Kb circular apicoplast DNA from *Plasmodium lophurae* [2]. However, the DNA was mistaken for the mitochondrial genome. The circular extrachromosomal DNA was misidentified even after the observation that the DNA formed a cruciform structure which is typical of chloroplast genomes rather than mitochondrial genomes [3]. A major advancement in the realization that the apicoplast genome was different from the mitochondrial genome came from sucrose density gradient experiments with *Plasmodium knowlesi* and *Plasmodium falciparum* lysates that revealed two extrachromosomal DNA bands [4, 5]. The two extrachromosomal DNA bands were a 35 Kb circular DNA and a 6 Kb DNA fragment. The 6 Kb DNA was sequenced and was found to contain typical mitochondrial genes encoding proteins cytochrome b and cytochrome c [6, 7]. This finding led to the reevaluation of the origin of the mysterious 35 Kb DNA. The 35 Kb DNA was further

analyzed and sequenced which surprisingly suggested the DNA to be of plastid origin due to an RNA polymerase sharing similarities to the RNA polymerase of chloroplasts [8-10]. The connection between the 35 Kb DNA and the multi-membrane organelle remained unrecognized until in-situ hybridization studies were performed on the DNA in *T. gondii*. The experiments proved that the plastid DNA belonged to the previously uncharacterized multi-membrane organelle now known as the apicoplast [11, 12]. Analysis of the 35 Kb DNA sequence demonstrated that the apicoplast genome is comprised of genes that control transcription and translation, but the genome did not encode proteins that are involved in photosynthesis [13]. The question that remained to be answered at the time is where did the apicoplast originate from and what is the role of a plastid organelle in non-photosynthetic parasites.

2.3 Endosymbiosis and the origin of complex plastids

Primary plastids are surrounded by two membranes reflecting the cyanobacteria origin. The apicoplast and other complex plastids are surrounded by additional membranes reflecting a more elaborate origin. A secondary endosymbiosis event involving an heterotrophic eukaryote acquiring an organism with a primary plastid gave rise to complex plastids [14]. The endosymbiont's genome is reduced through gene transfer to the host nucleus resulting in the host's control of the endosymbiont and the rise of the complex plastid. This secondary endosymbiosis event has occurred independently at least three times with the engulfment of red algae or green algae resulting in the diverse species that we have today. Euglenids and chlorarachinophytes were derived from two independent secondary endosymbiosis events involving the uptake of a green algae containing chlorophyll a+b pigmented plastids [15]. These phyla consist of diverse

organisms that occupy different ecological niches with some organisms completely dependent on photosynthesis while others have lost their plastid and are phagotropic. In contrast, the apicoplast lacks photosynthesis and likely originated from a red algal endosymbiosis event based on phylogeny studies of the plastid protein glyceraldehyde 3-phosphate hydrogenase (GAPDH) [16]. Where exactly does the apicoplast fit into the evolutionary history of complex plastids?

Phylogenetic analysis has suggested that a red alga (rhodophyte) was engulfed by a flagellated heterotropic protist which resulted in the ancestor for present day apicomplexans, cryptomonads, dinoflagellates, haptophytes and stramenopiles [17]. The idea that all these organisms originated from a single red algal endosymbiosis event followed by years of evolution and diversification is known as the "chromalveolate hypothesis" [17]. This finding was initially surprising as these organisms are quite diverse ranging from multicellular photosynthetic organisms to single cell non-photosynthetic parasites. Originally this hypothesis was supported by the finding that the plastid GAPDH gene has been replaced with a cytoplasmic GAPDH in a number of organisms in the chromalveolate super-phylum [16, 18]. A pivotal evolutionary finding was the discovery of two phototropic non-parasitic organisms, *Chromera veila* and *Vitrella brassicaformis*, which represent the closest relatives to apicomplexans [19, 20]. These organisms are coral symbionts and represent a missing link in the evolution of apicomplexans. This finding enabled the opportunity to compare genomes between organisms derived from red algal secondary enodsymbiosis to better understand the evolutionary history. Analysis of *C. veila and V. brassicaformis'* genomes and plastids confirm that apicomplexans and dinoflagellates share a red algal origin [21-23].

Although the “chromalveolate hypothesis” was the dominant hypothesis for many years, multiple investigations and studies have contradicted the idea [24-28]. Phylogenetic analysis of Calvin cycle markers and plastid import machinery proteins provide evidence that horizontal gene transfer events and higher order endosymbiosis events gave rise to cryptomonads, haptophytes, stramenopiles, apicomplexans and dinoflagellates instead of a single red algal endosymbiosis event followed by vertical inheritance of the complex plastid. An alternative model taking these new findings into account is the “rhodoplex hypothesis.” This hypothesis suggests that a single secondary endosymbiosis event of a red algal followed by tertiary and possibly quaternary endosymbiosis events resulted in the spreading of the complex plastid and the evolution of the current phyla [29]. This hypothesis is strengthened by the redundancy of metabolic processes including fatty acid, isoprenoid, and tetrapyrolle synthesis identified in dinoflagellate and apicomplexan common ancestors while absent in dinoflagellates and apicomplexans. This suggests that it is highly unlikely that these redundant pathways were maintained over a long evolutionary time [28]. However, these tertiary endosymbiosis events have only been conclusively demonstrated in some dinoflagellate lineages [30, 31]. The “serial endosymbiosis hypothesis” has expanded upon the “rhodoplex hypothesis” by proposing that all chromalveolate plastids originated from subsequent endosymbiosis events starting with the secondary endosymbiosis of a red algal resulting in cryptomonads, tertiary endosymbiosis leading to stramenopiles and quaternary endosymbiosis giving rise to haptophytes [32]. The hypothesis claims that these three groups originated from mutually exclusive endosymbiosis events arguing

against a monophyletic phylum. While the exact details of the origin of complex plastids remain to be reconciled, most evidence supports that the apicoplast is of red algal origin.

2.4 Plastid proteome

Primary plastids are not only the site of photosynthesis, but the source of numerous important metabolites such as fatty acids, starch, heme, aromatic amino acids and isoprenoids [33, 34]. The proteome of primary plastids is composed of many proteins that are needed to maintain the complex pathways that generate all the important metabolites. The plastid proteome typically consist of over 1,000 proteins, but the plastid genome is relatively small containing ~120 genes and at the most ~250 genes (*Porphyra* and *Pyropia*) [35-37]. The plastid genome is reduced due to horizontal gene transfer to the host nucleus. A secondary endosymbiosis event provides additional complexity as the host, algal and original plastid genomes were all originally contained in one cell. Horizontal gene transfer from the algal genome to the host genome is observed in complex plastids as most complex plastids have completely lost their algal genome while retaining the original reduced plastid genome. However, some phyla have maintained a reduced algal genome. Cryptomonads and chlorarachinophytes have retained a reduced algal genome which is referred to as the nucleomorph [38, 39]. The nucleomorphs of cryptomonads and chlorarachinophytes range in size from 373 KB to 703 KB and have characteristic nucleus features including nuclear envelopes and pores [40-42]. The nucleomorph is present between the second outermost and third outermost membrane known as the periplastid compartment (PPC) [42].

The beneficial metabolic processes of primary plastids have carried over to complex plastids after the secondary endosymbiosis event. The importance of these

processes is highlighted by the rare occurrences of plastid loss after a stable endosymbiosis event. There are only two traceable plastid loss events in apicomplexans and dinoflagellates which have occurred in *Cryptosporidium* and *Hematodinium* respectively [43, 44]. The apicoplast is the location of these advantageous metabolic processes including isoprenoid, iron-sulphur cluster, fatty acid and heme synthesis which have been shown to be essential for different apicomplexans during various life cycle stages [45]. The apicoplast genome of *T. gondii* is 35 KB which encodes 65 proteins [46]. These proteins are mostly transcription and translation related proteins. However, it is predicted that ~500 proteins are nuclear encoded and imported into the apicoplast [47]. The bulk of the plastid proteome needed to support a functional plastid is now encoded in the nucleus due to the horizontal gene transfer event [48].

This gene transfer event demands that a mechanism be implemented to import these proteins back into the organelle. An import mechanism is present in organelles that originated from primary endosymbiosis like the mitochondria and chloroplast. The translocase of the outer and inner membranes (TOM/TIM) and translocase of the outer and inner chloroplast membrane (TOC/TIC) machineries aid in protein translocation across the membranes of the mitochondria and chloroplast respectively [49, 50]. Both machineries are derived from the endosymbiont's outer membrane proteins that facilitated protein transport. Complex plastids including the apicoplast also utilize translocons and other proteins to aid plastid protein import into the organelle. However, complex plastids are surrounded by three or four membranes insinuating that the import mechanism will be more intricate than the import mechanism of primary plastids.

2.5 Bipartite N-terminal leaders act as signal for plastid import

Primary plastids utilize a transit peptide at the N-terminus of the protein that is necessary and sufficient for import into the plastid [51, 52]. Transit peptides show little primary sequence conservation and they tend to vary in length. Additionally, transit peptides typically possess an overall positive charge and are enriched with hydroxylated amino acids serine and threonine [53]. Complex plastids including the apicoplast employ a transit peptide which is also positively charged and similar to transit peptides of primary plastids [54]. However, the net charge of complex plastid transit peptides is considerably higher than the net charge of transit peptides from organisms with primary plastids [55, 56]. Experiments in apicomplexans have shown that the positive charge is necessary for apicoplast import, but the position of the positive charged amino acids in the transit peptide can be variable [54]. The transit peptide of complex plastids is sufficient to mediate import into plant chloroplasts demonstrating the similarity of the transit peptides between complex and primary plastids [57, 58]. One difference is that primary plastid transit peptides form alpha helices which are recognized by the import machinery [59, 60]. The transit peptide of apicoplast proteins differs in that it seems to lack a secondary structure. Mutations that increase the helical content of the transit peptide of a *P. falciparum* apicoplast protein block import into the organelle suggesting that recognition of the transit peptide requires an unstructured state [61].

The major difference in import between primary and complex plastids is that complex plastids have adopted an additional signal for import across the additional membranes. Nuclear encoded proteins targeted to complex plastids have a bipartite signal consisting of a signal peptide followed by a transit peptide at the N-terminus. The

bipartite signal peptide can be replaced with a canonical signal peptide without interfering with complex plastid localization [62]. It is believed that the signal peptide directs the protein to be cotranslationally inserted into the ER lumen through the Sec61 translocon similar to proteins traveling through the canonical secretory pathway [63]. It is predicted that the signal peptide is cleaved by a signal peptidase residing in the ER lumen which reveals the transit peptide. Subsequently, it is presumed that the transit peptide is recognized by an unknown facilitator that aids import into the complex plastid. The transit peptide is thought to be cleaved by a protease and the protein folds into its mature form when the protein reaches the innermost compartment of the complex plastid [64].

The complex plastid import mechanism is considerably different from primary plastids whose translocon of the outer chloroplast membrane (TOC) machinery directly recognizes and interacts with plastid proteins in the cytoplasm [65]. However, it has been reported that some primary plastid proteins are N-glycosylated in the ER and are transported via the secretory pathway to the chloroplast instead of the traditional TOC/TIC machinery [66]. It has been shown that the N-terminal leaders of certain periplastid and stromal proteins are glycosylated in the ER of the stramenopile *Phaeodactylum tricornutum* [67]. These glycosylated proteins are transported into the proper compartment of the complex plastid by crossing the remaining membranes. It is unclear whether complex plastids exploit specialized import machinery for glycosylated proteins as the diameters of the Toc75 and Tic20 pores are predicted to be too narrow for glycosylated proteins. This glycosylation event does not seem to interfere with proper localization or inhibit transport into the complex plastid. However, glycosylation does not

appear to be necessary for import into the complex plastid as the N-linked glycosylation inhibitor tunicamycin does not affect localization [67].

Further complexity is added to the system by the requirement to differentiate between proteins localized in different compartments of the plastid. Investigations into the process have unveiled that complex plastid proteins destined to the innermost compartment of the plastid typically have an aromatic amino acid, usually phenylalanine, at the +1 site of the transit peptide [68, 69]. This has been experimentally validated in *Guillardia theta* and *P. tricornutum* where a single amino acid substitution at the +1 site to a non-aromatic amino acid results in the mislocalization of stromal proteins to the periphery of the complex plastid [65, 70]. This finding was supported by a detailed comparison of known second outermost compartment or periplastid compartment (PPC) proteins and stromal plastid proteins which revealed that PPC proteins lack the aromatic amino acid at the +1 site while stromal proteins contain a conserved phenylalanine residue [65, 70]. This is consistent with data showing that the transit peptides for primary plastids of red algae and glaucocystophytes require aromatic amino acids for import [71-73]. Further work in glaucocystophytes has shown that the TOC machinery is responsible for the recognition of the phenylalanine residue [74]. The amino acid composition of transit peptides was analyzed in a multitude of apicomplexans and showed an enrichment of aromatic amino acids tyrosine and phenylalanine at the +1 site [75]. Recently, the sequences of forty-seven transit peptides from experimentally confirmed *T. gondii* apicoplast proteins were analyzed. This study revealed similar results with an enrichment of aromatic amino acids at the +1 site for stromal apicoplast proteins and an absence of aromatic amino acids at the +1 transit peptide site for peripheral proteins [76]. In addition,

experiments in *T. gondii* showed mutations to the aromatic amino acid at the +1 site to the stromal apicoplast protein acyl carrier protein resulted in mislocalization of the protein to the periphery of the apicoplast. However, this is not a consistent characteristic of apicomplexan stromal proteins as many of the stromal proteins are missing the +1 aromatic amino acid. It was also demonstrated that the deletion of the transit peptide of the *T. gondii* luminal apicoplast protein ferredoxin NADP+ reductase resulted in no observable change in targeting [77]. Altogether this suggests a possible alternative mechanism or signal for differentiating stromal from peripheral apicoplast proteins.

2.6 Crossing the outermost plastid membrane

Interestingly, there are complex plastids that are surrounded by three membranes while others are surrounded by four membranes. Although dinoflagellates and euglenids are derived from independent endosymbiosis events, they both contain plastids with three membranes. Nuclear encoded plastid proteins are cotranslationally imported into the ER where proteins are then transported as vesicles to the Golgi and transported across the outermost membrane of the plastid in both phyla [78-80]. Furthermore, cell free protein trafficking systems in euglenids have shown that fusion of vesicles to the plastid is independent of SNARE proteins [81]. Cryptomonads, haptophytes and stramenopiles contain plastids that reside inside the endomembrane system. These plastids are surrounded by four membranes counting the ER membrane as the outermost membrane. Plastid protein signal peptides fused to an epitope tag in these phyla resulted in ER localization [82, 83]. This validates the idea that import requires proteins to be co-translationally imported across the ER membrane in these phyla [84]. The apicoplast and the ER are in close proximity, but electron tomography shows that the two organelles are

distinct from each other [85]. The apicoplast is still surrounded by four membranes suggesting that an additional transport step is needed to cross the outermost membrane of the apicoplast. Our current model for protein import into the apicoplast is deeply informed by the previously discussed mechanisms for import into complex plastids. Nuclear encoded apicoplast proteins are predicted to be co-translationally imported into the ER and transported in vesicles to the outermost membrane where the vesicles most likely fuse with the outermost membrane. There are numerous reports of apicoplast proteins detected in vesicles by electron and light microscopy [86-89]. Chlorarchinophytes are similar to apicomplexans in that their plastid is surrounded by four membranes and separated from the ER. It is predicted that chlorarachniophytes also use vesicles for import across the outermost membrane [90].

The question of whether apicoplast proteins go through the Golgi to reach the apicoplast remains unclear. Early evidence in *T. gondii* and *P. falciparum* suggested that apicoplast proteins are transported independently of the Golgi apparatus [62, 91]. Reexamination of these studies in *P. falciparum* demonstrated that the addition of ER retrieval signals to nuclear encoded apicoplast proteins resulted in reduced trafficking to the apicoplast and transit peptide processing [92]. This contradicts previous findings and might suggest that the Golgi body plays a role in trafficking to the apicoplast. This finding prompted the hypothesis that a transit peptide receptor resides in the Golgi in order to segregate apicoplast proteins from non-apicoplast proteins for further shipment to the organelle. Further work needs to be done to elucidate whether this hypothesis is valid. While this study provides compelling evidence for a Golgi dependent trafficking mechanism, transit peptide processing is still visible after treatment with the Golgi

inhibitor Brefeldin A. This suggests import is still occurring after blocking Golgi dependent secretory processes and a Golgi dependent protein import model cannot fully explain the results.

There is also emerging evidence that not all nuclear encoded apicoplast proteins are imported in the same manner in *T. gondii*. The identification of nuclear encoded apicoplast proteins in *T. gondii* lacking a signal and transit peptide increases the probability of separate trafficking mechanisms to the apicoplast [86, 87, 93-96]. Peripheral plastid protein 1 (PPP1) is an essential apicoplast protein that does not contain a bipartite leader, but rather possesses a recessed hydrophobic domain [93]. Atrx2 is a membrane protein localized to the periphery of the apicoplast that contains a signal anchor sequence at the N-terminus which is cleaved instead of containing the canonical leader sequence [86]. FtsH1 is a protease with a single transmembrane domain which has been experimentally validated to be necessary for import into the apicoplast. FtsH1 is the only known apicoplast protein to undergo C-terminal processing upon the protein's arrival in the organelle [94, 95]. APT1 is an outermost apicoplast transporter with a tyrosine and glycine motif at the 16^{th} and 17^{th} amino acid residue upstream of the first transmembrane domain. The tyrosine/glycine motif resides in the cytoplasm and has been demonstrated to be necessary, but not sufficient for proper apicoplast import. APT1 is an anomaly in that it doesn't get processed [96]. A homolog for this phosphate transporter has been identified in *P. falciparum* (PfoTPT) and recently shown to have a cytosolic tyrosine residue essential for import along with a crucial transmembrane domain [97]. Another characteristic shared by these non-canonical leader sequence apicoplast proteins is that they have been identified in large vesicles which are abundant during apicoplast

division and accumulate in plastid deficient strains [98]. These large vesicles have been observed near the apicoplast or in some cases merging with the organelle through immunoelectron microscopy [86]. The argument for two separate trafficking mechanisms was strengthened by recent work illustrating that luminal apicoplast proteins are absent from these large vesicles and that the large vesicles seem to be independent of the Golgi [98]. However, the biological role of these large vesicles trafficking non-canonical leader sequence proteins to the apicoplast independent of canonical nuclear encoded apicoplast proteins is unknown. Further work is needed to tease apart the mechanism of import across the outermost membrane of the apicoplast.

Both autophagy related protein 8 (ATG8) and phosphatidylinositol 3-monophosphate (PI(3)P) have been implicated in potential roles in protein import across the outermost membrane of the apicoplast. Autophagy is a conserved process amongst eukaryotes that culminates in the formation of autophagasomes around proteins and organelles that will be digested by the lysosome to recycle cellular components [99]. This biological process is controlled by a plethora of autophagy related (ATG) proteins. Several core ATG proteins needed for ATG8 conjugation to membranes are conserved in apicomplexans including in *T. gondii*, but overall the autophagy machinery is reduced in its complexity in the phylum [100]. Double membrane structures covered with ATG8 have been identified in *T. gondii* under starvation conditions which is reminiscent of autophagasomes and the autophagy process [101, 102]. However, in *T. gondii* it was also shown that under standard growth conditions ATG8 and the autophagy machinery have functions independent from the conserved autophagy process. ATG8 was demonstrated to be localized to the apicoplast during standard intracellular growth conditions [103,

104]. ATG8 was similarly found to be associated with the apicoplast in *P. falciparum* [105, 106]. *T. gondii* mutants to the proteins ATG3 and ATG4 which act in the conjugation of ATG8 to the lipid phosphatidylethanolamine (PE) and the removal of ATG8 respectively resulted in apicoplast biogenesis defects and reduced cell survival [101, 103]. A recent study clarifies the connection between the apicoplast and the autophagy machinery by generating a conditional mutant for ATG8. Super-resolution imaging and proteinase K protection assays revealed that ATG8 resides on the cytoplasmic side of the outermembrane of the apicoplast. ATG8 mutants demonstrated that ATG8 is essential for tachyzoite growth due to the loss of the protein leading to segregation defects of the apicoplast during division. It was demonstrated that the segregation defect is most likely due to the role of ATG8 physically linking the apicoplast to the centrosome [107]. These results suggest that ATG8 and the autophagy machinery are crucial for apicoplast inheritance during division, but it is no longer believed that ATG8 mediates a role in apicoplast import across the outermost membrane.

The other component that is associated with the outermembrane of the apicoplast and apicoplast protein transport is phosphatidylinositol 3-monophosphate (PI(3)P). PI(3)P is a lipid that is normally associated with endosomes and plays a role in trafficking these compartments to the lysosome [108]. PI(3)P was found to be localized to the apicoplast and vesicles associated with nuclear encoded apicoplast proteins. In addition, overexpression of PI(3)P binding domains and the use of a PI3 kinase inhibitor resulted in apicoplast biogenesis defects and the accumulation of these vesicles around the apicoplast [109]. Therefore, it was proposed that PI(3)P was involved in vesicular trafficking of apicoplast proteins across the outermost membrane of the apicoplast. This

hypothesis was initially supported by PI(3)P localizing to food vacuoles and the apicoplast in *P. falciparum* [110]. The role for PI(3)P may be conserved as the use of PI3 kinsase inhibitors resulted in growth defects in *P. falciparum* blood stages and the inability to generate a PI3 kinase mutant in *Plasmodium berghei* suggests essentiality [110]. However, recent work in *T. gondii* demonstrates that PI(3)PK and PIKfyve (kinases that facilitate the synthesis of PI(3)P and PI(35)P_2 respectively) mutants do not display defects in the import peripheral or luminal apicoplast proteins [111]. It is alternatively suggested that PI(3)P and PI(35)P_2 are essential for the parasite by maintaining apicoplast morphology as the mutants for the kinases and regulator of PIKfyve resulted in enlarged apicoplasts and eventual organelle loss. The rigorous studies of ATG8 and PI(3)P have led to a broader picture of apicoplast biology, but the mechanism of transport of proteins across the outermost membrane is still unclear.

2.7 Algal derived translocons facilitate import in complex plastids

Protein import into primary plastids is mediated by the TOC and TIC machinery. The TOC/TIC machinery is composed of numerous components that range from membrane proteins that act as the pore to various other chaperones that recognize proteins and mediate import [112]. Although the machinery is composed of more than a dozen proteins, the key components of the TOC machinery are the GTP binding proteins Toc34 and Toc159 which recognize the transit peptide of proteins and the beta-barrel pore membrane protein Toc75 [113]. Once primary plastid proteins pass through the Toc75 translocon, they interact with Tic22 in the intermembrane space which putatively transfers proteins from the TOC machinery to the TIC machinery [114]. The identity of the proteins that form the translocon for the innermost membrane is contentious. One

model claims that proteins Tic20 and Tic110 form the channel together while another prominent hypothesis is that Tic110 is not directly involved in the formation of the pore which facilitates protein passage [115]. However, the most recent evidence suggests that Tic20 acts independent of Tic110 as the pore with the assistance of Tic56, Tic100 and Tic214 [116]. It is also established that the AAA ATPase ClpC which is located in the stroma interacts with Tic40 and Tic110 and facilitates protein transport across the innermost membrane [117]. How are proteins transported across the innermost membranes of complex plastids?

It was initially proposed that the TOC/TIC machinery would be conserved in complex plastids due to the algal plastid origin. Genome wide searches for homologs of TOC/TIC proteins in organisms with complex plastids derived from red and green algae revealed many homologs [39, 76, 88, 90, 118-120]. It was reported that TIC components are encoded in the nucelomorph genome of the cryptomonad *G. theta* and chlorarachniophyte *Bigelowiella natans* [39]. Also homologs of ClpC which is associated with the TIC machinery are encoded in the plastid genome of *T. gondii* and *P. falciparum* [13, 121]. This suggests that the TOC/TIC machinery is conserved in complex plastids and is critical for import. However, there are considerable differences in the conservation of the homologs between organisms with complex plastids. For instance, a detailed search for TOC/TIC components in the genome of *B. natans* identified eleven putative TOC/TIC homologs [90]. The TOC/TIC machinery is reduced in other complex plastids including *T. gondii* which revealed only three conserved TOC/TIC proteins through bioinformatic searches [76, 90, 122]. Additionally, the major transit peptide recognition factor Toc34 was found in *P. falciparum*, but was absent in the complex plastid of *B.*

natans using similar screens [90, 123]. While many proteins of the TOC/TIC machinery have been identified through bioinformatic screens, some of these proteins have also been experimentally characterized. The Toc75 translocon in *P. tricornutum* was localized to the complex plastid and characterized as a membrane bound protein with the N terminus and C terminus exposed to the periplastid compartment (PPC) of the plastid which is consistent with its topology in primary plastids [118]. Tic22 has been identified in the apicoplast of *P. falciparum* and was shown to be a soluble protein associated with the membrane of the apicoplast [119]. Although identifying TOC/TIC proteins experimentally and bioinformatically in complex plastids indicates that they have a role in protein import, functional proof was lacking. However, *T. gondii* has become a useful model to study complex plastids due to the ability to construct conditional mutants in this organism and the development of an apicoplast import assay. These advancements allowed us to test the importance and role of particular complex plastid proteins.

The three homologs of the TOC/TIC machinery identified in *T. gondii* are Toc75, Tic20 and Tic22. Split-GFP assays with the Tic20 protein in *T. gondii* have demonstrated that the protein is membrane bound with four transmembrane domains with the N and C terminus residing in the stromal compartment which is consistent with Tic20 of primary plastids [88]. Tic22 was shown to be a soluble protein in the apicoplast [120]. Tic20 and Tic22 were demonstrated to be essential for apicoplast biogenesis and protein import into the apicoplast in *T. gondii* [88, 120]. The Toc75 protein initially proved to be difficult to identify in *T. gondii* due to modest sequence conservation, but a recent study has identified the Toc75 translocon (see appendix 1). Toc75 was demonstrated to be essential for apicoplast biogenesis and protein import. The Toc75 translocon mutant resulted in the

loss of import for stromal proteins. Importantly, import of apicoplast proteins that reside in the periphery of the apicoplast is not affected. This supports the notion that Toc75 is mediating import of stromal apicoplast proteins and peripheral apicoplast proteins do not rely on the Toc75 translocon for proper import. I have contributed to this study by performing several of the apicoplast import assays with the Toc75 mutant (see appendix 1). Overall, experimental evidence has provided a model which claims nuclear encoded apicoplast proteins are transported through the Toc75 translocon then across the Tic20 translocon with the help of Tic22 to reach the stroma of the apicoplast.

2.8 ERAD derived machinery has been retooled for import across the periplastid membrane

The second outermost membrane or the periplastid membrane (PPM) of complex plastids is thought to be derived from the algal endosymbiont plasma membrane [124]. While the finding that the TOC/TIC translocons played a role in import across the innermost membranes of complex plastids was congruent with primary plastids, there was no immediate homologous model to explain on how proteins would be imported across the PPM. Protein import across the PPM remained a mystery until sequencing of the nucleomorph of the cryptomonad *G. theta* revealed that the algal genome encoded homologs of the endoplasmic reticulum associated degradation (ERAD) pathway [125]. This discovery was initially puzzling as the red algal endosymbiont is reduced and no longer has an ER compartment. The Maier group proposed that cryptomonads have retooled the ERAD proteins for import across the PPM. ERAD components were shown to be encoded in multiple genes in the genome in a number of organisms containing complex plastids supporting the initial hypothesis [119, 126-128]. It was observed that

these duplicated ERAD proteins are localized to a shared compartment in the periphery of the plastid using split-GFP assays in *P. tricornutum* [129]. Light and electron microscopic imaging of these ERAD components in apicomplexans, haptophytes and stramenopiles confirmed that they are localized to the periplastid compartment (PPC) [126, 127]. Phylogenetic analysis of the plastid ERAD and ER ERAD proteins in organisms containing complex plastids reveals two set of proteins of divergent evolutionary origin. Plastid ERAD proteins appear most closely related to and derived from the red algal while the ER ERAD proteins may have originated through vertical inheritance [127]. This discovery supports the secondary endosymbiosis model as these proteins are localized to the PPC which is predicted to be the remnant of the algal cytoplasm.

The hypothesis that ERAD proteins have been retooled for import across the PPM of complex plastids while initially surprising is consistent with the activity of the ERAD proteins in their classic function. The ERAD machinery provides quality control to the secretory pathway by recognizing misfolded proteins in the ER lumen, pulling proteins out of the ER and marking them with ubiquitin for degradation by the proteasome [130]. The plastid ERAD components consistently observed among complex plastids are Der1, CDC48, Ufd1, the ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2), the ubiquitin ligase (E3), a deubiquitinase and an ubiquitin-like protein. Der1 is an integral ER membrane protein that is most likely the site of protein retro-translocation from the ER lumen to the cytoplasm in the classic ERAD pathway. It is debated whether Der1 alone acts as the translocon or interacts with the RING finger ubiquitin ligase (Hrd1p) which is also an integral ER membrane protein to form a heterotetramer protein

that acts as the pore. Experiments using purified *S. cerviesea* proteins to reconstitute the ERAD pathway have shown that Hrd1p is part of a translocon and also critical in discriminating between folded and unfolded proteins in the ER lumen [131]. This close interaction between the Der1 translocon and the ubiquitin machinery allows misfolded proteins to be ubiquitinated as the proteins are translocated into the cytoplasm. The formation of polyubiquitin chains is crucial for recognition by CDC48 which is another pivotal protein in the ERAD pathway [132].

CDC48 is an AAA ATPase hexameric protein that is involved in a number of cellular activities including the ERAD pathway, membrane fusion and chromatin regulation [133]. CDC48 consists of an N-terminal domain and two ATPase domains. CDC48's N-terminal domain interacts with a number of chaperones and proteins depending on the specific cellular activity it is engaged in including ubiquitin [134, 135]. CDC48 extracts polyubiquitinated proteins from complexes in cellular functions independent of the ERAD pathway [136]. It has been shown in vitro and in yeast cells that during the ERAD pathway CDC48 along with the chaperones Ufd1 (ubiquitin fusion degradation protein 1) and Npl14 (Nuclear protein localization 4 homolog) can bind to ubiquitinated proteins [137, 138]. CDC48 and the aforementioned chaperones are recruited to the ER membrane and the ubiquitin ligase where they play a critical part in the translocation of misfolded proteins [139-141]. CDC48's ATPase activity is required to pull proteins out of the ER lumen into the cytoplasm for protein degradation. The loss of CDC48's ATPase activity due to specific point mutations prevents the degradation of ERAD substrates by the proteasome [142-144]. The UT3 motif at the N-terminus of Ufd1

interacts with polyubiquitin in the ERAD pathway also. This interaction is needed for proper protein translocation across the ER membrane [145, 146].

Although complex plastids share many of the same plastid ERAD components, there are subtle differences between organisms. For instance, genome mining identified two plastid Der1 genes in cryptomonads, stramenopiles, haptophytes and *P. falciparum* while only a single gene found in *T. gondii*. Immunoprecipitation assays using *P. tricornutum* have experimentally validated that the two plastid Der1 proteins (Der1-1 and Der1-2) form homo-oligomers and hetero-oligomers [129]. Also it has been reported that the plastid ERAD machinery is more complex in stramenopiles compared to apicomplexans as four new related ERAD proteins were identified in the PPC of *P. tricornutum* plastids which are not conserved in the plastids of apicomplexans [147]. These four proteins are sNpl4, sUBX, sUbq, and sPng1. Both sNpl4 and sUBX are predicted to bind and interact with CDC48 [148]. The complex plastid ubiquilin-like protein (sUbq) is related to Dsk2p which in the classic ERAD pathway is known to bind to polyubiquitin chains through its C-terminal UBA domain and interact with the proteasome [149]. However, the symbiont version is lacking the UBA domain and the current function of the protein is still unknown. The protein sPng1 shares sequence similarity to the deglycosylation enzyme PNGase which further illustrates that glycosylation occurs on plastid proteins in stramenopiles [150]. Also there have been a total of seven core 20S proteasomal proteins experimentally localized to the PPC of *P. tricornutum* [147, 151]. Cryptomonads and stramenopiles are the only phyla containing complex plastids with proteasomal subunits with a majority of the proteasomal subunits being encoded in the nucleomorph of the cryptomonad *G. theta* [38]. The proteasome is a

complex of proteins made up of the core or 20S subunit and a cap structure or 19S subunit of the proteasome. The 20S component consists of four rings with each ring comprised of seven alpha or seven beta subunits which functions as the proteolytic domain of the proteasome. The 19S cap structure is constructed from 19 subunits and is critical for ATP hydrolysis and removal of ubiquitin chains [152]. In silico comparison of putative plastid proteasomal units of *P. tricornutum*, *Thalassiosira pseudonana*, and *Fragilariopsis cylindrus* revealed that all these organisms contain varying proteins of the 20S proteasome (including proteolytic active subunits), but all lack a complete set of 20S proteins. All these organisms lack components of the 19S subunit of the proteasome further illustrating that the proteasome is most likely inactive in stramenopiles [147]. Overall, it is suggested that the ERAD derived proteins in complex plastids have been repurposed for pulling proteins out of the PPM instead of the ER membrane and have lost their role in protein degradation. Although all these components were consistently identified in complex plastids, there was no experimental evidence supporting the hypothesis at the time.

The first experimental evidence to demonstrate that the plastid ERAD proteins were mediating import occurred in *T. gondii.* A conditional Der1 mutant was demonstrated to be essential for parasite viability and the mutant resulted in the loss of protein import into the apicoplast [127]. In Chapter 3 of this dissertation I will show that $CDC48_{AP}$ is crucial not only for import into the apicoplast, but for successful import across the PPM of the apicoplast. It should be noted that no ERAD proteins have been identified in complex plastids that were derived from green algae [90]. However, not all proteins found in the PPC are homologs of the ERAD machinery. The protein identified

as PPP1 in *T. gondii* and *P. tricornutum* was validated as a PPC protein [93, 151]. A PPP1 conditional mutant resulted in very similar import defects across the PPM of the apicoplast in *T. gondii* [93]. It remains unclear how PPP1 fits into the protein import model, but it is suggested that PPP1 might act as a chaperone that facilities import across the PPM. Overall, these mutants suggest a direct role of the plastid ERAD machinery in import across the PPM thus validating the ERAD hypothesis. The current model proposes that all plastid proteins are transported through the Der1 translocon where proteins are conjugated by an ubiquitin-like modifier. It is hypothesized that the binding of an ubiquitin-like protein is a requirement for recognition by $CDC48_{AP}$ and its cofactor $Ufd1_{AP}$ which pull proteins through the PPM in a similar fashion as CDC48 in the classic ERAD pathway.

2.9 The role of ubiquitin and the ubiquitinating machinery in complex plastids

One of the more confounding elements of the proposed model for import across the PPM is the role of ubiquitin. Misfolded proteins in the ERAD pathway are ubiquitinated when they are transported into the cytoplasm which acts as a signal for degradation by the proteasome [130]. Ubiquitin is attached to substrates through an enzymatic cascade of proteins which consists of the ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2) and the ubiquitin ligase (E3). Ubiquitin is activated by the E1 enzyme and consequently passed to the E2 and E3 enzymes before being conjugated to the substrate [153]. The E1 and E2 components of the ubiquitinating machinery have been experimentally verified to be localized to the periphery of complex plastids in *T. gondii*, *P. falciparum*, *E. huxleyi* and *P. tricornutum* [125, 126, 154]. The E3 protein has also been identified in *P. falciparum* and *P. tricornutum* and both E3s were shown to have the

ability to auto-ubiquitinate itself through in vitro experiments [154, 155]. It was also later shown that the ubiquitin conjugating enzyme ($E2_{AP}$) is necessary for import into the apicoplast using a conditional mutant and biochemical apicoplast import assay in *T. gondii* [154]. The finding that the ubiquitinating machinery is involved in import into the plastid rather than protein degradation leaves the mechanistic role of the ubiquitin-like modifier unclear. However, a recent model for ubiquitination in the classic ERAD system proposes ubiquitin is required for dislocation of proteins out of the ER membrane in addition to acting as a marker for protein degradation [156-158]. Another critical aspect of ubiquitin biology is the ability to regulate ubiquitination through a deubiquitinase enzyme. Plastid specific deubiquitinases have been identified in *P. tricornutum* and in vitro deubiquitination assays has confirmed that the *P. tricornutum* plastid deubiquitinase is enzymatically active [155]. Experiments in mammalian cells and yeast have shown both addition and removal of ubiquitin is required for proper classic ERAD function [131, 159, 160]. YOD1 is a CDC48 associated deubiquitinase that cleaves polyubiquitinated proteins in the ERAD pathway and YOD1 has been found to be required for proper retro-translocation [157, 160]. Interestingly, there are even reports that YOD1 may be needed for translocation of misfolded proteins out of the ER lumen that were not ubiquitinated in the first place [161]. Recent reports demonstrated that YOD1 and CDC48 are required for deglycosylation and degradation events of ERAD substrates downstream of substrate translocation [162]. These studies suggest that in complex plastids deubiquitinases could be critical for protein import. However, no deubiquitinase mutant suitable to test this hypothesis has been generated in complex plastids yet.

In Chapter 3 of this dissertation I make the finding of a unique plastid ubiquitin-like protein (PUBL) in the PPC of *T. gondii* rather than ubiquitin. In silico searches revealed that homologs of PUBL are found throughout complex plastids containing four membranes besides *G. theta* which expresses a plastid specific ubiquitin [147]. PUBL has also been experimentally localized to the complex plastid in *P. tricornutum* and *P. falciparum* [128, 147]. Chapter 3 demonstrates that PUBL in *T. gondii* is essential for parasite survival and import into the apicoplast. Complementation assays with the PUBL mutant showed that PUBL can be replaced with an ubiquitin sent to the apicoplast suggesting that PUBL acts in a similar fashion as ubiquitin.

There is still a considerable lack of knowledge on how these organisms utilize the ubiquitination system for import across the PPM. The current model suggests that all imported plastid proteins are attached to PUBL which is a signal for the symbiont version of CDC48 to pull proteins through the PPM. This hypothesis would predict that PUBL visibly accumulates on plastid proteins, but experiments consistently fail to detect PUBL on any plastid protein. However, recent data in *P. tricornutum* and *T. gondii* has provided indirect evidence that PUBL is binding onto imported substrates. Predicted ubiquitination sites at the N-terminus of imported plastid proteins in *P. tricornutum* were shown to be essential for protein import. More excitingly mutations to these potential ubiquitination sites resulted in proteins being “frozen” in the periplastid membrane (PPM) of the plastid. They also showcase that the majority of imported proteins are enriched in potential ubiquitination sites compared to other proteins in the secretory pathway [163]. In Chapter 3 the PUBL mutant in *T. gondii* revealed that the C-terminal diglycine motif is essential for parasite survival. The diglycine motif is indispensable for ubiquitin-like proteins as

the motif is used to conjugate onto substrates [164]. Overall, this provides indirect evidence that PUBL is most likely binding onto substrates. However, no experimental evidence has formally demonstrated a plastid ubiquitin-like protein conjugated to any substrate.

One possible explanation is that this process is very transient due to an active plastid deubiquitinase resulting in the inability to visualize the linked plastid ubiquitin-like protein on a substrate. A thorough analysis of the ubiquitin proteome in *T. gondii* was recently performed which revealed that over 500 proteins with diverse localizations were ubiquitinated. They were able to map the ubiquitin proteome by performing pulldown experiments using diglycine antibodies which would potentially detect any plastid ubiquitin-like proteins attached to substrates. However, no apicoplast proteins were pulled down in these experiments [165]. Another possible reason for the absence of PUBL bound to numerous substrates is that PUBL is only binding one protein. Recently, it was illustrated through in vivo and in vitro experiments that auto-ubiquitination of lysine residues of the ERAD ubiquitin ligase, Hrd1p, is essential for misfolded proteins to be translocated across the ER membrane. This suggests that the ubiquitin ligase acts as a gate for protein translocation and that ubiquitination of the ubiquitin ligase controls when proteins are exported out of the ER lumen [166]. It was also discovered that the ubiquitin-proteasome system regulates the Toc75 pore during primary plastid development by ubiquitinating Toc75 to mark it for degradation which ultimately determines the fate of the plastid [167, 168]. If these processes are conserved in complex plastids, PUBL covalently binds to the ubiquitin ligase or Toc75 translocon as a mechanism to control

when proteins are transported. Overall, these findings present alternative models for the role of the plastid ubiquitin-like protein and its role in protein import.

2.10 Conclusion

The understanding of complex plastid biology has seen drastic advances in the past years because of breakthroughs in genome sequencing and genetic tools in organisms with complex plastids. Although the exact details on the origin and evolution of these complex plastids remain to be solved, these organisms share many of the same mechanisms for import into the organelles. It has been established that complex plastids are conserved in that nuclear-encoded plastid proteins typically require a signal and transit peptide at the N-terminus to direct the protein across the multi-membrane plastid. Complex plastids have also sustained the TOC/TIC machinery derived from the algal symbiont for import across the innermost membranes of the complex plastid. Red algal derived complex plastids have retooled the ERAD pathway for import across the PPM. Although there are still large amounts of gaps in the knowledge of protein import into these multi-membrane structures, current studies have provided a detailed molecular model for import.

References

1. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF: Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 2005, 15(14):1325-1330.
2. Kilejian A: Circular mitochondrial DNA from the avian malarial parasite Plasmodium lophurae. *Biochim Biophys Acta* 1975, 390(3):276-284.
3. Borst P, Overdulve JP, Weijers PJ, Fase-Fowler F, Van den Berg M: DNA circles with cruciforms from Isospora (Toxoplasma) gondii. *Biochim Biophys Acta* 1984, 781(1-2):100-111.
4. Williamson DH, Wilson RJ, Bates PA, McCready S, Perler F, Qiang BU: Nuclear and mitochondrial DNA of the primate malarial parasite Plasmodium knowlesi. *Mol Biochem Parasitol* 1985, 14(2):199-209.
5. Gardner MJ, Bates PA, Ling IT, Moore DJ, McCready S, Gunasekera MB, Wilson RJ, Williamson DH: Mitochondrial DNA of the human malarial parasite Plasmodium falciparum. *Mol Biochem Parasitol* 1988, 31(1):11-17.
6. Vaidya AB, Akella R, Suplick K: Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol* 1989, 35(2):97-107.
7. Aldritt SM, Joseph JT, Wirth DF: Sequence identification of cytochrome b in Plasmodium gallinaceum. *Mol Cell Biol* 1989, 9(9):3614-3620.

8. Gardner MJ, Williamson DH, Wilson RJ: A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Mol Biochem Parasitol* 1991, 44(1):115-123.

9. Howe CJ: Plastid origin of an extrachromosomal DNA molecule from Plasmodium, the causative agent of malaria. *J Theor Biol* 1992, 158(2):199-205.

10. Williamson DH, Gardner MJ, Preiser P, Moore DJ, Rangachari K, Wilson RJ: The evolutionary origin of the 35 kb circular DNA of Plasmodium falciparum: new evidence supports a possible rhodophyte ancestry. *Mol Gen Genet* 1994, 243(2):249-252.

11. McFadden GI, Reith ME, Munholland J, Lang-Unnasch N: Plastid in human parasites. *Nature* 1996, 381(6582):482.

12. Kohler S, Delwiche CF, Denny PW, Tilney LG, Webster P, Wilson RJM, Palmer JD, Roos DS: A plastid of probable green algal origin in apicomplexan parasites. *Science* 1997, 275(5305):1485-1489.

13. Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW *et al*: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. *J Mol Biol* 1996, 261(2):155-172.

14. Gould SB, Waller RF, McFadden GI: Plastid evolution. *Annu Rev Plant Biol* 2008, 59:491-517.

15. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ: The complete chloroplast genome of the chlorarachniophyte Bigelowiella natans: evidence for

independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol* 2007, 24(1):54-62.

16. Fast NM, Kissinger JC, Roos DS, Keeling PJ: Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 2001, 18(3):418-426.

17. Cavalier-Smith T: Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 1999, 46(4):347-366.

18. Harper JT, Keeling PJ: Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 2003, 20(10):1730-1735.

19. Moore RB, Obornik M, Janouskovec J, Chrudimsky T, Vancova M, Green DH, Wright SW, Davies NW, Bolch CJ, Heimann K *et al*: A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 2008, 451(7181):959-963.

20. Obornik M, Modry D, Lukes M, Cernotikova-Stribrna E, Cihlar J, Tesarova M, Kotabova E, Vancova M, Prasil O, Lukes J: Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* 2012, 163(2):306-323.

21. Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ: A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A* 2010, 107(24):10949-10954.

22. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michalek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A *et al*: Chromerid genomes reveal the

evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* 2015, 4:e06974.

23. Janouskovec J, Horak A, Barott KL, Rohwer FL, Keeling PJ: Environmental distribution of coral-associated relatives of apicomplexan parasites. *ISME J* 2013, 7(2):444-447.

24. Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H: Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 2010, 27(7):1698-1709.

25. Burki F, Okamoto N, Pombert JF, Keeling PJ: The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* 2012, 279(1736):2246-2254.

26. Bodyl A, Stiller JW, Mackiewicz P: Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol* 2009, 24(3):119-121; author reply 121-112.

27. Teich R, Zauner S, Baurain D, Brinkmann H, Petersen J: Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. *Protist* 2007, 158(3):263-276.

28. Waller RF, Gornik SG, Koreny L, Pain A: Metabolic pathway redundancy within the apicomplexan-dinoflagellate radiation argues against an ancient chromalveolate plastid. *Commun Integr Biol* 2016, 9(1):e1116653.

29. Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H: Chromera velia, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol* 2014, 6(3):666-684.

30. Burki F, Imanian B, Hehenberger E, Hirakawa Y, Maruyama S, Keeling PJ: Endosymbiotic Gene Transfer in Tertiary Plastid-Containing Dinoflagellates. *Eukaryotic cell* 2014, 13(2):246-255.

31. Hehenberger E, Imanian B, Burki F, Keeling PJ: Evidence for the Retention of Two Evolutionary Distinct Plastids in Dinoflagellates with Diatom Endosymbionts. *Genome biology and evolution* 2014, 6(9):2321-2334.

32. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J: The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun* 2014, 5:5764.

33. Neuhaus HE, Emes MJ: Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Phys* 2000, 51:111-140.

34. Borza T, Popescu CE, Lee RW: Multiple metabolic roles for the nonphotosynthetic plastid of the green alga Prototheca wickerhamii. *Eukaryotic cell* 2005, 4(2):253-261.

35. Janouskovec J, Liu SL, Martone PT, Carre W, Leblanc C, Collen J, Keeling PJ: Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One* 2013, 8(3):e59001.

36. Terashima M, Specht M, Hippler M: The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. *Curr Genet* 2011, 57(3):151-168.

37. Daniell H, Lin CS, Yu M, Chang WJ: Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome biology* 2016, 17.

38. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG: The highly reduced genome of an enslaved algal nucleus. *Nature* 2001, 410(6832):1091-1096.

39. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI: Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci U S A* 2006, 103(25):9566-9571.

40. Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM: Complete nucleomorph genome sequence of the nonphotosynthetic alga Cryptomonas paramecium reveals a core nucleomorph gene set. *Genome Biol Evol* 2011, 3:44-54.

41. Tanifuji G, Onodera NT, Brown MW, Curtis BA, Roger AJ, Ka-Shu Wong G, Melkonian M, Archibald JM: Nucleomorph and plastid genome sequences of the chlorarachniophyte Lotharella oceanica: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. *BMC Genomics* 2014, 15:374.

42. Grosche C, Hempel F, Bolte K, Zauner S, Maier UG: The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. *Curr Opin Microbiol* 2014, 22:88-93.

