Proteomic Profiling of *Plasmodium* Sporozoite Maturation Identifies New Proteins Essential for Parasite Development and Infectivity

Edwin Lasonder¹,²,⁹, Chris J. Janse³,⁹, Geert-Jan van Gemert⁴,⁹, Gunnar R. Mair²,⁸, Adriaan M. W. Vermunt¹,², Bruno G. Douradinha⁴,⁸, Vera van Noort²,², Martijn A. Huynen², Adrian J. F. Luty⁴, Hans Kroese³, Shahid M. Khan³, Robert W. Sauerwein⁴, Andrew P. Waters³,⁹, Matthias Mann⁵,⁶, Hendrik G. Stunnenberg¹,²

¹ Department of Molecular Biology, NCMLS, Radboud University Nijmegen, Nijmegen, The Netherlands, ² Center for Molecular and Biomolecular Informatics, NCMLS, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, ³ Leiden Malaria Research Group, Department of Parasitology, Centre for Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands, ⁴ Department of Medical Microbiology, NCMLS, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, ⁵ Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

Abstract

*Plasmodium falciparum* sporozoites that develop and mature inside an *Anopheles* mosquito initiate a malaria infection in humans. Here we report the first proteomic comparison of different parasite stages from the mosquito—early and late oocysts containing midgut sporozoites, and the mature, infectious salivary gland sporozoites. Despite the morphological similarity between midgut and salivary gland sporozoites, their proteomes are markedly different, in agreement with their increase in hepatocyte infectivity. The different sporozoite proteomes contain a large number of stage specific proteins whose annotation suggest an involvement in sporozoite maturation, motility, infection of the human host and associated metabolic adjustments. Analyses of proteins identified in the *P. falciparum* sporozoite proteomes by orthologous gene disruption in the rodent malaria parasite, *P. berghei*, revealed three previously uncharacterized *Plasmodium* proteins that appear to be essential for sporozoite development at distinct points of maturation in the mosquito. This study sheds light on the development and maturation of the malaria parasite in an *Anopheles* mosquito and also identifies proteins that may be essential for sporozoite infectivity to humans.

Citation: Lasonder E, Janse CJ, van Gemert G-J, Mair GR, Vermunt AMW, et al. (2008) Proteomic Profiling of *Plasmodium* Sporozoite Maturation Identifies New Proteins Essential for Parasite Development and Infectivity. PLoS Pathog 4(10): e1000195. doi:10.1371/journal.ppat.1000195

Received July 22, 2008; Accepted October 9, 2008; Published October 31, 2008

Copyright: © 2008 Lasonder et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Part of this work was supported by the Dutch Science Foundation (NWO/Genomics, grant number 050-10-053) and The Wellcome Trust Functional Genomics Initiative. EL and SMK were supported by a programme from the Dutch NWO/Genomics, AMWV and GJvG by Maltrans (EU Framework Programme VI), GRM was supported by BioMalPar and is a recipient of a Netherlands Genomics Initiative HORIZON Project Grant (050-71-061), and SMK is supported by Ti-Pharma (T4-102). This study was supported by the WHO.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: h.stunnenberg@ncmls.ru.nl

† Current address: Instituto de Medicina Molecular, Lisboa, Portugal

‡ Current address: Groen Agro Control, Delfgauw, The Netherlands

§ Current address: Biochemistry and Molecular Biology, Novartis Vaccines and Diagnostics, Siena, Italy

∥ Current address: European Molecular Biology Laboratory, Heidelberg, Germany

¶ Current address: Division of Infection and Immunity, Faculty of Biomedical Life Sciences, and Wellcome Centre for Molecular Parasitology, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow, Scotland, United Kingdom

¶ Current address: Max Planck Institute for Biochemistry, Department of Proteomics and Signal Transduction, Martinsried, Germany

These authors contributed equally to this work.

Introduction

The life cycle of human malaria parasite *Plasmodium falciparum* within the mosquito vector begins when gametocytes are taken up in an infected blood meal; after forming gametes and fertilisation, the resulting zygote differentiates into a motile ookinete that traverses the midgut epithelium and transforms within 36--48 hours into an oocyst (OOC) between the midgut epithelial cells and the basal lamina. The oocyst is an asexually replicating form of the parasite, which produces up to 2000--4000 sporozoites in about two weeks. Rupture of mature oocysts releases oocyst-derived sporozoites (ODS) into the hemocel of the mosquito. The movement of the hemolymph brings the ODS in contact with the salivary glands, which they then invade. The sporozoites mature inside the salivary glands and then are stored ready for transmission to the mammalian host upon the next blood meal. A limited number of the salivary gland sporozoites (SGS) are injected during a mosquito bite and only a few of these complete the necessary migration from the skin to the liver to establish an infection inside hepatocytes. Clearly, the sporozoite has to complete a number of functions and metabolic readjustments both before and after injection into a mammalian host. The sporozoite has to be capable of actively exiting an oocyst, travelling through the hemolymph (the mosquito circulatory system), and
invading salivary glands. Further, following a mosquito bite injection the sporozoites enters a very different physiological environment of the human host, and then has to traverse through human endothelial cells, possibly Kupffer cells and finally hepatocytes where they establish an infection; moving all the time using a specialized form of gliding motility. Despite all these events the general morphology of the sporozoite is not visibly altered at any stage (for general reviews on sporozoite biology please see the following references and the references therein [1–7]).

Since the sporozoite plays an essential role in the first phase of a malaria infection, an understanding of its biology is of great importance in order to develop intervention methods against initial infection and consequently disease. A wealth of gene expression data from high throughput studies exists on the intracellular erythrocytic growth and development of Plasmodium parasites [8–17], whereas far less is known about the genes/proteins involved in sporozoite development [10,11,16,18–22]. Indeed, only a few (less than 25) proteins have been characterized as being essential for sporozoite development and infectivity. These include several proteins that are currently under investigation as either potential subunit vaccines (such as circumsporozoite protein (CS) and thrombospondin related anonymous protein (TRAP)) or may serve in the generation of whole organism, protein (CS) and thrombospondin related anonymous protein (TRAP).

In this study we have performed a detailed proteomic comparison of sporozoites obtained from both oocysts and salivary glands which were obtained by hand-dissection of infected mosquito midguts and salivary glands. The proteome analysis was performed using essentially the same high throughput mass spectrometric analysis that we previously applied to generate the proteomes of the blood stages of P. falciparum [15] as well as the proteomes of male and female gametocytes of P. berghei [14]. Our analyses resulted in a proteome of oocysts (n = 127), oocyst-derived sporozoites (n = 450) and salivary gland sporozoites (n = 477), which represent 728 individual Plasmodium proteins, of which 250 were exclusively detected in the oocyst/sporozoite stages when compared to the P. falciparum blood stage proteome generated in a previous study [15]. The identification of proteins and their relative distributions within the different proteomes suggest specific metabolic adaptations and other biological functions of the maturing sporozoite. Moreover, we analyzed the function of eight sporozoite-specific proteins identified in our proteome analyses that were specifically annotated as hypothetical proteins, by targeted gene disruption of the orthologous genes of the rodent malaria parasite, P. berghei. We were able to demonstrate an essential and distinct role for three of these proteins in sporozoite development.

**Results**

**Mosquito stage proteome**

Protein samples derived from infected mosquito midguts and salivary glands were analyzed by nano–liquid chromatography tandem mass spectrometry (nLC-MS/MS) essentially as previously described [15]. The MS/MS spectra were searched against a combined database of all possible predicted tryptic peptides derived from all P. falciparum, human, and mosquito (Anopheles gambiae) proteins. The proteomic analysis of P. falciparum oocysts, oocyst-derived sporozoites, and salivary gland sporozoites resulted in a total of 4611 unique peptides mapping to 728 non redundant Plasmodium proteins; they are distributed over the three stages with 127, 450 and 477, respectively and depicted as a Venn diagram in Figure 1A. Identified tryptic peptides and corresponding Plasmodium proteins of the mosquito stages are provided as supplementary material (Table S1). In our previous analysis of infected human red blood cells we identified 741 asexual blood stage parasite proteins from a mixture of schizonts and trophozoites and an additional 931 gametocyte and 645 gamete proteins [15]. Merging these datasets with the proteomes of the mosquito stages resulted in the identification of 250 Plasmodium proteins (Table S1) that are specifically detected in mosquito stages and 809 proteins that are expressed only in the blood stages (Figure 1B). However, it is important to note that due to the incomplete nature of all proteome datasets, absence of proteins from one dataset may also be due to the limits of detection and not the actual absence of expression. Parasite samples derived from infected mosquitoes were considerably contaminated with mosquito proteins with total parasite protein fractions of 35% for ODS, 31% for SGS and for OOC only 11% of the sequenced proteins were parasite in origin. Therefore this relatively high degree of contamination resulted in overall lower numbers of proteins compared to our previous Plasmodium infected blood stage proteome study. In particular, only 127 P. falciparum proteins in a pool of 987 mosquito proteins were identified for the oocyst sample that presumably represents the more abundantly expressed parasite proteins. Therefore, further analysis of the identified proteins and additional functional analyses are mainly focused on the proteins identified in the ODS and SGS. In total, we analyzed six different stages of Plasmodium (both from this study and our previous work) and have identified a total of 1543 Plasmodium proteins. The proportion of ‘stage specific’ proteins in the different life cycle stages ranged from 12% (gametes) to 28% and the stage specificity of proteins in the mosquito stages ranged between 15–24% (Figure 1C).

**Comparison with existing RNA/protein mosquito stage studies**

Genome-wide proteome and transcriptome studies have previously been reported for salivary gland sporozoites of P. falciparum [10,16], for oocysts and sporozoites of P. berghei [11] and
P. falciparum
distribution of detected
different life-cycle stages. (A) Venn diagram depicting the
Figure 1. Distribution of identified

Oocyst (127 )

24

30

197

62

11

161

243

Oocyst derived

sporozoite (450)

Salivary gland

sporozoite (477)

Mosquito stage

proteome (728)

Blood stage

proteome (1287)

250

478

809

Proteomics of P. falciparum Sporozoite Maturation

recently for oocyst-derived sporozoites and salivary gland sporozoites of P. yoelii [22]. The Florens et al SGS proteome [10] identified a total of 1040 proteins of which 314 proteins include at least one peptide that is fully tryptic. It has been shown that selection of only fully tryptic peptides greatly increases the confidence in each protein within the proteome and was similarly applied to our dataset [31]. Comparison of these ‘fully-trypitc proteins’ (proteins identified by peptides conforming to proper tryptic cleavage) with the ‘fully-trypitc proteins’ from our SGS proteome (n = 477) shows that 166 proteins are present in both proteomes (i.e. 53% of the Florens’ data (Table S2)). Moreover, in order to further increase our confidence in the ‘protein-calling’ in both datasets, a comparison was made using only those proteins that were identified by 2 or more fully tryptic peptides (i.e. 346 proteins from our mosquito stage proteome and 82 from the Florens SGS proteome). In this analysis, we found that 72 proteins were in common (i.e. 88% of the Florens enriched SGS proteome). Interestingly, we fail to find any PIEMP-1 proteins, as had previously been reported in the Florens et al SGS proteome, in either dataset when we examine only the ‘fully-trypitc peptide proteomes’ [10].

