Do membrane proteins travel to the apicoplast in the apicomplexan parasites via the Golgi?
- A biological and evolutionary study.

Dissertation submitted by

Israt Jahan

165300002

In partial fulfilment of the requirements for the award of the Degree of

Master of Science (Biotechnology)

Under the Guidance of

Prof. Swati Patankar

Department of Bioscience and Bioengineering

Indian Institute of Technology Bombay

Mumbai, 400076

2018

Letter of Consent

The work reported in this dissertation entitled **"Do membrane proteins travel to the apicoplast in the apicomplexan parasites via the Golgi? - A biological and evolutionary study."** has been carried out by **Israt Jahan (165300002)** under my guidance in my laboratory. I hereby approve the submission of the project report.

Professor Swati Patankar

Date:

Place: IIT Bombay, Mumbai

Dissertation Approval Sheet

Indian Institute of Technology Bombay

Dissertation Approval Sheet

Dissertation entitled **"Do membrane proteins travel to the apicoplast in the apicomplexan parasites via the Golgi? - A biological and evolutionary study."** submitted by **Israt Jahan (165300002)** is approved for the degree of Master of Science (Biotechnology).

Guide:

Chairman:

Examiner 1:

Examiner 2:

Examiner 3:

Date:

Plagiarism undertaking

I, Israt Jahan, Roll No. 165300002, understand that plagiarism is defined as any one or the combination of the following:

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagrams, etc.).

2. Uncredited improper paraphrasing of pages or paragraphs (by changing a few words or phrases, or rearranging the original sentence order).

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what.

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the other authors of this paper, or guide of the thesis /dissertation may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Israt Jahan

Roll Number: 165300002

ACKNOWLEDGEMENT

I am immensely grateful to Prof. Swati Patankar for giving me the opportunity to learn and hone my skills under her guidance. She has inspired me to drive my scientific query from directionless inquisitiveness to methodological problem solving by careful scrutiny and has motivated me to ask the right questions. She has a great role in shaping my scientific thinking and has influenced my approach to science and beyond.

I express my gratitude towards Pragati for mentoring my work all throughout. I also thank my fellow lab mates Aparna, Chhaminder, and Vishakha for providing me with their valuable feedback and time. I thank Rahul for being a fun lab mate and keeping me indulged in scientific trivia.

I would also take the opportunity to thank IIT Bombay for providing us with top-notch facility and infrastructure that enables us to learn and explore in a stimulating and safe environment. I cannot fail to thank all the teachers of the department who have laid the foundation of knowledge and inculcated qualities of scientific rigor and jest into us.

I am indebted to my parents who have shaped by being, and kept me going with their undying love and support. I thank my siblings; who in fact were sad to know my choice of organism affects cats; for bringing me little bundles of joy.

CONTENTS

Sl. no.	Chapter
I.	Introduction
	1. Trafficking of proteins to the Apicoplast 1.1 Targeting sequences for transport 1.2 Targeting of membrane proteins to the apicoplast
	2. Models of protein trafficking to the apicoplast 2.1 Direct transport 2.2 Golgi independent vesicular transport 2.3 Golgi dependent vesicular transport
	3. Targeting of nuclear encoded proteins to both mitochondria and apicoplast in apicomplexans
II.	Hypothesis
III.	Materials and Methodology
IV.	Results
V.	Discussion
VI.	Future Directions
VII.	Appendix
VIII.	References

List of figures

Sl. No.	Figures
1	Ultrastructure of *T. gondii* tachyzoites
2	Transmission electron micrograph of a *T. gondii* apicoplast
3	Secondary endosymbiosis
4	Apicoplast localization of nuclear-encoded proteins S9 and ACP
5	Fusing N-terminal extension (predicted Signal sequence) of apicoplast targeted protein ACP leads to the secretion of GFP out of the cell
6	An ER retrieval signal does not block targeting to the apicoplast of luminal proteins
7	Apicoplast localization of FtsH1 in multiple membranes
8	A vesicle near apicoplast showing FtsH1 localized
9	Predicted results from the hypothesis
10	Plasmid vector map of pCTG-HA
11	Alanine Scanning Mutagenesis
12	Prediction of transmembrane domains of APT1 using TMHMM
13	Isolation of genomic DNA from *T. gondii*
14	PCR of APT1 using Phusion Hi-fi polymerase
15	Isolation of the plasmid pCTG-HA
16	Restriction digestion of plasmid vector
17	Screening of recombinant by colony PCR
18	Restriction digestion of the recombinant plasmid
19	Recombinant plasmid map of pCTG-APT1-HA
20	Primer designing strategy for addition of HDEL sequence
21	HDEL construct of the gene APT1 obtained from the recombinant plasmid
22	APT1-HDEL insert from recombinant plasmid
23	Restriction digestion of the pCTG-APT1-HA
24	Colony PCR for APT1-HDEL template using restreaked colonies of transformants
25	Restriction digestion of the isolated recombinant plasmids APT1-HDEL
26	Recombinant plasmid map of pCTG-APT1-HA-HDEL
27	Immunofluorescence of parasites transfected with APT1-HA
28	Immunofluorescence of parasites transfected with APT1-HA-HDEL

List of graphs

	SignalP and MitoProt scores for dually targeted proteins
7.1	PfLpLA2
7.2	PfTPxGl
7.3	TgSOD2
7.4	TgIRP/ACN
7.5	TgTPx1.2

I. Introduction

Apicomplexans are parasitic alveolates of the phylum Apicomplexa, the largest taxon of parasitic protists. There are more than 4000 species known, including the infamous *Plasmodium falciparum* and *Toxoplasma gondii* that are parasites of human and other animals. *Plasmodium falciparum* is the causative agent of the major global disease malaria (Diseases, 2010) whereas the *Toxoplasma gondii* causes Toxoplasmosis, a latent disease that is present in 30-50% of the world population (Fledgr et al, 2014) and is deadly in immunosuppressed individuals. Toxoplasmosis also causes birth defects if the infection happens during pregnancy (Desmonts and Couvreur 1974).

A characteristic feature of apicomplexans is the presence of a relict plastid called the apicoplast (Figure 1). The origin of this organelle has been attributed to secondary endosymbiosis of a red alga (Blanchard and Hicks 1999). As a result, the apicoplast is surrounded by four membranes (Figure 2) that have been visualised with electron microscopy and cryo-electron tomography (Lemgruber et al, 2013). The outermost membrane is proposed to have originated from the endomembrane system of the parasite and the periplastid membrane (third membrane) is believed to be the cell membrane of the alga. The inner membranes correspond to the inner and outer membranes of chloroplast from the red algae. (Kohler et al, 1997)

Each parasite contains one apicoplast per cell which is essential for its survival (Fischera, 1995). The apicoplast is home to many important metabolic pathways of the cell like synthesis of heme, isoprenoids, fatty acids and iron sulfur clusters (Van Dooren, Stimmler, and McFadden 2006). Addition of drugs (antibiotics) to the *P. falciparum* cells showed an interesting phenotype named delayed death, which was easier to detect in *T. gondii* because of its replication by endodyogeny. In these cells, the parasite multiplied vigorously in the first vacuole but failed to do so while establishing the second one. This effect does not reverse itself upon the removal of the drug in the second cycle (Dahl and Rosenthal, 2008). Thus, *T. gondii* lacking an apicoplast only survives temporarily and is incapable of host invasion and proliferation (He and Roos, 2001) whereas *P. falciparum* that are functionally apicoplast minus survive upon supplementation of isopentenyl pyrophosphate (IPP), product of the isoprenoid precursor biosynthesis (Yeh and DeRisi, 2011).

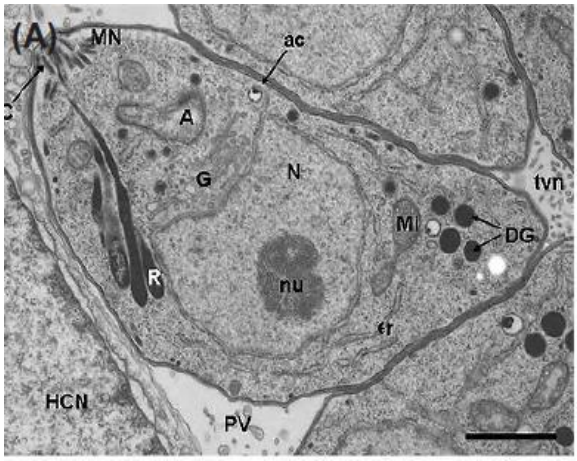


Fig. 1 Ultrastructure of *T. gondii* tachyzoites. A, apicoplast; G, Golgi; R, rhoptries; M, mitochondria; N, nucleus; DG, dense granules; ER, endoplasmic reticulum; MN, micronemes; PV, parasitophorous vacuole; HCN, host cell nucleus

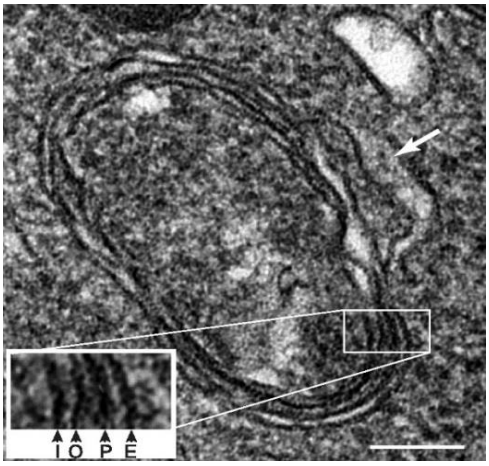


Fig. 2 Transmission electron micrograph of a *T. gondii* apicoplast. The inset shows an enlargement of the four membranes visible in the image. (Fig 1, Parsons et al, 2007)

Even though this vestigial plastid has lost photosynthetic abilities after transferring all of its genes to the nucleus of the parasite, it still carries a small 35 kb remnant chloroplast genome (Wilson et al, 1996). This reduced genome codes for genes that are mostly involved in housekeeping functions such as replication, transcription, translation and possibly plastid retention (Waller et al, 1998). As shown in Figure 3, the transfer of genes to the nucleus of the host is accompanied by eventual degeneration of the cytoplasm and other machinery of the endosymbiont, making the crucial reactions and metabolism of the endosymbiont highly dependent on the genes controlled by the host cell nucleus. The apicoplast genome of *T. gondii* was sequenced (GenBank accession U87145) and found strikingly similar to multiple *Plasmodium* species. This similarity in gene content and organisation between these organisms also run in parallel with their functional need to import proteins that are encoded by the nuclear genes. Incidentally, nuclear encoded apicoplast targeted (NEAT) proteins have been characterised to participate in the crucial metabolic processes undergoing inside the apicoplast. Thus, studying the transport of NEAT proteins is the next hot domain of study in the apicomplexans.

Unlike *P. falciparum* which has a 90% AT rich genome, the *T. gondii* DNA is unbiased. A higher efficiency in transfection, both stable and transient, makes the cell biology studies in *T. gondii* more feasible. The availability of cell markers make advanced microscopic techniques in the parasite relatively easy. With these factors contributing to the ease of manipulation, *T. gondii* remains the choice of organism to study apicomplexans.

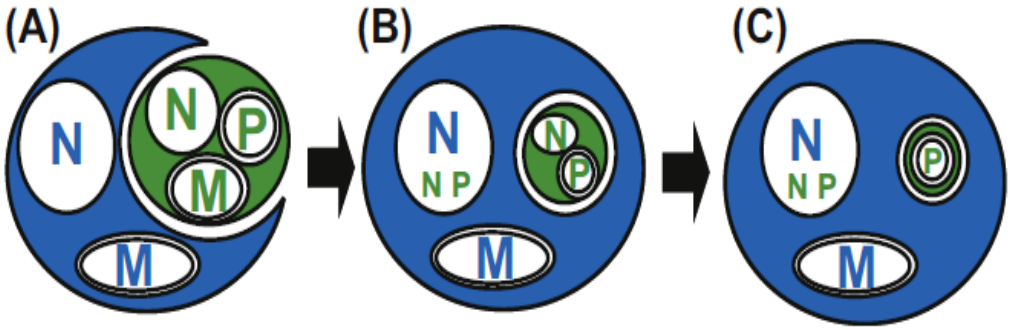


Fig. 3 Secondary endosymbiosis showing engulfment of an algal cell by a eukaryote (A), followed by transfer of genes to the host nucleus (B) and degeneration of the endosymbiont nucleus and cytosol(C). Seeber et al, 2007.

1. Trafficking of proteins to the Apicoplast

1.1 Targeting sequences for transport

The first step towards finding NEAT proteins was to exploit the available expressed sequence tag databases to identify genes for the plausible plastid proteins. Two of the apicoplast proteins thus identified, ACP and ribosomal protein S9, were shown to be localized in the apicoplast and became instrumental in studying targeting of several other proteins to the apicoplast (Waller et al, 1998). Sequence analysis helped in identification of an N-terminal extension in the proteins predicted to reside in the lumen. This N-terminal extension consists of a signal sequence followed by a transit peptide that together is termed as a "bipartite sequence" and mediates targeting to apicoplast (Waller et al, 1998; Bruce 2001). In eukaryotic cells, a signal peptide is used to target a protein to the secretory system while a transit peptide facilitates the targeting of proteins to a primary plastid. These targeting sequences are cleaved at their respective sites of localization upon import. In fact, Waller et al. could localize GFP to the apicoplast using the N-terminal sequence in *T. gondii* (Waller et al 1998) as seen in Figure 4. Furthermore, a GFP construct of the signal sequence of S9 lacking the transit peptide targeted the GFP to be secreted into the parasitophorous vacuole (Figure 5), confirming that the transport pathway starts by entering the secretory system (DeRocher et al, 2000). Other studies in *T. gondii* show that the targeting of proteins to the apicoplast depends on the processing of both the sequences such that deletion constructs and insertional mutagenesis leads to mis-targeting of the NEAT protein (Yung, 2001; Tonkin et al, 2006a).

Interestingly, the transit peptides of NEAT proteins are diverse in sequence and differ greatly in their lengths. These 50-200 amino acid residue chains are analogous to chloroplast transit

peptides with very few acidic and hydrophobic residues present in them (Foth et al, 2003). Studies in *T. gondii* showed while the basic residues are required for targeting of ACP, the overall charge of the peptide was important and not the exact position of the basic residues (Tonkin et al, 2006b). But at the same time, mutational analysis revealed that the basic acidic towards the N-terminal of the sequence had greater contribution in targeting than the distal ones. Surprisingly, a predicted HSP binding site in *P. falciparum* transit peptide important in targeting was complemented by other studies showing these transit peptides indeed exist in a disordered state (Gallagher et al., 2011). Taken together, it shows that unfolded structure might be important in targeting NEAT proteins. The transit peptides of NEAT proteins are processed rather slowly which makes visualization of the native as well as the mature processed form of the proteins possible (DeRocher et al, 2005).