43. Xu P, Widmer G, Wang YP, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ *et al*: The genome of Cryptosporidium hominis. *Nature* 2004, 431(7012):1107-1112.

44. Gornik SG, Febrimarsa, Cassin AM, MacRae JI, Ramaprasad A, Rchiad Z, McConville MJ, Bacic A, McFadden GI, Pain A *et al*: Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(18):5767-5772.

45. McFadden GI, Yeh E: The apicoplast: now you see it, now you don't. *Int J Parasitol* 2017, 47(2-3):137-144.

46. Arisue N, Hashimoto T: Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol Int* 2015, 64(3):254-259.

47. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF, McFadden GI: Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. *Science* 2003, 299(5607):705-708.

48. Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV: Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 1998, 393(6681):162-165.

49. Straub SP, Stiller SB, Wiedemann N, Pfanner N: Dynamic organization of the mitochondrial protein import machinery. *Biol Chem* 2016, 397(11):1097-1114.

50. Paila YD, Richardson LG, Schnell DJ: New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *J Mol Biol* 2015, 427(5):1038-1060.

51. Anderson S, Smith SM: Synthesis of the Small Subunit of Ribulose-Bisphosphate Carboxylase from Genes Cloned into Plasmids Containing the Sp6 Promoter. *Biochemical Journal* 1986, 240(3):709-715.

52. Mishkind ML, Wessler SR, Schmidt GW: Functional Determinants in Transit Sequences - Import and Partial Maturation by Vascular Plant Chloroplasts of the Ribulose-1,5-Bisphosphate Carboxylase Small Subunit of Chlamydomonas. *Journal of Cell Biology* 1985, 100(1):226-234.

53. Bolter B, Soll J: Once upon a Time - Chloroplast Protein Import Research from Infancy to Future Challenges. *Mol Plant* 2016, 9(6):798-812.

54. Tonkin CJ, Roos DS, McFadden GI: N-terminal positively charged amino acids, but not their exact position, are important for apicoplast transit peptide fidelity in Toxoplasma gondii. *Mol Biochem Parasitol* 2006, 150(2):192-200.

55. Patron NJ, Waller RF: Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2007, 29(10):1048-1058.

56. Garg SG, Gould SB: The Role of Charge in Protein Targeting Evolution. *Trends in cell biology* 2016, 26(12):894-905.

57. DeRocher A, Hagen CB, Froehlich JE, Feagin JE, Parsons M: Analysis of targeting sequences demonstrates that trafficking to the Toxoplasma gondii plastid branches off the secretory system. *J Cell Sci* 2000, 113 (Pt 22):3969-3977.

58. Wastl J, Maier UG: Transport of proteins into cryptomonads complex plastids. *J Biol Chem* 2000, 275(30):23194-23198.

59. Bruce BD: Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol* 2000, 10(10):440-447.

60. Oreb M, Tews I, Schleiff E: Policing Tic 'n' Toc, the doorway to chloroplasts. *Trends Cell Biol* 2008, 18(1):19-27.

61. Gallagher JR, Matthews KA, Prigge ST: Plasmodium falciparum apicoplast transit peptides are unstructured in vitro and during apicoplast import. *Traffic* 2011, 12(9):1124-1138.

62. Tonkin CJ, Struck NS, Mullin KA, Stimmler LM, McFadden GI: Evidence for Golgi-independent transport from the early secretory pathway to the plastid in malaria parasites. *Mol Microbiol* 2006, 61(3):614-630.

63. Rapoport TA: Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 2007, 450(7170):663-669.

64. van Dooren GG, Su V, D'Ombrain MC, McFadden GI: Processing of an apicoplast leader sequence in Plasmodium falciparum and the identification of a putative leader cleavage enzyme. *J Biol Chem* 2002, 277(26):23612-23619.

65. Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier UG: Protein targeting into the complex plastid of cryptophytes. *J Mol Evol* 2006, 62(6):674-681.

66. Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N *et al*: Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat Cell Biol* 2005, 7(12):1224-1231.

67. Peschke M, Moog D, Klingl A, Maier UG, Hempel F: Evidence for glycoprotein transport into complex plastids. *Proc Natl Acad Sci U S A* 2013, 110(26):10860-10865.

68. Deane JA, Fraunholz M, Su V, Maier UG, Martin W, Durnford DG, McFadden GI: Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. *Protist* 2000, 151(3):239-252.

69. Waller RF, Keeling PJ, Donald RGK, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS, McFadden GI: Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(21):12352-12357.

70. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG: Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol Biol* 2007, 64(5):519-530.

71. Steiner JM, Yusa F, Pompe JA, Loffelhardt W: Homologous protein import machineries in chloroplasts and cyanelles. *Plant J* 2005, 44(4):646-652.

72. Patron NJ, Waller RF, Archibald JM, Keeling PJ: Complex protein targeting to dinoflagellate plastids. *J Mol Biol* 2005, 348(4):1015-1024.

73. Kohler D, Dobritzsch D, Hoehenwarter W, Helm S, Steiner JM, Baginsky S: Identification of protein N-termini in Cyanophora paradoxa cyanelles: transit peptide composition and sequence determinants for precursor maturation. *Frontiers in plant science* 2015, 6.

74. Wunder T, Martin R, Loffelhardt W, Schleiff E, Steiner JM: The invariant phenylalanine of precursor proteins discloses the importance of Omp85 for protein translocation into cyanelles. *BMC Evol Biol* 2007, 7:236.

75. Ralph SA, Foth BJ, Hall N, McFadden GI: Evolutionary pressures on apicoplast transit peptides. *Mol Biol Evol* 2004, 21(12):2183-2194.

76. Sheiner L, Fellows JD, Ovciarikova J, Brooks CF, Agrawal S, Holmes ZC, Bietz I, Flinner N, Heiny S, Mirus O *et al*: Toxoplasma gondii Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins. *Traffic* 2015.

77. Harb OS, Chatterjee B, Fraunholz MJ, Crawford MJ, Nishi M, Roos DS: Multiple functionally redundant signals mediate targeting to the apicoblast in the apicomplexan parasite Toxoplasma gondii. *Eukaryotic cell* 2004, 3(3):663-674.

78. Sulli C, Fang Z, Muchhal U, Schwartzbach SD: Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J Biol Chem* 1999, 274(1):457-463.

79. Vacula R, Slavikova S, Schwartzbach SD: Protein trafficking to the complex chloroplasts of Euglena. *Methods Mol Biol* 2007, 390:219-237.

80. Nassoury N, Cappadocia M, Morse D: Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* 2003, 116(Pt 14):2867-2874.

81. Slavikova S, Vacula R, Fang Z, Ehara T, Osafune T, Schwartzbach SD: Homologous and heterologous reconstitution of Golgi to chloroplast transport and protein import into the complex chloroplasts of Euglena. *J Cell Sci* 2005, 118(Pt 8):1651-1661.

82. Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ, Maier UG: Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol Biol Evol* 2006, 23(12):2413-2422.

83. Kilian O, Kroth PG: Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J* 2005, 41(2):175-183.

84. Bhaya D, Grossman A: Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. *Mol Gen Genet* 1991, 229(3):400-404.

85. Tomova C, Humbel BM, Geerts WJ, Entzeroth R, Holthuis JC, Verkleij AJ: Membrane contact sites between apicoplast and ER in Toxoplasma gondii revealed by electron tomography. *Traffic* 2009, 10(10):1471-1480.

86. DeRocher AE, Coppens I, Karnataki A, Gilbert LA, Rome ME, Feagin JE, Bradley PJ, Parsons M: A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in Toxoplasma gondii. *Eukaryot Cell* 2008, 7(9):1518-1529.

87. Karnataki A, Derocher A, Coppens I, Nash C, Feagin JE, Parsons M: Cell cycle-regulated vesicular trafficking of Toxoplasma APT1, a protein localized to multiple apicoplast membranes. *Mol Microbiol* 2007, 63(6):1653-1668.

88. van Dooren GG, Tomova C, Agrawal S, Humbel BM, Striepen B: Toxoplasma gondii Tic20 is essential for apicoplast protein import. *Proc Natl Acad Sci U S A* 2008, 105(36):13574-13579.

89. van Dooren GG, Reiff SB, Tomova C, Meissner M, Humbel BM, Striepen B: A novel dynamin-related protein has been recruited for apicoplast fission in Toxoplasma gondii. *Curr Biol* 2009, 19(4):267-276.

90. Hirakawa Y, Burki F, Keeling PJ: Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. *Eukaryot Cell* 2012, 11(3):324-333.

91. DeRocher A, Gilbert B, Feagin JE, Parsons M: Dissection of brefeldin A-sensitive and -insensitive steps in apicoplast protein targeting. *J Cell Sci* 2005, 118(Pt 3):565-574.

92. Heiny SR, Pautz S, Recker M, Przyborski JM: Protein Traffic to the Plasmodium falciparum apicoplast: evidence for a sorting branch point at the Golgi. *Traffic* 2014, 15(12):1290-1304.

93. Sheiner L, Demerly JL, Poulsen N, Beatty WL, Lucas O, Behnke MS, White MW, Striepen B: A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. *PLoS Pathog* 2011, 7(12):e1002392.

94. Karnataki A, DeRocher AE, Feagin JE, Parsons M: Sequential processing of the Toxoplasma apicoplast membrane protein FtsH1 in topologically distinct domains during intracellular trafficking. *Mol Biochem Parasitol* 2009, 166(2):126-133.

95. Karnataki A, Derocher AE, Coppens I, Feagin JE, Parsons M: A membrane protease is targeted to the relict plastid of toxoplasma via an internal signal sequence. *Traffic* 2007, 8(11):1543-1553.

96. DeRocher AE, Karnataki A, Vaney P, Parsons M: Apicoplast targeting of a Toxoplasma gondii transmembrane protein requires a cytosolic tyrosine-based motif. *Traffic* 2012, 13(5):694-704.

97. Lim L, Sayers CP, Goodman CD, McFadden GI: Targeting of a Transporter to the Outer Apicoplast Membrane in the Human Malaria Parasite Plasmodium falciparum. *PLoS One* 2016, 11(7):e0159603.

98. Bouchut A, Geiger JA, DeRocher AE, Parsons M: Vesicles bearing Toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption. *PLoS One* 2014, 9(11):e112096.

99. Kaur J, Debnath J: Autophagy at the crossroads of catabolism and anabolism. *Nat Rev Mol Cell Bio* 2015, 16(8):461-472.

100. Brennand A, Gualdron-Lopez M, Coppens I, Rigden DJ, Ginger ML, Michels PAM: Autophagy in parasitic protists: Unique features and drug targets. *Mol Biochem Parasit* 2011, 177(2):83-99.

101. Besteiro S, Brooks CF, Striepen B, Dubremetz JF: Autophagy Protein Atg3 is Essential for Maintaining Mitochondrial Integrity and for Normal Intracellular Development of Toxoplasma gondii Tachyzoites. *PLoS pathogens* 2011, 7(12).

102. Ghosh D, Walton JL, Roepe PD, Sinai AP: Autophagy is a cell death mechanism in Toxoplasma gondii. *Cell Microbiol* 2012, 14(4):589-607.

103. Kong-Hap MA, Mouammine A, Daher W, Berry L, Lebrun M, Dubremetz JF, Besteiro S: Regulation of ATG8 membrane association by ATG4 in the parasitic protist Toxoplasma gondii. *Autophagy* 2013, 9(9):1334-1348.

104. Lavine MD, Arrizabalaga G: Analysis of Monensin Sensitivity in Toxoplasma gondii Reveals Autophagy as a Mechanism for Drug Induced Death. *PloS one* 2012, 7(7).

105. Tomlins AM, Ben-Rached F, Williams RA, Proto WR, Coppens I, Ruch U, Gilberger TW, Coombs GH, Mottram JC, Muller S *et al*: Plasmodium falciparum ATG8 implicated in both autophagy and apicoplast formation. *Autophagy* 2013, 9(10):1540-1552.

106. Kitamura K, Kishi-Itakura C, Tsuboi T, Sato S, Kita K, Ohta N, Mizushima N: Autophagy-related Atg8 localizes to the apicoplast of the human malaria parasite Plasmodium falciparum. *PLoS One* 2012, 7(8):e42977.

107. Leveque MF, Berry L, Cipriano MJ, Nguyen HM, Striepen B, Besteiro S: Autophagy-Related Protein ATG8 Has a Noncanonical Function for Apicoplast Inheritance in Toxoplasma gondii. *mBio* 2015, 6(6).

108. Di Paolo G, De Camilli P: Phosphoinositides in cell regulation and membrane dynamics. *Nature* 2006, 443(7112):651-657.

109. Tawk L, Dubremetz JF, Montcourrier P, Chicanne G, Merezegue F, Richard V, Payrastre B, Meissner M, Vial HJ, Roy C *et al*: Phosphatidylinositol 3-monophosphate is involved in toxoplasma apicoplast biogenesis. *PLoS Pathog* 2011, 7(2):e1001286.

110. Tawk L, Chicanne G, Dubremetz JF, Richard V, Payrastre B, Vial HJ, Roy C, Wengelnik K: Phosphatidylinositol 3-phosphate, an essential lipid in Plasmodium, localizes to the food vacuole membrane and the apicoplast. *Eukaryot Cell* 2010, 9(10):1519-1530.

111. Daher W, Morlon-Guyot J, Sheiner L, Lentini G, Berry L, Tawk L, Dubremetz JF, Wengelnik K, Striepen B, Lebrun M: Lipid kinases are essential for apicoplast homeostasis in Toxoplasma gondii. *Cell Microbiol* 2015, 17(4):559-578.

112. Bolter B, Soll J: Once upon a Time - Chloroplast Protein Import Research from Infancy to Future Challenges. *Molecular plant* 2016, 9(6):798-812.

113. Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R: Characterization of the translocon of the outer envelope of chloroplasts. *J Cell Biol* 2003, 160(4):541-551.

114. Rudolf M, Machettira AB, Gross LE, Weber KL, Bolte K, Bionda T, Sommer MS, Maier UG, Weber AP, Schleiff E *et al*: In vivo function of Tic22, a protein import component of the intermembrane space of chloroplasts. *Mol Plant* 2013, 6(3):817-829.

115. Nakai M: The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim Biophys Acta* 2015, 1847(9):957-967.

116. Kovacs-Bogdan E, Benz JP, Soll J, Bolter B: Tic20 forms a channel independent of Tic110 in chloroplasts. *BMC Plant Biol* 2011, 11:133.

117. Kovacheva S, Bedard J, Patel R, Dudley P, Twell D, Rios G, Koncz C, Jarvis P: In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. *Plant J* 2005, 41(3):412-428.

118. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E: Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. *J Biol Chem* 2010, 285(9):6848-6856.

119. Kalanon M, Tonkin CJ, McFadden GI: Characterization of two putative protein translocation components in the apicoplast of Plasmodium falciparum. *Eukaryot Cell* 2009, 8(8):1146-1154.

120. Glaser S, van Dooren GG, Agrawal S, Brooks CF, McFadden GI, Striepen B, Higgins MK: Tic22 is an essential chaperone required for protein import into the apicoplast. *J Biol Chem* 2012, 287(47):39505-39512.

121. Kohler S, Delwiche CF, Denny PW, Tilney LG, Webster P, Wilson RJ, Palmer JD, Roos DS: A plastid of probable green algal origin in Apicomplexan parasites. *Science* 1997, 275(5305):1485-1489.

122. McFadden GI, van Dooren GG: Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol* 2004, 14(13):R514-516.

123. Waller RF, McFadden GI: The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr Issues Mol Biol* 2005, 7(1):57-79.

124. Grosche C, Hempel F, Bolte K, Zauner S, Maier UG: The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. *Current opinion in microbiology* 2014, 22:88-93.

125. Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM, Maier UG: Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Mol Biol Evol* 2007, 24(4):918-928.

126. Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, Martin W, Maier UG: ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biol Evol* 2011, 3:140-150.

127. Agrawal S, van Dooren GG, Beatty WL, Striepen B: Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. *J Biol Chem* 2009, 284(48):33683-33691.

128. Spork S, Hiss JA, Mandel K, Sommer M, Kooij TW, Chu T, Schneider G, Maier UG, Przyborski JM: An unusual ERAD-like complex is targeted to the apicoplast of Plasmodium falciparum. *Eukaryot Cell* 2009, 8(8):1134-1145.

129. Hempel F, Bullmann L, Lau J, Zauner S, Maier UG: ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol Biol Evol* 2009, 26(8):1781-1790.

130. Smith MH, Ploegh HL, Weissman JS: Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. *Science* 2011, 334(6059):1086-1090.

131. Stein A, Ruggiano A, Carvalho P, Rapoport TA: Key steps in ERAD of luminal ER proteins reconstituted with purified components. *Cell* 2014, 158(6):1375-1388.

132. Goder V: Roles of Ubiquitin in Endoplasmic Reticulum-Associated Protein Degradation (ERAD). *Current protein & peptide science* 2012, 13(5):425-435.

133. Xia D, Tang WK, Ye YH: Structure and function of the AAA + ATPase p97/Cdc48p. *Gene* 2016, 583(1):64-77.

134. Niwa H, Ewens CA, Tsang C, Yeung HO, Zhang X, Freemont PS: The role of the N-domain in the ATPase activity of the mammalian AAA ATPase p97/VCP. *J Biol Chem* 2012, 287(11):8561-8570.

135. Yeung HO, Kloppsteck P, Niwa H, Isaacson RL, Matthews S, Zhang X, Freemont PS: Insights into adaptor binding to the AAA protein p97. *Biochem Soc Trans* 2008, 36(Pt 1):62-67.

136. Meyer H, Bug M, Bremer S: Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat Cell Biol* 2012, 14(2):117-123.

137. Richly H, Rape M, Braun S, Rumpf S, Hoege C, Jentsch S: A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* 2005, 120(1):73-84.

138. Ye Y, Meyer HH, Rapoport TA: Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. *J Cell Biol* 2003, 162(1):71-84.

139. Meyer HH, Shorter JG, Seemann J, Pappin D, Warren G: A complex of mammalian ufd1 and npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J* 2000, 19(10):2181-2192.

140. Ye Y, Meyer HH, Rapoport TA: The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. *Nature* 2001, 414(6864):652-656.

141. Schuberth C, Buchberger A: Membrane-bound Ubx2 recruits Cdc48 to ubiquitin ligases and their substrates to ensure efficient ER-associated protein degradation. *Nature cell biology* 2005, 7(10):999-U102.

142. Claessen JH, Mueller B, Spooner E, Pivorunas VL, Ploegh HL: The transmembrane segment of a tail-anchored protein determines its degradative fate

through dislocation from the endoplasmic reticulum. *J Biol Chem* 2010, 285(27):20732-20739.

143. Kothe M, Ye Y, Wagner JS, De Luca HE, Kern E, Rapoport TA, Lencer WI: Role of p97 AAA-ATPase in the retrotranslocation of the cholera toxin A1 chain, a non-ubiquitinated substrate. *J Biol Chem* 2005, 280(30):28127-28132.

144. Rutledge AC, Qiu W, Zhang R, Kohen-Avramoglu R, Nemat-Gorgani N, Adeli K: Mechanisms targeting apolipoprotein B100 to proteasomal degradation: evidence that degradation is initiated by BiP binding at the N terminus and the formation of a p97 complex at the C terminus. *Arterioscler Thromb Vasc Biol* 2009, 29(4):579-585.

145. Walters KJ: Ufd1 exhibits dual ubiquitin binding modes. *Structure* 2005, 13(7):943-944.

146. Park S, Isaacson R, Kim HT, Silver PA, Wagner G: Ufd1 exhibits the AAA-ATPase fold with two distinct ubiquitin interaction sites. *Structure* 2005, 13(7):995-1005.

147. Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG: Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryot Cell* 2012, 11(12):1472-1481.

148. Schuberth C, Buchberger A: UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. *Cell Mol Life Sci* 2008, 65(15):2360-2371.

149. Funakoshi M, Sasaki T, Nishimoto T, Kobayashi H: Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proc Natl Acad Sci U S A* 2002, 99(2):745-750.

150. Suzuki T, Park H, Lennarz WJ: Cytoplasmic peptide:N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions. *FASEB J* 2002, 16(7):635-641.

151. Moog D, Stork S, Zauner S, Maier UG: In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. *Genome Biol Evol* 2011, 3:375-382.

152. Bhattacharyya S, Yu HQ, Mim C, Matouschek A: Regulated protein turnover: snapshots of the proteasome in action. *Nat Rev Mol Cell Bio* 2014, 15(2):122-133.

153. Pickart CM, Eddins MJ: Ubiquitin: structures, functions, mechanisms. *Bba-Mol Cell Res* 2004, 1695(1-3):55-72.

154. Agrawal S, Chung DW, Ponts N, van Dooren GG, Prudhomme J, Brooks CF, Rodrigues EM, Tan JC, Ferdig MT, Striepen B *et al*: An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins. *PLoS Pathog* 2013, 9(6):e1003426.

155. Hempel F, Felsner G, Maier UG: New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Mol Microbiol* 2010, 76(3):793-801.

156. Hassink GC, Barel MT, Van Voorden SB, Kikkert M, Wiertz EJ: Ubiquitination of MHC class I heavy chains is essential for dislocation by human

cytomegalovirus-encoded US2 but not US11. *J Biol Chem* 2006, 281(40):30063-30071.

157. Ernst R, Claessen JH, Mueller B, Sanyal S, Spooner E, van der Veen AG, Kirak O, Schlieker CD, Weihofen WA, Ploegh HL: Enzymatic blockade of the ubiquitin-proteasome pathway. *PLoS Biol* 2011, 8(3):e1000605.

158. Claessen JH, Kundrat L, Ploegh HL: Protein quality control in the ER: balancing the ubiquitin checkbook. *Trends Cell Biol* 2012, 22(1):22-32.

159. Sanyal S, Claessen JH, Ploegh HL: A viral deubiquitylating enzyme restores dislocation of substrates from the endoplasmic reticulum (ER) in semi-intact cells. *J Biol Chem* 2012, 287(28):23594-23603.

160. Ernst R, Mueller B, Ploegh HL, Schlieker C: The otubain YOD1 is a deubiquitinating enzyme that associates with p97 to facilitate protein dislocation from the ER. *Mol Cell* 2009, 36(1):28-38.

161. Bernardi KM, Williams JM, Inoue T, Schultz A, Tsai B: A deubiquitinase negatively regulates retro-translocation of nonubiquitinated substrates. *Mol Biol Cell* 2013, 24(22):3545-3556.

162. Sasset L, Petris G, Cesaratto F, Burrone OR: The VCP/p97 and YOD1 Proteins Have Different Substrate-dependent Activities in Endoplasmic Reticulum-associated Degradation (ERAD). *Journal of Biological Chemistry* 2015, 290(47):28175-28188.

163. Lau JB, Stork S, Moog D, Sommer MS, Maier UG: N-terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids. *Mol Microbiol* 2015, 96(3):609-620.

164. Vierstra RD: The expanding universe of ubiquitin and ubiquitin-like modifiers. *Plant Physiol* 2012, 160(1):2-14.

165. Silmon de Monerri NC, Yakubu RR, Chen AL, Bradley PJ, Nieves E, Weiss LM, Kim K: The Ubiquitin Proteome of Toxoplasma gondii Reveals Roles for Protein Ubiquitination in Cell-Cycle Transitions. *Cell Host Microbe* 2015, 18(5):621-633.

166. Baldridge RD, Rapoport TA: Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD. *Cell* 2016, 166(2):394-407.

167. Ling Q, Jarvis P: Functions of plastid protein import and the ubiquitin-proteasome system in plastid development. *Biochim Biophys Acta* 2015, 1847(9):939-948.

168. Ling Q, Huang W, Baldwin A, Jarvis P: Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science* 2012, 338(6107):655-659.

Figures

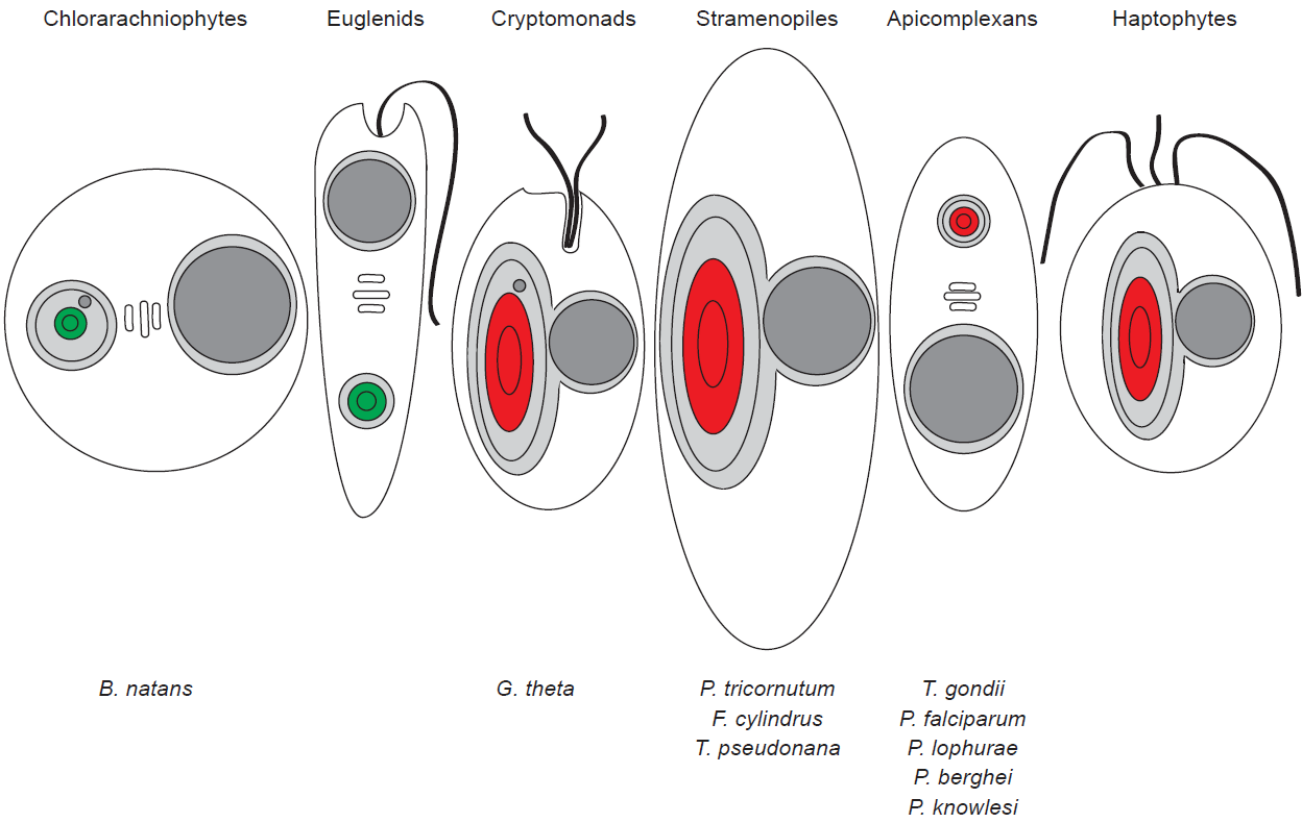


Figure 2.1. Overview of phyla and organisms. This literary review discusses the complex plastids of phyla: chlorarachniophytes, euglenids, cryptomonads, stramenopiles, apicomplexans and haptophytes. Chlorarchinophytes and euglenids are of green algal origin. Chlorarchniophytes contain a four membrane plastid with a nucleomorph genome while euglenids contain a 3 membrane plastid. Cryptomonads, stramenopiles and haptophytes contain four membrane plastids of red algal origin that reside in the endomembrane system. Cryptomonads contain a nucleomorph depicted by the dark gray circle in the second outermost compartment of the plastid. Apicomplexans have a four membrane plastid of red algal origin that is distinct from the ER membrane. Listed below the diagrams are organisms belonging to the phyla which have been useful experimental models reviewed in this chapter.

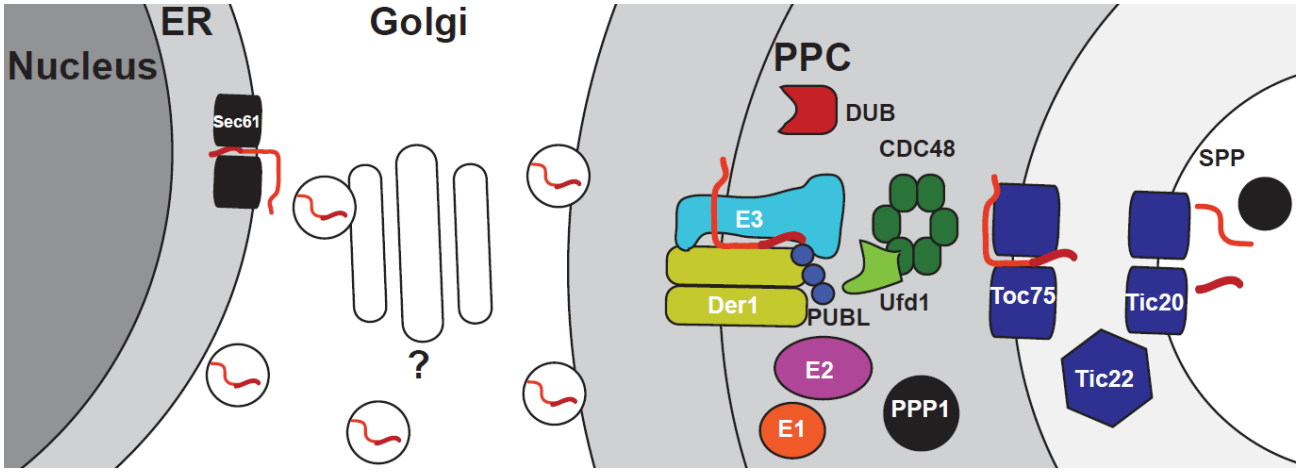


Figure 2.2. Nuclear encoded apicoplast protein import model. Schematic outline of the apicoplast protein import machinery. Proteins destined to the apicoplast are transported into the ER through the Sec61 complex (in black) and it is thought they travel by vesicles to the outermost membrane of the apicoplast. It is unclear whether proteins travel through the Golgi apparatus. Proteins then use an ERAD-like translocon that includes $Der1_{AP}$ and most likely the ubiquitin ligase (E3) to enter the periplastid compartment (PPC). We hypothesize that plastid ubiquitin-like protein (PUBL) is transferred from the ubiquitin activating enzyme (E1) to the ubiquitin conjugating enzyme (E2) which adds PUBL onto the imported protein. $Ufd1_{AP}$ recognizes the PUBL modification which allows $CDC48_{AP}$ to pull the protein through the translocon. It is believed that PPP1 (in black) acts as a chaperone to facilitate import across the periplastid membrane. We hypothesize that a deubiquitinase (DUB) subsequently cleaves PUBL followed by transport through the TOC and TIC translocons to reach the innermost compartment. The transit peptide is cleaved by a signal peptide peptidase (SPP) and the protein correctly folds to reach its mature form.

CHAPTER 3

A plastid ubiquitin-like protein evolved from ubiquitin and is required for apicoplast protein import in *Toxoplasma gondii*[1]

[1]Justin D. Fellows, Michael J. Cipriano, Swati Agrawal and Boris Striepen

Submitted to mBio 2017.

3.1 Abstract

Apicomplexan parasites cause a variety of important infectious diseases including malaria, Toxoplasma encephalitis, and severe diarrhea due to *Cryptosporidium*. Most apicomplexans depend on an organelle called the apicoplast which is derived from a red algal endosymbiont. The apicoplast is essential for the parasite as the compartment of fatty acid, heme, and isoprenoid biosynthesis. The majority of the approximate 500 apicoplast proteins are nuclear encoded and have to be imported across the four membranes that surround the apicoplast. Import across the second outermost membrane of the apicoplast, the periplastid membrane, depends on an apicoplast specific endoplasmic reticulum associated protein degradation (ERAD) complex and enzymes of the associated ubiquitination cascade. However, an apicoplast ubiquitin associated with this machinery has long been elusive. We identify a plastid ubiquitin-like protein (PUBL) an apicoplast protein that is derived from an ubiquitin ancestor, but has significantly changed in primary sequence. PUBL is distinct from known ubiquitin-like proteins and phylogenomic analyses suggest a clade specific to apicomplexans. We demonstrate that PUBL and the AAA ATPase $CDC48_{AP}$ both act to translocate apicoplast proteins across the periplastid membrane during protein import. Conditional null mutants and genetic complementation show that both proteins are critical for this process and parasite survival. PUBL residues homologous to those that are required for conjugation of a single ubiquitin onto target protein are essential for this function while those required for polyubiquitination and preprotein processing are dispensable. Our experiments provide mechanistic understanding of the molecular machinery that drives protein import across the membranes of the apicoplast.

3.2 Importance

Apicomplexan parasites are responsible for important human diseases. There are no effective vaccines for use in humans and drug treatment faces multiple challenges including emerging resistance, lack of efficacy across the lifecycle, and adverse drug effect. The apicoplast is a promising target for novel treatments: this chloroplast-like organelle is derived from an algal symbiont, absent from the host and essential for parasite growth and pathogenesis. We use *T. gondii* as model to study the apicoplast due to its strong genetic tools and established functional assays. We identify a plastid ubiquitin-like protein (PUBL) which is a novel ubiquitin-like protein and demonstrate its importance and that of the motor protein $CDC48_{AP}$ for apicoplast protein import. These findings broaden our understanding of the evolution and mechanistic workings of a unique parasite organelle and may lead to new opportunities for treatments against important human pathogens.

3.3 Introduction

Several important human and animal pathogens belong to the phylum Apicomplexa including the parasites that cause malaria, cryptosporidiosis and toxoplasmosis. *Toxoplasma gondii* is an intracellular parasite that infects about one third of the human population [1]. Infection usually persists throughout a person's life, but cellular immunity restricts the parasites to chronic tissue cysts. However, loss of immune function due to various types of immunosuppression results in reactivation of the infections which can have dire consequences to an individual's health including encephalitis and myocarditis [2]. Another major concern is transmission of *T. gondii* from mother to fetus when initial

infection of the woman occurs during pregnancy. Congenital toxoplasmosis can result in birth defects including hydrocephalus, blindness and still births [3].

Chloroplasts are the home of photosynthesis, and the hallmark of plants and numerous multicellular algae and single celled algal protists. Plastids evolved through endosymbiosis in which a free living photosynthetic prokaryote was taken up by an eukaryotic cell. Subsequently different unicellular algae were engulfed by a range of eukaryotes producing a remarkable diversity of photosynthetic organisms. Apicomplexans are one of the more surprising offshoots of this evolutionary tree and have a organelle called the apicoplast which is derived from a secondary endosymbiotic event between a red alga and a flagellated heterotrophic protist [4]. While apicomplexan parasites are no longer photosynthetic, the apicoplast is essential for the parasite as the location of fatty acid, heme and isoprenoid biosynthesis. The relative importance of each pathway differs between species and lifecycle stages and appears dictated by the parasites opportunity to scavenge host metabolites [4]. The vast majority of apicoplast proteins are nuclear encoded and thus must be imported across the four membranes that surround the apicoplast in order to maintain proper organelle function [5]. Nuclear encoded apicoplast proteins are targeted to the organelle through the secretory pathway, most often this depends on the presence of an N-terminal bipartite leader peptide [6]. The leader is made up of a signal peptide which is believed to facilitate co-translational insertion into the ER and a transit peptide which directs proteins to the apicoplast [7]. Vesicles carrying apicoplast proteins were described in multiple reports and are thought to bud from the ER to subsequently fuse with the outermost membrane of the apicoplast [8-10]. It has also been established that machinery homologous to Tic and Toc, the translocons of the inner

and outer chloroplast membrane which import proteins into primary plastids are found in the apicoplast and mediate import across the innermost and second innermost membrane of the apicoplast respectively [11-13]. The second outermost or periplastid membrane (PPM) is currently believed to be crossed using a specialized set of proteins that are derived from endoplasmic reticulum associated degradation (ERAD) proteins [14]. The ERAD machinery typically acts as a quality control system for protein folding in the ER and the secretory pathway. Misfolded proteins are recognized and exported across the ER membrane where they are marked by ubiquitination leading to subsequent degradation by the proteasome [15].

We have previously provided genetic evidence for the model that the apicoplast ERAD machinery, including the ubiquitin conjugating enzyme ($E2_{AP}$), has been retooled for protein import rather than protein degradation [16]. While multiple ERAD components have been identified in the apicoplast of *T. gondii*, including $CDC48_{AP}$ and $Ufd1_{AP}$, their function and potential interaction with the ubiquitin machinery is still unclear [14]. In the ERAD system CDC48 is a hexameric AAA ATPase that provides the mechanical force to unfold and extract misfolded proteins across the ER membrane [17]. At the moment it is unclear how apicoplast imported proteins are recognized by the import machinery as the apicoplast ERAD machinery is reduced compared to other complex plastids which contain additional plastid ERAD components [18]. We propose that $CDC48_{AP}$ and the ubiquitination machinery act in recognizing proteins at the periplastid membrane and transporting apicoplast proteins across the membrane. This idea is supported by the finding that CDC48 and its cofactor Ufd1 have ubiquitin binding domains and that ubiquitin recognition is necessary for proper translocation of misfolded

proteins across the ER membrane [19, 20]. Similarly, studies suggest that ubiquitination in the ERAD system has an initial mechanistic role in protein translocation across the ER membrane in addition to its subsequent role in protein degradation [21]. In addition, the ER membrane spanning ubiquitin ligase is key in substrate recognition and its autoubiquitination is mandatory for translocation of substrates [22, 23].

In this study we identify an ubiquitin-like protein that is localized to the apicoplast and differs in its amino acid sequence significantly from known ubiquitin-like proteins. We provide genetic evidence that the ubiquitin-like protein and $CDC48_{AP}$ are critical for parasite survival and import across the periplastid membrane (PPM) of the apicoplast. We also demonstrate that the C-terminal diglycine motif of this ubiquitin-like protein is critical to its function. The data suggests that conjugation of the ubiquitin-like protein onto imported proteins and the ATPase domain of $CDC48_{AP}$ are a mechanistic requirement for import into the apicoplast.

3.4 Results

$CDC48_{AP}$ is critical for parasite survival and protein import into the apicoplast

CDC48 is a highly conserved protein found in a wide array of eukaryotic organisms. CDC48 is an AAA ATPase typically localized to the cytoplasm where it is involved in a multitude of functions including cell cycle regulation, transcriptional activation, apoptosis, autophagy, endolysosomal sorting, and ER associated degradation (ERAD) [24]. In these varied contexts CDC48 uses the energy of ATP hydrolysis to unfold proteins, disassemble protein complexes, or translocate proteins across membranes. We have previously shown that there are two copies of CDC48 in *T. gondii* with one localizing to the cytoplasm while the other ($CDC48_{AP}$) is localized to the periplastid

compartment (PPC) of the apicoplast [14]. This is consistent with the idea that $CDC48_{AP}$ is part of the ERAD derived complex that aids proteins to cross the periplastid membrane (PPM) of the apicoplast. Specifically, we hypothesize that $CDC48_{AP}$ acts as the motor of the translocon. However, our initial attempts to test this and to generate a conditional mutant using a regulated ectopic copy and targeting plasmids failed. We thus modified a fosmid containing the $CDC48_{AP}$ locus to replace its promoter [25]. The engineered fosmid was transfected into a parasite line that is limited to homologous recombination and carries a tetracycline repressible transactivator (ΔKu80/TATi), and selected in the presence of pyrimethamine to isolate a stable line carrying a $CDC48_{AP}$ conditional mutant locus ((i)Δ$CDC48_{AP}$, Fig. S3.1A). PCR analyses were performed to determine whether in the (i)Δ$CDC48_{AP}$ line the endogenous promoter was indeed replaced with the regulatable t7s4 promoter, experiments with the (i)Δ$CDC48_{AP}$ line amplified the t7s4 promoter while showing loss of the endogenous promoter (Fig. S3.1B). In this mutant line $CDC48_{AP}$ gene expression was ablated upon the addition of anhydrous tetracycline (ATc). Western blots revealed that levels of the larger band or $CDC48_{AP}$ with its N-terminal bipartite leader sequence uncleaved is diminished after 1 day of ATc treatment while the smaller or cleaved mature form of $CDC48_{AP}$ is lost entirely after 3 days of treatment (Fig. 3.1A). The level of control proteins remained unchanged. The loss of the unprocessed version of an apicoplast protein prior to the mature form of the protein is typical in conditional mutants as the mature forms of apicoplast proteins remain in a steady state in the apicoplast while the unprocessed forms are no longer expressed or rapidly cleaved to the mature form [14].

The development of a conditional mutant allowed us to test whether $CDC48_{AP}$ is necessary for parasite survival. We conducted plaque assays to examine the importance of the protein. In the absence of ATc (i)$\Delta CDC48_{AP}$ parasites invade, replicate and egress which result in plaque formation. However, in the presence of ATc no plaques form suggesting that $CDC48_{AP}$ is critical for parasite growth (Fig. 3.1B). We constructed a (i)$\Delta CDC48_{AP}$ strain expressing an ectopic copy of $CDC48_{AP}$ tagged with an epitope tag ($CDC48_{AP}$-myc) and repeated the plaque assay. The complemented mutant line was able to form plaques even upon the addition of ATc demonstrating that the loss of plaque formation of the (i)$\Delta CDC48_{AP}$ line is directly linked to the loss of $CDC48_{AP}$. CDC48 has two ATPase domains and each contains a Walker A (GXXXXGKT/S, where X is any amino acid) and a Walker B motif ($\Phi\Phi\Phi\Phi$DE, where Φ is a hydrophobic amino acid) for ATP binding and hydrolysis, respectively and critical to the role of the protein as part of the ERAD system [26]. Our annotation of the apicoplast $CDC48_{AP}$ shows homologous ATPase domains with Walker motifs (Fig. S3.2A-B). We constructed point mutations in the critical residues (502 K/A and 829 E/Q). In transient transfections mutant proteins properly localize to the apicoplast (Fig. S3.2C-D). However, we consistently failed to isolate stable transgenic parasites suggesting a strong dominant negative effect of these mutations.

We next tested the ability of the (i)$\Delta CDC48_{AP}$ line to import apicoplast proteins in the presence and absence of ATc. The majority of nuclear encoded apicoplast proteins have a N-terminal bipartite leader peptide which consist of a signal peptide followed by a transit peptide [27]. The transit peptide which directs the protein to the apicoplast is cleaved by an unknown protease in the lumen of the apicoplast, proteins that remain in

the periphery show similar maturation that likely depends on a peripheral maturase. Western blot analysis of apicoplast proteins thus often results in two bands. A larger band represents the apicoplast protein *en route* to the organelle and a smaller band represents the mature protein that has been processed in the lumen of the apicoplast. The loss of apicoplast import results in the loss of this processing of apicoplast protein and we have previously exploited this to measure apicoplast import in several mutants [13, 14, 16, 28].

Acyl carrier protein (ACP) a protein that targets to the apicoplast lumen was endogenously tagged with YFP in the (i)ΔCDC48$_{AP}$ line (Fig. 3.1C). The parasite strain was grown from 0 to 3 days under ATc treatment and then harvested for Western blot analysis. The mature form of the protein was lost after 3 days of ATc treatment while the level of the control and larger precursor band remained unchanged throughout the experiment (Fig. 3.1E). This suggests that the protein does not reach the lumen for processing and that parasites lacking CDC48$_{AP}$ thus display an apicoplast import defect. To test whether CDC48$_{AP}$ is also important for the import of apicoplast proteins that reside in the periphery of the organelle we followed the protein encoded by TGME49_201270 [29]. The transit peptide of the peripheral apicoplast protein encoded by TGME49_201270 is cleaved in the periplastid compartment (PPC) after import across the periplastid membrane (PPM) rather than in the lumen of the apicoplast [11, 29]. TgMe49_201270 was endogenously tagged with an HA epitope tag in the (i)ΔCDC48$_{AP}$ parasite line (Fig. 3.1D), grown under ATc and analyzed by Western blot. Treatment with ATc resulted in the loss of the mature form of the protein after 2 days (Fig. 3.1F). The loss of processing of peripheral and luminal apicoplast proteins suggests that (i)ΔCDC48$_{AP}$ acts early in protein import and is consistent with its site of residence and

activity in the periplastid compartment (PPC) rather than the lumen. To control for the possibility that loss of the mature form of import reporters may be due to loss of the organelle as the consequence of a broader role of $CDC48_{AP}$ in apicoplast biology, we monitored the numbers of apicoplasts in the (i)$\Delta CDC48_{AP}$ strain under knock down conditions through loss of immunofluorescence staining and quantitative PCR (Fig. 3.1G, Fig. S3.3 and Fig. S3.4A). No other structural abnormalities were observed in the mutant line and the parental line showed no apicoplast biogenesis defects upon the addition of ATc (Fig. 3.1H). There was no significant decrease in apicoplast numbers in the (i)$\Delta CDC48_{AP}$ line until the fourth day or five day of ATc treatment, in contrast protein import was already affected on day two and three, arguing for a direct role of $CDC48_{AP}$ in apicoplast protein import.

A novel ubiquitin-like protein is localized to the apicoplast

We have previously demonstrated the importance of the ubiquitin conjugating enzyme $E2_{AP}$ in the PPC of the apicoplast in *T. gondii* [16]. A conditional mutant for this enzyme resulted in apicoplast import defects similar to those seen here for $CDC48_{AP}$. Ubiquitination machinery and $CDC48_{AP}$ likely work together in the PPC to import proteins. CDC48 and its many cofactors have ubiquitin binding domains which have been shown to bind and interact with polyubiquitin and monoubiquitin [19, 20, 24, 30]. $CDC48_{AP}$ retained the conserved N domain with its double-psi β barrel motif that is a known ubiquitin interaction site (Fig. S3.2E). However, so far we have been unable to demonstrate import of cytoplasmic ubiquitin into the apicoplast or the presence of a specific apicoplast targeted ubiquitin. Therefore, we broadened our search for an apicoplast specific modifier to include ubiquitin-like proteins by using BLAST searches

to identify proteins with a putative signal peptide and an ubiquitin-like domain, we also systematically reevaluated the gene models and transcription start sites of all ubiquitin-like genes. This effort yielded gene TgMe49_223125 which is predicted to encode a protein with a C-terminal domain with similarity to ubiquitin. For brevity we will from here on refer to the novel ubiquitin encoded by TgMe49_223125 as plastid ubiquitin-like protein (PUBL).

PUBL is 311 amino acids long, which is considerably larger than the 76 amino acid ubiquitin found in the cytoplasm of *T. gondii*. A sequence alignment between PUBL and *T. gondii* ubiquitin reveals a C-terminal ubiquitin-like domain with 56% identity (Fig. 3.2A). Note for comparison that *T. gondii* ubiquitin and human ubiquitin differ by a single residue. However, when we used *de novo* protein structure prediction algorithms we readily discern the ability of PUBL to form the beta-grasp fold (Fig. 3.2B) consisting of two beta sheets, an alpha helix and three additional beta sheets characteristic of ubiquitin and ubiquitin-like proteins [31]. The structure of human ubiquitin is shown for comparison and we note a high degree of similarity despite significant difference in primary sequence. We identified putative homologs of PUBL in numerous apicomplexans and similar to PUBL these proteins carry N-terminal extensions (Fig. S3.5 and S3.6 also see [18]). No homolog was identified in the *Cryptosporidium* species which lack an apicoplast. Figure 2C shows a phylogenetic tree of PUBL and homologs with identifiable leader peptides from a selection of organisms with a secondary red algal plastid. The PUBL of *P. tricornutum* has been previously experimentally validated to be targeted to the plastid [32]. Note that while cytoplasmic ubiquitins are extremely

conserved plastid homologs show variation. We conclude that PUBL and its homologs are derived from ubiquitin due to its sequence similarity and conserved structure.