The oocyst proteome of P. berghei described by Hall et al [11] detected a total of 220 proteins of which 175 proteins have an orthologue in P. falciparum and 87 of these (i.e. 50%) were also detected in our mosquito proteomes (Table S2). Again consideration of only fully tryptic peptides revealed that 60 of the resulting 111 P. berghei orthologs (i.e. 54%) were found in common. Similarly, of the 108 proteins identified in the P. berghei SGS proteome 86 proteins have an orthologue in P. falciparum (Table S2) of which 46 (i.e. 53% of the Hall SGS proteome) were detected in our SGS proteome of P. falciparum. There were only 20 fully tryptic proteins in the Hall SGS proteome of which 75% (n = 15) were also detected in our P. falciparum SGS proteome. Selecting the 202 genes that were commonly expressed in our SGS proteome and in the published SGS proteome of P. falciparum [10], the relative abundance of protein in the two datasets was examined using a Pearson correlation. The emPAI peptide counting method using the number of observed peptides detected per protein and corrected to the number of expected tryptic peptides was applied to compute relative protein levels [32,33]. A good correlation (r = 0.73) existed between protein abundance levels (emPAI values; see Materials and Methods section) in our SGS proteome and the published SGS proteome of P. falciparum [10], the absolute abundance of protein in the two datasets was examined using a Pearson correlation. The emPAI peptide counting method using the number of observed peptides detected per protein and corrected to the number of expected tryptic peptides was applied to compute relative protein levels [32,33]. A good correlation (r = 0.73) existed between protein abundance levels (emPAI values; see Materials and Methods section) in our SGS proteome and the published SGS proteome of P. falciparum [10], the absolute abundance of protein in the two datasets was examined using a Pearson correlation. The emPAI peptide counting method using the number of observed peptides detected per protein and corrected to the number of expected tryptic peptides was applied to compute relative protein levels [32,33]. A good correlation (r = 0.73) existed between protein abundance levels (emPAI values; see Materials and Methods section) in our SGS proteome and the published SGS proteome of P. falciparum [10], the absolute abundance of protein in the two datasets was examined using a Pearson correlation. The emPAI peptide counting method using the number of observed peptides detected per protein and corrected to the number of expected tryptic peptides was applied to compute relative protein levels [32,33].

Several (smaller scale) studies have been reported that using either subtractive hybridization or cDNA quantification methods (i.e. Serial Analysis of Gene Expression (SAGE)) to identify sets of genes transcribed in sporozoites in the rodent malaria parasites, P. berghei [20,21] and P. yoelii [18]. Comparison of the identified P. yoelii mRNAs with our proteomes showed that for nearly all genes transcribed in sporozoites (20 out of 23 sporozoite (S) genes), proteins were detected in our sporozoite proteomes (Table S4). This may suggest that for a significant proportion of genes transcription and protein expression coincide within the sporozoite. However, a weaker correlation was found between transcription in P. berghei sporozoites and the presence of protein in our proteomes. Specifically, we were able to detect protein for 34 of the 98 genes identified in the P. berghei sporozoite SAGE analysis (i.e. the Sporozoite expressed gene Identified by SAGE (SIS) genes (Table S4)) but only 5 out of 26 transcribed genes in the Suppression Subtractive Hybridization (SSH) analysis (i.e. the
Upregulated in Sporozoites (UIS) genes (Table S4). It is however interesting to observe that between the two SSH studies only 2 out of 30 genes appear clearly up-regulated in both *P. yoelii* and *P. berghei* sporozoites.

**Functional annotation of mosquito stage proteins**

A global functional characterization of the ‘mosquito stage’ proteome was performed by an enrichment analysis of Gene Ontology (GO) annotations, for both the proteins that are shared between blood stages and mosquito stages (n = 478) and for the mosquito stage specific proteins (n = 250). The set of 478 genes commonly expressed in both mosquito and blood stages showed enrichment in GO annotations in all classes (i.e. Molecular Function, Cellular Component and Biological process (Figure S1)) and this enrichment is principally associated with housekeeping genes (Figure 2). The mosquito stage specific proteome did not reveal significant (p<0.01) enrichment in GO annotations nor did additional analyses for GO enrichment of the mosquito stage specific proteins using BINGO [34] and Ontologizer [35] (data not shown). In Figure 2B GO categories (Molecular Function) are shown for the mosquito stage specific proteome that contain more than 5 proteins. The lack of enrichment could be caused by the high proportion of genes annotated as hypothetical (300 out of 728) and consequently the relatively large number of proteins in the mosquito stage specific proteome (124 out of 250) without a GO annotation. Since our analysis did not reveal a significant GO enrichment for proteins known to be important in sporozoite function (e.g. motility and motor activity (Figure 2)) we analyzed our mosquito stage proteome for previously reported proteins, for which a function during sporozoite development is described and supported by strong experimental evidence (e.g. gene-knockout and/or antibody-inhibition studies). These proteins, in total 23, are listed in Table 1 and 15 out of 23 proteins are present in the mosquito stage proteome reported here. Based on a total number of 5410 genes in the genome of *P. falciparum* and 728 proteins in our mosquito stage proteome, these 15 proteins represent a 4.8 fold functional enrichment relative to the annotated genome and is highly significant (p<0.001 using Ontologizer). A good agreement exists between the function of the sporozoite proteins as shown in Table 1 and their expression pattern in the different mosquito stages. For example, proteins with multiple roles during sporozoite maturation (e.g. CS and TRAP) were identified in all stages (OOC, ODS and SGS) whereas proteins involved in hepatocyte traversal, such as SPECT1, SPECT2 (sporozoite microneme protein essential for cell traversal 1 and 2) and CelTOS (cell-traversal protein for ookinetes and sporozoites) were exclusively identified in mature SGS.

Sporozoites, like other motile stages (except male gametes) of Apicomplexan organisms, move on substrates by a mechanism known as gliding motility which is driven by an actomyosin motor complex [3,36,37]. Although there was no enrichment with high confidence (p<0.05) of the GO Molecular Function category ‘motor activity’ for mosquito stage specific proteins (Figure 2), several proteins known to be involved in the actomyosin motor complex are well represented and include TRAP, myosin A, MyoATail Domain Interacting Protein (MTIP), actin and F-1,6-BP aldolase (3.6 fold enrichment with low confidence). Additionally, sporozoites encode a variety of surface molecules for both motility and invasion of host cells. For apicomplexan parasites members of the TRAP/MIC2 family have been shown to be important for host cell recognition and motility. The general architecture of this family is typified by one or more thrombospondin type I (TSP1) domains in their extracellular regions which may in addition also posses von Willebrand factor A (vWA) extracellular domains [38]. Our sporozoite proteome shows a 4.7 fold enrichment for proteins that contain one or multiple TSP1 domains (Table 2) compared to the *P. falciparum* proteome of 5410 proteins.

**Two sporozoite proteomes – ODS versus SGS**

Although the morphology of oocyst-derived and salivary gland sporozoites is identical at the level of light microscopy, ODS of *P. berghei* are significantly less infective to the mammalian host than SGS [39]. This marked difference in infectivity suggests significant developmental changes between these forms and was indicated by the analyses of gene transcription of different sporozoite stages by either SSH screens or SAGE analysis, which alludes to changes in protein expression in the sporozoite during the period of egress from the oocysts and the establishment of infection of the salivary glands [18,20,21]. In agreement with these observations, we found a large number of proteins expressed in SGS that were absent or relatively low expressed in ODS (Table S1). Several proteins involved in metabolic pathways show clear differences in distribution between ODS and SGS (Figure S2). For example, 8 out of 9 enzymes of the glycolytic pathway for ATP production were detected, all which were either more abundant or exclusive to SGS (SGS 8 proteins with 140 peptides; ODS 4 proteins and 48 peptides). A similar profile is observed for proteins involved in the production of NADPH via the pentose phosphate pathway with an up-regulation of these proteins in SGS (5 proteins and 26 peptides) compared to ODS (1 protein and 2 peptides). A third up-regulated metabolic pathway is the tricarboxylic acid (TCA) cycle (7 proteins and 85 peptides in SGS compared to 4 proteins and 26 peptides in ODS). Interestingly, several genes (4 out of 10) of the TCA cycle are most abundantly expressed in SGS, not only when compared to ODS but also in comparison with the blood stages, indicating an important role of the TCA cycle in mature sporozoites. Also the enzyme phosphoenolpyruvate carboxykinase (PF13_0234) is upregulated in the salivary gland sporozoites (3 peptides in ODS and 17 in SGS), which is again in agreement with the upregulation of enzymes involved in the TCA cycle and glycolysis [40]. It also appears that SGS prepare for enhanced protein synthesis: 9 of the 11 detected tRNA ligases are only detected in the SGS proteome and not in the ODS proteome (Table S1) as are ribosomal proteins, translation elongation factors and the TCP chaperon complex proteins, which are either exclusively detected in SGS or are represented in the SGS proteome by substantially more peptides compared to the ODS proteome. As is shown in Table 1, proteins that are known to play a role in traversal and invasion of hepatocytes are highly enriched in SGS. On the other hand, the expression of MAEBL that is expressed along with CS and well before AMA-1 [41] and is known to function in attachment and invasion of the salivary gland [42] is more abundantly expressed in ODS. Therefore, it would appear that the proteomes of the sporozoite characterised by this study at different stages of development accurately reflect the functionality of either the ODS or SGS.

Consequently, based on the expression pattern and relative abundance of the peptides in the proteomes from OOC, ODS and SGS (see Materials and Methods section) the mosquito stage specific proteins can be regarded as belonging to one of 3 distinct groups (Table S1): Group I consists of 112 ODS proteins highly enriched for the ODS stage, putatively involved in sporozoite maturation inside the oocyst and in salivary gland invasion; similarly Group II which contains 74 proteins up-regulated in SGS potentially involved in infection of the mammalian host; and finally Group III that contains 59 proteins that are shared between the different mosquito stage proteomes and therefore may be
involved in sporozoite functions necessary both in the mosquito vector and the mammalian host (e.g. proteins involved in gliding motility and invasion such as CS [43,44] and TRAP [45,46] (Table 1)). These three groups formed the basis for selection of genes for further functional analysis of their encoded proteins through targeted disruption of the orthologous genes in the rodent malaria parasite, \emph{P. berghei}. The three groups were further refined for subsequent functional analysis using the following criteria (see also Materials and Methods section): i) high expression level as determined by the number of uniquely detected peptides per protein, ii) presence of gene sequences encoding putative transmembrane regions, signal peptides and/or GPI anchors, and iii) presence exclusively in the mosquito stage proteomes. This resulted in selection of genes as shown in Table 3. Further, in order to enrich for proteins that may define \emph{Plasmodium} specific functions, we preferentially selected not only genes that were annotated as hypothetical but also had no domains predicted by either the SMART or Pfam algorithms (i.e. with no indication of predicted function).

