Targeting Malaria Virulence and Remodeling Proteins to the Host Erythrocyte

Matthias Marti,* Robert T. Good,* Melanie Rug, Ellen Knuepfer, Alan F. Cowman†

To establish infection in the host, malaria parasites export remodeling and virulence proteins into the erythrocyte. These proteins can traverse a series of membranes, including the parasite membrane, the parasitophorous vacuole membrane, and the erythrocyte membrane. We show that a conserved pentameric sequence plays a central role in protein export into the host cell and predict the exported proteome in *Plasmodium falciparum*. We identified 400 putative erythrocyte-targeted proteins corresponding to ~8% of all predicted genes, with 225 virulence proteins and a further 160 proteins likely to be involved in remodeling of the host erythrocyte. The conservation of this signal across *Plasmodium* species has implications for the development of new antimalarials.

Species of the genus *Plasmodium* are obligate intracellular parasites of the phylum Apicomplexa that switch between an arthropod vector and a vertebrate host, where they undergo cycles of asexual reproduction in blood cells. Each year, several hundred million people become infected with *P. falciparum*, which causes the most severe form of malaria in humans, and 1 to 2 million die from complications. Once the parasite is in the blood, its continuous asexual multiplication inside erythrocytes is responsible for clinical symptoms of malaria and the associated morbidity and mortality. Asexual stages reside in a parasitophorous vacuole, and they elaborate a membranous network in the erythrocyte cytoplasm, including Maurer's clefts. Moreover, the surface of the infected erythrocyte membrane is also remodeled with electron-dense elevations called knobs. The 85- to 110-kD knob-associated histidine-rich protein (KAHRP) is present on the cytoplasmic side of the knob structure (*1*), and the 200- to 300-kD antigen PfEMP1 is located on the exterior surface (*2*). PfEMP1 is encoded by ~60 *var* genes per parasite (*3*), and monoallelic expression of variant forms of the protein is responsible for antigenic variation (*4*). The primary structure of PfEMP1 consists of a large N-terminal ecto-

Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: cowman@wehi.edu.au

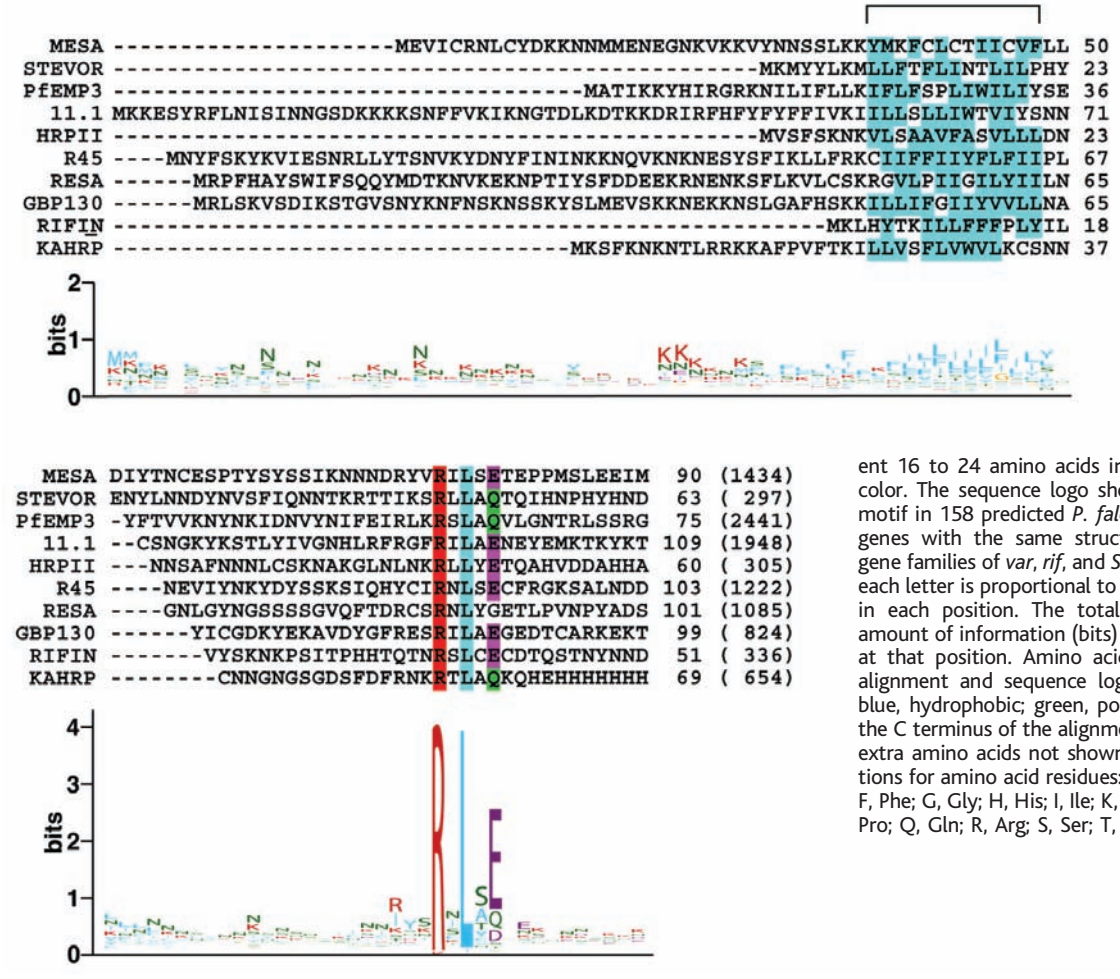

Fig. 1. Identification of a novel motif in exported *P. falciparum* proteins. Multiple sequence alignment of the N-terminal portion of 10 exported *P. falciparum* proteins shows that gene structure is conserved with a hydrophobic signal sequence (see bracket; hydrophobic amino acids are highlighted in blue) encoded in exon 1 (upper panel) and a highly variable sequence encoded in exon 2 (lower panel). However, a conserved pentameric motif is present 16 to 24 amino acids into exon 2, as highlighted in color. The sequence logo shows the conservation of this motif in 158 predicted *P. falciparum* proteins encoded by genes with the same structure, excluding the multiple gene families of *var*, *rif*, and *Stevor* (fig. S1A). The height of each letter is proportional to the frequency of amino acids in each position. The total letter height indicates the amount of information (bits) contained in the amino acids at that position. Amino acid color coding for both the alignment and sequence logo: red, basic; purple, acidic; blue, hydrophobic; green, polar amino acids. Numbers at the C terminus of the alignment (lower panel) indicate the extra amino acids not shown in the alignment. Abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

domain containing a variable number of duffy-binding-like (DBL) domains that mediate cytoadherence to various host cell receptors (*5*). The cytoadhesion of parasite-infected erythrocytes to a number of host cells is a causative factor in severe pathology of malaria, and PfEMP1 is considered the major virulence determinant of *P. falciparum*. Remodeling of the anucleated erythrocyte requires targeting of parasite proteins beyond the plasma membrane and translocation across the parasitophorous vacuole membrane. Proteins exported through this route, such as KAHRP, are trafficked through the parasite's secretory pathway by virtue of a hydrophobic signal and secreted into the parasitophorous vacuole (*6*). Further transport across the parasitophorous vacuole membrane requires additional targeting signals located within 60 amino acids downstream of the signal sequence, at least in the case of KAHRP (*6*) or HRPII (*7*). Interestingly, in most known exported proteins (such as KAHRP, GBP130, the Rifin and Stevor families, RESA, MESA, HRPII/III, R45, and protein 11.1), the signal sequence is encoded in a first exon while the remaining portion of the protein is encoded in the second exon (*8*).

