N-terminal targeting sequences and coding sequences act in concert to determine the localization and trafficking pathway of apicoplast proteins in *Toxoplasma gondii*

Sofia Anjum [1,*], Aparna Prasad[1,2,*], Pragati Mastud[1,3] and Swati Patankar[1,4]

1. Department of Biosciences & Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India

2: Present address: Ismar Healthcare, Lier, Belgium

3: Present address: Cactus Communications, Mumbai, India

4: Corresponding author: patankar@iitb.ac.in

* Contributed equally

Abstract

Toxoplasma gondii has a relict plastid, the apicoplast; for most proteins that are localized in the apicoplast, N-terminal signal sequences precisely direct this localization. Proteins exclusively found in the apicoplast use a Golgi-independent route for trafficking, while dually targeted proteins found in both the apicoplast and the mitochondrion use a Golgi-dependent route. This report shows that the N-terminus of a dually targeted protein, *Tg*TPx1/2, has all the signals for targeting both to the apicoplast and the mitochondrion yet exhibits flexibility in organellar targeting depending on the reporter protein. When enhanced Green Fluorescent Protein (eGFP) is used as a reporter, *Tg*TPx1/2 targeting signals direct the protein to the mitochondrion, while the same targeting signals direct *T. gondii* endogenous proteins *Tg*SOD2 and *Tg*SOD3 to the apicoplast. Chimeric proteins consisting of N-termini and coding sequences from different proteins were tested for both their apicoplast localization and their trafficking route to the apicoplast. This report shows, for the first time, that in addition to the N-terminal signal sequences, targeting and trafficking signals also reside within the coding sequences of apicoplast proteins. Our results hint that during evolution when the apicoplast endosymbiont was transferring its genes to the host nucleus, signal sequences such as those of *Tg*TPx1/2 enabled sampling of different sub-cellular compartments before the final destination of their protein products was fixed.

Keywords: *Toxoplasma gondii*, apicoplast, trafficking pathways, signal sequences

Introduction

The phylum Apicomplexa includes the obligate intracellular parasites *Toxoplasma gondii* and *Plasmodium* species that are clinically relevant. Most of these parasites, including *T. gondii*, contain a relict plastid, the apicoplast (McFadden *et al.*, 1996), which though non-photosynthetic, harbors several essential metabolic pathways and is indispensable for parasite growth and survival (Jomaa *et al.*, 1999; Ralph *et al.*, 2004; Mazumdar *et al.*, 2006). The apicoplast was procured by secondary endosymbiosis, where an autotrophic alga containing a primary plastid was engulfed by another eukaryote (Bodył, 1999; Parsons *et al.*, 2007). Acquisition of the apicoplast was followed by gene transfer to the host nucleus, an event that conferred host control over the endosymbiont and confined the free-living eukaryote into a semi-autonomous organelle (reviewed in Waller & McFadden, 2005). Due to the transfer of genes to the nucleus, the apicoplast relies on the cell for translation of the newly acquired genes; therefore, apicomplexan parasites must use the cellular trafficking machinery to send nuclear-encoded proteins to the apicoplast.

Trafficking of nuclear-encoded proteins to a particular organelle is achieved by the presence of specific signals encoded within the protein sequence. For example, N-terminal signal peptides (SP) direct secretory proteins through the endoplasmic reticulum (ER) and eventually to the secretory pathway (von Heijne, 1990); an N-terminal mitochondrial targeting sequence (MTS) direct proteins to the mitochondrion (Roise and Schatz, 1988). For most apicoplast proteins, an N-terminal bipartite signal, an SP followed by a transit peptide (TP), serves as a specific tag for organellar localization (Waller *et al.*, 1998). The SP directs proteins into the ER, where it gets cleaved, exposing the TP, which then directs proteins to the apicoplast and is removed once inside the organelle, releasing the mature protein (Waller *et al.*, 1998, 2000; Ralph, 2004; Parsons *et al.*, 2007). Hence, the ER is the first step in apicoplast protein targeting. Interestingly, proteins from the ER can take two routes to the apicoplast, i.e., the classical secretory pathway via the Golgi or a Golgi-independent pathway (DeRocher *et al.*, 2005; Prasad, Mastud and Patankar, 2021).

Even though specific signals ensure localization to a particular compartment, there are reports of a single protein localizing in multiple organelles (Chew *et al.*, 2003; Pino *et al.*, 2007; Baudisch *et al.*, 2014). In *T. gondii*, thioredoxin peroxidase, superoxide dismutase, and aconitase (*Tg*TPx1/2, *Tg*SOD2, and *Tg*ACN) dually localize to the apicoplast and the mitochondrion (Pino *et al.*, 2007). A detailed study of the N terminal signal sequences of *Tg*TPx1/2 revealed that this dually localized protein has a single targeting sequence for both the apicoplast and the mitochondrion, as opposed to two distinct sequences (Mastud and Patankar, 2019).

Organelle-targeting signal sequences have been used to deliver reporters and cellular proteins to particular organelles (Huttly, 2009). In *T. gondii,* enhanced Green Fluorescent Protein (eGFP), among others, is a commonly employed fluorescent reporter for studying various aspects of parasite biology (Gubbels and Striepen, 2004; Striepen and Soldati, 2007; Hanig, Entzeroth and Kurth, 2012). The N-terminal targeting signals of exclusively apicoplast-targeted proteins are capable of localizing GFP to the apicoplast (Waller *et al.*, 1998; DeRocher *et al.*, 2005), however, the N-terminal sequences of *Tg*SOD2 (a protein dually localized to the apicoplast and mitochondrion) were unable to target GFP to the apicoplast, with the fusion protein showing only mitochondrial localization (Brydges and Carruthers, 2003; Pino *et al.*, 2007)

This report extends these published observations to another dually localized protein, *Tg*TPx1/2. Like *Tg*SOD2, a loss of apicoplast localization was observed when enhanced GFP (eGFP) was fused with the N-terminal signal sequences of *Tg*TPx1/2. However, apicoplast localization was restored when eGFP was replaced with endogenous proteins of *T. gondii*. Differential localization depending on different reporter proteins led us to investigate whether the choice of the reporter protein can also affect the choice of the trafficking pathway to the apicoplast. There are two routes to the apicoplast: proteins exclusively localized to the apicoplast follow a Golgi-independent pathway, while proteins dually localized to the apicoplast and the mitochondrion follow a Golgi-dependent pathway (Prasad, Mastud and Patankar, 2021). Therefore, we tested whether the choice of the trafficking pathway to the apicoplast is determined by N-terminal targeting sequences alone.

Here we show that the N-terminal sequence of the dually localized protein *Tg*TPx1/2 when used to drive apicoplast localization of *Tg*ACP (an apicoplast targeted protein) or *Tg*SOD2 (a protein dually targeting to the apicoplast and mitochondrion), results in both proteins taking the Golgi-independent route; for *Tg*SOD2, the trafficking route has been changed by swapping its N-terminal sequences with those of *Tg*TPx1/2. Similarly, the N-terminal sequence of the apicoplast protein *Tg*ACP when fused with the coding sequences of the dually targeted proteins *Tg*TPx1/2 and *Tg*SOD2 changes the trafficking pathway of both proteins from Golgi-dependent to Golgi-independent. Therefore, choice of the trafficking pathway of chimeric proteins depends on specific combinations of signal sequences and coding sequences, indicating for the first time, that apicoplast targeting motifs are present throughout the entire protein coding sequence. This work has implications for the field of apicoplast biology, which has been largely dominated by studies of apicoplast proteins where the N-terminal sequences are sufficient to drive proteins to their destinations.

Materials and Methods

Plasmid construction of *T. gondii* genes in expression vectors

For constructing C-terminal eGFP fusion proteins of *Tg*TPx1/2 (ToxoDB gene ID: TGGT1_266120), pCTG-TPx1/2-HA (Mastud and Patankar, 2019) was used as a template for amplifying the full-length protein (TPx1/2_{FL}) and the N-terminus encoding the 1-150 amino acids (TPx1/2_L). These DNA sequences were cloned between the *NheI* and *AvrII* sites of pCTG-eGFP (van Dooren *et al.*, 2008), generating pCTG-TPx1/2_{FL}-eGFP and pCTG-TPx1/2_L-eGFP, respectively. The TPx1/2(R24A)$_L$-eGFP plasmid was generated by site-directed mutagenesis, replacing the arginine at the position 24 with alanine.

For the generation of pCTG-TPx1/2_L-SOD2-HA and pCTG-TPx1/2_L-SOD3-HA plasmids, first, the N-terminus of TPx1/2_L(1-150 amino acids) was amplified from pCTG-TPx1/2-HA and cloned between *MfeI* and *NdeI* sites in the pCTG-HA vector. The coding sequences of *Tg*SOD2 (ToxoDB gene ID: TGGT1_316330) from 92-287 amino acids and *Tg*SOD3 (ToxoDB gene ID: TGGT1_316190) from 57-258 amino acids were amplified from pCTG-SOD2-HA and pCTG-SOD3-HA (Prasad, Mastud and Patankar, 2021) and inserted downstream of the TPx1/2_L in the *NdeI* site of the pCTG-TPx1/2_L-HA vector. For generating TPx1/2_L-SOD2-HA-HDEL and TPx1/2_L-SOD3-HA-HDEL constructs, the respective genes were amplified using a common forward primer and a reverse primer containing an HDEL sequence and were cloned in the pCTG-HA vector.

For the construction of ACP_L-eGFP, the N-terminal 104 amino acids of ACP (ToxoDB gene ID: TGGT1_264080) were amplified from cDNA and cloned between the *MfeI* and *NdeI* sites of pCTG-HA vector. This ACP_L (1-104 amino acids) was then amplified and cloned upstream of eGFP between the *NheI* and *AvrII* sites of the pCTG-eGFP vector. The HA and HDEL constructs of TPx1/2_L-ACP were constructed by replacing the coding sequence of SOD2 in the *NdeI* site from the HA, and HA-HDEL tagged pCTG-TPx1/2_L-SOD2 vector with the coding sequence of ACP (105-183 amino acids) amplified from cDNA.

