Identification of a new export signal in *Plasmodium Yoelii*: identification of a new exportome

Anthony Siau[1*], Ximei Huang[1§], Xue Yan Yam[1§], Ndeye Sakha Bob[1], Hequan Sun[2], Jagath C Rajapakse[2], Laurent Renia[3] and Peter R Preiser[1*]

[1]: Nanyang Technological University, School of Biological Sciences, 637551, Singapore

[2]: Nanyang Technological University, School of Computer Engineering, 639798, Singapore

[3]: Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Biopolis, Singapore

[§] Ximei Huang and Xue Yan Yam contributed equally to this work.

To whom the correspondence should be addressed:
*Anthony Siau (asiau@ntu.edu.sg) and Peter R Preiser (prpreiser@ntu.edu.sg)

Tel: (65) 6316-2822/6316-2869 GMT+8h | Fax: (65) 6791-9516

Running title

P. yoelii exportome exposed by a new export signal

Summary

Development of the erythrocytic malaria parasite requires targeting of parasite proteins into multiple compartments located within and beyond the parasite confine. Beyond the PEXEL/VTS pathway and its characterized players, increasing amount of evidence has highlighted the existence of proteins exported using alternative export-signal(s)/pathway(s); hence, the exportomes currently predicted are incomplete. The nature of these exported proteins which could have a prominent role in most of the *Plasmodium* species remains elusive.

Using *P. yoelii* variant proteins, we identified a signal associated to lipophilic region that mediates export of *Plasmodium yoelii* proteins. This non-PEXEL signal termed PLASMED is defined by semi-conserved residues and possibly a secondary structure. *In vivo* characterization of exported-proteins indicated that PLASMED is a *bona fide* export-signal that allowed us to identify an unseen *P. yoelii* exportome.

The repertoire of the newly predicted exported proteins opens up perspectives for unraveling the remodeling of the host-cell by the parasite, against which new therapies could be elaborated.

Introduction

Specific targeting of proteins in eukaryotic cells is defined by unique sequence motifs. To achieve targeting to different compartments, cells use a variety of strategies. Similarly, malaria parasites need to traffic proteins to multiple compartments including the parasitophorous vacuole lumen and/or the host-cell-cytoplasm (HCC), located beyond the parasite plasma membrane or the PV membrane (PVM), respectively.

The identification of a short host-cell targeting motif named PEXEL or VTS, located downstream of a lipophilic signal peptide (SP), allowed the prediction of a repertoire of ~300 exported proteins in *P. falciparum* (Marti et al, 2004; Sargeant et al, 2006; van Ooij et al, 2008). Nevertheless, it is clear that these PEXEL/VTS proteins reflect only a fraction of the parasite exportome as supported by the increasing number of PEXEL-Negative-Exported-Proteins (PNEP) reported in *P. falciparum* (Heiber et al, 2013; Spielmann & Gilberger, 2010). This indicates that alternative export-signal(s) should exist. For the other malaria *spp.*, the repertoire of exported proteins is more elusive, as most of the known *P. falciparum* exported proteins do not have a direct homologue. Moreover, only few PEXEL proteins (Marti et al, 2004; Sargeant et al, 2006; van Ooij et al, 2008) are found in non-*falciparum spp.*, suggesting that PNEP play a more prominent role in these species (Spielmann & Gilberger, 2010). This is supported by the existence in all the non-*falciparum spp.* of PEXEL negative multigene families. These latter are localized in subtelomeric regions which are known to contain genes encoding exported proteins involved in the remodeling of the host-cell.

The largest multigene family identified so far in *Plasmodium spp.* is collectively named *pir* (*Plasmodium* interspersed repeats). The *yir* genes of *P. yoelii* constitute the largest *pir* subset with more than 830 members. Members of *pir* have also been identified in *P. vivax*, *P. berghei*, *P. chabaudi*, *P. knowlesi* and *P. cynomolgi* (Carlton et al, 2005; Carlton et al, 2002;

Janssen et al, 2004; Jemmely et al, 2010; Lawton et al, 2012; Neafsey et al, 2012). Most PIR proteins have a predicted transmembrane (TM) domain while some also contain an additional predicted SP. Except for the VIR-D subset (Sargeant et al, 2006), PIR proteins lack an obvious PEXEL/VTS motif, despite being exported to the host-cell-cytoplasm (Di Girolamo et al, 2008; Pasini et al, 2012; Sijwali & Rosenthal, 2010) and expressed on the infected red-blood-cell (iRBC) surface (Cunningham et al, 2005; del Portillo et al, 2001; Janssen et al, 2004).

After *pir*, one of the largest multigene families found in rodent malaria parasites, with ~220 members in *P. yoelii,* is *pyst.* The PYST-A subfamily contains a predicted SP only, PYST-B contains SP and PEXEL/VTS motifs, while PYST-C and PYST-D contain both predicted SP and TM domains (Carlton et al, 2002). Although only PYST-B contains the predicted PEXEL/VTS motif, PYST-A of *P. berghei* can be exported to the HCC, and one PYST-A member is localized at the iRBC membrane (Franke-Fayard et al, 2010; Pasini et al, 2012).

Here, we investigated the export sequence of these variant PNEP *in P. yoelii*. This allowed us to identify a new PNEP export signal and present a significantly expanded exportome.

Results

YIR and PYST are exported in the host-cell-cytoplasm.

The subcellular localization of a YIR and a PYST protein without predicted PEXEL/VTS motifs was characterized in *P. yoelii* by GFP-tagging. The YIR protein PY01872 (282 residues) have a canonical PIR pattern with an annotated TM domain at the C-terminal end while the PYST protein PY04313 (289 residues) is predicted to contain a SP but no TM domain. To ensure that the GFP-tag did not affect the localization of the different proteins, GFP was inserted either internally (yir-iGFP or pyst-iGFP) or at the C-terminal end of the PIR and PYST proteins (yir-CtGFP or pyst-CtGFP).

Localization of YIR and PYST full-length chimeras was ascertained using live-cell fluorescence microscopy performed on mixed-stage blood samples. For this, we took advantage of the observation that DAPI staining coupled to a longpass UV filter allowed us to discern the internal structures and membranes of live parasites, including the PVM (Supplementary Figure S1A). The export patterns in transfected parasites were evaluated by classifying the export efficiency into five representative groups named "efficient, positive, moderate, low and no" export patterns with ≥100%, ~75%, ~50%, ≤25% and 0% of the GFP signal in the HCC compared to that measured within the parasite (Figure 1 III).

Both GFP-tagged YIR and PYST were localized in the parasite cytoplasm as well as in the HCC, independent of the position of the GFP-tag (figure 1). The punctate green pattern observed in the HCC of YIR transfectants suggested that YIR protein could be localized in specialized structures in the HCC (Figure 1A) while the smooth green pattern observed for PYST transfectants supports that they are mainly soluble inside the HCC (Figure 1B). These localization patterns were supported by solubility assays data which indicated that YIR and PYST chimeras are mostly found associated with integral membrane and soluble protein fractions, respectively (Supplementary Figure S1B). Finally, a smooth staining surrounding

the periphery of the parasite was observed in a subset of iRBC, suggesting that the YIR/PYST fused proteins also accumulated in the parasite PV (white arrow, Figure 1).

Analysis of the export efficiency revealed a decrease of the export efficiency when the GFP-tag was fused downstream of the plasmodial protein, suggesting that GFP had an adverse effect on chimera export (Figure 1, PY01872-CtGFP and PY04313-CtGFP). To confirm the ultimate localization of these proteins, we generated transgenic parasite lines expressing PY01872 or PY04313 fused with a myc or a HA tag placed at the C-termini, respectively (Supplementary Figure S1C). Immunolabeling using specific anti-tag and anti-PyEXP1 antibodies showed PY01872 and PY04313 being exported to the HCC, in line with what was observed previously in live parasite expressing the cognate proteins tagged with an internal GFP.

Validating these data, transfected parasites expressing GFP chimeras including the YIR proteins PY03906 (1 TM domain) and PY03632 (1 SP and 1 TM domain) or the PYST protein PY01395 (1 SP) showed localization patterns similar to those observed for YIR and PYST transfectants, respectively (Supplementary Figure S1D). Altogether, this data provides the identification of two groups of *P. yoelii* proteins exported beyond the parasite confine.