To identify sequences involved in protein translocation through the parasitophorous vacuole membrane and into the erythrocyte cytoplasm, we aligned the N-terminal sequences of these proteins (Fig. 1). Although the N-terminal region of the mature proteins shows no overall conservation (Fig. 1), we identified a pentameric motif present in the N-terminal portion of all these proteins. It consists of a positively charged, hydrophilic amino acid in position 1 (Arg or Lys), a hydrophobic amino acid in position 3 (Leu or Ile), and another less conserved amino acid in position 5 (predominantly Asp, Glu, or Gln), with noncharged amino acids in positions 2 and 4 (Fig. 1). To investigate the role of this motif in protein export, we fused the N termini of two exported proteins, KAHRP and GBP130 (i.e., the portions shown in the alignment of Fig. 1), to green and yellow fluorescent reporter proteins (GFP and YFP), respectively, and generated a series of modifications (Fig. 2). The soluble fusions that included the wild-type motif were efficiently exported and evenly distributed throughout the erythrocyte cytoplasm. When the motif was either mutated or deleted, the reporters were accumulated in the parasitophorous vacuole; this result shows that the motif is required for export of soluble proteins in *P. falciparum* (*9*). We termed this motif Pexel (short for *Plasmodium* export element).

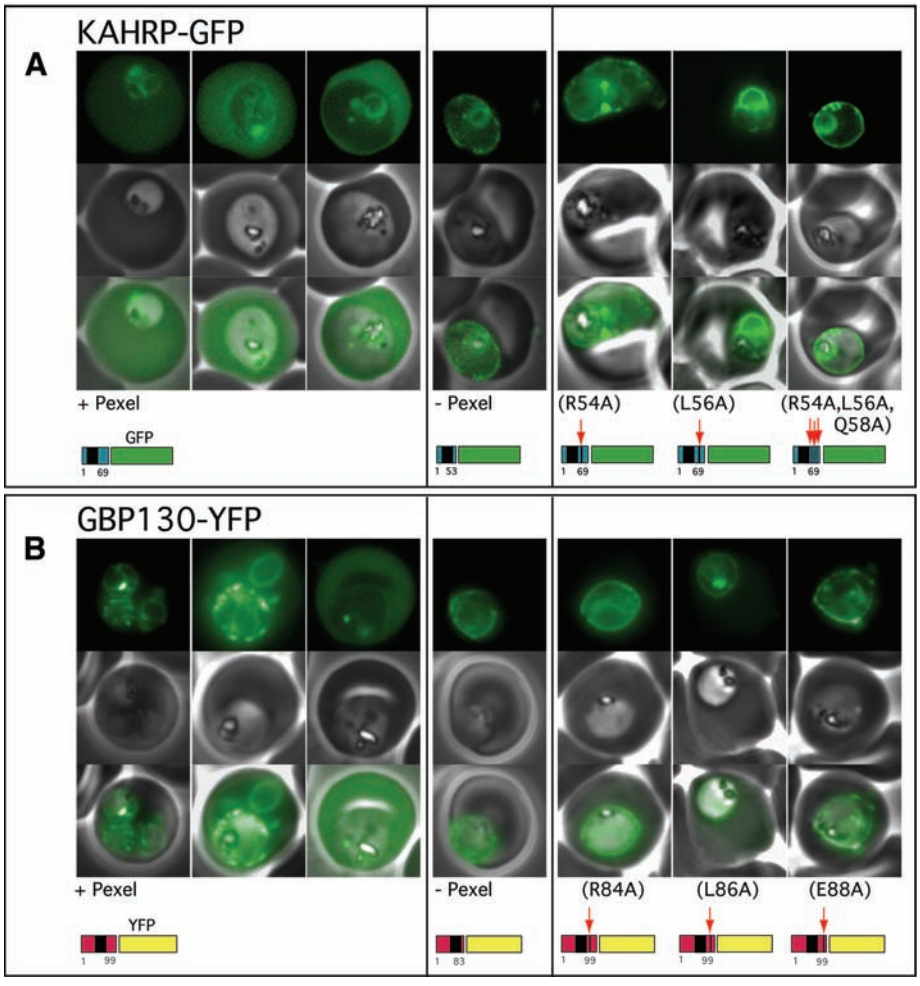

Fig. 2. The Pexel motif is necessary for export of soluble proteins in *P. falciparum*. The reporter is exported only when fused to the N terminus of KAHRP (**A**) or GBP130 (**B**) including the complete Pexel motif (left panel, +Pexel); complete truncations of the motif result in accumulation of the reporter in the parasitophorous vacuole (for construct details, see fig. S2). Representative cells expressing the exported reporter (+Pexel) are shown in ring stage (left), early (center), and late trophozoites (right). Note that KAHRP-GFP is already exported into the host cell in early stages, whereas GBP130-YFP is retained in the parasitophorous vacuole and released after development into mature trophozoite stages, perhaps because of its unusually recessed or embedded signal sequence (see also fig. S3). Coloring of the schematic representation of fusion proteins: black, hydrophobic signal sequence; blue, KAHRP N terminus; green, GFP; red, GBP130 N terminus; yellow, YFP. Arrows indicate positions of mutated amino acids, e.g., R54A for mutation of R (Arg) to A (Ala) in position 54 of KAHRP.