For the generation of HA and HA-HDEL constructs of ACP_L-TPx1/2 and ACP_L-SOD2, first, pCTG-ACP_L-HA and pCTG-ACP_L-HA-HDEL were constructed by amplifying the first 104 amino acids of ACP_L from pCTG-ACP_L-eGFP and replacing it in the location of the ATrx1 gene between the *MfeI* and *NdeI* sites in the pCTG-ATrx1-HA and pCTG-ATrx1-HA-HDEL vectors (Prasad, Mastud and Patankar, 2021). The coding sequences of *Tg*TPx1/2 (151-333 amino acids) and *Tg*SOD2 (92-287 amino acids) were then amplified from pCTG-TPx1/2-HA and pCTG-

SOD2-HA and inserted downstream of ACP_L in the *NdeI* site of pCTG-ACP_L-HA and pCTG-ACP_L-HA-HDEL vector. All the clones were confirmed by sequencing. The primers used are listed in Table 1.

Table 1: List of primers used in this study. All the primers are shown in the 5' to 3' direction with the restriction sites underlined, and the desired mutations shown in bold.

Construct	**Primers**
TPx1/2_{FL}-eGFP	Forward: TACAGCTAGCGACAAAATGCTTCCTCTCTGCG Reverse: GGATCCTAGGGCCGCCGCCGAGAGGCTTCAGCTCGCCTGC
TPx1/2_L-eGFP	Forward: TACAGCTAGCGACAAAATGCTTCCTCTCTGCG Reverse: TAATCCTAGGGCCGCCGCCCGAGACCGGAATTTGGTCC
TPx1/2(R24A)$_L$-eGFP	Forward: CAGTGCTGCCG**GCG**CTTTGCTTGGCTACATG Reverse: GGCAAGACAGGCACTCCTGTCCGAGATACATC
TPx1/2_L-HA	Forward: TACACAATTGGACAAAATGCTTCCTCTCTGCGCGTCT Reverse: GGATCATATGCGAGACCGGAATTTGGTCC
TPx1/2_L-SOD2-HA	Forward: TACACATATGCTGCCCTACGACTACAATGCGC Reverse: TAGTCATATGCTGGCCTCCACCGACTGTCTCATCG
TPx1/2_L-SOD3-HA	Forward: TACACATATGCTCCCGTACGCCGAAGACG Reverse: GGATCATATGGTTGCTTTCAAGTGCTTTCACCAAGTTC
TPx1/2_L-SOD2-HA -HDEL	Forward: TACACAATTGGACAAAATGCTTCCTCTCTGCGCGTCT Reverse: ATTACTGCAGCTACAACTCGTCGTGCGCGTAGTCAGGGACGTCG
TPx1/2_L-SOD3-HA-HDEL	Forward: TACACAATTGGACAAAATGCTTCCTCTCTGCGCGTT Reverse: ATTACTGCAGCTACAACTCGTCGTGCGCGTAGTCAGGGACGTCG
ACP_L-HA	Forward: ATACAATTGACAAAATGGAGATGCATCCCC Reverse: GCTCATATGCCGATCATCAGAACTCG
ACP_L-eGFP	Forward: GAAGCTAGCAAAATGGAGATGCATCCCCGCAACGC Reverse: ATTCCTAGGGCCGCCGCCCCGATCATCAGAACTCG
TPx1/2_L-ACP-HA	Forward: TGCGCATATGCCCCTTCTTGAGAGAGTAAAGGACG Reverse: GTATCATATGCGCTGTAGCGCTCTTGGCTTTCTCGATATAGC
ACP_L-SOD2-HA	Forward: ATACAATTGACAAAATGGAGATGCATCCCC Reverse: GCTCATATGCCGATCATCAGAACTCG
ACP_L-TPx1/2-HA	Forward: GCTCATATGTTCAGCACCATTACTTTC

	Reverse: TATCATATGGAGAGGCTTCAGCTCG

Parasite maintenance and transient transfections

T. gondii wild-type RH strain were cultured and maintained in primary human foreskin fibroblast (HFF) cells as described previously (Striepen and Soldati, 2007). For transient transfection of tachyzoites by a single plasmid, parasites were harvested and resuspended in a reaction mixture of 400 µl of incomplete DMEM (without serum and gentamicin) with 50 µg DNA of plasmid. For transient co-transfection, extracellular parasites were resuspended in Cytomix buffer containing 120 mM KCl (Merck), 0.15 mM CaCl (Merck), 10 mM K_2HPO_4 (SRL Chemical), 10 mM KH_2PO_4 (Sisco Research Laboratories) pH 7.6, 25 mM HEPES pH 7.6 (Sigma-Aldrich), 2 mM EGTA (Merck), and 5 mM $MgCl_2$ (Merck) (Farhat *et al.*, 2021) with 35 µg of DNA of each plasmid. These reaction mixtures were then added in a 2 cm Bio-Rad GenePulser electroporation cuvette and electroporated at 1.5 kV, 50 Ω, and 25 µF using a Bio-Rad GenePulser Xcell system. Transfected parasites were grown in HFF cells for 24 hours, followed by immunostaining or western blot.

Immunostaining and microscopy

For immunostaining, transiently transfected parasites were grown in chamber slides (SPL Life Science Co.). After 24 hours, intracellular parasites were fixed in 4% formaldehyde (Merck) and 0.0075% glutaraldehyde in 1X PBS for 30 minutes. After fixing, parasites were washed twice with 1X PBS and were permeabilized by treating with 0.25% Triton X-100 (Sigma-Aldrich) for 10 minutes. Blocking was performed with 3% BSA (GeNei™) for one hour. After blocking, parasites were incubated with primary antibody, 1:1000 anti-HA rabbit monoclonal antibody (C29F4; Cell Signaling Technology), 1:1000 anti-HA mouse monoclonal antibody (Clone HA-7, mouse IgG1 Isotype; Sigma-Aldrich) or 1:1000 anti-ACP antibodies (a kind gift from Prof. Dhanasekaran Shanmugam, NCL, Pune) for 2 hours at room temperature followed by washing with 1X PBS. After three washes, parasites were incubated with appropriate secondary antibody for 1.5 hours at room temperature (anti-rabbit Alexa 488 (1:400), anti-rabbit 568 (1:400) and anti-mouse Alexa 594 (1:250) from Invitrogen™). Parasites were washed thrice in 1X PBS and mounted in ProLong Diamond Antifade mountant (Thermo Fisher Scientific). MitoTracker® red CMXRos (300 nM) from Invitrogen™ was used for staining mitochondria as described previously (Salunke *et al.*, 2018)

Confocal images were acquired by a Zeiss LSM 780 NLO confocal microscope using a Plan-Apochromat 100X, 1.4 NA objective. Optical sections along the Z-axis were captured and projected at maximum intensity.

Image processing and analysis

For colocalization studies, at least 40-50 vacuoles for each transfectant were observed microscopically, and 20-30 parasite vacuoles were imaged randomly for documentation. The images for the HA and HA-HDEL were taken using the same settings. Representative images were later processed for brightness and contrast; however, no non-linear adjustments were made, such as changes in the gamma settings. Unprocessed images were taken to measure Pearson's correlation coefficient (PCC) using the JaCoP plugin in Image J and Imaris software. A region of interest was drawn around the apicoplast marker in the entire Z-stack of the apicoplast, followed by automatic thresholding for measuring the PCC value. Quantitative analysis of the colocalization of proteins with the apicoplast marker was done using PCC values. In this study, based on the box plot (Figure S1), PCC values ranging between 0.4 to 1.0 were considered to have a strong correlation, and hence those proteins were said to be colocalized with the apicoplast marker protein. The PCC values ranging between 0.4 to -1 corresponded to a weak correlation, indicating a lack of colocalization of the protein with the apicoplast marker (Figure S1). A dot plot of PCC values showing the median with a 95% Confidence Interval was then plotted using GraphPad Prism version 8.0.1(244) for Windows, GraphPad Software, La Jolla, California, USA. The statistical significance of the change in the PCC values of HA and HA-HDEL constructs was analyzed using the Mann-Whitney test.

Comparison between different approaches for studying apicoplast localization with different apicoplast markers

Due to logistics issues regarding antibodies recognizing *Tg*ACP, the apicoplast marker, we were forced to use ACP_L-eGFP as the apicoplast marker midway through the study. First, we checked the co-transfection efficiency of ACP_L-eGFP with ACP_L-SOD2-HA. The co-transfection efficiency ranged between ~30-40% of the transfected parasites, showing that doubly transfected parasites could be easily detected for further immunofluorescence analysis. To compare the co-transfection strategy to the experiments that used anti-*Tg*ACP antibodies for staining the apicoplast, we co-transfected plasmids expressing *Tg*ACP_L-eGFP with all the constructs. No significant difference in the immunofluorescence or the median PCC values was observed between

the two strategies; therefore, the co-transfection strategy was deemed suitable for the remaining experiments.

Western blotting

For stable transfectants, ~10 million parasites were lysed in SDS-PAGE gel loading buffer (1X) and loaded on a 15% SDS-PAGE gel. The proteins were transferred onto a PVDF membrane at 35mA for 16 hours. The membrane was blocked with 3% BSA in PBS for 1 hour and incubated with primary antibody for 2 hours at a dilution of 1:1000 in 3% BSA-PBS. Primary antibodies used anti-HA Rabbit monoclonal antibody (C29F4; Cell Signaling Technology), anti-HA Mouse monoclonal antibody (Clone HA-7, mouse IgG1 Isotype; Sigma-Aldrich). The membrane was washed thrice in PBS and incubated with HRP-conjugated secondary antibody (anti-rabbit and anti-mouse; GeNei™) for 1 hour at a dilution of 1:2000 in 3% BSA-PBS. The blot was developed as per instructions for chemiluminescence using Clarity Western ECL Substrate (1705061; Bio-Rad).