**Figure 2. Gene Ontology term enrichment analysis of mosquito stage proteome.** (A) Enrichment for GO ‘Biological Process’ terms of proteins detected in mosquito and blood stages. The figure shows terms on the x-axis that are significantly enriched (p<0.004) by more than four fold. GO terms of the shared set of proteins (n=478, purple bars) is compared to terms of all predicted \emph{P. falciparum} proteins (5410, green bars). The y-axis displays the fraction relative to all GO Biological Process terms. (B) Enrichment for GO ‘Molecular Function’ main terms of proteins detected specifically in mosquito stages (and blood stages). GO terms of the mosquito specific set of proteins (n=250, blue bars) is compared terms of all predicted \emph{P. falciparum} proteins (5410, green bars). The y-axis displays the fraction relative to all GO Molecular Function terms. These terms do not show a significant enrichment (p>0.5).

doi:10.1371/journal.ppat.1000195.g002

Functional characterization of sporozoites-specific proteins

In total eight genes identified in this study were selected (Table 4) for functional analysis by targeted gene disruption of their corresponding orthologs in \emph{P. berghei}, specifically, 3 ODS specific

---

**Proteomics of \emph{P. falciparum} Sporozoite Maturation**

PLoS Pathogens | www.plospathogens.org | 5 October 2008 | Volume 4 | Issue 10 | e1000195
Table 1. Characterized proteins involved in sporozoite development and invasion of host cells.

| Accession nr | Protein name (1) | Protein involved in (2) | nr unique pept/protein in life cycle stages (3) | Reference |
|--------------|------------------|------------------------|-----------------------------------------------|------------|
| PF14_0067    | CCP3             | sporozoite development | - 31 10 - - -                                  | Claudianos [72], Pradel [73] |
| PF14_0532    | CCP2             | sporozoite development | - 39 20 - - -                                  | Pradel [73] |
| PFC0180c     | IMC1             | sporozoite development (cell shape) | - - - 0 9 25 | Khatre [74] |
| PFC0210c     | CS               | sporozoite development, salivary gland and hepatocyte invasion | - - - 2 5 9 | Menard [43], Wang [44] |
| PFB0325c     | Cysteine protease | egress from oocyst     | - - - - 16 23 | Aly [56] |
| PF13_0233    | Myosin A         | sporozoite gliding motility | - - - 2 39 49 | Bergman [73] |
| PFL2225w     | MTIP             | sporozoite gliding motility | 1 1 - 2 4 3 | Bergman [73] |
| PF13_0201    | SSP2/TRAP        | sporozoite gliding motility, salivary gland and hepatocyte invasion | - - - 1 6 35 | Rogers [45], Sultan [46] |
| PF10550w     | CRMP1            | salivary gland invasion | - 1 - - - -                                  | Thompson [76] |
| MAL7P1.92    | CRMP2            | salivary gland invasion | - - - - - -                                  | Thompson [76] |
| PF11_0486    | MAEBL            | salivary gland invasion | - - - - 33 1 | Kariu [42] |
| MAL13P1.212  | SPECT1           | cell traversal hepatocytes | 1 - - - - - 16 | Ishino [77] |
| PFDo430c     | SPECT2           | cell traversal hepatocytes | 1 - - - - - 26 | Ishino [78] |
| PEL0800c     | CelTOS           | cell traversal hepatocytes | - - - - - 6 | Kariu [79] |
| PFF1420w     | PL               | cell traversal hepatocytes | - - - - - - - | Bhanot [80] |
| PF11_0344    | AMA-1            | hepatocyte invasion     | 1 - - - - - 2 19 | Silve [81] |
| PFB0570w     | SPATR            | hepatocyte invasion     | - - - - - - 5 | Chattopadhyay [82] |
| PFB0095c     | PEMP3            | hepatocyte invasion     | 3 - - - - - - - | Gruuner [83] |
| PF07_0006    | STARp            | hepatocyte invasion     | - - - - - - - | Pasquetto [84] |
| GI1477963    | SALSA            | hepatocyte invasion     | - - - - - - - | Puentes [85] |
| PFAA200w     | TRSP             | hepatocyte invasion     | - - - - - - 1 3 | Labaerd [86] |
| PFD2015c     | Pf52 protein/P36p | hepatocyte invasion - development | - - - - - - - 5 | Ishino [29], van Dijk [30] |
| PFD2010c     | Pbs36 homologue/P36 | hepatocyte invasion - development | - - - - - - - 8 | Ishino [29] |

1CCP (members of the LCCL protein family), IMC (inner membrane complex), CS (circumsporozoite), MTIP (myosin A tail domain interacting protein), SSP2 (sporozoite surface protein), TRAP (Thrombospondin related anonymous protein), CRMP (cysteine repeat modular protein), MAEBL (membrane antigen erythrocyte binding like protein), SPECT (sporozoite microneme protein essential for cell traversal), CelTOS (cell-traversal protein for ookinete and sporozoite), PL (phospholipase), AMA (apical membrane antigen), SPATR (secreted protein with altered thrombospondin domain), EMP (erythrocyte membrane protein), STARp (sporozoite threonine and asparagine-rich protein), SALSA (sporozoite and liver stage antigen), TRSP (thrombospondin related protein).

2The role of these proteins has been determined by functional analysis in gene-knockout and/or antibody-inhibition studies.

3ASX - asexual blood stages; GCT – gametocytes; GAM – gametes; OOC – oocysts; ODS - oocyst-derived sporozoites; SGS - salivary gland sporozoites.

It was not possible to select mutant parasites for two genes, one belonging to Group I (orthologous to PF14_0607) and the other belonging to Group III (orthologous to PFA0205w) in 3 independent transfection experiments, suggesting that both these proteins may have an additional and essential role during blood stage development. For the remaining 6 genes mutants were generated in two independent transfection experiments per gene (Table 4) and correct disruption of the target genes was shown for all mutants (Figure S3). All 6 mutant lines showed normal asexual growth and also gametocyte and ookinet production that was comparable to wild type parasites (data not shown). As an initial phenotype screen of mosquito stage development, uncloned parental populations of the 6 mutant lines were allowed to infect mosquitoes. Oocyst numbers and salivary gland sporozoite numbers were determined at day 6 and 20 after infection, respectively, and infected mosquitoes were allowed to feed at day 20–22 on naive mice. In 3 out of the 6 mutant lines (orthologous to ΔPF11_0528, ΔPF14_0074 and ΔPFF1195c) parasite development inside the mosquito (oocyst number and salivary gland sporozoites number) was not significantly different from wild type
parasites (Table 4). After infection of mice by bite of mosquitoes infected with any of these three mutant lines, all mice developed parasitemias between 0.1 and 0.5 at day 4 after infection, indicating ‘wild type’ infectivity of the sporozoites of these 3 mutants. Genotype characterization by Field Inverse Gel Electrophoresis (FIGE) analysis and diagnostic PCR of blood stage parasites after mosquito transmission of these 3 mutants revealed the correct gene disruption genotype in blood stages of all 3 mutants, demonstrating normal mosquito transmission of the mutant, rather than breakthrough of wild type parasites (Figure S3). The lack of a clear effect of disruption of these 3 genes on sporozoite production and infectivity to the mammalian host suggests the existence of significant redundancy in the function of these mosquito stage specific proteins.

The remaining 3 mutant lines (orthologous to ΔPF14_0435, ΔPFD0425w and ΔMAL8P1.66) showed an aberrant development during mosquito development. The phenotypes of cloned lines of these mutants were therefore analyzed in more detail. Clones of all 3 gene-disrupted lines produced wild type numbers of oocysts ranging from 150–250 oocysts per mosquito on day7/8 post infection. The development of parasites deficient in PB00029.02.0 (orthologue of PF14_0435; line 802cl1) was largely blocked at the oocyst stage. However, low numbers of sporozoites were formed that were able to invade the salivary gland (2750–6250 oocyst sporozoites per mosquito and 875–6600 SGS per salivary gland). Despite the low numbers of sporozoites that emerge from the oocyst, salivary gland invasion appears not to be affected since ODS and SGS numbers were comparable. In contrast to sporozoites of mutant 841cl1, salivary gland sporozoites of 843cl1 injected either intravenously (1 × 10⁷ sporozoites) or by mosquito bite were not infective for mice (2 experiments with 2 mice). Interestingly, 843cl1 sporozoites demonstrated the same or greater hepatocyte traversal rate than wild type sporozoites and they were also able to traverse and invade hepatocytes in vitro (Figure 3). This suggests that the lack of sporozoite infectivity to mice may be due to a defect in liver stage development after invasion of the hepatocyte.

| Accession nr | Protein name | domain | nr unique pept/prot in life cycle stages (1) | Reference |
|--------------|--------------|--------|---------------------------------------------|-----------|
| PF13_0201    | TRAP         | TSP1   | VWA                                         | 1 6 35    |
| PFCD010c     | CS           | TSP1   | -                                           | 2 5 9     |
| PFAO200w     | TRSP         | TSP1   | -                                           | 1 3       |
| MAL8P1.45    | hypothetical protein | TSP1   | -                                           | 1 - -     |
| PB0570w      | SPATR        | TSP1   | -                                           | 5         |
| PFF0800w     | TRAP-like protein | TSP1   | 1                                           | 2         |
| PFL0870w     | PTRAMP       | TSP1   | -                                           | 1         |
| PF10_0281    | MTRAP        | TSP1   | -                                           | - - - -   |
| PF08_0136b   | WARP         | TSP1   | -                                           | - - - -   |
| PFC0640w     | CTRP         | TSP1   | VWA                                         | 1 2       |
| PFL0875w     | hypothetical protein | TSP1   | -                                           | - - - Aravind [38] |

1 ASX - asexual blood stages; GCT – gametocytes; GAM – gametes; OOC – oocysts; ODS - oocyst-derived sporozoites; SGS - salivary gland sporozoites.

Table 2. Expression of Plasmodium proteins containing thrombospondin type 1 (TSP1) and/or von Willebrand factor A (vWA) domains in different life cycle stage proteomes.