The C-terminal part of YIR, including a TM domain, is required for YIR export.

We next determined the export requirements of YIR protein through the expression of chimeras containing 1) the N-terminal region only or 2) the C-terminal region only including a predicted TM domain.

When the N-terminal region of the YIR protein PY1872 (residue 1 to 217) was fused upstream of GFP, no export was observed (Figure 2A). This suggests that insufficient or no export cues are present in these N-terminal domains. In contrast, chimera including a GFP-tag fused upstream to the C-terminal part of PY01872 (residue 218 to 282) and representing ~25% of

the total protein was efficiently exported to the HCC (Figure 2B). Additional shortening of the YIR C-terminal sequence resulted in no export to the HCC (Supplementary Figure S2A). Finally, solubility assays confirmed the lipophilic nature of the C-terminal domain of YIR and also suggested a possible solubility shift resulting of the export of the chimera as previously observed for *P. falciparum* PNEP (Gruring et al, 2012). (Supplementary Figure S2B). This data along with those obtained with chimeras including N- and C-terminal part the YIR PY03906 and PY03632 (Supplementary Figure S2C), indicates that the YIR export information is located in the C-terminal region and includes a TM domain.

The N-terminal sequence of PYST, including a SP, is responsible for export

Unlike YIR, the analyzed PYST have a SP but no TM domain, suggesting that PYST trafficking signals might be located in a different part of the protein. To map these regions, parasites expressing truncated parts of the PYST protein PY04313 fused to GFP were generated. Nested C-terminal deletion constructs generated for PY04313 showed that export was still maintained in proteins expressing the first 44 residues (Figure 2C). Additional deletions of the PYST N-terminal region or the SP abolished chimera export (Figure 2D). These results are in line with those indicating that the export sequence of PY01395 is localized in its first 50 residues (Supplementary Figure S2D) and support that the PYST trafficking information is located within its N-terminal part, which includes a SP.

Identification of a PNEP export-signal common to both YIR and PYST

The minimal export sequence of YIR and PYST were analyzed. For this, we compared the sequence of the C-terminal sequence of the YIR PY03906, PY01872 and PY03906 and the N-terminal sequence of the PYST PY04313 and PY01395. Although there was no overall conservation, we identified a semi-conserved sequence motif ranging from 21 (PY01395) to

26 (PY03632) residues in all these proteins (Figure 3A). The first residue of this motif is a serine (S) and the last residue is a lysine (K). Near the middle of the motif, two semi-conserved residues including a phenylalanine (F) and an alanine (A) are found. In addition, valine (V) and serine (S) residues can be found with a high prevalence in the first and second half of the motif, respectively. In addition, the semi-conserved motif is associated with a lipophilic region: the semi-conserved motif of YIR (YIR motif) is located mostly within the TM domain with a few residues extending into the N-terminal flanking region, whereas the semi-conserved motif of PYST (PYST motif) is located within the second half of the SP, ending a few residues downstream of the SP (Figure 3B).

In parallel, secondary structure analysis of the YIR/PYST minimal export region using Jpred (Cole et al, 2008) showed that a significant part of the YIR motif, mainly localized within the TM domain, is predicted to form an α-helix (Figure 3B). Similarly, 15 residues spanning the N-terminal part of the SP and the PYST motif are also predicted to form an α-helix, with the last six amino acids of the α-helix corresponding to the first six residues of the PYST motif (Figure 3B).

To validate the features identified here, we first assessed whether the PYST motif of 22 residues localized between position 14 and 35 of PY04313 (Figure 3B lower) could confer export on an unrelated peptide of 38 residues derived from the commercial plasmid pDisplay and named pD (Figure 3C). As expected, this peptide which contains a Mouse Ig K-chain SP and an HA tag was unable to enable GFP export ($pD_{1\text{-}38}$-GFP) in *P. yoelii*. In contrast, the replacement of pD residues 14 to 35 by the PYST motif resulted in efficient GFP export ($pD_{PYST\ domain}$-GFP). This indicates that the semi-conserved motif identified here is sufficient for protein export.

To characterize the residues required for trafficking, an increasing number of residues found in the PYST motif were introduced into the $pD_{1\text{-}38}$-GFP chimera until an efficient export was

obtained. All the sequences of the chimeras generated were analyzed using SignalP (Petersen et al, 2011) to ensure that the mutations introduced in the pD sequence do not interfere with SP processing and function (Figure 4).

Introduction of nine PYST residues (Ser in position 1, Asp in position 8, Val in position 9, Ala in position 10 and 12, Phe in position 11, Ser in position 13, Glu in position 14 and Lys in position 22) led to a low protein export (Figure 4A, pD.A). This indicates that these residues are sufficient to trigger a basal protein export. Interestingly, the export efficiency was increased to a level similar to the wild-type domain after inclusion of Gly and Tyr in position 4 and 5, respectively (Figure 4A, pD.B). This increased export efficiency could be partly due to a change in secondary structure rather than the requirement for specific residues. Indeed, the insertion of Gly and Tyr leads to an extension of the predicted α-helix formed by the SP and the PYST sequence, from 10-11 residues (pD.A) to twelve residues (pD.B), with the last three residues (Ser-Trp-Val) forming the α-helix reaching into the motif (Figure 4A). To test this, unrelated residues were included in position 4 and 5 of the weakly exported pD.A chimera. All the residue pairs predicted to increase the length of the α-helix to 12-15 residues were associated with efficient GFP export (Figure 4B, pD.C and Supplementary Figure S3 for additional constructs) whereas those expected to not increase the α-helix beyond 11-residue-long resulted in low protein export (Figure 4B, pD.D). These results suggest that an α-helix with a minimal length of twelve residues is required to obtain efficient chimera export and thus, the export-signal for YIR/PYST includes a secondary structure element.

To further define the residues associated with export, a mutational scan was performed across the residues introduced in the efficiently exported pD.B chimera by replacing these with the cognate residues found in the original pD sequence. Changes of six residues out of eleven led to a decrease in protein export with chimeras accumulated in the parasite perinuclear region, indicative of the ER, and/or in the PV (Figure 4C and Supplementary Figure S4 for detailed

pictures). These include the Ser in position 1 (pD.E); the Glu added in position 4 to increase the length of the α-helix (pD.F); the Val in position 9 (pD.I); the Phe in position 11 (pD.K); the Ser in position 13 (pD.M) and the Lys in position 22 (pD.O). Surprisingly, mutations of each of the two Ala at the midpoint of the motif (position 10 and 12) did not alter the chimera export efficiency (pD.J and pD.L). This is possibly due to the presence of a second Ala in close proximity, as a mutation of both Ala led to low export (pD.P).

Altogether, the mutation scan highlighted that in addition to a lipophilic domain and a possible conformational requirement, six specific residues are associated with export. These include four semi conserved residues (Ser located at the beginning, Ala and Phe located at the midpoint and Lys located at the end) and two residues (Val and Ser) which can hold variable positions within the defined export region.

Importantly, introduction of these specific residues along with Glu placed to increase the length of the α-helix into the unrelated pD $_{1\text{-}38}$–GFP chimera, resulted in moderate GFP export (Figure 4D, pD$_{PLASMED}$-GFP) and thus validated these findings. Finally, to establish whether these residues could also confer export when associated with a TM domain, two unrelated TM domains derived from mouse TMIE and TMED10 protein were fused downstream to a GFP (Figure 4E). As expected, these chimeras were not exported (GFP-TMIE, GFP-TMED10) despite the presence of residues associated with export in the TM domain region. In contrast, the complementation of these partial signals in the TMIE and TMED10 peptides enabled a moderate GFP export (GFP-TMIE mutated, GFP-TMED10 mutated). These provide strong evidence that the signal identified here is sufficient to target proteins for export into the HCC, when associated with any SP or TM domain and thus, it could be widely use by *Plasmodium*. We named this export-signal Plasmodium Lipophilic And Secondary structure Mediated Export Domain (PLASMED) (Figure 4F).

***P. yoelii* PLASMED exportome**

A reiterative process of interrogating the *P. yoelii* protein database was performed using the PLASMED signal features identified here (see Supplementary Information). Overall, 1,277 *P. yoelii* proteins (i.e. ~17% of the proteome) containing at least one PLASMED were identified (Supplementary Table S1 PY PLASMED based).