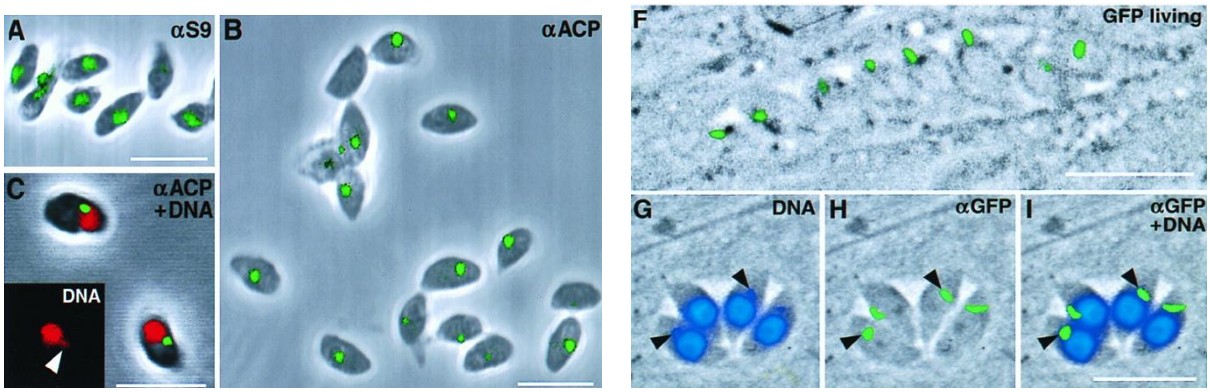


Fig.4 Apicoplast localization of nuclear-encoded proteins S9 and ACP. (*A* and *B*). (*C*) Counterstaining with propidium iodide (red) confirms co-localization of ACP (green) with the extranuclear apicoplast DNA. (*F*) The N-terminal domain of ACP (*Tg*ACP$_{leader}$-GFP) is sufficient to target GFP to the apicoplast, and the recombinant protein can be visualized in living cells. (*G–I*) In fixed cells labelled with anti-GFP (green) and counterstained with DAPI (blue), *Tg*ACP$_{leader}$-GFP can be seen to colocalize with the apicoplast DNA (arrowheads indicate apicoplasts in two cells). Color images were collected independently and overlaid on top of phase-contrast micrographs. (White scale bars = 10 μm, black scale bars = 200 nm). (Fig 2, Waller et al, 1998)

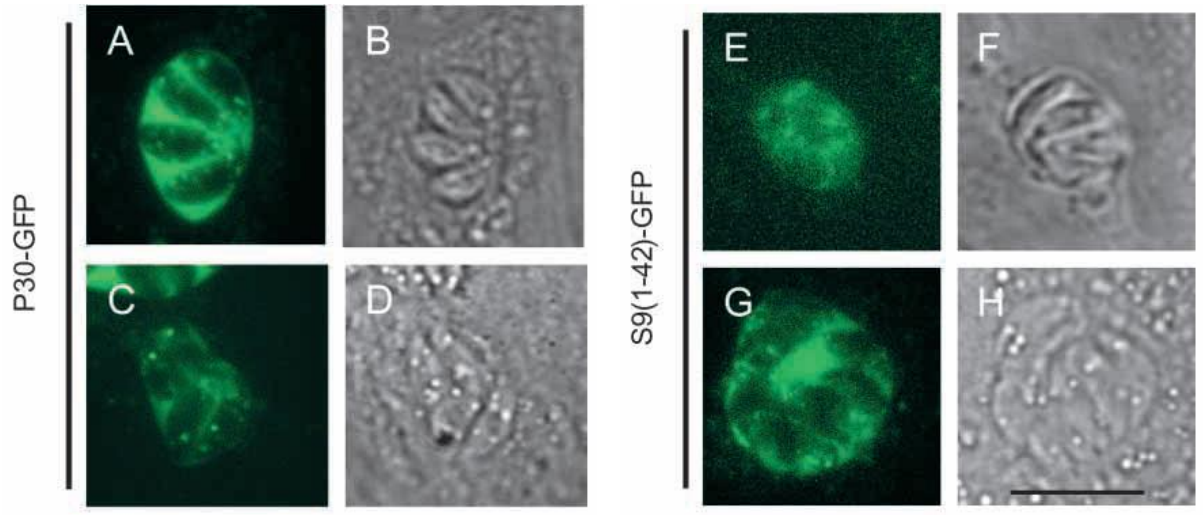


Fig. 5 Fusing N-terminal extension (predicted Signal sequence) of apicoplast targeted protein ACP leads to the secretion of GFP out of the cell. The S9 signal sequence directs secretion of GFP. *T. gondii* were stably transformed with constructs encoding P30-GFP (A-D) and S9 (1-42)-GFP (E-H). Cells shown in (C, D, G, H) were treated with cycloheximide for 1 hour prior to viewing. Green fluorescence and transmitted light pictures of live cells are shown. Bar, 10 mm. (Fig 3, DeRocher et al, 2000)

1.2 Targeting of membrane proteins to the apicoplast

Studies concerning the transport and localization of NEAT luminal proteins generated interest towards the proteins that worked as the machinery for the import of these luminal proteins as well as the export of products across the four membranes. Under immuno-electron microscope, certain proteins were observed to show a ring-like staining pattern around a luminal marker as well as to localize near the apicoplast. These studies involved deconvolution microscopy and immunofluorescence analysis. The characterization of these membrane proteins and locating them to individual membranes was challenging due to the close proximity with which the four apicoplast membranes lie next to each other.

The proteins predicted to reside in the membranous compartments of the apicoplast when studied for their targeting sequences showed interesting results. The *P. falciparum* Tic22 and Der1-1, along with *T. gondii* Tic20 were the first ones studied to have a recognizable bipartite sequence (Kouranov et al, 1999; Kalanon et al, 2009; Spork et al, 2009). The Tic20 and Tic22 proteins, homologous to the components of TIC complex of the chloroplast translocon, are highly likely to reside in the innermost membrane of the apicoplast and the adjacent intermembrane space whereas PfDer1-1 is likely to be present on the periplastid membrane of the apicoplast.

On the other hand, several other proteins predicted to be localized on the apicoplast membrane do not possess the recognizable bipartite signal sequence required for transport. Though such proteins contain sequences in their N-termini that mediate their entry into the secretory system, these sequences are unlike the conventional signal peptide that is cleaved after the protein is imported into the ER. In *T. gondii,* the proteins ATrx1 and FtsH1, carry anchor sequences in their N-terminal by virtue of which they enter the ER but lack a transit peptide (Karnataki et al, 2007a; DeRocher et al, 2008). Surprisingly, a phosphate transporter in *T. gondii* APT1 as well as a sugar phosphate transporter of the outer membrane in *P. falciparum*, both have been shown to not possess any recognizable signal sequence (Karnataki et al, 2007a; Mullin et al, 2006). It is hypothesized that the multiple transmembrane domains in the N-terminal of the protein compensate for the absence of signal and transit peptides. Another interesting finding is the presence of bipartite sequence in the paralogue iTPT of the same protein in *P. falciparum*, which most likely resides in the innermost membrane in the organelle (Mullin et al, 2006).

The molecular mechanism underlying the transport of proteins that lack the transit peptide to the apicoplast remains unaddressed. ATrx1 has been shown to require additional 200 residues important for its targeting, in addition to the anchor sequence present. Serendipitously, more information is present with respect to APT1 localization and targeting to the apicoplast. The N-terminal region prior to the first transmembrane domain of the APT1 protein has been shown to be necessary for its localization (DeRocher et al, 2012). A tyrosine-glycine motif was identified, the mutation of which led to APT1 mislocalization. When the tyrosine residue was mutated to alanine, the protein localized to the Golgi but a mutation in glycine led to partial mislocalization. While it is striking to find specificity offered by a motif so small as compared to the eukaryotic tyrosine based motifs that function in secretory systems, tyrosine potentially participates in protein-protein interactions. When an intact YG motif is placed within the first 7 amino acids, the protein was shown to mislocalize. This motif faces the cytosol and has not been identified in any other membrane protein studied yet.

The peculiarities in the proteins mentioned make them interesting targets for further studies involving trafficking of membrane proteins to the apicoplast.

2. Models of NEAT protein trafficking to the apicoplast

Protein targeting studies about signal sequences clearly account for NEAT proteins entering the ER, and their passage into the apicoplast lumen via the transit peptide. There remain unanswered questions about the transportation events in between. This has led to the proposal of three pathways that can be taken by a NEAT protein from the ER to the apicoplast:

2.1 Direct transport without vesicles

2.2 Golgi independent vesicular transport

2.3 Golgi dependent vesicular transport

Experimental evidence is available for the second and third pathways and will be discussed.

2.1 Direct transport without vesicles

This hypothesis proposes that the apicoplast literally lies in the secretory system with the apicoplast membranes being extensions of the ER. This would facilitate the entry of the proteins directly to the apicoplast without the requirement of any vesicles. According to this, the NEAT proteins would be getting trapped inside the organelle by virtue of their transit peptide. This model is based on the observation of ribosomes on the surface of secondary plastids in some organisms (Gibbs, 1981). While the apicoplast lies in close proximity with the nucleus of the cell (Tomova et al, 2009), where the nuclear envelope is indeed an extended ER (Hagger et al, 1999), there has been no evidence showing the presence of ribosome on the outermost apicoplast membrane, thus nullifying the theory. Moreover, the absence of transit peptide in some of the proteins as described earlier would make discrimination between NEAT and proteins targeted to other destinations quite confusing according to this model. In fact, the apicoplast membrane has been shown to not fuse with ER (Tomova et al, 2009) or nuclear envelope of the cell which lies in 10 nm distance (Kudreyashev et al, 2010). But at the same time, mitochondria, which lies in great proximity to the apicoplast such that these organelles are called the sticky duo, has been shown to transiently interact with the ER. Moreover, the similarity of lipid composition of the ER extension with the apicoplast membranes hint towards speculative transient contact between the ER and apicoplast.

While this model is highly unlikely, indirect correlations demand more studies to disprove this hypothesis.

2.2 Golgi independent vesicular transport

The speculations surrounding the above model led to ideas about the proteins being transported to the apicoplast from the ER in vesicles, either directly or indirectly. Studies on trafficking of luminal proteins indicated a vesicular transport that is independent of Golgi i.e. it doesn't travel via Golgi to the apicoplast lumen from the ER. The vesicles carrying these proteins have not been described but there is ample evidence to suggest that luminal proteins are exempted from the Golgi route of the secretory system. DeRocher et al, elucidated that disruption of Golgi trafficking had no effect on the localization of a marker protein. This could be achieved by exposure to 16°C or by subjecting the cells to Brefeldin A, an inhibitor of Golgi trafficking (DeRocher et al, 2005). In order to make sure there are no false results due to the proteins level already present in the cells, the exit of proteins from the ER was blocked using a conditional aggregation domain (CAD) tag which allows transport of the marker protein upon binding to a synthetic ligand. Depletion of the protein in apicoplast for two days followed by addition of the ligand still localized the protein to the apicoplast showing no effect of the inhibitor or low temperature.

Additional lines of evidence came from studies involving the addition of ER retrieval sequences to the luminal proteins. The ER performs the role of sorting and assembling the proteins to be targeted to various locations in the cell. While all proteins that travel via the secretory pathway are transported to the Golgi in bulk from the ER, the proteins that are resident of the ER, have been mistargeted or are assembled incorrectly are localized back to the ER by virtue of an ER retrieval sequence.

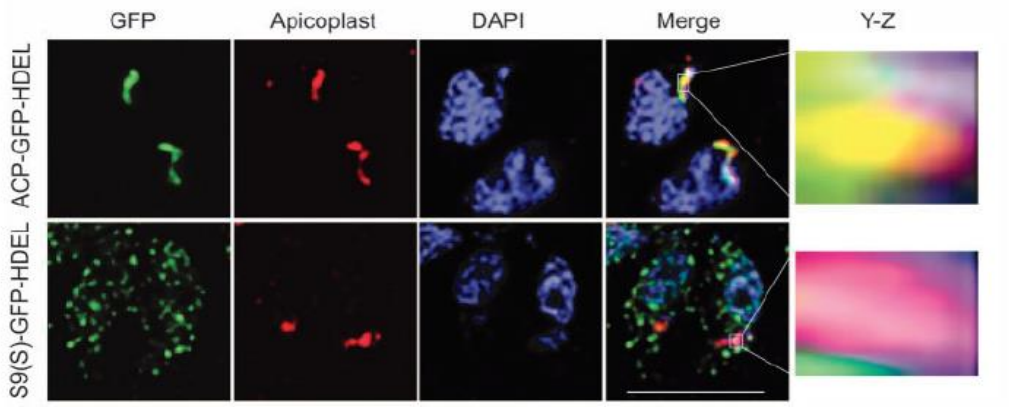


Fig. 6 An ER retrieval signal does not block targeting to the apicoplast of luminal proteins. (A) *T. gondii* expressing ACP-GFP-HDEL and S9(S)-GFP-HDEL were analyzed by IFA and visualized with anti-GFP, and anti-BiP [S9(S)-GFP-HDEL only], the apicoplast marker quantum red streptavidin and DAPI. A merge of anti GFP or anti BiP, quantum red streptavidin and DAPI is also shown. A merged projection of the marked apicoplasts in the *y-z* dimension is shown. Fusing N-terminal extension (predicted Signal sequence) of apicoplast targeted protein ACP leads to the secretion of GFP out of the cell. (Fig 1, DeRocher et al, 2005)

Incorporation of an ER retrieval sequence to the luminal proteins ACP and S9 had no effect on the localization of the proteins (Figure 6) thus confirming that the luminal proteins are not targeted to the apicoplast *en route* Golgi (DeRocher et al, 2005; Tonkin et al, 2006c).

While it is tempting to assume that the proteins are not taking the secretory pathway because of their apparent non-involvement with Golgi, it is very much plausible that the vesicles speculated are not the ones originated form the ER-Golgi route and are separate bizarre vesicles that take part specifically in apicoplast protein transport.

2.3 Golgi dependent vesicular transport

Unlike the luminal proteins, there are vesicles containing proteins shown to localize at the outer compartments of the apicoplast in *T. gondii*. These structures have been visualised by immunofluorescence and solved using immunoelectron microscopy. The tagged forms of APT1, FtsH1 and ATrx1 appear as dots upon immunofluorescence. These vesicles as well as the apicoplast membrane have been shown to bear phosphatidylinositol-3-phosphate. Although the importance of PI3P in apicoplast biogenesis has been demonstrated, its function is not elucidated in *T. gondii* (Tawk et al, 2011). Surprisingly, these vesicles bearing the tagged proteins still persist in cells that have lost their apicoplast due to overexpression of PI3P binding domain indicating towards their improbable origin from apicoplast. The over-expression of these domains is thought to interfere with the biogenesis of apicoplast. In mutational studies of APT1, the protein has been seen to localize at the apicoplast as well as in these large vesicles. This observation makes it tempting to denote these vesicles as carriers of membrane proteins, though such exclusivity would need more evidence for such a statement.

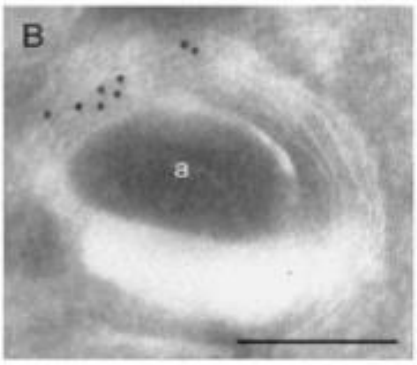


Fig. 7 Apicoplast localization of FtsH1 (Fig 4, Karnataki et al, 2007)

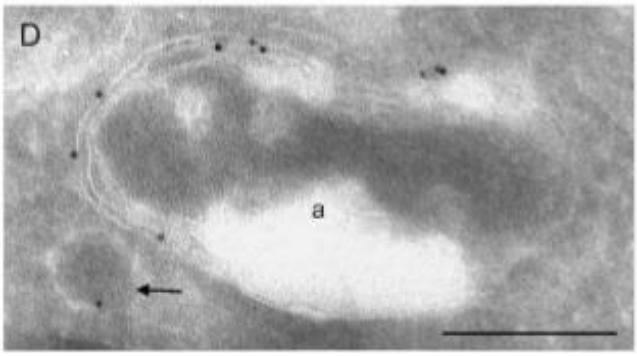


Fig. 8 A vesicle near apicoplast showing FtsH1 localized (Fig 4, Karnataki et al, 2007)

The Golgi pathway has been elucidated better in other secondary endosymbionts like *Euglena*. Immunoelectron microscopy of Rubisco, a nuclear encoded protein to the three membraned

plastid of *Euglena*, showed its trafficking via the Golgi (Nassoury, Cappadocia et al. 2003). In *P. falciparum*, the antioxidant enzyme glutathione-like thioredoxin peroxidase (PfTPxGl) which localizes to the apicoplast outermost membrane, showed insights about the Golgi dependent pathway. This protein was not transported when the parasites are treated with BFA suggesting its possible trafficking *en route* Golgi. Moreover, studies involving disruption of vesicular transport completely by inhibitor Aluminium Tetrafluoride (AlF_4-) showed PfTPxGl to be absent and antibodies against the protein gave signals around the periphery of the apicoplast. This inhibitor works by blocking the fusion of membranes of the vesicles and the target organelle, thus stopping vesicular transport entirely (Taraschi, Trelka et al. 2001). Interestingly, PfTPxGl shows dual targeting and localization. Even though it has one bipartite sequence, it is localized to apicoplast as well as mitochondria (Chaudhari et al, 2012).

Similar results have been generated from the lab with the *T. gondii* homologue of the same protein. TgTPxGl has been shown to localize to the apicoplast as well as the mitochondria. An ER retrieval sequence when attached to the C-terminal protein led to its trafficking back and capture in the ER showing that it reaches apicoplast via the Golgi.

3. Dual Targeting to endosymbiotic organelles: Targeting of nuclear encoded proteins to both mitochondria and apicoplast in apicomplexans

Mitochondrial targeting sequences resemble plastid transit peptides. These topogenic signals are rich in basic amino acids and show considerable flexibility in sequence. The mitochondrial targeting sequences generally adopt an amphipathic structure, which is not essential for transit peptides (reviewed in Haucke and Schatz, 1997). The NEAT protein S9 was one of the first sequences shown to function as a mitochondrial targeting sequence in *T. gondii*. A signal sequence and a predicted plastid targeting sequence are encoded at the N-terminal by the *T. gondii* S9 gene. These sequences together enable the targeting of GFP to the apicoplast. However, in the presence of only the transit sequence, GFP was targeted to the mitochondrion. The diversion of *T. gondii* apicoplast proteins into the secretory system aides the delivery of proteins to the correct location. Thus, the cleavage of signal sequence is involved in revealing the cryptic mitochondrial targeting, but mitochondrial targeting can no longer occur at this point.

In a study performed by Soldati and colleagues, the targeting of *T. gondii* SOD2 was investigated. Instead of using the targeting sequence fused GFP, the full protein was fused to a

C-terminal epitope tag. The protein encoded by this construct localized to both the apicoplast and the mitochondrion (Pino et al., 2007). Because the construct employed a cDNA sequence, alternative splicing was ruled out, and because the presequence lacked additional methionines after the start codon, alternative translation initiation was also ruled out. Although common in plants, very few proteins are currently known to be localized to both the plastid and the mitochondrion in apicomplexans. Use of alternative start codons is another means by which dual localization to the apicoplast and mitochondrion can be achieved (Saito et al., 2008). Here, the predicted presequence of pyruvate kinase II contains multiple methionines. When the third methionine was mutated to alanine, the protein was no longer found in the mitochondrion.