Discussion

The proteome analyses of the three mosquito stages of Plasmodium falciparum, oocysts, oocyst-derived sporozoites and salivary gland sporozoites, resulted in the identification of 728 proteins of which 250 are ‘mosquito stage specific’, having not been detected in our previous analysis of blood stage parasites [15]. Although the total number of proteins identified in the mosquito stages is lower compared to blood stages [15], which is in all likelihood due to sample purity and not reduced protein expression, we show a clear developmental progression of the
Table 3. Characteristics of mosquito stage specific proteins selected on the basis of expression pattern during sporozoite development.

| Accession nr (1) | Protein Name                  | MW (kDa) | Group (2) | MOTIF PREDICTION (3) | ORTHOLOGY PREDICTION (4) | nr unique peptide/protein (5) |
|------------------|-------------------------------|----------|-----------|----------------------|---------------------------|-------------------------------|
|                  |                               |          |           |                      |                           |                               |
| PF11_0528        | hypothetical protein          | 710      | I         | 5.3                  |                           |                               |
| PF11_0486        | MAEBL                        | 243      | I         | 1.3                  |                           |                               |
| MAL13P1.66       | hypothetical protein          | 346      | I         | 1.8                  |                           |                               |
| PFF0645c         | integral membrane protein     | 163      | I         | 6.5                  |                           |                               |
| PF14_0404        | hypothetical protein          | 408      | I         | 0.8                  |                           |                               |
| PF14_0722        | CRMP4                        | 699      | I         | 9.0                  |                           |                               |
| PF14_0435        | hypothetical protein          | 144      | I         | 6.8                  |                           |                               |
| MAL8P1.126       | serine protease               | 101      | I         | 1.5                  |                           |                               |
| PF14_0607        | hypothetical protein          | 127      | I         | 4.8                  |                           |                               |
| MAL8P1.135       | hypothetical protein          | 116      | I         | 8.8                  |                           |                               |
| PF14_0694        | protein disulfide isomerase   | 66       | I         | 0.8                  |                           |                               |
| PF11_0229        | hypothetical protein          | 57       | I         | 1.1                  |                           |                               |
| PFE0340c         | hypothetical protein          | 87       | I         | 5.0                  |                           |                               |
| MAL13P1.119      | cAMP-specific 3’,5’-cyclic    | 86       | I         | 4.5                  |                           |                               |
| PFD0295c         | hypothetical protein          | 85       | I         | 1.0                  |                           |                               |
| PF11_0064        | hypothetical protein          | 85       | I         | 1.5                  |                           |                               |
| PF11_0141        | UDP-galactose transporter     | 39       | I         | 8.0                  |                           |                               |
| PF10_0223        | hypothetical protein          | 49       | I         | 4.3                  |                           |                               |
| PF11_0344        | AMA-1                         | 72       | II        | 1.5                  |                           |                               |
| PFD0430c         | SPECT2 (PLP1)                | 94       | II        | 1.5                  |                           |                               |
| PF14_0074        | hypothetical protein          | 47       | II        | 0.5                  |                           |                               |
| MAL13P1.212      | SPECT1                       | 28       | II        | 0.5                  |                           |                               |
| PFL0800c         | CalTOS                       | 20       | II        | 0.5                  |                           |                               |
| PF80570w         | SPATR                        | 29       | II        | 0.8                  |                           |                               |
| PFF1195c         | hypothetical protein          | 75       | II        | 1.1                  |                           |                               |
| PFE0395c         | hypothetical protein          | 41       | II        | 1.3                  |                           |                               |
| PFC0905c         | hypothetical protein          | 360      | III       | 1.0                  |                           |                               |
| PF13_0233        | myosin a                     | 92       | III       | 0.8                  |                           |                               |

Proteomics of P. falciparum Sporozoite Maturation
| Accession nr (1) | Protein Name | MW (kDa) | Group (2) | MOTIF PREDICTION (3) | ORTHOLOGY PREDICTION (4) | nr unique peptide/protein (5) |
|-----------------|--------------|----------|-----------|----------------------|--------------------------|-------------------------------|
|                 |              |          |           |                      | Ortho MCL group | nr species | additional nr species | OOC | ODS | SGS | Mosq. fr. |
|                 |              |          |           |                      | Plasm. | Api | other |      |      |      |          |
| PFD0425w        | hypothetical protein | 113   | III       | 1 | 1.8 | OG2_86498 | 6 | 2 | 0 | 3 | 42 | 38 | 1 |
| PF13_0201       | TRAP         | 65      | III       | 1 | 1.8 | VWA; TSP1 | OG2_79098 | 6 | 3 | 5 | 1 | 6 | 35 | 1 |
| PFC0210c        | CS           | 43      | III       | 1 | 1 | TSP1 | OG2_87352 | 1 | 2 | 4 | 2 | 5 | 9 | 1 |
| PFB0325c        | cysteine protease | 78     | III       | 0.5 | 1 | Pept_C1 | OG2_93012 | 6 | 0 | 0 | 16 | 23 | 1 |
| PFC0180c        | membrane skeletal protein | 98    | III       | 0.8 | 1 | TSP1 | OG2_88543 | 6 | 1 | 0 | 9 | 25 | 1 |
| MALBPI.73       | hypothetical protein | 141  | III       | 3.3 | 1 | Plasm. | OG2_88418 | 6 | 2 | 0 | 16 | 13 | 1 |
| PFA0205w        | hypothetical protein | 93    | III       | 4.0 | 1 | Plasm. | OG2_92998 | 6 | 0 | 0 | 7 | 22 | 0.97 |
| PF13_0234       | phosphoenol pyruvate carboxykinase | 66   | III       | 1.0 | 1 | PEPCK_ATP | OG2_72511 | 6 | 3 | >10 | 8 | 17 | 0.93 |
| PF08_0008       | hypothetical protein | 85    | III       | 1 | 1 | 2.0 | OG2_88440 | 6 | 1 | 0 | 3 | 22 | 0.93 |
| PFL1165w        | hypothetical protein | 96    | III       | 0.8 | 1 | Plasm. | OG2_93001 | 6 | 0 | 0 | 15 | 6 | 1 |
| MAL13P1254      | hypothetical protein | 31    | III       | 1 | 2.0 | Plasm. | OG2_92954 | 6 | 0 | 0 | 11 | 9 | 1 |
| PF14_0495       | hypothetical protein | 249   | III       | 1 | 2.0 | Plasm. | OG2_78524 | 6 | 3 | 5 | 12 | 6 | 1 |
| PF14_0286       | glutamate dehydrogenase | 57   | III       | 1 | 1 | 0.5 | ELFV_dehydrog N, ELFV_dehydrog | OG2_70937 | 6 | 1 | >10 | 3 | 12 | 1 |
| MALBPI.66       | hypothetical protein | 26    | III       | 1 | 1.3 | 1 | Plasm. | OG2_72432 | 5 | 3 | >10 | 5 | 4 | 1 |

1Genes shown in bold been further characterized by functional analysis through targeted gene disruption (see Table 4).
2Classification of proteins based on their protein expression patterns: group I (ODS-enriched), group II (SGS-enriched), and group III (ODS/SGS-enriched).
3Abbreviations used for motif predictions: are Signal Peptide (SP), Glycosylphosphatidylinositol anchored proteins (GPI), Trans Membrane regions (average of number of TM regions predicted by four algorithms) and N-terminal PEXEL (Plasmodium export element within first 200 amino acids obtained from http://www.plasmodb.org). Abbreviations for predicted SMART domains: Apical membrane antigen (AMA), Complement control proteins (CCP), Domain of Unknown Function (DUF), Glu/Leu/Phe/Val dehydrogenase dimerisation domain (ELFV_dehydrog_N), Glutamate/Leucine/Phenylalanine /Valine dehydrogenase (ELFV_dehydrog). Domain containing 5 cysteine conserved residues (GCC2_GCC3), Metal dependent phosphohydrolases with conserved ‘HD’ motif (HDc), Membrane-attack complex /perforin domain family (MACPF), Phosphoenolpyruvate carboxykinase (PEPCK_ATP), Papain family cysteine protease (Pept_C1), Sexual stage antigen s48/45 domain (s48-45), Thrombospondin type 1 (TSP1), Triose-phosphate Transporter family (TPT), UAA transporter family (UAA), Von Willibrand factor type A domain (VWA).
4Orthology prediction from OrthoMCL database (http://www.orthomcl.org) shows OrthoMCL group, nr species Plasm (the number of Plasmodium species, additional nr species Api (number of additional Apicomplexan species without Plasmodium species), additional species other (number of species without Apicomplexan species (including Plasmodium)).
5OOC - oocysts; ODS - oocyst-derived sporozoites; SGS - salivary gland sporozoites. The mosquito fraction (mosq fr.) of a protein is calculated by the number of detected peptides in mosquito stages divided by the sum of the number of peptides in mosquito and blood stages.

doi:10.1371/journal.ppat.1000195.t003

Table 3. Cont.
Table 4. Genes selected for further functional analysis of their encoded proteins through targeted disruption of their orthologs in the *P. berghei* rodent model.

| Protein | P. falciparum | P. berghei accessions | Exp nr (mutant) | Successful disruption | Oocyst no. mean (s.d.) (2–4 exp) | spor. no. mean (s.d.) (2–3 exp) | spor. infectivity to mice (mosquito bite) | Phenotype | Remarks |
|---------|---------------|-----------------------|-----------------|-----------------------|-------------------------------|-------------------------------|----------------------------------|-----------|---------|
| Control (WT) | - | - | - | 169 (39) | 104000 (14500) | 4 (n = 4) | wild type | - |
| PF11_0528 | I | PB001073.00.0/PB00529.02.0/PB00863.00.0 | 800, 836 | Yes (2 ×) | 160 (48) | 64800 (12500) | 2 (n = 2) | Wild type | Redundant function? |
| PF14_0435 | I | PB101363.00.0/PB00829.02.0/PB105739.00.0 | 802, 838 | Yes (2 ×) | 181 (60) | 0 | 0 (n = 4) | Complete block of sporozoite development inside oocyst |
| PF14_0607 | I | PB000015.03.0 | 803, 829, 839 | No (3 ×) | - | - | - | Essential for blood stage? |
| PF14_0074 | II | PB000372.00.0 | 807, 842 | Yes (2 ×) | 153 (27) | 80200 (16000) | 2 (n = 2) | Wild type | Redundant function? |
| PF01198c | II | PB107027.00.0/PB107193.00.0/PB001101.03.0 | 801, 837 | Yes (2 ×) | 165 (47) | 97300 (12200) | 2 (n = 2) | Wild type | Redundant function? |
| PF0425w | III | PB000251.01.0 | 806, 841 | Yes (2 ×) | 154 (59) | 375 (125) | 0 (n = 4) | Arrest of sporozoite egress from oocyst |
| PF0425w | III | PB301531.01.0 | 809, 830, 844 | No (3 ×) | - | - | - | Essential for blood stage? |
| MAL8P1.66 | III | PB402680.00.0 | 808, 843 | Yes (2 ×) | 188 (51) | 2625 (415) | 0 (n = 4) | (Partial) block of sporozoite development inside oocyst |

1 Classification of proteins based on their protein expression patterns. Group I ODS-enriched, group II SGS-enriched, and group III ODS/SGS-enriched.
2 The experiment numbers (= mutant number) of 2–3 independent transfection experiments.
3 Successful disruption was determined by Southern analysis of separated chromosomes and diagnostic PCR of selected, pyrimethamine resistant parasites as shown in Figure S3 and Table S5.
4 Mean number of oocysts per mosquito and standard deviation (s.d.).
5 Mean number of salivary gland sporozoites per mosquito and standard deviation (s.d.).
6 Infectivity of sporozoites was tested by infecting mice by mosquito bite. The number of mice that showed blood stage infection comparable to wild type infections is shown.
7,8 See Figure 3 for details of the characterization of the 3 mutants with a phenotype different from wild type.

OOC – oocysts; ODS – oocyst-derived sporozoites; SGS – salivary gland sporozoites.
parasite through the mosquito that is reflected in changes of its protein repertoire.

