12. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419,** 537–542 (2002).
13. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).
14. Oefner, P. J. *et al.* Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res.* **24,** 3879–3886 (1996).
15. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–185 (1998).
16. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8,** 195–202 (1998).
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
18. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29,** 37–40 (2001).
19. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30,** 276–280 (2002).
20. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
21. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300,** 1005–1016 (2000).
22. Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241,** 779–786 (1996).
23. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
24. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6 (1997).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We acknowledge the generosity of the participating scientists at Stanford University, TIGR, the WTSI, the NMRC and Oxford University. We also thank N. Hall, M. Berriman, A. Pain and B. Barrell for their time and expertise during the gene-calling annotation process, and are grateful to the members of our Stanford Genome Technology Center for their assistance throughout this project. We thank the Burroughs Wellcome Fund for support that allowed us to participate in the international Malaria Genome Project.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W.H. (e-mail: hyman@sequence.stanford.edu). The GenBank accession number of the sequence of *P. falciparum* (clone 3D7) chromosome 12 is AE014188.

Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry

Edwin Lasonder*†, Yasushi Ishihama*, Jens S. Andersen*, Adriaan M. W. Vermunt†, Arnab Pain‡, Robert W. Sauerwein§, Wijnand M. C. Eling§, Neil Hall‡, Andrew P. Waters‖, Hendrik G. Stunnenberg† & Matthias Mann*

* *Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*
† *Department of Molecular Biology, NCMLS, University of Nijmegen, Geert Grooteplein 26, 6525 GA Nijmegen, The Netherlands*
‡ *The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
§ *Department of Medical Microbiology, NCMLS, University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands*
‖ *Leiden Malaria Research Group, Department of Parasitology, Centre for Infectious Disease, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands*

The annotated genomes of organisms define a 'blueprint' of their possible gene products. Post-genome analyses attempt to confirm and modify the annotation and impose a sense of the spatial, temporal and developmental usage of genetic information by the organism. Here we describe a large-scale, high-accuracy (average deviation less than 0.02 Da at 1,000 Da) mass spectrometric proteome analysis[1–3] of selected stages of the human malaria parasite *Plasmodium falciparum*. The analysis revealed 1,289 proteins of which 714 proteins were identified in asexual blood stages, 931 in gametocytes and 645 in gametes. The last two groups provide insights into the biology of the sexual stages of the parasite, and include conserved, stage-specific, secreted and membrane-associated proteins. A subset of these proteins contain domains that indicate a role in cell–cell interactions, and therefore can be evaluated as potential components of a malaria vaccine formulation. We also report a set of peptides with significant matches in the parasite genome but not in the protein set predicted by computational methods.

The *Plasmodium falciparum* parasite is pre-committed to one of three different developmental pathways on re-invasion of a host erythrocyte[4]. Either it develops asexually, resulting in proliferation, or it develops into a male or a female gametocyte—sexual precursor forms that maintain a stable G1 cell cycle arrest and circulate in the peripheral blood stream when mature. Gametocytes are activated in the mid-gut of the mosquito when ingested by the vector through the consumption of a blood meal, and rapidly develop into mature gametes that fertilize to form zygotes. The zygotes mature into a motile invasive form, the ookinete, which is adapted for colonization of the mosquito. Clearly, sexual development and fertilization are essential processes within the parasite life cycle and one strategy of vaccination (transmission blocking) seeks their interruption through antibody-based blockades. Although good candidates for such vaccines exist, it is an accepted view that an effective vaccine will need to target several stages of the parasite and several components of the different forms of the parasite[5]. High-throughput proteome studies on pure parasite forms are a rapid and sensitive means to discover such vaccine candidates.