The reason we overlooked this protein previously is that it is not recognized by algorithms typically used to detect apicoplast proteins due to the lack of an N-terminal signal peptide. However, PUBL contains a predicted transmembrane domain from amino acid 52 to 74 and we considered that this portion of the protein might serve as a recessed signal peptide (Fig. 3.2A). We expressed tagged versions as transgenes to localize the protein. First we tested whether the N-terminus of PUBL was capable to direct apicoplast import by expressing the first 180 amino acids with a GFP tag in *T. gondii*. The immunofluorescence assay showed GFP to colocalize with the apicoplast marker CPN60 [14] demonstrating that the N-terminus of PUBL acts as an apicoplast leader (Fig. 3.3A). Next we wanted to tag the entire protein with an epitope. This was complex as tagging the N-terminus would likely have interfered with the N-terminal apicoplast trafficking information. Conversely tagging ubiquitin-like proteins at the C-terminus is not practical either as this often results in the removal of the epitope tag by deubiquitinases and interferes with ubiquitination [33]. We therefore amplified the coding sequence of PUBL from cDNA by PCR and ligated it into an expression plasmid in a way that introduced an internal Ty-1 epitope tag (Fig. S3.8A and S3.8B). An immunofluorescence assay was performed on a parasite line expressing the tagged protein which showed a single punctate structure per cell. We counter stained with an antibody to CPN60, which produced labeling that overlapped with PUBL suggesting that PUBL is localized to the apicoplast (Fig. 3.3B). To validate this assignment independent of transgene overexpression we expressed and purified the ubiquitin-like domain of PUBL in bacteria

and raised monoclonal antibodies (Fig. 3.3E). Immunofluorescence assays using this new antibody again produced apicoplast labeling (Fig. 3.3C). While PUBL clearly is localized to the apicoplast we observed slight differences in staining of these reagents when compared to CPN60. We had previously noted comparable differences for proteins localized to the PPC [14, 16, 29]. While we cannot fully resolve the four membranes the images are consistent with residence of PUBL in the PPC of the apicoplast alongside the previously characterized ubiquitinating machinery. Overall, we concluded that PUBL is a novel ubiquitin-like protein found in the apicoplast and conserved among Apicomplexans.

As pointed out most apicoplast proteins are proteolytically processed to remove the transit peptide. In addition PUBL also has a RGG motif that immediately precedes its C-terminal ubiquitin-like domain. Deubiquitinases typically cleave polyubiquitin and other precursors at this position releasing the 76 amino acid ubiquitin domain [34]. Based on this precedence, we expected the size of mature PUBL to be 8.5 KDa similar to the size of ubiquitin. Multiple tagged versions of PUBL were generated to test this hypothesis including a form that mutated the two glycine residues at amino acid position 232 to alanine which should prevent cleavage. However, Western blot analysis revealed a single 26.5 kDa or 29.5 kDa band corresponding to the full-length PUBL protein for every version of the epitope tagged PUBL (Fig. 3.3D). This suggests lack of processing in PUBL. It is conceivable that we may fail to detect mature PUBL due to folding or steric hindrance or that epitope-tagging blocks processing. However, we note that this was a highly reproducible observation not only using a variety of epitope tags and tagging positions within the protein but also using native protein detected by our new antibody (Fig. 3.3F).

PUBL is essential for parasite survival and import across the PPM of the apicoplast

In order to test PUBL's importance in the apicoplast we constructed a conditional mutant ((i)ΔPUBL) following the strategy described for the $CDC48_{AP}$ mutant. PCR analysis indicated that in this PUBL mutant the tetracycline regulated promoter replaced the endogenous promoter (Fig. S3.1B) and Western blot experiments showed that PUBL is no longer detectable after the second day of ATc treatment (Fig. 3.4A). We performed plaque assays with the (i)ΔPUBL line. Addition of ATc to the mutant strain resulted in the loss of plaque formation (Fig. 3.4B), demonstrating that PUBL is required for parasite growth.

Previous worked showed that the PPC ubiquitinating machinery is crucial for import of proteins into the apicoplast [16]. We thus asked whether PUBL would also be essential for import. Ferredoxin NADPH reductase (FNR) is a luminal apicoplast protein which we are able to tag with red fluorescent protein [35]. A tagged version of FNR was used instead of ACP as in the $CDC48_{AP}$ mutant due to the difficulty in isolating a stable line. A stable line was isolated that ectopically expresses FNR-RFP in the PUBL mutant (Fig. 3.4C). This parasite line was treated with ATc for 0 to 4 days and prepared for Western blot analysis. The smaller mature band is diminished after two days and lost after 3 days of ATc treatment while the levels of control and precursor protein remained constant (Fig. 3.4E). We also endogenously tagged the peripheral apicoplast gene TgMe49_201270 with an HA epitope tag in the (i)ΔPUBL line (Fig. 3.4D). Again Western blot analysis showed reduced maturation after 2 days of ATc treatment and loss of the mature band after 3 days. We thus find loss of processing of luminal and peripheral proteins which suggests that PUBL is critical for an early step of protein import into the

apicoplast, likely translocation of the membrane that bounds the PPC. Apicoplast numbers were measured in the mutant strain via immunofluorescence loss and quantitative PCR and showed no significant loss of apicoplast numbers prior to loss of apicoplast protein import (Fig. 3.4G-H and Fig. S3.4B). No other structural abnormalities were observed in the mutant line.

Genetic complementation of mutant reveals terminal glycines to be critical for PUBL function

We tested whether the (i)ΔPUBL line could be complemented with an ectopic copy of PUBL. We introduced an extra copy of Ty-1 epitope tagged PUBL into the uracil-phosporibosyltransferase (UPRT) locus disrupting UPRT function. Loss of UPRT confers resistance to 5-fluorodeoxyuridine (FUDR) which we exploited for selection. We confirmed expression and correct localization of the ectopic Ty-1 epitope tagged PUBL through immunofluorescence assay (Fig. 3.5K-M and Fig. S3.8). The line is able to form plaques when cultured in the presence of ATc demonstrating that the extra copy was able to rescue the growth defect of the mutant line (Fig. 3.5A-B). Lysine residues are of particular importance in the biology of ubiquitin as they serve as sites for linkage to form polyubiquitin chains. Different lysine chain linkages are recognized as signals for distinct biological processes [36]. PUBL only has five lysine residues in the ubiquitin-like domain of the protein with four of the lysines being shared with ubiquitin and one non-conserved lysine at the 288th position. We systematically mutated the five lysine residues in the ubiquitin domain to the similarly charged arginine. Previous research has shown that mutating lysine to arginine prevents the formation of polyubiquitin lysine linked chains [37]. Mutations were engineered to the lysines at position 239, 244, 282, and 288

respectively of the ubiquitin domain of Ty-1 epitope tagged PUBL. These extra copies of PUBL were introduced into the UPRT locus of the (i)ΔPUBL line. Plaque assays were performed with these lines and expression of the mutated PUBL was able to complement the (i)ΔPUBL line for all lysine mutations suggesting that polyPUBL does not form or is not important (Fig. 3.5C-5F). We considered that PUBL may utilize lysine residues in a redundant fashion and that point mutations to single lysine residue thus would not affect polyPUBL chain formation. Therefore, we introduced into the (i)ΔPUBL line a tagged PUBL which contained all four lysine point mutations simultaneously. This line was able to fully complement the loss of endogenous PUBL (Fig. 3.5G). The experiment suggests that these four lysine residues are not critical for PUBL's function in the apicoplast. We note that we were unable to establish a stable line expressing a point mutation to the lysine at position 261 of the ubiquitin domain suggesting a dominant negative effect of this mutation. We were thus unable to test the importance of this residue.

One of the key features of ubiquitin and ubiquitin-like proteins is the ability to bind to other proteins via a C-terminal glycine and PUBL shares the requisite Di-glycine motif ([38] and see Fig. 3.2A). It is well documented that mutating these glycines to alanine residues results in conjugation-deficient ubiquitin and ubiquitin-like protein [39, 40]. We examined the importance of such conjugation by introducing an PUBL with mutated C-terminal glycines into the (i)ΔPUBL line. This mutant was unable to complement the (i)ΔPUBL line (Fig. 3.5H). We thus conclude that the PUBL C-terminal diglycine motif is indispensable for its function indicating PUBL transfer to substrate proteins similar to ubiquitination. As discussed previously, PUBL harbors a second diglycine motif immediately preceding the ubiquitin-like domain. Mutation of these

residues did not affect complementation (Fig. 3.5I). This is consistent with our observation of lack of processing at this site. We hypothesize that PUBL acts like ubiquitin and is transferred onto proteins. PUBL diverged from ubiquitin considerably but retained the stereotypical ubiquitin fold. We therefore tested whether ubiquitin, when appropriately localized, could rescue the loss of PUBL. An expression vector was engineered which replaced the ubiquitin domain of PUBL with *T. gondii* ubiquitin. As shown in Figure 3.5J, expression of the PUBL/ubiquitin chimera in the (i)ΔPUBL line (PUBL-UB-Ty Chimera) resulted in the ability of the line to form plaques under ATc. We noted a slight but consistent reduction of plaque size for this strain upon addition of ATc (Fig. 3.5J). We interpret this result to indicate that PUBL acts in a fashion highly similar to ubiquitin and that replacement of PUBL with ubiquitin thus only produces minor attenuation of that function.

3.5 Discussion

It is difficult to overrate the importance of endosymbiosis in the early evolution of eukaryotes. Many (but certainly not all) evolutionary biologists have come to view the initial endosymbiotic acquisition of an α-proteobacterium not only as the point of origin of the mitochondrion but as the very birth of the eukaryotic cell. Chloroplasts have a similar endosymbiotic genesis, this event harnessed the ability to photosynthesize and allowed eukaryotes to conquer primary production. Since then a series of secondary and tertiary events of uptake and loss has given rise to tremendous organismic diversity. Horizontal gene transfer from the newly acquired organelle to the host nucleus is a hallmark of all endosymbiosis events [41]. This process endows the host with control over its symbiont but also requires the concurrent evolution of post-translational

mechanisms to route symbiont proteins now encoded and produced by the host into the organelle. For mitochondria and primary chloroplasts specific cargo recognition and translocation complexes facilitate protein translocation across each of the membranes of the respective organelle and this is understood in considerable mechanistic detail [42, 43]. Apicomplexa and other phyla including chromera, dinoflagellates, haptophytes, cryptophytes and diatoms possess a secondary plastid of red algal origin [44, 45]. Consistent with its evolutionary origin the apicoplast shares import machinery with primary plastids, specifically the TOC and TIC complex which facilitate protein import across the outer and inner membrane of the chloroplast. In previous studies we demonstrated that the *T. gondii* apicoplast relies on a homolog of Toc75 for protein import across the second innermost membrane while Tic20 and Tic22 homologs are required for import across the innermost membrane [11-13]. The apicoplast is surrounded by four membranes and nuclear encoded proteins thus have to cross two membranes before they encounter the TOC complex.

Endosomal and autophagic pathways are candidate mechanisms that may guide to and fuse vesicles with the outermost apicoplast membrane. Intriguingly, autophagy related protein 8 as well as phosphatidylinositol (3) phosphate heavily accumulate on the surface of the apicoplast. However, genetic interference with these mechanisms in *T. gondii* and *P. falciparum* has not produced an unequivocal link to protein import but rather pointed to a broader role of autophagy in apicoplast morphogenesis and inheritance [46-48]. While transport across the outermost membrane thus remains to be unraveled, we know more about the periplastid membrane imported proteins encounter next. Pioneering observations by Maier and colleagues in cryptomonads led to the model that

the symbiont's ERAD pathway was retooled to serve as an import mechanism into the periplastid space – the former cytoplasm of the red alga [18, 32, 49-52]. This model received robust experimental support from genetic studies in *T. gondii* that demonstrated the requirement of the ERAD components Der1 and ubiquitin conjugating enzyme for apicoplast protein import [14, 16]. However, not all secondary plastids utilize such ERAD derived proteins. Secondary plastids derived from endosymbiosis of green algae appear to rely on an ERAD independent and yet to be fully characterized mechanism for import across the periplastid membrane [53].

In the current study we demonstrate that $CDC48_{AP}$ is a critical component of the ERAD-derived apicoplast import machine. Loss of the protein results in loss of import. We probed the mutant with different cargo proteins and found import of luminal and periplastid proteins to be equally blocked. This is in contrast to observations we recently reported for a *T. gondii* Toc75 mutant [11], in that mutant only the import of luminal proteins was ablated. This difference genetically establishes the order of the translocons in the apicoplast. $CDC48_{AP}$ likely acts as the motor of the translocon; point mutations in the Walker A and B motifs ablating $CDC48_{AP}$'s ATPase function proved to be highly deleterious and not tolerated by the parasite. The translocation activity of cytoplasmic CDC48 in the context of ERAD at the ER membrane requires ubiquitination of cargo [15, 19, 30]. $CDC48_{AP}$ like its cytoplasmic homolog features ubiquitin binding domains and we thus hypothesize that such an interaction occurs in the apicoplast.

There are numerous ubiquitin-like proteins recognized in eukaryotes that act in an array of diverse cellular processes and we propose PUBL as a new member of this family [54]. Ubiquitin is famous for being one of the most conserved proteins amongst

eukaryotes –there is only a single amino acid difference in the sequence of the cytoplasmic ubiquitin of *T. gondii* and *H. sapiens*. However, there is considerable sequence difference between *T. gondii* PUBL and ubiquitin. Similarly, homologs of PUBL and ubiquitin that carry identifiable plastid leaders are quite diverse (Fig. 3.2C). The absence of Plasmodium is conspicuous in our homolog search. A putative apicoplast specific ubiquitin-like protein (PF3D7_081570) has been previously identified to be localized to the apicoplast of *P. falciparum*. However, the gene was excluded from our study due to the gene missing key components of PUBL homologs including a leader sequence directly upstream of the ubiquitin domain and any diglycine motifs typical of all ubiquitin-like proteins. This gene also contains five different ubiquitin domains which makes it challenging to choose which ubiquitin domain to be used for the alignment. It should be noted that the authors also were unable to identify any substrates of this ubiquitin-like protein [55]. We believe this diversity of PUBL reflects a relaxation of functional constraint. Ubiquitin has to interact with literally hundreds of cellular proteins to fulfill its multitude of functions, these multifaceted interactions ensure conservation. Evolution streamlined and simplified the red alga into the apicoplast gradually reducing ubiquitin interactions and thus released the leash on sequence conservation.

Alternatively, changes in PUBL sequence could be the consequence of change in function. We conducted genetic experiments to address this question. Loss of PUBL was not tolerated by the parasite as it blocks apicoplast protein import. We note that as for $CDC48_{AP}$ this block applied to luminal and periplastid proteins associating both with the peripheral ERAD derived translocon. We used complementation analysis to further explore the functional relevance of specific residues. PUBL has five lysine residues that

potentially could enable poly-PUBL chain formation. We replaced four of these individually and collectively with arginine and those mutants fully complemented ATc induced gene ablation in trans. The C-terminal RGG motif typically required for conjugation onto substrate protein is a firm requirement of complementation whereas the RGG preceding the ubiquitin-like domain is not. Lastly, a chimera in which the C-terminus of PUBL is replaced with cytoplasmic ubiquitin complements. Overall this data strongly supports a model under which PUBL is conjugated via a C-terminal glycine onto cargo protein in a fashion analogous to ubiquitin. This is in agreement with our previous analysis of the apicoplast ubiquitinating enzymes in *T. gondii* and *P. falciparum* [16] and with studies on the diatom *P. tricornutum* which share ancestry and a secondary plastid with apicomplexans [56]. Polyubiquitin chains with specific lysine linkages have recently been identified in *T. gondii* and are shown to be recognized by specific deubiquitinases and accumulate at different points of the cell cycle emphasizing the complexity of polyubiquitin chains [57]. Poly-chain formation is apparently not critical for PUBL's function, this is fitting as this typically is associated with protein degradation. However, we caution that a K261R mutation was dominant negative which limited our ability to test all residues. K261 is highly conserved among PUBL and its homologs (Fig. S3.5 and [18]). Interestingly, proteolytic processing at the N-terminal RGG to release a 'mature' ubiquitin maturation from larger precursors [38] was not required for PUBL function, which is consistent with our Western blot measurements of the molecular mass of PUBL which suggest absence of processing (Fig. 3.3D and 3.3F).

What is the target of PUBL? Recent studies by Lau and colleagues highlighted the transit peptide as a likely site of cargo ubiquitination during plastid import in *P.*

tricornutum [58, 59]. Mutation of all lysines in this region abolish import, reintroduction of lysine in a different position of the transit peptide restores import. Revealingly, imported proteins lacking lysine residues in the leader peptide appeared “frozen” in the PPM where they associate with a 540 kDa protein complex. This complex appears to contain Der1 homologs and may further interact with homologs of UBX a protein that recognizes ubiquitin in the context of CDC48. Despite significant effort we were unable to directly biochemically demonstrate PUBL in peptide linkage to cargo proteins in *T. gondii* through pull down experiments with the PUBL antibody followed by Western blot or mass spectrometry (data not shown). A comprehensive study on ubiquitination in *T. gondii* highlights the difficulty in identifying a substrate as pulldown experiments showed the absence of ubiquitination or PUBL modifications in the apicoplast [60]. To our knowledge PUBL modifications have also not been formally demonstrated in diatoms [56]. We believe this to reflect the very transient nature of these modification, which may be restricted to the translocation event itself thus severely limiting the conjugated pool available for detection. Apicoplast specific deubiquitinases may act swiftly to remove PUBL upon translocation, PUBL transfer and removal could even be physically linked as part of a multiprotein translocation complex. The adaptation of ERAD to protein import required its dissociation from protein degradation. One satisfying hypothesis to potentially explain this is loss of the K48 polyubiquitination site on ubiquitin that typically drive proteasome interaction which was observed in diatoms [18, 61]. However, this residue is still present in PUBL (and its apicomplexan homologs) but is dispensable for its function. A translocation complex in which ubiquitination is transient and deubiquitination a requirement for cargo release may also protect cargo. These

hypotheses could be tested by identifying and genetically ablating apicoplast deubiquitinases. Loss of the activity could result in the accumulation of PUBL modified proteins. Such accumulation of ubiquitin modification has been observed for deubiquitinase mutants in other cellular contexts [62, 63].

We now understand PUBL to be an essential part of the apicoplast protein import machinery but important mechanistic aspects of its addition and particularly removal remain to be worked out. Pursuing these mechanisms may serve purpose beyond the obvious interest to evolutionary cell biology. Interference with ubiquitination and deubiquitination has emerged as a rich ground for the development of drugs targeting cancer and infection [64].

3.6 Materials and methods

Cell Culture and Transfection. *T. gondii* RH and ΔKu80/TATi strains were cultivated in human foreskin fibroblasts (HFFs) in Dulbecco's Modified Eagles Medium supplemented with fetal bovine serum, penicillin-streptomycin, and L-glutamine. Transfections were carried out by resuspending parasites in cytomix supplemented with 2mM ATP and 5 mM glutathione to 3.3 X 10^7 parasites per ml. 300 µl of the parasite suspension and 30 µg of plasmid were mixed and transferred to a 2 mm gap cuvette and electroporated using a single 1.5 kV pulse, a resistance of 25 **Ω**, and a capacitor setting of 25 µF. Parasites were selected in the presence of 1 µM of pyrimethamine, 20 µM of chloramphenicol, or 5 µM of FUDR respectively.

Tagging of Genes and Genetic Complementation. Vector pDC was constructed by replacing the tubulin promoter with *T. gondii* DHFR promoter in expression plasmid pTC [65]. The coding region of PUBL (TgME49_223125) was amplified from *T. gondii*

cDNA using primers that introduced a Ty-1 epitope tag immediately before the C-terminal ubiquitin domain (see Fig. S3.7 for detail). This internally tagged PUBL was inserted into vector pDC using BglII and EcoRV restriction cut sites. The vector expressing *T. gondii* cytoplasmic ubiquitin (TGME49_219820) fused to the N-terminus of PUBL (PUBL/ubiquitin vector) was constructed using a similar approach (see Table S3.1 in the supplemental material). The $CDC48_{AP}$ complementation vector was constructed by amplifying the coding sequence of $CDC48_{AP}$ from cDNA and inserting the amplicon into vector $pTCM_3$ which introduces a 3x myc epitope tag at the C-terminus of the $CDC48_{AP}$ protein. The peripheral apicoplast protein encoded by gene TGME49_201270 was endogenously tagged with a HAx3 tag by introducing a linearized p3HA.LIC.CATΔpacI-201270 vector into parasites [29]. Acyl carrier protein was endogenously tagged with an YFP tag by transfecting parasites with a linearized pLicCATYFPΔpacI-ACP vector. Ferredoxin NADPH reductase tagged with RFP was introduced into the (i)ΔPUBL line, subjected to flow cytometry to transgenics. If not stated otherwise all transgenic parasites used here were clonal lines established by limited dilution. Complementation assays were performed by transfecting 30 µg of a Cas9 plasmid that introduces a cut in the UPRT gene along with 30 µg of a PCR product that included the gene of interest marked with an epitope tag and suitable flanks to guide insertion into the UPRT locus by homologous recombination [66]. Transfectants were cultured in 5 µM of FUDR to select for disruption of the UPRT gene due to insertion of the ectopic complementation cassette, expression and proper localization of the complementing transgene was tested following the epitope tag by immunofluorescence assay (IFA, see Fig. 3.4K-M and Fig. S3.8). The QuikChange II Site-Directed

Mutagenesis Kit (Stratagene) was used to generate point mutations using the manufacturer's protocols and primers listed in Table S3.1.

Construction of Conditional Mutant. Mutants were constructed by replacing the endogenous promoter with the conditional promoter in a ΔKu80/TATi parasite line. A fosmid construct was engineered as previously described to replace the promoter of the gene of interest with the tetracycline conditional t7s4 promoter [25] (See Fig. S3.1). Primers used are listed in Table S3.1. The fosmids used here were RHfos08E17 and RHfos22J15 for the development of the $CDC48_{AP}$ and PUBL mutants respectively. The modified fosmids were transfected into the ΔKu80/TATi strain and pyrimethamine drug selection was used to isolate stable parasite lines. Parasite lines which successfully replaced the endogenous promoter with the conditional t7s4 promoter were identified by PCR mapping the genomic locus of the targeted gene (see Fig. S3.1). Mutants were tested by plaque assay as described [13].

Microscopy. HFF coverslip cultures were infected with parasites in the absence of ATc (unless otherwise stated), and 24 hours after infection fixed for 20 minutes with 4% paraformaldehyde, blocked with 3% BSA for 10 minutes, and permeabilized with 0.2% Triton X-100 in 3% BSA in PBS for 20 minutes. Primary antibodies used were mouse anti-PUBL at 1:200 (generated in this study), rabbit anti-CPN60 [14] at 1:2000, rat anti-HA at 1:400 (clone 3F10 Roche Applied Sciences), mouse anti-GFP at 1:500 (Torry Pines Biolab) and mouse anti-Ty-1 at 1:20 (a gift from Keith Gull, Oxford University). The secondary antibodies used were goat anti-mouse Alexa Fluor 488, goat anti-rat Alexa Fluor 488, goat anti-rabbit Alexa Fluor 488 and goat anti-rabbit Alexa Fluor 546 at

1:2000 (Invitrogen). Images were collected using an Applied Precision Delta Vision microscope. Images were deconvolved and adjusted for contrast using Softworx software.

Western Blot. *T. gondii* parasites were harvested (1 X 10^6), lysed in RIPA lysis buffer, boiled in 1x NuPAGE LDS Sample Buffer, loaded onto precast 10% and Any-KD Mini-PROTEAN TGX gels (Bio-Rad) and ran at 150 V [13]. Proteins were transferred to nitrocellulose membranes and probed with antibodies mouse anti-PUBL at 1:200 (generated in this study), rabbit anti-$CDC48_{AP}$ at 1:500 [14], mouse anti-tubulin at 1:2000 (12G10, a gift from Jacek Gaertig, University of Georgia), rat anti-HA at 1:400 (clone 3F10 Roche Applied Sciences), mouse anti-GFP at 1:500 (Torry Pines Biolab), rabbit anti-RFP at 1:1000 (Rockland Immunochemicals) followed by incubating the membranes with horseradish peroxidase-conjugated anti-mouse, anti-rat or anti-rabbit antibodies at a dilution of (1:10,000 dilution; Bio-Rad) respectively. Bands were detected by incubating the membrane with Pierce ECL Western Blotting Substrate and exposing the membrane to film.

qPCR. *T. gondii* genomic DNA was extracted and prepared from parasites treated with ATc or in the absence of ATc using QIAGEN's DNeasy Blood & Tissue Kit. 150 ng of the genomic DNA was used for the qPCR reaction. The qPCR was performed using SYBR Green Mix (Biorad) and primers UPRT-qPCR-F/R to amplify the nuclear genome and primers Apg-qPCR-F/R to amplify the apicoplast genome as previously described [11]. Copy number control was performed by making a standard curve for each qPCR based on the serial dilution of plasmids (10^8 copies - 10^3 copies) containing the UPRT locus or the apicoplast genome as previously described [11]. All reactions were performed in triplicate in a Bio-Rad iQ5 real-time PCR detection system. The copy

number of the apicoplast and nuclear DNA was normalized such that parasites grown in the absence of ATc is 1.

Phylogenetic Analysis & 3D Protein Modeling. A maximum likelihood phylogenetic tree was constructed using the software tools offered through Phylogeny.fr [67] and visualized using figtree. *T. gondii* ubiquitin was arbitrarily chosen to root the tree. An ab initio protein folding and structure predictions was generated using the Quark algorithm [68]. The resulting 3D protein model as well as the established human ubiquitin structure were visualized using UCSF Chimera [69]. Sequences were aligned using default T-Coffee settings and viewed through Jalview [70]. Analysis of the conservation between sequences was performed using JABAWS [71].

Antibody Development. The C-terminal ubiquitin-like domain was inserted into vector pAVA421 to encode a fusion protein in which the last 77 amino acids of PUBL are expressed with a 6x His tag at the N-terminus. The resulting PUBL expression plasmid was transformed into BL21 *E. coli* cells. Protein expression was induced with 1mM of IPTG at 37 C for 4 hours, cells were lysed and proteins were purified by affinity chromatography using NI+NTA resin as previously described [72]. Mice were injected with 100 μg of recombinant protein with incomplete Freund's adjuvant. Additional booster injections of 50 μg of recombinant protein with incomplete Freund's adjuvant were given every two weeks. After 8 weeks mice were sacrificed and B cells were fused with myeloma cells. Hybridomas were tested to identify clones that expressed antibody to PUBL.

Acknowledgments

We would like to thank Carrie Brooks for technical assistance, Julie Nelson for help with flow cytometry, Melissa Storey for assistance with immunization, Ruth Davis for monoclonal antibody production, and Liz Hedstrom for discussion.

Funding information

This work was supported by a grant from the National Institutes of Health (NIH) to B.S. (RO1 AI 64671), J.D.F. received a predoctoral fellowship from the American Heart Association, and B.S. is a Georgia Research Alliance Distinguished Investigator. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Tenter, A.M., A.R. Heckeroth, and L.M. Weiss, *Toxoplasma gondii: from animals to humans.* Int J Parasitol, 2000. **30**(12-13): p. 1217-58.
2. Weiss, L.M. and J.P. Dubey, *Toxoplasmosis: A history of clinical observations.* Int J Parasitol, 2009. **39**(8): p. 895-901.
3. Jones, J.L., et al., *Congenital toxoplasmosis: a review.* Obstet Gynecol Surv, 2001. **56**(5): p. 296-305.
4. van Dooren, G.G. and B. Striepen, *The Algal Past and Parasite Present of the Apicoplast.* Annu Rev Microbiol, 2013.
5. Foth, B.J., et al., *Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum.* Science, 2003. **299**(5607): p. 705-8.
6. van Dooren, G.G., et al., *Processing of an apicoplast leader sequence in Plasmodium falciparum and the identification of a putative leader cleavage enzyme.* J Biol Chem, 2002. **277**(26): p. 23612-9.
7. Tonkin, C.J., D.S. Roos, and G.I. McFadden, *N-terminal positively charged amino acids, but not their exact position, are important for apicoplast transit peptide fidelity in Toxoplasma gondii.* Mol Biochem Parasitol, 2006. **150**(2): p. 192-200.
8. Heiny, S.R., et al., *Protein Traffic to the Plasmodium falciparum apicoplast: evidence for a sorting branch point at the Golgi.* Traffic, 2014. **15**(12): p. 1290-304.

9. Bouchut, A., et al., *Vesicles bearing Toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption.* PLoS One, 2014. **9**(11): p. e112096.

10. DeRocher, A.E., et al., *A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in Toxoplasma gondii.* Eukaryot Cell, 2008. **7**(9): p. 1518-29.

11. Sheiner, L., et al., *Toxoplasma gondii Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins.* Traffic, 2015.

12. Glaser, S., et al., *Tic22 is an essential chaperone required for protein import into the apicoplast.* J Biol Chem, 2012. **287**(47): p. 39505-12.

13. van Dooren, G.G., et al., *Toxoplasma gondii Tic20 is essential for apicoplast protein import.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13574-9.

14. Agrawal, S., et al., *Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins.* J Biol Chem, 2009. **284**(48): p. 33683-91.

15. Smith, M.H., H.L. Ploegh, and J.S. Weissman, *Road to ruin: targeting proteins for degradation in the endoplasmic reticulum.* Science, 2011. **334**(6059): p. 1086-90.

16. Agrawal, S., et al., *An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins.* PLoS Pathog, 2013. **9**(6): p. e1003426.

17. Ye, Y., H.H. Meyer, and T.A. Rapoport, *The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol.* Nature, 2001. **414**(6864): p. 652-6.

18. Stork, S., et al., *Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation.* Eukaryot Cell, 2012. **11**(12): p. 1472-81.

19. Ye, Y., H.H. Meyer, and T.A. Rapoport, *Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains.* J Cell Biol, 2003. **162**(1): p. 71-84.

20. Flierman, D., et al., *Polyubiquitin serves as a recognition signal, rather than a ratcheting molecule, during retrotranslocation of proteins across the endoplasmic reticulum membrane.* J Biol Chem, 2003. **278**(37): p. 34774-82.

21. Ernst, R., et al., *Enzymatic blockade of the ubiquitin-proteasome pathway.* PLoS Biol, 2011. **8**(3): p. e1000605.

22. Stein, A., et al., *Key steps in ERAD of luminal ER proteins reconstituted with purified components.* Cell, 2014. **158**(6): p. 1375-88.

23. Baldridge, R.D. and T.A. Rapoport, *Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD.* Cell, 2016. **166**(2): p. 394-407.

24. Meyer, H., M. Bug, and S. Bremer, *Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system.* Nat Cell Biol, 2012. **14**(2): p. 117-23.

25. Vinayak, S., et al., *Genetic manipulation of the Toxoplasma gondii genome by fosmid recombineering.* MBio, 2014. **5**(6): p. e02021.

26. Kobayashi, T., et al., *Functional ATPase activity of p97/valosin-containing protein (VCP) is required for the quality control of endoplasmic reticulum in neuronally differentiated mammalian PC12 cells.* J Biol Chem, 2002. **277**(49): p. 47358-65.

27. Waller, R.F., et al., *Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12352-7.

28. Harb, O.S., et al., *Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite Toxoplasma gondii.* Eukaryot Cell, 2004. **3**(3): p. 663-74.

29. Sheiner, L., et al., *A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor.* PLoS Pathog, 2011. **7**(12): p. e1002392.

30. Park, S., et al., *Ufd1 exhibits the AAA-ATPase fold with two distinct ubiquitin interaction sites.* Structure, 2005. **13**(7): p. 995-1005.

31. Burroughs, A.M., L.M. Iyer, and L. Aravind, *The natural history of ubiquitin and ubiquitin-related domains.* Front Biosci (Landmark Ed), 2012. **17**: p. 1433-60.

32. Sommer, M.S., et al., *Der1-mediated preprotein import into the periplastid compartment of chromalveolates?* Mol Biol Evol, 2007. **24**(4): p. 918-28.

33. Ellison, M.J. and M. Hochstrasser, *Epitope-tagged ubiquitin. A new probe for analyzing ubiquitin function.* J Biol Chem, 1991. **266**(31): p. 21150-7.

34. Clague, M.J., et al., *Deubiquitylases from genes to organism.* Physiol Rev, 2013. **93**(3): p. 1289-315.

35. Striepen, B., et al., *The plastid of Toxoplasma gondii is divided by association with the centrosomes.* J Cell Biol, 2000. **151**(7): p. 1423-34.

36. Komander, D., *The emerging complexity of protein ubiquitination.* Biochem Soc Trans, 2009. **37**(Pt 5): p. 937-53.

37. Lauwers, E., C. Jacob, and B. Andre, *K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway.* J Cell Biol, 2009. **185**(3): p. 493-502.

38. Vierstra, R.D., *The expanding universe of ubiquitin and ubiquitin-like modifiers.* Plant Physiol, 2012. **160**(1): p. 2-14.

39. Dorval, V. and P.E. Fraser, *Small ubiquitin-like modifier (SUMO) modification of natively unfolded proteins tau and alpha-synuclein.* J Biol Chem, 2006. **281**(15): p. 9919-24.

40. Yeh, H.M., et al., *Ubiquitin-specific protease 13 regulates IFN signaling by stabilizing STAT1*. J Immunol, 2013. **191**(6): p. 3328-36.

41. Kleine, T., U.G. Maier, and D. Leister, *DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis.* Annu Rev Plant Biol, 2009. **60**: p. 115-38.

42. Paila, Y.D., L.G. Richardson, and D.J. Schnell, *New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development.* J Mol Biol, 2015. **427**(5): p. 1038-60.

43. Straub, S.P., et al., *Dynamic organization of the mitochondrial protein import machinery.* Biol Chem, 2016.

44. Petersen, J., et al., *Chromera velia, Endosymbioses and the Rhodoplex Hypothesis-Plastid Evolution in Cryptophytes, Alveolates, Stramenopiles, and Haptophytes (CASH Lineages).* Genome Biology and Evolution, 2014. **6**(3): p. 666-684.

45. Gould, S.B., U.G. Maier, and W.F. Martin, *Protein Import and the Origin of Red Complex Plastids.* Current Biology, 2015. **25**(12): p. R515-R521.

46. Leveque, M.F., et al., *Autophagy-Related Protein ATG8 Has a Noncanonical Function for Apicoplast Inheritance in Toxoplasma gondii.* MBio, 2015. **6**(6): p. e01446-15.

47. Daher, W., et al., *Lipid kinases are essential for apicoplast homeostasis in Toxoplasma gondii.* Cell Microbiol, 2015. **17**(4): p. 559-78.

48. Voss, C., et al., *Overexpression of Plasmodium berghei ATG8 by Liver Forms Leads to Cumulative Defects in Organelle Dynamics and to Generation of Noninfectious Merozoites.* MBio, 2016. **7**(3).

49. Hempel, F., G. Felsner, and U.G. Maier, *New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms.* Mol Microbiol, 2010. **76**(3): p. 793-801.

50. Felsner, G., et al., *ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane.* Genome Biol Evol, 2011. **3**: p. 140-50.

51. Spork, S., et al., *An unusual ERAD-like complex is targeted to the apicoplast of Plasmodium falciparum.* Eukaryot Cell, 2009. **8**(8): p. 1134-45.

52. Kalanon, M., C.J. Tonkin, and G.I. McFadden, *Characterization of two putative protein translocation components in the apicoplast of Plasmodium falciparum.* Eukaryot Cell, 2009. **8**(8): p. 1146-54.

53. Hirakawa, Y., F. Burki, and P.J. Keeling, *Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga.* Eukaryot Cell, 2012. **11**(3): p. 324-33.

54. van der Veen, A.G. and H.L. Ploegh, *Ubiquitin-like proteins.* Annu Rev Biochem, 2012. **81**: p. 323-57.

55. Spork, S., et al., *An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum.* Eukaryotic Cell, 2009. **8**(8): p. 1134-1145.

56. Maier, U.G., S. Zauner, and F. Hempel, *Protein import into complex plastids: Cellular organization of higher complexity.* Eur J Cell Biol, 2015. **94**(7-9): p. 340-8.

57. Dhara, A. and A.P. Sinai, *A Cell Cycle-Regulated Toxoplasma Deubiquitinase, TgOTUD3A, Targets Polyubiquitins with Specific Lysine Linkages.* mSphere, 2016. **1**(3).

58. Lau, J.B., et al., *N-terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids.* Mol Microbiol, 2015. **96**(3): p. 609-20.

59. Lau, J.B., et al., *Protein-protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids.* Mol Microbiol, 2016. **100**(1): p. 76-89.

60. de Monerri, N.C.S., et al., *The Ubiquitin Proteome of Toxoplasma gondii Reveals Roles for Protein Ubiquitination in Cell-Cycle Transitions.* Cell Host & Microbe, 2015. **18**(5): p. 621-633.

61. Maier, U.G., S. Zauner, and F. Hempel, *Protein import into complex plastids: Cellular organization of higher complexity.* European Journal of Cell Biology, 2015. **94**(7-9): p. 340-348.

62. Sanyal, S., J.H. Claessen, and H.L. Ploegh, *A viral deubiquitylating enzyme restores dislocation of substrates from the endoplasmic reticulum (ER) in semi-intact cells.* J Biol Chem, 2012. **287**(28): p. 23594-603.

63. Nakamura, N., et al., *Ubiquitin-specific protease 19 regulates the stability of the E3 ubiquitin ligase MARCH6.* Exp Cell Res, 2014. **328**(1): p. 207-16.

64. Edelmann, M.J., B. Nicholson, and B.M. Kessler, *Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases.* Expert Rev Mol Med, 2011. **13**: p. e35.

65. Eaton, M.S., L.M. Weiss, and K. Kim, *Cyclic nucleotide kinases and tachyzoite-bradyzoite transition in Toxoplasma gondii.* Int J Parasitol, 2006. **36**(1): p. 107-14.

66. Sidik, S.M., et al., *Efficient genome engineering of Toxoplasma gondii using CRISPR/Cas9.* PLoS One, 2014. **9**(6): p. e100450.

67. Dereeper, A., et al., *Phylogeny.fr: robust phylogenetic analysis for the non-specialist.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W465-9.

68. Xu, D. and Y. Zhang, *Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field.* Proteins, 2012. **80**(7): p. 1715-35.

69. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

70. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.

71. Troshin, P.V., J.B. Procter, and G.J. Barton, *Java bioinformatics analysis web services for multiple sequence alignment--JABAWS:MSA.* Bioinformatics, 2011. **27**(14): p. 2001-2.

72. Francia, M.E., et al., *Cell division in Apicomplexan parasites is organized by a homolog of the striated rootlet fiber of algal flagella.* PLoS Biol, 2012. **10**(12): p. e1001444.

Figures

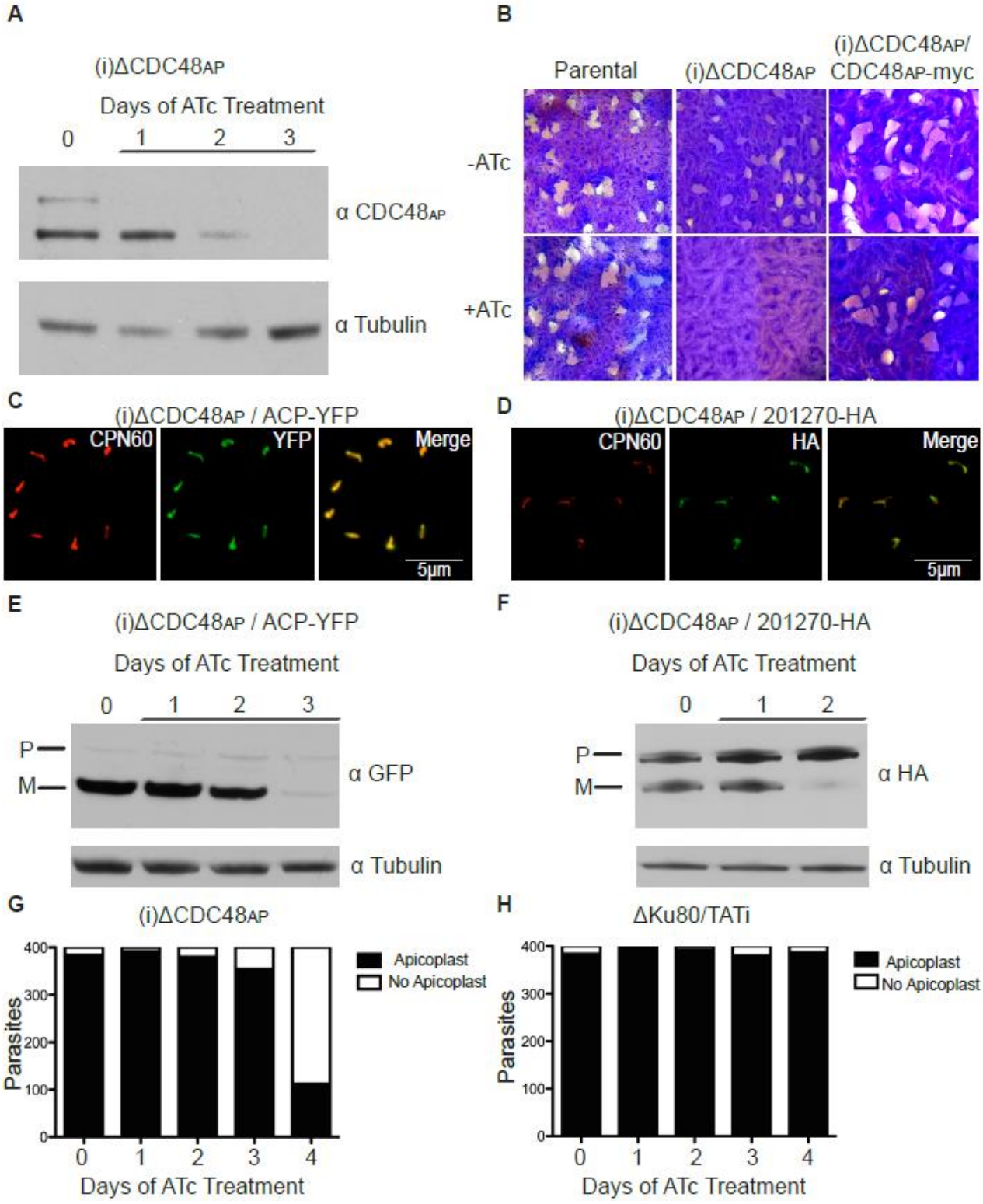

A
(i)ΔCDC48AP
Days of ATc Treatment
0 1 2 3
α CDC48AP
α Tubulin
B
Parental
(i)ΔCDC48AP
(i)ΔCDC48AP/ CDC48AP-myc
-ATc
+ATc
C
(i)ΔCDC48AP / ACP-YFP
CPN60
YFP
Merge
5μm
D
(i)ΔCDC48AP / 201270-HA
CPN60
HA
Merge
5μm
E
(i)ΔCDC48AP / ACP-YFP
Days of ATc Treatment
0 1 2 3
P
M
α GFP
α Tubulin
F
(i)ΔCDC48AP / 201270-HA
Days of ATc Treatment
0 1 2
P
M
α HA
α Tubulin
G
(i)ΔCDC48AP
Parasites
400
300
200
100
0
0 1 2 3 4
Days of ATc Treatment
Apicoplast
No Apicoplast
H
ΔKu80/TATi
Parasites
400
300
200
100
0
0 1 2 3 4
Days of ATc Treatment
Apicoplast
No Apicoplast

Figure 3.1. CDC48$_{AP}$ is critical for parasite survival and import across the PPM (A) Western blot analysis using CDC48$_{AP}$ antibody on the (i)ΔCDC48$_{AP}$ line after ATc treatment for the indicated times. Levels of CDC48$_{AP}$ are diminished after two days of ATc treatment and completely ablated after four days of ATc treatment. (B) Plaque assays were performed on the parental ΔKu80/TATi line, (i)ΔCDC48$_{AP,}$ and the complemented (i)ΔCDC48$_{AP}$ line in the absence or presence of 0.5 μg/ml of ATc. Note lack of plaque formation for (i)ΔCDC48$_{AP}$ under ATc. (C, D) Immunofluorescence assays performed on stable (i)ΔCDC48$_{AP}$ lines expressing ACP-YFP (C) or 201270-HA (D). CPN60 (red) serves as marker for the apicoplast. ACP-YFP and 201270-HA (both green) are properly localized to the apicoplast and show patterns of overlap with CPN60 typical for *T. gondii* luminal or PPC apicoplast proteins [14, 29]. (E-F) Apicoplast import assays were performed with the (i)ΔCDC48$_{AP}$/ACP-YFP (E) line and the (i)ΔCDC48$_{AP}$/ 201270-HA (F) line respectively. Parasites were treated with ATc for the time indicated and harvested for Western blot analysis using YFP or HA antibody respectively. Note the loss of the mature band (M) for both reporters while the levels of precursor (P) levels remain unchanged under ATc treatment. (G) The presence of apicoplasts was scored daily by IFA using anti-CPN60 for 400 ATc treated (i)ΔCDC48$_{AP}$ parasites. (H) The same assay was performed on the parental ΔKu80/TATi lines. Note that there is no significant difference in apicoplast numbers after ATc treatment. αTubulin serves as loading control in (A, E and F).