One of the more complex examples is a thioredoxin dependent peroxidase (TgTPx1.2) which is alternatively spliced to yield different 50 transcript ends (Pino et al., 2007). cDNAs corresponding to these splicing variants, when transfected into *T. gondii*, yield proteins localized to the cytosol (shorter cDNA) or, apicoplast and mitochondrion (longer cDNA). Brydges and Carruthers (2003) reported an unexpected yet interesting result when assessing the subcellular location of a *T. gondii* iron superoxide dismutase (FeSOD20). Changing a single amino acid in the 'signal' sequence from arginine to alanine was sufficient to render it functional, diverting GFP to the apicoplast (Brydges and Carruthers, 2003).

Several other examples of proteins dually targeted to the apicoplast and mitochondrion have been identified since then in *T. gondii* (Pino et al., 2007) and *P. falciparum* (Günther et al., 2007; Kehr et al., 2010). One such protein is *P. falciparum* glutathione peroxidase-like thioredoxin peroxidase (PfTPxGl) which showed interesting stochastic results in a study in our lab (Chaudhari et. al., 2012). This protein has been shown to localize to the apicoplast, the mitochondrion and the cytosol by immunofluorescence imaging and bioinformatics predictions. The distribution of PfTPxGl was random in the population, with the protein localizing to any one organelle in some parasites and to both in others.

These experiments suggest that the context of the N-terminal targeting sequence also affects localization and point to the need for caution in ascribing cellular localizations on the basis of predicted trafficking sequences and experiments with tagged proteins. They also indicate that dual targeting of nuclear encoded proteins is a factor in organelle function. The study we perform here in aimed at understanding the targeting sequences of such dually localized proteins with an aim of identifying mutations that can yield a stochastic distribution to the mitochondria and apicoplast of the cell.

II. Hypothesis

The membrane proteins take the evolutionarily conserved secretory pathway for apicoplast targeting

The secretory system of the eukaryotic cell is responsible for trafficking of proteins that are to be secreted out of the cell or the membrane proteins of intracellular organelles. It starts off with the ER where the proteins are sorted based on their destinations and transported to the Golgi in vesicles. These vesicles fuse with the target membranes, thus delivering the destined proteins. The proteins in these vesicles can be present in two ways: they can either reside inside the lumen of the vesicle or they can be integrated to the membrane of the vesicle itself. During an early period of endosymbiont evolution, the translocons and transporters on cell membrane as well as on the membranes of intracellular compartments are believed to have reached through ER-Golgi vesicular transport. This is speculated to have happened across all cell types and marks a crucial event in the formation of organelles in early evolution.

The available evidence in the trafficking of proteins to the apicoplast suggest a hypothetical trend wherein only the membrane proteins have been associated with the trafficking route involving Golgi. Our hypothesis is the apicoplast membrane proteins also follow the classical evolutionarily conserved secretory pathway taken by proteins to the membranous compartments in eukaryotic cells.

To test this hypothesis, we are exploiting the ER retrieval pathway in eukaryotes by which the ER-resident proteins and the proteins with incorrect folding, assembly or mis-targeting are trafficked back to the ER. In fact, the translocon Sec61 of ER and other proteins involved in vesicle formation, which is the transporter for the proteins targeted to intracellular compartments, is retained in the ER by virtue of its ER- retrieval sequence. The most common ER retrieval sequences are the KDEL and HDEL sequences that are present in the C-terminal of the targeted proteins.

Unpublished data from the lab has shown that attaching an HDEL sequence to the luminal protein of apicoplast SOD2 doesn't localize it to the ER whereas when the same is done for the membrane protein TgTPx1.2, the protein can be visualised to be present in the ER under immunofluorescence. This suggests exclusivity of the ER-Golgi pathway for at least one membrane protein. However, to generalize it to the other proteins, proteins with varied targeting sequences are chosen, which also have been shown to localize at different

compartments of the apicoplast periphery. The chosen proteins APT1, Tic20, ATrx1, and FtsH1 should be localized to the apicoplast in the absence of C-terminal HDEL sequence whereas addition of the same is expected to localize the proteins in the ER of the cell.

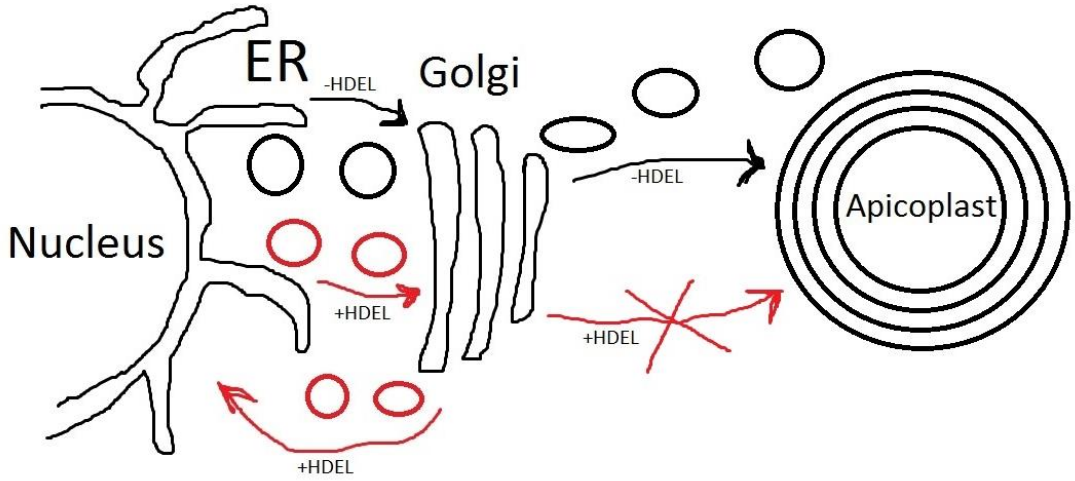


Fig. 9 The figure shows the predicted path taken by the membrane proteins in the presence and absence of ER-Retrieval (HDEL) sequence. The black circles represent vesicles carrying proteins without HDEL and red circles represent vesicles with proteins having HDEL attached to them.

III. Materials and Methods

1. Cloning of the gene APT1 into the vector pCTG-HA

1.1. Isolation of genomic DNA from the parasite cells

APT1 gene is devoid of introns and hence can be directly cloned from the genomic DNA. To obtain the gene and insert into a plasmid vector, the genome of *Toxoplasma gondii* RH was isolated using the *T. gondii* RH pellet stored in -80°C. The pellet was dissolved in 200 µl 1X PBS (pH 7.4) buffer, followed by addition of 700 µl Lysis Buffer [50 mM Tris HCl (pH 8.0), 200 mM NaCl, 20 mM EDTA (pH 8.0), 1% SDS] containing 20 µg/ml RNase A. The solution was incubated at 37°C for an hour during which the tube was inverted intermittently. 20 µl of Proteinase K (20 mg/ml) (VWR, Amresco Life Science) was added following which the tube was incubated at 56°C for an hour with intermittent mixing. To this solution, 1 ml of Phenol: Chloroform: Isoamylalcohol (PCI) (25:24:1) was added and mixed by inverting the tube. Centrifugation was done at 13800 rpm for 10 min. The upper aqueous phase of the solution was transferred to a fresh microfuge tube with an addition of 3M Sodium Acetate (pH 5.2) corresponding to $1/10^{th}$ of the volume. Absolute ethanol 2.5 times the volume of the resulting solution was added. The solution was kept at -20°C overnight for precipitation of the nucleic acid. Centrifugation at 13800 rpm for 30 min was done the following day. The pellet obtained was washed with 500 µl of 70% ethanol and centrifuged at 13800 rpm for 20 min. The supernatant was discarded and DNA pellet kept for drying on a heat block at 45 °C. The dried pellet was dissolved in 10 µl of sterile distilled water and stored at -20°C.

1.2. PCR amplification of APT1 from genomic DNA

The gene sequence of APT1 was obtained from the ME49 strain (RH strain has not been sequenced) and used for designing primers.

Primers for APT1:

TgAPT1F: TAAACAATTG**GACAAA**ATGGAGGAATCGAAACGCTTGGG

TgAPT1R: GGATCATATGTCCGTACTTGGTCTTCGAGAGAG

Conditions: 0.2 µM oligo, 2 mM Mg^{2+}, 0.2 mM dNTP

CAATTG is the restriction site for Mfe1 and CATATAG is the restriction site for Nde1. GACAAA is the general *T. gondii* Kozak sequence.

A PCR reaction of 25 µl containing 5 µl of 5X Phusion HiFi buffer, 0.5 µl of 10 pM of each primer, 0.5 µl of dNTPs and 1 µl of genomic DNA (16.5 ng/µl) template was set up for 30 cycles with the following conditions:

	Temperature(°C)	Time
Initial denaturation	98	5 mins
Denaturation	98	20 secs
Annealing	72	15 secs
Extension	72	45 secs
Final Extension	72	10 mins

1.3. Isolation of the plasmid vector from *E. coli* DH5α cells

E. coli DH5α cells containing the plasmid vector were grown overnight in 5 ml LB containing 100 µg/ml Ampicillin in a shaker incubator at 37°C. The following day, 3 ml culture was pelleted at 10000 rpm for 5 min. The culture was used for plasmid isolation by the alkaline lysis method as per Maniatis protocol. The isolated plasmid was stored at -20°C till further use.

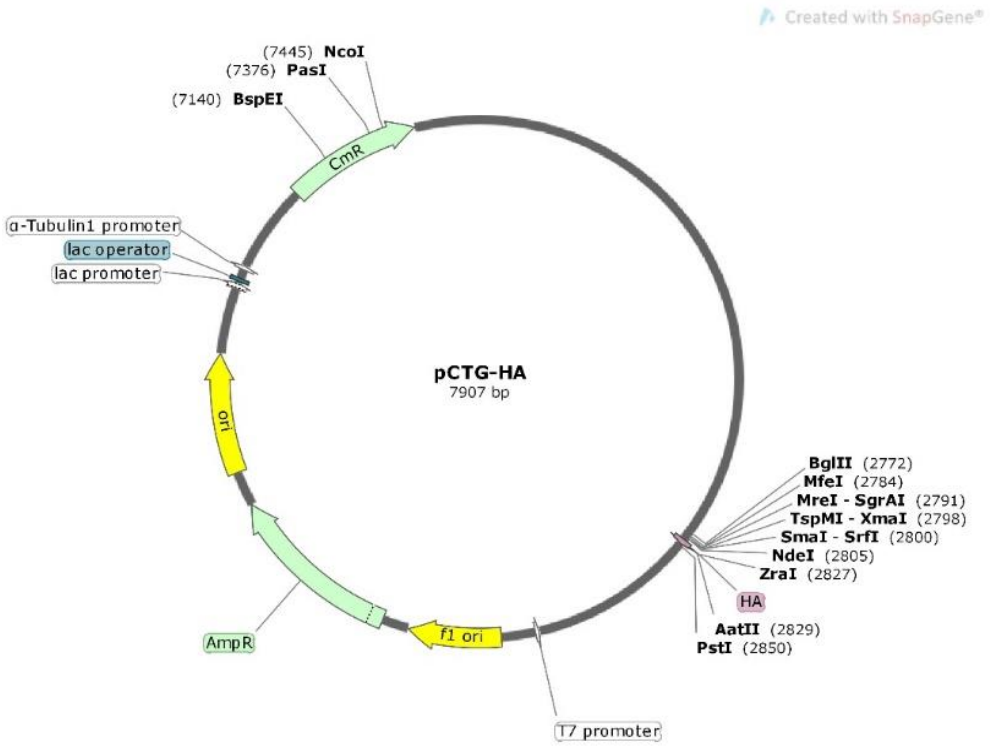


Fig. 10 Plasmid vector map of pCTG-HA

a. Restriction Digestion of the plasmid and insert

The PCR product of the insert APT1 and the isolated plasmid DNA, both were digested using the enzymes MfeI (FD0754, ThermoFisher Scientific) and NdeI (FD0584, ThermoFisher Scientific). 1 µg of vector DNA and 350 µg of the insert DNA were doubly digested in separate 30 µl reaction mixtures. Additionally, the vector was dephosphorylated using alkaline phosphatase (EF0654, ThermoFisher Scientific) to prevent re-ligation of the sticky ends of the plasmid. The reaction mixtures were kept at 37°C for 6 hours and the digested products were precipitated using PCI purification method.

b. Ligation of the digested vector and insert

The ligation reaction was set using 30 µg of vector with a 1:3 ratio of vector: insert. In addition to this, two control reactions, the vector alone without ligase and the vector with ligase were also set. All three reaction mixtures were kept at 22°C for 5 hours using T4 DNA ligase.

c. Transformation in *E.coli* DH5α cells

E. coli competent cells (stored at -80°C, 200 µl) were thawed on palm and immediately incubated on ice for 10 min. The ligation mixture was added to the competent cells and incubated on ice for 30 min. The samples were given heat shock for 90 sec on a heat block at 42°C. Incubation was done for 10 min on ice. 900 µl LB was added to the sample and incubated at 37°C for an hour with a continuous shaking at 200 rpm. The samples were centrifuged at 6000 g for 3 min at 4°C and the pellet dissolved in 100 µl of the sample remaining after discarding 900 µl of the supernatant. This small volume was spread on an LA plate containing Ampicillin. The plates were kept for incubation at 37°C overnight.

d. **Screening for recombinants**

The colonies obtained on the LA plate is a mixed population of cells that have taken up the plasmid containing the desired insert APT1 (recombinants) and the ones which took up the plasmid without the gene of interest. The presence of recombinants can be screened with colony PCR. The reaction mixture of 25 µl contained a pinch of colonies, 25 mM $MgCl_2$, 10 mM dNTPs, 10 pM primers specific for the gene APT1, and the Taq polymerase enzyme. The colonies showing a positive result in the PCR were inoculated in LB with Ampicillin and mini-prep plasmid isolation was performed, following which the plasmids were digested with the enzymes Mfe1 and Nde1 to confirm the release.

2. Creating the APT1-HDEL construct and cloning it into the vector pCTG-HA

2.1 Obtaining the APT1-HDEL insert

The recombinant plasmid vector pCTG-APT1-HA was used as a template to obtain the insert APT1-HDEL via PCR. The reverse primer contains the HDEL sequence which is incorporated into the C-terminal of the gene in the PCR.

The primers used for this PCR were:

TgAPT1F: TAAACAATTG**GACAAA**ATGGAGGAATCGAAACGCTTGGG

TgHDEL: CGACGTCCCTGACTACGCGCACGACGAGTTG**TAG**CTGCAGTAAT

Conditions: 0.2 µM oligo, 2 mM Mg^{2+}, 0.2 mM dNTP

CAATTG is the restriction site for Mfe1 and CTGCAG is the restriction site for Pst1. GACAAA is the general *T. gondii* Kozak sequence.

	Temperature(°C)	Time
Initial denaturation	98	5 mins
Denaturation	98	20 secs
Annealing	72	15 secs
Extension	72	45 secs
Final Extension	72	10 mins

2.2 Isolation of the plasmid vector from *E. coli* DH5α cells

E. coli DH5α cells were used for plasmid isolation by the alkaline lysis method as per Maniatis protocol. The isolated plasmid was stored at -20°C till further use.

2.3 Restriction Digestion of the plasmid and insert

The PCR product of the insert APT1-HDEL and the isolated plasmid DNA, both were digested using the enzymes MfeI (FD0754, ThermoFisher Scientific) and PstI (FD0614, ThermoFisher Scientific). 1 µg of vector DNA and 300 µg of the insert DNA were doubly digested in separate 30 µl reaction mixtures. Additionally, the vector was dephosphorylated using alkaline phosphatase (EF0654, ThermoFisher Scientific) to prevent re-ligation of the sticky ends of the plasmid. The reaction mixtures were kept at 37°C for 14 hours to ensure complete digestion and the digested products were precipitated using PCI purification method.

2.4 Ligation of the digested vector and insert

The ligation reaction was set using 50 µg of vector with a 1:5 ratio of vector: insert. In addition to this, two control reactions, the vector alone without ligase and the vector with ligase were also set. All three reaction mixtures were kept at 22°C overnight using T4 DNA ligase.

2.5 Transformation in *E.coli* DH5α cells

E. coli competent cells (stored at -80°C, 200 µl) were thawed on palm and immediately incubated on ice for 10 min. The ligation mixture was added to the competent cells and incubated on ice for 30 min. The samples were given heat shock for 90 sec on a heat block at 42°C. Incubation was done for 10 min on ice. 900 µl LB was added to the sample and incubated at 37°C for an hour with a continuous shaking at 200 rpm. The samples were centrifuged at 6000 g for 3 min at 4°C and the pellet dissolved in 100 µl of the sample remaining after discarding 900 µl of the supernatant. This small volume was spread on an LA plate with Ampicillin in it. The plates were kept for incubation at 37°C overnight.