Analysis of the ‘stage specificity’ of proteins in six different life cycle (mammalian and mosquito) stage proteomes demonstrated that expression of proteins restricted to a single stage ranges from 12 to 28% with the highest percentage of ‘stage specificity’ in the gametocyte and reaching 24% in ODS. The 478 proteins common to blood and mosquito stages are significantly enriched in house keeping proteins involved in metabolic processes. The absence of specific enrichment of GO annotations in the 250 proteins of the mosquito stage specific proteome can most likely be ascribed to the fact that a relatively small number of these proteins posses a GO designation. Many of the mosquito stage specific proteins are still annotated as hypothetical and probably have functions that are specific for sporozoites and/or Plasmodium. This concept is supported by the observation that 15 of the 23 Plasmodium proteins known to have a sporozoite specific function are present in the 250 mosquito stage proteins identified in this study, a 4–5 fold enrichment. Moreover, their stage specific expression in our different proteomes also confirms that in general the timing of protein expression coincides with observation of function as inferred from gene deletion studies. For example, proteins involved in the traversal and invasion of the hepatocyte (e.g. SPECT1/2, CelTOS, AMA-1, STARp, TRSP, P36p and P36 (Table 1)) are either exclusively or much more highly expressed in SGS than ODS. Such changes in protein composition and abundance demonstrate that sporozoites go through dynamic changes and may exist as clearly defined developmental stages – currently ODS and SGS – that express stage specific proteins. These clear differences seem unexpected in the light of the morphological similarity of the two stages but on the other hand are in good agreement with the significant rise in mammalian host infectivity observed during the maturation and migration of sporozoites from oocysts to salivary glands [20,39]. These changes are not only restricted to proteins directly involved in these processes, but extend also to enzymes implicated in metabolic housekeeping processes such as glycolysis, production of NADPH and the TCA cycle that might be expected to coincide with subcellular reorganisation at the level of the organelles. Mature, salivary gland sporozoites might be considered to be in the resting phase (G0) of the cell cycle and are able to persist and remain infectious within the salivary glands of the mosquito for the remainder of its life. Therefore, the abundance and storage of these proteins may suggest that the salivary gland sporozoite contains stockpiles of proteins which are deployed only upon activation in the vertebrate host and growth (G1) and multiplication (S, M phases) inside the hepatocyte. Alternatively, some of these proteins could specifically be required by the parasite in the salivary glands of the mosquito host and therefore do not depend on activation in the vertebrate host.

Protein and gene expression studies of SGS have previously been performed in *P. falciparum* [10,16] as well as for the rodent parasites *P. berghei* [11,20,21] and *P. yoelii* [18,19]. The relatively low overlap between the proteins detected in the various proteomes of sporozoites can in part be ascribed to the difficulties in collecting material of sufficient purity and quantity. This limitation results in the frequent sequencing of peptides derived from mosquito proteins which reduces the total number of identified parasite proteins. However, both the degree of overlap between the proteomes and the degree of certainty in protein calling can be improved if more strict selection criteria are used for protein calling [31]. When we compared only proteins that were identified by at least 2 or more fully-tryptic peptides in all datasets (i.e. ours, Flores [10] *P. falciparum* SGS and Hall [11] *P. berghei* SGS) we found a greater than 50% overlap in proteins. Moreover, in the Hall *P. berghei* SGS and OOC proteomes it is observed that more than 80% of these proteins have a direct ortholog in *P. falciparum*. Further, when we again only compare ‘fully tryptic proteomes’ we find 75% of the *P. berghei* SGS proteins are also
expressed in the SGS of *P. falciparum* indicating that sporozoites of different *Plasmodium* species employ similar processes of maturation and invasion. Despite the relatively low overlap in total numbers of proteins detected in the different proteomes, there is a good correlation of protein abundance between our SGS proteome and the previously reported SGS proteome of *P. falciparum* [10] based on peptide counting methods.

Interestingly, nearly all the expressed genes of *P. yoelii* sporozoites detected by EST analyses [18] are also present in different proteome stages we find that all of our 601 (i.e. 601 of the 728 genes that have a *P. falciparum* ortholog) mosquito stage specific *P. falciparum* proteins are also detected as mRNA [22].

We found a lower percentage of shared proteins between our proteome and the transcripts detected in sporozoites of *P. berghei* [11,20,21]. The variation in overlap between the various proteome and transcriptome studies is certainly influenced by the varying and often small number of identified genes/proteins and indicates that a comprehensive expression profile of the salivary gland sporozoite has still to be realized. Comparison of mRNA species detected in *P. falciparum* SGS [22] with our proteome showed that for a large percentage of genes, mRNA production coincides with the presence of its protein (463 mRNA species for 477 proteins; 97%). The simultaneous presence of transcripts and protein expression has also been observed during blood stage development, supporting the 'just in time' model [9]. However, more than 2100 genes demonstrate an up-regulation of transcription in sporozoites [22], many of which were not detected as a protein in the various proteome studies. Moreover, a low correlation exists between the abundance levels of our SGS proteins (i.e. by emPAI) and the mRNA abundance of previously reported large-scale SGS transcriptome studies (i.e. r = 0.31–0.33; [16,22]). This is in line with the observations made by Le Roch et al (2004) where transcript levels are not always well correlated with same stage protein expression, suggesting a delay between mRNA and protein accumulation [49]. It is interesting to speculate whether these differences in expression between RNA and protein could be in part explained by translation repression as is observed in gametocytes that contain pools of translationally repressed transcripts that are only translated following zygote formation [11,50,51]. However, as discussed above, the proteome of sporozoites may not be comprehensive enough to draw conclusions about the relationship between specific mRNA and protein expression patterns.

The sporozoite proteomes, despite not being exhaustive, provide for the first time information on parasite protein expression both at the mosquito midgut and salivary gland stages. This has allowed for the identification of hitherto uncharacterized proteins which in turn has informed the selection of genes for targeted orthologous gene disruption studies in the rodent malaria parasite, *P. berghei*. Mutant *P. berghei* parasites lacking mosquito stage specific proteins have proven to be an efficient way to obtain an understanding into the function of such proteins [32]. We were able to generate 6 mutants for 8 hypothetical proteins that were selected from our proteomes for further functional analysis in *P. berghei* of which 3 showed distinct phenotypes, demonstrating an important and essential role of these proteins in sporozoite development and maturation. The knock-out parasite lines of 3 genes that do not exhibit a clear phenotypic difference from wild type parasites indicate either a redundancy in function for the proteins encoded by these genes or else phenotypes that are presently too subtle for us to detect with our current methodologies. However, functional redundancy is a well-established phenomenon for a number of *Plasmodium* proteins that are expressed in the blood and sexual stages of the parasite [53–55].

The *P. falciparum* protein PFD0425w is highly and exclusively expressed in sporozoites obtained from the oocyst stage and the phenotype of the orthologous gene knock-out mutant in *P. berghei*, 802cl1, is an abnormal development of the oocyst and the complete absence of sporozoite production. This example demonstrates not only the validity of the orthologous gene studies in *P. berghei* but also the informative power of this combination of proteome-reverse genetic approach in the characterization of proteins at discreet stages of the parasite life-cycle. Furthermore, the number of oocysts produced by the 802cl1 mutant is not different from wild type levels and a defect appears to occur prior to sporozoite development indicating that the role of PFD0425w is upstream of sporozoite production. The phenotype of a second *P. berghei* mutant, line 841cl1, which lacks the orthologue of PFD0425w closely resembles the egress defects observed with the cysteine protease ECP1 (or SERA8 in *P. falciparum*) and CS mutants that are mutated in their thrombospondin repeat; where sporozoites are unable to exit from midgut oocytes [44,56]. Although ECP1 mutant sporozoites are not infectious, it has been suggested that ECP1 may be involved in the cleavage of CS and thereby release of sporozoites from the oocyst [4]. Interestingly, while oocyst-derived sporozoites that lack ECP1 or express mutated CS are not infective to mice, the mechanically liberated oocyst-derived sporozoites of mutants lacking PFD0425w are able to establish an infection in mice by i.v. inoculation, and this implies that PFD0425w - in contrast to ECP1 - has no additional function during infection of the mammalian host. Its role appears to be restricted and directly involved in sporozoite release from the oocyst and has a more immediate/causative function in the release of sporozoites from oocysts. In contrast, the protein MAL8P1.66 appears to have multiple roles during sporozoite development within the oocyst and infectivity to the mammalian host. Mutants lacking this protein (i.e. line 843) are affected in the production of sporozoites within oocysts. However, the low numbers of sporozoites formed are able to invade salivary glands and hepatocytes in *vivo* but are unable to infect mice, suggesting an additional role during further development inside the hepatocyte. Interestingly and in line with the expectation, the expression of MAL8P1.66 has recently been identified in liver stage of *Plasmodium* [57]. The exact role during development of the liver stages awaits further analysis.

This study sheds light not only on the development and maturation of the malaria parasite in an *Anopheles* mosquito but also identifies proteins that are uniquely synthesized as the sporozoite becomes increasingly infectious to humans. Infection initiated by injection of *P. falciparum* sporozoites into humans represents the culmination of many precise, sequential and critical developmental steps of the malaria parasite through the mosquito. Moreover, transmission is a bottleneck in the life cycle of *Plasmodium* and the full maturation of sporozoites is essential in the survival of the parasite. The changes in the different sporozoite proteomes documented here emphasize that each event from oocyst development to egress and invasion of salivary gland and injection is tightly regulated. Intervention studies are now being conducted that aim to exploit the tightly regulated pathways that the parasite has evolved to ensure transmission. This has been recently demonstrated with the use of genetically attenuated sporozoites that have rapidly become an important focus in the development of new vaccines. The disruptions of individual genes that encode sporozoite proteins sufficiently weakens the parasite such that development in the liver is blocked, enabling the mammalian host to generate a strong protective immunity against...
subsequent infection. Clearly, the targeted disruption of genes encoding proteins identified in this study, which are involved in essential mature sporozoites functions, namely hepatocyte traversal, invasion and intracellular survival may also accelerate the identification of new protective attenuated parasite lines. Understanding the sporozoite and all its various developmental steps during the establishment of an infection continues to represent a promising approach in the hunt for new weapons in the fight against malaria.

Materials and Methods

Collection of *P. falciparum* oocysts and sporozoites

*Anopheles stephensi* mosquitoes (Sind-Kasur strain, 3–5 days old) [58] were infected with *P. falciparum* gametocytes (NF54) [59] by membrane feeding. Unfed and partially fed mosquitoes were removed and fully fed mosquitoes were kept at 26±1°C at 80% humidity. After one day, a 5% glucose solution soaked in cotton wool was offered to the mosquitoes and mosquitoes were allowed to take an extra (uninfected) blood meal at day 8–10 after infection [60]. Oocysts and oocyst-derived sporozoites were collected from midguts at 7–8 and 13–14 days after infection, respectively. Approximately 100–200 mosquito midguts were hand-dissected and homogenized in a home made glass tissue grinder in 200 μl of PBS pH 7.2 at 4°C. Salivary gland sporozoites were collected from salivary glands 18–22 days after infection. Approximately 70 salivary glands were hand-dissected and treated in a similar way as the oocyst samples. For the parasite preparations (OOC, ODS and SGS), four, three and two batches respectively were generated and processed further for analysis by nLC-MS/MS.