The exportome was then refined using a machine-learning-approach that was trained with a cognate dataset including exported and non-exported proteins. To develop these datasets, the subcellular localization data of 14 *P. yoelii* PLASMED proteins were first determined (Figure 5A). To allow rapid screening, proteins with a size below ~430 residues and encoded by genes without a predicted intron were preferentially selected; without further consideration for the annotations available in the parasite databases. The localization of these proteins were determined using C-terminal GFP-tagging with the exception of the large *P. yoelii* PY04355 protein (513 residues), which was detected with a specific anti-serum. Of the 14 proteins assayed, seven GFP-tagged proteins were exported to the HCC including PY04355 which was detected beyond the PVM using specific anti-serum and anti-PyEXP1 antibody (Figure 5A and Supplementary Figure S4 for detailed pictures).

The dataset was then enriched using orthology based information. The unrefined exportome contained proteins previously characterized in the literature; either in the endogenous parasite or as orthologues in other *Plasmodium spp.* We searched the literature for PLASMED-bearing proteins for which experimental evidence pertaining to their subcellular location was available. Only proteins with clear localization data obtained by immunolabeling or protein tagging approaches using non invasive asexual blood stage parasites were retained. Reflecting the low number of exported proteins with orthologues in other *plasmodium spp.* currently known, most of the proteins were annotated as non-exported. We identified 20 exported and 44 non exported proteins including the five YIR/PYST analyzed here, two proteins that were

previously characterized in *P. yoelii* iRBC and 57 proteins with localization data available in orthologues (Supplementary Table S1 PY dataset).

It is worth noting that the utilization of a refinement set characterized by an imbalanced distribution of exported *vs.* non exported proteins may lead to suboptimal refinement. Therefore to improve the refinement outcome, the distribution of exported *vs.* non exported proteins in the datasets was adjusted by the addition of ten artificial PLASMED bearing chimeras generated throughout this study and showing moderate to efficient GFP export (Supplementary Table S1 PY dataset).

Altogether, this allowed us to generate an expanded dataset including 38 exported and 50 non-exported proteins. These were used for *in silico* refinement using a machine-learning-approach named C4.5 decision trees (Acquaah-Mensah et al, 2006; Horton & Nakai, 1997; Quinlan, 1993). This has predicted a putative PLASMED exportome of 676 *P. yoelii* proteins (Supplementary Table S1 C4.5_PY_exportome).

The ability of the prediction algorithm to identify new exported proteins was assessed. For this, the subcellular localizations of eight proteins randomly selected among those predicted to be exported were determined using C-terminal GFP-tagging (Figure 5B). For technical reasons, only small proteins encoded by genes mostly without intron were selected and the annotations available in the parasite databases was not used as a selection criteria. Six out of the eight *P. yoelii* PLASMED proteins screened were exported to the HCC, representing a precision rate of 75% (Figure 5B). These results validated the overall approach used to predict new PLASMED exported proteins (Figure 5C). The identification of new *P. yoelii* exported proteins beyond multigene family proteins, suggests that PLASMED is a *bona fide* export-signal widely used by the parasite and not solely restricted to variant proteins.

Discussion

The survival of asexual blood stages of *Plasmodium* depends crucially on the export of parasite proteins across the PVM. Here, we showed that both YIR and PYST proteins are exported to the HCC. Once exported, PYST proteins remain soluble in the host cell while YIR proteins are targeted to dot-like structures which could be integral membrane associated. The punctate green pattern observed for YIR (Figure 1) is reminiscent of the dot-like structures (Supplementary Figure S1, white arrow) displayed by DAPI-stained parasites observed under a longpass UV filter. These structures could be the HCC motile membranous structures, recently described in *P. berghei* iRBC, which have been suggested to share some functions with *P. falciparum* Maurer's clefts despite morphological differences (Ingmundson et al, 2012).

The export signal identified here defines a new type of translocation signal, not solely based on a conserved motif such as for PEXEL but including semi-conserved residues along with other features including hydrophobic a region and possibly a secondary structure requirement. Mutational scan results also suggested that variations in PLASMED resulted in multiple export efficiencies. Altogether, these distinguish PLASMED from a typical "canonical" N-terminal trafficking signal. Interestingly, the absence of complete export abolition during the mutational scan (Figure 4, pD.C to pD.P) suggests a degree of redundancy in the export signal and the possible existence of other undetected export factors. This could explain the moderate export efficiency obtained with PLASMED bearing unrelated peptides (Figure 4D), while endogenous plasmodial sequences showed efficient export (Figure 2B, GFP-CtPY01872$_{218\text{-}282}$ and 3C, pD$_{\text{PYST domain}}$-GFP). Finally, the multiple export efficiencies observed due to the variations in the PLASMED sequence might be the parasite's mechanism of regulating the amount of protein being targeted to the PV, the HCC or internally within the parasite. Hence, proteins with a degenerate PLASMED might be likely associated with HCC remodelling,

while those with an incomplete PLASMED signal might have a more important role in the parasite and / or the PV. The data generated here is now an ideal starting point to further refine the PLASMED signal and investigate the role of PLASMED proteins during erythrocyte development.

Sorting of proteins to the HCC involves two main steps: 1) targeting of proteins to the parasite PV lumen using secretory pathways and 2) translocation of proteins across the PVM by an machinery. The PLASMED signal identified here includes the information recognized by both the secretion and translocation machineries. First, the parasite sorts and targets newly synthesized proteins to the secretory pathway using a lipophilic secretory signal. In line with this, we showed that plasmodial SP (Supplementary Figure S2C, PYST SP) and TM domain (Supplementary Figure S2B, YIR TM domain) as well as mammalian SP (Figure 3C, $pD_{1\text{-}38}$-GFP) and TM domain (Supplementary Figure S4, GFP-TMIE and GFP-TMED10) are sufficient to enable the translocation of the GFP up to the PV. The significance of the N-terminal processing by plasmepsin V or signal-peptidase during protein journey remains to be determined as well as the role of the golgi. Second, the translocation of proteins beyond the PVM requires six conserved residues and potentially a proper secondary structure, all localized in the secretory signal region. Importantly, none of the mutated chimeras used here to characterize PLASMED was predicted by SignalP to interfere with SP processing and function (Figure 4). In addition, both the mutations in the PLASMED residues as well as those altering the secondary structure impaired the protein export, but retained targeting to the secretory pathway with chimeras accumulating mostly in the ER (pD.D, pD.E, pD.F, pD.I, pD.K and pD.O) and/or the PV (pD.M and pD.P) (Supplementary Figure S4 for detailed pictures of the transfectants). Altogether, these indicate that PLASMED contains export information. Finally, while TM/PLAMSED chimeras fused to GFP showed efficient export (Figure 2B), replacement of the GFP by superfolder GFP, an engineered GFP with increased

resistance to denaturation and enhanced folding property (Pedelacq et al, 2006), abolished export with chimera accumulating in the parasite periphery (Supplementary Figure S5). This supports that unfolding of the protein is required for PLASMED export as previously shown for *P. falciparum* PNEP (Gruring et al, 2012; Heiber et al, 2013).