2.6 Screening for recombinants

The colonies obtained on the LA plate is a mixed population of cells that have taken up the plasmid containing the desired insert APT1 (recombinants) and the ones which took up the plasmid without the gene of interest. The presence of recombinants can be screened with colony PCR. The reaction mixture of 25 µl contained a pinch of colonies, 25 mM $MgCl_2$, 10 mM dNTPs, 10 pM primers specific for the gene APT1, and the Taq polymerase enzyme. The colonies showing a positive result in the PCR were inoculated in LB with Ampicillin and mini-prep plasmid isolation was performed, following which the plasmids were digested with the enzymes Mfe1 and Nde1 to confirm the release of the insert.

3. Maintenance of *Toxoplasma gondii* parasites, transfection of recombinant vectors and visualization for localization of proteins

Toxoplasma gondii is an obligate parasite capable of infecting almost all nucleated cells. The cell line used for culturing the parasites is derived from Human Foreskin Fibroblast (HFF) cells. Even though other large spread out cells like Vero, 3T3 etc. can be used to culture, we have chosen Human Foreskin Fibroblast (HFF) cells for growing the parasites. The immortal cell lines are easily available and give higher parasite yield but our choice of the primary cell line of HFF cells is based on the fact that these cells show contact inhibition and can be maintained for longer durations because of their slow growth. These cells have a definite life duration and die with increasing passage number and hence have to be maintained more carefully than the immortal cell lines.

3.1 Culturing of HFF cells

The HFF cells are grown in complete Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum pH 7.4 and gentamycin, using TC treated (collagen coated) T25 flasks. The cultures are maintained in a 37° C incubator at 5% CO_2 levels. Since a limited number of cell divisions is allowed for the primary cell lines, early passage cells are preserved in liquid nitrogen (cryopreservation) and are used when required by thawing. When the cells cover 80-90% of the surface of the bottom of the flask, they are said to have attained confluency, or in other terms are confluent, the cells are distributed further into new TC treated T25 flasks by a procedure called splitting. The cells from one flask are split into 5 new flasks, thus the new flasks have the cell content of the parent culture diluted in 1:5 ratio. This splitting of cells was performed as follows:

A T25 flask with confluent culture was taken, and the media aspirated. The bottom of the flask was washed with 5ml 1X PBS buffer. After removing this solution, 1 ml 1X PBS buffer and 100 µl Trypsin was added into the flask. Trypsin is a protease which detaches the cells from the flask making them non-adherent to the surface. The cells suspend into the solution and can be transferred to other flasks. 2-3 min incubation at 37°C is required for the trypsinization to occur efficiently. This can be confirmed by looking at the flask under an inverted microscope at 40X magnification. The cells that were spread out initially, now appear nearly circularized and move in the solution. 4 ml of media pre-warmed at 37°C was added to the 1ml solution in the flask and mixed by gently moving the flask. 1 ml from this flask distributed to the other 5 flasks containing 4 ml pre-warmed DMEM each.

The nutrients in the media get depleted as the cells use it up for growth. To maintain the confluent cells for longer durations when not split, the media must be changed every 4-6 days. Confluent cultures are infected with parasites to maximize the yield upon harvesting of parasites.

3.2 Growing *Toxoplasma gondii* parasites in the HFF cells

The *T. gondii* cultures are kept for cryopreservation and thawed upon requirement. A host cell culture infected with the parasites can be used as inoculum for infecting new flasks. This requires the parasites to be in a freshly lysed state such that their viability is maintained. The cryopreserved culture stocks are not disturbed frequently in this manner.

Typically, a 0.5-1 ml of *T. gondii* RH infection takes around 2-3 days to lyse completely. Under an inverted microscope, the parasites can be seen to form rosette like structures inside the parasitophorous vacuole it forms in the infected cell. Egression of the vacuole releases the parasites from the cell enabling them to enter the neighbouring cells for infection and thus increase in number. When almost all cells of the HFF culture appear to be infected in large numbers suggesting a good expected yield of the parasites, the HFF cells are subjected to "harvesting" i.e. isolation of the parasites from the infected cells. This was carried out as follows:

The flask bottom was scraped using a sterile cell scraper to detach the HFF cells adhered to its surface and rinsed with 1X PBS buffer. The flask was and this solution transferred to the tube as well. The resulting solution in the polystyrene tube was passed through a membrane filter twice using a 10 ml syringe, first with a 20 gauge needle followed by a 23 gauge needle into the tube itself. This pressurised filtration of the cell suspension through a small opening causes the cells to rupture releasing the parasites from the host cells. 5 ml 1X PBS buffer was passed through the filter to collect the residual parasites and dilute parasite suspension. This was collected in a fresh polystyrene tube. The sample obtained was centrifuged at 1500 rpm for 10 mins and the pellet suspended in 1 ml 1X PBS.

To count the number of parasite obtained, 10 µl of the resuspended solution was added to 90 µl 1X PBS making it 1:10 diluted. The cells were counted using a haemocytometer and calculations done. The remaining sample was centrifuged at 1500 rpm for 10 min and the pellet stored at -80°C.

3.3 Transfection of the recombinant vectors pCTG-APT1-HA and pCTG-APT1

Freshly growing *T. gondii* tachyzoites were harvested from confluent HFF flask and resuspended in 400 µl of incomplete DMEM (without serum and gentamycin). This suspension was transferred to a 2 mm gap electrocuvette and 50 µg of plasmid DNA (34 µl of pCTG-APT1-HA and 13 µl of pCTG-APT1-HA-HDEL as calculated after maxi prep of the plasmid DNA) was added. The parasites were electroporated with a single pulse of 1500 V, 25 F capacitance and 50 Ω resistance in a BioRad electroporator. This was done separately for both the DNA transfections. The transfected parasites were transferred to a 4-well chamber slide (which had HFF cells in a confluent state) and incubated at 37°C for 24 hours.

3.4 Preparation of slides for visualising the localization of the proteins via Immunofluorescence

A confluent chamber slide with one of its adjacent compartments transfected with pCTG-APT1-HA and pCTG-APT1-HA-HDEL plasmids, respectively was taken for preparation of imaging slides. To use antibodies to stain cells, medium was removed from the chamber slides. The cells were fixed with 400 µl of 4% paraformaldehyde solution in PBS and kept in dark at RT for 30 min. The fixative was removed and cells permeabilized in 400 µl of 0.25% Triton X100 (in PBS) for 10 minutes. Blocking was done with 300 µl of 3% w/v BSA in PBS for 1 hour on a rocker. After removing the BSA, 400 µl of primary antibody (rabbit raised anti-HA, 1:500 diluted in BSA) was added to react for 2.5 hours on the rocker. 400 µl of secondary antibody (rabbit raised Alexa 568, 1:400 diluted in BSA) was added and kept on the rocker for 1.5 hours at RT. Place back into six-well dish (cell side up) and wash three times with 3 ml PBS (5 minutes each). 300 µl of DAPI stain (1:400 diluted) was added to the chambers after washing and kept for 20 minutes on the rocker. Mounting solution was added to the slides (dismantling the chamber) and a cover slip placed on top of it. The slides was kept in the dark at RT to be imaged after 24 hours.

4. Bioinformatic analysis of dually localized proteins using Alanine Scanning Mutagenesis

The first 50 amino acids in the N-terminus of these proteins were mutated with Alanine (A), one at a time; a strategy that can be referred to as Alanine Scanning Mutagenesis.

These mutations were generated using a C++ code (Appendix) and compiled using standard C++ ide. The code replaces each residue to A, one at a time, starting from the second position

of the full length protein till the 50th. The resultant singly mutated unique sequences have A in place of an amino acid till the first 50 residues with the rest of the protein sequence same as the wild type.

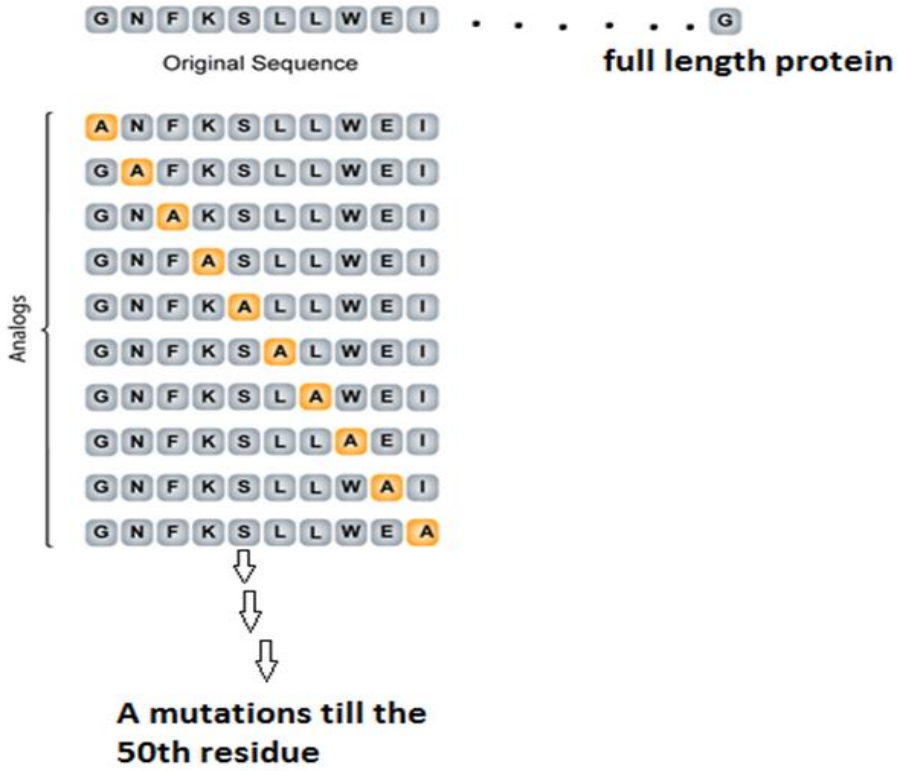


Fig. 11 Alanine Scanning Mutagenesis

SignalP 3.0 NN (Neural Network) was used to predict the presence of signal peptide, and location of the cleavage site. The 50 mutated sequences for each protein were entered in the input, with a default truncation of 70 residues since the length and cleavage site of the signal peptide of the proteins is not known. Mitochondrial targeting sequence prediction for the mutated sequence that showed a significant change in SignalP 3.0 NN score from the wild type protein, i.e. a variation of at least 0.1, was done using MitoProt II - v1.101. The SignalP 3.0 NN and MitoProt scores for each protein (all 50 mutated sequences) were plotted separately.

IV. Results

In order to test our hypothesis, we have chosen the membrane protein apicoplast phosphate transporter (APT1) to be studied for Golgi dependency in trafficking to the apicoplast. While there are multiple proteins that are equally relevant for this study, we have started with APT1 attributing to its interesting features. APT1 lacks a canonical bipartite sequence and has been predicted to have 6 transmembrane domains by TMHMM.

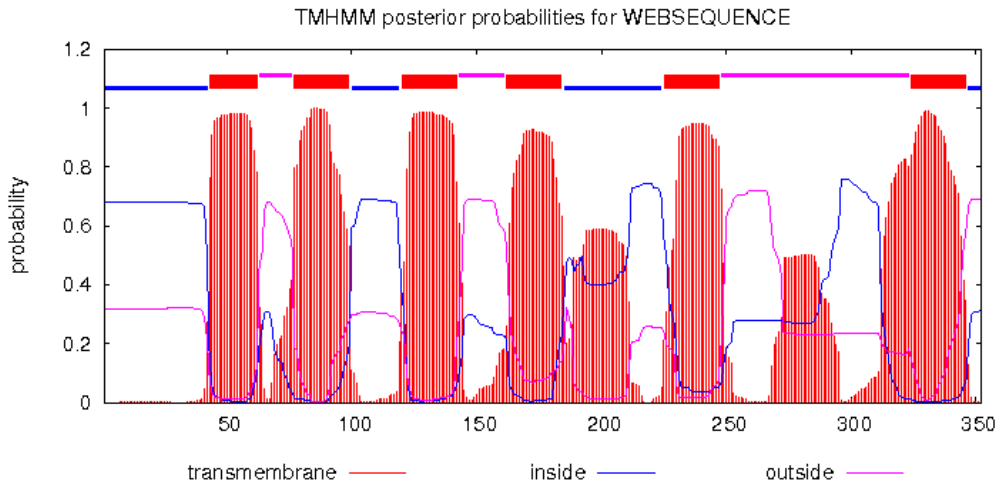


Fig. 12 Prediction of transmembrane domains of APT1 using TMHMM

Interestingly, APT1 has been shown to localize to multiple membranes of the apicoplast in *T. gondii* unlike its homologs in *P. falciparum*, namely PfoTPT and PfiTPT which are localized to the outermost and innermost membranes of the apicoplast respectively.

To check if the protein takes a Golgi dependent pathway in reaching the apicoplast in *T. gondii*, manipulation in the parasite was done using a plasmid vector.

The vector chosen for cloning is a plasmid vector pCTG-HA that was available in the lab. This plasmid acts as a shuttle vector with selection marker for both *E. coli* DH5α (ampicillin-resistance) as well as *T. gondii* (chloramphenicol resistance). The 9 amino acid long HA tag acts as an epitope for visualization of the in-frame expressed proteins by immunofluorescence. The HA tag follows the Multiple Cloning site (MCS) where the genes of interest are cloned using unique restriction cutters. The expression of genes in *T. gondii* when the clone is transfected is under the tubulin promoter.

APT1 was cloned in the multiple cloning site of the vector as described below. APT1 lacks introns and hence was obtained from genomic DNA

1. Cloning of the gene APT1 into the vector pCTG-HA

1.1. Isolation of the insert APT1

The genome of *T. gondii* RH was isolated using a pellet stored in -80°C to obtain the insert APT1, which doesn't contain any introns.

1.1.1. Isolation of genomic DNA

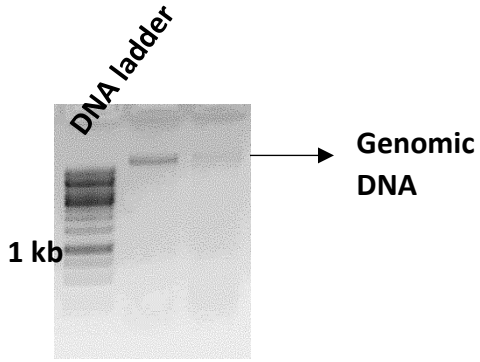


Fig 13. 0.8% agarose gel showing isolated genomic DNA from *T. gondii*

The DNA isolated using PCI method was run on an agarose gel, as shown in the figure above. The isolated DNA showed the following data concentration of DNA= 16.5 ng/ µl upon Nano drop quantification. This DNA was used further for PCR and a product was obtained as shown below in Figure 13.

1.1.2. PCR product of APT1 with Phusion Hi-fi Polymerase

The PCR product obtained from the DNA isolated showed an appropriate band of 1 kb when 5 µl of the sample was loaded on the agarose gel. This product is the gene APT1 to be inserted into the vector. The PCR product was digested using the enzymes Mfe1 and Nde1to generate compatible ends.

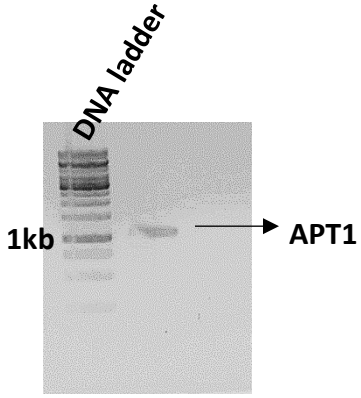
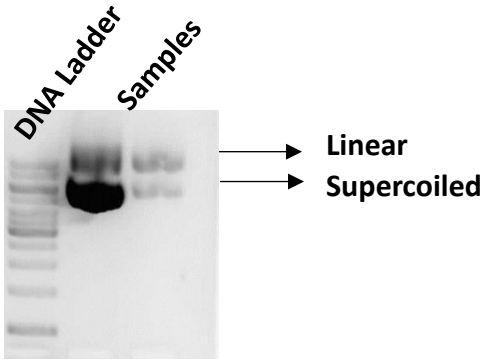


Fig 14. 0.8% agarose gel showing PCR amplified APT1 from genomic DNA template

1.2. Isolation of the plasmid vector pCTG-HA

The plasmid DNA isolated from *E. coli* DH5α cells by alkaline lysis method using 3 ml culture into two eppendorf tubes was run on a gel, with 2 µl of the sample loaded, as shown in the Figure 11. The concentration of the plasmid DNA found from the ImageJ software was 161.18 ng/µl. Two distinct bands observed on the gel stacked onto each other was very characteristic of plasmids and shows different conformations.



Concentration of DNA= 161.18 ng/ µl

Fig 15. 0.8% agarose gel showing isolated pCTG-HA plasmid from *E.coli* DH5α

1.3. Restriction digestion of plasmid vector

The isolated plasmid was digested using the enzymes Mfe1 and Nde1. When cut singly the vector showed one band unlike the two bands of an uncut intact supercoiled plasmid DNA. Since, the double digestion wasn't visible properly on the gel with 2 µl sample loaded, 5 µl of the samples were loaded again as shown in the figure 14(b) which showed a clean band for double digestion of the required plasmid size ~8 kb.