Results

For many apicoplast proteins, N-terminal signal sequences can direct proteins to their destination (Waller *et al.*, 1998; DeRocher *et al.*, 2000; He *et al.*, 2001); however, for *Tg*SOD2, a protein dually localized to both the apicoplast and the mitochondrion, the N-terminus drives different reporter proteins to different organelles (Pino *et al.*, 2007). To understand the inherent nature of targeting signals in dually localized proteins, we made several chimeric constructs of *Tg*TPx1/2 and *Tg*SOD2 with heterologous reporters or native proteins of *T. gondii* and analyzed their effects on localization and trafficking routes to the apicoplast.

The N-terminus of *Tg*TPx1/2 drives eGFP to the mitochondrion, with a loss of apicoplast targeting

*Tg*TPx1/2 is a dually localized protein, which localizes to both the apicoplast and the mitochondrion, when tagged with small epitope tags such as hemagglutinin (HA) (Pino *et al.*, 2007). Due to their advantages during live-cell imaging, reporter proteins like GFP have been used in *T. gondii* for visualizing organelles (Waller *et al.*, 1998; Brydges and Carruthers, 2003). Therefore, we assessed the localization of an in-frame fusion of *Tg*TPx1/2$_{FL}$ with eGFP at the C-terminus and found that this protein showed a distinct mitochondrial localization, with no colocalization with the apicoplast marker (Figure 1A). This observation suggested that the eGFP reporter protein interferes with the apicoplast targeting of *Tg*TPx1/2.

Organellar targeting signals for apicoplast and mitochondrial import are present at the N-terminus of the protein (Sheiner and Soldati-Favre, 2008). For *Tg*TPx1/2, it was previously demonstrated that the first 50 amino acids of the protein fused to eGFP showed faint apicoplast and mitochondrial localization (Mastud and Patankar, 2019). Interestingly, a Western blot of this protein revealed that the signal cleavage sites for organellar uptake are present beyond the first 50 amino acids of the protein. An analysis of *Tg*TPx1/2 by the Simple Modular Architecture Research Tool (SMART) Version 9 (Letunic, Doerks and Bork, 2015; Letunic and Bork, 2018) predicted a glutathione peroxidase domain starting at amino acid 151. These data suggest that amino acids from 1-150 of *Tg*TPx1/2 should possess signals for targeting to both the apicoplast and the mitochondrion. To test this, we generated a chimeric construct encoding the first 150 amino acids of *Tg*TPx1/2 fused to eGFP (Figure 1B) and observed that parasites expressing *Tg*TPx1/2$_L$-eGFP showed colocalization with the mitochondrial marker SP-TP-SOD2-DsRed (Pino *et al.*, 2007). Apart from mitochondrial localization, the protein also showed a faint signal colocalizing with the apicoplast marker, acyl carrier protein (*Tg*ACP) in 42% of the population. However, quantitative analysis using the Pearson correlation coefficient (PCC) gave a median PCC value of 0.07, indicating no colocalization with the apicoplast marker. Charged amino acid mutants in the signal peptide region *Tg*TPx1/2(R24A) completely shift the localization of *Tg*TPx1/2 to the apicoplast (Mastud and Patankar, 2019). We generated an R24A mutant of *Tg*TPx1/2$_L$-eGFP and expected it to localize exclusively to the apicoplast. Contrary to expectations, *Tg*TPx1/2(R24A)$_L$-eGFP localized to the mitochondrion (Figure 1C), with only 30% of parasites showing a weak apicoplast signal. However, as the median PCC value for this signal was 0.10, we conclude that there is no localization of this protein to the apicoplast

As the full-length protein, the first 150 amino acids of *Tg*TPx1/2 and an R24A mutant of *Tg*TPx1/2$_L$ when fused to eGFP were all localized to the mitochondrion, we infer that eGFP interferes with the apicoplast localization of this dually targeted protein. This result is consistent with data seen for another dually targeted protein, *Tg*SOD2 (Brydges and Carruthers, 2003; Pino *et al.*, 2007), and in complete contrast to data for proteins that are localized only to the apicoplast (Waller *et al.*, 1998; DeRocher *et al.*, 2000; Wu *et al.*, 2015)

<u>The N-terminus of *Tg*TPx1/2 targets endogenous proteins to the apicoplast and weakly to the mitochondrion.</u>

As the heterologous reporter protein, eGFP affects the localization of *Tg*TPx1/2 to the apicoplast, we employed an endogenous protein of *T. gondii*, superoxide dismutase 3 (*Tg*SOD3), with a C-terminal HA tag as a reporter protein. *Tg*SOD3 is a mitochondrial protein containing a

mitochondrial targeting sequence from residues 1-56 (Brydges and Carruthers, 2003). The first 150 amino acids of *Tg*TPx1/2 were fused to *Tg*SOD3 (amino acids 57-258), devoid of its targeting signals. In contrast to *Tg*TPx1/2_L-eGFP, *Tg*TPx1/2_L-SOD3 was targeted to the apicoplast in the entire population of transiently transfected parasites, with a faint mitochondrial signal observed in a subset of the vacuoles (Figure 2B). The apicoplast localization of *Tg*TPx1/2_L-SOD3 was confirmed by colocalization with *Tg*ACP, showing a median PCC value of 0.50 (Figure 2C).

When the N-terminus of *Tg*TPx1/2 was fused to the coding sequence of another dually targeted protein, *Tg*SOD2 (92-287), the fusion protein *Tg*TPx1/2_L-SOD2 was targeted to the apicoplast in the entire population showing a median PCC value of 0.60 (Figure 2D). This fusion protein was present in the apicoplast and the mitochondrion in ~65% of the population (Figure 2A). Similar to the observations for *Tg*SOD3, the mitochondrial signal was weak (Figure 2A). Therefore, unlike eGFP, both *Tg*TPx1/2_LSOD3 and *Tg*TPx1/2_LSOD2 show robust apicoplast targeting and weak mitochondrial targeting. This result is different from another dually targeted protein, *Tg*SOD2. The N-terminus of *Tg*SOD2, when fused to GFP, is targeted only to the mitochondrion, but when fused to the coding sequence of *Tg*SOD3, dual targeting is restored (Pino *et al.*, 2007).

To confirm that the chimeric proteins are correctly processed, we performed western blots of the transiently transfected parasites using antibodies against the HA tag (Figure 2E). The chimeric constructs *Tg*TPx1/2_LSOD2, *Tg*TPx1/2_LSOD3 and the wild type *Tg*TPx1/2 showed pre-processed and processed bands of the expected sizes (~40-41 kDa and ~27-30 kDa, respectively), indicating that the N-terminal targeting signals have the appropriate sequences for cleavage in the apicoplast and mitochondrion.

<u>The coding sequences of apicoplast proteins contain signals that drive them through the Golgi-independent pathway</u>

Since the presence of endogenous proteins restored the apicoplast localization of fusion proteins having *Tg*TPx1/2 N-terminal signal sequences, we next checked the choice of pathway of these fusion proteins to the apicoplast. Dually targeted proteins are known to take a Golgi-dependent pathway to the apicoplast, while exclusively apicoplast-localized proteins are targeted independent of the Golgi (Prasad, Mastud and Patankar, 2021). We hypothesized that if the N-terminal 150 amino acids of *Tg*TPx1/2 with endogenous proteins can reach the apicoplast, the fusion protein should contain all the signals for trafficking through a Golgi-dependent pathway.

To test this hypothesis, an HDEL sequence was added at the C-terminal to *Tg*TPx1/2$_L$-SOD3-HA and *Tg*TPx1/2$_L$-SOD2-HA (Figures 3A and 3B). In *T. gondii*, the HDEL motif serves as a retrieval signal for ER proteins that have escaped to the Golgi apparatus by bulk flow (Hager *et al.*, 1999), and if a protein employs a Golgi-dependent route for trafficking, the addition of the HDEL motif results in a reduced localization of the protein to the destined compartment and an increase in ER retention (DeRocher *et al.*, 2000; Prasad, Mastud and Patankar, 2021). However, for *Tg*SOD2, a dually targeted protein that employs an ER-Golgi route for apicoplast trafficking, the addition of the HDEL sequence resulted in a reduced apicoplast localization without any increase in ER retention (Prasad, Mastud and Patankar, 2021). Therefore, in this study, ER retention was not assayed; instead, a reduced signal of the HDEL-containing protein in the apicoplast was taken to indicate that the protein traverses through the Golgi. No reduction in the apicoplast localization of HDEL-containing proteins would indicate that the protein employs a Golgi-independent route. This strategy has been used previously to understand the trafficking pathway of apicoplast proteins in *T. gondii* (Prasad, Mastud and Patankar, 2021)

For *Tg*TPx1/2$_L$-SOD3-HA, a significant decrease in the PCC values of the HDEL construct was observed when compared to the construct without an HDEL motif (P-value: 0.0104), indicating that the N-terminus of *Tg*TPx1/2 is able to drive the chimeric protein through the Golgi-dependent pathway to the apicoplast. Contrastingly, for *Tg*TPx1/2$_L$-SOD2-HA, no significant difference (P-value: 0.7665) was observed in the PCC value between the HA and the HDEL constructs, indicating a Golgi-independent route for this protein. As *Tg*SOD3 is a mitochondrial protein while *Tg*SOD2 is a dually targeted protein, we created a construct with N-terminal signal sequences of *Tg*TPx1/2 fused with the coding sequence (105-183 amino acids) of the apicoplast protein *Tg*ACP. Similar to *Tg*TPx1/2$_L$-SOD2, parasites expressing *Tg*TPx1/2$_L$-ACP with or without an HDEL motif showed no significant difference between the median PCC values (P-value: 0.9333) with the apicoplast marker *Tg*ACP$_L$-GFP (Figures 4D and 4F), indicating a Golgi-independent route for apicoplast trafficking. Therefore, the N-terminal targeting sequences of *Tg*TPx1/2 are able to target *Tg*SOD3 coding sequences to the apicoplast through the Golgi, while coding sequences of the apicoplast proteins *Tg*ACP and *Tg*SOD2 appear to contain signals for trafficking through the Golgi-independent route.