To define the proteome of the asexual and sexual blood stages of the malaria parasite *P. falciparum* (NF54 isolate), purified asexual (trophozoites and schizonts, Fig. 1a, left panel) and sexual stage parasites (gametocytes, right panel) or gametes (not shown) were extracted by freeze–thawing and centrifugation, yielding soluble and insoluble (pellet) fractions (Fig. 1b). The result of a typical gametocyte extraction is shown in Fig. 1c, revealing that most of the *P. falciparum* and red blood cell (RBC) proteins were present in the soluble fraction, whereas membrane proteins such as the gametocyte-specific cell surface protein Pfs48/45 (ref. 6) were found exclusively in the pellet fraction, as revealed by western blotting (Fig. 1c). These complex protein mixtures were then analysed 'gel free' (differentially extracted membrane fractions) or separated into ten molecular mass fractions by one-dimensional gel electrophoresis followed by excision of equally spaced bands after precisely removing haemoglobin and globin (Fig. 1c, right panel), and tryptic digestion. The tryptic peptides were separated by reversed phase liquid chromatography coupled to quadrupole time-of-flight mass spectrometry for peptide sequencing (nanoLC-MS/MS). Iterative calibration algorithms were used to achieve a final, average absolute mass accuracy of better than 20 parts per million (p.p.m.) in both the precursor and fragment ions, or a mass deviation of 0.03 Da for a typical tryptic peptide of mass 1,300 Da.

These high-accuracy spectra were searched against a combined human and draft *P. falciparum* database, using probability-based scoring in which the fragment ions are matched against the calculated fragments of all tryptic peptides from the human and parasite sequences[7]. A total of 7,548 distinct peptides from the putative set of malaria proteins were matched with significant probability scores (Supplementary Table A). These peptides mapped to 1,709 malaria proteins. Additional constraints were applied to the peptides, including peptide size, discrimination to the next best match, and features of the tandem mass spectra such as

the dominant amino-terminal and inhibited carboxy-terminal fragmentation of proline. These criteria resulted in a list of 1,289 unique proteins that were identified with high confidence, comprising about 23% of the current proteome prediction[8] (Supplementary Table B). Roughly one-third of the malaria proteins identified appeared in both the soluble and insoluble fractions; the remaining two-thirds were present either in the insoluble or in the soluble fraction (Fig. 2a). Taken together, the analysis revealed 714 malaria proteins in asexual stage parasite preparations, 931 in gametocytes and 645 in gametes. The analyses also resulted in identification of more than 1,000 human proteins.

At present, 315 malaria proteins have been found solely in gametocytes, 226 in trophozoites and schizonts, and 97 in gametes (Fig. 2b). A total of 575 proteins were found only in sexual stages (gametocytes plus gametes) and 488 proteins were present in both asexual and sexual stages. The definitive classification of a protein as stage-specific is tentative however, and awaits full analysis of all life cycle stages. Moreover, biological samples are rarely completely pure or synchronous (see Methods). The well-characterized, sexual stage-specific surface antigen, Pfs48/45, highlights this point. Three Pfs48/45-derived peptides were detected in the asexual preparations as opposed to 57 in gametocytes. Integration of the peptide ion currents, a quantity roughly proportional to molar amount of peptide species (see Supplementary Fig. A), yielded a protein abundance ratio of more than eight between the two stages. The unexpected presence of Pfs48/45 in asexual stages is probably due to a small percentage of committed gametocytes in the asexual preparations from previous cycles of growth that cannot be physically separated from asexual-stage parasites. This interpretation is supported by western and northern blot analyses[9] and was further corroborated by quantitative polymerase chain reaction with reverse transcription (RT–PCR) analysis (Fig. 1c and Table 1). Notwithstanding the restrictions, the data are of sufficiently high quality and coverage to justify the classification of a large number of known and new proteins as 'stage-specific'. In fact, the small gametocyte contaminations in the asexual preparation strengthens the stage-specific classification of those proteins that are found exclusively in sexual stage parasites such as Pfg377, gene 11-1 and the hypothetical protein PF14_0039. A total of 517 peptides from the sexual stage antigen Pfg377 (ref. 10) were detected in the gametocytes and gamete preparations whereas not a single Pfg377 peptide was detected in the asexual stage, illustrating the exquisite sensitivity of the analytical procedure as well as the quality of the biological sample. Collectively, mass spectrometric and quantitative RT–PCR analysis (see below) indicates that expression of Pfg377, gene 11-1 and the hypothetical protein PF14_0039 is turned on later