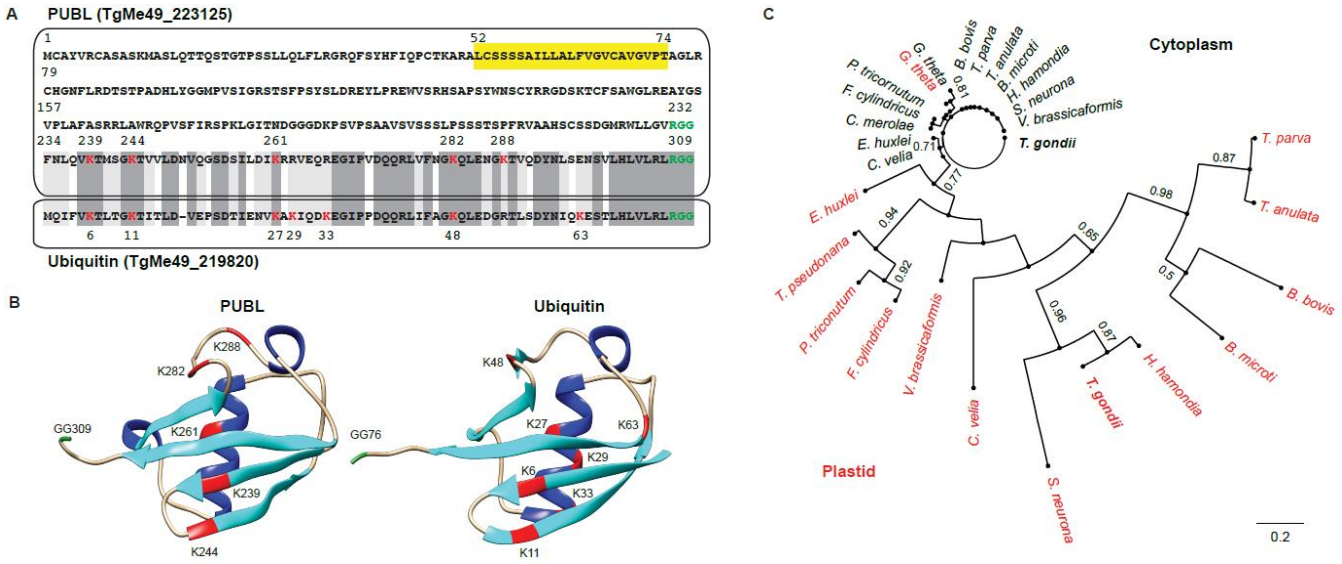
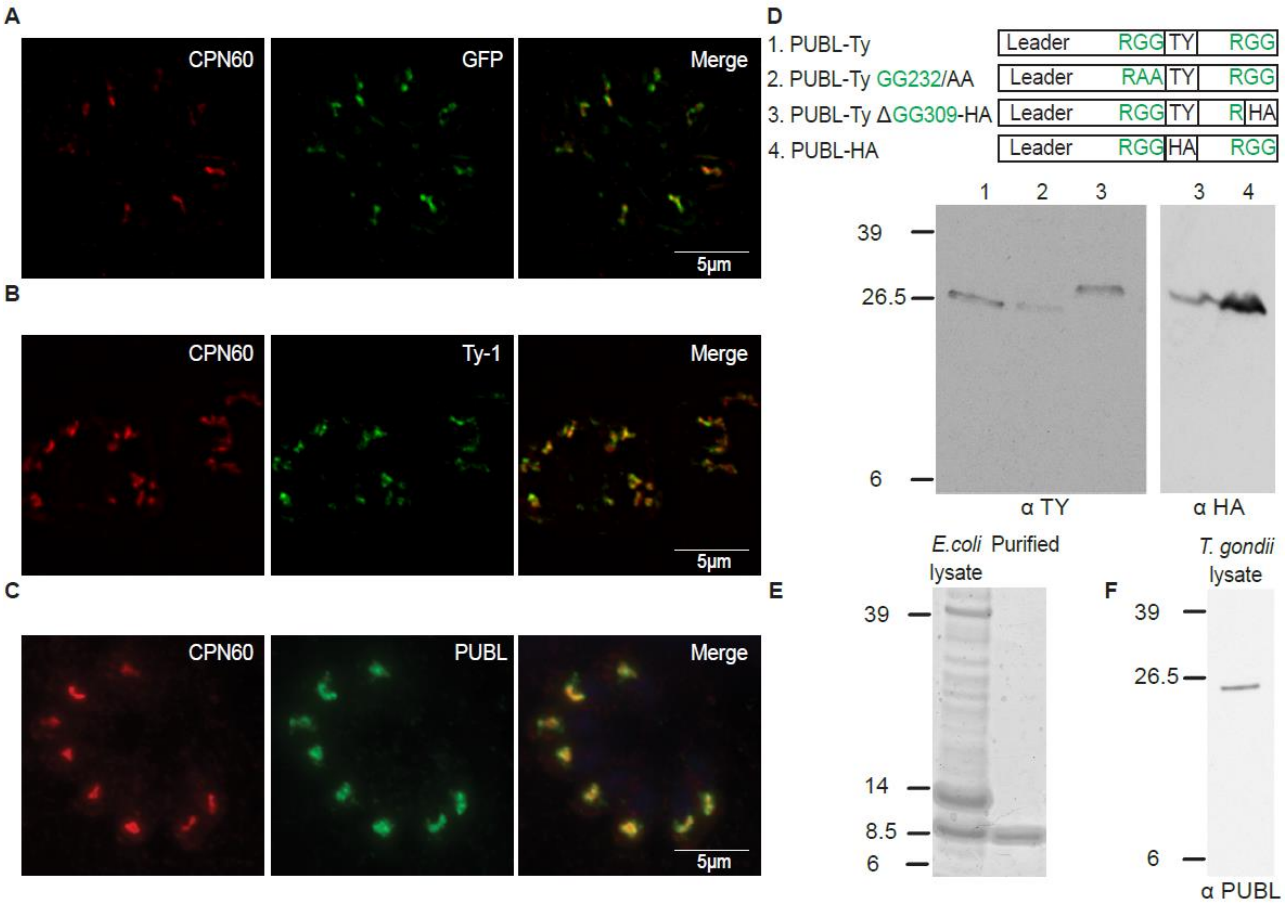


Figure 3.2. The C-terminal domain of PUBL resembles ubiquitin. (A) Sequence alignment of *T. gondii* PUBL and ubiquitin. A sequence that could serve as transmembrane domain or recessed signal peptide is highlighted in yellow, the lysine residues of the ubiquitin domain in red and potential deubiquitinase cleavage RGG motifs are shown in green. In the lower box is *T. gondii* cytoplasmic ubiquitin. Amino acids in the alignment that show identity are boxed in dark grey while light grey indicates conservative substitutions. (B) Ribbon diagram representation of the structure of the C-terminal domain of PUBL predicted *de novo* using Quark compared to the experimentally established structure of human ubiquitin (MMDB ID: 57540PDB, ID: 1UBQ). Beta sheets are colored cyan, alpha helices blue, lysines red and the diglycine motif green. (C) Maximum likelihood tree depicting the phylogenetic relationship of PUBL and selected plastid (red) and cytoplasmic (black) homologs, bootstrap values for 100 replicates are shown (Apicomplexans: *B. microti*, BBM_I02580, BBM_III01010; *B. bovis*, BBOV_III010050, BBOV_1V010030; *C. velia*, CVEL_26518, CVEL 26884; *H. hammondi*, HHA_223125, HHA_289750; *S. neurona*, SRCN_6530, SRCN_6527; *T.*

annulata, TA11575, TA16165; *T. gondii* TgME49_223125, TgME49_219820; *T. parva* TP02_0142, TP01_1070; *V. brassicaformis*, Vbra_4789, Vbra_15758. Cryptophytes: *G. theta*, 155024, 152873. Diatoms: *F. cylindrus*, 270635, 2686161 *P. tricornutum*, 54323, 51931; *T. pseudonana*, 1536, 259049. Haptophytes: *E. huxleyi*, 428400, 349903. Rhodophyta: *C. merolae* CMK296C).

Figure 3.3. PUBL is an apicoplast specific ubiquitin-like protein. (A)



Immunofluorescence assays depicting parasites expressing the N-terminal 180 amino acids of PUBL fused to GFP (anti-GFP, green), (B) a full length version of PUBL with an internal Ty-1 epitope tag (Api-Ty1 as shown in Fig. 3.3D-1, anti-Ty1, green) and (C) ΔKu80/TATi parental parasites stained with a monoclonal antibody raised against the ubiquitin-like domain of PUBL (green). Counterstaining for CPN60 is shown in red. (D)

The first panel depicts four constructs transfected into parasite lines to express tagged and/or mutated versions of PUBL (1. PUBL-Ty, 2. PUBL-Ty GG232/AA, 3. PUBL-TyΔGG309-HA and 4. PUBL-HA). Western blot analysis of the protein lysates prepared from the four different lines as numerically indicated in the first panel. All versions of PUBL are predicted to be 26.5 Kda except for construct 3 which is predicted to be 29.5 KDa due the extra 3x HA tag at the C terminus. (E) Coomassie stained protein gel of protein extract and purified protein derived from *E. coli* cell expressing a recombinant version of the c-terminal ubiquitin-like domain of PUBL. 8.5 kDa band is visible and matches the expected size of recombinant PUBL carrying a 6xHis tag. (F) Western blot analysis of protein extracts of ΔKu80/TATi parasites stained with monoclonal antibody obtained through immunization with recombinant protein shown in (E). Note single bands in (D) and (F) that are considerably larger than the 8.5 KDa protein shown in (E), slight differences in apparent molecular weight are due to epitope tags.

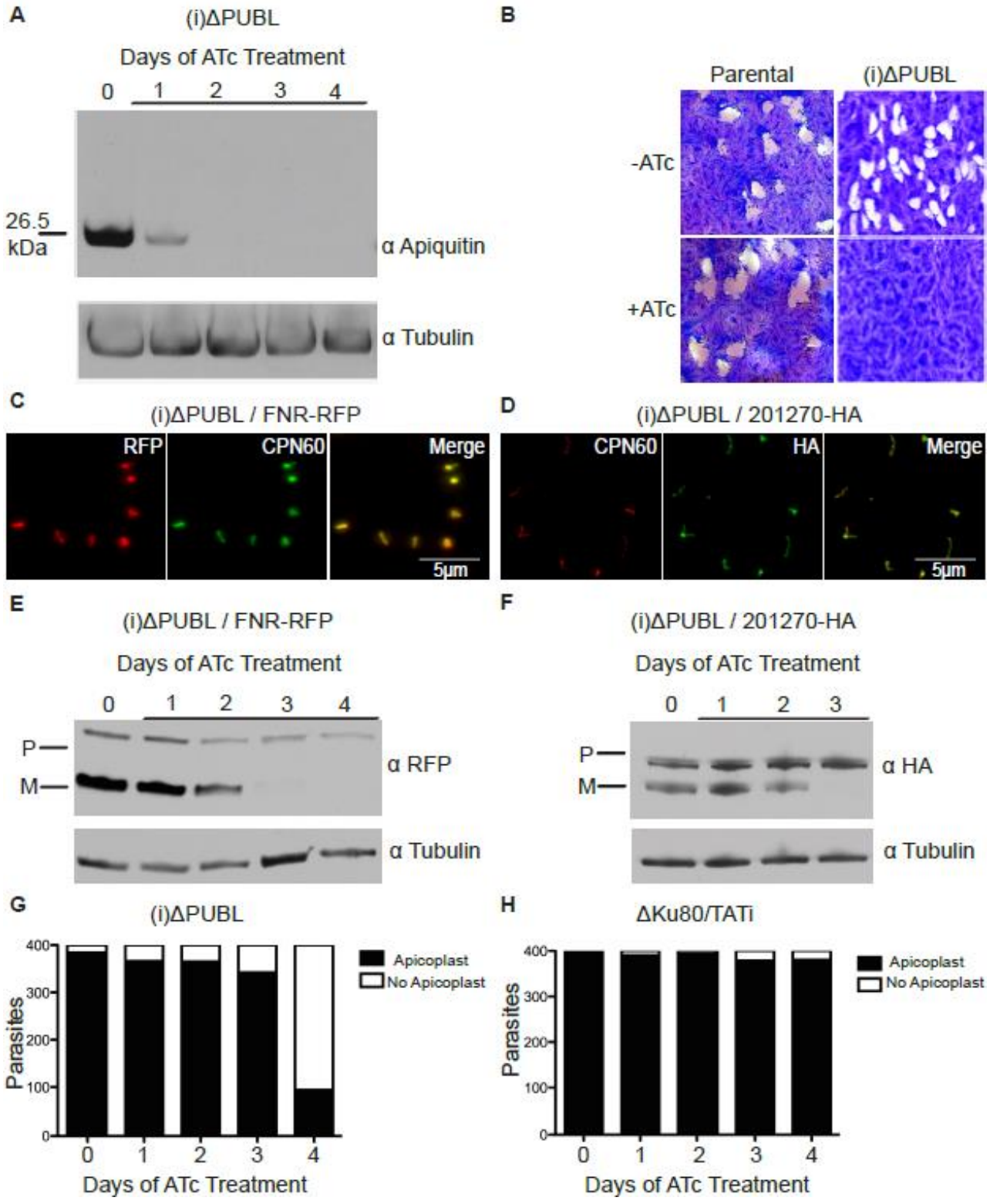

A
(i)ΔPUBL
Days of ATc Treatment
0 1 2 3 4
26.5 kDa
α Apiquitin
α Tubulin
B
Parental
(i)ΔPUBL
-ATc
+ATc
C
(i)ΔPUBL / FNR-RFP
RFP
CPN60
Merge
5μm
D
(i)ΔPUBL / 201270-HA
CPN60
HA
Merge
5μm
E
(i)ΔPUBL / FNR-RFP
Days of ATc Treatment
0 1 2 3 4
P
M
α RFP
α Tubulin
F
(i)ΔPUBL / 201270-HA
Days of ATc Treatment
0 1 2 3
P
M
α HA
α Tubulin
G
(i)ΔPUBL
Parasites
Apicoplast
No Apicoplast
Days of ATc Treatment
H
ΔKu80/TATi
Parasites
Apicoplast
No Apicoplast
Days of ATc Treatment

Figure 3.4. Loss of PUBL leads to loss of protein import across the PPM and block of parasite growth. (A) Western blot analysis of (i)ΔPUBL parasite grown under ATc for the indicated time. (B) Plaque assay measuring growth of the (i)ΔPUBL line in the presence and absence of ATc. (C-D) Immunofluorescence assay performed on stable (i)ΔPUBL lines expressing FNR-RFP (red) and 201270-HA (green) respectively. CPN60 is shown as apicoplast marker. (E-F) Apicoplast import assays were performed on the (i)ΔPUBL/FNR-RFP (E) and (i)ΔPUBL/201270-HA (F). Note loss of the mature band (M) under ATc treatment while the precursor band (P) remains unchanged. (G) Presence of apicoplast was scored by IFA for indicated times of ATc treatment as detailed in Fig. 3.1 for the $CDC48_{AP}$ mutant. (H) The same assay was performed on the parental ΔKu80/TATi lines. Note that there is no significant difference in apicoplast numbers after ATc treatment. αTubulin serves as loading control in (A, E and F).

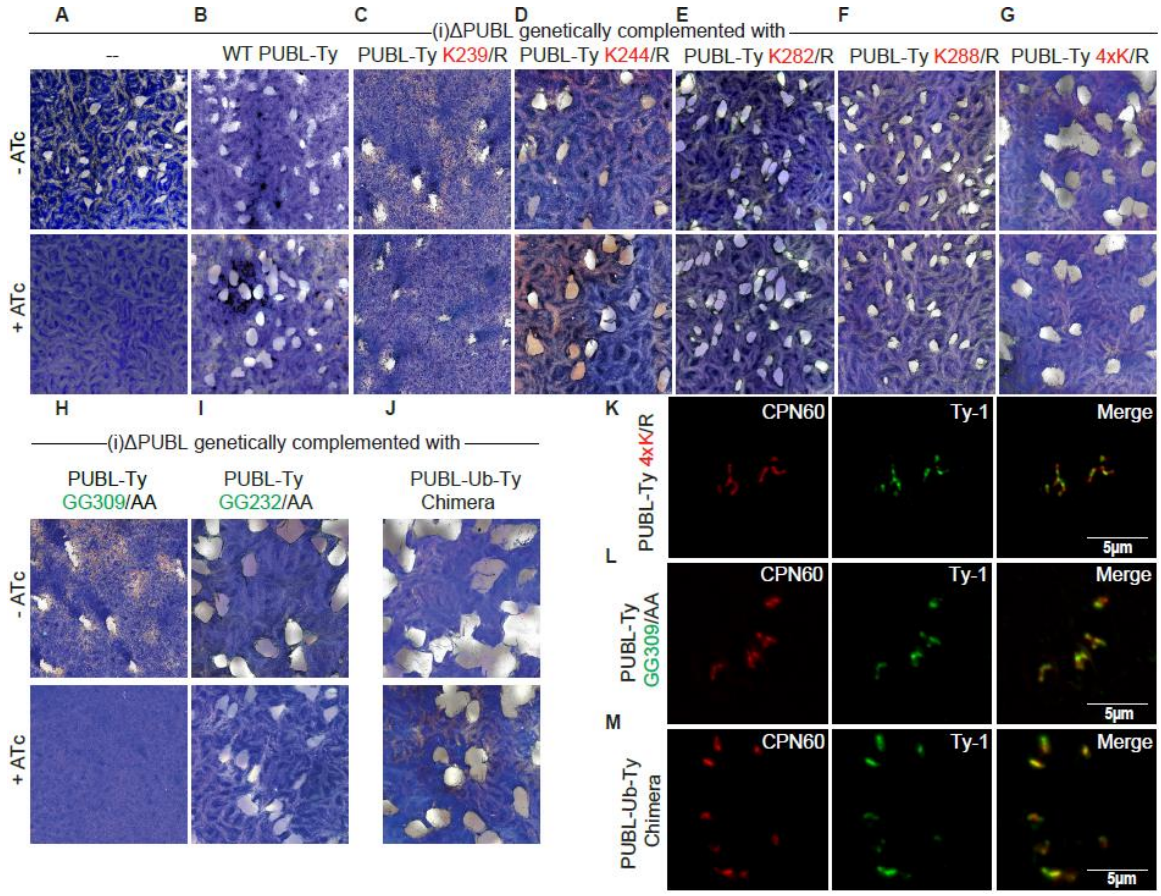


Figure 3.5. Genetic complementation analysis reveals essential PUBL residues.

Wildtype as well as a range of point-mutations of the PUBL coding sequence (all marked with an internal Ty tag) were introduced into the UPRT locus of the (i)ΔPUBL line. (A-I) Plaque assays were performed in the absence or presence of ATc to test the ability of each mutant to complement the loss of PUBL expression from the native locus. Lysine was replaced with asparagine, glycine with alanine, mutant 4xK was mutated at all four previously indicated positions. (J) Complementation experiment as describe above using a chimera that encodes a chimera in which the c-terminal domain of PUBL is replaced with the *T. gondii* ubiquitin sequence. (K-M) Immunofluorescence assays of complemented mutant (additional data shown in supplementary Fig. 3.5. Note that all

transgenes encode proteins that show localization indistinguishable from the wildtype protein. CPN60 is used as apicoplast marker for counterstain.

Supplemental Material

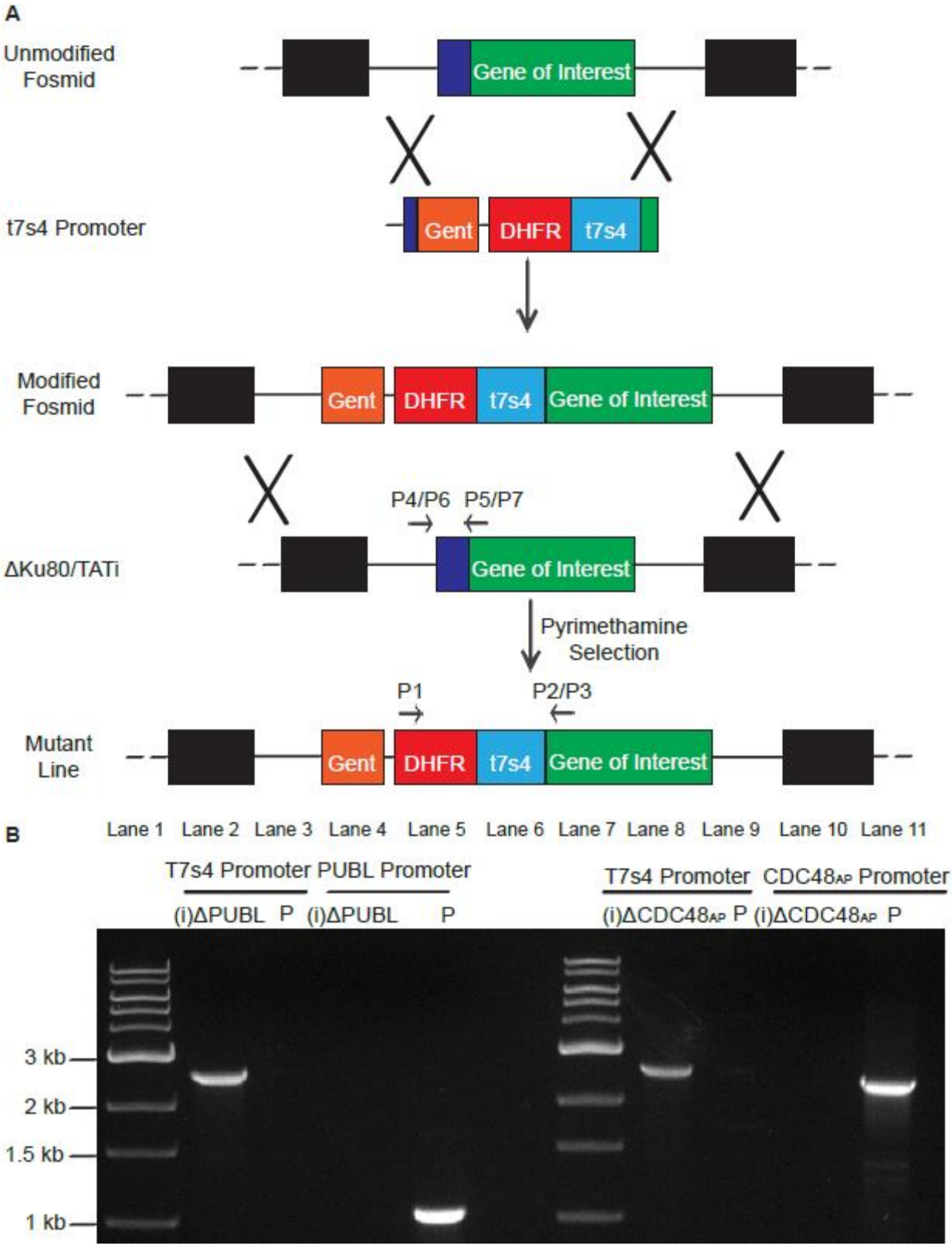


Figure S3.1. Construction of conditional mutants (A) Graphical representation of the fosmid recombineering strategy used to construct conditional mutants. The bacterial selection marker gentamycin (Gent) is shown in orange, the *T. gondii* selection marker dihydrofolate reductase (DHFR) in red, and the conditional tetracycline promoter (t7s4) in light blue. (B) Diagnostic PCR to confirm the correct replacement of the endogenous PUBL and $CDC48_{AP}$ promoter with the t7s4 promoter. (i)ΔPUBL and the parental ΔKu80/TATi line (P) were used in lanes 2 and 3 respectively with a primer set that amplifies 2.6 KB of the t7s4 promoter (P1 and P2). Lanes 4 and 5 used primers that amplify 1 KB of the native PUBL promoter (P4 and P5). (i)Δ$CDC48_{AP}$ and ΔKu80/TATi (P) were used in lanes 8 and 9 respectively with a primer set that amplifies 2.6 KB of the t7s4 promoter (P1 and P3). Lanes 10 and 11 were used with primers designed to amplify 2 KB of the $CDC48_{AP}$ promoter (P6 and P7). We observed a PCR product of the correct size for the (i)ΔPUBL and (i)Δ$CDC48_{AP}$ lines with the t7s4 primers but witnessed no PCR product for the endogenous promoter which suggests that the mutant promoters are correctly disrupted. All primers used are in table S3.1.

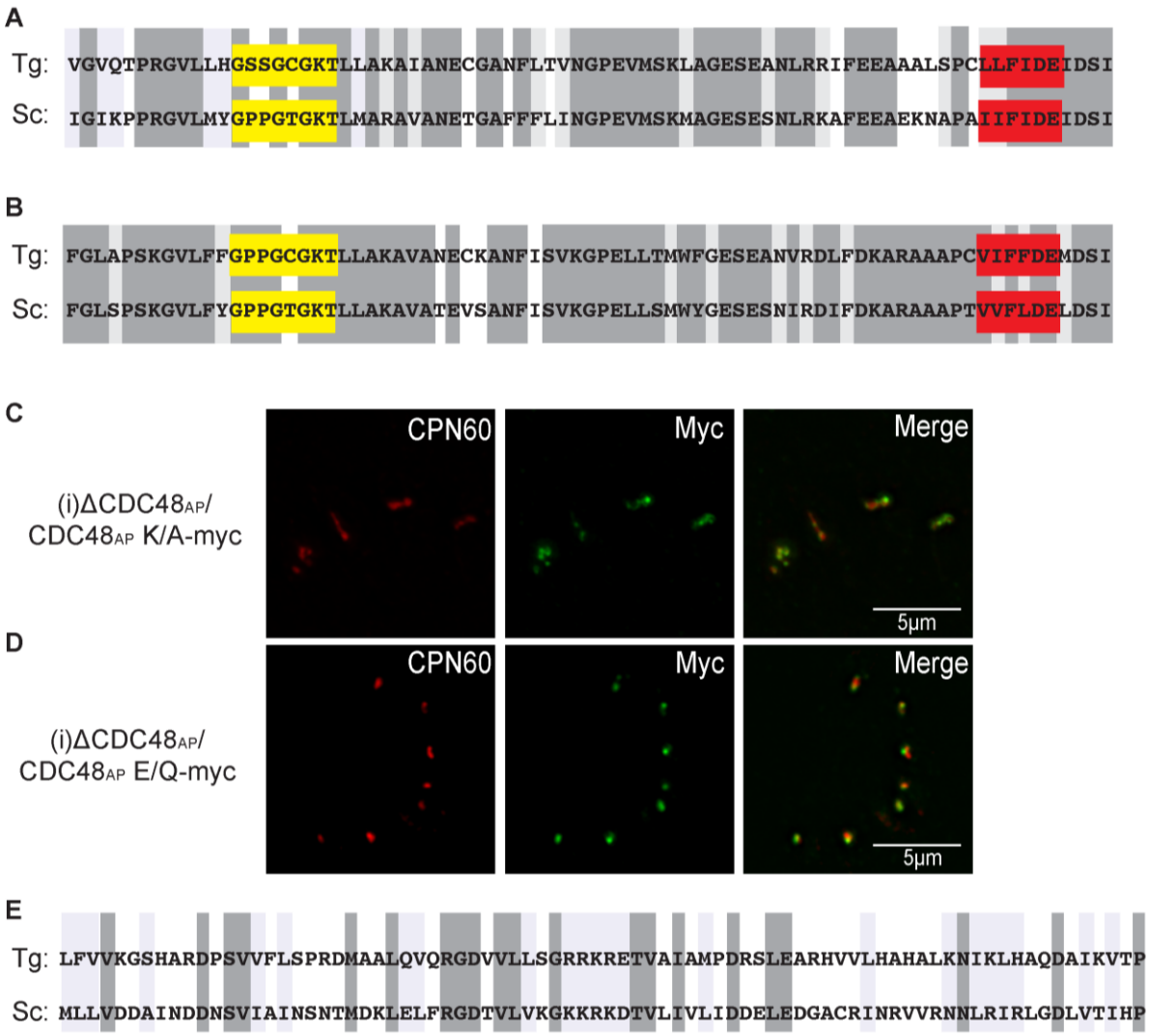


Figure S3.2. CDC48$_{AP}$ D1 and D2 ATPase domains and N-terminal domain.

Alignment of *Toxoplasma gondii* CDC48$_{AP}$ with CDC48 of *Saccharomyces cerevisiae*. Domains were identified using blastp algorithm from the National Center for Biotechnology Information. Domains were then aligned using Clustal Omega. Amino acids in the alignment that show identity are boxed in dark grey while light grey indicates conservative substitutions. (A-B) Partial alignment of the D1 and D2 ATPase domains between CDC48$_{AP}$ and *S. cerevisiae* CDC48 respectively. Highlighted in yellow are the Walker A motifs and highlighted in red are the Walker B motifs. (C-D) Immunofluorescence assay of ΔCDC48$_{AP}$ line expressing an ectopic myc tagged version of CDC48$_{AP}$ (green) with point mutations to the Walker A (ΔCDC48$_{AP}$ K/A-myc) and

Walker B (ΔCDC48$_{AP}$ E/Q-myc) motif. In red is the apicoplast luminal marker Chaperonin 60 (CPN60). Images demonstrate that the ectopic CDC48$_{AP}$ with point mutations localize to the apicoplast. (E) Alignment of the N-terminal domain of CDC48$_{AP}$ and *S. cerevisiae* CDC48 which is a known ubiquitin and chaperone binding site.

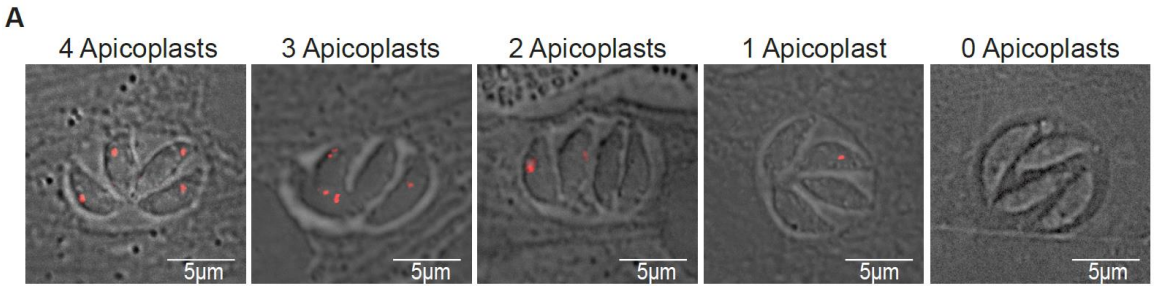


Figure S3.3. Measurement of apicoplast loss through immunofluorescence. (A) Representative images of ΔCDC48$_{AP}$ with 4 to 0 apicoplasts after being treated with ATc for 3 days. CPN60 in red was used as the apicoplast marker.

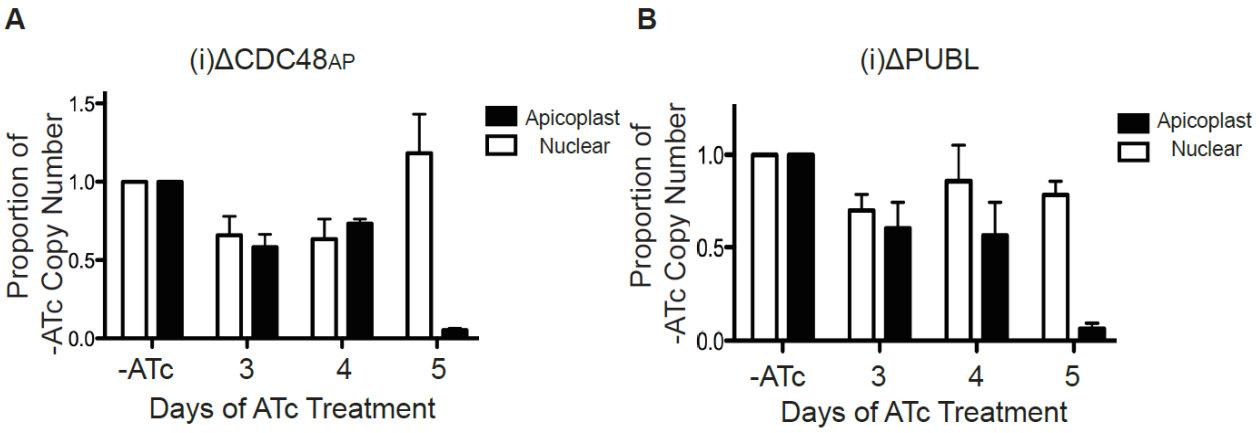


Figure S3.4. Quantification of apicoplast loss through qPCR. (A) The number of apicoplasts was quantified for the ΔCDC48$_{AP}$ line by comparing nuclear and apicoplast genome copy number through qPCR. After 5 days of ATc treatment the average proportion of apicoplast copy number is 0.05 while the average proportion of nuclear genomic copy number average is 1.18 compared to untreated parasites. (B) A qPCR was similarly performed on the ΔPUBL line. After 5 days of ATc treatment the average

proportion of apicoplast copy number is 0.05 while the average proportion of nuclear genomic copy number average is 0.78 compared to untreated parasites.

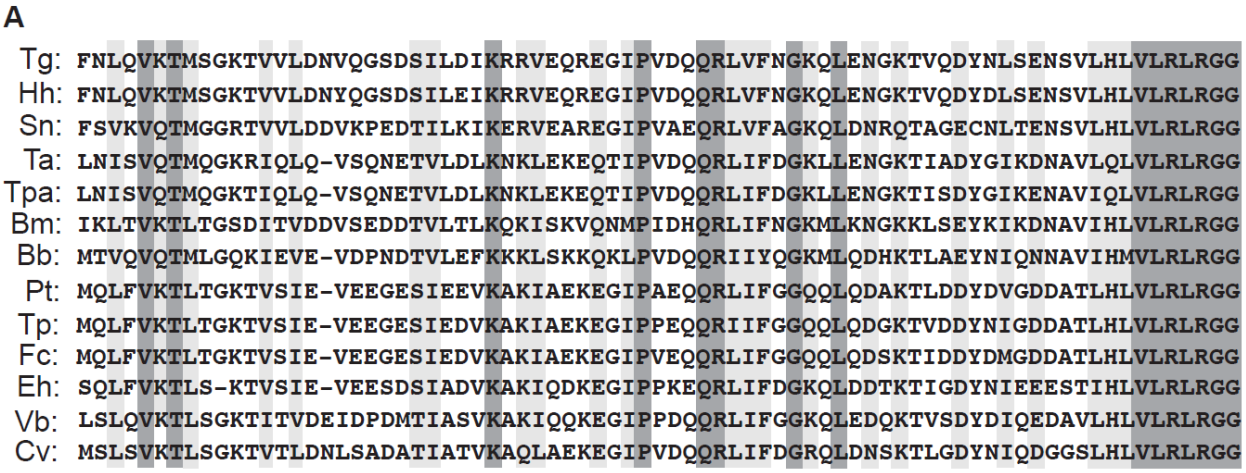
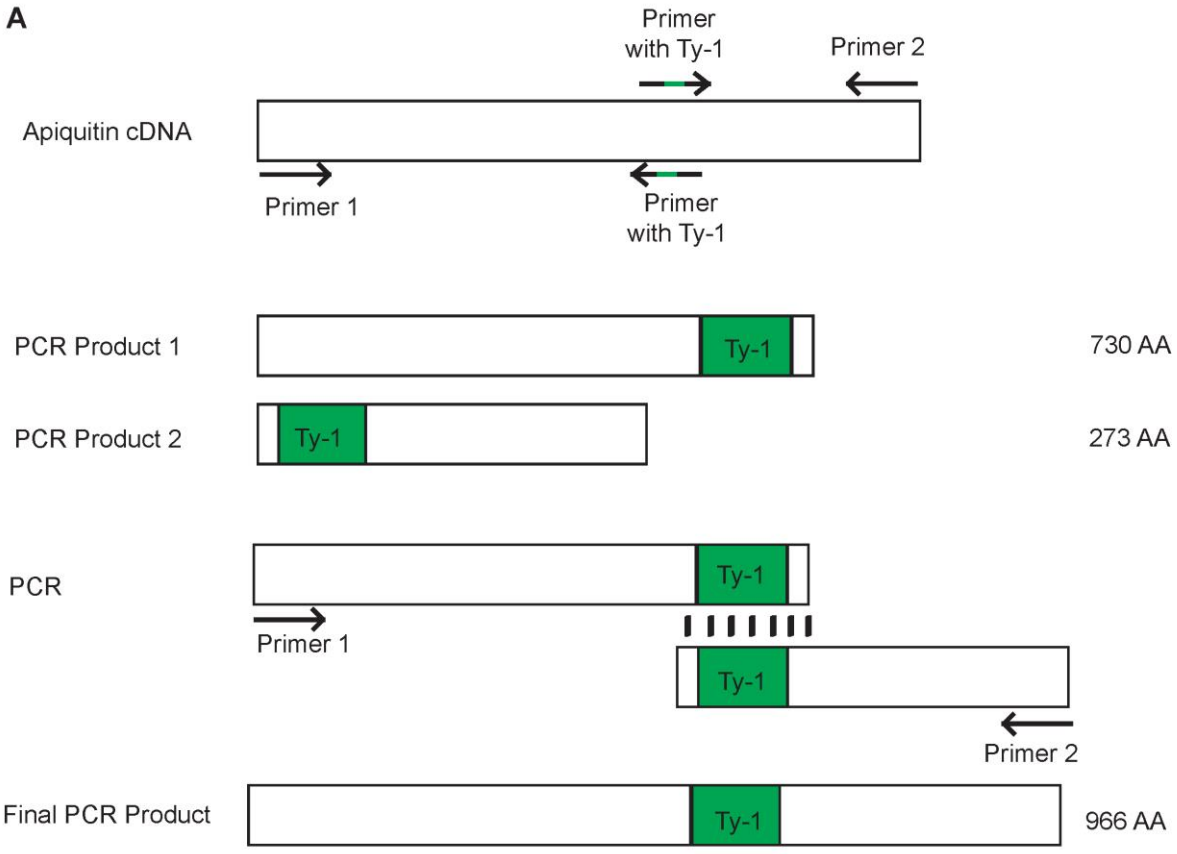


Figure S3.5. Alignment of ubiquitin domain of PUBL homologs identified in red algal derived plastid containing organisms. (A) Alignment of multiple ubiquitin domains of putative PUBL proteins in apicomplexan parasites and diatoms. Homologs were identified by a BLAST search using *T. gondii* PUBL's amino acid sequence. Species used in alignment *Toxoplasma gondii* (Tg), *Hammondia hammondia* (Hh), *Sarcocystis neurona* (Sn), *Theileria annulata* (Ta), *Theileria parvum* (Tpa), *Babesia microti* (Bm), *Babesia bovis* (Bb), *Phaeodactylum tricornutum* (Pt), *Thalassiosira pseudonana* (Tp), *Fragilariopsis cylindrus* (Fc), *Emiliania huxleyi* (Eh), *Vitrella brassicaformis* (Vb) and *Chromera velia* (Cv). Highlights as detailed in Supplementary Fig 2. Gene IDs of putative PUBL homologs used in alignment: BBOV_III010050, SRCN_6530, HHA_223125, BBM_I02580, TA11575, TP02_0142, 54323 for *P. tricornutum*, 270635 for *F. cylindrus*, 428400 for *E. huxleyi*, 1536 for *T. pseudonana*, CVEL_26518 and Vbra_4789. Note that all genes have a putative leader sequence and additional amino acids following the C-terminal "RGG" motif.

A

Figure S3.6. Full sequence alignment of PUBL homologs. The same species are used in this alignment as supplementary figure 5 except the N-terminal leader region is included in the alignment. No sequence conservation is seen between the homologs N-terminal regions. “RGG” motif is colored in green. Note that most sequences have this “RGG” motif immediately before the C-terminal ubiquitin domain.


A
Primer with Ty-1
Primer 2
Apiquitin cDNA
Primer 1
Primer with Ty-1
PCR Product 1
Ty-1
730 AA
PCR Product 2
Ty-1
273 AA
PCR
Ty-1
Primer 1
Ty-1
Primer 2
Final PCR Product
Ty-1
966 AA

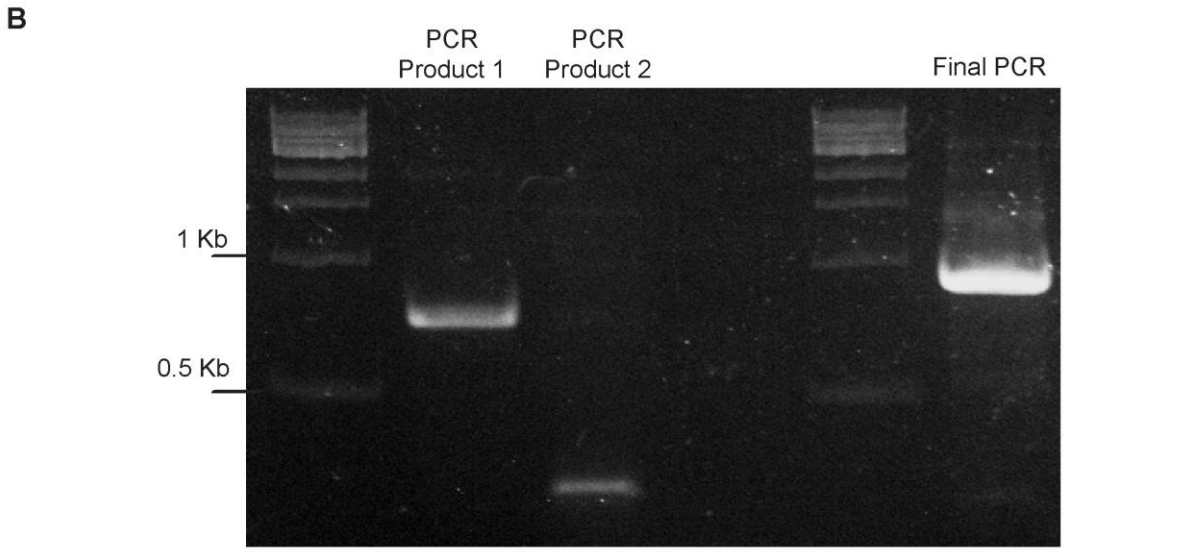

B
PCR Product 1
PCR Product 2
Final PCR
1 Kb
0.5 Kb

Figure S3.7. Construction of internal Ty-1 tagged PUBL. (A) Diagram of the procedure used for internally tagging the PUBL coding sequence. Two initial PCR reactions were implemented to amplify the 5' end of PUBL (730 amino acids) and the 3' end (273 amino acids). The PCR products were used as a template for a third PCR reaction which results in the final internal Ty-1 tagged PUBL gene. (B) Agarose gel of PCR products. Lane 2 is the amplification of the 5' end of PUBL. Lane 3 is the amplification of the 3' end of PUBL. Lane 6 is the final PCR amplification which was inserted into a pDC vector.

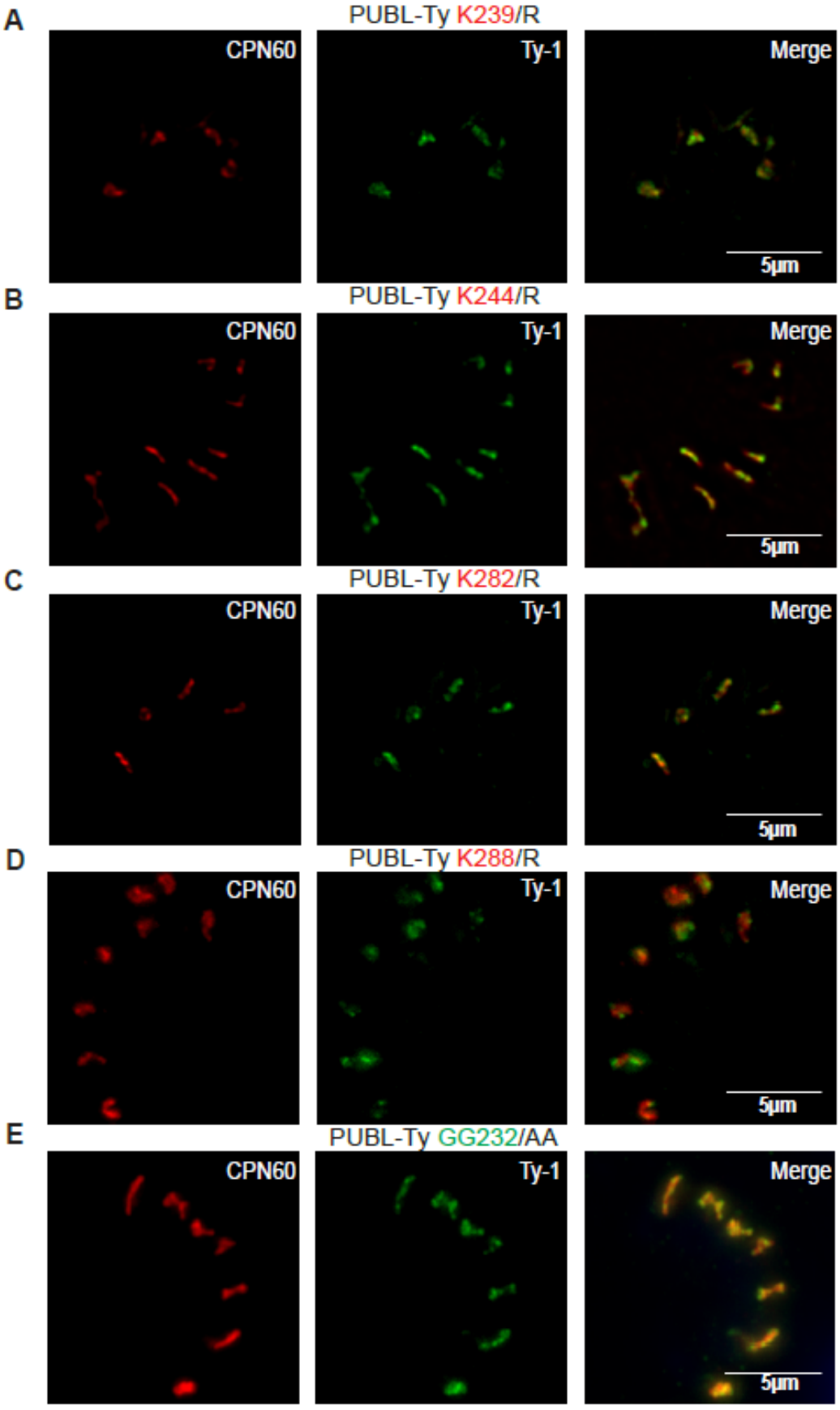

A
PUBL-Ty K239/R
CPN60
Ty-1
Merge
5µm
B
PUBL-Ty K244/R
CPN60
Ty-1
Merge
5µm
C
PUBL-Ty K282/R
CPN60
Ty-1
Merge
5µm
D
PUBL-Ty K288/R
CPN60
Ty-1
Merge
5µm
E
PUBL-Ty GG232/AA
CPN60
Ty-1
Merge
5µm

Figure S3.8. Immunofluorescence assay on (i)ΔPUBL expressing ectopic mutated versions of epitope tagged PUBL. (A-E) Immunofluorescence assays were performed on the stable complemented (i)ΔPUBL lines PUBL-Ty K239/R, PUBL-Ty K244/R, PUBL-Ty K282/R, PUBL-Ty K288/R and PUBL-Ty GG232/AA respectively. All ectopic versions of PUBL were tagged internally with the Ty-1 epitope (green) while CPN60 (red) was used as apicoplast marker. Correct localization of the ectopic PUBL in the apicoplast was observed for all mutants.