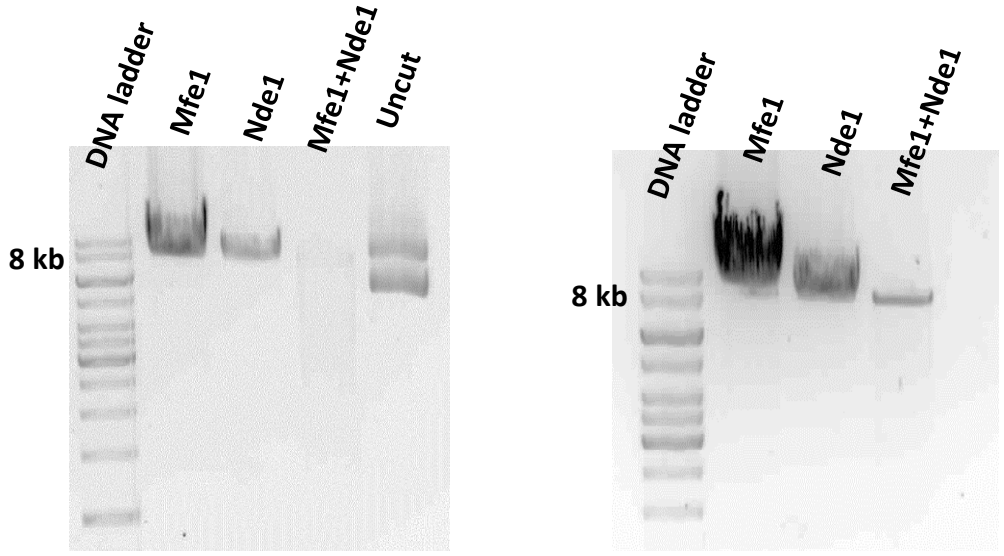


Fig 16. (a) 0.8% agarose gel showing restriction digestion of the plasmid vector with the appropriate enzymes along with uncut plasmid **(b)** 5µl loaded to visualise double digestion clearly.

1.4. Screening for recombinant plasmids

The restricted samples were ligated at 22°C for 5 hours with a 1:3 vector: insert ratio using T4 DNA ligase that after transformation gave 112 colonies on a LA plate with ampicillin.

1.4.1. Colony PCR of 8 randomly picked isolated colonies from the LA plate using Taq Polymerase

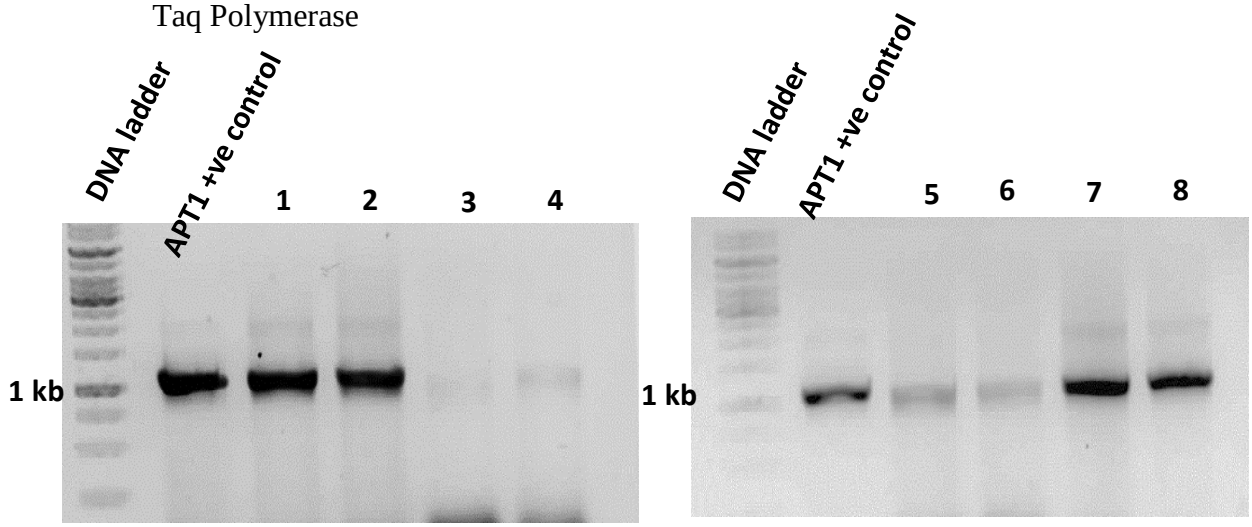


Fig 17. 0.8% agarose gel showing PCR with the template using restreaked colonies of transformants.

The colony PCR showed the desired band of 1 kb on the agarose gel indicating the presence of the inserted gene in the vector. The colonies that showed a band on gel were cultured in LB for the plasmid to be isolated.

1.4.2. Restriction digestion of the recombinant plasmid to check for the insert

Even though the desired size of the band was obtained, to check that the insert is the APT1 gene inserted at the right site in the MCS, restriction digestion was performed.

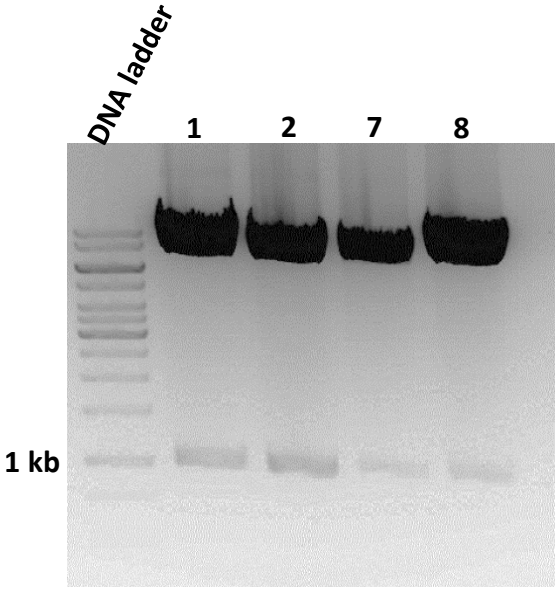


Fig 18. 0.8% agarose gel showing restriction digestion of the isolated recombinant plasmids.

The release of 1 kb band corresponding to APT1 upon gene specific restriction digestion confirms the colonies to be recombinants. Out of these, 3 colonies were picked from the restreaked plate and plasmid isolated using column purification to be sent for sequencing to check for mutations, if any.

pCTG-APT1-HA: The recombinant plasmid with APT1 gene

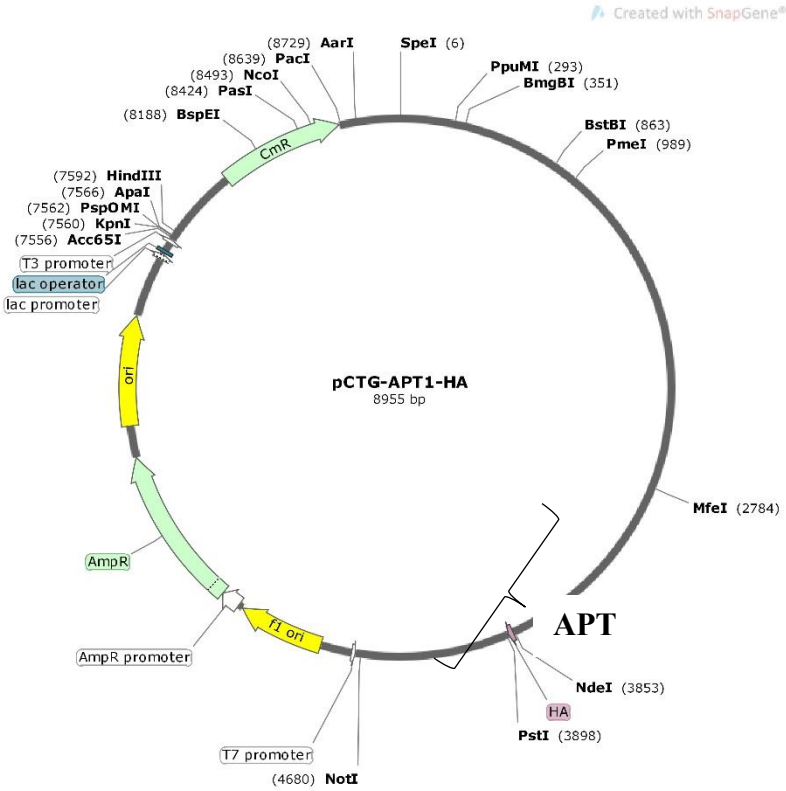


Fig 19. pCTG-APT1-HA

5. Cloning the HDEL construct of APT1 into the vector pCTG-HA

5.1. Designing the reverse primer containing HDEL sequence

The reverse primer was designed such that it can be used to add an HDEL sequence to any gene sequence cloned into the MCS of pCTG-HA vector. The strategy is such that the reverse primer 3'-5' is largely complementary to the features of the plasmid pCTG-HA, namely the HA Tag (partially), the stop codon and the Pst1 restriction enzyme site. Between these complementary sequences, the sequence for the HDEL residues is interspersed. In the first cycle of PCR, the complementary sequences help in anchoring of the template and the HDEL sequence loops out. But in the further cycles, the HDEL sequence is also amplified as a template and is thus incorporated into the gene.

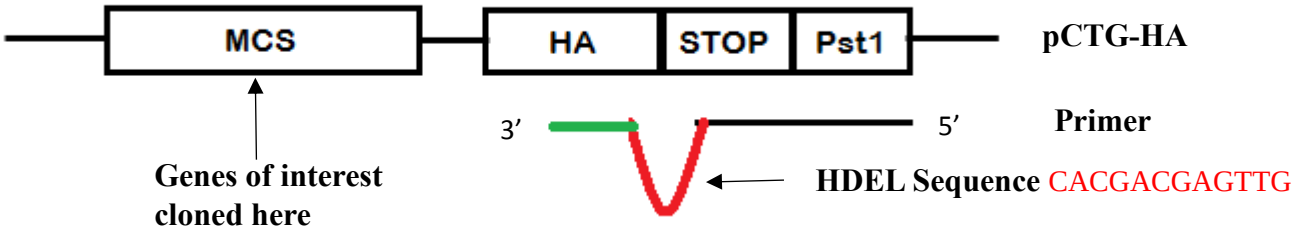


Fig.20 Primer designing strategy for addition of HDEL sequence

5' CGACGTCCCTGACTACGCGCACGACGAGTTGTAGCTGCAGTAAT 3' **pCTG-HA template**

Reverse complementary of the above sequence is the reverse primer:

5' ATTACTGCAGCTACAACTCGTCGTGCGC GTA GTC AGG GAC GTCG 3'

CGACGTCCCTGACTACGCG – Part of the HA Tag

CAC GAC GAG TTG – HDEL Sequence

TAG – Stop Codon

5.2. Obtaining the insert APT1-HDEL

The pCTG-APT1-HA vector (Fig. 18) was used as the template for the PCR to obtain APT1-HDEL insert, with the primers as mentioned in Page 24. The reverse primer contains the HDEL sequence to be added to the C-terminal of the gene by PCR as described in the section above. This PCR was done using Q5 Hi-fi polymerase and the product loaded on a 1% agarose gel.

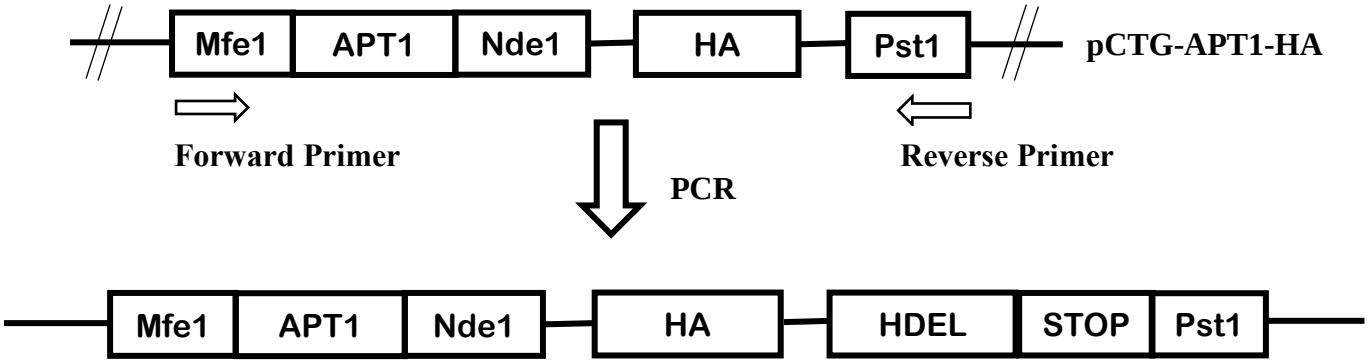


Fig. 21 HDEL construct of the gene APT1 obtained from the recombinant plasmid

The PCR product showed appropriate band size on the gel, and was precipitated with PCI method. Restriction digestion was done for the insert using the enzymes Mfe1 and Pst1. The digested product was loaded on a 1% gel as shown in Fig. 20 (b) lane 3.

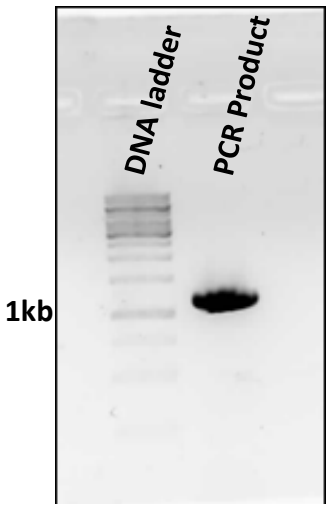


Fig 22. 1% agarose gel showing APT1-HDEL insert from recombinant plasmid

5.3. Obtaining the cut vector pCTG-HA for the insertion of APT1-HDEL

The plasmid pCTG-HA was used as a vector and was also cut with the enzymes Mfe1 and Pst1 to generate compatible ends with the insert. The digested products were loaded on a 0.8% agarose gel and showed single bands unlike the supercoiled conformations of the uncut plasmid. Both the vector and insert were digested with Pst1 (star activity post 6 hours) first for 5 hours, followed by an overnight digestion reaction with the enzyme Mfe1.

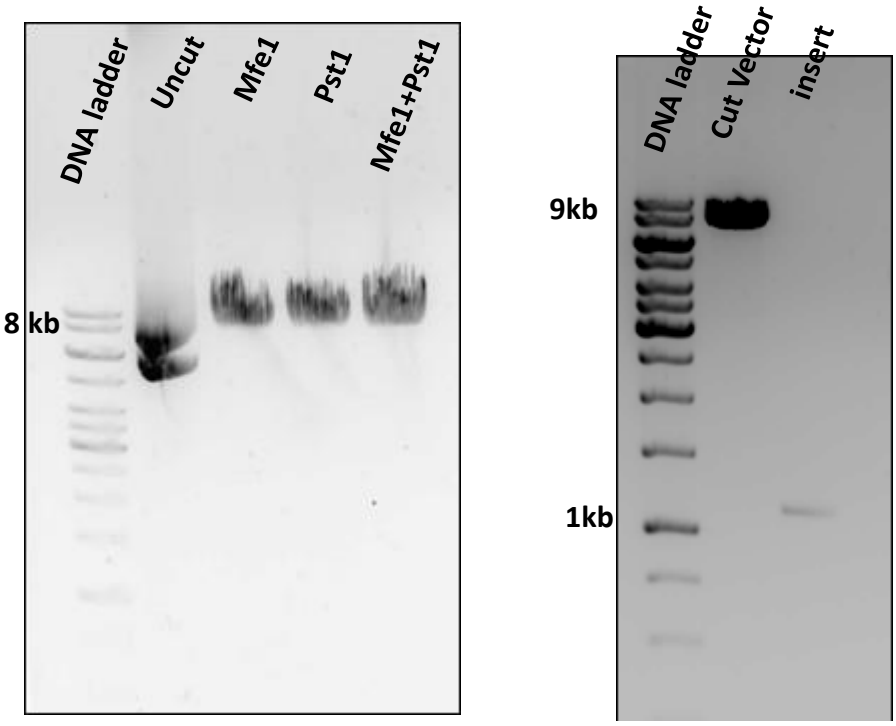


Fig. 23 (a) 0.8% agarose gel showing restriction digestion of the plasmid vector with the appropriate enzymes along with uncut plasmid **(b)** The cut vector and insert loaded after precipitation with PCI.

5.4. Screening for the recombinant plasmids

The digested products were ligated using T4 DNA ligase with a vector: insert ratio of 1:5 at 22°C overnight. The ligation reaction, when plated on LA plates containing ampicillin, gave 110 colonies. A colony PCR was done with randomly picked colonies to screen for the presence of the insert APT1-HDEL in the transformants. Out of the 11 colonies restreaked, 6 colonies showed a positive result with a band of desired size in the colony PCR as shown in Fig. 22.