To determine whether the Pexel motif was also required for transport of membrane-bound surface proteins out of the parasitophorous vacuole, we replaced the ectodomain of a Rifin variant, a member of a superfamily (~160 genes in the *P. falciparum* genome), with YFP (Fig. 3A). As expected, the reporter was transported into the host cell only when the Pexel motif was present, whereas truncations of the motif completely blocked export. Mutation of the Pexel motif also blocked export; hence, it is required for transport of the Rifin proteins across the parasitophorous vacuole membrane and into the erythrocyte.

These data show that the Pexel motif is required for protein export in *P. falciparum* for both soluble and membrane proteins such as KAHRP, GBP130, and Rifin family members. The virulence protein PfEMP1 was also exported; however, it lacks a signal sequence at the N terminus but has a putative C-terminal transmembrane region. Alignment of the N terminus (preceding the first DBL domain) of the 60 PfEMP1 proteins encoded in the *P. falciparum* genome sequence revealed the presence of a conserved motif that shares features of the Pexel motif (Fig. 3C) (fig. S1, B and C). To determine whether export of PfEMP1 was dependent on the presence of a functional Pexel motif, we fused portions of the *var* gene sequence (PFL1960w) to YFP. Specifically, the N-terminal portion of PfEMP1 including the putative Pexel motif was fused to the PfEMP1 transmembrane domain and the acidic terminal segment (ATS), and YFP (added at the C terminus). Control constructs lacking the putative Pexel motif were used, as in the pre-

vious experiments (Figs. 2 and 3A). Additionally, we specifically mutated the Pexel motif to Ala residues. Again, in support of the hypothesis that this motif is a general export signal for translocation across the parasitophorous vacuole membrane and into the host erythrocyte, only PfEMP1-YFP fusions containing a functional signal were exported into the erythrocyte while the control proteins accumulated in the parasitophorous vacuole (Fig. 3B). Similar to Rif-YFP, PfEMP1-YFP was initially localized to Maurer's clefts in the erythrocyte cytosol, and a rim fluorescence pattern suggestive of surface localization later emerged at the erythrocyte membrane, although we have not shown that this PfEMP1 chimera was in the correct orientation on the red cell membrane surface. Our data support previous evidence that Maurer's clefts function as intermediate compartments for proteins destined for the erythrocyte surface (*6*). Recently, we found that the KAHRP Pexel sequence can complement the equivalent PfEMP1 motif for export across the parasitophorous vacuole membrane and (with the putative transmembrane region) can provide the information for insertion into the erythrocyte membrane and localization of the exodomain on the outside surface of the infected host cell (*10*).