<u>The coding sequences of dually targeted proteins play a role in apicoplast trafficking through the Golgi.</u>

The exclusively apicoplast-targeted protein *Tg*ACP has an N-terminal leader (*Tg*ACP$_L$) that is sufficient to drive exogenous reporters like GFP and DsRed to the apicoplast (Waller *et al.*, 1998;

Pino *et al.*, 2007) using a Golgi-independent route (DeRocher *et al.*, 2005). We asked whether this leader sequence can drive dually targeted proteins to the apicoplast, and if so, whether they go through a Golgi-dependent or Golgi-independent pathway.

The *Tg*ACP leader sequence (1-104 amino acids) was fused with the coding sequence of the dually targeted protein *Tg*TPx1/2 (151-333 amino acids). Parasites expressing *Tg*ACP$_L$-TPx1/2-HA showed apicoplast localization in 60% of the population and apicoplast localization with perinuclear staining in 40% of the population (Figure 4A), similar to previous reports with other proteins (Waller *et al.*, 1998; He *et al.*, 2001; van Dooren *et al.*, 2002; DeRocher *et al.*, 2005; Karnataki *et al.*, 2009). To check the trafficking pathway of *Tg*ACP$_L$-TPx1/2-HA, we added an HDEL motif to its C-terminus. The median PCC values of the *Tg*ACP$_L$-TPx1/2-HA protein, with or without an HDEL motif showed no significant difference with a P-value of 0.7688, indicating a Golgi-independent route.

We next tested the ability of the *Tg*ACP leader sequence to drive the coding sequence of another dually localized protein, *Tg*SOD2, to the apicoplast. *Tg*ACP$_L$-SOD2 proteins with and without the HDEL motif were transiently expressed in *T. gondii*. Parasites expressing *Tg*ACP$_L$-SOD2-HA showed ~68% apicoplast localization and 32% apicoplast and perinuclear staining, while the HDEL construct showed 60% apicoplast localization and 40% apicoplast and a perinuclear signal throughout the cell. Interestingly, PCC analysis of these results showed that in two separate experiments, the trafficking pathway appeared to be different. In one experiment, the data indicated that the Golgi is involved in the trafficking of *Tg*ACP$_L$-SOD2-HA to the apicoplast (significant difference between the constructs with or without HDEL; P-value < 0.0001) while another experiment indicated that the Golgi is not involved (no significant difference between the constructs with or without HDEL; P-value: 0.8361). It is important to note that the apicoplast marker used in these two experiments was different (antibodies against *Tg*ACP vs. co-transfection with *Tg*ACP$_L$-eGFP). For all other experiments, the two strategies and repeated experiments gave identical results, therefore we conclude that *Tg*SOD2 behaves differently from the other reporters being tested. These data strengthen our hypothesis that a combination of both the N-terminal signal sequence and the coding sequence determines the pathway taken to reach the apicoplast.

Discussion

The ambiguous N-terminal signals of *Tg*TPx1/2 enable reporter proteins to sample multiple compartments

Enhanced GFP has been extensively used as a label for studying protein-protein interactions, measuring gene expression in real-time, and generating fluorescent organelle markers with well-defined signal sequences to facilitate the localization of newly identified proteins (Lippincott-Schwartz, Altan-Bonnet and Patterson, 2003; Snapp, 2005; Nelson, Cai and Nebenführ, 2007). In *T. gondii,* signal sequences of exclusive apicoplast proteins have been studied substantially using their eGFP fusions as organellar markers. In this report, we found that when eGFP is used as a reporter driven by the N-terminal signal sequences of *Tg*TPx1/2, instead of dual localization to both the mitochondrion and apicoplast, it showed only mitochondrion localization. Surprisingly, a charged residue mutation, R24A, in the N-terminus of *Tg*TPx1/2$_{1\text{-}150}$-eGFP should have resulted in apicoplast targeting (Mastud and Patankar, 2019), yet was also unable to drive eGFP to the apicoplast (Figure 2). As similar results were reported for *Tg*SOD2 (Pino *et al.*, 2007), the N-terminal targeting sequences of at least two dually targeted proteins are different from those of exclusive apicoplast proteins. *Tg*TPx1/2 and *Tg*SOD2 have been shown to evolve from similar evolutionary lineages and are more closely related to bacteria and mitochondrion homologues (Prasad, Mastud and Patankar, 2021), suggesting similarities in their N-terminal signal sequences.

*Tg*TPx1/2 contains ambiguous signals recognized by both the mitochondrial translocons as well as the signal recognition particle (SRP) for uptake in the ER, which is the first step in apicoplast trafficking. Earlier, a contest of strength or 'tug-of-war' was proposed between the two receptors for signal recognition of dually localized proteins with an ambiguous signal (Mastud and Patankar, 2019). After this early step, dually localized proteins use a Golgi-dependent pathway to reach the apicoplast. The presence of the eGFP reporter appears to interfere with one or more steps in this pathway to the apicoplast. Based on the complete lack of apicoplast or ER localization of eGFP fusions with the N-terminus of *Tg*TPx1/2, we propose that the very first step of apicoplast trafficking (SRP recognition and insertion into the ER) may be affected, allowing the mitochondrial translocons to win the tug-of-war.

Apicoplast localization was restored when eGFP was replaced with endogenous proteins, *Tg*SOD2 and *Tg*SOD3, as reporters (Figure 3). These results suggest that dual organellar targeting of fusion proteins having the *Tg*TPx1/2(1-150) N-terminus is affected by both the ambiguous N-terminal signal sequence of *Tg*TPx1/2 and the coding sequences of the mature proteins, which might

contain signals that are missing in eGFP. A surprising result obtained with these constructs was their weak mitochondrial localization. Mitochondrial targeting sequences are found in the N-terminus of the protein and should not be affected by the coding sequence of the fusion protein. Again, one could invoke the unique feature of dually targeted proteins: competition between the receptors at the apicoplast and mitochondrion. *Tg*SOD2 and *Tg*SOD3 as reporters may tip this balance in favour of the SRP rather than mitochondrial translocons.

What are the implications of fusion proteins showing different organellar localization? It has been proposed that evolutionarily, the acquired gene products from the endosymbiont do not attain signals for specific protein targeting to the new organelle immediately. Instead, these proteins explore different organellar environments before choosing the compartment where the proteins can form a complete biochemical pathway (Martin, 2010). For these expeditions, it would be favourable if the signal sequences and the coding sequences showed plasticity to be able to sample different organelles. Our work provides a glimpse into the evolution of protein trafficking to the endosymbiotic organelles, where the composite plasticity of the N-terminal and coding sequences influence the final destination of the protein: the trafficking of *Tg*TPx1/2 fused with either EGFP or endogenous proteins to different compartments represents a molecular mechanism of exploring different destinations during the evolutionary process.

<u>Coding sequences of apicoplast proteins play a role in their choice of trafficking pathways.</u>

Proteins are trafficked to the apicoplast via two distinct pathways (reviewed in Mallo *et al.*, 2018; Prasad, Mastud and Patankar, 2021); however, what factors determine which pathway to follow is still unclear. It was earlier suggested that the difference in the strength of the transit peptides present at the N-terminus of exclusive apicoplast proteins and the dually localized proteins could be the determining factor for the choice of the pathway (Prasad, Mastud and Patankar, 2021). Results presented in this report show that signals present at the N-terminus of *Tg*TPx1/2 are strong enough to take a protein to the apicoplast via the Golgi-dependent pathway, as shown by *Tg*TPx1/2_L-SOD3 (Figure 4B). But, when fused with the coding sequence of exclusive apicoplast protein *Tg*ACP, the chimeric proteins use the Golgi-independent pathway (Figure 4C), suggesting there are signals in the coding sequence of *Tg*ACP which supersede the signals in *Tg*TPx1/2. Strangely, the coding sequences of the dually localized protein *Tg*SOD2, which originally takes a Golgi-dependent route to the apicoplast, when driven by the N-terminal targeting sequences of *Tg*TPx1/2, follow a Golgi-independent pathway (Figure 4A). This result suggests that the combination of these two dually localized proteins behaves antagonistically, changing the native

pathway. It is interesting to note that the signal sequences of *Tg*TPx1/2 can drive a mitochondrial protein to the apicoplast through the Golgi-dependent pathway, but two apicoplast proteins (*Tg*SOD2 and *Tg*ACP) appear to contain signals in their coding regions that dominate over the signal sequences of the *Tg*TPx1/2 N-terminus leading to a Golgi-independent route to the apicoplast. Clearly, coding sequences also play a role in apicoplast trafficking in *T. gondii*, a phenomenon that has been reported in the plastid of the diatom *Phaeodactylum tricornutum* (Felsner, Sommer and Maier, 2010).

In contrast to the results seen with the N-terminus of *Tg*TPx1/2, the N-terminus of ACP was able to drive all reporter proteins through the Golgi-independent pathway. These data suggest that there are recognition sequences in the ACP leader that are strongly recognized by receptors in the ER. It is important to note that *Tg*SOD2 shows both Golgi-dependent and Golgi-independent trafficking to the apicoplast when fused with the *Tg*ACP N-terminus. Further studies are required to understand the mechanistic explanation underlying this result; *Tg*SOD2 being a superoxide dismutase, this enzyme may choose different routes depending on local micro-environments with different oxidative stress levels. Interestingly, *Tg*SOD2 displays unusual properties in another set of experiments: when fused with HDEL, the native protein is not retained in the ER and instead shows a complete loss of apicoplast localization (Pino *et al.*, 2007; Prasad, Mastud and Patankar, 2021)

The results lead to a model (summarized in Figure 5B) that indicates a hierarchy of recognition by receptors in different organelles of *T. gondii*. As an apicoplast protein enters the ER, its SP is cleaved, revealing the TP. For the exclusive apicoplast protein, *Tg*ACP, this TP is recognized by TP receptors in the ER and can drive any reporter protein through the Golgi-independent route. Therefore, the *Tg*ACP TP is the strongest signal for the Golgi-independent trafficking pathway. In this report, we show, for the first time, that coding sequences of apicoplast proteins also have signals for trafficking via the Golgi-independent pathway, which are used when driven by the N-terminal sequences of *Tg*TPx1/2. Finally, the TP sequences of *Tg*TPx1/2 differ from the TP of *Tg*ACP, thereby directing the protein to follow a default route into the Golgi. This default route can be explained by the bulk flow of proteins from the ER to the Golgi (Wieland *et al.*, 1987). Therefore, the TP of *Tg*TPx1/2 must be recognized by TP receptors in the Golgi, that then carry the protein to the apicoplast. There is evidence for receptors for apicoplast proteins in the Golgi (Cao *et al.*, 2021), although *Tg*TPx1/2 was not tested (Prasad and Patankar, 2021) in the study. We emphasize that this model is based on the N-terminal sequences of *Tg*TPx1/2 and extrapolation to other proteins would require further experiments using their N-terminal sequences to generate chimeric proteins as done in this study.