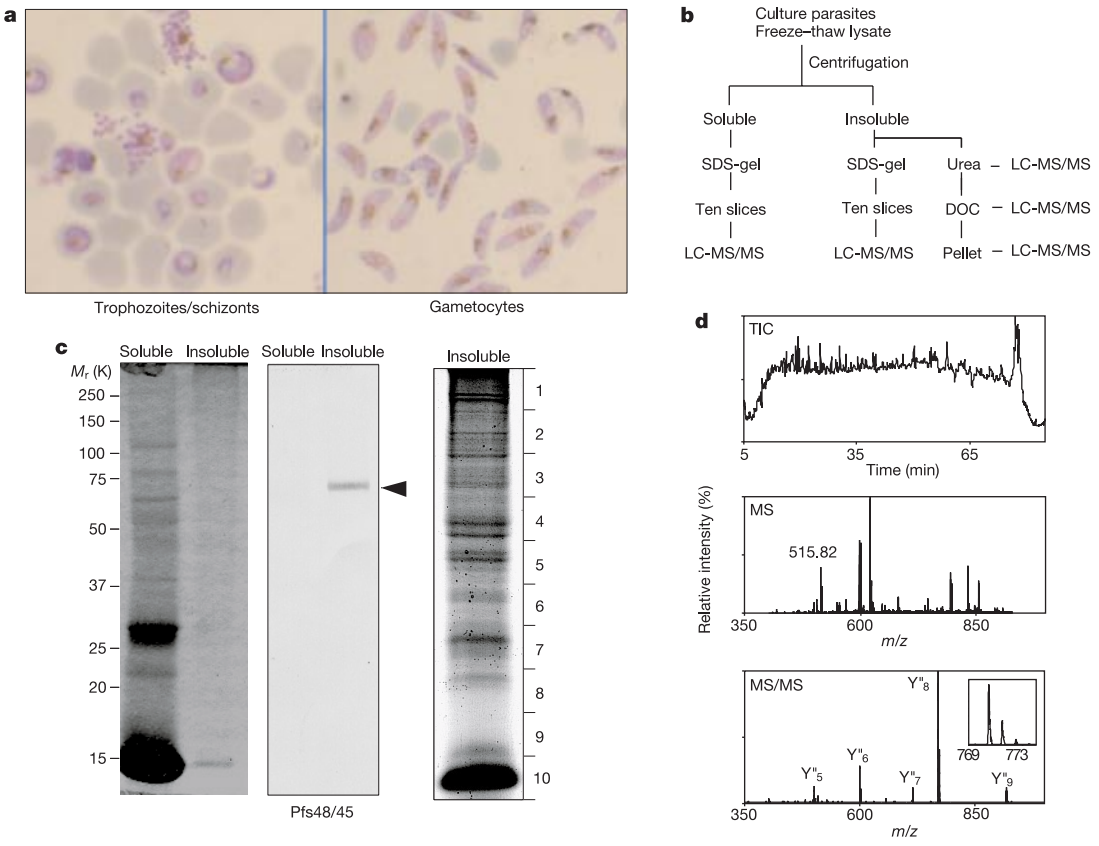

Figure 1 Differential extraction of parasite-infected red blood cells (RBCs) and flow chart of MS analysis. **a**, Giemsa staining of purified trophozoites and schizonts (left panel; asexual stage parasites), and gametocytes (right panel; sexual stage parasites). **b**, Extraction procedure for infected RBCs using freeze–thaw lyses and centrifugation, yielding a soluble and insoluble (pellet) fraction. **c**, The soluble and insoluble fraction from 5×10^5 gametocytes were separated on a 10% SDS gel and stained with Coomassie blue (left panel) or processed for western blotting and developed with a rabbit polyclonal antibody to Pfs48/45. The arrowhead indicates Pfs48/45. The right hand panel shows the insoluble fraction of 2×10^7 gametocytes. **d**, Mass spectrometric analysis of one of the gel slices shown in **c**. The top panel shows the summed ion current of all peptides eluting at a particular time. The middle panel shows a typical mass spectrum obtained from the eluting peptides. The bottom panel shows a fragmentation spectrum of the peptide with a mass of 515.82 indicated in the middle panel and magnification (inset) of one of the fragments. Database searching with this fragmentation information results in identification of peptide LFGIVGSIPK from Pfs48/45.