If we assume that parasite protein export via a Pexel-mediated translocation process is the predominant entry into the erythrocyte cytoplasm, the in vivo experiments also confirm that the exported *P. falciparum* proteome can be predicted in silico. Some proteins such as SBP1 (*11*) are exported, although they show a different gene structure and appear to lack a Pexel motif; this observation suggests other mechanisms such as escorters (*12*). Apart from the proteins encoded by the *var* (60 paralogs), *Rifin* (140 paralogs), and *Stevor* (25 paralogs) gene families, a motif search of the *P. falciparum* genome (*13*) at www.plasmodb.org revealed the existence of ~160 genes with (i) a short first exon encoding a signal sequence, and (ii) a Pexel motif encoded close to the start of the second exon (Fig. 1) (fig. S1A). Together, these findings suggest that ~400 *P. falciparum* proteins (i.e., ~8% of its proteome) are exported into the host erythrocyte, where they are involved in antigenic and structural alterations of the erythrocyte membrane and cytoplasm, mediate nutrient import from the red blood cell into the parasite, and provide the machinery for protein export to the erythrocyte (*14*). These exported proteins appear to be unique to *Plasmodium*, because most show no sequence homology to known protein domains. An exception is the RESA protein family, whose members have a DNAJ-like protein-binding domain. Interestingly, the vast majority of the corresponding genes are located in subtelomeric regions of the 14 chromosomes in close apposition to members of the *var* and *Rifin* gene families (Fig. 4A). This finding supports the idea that genes encoding factors involved in host-parasite interactions are located on chromosome ends,