Conclusion

In sum, we show that the N-terminal targeting sequences of dually targeted proteins are able to drive different reporters to the apicoplast or mitochondrion with varying efficiency. These results indicate that targeting signals exist in the coding sequences of apicoplast proteins, and intriguingly, these targeting signals can also drive the choice of the trafficking pathways to the apicoplast. This work has technical implications as it suggests that fusions of the N-terminal signal sequences of apicoplast proteins with reporters should be done with caution. The work also has biological implications as it provides evidence of N-terminal signal sequences having the flexibility to drive different reporters to different organelles, a "mix and match" strategy that may have evolutionary benefits.

References

Baudisch, B. *et al.* (2014) 'The exception proves the rule? Dual targeting of nuclear-encoded proteins into endosymbiotic organelles', *New Phytologist*, 201(1), pp. 80–90. Available at: https://doi.org/10.1111/nph.12482.

Bodył, A. (1999) 'Evolutionary pathway of the apicomplexan plastids and its implications', *Trends in Microbiology*, 7(7), pp. 266–267. Available at: https://doi.org/10.1016/S0966-842X(99)01520-6.

Brydges, S.D. and Carruthers, V.B. (2003) 'Mutation of an unusual mitochondrial targeting sequence of SODB2 produces multiple targeting fates in *Toxoplasma gondii*', *Journal of Cell Science*, 116(22), pp. 4675–4685. Available at: https://doi.org/10.1242/jcs.00750.

Cao, S. *et al.* (2021) 'The Dissection of SNAREs Reveals Key Factors for Vesicular Trafficking to the Endosome-like Compartment and Apicoplast via the Secretory System in *Toxoplasma gondii*', *mBio*, 12(4), p. e0138021. Available at: https://doi.org/10.1128/mBio.01380-21.

Cavalier-Smith, T. (1999) 'Principles of Protein and Lipid Targeting in Secondary Symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan Plastid Origins and the Eukaryote Family Tree, [2]', *The Journal of Eukaryotic Microbiology*, 46(4), pp. 347–366. Available at: https://doi.org/10.1111/j.1550-7408.1999.tb04614.x.

Chew, O. *et al.* (2003) 'Characterization of the targeting signal of dual-targeted pea glutathione reductase', *Plant Molecular Biology*, 53(3), pp. 341–356. Available at: https://doi.org/10.1023/B:PLAN.0000006939.87660.4f.

DeRocher, A. *et al.* (2000) 'Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system', *Journal of Cell Science*, 113 (Pt 22), pp. 3969–3977. Available at: https://doi.org/10.1242/jcs.113.22.3969.

DeRocher, A. *et al.* (2005) 'Dissection of brefeldin A-sensitive and -insensitive steps in apicoplast protein targeting', *Journal of Cell Science*, 118(3), pp. 565–574. Available at: https://doi.org/10.1242/jcs.01627.

van Dooren, G.G. *et al.* (2002) 'Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme', *The Journal of Biological Chemistry*, 277(26), pp. 23612–23619. Available at: https://doi.org/10.1074/jbc.M201748200.

van Dooren, G.G. *et al.* (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), pp. 13574–13579. Available at: https://doi.org/10.1073/pnas.0803862105.

Farhat, D.C. *et al.* (2021) 'A plant-like mechanism coupling m6A reading to polyadenylation safeguards transcriptome integrity and developmental gene partitioning in *Toxoplasma* ', *eLife*. Edited by D. Bourc'his and D. Soldati-Favre, 10, p. e68312. Available at: https://doi.org/10.7554/eLife.68312.

Felsner, G., Sommer, M.S. and Maier, U.G. (2010) 'The physical and functional borders of transit peptide-like sequences in secondary endosymbionts', *BMC Plant Biology*, 10, p. 223. Available at: https://doi.org/10.1186/1471-2229-10-223.

Gubbels, M.-J. and Striepen, B. (2004) 'Studying the Cell Biology of Apicomplexan Parasites Using Fluorescent Proteins', *Microscopy and Microanalysis*, 10(5), pp. 568–579. Available at: https://doi.org/10.1017/S1431927604040899.

Hager, K.M. *et al.* (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(16), pp. 2631–2638. Available at: https://doi.org/10.1242/jcs.112.16.2631.

Hanig, S., Entzeroth, R. and Kurth, M. (2012) 'Chimeric fluorescent reporter as a tool for generation of transgenic Eimeria (Apicomplexa, Coccidia) strains with stage specific reporter gene expression', *Parasitology International*, 61(3), pp. 391–398. Available at: https://doi.org/10.1016/j.parint.2012.01.010.

He, C.Y. *et al.* (2001) 'Targeting and Processing of Nuclear-encoded Apicoplast Proteins in Plastid Segregation Mutants of *Toxoplasma gondii**', *Journal of Biological Chemistry*, 276(30), pp. 28436–28442. Available at: https://doi.org/10.1074/jbc.M102000200.

von Heijne, G. (1990) 'The signal peptide', *The Journal of Membrane Biology*, 115(3), pp. 195–201. Available at: https://doi.org/10.1007/BF01868635.

Huttly, A. (2009) 'Reporter Genes', in H.D. Jones and P.R. Shewry (eds) *Transgenic Wheat, Barley and Oats: Production and Characterization Protocols*. Totowa, NJ: Humana Press (Methods in Molecular Biology™), pp. 39–69. Available at: https://doi.org/10.1007/978-1-59745-379-0_3.

Jomaa, H. *et al.* (1999) 'Inhibitors of the Nonmevalonate Pathway of Isoprenoid Biosynthesis as Antimalarial Drugs', *Science*, 285(5433), pp. 1573–1576. Available at: https://doi.org/10.1126/science.285.5433.1573.

Karnataki, A. *et al.* (2009) 'Sequential processing of the *Toxoplasma* apicoplast membrane protein FtsH1 in topologically distinct domains during intracellular trafficking', *Molecular and Biochemical Parasitology*, 166(2), pp. 126–133. Available at: https://doi.org/10.1016/j.molbiopara.2009.03.004.

Letunic, I. and Bork, P. (2018) '20 years of the SMART protein domain annotation resource', *Nucleic Acids Research*, 46(D1), pp. D493–D496. Available at: https://doi.org/10.1093/nar/gkx922.

Letunic, I., Doerks, T. and Bork, P. (2015) 'SMART: recent updates, new developments and status in 2015', *Nucleic Acids Research*, 43(Database issue), pp. D257-260. Available at: https://doi.org/10.1093/nar/gku949.

Lippincott-Schwartz, J., Altan-Bonnet, N. and Patterson, G.H. (2003) 'Photobleaching and photoactivation: following protein dynamics in living cells', *Nature Cell Biology*, Suppl, pp. S7-14.

Mallo, N. *et al.* (2018) 'Protein Import into the Endosymbiotic Organelles of Apicomplexan Parasites', *Genes*, 9(8), p. 412. Available at: https://doi.org/10.3390/genes9080412.

Martin, W. and Herrmann, R.G. (1998) 'Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why?1', *Plant Physiology*, 118(1), pp. 9–17. Available at: https://doi.org/10.1104/pp.118.1.9.

Mastud, P. and Patankar, S. (2019) 'An ambiguous N-terminus drives the dual targeting of an antioxidant protein Thioredoxin peroxidase (TgTPx1/2) to endosymbiotic organelles in *Toxoplasma gondii*', *PeerJ*, 7, p. e7215. Available at: https://doi.org/10.7717/peerj.7215.

Mazumdar, J. *et al.* (2006) 'Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in *Toxoplasma gondii*', *Proceedings of the National Academy of Sciences of the United States of America*, 103(35), pp. 13192–13197. Available at: https://doi.org/10.1073/pnas.0603391103.

McFadden, G.I. *et al.* (1996) 'Plastid in human parasites', *Nature*, 381(6582), pp. 482–482. Available at: https://doi.org/10.1038/381482a0.

Nelson, B.K., Cai, X. and Nebenführ, A. (2007) 'A multicolored set of in vivo organelle markers for colocalization studies in Arabidopsis and other plants', *The Plant Journal*, 51(6), pp. 1126–1136. Available at: https://doi.org/10.1111/j.1365-313X.2007.03212.x.

Parsons, M. *et al.* (2007) 'Protein Trafficking to the Apicoplast: Deciphering the Apicomplexan Solution to Secondary Endosymbiosis', *Eukaryotic Cell*, 6(7), pp. 1081–1088. Available at: https://doi.org/10.1128/EC.00102-07.

Pino, P. *et al.* (2007) 'Dual Targeting of Antioxidant and Metabolic Enzymes to the Mitochondrion and the Apicoplast of *Toxoplasma gondii*', *PLOS Pathogens*, 3(8), p. e115. Available at: https://doi.org/10.1371/journal.ppat.0030115.

Prasad, A., Mastud, P. and Patankar, S. (2021) 'Dually localized proteins found in both the apicoplast and mitochondrion utilize the Golgi-dependent pathway for apicoplast targeting in *Toxoplasma gondii*', *Biology of the Cell*, 113(1), pp. 58–78. Available at: https://doi.org/10.1111/boc.202000050.