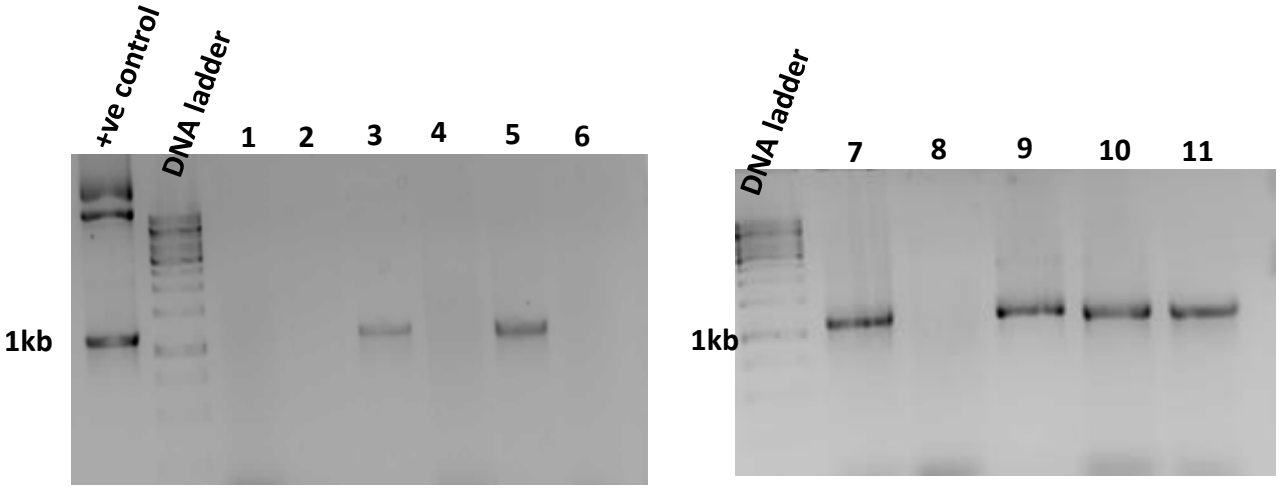


Fig 24. 0.8% agarose gel showing PCR with the template using restreaked colonies of transformants.

These plasmids were isolated and digested with the specific enzymes Mfe1 and Pst1 to check for the release of the insert of interest. The digested product was loaded on a 1% gel as shown in Fig. 23.

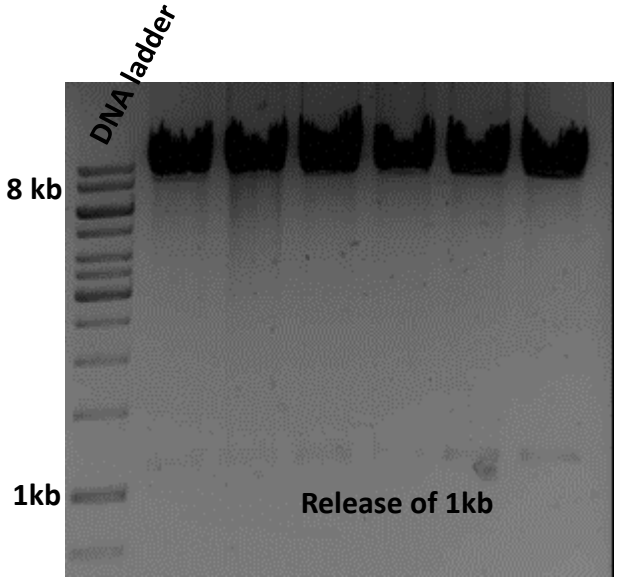


Fig 25. 0.8% agarose gel showing restriction digestion of the isolated recombinant plasmids APT1-HDEL

pCTG-APT1-HA-HDEL: The recombinant plasmid with HDEL construct of APT1 gene

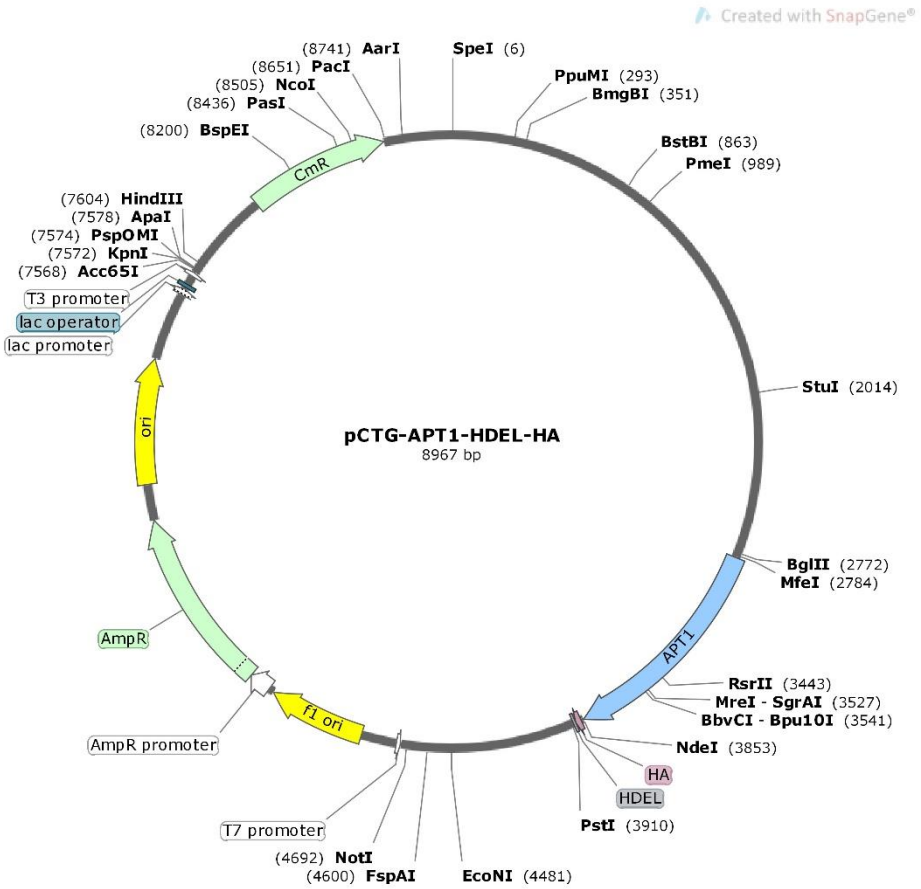


Fig 26. pCTG-APT1-HA-HDEL

6. Transfection of the recombinant plasmid DNA into the *T. gondii* parasites

The recombinant plasmids, both pCTG-APT1-HA and pCTG-APT1-HA-HDEL, were isolated using the QIAGEN maxi prep kit and quantified using UV-Visible spectrophotometer.

The concentration of DNA quantified:

pCTG-APT1-HA = 1.53 µg/ µl

pCTG-APT1-HA-HDEL = 4.35 µg/ µl

The appropriate volume of the recombinant plasmids amounting to 50 µg DNA was taken for both. For the expression of the proteins in the parasites, the cells were kept undisturbed after transient transfection. A slide was prepared using anti-HA antibodies which is specific for the HA tag in the plasmids.

APT1-HA was localized to the apicoplast, as expected (Fig. 27). This served as a control for checking the Golgi dependency of the trafficking of the protein.

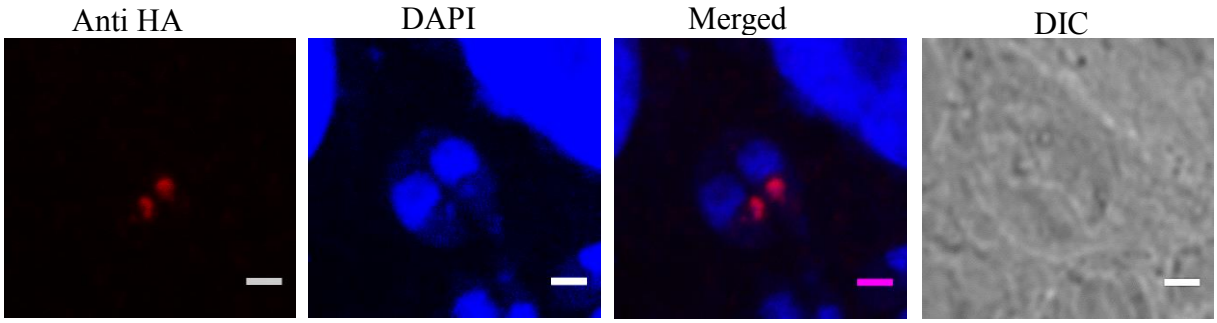


Fig 27. Immunofluorescence of parasites transfected with APT1-HA using anti-HA antibodies (scale- 2 µm). Localization of APT1-HA to the apicoplast of *T. gondii*.

The red spot observed in the first panel corresponds to the apicoplast of the cell, which is the location of the protein APT1. This is confirmed by its co-localization with DAPI that stains DNA. In the second panel of the figure, DAPI can be seen to stain the host cell as well as both the nucleus and a small extra-chromosomal DNA in the parasite, which is the apicoplast DNA. The third panel shows the merged signal from both the channels. The DIC image shows the two intracellular parasites.

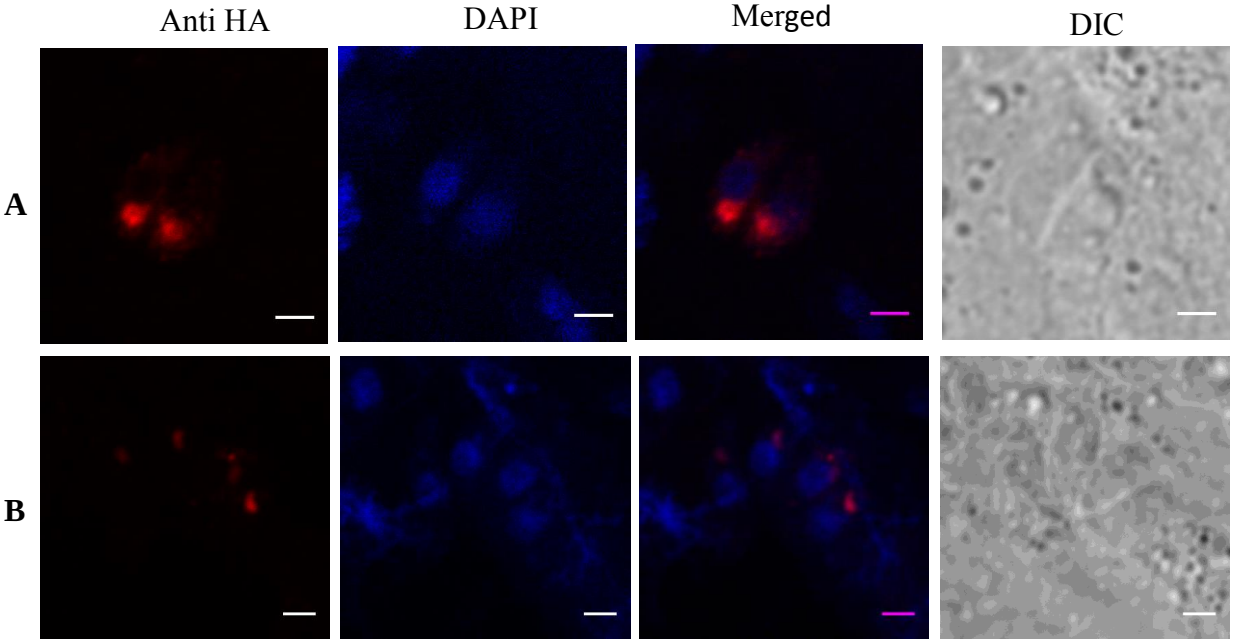


Fig 28. Immunofluorescence of parasites transfected with APT1-HA-HDEL using anti-HA antibodies (scale- 2 µm). Localization of APT1-HA-HDEL in the *T. gondii* cells.

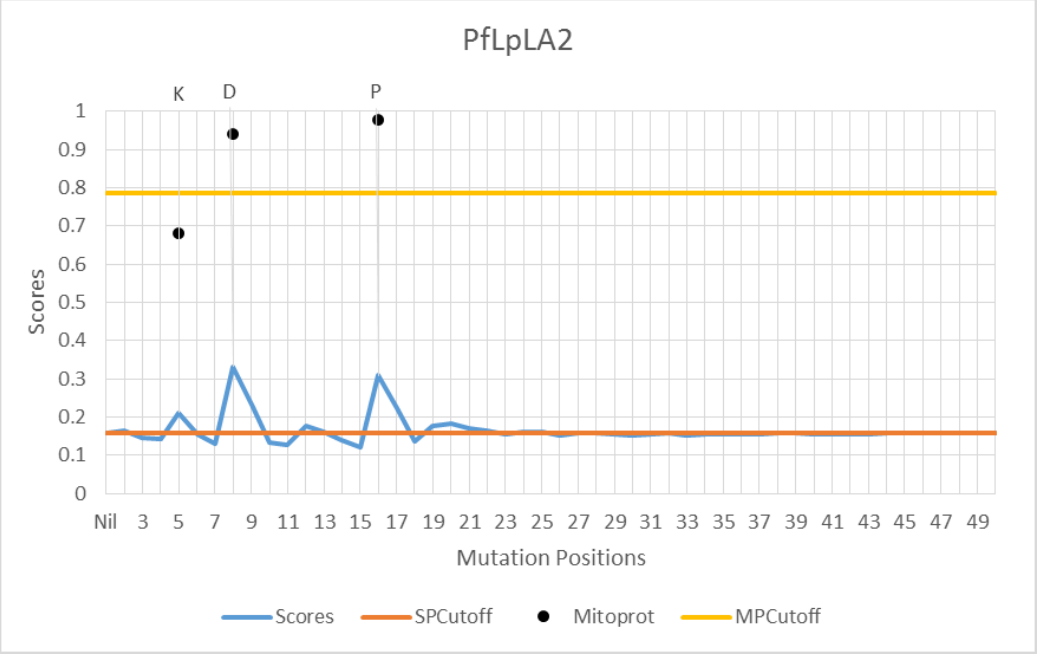
Interestingly, the APT1-HA protein with a C-terminal HDEL sequence still localized to the apicoplast of the cells which can be seen with DAPI co-localization. The HDEL sequence in the C-terminal of the proteins marks them for retrieval to the ER of the cell. However, the addition of this retrieval sequence seems to have no effect on the trafficking of the protein

APT1. In addition to the apicoplast signal, some parasites can be observed to stain in the regions peripheral to the apicoplast as shown in the panel A of Figure 27. In contrary to our hypothesis, an ER characteristic fluorescence was not observed in the APT1 protein containing an HDEL sequence which indicates a trafficking pathway independent of the Golgi.

7. Bioinformatics analysis of dually targeted apicomplexan proteins by alanine scanning mutagenesis of N-terminal sequence

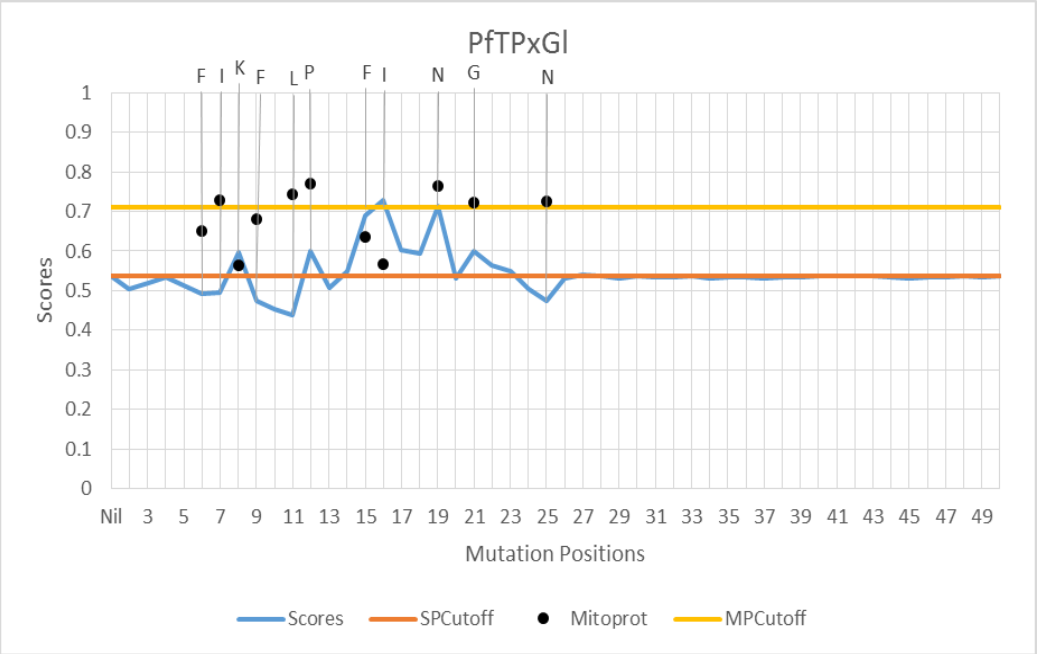
The five known dually localized proteins, i.e. targeted to both mitochondria and apicoplast, in the apicomplexan parasites *P. falciparum* and *T. gondii*, namely PfLPLa2, PfTPxGl, TgSOD2, TgIRP/ACN, and TgTPx1.2 were chosen for the study of the N terminal targeting sequence. For each of the five proteins, 50 mutated sequences were generated by Alanine Scanning Mutagenesis. Alanine was chosen as the substitution residue because it eliminates the side chain beyond the β carbon but does not alter the main-chain conformation (as can glycine or proline). In addition to that, Alanine does not impose extreme electrostatic or steric effects to the mutated protein. The prediction score of signal peptide was found using SignalP 3.0NN. The positions at which mutation to Alanine increased or decreased the score from the wild type, the full length sequences of those were predicted for a mitochondrial targeting sequence as well and the scores plotted on the same graph.