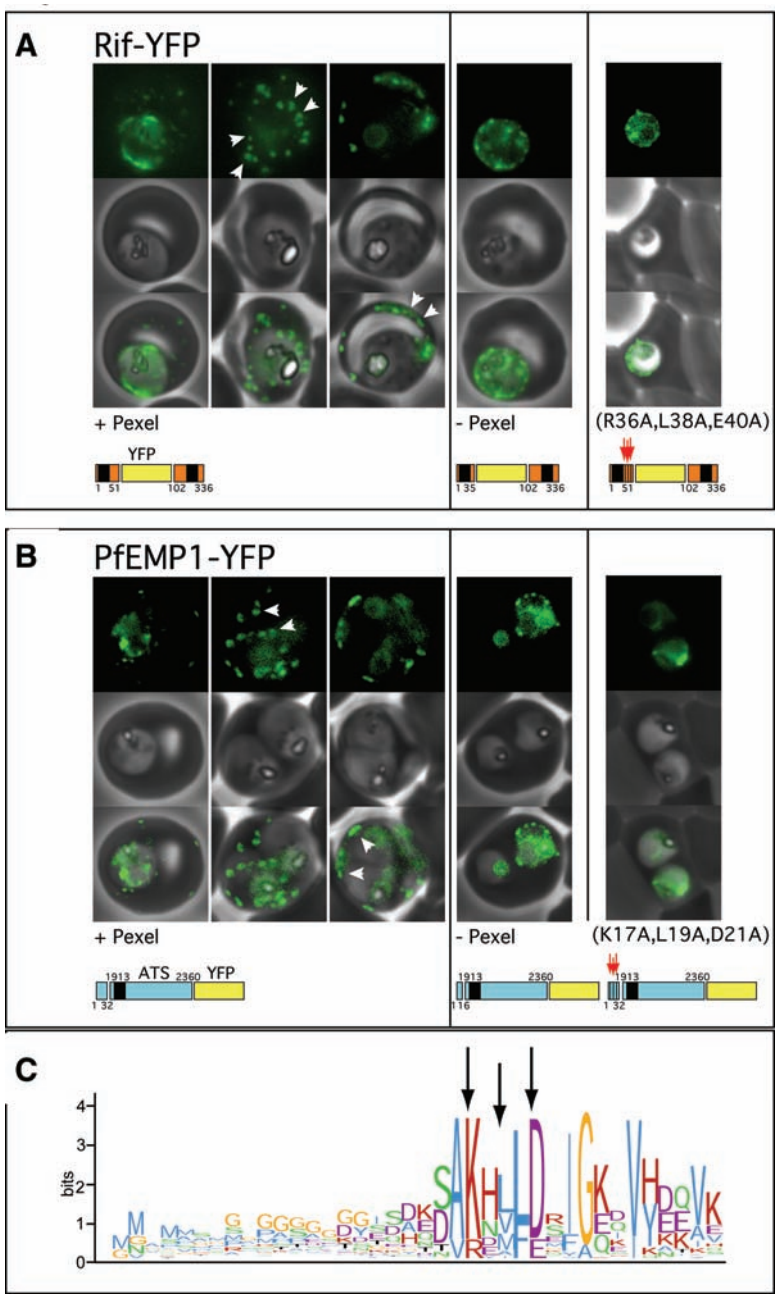

Fig. 3. Export of Rifin surface antigens and of the major *P. falciparum* virulence factor, PfEMP1, depends on a functional Pexel motif. (**A**) A Rifin-YFP fusion including the Pexel motif is exported into the erythrocyte while truncated versions accumulate in the parasitophorous vacuole membrane. Arrowheads indicate the localization of the reporter to punctate structures reminiscent of Maurer's clefts (left panel, center) and later to the surface. Coloring of the schematic representation of fusion proteins: black, hydrophobic stretches (N-terminal signal sequence and C-terminal transmembrane domain); orange, N-terminal and C-terminal portion of the Rifin protein; yellow, YFP. Arrows indicate positions of mutated amino acids. (**B**) Export of PfEMP1-YFP also depends on a functional Pexel motif. As for the Rifin-YFP fusion, the reporter transiently localizes to structures similar to Maurer's clefts, followed by a rim fluorescence suggestive of surface localization. For construct details, see fig. S2. Coloring of the schematic representation of fusion proteins: black, hydrophobic PfEMP1 transmembrane domain; blue, PfEMP1 N terminus and ATS; yellow, YFP. Arrows indicate positions of mutated amino acids. (**C**) Logo alignment (based on the ClustalX alignment of fig. S1B) of all 60 PfEMP1 variants encoded in the genome of *P. falciparum* strain 3D7, indicating the conservation of the Pexel motif. Note that the amino acid residues surrounding the Pexel motif are more conserved in PfEMP1 because it is a more recently radiated gene family. Arrows indicate positions 1, 3, and 5 of the Pexel motif. Amino acid color coding is the same as in Fig. 1.