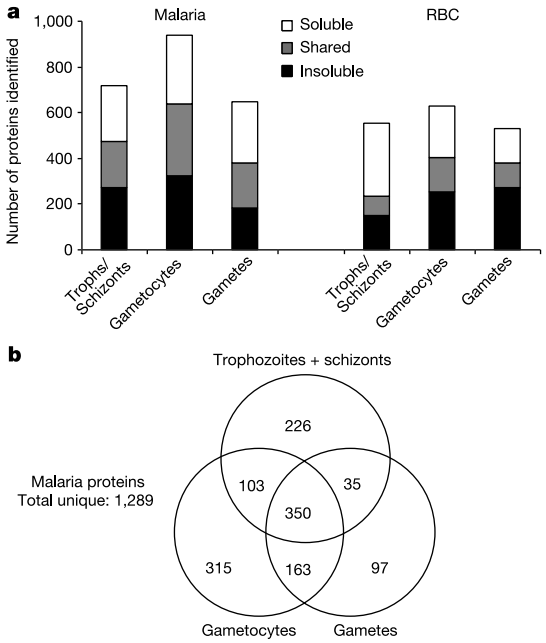

Figure 2 Schematic representation of proteomic data. **a**, Compilation of unique and common proteins in the soluble and insoluble fractions. Bars represent the number of proteins in the different preparations (trophozoites/schizonts, gametocytes and gametes) that were assigned to malaria or human RBCs. Proteins present solely in soluble (open bars) or insoluble (black bars) fractions as well as protein common to both fractions (shaded bars) are indicated. **b**, Venn diagram of the distribution of identified malaria proteins over the three blood stages. The distribution between trophozoites plus schizonts, gametocytes and gametes is shown.

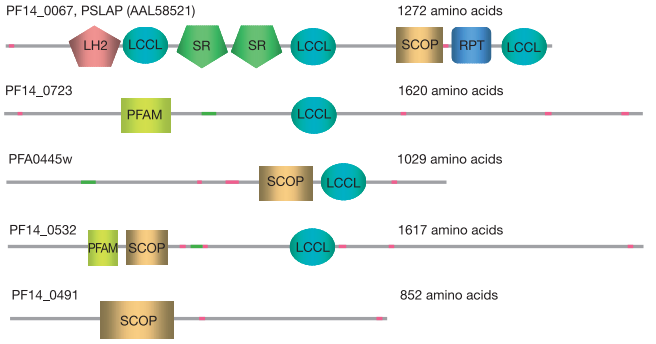

Figure 3 LCCL/lectin domain proteins expressed in sexual stages of *P. falciparum*. The SMART[22]-generated graphic descriptions of the five proteins expressed in sexual stages that contain LCCL/lectin domains are shown. The *P. falciparum* homologue of PSLAP[16] (GenBank accession number AAL58521), PF14_0067, contains a signal sequence (cleavage: amino acids, aa, 22 and 23); 3 LCCL domains (position: aa, 275–362, E value 1.28×10^{-3}; 670–676, E value 8.11×10^{-5}; and 1169–1255, E value 6.89×10^{-16}); 2 scavenger receptor (SR) domains (aa 401–515, E value 3.41×10^{-6} and 528–642, E value 3.10×10^{-7}); 1 lipoxygenase homology 2 (LH2) domain (aa 157–265, E value 8.20×10^{-3}); and one SCOP ConA lectin-like domain (aa 911–1013, E value 1×10^{-4}). PF14_0723 contains a signal peptide (cleavage: aa 19 and 20); 1 LCCL domain (aa 752–843, E value 1.4×10^{-21}); and a PFAM predicted discoidin domain (aa 296–420, E value 8.7×10^{-2}). PFA0445w contains a predicted signal sequence (cleavage: aa 24 and 25); 1 LCCL domain (aa 740–827, E value 4.49×10^{-6}); 1 SCOP kringle family domain (aa 43–95, E value 1.61×10^{-2}; not shown); and 1 SCOP dld7pm galactose-binding motif (aa 593–714, E value 4.01×10^{-3}). PF14_0532, also recognized as a gametocyte-specific transcript in *P. berghei* (AF491294), contains signal sequence (cleavage: aa 23 and 24); 1 LCCL domain (aa 724–815, E value 8.87×10^{-4}); and 1 PFAM ricin domain (aa 222–269, E value 4.0×10^{-2}). PF14_0491 is not predicted to contain a signal sequence but contains a SCOP dlacc anthrax protective antigen family with an immunoglobulin fold (aa 204–372, E value 3×10^{-9}) and SCOP/FN2 kringle-like (aa 30–80, E value 6.7×10^{-2}; not shown).