The approach of a genome-wide screen for PLASMED proteins combined with a machine learning prediction approach enabled us to predict an expanded exportome with a precision of 75% as supported by *in vivo* characterization of newly predicted exported-proteins. This is comparable to the PEXEL/VTS based prediction algorithm for which a maximum positive prediction rate of ~70% has been reported (van Ooij et al, 2008) and confirms that PLASMED is a *bona fide* export signal in *P. yoelii* that can be used to reveal unseen PNEP. Overall, 676 *P. yoelii* proteins were predicted to be exported. Most of these proteins belong to the *P. yoelii* YIR (n=473) and PYST (n=45) families (~73% the *P. yoelii* putative PLASMED exportome). This reflects the expansion of proteins belonging to the subtelomeric multigene families and indicates their importance in remodeling of the erythrocyte. Importantly, a part of the variant proteins assayed here were not exported to the HCC. This indicates that proteins belonging to the same multigene family can have multiple localizations and functions, as previously shown for *P. falciparum* RIFIN (Petter et al, 2007) (Figure 5, YIR PY04006, PYST PY02138). While PIR are shared by all non-*falciparum* species, the expansion of PYST in rodent plasmodium species suggests that rodent malaria parasites have developed some unique remodeling features that are distinct from the other non-*falciparum* spp. Of the ~160 remaining proteins, most of them have orthologues in other *Plasmodium spp*. (Supplementary Table S1 C4.5_PY_exportome) and it is likely that these predicted proteins play an important role in enabling the parasite to modulate the host-cell. Indeed, a number of these proteins are predicted to be involved in processes such as parasite motility, transport, signaling pathways and pathogenesis, and are encoded by genes displaying an expression

pattern typical for host-cell remodeling factors with a maximum of expression found in early or late blood stage parasites (Bozdech et al, 2003; Bozdech et al, 2008). Considering that the ER-GOLGI export system in eukaryotic cells contains > 1000 proteins (Gilchrist et al, 2006), it is not surprising that the parasite may need a large number of peptides to establish a completely independent export machinery in the HCC. Finally, it is noteworthy that 21 *P. yoelii* proteins contain a SP/PEXEL motif (Sargeant et al, 2006) as well and it will be interesting to establish the reason behind these redundant export sequences.

In conclusion, we presented the proof of concept of a new type of translocation signal that enables the export of the protein when placed either at the N-terminal or internal position. Having such kind of universal export-signal represents an evolutionary advantage, as it allows the translocation of a wide repertoire of proteins including proteins without ER signal sequence to different compartments. Whilst the complexity of the signal is currently preventing us from further refining the PLASMED signal using a classical motif search approach, the features defined here are broadly used to mediate export as well as to predict a new and reliable exportome. The utilization of *P. yoelii* not only provides insights about the remodeling events initiated by this species, but importantly allows us to better understand how these events work in other *Plasmodium spp.* such as the intractable human parasite *P. vivax,* the most frequent and widely distributed cause of recurring (tertian) malaria. The molecular events which allow *P. vivax* to remodel its host-cell are currently unknown and require an animal substitute, as *P. vivax* cannot be cultured *in vitro*. However, the fact that *P. vivax* shares the PIR subtelomeric multigene families with rodent malaria parasites suggests that it may use the same export mechanism as *P. yoelii*. This can be investigated further in the future. Insights into parasite trafficking pathways as well as host-cell remodeling will provide guidance in the selection of new therapeutic targets that are able to interfere with conserved aspects essential for the development of all *Plasmodium spp.*

Materials and Methods

Plasmid construction

DNA corresponding to the full coding sequence of *pir* and *pyst* was generated by chemical gene synthesis (GenScript). The entire coding region or parts thereof were amplified by PCR. The sequence encoding a Mouse Ig K-chain SP and HA tag was amplified by PCR from a pDisplay plasmid (Invitrogen). The chimeric sequences encoding the first 13 residues of mouse Ig K-chain SP and part of *pyst*/*yir* sequence fused upstream of the *egfp*-tag (Figure 4 to 6) were obtained by PCR walking following two successive PCR amplifications of an *egfp* sequence using two different long forward primers and a reverse primer with sequence complementary to *egfp*. The first PCR allowed the synthesis of PCR products containing ~ the *pyst*/*yir* sequence fused upstream of the *egfp*. A second PCR using long forward primers with sequence complementary to the *pyst*/*yir* sequence previously added was then performed to insert the Mouse Ig K-chain SP sequence. Finally, all the non-*pir*/*pyst* sequences were obtained by PCR amplification of genomic DNA/cDNA or chemical synthesis. All these sequence were then cloned into the plasmid ePL (containing both the *P. berghei* EF1 constitutive promoter, the 3' UTR region of *P. berghei* DHFR/TS gene and the selectable marker *T. gondii* DHFR gene), either upstream or downstream of an *egfp*-tag.

P. yoelii parasite preparation and transfection

BALB/c mice were infected with *Plasmodium yoelii* 17X 1.1 parasites by intraperitoneal injections. Transfections were carried out as previously described (Janse et al, 2006).

Live cell microscopy

Infected blood samples obtained after *in vivo* drug selection, were stained with DAPI (1µg/ml) for 5 min. Transfected parasites expressing GFP-tagged chimeras were then observed with an Olympus IX71 fluorescent microscope using a 100X oil immersion

objective. DAPI was detected using a Chromas 11000v3 filter set whereas eGFP was detected using a Chroma 49011 Filter Set. Pictures were captured using an Olympus DP30BW camera and processed using ImageJ 1.42a. Throughout the study, representative examples of at least 20 independently imaged trophozoite/early schizont stages parasites were collected and analyzed. The chimera export phenotype determined after analysis of these 20 parasites is shown either with the symbol "yes" or "no" standing for positive or negative export phenotype, respectively. When not all the transfectants showed a similar export phenotype, the fraction of transfectants showing proper protein export is indicated on the right of the export phenotype symbol. For each construct assayed in the Figures 2, 3, and 4, the merged picture of a representative parasite is shown. The corresponding extent of export is quantified by measuring the intensities of pixels corresponding to the green fluorescence signal along the yellow line from A to B in the picture depicted using ImageJ and represented using a profile plot using a scale ranging from 0 to 80 pixel intensities.

Prediction of the *P. yoelii* PLAMED exportome.

P. yoelii PLASMED exportomes were first generated by a reiterative process of interrogating parasite protein database and were then refined using machine-learning-approaches. See extended supplementary method for additional details.

Characterization of protein localization using specific antisera

Parasite expressing PY01872 and PY04313 fused to a HA-tag were air-dried, fixed with methanol and stained with rabbit anti-tag (Abcam) in conjunction mouse anti-PyEXP1 antibody (gift from Prof. D. Mazier). Anti-rabbit secondary antibody coupled to Alexa Fluor 594 (Invitrogen) and with anti-mouse secondary antibody coupled to Alexa Fluor 488 (Invitrogen) were then used to reveal the localization of the HA-tagged proteins as well as the PVM, respectively. To localize PY04355, DNA corresponding to coding sequences of PY04355 were amplified from *P. yoelii* genomic DNA/cDNA and inserted into the His-tag

expression vector pET-24. Recombinant proteins were then expressed in BL21 *E. coli* and purified. Proteins were injected intra-peritoneally into BALB/c mice first with complete and then with incomplete Freund's adjuvant at two-week intervals. The mice were bled one week after the fifth immunization. Specificity of the antisera was confirmed by western blot using whole parasite extract prepared from a mixed stages parasite sample (Supplementary Figure S6). To localize these proteins, air-dried *P. yoelii* iRBC fixed with methanol were first stained with specific anti-serum (dilution 1/200) in conjunction with anti-mouse secondary antibody coupled to Alexa Fluor 594. The Parasite PVM was then revealed using mouse anti-PyEXP1 antibody in conjunction with anti-mouse secondary antibody coupled to Alexa Fluor 488. Non-immune mouse serum was used as control. Slides were then examined by fluorescence microscopy after a brief incubation with DAPI 1mg/ml.

Acknowledgments

The authors are grateful to Yen Hoon Luah, Ramya Ramadoss and Chee Sheng Ng for their technical help and Dr. Till Voss and Prof. Mark Featherstone for the critical reading of the manuscript.

Conflict of interest:

The authors declare that they have no conflict of interest.