Supplementary Tables

Table S3.1. Primer list

Gene Used	Purpose	Name	Primer Sequence
Apiquitin	Apiquitin-Ty pDC	ApiqEcoRVR	agtgatatcctatctagaaccaccacca
Apiquitin	Apiquitin-Ty pDC	Apiq5TYR	ctgcatatctagtggatcttgattagtatgaacttcgcctccgccgacgccaagga
Apiquitin	Apiquitin-Ty pDC	Apiq3TYF	ggaggcgaagttcatactaatcaagatccactagatatgcagatctttgtgaagactcttacc
Apiquitin/ Ubiquitin	Api/Ub-Ty Chimera pDC	UB5TYR	gttgaaatctagtggatcttgattagtatgaacttcgcctccgcggacgccaagga
Apiquitin/ Ubiquitin	Api/Ub-Ty Chimera pDC	UB3TYF	ggaaggcgaagttcatactaatcaagatccactagatttcaacttgcaagtcaagacaatgagc
Apiqutin/ Ubiquitin	Api/Ub-Ty Chimera pDC	UB5BglIIF	gatagatctatgtgcgcgtatgtgaggtgtg
Apiquitin/ Ubiquitin	Api/Ub-Ty Chimera pDC	UB3EcoRVR	agtgatatctcatgcgcctcctcggagt
Apiquitin	Apiquitin180AApTG	223125AvrII	agtcctaggtgcgcctcctcggagtctga
Apiquitin	Apiquitin-Ty 239K/R	Api6KRF	gtcttcccgctcattgtcctgacttgcaagttgaaat
Apiquitin	Apiquitin-Ty 239K/R	Api6KRR	atttcaacttgcaagtcaggacaatgagcgggaagac
Apiquitin	Apiquitin-Ty 282K/R	ApiKtoR48F	accaacagcgtcttgtattcaacggacgacaactggagaatgg
Apiquitin	Apiquitin-Ty 282K/R	ApiKtoR48R	ccattctccagttgtcgtccgttgaatacaagacgctgttggt
Apiquitin	Apiquitin-Ty 288K/R	UBKtoR54F	cggaaaacaactggagaatgggcgaacagtccaagattacaatctc
Apiquitin	Apiquitin-Ty 288K/R	UBKtoR54R	gagattgtaatcttggactgttcgcccattctccagttgttttccg
Apiquitin	Apiquitin-Ty 244K/R	ApiK11toRF	gtcaagcacgacagtcctcccgctcattgtctt
Apiquitin	Apiquitin-Ty 244K/R	ApiK11toRR	aagacaatgagcgggaggactgtcgtgcttgac
Apiquitin	Apiquitin-Ty 232GG/AA	UBinternalggFty	gattagtatgaacttcgactgcgcggacgccaaggag
Apiquitin	Apiquitin-Ty 232GG/AA	UBinternalggRty	ctccttggcgtccgcgcagtcgaagttcatactaatc

Apiquitin	Apiquitin-Ty 309GG/AA	ApiqGGAAF	tgtgatatctcatgcggctgctcggagtctgaggac
Apiquitin	Apiquitin-Ty 309GG/AA	ApiqGGAAR	tcctcagactccgagcagccgcatgagatatcaca
Apiquitin	Apiquitin pAVA	UBLicF	ccagggtcctggttcgttcaacttgcaagtcaagac
Apiquitin	Apiquitin PAVA	UBLicR	cttgttcgtgctgtttatcatgcgcctcctcgga
Apiquitin	(i)Δapiquitin	223125PRF	acgactaccaaaattaatgcctggaagagaaccgaagctacagagaaaaggaatggtaaccgacaaacgcgttc
Apiquitin	(i)Δapiquitin	223125PRR	gcaacttgacgggggattaatgcatttcactcttgcagatagtttttcatagatctggttgaagacagacgaaag
Apiquitin	(i)Δapiquitin Promoter	UBPRcheckF	ccatccactatgttgccttc
Apiquitin	(i)Δapiquitin Promoter	UBPRcheckR	tcctttgtggagaaccgct
$CDC48_{AP}$	(i)Δ$CDC48_{AP}$	CDC48PRF	tctctattatcacagcacacggtcgtcaggagcttttcgacagtacgtgcgaatggtaaccgacaaacgcgttc
$CDC48_{AP}$	(i)Δ$CDC48_{AP}$	CDC48PRR	gctgaggagtcccagcccgcaccctcgagagggcaccacgcagtccccatagatctggttgaagacagacgaaagc
$CDC48_{AP}$	(i)Δ$CDC48_{AP}$ Promoter	CDC48PRcheckF	gtacacggcgttaatcctg
$CDC48_{AP}$	(i)Δ$CDC48_{AP}$ Promoter	CDC48PRcheckR	tccctctctgagagcgaa
T7s4	T7s4 Promoter	DHFRseqF	cgtttcctcttccctcaaa
T7s4/ $CDC48_{AP}$	T7s4 Promoter for (i)Δ$CDC48_{AP}$	CDC48PRR	gctgaggagtcccagcccgcaccctcgagagggcaccacgcagtccccatagatctggttgaagacagacgaaagc
T7s4/ Apiquitin	T7s4 Promoter for (i)Δapiquitin	223125PRR	gcaacttgacgggggattaatgcatttcactcttgcagatagtttttcatagatctggttgaagacagacgaaag
DHFR	pDC	DHFRSpeIF	agtactagtccaggctgtaaatcccgt
DHFR	pDC	DHFRBglIIR	gatagatctcttcccagacacgacaac
$CDC48_{AP}$	Δ$CDC48_{AP}$ K/A-myc	CDC48KAF	cgagcggctgtggagcgaccttgctggcga
$CDC48_{AP}$	Δ$CDC48_{AP}$ K/A-myc	CDC48KAR	tcgccagcaaggtcgctccacagccgctcg
$CDC48_{AP}$	Δ$CDC48_{AP}$ E/Q-myc	Cdc48E829Q rev	ttcgcgatcgagtccatctggtcgaagaagatcacgc
$CDC48_{AP}$	Δ$CDC48_{AP}$ E/Q-myc	Cdc48E829Q sense	gcgtgatcttcttcgaccagatggactcgatcgcgaa

CHAPTER 4

CHARACTERIZATION OF PUTATIVE APICOPLAST DEUBIQUITINASES AND UBIQUITIN LIGASES IN *TOXOPLASMA GONDII*

4.1 Introduction

Toxoplasma gondii is a protozoan parasite that chronically infects about 22.5% of the U.S. population [1]. Once established the infection is typically maintained for the remainder of the life of a person. *T. gondii* can cause encephalitis and infectious myocarditis in immunosuppressed individuals while blindness, hydrocephalus and still births can occur in the context of congenital toxoplasmosis [2, 3]. *T. gondii* has a unique organelle called the apicoplast which is derived from a red algal endosymbiont and is essential for the parasite as the location of fatty acid, heme and isoprenoid biosynthesis [4-6]. It is estimated that essentially all 500 apicoplast proteins are nuclear encoded and thus must be imported across the four membranes that surround the apicoplast in order to maintain proper organelle function [7]. *T. gondii* is a strong model to study the apicoplast due to its strong genetic tools and well established functional assays that have been previously developed by our laboratory [8]. The apicoplast is an established drug target due to the fact that drugs including azithromycin, spiramycin and clindamycin act on the organelle [9, 10]. However, our understanding of the apicoplast is limited at this time and a question of particular interest is how proteins are imported across the four membranes that surround the organelle.

The apicoplast is derived from a secondary endosymbiotic event between a red algae and a flagellated heterotrophic protist resulting in the four membranes that surround the apicoplast [6]. Nuclear encoded apicoplast proteins typically contain a bipartite leader peptide (consisting of a signal and transit peptide) which acts as a signal for the protein to be transported into the ER through the Sec61 complex [11, 12]. Subsequently, it is believed proteins travel by vesicles to the outermost membrane of the apicoplast where the vesicle fuses with the outermost membrane [13-15]. Proteins interact with endoplasmic reticulum associated degradation (ERAD) homologs that include the $Der1_{AP}$ translocon to cross the periplastid membrane (PPM) into the periplastid compartment (PPC) [16]. We have demonstrated that the plastid ERAD AAA ATPase $CDC48_{AP}$, ubiquitin conjugating enzyme ($E2_{AP}$) and plastid ubiquitin-like protein (PUBL) are essential in protein trafficking across the PPM [17]. Nuclear encoded apicoplast proteins then traverse the two innermost membranes through the Toc75 and Tic20 translocons respectively to reach the innermost compartment of the apicoplast [8, 18]. The classic ERAD machinery acts as a quality control system for protein folding in the ER and the secretory pathway. Misfolded proteins are recognized and exported across the ER membrane where they are marked by ubiquitination leading to subsequent degradation by the proteasome [19]. However, experimental evidence suggests that the ERAD derived ubiquitinating machinery of the apicoplast has been retooled to participate in the transfer of proteins across the PPM rather than a signal for protein degradation [16, 20]. One of the remaining gaps of knowledge in this import model is how PUBL and the ubiquitin-like machinery are facilitating the transport of proteins across the PPM.

The leading hypotheses for the role of PUBL is that PUBL is transferred from the ubiquitin activating enzyme ($E1_{AP}$) to the ubiquitin conjugating enzyme ($E2_{AP}$) which adds ubiquitin onto the imported protein. $CDC48_{AP}$ and its cofactor $Ufd1_{AP}$ recognize the protein conjugated by PUBL which allows $CDC48_{AP}$ to pull the protein through the translocon at the PPM. This hypothesis is supported by multiple independent observations from studies of complex plastids and the classic ERAD machinery. It has been demonstrated that CDC48 and Ufd1 in the classic ERAD pathway contain ubiquitin binding domains which are necessary for transport of proteins across the ER membrane [21, 22]. Additional experiments have suggested that ubiquitin has an early mechanistic role in protein translocation across the ER membrane in addition to its later role in protein degradation [23]. Perhaps the most intriguing evidence comes from the diatom *P. tricornutum* which also contains a complex plastid that is evolutionary derived from a red algal secondary endosymbiosis event. *P. tricornutum* similarly retooled the ERAD machinery for import across the PPM [24]. Biochemical and molecular experiments show that nuclear encoded plastid proteins require lysine residues at the leader peptide for import across the PPM and the lack of lysines result in imported proteins "frozen" in the PPM [25]. Lysine residues are the main sites for ubiquitination and the formation of polyubiquitin chains suggesting that plastid ubiquitin-like proteins binding onto substrates are critical for import. An alternative hypothesis is that the ubiquitin ligase auto-ubiquitinates itself as mechanism to control when proteins are transported across the PPM. The ubiquitin ligase in the ERAD system is a transmembrane protein in the ER membrane that is key in recognizing misfolded proteins in the lumen of the ER and most likely forms a pore with Der1 [26, 27]. Recently, it was shown through in vivo and in

vitro experiments that auto-ubiquitination of lysine residues of the ERAD ubiquitin ligase is essential for misfolded proteins to be translocated across the ER membrane. This suggests that the ubiquitin ligase acts as a gate for protein translocation and that ubiquitination of the ubiquitin ligase controls when proteins are exported out of the ER membrane [27]. Therefore, we expect PUBL to be attached to the apicoplast specific ubiquitin ligase or a multitude of imported apicoplast proteins based on these hypotheses.

These hypotheses would predict covalent modifications of proteins by PUBL attachment to the substrates. However, this has not been formally demonstrated. A possible explanation is that this process is transient due to an active plastid deubiquitinase resulting in the inability to visualize ubiquitination. Loss of apicoplast specific deubiquitinase activity due to ablation of a suitable gene could potentially result in the accumulation of PUBL on substrates in a similar manner a deubiquitinase mutant increase ubiquitination in the cytoplasm [28, 29]. However, currently we cannot test this hypothesis because we do not know the identity of a known apicoplast specific deubiquitinase in *T. gondii*. Similarly an apicoplast specific ubiquitin ligase has not been identified which complicates testing whether PUBL is solely bound to the ubiquitin ligase as a way to regulate import. Therefore, here we strive to identify a substrate of PUBL in the apicoplast of *T. gondii* to better understand the protein import mechanism. We attempt this by tagging putative apicoplast specific deubiquitinases and ubiquitin ligases. We also seek to identify apicoplast specific deubiquitinases using biochemical probes which irreversibly bind to deubiquitinases that attempt to cleave PUBL. We were able to tag several candidates and identify their localization in an effort to narrow down our search for the correct candidate. We demonstrate that there is a difference in

deubiquitinase recognition and activity between ubiquitin and PUBL through in vitro assays.

4.2 Results

Localization of potential apicoplast specific ubiquitin ligases

We have previously localized the apicoplast ubiquitin activating enzyme ($E1_{AP}$) and the ubiquitin conjugating enzyme ($E2_{AP}$) and demonstrated the importance of $E2_{AP}$ in the PPC of the apicoplast [17]. It is suggested that the ubiquitination machinery and other ERAD components are essential for nuclear encoded apicoplast protein import across the PPC. However, the ubiquitination machinery also consists of the ubiquitin ligase (E3) which is the enzyme that recognizes substrates to be ubiquitinated and aids in the final step of conjugating ubiquitin to the substrate. Previous attempts at identifying an apicoplast specific ubiquitin ligase ($E3_{AP}$) in *T. gondii* have fallen short despite the confirmation of plastid specific ubiqutin ligases in related organisms [17, 30, 31]. A comprehensive study of ubiquitin and the ubiquitination machinery discovered 94 ubiquitin ligases in the genome of *T. gondii* by using PFAM domain models for E3s and the latest *T. gondii* proteome [32]. Attempting to tag close to a hundred ubiquitin ligases would be a monumental task. Therefore, we concentrated our strategy to search for potential $E3_{AP}$ candidates by searching for ubiquitin ligases with putative signal peptides and multiple transmembrane domains. We narrowed the candidates using these parameters due to the reasoning that the majority of apicoplast proteins need a signal peptide for import into the apicoplast and E3s previously identified contain multiple transmembrane domains [17, 31, 33]. There are 11 E3's that contain a signal peptide and

23 E3's with transmembrane domains. Here we list 25 of the best apicoplast specific ubiquitin ligase candidates (Table 4.1).

We next wanted to test whether any of these candidates are localized to the apicoplast. We amplified the leader sequences of five $E3_{AP}$ candidates from cDNA by PCR and ligated it into expression plasmids that introduce a GFP epitope tag at the C-terminus. Immunofluorescence assays (IFA) were performed on these five parasite lines transiently expressing the tagged gene which showed uniform labeling throughout the cell. We counterstained with an apicoplast specific antibody (CPN60) which is visualized as a single punctate structure per cell (Fig. 4.1A-E). This suggests that the five $E3_{AP}$ candidates tagged are not apicoplast proteins, but are most likely cytoplasmic based on the staining. The experimentally localized ubiquitin ligase in the apicoplast of *P. falciparum* (PF3D7_0312100) contains five transmembrane domains suggesting that the $E3_{AP}$ in the apicoplast of *T. gondii* may also contain multiple transmembrane domains [17]. Genes TgMe49_248450 and TgMe49_304460 both contain five transmembrane domains near the N-terminus and are closely related to the E3's identified in the complex plastids of other organisms. Therefore, both genes were endogenously tagged at the C-terminus with an HA epitope tag to avoid interfering with the multiple transmembrane domains and a stable line was selected for in the presence of chloramphenicol. Immunofluorescence assays revealed punctate structures for TgMe49_248450 in the proximity of the apicoplast marker CPN60, but not fully colocalizing with the marker. This suggests TgMe49_248450 has a unique localization, but is not in the apicoplast. Several attempts at identifying the localization of stable tagged lines of gene TgMe49_304460 did not show any expression or staining.

Localization of putative apicoplast specific deubiquitinases

A detailed analysis of the *T. gondii* ubiquitin proteome suggested that there are 46 deubiquitinases in the genome with only six of these genes containing a predicted signal peptide [32]. As previously mentioned, most apicoplast proteins contain a signal peptide which narrows down the list of apicoplast specific deubiquitinase candidates. Additionally, a genome wide CRISPR/Cas9 screen was performed in *T. gondii* which measured a gene's contribution to the parasites overall fitness [34]. The lower the phenotypic score given in that study the less dispensable the gene is to the parasite's fitness. Proteins involved in apicoplast import including the ubiquitinating machinery are typically essential for the parasite suggesting that a deubiquitinase may also be important for parasite fitness [16, 17]. 17 of the 46 deubiquitinases have a score of -2 or lower which suggests they are dispensable. Listed in Table 4.2 are the 15 best candidates based on the previously mentioned criteria. We began to systematically analyze these candidates by endogenously tagging the gene or expressing the leader peptide with an epitope tag at the C-terminus. We performed immunofluorescence assays on parasites expressing these tagged gene candidates which displayed varied putative staining patterns including cytoplasmic, plasma membrane, ER and nuclear staining. However, none of the proteins colocalized with the apicoplast marker CPN60 (Fig. 4.2A-H). This suggests that none of the deubiquitinase candidates are apicoplast specific or that our experimental approach interferes with their function or localization.

A deubiquitinase has been identified in the complex plastid of *P. tricornutum* which is related to the apicoplast [30]. Blast searches show that TgMe49_289330 is the closest homolog to the plastid specific deubiquitinase. Interestingly, this gene is one of

the three genes that is considered indispensable for the parasites and has a predicted signal peptide. This deubiquitinase was also identified in experiments from an ATG8 pull down which is a known apicoplast associated protein (unpublished data). This scenario makes the gene the strongest apicoplast specific deubiquitinase candidate. However, we were unable to endogenously tag this gene and we were only able to see cytoplasmic and plasma membrane localization from expressing the first 284 amino acids with a GFP epitope tag. We thought it was worthwhile to generate a conditional mutant and to test whether the gene had an apicoplast biogenesis defect. PCR analysis shows that we were able to successfully replace the endogenous promoter of TgMe49_289330 with the tetracycline regulatable promoter (t7s4) and produce a stable monoclonal line (Fig. 4.3A). The development of a conditional mutant allowed us to test whether gene TgMe49_289330 which we will refer to as UCH (ubiquitin C-terminal hydrolase) is necessary for parasite survival. Plaque assays were performed to examine whether parasites can invade, replicate and egress out of the host cell resulting in visible plaques upon the ablation of gene expression. (i)ΔUCH grown in the absence of anhydrous tetracycline (ATc) results in plaque formation, but upon the addition of ATc no plaques are visible suggesting that TgMe49_289330 is crucial for parasite growth (Fig. 4.3B). This finding correlates with the gene's phenotypic score of -4.51.

Conditional mutants generated for genes involved in the apicoplast specific ubiquitinating machinery all have shown import defect phenotypes. The loss of apicoplast import results in the loss of this processing of apicoplast protein and we have previously exploited this to measure apicoplast import in several mutants [8, 16, 17, 35]. Therefore, we tested the ability of the (i)ΔUCH line to import apicoplast proteins in the presence and

absence of ATc. The loss of processing of apicioplast proteins can be measured as the majority of nuclear encoded apicoplast proteins contain an N-terminal transit peptide which is cleaved by an unknown protease in the lumen of the apicoplast [36]. Western blots reveal two bands, a larger band representing the apicoplast protein *en route* to the organelle and a smaller band represents the mature protein that has been processed in the lumen of the apicoplast. Acyl carrier protein (ACP) a protein that targets to the apicoplast lumen was endogenously tagged with YFP in the (i)ΔUCH line. A stable (i)ΔUCH/ACP-YFP line was generated and an IFA was performed to test whether ACP-YFP was localized to the apicoplast. The IFA showed ACP colocalizes with the apicoplast marker CPN60 demonstrating ACP-YFP is correctly trafficked to the apicoplast (Fig. 4.3C). The (i)ΔUCH/ACP-YFP line was grown from 0 to 5 days under ATc treatment and then harvested for Western blot analysis. The mature and precursor form of the protein was seen throughout ATc treatment (Fig. 4.3D). This suggests that the protein is processed upon the genetic ablation of TgMe49_289330 showing that the gene does not affect apicoplast import. We next wanted to test whether gene TgMe49_289330 is essential for apicoplast biogenesis. This was tested by counting the number of apicoplasts upon the addition of ATc in the (i)ΔUCH line and monitoring whether there was a loss of the organelle (Fig 4.3E-F). There was no significant decrease in apicoplast numbers in the (i)ΔUCH line compared to the parental line suggesting that UCH does not have a role in apicoplast biogenesis. This evidence makes it highly unlikely that gene TgMe49_289330 is associated with the apicoplast ubiquitin machinery.

Plastid ubiquitin-like protein probes to identify interacting deubiquitinases

The unsuccessful attempts at tagging and identifying a putative apicoplast specific deubiquitinase proved the necessity for other methods to identify deubiquitinases. A useful tool to study deubiquitinase activity are ubiquitin probes with electrophilic groups at the C-terminus that irreversibly bind to the cysteine moiety of active deubiquitinases [37]. Ubiquitin-like protein probes Nedd8, ISG15 and SUMO-1 have also been synthesized and demonstrated to irreversible form thioether bonds with specific proteases [38]. Therefore, various epitope tagged PUBL probes were synthesized that contained a vinylmethylester (VME), vinyl sulfone (VS) and propargylamine (PA) electrophilic group at the C-terminus. The synthesis of probes fusing propargylamine to the C-terminus of ubiquitin display better reactivity to deubiquitinases and provide more stringent activity-based proteomic studies than probes with the vinyl groups [39]. An inactive PUBL probe was also generated with a PA electrophilic group which contains the PUBL sequence lacking the diglycine motif preventing covalent bond formation between PUBL and a deubiquitinase.

We first used PUBL and ubiquitin probes on *T. gondii* lysates and saw different banding patterns between the two different probes (Fig. 4.4A). This suggests a difference between PUBL and ubiquitin and how deubiquitinases recognize these two ubiquitin-like proteins. The PUBL probes consistently showed two bands around 40 KDa and 55 KDa indicating that two different deubiquitinases interact with PUBL. However, the 40 KDa band is seen in both ubiquitin and PUBL probes suggesting that the 40 KDa deubiquitinase is not PUBL specific. The 55 KDa band is absent from the ubiquitin probe Western blot suggests that the 55 KDa deubiqutinase is specific to PUBL and a good

apicoplast specific candidate. The C-terminal electrophilic groups have different activity towards deubiqutinases explaining why the 55 KDa is faint using the VME probe compared to the VS probe. The biotin tagged PUBL-PA probe was also incubated with *T. gondii* lysates revealing three bands representing deubiquitinases (Fig. 4.4B). PUBLΔGG-PA probes were used as a control to verify that these bands represent active deubiquitinases and are not the result of nonspecific binding. The largest band is still visible in the inactive probe suggesting that the band is not a deubiquitinase, but most likely the mitochondrial biotinylated protein carboxylase. Two of the three bands were the previously identified 40 KDa and 55 KDa deubiquitinases. However, PUBL-PA also displayed a fainter ~35 KDa band suggesting that the PUBL-PA probe is more sensitive than the other probes. These probes were also used in samples generated from human foreskin fibroblast host cells that *T. gondii* are grown in to examine whether the active deubiquitinases maybe from the host cell (Fig. 4.4A). Western blot analysis reveals that the band sizes from the host cells are different from the parasite lysates suggesting that the deubiquitinases interacting with PUBL are *T. gondii* specific. The two bands observed for the PUBL probe are also present in the ubiquitin probe suggesting that those two host cell deubiquitinases are promiscuous and recognize multiple types of ubiquitin-like proteins.

4.3 Discussion

Ubiquitin is a 76 amino acid protein found throughout the eukaryotic tree of life. Ubiquitin is one of the most conserved proteins with only a single amino acid (an aspartic acid to glutamic acid substitution) difference between *T. gondii* ubiquitin and human ubiquitin. Ubiquitin regulates a number of critical cell processes including cell

degradation, transcription regulation, DNA repair, signaling and the endocytic pathway by covalently binding onto lysine residues of proteins [40]. Ubiquitin is encoded as three forms in the genome of *T. gondii*. Two of the ubiquitin genes are fused to ribosomal units (RPL40 and S27A) and the third is encoded as a linear polyubiquitin chain. *T. gondii* encodes many of the other conserved ubiquitin-like proteins including a single small ubiquitin like modifier (SUMO) which has been previously characterized in the parasite [41]. A surprising discovery was that *T. gondii*, along with organisms with a four membrane bound complex plastid originating from a red algae, encodes a unique plastid ubiqutin-like protein (PUBL). Genetic studies demonstrated that PUBL and the related ubiquitin conjugating enzyme ($E2_{AP}$) are required for nuclear encoded apicoplast protein import across the periplasmid membrane [17]. This study focuses the search for potential apicoplast specific deubiquitinases and ubiquitin ligases to garner a better understanding of a process essential for parasite growth and pathogenesis.

The ubiquitin machinery consists of the ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2) and the ubiquitin ligase (E3). Most ubiquitin-like proteins utilize a similar enzymatic cascade of proteins that facilitate the transfer of the ubiquitin-like protein to the substrate [42]. The enzymatic cascade begins with ubiquitin forming a thioester linkage with a catalytic cysteine residue of E1 and subsequently transferring ubiquitin to the E2 protein. The final step involves E3 recognizing substrates and catalyzing the transfer of ubiquitin to the lysine residue of the substrate [43]. RING between RING (RBR), really interesting new gene (RING) finger and homolog to E6AP C terminus (HECT) are the three family types of ubiquitin ligases [44]. Ubiquitin is directly transferred to a catalytic cysteine residue of the E3 before being transferred to the

substrate when RBR and HECT are the E3s of the enzymatic cascade. RING family E3s are unique in that they bind both the ubiquitin bound E2 and the substrate while facilitating the transfer of ubiquitin from the E2 subunit directly to the substrate [45].

The E3 that is associated with the ERAD pathway called Hrd1p is a RING finger class E3, resides in the ER membrane and has six transmembrane domains [46]. The E3 involved in protein import into complex plastids has been identified in *P. falciparum* and *P. tricornutum* [17, 30]. The E3s of *P. falciparum* and *P. tricornutum* are both RING finger proteins with transmembrane domains and have been validated to be active enzymes by in vitro auto-ubiquitination assays [17, 30]. There are a reported 94 putative E3 proteins in the genome of *T. gondii* which consist of 68 proteins containing RING domains and 10 containing the HECT ubiquitin ligase domain [32]. We were able to tag and localize six ubiquitin ligases showing diverse localizations. Unfortunately, none of the ubiquitin ligases tagged showed a clear apicoplast specific localization. Recently, activity based probes for ubiquitin ligases and the whole ubiquitin machinery have been utilized by designing E2-ubiquitin conjugates or ubiquitin-dehydroalanine conjugates respectively [47, 48]. A similar plastid ubiquitin-like protein probe could be synthesized to identify the plastid E3 and demonstrate activity. TgMe49_248450 and TgMe49_304460 were the leading candidates to be the plastid specific E3 because they both contain multiple transmembrane domains and are closely related to the previously identified E3s of the complex plastid. However, the lack of a signal peptide, a positive phenotypic score and non-apicoplast localization eliminates gene TgMe49_248450 as a strong candidate. TgMe49_304460 contains a signal peptide and has a low phenotypic score similar to other apicoplast proteins involved in import suggesting that the gene is

the leading plastid E3 candidate. Multiple attempts at tagging the gene and generating a conditional mutant came up short resulting in failure to characterize the gene. The inability to visualize a tagged version of the protein might be due to incorrect gene models or the addition of an epitope tag interferes with proper protein function.

Another important factor in the ubiquitination process is the ability to cleave ubiquitin and polyubiquitin chains to recycle ubiquitin molecules. Ubiquitin is cleaved by a deubiquitinase enzyme. There are five classes of deubiquitinases: ubiquitin C-terminal hydrolase (UCH), ubiquitin specific protease (USP), ovarian tumor (OTU), Josephin and JAB1/MPN/MOV34 (JAMM) domain deubiquitinases [49]. All of the deubiquitinase families are cysteine proteases with the exception of JAMM which is a zinc dependent metalloprotease. The deubiquitinase that is linked with the classic ERAD system is an OTU type deubiquitinase called YOD1 that is associated with CDC48. Experiments have shown that not only is the ubiquitin process required for proper ERAD function, but deubiquitinating activity is also critical [28]. It has been specifically demonstrated that the deubiquitinase YOD1 is required for proper retro-translocation [23, 50]. There is even a report that the deubiquitinase YOD1 is needed for translocation of misfolded proteins that are not ubiquitinated across the ER membrane [51]. These findings suggest that the complex plastid deubiquitinase could also be critical for import of proteins. There are 46 predicted deubiquitinase enzymes encoded in the genome of *T. gondii* with a majority of them being cell cycle regulated [32]. A plastid specific deubiquitnase has been identified in *P. tricornutum* and an in vitro deubiquitination assay has confirmed that the protein is a functional deubiquitinase [30]. In this study we tagged nine deubiquitinases which revealed putative cytoplasmic, ER and nuclear localizations. However, we were unable to

identify an apicoplast specific deubiquitinase. The leading candidate at this time is gene TgMe49_260510 because it is one of the few deubiquitinases with a signal peptide and is an OTU class deubiquitinase similar to the ERAD YOD1 deubiquitinase. We transiently expressed the leader peptide with an epitope tag and saw cytoplasmic localization, but overexpression could result in mislocalization. Therefore, we set out to endogenously tag this gene, but we were consistently unsuccessful in our attempts. We have also attempted to generate a conditional mutant for this gene, but we were unable to isolate a stable line. A different strategy to identify the localization of the protein and generate a mutant is needed to further characterize this protein.

Deubiquitinases contain a catalytic triad domain which consist of a cysteine residue that performs a nucleophilic attack on the isopeptide bond of the lysine residue, a histidine residue and an asparagine or aspartic acid residue that enable the nucleophilic attack [52]. The understanding of this process has been exploited with ubiquitin probes with electrophilic groups at the C-terminus. Probes were generated using the last 77 amino acids of the plastid ubiquitin-like sequence and were incubated with *T. gondii* lysates. Two bands were consistently visualized around the size of 40 KDa and 55 KDa. Ubiquitin and ubiquitin-like probes containing electrophilic groups have been used previously in *T. gondii* and *P. falciparum* to identify deubiquitinases [53, 54]. It was shown that the *P. falciparum* deubiquitinase PfUCH54 recognizes ubiquitin and the ubiquitin-like protein Nedd8 [54]. However, the most prominent deubiquitinase in *T. gondii* and *P. falciparum* screens was UCHL3 which was shown to recognize ubiquitin and the ubiquitin-like protein Nedd8. UCHL3 bound to the UB-VME probe is the size of ~ 40 KDa on a Western blot. UCHL3 is a promiscuous deubiquitinases suggesting that

the 40 KDa band seen in PUBL probes could represent UCHL3 attached to the probe. UCHL3 was demonstrated to reside in the nucleus and cytoplasm of *T. gondii* suggesting that UCHL3 is most likely not interacting with the plastid ubiquitin machinery [53]. The strongest deubiquitinase candidate TgMe49_260510 is predicted to encode a protein the size of 44.4 KDa which is roughly equivalent to the 55 KDa band if the enzyme is bound by the 8.5 KDa PUBL probe. This strengthens the claim that TgMe49_260510 is the leading candidate. Performing pull downs on the bound probes and performing mass spectrometry with the pull down sample is a top priority to determine the active deubiquitinases.

What is the role of ubiquitin and PUBL in *T. gondii*? A detailed analysis of the *T. gondii* ubiquitin proteome reveals ubiquitination of a number of transcription factors, histones, chromatin modifying proteins suggesting that ubiquitin in *T. gondii* has synonymous functions to other eukaryotes [32]. Over 500 proteins with diverse localizations were modified by ubiquitin. It was shown that 35% of these proteins have a cell cycle regulated transcription profile which is consistent with the observation that the ubiquitin-proteasome system degrades cell cycle checkpoint proteins. It was also shown that 18% of the ubiquitinated proteins localize to the inner membrane complex (IMC). The mapping of the ubiquitin proteome was made possible by performing pulldown experiments using diglycine antibodies which would potentially detect any proteins bound by PUBL. However, no apicoplast proteins were shown to be ubiquitinated or conjugated to PUBL [32]. The lack of an apparent PUBL protein modification is possibly due to the process being transient in nature due to an active deubiquitinase. We hypothesize that knocking out a deubiquitinase results in the accumulation of PUBL on

the target protein. In this study we narrow down the list of apicoplast specific deubiquitinase candidates and come one step closer to testing this hypothesis. Another immediate question that remains unanswered is which proteins are targeted by PUBL. One possible model is that the ubiquitin ligase auto-ubiquitinates itself acting as a PUBL gated channel controlling protein import across the PPM. Here we narrow down the search for the apicoplast specific ubiquitin ligase and provide evidence that gene TgMe49_304460 is the strongest candidate. Overall, uncovering the mechanistic role of PUBL facilitating import across the PPM of the apicoplast is dependent on identifying a plastid specific deubiquitinase and ubiquitin ligase.

4.4 Materials and Methods

Cell Culture and Transfection. *T. gondii* RH and ΔKu80/TATi strains were cultivated in human foreskin fibroblasts (HFFs) in Dulbecco's Modified Eagles Medium supplemented with fetal bovine serum, penicillin-streptomycin, and L-glutamine. Parasites were resuspended in cytomix supplemented with 2mM ATP and 5 mM glutathione to 3.3 X 10^7 parasites per ml prior to transfections. 300 µl of the parasite suspension and 30 µg of plasmid were mixed and transferred to a 2 mm gap cuvette. Parasites were electroporated using a single 1.5 kV pulse, a resistance of 25 Ω, and a capacitor setting of 25 µF. Parasites were selected in the presence of 1 µM of pyrimethamine or 20 µM of chloramphenicol.

Tagging of Genes. The coding region of genes TgME49_0205770, TgME49_277400, TgME49_258790, TgME49_209000, TgME49_295658, TgME49_260510, TgME49_289330, TgME49_251500 and TgME49_308590 was amplified from *T. gondii* cDNA using primers (See Table 4.3) that amplified anywhere from the first ~150

nucleotides to the full gene and inserted into vectors PTDHA$_{x3}$ or pTDGFP. The cDNA of the gene of interest was inserted into the vector using BglII and AvrII restriction cut sites. The gene of interest was ectopically expressed and visualized. Genes TgME49_248450, TgME49_243510, TgME49_294360, TgME49_255180, TgME49_229710 and TgME49_263420 were endogenously tagged with a HAx3 tag by using vector pLIC-HA$_3$-CAT as previously described [55, 56]. Stables lines were established through selecting parasites in 20 μM of chloramphenicol. Acyl carrier protein was endogenously tagged with an YFP tag by transfecting the TgMe49_289330 mutant line with a linearized pLicCATYFPΔpacI-ACP vector. Transfectants were cultured in 20 μM of chloramphenicol to select for clones with the inserted plasmid. Expression and proper localization of the endogenously tagged ACP was tested by immunofluorescence assay (Fig. 4.3C).

Microscopy. Parasites were used to infect HFF coverslip for 24 hours and then infections were fixed for 20 minutes with 4% paraformaldehyde, blocked with 3% BSA for 10 minutes, and permeabilized with 0.2% Triton X-100 in 3% BSA/PBS for 20 minutes. Primary antibodies used were rabbit anti-CPN60 [16] at 1:2000, rat anti-HA at 1:400 (clone 3F10 Roche Applied Sciences) and mouse anti-GFP at 1:500 (Torry Pines Biolab). The secondary antibodies used were goat anti-mouse Alexa Fluor 488, goat anti-rat Alexa Fluor 488 and goat anti-rabbit Alexa Fluor 546 at 1:2000 (Invitrogen). An Applied Precision Delta Vision microscope was used for imaging. Images were deconvolved and adjusted for contrast using Softworx software.

Construction of Conditional Mutant. TgMe49_289330 or the UCH mutant was constructed by replacing the endogenous promoter with the conditional (t7s4) promoter

in a ΔKu80/TATi parasite line. A fosmid construct was engineered as previously described to replace the promoter of the gene of interest with the tetracycline conditional t7s4 promoter [57]. Modified fosmid RHfos09L19 was transfected into the ΔKu80/TATi strain and pyrimethamine drug selection was used to isolate stable parasite lines. PCR mapping of the genomic locus of TgMe49_289930 was used to analyze whether the parasite lines had successfully replaced the endogenous promoter with the conditional t7s4 promoter (see Fig. 4.3A).

Plaque Assay. Confluent HFF cells in T-25 flasks were infected with 100 parasites from the designated parasite line and treated in the presence or absence of 0.5 uM of ATc for ten days. After ten days flasks were washed with 1x PBS and fixed with 100% ethanol. Cells were stained with crystal violet as previously described [8].

Western Blot. Parasites were harvested (1 X 10^6), lysed in RIPA lysis buffer, boiled in 1x NuPAGE LDS Sample Buffer, loaded onto precast 10% and Any-KD Mini-PROTEAN TGX gels (Bio-Rad) and ran at 150 V [8]. Proteins were transferred to nitrocellulose membranes and probed with mouse anti-GFP at 1:500 (Torry Pines Biolab) and mouse anti-tubulin at 1:2000 (12G10, a gift from Jacek Gaertig, University of Georgia) antibodies followed by incubating the membranes with horseradish peroxidase-conjugated anti-mouse antibody at 1:10,000 dilution (Bio-Rad) or streptavidin horseradish peroxidase conjugated antibody at 1:500 (Pierce). Membranes were treated with Pierce ECL Western Blotting Substrate to detect bands after exposing the membrane to film.

Ubiquitin and PUBL electrophilic probe reaction. *T. gondii* parasites were harvested (1 X 10^6) and resuspended in freshly made ubiquitin buffer (50 mM Tris PH 7.4, 5 mM

$MgCl_2$, 250 mM sucrose, 1 mM DTT and 2mM ATP) supplemented in 0.1% Triton-X and Roche Complete Protease Inhibitor Cocktail. 1 uM of the PUBL or ubiquitin probe was added to the buffer and parasites were incubated for 30 minutes at room temperature. Parasites were prepared for Western blots as previously mentioned. Biotin PUBL probes were synthesized and received from the lab of Boris Rodenko and PUBL probes with the HA epitope tag were synthesized from the lab of Liz Hedstrom. The 77 amino acid ubiquitin domain of PUBL was used in synthesizing the PUBL probes.

References

1. Jones, J.L., M.E. Parise, and A.E. Fiore, *Neglected Parasitic Infections in the United States: Toxoplasmosis.* American Journal of Tropical Medicine and Hygiene, 2014. 90(5): p. 794-799.
2. Weiss, L.M. and J.P. Dubey, *Toxoplasmosis: A history of clinical observations.* Int J Parasitol, 2009. 39(8): p. 895-901.
3. Jones, J.L., et al., *Congenital toxoplasmosis: a review.* Obstet Gynecol Surv, 2001. 56(5): p. 296-305.
4. Mazumdar, J., et al., *Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in Toxoplasma gondii.* Proc Natl Acad Sci U S A, 2006. 103(35): p. 13192-7.
5. Nair, S.C., et al., *Apicoplast isoprenoid precursor synthesis and the molecular basis of fosmidomycin resistance in Toxoplasma gondii.* J Exp Med, 2011. 208(7): p. 1547-59.
6. van Dooren, G.G. and B. Striepen, *The Algal Past and Parasite Present of the Apicoplast.* Annu Rev Microbiol, 2013.
7. Foth, B.J., et al., *Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum.* Science, 2003. 299(5607): p. 705-8.
8. van Dooren, G.G., et al., *Toxoplasma gondii Tic20 is essential for apicoplast protein import.* Proc Natl Acad Sci U S A, 2008. 105(36): p. 13574-9.

9. Goodman, C.D. and G.I. McFadden, *Targeting apicoplasts in malaria parasites.* Expert Opin Ther Targets, 2013. 17(2): p. 167-77.

10. Wiesner, J., et al., *The plastid-like organelle of apicomplexan parasites as drug target.* Curr Pharm Des, 2008. 14(9): p. 855-71.

11. van Dooren, G.G., et al., *Processing of an apicoplast leader sequence in Plasmodium falciparum and the identification of a putative leader cleavage enzyme.* J Biol Chem, 2002. 277(26): p. 23612-9.

12. Tonkin, C.J., D.S. Roos, and G.I. McFadden, *N-terminal positively charged amino acids, but not their exact position, are important for apicoplast transit peptide fidelity in Toxoplasma gondii.* Mol Biochem Parasitol, 2006. 150(2): p. 192-200.

13. Heiny, S.R., et al., *Protein Traffic to the Plasmodium falciparum apicoplast: evidence for a sorting branch point at the Golgi.* Traffic, 2014. 15(12): p. 1290-304.

14. Bouchut, A., et al., *Vesicles bearing Toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption.* PLoS One, 2014. 9(11): p. e112096.

15. DeRocher, A.E., et al., *A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in Toxoplasma gondii.* Eukaryot Cell, 2008. 7(9): p. 1518-29.

16. Agrawal, S., et al., *Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins.* J Biol Chem, 2009. 284(48): p. 33683-91.

17. Agrawal, S., et al., *An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins.* PLoS Pathog, 2013. 9(6): p. e1003426.

18. Sheiner, L., et al., *Toxoplasma gondii Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins.* Traffic, 2015.

19. Smith, M.H., H.L. Ploegh, and J.S. Weissman, *Road to ruin: targeting proteins for degradation in the endoplasmic reticulum.* Science, 2011. 334(6059): p. 1086-90.

20. Hempel, F., et al., *ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms.* Mol Biol Evol, 2009. 26(8): p. 1781-90.

21. Ye, Y., H.H. Meyer, and T.A. Rapoport, *Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains.* J Cell Biol, 2003. 162(1): p. 71-84.

22. Flierman, D., et al., *Polyubiquitin serves as a recognition signal, rather than a ratcheting molecule, during retrotranslocation of proteins across the endoplasmic reticulum membrane.* J Biol Chem, 2003. 278(37): p. 34774-82.

23. Ernst, R., et al., *Enzymatic blockade of the ubiquitin-proteasome pathway.* PLoS Biol, 2011. 8(3): p. e1000605.

24. Maier, U.G., S. Zauner, and F. Hempel, *Protein import into complex plastids: Cellular organization of higher complexity.* European Journal of Cell Biology, 2015. 94(7-9): p. 340-348.

25. Lau, J.B., et al., *N-terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids.* Mol Microbiol, 2015. 96(3): p. 609-20.

26. Stein, A., et al., *Key steps in ERAD of luminal ER proteins reconstituted with purified components.* Cell, 2014. 158(6): p. 1375-88.

27. Baldridge, R.D. and T.A. Rapoport, *Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD.* Cell, 2016. 166(2): p. 394-407.

28. Sanyal, S., J.H. Claessen, and H.L. Ploegh, *A viral deubiquitylating enzyme restores dislocation of substrates from the endoplasmic reticulum (ER) in semi-intact cells.* J Biol Chem, 2012. 287(28): p. 23594-603.

29. Nakamura, N., et al., *Ubiquitin-specific protease 19 regulates the stability of the E3 ubiquitin ligase MARCH6.* Exp Cell Res, 2014. 328(1): p. 207-16.

30. Hempel, F., G. Felsner, and U.G. Maier, *New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms.* Mol Microbiol, 2010. 76(3): p. 793-801.

31. Sommer, M.S., et al., *Der1-mediated preprotein import into the periplastid compartment of chromalveolates?* Mol Biol Evol, 2007. 24(4): p. 918-28.

32. Silmon de Monerri, N.C., et al., *The Ubiquitin Proteome of Toxoplasma gondii Reveals Roles for Protein Ubiquitination in Cell-Cycle Transitions.* Cell Host Microbe, 2015. 18(5): p. 621-33.

33. DeRocher, A., et al., *Analysis of targeting sequences demonstrates that trafficking to the Toxoplasma gondii plastid branches off the secretory system.* J Cell Sci, 2000. 113 (Pt 22): p. 3969-77.

34. Sidik, S.M., et al., *A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes.* Cell, 2016. 166(6): p. 1423-+.

35. Harb, O.S., et al., *Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite Toxoplasma gondii.* Eukaryot Cell, 2004. 3(3): p. 663-74.

36. Waller, R.F., et al., *Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1998. 95(21): p. 12352-7.

37. Love, K.R., et al., *Ubiquitin C-Terminal Electrophiles Are Activity-Based Probes for Identification and Mechanistic Study of Ubiquitin Conjugating Machinery.* Acs Chemical Biology, 2009. 4(4): p. 275-287.

38. Hemelaar, J., et al., *Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins.* Mol Cell Biol, 2004. 24(1): p. 84-95.

39. Ekkebus, R., et al., *On Terminal Alkynes That Can React with Active-Site Cysteine Nucleophiles in Proteases.* Journal of the American Chemical Society, 2013. 135(8): p. 2867-2870.

40. Huang, T.T., et al., *Regulation of monoubiquitinated PCNA by DUB autocleavage.* Nat Cell Biol, 2006. 8(4): p. 339-47.

41. Braun, L., et al., *The small ubiquitin-like modifier (SUMO)-conjugating system of Toxoplasma gondii.* Int J Parasitol, 2009. 39(1): p. 81-90.

42. Vierstra, R.D., *The expanding universe of ubiquitin and ubiquitin-like modifiers.* Plant Physiol, 2012. 160(1): p. 2-14.

43. Pickart, C.M. and M.J. Eddins, *Ubiquitin: structures, functions, mechanisms.* Biochimica Et Biophysica Acta-Molecular Cell Research, 2004. 1695(1-3): p. 55-72.

44. Berndsen, C.E. and C. Wolberger, *New insights into ubiquitin E3 ligase mechanism.* Nature Structural & Molecular Biology, 2014. 21(4): p. 301-307.

45. Deshaies, R.J. and C.A.P. Joazeiro, *RING Domain E3 Ubiquitin Ligases.* Annual Review of Biochemistry, 2009. 78: p. 399-434.

46. Carvalho, P., A.M. Stanley, and T.A. Rapoport, *Retrotranslocation of a Misfolded Luminal ER Protein by the Ubiquitin-Ligase Hrd1p.* Cell, 2010. 143(4): p. 579-591.

47. Pao, K.C., et al., *Probes of ubiquitin E3 ligases enable systematic dissection of parkin activation.* Nature Chemical Biology, 2016. 12(5): p. 324-+.

48. Mulder, M.P.C., et al., *A cascading activity-based probe sequentially targets E1-E2-E3 ubiquitin enzymes.* Nature Chemical Biology, 2016. 12(7): p. 523-+.

49. Neutzner, M. and A. Neutzner, *Enzymes of ubiquitination and deubiquitination.* Lysine-Based Post-Translational Modification of Proteins, 2012. 52: p. 37-50.

50. Ernst, R., et al., *The otubain YOD1 is a deubiquitinating enzyme that associates with p97 to facilitate protein dislocation from the ER.* Mol Cell, 2009. 36(1): p. 28-38.

51. Bernardi, K.M., et al., *A deubiquitinase negatively regulates retro-translocation of nonubiquitinated substrates.* Mol Biol Cell, 2013. 24(22): p. 3545-56.

52. Clague, M.J., et al., *Deubiquitylases from Genes to Organism.* Physiological Reviews, 2013. 93(3): p. 1289-1315.

53. Frickel, E.M., et al., *Apicomplexan UCHL3 retains dual specificity for ubiquitin and Nedd8 throughout evolution.* Cellular Microbiology, 2007. 9(6): p. 1601-1610.

54. Artavanis-Tsakonas, K., et al., *Identification by functional proteomics of a deubiquitinating/deNeddylating enzyme in Plasmodium falciparum.* Molecular Microbiology, 2006. 61(5): p. 1187-1195.

55. Huynh, M.H. and V.B. Carruthers, *Tagging of Endogenous Genes in a Toxoplasma gondii Strain Lacking Ku80.* Eukaryotic Cell, 2009. 8(4): p. 530-539.

56. Fox, B.A., et al., *Efficient Gene Replacements in Toxoplasma gondii Strains Deficient for Nonhomologous End Joining.* Eukaryotic Cell, 2009. 8(4): p. 520-529.

57. Vinayak, S., et al., *Genetic manipulation of the Toxoplasma gondii genome by fosmid recombineering.* MBio, 2014. 5(6): p. e02021.