during gametogenesis than Pfs48/45.

Among the identified asexual proteins described previously are stage-specific antigens KAHRP, PfEMP3, PfSAR1 and PfSBP1, which are involved in the assembly of structures specific to the blood stage, such as knobs[11]. Schizont-specific proteins associated with the merozoite surface and invasive organelles such as MSP1, -2, -3, -6, -7 and -7a, or AMA1 and rhoptry-associated proteins RAP1, RAP2 and Rhop1 (for review of the proteins involved, see ref. 12) were also readily detected. Additionally, numerous hypothetical proteins containing a predicted signal sequence (61) and/or a transmembrane domain (189) were discovered, thus extending the repertoire of confirmed gene products and of potential new vaccine candidates (Supplementary Table B). Our analysis did not reveal any high-scoring peptides in the asexual blood stage preparations that could be assigned unambiguously to a member of the highly variable surface molecules of the PfEMP1 family, which are

Table 1 **Quantification of stage-specific proteins**

Protein name	Accession no.	RT–PCR*	MS-XIC†
Asexual stage-specific proteins (known)			
Apical membrane antigen 1 (AMA1)	PF11_0344	3 (1)	>2
Cyto-adherence-linked asexual protein (CLAG9)	PFI1730w	13 (3)	>5
Mature parasite-infected erythrocyte surface antigen (MESA)	PFE0040c	7 (4)	>12‡
Merozoite surface protein 1 (MSP1)	PFI1475w	66 (37)	>60
Cysteine protease	PFB0340c	221 (61)	>8
Sexual stage-specific proteins (known)			
Actin II	PF14_0124	26 (4)	>10
Pfs16 surface antigen	PF11_0318	8 (2)	>6
Sexual stage-specific surface antigen (Pfs48/45)	PF13_0247	43 (19)	8.2 (4.0)‡
Transmission-blocking target antigen (Pfs230)	PFB0405w	61 (20)	>6
Plasmodium falciparum gametocyte antigen (Pfg377)	PFL2405c	289 (79)	>20
Gene 11-1	PF10_0374	160 (14)	>8
Sexual stage antigen Pfs25	PF10_0303	858 (290)	>4
Sexual stage-specific proteins (novel)			
Hypothetical protein	PF14_0039	597 (208)	>17
Hypothetical protein	PF11_0310	9 (5)	>2
Hypothetical protein	PF11_0413	5 (1)	2.4 (0.5)

*Ratio between stages. RT–PCR normalized to heat shock protein 86, HSP86. Standard deviations are indicated in parentheses.
†Ratio between stages. Values were calculated from ion currents for the precise peptide molecular mass using maximum peak heights of peptides with more than 30 ion counts.
‡Extracted ion chromatogram (XIC) of the peptides of these proteins are presented in Supplementary Fig. A.

knob-associated and primarily responsible for cyto-adherence of parasites to the peripheral vascular endothelium. PfEMP1 may be expressed at very low abundance in *P. falciparum* strain NF54, or poorly extracted by our methods. Our mass spectrometry analyses revealed almost all of the known proteins specific to *P. falciparum* gametocytes and gametes (Supplementary Table B). For example, four of the six members of the P48/45 family that are transcribed in sexual stages were readily identified. In addition we detected the sexual stage-specific Pfs48/45 paralogue, Pf47, for the first time.