Sample preparation for Mass Spectrometry analysis

In order to estimate the number of sporozoites in the samples described above the total number of oocyst and salivary gland sporozoites per mosquito was determined as follows: midguts and salivary glands were dissected from 10 mosquitoes at day 13 and day 22 after feeding respectively. The midguts/salivary glands were homogenized in a home made glass chamber in 1000 μl of PBS pH 7.2 and sporozoites were counted in a Burker-Türk counting chamber using phase-contrast microscopy (1–1.6×10^6 sporozoites obtained from salivary glands of one mosquito, and 0.5–5×10^3 sporozoites per mosquito midgut). Parasites samples from mosquito midguts and salivary glands (approx. 1–4×10^7 ODS and SGS sporozoites, and 1–2×10^6 oocysts from 65–200 mosquito midguts) were divided into a soluble and insoluble fraction by a freeze–thawing procedure similar to the oocyst samples. For the parasite preparations (OOC, ODS and SGS), four, three and two batches respectively were generated and processed further for analysis by nLC-MS/MS.

Identified peptide count analysis to determine protein abundance index values

To determine the protein abundance in our samples, mass spectrometric data was analyzed using an identified peptide per protein count analysis to compute the exponentially modified Protein Abundance Index (emPAI) values [16,22]. EmPAI values for all proteins in Table S1 were calculated as 10^PAI-1 (PAI = nobserved peptides/nobservable peptides). The number of ‘observable’ peptides per protein was calculated from the output of the program Protein Digestion Simulator (http://ncrr.pnl.gov/), Protein Abundance Index and duty-cycle enhancement. Since the parasite samples were contaminated with host (mosquito) proteins, we measured samples up to four times with exclusion lists to acquire MS/MS spectra of *P. falciparum* peptides. Peptides sequenced in the first run were excluded for sequencing in subsequent runs, peptides from the 2nd run were excluded in the 3rd run etc. This procedure results in an enrichment of low abundant peptides in the second, third and fourth LC-MS/MS run. In total, more than 750 LC-MS/MS runs were acquired resulting in at least 200,000 MS/MS spectra per parasite stage. *Plasmodium* proteins were identified by searching combined protein databases of *P. falciparum* (http://www.plasmodb.org), *Anopheles gambiae* (http://ftpensembl.org/pub/) and human IPI (ftp://ftp.ebi.ac.uk/pub/databases/IPI/) using the Mascot search algorithm (Matrix Science) with tryptic requirement and 0.2 Da mass tolerance for precursor mass and fragment masses. First ranked peptides (Mascot peptide scores>15) were parsed from Mascot database search html-files with MSQuant (http://www.mquant.sourceforge.net) to generate unique first ranked peptide lists. *Plasmodium* proteins identified by 1–3 three first ranked peptides were verified by manual inspection of the MS/MS spectra in MSQuant or in Mascot. An initial validation filter was applied to the dataset after reversed database searches. A minimal Mascot peptide score of 30 was determined by a reverse database search, which revealed a false positive rate of 17% for proteins identified by 1 peptide with a Mascot peptide score>30, delta score>5, 5% for proteins identified by 2 peptides (average Mascot score>30) and 0.3% for proteins identified by more than 2 peptides. Manual verification for proteins detected by less than 4 peptides substantially decreased the false positive rates and included proteins below this filter. After internal calibration of the peptide masses by MSQuant, an average absolute mass accuracy of 23.5 ppm was obtained for the entire dataset of *P. falciparum* peptides. To remove redundancy on the protein level and to uniquely assign peptides to one protein, the peptides were remapped to PlasmoDB 5.3 annotated genome using the program Protein Coverage Summarizer (http://ncrr.pnl.gov/software/). The collected peptide list of this study (malaria peptides identified in the mosquito stages) is available in Table S1.

Nanofluidic chromatography tandem mass spectrometry

The nLC-MS/MS procedure as described for the analysis of blood stages [13] was used with minor adjustments. Peptide mixtures were loaded onto 100 μm ID columns packed with 3 μm C18 particles (Vydac) and eluted into a quadrupole time-of-flight mass spectrometer (QSTAR, Sciex-Applied Biosystems). Fragment ion spectra were recorded using information-dependent acquisition and duty-cycle enhancement. Since the parasite samples were contaminated with host (mosquito) proteins, we measured samples up to four times with exclusion lists to acquire MS/MS spectra of *P. falciparum* peptides. Peptides sequenced in the first run were excluded for sequencing in subsequent runs, peptides from the 2nd run were excluded in the 3rd run etc. This procedure results in an enrichment of low abundant peptides in the second, third and fourth LC-MS/MS run. In total, more than 750 LC-MS/MS runs were acquired resulting in at least 200,000 MS/MS spectra per parasite stage. *Plasmodium* proteins were identified by searching combined protein databases of *P. falciparum* (http://www.plasmodb.org), *Anopheles gambiae* (http://ftpensembl.org/pub/) and human IPI (ftp://ftp.ebi.ac.uk/pub/databases/IPI/) using the Mascot search algorithm (Matrix Science) with tryptic requirement and 0.2 Da mass tolerance for precursor mass and fragment masses. First ranked peptides (Mascot peptide scores>15) were parsed from Mascot database search html-files with MSQuant (http://www.mquant.sourceforge.net) to generate unique first ranked peptide lists. *Plasmodium* proteins identified by 1–3 three first ranked peptides were verified by manual inspection of the MS/MS spectra in MSQuant or in Mascot. An initial validation filter was applied to the dataset after reversed database searches. A minimal Mascot peptide score of 30 was determined by a reverse database search, which revealed a false positive rate of 17% for proteins identified by 1 peptide with a Mascot peptide score>30, delta score>5, 5% for proteins identified by 2 peptides (average Mascot score>30) and 0.3% for proteins identified by more than 2 peptides. Manual verification for proteins detected by less than 4 peptides substantially decreased the false positive rates and included proteins below this filter. After internal calibration of the peptide masses by MSQuant, an average absolute mass accuracy of 23.5 ppm was obtained for the entire dataset of *P. falciparum* peptides. To remove redundancy on the protein level and to uniquely assign peptides to one protein, the peptides were remapped to PlasmoDB 5.3 annotated genome using the program Protein Coverage Summarizer (http://ncrr.pnl.gov/software/). The collected peptide list of this study (malaria peptides identified in the mosquito stages) is available in Table S1.
Correlation between protein expression data from different studies

Values for the level (abundance) of protein expression from different datasets were obtained for all individual proteins by calculated emPAI values. EmpAI values and mRNA levels of microarray analyses were log2 transformed before regression analysis to obtain normal distributions. Pearson correlation between datasets was performed using R (http://www.r-project.org/).

Gene ontology annotation

Gene Ontology SLIM terms were assigned using “Generic GO (http://go.princeton.edu/cgi-bin/GOTermMapper). A GO enrichment analysis for ‘Biological Process’, ‘Cellular Component’ and ‘Molecular Function’ using default GO association files was performed with “GO Term Finder” (http://go.princeton.edu/cgi-bin/GOTermFinder) where statistical significance (p-value) is calculated based on hypergeometric distribution with Bonferroni multiple testing correction and false discovery rate calculation as described [62]. To perform a GO enrichment analysis with adjusted GO association files, Ontologizer (http://www.charite.de/ch/medgen/ontologizer/) was used where statistical significance (p-value) is calculated as in ‘GO Term Finder’[see above and [63]].

Selection criteria for ODS, SGS and ODS/SGS proteins

Proteins with more than 90 percent of the peptides detected in the mosquito stages (mosquito fraction>0.9) were divided into three groups (OOC, ODS and ODS/SGS) based on their expression patterns. The mosquito fraction equals \( n_{\text{mosq}} / (n_{\text{mosq}} + n_{\text{blood}}) \) where \( n \) is the number of unique peptides per protein at the mosquito and blood stages, respectively. The mosquito enriched proteins were further subdivided into 112 ODS-specific proteins expressed in the ODS stage and not SGS (Group I); 74 SGS-specific proteins not expressed in ODS (Group II); and finally 59 Group III proteins that are shared between several mosquito life cycle stages. A further refinement of these groups was based on the following criteria. Only proteins with more than two peptides detected in the mosquito stages \( n_{\text{mosq}} \geq 3 \) were considered. In addition, only proteins were selected that contained Signal peptide (SP), Transmembrane (TM) or Glycosylphosphatidylinositol (GPI) domains and combinations of these motifs. Sequence-based prediction data for these domains was retrieved from PlasmoDB (http://www.plasmodb.org) for SP and TM domain predictions based on TMHMM, TMAP, TMHMM2 and TOPRED2 algorithms; from http://gpi.umibe.ch/ for GPI predictions by a Kohonen Self Organizing Map; and from http://smart.embl-heidelberg.de/ for SMART protein domain searches. The number of TM domains is the average of four values obtained from the different TM prediction algorithms. Different criterions were set for combinations of predicted motifs. For less abundant proteins without predicted signal peptide (SP = 0, and 3≤\( n_{\text{mosq}} \leq 15 \), only proteins with at least 4 predicted TM regions were included (average TM>4). For abundant proteins without signal peptide (SP = 0, \( n_{\text{mosq}} > 15 \) ) proteins with at least 0.5 predicted TM regions (average TM>0.5) were included. For proteins with predicted signal peptide (SP = 1), all proteins with at least 0.5 predicted TM regions (average TM>0.5) were included. Finally, all proteins with a predicted GPI anchor (GPI = 1) were selected independent of the presence of predicted signal peptide or TM regions.

Generation and characterization of gene knockout P. berghei parasite mutants

Eight P. falciparum proteins were selected for functional analysis by targeted gene disruption of their corresponding orthologs in P. berghei. The sequences of the eight P. berghei gene orthologs (as well their corresponding up and downstream sequences) were retrieved from the on-line Plasmodium genome databases, http://www.plasmodb.org and http://www.genedb.org/genedb/Pberghei. For P. berghei genes with incomplete sequence information in the database (4 out of 8), the complete genes were manually assembled from a number of different P. berghei sequences by performing BLAST sequence searches of the full length P. falciparum genes against the P. berghei genome and closing gaps by PCR and DNA sequencing (see for details Figure S3 and Table S3). Standard plasmid vectors were designed for targeted gene disruption by double cross-over homologous recombination [48]. To replace the protein coding sequences of the target genes with the dhfr/ts pyrimethamine resistance marker from Toxoplasma gondii, we cloned the 5′ and 3′ flanking regions of the gene of interest up- and downstream of the selection cassette of p0001; also in MR4 (http://www.m4r.org/). Briefly described for one candidate gene, to generate a PB0000829.02.0/PF14_0435 disruption vector, an upstream region (position 74–436 on singleton berg-2274h02.p1.k) and a downstream region (position 516–1016 on contig PB_PP2658) - the latter containing the P. berghei orthologous gene PB0000829.02.0 - were amplified from P. berghei genomic DNA using primer-pairs 2666–2653 and 2654–2655, respectively. The PCR products were digested with Aap718 and HindIII, or EcoRI and NolI, respectively, and ligated into plasmid p0001 yielding targeting plasmid pL1175 (see for further details of all plasmids and the sequence of the primers, Figure S3 and Table S5). All plasmids generated were sequence analysed. Transfection of GFP-expressing ‘wild type’ parasites from the P. berghei reference line 507cl1 [64] with linearised targeting constructs, selection and cloning of the mutant parasites were performed according to procedures previously described [48]. Genotypic analysis of transfected parasites was performed by Southern analysis of FIGE separated chromosomes and diagnostic PCR on genomic DNA (details of the primers used for PCR are shown in Table S5).