References

Acquaah-Mensah GK, Leach SM, Guda C (2006) Predicting the subcellular localization of human proteins using machine learning and exploratory data analysis. *Genomics Proteomics Bioinformatics* **4**(2)**:** 120-133

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol* **1**(1)**:** E5

Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, Ginsburg H, Nosten F, Day NP, White NJ, Carlton JM, Preiser PR (2008) The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc Natl Acad Sci U S A* **105**(42)**:** 16290-16295

Carlton J, Silva J, Hall N (2005) The genome of model malaria parasites, and comparative genomics. *Curr Issues Mol Biol* **7**(1)**:** 23-37

Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature* **419**(6906)**:** 512-519

Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**(Web Server issue)**:** W197-201

Cunningham DA, Jarra W, Koernig S, Fonager J, Fernandez-Reyes D, Blythe JE, Waller C, Preiser PR, Langhorne J (2005) Host immunity modulates transcriptional changes in a multigene family (yir) of rodent malaria. *Mol Microbiol* **58**(3)**:** 636-647

del Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP, Schneider NK, Villalobos JM, Rajandream MA, Harris D, Pereira da Silva LH, Barrell B, Lanzer M (2001) A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. *Nature* **410**(6830)**:** 839-842

Di Girolamo F, Raggi C, Birago C, Pizzi E, Lalle M, Picci L, Pace T, Bachi A, de Jong J, Janse CJ, Waters AP, Sargiacomo M, Ponzi M (2008) Plasmodium lipid rafts contain proteins implicated in vesicular trafficking and signalling as well as members of the PIR superfamily, potentially implicated in host immune system interactions. *Proteomics* **8**(12)**:** 2500-2513

Franke-Fayard B, Fonager J, Braks A, Khan SM, Janse CJ (2010) Sequestration and tissue accumulation of human malaria parasites: can we learn anything from rodent models of malaria? *PLoS Pathog* **6**(9)**:** e1001032

Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, Nagaya H, Roy L, Gosline SJ, Hallett M, Paiement J, Kearney RE, Nilsson T, Bergeron JJ (2006) Quantitative proteomics analysis of the secretory pathway. *Cell* **127**(6)**:** 1265-1281

Gruring C, Heiber A, Kruse F, Flemming S, Franci G, Colombo SF, Fasana E, Schoeler H, Borgese N, Stunnenberg HG, Przyborski JM, Gilberger TW, Spielmann T (2012) Uncovering common principles in protein export of malaria parasites. *Cell Host Microbe* **12**(5)**:** 717-729

Heiber A, Kruse F, Pick C, Gruring C, Flemming S, Oberli A, Schoeler H, Retzlaff S, Mesen-Ramirez P, Hiss JA, Kadekoppala M, Hecht L, Holder AA, Gilberger TW, Spielmann T (2013) Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in Plasmodium falciparum protein export. *PLoS Pathog* **9**(8)**:** e1003546

Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proc Int Conf Intell Syst Mol Biol* **5:** 147-152

Ingmundson A, Nahar C, Brinkmann V, Lehmann MJ, Matuschewski K (2012) The exported Plasmodium berghei protein IBIS1 delineates membranous structures in infected red blood cells. *Mol Microbiol*

Janse CJ, Ramesar J, Waters AP (2006) High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite Plasmodium berghei. *Nat Protoc* **1**(1)**:** 346-356

Janssen CS, Phillips RS, Turner CM, Barrett MP (2004) Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res* **32**(19)**:** 5712-5720

Jemmely NY, Niang M, Preiser PR (2010) Small variant surface antigens and Plasmodium evasion of immunity. *Future Microbiol* **5**(4)**:** 663-682

Lawton J, Brugat T, Yan YX, Reid AJ, Bohme U, Otto TD, Pain A, Jackson A, Berriman M, Cunningham D, Preiser P, Langhorne J (2012) Characterization and gene expression analysis of the cir multi-gene family of Plasmodium chabaudi chabaudi (AS). *BMC Genomics* **13:** 125

Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**(5703)**:** 1930-1933

Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q, Chapman SB, Dash AP, Anvikar AR, Sutton PL, Birren BW, Escalante AA, Barnwell JW, Carlton JM (2012) The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nat Genet* **44**(9)**:** 1046-1050

Pasini EM, Braks JA, Fonager J, Klop O, Aime E, Spaccapelo R, Otto TD, Berriman M, Hiss JA, Thomas AW, Mann M, Janse CJ, Kocken CH, Franke-Fayard B (2012) Proteomic and genetic analyses demonstrate that Plasmodium berghei blood stages export a large and diverse repertoire of proteins. *Mol Cell Proteomics*

Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**(1)**:** 79-88

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**(10)**:** 785-786

Petter M, Haeggstrom M, Khattab A, Fernandez V, Klinkert MQ, Wahlgren M (2007) Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol* **156**(1)**:** 51-61

Quinlan JR (1993) C4.5: Programs for Machine Learning. *Morgan Kaufmann; San Mateo, USA*

Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP, Cowman AF (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* **7**(2)**:** R12

Sijwali PS, Rosenthal PJ (2010) Functional evaluation of Plasmodium export signals in Plasmodium berghei suggests multiple modes of protein export. *PLoS One* **5**(4)**:** e10227

Spielmann T, Gilberger TW (2010) Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? *Trends Parasitol* **26**(1)**:** 6-10

van Ooij C, Tamez P, Bhattacharjee S, Hiller NL, Harrison T, Liolios K, Kooij T, Ramesar J, Balu B, Adams J, Waters AP, Janse CJ, Haldar K (2008) The malaria secretome: from algorithms to essential function in blood stage infection. *PLoS Pathog* **4**(6)**:** e1000084

Figures legend

Figure 1. YIR and PYST are exported to the red-blood-cell cytoplasm.

Representative images of live *P. yoelii* infected red-blood-cell expressing full-length YIR (A) or PYST (B) as determined by fluorescence microscopy of GFP-tagged proteins. The white arrows indicate the position of the parasite PV. The schematic diagram of the construct transfected is shown on the left of the images. The ability of each chimera to be exported (yes/no) is shown to the right of the images. The extent of export was quantified based on the pixel intensity of the GFP signal along the yellow line in the representative picture. Results are shown in the accompanying fluorescence plots (See Material and Methods). The horizontal line indicates the % GFP intensity threshold of 0, ≤25, 75 and ≥100% corresponding to "no export", "low", "positive" and "efficient" export pattern, respectively. The position of the parasite and the HCC are noted using black and orange double arrows, respectively.

Figure 2. Defining essential regions of YIR and PYST required for export to the HCC.

Localization of GFP chimeras fused to N-terminal part of YIR (A), C-terminal part of YIR (B) or N-terminal parts of PYST (C and D) in live *P. yoelii* infected red-blood-cell. Left panel shows the schematic diagram of the constructs used. Right panels show the corresponding representative merged photographs of infected erythrocytes, the ability of each chimera to be exported (yes/no) and the corresponding export ability as analyzed for Figure 1. The representative GFP, DAPI and bright field images for all constructs are shown in Supplementary Figure S4.

Figure 3. Identification of a related export motif in YIR and PYST . A. Alignment of the amino acid sequences of the minimal region mediating YIR and PYST export. Residues are colored according to their physicochemical properties. The positions of semi-conserved residues (Ser (S), Phe (F), Ala (A), Lys (K)) are indicated by vertical arrows. The position of the residues commonly found in all the sequence with variable positions in the first (Val (V)) or the second half (Ser (S)) of the export domain are indicated by *. The sequence shows the consensus found in YIR and PYST sequences; only the YIR protein PY03632 has residues at all 26 positions displayed. The height of each letter is proportional to the frequency of amino acid identity at each position.

B. Analysis of the secondary structure of the minimal region of YIR/PYST proteins required for export to the HCC as defined by Jpred. The regions predicted as part of the PNEP export domain were highlighted in blue (YIR) or red (PYST) and the position of the predicted TM domain (YIR) or the SP (PYST) were noted using double arrows, respectively. The letters "H" and "E" are aligned beneath residues predicted to form an α-helix or β-strand, respectively.

C. Chimeras containing parts of pDisplay coupled to PYST sequences and fused to a GFP-tag were tested for their ability to be exported in transfected *P. yoelii*-iRBC. I. N-terminal pDisplay sequence in relation to the GFP-tag. Residues from pDisplay are colored in black ($pD_{1\text{-}38}$–GFP), whereas substitutions in the pDisplay sequence derived from the PYST export domain are colored in red. Secondary structure predictions are shown below as for Figure 3B. The positions of the Mouse Ig K-chain SP and the HA tag are indicated above the sequence of $pD_{1\text{-}38}$–GFP. II. Merged pictures representative of the transfected parasites, export phenotype and the corresponding export ability analyzed as for Figure 1. The representative GFP, DAPI and bright field images for all constructs are shown in Supplementary Figure S4.