To corroborate and extend the stage specificity, we compared not only the absence or presence of particular peptides in the respective fractions, but also integrated the ion currents of peptides (see Supplementary Fig. A). This is important because it is possible that a peptide is not selected for sequencing owing to the complexity of the sample, but is nevertheless present at significant amounts. For example, MESA was found with 18 peptides in the asexual stage. The ion currents at the corresponding elution times in the preparation of a sexual stage were inspected but did not show any signal at the respective peptide masses. Similarly, analysis of peptide ion currents corresponding to hypothetical proteins such as PF14_0039 revealed a protein abundance ratio of at least 17:1 between the sexual and asexual stages. We next performed quantitative RT–PCR using messenger RNA from parallel asexual, gametocyte and gamete parasite preparations and gene-specific primer sets for a group of hypothetical and known proteins and an arbitrary selection of putative sexual and asexual stage-specific proteins. RT–PCR ratios of signals obtained in asexual versus sexual parasites ranged from 3-fold to over 850-fold (Table 1). In accord with the mass spectrometric findings, mRNA from Pfg377, gene 11-1 and the protein PF14_0039, but not Pfs48/45, were detected exclusively in gametocyte and gamete preparations. Comparison between the results obtained with mass spectrometry and quantitative RT–PCR shows that the changes observed in protein abundance between the asexual and sexual stages were also reflected in their mRNA levels.

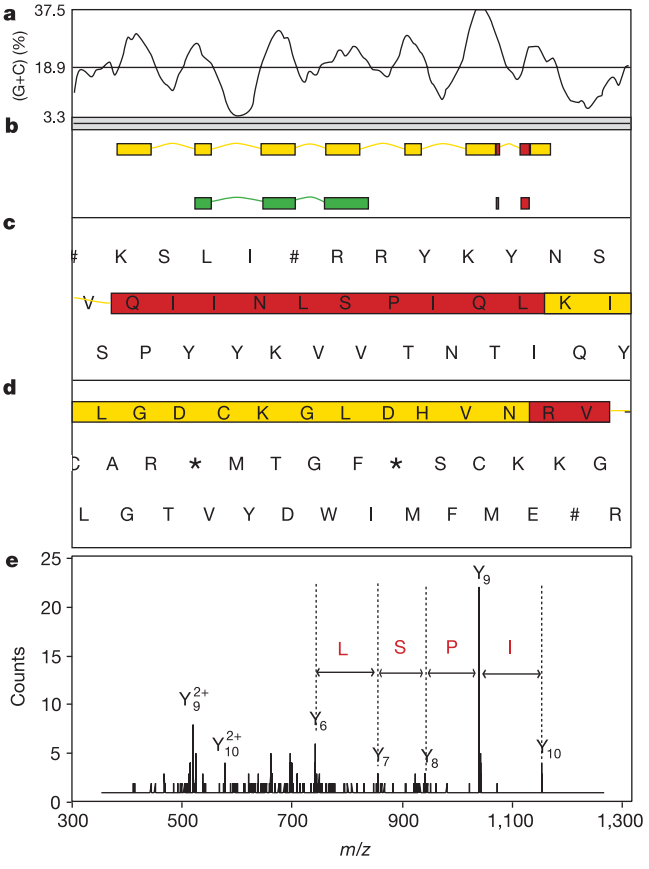

Figure 4 Refinement of gene structure and re-annotation using proteomic data. **a**, (G + C) plot in an ARTEMIS window corresponding to the genomic region of the gene PF11_0169 on chromosome 11 of *P. falciparum*. **b**, Alternative gene models for PF11_0169. The green and yellow gene models represent the original and refined gene structure of PF11_0169, respectively. The red blocks represent the peptide LQIPSLNIIQVR and its occurrence in respect to the green and the yellow gene models. **c**, **d**, Sections from an ARTEMIS window showing the splice-donor site of exon I and the splice-acceptor site of exon II of the yellow gene model. The three-frame translation is shown with hash or asterisk, denoting stop codons. The red blocks together represent the full-length peptide LQIPSLNIIQVR. **e**, Tandem mass spectrum of the orphan peptide LQIPSLNIIQVR. C-terminal fragment ions (Y ions) are indicated as well as a partial sequence.