Figures

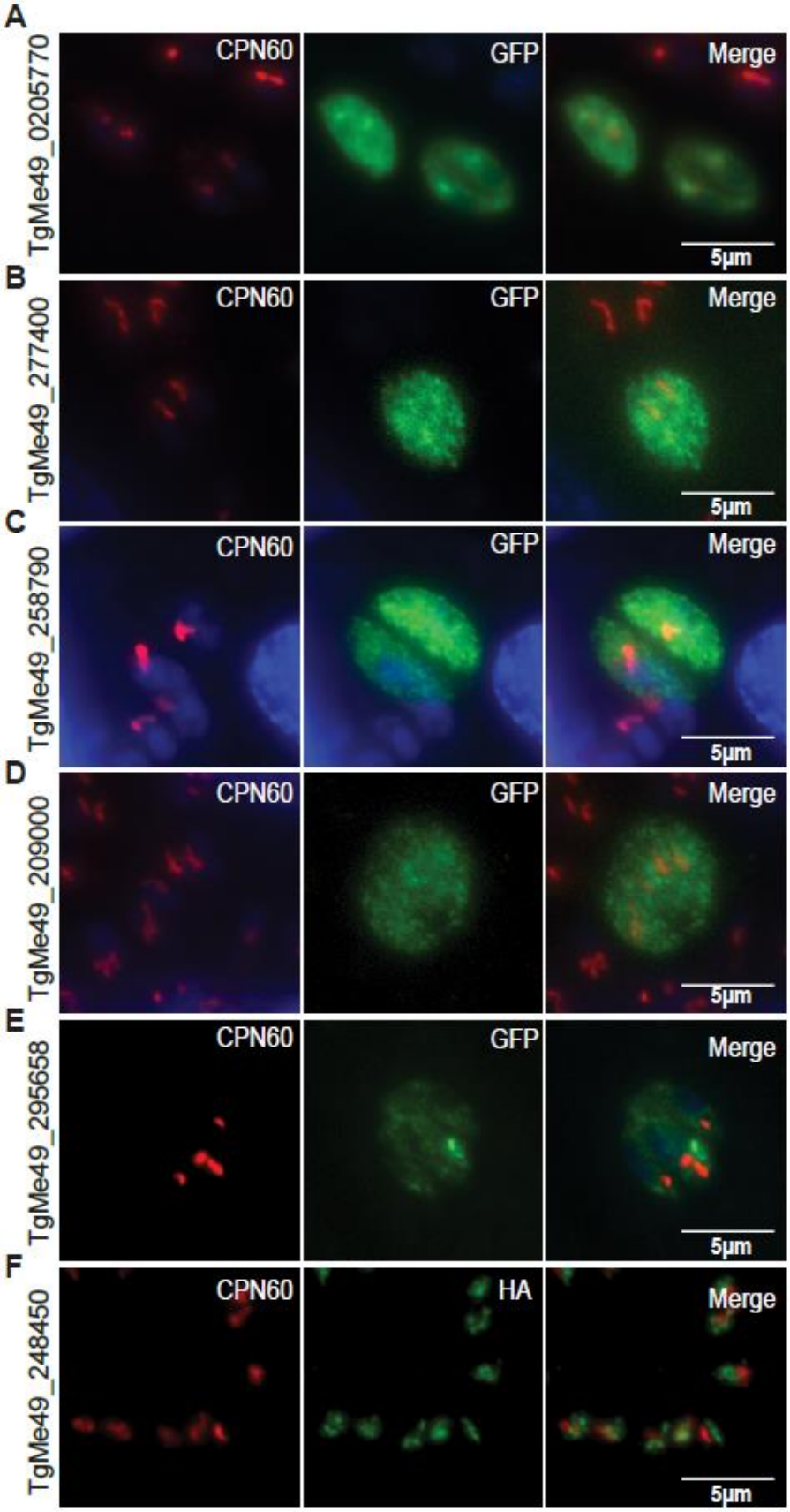

A
TgMe49_0205770
CPN60
GFP
Merge
5µm
B
TgMe49_277400
CPN60
GFP
Merge
5µm
C
TgMe49_258790
CPN60
GFP
Merge
5µm
D
TgMe49_209000
CPN60
GFP
Merge
5µm
E
TgMe49_295658
CPN60
GFP
Merge
5µm
F
TgMe49_248450
CPN60
HA
Merge
5µm

Figure 4.1. Tagging and staining of apicoplast specific ubiquitin ligase candidates. (A-E) Immunofluorescence assays (IFA) of parasites expressing the N-terminal leader of the gene of interest fused to GFP at the C-terminus (green). All genes showed cytoplasmic localization. (F) IFA of gene TgMe49_248450 endogenously tagged with an HA epitope tag at the C-terminus (green). Localization displays punctate sturctures, but does not colocalize with the apicoplast marker. Counterstaining for apicoplast marker CPN60 is shown in red for all panels. DAPI staining of the nucleus is in blue. No candidate showed colocalization with the CPN60 apicoplast marker.

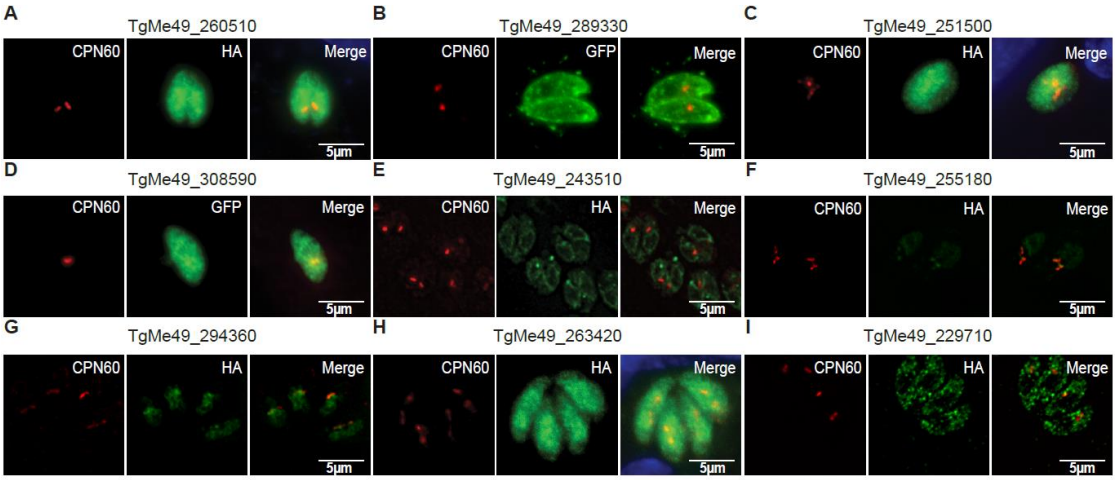


Figure 4.2. Tagging and staining of apicoplast specific deubiquitinase candidates. (A) Immunofluorescence assays (IFA) depicting parasites expressing the first 145 amino acids of gene TgMe49_260510 fused to an HA epitope tag (green). (B) IFA showing parasites expressing the N-terminal 284 amino acids of gene TgMe49_289330 fused to GFP (green). (C) IFA performed on parasites expressing the first 198 amino acids of gene TgMe49_251500 fused to GFP (green). (D) IFA performed on a parasite line expressing the full length cDNA of TgMe49_308590 fused to an GFP epitope tag (green). (E-I) IFAs depicting parasites with genes TgMe49_243510, TgMe49_294360, TgMe49_255180,

TgMe49_251500 and TgMe49_229710 endogenously tagged with an HA epitope tag (green). The IFA of the tagged version of TgMe49_243510 and TgMe49_229710 show putative ER localization. The staining of genes TgMe49_294360 and TgMe49_255180 tagged with an epitope tag both demonstrate nuclear localization. IFA of gene TgMe49_251500 displays a cytoplasmic localization. Counterstaining for apicoplast marker CPN60 is shown in red for all panels. DAPI staining of the nucleus is in blue. No candidate showed colocalization with the CPN60 apicoplast marker.

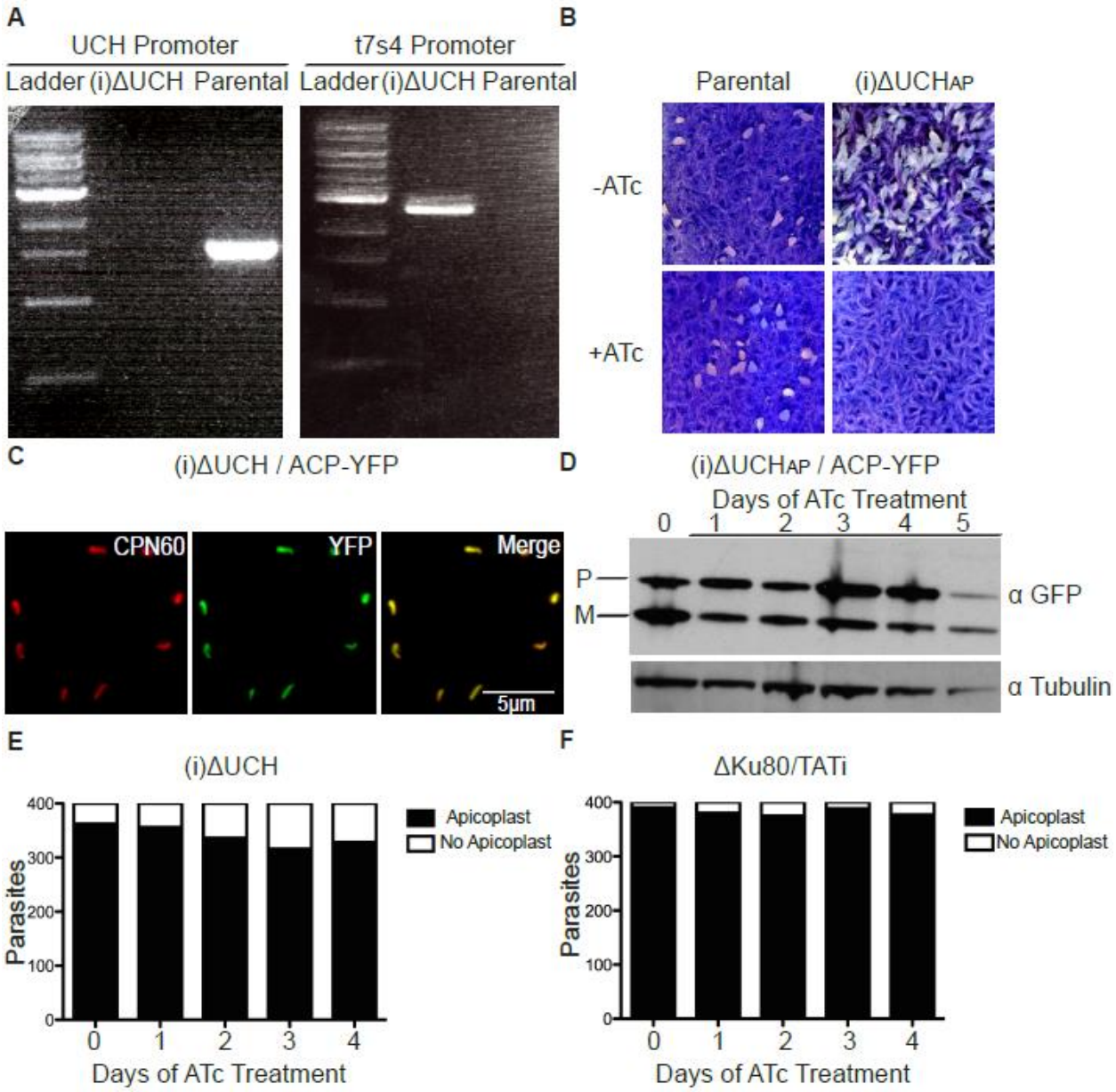


Figure 4.3. A deubiquitinase is essential for parasite survival but not apicoplast biogenesis. (A) Diagnostic PCR to validate the generation of the (i)ΔTgMe49_289330 or (i)ΔUCH line. PCR was performed to amplify 1.5 KB of the endogenous UCH promoter and 2.5 KB of the t7s4 promoter. (i)ΔUCH no longer amplifies the endogenous promoter, but amplifies the t7s4 promoter suggesting the t7s4 promoter has correctly replaced the

endogenous promoter. (B) Plaque assays were performed on the parental ΔKu80/TATi line and the (i)ΔUCH line in the absence or presence of 0.5 μg/ml of ATc. Plaques are not visible in the (i)ΔUCH line after ATc treatment suggesting that the gene is critical for the parasites. (C) Immunofluorescence assays performed on a stable monoclonal (i)ΔUCH line expressing ACP-YFP. ACP-YFP (green) is colocalized to the apicoplast marker CPN60 (red) suggesting that ACP-YFP is properly imported into the apicoplast. (D) An apicoplast import assay were performed with the (i)ΔUCH/ACP-YFP line. Parasites were treated with ATc from zero to five days, harvested for Western blot analysis and bands were visualized using a GFP antibody. There is no loss of the mature band (M) compared to the precursor band (P) after ATc treatment suggesting that the deubiquitinase does not facilitate apicoplast protein import. αTubulin serves as a loading control. (E) In the (i)ΔUCH line apicoplast numbers were scored daily under ATc treatment by IFA using anti-CPN60 to visualize the organelle. (F) The same assay was performed on the parental ΔKu80/TATi line. There is only a slight decrease in apicoplast numbers in the mutant line. The lack of a significant difference in apicoplast numbers after ATc treatment suggests that the gene has no apicoplast biogenesis defect.

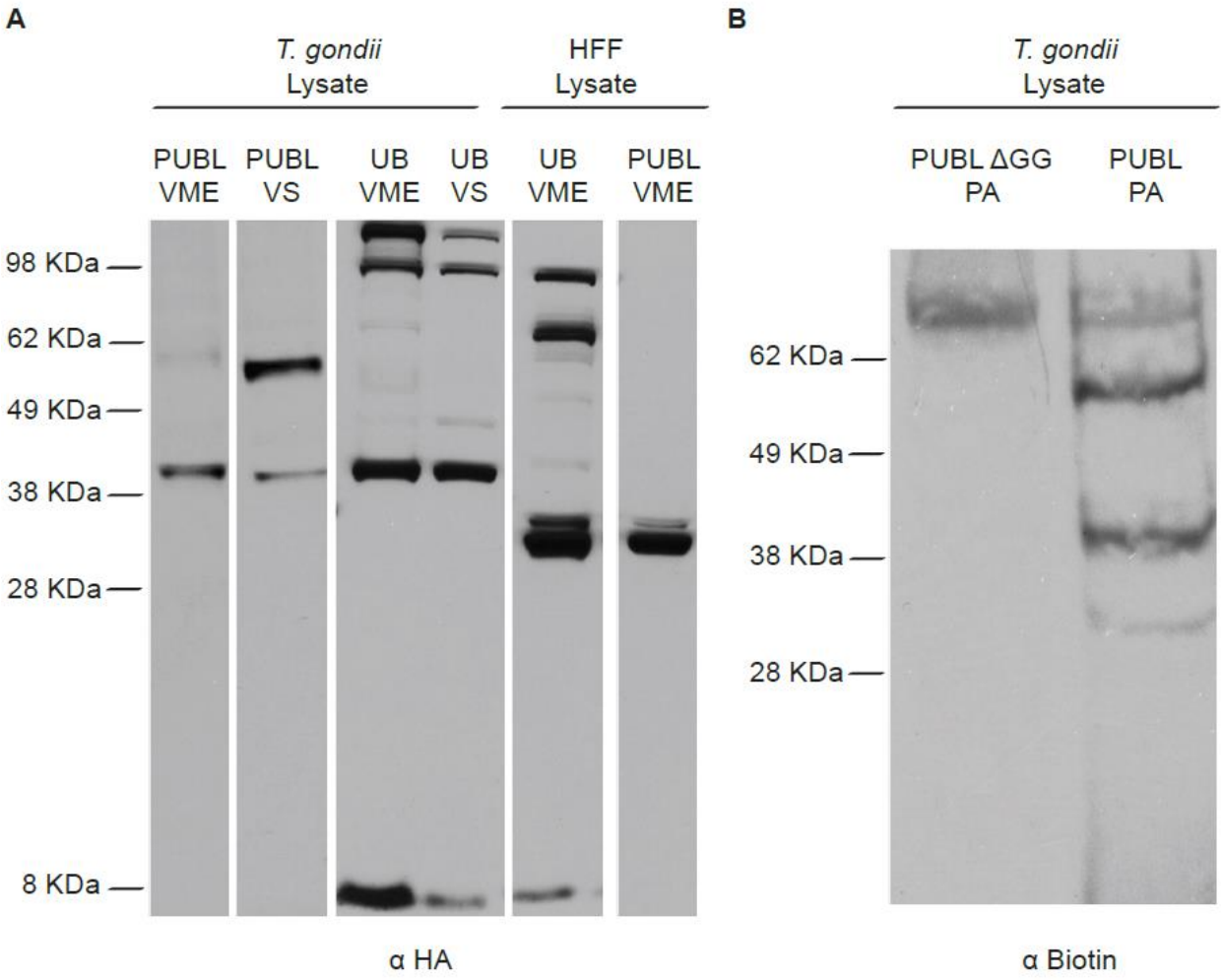


Figure 4.4. Two deubiquitinases interact with the plastid ubiquitin-like protein probes. (A) The ΔKu80/TATi *T. gondii* parasites and human foreskin fibroblasts (HFFs) were incubated with probes PUBL-VME, PUBL-VS, UB-VME and UB-VS for 30 minutes and then harvested for Western blot analysis. Two bands were seen in *T. gondii* lysates with the PUBL probe around ~40 KDa and ~55 KDa suggesting that there are two deubiquitinases in the *T. gondii* genome that are active against PUBL. Multiple bands were seen when *T. gondii* was incubated with the ubiquitin probes. The ~40 KDa band is seen in both ubiquitin and PUBL probes suggesting that the ~40 KDa deubiquitinase is promiscuous against ubiquitin-like proteins. The VME and VS electrophilic group at the C-terminus have different activity towards different deubiqutinases explaining the

differences in the banding patterns between the two different probes. The UB-VME and PUBL-VME probe were incubated with lysed HFF cells revealing several bands for the ubiquitin probe and two bands for the PUBL probe. The two bands seen for the PUBL probe are also visualized for the ubiquitin probe suggesting that those two deubiquitinases are promiscuous for ubiquitin-like proteins. Bands the size of 8.5 KDa seen in the Western blots represent ubiquitin or PUBL probes that have not reacted with any deubiquitinases. All probes contain an HA epitope tag allowing for Western blot analysis. (B) PUBL-PA and PUBLΔGG-PA probes were incubated with *T. gondii* lysates. Both probes contain a biotin tag at the N-terminus. The same ~40 KDa and ~55 KDa bands were seen with the PUBL-PA probe. A fainter band was also seen the size of ~35 KDa. These three bands were missing in the Western blot for the inactive PUBLΔGG-PA probe suggesting that these three bands are active deubiquitinases. The larger band seen in both Western blots is most likely the endogenous mitochondrial biotinylated pyruvate carboxylase which is predicted to be ~130 KDa.

Tables

Table 4.1. **Putative apicoplast specific ubiquitin ligases**

ToxoDB ID	Product Description	Signal Peptide	TM Domain	Phenotype Score	Localization
TgMe49_248450	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	5	0.28	Punctate Structure
TgMe49_277740	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	Yes	No	-2.11	Cytoplasm
TgMe49_209000	HECT Domain (Ubiquitin-Transferase) Domain-Containing Protein	Yes	No	-4.49	Cytoplasm
TgMe49_020570	Hypothetical Protein	No	1	0.68	Cytoplasm
TgMe49_258790	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	No	-3.52	Cytoplasm
TgMe49_295658	Zinc Finger In N-Recognin Protein	No	No	-1.25	Cytoplasm
TgMe49_304460	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	Yes	5	-4.44	No signal
TgMe49_294020	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	2	-2.09	N/A
TgMe49_236640	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	2	0.61	N/A
TgMe49_261990	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	1	-1.3	N/A
TgMe49_232160	Zinc Finger, C3HC4 Type (RING Finger)	No	4	-1.52	N/A

	Domain-Containing Protein				
TgMe49_294020	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	No	2	-2.09	N/A
TgMe49_245660	Hypothetical Protein	Yes	No	-2.59	N/A
TgMe49_223880	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	Yes	No	-2.25	N/A
TgMe49_289730	Pep3/Vps18/Deep Orange Family Protein	Yes	No	-3.24	N/A
TgMe49_230440	CHY Zinc Finger Domain-Containing Protein	No	2	-2.49	N/A
TgMe49_246160	Hypothetical Protein	No	6	-1.2	N/A
TgMe49_226740	Zinc Finger, C3HC4 Type (RING Finger) Domain-Containing Protein	Yes	3	1.4	N/A
TgMe49_289310	Cullin Family Protein	Yes	No	-4.44	N/A
TgMe49_300290	SNARE Domain-Containing Protein	No	1	-1.71	N/A

Table 4.2 Putative apicoplast specific deubiquitinases

ToxoDB ID	Product Description	Signal Peptide	Phenotype Score	Localization
TgMe49_289330	Ubiquitin Carboxyl-Terminal Hydrolase	Yes	-4.51	Plasma Membrane
TgMe49_308590	Mov34/MPN/PAD-1 Family Protein	Yes	-1.1	Cytoplasm
TgMe49_251500	Eukaryotic Initiation Factor-3, Subunit 3, Putative	Yes	-3.35	Cytoplasm
TgMe49_229710	OTU Family Cysteine Protease	Yes	-3.46	ER/Cytoplasm
TgMe49_260510	Ubiquitin Thioesterase Otubain-Like Family Protein	Yes	1.48	Cytoplasm
TgMe49_243510	OTU Family Cysteine Protease	No	-2.99	ER
TgMe49_294360	Ubiquitin Specific Protease 39 Isoform 2, Putative	No	-4.82	Nucleus
TgMe49_255180	Ubiquitin Carboxyl-Terminal Hydrolase	No	-1.71	Nucleus
TgMe49_263420	Ubiquitin-Specific Protease USP4	No	0.78	Cytoplasm
TgMe49_213870	UBA/TS-N Domain-Containing Protein	No	-4.83	N/A
TGME49_277895	Ubiquitin Carboxyl-Terminal Hydrolase	No	-4.79	N/A
TgMe49_243430	OTU Family Cysteine Protease	Yes	0.43	N/A
TgMe49_323200	OTU Family Cysteine Protease	Yes	N/A	N/A
TGME49_269250	Mov34/MPN/PAD-1 Family Protein	No	-4.99	N/A
TGME49_207650	OTU Family Cysteine Protease	No	-2.75	N/A

Table 4.3: Primers used for the study

Gene ID Used	Purpose	Name	Primer
TgMe49_248450	248450-HAx3	248450F	tacttccaatccaatttaatgcacctgagtgtgc gatctgc
TgMe49_248450	248450-HAx3	248450R	tcctccacttccaattttagcagggaatggagg caggg
TgMe49_277740	277740-GFP	277740F	gatcagatctatgcttcctgtagcttac
TgMe49_277740	277740-GFP	277740R	agtcctaggcacgattggctgcgtatagg
TgMe49_209000	209000-GFP	209000F	gatcagatctatgaggcctcgactcccg
TgMe49_209000	209000-GFP	209000R	agtcctaggacagttcctgggagaatctg
TgMe49_020570	020570-GFP	020570F	gatcagatctatgagattcagggaaaacga
TgMe49_020570	020570-GFP	020570R	agtcctaggagaaggcaggaaagaaaacg
TgMe49_258790	258790-GFP	258790F	gatagatctatggcatcctcgcacgg
TgMe49_258790	258790-GFP	258790R	agtcctaggggcgaccttgtctcct
TgMe49_295658	295658-GFP	295658F	gatcagatctatggatccttcctcatcg
TgMe49_295658	295658-GFP	295658R	agtcctagggagtctctgagtgacgaag
TgMe49_289330	289330-GFP	289330F	gatcagatctatgtctcgtcgcgagaga
TgMe49_289330	289330-GFP	289330R	agtcctaggctcttccagtccttcgagga
TgMe49_308590	308590-GFP	308590F	gatagatctatgccgccgactctgtcacttgaa gact
TgMe49_308590	308590-GFP	308590R	agtcctagggagccgtacggagcctgg
TgMe49_251500	251500-HAx3	251500F	gatggatccatggtcgagagcgacg
TgMe49_251500	251500-HAx3	251500R	agtcctaggggcgtctgcttctttcg
TgMe49_260510	260510-HAx3	260510F	gatagatctatgctgcgaagtatctcg
TgMe49_260510	260510-HAx3	260510R	agtcctaggcgtgctctggtgagc
TgMe49_243510	243510-HAx3	243510F	ggtaccctgtacttccaatccaatttaatcttgc gtctttttcgtgctct
TgMe49_243510	243510-HAx3	243510R	gaattcccgtcctccacttccacttccaattttac gattcagatacatacgc

TgMe49_294360	294360-HAx3	294360F	ggtaccctgtacttccaatccaatttaatgcgatttgtggtgtgtgtgt
TgMe49_294360	294360-HAx3	294360R	gaattcccgtcctccacttccaattttacgattcagatacatacgc
TgMe49_229710	229710-HAx3	229710F	tacttccaatccaatttaatgcacagggaagatgctcactc
TgMe49_229710	229710-HAx3	229710R	tcctccacttccaattttagccacgcacaccagcggcga
TgMe49_263420	263420-HAx3	263420F	ggtaccctgtacttccaatccaatttaatgaaactcggcaggtcaag
TgMe49_263420	263420-HAx3	263420R	gaattcccgtcctccacttccaattttaacacacaaggtgactgctaa
TgMe49_255180	255180-HAx3	255180F	ggtaccctgtacttccaatccaatttaatgagcaagacagaaggagc
TgMe49_255180	255180-HAx3	255180R	gaattcccgtcctccacttccaattttaagacggtgctgctcgatcatcct
TgMe49_304460	304460-HAx3	304460F	tacttccaatccaatttaatgcaaatcactggaccaacgtc
TgMe49_304460	304460-HAx3	304460R	tcctccacttccaattttagcgattgtccgctcctcttg
TgMe49_289330	UCH Promoter Replacement	289330F	tgttgcgctgctgcctctcttggtgtgcgaggaagggagttttcaaccaagaatggtaaccgacaaacgcgttc
TgMe49_289330	UCH Promoter Replacement	289330R	ggatcacttttgttcaactcatttgccccctttctctcgcgacgagacatagatctggttgaagacagacgaaagc
TgMe49_223125	Apiquitin pAVA	UBLicF	ccagggtcctggttcgttcaacttgcaagtcaagac
TgMe49_223125	Apiquitin PAVA	UBLicR	cttgttcgtgctgtttatcatgcgcctcctcgga

CHAPTER 5

CONCLUSION

5.1 Introduction

In this dissertation I set out to understand how nuclear encoded apicoplast proteins are transported across the four membranes of the apicoplast in *T. gondii*. Previous studies in *T. gondii* and other organisms described nuclear encoded apicoplast proteins requiring a bipartite leader peptide for targeting to the organelle. The TOC/TIC machinery has been characterized to regulate import across the two innermost membranes of the apicoplast similar to the membranes of the chloroplast. The molecular mechanism regulating protein import across the second outermost membrane or the periplastid membrane (PPM) was less established. At the commencement of this thesis only a $Der1_{AP}$ translocon and an ubiquitin conjugating enzyme have been demonstrated to reside in the periphery of the apicoplast and direct protein import into the organelle. Other ERAD derived proteins have been localized to the apicoplast, but their potential role in protein translocation across the PPM had not been characterized. Here we unravel the details of apicoplast import across the PPM by: 1) characterizing previously identified ERAD derived proteins localized to the apicoplast 2) identifying an ubiquitin-like modifier in the apicoplast and determining how the ubiquitin machinery has been retooled for import 3) identifying apicoplast specific enzymes that regulate ubiquitination or ubiquitin-like modifications.

5.2 CDC48$_{AP}$ is important for apicoplast biogenesis and protein import

CDC48$_{AP}$ was previously localized to the periplastid compartment of the apicoplast and is evolutionary derived from the ERAD machinery of the red algal endosymbiont. We analyzed the role of CDC48$_{AP}$ in the apicoplast by generating a conditional mutant for the gene. We effectively demonstrate that CDC48$_{AP}$ is essential for parasite survival and apicoplast biogenesis through plaque assays and measuring apicoplast numbers. We were also able to determine that CDC48$_{AP}$ is crucial for import of apicoplast proteins across the PPM by utilizing different apicoplast markers in an import assay. We propose a model in which CDC48$_{AP}$ is the motor of the import machinery that utilizes its ATPase domains to mechanically pull proteins through the PPM (Fig. 5.1A-B).

5.3 A plastid ubiquitin-like protein is essential and required for import across the PPM of the apicoplast

An ubiquitinating machinery had been previously identified in the apicoplast and demonstrated to be retooled for protein import rather than protein degradation. We were interested to examine exactly how this machinery acts in protein trafficking. We were able to identify the plastid specific ubiquitin-like protein (PUBL) which has been evasive for many years. PUBL is unique in that it doesn't contain the typical bipartite leader sequence, but instead has a recessed hydrophobic patch most likely involved in mediating its transport to the apicoplast. PUBL has the typical ubiquitin-like fold at its C-terminus, but PUBL is unlike any other ubiquitin-like protein. PUBL is 311 amino acids long with an ubiquitin domain at the C-terminus which is only 56% conserved to ubiquitin. It was also observed that PUBL is conserved amongst organisms containing a complex plastid with four membranes derived from a red algae engulfed during a secondary

endosymbiosis event. A conditional mutant for PUBL was generated which demonstrated that PUBL is essential for apicoplast biogenesis and protein import across the PPM of the organelle. Complementation assays revealed that PUBL most likely functions in a similar manner to ubiquitin due to the results showing that PUBL's diglycine motif is essential and PUBL can be replaced with ubiquitin with limited growth defects. The finding that complex plastids utilize a unique ubiquitin-like protein rather than ubiquitin in protein trafficking is revealing. Ubiquitin is involved in a multitude of processes and interacts with numerous proteins resulting in a lack of change in the amino acid sequence across organisms. We may be seeing the release of evolutionary constraints of the endosymbiont's ubiquitin resulting in novel ubiquitin-like protein PUBL. Overall, experimental results discussed here provides us with two models for PUBL's role in protein import across the PPM. One model postulates that PUBL is binding onto all imported proteins which mediates import by acting as a signal for $CDC48_{AP}$ and its cofactor $Ufd1_{AP}$ to pull proteins through the membrane (Fig. 5.1A). An alternative model is that PUBL is attached to the ubiquitin ligase which is a mechanism to control when proteins are transported across the PPM (Fig. 5.1B).

5.4 *T. gondii* encode several deubiquitinases and ubiquitin ligases

In this dissertation I sought to identify a plastid specific ubiquitin ligase and deubiquitinase. I classify the top candidates by searching for proteins with a signal peptide, a low phenotypic score from the CRISPR screen and homologs of experimentally confirmed plastid proteins. We generated parasite strains expressing tagged versions of these genes and confirmed their localizations through immunofluorescense assays. The staining revealed mostly cytoplasmic localization and

none of the tagged candidates fully localized to the apicoplast. We also generated a conditional mutant for a particularly strong deubiquitinase candidate (TgMe49_289330). The protein is essential for parasite viability, but the mutant did not display any apicoplast biogenesis defects suggesting that the deubiquitinase does not regulate PUBL in the apicoplast. Also the use of ubiquitin and PUBL probes showed that there is a difference in deubiquitinase recognition between the two ubiquitin-like proteins.

5.5 Future work

Although the study presented in this dissertation makes great strides in understanding the apicoplast protein import mechanism in *T. gondii*, the substrates of PUBL still remain a mystery. A potential reason for the inability to detect PUBL covalently attached to a substrate is that this process is very transient due to an active plastid deubiquitinase. However, a plastid specific deubiquitinase has not been identified in *T. gondii*. We are currently in the process of generating conditional mutants and tagging the remaining deubiquitinases candidates. Gene TgMe49_260510 is the leading candidate and previous attempts at endogenously tagging the gene resulted in no fluorescence signal. The lack of a fluorescence signal could be from an incorrect gene model or the epitope tag interfering with the protein. It would be advantageous to generate an antibody to avoid these potential pitfalls of tagging the gene with an epitope tag. We are simultaneously repeating the experiments utilizing PUBL probes that capture active deubiquitinases and pulling down the probes for mass spectrometry analysis. In the future we will also use deubiquitinase inhibitors to determine whether inhibitors result in the accumulation of PUBL onto substrates. Inhibitors like NEM are frequently used to block deubiquitination and result in the accumulation of ubiquitin on substrates. Another possible explanation is

that PUBL is only bound to the plastid specific ubiquitin ligase. Future experiments will be done to generate a mutant and localize the leading plastid ubiquitin ligase candidate (TgMe49_304460).

Another significant question going forward is whether $CDC48_{AP}$ interacts with PUBL in a similar fashion as CDC48 and its chaperones interact with polyubiquitinated substrates in the ERAD pathway. Antibodies for $CDC48_{AP}$ and PUBL have been generated in this study and previous studies giving us the ability to detect $CDC48_{AP}$ and PUBL through co-immunoprecipitation experiments. It is also feasible that the chaperone of $CDC48_{AP}$, $Ufd1_{AP}$, is interacting with PUBL. $Ufd1_{AP}$ has previously been identified in the periphery of the apicoplast in *T. gondii*, but $Ufd1_{AP}$ has not been further characterized. In the future we plan to generate a conditional mutant for $Ufd1_{AP}$ and test whether $Ufd1_{AP}$ is essential for apicoplast biogenesis and import. In conclusion, the research presented in this dissertation provides insight into the molecular mechanism regulating protein trafficking into the apicoplast of *T. gondii*.

Figures

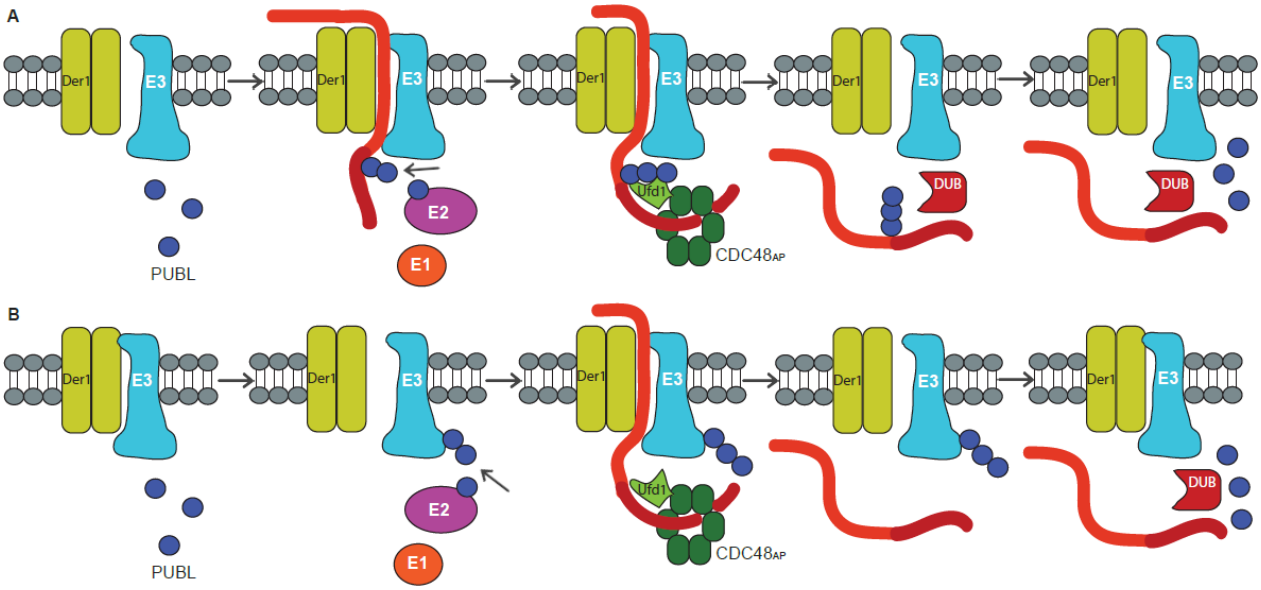


Figure 5.1 Model for protein import across the periplastid membrane. A) The enzymatic cascade of the ubiquitinating proteins ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2) and ubiquitin ligase (E3) covalently attach PUBL (in dark blue) to apicoplast proteins (red). Der1 and E3 are integral membrane proteins that most likely act as the translocon of the periplastid membrane (PPM) which proteins use to transverse the PPM. $CDC48_{AP}$ and its cofactor $Ufd1_{AP}$ recognize PUBL attached to the imported protein. $CDC48_{AP}$ provides the energy to mechanically pull proteins through the PPM. A deubiquitinase (DUB) cleaves PUBL freeing the protein for further import across the two innermost membranes. B) I present an alternative model for protein import across the PPM. PUBL is bound to E3 using the same enzymatic cascade of ubiquitinating proteins. PUBL attached to E3 opens up the Der1/E3 translocon allowing proteins to pass through the PPM. $CDC48_{AP}$ and $Ufd1_{AP}$ mediate protein import by $CDC48_{AP}$'s ATPase domains to pull proteins through the PPM. A DUB cleaves PUBL which results in the closing of the translocon.

APPENDIX 1

Toxoplasma gondii Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins[2]

[2] L. Sheiner, J.D. Fellows, J. Ovciarikova, C.F. Brooks, S. Agrawal, Z.C. Holmes, I. Bietz, N. Flinner, S. Heiny, O. Mirus, J.M. Przyborski and B. Striepen. 2015. In *Traffic* Volume 16. 1254-1269.

Abstract

Apicomplexa are unicellular parasites causing important human and animal diseases, including malaria and toxoplasmosis. Most of these pathogens possess a relict but essential plastid, the apicoplast. The apicoplast was acquired by secondary endosymbiosis between a red alga and a flagellated eukaryotic protist. As a result the apicoplast is surrounded by four membranes. This complex structure necessitates a system of transport signals and translocons allowing nuclear encoded proteins to find their way to specific apicoplast sub-compartments. Previous studies identified translocons traversing two of the four apicoplast membranes. Here we provide functional support for the role of an apicomplexan Toc75 homolog in apicoplast protein transport. We identify two apicomplexan genes encoding Toc75 and Sam50, both members of the Omp85 protein family. We localize the respective proteins to the apicoplast and the mitochondrion of *Toxoplasma* and *Plasmodium*. We show that the *Toxoplasma* Toc75 is essential for parasite growth and that its depletion results in a rapid defect in the import of apicoplast stromal proteins while the import of proteins of the outer compartments is affected only as the secondary consequence of organelle loss. These observations along with the homology to Toc75 suggest a potential role in transport through the second innermost membrane.

A1.1 Introduction

Apicomplexan parasites are the cause of important human and animal diseases, including malaria and toxoplasmosis. Most of these pathogens possess a relict plastid named the apicoplast. While the apicoplast is no longer photosynthetic, it has important metabolic roles and supplies the parasite with fatty acids, isoprenoids and heme [1, 2]. The

apicoplast is the product of secondary endosymbiosis whereby a single celled red alga was engulfed by a flagellated eukaryote and a stable endosymbiotic relation ensued. This event gave rise to a large and diverse group of photosynthetic and non-photosynthetic eukaryotes referred to by some authors as the chromalveolates [3, 4]. The apicoplast and the plastids of other chromalveolates are surrounded by four membranes reflecting their complex endosymbiotic origin. The innermost membrane and second innermost membrane (SIMM) originate from the algal primary plastid. The next membrane out, bounding the periplastid compartment (PPC), originates from the algal plasma membrane and the outermost membrane is believed to be derived from the host endomembrane system [5].

Key to the conversion of the algal endosymbiont into a plastid was the transfer of the symbiont's genes to the nucleus of the host, allowing far reaching transcriptional and translational control by the host. This transfer of genetic material from the endosymbiont to the host is only possible upon coevolving systems that allow the import of host-translated proteins into the endosymbiont. In the case of the apicoplast this requires translocation across its four delineating membranes to reach the stroma. Apicomplexan parasites target large numbers of nuclear encoded proteins to the apicoplast. Ten percent of the *Plasmodium falciparum* proteome is predicted to be transported to the apicoplast, underscoring the importance of this trafficking pathway [6]. Our current model (Figure A1.1A) assumes this pathway to start with signal sequence guided entry into the ER lumen, likely via the Sec61 translocon. Trafficking from the ER to and across the outer apicoplast membrane remains poorly understood, but potentially depends on signals typically involved in endocytosis or autophagy [7-9] and may take place by more than

one route [10, 11]. Translocation across the periplastid membrane is mediated by machinery evolved from the endosymbiont's ER-associated protein degradation (ERAD) system [12-16]. Finally, based on their evolutionary origin in chloroplast membranes, it is believed that homologs of the translocons of the inner and outer chloroplast membrane (TIC/TOC) function in translocation of proteins through the apicoplast's two innermost membranes. Experimental evidence supports the role of the TIC complex in apicoplast protein import [17, 18] but is lacking in the case of the putative TOC machinery.

Most stromal proteins possess a bipartite signal, comprised of a signal and a transit peptide [6]. Upon translocation to the ER lumen the N-terminal signal peptide portion of the leader is cleaved off, exposing the transit peptide that is required for further trafficking [19]. In diatoms, a group likely descended from the same endosymbiotic event as Apicomplexa, subplastidal targeting depends on the first amino acid (position +1) of the transit peptide [20, 21]. An aromatic amino acid at this position targets the protein through the SIMM *en route* to the stroma; otherwise, the proteins are retained in the periplastid space. Incidentally, an aromatic residue is also required for import through the outer membrane of primary plastids of red algae [20, 22, 23] and of glaucocystophytes [24]. In the primary plastids of glaucocystophytes, recognition of the aromatic residue depends on an outer-membrane proteins of 85 kDa (Omp85) family member that functions as the translocation pore of their primitive TOC machinery [25].

Abundance of aromatic residues at position +1 was reported for the transit peptides of additional Chromalveolates [26, 27]. These studies include *Toxoplasma* and *Plasmodium* spp where enrichment of phenylalanine was reported at this position [27]. Nevertheless the role of this amino acid was so far not supported experimentally. The

targeting sequence of a *Toxoplasma* apicoplast stromal protein, ferredoxin-NADP+ reductase (FNR), was studied in detail [30]. An extensive series of deletion in the N-terminal sequence suggested the presence of two targeting signal and the data did not align with involvement of the residue at position +1 [28]. Whether this is true for all apicoplast stromal proteins remains unaddressed.

Omp85 is a protein that catalyzes insertion and assembly of β-barrel proteins into the outer membrane of gram-negative bacteria. The more widely distributed superfamily of Omp85-related proteins shares a conserved domain organization that includes N-terminal polypeptide-transport-associated (POTRA) repeats and a C-terminal transmembrane β-barrel. Three main eukaryotic representatives are well described: the mitochondrial sorting and assembly machinery of 50 kDa (Sam50/Tob55) and the chloroplast proteins Toc75-III and Toc75-V [29, 30]. Like its bacterial ancestor, Sam50/Tob55 recognizes mitochondrial outer-membrane proteins in the intermembrane space after they fully translocate across the outer membrane and catalyzes their insertion into it [31, 32]. Toc75V (or Oep80 for outer envelope protein 80) is hypothesized to perform a similar role in the outer chloroplast membrane [33-35]. Toc75III functions as the channel of the TOC translocon in the outer chloroplast membrane that allows proteins to fully translocate through it [36].

In diatoms, an Omp85-like protein, *Pt*Omp85, was identified that possesses a bipartite plastid targeting signal and two POTRA domains [37]. This protein is localized to the diatom complex plastid and both its N and C terminal domain face the periplastid compartment [37]. Using the sequence of *Pt*Omp85, Bullmann and co-workers were able to identify putative apicomplexan homologs [37, 38], and this identification gained

further support from Hirakawa et al. [39]. These homologs possess features supporting their Omp85 affiliation such as a signal sequence, the typical N-terminal POTRA signature, and a C-terminus that likely forms a beta-barrel. However, their putative role in apicoplast protein import has not been evaluated experimentally.

Here we seek to gain new insights about the pathways by which apicoplast proteins traverse the SIMM. We analyze the targeting sequences of a large group of experimentally confirmed apicoplast proteins (summarized in [40]). Our findings suggest that an aromatic residue at position +1 is more abundant in stromal proteins compared to peripheral proteins. Additionally, a mutation of the phenylalanine at position +1 of the transit peptide of the stromal Acyl Carrier Protein (ACP) to an alanine results in peripheral retention. Nevertheless, it is clear that an aromatic +1 is not a uniform feature and other signals likely play a role in the apicoplast sub-compartmental sorting. We confirm the identity and localization of Omp85 proteins from both *P. falciparum* and *Toxoplasma gondii* and demonstrate that the *Toxoplasma* Toc75 functions in the import of proteins into the stroma of the apicoplast. Finally, we show that import of peripheral apicoplast protein is not dependent on TgToc75 activity, which is consistent with the potential assignment of TgToc75 to the second innermost of the four apicoplast membranes.

A1.2 Results

Sequence analysis reveals enrichment of aromatic residues at position +1 of stromal proteins and the presence of two omp85-like proteins in apicomplexan genomes

We used sequence analysis to identify signals and machinery potentially involved in traversal of the apicoplast SIMM membrane. We have recently substantially expanded

the repertoire of experimentally confirmed apicoplast proteins in *T. gondii* [40] now counting 47 proteins (Table A1.S1). We utilized the online prediction algorithm SignalP (http://www.cbs.dtu.dk/services/SignalP‐3.0/) to predict the signal peptide cleavage site of all 47 proteins. Using SignalP 3.0 server we were able to define with high certainty the amino acid at position +1 of the transit peptide of 29 of these proteins (Table A1.S1 shows the predictions obtained with both SignalP servers: 3.0 and 4.1). Figure A1.1B shows the distribution of +1 residue abundance in (i) 22 stromal and (ii) 7 peripheral proteins. We found that 27% of stromal proteins have an aromatic residue (mostly a phenylalanine) at the predicted position +1 while none of the non-stromal proteins feature an aromatic amino acid at this position (Table A1.S1, Figure A1.1B).

Next, we revisited the repertoire of potential apicomplexan Omp85-like encoding genes. Using jackhmmer to mine the NCBI non-redundant database, and subsequent reciprocal BLAST searches against the EupathDB, we identified two Omp85-like proteins in *T. gondii* (TGME49_205570, TGME49_272390), *P. falciparum* (PF3D7_0608310, PF3D7_1234600), and several other apicomplexan species (Table A1.1).

To determine the respective affiliation of these genes, we selected representative species across the eukaryotic tree of life and reconstructed a majority rule consensus tree from 1000 bootstrap trees (Figure A1.2; see also maximum likelihood tree Figure A1.S1A, Supporting Information). The tree shows a clear split into Sam50 and Toc75 clades supported by a bootstrap of 100. We classified sequences TGME49_205570 and PF3D7_0608310 as Sam50 (herein named TgSam50 and PfSam50, respectively) and sequence TGME49_272390 as Toc75 (named TgToc75). PF3D7_1234600 could not be

resolved with certainty in this analysis, however subsequent analysis included PF3D7_1234600 (Figure A1.S1B) and used the POTRA region only (Figure A1.S1C) to construct a maximum likelihood tree which shows that PF3D7_1234600 is affiliated with the Toc75 homologs from Chromalveolates (herein named PfToc75). The presence of two POTRA domains in PfToc75 and TgToc75 and a predicted apicoplast targeting signal in PfToc75 support this affiliation.

Mutagenesis of a Phenylalanine at position +1 of the transit peptide of the stromal protein ACP to alanine results in peripheral retention

The putative role of the aromatic residue at position +1 of the stromal protein ACP in trafficking was teased via mutagenesis. YFP-tagged ACP with the wild type sequence ($ACP_{WT}YFP$) and YFP-tagged ACP with the phenylalanine replaced to an alanine ($ACP_{F/A}YFP$) were transiently transfected and their localization was assessed by high-resolution microscopy. While $ACP_{WT}YFP$ showed precise colocalization with the stromal marker CPN60 [12, 40], $ACP_{F/A}YFP$ showed very little overlap with it (Figure A1.1C). Similarly, the signal from $ACP_{WT}YFP$ did not overlap with signal from the HA-tagged periplastid marker ATrx2 [40], while $ACP_{F/A}YFP$ showed substantial colocalization with this periplastid marker (Figure A1.1C).

The signal peptide cleavage prediction by SignalP 3.0 differs from that obtained by SignalP 4.1. While both suggest the phenylalanine at position +1 with high likelihood, the latter predicts an upstream tyrosine to be at this position. We generated YFP-tagged ACP with the tyrosine replaced to an alanine ($ACP_{Y/A}YFP$) and examined its localization upon transient expression by high-resolution microscopy. Similar to wild type ACP,

$ACP_{Y/A}$YFP showed full colocalization with the stromal marker CPN60 and little overlap with the periplastid marker ATrx2 (Figure A1.1C).

Localization of the *T. gondii* and *P. falciparum* Omp85 proteins to the apicoplast and mitochondrion supports their assignments as Toc75 and Sam50

The assignment of Omp85 proteins to their respective families as determined by the phylogeny was tested by localization studies. Both the first 78, and the first 95 N-terminal amino acids derived from both TgToc75 (Figure A1.S2) and PfToc75 (Figure A1.3A) target to a punctate structure within the parasite that colocalized with the *Toxoplasma* or *Plasmodium* apicoplast markers FNR-RFP or ACP, respectively. We conclude that these N-termini serve in apicoplast localization of these proteins. Moreover, full-length TgToc75 also colocalizes with the apicoplast marker FNR-RFP further supporting the Toc75 affiliation (Figure A1.3A). High-resolution microscopy and co-staining with the stromal marker CPN60 suggested TgToc75 localization is peripheral to the apicoplast stroma (Figure A1.3B), In line with the expected peripheral localization of a Toc75 homolog.