Figure 4. Determination of the minimal sequence and secondary structure required for export. I. Mutated pDisplay sequences fused to a C-terminal GFP-tag were tested for their ability to be exported. The positions from 1 to 22 of the residues used for mutational analysis are indicated above each set of sequences in A through D. The presence and location of SP cleavage sites in each chimera were investigated by SignalP and indicated by an arrow above the sequence. The corresponding secondary structure predictions are shown below the sequence as for Figure 3. II. Qualitative assessment (yes/no) and extent of export for each construct. III. Merged pictures representative of the export pattern observed. The extent of export was accessed as for Figure 1. The white arrows show the accumulation of GFP chimeras in the parasite perinuclear region and/or in the PV. The representative GFP, DAPI and bright field images for all constructs are shown in Supplementary Figure S4.

A. Increasing substitution of residues from the PYST domain into $pD_{1\text{-}38}$-GFP lead to chimera export. The residues sequentially introduced are underlined in red. B. The role of the secondary structure in protein export. Residues added to alter the length of the α-helix are colored and underlined in blue. C. Determination of the minimal export consensus by mutational scan. The mutations introduced in pD.C are colored in green and labeled by an asterisk above the residue. D. Validation of the PLASMED signal. The mutations introduced in $pD_{1\text{-}38}$-GFP are colored in green and labeled by an asterisk above the residue. The two and four residues associated with PLASMED export originally present in the TM domain region of TMIE and TMED10 peptide are colored in blue whereas the mutations introduced in these sequences to generate a complete PLASMED signal are colored in green and labeled by an asterisk above the residue. The position of the predicted TM domain was noted using double arrows. E. Summary of the PYST and YIR export-signal that we named PLASMED, with the conserved residues having a semi-conserved position colored in red, and those with variable position in blue.

Figure 5. Prediction of *P. yoelii* PLASMED exportome. A Creation of the *P. yoelii* dataset. Summary of the export pattern observed in *P. yoelii* infected red-blood-cell expressing PLASMED-bearing proteins as determined by fluorescence microscopy of fluorescently-tagged proteins or immunofluorescently labeled protein (PY04355). The ability of each chimera to be exported (yes/no) is shown on the right in addition to the export ability. The pictures illustrating the localization pattern of the PLASMED-bearing proteins assayed are shown in Supplementary Figure S4. B. Validation of *P. yoelii* PLASMED exportome. The localization of *P. yoelii* PLASMED-bearing proteins predicted to be exported in live *P. yoelii* infected red blood cell was determined by fluorescence microscopy of GFP-tagged proteins. The ability of each chimera to be exported (yes/no) is shown on the right of the images in addition to the export ability. C. Strategy used to predict *P. yoelii* exportome.

Figure 1

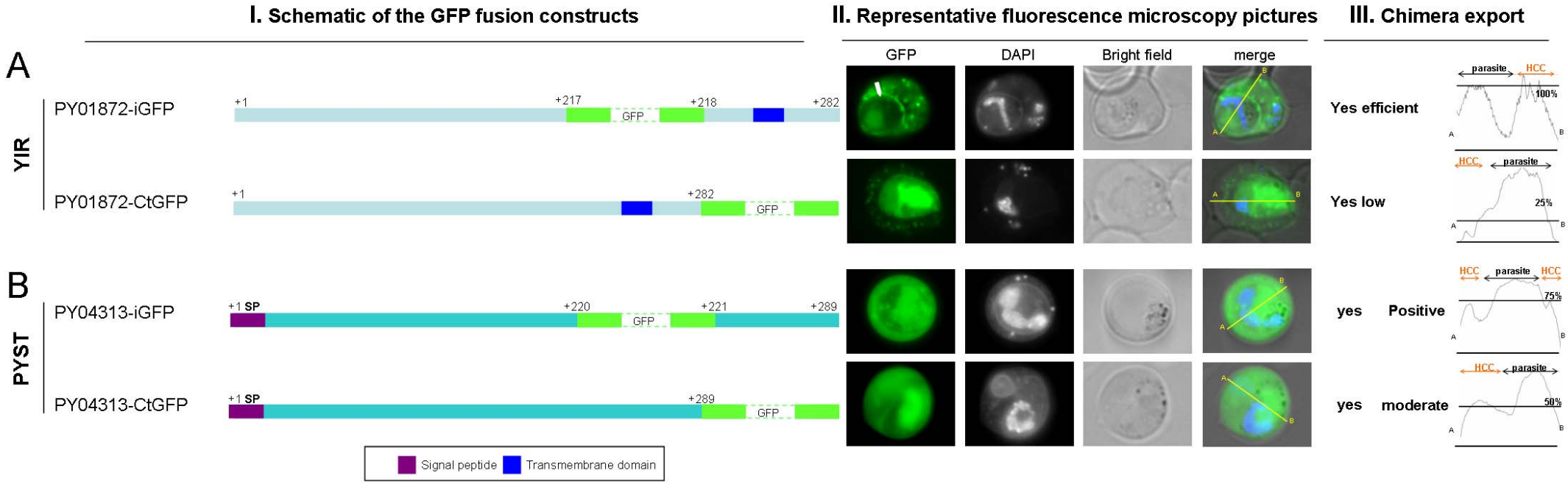

I. Schematic of the GFP fusion constructs
II. Representative fluorescence microscopy pictures
III. Chimera export
A
YIR
PY01872-iGFP
+1
+217
GFP
+218
+282
PY01872-CtGFP
+1
+282
GFP
B
PYST
PY04313-iGFP
+1 SP
+220
GFP
+221
+289
PY04313-CtGFP
+1 SP
+289
GFP
Signal peptide
Transmembrane domain
GFP
DAPI
Bright field
merge
Yes efficient
parasite
HCC
100%
Yes low
HCC
parasite
25%
yes
Positive
HCC
parasite
HCC
75%
yes
moderate
HCC
parasite
50%