The graphs for the analysis of each protein is shown below:

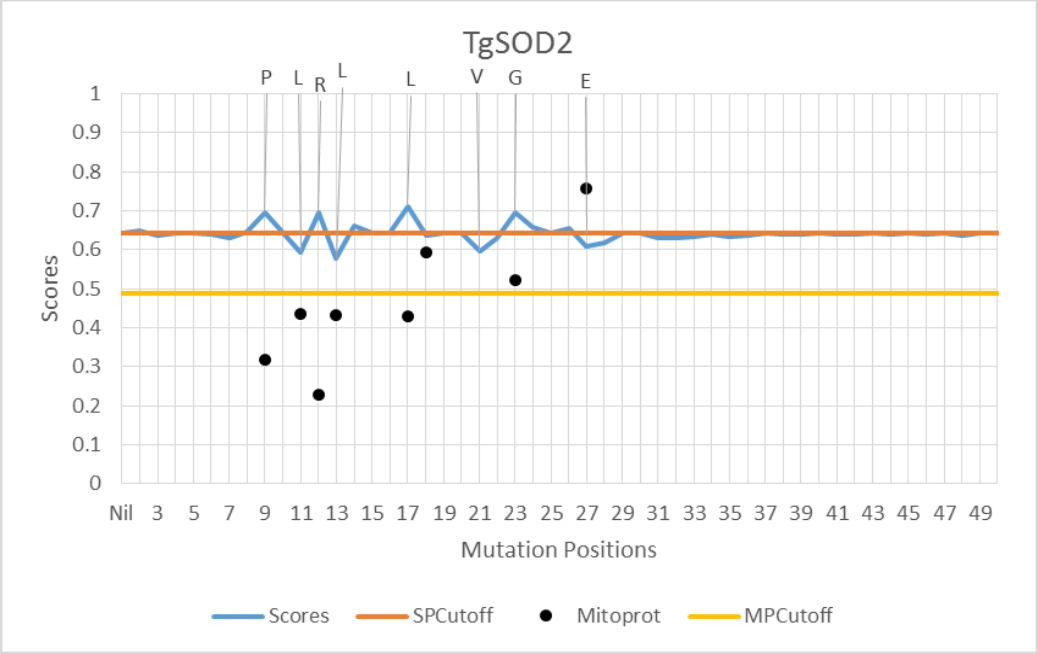


7.1 PfLplA2: Lipoate protein ligase 2 of *P. falciparum* (PfLplA2), randomly targeted to mitochondria and apicoplast of the cell, has a low score of 0.157 and a much higher

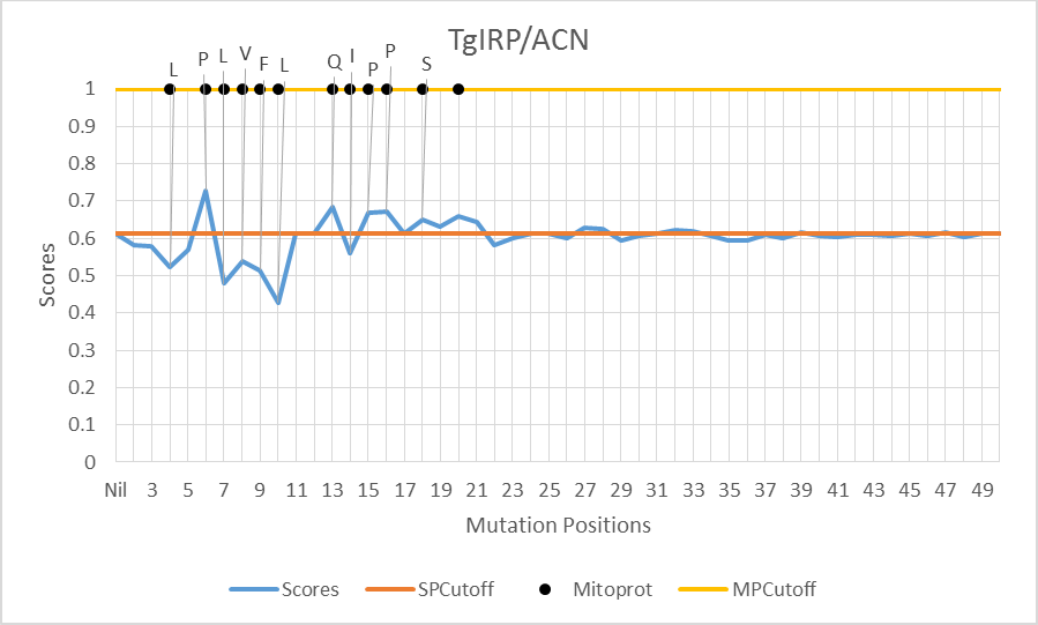
MitoProt score of 0.786 for the wild type protein (cut off). The Alanine scanning shows an increase in the SignalP 3.0 NN score when the 5th Lysine (K), 8th Aspartate (D) and the 16th Proline (P) are changed to Alanine (A). The MitoProt score for these mutated full length proteins, with a cut-off of 0.786 (wildtype protein), was also plotted. The K-A mutation showed a decrease in MitoProt score whereas the D-A and P-A mutations showed an increase in the mitochondrial targeting probability. There is no change in the SignalP scores post the 22nd residue.



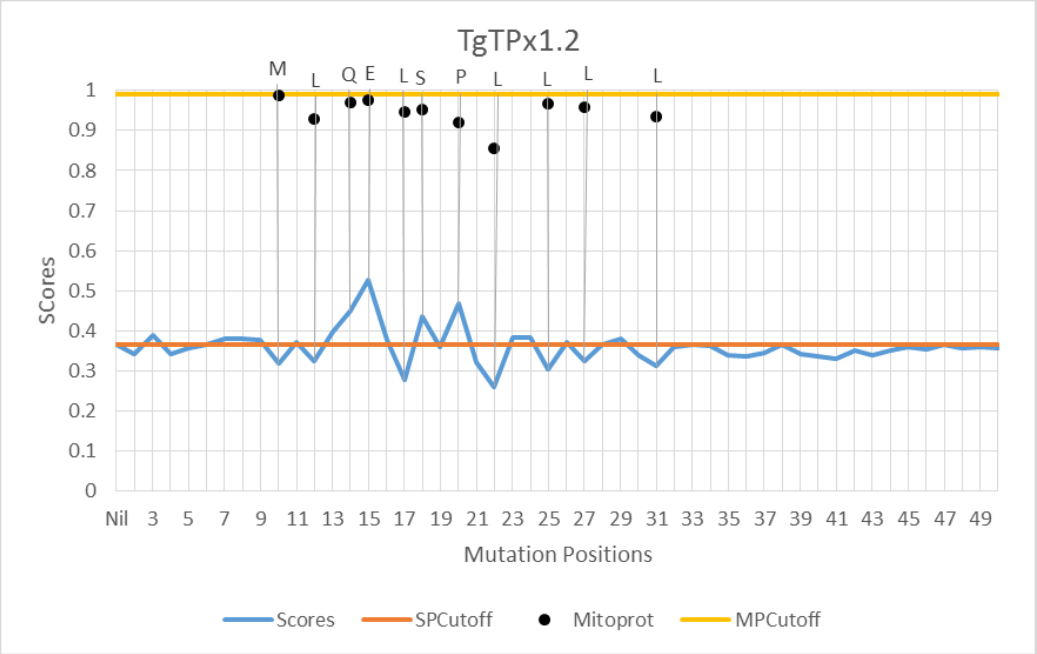
7.2 PfTPxGl: *P. falciparum* glutathione peroxidase-like thioredoxin peroxidase (PfTPxGl) has a random distribution in mitochondria and apicoplast of the cell. The cut off for SignalP 3.0 NN is 0.536 and show a much more dynamic trend in the graph when the mutated sequence scores are plotted. The Alanine scanning showed no effect on the SignalP score post the 26th residue. The MitoProt score for the mutated sequences with an increase/decrease in SignalP scores was also plotted, with a cut-off of 0.711 (wildtype protein), and again showed a diverging trend in the graph. The major residues that contributed to the shift in the scores are uncharged (F, L, I, P, N, G) whereas only one charged residue was present at the 8th position- a Lysine (K).



7.3 TgSOD2: *T. gondii* Superoxide Dismutase 2 (TgSO2) is a dually targeted protein which unlike PfTPxGl is not distributed randomly to mitochondria and apicoplast. The wild type protein showed a higher SignalP score than the MitoProt score (0.486). The mutated sequences showed little variation as compared to the wild type protein in the SignalP score. Whereas the MitoProt showed a much greater differences in the wild type score and the scores of the mutated sequence. There are only two charged residues out of the mutations checked, a positively charged R and a negatively charged E, at positions 11 and 27 respectively. The change in residues to Alanine show no effect on the score from the 29th residue onwards.



7.4 TgIRP/ACN: *T. gondii* Iron regulatory protein/Aconitase (IRP/ACN) is a dually targeted protein which shows a very strong probability for mitochondrial targeting with scores 0.998 for all the mutated sequence. The SignalP scores for the mutated sequences vary till the 22nd residue as compared to the wild type sequence (0.612). Interestingly, no charged residues are involved in contributing to a change in SignalP prediction score from the cut-off value. The protein is not distributed randomly to the endosymbiotic organelles, and show a much higher probability for the mitochondria (0.998) as compared to apicoplast (0.612).



7.5 TgTPx1.2: *T. gondii* thioredoxin-like peroxidase 1.2, a dually targeted protein with no random distribution, showed a much lower score for SignalP predition (0.364) as compared to the MitoProt (0.989) for the wild type protein. The alanine mutation shows a change in the scores from the cut-off value for a larger stretch (till 41 residues) unlike the other proteins studied above. The mutated sequences with variations in the SignalP score when checked for mitochondrial targeting sequence probability in MitoProt showed little variation from the cut-off score, barring a L at 22nd position for which the score went down from 0.989 to 0.852. The sequence with E at 15th position showed a leap in the SignalP score from the cut-off value.

V. Discussion

1. Trafficking of APT1 is via a Golgi independent route

The APT1-HA protein expressed under the strong tubulin promoter showed varied localisation of the protein when visualised with anti-HA antibodies by immunofluorescence. A vacuole specific pattern was observed in imaging of the cells with varied distribution. The parasites in a vacuole in the host cell have synchronised cell cycle. In some of the vacuoles, a punctuate localisation very characteristic of early stage apicoplast was seen. On the other hand, some other vacuoles also showed circumplastid localization. The reticular distribution seen in some of the cells has been suggested to be reminiscent of ER by Karnataki et al. in 2007, in their study of TgAPT1 expression regulated by the cell cycle of the parasite.

The ER retrieval sequence added to the C-terminal of the protein APT1 was expected to localize it to the ER of the parasite in case of the protein being trafficked via the Golgi to the apicoplast. This strategy has shown another membrane protein TgTPx1.2 to localise to the ER establishing the involvement of Golgi in the trafficking of the protein. In our attempt to gain insights from evolutionary implications and relevance of the primitive and conserved ER-Golgi pathway of trafficking, we expected a similar trend for the membrane protein APT1. However, the HDEL construct of APT1 showed no difference in staining and hence localization as compared to the control protein which lacked the HDEL sequence. This is suggestive of the trafficking of APT1 to be independent of Golgi. In this regard, APT1 follows a path taken by other luminal proteins like SOD2 which have also been shown to target to apicoplast with no involvement of Golgi.

In their study of APT1 targeting to multiple membranes of the apicoplast, Karnataki et al. (2007) showed vesicles decorated with APT1 spread across post-Golgi space with the help of immunoelectron microscopy. In addition to that, they also observed protrusions of apicoplast in different stages of the cell containing with the same electron density to the APT1HA bearing vesicles. Moreover, they also show the nuclear envelope with APT1 signal with gold particles.

Our observations of Golgi independent pathway of trafficking and studies from Karnataki et al suggest a probable trafficking route for APT1 via vesicles emerging from the ER and eventually fusing to the outer membrane of apicoplast. The internal signal sequence plays a cryptic role in the initial import of the protein into the ER and moves to the nuclear envelope, accumulating in vesicles near the apicoplast.

2. Bioinformatic analysis of dually targeted proteins in apicomplexans

The N-terminal targeting sequence of dually localized proteins studied by Alanine scanning mutagenesis, although performed over a very small data set, has led to some insights into the distribution of these proteins in the two endosymbiotic organelles of the cell.

The Signal peptide functions in targeting of the protein upon recognition by a signal recognition particle, a protein, and this interaction is crucial for the transport of a protein and its processing which involves the cleavage of signal peptide. All the proteins, whether sampled randomly or not, showed a trend of the score varying till a certain length in the N-terminal further which changing the residues do not seem to affect the probability of a sequence being a signal peptide. Evolutionarily, this has a probable relevance in keeping the signal peptide separable from the rest of the protein and could have been instrumental in deciding where the cleavage site of a protein should be placed. This seems to follow the neutral hypothesis of mutations in terms of evolution.

The fluctuation in the scores of signal peptide prediction brings us to speculate that the N-terminus carries some intrinsic properties allowing the random distribution of the proteins to different organelles and are attributed to some of the key residues in the sequence. The superimposition of MitoProt score gives additional information about this property of distribution of the proteins to two organelles of the cell.

In case of the randomly targeted protein PfLpLA2, the K5A mutation shows a decrease in the mitochondrial targeting probability but simultaneously increase in the SignalP score suggesting a higher probability in the targeting of the protein to apicoplast. Thus, single mutations can be sufficient in distributing the protein to one of the organelles in the cell. Similar observations were made for the other randomly distributed protein PfTPxGl with mutations including K8A, L11A, F15A, I16A, and N25A. These mutations are possibly responsible for the random distribution of the protein in mitochondria and apicoplast where mutations lead to the targeting to any one of the organelles. Looking at these trends, some predictions can be made for the proteins that are dually targeted but not random in distribution. For example, in the N-terminus of the protein TgSOD2, a very dynamic graph for the prediction scores of MitoProt and SignalP can be observed where the mutations P9A, R12A, L17A, and E26A increase the probability of targeting to one of the organelles but not the other. These mutations can be exploited and determined experimentally for localization. For the protein TgTPx1.2, experimental data has

been obtained in our lab where the mutation E15A targets the otherwise dually localized protein to have a random distribution in the parasite vacuoles of the host cell. The mutation R24A targeted TgTPx1.2 to only the apicoplast whereas L17A, and L27A targets the protein to only the mitochondrion.

However, when the graph of TgIRP/ACN is observed the protein has a very strong preference to the mitochondria where no mutations affect the mitochondrial targeting probability even though the SignalP varies with some of the mutations. This could be suggestive of TgIRP/ACN being present in the mitochondrion at all times, even if the mutations cause the protein to leave apicoplast due to lesser signal strength.

The involvement of the charged residues K, R and E in affecting the predction scores for the targeting of proteins to one of the proteins can be attributed to their charge which can possibly be involved in interactions with the receptors for targeting. Similarly the residues L, F and I could be interfering with the interactions due to their hydrophobicity.

These mutations are handy in directing a protein to an alternate compartment with just a minor change and must have been evolutionarily instrumental in circumventing the need for another trafficking pathway altogether, thereby reducing the energy burden of the cell. This can involve additional aspects like the affinity of binding, interactions and available concentration in the cell; but the mutational sampling is one of the important things that one can think about.

VI. Future directions

The study cannot possibly even scratch the surface of the greater question that we tried to address. The data obtained in this study can be further refined with experiments using specific markers for the ER and apicoplast. A split GFP assay will give insight into the orientation of the protein in a membrane. There are always more questions to ask including what motifs are responsible for the targeting TgAPT1 to the apicoplast? Are there any motifs present in the N terminus of the protein?

This data could be indicative of APT1 being an exception the generalization of membrane protein trafficking due to its undefined internal targeting sequence which is not processed even in the matured form of the protein. Taken together, this study in another step in recognizing the complexities inherent in the problem. There could also be a possibility of multiple mechanisms in targeting proteins to the apicoplast membranes.

To make concrete conclusions, more membrane proteins have to be studied and trends observed carefully without jumping into generalizations. A hypothesis for the evolutionary implication of the trafficking pathway, and the involvement of signal sequences in the context of compartmentalisation in endosymbiosis can be formulated by asking these questions to a statistically significant dataset.