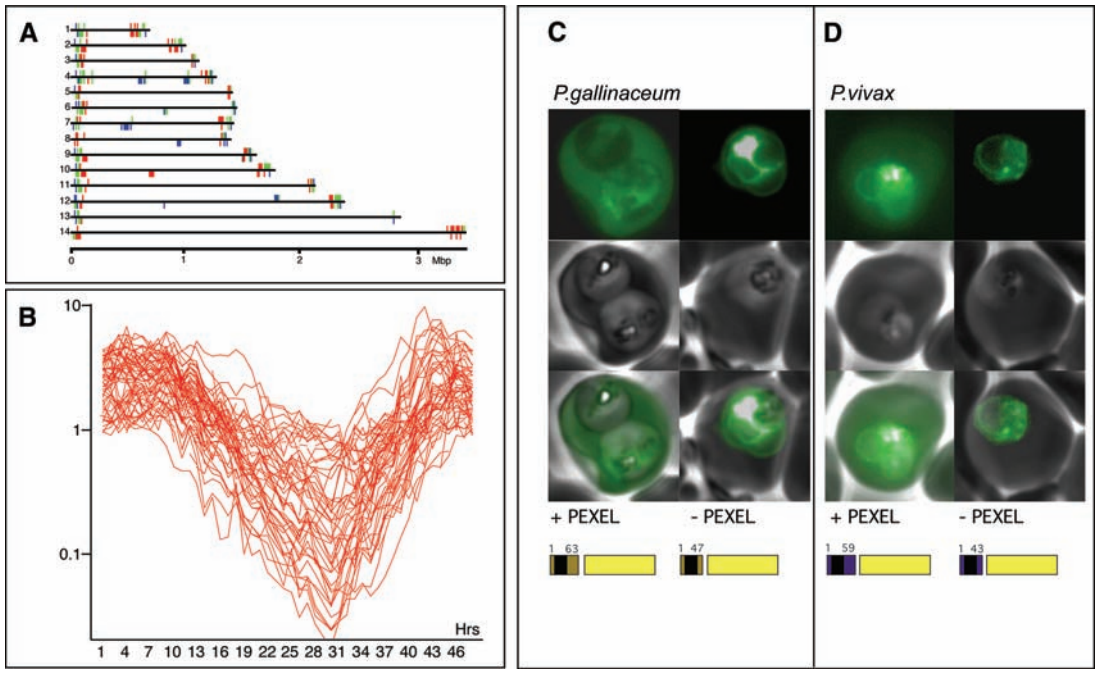

Fig. 4. Defining *Plasmodium* proteins exported into host erythrocytes. (**A**) Chromosomal map of *P. falciparum* strain 3D7, showing the positions of predicted proteins containing Pexel motifs, demonstrates the subtelomeric location of the majority of exported proteins. Bars represent individual genes: 60 *var* (blue), 165 *rif/stevor* (green), and 158 other proteins containing a Pexel motif (red). (**B**) Normalized transcriptional expression profiles during the asexual cycle of 90 Pexel genes [red bars in (A)], visualized on a time course as described (*17*). The set was selected on the criterion that the control expression value of each gene was at least 10 in the Affymetrix microarray experiments (for original microarray data, see fig. S4). Export of predicted surface antigens from *P. gallinaceum* (**C**) and *P. vivax* (**D**) depends on a functional Pexel motif when expressed in *P. falciparum*. Coloring of the schematic representation of fusion proteins: black, hydrophobic signal sequence; olive, N terminus of a predicted *P. gallinaceum* surface antigen; dark blue, N terminus of a predicted *P. vivax* surface antigen; yellow, YFP.

because these regions exhibit the highest genomic plasticity and therefore allow rapid adaptation of the parasite to the host (*15*).

Microarray experiments reveal that the overall transcription patterns of these ~400 genes show peak levels between late ring and early trophozoite stages of development (Fig. 4B) (fig. S4), a time during which most host cell remodeling occurs; this is in agreement with other microarray studies of the *P. falciparum* transcriptome (*16*, *17*). However, some predicted exported proteins are expressed exclusively in sporozoite or gametocyte stages (*18*, *19*), and there is evidence that protein 11.1 is exported in early gametocytes (*20*) (tables S1 and S2). The presence of Pexel-containing proteins in intrahepatocytic stages indicates that the machinery for protein translocation across the parasitophorous vacuole membrane is functional in different host cell types.