Figure 2

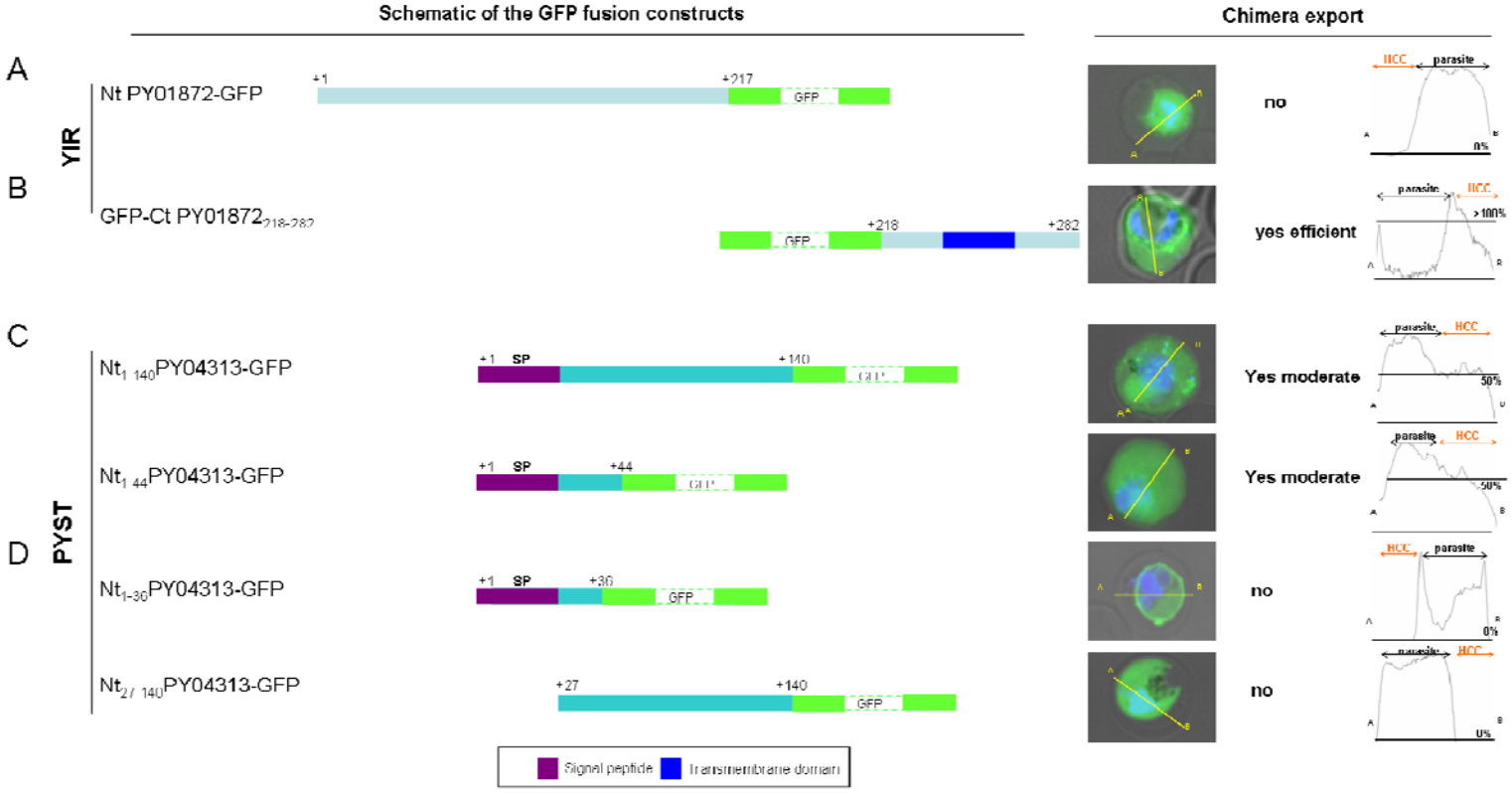

Schematic of the GFP fusion constructs
Chimera export
A
YIR
Nt PY01872-GFP
+1
+217
GFP
no
B
GFP-Ct PY01872218-282
+218
+282
GFP
yes efficient
C
PYST
Nt1 140PY04313-GFP
+1 SP
+140
Yes moderate
Nt1 44PY04313-GFP
+44
Yes moderate
D
Nt1-30PY04313-GFP
+30
no
Nt27 140PY04313-GFP
+27
no
Signal peptide
Transmembrane domain
parasite
HCC

Figure 3

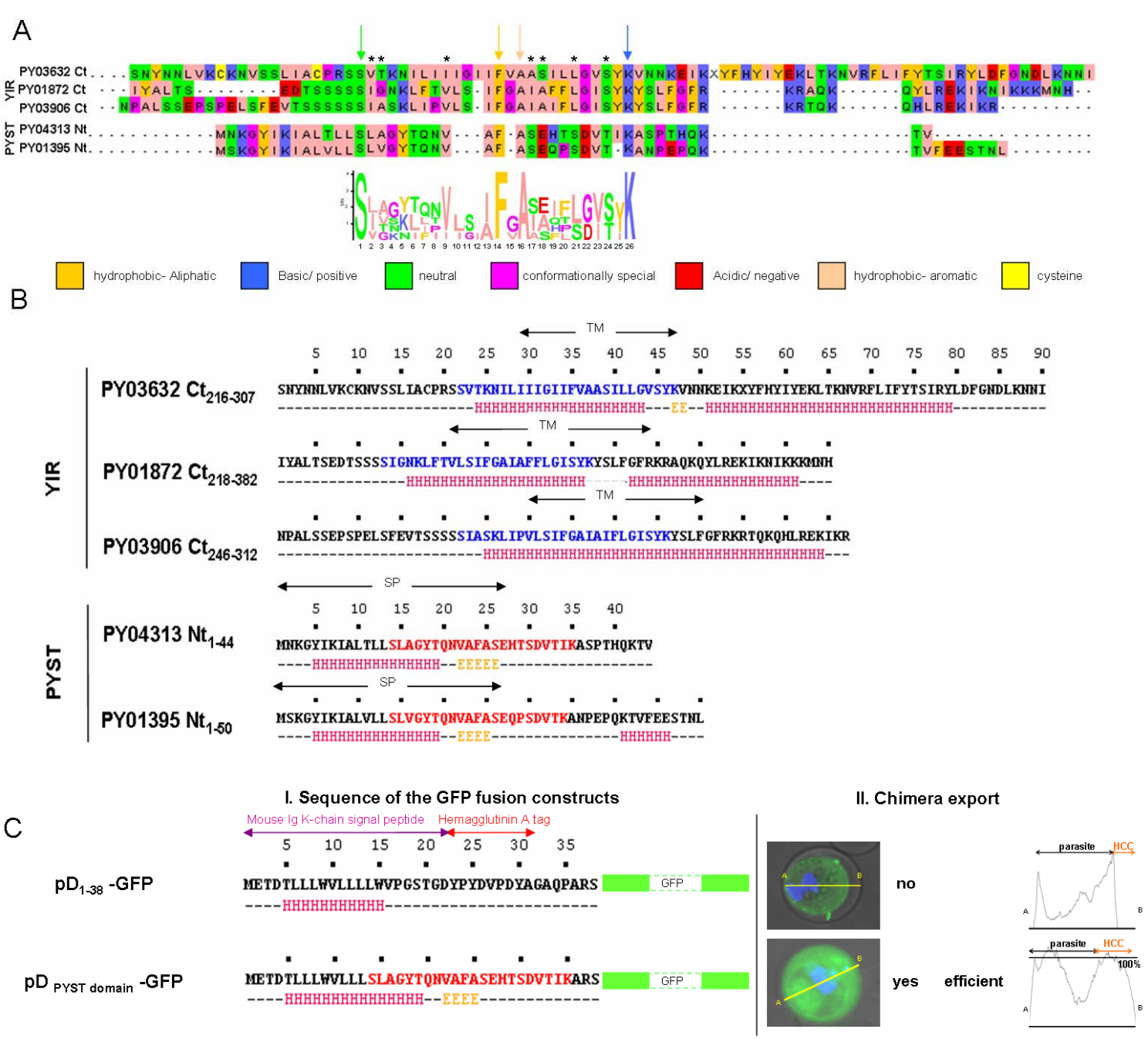

A
YIR
PYST
PY03632 Ct
PY01872 Ct
PY03906 Ct
PY04313 Nt
PY01395 Nt
hydrophobic- Aliphatic
Basic/ positive
neutral
conformationally special
Acidic/ negative
hydrophobic- aromatic
cysteine
B
TM
SP
PY03632 Ct216-307
PY01872 Ct218-382
PY03906 Ct246-312
PY04313 Nt1-44
PY01395 Nt1-50
C
I. Sequence of the GFP fusion constructs
II. Chimera export
Mouse Ig K-chain signal peptide
Hemagglutinin A tag
pD1-38 -GFP
pD PYST domain -GFP
GFP
no
yes
efficient
parasite
HCC
100%