VII. Appendix

C++ code for Alanine Scanning Mutagenesis

```
#include <iostream>
#include <fstream>
#include <string>
using namespace std;

int main() {
    int start,end;
    cout<<"\n Enter start index:";
    cin>>start;
    cout<<"\n Enter end index:";
       cin>>end;
    ifstream readfile;
       readfile.open("file.txt");
       string s;
       readfile>>s;
       string str=s.substr(start,end-start+1);
       readfile.close();

       char ch;
       cout<<"\n Enter the letter by which you want to replace:";
       cin>>ch;

    for(int i=start-1;i<=end-start+1;i++)
       {
           ofstream newfile;

newfile.open(to_string(i+1)+"_letter_replaced_by_"+ch+".txt");

           string temp=s;
           temp[i]=ch;
           newfile<<temp;

           newfile.close();
       }
       cout<<"\n The files are in the same directory where this program
is!";
       cout<<"\n Check them!!";

       return 0;
}
```

APT1 Protein Sequence (352 aa)

MEESKRLGVSALPPQYGTVSTGGARPAKDLESQASPASGDQTAFYAQLGVMLLFWYALNVMYNLDNKL
ALIMLPLPWTVSTFQLFFGWLFFGFAWATGLRPVPRIHTTELFVTRIAPQGLCHFFVHIGAVISMGCG
AVSFTHIVKASEPVLTALLSGLALHQVFSWQTYLSLVPIVAGVIMASVTELSFTWKAFGCALVSALGS
SARAVFAKLAMADRKQVGENLSSANMYALLTIVASLVSLPPAIFAEGAKVAAVWEACTGPDSPWTGQQ
IIAKLCFSGLWYYMYNEVAYLCLEKINQVTHAVANTLKRVVIIVASVLFFQTPVTALGATGSFVAIAG
TLIYSLSKTKYG

APT1 Gene sequence (NO INTRONS) 1059 bp

5’

ATGGAGGAATCGAAACGCTTGGGTGTCTCGGCTCTCCCGCCGCAGTACGGGACTGTTTCGACAGGCGG
CGCTCGGCCGGCGAAGGATTTAGAGTCGCAGGCGAGTCCCGCGTCTGGAGATCAGACGGCGTTCTACG
CGCAACTCGGCGTGATGTTGCTCTTCTGGTACGCGCTGAACGTGATGTACAATCTGGACAACAAGCTG
GCCCTGATCATGCTTCCACTTCCCTGGACGGTGTCGACGTTTCAGCTCTTCTTCGGATGGCTGTTCTT
CGGCTTCGCGTGGGCAACTGGGCTGCGGCCGGTCCCGCGAATCCACACCACGGAGCTGTTCGTCACGC
GCATCGCCCCGCAGGGTCTCTGTCACTTCTTCGTCCACATCGGCGCGGTGATTTCAATGGGCTGTGGC
GCTGTCAGCTTCACACACATCGTGAAGGCGTCGGAGCCAGTCCTCACGGCGCTGCTCTCTGGACTGGC
TCTTCACCAGGTCTTCAGCTGGCAGACGTACCTCTCCCTCGTCCCGATCGTCGCAGGAGTGATCATGG
CCTCGGTCACGGAACTCTCCTTCACGTGGAAGGCGTTCGGATGCGCCTTGGTCAGTGCGCTGGGCAGC
TCGGCTCGCGCAGTGTTCGCGAAACTCGCGATGGCGGACCGCAAACAAGTAGGAGAAAATCTGTCGTC
TGCAAACATGTACGCGCTTCTAACCATCGTCGCGTCCCTCGTCTCCCTCCCGCCGGCGATCTTCGCTG
AGGGCGCCAAAGTCGCCGCTGTCTGGGAGGCATGCACCGGACCCGACAGCCCCTGGACGGGCCAGCAG
ATCATCGCGAAACTGTGCTTCTCAGGCTTGTGGTATTACATGTACAACGAAGTCGCCTACCTCTGCCT
CGAGAAAATCAACCAGGTCACGCATGCAGTAGCGAACACCCTCAAACGCGTCGTCATCATCGTCGCGT
CTGTCCTCTTCTTCCAGACTCCCGTGACGGCCCTTGGAGCAACTGGGTCGTTCGTCGCCATTGCAGGC
ACCCTGATCTATTCTCTCTCGAAGACCAAGTACGGATAA

3’

VIII. References

- Archibald, J. M., & Keeling, P. J. (2002). Recycled plastids: A "green movement" in eukaryotic evolution. Trends in Genetics, 18(11), 577–584. https://doi.org/10.1016/S0168-9525(02)02777-4
- Blanchard, J L, and J S Hicks. 1999. "The Non-Photosynthetic Plastid in Malarial Parasites and Other Apicomplexans Is Derived from Outside the Green Plastid Lineage." *J Eukaryot Microbiol* 46(4): 367–75.
- Bouchut, A., Geiger, J. A., Derocher, A. E., & Parsons, M. (2014). Vesicles bearing toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption. PLoS ONE, 9(11), 1–14. https://doi.org/10.1371/journal.pone.0112096
- Bruce, B. D. (2001). The paradox of plastid transit peptides: Conservation of function despite divergence in primary structure. Biochimica et Biophysica Acta - Molecular Cell Research, 1541(1–2), 2–21. https://doi.org/10.1016/S0167-4889(01)00149-5
- Chaudhari, R., Dey, V., Narayan, A., Sharma, S., & Patankar, S. (2017). Membrane and luminal proteins reach the apicoplast by different trafficking pathways in the malaria parasite *Plasmodium falciparum*. PeerJ, 5, e3128. https://doi.org/10.7717/peerj.3128
- Cheresh, P., Harrison, T., Fujioka, H., & Haldar, K. (2002). Targeting the Malarial Plastid via the Parasitophorous Vacuole* □ S. https://doi.org/10.1074/jbc.M109331200
- DeRocher, A. E., Coppens, I., Karnataki, A., Gilbert, L. A., Rome, M. E., Feagin, J. E., … Parsons, M. (2008). A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in *Toxoplasma gondii*. Eukaryotic Cell, 7(9), 1518–1529. https://doi.org/10.1128/EC.00081-08
- DeRocher, A., Gilbert, B., Feagin, J. E., & Parsons, M. (2005). Dissection of brefeldin A- sensitive and insensitive steps in apicoplast protein targeting. Journal of Cell Science, 118(3), 565–574. https://doi.org/10.1242/jcs.01627
- DeRocher, Hagen, C. B., Froehlich, J. E., Feagin, J. E., & Parsons, M. (2000). Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system. Journal of Cell Science, 113 (Pt 2, 3969–3977. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11058084
- Desmonts, Georges, and Jacques Couvreur. 1974. "Congenital Toxoplasmosis - A Prospective Study of 378 Pregnancies." *New England Journal of Medicine* 290(20):

1110–16. http://www.nejm.org/doi/abs/10.1056/NEJM197405162902003.

- Diseases, Global Health Division Of Parasitic. 2010. "About Malaria - History - Laveran and the Discovery of the Malaria ParasGlobal Health - Division of Parasitic Diseasesite." *CDC Malaria*.
- Dubey, J. P., Lindsay, D. S., & Speer, C. A. (1998). Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clinical Microbiology Reviews, 11(2), 267–299. https://doi.org/PMC106833
- Foth, B. J. (2003). Dissecting Apicoplast Targeting in the Malaria Parasite *Plasmodium falciparum*. Science, 299(5607), 705–708. https://doi.org/10.1126/science.1078599
- Gallagher, J. R., Matthews, K. A., & Prigge, S. T. (2011). *Plasmodium falciparum* apicoplast transit peptides are unstructured in vitro and during apicoplast import. Traffic, 12(9), 1124–1138. https://doi.org/10.1111/j.1600-0854.2011.01232.x
- Gomord, V., Denmat, L. A., Fitchette-Lainé, A. C., Satiat-Jeunemaitre, B., Hawes, C., & Faye, L. (1997). The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. Plant Journal. https://doi.org/10.1046/j.1365-313X.1997.11020313.x
- Hager, K. M., Striepen, B., Tilney, L. G., & Roos, D. S. (1999). The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. Journal of Cell Science, 112 (Pt 1, 2631–2638.
- Jomaa, H. (1999). Inhibitors of the Nonmevalonate Pathway of Isoprenoid Biosynthesis as Antimalarial Drugs. Science, 285(5433), 1573–1576. https://doi.org/10.1126/science.285.5433.1573
- Kalanon, M., Tonkin, C. J., & McFadden, G. I. (2009). Characterization of two putative protein translocation components in the apicoplast of *Plasmodium falciparum*. Eukaryotic Cell, 8(8), 1146–1154. https://doi.org/10.1128/EC.00061-09
- Karnataki, A., DeRocher, A., Coppens, I., Nash, C., Feagin, J. E., & Parsons, M. (2007). Cell cycle-regulated vesicular trafficking of Toxoplasma APT1, a protein localized to multiple apicoplast membranes. Molecular Microbiology, 63(6), 1653–1668. https://doi.org/10.1111/j.1365-2958.2007.05619.x
- Karnataki, A., Derocher, A. E., Coppens, I., Feagin, J. E., & Parsons, M. (2007). A membrane protease is targeted to the relict plastid of Toxoplasma via an internal signal

sequence. Traffic, 8(11), 1543–1553. https://doi.org/10.1111/j.1600-0854.2007.00637.x

- Karnataki, A., Derocher, A. E., Feagin, J. E., & Parsons, M. (2009). Sequential processing of the Toxoplasma apicoplast membrane protein FtsH1 in topologically distinct domains during intracellular trafficking. Mol Biochem Parasitol, 166(2). https://doi.org/10.1016/j.molbiopara.2009.03.004
- Kohler, S. (1997). A Plastid of Probable Green Algal Origin in Apicomplexan Parasites. Science, 275(5305), 1485–1489. https://doi.org/10.1126/science.275.5305.1485
- Lemgruber, L., Kudryashev, M., Dekiwadia, C., Riglar, D. T., Baum, J., Stahlberg, H., Frischknecht, F. (2013). Cryo-electron tomography reveals four-membrane architecture of the Plasmodium apicoplast. Malaria Journal, 12(1), 25. https://doi.org/10.1186/1475-2875-12-25
- Lim, L., Kalanon, M., & McFadden, G. I. (2009). New proteins in the apicoplast membranes: time to rethink apicoplast protein targeting. Trends in Parasitology, 25(5), 197–200. https://doi.org/10.1016/j.pt.2009.02.001
- Lim, L., McFadden, G. I., Aikawa, M., Azuma, M., Kabe, Y., Kuramori, C., Schneider, G. (2010). The evolution, metabolism and functions of the apicoplast. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 365(1541), 749–63. https://doi.org/10.1098/rstb.2009.0273
- Mcfadden, G. I. (n.d.). The apicoplast. https://doi.org/10.1007/s00709-010-0250-5
- Morrison, D. A. (2009). Evolution of the Apicomplexa: where are we now? Trends in Parasitology, 25(8), 375–382. https://doi.org/10.1016/j.pt.2009.05.010
- Mullin, K. A., Lim, L., Ralph, S. A., Spurck, T. P., Handman, E., & McFadden, G. I. (2006). Membrane transporters in the relict plastid of malaria parasites. Proceedings of the National Academy of Sciences of the United States of America, 103(25), 9572–7. https://doi.org/10.1073/pnas.0602293103
- Narayan, A., Kumar, M., Singh, M., Babar, P., Chaudhari, R., & Patankar, S. (2012). Regulation of gene expression in *Plasmodium falciparum*. Current Science, 102(5), 712–724.
- Pagny, S., Lerouge, P., Faye, L., & Gomord, V. (1999). Signals and mechanisms for protein retention in the endoplasmic reticulum. J. Exp. Bot., 50(331), 157–164. https://doi.org/10.1093/jxb/50.331.157

- Parsons, M., Karnataki, A., Feagin, J. E., & DeRocher, A. (2007). Protein trafficking to the apicoplast: Deciphering the apicomplexan solution to secondary endosymbiosis. Eukaryotic Cell, 6(7), 1081–1088. https://doi.org/10.1128/EC.00102-07
- Ralph, S., D'Ombrain, M., & McFadden, G. (2001). The apicoplast as an antimalarial drug target. Drug Resistance Updates, 4, 145–151. https://doi.org/S1368-7646(01)90205-5 [pii] 10.1054/drup.2001.0205 [doi]
- Sheiner, L., Fellows, J. D., Ovciarikova, J., Brooks, C. F., Agrawal, S., Holmes, Z. C., Striepen, B. (2015). *Toxoplasma gondii* Toc75 Functions in Import of Stromal but not Peripheral Apicoplast Proteins. Traffic, 16(12), 1254–1269. https://doi.org/10.1111/tra.12333
- Sheiner, L., & Soldati-Favre, D. (2008). Protein trafficking inside *Toxoplasma gondii*. Traffic, 9(5), 636–646. https://doi.org/10.1111/j.1600-0854.2008.00713.x
- Sheiner, L., & Striepen, B. (2014). The apicoplast: A parasite's symbiont. In Plastid Biology (pp. 209–238). https://doi.org/10.1007/978-1-4939-1136-3_8
- Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. (1987). Analytical Biochemistry, 162(1), 156–159. https://doi.org/10.1016/0003-2697(87)90021-2
- Spork, S. et al. (2009). An unusual ERAD-like complex is targeted to the apicoplast of *Plasmodium falciparum*. Eukaryotic Cell, 8(8), 1134–1145. https://doi.org/10.1128/EC.00083-09
- Striepen, B., He, C. Y., Matrajt, M., Soldati, D., & Roos, D. S. (1998). Expression, selection, and organellar targeting of the green fluorescent protein in *Toxoplasma gondii*. Molecular and Biochemical Parasitology, 92(2). https://doi.org/10.1016/S0166-6851(98)00011-5
- Tawk, L., Dubremetz, J. F., Montcourrier, P., Chicanne, G., Merezegue, F., Richard, V., Lebrun, M. (2011). Phosphatidylinositol 3-monophosphate is involved in Toxoplasma apicoplast biogenesis. PLoS Pathogens, 7(2), 1–16. https://doi.org/10.1371/journal.ppat.1001286
- Tonkin, C. J., Foth, B. J., Ralph, S. a, Struck, N., Cowman, A. F., & McFadden, G. I. (2008). Evolution of malaria parasite plastid targeting sequences. Proceedings of the National Academy of Sciences of the United States of America, 105, 4781–4785. https://doi.org/10.1073/pnas.0707827105

- Tonkin, C. J., Pearce, J. A., McFadden, G. I., & Cowman, A. F. (2006)a. Protein targeting to destinations of the secretory pathway in the malaria parasite *Plasmodium falciparum*. Current Opinion in Microbiology, 9(4), 381–387. https://doi.org/10.1016/j.mib.2006.06.015
- Tonkin, C. J., Roos, D. S., & McFadden, G. I. (2006) b. N-terminal positively charged amino acids, but not their exact position, are important for apicoplast transit peptide fidelity in *Toxoplasma gondii*. Molecular and Biochemical Parasitology, 150(2), 192–200. https://doi.org/10.1016/j.molbiopara.2006.08.001
- Tonkin, C. J., Struck, N. S., Mullin, K. A., Stimmler, L. M., & McFadden, G. I. (2006) c. Evidence for Golgi-independent transport from the early secretory pathway to the plastid in malaria parasites. Molecular Microbiology, 61(3), 614–630. https://doi.org/10.1111/j.1365-2958.2006.05244.x
- Van Dooren, Giel G., Luciana M. Stimmler, and Geoffrey I. McFadden. 2006. "Metabolic Maps and Functions of the Plasmodium Apicoplast." *FEMS Microbiology Reviews* 30(4): 596–630.
- Van Dooren, G. G., Schwartzbach, S. D., Osafune, T., & McFadden, G. I. (2001). Translocation of proteins across the multiple membranes of complex plastids. Biochimica et Biophysica Acta - Molecular Cell Research, 1541(1–2), 34–53. https://doi.org/10.1016/S0167-4889(01)00154-9
- Van Dooren, G. G., Tomova, C., Agrawal, S., Humbel, B. M., & Striepen, B. (2008). *Toxoplasma gondii* Tic20 is essential for apicoplast protein import. Proceedings of the National Academy of Sciences of the United States of America, 105(36), 13574–9. https://doi.org/10.1073/pnas.0803862105
- Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., McFadden, G. I. (1998). 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, 95(21), 12352–7. https://doi.org/10.1073/PNAS.95.21.12352
- Waller, R. F., & McFadden, G. I. (2005). The apicoplast: A review of the derived plastid of apicomplexan parasites. Current Issues in Molecular Biology, 7(1), 57–80.
- Waller, R. F., Reed, M. B., Cowman, A. F., & McFadden, G. I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. The EMBO Journal, 19(8), 1794–802. https://doi.org/10.1093/emboj/19.8.1794

- Yeh, E., & DeRisi, J. L. (2011). Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. PLoS Biology, 9(8). https://doi.org/10.1371/journal.pbio.1001138
- Yung, S., Unnasch, T. R., & Lang-Unnasch, N. (2001). Analysis of apicoplast targeting and transit peptide processing in *Toxoplasma gondii* by deletional and insertional mutagenesis. Molecular and Biochemical Parasitology, 118(1), 11–21. https://doi.org/10.1016/S0166-6851(01)00359-0