A bioinformatic survey of the (uncompleted) genomes of other *Plasmodium* species—the human parasite *P. vivax*, the rodent parasites *P. yoelii* (*21*) and *P. berghei*, and the avian parasite *P. gallinaceum*—revealed a remarkable radiation of the "export element" consisting of the first exon encoding the signal sequence, the intron, and the second exon encoding the Pexel motif (fig. S5). Moreover, we identified a number of large species-specific gene families encoding putative surface antigens in the genomes of *P. vivax*, *P. berghei*, and *P. gallinaceum* (fig. S5). On the basis of their structural similarity to PfEMP1, Rifin, and Stevor in *P. falciparum* (i.e., a highly variable N-terminal exodomain, a putative transmembrane domain, and a conserved cytoplasmic tail), these antigens are predicted to be virulence factors on the surface of the corresponding host cells. Accordingly, we show that the N termini including the Pexel motif of members of these families from *P. gallinaceum* (Fig. 4C) and *P. vivax* (Fig. 4D) are necessary to target YFP to the erythrocyte when transfected into *P. falciparum*. Hence, the Pexel motif is conserved in *Plasmodium* spp. Although many exported proteins are prime candidates for vaccine development or therapeutic intervention, their usefulness is limited by antigenic variation and strain-specific differences. The identification of an export mechanism unique to *Plasmodium* spp. raises the possibility of developing completely novel strategies to interfere with multiple aspects of parasite development through a single target.

References and Notes

1. L. G. Pologe, A. Pavlovec, H. Shio, J. V. Ravetch, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7139 (1987).
2. D. I. Baruch *et al.*, *Cell* **82**, 77 (1995).
3. X. Z. Su *et al.*, *Cell* **82**, 89 (1995).
4. J. D. Smith *et al.*, *Cell* **82**, 101 (1995).
5. S. Kyes, P. Horrocks, C. Newbold, *Annu. Rev. Microbiol.* **55**, 673 (2001).
6. M. E. Wickham *et al.*, *EMBO J.* **20**, 1 (2001).
7. C. Lopez-Estrano, S. Bhattacharjee, T. Harrison, K. Haldar, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12402 (2003).
8. M. Foley, L. Tilley, *Int. J. Parasitol.* **28**, 1671 (1998).
9. See supporting data on *Science* Online.
10. E. Knuepfer, M. Rug, N. Klonis, L. Tilley, A. F. Cowman, unpublished data.
11. T. Blisnick *et al.*, *Mol. Biochem. Parasitol.* **111**, 107 (2000).
12. D. L. Baldi *et al.*, *EMBO J.* **19**, 1 (2000).
13. M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
14. K. Haldar *et al.*, *Cell. Microbiol.* **4**, 383 (2002).
15. L. H. Freitas-Junior *et al.*, *Nature* **407**, 1018 (2000).
16. K. G. Le Roch *et al.*, *Science* **301**, 1503 (2003); published online 31 July 2003 (10.1126/science.1087025).
17. Z. Bozdech *et al.*, *Genome Biol.* **4**, R9 (2003).
18. L. Florens *et al.*, *Nature* **419**, 520 (2002).
19. E. Lasonder *et al.*, *Nature* **419**, 537 (2002).
20. A. Scherf *et al.*, *EMBO J.* **11**, 2293 (1992).
21. J. M. Carlton *et al.*, *Nature* **419**, 512 (2002).

22. The genome sequence data for *P. berghei* and *P. gallinaceum* were produced by the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium). Sequence data for *P. vivax* and *P. yoelii* were obtained from The Institute for Genomic Research (TIGR; www.tigr.org). The *P. vivax* sequences used in this study were performed by TIGR and funded by the National Institute of Allergy and Infectious Diseases and the U.S. Department of Defense. Corresponding PlasmoDB (for 3D7) and GenBank (for all other strains) accession numbers can be found at *Science* Online. We thank J. Healer, A. B. Hehl, G. I. McFadden, L. Tilley, and B. S. Crabb for comments, and the Red Cross Blood Service (Australia) for blood. Supported by a postdoctoral fellowship from the Swiss National Science Foundation (M.M.), a Wellcome Trust International Traveling Fellowship (E.K.), NIH grant R01-A144008-04A1, the Wellcome Trust, and the National Health and Medical Research Council (Australia). A.F.C. is an International Scholar of the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5703/1930/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References

19 August 2004; accepted 22 October 2004
10.1126/science.1102452