Prasad, A. and Patankar, S. (2021) 'Recognition of Two Distinct Pathways for Trafficking of Proteins to the Apicoplast', *mBio*, 12(6), pp. e02634-21. Available at: https://doi.org/10.1128/mBio.02634-21.

Ralph, S.A. (2004) 'Evolutionary Pressures on Apicoplast Transit Peptides', *Molecular Biology and Evolution*, 21(12), pp. 2183–2194. Available at: https://doi.org/10.1093/molbev/msh233.

Ralph, S.A. *et al.* (2004) 'Metabolic maps and functions of the *Plasmodium falciparum* apicoplast', *Nature Reviews Microbiology*, 2(3), pp. 203–216. Available at: https://doi.org/10.1038/nrmicro843.

Roise, D. and Schatz, G. (1988) 'Mitochondrial presequences.', *Journal of Biological Chemistry*, 263(10), pp. 4509–4511. Available at: https://doi.org/10.1016/S0021-9258(18)68809-X.

Salunke, R. *et al.* (2018) 'Highly diverged novel subunit composition of apicomplexan F-type ATP synthase identified from *Toxoplasma gondii*', *PLoS Biology*, 16(7), p. e2006128. Available at: https://doi.org/10.1371/journal.pbio.2006128.

Sheiner, L. and Soldati-Favre, D. (2008) 'Protein trafficking inside *Toxoplasma gondii*', *Traffic (Copenhagen, Denmark)*, 9(5), pp. 636–646. Available at: https://doi.org/10.1111/j.1600-0854.2008.00713.x.

Snapp, E. (2005) 'Design and use of fluorescent fusion proteins in cell biology', *Current Protocols in Cell Biology*, Chapter 21, p. 21.4.1-21.4.13. Available at: https://doi.org/10.1002/0471143030.cb2104s27.

Striepen, B. and Soldati, D. (2007) '15 - Genetic Manipulation of *Toxoplasma gondii*', in L.M. Weiss and K. Kim (eds) *Toxoplasma gondii*. London: Academic Press, pp. 391–418. Available at: https://doi.org/10.1016/B978-012369542-0/50017-9.

Waller, R.F. *et al.* (1998) 'Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*', *Proceedings of the National Academy of Sciences*, 95(21), pp. 12352–12357. Available at: https://doi.org/10.1073/pnas.95.21.12352.

Waller, R.F. *et al.* (2000) 'Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway', *The EMBO Journal*, 19(8), pp. 1794–1802. Available at: https://doi.org/10.1093/emboj/19.8.1794.

Waller, R.F. and McFadden, G.I. (2005) 'The Apicoplast: A Review of the Derived Plastid of Apicomplexan Parasites', *Current Issues in Molecular Biology*, 7(1), pp. 57–80. Available at: https://doi.org/10.21775/cimb.007.057.

Wieland, F.T. *et al.* (1987) 'The rate of bulk flow from the endoplasmic reticulum to the cell surface', *Cell*, 50(2), pp. 289–300. Available at: https://doi.org/10.1016/0092-8674(87)90224-8.

Wu, Liang *et al.* (2015) '*Toxoplasma gondii* nucleus coding apicoplast protein ACP synthesis and trafficking in delayed death', *Parasitology Research*, 114(3), pp. 1099–1105. Available at: https://doi.org/10.1007/s00436-014-4281-2.

Acknowledgements

We thank the Department of Biotechnology, Govt. of India, for research funding (Grant no. BT/PR13546/BRB/10/1423/2015) to SP and the PhD fellowship (DBT/2018/IITB/1149) to SA. We thank the Ministry of Education, Govt. of India and IIT Bombay for the PhD fellowships to AP and PM. We acknowledge the Indian Institute of Technology Bombay for their Confocal Laser Scanning Microscope facility. We appreciate Manesh Joshi for constructing the pCTG-ACP$_L$-HA plasmid and Geetanjali Mishra for her constructive feedback on the manuscript.

Authors contributions

SA, AP, PM and SP conceived and designed the experiments. SA, AP, PM performed the experiments. SA, AP and SP analyzed the data and prepared the manuscript. SA prepared the figures. SA, AP, PM and SP reviewed drafts of the manuscript.

Figure legends

Figure 1: GFP reporter results in the localization of *Tg*TPx1/2 to the mitochondrion but not the apicoplast. Microscopic images of parasites transiently expressing **A)** *Tg*TPx1/2-eGFP (green), **B)** *Tg*TPx1/2$_L$-eGFP (green), with apicoplast marker anti-*Tg*ACP antibodies (red) and mitochondrion marker, SP-TP-SOD2-DsRed (red), **C)** *Tg*TPx1/2(R24A)$_L$-eGFP (green), with apicoplast marker anti-*Tg*ACP antibodies (red) and mitochondrion marker, Mitotracker.

Figure 2: The N-terminus of the dually targeted protein *Tg*TPx1/2 can target endogenous proteins to the apicoplast and the mitochondrion. The microscopic images of parasites expressing HA-tagged protein **A)** *Tg*TPx1/2$_L$-SOD2, **B)** *Tg*TPx1/2$_L$-SOD3 with apicoplast marker anti-*Tg*ACP antibodies (green) and mitochondrial marker MitoTracker (red), **(C, D)** PCC values showing median value with 95% confidence interval of *Tg*TPx1/2$_L$-SOD2 and *Tg*TPx1/2$_L$-SOD3 with apicoplast marker anti-*Tg*ACP antibodies, **E)** Western blot analysis using anti-HA antibodies of the cell lysate of parasites transiently transfected with *Tg*TPx1/2-HA, *Tg*TPx1/2$_L$-SOD2, and *Tg*TPx1/2$_L$-SOD3.

Figure 3: Evaluation of the pathway taken to the apicoplast by chimeric constructs consisting of the N-terminal of *Tg*TPx1/2 fused with endogenous reporters. Each panel in red represents microscopic images of parasites transiently expressing HDEL-tagged protein of **A)** *Tg*TPx1/2$_L$-SOD2 **B)** *Tg*TPx1/2$_L$-SOD3 with apicoplast marker anti-*Tg*ACP antibodies (green) **C, D)** Pooled PCC values of HA and HDEL tagged proteins of *Tg*TPx1/2$_L$-SOD2, *Tg*TPx1/2$_L$-SOD3 with apicoplast marker, anti-*Tg*ACP antibodies showing median value with

95% confidence interval respectively **E)** Microscopic images of *Tg*TPx1/2$_L$-ACP (red) with apicoplast marker *Tg*ACP-eGFP (green). **F)** Pooled PCC values of HA and HDEL tagged proteins of *Tg*TPx1/2$_L$-ACP with apicoplast marker, *Tg*ACP-eGFP showing median value with 95% confidence interval.

Figure 4: Analysis of the role of coding sequences of dually localized proteins in deciding the trafficking route. The microscopic images of parasites expressing tagged protein with or without HDEL motif are presented in red, and the apicoplast marker is shown in green **A)** *Tg*ACP$_L$-TPx1/2, with apicoplast marker *Tg*ACP-eGFP **B)** PCC values of *Tg*ACP$_L$-TPx1/2 constructs with and without HDEL motif showing median value with 95% confidence interval of *Tg*TPx1/2$_L$-ACP with *Tg*ACP-eGFP **C) and D)** Microscopic images of *Tg*ACP$_L$-SOD2 with and without HDEL motif from two independent experiment with apicoplast marker anti-*Tg*ACP antibodies and *Tg*ACP-eGFP respectively. **E and F**) PCC values of HA and HDEL constructs showing median value with 95% confidence interval of *Tg*ACP$_L$-SOD2 with anti-*Tg*ACP antibodies, and with the apicoplast marker *Tg*ACP-eGFP, respectively.

Figure 5: A summary of the localization and trafficking pathway of the chimeric proteins reported in this study A) Schematic representation of chimeric proteins. The diagram represents the amino acid residues used to generate the chimeric proteins and the name assigned to each construct. All the chimeric proteins contain either eGFP or HA at the C-terminal to study their localization in the parasite. An HDEL motif was fused in-frame at the C-terminal end to study the trafficking pathway. SP-signal peptide, TP-transit peptide, LP-leader peptide. **B)** Model representing the trafficking pathway taken by the fusion constructs. ACP$_L$-SOD2* construct follows different trafficking pathways in two independent experiments. ACP-TP: ACP transit peptide, TPx1/2 TP- Transit peptide of *Tg*TPx1/2, ER receptor: transit peptide receptor in the Endoplasmic Reticulum, Golgi TP receptor: Transit peptide receptor in Golgi (Created with BioRender.com)