Phenotype analysis of mutant parasites during blood stage development, quantification of gametocyte production and ookinete development in vitro was performed using standard methods as previously described [14, 65]. Mosquito stage development was analysed in A. stephensi mosquitoes using standard methods of infection of mosquitoes and analysis of oocyst and sporozoite production and analysis of sporozoite infectivity to C57Bl6 mice [66]. The number of sporozoites in oocysts of mosquito midguts and in salivary glands derived from 10 mosquitoes was determined in quadruplicate as described above for counting P. falciparum sporozoites and represented as mean number with standard deviation per stage per mosquito. The capacity of the mutant parasites to infect mice by mosquito interrupted feeding was determined by exposure of female C57Bl6
antibody 3D11 against CS was used. 2 hours after adding sporozoites to HepG2 cells, and compared to determined by counting the percentage of dextran-positive cells in infection of wild type sporozoites. Hepatocyte cell traversal was monoclonal antibody 2E6 against HSP70 [69] and compared to HepG2 cells, infection was quantified by staining EEFs with fully in wounded dextran-positive cells [68]. After fixation of the dextran-negative cells because parasites do not develop successfully exoerythrocytic forms (EEFs) of the parasite. Hepatocyte invasion for an additional 24 hours to determine the development of HepG2 cells in the presence of fluorescent dextran for 2 hr, cell traversal and invasion was accomplished by using a cell-membrane marker molecule, rhodamine-dextran at 1 mg/ml that will visualize parasitized wounded cells specifically but not uninfected HepG2 cells. Sporozoites were incubated with HepG2 cells in the presence of fluorescent dextran for 2 hr, followed by washing the cells to remove the marker and incubation for an additional 24 hours to determine the development of exoerythrocytic forms (EEFs) of the parasite. Hepatocyte invasion was determined by counting the percentage of sporozoites inside dextran-negative cells because parasites do not develop successfully in wounded dextran-positive cells [68]. After fixation of the HepG2 cells, infection was quantified by staining EEFs with monoclonal antibody 2E6 against HSP70 [69] and compared to infection of wild type sporozoites. Hepatocyte cell traversal was determined by counting the percentage of dextran-positive cells 2 hours after adding sporozoites to HepG2 cells, and compared to wild type sporozoite cell traversal. In this procedure, monoclonal antibody 3D11 against CS was used.

In vitro hepatocyte traversal and invasion experiments were performed as described elsewhere [67,68] by adding purified sporozoites (5×10⁴) to confluent monolayers of HepG2 cells in DMEM medium (note: medium had 10% FCS and 1% PenStrep). Mutant sporozoites were obtained as described above from either oocysts (day 20) or from salivary glands (day 27). Quantification of cell traversal and invasion was accomplished by using a cell-impermeable fluorescent marker molecule, rhodamine-dextran at 1 mg/ml that will visualize parasitized wounded cells specifically but not uninfected HepG2 cells. Sporozoites were incubated with HepG2 cells in the presence of fluorescent dextran for 2 hr, followed by washing the cells to remove the marker and incubation for an additional 24 hours to determine the development of exoerythrocytic forms (EEFs) of the parasite. Hepatocyte invasion was determined by counting the percentage of sporozoites inside dextran-negative cells because parasites do not develop successfully in wounded dextran-positive cells [68]. After fixation of the HepG2 cells, infection was quantified by staining EEFs with monoclonal antibody 2E6 against HSP70 [69] and compared to infection of wild type sporozoites. Hepatocyte cell traversal was determined by counting the percentage of dextran-positive cells 2 hours after adding sporozoites to HepG2 cells, and compared to wild type sporozoite cell traversal. In this procedure, monoclonal antibody 3D11 against CS was used.

Accession Numbers

All datasets will become available through the official Web site of the Plasmodium genome project, PlasmoDB [http://www.plasmodb.org] [70,71]]. In the text and tables most genes and gene products are accompanied with their PlasmoDB Accession Number.

The PlasmoDB accession numbers for other genes and gene products discussed in this paper are for *P. falciparum* CS (PF00210c), TRAP (PF13_0201), UIS3 (PF13_0012), P36 (PF00210c), P36p (PF00215c), myosin A (PF13_0233), MTIP (PF12225w), actin (PF1221w) and F-1,6-BP aldolase (PF14_0245), AMA-1 (PF11_0344), TRSP (PF0A000w), RESA3 (PF00235c), SPECT1 (MAL13P1.212), SPECT2 (PF00430c), CcITOS (PF00800c), STAR (PF07_0006), and for *P. berghei*: UIS4 (PB100551.00.0), ECPI (PB000649.01.0).

The sequences of the eight *P. berghei* gene orthologs (as well their corresponding up and downstream sequences) that have been analysed in gene-disruption studies were retrieved from the PlasmoDB database [http://www.plasmodb.org] and from the GeneDB database [http://www.genedb.org/].

For 4 out of 8 *P. berghei* genes with incomplete sequence information in the database, the complete genes were manually assembled from a number of different *P. berghei* sequences by performing BLAST sequence searches, PCR and DNA sequencing (see for details Figure S3 and Table S3). Primer sequences used in contig gap closure and location of primers relating to contigs and reads of the revised *P. berghei* gene models have been submitted to GenBank and are provided in Figure S3 and Table S3.

Supporting Information

Figure S1 Gene Ontology (GO) annotation for proteins from proteomes from two mosquito stages of *P. falciparum*, oocyst-derived sporozoites and salivary gland sporozoites. Found at: doi:10.1371/journal.ppat.1000195.s001 (0.05 MB DOC)

Figure S2 Pathway profiling with the number of unique peptides/protein detected in 5 different life-cycle stages (data obtained from this study and from Lasonder et al. [15]). Found at: doi:10.1371/journal.ppat.1000195.s002 (0.09 MB DOC)

Table S1 Peptides and proteins of *P. falciparum* identified in proteomes of oocysts, oocyst-derived sporozoites and salivary gland sporozoites. Page ‘Peptides’: peptides identified by nLC-MS/MS. Information provided in the table: 1) life cycle stage, 2) peptide sequence, 3) MCR (mass of charge of parent ion), 4) charge of parent ion, 5) measured mass of peptide, 6) calibrated mass of peptide after internal mass calibration, 7) peptide rank in Mascot searches, 8) peptide score in Mascot searches, 9) Mascot emPAI value. Page ‘Proteins’: The corresponding proteins identified by the sequenced peptides, listing: 1) accession number (PlasmoDB version 5.3), 2) protein name, 3) reannotation PlasmoDB 2008/07/15 (genes with the most recently modified annotation by PlasmoDB), 13) protein mass (molecular weight), 14) protein pl (iso-electric point), 15) number of unique peptides per protein, 16) Residue Start (residue nr in the protein sequence of the N-terminal amino acid), 17) Residue End (residue nr in the protein sequence of the C-terminal amino acid), 18) sequence coverage (percentage of the protein covered by the identified peptides), and 19) protein emPAI value. Page ‘Proteins’: The corresponding proteins identified by the sequenced peptides, listing: 1) accession number (PlasmoDB version 5.3), 2) protein name, 3) reannotation PlasmoDB 2008/07/15 (genes with the most recently modified annotation by PlasmoDB), 4) protein mass (molecular weight), 5) protein pl (iso-electric point), 6) number of unique identified peptides, 7) protein emPAI value, and 8) mosquito fraction (which is calculated by the number of detected peptides in mosquito stages divided by the sum of the number of peptides in mosquito and blood stages). Page ‘OOC’: proteins detected in OOC; column headings the same as for Page ‘Proteins’. Page ‘ODS’: proteins detected in ODS; column headings the same as for Page ‘Proteins’. Page ‘SGS’: proteins detected in SGS; column headings the same as for Page ‘Proteins’. Page ‘Mosquito stage specific’: proteins exclusively detected in mosquito stages (mosquito fraction = 1); column headings the same as for Page ‘Proteins’. Page ‘OOC-enriched’: proteins that are highly
enriched’ in OOC (mosquito fraction>0.9); column headings the same as for Page ‘Proteins’. Page ‘ODS-enriched’: proteins that are ‘highly enriched’ in ODS (mosquito fraction>0.9); column headings the same as for Page ‘Proteins’. Page ‘SGS-enriched’: proteins that are ‘highly enriched’ in SGS (mosquito fraction>0.9); column headings the same as for Page ‘Proteins’. Page ‘tRNA ligase’: overview of all tRNA proteins detected in mosquito stages, column headings the same as for Page ‘Proteins’.

Table S2 Comparison of proteins identified in the mosquito stage proteomes of this study with the proteomes of salivary gland sporozoites (SGS) of *P. falciparum* as reported by Florens and colleagues [10] and the proteomes of oocysts (OOC) and SGS of *P. berghei* as reported by Hall and coworkers [11]. The tables contain information about proteins shared by our analysis and others, where our expression data is presented by the number of unique tryptic peptides per protein. The other data sets are presented in a similar way, but a distinction is made in expression data for detection either by all peptides (tryptic, half (non) tryptic), or by tryptic peptides.

Table S3 Correlation of protein abundance (emPAI approach) identified in the SGS stage proteome of this study with mRNA levels of the SGS transcripts of *P. falciparum* as reported by Le Roch and coworkers [16] and Zhou and coworkers [22]. Tables show Pearson correlation coefficients (r), probabilities (p) and number of shared genes/proteins (n).

Table S4 Comparison of proteins identified in the mosquito stage proteomes with genes transcribed in sporozoites in *P. berghei* and *P. yoelii* as identified by either subtractive hybridization (SSH) or cDNA quantification methods (SAGE). 5 genes: 23 sporozoite (S) genes identified in a *P. yoelii* SSH screen [18]. UIS:genes: 30 UIS genes (Upregulated In Sporozoites) identified in a *P. berghei* SSH screen [20]. SIS genes: 123 SIS genes (Sporozoite expressed gene Identified by SAGE (SIS) genes) identified in a *P. berghei* SAGE analysis [21].

Table S5 Primer sequences used in this study. Primers used in KO targeting plasmid construction. Primers used to check for plasmid integration in mutant (KO) parasites. Primers used in contig gap closure and wild type (WT) PCR analysis.

Acknowledgments

We would like to thank Margar van de Vege-Bolmer, Jolanda Klaassen, Laura Pelsor-Posthumus, Astrid Pouwelsen, Jacqueline Kühnen, and Jai Ramesar for technical assistance with the preparation of parasite samples, mosquito infection and transfection experiments. We would also like to thank the Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K. for making publicly available sequence data for *Plasmodium berghei*.

Author Contributions

Conceived and designed the experiments: EL CJJ APW MM HGS. Production of *P. falciparum* and *P. berghei* KO targeting plasmid construction. Primers used to check for expression analysis [21]. Table S5: Primer sequences used in this study. Primers used in KO targeting plasmid construction. Primers used to check for plasmid integration in mutant (KO) parasites. Primers used in contig gap closure and wild type (WT) PCR analysis. Primers used in KO targeting plasmid construction. Primers used to check for plasmid integration in mutant (KO) parasites. Primers used in contig gap closure and wild type (WT) PCR analysis.