Erythrocytic stages of the parasites are under enhanced oxidative stress and are particularly vulnerable to exogenous challenges by reactive oxygen species. Therefore it is interesting to note that proteins involved in protection against oxidative stress—the glyoxalase enzymatic system associated with glutathione—seem to be upregulated in gametes and gametocytes, such as glyoxalases I (PF11_0145) and II (PFL0285w), and two different glutaredoxins (MAL6P1.72 and PFC0271c)[13]. Some of the enzymes involved in the synthesis and metabolism of glutathione such as glutathione *S*-transferase (PF14_0187) and γ-glutamylcysteine synthetase (PFI0925w) are exclusively present in sexual stages. Others were found in sexual as well as asexual stages; these include glutathione peroxidase (PFL0595c) and glutathione reductase (PF14_0192). Furthermore, thioredoxin systems are also readily found in sexual stages (thioredoxin, PF14_0545 and PF13_0272 (secreted); thioredoxin peroxidase, PFL0725w; and thioredoxin reductase, PFL0725w).

Another stage-specific feature is the motility of male gametes. This involves the axoneme, a specialized structure consisting of two parallel microtubules thought to be connected by dynein complexes, which so far have only been demonstrated in mature merozoites[14]. The full *P. falciparum* genome reveals 12 genes that are expected to encode different dynein forms. They are conserved throughout *Plasmodium* species and have maximum homology to flagellar dyneins (for example, PF11_0240). Our analysis indicates that at least six of these (three light and three heavy chains) are expressed exclusively in sexual stages. Further specialized motility-associated proteins that may assist in the formation of the axoneme, such as actin II, α-tubulin II, β-tubulin, putative kinesins, as well as predicted tubulin-specific chaperones, are also evident.

The male gamete has been shown to bind to sialic acid on erythrocytes through a lectin-like activity[15]. Therefore, one anticipated finding within the proteome of the sexual stage was the presence of proteins (protein families) that exhibit appropriate adhesion properties. Our analysis revealed five proteins expressing predicted lectin domains of which four contain an additional LCCL motif thought to be involved in protein–protein interactions—all five are expressed exclusively in sexual stages (Fig. 3). One of these proteins, PSLAP (PF14_0067; AAL58521), has been characterized previously as gametocyte-specific and has been shown to contain multiple different domains suggestive of a role in cell–cell interactions[16]. A second (PF14_0723) has been characterized as a gametocyte-specific transcript in the rodent parasite *Plasmodium berghei* (GenBank accession number AF491294). Four of the five are predicted to be secreted proteins (PF14_0723, PFA0445w, PF14_0532, PSLAP). A fifth protein (PF14_0491) contains only a lectin domain and is predicted to be non-secreted; however, we believe that this is due to annotation based on a short exon-splice model, because alternative models would create a signal peptide. Interrogation of the *Plasmodium* genomes (*P. falciparum* and *P. y. yoelii*) suggests that these proteins are conserved in both models

and in other *Plasmodium* species that infect humans, and that expression of LCCL-domain-containing proteins is restricted to the parasite stages found in the mid-gut of the mosquito. Future investigations will determine the precise role of these proteins in the complexity of gamete fertilization, erythrocyte binding by male gametes, interactions with the host and vector immune systems, and the surface carbohydrate structures of the mosquito mid-gut.

The gel slice approach chosen for proteome analysis allowed us to correlate apparent molecular masses with gene annotation derived from theoretical mass predictions, revealing outliers that may be due to protein processing or to incomplete gene annotation. Protein Pfg377, for example, probably represents a case of protein processing (Supplementary Fig. B). The C-terminal portion of Pfg377 was covered with 69 sequenced peptides solely in the protein mixture from gel slice 5 (relative molecular mass (M_r) range of 100,000 to 140,000 (100K–140K)), whereas 15 peptides matching the N-terminal portion were found exclusively in gel 7 (65K–84K). Northern blot analysis[10] and quantitative RT–PCR using 5′ and 3′ specific primer sets (data not shown) indicate that the proteins are encoded by a single gene that is processed after translation.