Figures

Figure 1

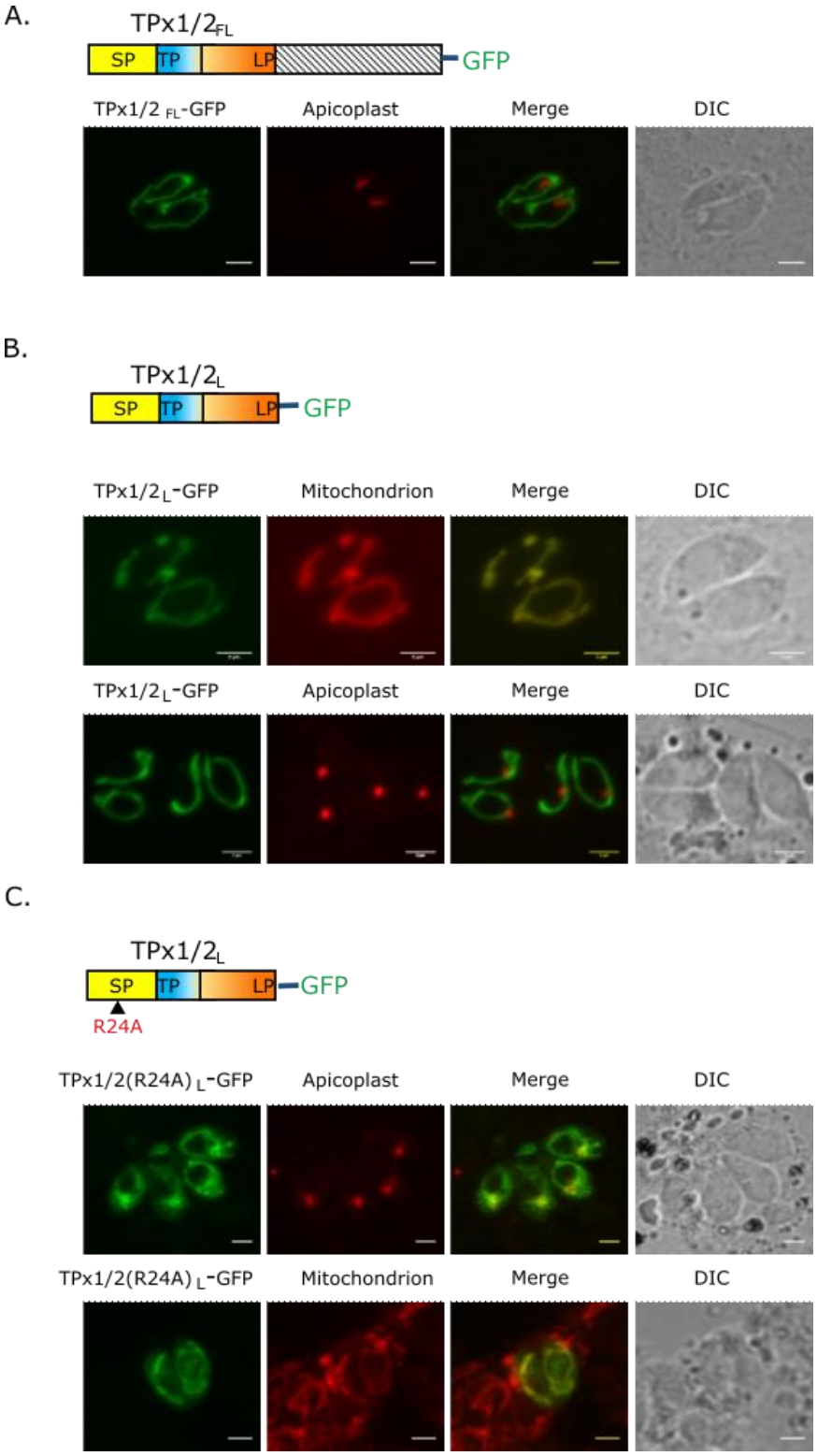


Figure 2

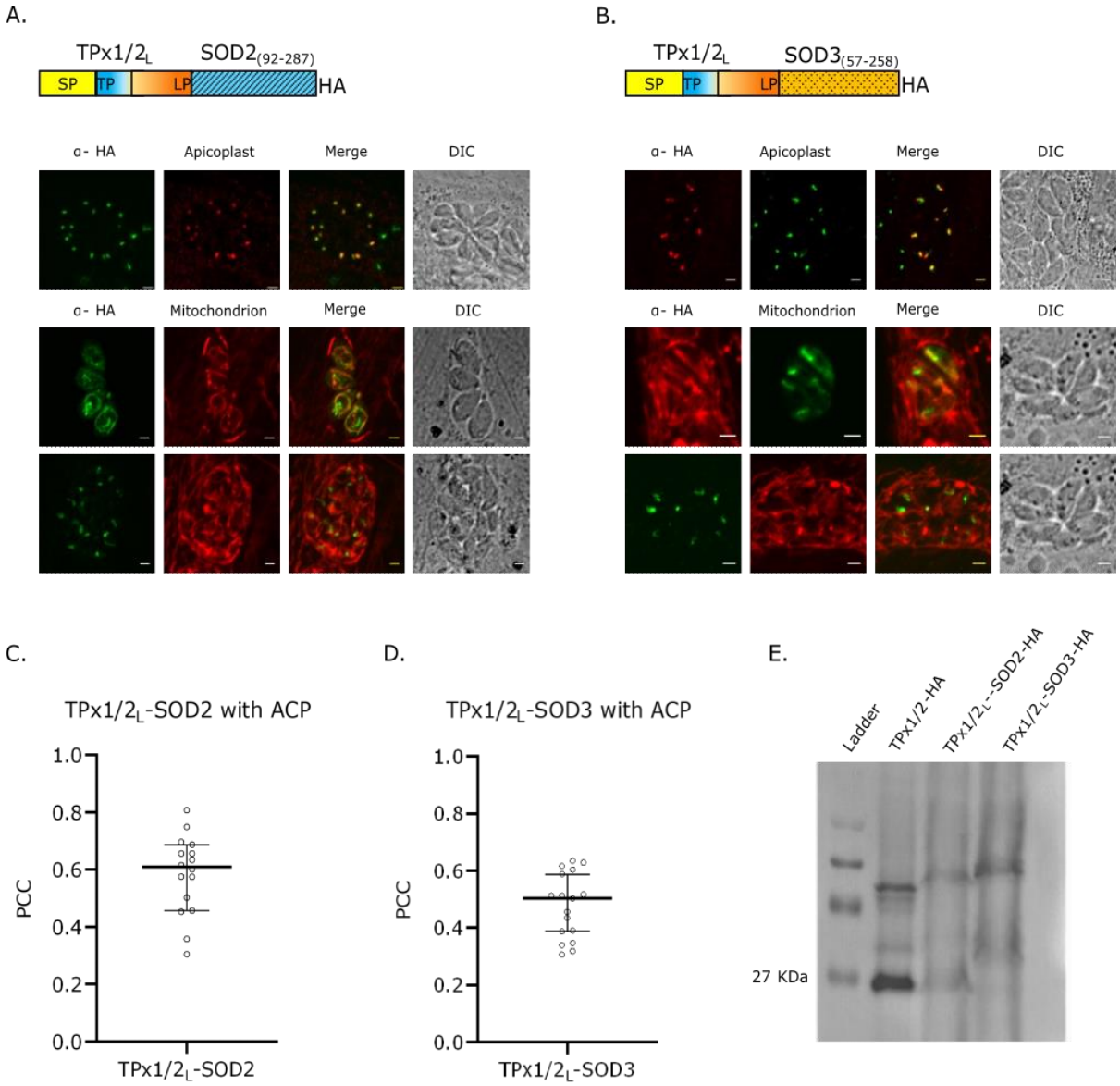

A.
TPx1/2L
SOD2(92-287)
SP
TP
LP
HA
α- HA
Apicoplast
Merge
DIC
α- HA
Mitochondrion
Merge
DIC
B.
TPx1/2L
SOD3(57-258)
SP
TP
LP
HA
α- HA
Apicoplast
Merge
DIC
α- HA
Mitochondrion
Merge
DIC
C.
TPx1/2L-SOD2 with ACP
PCC
1.0
0.8
0.6
0.4
0.2
0.0
TPx1/2L-SOD2
D.
TPx1/2L-SOD3 with ACP
PCC
1.0
0.8
0.6
0.4
0.2
0.0
TPx1/2L-SOD3
E.
Ladder
TPx1/2-HA
TPx1/2L-SOD2-HA
TPx1/2L-SOD3-HA
27 KDa