We next assessed the localization of the second Omp85 homolog identified in each of the species. A mitochondrial-targeting signal was predicted for PfSam50 but not for TgSam50 (Table A1.S1). The first 60 amino acids of PfSam50 targeted GFP to a ribbon-like structure within *P. falciparum* parasites that colocalized with the signal obtained through the use of MitoTracker (Figure A1.3C). Likewise, full-length TgSam50 colocalized with the mitochondrial marker Hsp60-RFP (Figure A1.3C). Finally, co-transfection of both full-length HA-tagged TgToc75 and full-length Ty-tagged TgSam50 in *T. gondii* reveals two distinct patterns of fluorescence with minimal signal overlap.

This demonstrates the existence of two Omp85-like proteins in *T. gondii* in the two distinct endosymbiotic compartments; the apicoplast and the mitochondrion (Figure A1.3D). Taken together, our localization experiments are entirely consistent with the classification proposed by phylogenetic analysis.

TgToc75 is required for parasite growth and apicoplast maintenance

To test whether TgToc75 functions in apicoplast protein import we generated a mutant in which its expression level can be manipulated. First we constructed the *TATiΔKu80iToc75pi* line: in this parasite the TgToc75 open reading frame is separated from its native promoter by a tetracycline-regulatable promoter cassette [40]. This parasite line was established using cosmid (PSBL491) recombineering [41, 42] (Figure A1.S3). Our analysis of this line suggested that TgToc75 is essential for parasite growth (Figure A1.S3), and that its downregulation results in apicoplast demise and in a stromal protein modification defect (Figure A1.S3), as expected from interference in the apicoplast protein import machinery [12, 18, 40]. However, this mutant proved unstable resulting in loss of regulation. We thus utilized recombineering to construct the *TATiΔKu80iToc75pr* line: in this line we replaced the TgToc75 native promoter with the tetracycline-regulatable promoter cassette (Figure A1.4A). This line is stable and was used for the remainder of the analyses. We found downregulation of TgToc75 (Figure A1.4B) to result in a growth defect as observed by plaque assay (Figure A1.4C). Additionally, TgToc75 depletion resulted in loss of the apicoplast evident by loss of plastid DNA, which was quantified via quantitative PCR, as well as by loss of immunofluorescence staining of the apicoplast stromal protein CPN60 (Figure A1.4D,E).

Organelle loss gradually started with 28% loss at 24 h of Toc75 downregulation and 99.5% loss by 96 h.

Loss of TgToc75 results in loss of import with more rapid impact on stromal when compared to peripheral apicoplast proteins

To examine apicoplast protein import under TgToc75 downregulation we followed the maturation of the plastid stromal protein ACP [12, 18, 40]. Typically two bands can be observed for this protein by western blot, a larger precursor protein *en route* to the plastid, and a mature protein lacking the leader peptide due to the activity of stromal signal peptidase [18, 19, 43, 44]. By following endogenously-tagged ACP [40] we detected accumulation of un-cleaved precursor starting at 48 h after Toc75 downregulation (Figure A1.5A). Interestingly, the precursor of the protein encoded by TGME49_001270, an outer apicoplast membrane protein [40], does not accumulate even as late as 72 h (Figure A1.5B). To assess whether this difference is specific to TgToc75 depletion we conducted control experiments with a regulated mutant of the periplastid protein 1 (PPP1). PPP1 is a periplastid compartment resident protein that plays an essential role in apicoplast protein import and that is likely required for the translocation of proteins across the periplastid membrane [40]. Here we show that upon downregulation of PPP1 both the stromal ACP and the outer-membrane protein encoded by TGME49_001270 show precursor accumulation ([40] and Figure A1.5C,D). We conclude that proteins pass through the PPP1 associated translocon first and the Toc75 translocon second and that the outer translocon can act and assemble (at least for a limited time) independently of Toc75.

To test whether these observations hold true for other stromal and non-stromal proteins we examined two additional markers, the stromal protein LytB [45] and the

periplastid protein PPP1. In order to follow protein maturation in real time we measured maturation of LytB and PPP1 expressed transiently at different time points after TgToc75 downregulation. In agreement with the above observations, newly synthesized stromal LytB, shows precursor accumulation starting as early as 24 h after TgToc75 downregulation (Figure A1.5E), while the newly synthesized periplastid protein PPP1 shows precursor accumulation only late into suppression (72 h, Figure A1.5F) when many apicoplasts are lost due to secondary effects (Figure A1.4D,E).

A1.3 Discussion

The acquisition of secondary plastids went hand in hand with the development of appropriate machineries for protein import [3]. The complex nature of these plastids requires a set of signals allowing precursor protein trafficking to their final sub-organellar destination. An elevated abundance of aromatic amino acids, particularly phenylalanine, at position +1 downstream of the predicted signal peptide cleavage site, was reported in several chromealveolates and was proposed to be a functional feature of the transit peptide in these organisms [26, 46]. A signature residue, most frequently a phenylalanine (but also tyrosine and tryptophan), at position +1 of the transit peptide, was proposed to serve as forward signaling from the periplastid space through the two innermost membranes in several groups of organisms with secondary plastids. A similar requirement is found for import into the primary plastids of red algae [20, 22, 23, 47]. Gould et al. suggested a model according to which all proteins targeted to a complex plastid of red origin gain entry to the periplastid compartment by a common indiscriminate mechanism [47]. We analyzed 45 *T. gondii* sequences of proteins experimentally shown to target to the apicoplast periphery or the apicoplast stroma. Of

those we could assess 29 proteins with high certainty. This analysis does not align with the notion of a uniform mechanism. On one hand we show enrichment of an aromatic residue at position +1 in the putative transit peptides of proteins that cross all four apicoplast membranes (Table A1.S1, Figure A1.1B). Further, our mutagenesis experiments support the notion that this aromatic +1 plays a role in the targeting of the stromal ACP (Figure A1.1C). However, on the other hand, not all stromal proteins obey this rule. In fact, the vast majority (73%) of stromal proteins we are not predicted to have a +1 aromatic residue, supporting an alternative or an additional explanation to the signals involved in SIMM traversal. Indeed, in the case of FNR, for which most computational analyzes ([28] + Table A1.S1) do not predict an aromatic residue at position +1, other signals were shown responsible for stromal localization [28]. We also observed the lack of aromatic residues at position +1 of peripheral proteins, however, the repertoire of well documented residents of these outer compartments is still limited (only 7 predicted with certainty). Overall we provide preliminary observations that are consistent with the previously proposed [20, 21, 28] broader conservation of the +1 aromatic signal as one of the mechanisms for stromal import. However, further experiments of mutagenesis and deletion of a substantial number of proteins, and assessment of their direct interaction with the translocon, will shed further light on this topic.

The secondary plastid of Apicomplexa and related taxa was shaped by contributions from three organisms: a cyanobacterium, a red alga and a flagellated heterotrophic eukaryote. The current model of protein import suggests that each membrane is traversed with the help of machinery derived from its organism of origin.

This model gradually gained support with the identification and functional characterization of TIC components [18, 48] and of ERAD/SELMA components [12, 13, 16]. The confirmation of the TOC link in this model was slow to emerge, most likely due to significant primary sequence divergence of the TOC components in organisms with complex plastids. An important breakthrough was made by the identification of an Omp85-like protein in the diatom *Phaeodactylum tricornutum*, for which phylogeny, subcellular localization and electrophysiology support affiliation with Toc75 [49]. Here we provide experimental support for the general conservation of this transport pathway by localization of the apicomplexan homologs of PtToc75 to the apicoplast (Figure A1.3) and by functional analysis of TgToc75 (Figures A1.4 and A1.5).

Aside from TgToc75/PfToc75, our search for members of the polypeptide-transporting β-barrel protein superfamily in the genomes of Apicomplexa identified only one additional gene in each species, which encodes a Sam50/Tob55 homolog. We supported this assignment by localizing these proteins to the mitochondrion (Figure A1.3B). TgToc75/PfToc75 thus likely represent the only plastid Omp85 in these parasites, an observation that joins a growing line of evidence for a single Toc75 in the red lineage of plastids. The genome of the red alga *Cyanidioschyzon merolae* encodes a single Toc75 [50]. Bullmann et al. similarly report a single Toc75 in their analysis of the genomes of the diatoms *P. tricornutum* and *Thalassiosira pseudonana*, and the haptophyte *Emiliania huxleyi* [49]. In contrast, higher plants possess two functional Toc75 homologs: Toc75V/Oep80, which mediates assembly of proteins into the outer membrane of the chloroplast [51], and Toc75-III [36], which is the central component of the TOC machinery. At least two plastidial Toc75 proteins were identified in other members of the

green lineage, and in all cases at least one ortholog of Toc75V/Oep85 was identified [50]. Whether the Toc75 found in the red lineage serves the roles of both of its green algal counterparts is unclear at this point.

Others [49] and we herein hypothesize that TgToc75 plays a role in precursor transit through the SIMM. To test TgToc75's involvement in apicoplast protein import we generated a conditional *TgToc75* mutant parasite cell using our recently described tetracycline-based promoter replacement system [40]. We demonstrated *TgToc75* to be a firm requirement for apicoplast protein import, apicoplast maintenance and parasite growth consistent with the hypothesis that this protein is an essential component of the apicoplast protein import machinery.

In agreement with a role for TgToc75 in stromal protein import, we observed a defect in precursor processing for the endogenously YFP-tagged stromal protein ACP (Figure A1.5A). The slow onset of this defect may reflect an overall slow impact of Toc75 mutants as previously noted in primary chloroplast [52], or could result from the long half-life of mature ACP as noted before [18, 40]. We therefore tested an independent stromal protein (LytB) by transient transfection to follow the protein synthesized at various time points after Toc75 downregulation was ongoing. This assay showed impaired precursor processing as early as 24 h after TgToc75 downregulation (Figure A1.5E) and before secondary defects due to loss of the organelle (Figure A1.4D,E). Overall the TgToc75 mutant produces a phenotype similar to previously studied inducible mutants in components of the apicoplast protein import machinery [12, 18, 40, 48] supporting the proposed role of TgToc75 in mediating stromal precursor protein import.

Interestingly, unlike the stromal proteins, only a mild processing defect was observed for outer compartment proteins (Figure A1.5B,F). This is specific to TgToc75 depletion, as processing defect is observed for an outer compartment protein upon disruption of the periplastid import machinery (Figure A1.5D). These experiments support the notion that proteins of the apicoplast outer compartments (periplastid and outer-membrane compartments) are not dependent on TgToc75 for their transport into the organelle while stromal proteins (ACP and LytB) are. Taken together with ours and previous phylogenetic analyses these observations support the likelihood that TgToc75 is a component of the apicoplast TOC channel, however, direct experimental demonstration for its activity at the SIMM is yet to be established.

Finally, seeing that outer compartment protein processing occurs under depletion of TgToc75, our findings predict the existence of at least two apicoplast transit peptide peptidases: one in the lumen and one in the outer compartments upstream of the TOC machinery.

While we provide functional support for the role of Toc75 in protein import into complex plastids of red origin, we were unable to identify other components of the TOC machinery in the genomes of Apicomplexa by using BLAST searches, in line with previous reports [48, 50]. Most striking is the apparent absence of homologs for the receptor components Toc159/Toc120/Toc132 and Toc34/Toc33 [53] that are found in primary plastids of both the green and the red lineage [50, 54]. Interestingly, a similar finding was recently reported for the secondary plastid of green origin of *Bigelowiella natans* [41]. Hirakawa et al. suggest an explanation whereby unlike primary plastids where the TOC machinery has to distinguish plastid proteins from all other cytoplasmic

and mitochondrial proteins, the TOC machinery of secondary plastids interacts with a more focused repertoire of precursors that was already screened by previous translocation machineries. This notion is supported by the identification that secondary plastid transit peptide lack features that differentiate between mitochondria and plastid targeting in organisms with primary plastids [27]. In agreement with this model it was proposed before that in membranes with a primitive, reduced TOC machinery, the Omp85-like component is involved in precursor selection that is based on the presence of an aromatic residue [25]. While it is clear that apicoplast stromal import could not be explained by this simple model ([28], Table A1.S1), our finding provide grounds for further investigation of the potential role of such a pathway in the trafficking of at least some of the stromal proteins.

One of the soluble components that interact with the TOC machinery is Tic22 [55]. TgTic22 was identified and functionally characterized using a similar genetic system [48]. TgTic22 downregulation results in a phenotype similar to our observations here, whereby the maturation of a stromal marker (FNR-DHFR-cMyc) was reduced at 24 h after addition of Anhydrous tetracycline (ATc, [48]), supporting their potential cooperation in a common pathway. Interestingly, Toc75 and Tic22 are the sole TOC components found so far in secondary red plastids. They are also the only TOC components for which a clear homology with their cyanobacterial ancestors was demonstrated [55, 56].

A1.4 Materials and methods

Search for Omp85 homologs

The non-redundant protein database was downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/) and screened with jackhmmer [57] using AtToc75-III as query for members of the Omp85 superfamily. Then *T. gondii* ME49 and *P. falciparum* 3D7 genomes were screened (i) with BLAST for homologs of *Toxoplasma* and *Plasmodium* sequences detected by jackhmmer and (ii) with hmmsearch [57] for proteins with at least one of the following PFAM [58] domains: Surface_Ag_VNR (PF07244), Bac_surface_Ag (PF01103), POTRA_2 (PF08479), and ShlB (PF03865). Finally the resulting *Plasmodium* and *Toxoplasma* omp85 homologs were used as query for BLASTs of EupathDB (http://eupathdb.org/eupathdb/) to identify the respective homologs in other Apicomplexan genomes.

Phylogenetic analysis

A multiple sequence consensus alignment was constructed as described in [59] from a subset of Sam50 and Toc75 homologs. From this alignment a maximum likelihood phylogeny was reconstructed with RAxML [60] using the WAG model [61] and gamma-distributed rate heterogeneity. Branch support values were derived from 1000 rapid bootstrap trees and a majority rule consensus tree was constructed from them. Note that the older gene model for TgToc75, TGME49_072390, predicts a longer protein, which includes the extreme C-terminal part of the β-barrel (missing in the new gene model: TGME49_272390). The alignment and phylogeny were done with the longer older gene model. Similarly, an older gene model of PfSam50, PFF0410w, spans two new predicted genes: PF3D7_0608300/0608310. We experimentally confirmed the older gene model

(see Supporting Information). The alignment and phylogeny were performed using the new and shorter version PF3D7_0608310, containing the conserved domain.

Constructs

Toxoplasma gondii

Total RNA was extracted from *T. gondii* (strain RH) using Trizol (Invitrogen). Overlapping cDNA fragments encoding the entire TgToc75 and TgSam50 genes were amplified from total RNA using the SuperScript III One-Step RT-PCR kit (Invitrogen) and primers shown in Table S2. All resulting PCR products were cloned using the ZeroBlunt PCR cloning kit (Invitrogen) and sequenced (GATC, Konstanz). TgToc75^{78}, TgToc75^{95}: Fragments encoding the noted amino acids were amplified from cDNA (primers in Table S2), and inserted into the TUB8mycGFPMyoATy vector resulting in expression of these N-terminal amino acids directly fused to a Ty tag. Using *Eco*RI/*Nsi*I allowed for an in frame C-terminal Ty tag. TgToc75$^{full\text{-}Ty}$: A full-length cDNA version (based on the older gene model – TGME49_072390) of the *TgToc75* gene (removing an internal *Eco*RI restriction site) was synthesized by Geneart (Regensburg) and cloned into the TUB8mycGFPMyoATy vector as above. TgToc75$^{full\text{-}HA}$: TgToc75$^{full\text{-}Ty}$ was digested with *Nsi*I/*Pac*I and a 3× hemagglutinin (HA) tag was inserted, having been generated by amplification (primers in Table S2). TgSam50$^{full\text{-}HA}$: A full-length cDNA version of the *TgSam50* gene was synthesized by Geneart (Regensburg) and cloned into the TgToc75$^{full\text{-}HA}$ vector using *Eco*RI/*Nsi*I.

Inducible knock-down cosmids

pGDT7S4 [62] was used as templates to PCR amplify a 4-Kb promoter modification cassette (primers in Table S2) containing a gentamycin resistance marker for selection in

bacteria, a DHFR marker for the subsequent selection in *T. gondii* and the T7S4 promoter to be inserted upstream of TgToc75 start site (*pi*) or to replace the TgToc75 endogenous promoter sequence (*pr*). This cassette was used for PSBL491 recombineering as done before [41].

Site-directed mutagenesis

To mutate the putative F or Y at position +1 of the transit peptide of ACP within the plasmids pTUB8ACP$_{WT}$YFP, primers (ACP$_{F/A}$mutR or ACP$_{Y/A}$mutF/R) (Table S2) were used in a site-directed mutagenesis reaction using the commercial QuikChange II Site-Directed Mutagenesis Kit (Stratagen) and according to manufacturers' instructions.

Plasmodium falciparum

Total RNA was extracted from *P. falciparum* (3D7) using Trizol (Invitrogen). PfToc75[78], PfToc[95] and PfSam50[60]: Fragments encoding the noted amino acids were amplified from total RNA (primers in Table S2), and inserted into the *Xho*I/*Avr*II sites of pARL2-GFP. All final constructs were verified by restriction digest and automated sequencing (GATC, Konstanz).

Cell culture and transfection of T. gondii and P. falciparum

Cultivation and transfections of *T. gondii* (strain RH delta hxgprt, a kind gift of Markus Meissner, and our TATi/ΔKu80strain, (published in [40]) in human foreskin fibroblasts and *P. falciparum* (3D7) in human erythrocytes was carried out under standard conditions. *P. falciparum* transfectants were selected with 2.5 nm WR99210 (a kind gift of Jacobus Pharmaceuticals). Promoter replacement or insertion in TATi/ΔKu80strain was selected with 1 µm Pyrimethamin as described in [62]. FNR^{RFP} and Hsp60^{RFP} constructs were a kind gift of Markus Meissner.

Plaque assay

Fresh monolayers of HFF were infected with parasites in the presence or absence of 1.5 µg/mL ATc for 7 days. Fixation, staining and visualization were performed as previously described [40].

RT-PCR and qPCR

RNA was prepared from cultures grown without ATc or with ATc for 24 and 72 h using RNeasy® (QIAGEN) and reverse transcriptase reaction was performed using SuperScript® III First-Strand Synthesis (Invitrogen) (both according to the manufacturer's instructions). About 300 ng of the resulting template was used for qPCR reaction using SYBR Green Mix (Bio-Rad) and primers TOC75RTPCRf2 and TOC75RTPCRr2. Copy number control was performed using cosmid PSBL491 as template.

Genomic DNA was prepared from cultures grown without ATc or with ATc for 24 and 72 h using DNeasy® (QIAGEN). The 100 ng of the resulting template was used for qPCR reaction using SYBR Green Mix (Bio-Rad) and primers Apg-qPCR-F/R for apicoplast and UPRT-qPCR-F/R for nuclear genomes. Copy number control was performed using specific plasmids as described in [63].

IFA and imaging

T. gondii: Immunofluorescence was carried out on infected HFF cells seeded onto glass cover slips. Cells were fixed with 4% paraformaldehyde/PBS (15 min, room temperature), permeabilized with 0.5% T-X-100/PBS (15 min, room temperature), blocked with 5% BSA/PBS (30 min, room temperature), incubated with primary antibodies diluted in 5% BSA/PBS (1 h, room temperature), washed three times in PBS, incubated with suitable

fluorescent-conjugated secondary antibodies (1 h, room temperature), washed three times in PBS, incubated with 50 ng/mL Hoechst 33258/PBS (5 min, room temperature), washed in distilled water and cover slips were mounted onto glass slides using Flouromount (SouthernBiotech). *P. falciparum*: Cells were fixed in 4% Paraformaldehyde/0.00075% Glutaraldehyde (37°C, 30 min), quenched in 125 mm Glycine/PBS, Hoechst 33258 (Molecular probes) was used at 50 ng/mL for fixed parasites or 10 mg/mL for live parasites.

Images were acquired on Carl Zeiss Axio Observer inverse epifluorescence microscope (Figures A1.3 and A1.S2). Individual images were imported into ImageJ64 (version 1.46r, available at http://rsb.info.nih.gov/ij), converted to 8-bit grayscale, subjected to background subtraction, and overlaid. Image in Figure A1.S3 was taken using a Delta Vision microscope as described [12]. Antibodies and concentrations used were: rabbit anti-HA (Sigma-Aldrich, 1:50); mouse anti-Ty tag (a kind gift of Keith Gull, 1:20); anti-ACP (a kind gift of Geoff McFadden, 1:500), Cy2 goat anti-rabbit, Cy3 goat anti-Rabbit, Cy2 goat anti-mouse and Cy3 goat anti-mouse (all Jackson Immuno Research Laboratories, 1:2000).

For Superresolution structural illumination microscopy (SR-SIM), stacks of 30–40 images were taken with increments of 0.091 μm in a Zeiss Elyra Superresolution microscope with a 63× oil immersion objective and an immersion oil with a refractive index of 1.518 (Zeiss). Superresolution images were generated using zen software (version Zen 2012 SP1, Zeiss) and processed into their final form using fiji software [64].

Apicoplast protein import assay and western blot

Western blot of steady-state levels of proteins

Clonal parasite lines grown in the presence or absence of ATc and collected (1500 × *g*, 10 min, room temperature), lysed in sample buffer, separated by SDS-PAGE and blotted using anti-GFP (ROCHE) antibody for ACP-YFP and anti-HA antibody (Sigma) for TGME49_001270.

Western blot of transiently expressed proteins

TATiΔKu80iToc75pi parasites were grown in ATc for a given period of time, then transiently transfected with pBT_LytB or pTUB8-PPP1-HA, and let to grow for an additional 24 h to reach the total desired time of downregulation (for example for 72 h +ATc time point, parasites were grown for 48 h in ATc, transfected and then grown for an additional 24 h in ATc). Transfected and treated parasites were collected, separated by SDS-PAGE and blotted using anti-HA or anti-Ty antibodies.

Pulse/chase analysis was performed as described before [12, 18, 40].

Acknowledgments

This work was supported in part by U.S. National Institutes of Health RO1 grants AI064671, AI084415 (to B. S.) and K99 grant AI103032 (to L. S.). B. S. is a Georgia Research Alliance distinguished investigator. J. D. F. is a predoctoral fellow of the American Heart Association. I. B. and S. H. were supported by DFG grant PR1099/2-1 (to J. M. P.). N. F. was supported by CEF and SFB807 to Enrico Schleiff and we would like to thank him for critical discussions. J. M. P. and L. S. wishes to thank Markus Meissner for initial assistance with cultivation and transfection of *T. gondii* and for the kind contribution of antibodies and markers.

A1.S1 Supplementary materials and methods

Sequence analysis of omp85-like proteins in Apicomplexa genomes

We used jackhmmer to mine the NCBI non-redundant database with *Arabidopsis thaliana* Toc75-III as query sequence. This search revealed two Omp85-like protein coding genes in *T. gondii* ME49 (TGME49_205570, TGME49_272390) and in several *Plasmodium* spp (Table A1.1). Subsequent Omp85-related pHMM searches in the PFAM database, and a BLAST search using the above detected *Toxoplasma* and *Plasmodium* Omp85 proteins unraveled two Omp85-like proteins also in *P. falciparum* 3D7 (PF3D7_0608310 and PF3D7_1234600). Reciprocal BLASTs against the apicomplexan databases in EupathDB (http://eupathdb.org/eupathdb/) identified further homologs of both proteins encoded by several species (Table A1.1).

The predicted gene models for TgToc75 and PfToc75 as found on EupathDB were changed since we first identified these genes: TgToc75 older version, TGME49_072390 includes an extreme C-terminal domain, which is part of the predicted β-barrel. This C-terminal domain is missing in the new version (TGME49_272390). Our RT-PCR and localization of full-length protein supports the old gene models. Similarly, PfToc75's previous model (PFF0410w) predicts one continuous gene, which is now predicted to be two separate genes (PF3D7_0608300/0608310). Our RT-PCR confirms the old model. Prediction of organelle targeting signals as shown in table 1 used the older experimentally confirmed gene models. User comments were added to the respective gene pages in ToxoDB and PlasmoDB.

To determine the affiliations of the four identified Omp85-like sequences from *Plasmodium falciparum* and *Toxoplasma gondii*, we selected a subset of species across

the eukaryotic tree of life and generated a sequence alignment as described in [65]. A majority rule consensus tree was then constructed from 1,000 bootstrap trees based on this alignment (Figure A1.2). A maximum likelihood (ML) phylogeny was reconstructed from the same dataset with RAxML. Branch support values were determined from 1,000 bootstrap trees (Figure A1.S1A). We then added the second *Plasmodium falciparum* Omp85 sequence that was not originally identified via the jackhmmer search (PF3D7_1234600) to the dataset and reconstructed another ML tree (Figure A1.S1B). However, the classification of this sequence is ambiguous. We set out to clarify its affiliation by constructing a phylogenetic tree of the excised POTRA region, which is more conserved than the β-barrel region and thus more suitable for the tree reconstruction. This tree shows that PF3D7_1234600 is located within the sub-tree containing the other Toc75 homologs from Chromalveolates (Figure A1.2B). Furthermore, the bootstrap between the Sam50 and Toc75 clades is reliable with a value of 88. In our alignment we could identify two POTRA domains in PfToc75 (residues 118-192, 193-454) and TgToc75 (189-336, 337-451), which are in agreement with fold recognition results except that the HHpred webserver [66] does not detect the 1st β-strand of PfToc75's 2nd POTRA domain (399-454). In agreement with the assignment based on the phylogenetic trees PfToc75 possesses a predicted apicoplast-targeting signal and for PfSam50 a mitochondrial targeting sequence was predicted (Table A1.1).

References

1. Sheiner L, Vaidya AB, McFadden GI: The metabolic roles of the endosymbiotic organelles of Toxoplasma and Plasmodium spp. *Current opinion in microbiology* 2013, 16(4):452-458.
2. van Dooren GG, Striepen B: The Algal Past and Parasite Present of the Apicoplast. *Annual review of microbiology* 2013.
3. Cavalier-Smith T: Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *The Journal of eukaryotic microbiology* 1999, 46(4):347-366.
4. Koreny L, Sobotka R, Janouskovec J, Keeling PJ, Obornik M: Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *The Plant cell* 2011, 23(9):3454-3462.
5. Sheiner L, Striepen B: Protein sorting in complex plastids. *Biochimica et biophysica acta* 2013, 1833(2):352-359.
6. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF, McFadden GI: Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. *Science* 2003, 299(5607):705-708.
7. Jayabalasingham B, Voss C, Ehrenman K, Romano JD, Smith ME, Fidock DA, Bosch J, Coppens I: Characterization of the ATG8-conjugation system in 2 Plasmodium species with special focus on the liver stage Possible linkage

between the apicoplastic and autophagic systems? *Autophagy* 2014, 10(2):269-284.

8. Tawk L, Dubremetz JF, Montcourrier P, Chicanne G, Merezegue F, Richard V, Payrastre B, Meissner M, Vial HJ, Roy C *et al*: Phosphatidylinositol 3-Monophosphate Is Involved in Toxoplasma Apicoplast Biogenesis. *PLoS pathogens* 2011, 7(2).

9. Tomlins AM, Ben-Rached F, Williams RAM, Proto WR, Coppens I, Ruch U, Gilberger TW, Coombs GH, Mottram JC, Muller S *et al*: Plasmodium falciparum ATG8 implicated in both autophagy and apicoplast formation. *Autophagy* 2013, 9(10):1540-1552.

10. Bouchut A, Geiger JA, DeRocher AE, Parsons M: Vesicles bearing Toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption. *PloS one* 2014, 9(11):e112096.

11. Heiny SR, Pautz S, Recker M, Przyborski JM: Protein Traffic to the Plasmodium falciparum apicoplast: evidence for a sorting branch point at the Golgi. *Traffic* 2014, 15(12):1290-1304.

12. Agrawal S, van Dooren GG, Beatty WL, Striepen B: Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. *The Journal of biological chemistry* 2009, 284(48):33683-33691.

13. Hempel F, Felsner G, Maier UG: New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Molecular microbiology* 2010, 76(3):793-801.

14. Kalanon M, Tonkin CJ, McFadden GI: Characterization of Two Putative Protein Translocation Components in the Apicoplast of Plasmodium falciparum. *Eukaryotic cell* 2009, 8(8):1146-1154.

15. Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG, Przyborski JM: An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic cell* 2009, 8(8):1134-1145.

16. Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG: Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic cell* 2012, 11(12):1472-1481.

17. Glaser S, van Dooren GG, Agrawal S, Brooks CF, McFadden GI, Striepen B, Higgins MK: Tic22 Is an Essential Chaperone Required for Protein Import into the Apicoplast. *Journal of Biological Chemistry* 2012, 287(47).

18. van Dooren GG, Tomova C, Agrawal S, Humbel BM, Striepen B: Toxoplasma gondii Tic20 is essential for apicoplast protein import. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105(36):13574-13579.

19. van Dooren GG, Su V, D'Ombrain MC, McFadden GI: Processing of an apicoplast leader sequence in Plasmodium falciparum and the identification of a putative leader cleavage enzyme. *The Journal of biological chemistry* 2002, 277(26):23612-23619.

20. Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier UG: Protein targeting into the complex plastid of cryptophytes. *Journal of molecular evolution* 2006, 62(6):674-681.

21. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG: Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant molecular biology* 2007, 64(5):519-530.

22. Kilian O, Kroth PG: Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *The Plant journal : for cell and molecular biology* 2005, 41(2):175-183.

23. Patron NJ, Waller RF, Archibald JM, Keeling PJ: Complex protein targeting to dinoflagellate plastids. *Journal of molecular biology* 2005, 348(4):1015-1024.

24. Steiner JM, Yusa F, Pompe JA, Loffelhardt W: Homologous protein import machineries in chloroplasts and cyanelles. *Plant Journal* 2005, 44(4):646-652.

25. Wunder T, Martin R, Loffelhardt W, Schleiff E, Steiner JM: The invariant phenylalanine of precursor proteins discloses the importance of Omp85 for protein translocation into cyanelles. *BMC evolutionary biology* 2007, 7.

26. Deane JA, Fraunholz M, Su V, Maier UG, Martin W, Durnford DG, McFadden GI: Evidence for nucleomorph to host nucleus gene transfer: Light-harvesting complex proteins from cryptomonads and chlorarachniophytes. *Protist* 2000, 151(3):239-252.

27. Ralph SA, Foth BJ, Hall N, McFadden GI: Evolutionary pressures on apicoplast transit peptides. *Mol Biol Evol* 2004, 21(12):2183-2194.

28. Harb OS, Chatterjee B, Fraunholz MJ, Crawford MJ, Nishi M, Roos DS: Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite Toxoplasma gondii. *Eukaryotic cell* 2004, 3(3):663-674.

29. Schleiff E, Becker T: Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat Rev Mol Cell Bio* 2011, 12(1):48-59.

30. Simmerman RF, Dave AM, Bruce BD: Structure and Function of POTRA Domains of Omp85/TPS Superfamily. *Int Rev Cel Mol Bio* 2014, 308:1-34.

31. Gentle I, Gabriel K, Beech P, Waller R, Lithgow T: The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *Journal of Cell Biology* 2004, 164(1):19-24.

32. Kozjak V, Wiedemann N, Milenkovic D, Lohaus C, Meyer HE, Guiard B, Meisinger C, Pfanner N: An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *Journal of Biological Chemistry* 2003, 278(49):48520-48523.

33. Eckart K, Eichacker L, Sohrt K, Schleiff E, Heins L, Soll J: A Toc75-like protein import channel is abundant in chloroplasts. *EMBO reports* 2002, 3(6):557-562.

34. Inoue K, Potter D: The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. *Plant Journal* 2004, 39(3):354-365.

35. Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R: Characterization of the translocon of the outer envelope of chloroplasts. *Journal of Cell Biology* 2003, 160(4):541-551.

36. Schnell DJ, Kessler F, Blobel G: Isolation of Components of the Chloroplast Protein Import Machinery. *Science* 1994, 266(5187):1007-1012.

37. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E: Filling the Gap, Evolutionarily Conserved Omp85 in Plastids of Chromalveolates. *Journal of Biological Chemistry* 2010, 285(9):6848-6856.

38. Agrawal S, Striepen B: More Membranes, more Proteins: Complex Protein Import Mechanisms into Secondary Plastids. *Protist* 2010, 161(5):672-687.

39. Hirakawa Y, Burki F, Keeling PJ: Genome-Based Reconstruction of the Protein Import Machinery in the Secondary Plastid of a Chlorarachniophyte Alga. *Eukaryotic cell* 2012, 11(3):324-333.

40. Sheiner L, Demerly JL, Poulsen N, Beatty WL, Lucas O, Behnke MS, White MW, Striepen B: A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. *PLoS pathogens* 2011, 7(12):e1002392.

41. Brooks CF, Johnsen H, van Dooren GG, Muthalagi M, Lin SS, Bohne W, Fischer K, Striepen B: The toxoplasma apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. *Cell host & microbe* 2010, 7(1):62-73.

42. Francia ME, Jordan CN, Patel JD, Sheiner L, Demerly JL, Fellows JD, de Leon JC, Morrissette NS, Dubremetz JF, Striepen B: Cell division in Apicomplexan parasites is organized by a homolog of the striated rootlet fiber of algal flagella. *PLoS biology* 2012, 10(12):e1001444.

43. Waller RF, Keeling PJ, Donald RG, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS, McFadden GI: Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(21):12352-12357.

44. Waller RF, Reed MB, Cowman AF, McFadden GI: Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. *Embo Journal* 2000, 19(8):1794-1802.

45. Nair SC, Brooks CF, Goodman CD, Sturm A, McFadden GI, Sundriyal S, Anglin JL, Song Y, Moreno SN, Striepen B: Apicoplast isoprenoid precursor synthesis and the molecular basis of fosmidomycin resistance in Toxoplasma gondii. *The Journal of experimental medicine* 2011, 208(7):1547-1559.

46. Ralph SA, Foth BJ, Hall N, McFadden GI: Evolutionary pressures on apicoplast transit peptides. *Molecular biology and evolution* 2004, 21(12):2183-2194.

47. Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ, Maier UG: Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Molecular biology and evolution* 2006, 23(12):2413-2422.

48. Glaser S, van Dooren GG, Agrawal S, Brooks CF, McFadden GI, Striepen B, Higgins MK: Tic22 is an essential chaperone required for protein import into the apicoplast. *The Journal of biological chemistry* 2012, 287(47):39505-39512.

49. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E: Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. *The Journal of biological chemistry* 2010, 285(9):6848-6856.

50. Kalanon M, McFadden GI: The chloroplast protein translocation complexes of Chlamydomonas reinhardtii: A bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. *Genetics* 2008, 179(1):95-112.

51. Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R: Characterization of the translocon of the outer envelope of chloroplasts. *The Journal of cell biology* 2003, 160(4):541-551.

52. Huang WH, Ling QH, Bedard J, Lilley K, Jarvis P: In Vivo Analyses of the Roles of Essential Omp85-Related Proteins in the Chloroplast Outer Envelope Membrane. *Plant physiology* 2011, 157(1):147-159.

53. Jarvis P, Chen LJ, Li HM, Pete CA, Fankhauser C, Chory J: An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science* 1998, 282(5386):100-103.

54. Reumann S, Inoue K, Keegstra K: Evolution of the general protein import pathway of plastids (Review). *Molecular membrane biology* 2005, 22(1-2):73-U20.

55. Tripp J, Hahn A, Koenig P, Flinner N, Bublak D, Brouwer EM, Ertel F, Mirus O, Sinning I, Tews I *et al*: Structure and Conservation of the Periplasmic Targeting Factor Tic22 Protein from Plants and Cyanobacteria. *Journal of Biological Chemistry* 2012, 287(29):24164-24173.

56. Bredemeier R, Schlegel T, Ertel F, Vojta A, Borissenko L, Bohnsack MT, Groll M, Von Haeseler A, Schleiff E: Functional and phylogenetic properties of the pore-forming beta-barrel transporters of the Omp85 family. *Journal of Biological Chemistry* 2007, 282(3):1882-1890.

57. Eddy SR: Accelerated Profile HMM Searches. *Plos Comput Biol* 2011, 7(10).

58. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J *et al*: Pfam: the protein families database. *Nucleic acids research* 2014, 42(D1):D222-D230.

59. Flinner N, Ellenrieder L, Stiller SB, Becker T, Schleiff E, Mirus O: Mdm10 is an ancient eukaryotic porin co-occurring with the ERMES complex. *Bba-Mol Cell Res* 2013, 1833(12):3314-3325.

60. Stamatakis A: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014, 30(9):1312-1313.

61. Whelan S, Goldman N: A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001, 18(5):691-699.

62. Francia ME, Jordan CN, Patel JD, Sheiner L, Demerly JL, Fellows JD, de Leon JC, Morrissette NS, Dubremetz JF, Striepen B: Cell Division in Apicomplexan Parasites Is Organized by a Homolog of the Striated Rootlet Fiber of Algal Flagella. *PLoS biology* 2012, 10(12).

63. Reiff SB, Vaishnava S, Striepen B: The HU protein is important for apicoplast genome maintenance and inheritance in Toxoplasma gondii. *Eukaryotic cell* 2012, 11(7):905-915.

64. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al*: Fiji: an open-source platform for biological-image analysis. *Nature methods* 2012, 9(7):676-682.

65. Flinner N, Ellenrieder L, Stiller SB, Becker T, Schleiff E, Mirus O: Mdm10 is an ancient eukaryotic porin co-occurring with the ERMES complex. *Biochim Biophys Acta* 2013, 1833(12):3314-3325.

66. Soding J, Biegert A, Lupas AN: The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005, 33(Web Server issue):W244-248.

67. Agrawal S, Chung DWD, Ponts N, van Dooren GG, Prudhomme J, Brooks CF, Rodrigues EM, Tan JC, Ferdig MT, Striepen B *et al*: An Apicoplast Localized Ubiquitylation System Is Required for the Import of Nuclear-encoded Plastid Proteins. *PLoS pathogens* 2013, 9(6).

Figures

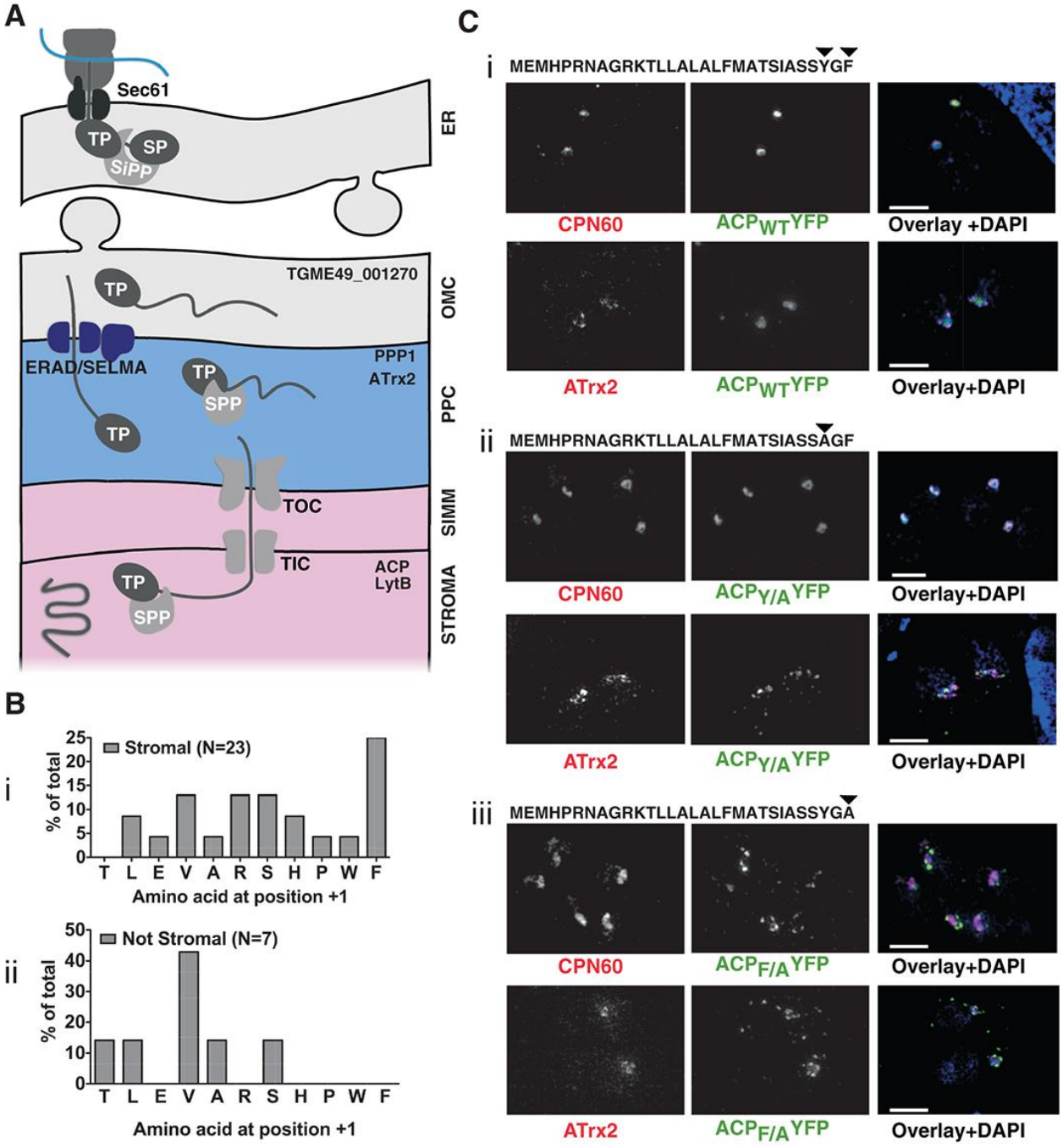


Figure A1.1. The translocation of precursor protein through the apicoplast membrane is mediated by machineries and signals that are conserved among organisms with complex plastids of the red lineage. A) Schematic representation of the

translocation machineries responsible for protein import through the four membranes of the apicoplast. Endomembrane system is in grey, former red algal cytosol in blue and former primary plastid in pink. According to the current model of transport apicoplast precursor proteins are first co-translationally transported into the ER via the SEC61 complex courtesy of their signal peptide (SP). In the ER lumen, the SP is cleaved by a signal peptide peptidase (SiPP). The now exposed transit peptide (TP) signals for transport from the ER to the apicoplast. Next, the translocation through ERAD/SELMA is possibly mediated by ubiquitination [67]. The protein then moves through the TOC and TIC complexes and its TP is cleaved in the stroma. Compartmental markers used in this study are depicted at the upper right corner of their corresponding compartment. B) Abundance of residues at position +1 of 29 proteins experimentally confirmed to localize to the apicoplast stroma, or peripheral compartments based on SP cleavage prediction by SignalP (the detailed analysis is provided in Table A1.S1). C) High-resolution microscopy analysis of the localization of transiently expressed ACP-YFP with the wild type phenylalanine and tyrosine at the two predicted potential position +1 (i) with tyrosine to alanine mutation (ii) and with phenylalanine to alanine mutation (iii). The schemes show the full SP sequence with both potential +1 residues highlighted in yellow. Scale bars 2 μM.

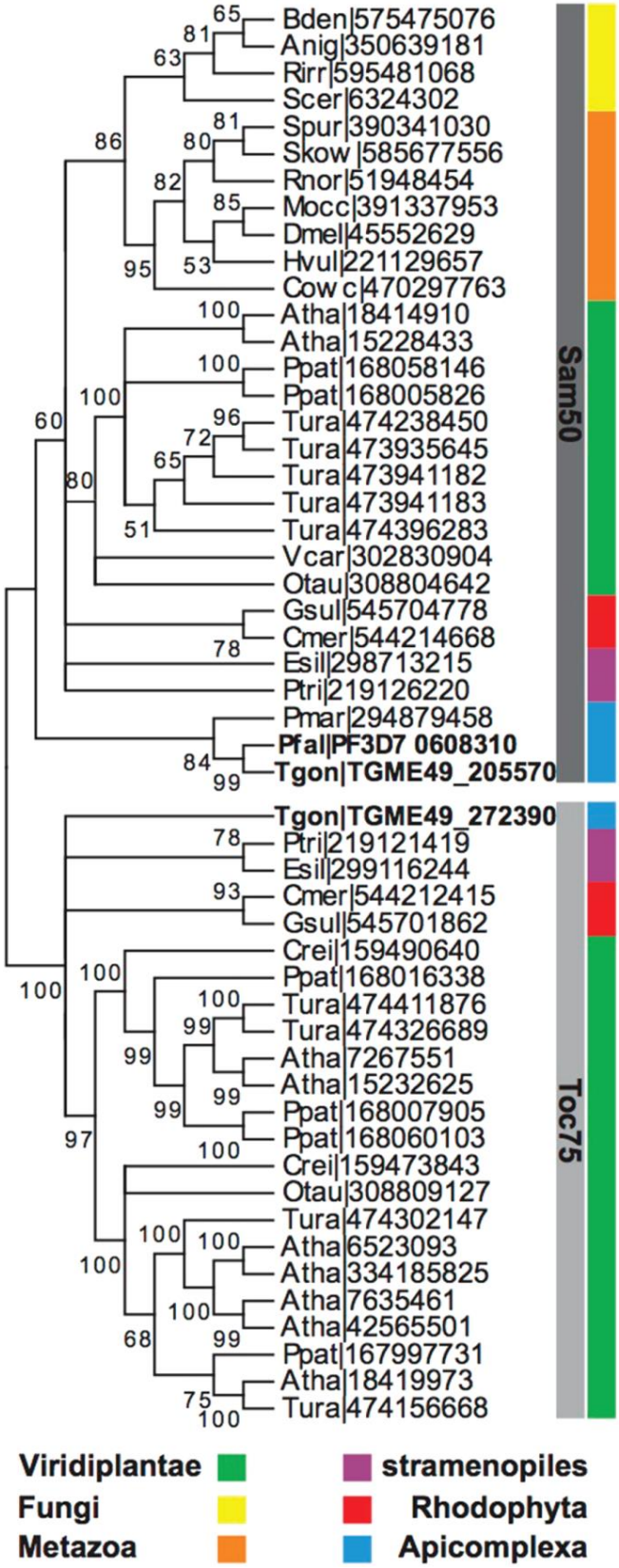

Bden|575475076
Anig|350639181
Rirr|595481068
Scer|6324302
Spur|390341030
Skow|585677556
Rnor|51948454
Mocc|391337953
Dmel|45552629
Hvul|221129657
Cowc|470297763
Atha|18414910
Atha|15228433
Ppat|168058146
Ppat|168005826
Tura|474238450
Tura|473935645
Tura|473941182
Tura|473941183
Tura|474396283
Vcar|302830904
Otau|308804642
Gsul|545704778
Cmer|544214668
Esil|298713215
Ptri|219126220
Pmar|294879458
Pfal|PF3D7 0608310
Tgon|TGME49_205570
Tgon|TGME49_272390
Ptri|219121419
Esil|299116244
Cmer|544212415
Gsul|545701862
Crei|159490640
Ppat|168016338
Tura|474411876
Tura|474326689
Atha|7267551
Atha|15232625
Ppat|168007905
Ppat|168060103
Crei|159473843
Otau|308809127
Tura|474302147
Atha|6523093
Atha|334185825
Atha|7635461
Atha|42565501
Ppat|167997731
Atha|18419973
Tura|474156668
Sam50
Toc75
Viridiplantae
Fungi
Metazoa
stramenopiles
Rhodophyta
Apicomplexa

Figure A1.2. Phylogenetic classification of Omp85-like proteins in *T. gondii* and *P. falciparum*. A majority rule consensus tree of selected Sam50 and Toc75 homologs was constructed with RAxML from 1000 bootstrap trees. The corresponding maximum likelihood (ML) tree is given in Figure A1.S1A. The proteins are referenced by their ID (GenBank, or EupathDB); species abbreviations are as follows: Anig (*Aspergillus niger* ATCC 1015), Atha (*Arabidopsis thaliana*), Bden (*Batrachochytrium dendrobatidis* JAM81), Cmer (*Cyanidioschyzon merolae* strain 10D), Cowc (*Capsaspora owczarzaki* ATCC 30864), Crei (*Chlamydomonas reinhardtii*), Dmel (*Drosophila melanogaster*), Esil (*Ectocarpus siliculosus*), Gsul (*Galdieria sulphuraria*), Hvul (*Hydra vulgaris*), Mocc (*Metaseiulus occidentalis*), Otau (*Ostreococcus tauri*), Pfal (*Plasmodium falciparum* 3D7), Pmar (*Perkinsus marinus* ATCC 50983), Ppat (*Physcomitrella patens*), Ptri (*Phaeodactylum tricornutum* CCAP 1055/1), Rirr (*Rhizophagus irregularis* DAOM 197198w), Rnor (*Rattus norvegicus*), Scer (*Saccharomyces cerevisiae* S288c), Skow (*Saccoglossus kowalevskii*), Spur (*Strongylocentrotus purpuratus*), Tgon (*Toxoplasma gondii* ME49), Tura (*Triticum urartu*) and Vcar (*Volvox carteri* f. nagariensis). The full alignments used for this analysis are provided in Figure A1.S1.