Figure 4

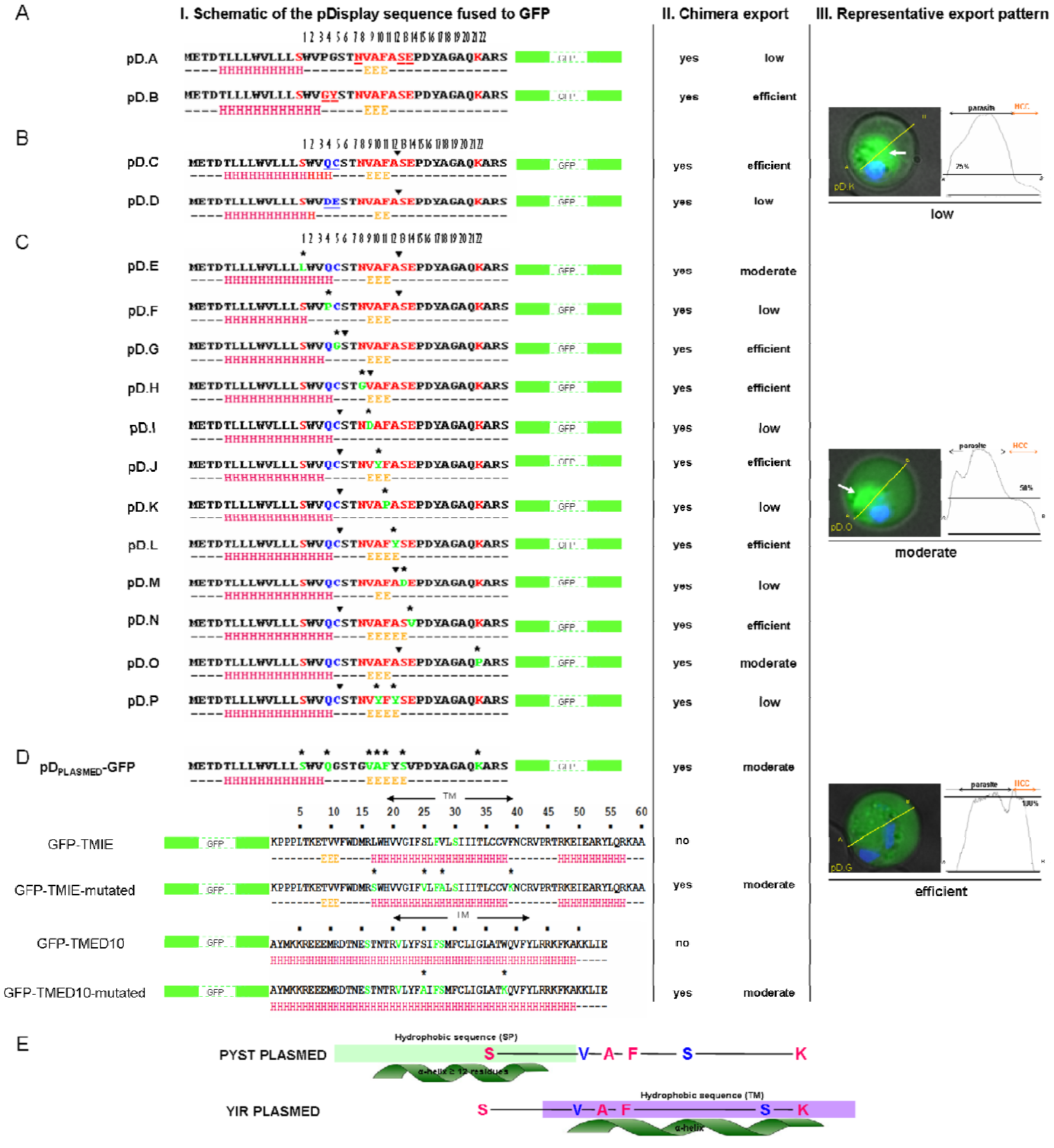
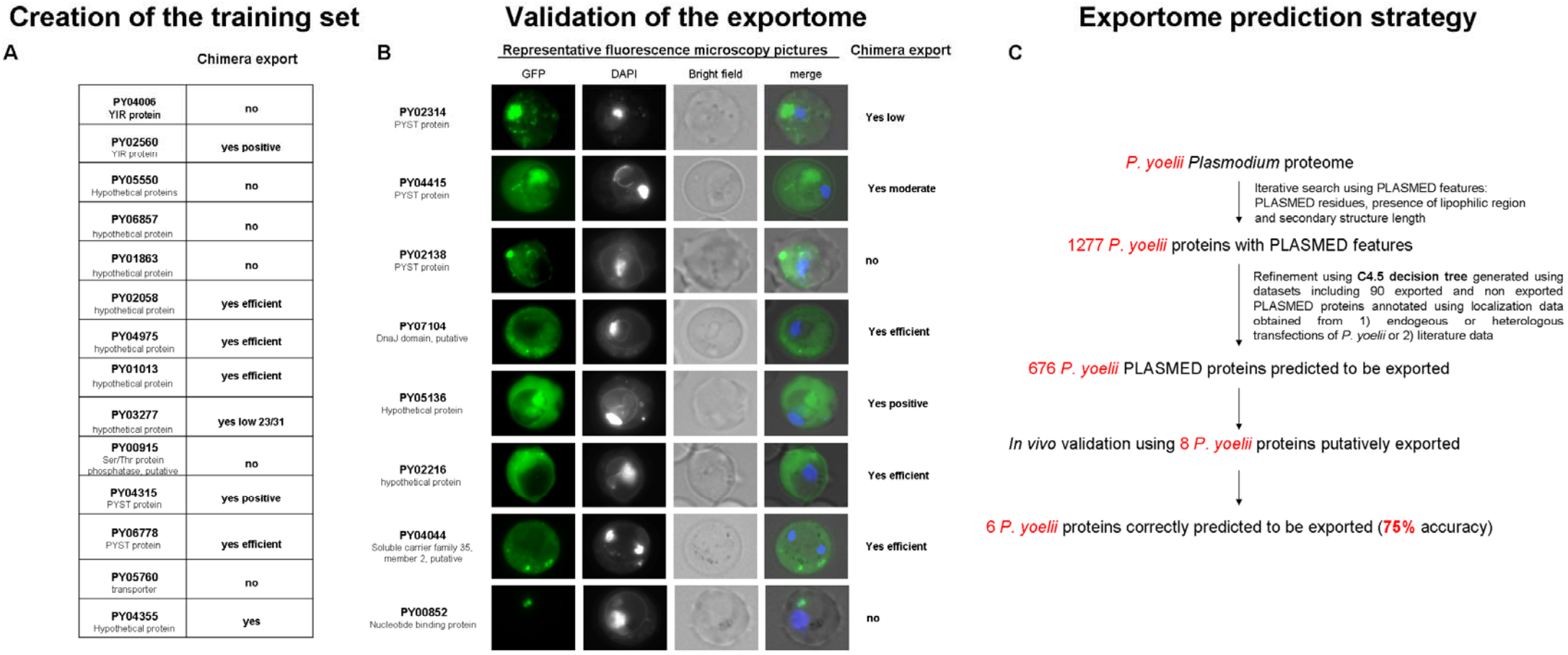

A
I. Schematic of the pDisplay sequence fused to GFP
II. Chimera export
III. Representative export pattern
pD.A
yes
low
pD.B
yes
efficient
B
pD.C
yes
efficient
pD.D
yes
low
low
C
pD.E
yes
moderate
pD.F
yes
low
pD.G
yes
efficient
pD.H
yes
efficient
pD.I
yes
low
pD.J
yes
efficient
pD.K
yes
low
pD.L
yes
efficient
pD.M
yes
low
pD.N
yes
efficient
pD.O
yes
moderate
pD.P
yes
low
moderate
D
pDPLASMED-GFP
yes
moderate
TM
GFP-TMIE
no
GFP-TMIE-mutated
yes
moderate
IM
GFP-TMED10
no
GFP-TMED10-mutated
yes
moderate
efficient
E
Hydrophobic sequence (SP)
PYST PLASMED
α-helix ≥ 12 residues
Hydrophobic sequence (TM)
YIR PLASMED
α-helix

Figure 5


Creation of the training set
A
Chimera export
PY04006 YIR protein | no
PY02560 YIR protein | yes positive
PY05550 Hypothetical proteins | no
PY06857 hypothetical protein | no
PY01863 hypothetical protein | no
PY02058 hypothetical protein | yes efficient
PY04975 hypothetical protein | yes efficient
PY01013 hypothetical protein | yes efficient
PY03277 hypothetical protein | yes low 23/31
PY00915 Ser/Thr protein phosphatase, putative | no
PY04315 PYST protein | yes positive
PY06778 PYST protein | yes efficient
PY05760 transporter | no
PY04355 Hypothetical protein | yes
Validation of the exportome
B
Representative fluorescence microscopy pictures
Chimera export
GFP
DAPI
Bright field
merge
PY02314 PYST protein | Yes low
PY04415 PYST protein | Yes moderate
PY02138 PYST protein | no
PY07104 DnaJ domain, putative | Yes efficient
PY05136 Hypothetical protein | Yes positive
PY02216 hypothetical protein | Yes efficient
PY04044 Soluble carrier family 35, member 2, putative | Yes efficient
PY00852 Nucleotide binding protein | no
Exportome prediction strategy
C
P. yoelii Plasmodium proteome
Iterative search using PLASMED features: PLASMED residues, presence of lipophilic region and secondary structure length
1277 P. yoelii proteins with PLASMED features
Refinement using C4.5 decision tree generated using datasets including 90 exported and non exported PLASMED proteins annotated using localization data obtained from 1) endogeous or heterologous transfections of P. yoelii or 2) literature data
676 P. yoelii PLASMED proteins predicted to be exported
In vivo validation using 8 P. yoelii proteins putatively exported
6 P. yoelii proteins correctly predicted to be exported (75% accuracy)