Figure 3

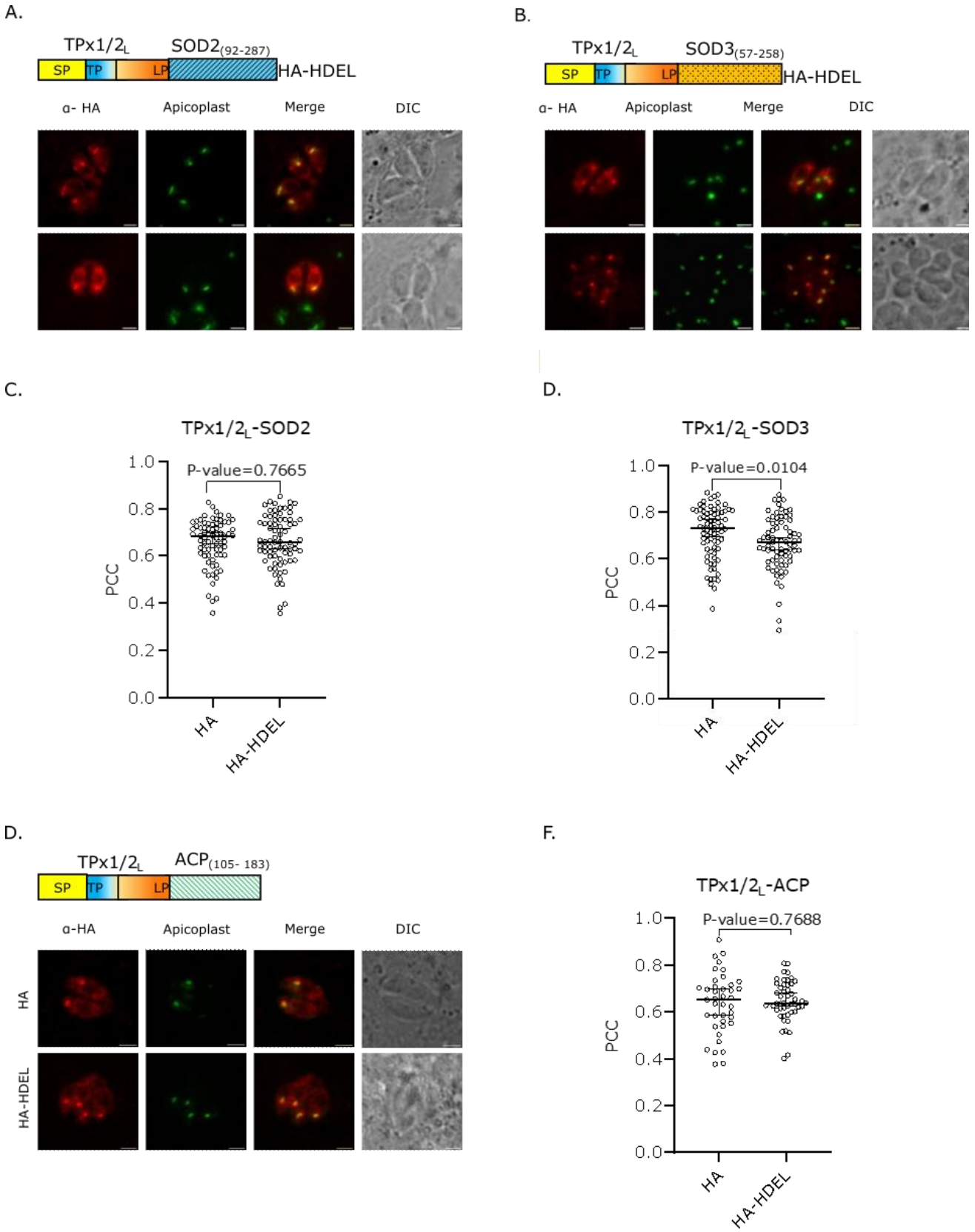

A.
TPx1/2L
SOD2(92-287)
SP
TP
LP
HA-HDEL
α- HA
Apicoplast
Merge
DIC
B.
TPx1/2L
SOD3(57-258)
SP
TP
LP
HA-HDEL
α- HA
Apicoplast
Merge
DIC
C.
TPx1/2L-SOD2
P-value=0.7665
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HA-HDEL
D.
TPx1/2L-SOD3
P-value=0.0104
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HA-HDEL
D.
TPx1/2L
ACP(105- 183)
SP
TP
LP
α-HA
Apicoplast
Merge
DIC
HA
HA-HDEL
F.
TPx1/2L-ACP
P-value=0.7688
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HA-HDEL

Figure 4

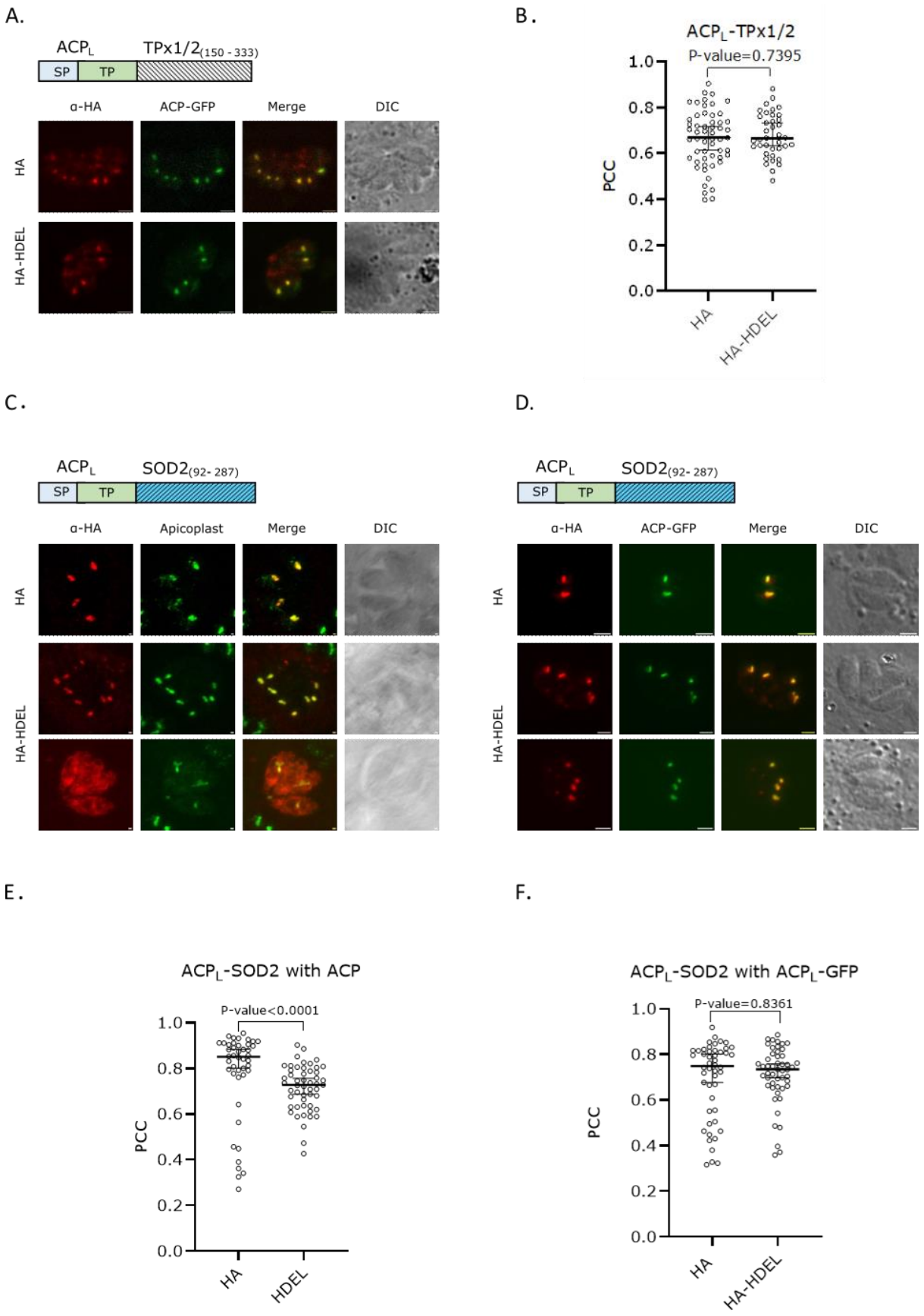
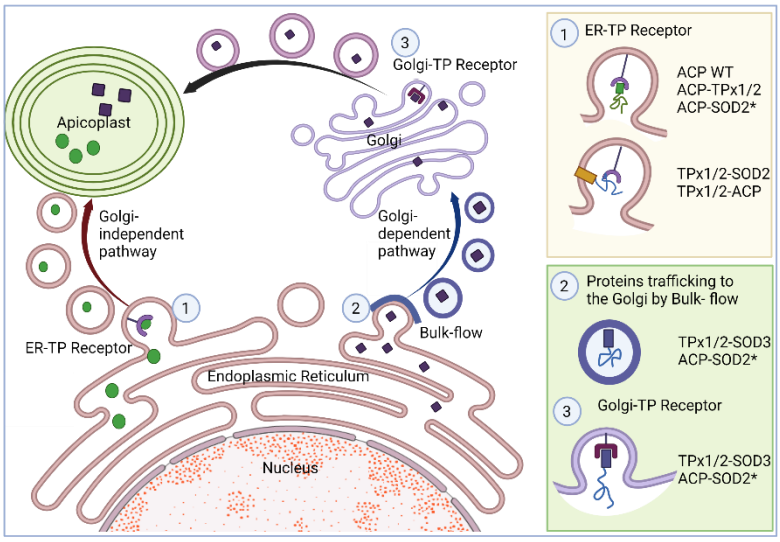

A.
ACPL
TPx1/2(150 - 333)
SP
TP
α-HA
ACP-GFP
Merge
DIC
HA
HA-HDEL
B.
ACPL-TPx1/2
P-value=0.7395
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HA-HDEL
C.
ACPL
SOD2(92- 287)
SP
TP
α-HA
Apicoplast
Merge
DIC
HA
HA-HDEL
D.
ACPL
SOD2(92- 287)
SP
TP
α-HA
ACP-GFP
Merge
DIC
HA
HA-HDEL
E.
ACPL-SOD2 with ACP
P-value<0.0001
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HDEL
F.
ACPL-SOD2 with ACPL-GFP
P-value=0.8361
PCC
1.0
0.8
0.6
0.4
0.2
0.0
HA
HA-HDEL

Figure 5

A.

Constructs	Schematic representation of constructs	Localization	Pathway
TPx1/2$_{FL}$- GFP	TPx1/2$_{FL}$ SP TP LP GFP	Mitochondrion	
TPx1/2$_{L}$-GFP	TPx1/2$_{L}$ SP TP LP GFP	Mitochondrion	
TPx1/2(R24A)$_{L}$-GFP	TPx1/2$_{L}$ SP TP LP GFP R24A	Mitochondrion	
TPx1/2$_{L}$-SOD2-HA	TPx1/2$_{L}$ SOD2$_{92-287}$ SP TP LP HA	Apicoplast + weak mitochondrion	Golgi-independent
TPx1/2$_{L}$-SOD2-HA-HDEL	TPx1/2$_{L}$ SOD2$_{92-287}$ SP TP LP HA-HDEL	Apicoplast + Peri-nuclear	Golgi-independent
TPx1/2$_{L}$-SOD3-HA	TPx1/2$_{L}$ SOD3$_{57-258}$ SP TP LP HA	Apicoplast + weak mitochondrion	Golgi-dependent
TPx1/2$_{L}$-SOD3-HA-HDEL	TPx1/2$_{L}$ SOD3$_{57-258}$ SP TP LP HA-HDEL	Apicoplast + Peri-nuclear	Golgi-dependent
TPx1/2$_{L}$-ACP-HA	TPx1/2$_{L}$ ACP$_{105-183}$ SP TP LP HA	Apicoplast + Peri-nuclear	Golgi-independent
TPx1/2$_{L}$-ACP-HA-HDEL	TPx1/2$_{L}$ ACP$_{105-183}$ SP TP LP HA-HDEL	Apicoplast + Peri-nuclear	Golgi-independent
ACP$_{L}$-TPx1/2-HA	ACP$_{L}$ TPx1/2$_{150-333}$ SP TP HA	Apicoplast	Golgi-independent
ACP$_{L}$-TPx1/2-HA-HDEL	ACP$_{L}$ TPx1/2$_{150-333}$ SP TP HA-HDEL	Apicoplast + Peri-nuclear	Golgi-independent
ACP$_{L}$-SOD2-HA	ACP$_{L}$ SOD2$_{92-287}$ SP TP HA	Apicoplast	Golgi-independent and Golgi-dependent (differences b/w experiments)
ACP$_{L}$-SOD2-HA-HDEL	ACP$_{L}$ SOD2$_{92-287}$ SP TP HA-HDEL	Apicoplast + Peri-nuclear	

B.