References

1. Baldacci P, Menard R (2004) The elusive malaria sporozoite in the mammalian host. Mol Microbiol 54: 298-306.
2. Revet U, Usynin I, Barf K, Klotz C (2006) Nonadecapeptide: can Kupffer cell function as portals for malaria sporozoites to the liver? Cell Microbiol 8: 1537-1546.
3. Kappe SH, Buscaglia CA, Bergman LW, Coppen I, Nussenzweig V (2004) Apicomplexan gliding motility and host cell invasion: overhauling the motor model. Trends Parasitol 20: 13-16.
4. Matuschewski K (2006) Getting infectious: formation and maturation of *Plasmodium sporozoites* in the Anopheles vector. Cell 125: 1547-1556.
5. Mota MM, Rodriguez A (2004) Migration through host cells: the first steps of *Plasmodium* motility in the mammalian host. Cell Microbiol 6: 1113-1118.
6. Sinnis P, Coppi A (2007) A long and winding road: the *Plasmodium* sporozoite’s journey in the mammalian host. Parasitol Int 56: 176-178.
7. Yuda M, Ishino T (2004) Liver invasion by malarial parasites: how do malarial parasites break through the host barrier? Cell Microbiol 6: 1119-1125.
8. Ben Mamoun C, Gluzman IY, Host C, MacMillan SK, Amarakone AS, et al. (2001) Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. Mol Microbiol 39: 26-36.
9. Boozdech Z, Llanas M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The Transcriptome of the Intracellular Developmental Cycle of *Plasmodium falciparum*. PLoS Biol 1: e5. doi:10.1371/journal.pbi.0000005.
10. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, et al. (2004) A proteomic view of the *Plasmodium falciparum* life cycle. Nature 419: 537-542.
11. Hall N, Marras M, Raine JD, Carlton JM, Kooij TW, et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 307: 82-90.
12. Haywood RE, Deris JL, Alathurai S, Kaslow DG, Brown PO, et al. (2000) Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. Mol Microbiol 35: 6-14.
13. Isokpihi RD, Hidt WA (2003) Integrative analysis of intracellularly differentially expressed transcripts yields novel insights into the biology of *Plasmodium falciparum*. Malar J 2: 38.
14. Khan SM, Franke-Fayard B, Mair GR, Laoseder E, Janse CJ, et al. (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific expression patterns. Mol Biochem Parasitol 131: 675-683.
15. Laoseder E, Ishihama Y, Andersson JS, Vermeent AM, Pain A, et al. (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature 419: 537-542.
30. van Dijk MR, Mouradina B, Franke-Fayard B, Heusler V, van Dooren MW, et al. (2005) Genetically attenuated, Pfδ6-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. Proc Natl Acad Sci U S A 102: 12194–12199.

31. Ohn JY, Ong SE, Mann M (2004) Trypsin cleaves exclusively C-terminal to Arginine and lysine residues. Mol Cell Proteomics.

32. Ishihama Y, Oda Y, Tabata T, Sato T, Naganu T, et al. (2003) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomes by the number of sequenced peptides per protein. Mol Cell Proteomics 4: 1265–1272.

33. Ishihama Y, Schmidt T, Rapplinger J, Mann M, Hartl FU, et al. (2008) Computer-assisted gene ontology annotation. Nucleic Acids Res 36: 3463–3470.

34. Morrisett NS, Smylie LD (2002) Cytoskeleton of apicomplexan parasites. Microbiol Mol Biol Rev 66: 21–38. table of contents.

35. Smylie LD (2004) Intracellular parasite invasion strategies. Science 304: 248–253.

36. Aravin I, Iyer LM, Welleris TE, Miller LH (2003) Plasmodium biology: genomic gleanings. Cell 115: 771–785.

37. Martschenko Y, Egan DA, Nussenzweig V, Menard R (2002) Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. Environ J 21: 1597–1606.

38. Hayward RE (2000) Plasmodium falciparum phosphoproteinvaccine cytotoxicity is developmentally regulated in gametocytes. Mol Biochem Parasitol 107: 227–240.

39. Srinivasan P, Abraham EG, Ghosh AK, Valenzuela J, Ribeiro JM, et al. (2004) Global proteome analysis of Plasmodium falciparum female specific surface protein in Plasmodium falciparum. Mol Biochem Parasitol 133: 25–36.

40. Janse CJ, Franke-Fayard B, Waters AP (2006) Selection by flow-sorting of different parasite life cycle stages. J Biol Chem 281: 35081–35087.

41. Sultan AA, Thathy V, Frevert U, Robson KJ, Crisanti A, et al. (1997) TRAP is essential for transmission of Plasmodium berghei to mosquitoes. Nat Protoc 1: 346–356.

42. Kariu T, Ishino T, Yano K, Chinzei Y, Yuda M (2006) CelTOS, a novel cell-passage activity is required for Plasmodium berghei. Nat Protoc 1: 614–623.

43. Menard R, Sultan AA, Cortes C, Altszuler R, van Dijk MR, et al. (1997) Malaria female gametocytes. Nucleic Acids Research 25: 30089–30090.

44. Rogers WO, Malik A, McLoek S, Nakamura K, Rogers MD, et al. (1992) Characterization of Plasmodium falciparum sporozoite surface protein 2. Proc Natl Acad Sci U S A 89: 9716–9720.

45. Sahu AA, Thathy V, Frevert U, Robson KJ, Grisanti A, et al. (1997) TRAP is necessary for glding motility and infectivity of plasmodium sporozoites. Cell 90: 511–522.

46. Jane CJ, Franke-Fayard B, Waters AP (2006) Selection by flow-sorting of genetically transformed, GFP-expressing blood stages of the rodent malaria parasite, Plasmodium berghei. Nat Proto 1: 614–623.

47. Jane CJ, Ramesar J, Waters AP (2004) High-efficiency transfection and drug screening of Plasmodium falciparum by transient gene expression. Mol Biochem Parasitol 133: 25–36.

48. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, et al. (2004) Global transcriptome analysis of Plasmodium falciparum parasites, Plasmodium berghei. Cell 119: 667–669.

49. Mota MM, Hafalla JC, Rodriguez A (2002) Migration through host cells activates Plasmodium sporozoites for infection. Nat Med 8: 1318–1322.

50. Mota MM, Pradel G, Vanderberg JP, Hafalla JC, Frevert U, et al. (2001) Migration of Plasmodium sporozoites through cells before infection. Science 291: 141–144.

51. Tsudo M, Matsuura Y, Nussenzweig RS, Eichinger D, Zavala F (1994) Demonstration of heat-shock protein 70 in the sporozoite stage of malaria parasites. Parasitology 108: 16–21.

52. Balda M, Brunk B, Coppol R, de la Osa-Cobos J, Dikmen SJ, et al. (2002) PlasmDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping sequence and gene data (both finished and unfinished). Nucleic Acids Res 30: 87–90.

53. Balda M, Brunk B, Crabtree J, Frevert U, Gajria B, et al. (2003) PlasmDB: the Plasmodium genome resource. A database integrating experimental and computational data. Nucleic Acids Res 31: 212–215.

54. Claudianos C, Desmont JM, Trueman HE, Arau M, Mendoza J, et al. (2002) A malaria scavenger receptor-like protein essential for parasite development. Mol Microbiol 45: 1473–1480.

55. Pradel G, Hayton K, Reruyer J, Iyer LM, Abraham MS, et al. (2004) A multidomain adhesion protein family expressed in Plasmodium falciparum is essential for transmission to the mosquito. Science 305: 1533–1544.

56. Khater EJ, Sinden RE, Dessens JT (2004) A malaria membrane skeletal protein is essential for normal morphogenesis, motility, and infectivity of sporozoites. J Cell Biol 167: 425–432.

57. Bergman LW, Kaiser K, Fujikawa H, Coppens I, Daly TM, et al. (2003) Myosin A tail domain interacting protein (MTIP) localizes to the inner membrane complex of Plasmodium sporozoites. J Cell Sci 116: 39–49.

58. Thompson J, Fernandez-Reyes D, Sharling I, Moore SG, Eling W, et al. (2006) Plasmodium carboxydrolase module 1 protein: 14 complex with proteins throughout the malaria parasite life cycle. Cell Microbiol 9: 1466–1480.

59. Ishino T, Yano K, Zavala F, Mota MM (2004) Cell-passage activity is required for the malaria parasite to cross the liver sinusoidal cell layer. PLoS Biol 2: e215. doi:10.1371/journal.pbio.0020021.

60. Ishino T, Zavala F, Mota MM (2005) A Plasmodium sporozoite protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection. Cell Microbiol 7: 199–208.

61. Kariu T, Ishino T, Yano K, Chinzei Y, Yuda M (2006) GCTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. Mol Microbiol 59: 1369–1379.

62. Bhartia P, Schauer K, Coppens I, Nussenzweig V (2005) A surface phospholipase is involved in the migration of plasmodium sporozoites through cells. J Biol Chem 280: 6752–6760.

63. Silve O, Franzchic JC, Charrin S, Mueller MS, Siau A, et al. (2004) A role for apical membrane antigen 1 during invasion of hepatocytes by Plasmodium falciparum sporozoites. J Biol Chem 279: 9490–9496.

64. Chattopadhayay R, Rathore D, Fujikawa H, Kumar S, de la Vega P, et al. (2003) PSMAPTR, a Plasmodium falciparum protein containing an altered thrombospondin type I repeat domain is expressed at several stages of the parasite life cycle and is the target of inhibitory antibodies. J Biol Chem 278: 25977–25981.

65. Groener AC, Brahami K, Eling W, Koningis R, Meis J, et al. (2001) The Plasmodium falciparum knob-associated PEINS antigen is also expressed at pre-erythrocytic stages and induces antibodies which inhibit sporozoite invasion. Mol Biochem Parasitol 115: 147–159.

66. Pasquotto V, Fildeck DA, Gras H, Badell E, Eling W, et al. (1997) Plasmodium falciparum sporozoite invasion is inhibited by naturally acquired or experimentally induced polyclonal antibodies to the STARP antigen. Eur J Immunol 27: 2522–2533.

67. Puentes A, Garcia J, Vera R, Lopez R, Suarez J, et al. (2004) Sporozoite and liver stage antigen Plasmodium falciparum peptides bind specifically to human hepatocytes. Vaccine 22: 1150–1156.
86. Labaied M, Camargo N, Kappe SH (2007) Depletion of the Plasmodium berghei thrombospondin-related sporozoite protein reveals a role in host cell entry by sporozoites. Mol Biochem Parasitol 153: 158–166.

87. Moreira CK, Templeton TJ, Lavazec C, Hayward RE, Hobbs CV, et al. (2008) The Plasmodium TRAP/MIC2 family member, TRAP-Like Protein (TLP), is involved in tissue traversal by sporozoites. Cell Microbiol 10: 1505–1516.

88. Baker RP, Wijetilaka R, Urban S (2006) Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. PLoS Pathog 2: e113. doi:10.1371/journal.ppat.0020113.

89. Yuda M, Yano K, Tsuboi T, Torii M, Chinzei Y (2001) von Willebrand Factor A domain-related protein, a novel microneme protein of the malaria ookinete highly conserved throughout Plasmodium parasites. Mol Biochem Parasitol 116: 65–72.