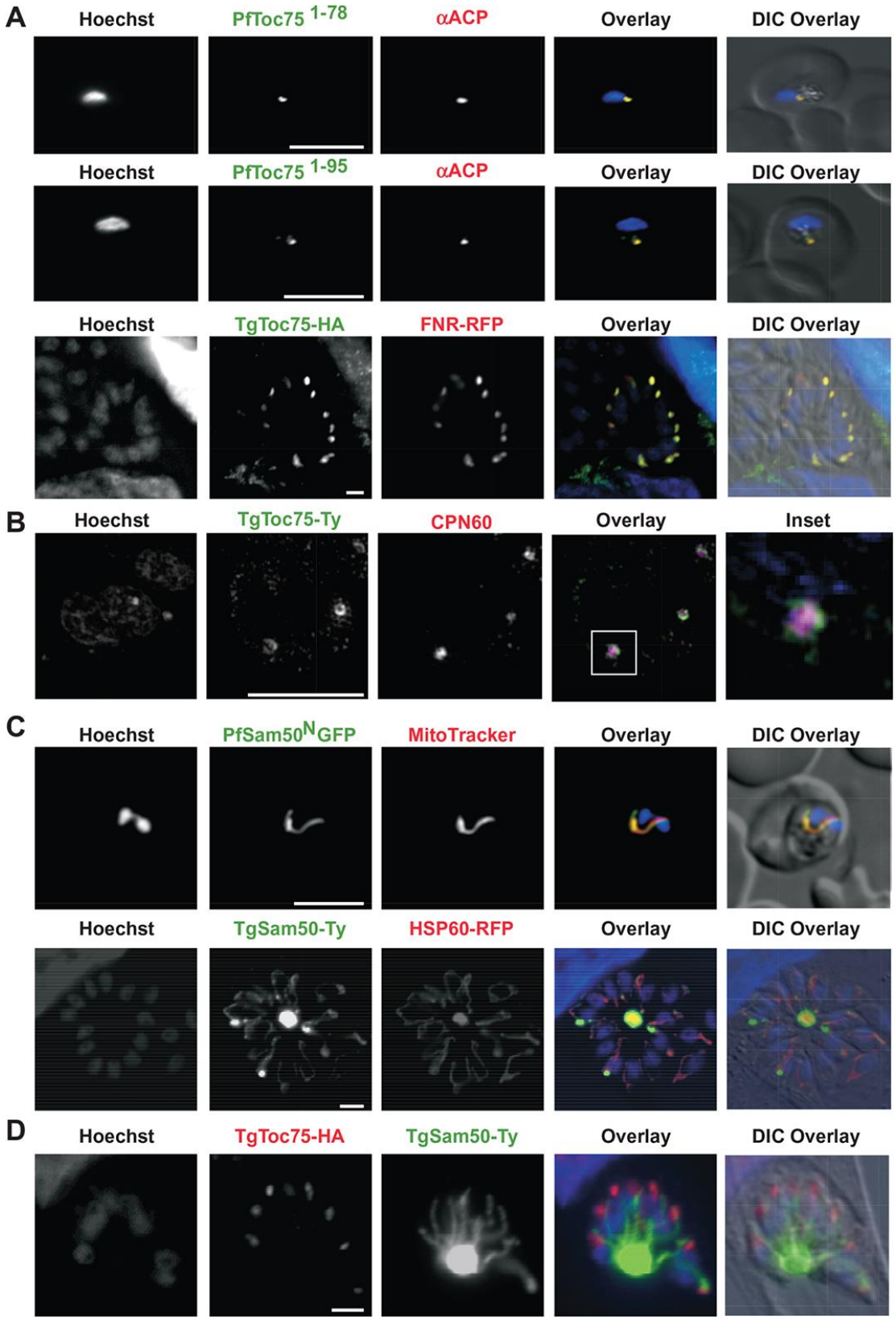

A
Hoechst
PfToc75 1-78
αACP
Overlay
DIC Overlay
Hoechst
PfToc75 1-95
αACP
Overlay
DIC Overlay
Hoechst
TgToc75-HA
FNR-RFP
Overlay
DIC Overlay
B
Hoechst
TgToc75-Ty
CPN60
Overlay
Inset
C
Hoechst
PfSam50NGFP
MitoTracker
Overlay
DIC Overlay
Hoechst
TgSam50-Ty
HSP60-RFP
Overlay
DIC Overlay
D
Hoechst
TgToc75-HA
TgSam50-Ty
Overlay
DIC Overlay

Figure A1.3. Localization of the omp85-like proteins supports their affiliation as Toc75 and Sam50 in Apicomplexa. A) Fluorescence microscopy analysis of *P. falciparum* parasites expressing ectopic GFP fusions of the 78 (upper panel) and 95 (middle panels) N-terminal amino acids of PfToc75, and of *T. gondii* parasites expressing ectopic HA-tagged full-length TgToc75 (lower panel). The tagged proteins are shown in green. Co-staining is done with ACP and FNR-RFP for *P. falciparum* and *T. gondii*, respectively and shown in red. B) High-resolution microscopy of transiently expressed full-length Ty-tagged TgToc75 and its localization with respect to the stromal marker CPN60. C) *P. falciparum* parasite expressing ectopic GFP fusion of the 60 N-terminal amino acids of PfSam50 (green) co-stained with mito-tracker (red) (upper panel); *T. gondii* parasites expressing ectopic Ty-tagged TgSam50 (green) co-stained with the mitochondrial marker HSP60-RFP (red) (lower panel). D) *T. gondii* parasites co-expressing ectopic HA-tagged TgToc75 (red) and Ty-tagged TgSam50 (green). Scale bars 5 μM.

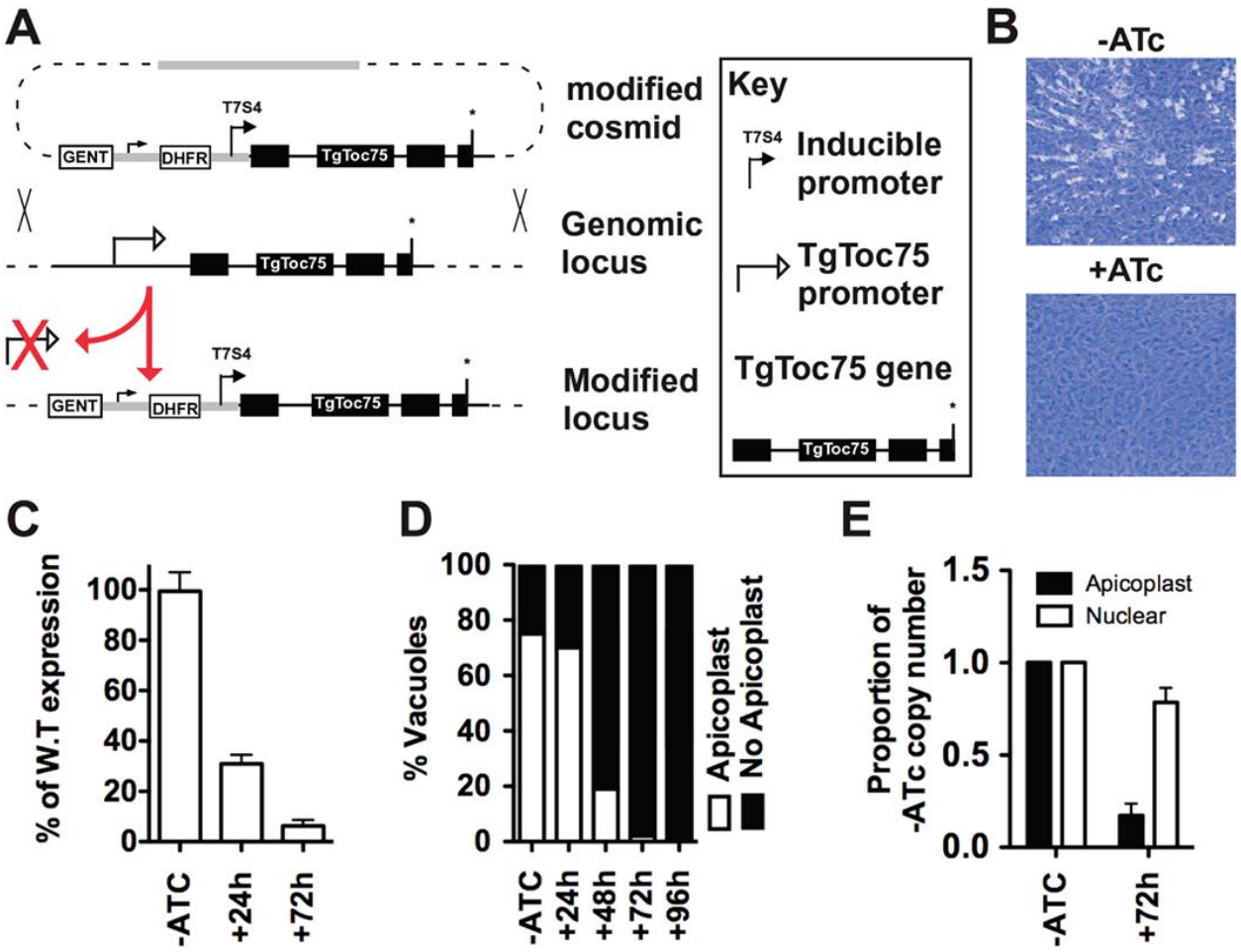


Figure A1.4. TgToc75 is essential for parasite growth and apicoplast maintenance. A) Schematic representation of the manipulation of the *TgToc75* locus to replace the native promoter with the tetracycline inducible promoter. Black boxes – exons; asterisk – stop codon; empty boxes – minigenes; solid lines – TgToc75 locus non-coding sequences; dashed line – genomic sequence; grey thick line – backbone of cosmid or of modification cassette. B) Plaque assays performed with the *TATiΔKu80iToc75pr* parasite line in the absence (−) or presence (+) of ATc. C) qRT-PCR analysis with RNA extracts from *TATiΔKu80iToc75pr* grown in the absence of ATc (−ATc) or upon downregulation of TgToc75 for 24 (+24 h) and 72 (+72 h) hours. TgToc75 mRNA levels decline swiftly upon ATc treatment. Y-axis shows the percentage of wild type copy numbers. D)

TATiΔKu80iToc75pr parasites were grown in ATc as indicated and plastids were counted based on immunofluorescence signal obtained via staining with anti-CPN60 antibody in 100 four-parasites vacuoles for each sample. Y-axis shows percentage of 4-parasites-vacuoles. E) Apicoplast loss was confirmed using qPCR comparing nuclear genome and apicoplast genome copy numbers. The data was normalized such that copy number from each genome from no ATc treatment is 1. In support of apicoplast loss the proportion of apicoplast copy number after TgToc75 downregulation for 72 h is in average 0.17 while genomic copy number average proportion is 0.78.

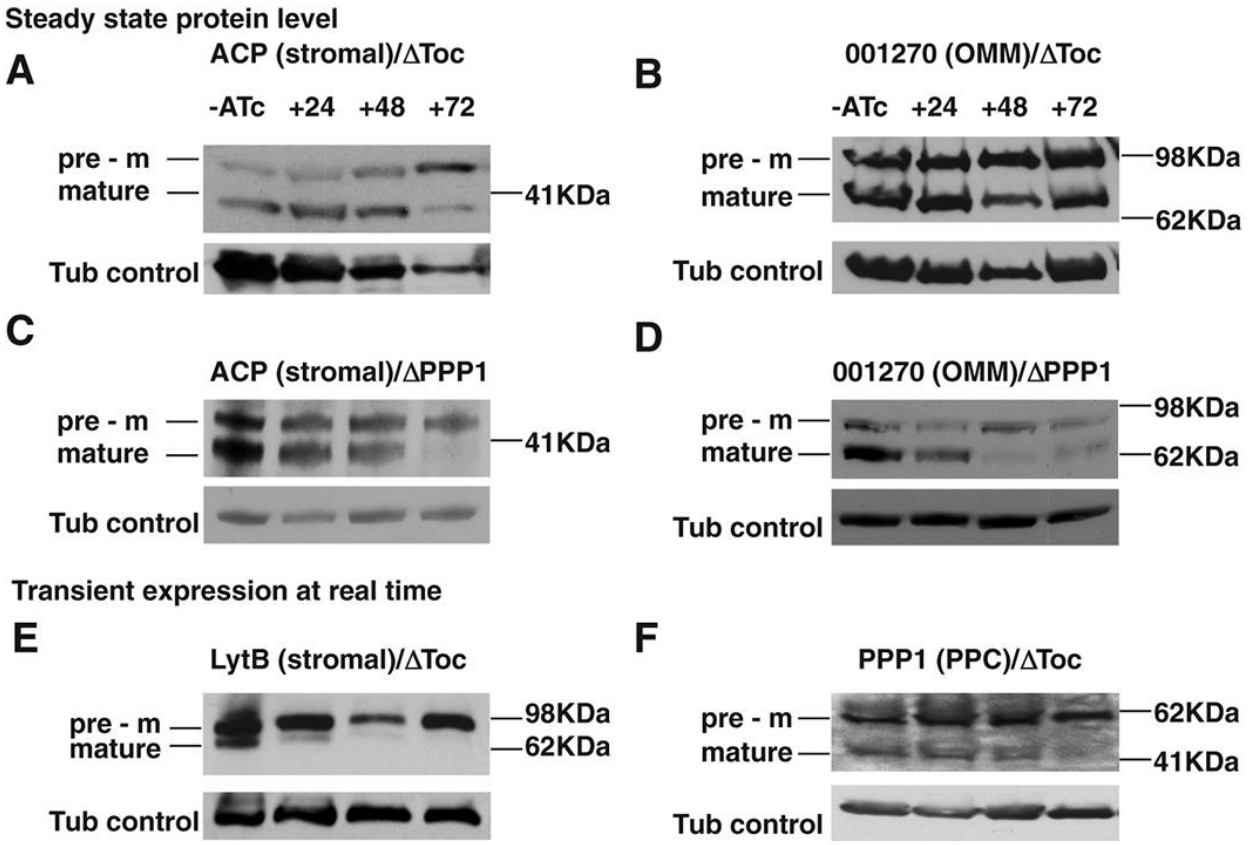


Figure A1.5. TgToc75 downregulation results in deficient import of stromal but not PPC or OMC apicoplast proteins. We preformed western blot analysis to follow the maturation of apicoplast proteins under the downregulation of apicoplast import components. The steady-state expression of endogenously YFP-tagged ACP [40] (A) and of endogenously HA-tagged TGME49_001270 (B) was monitored at each time point of *TgToc75* downregulation showing maturation defect in ACP but not TGME49_001270 at 72 h. Western blot analysis following the maturation of the same makers (YFP-tagged ACP (C) and endogenously HA-tagged TGME49_001270 (D)), but this time under downregulation of the PPC import component *TgPPP1*, shows maturation defect for both at 48 h. We then performed western blot analysis following the maturation of the stromal protein LytB-Ty (E) and the periplastid protein PPP1 (F). In this experiment LytB or PPP1 are transiently expressed for 24 h at each time point of *TgToc75* downregulation.

This analysis shows defect in LytB maturation that is detected at 24 h and is completed by 48 h. In contrast, maturation defect of PPP1 is only observed at 72 h. Loading control performed with anti-alpha-tubulin antibody is shown for each blot.

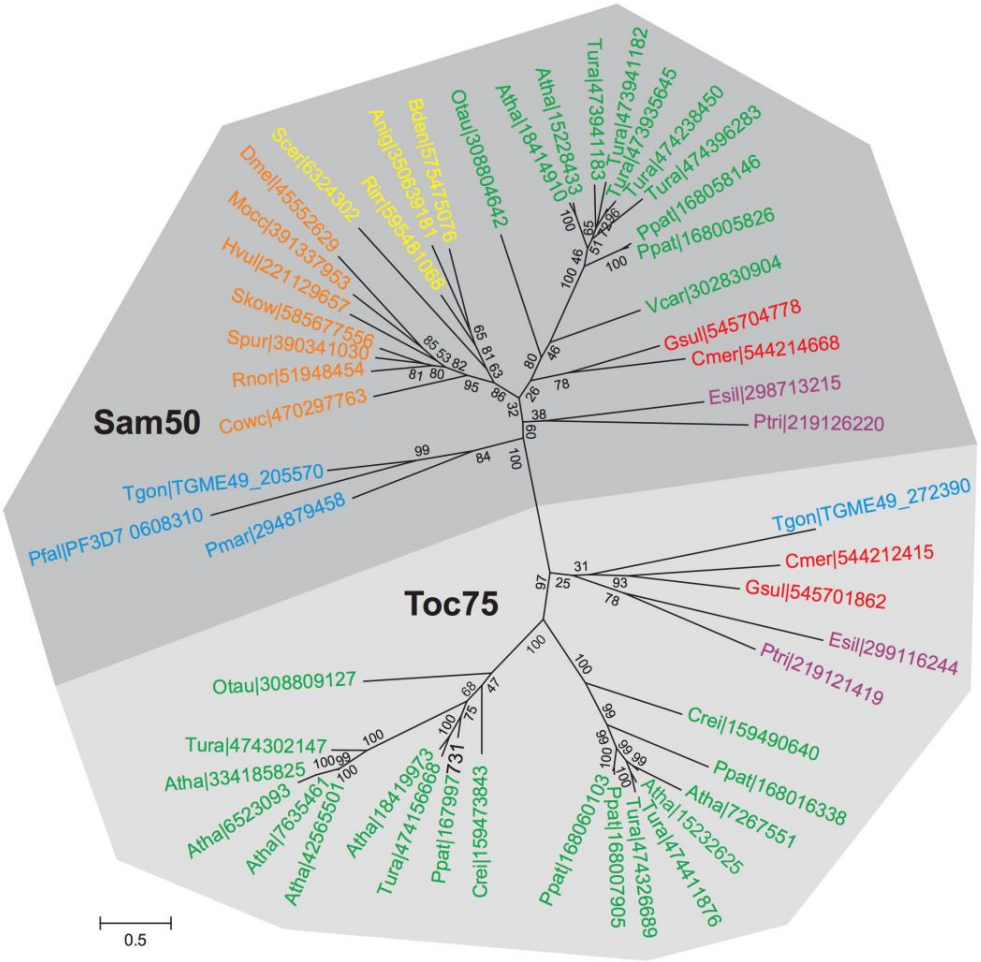
Tables

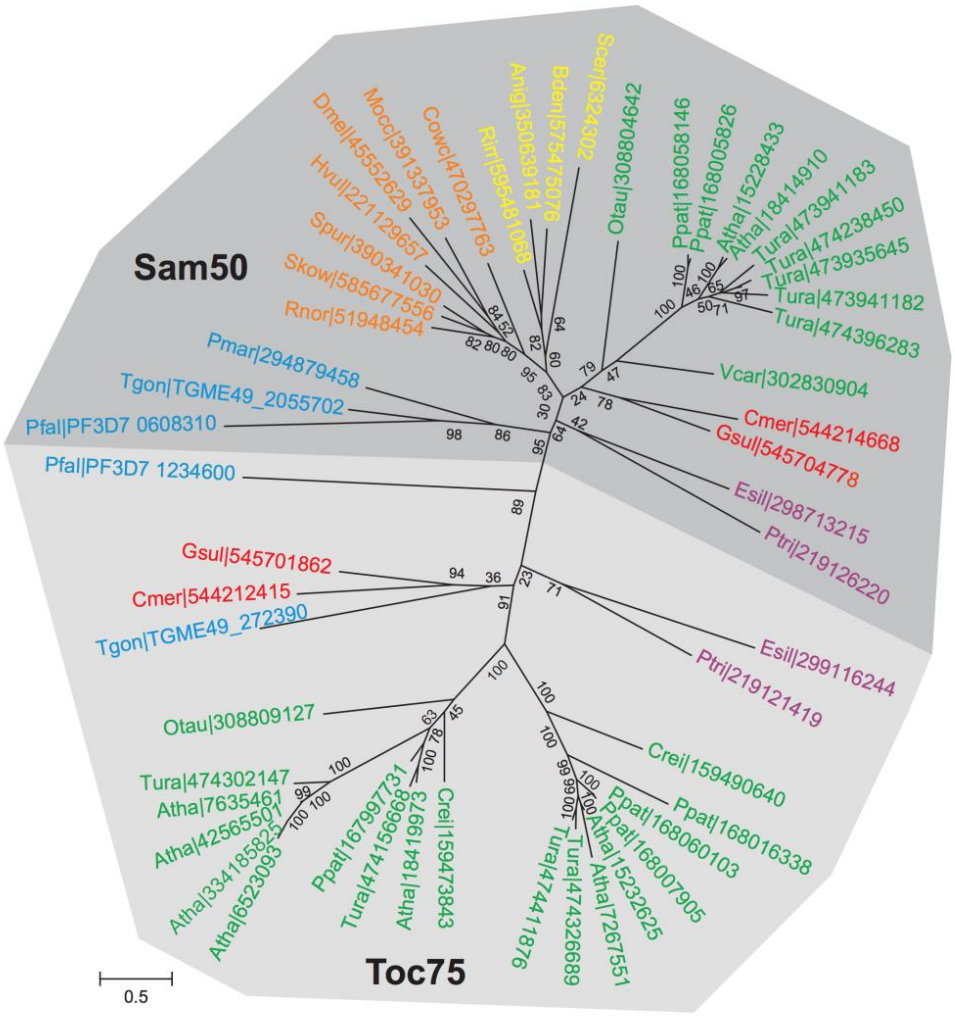
Table A1.1. GeneIDs and summary of targeting prediction for apicomplexan omp85-like protein encoding genes

	Toc75	SignalP	PlasmoAP	PATS	Sam50	Mito Prot	Plas Mit
Toxoplasma gondii	[3]TGME49_072390	N	N	N	TGME49_205570	N	N
Neospora caninum	NCLIV_034910	N	N	N	NCLIV_020120	N	Y
Eimeria falciformis	EfaB_MINUS_25252.g2122	N	N	N	NF	-	-
Eimeria praecox	EPH_0025670	Y	N	N	NF	-	-
Eimeria necatrix	ENH_00027930	N	N	N	ENH_0075630	N	N
Plasmodium falciparum	PF3D7_1234600	Y	Y	Y	*pff04010W	N	Y
Plasmodium chabaudi	PCYB_146020	N	Y	Y	PCHAS_010750	N	Y
Plasmodium_berghei	PBANKA_144920	N	Y	Y	PBANKA_010690	N	Y
Plasmodium_yoelii	PY17X_1451700	N	Y	Y	PY17X_010690	N	Y
Plasmodium_cynomolgi	PCYB_146020	N	N	N	PCYB_114820	N	Y
Plasmodium_knowlesi	PKH_145170	Y	Y	N	PKH_114100	N	Y
Plasmodium_vivax	PVX_100680	Y	Y	Y	PVX_113574	N	N
Theileria_equi	NF	-	-	-	BEWA_051860	N	N
Babesia_bovis	NF	-	-	-	BBOV_III000300	N	N

[3]Newer gene model does not agree with our experimental data. See GeneBank accession numbers: TgToc75 KT27 1755. PfSam50 KT271756

Supplementary Figures




Sam50
Toc75
Scer|6324302
Otau|30880642
Bden|575475076
Anig|350639181
Rirr|595481068
Cowc|470297763
Mocc|391337953
Dmel|45552629
Hvul|221129657
Spur|390341030
Skow|585677556
Rnor|51948454
Ppat|168058146
Ppat|168005826
Atha|15228433
Atha|18414910
Tura|473941183
Tura|474238450
Tura|473935645
Tura|473941182
Tura|474396283
Vcar|302830904
Cmer|544214668
Gsul|545704778
Esil|298713215
Ptri|219126220
Pmar|294879458
Tgon|TGME49_2055702
Pfal|PF3D7 0608310
Pfal|PF3D7 1234600
Gsul|545701862
Cmer|544212415
Tgon|TGME49_272390
Esil|299116244
Ptri|219121419
Otau|308809127
Crei|159490640
Ppat|168016338
Ppat|168060103
Ppat|168007905
Atha|1523625
Atha|7267551
Tura|474326689
Tura|474411876
Crei|159473843
Atha|18419973
Tura|474156668
Ppat|167997731
Atha|6523093
Atha|33418582
Atha|4256550
Atha|7635461
Tura|474302147
0.5

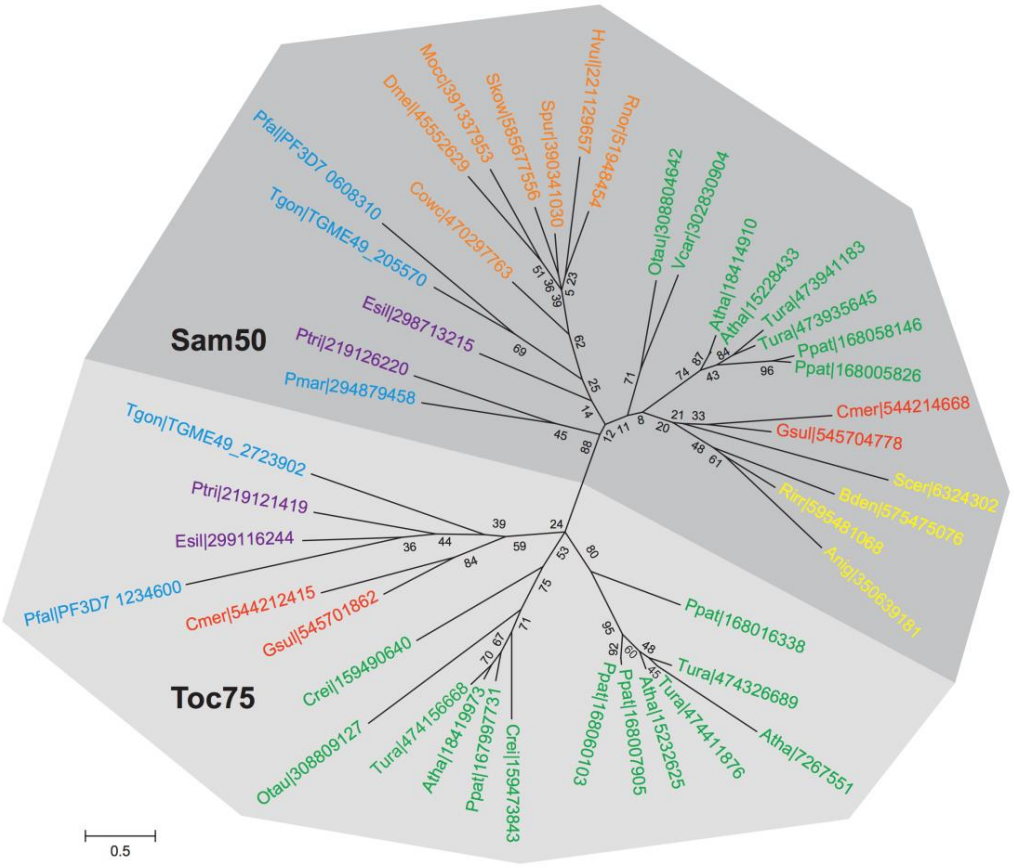


Figure A1.S1 – **Phylogenetic classification of Omp85-like proteins in T. gondii and P. falciparum**. Maximum likelihood (ML) phylogenies were reconstructed with RAxML. Branch support values were determined from 1,000 bootstrap trees. Sam50 and Toc75 clades are marked by dark and light gray areas, respectively. The sequence labels are colored according to their taxonomy (color code given in Figure A1.2). (A) The ML tree was reconstructed from the same set of sequences as used for the majority rule consensus tree in Figure A1.2. (B) This ML tree was reconstructed with the same sequences as (A) while including PfToc75. (C) From the multiple sequence consensus alignment as used for the trees above the N-terminal part containing the POTRA domains was excised and a ML tree reconstructed.

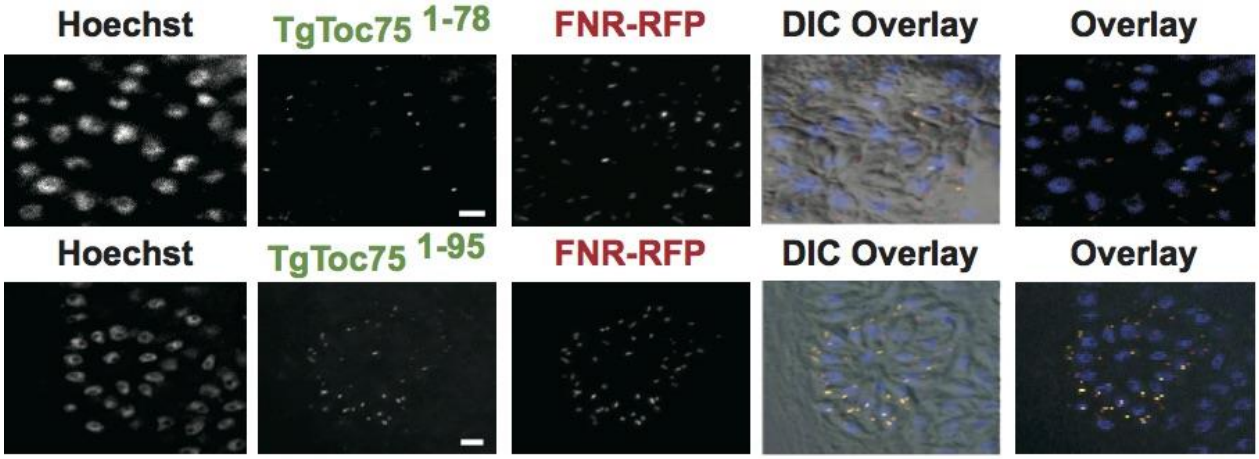


Figure A1.S2 – The N-terminal domain of TgToc75 is sufficient for apicoplast localization. Fluorescence microscopy analysis of parasites expressing ectopic Ty-tagged fusions of the 78 (upper panel) and 95 (lower panels) N-terminal amino acids of TgToc75. Scale bars 5 μM.

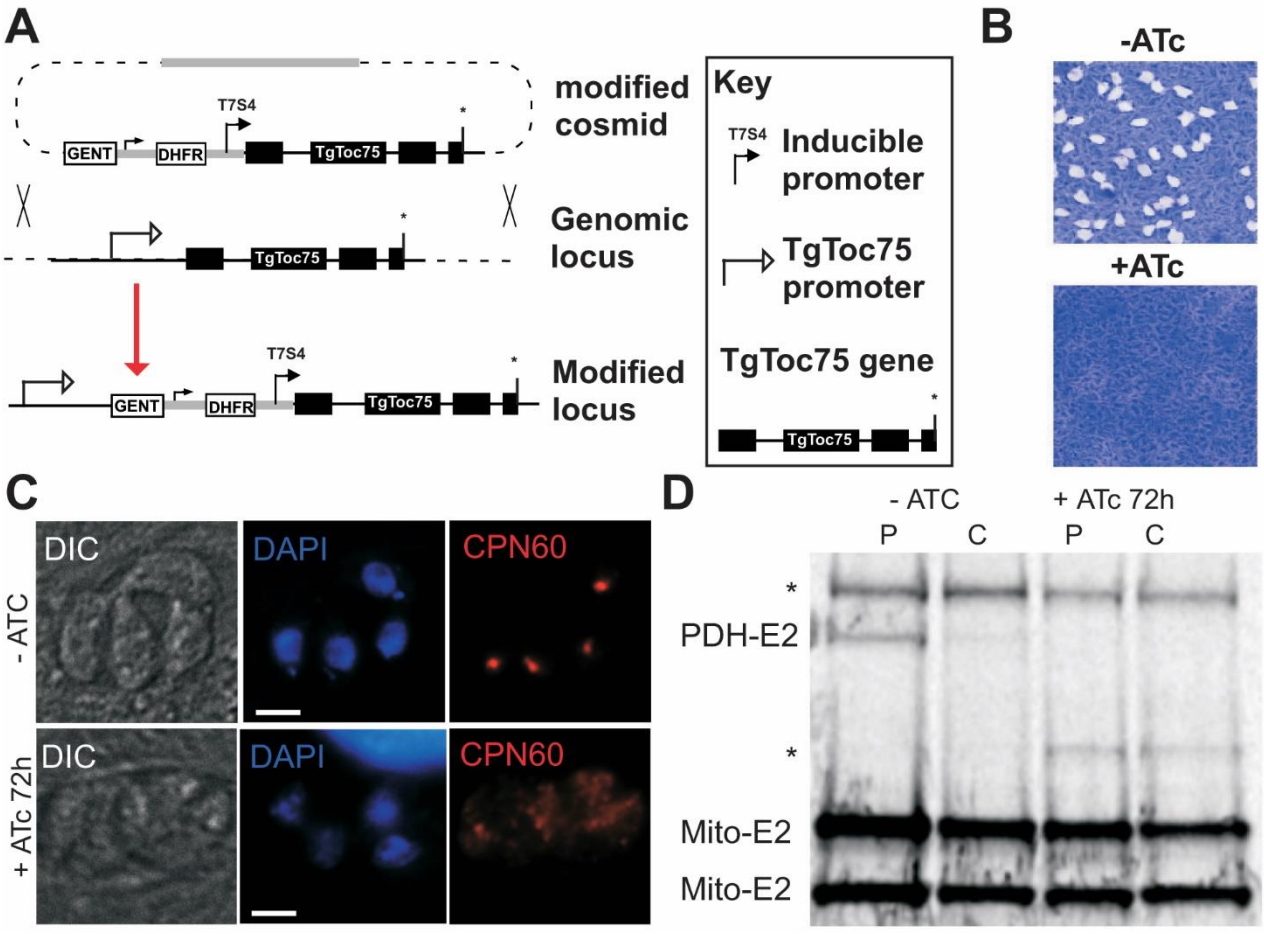


Figure A1.S3 – **TgToc75 is essential for parasite growth and apicoplast biogenesis**. (A) Schematic representation of the manipulation of the *TgToc75* locus to insert the tetracycline inducible promoter between the native promoter and the ORF. (B) Plaque assays performed with the *TATiΔKu80iToc75pi* parasite line in the absence (-) or presence (+) of ATc, namely TgToc75 constitutive levels or downregulation respectively. (C) Fluorescence microscopy of *TATiΔKu80iToc75pi* grown in absence of ATc (-ATc) or upon down regulation of TgToc75 for 72 hours (+ATc 72h) stained with the apicoplast marker CPN60 [12] showing loss of apicoplast in most parasites at this time point. Scale bars 5 μM (D) Pulse-chase (*P/C*) analysis of protein synthesis and post-translational lipoylation of apicoplast (PDH-E2) and mitochondrial (mito-E2) proteins. Parasites were metabolically labeled as detailed in [12, 18, 40] and lipoylated proteins were isolated by immunoprecipitation using a specific antibody. Lipoylation of PDH-E2 is lost upon ATc treatment. Bands labeled with an asterisk likely represent lipoylated host cell proteins

Supplementary Tables

Table A1.S1 – prediction of signal peptide and amino acid at position +1 of putative transit peptide for 45 experimentally confirmed apicoplast proteins.

Localization	Name	GeneID	SignalP4.1						SignalP3.0					
			C-Score	Y-Score	S-Score	D-Score	SP?	AA at +1	C-Score	Y-Score	S-Score (max)	D-Score	SP?	AA at +1
PPC	PPP1	TGME49_287270	0.11	0.115	0.207	0.123	NO		0.043	0.056	0.269	0.103	NO	
PPC	ATrx2	TGME49_310770	0.198	0.303	0.709	0.369	YES	A	0.299	0.449	0.985	0.665	YES	A
PPC	TgApicE2	TGME49_295990	0.356	0.251	0.405	0.224	NO		0.294	0.405	0.993	0.500	YES	S
PPC	CDC48AP	TGME49_321640	0.356	0.251	0.405	0.224	NO		0.288	0.410	0.944	0.433	YES	V
PPC	TgApicE1	TGME49_314890	0.353	0.188	0.133	0.143	NO		0.200	0.033	0.165	0.089	NO	
PPC	Ubiquitin	TGME49_223125	0.137	0.145	0.215	0.149	NO		0.105	0.085	0.498	0.155	NO	
PPC	UDF1AP	TGME49_285700	0.128	0.136	0.195	0.137	NO		0.067	0.064	0.371	0.100	NO	
outermostmembrane	FtsH	TGME49_259260	0.112	0.12	0.147	0.124	NO		0.087	0.042	0.177	0.069	NO	
outermostmembrane	TGME49_201270	TGME49_201270	0.111	0.106	0.111	0.102	NO		0.034	0.017	0.060	0.017	NO	
outermostmembrane	ATrx1	TGME49_312110	0.265	0.371	0.734	0.418	YES	V	0.370	0.490	0.890	0.468	YES	V
outermostmembrane	APT	TGME49_261070	0.108	0.105	0.157	0.101	NO		0.063	0.044	0.306	0.074	NO	
second inner-most	Toc75/Omp85	TGME49_272390	0.209	0.242	0.685	0.352	YES	V	0.444	0.326	0.993	0.453	YES	V
Innermost	Tic20	TGME49_255370	0.611	0.753	0.99	0.849	YES	Q	0.633	0.701	0.989	0.815	YES	T
Innermost	Tic22	TGME49_286050	0.694	0.58	0.691	0.519	YES	L	0.740	0.755	0.919	0.735	YES	L
luminal	PDH E1b	TGME49_272290	0.415	0.608	0.963	0.77	YES	S	0.539	0.442	0.967	0.653	YES	L
luminal	FabH	TGME49_231890	0.118	0.121	0.157	0.119	NO		0.072	0.073	0.409	0.124	NO	
luminal	PDH E2	TGME49_206610	0.524	0.662	0.961	0.754	YES	Q	0.599	0.625	0.961	0.740	YES	L
luminal	PGKII	TGME49_225990	0.129	0.15	0.279	0.149	NO		0.124	0.080	0.846	0.127	NO	
luminal	FabD	TGME49_225990	0.379	0.37	0.53	0.391	YES	P	0.532	0.530	0.885	0.586	YES	V
luminal	PDH E3	TGME49_305980	0.125	0.23	0.56	0.312	NO		0.127	0.192	0.924	0.495	YES	R
luminal	GyraseB	TGME49_297780	0.109	0.117	0.137	0.117	NO		0.030	0.039	0.310	0.130	NO	
luminal	GyraseA	TGME49_221330	0.609	0.693	0.956	0.763	YES	R	0.679	0.614	0.927	0.069	YES	R
luminal	FNR	TGME49_298990	0.198	0.262	0.641	0.321	NO		0.204	0.358	0.950	0.600	YES	V
luminal	ACP	TGME49_264080	0.47	0.629	0.929	0.744	YES	Y	0.628	0.678	0.970	0.715	YES	F
luminal	TGME49_239680	TGME49_239680	0.275	0.286	0.382	0.291	NO		0.647	0.379	0.985	0.525	YES	S
luminal	TPI-II	TGME49_233500	0.437	0.46	0.765	0.489	YES	F	0.236	0.320	0.913	0.401	YES	F
luminal	FabI	TGME49_251930	0.271	0.399	0.914	0.53	YES	F	0.470	0.402	0.974	0.522	YES	F
luminal	PYKII	TGME49_299070	0.12	0.112	0.132	0.114	NO		0.438	0.079	0.110	0.058	NO	
luminal	RP128	TGME49_209710	0.558	0.477	0.7	0.474	YES	F	0.937	0.707	0.958	0.753	YES	F
luminal	CPN60	TGME49_240600	0.139	0.192	0.349	0.224	NO		0.259	0.359	0.861	0.465	YES	F
luminal	ACC1	TGME49_221320	0.221	0.276	0.52	0.313	NO		0.533	0.446	0.748	0.493	YES	A
luminal	PDH Ea	TGME49_245670	0.128	0.151	0.254	0.148	NO		0.274	0.137	0.876	0.213	YES	P
luminal	ICDH2	TGME49_266760	0.822	0.736	0.892	0.699	YES	S	0.928	0.711	0.901	0.709	YES	R
luminal	YbaK	TGME49_255680	0.107	0.102	0.111	0.099	NO		0.027	0.022	0.072	0.039	NO	
luminal	UROD	TGME49_289940	0.116	0.114	0.163	0.111	NO		0.102	0.139	0.346	0.138	NO	
luminal	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase domain-containing protein	TGME49_055690	0.503	0.653	0.944	0.654	YES	P	0.352	0.501	0.943	0.576	YES	S
luminal	Product: 1-deoxy-D-xylulose-5-phosphate synthase	TGME49_008820	0.11	0.107	0.13	0.106	NO		0.062	0.033	0.119	0.035	NO	
luminal	HU	TGME49_027970	0.235	0.468	0.968	0.72	YES	E	0.364	0.529	0.995	0.700	YES	E
luminal	LipB	TGME49_115640	0.14	0.145	0.34	0.192	NO		0.175	0.116	0.412	0.128	NO	
luminal	RNA helicase	TGME49_291670	0.387	0.497	0.95	0.678	YES	L	0.615	0.572	0.979	0.714	YES	V
luminal	hypothetical	TGME49_059230	0.708	0.556	0.712	0.467	YES	R	0.458	0.626	0.986	0.728	YES	H
luminal	hypothetical	TGME49_039320	0.112	0.132	0.207	0.143	NO		0.052	0.060	0.617	0.172	NO	
luminal	hypothetical	TGME49_039680	0.275	0.286	0.382	0.29	NO		0.647	0.379	0.985	0.525	YES	S
luminal	hypothetical	TGME49_002440	0.178	0.239	0.412	0.24	NO		0.126	0.296	0.947	0.391	YES	W
luminal	hypothetical	TGME49_001270	0.111	0.106	0.111	0.102	NO		0.034	0.017	0.060	0.017	NO	
luminal	NFU	TGME49_021920	0.148	0.222	0.482	0.218	NO		0.415	0.282	0.902	0.338	YES	H
luminal	DNAdDNAh	TGME49_008840	0.111	0.11	0.134	0.104	NO		0.055	0.054	0.527	0.291	NO	

Table A1.S2 – primers used in this study

Primer name	Primer sequence	Purpose
TgToc75_EcoRI_F	ccgaattcatggcggaggaagaaagac	Forward to amplify TgToc75^{78} for ectopic expression in Toxoplasma
TgToc75_78_NsiI_R	ccatgcatagaaactggagaagaccc	Reverse to amplify TgToc75^{78} for ectopic expression in Toxoplasma
TgToc75_95_NsiI_R	ccatgcataagaggggcgggggtgc	Reverse to amplify TgToc75^{95} for ectopic expression in Toxoplasma
TgToc75_277_NsiI_R	ccatgcattcacgatatccacgaaggtacg	Reverse to amplify TgToc75^{277} for ectopic expression in Toxoplasma
TgToc75_512_NsiI_R	ccatgcataaactgcgtcgtctgtcgtctg	Reverse to amplify TgToc75^{512} for ectopic expression in Toxoplasma
TgToc75_790_NsiI_R	ccatgcatagcctgcgaacgacgacgcctc	Reverse to amplify TgToc75^{790} for ectopic expression in Toxoplasma
TgToc75_FL_NsiI_R	ccatgcattgaagctgttgtcggccacg	Reverse to amplify TgToc75$^{full\text{-}HA/Ty}$ for ectopic expression in Toxoplasma
TgSam50_EcoRI_F	ccgaattcatggcggggtcagctcc	Forward to amplify TgSam50$^{full\text{-}HA}$ for ectopic expression in Toxoplasma
TgSam50_NsiI_R	ggatgcatactactcggggagtcttcc	Forward to amplify TgSam50$^{full\text{-}HA}$ for ectopic expression in Toxoplasma
PfOToc75_X_F	aactcgagatgaaaaatgttttaagaaaatatac	Forward to amplify PfToc75^{78} for ectopic expression in Plasmodium
PfToc75_78_A_R	ggcctaggtcttgttgttagcttattccataattc	Reverse to amplify PfToc75^{78} for ectopic expression in Plasmodium
PfSam50_X_F	ctcgagatgtttaattatttttttaagaagc	Forward to amplify PfSam50^{N} for ectopic expression in Plasmodium
PfSam50_60_A_R	aacctaggtaaacaaaaatgcttccaaaataatgg	Reverse to amplify PfSam50^{N} for ectopic expression in Plasmodium
PfOmp85_95_A_R	ggcctaggtcctgtttcttcattttctgtttcc	Reverse to amplify PfToc^{95} for ectopic expression in Plasmodium
Toc75prorepcosf	gtatgcacatgtctcttttctgaatcttttcgcatgagaagcaatgctccatcgaatggtaaccgacaaacgcgttc	Cosmid recombineering to create promoter replacement vector
Toc75prorepcosr	agtccacgactcaaaagagcgaaacgtgtgtttctacggtcgctcaacgtagatctggttgaagacagacgaaagc	Cosmid recombineering to create promoter replacement vector
toc75cosproinserf	acgttgagcgaccgtagaaacacacgtttcgctcttttgagtcgtggactgaatggtaaccgacaaacgcgttc	Cosmid recombineering to create promoter insertion vector.
toc75cosprionsrev	attgaacaccgccgccgtggcga	Cosmid recombineering to create

	cgatgcctgtctttcttcctccgccattttagatctggttgaagacagacgaaagc	promoter insertion vector.
HA_NsiI_F	ccatgcattacccgtacgac	Primer to amplify 3xHA tag
HA_PacI_R	ggttaattaattagagctcggc	Primer to amplify 3xHA tag
Apg-qPCR-F	tctattgcaatggaaaaaggtatg	qPCR to score apicoplast genome
Apg-qPCR-R	tcaatggtagagcaaaggactg	qPCR to score apicoplast genome
UPRT-qPCR-F	actgcgacgacatactggagaac	qPCR to score nuclear genome
UPRT-qPCR-R	aagaaaacaaagcggaacaacaa	qPCR to score nuclear genome
$ACP_{F/A}$mutF	ctgatcaggcctggtgacacagcaccgtaggaagaagcaatgg	Mutagenesis of F at position +1 of ACP to A
$ACP_{F/A}$mutR	ccattgcttcttcctacggtgctgtgtcaccaggcctgatcag	Mutagenesis of F at position +1 of ACP to A
$ACP_{Y/A}$mutF	cctggtgacacaaaaccggcggaagaagcaatggatg	Mutagenesis of Y at alternative position +1 of ACP to A
$ACP_{Y/A}$mutR	catccattgcttcttccgccggtttgtgtcaccagg	Mutagenesis of Y at alternative position +1 of ACP to A