The parasite genome is highly (A + T)-rich and has proven difficult to assemble and annotate[8]. For this reason, we also performed direct genome searches[17] where the malaria genome rather than the predicted set of proteins was searched by the mass spectrometric data. A large number of additional hits emerged. After analysing these 'orphan' peptide hits in the genome of *P. falciparum*, we were able to identify additional exons or assign different exon–intron boundaries of previously annotated genes using the ARTEMIS annotation tool[18]. One such orphan peptide, LQIPSLNIIQVR, identified two additional exons of the gene PF11_0169 on chromosome 11, annotated as a hypothetical protein. With this information we were able to identify two further exons and modify the gene structure appropriately. This led to re-annotation of the previously annotated hypothetical protein PF11_0169 as a putative homologue of sno-type pyridoxin biosynthesis protein (Fig. 4). The orphan peptide not only identified the two exons of the gene PF11_0169 but also verified the boundaries of the intron between the exons I and II. With the refined gene model we were able to re-annotate the gene with appropriate Gene Ontology (GO) terms. Initial analysis similar to the one outlined above led to eight definite and six probable gene annotation changes. Supplementary Table C lists more than 100 high-quality orphan peptides that can be used in gene annotation and the finding of new genes.

We have applied state-of-the-art proteomics techniques to obtain valuable information about stage-specific expression and localization of malaria proteins as well as information useful in confirming and extending the bioinformatics analysis of the proteome. Further work could encompass additional stages of the parasite life cycle and make use of the rapidly evolving mass spectrometric technology, associated bioinformatics, as well as direct comparison of stages using stable isotope techniques. □

Methods

Preparation of parasites

Plasmodium falciparum parasites were cultured using a semi-automated culture system. Asexual stages (trophozoites and schizonts) were purified using the Variomacs system essentially as described[14]. For the production of gametocytes, parasites were cultured for 14 days without addition of red blood cells. The gametocyte cultures were treated with 50 mM *N*-acetyl-glucosamine from day 8 to 12 after the start of the culture to remove asexual stage parasites. After a total of 14 days culture the gametocytes were collected and purified using the Variomacs system as described[14]. For the production of gametes, purified gametocytes were pelleted and incubated in G_m buffer containing 100 μM xanthurenic acid for a period of 3 h at 21 °C (ref. 19). (G_m buffer contains 1.67 mg ml^{-1} glucose, 8 mg ml^{-1} NaCl, 1 mg ml^{-1} Tris, pH 8.1.) Every effort was made to minimize enzymatic activity and protein degradation during sampling and the subsequent isolation of the parasites. However, we cannot exclude that some of the differences in protein profiles that we observe between the different life-cycle stages may be a consequence of the sample-handling procedures.

Sample preparation

The infected red blood cells (10^7 cells) of each stage were divided into a soluble and insoluble fraction by freeze–thawing several times and pelleted by centrifugation at 13,000 r.p.m. (16,100 *g*). Proteins were extracted in SDS–polyacrylamide gel electrophoresis (PAGE) loading buffer and separated into ten fractions on a 10% protein gel. The separated proteins were treated with dithiothreitol (DTT) and iodoacetamide, and in-gel digested with trypsin[20]. Proteins were also extracted by 8 M urea, 100 mM Tris-HCl, pH 8.0, treated with DTT and iodoacetamide, and digested in-solution with endoprotease Lys-C and trypsin after dilution to 2 M urea. Proteins from the insoluble fractions were further extracted with 1% deoxycholic acid (DOC), 100 mM Tris-HCl, pH 8.0, and digested together with the remaining pellet as above in the presence of 0.05% DOC.

NanoLC-MS/MS analysis of malaria proteins

Peptide mixtures were loaded onto 75-μm ID columns packed with 3-μm C18 particles (Vydac) and eluted into a quadruple time-of-flight mass spectrometer (QSTAR, Sciex-Applied Biosystems). Fragment ion spectra were recorded using information-dependent acquisition and duty-cycle enhancement (see ref. 21 and references therein for a more detailed description of the method). The three malaria stages studied were analysed at least in duplicate. In total, more than 100 nanoLC-MS/MS runs were analysed, yielding more than 200,000 peptide-sequencing events.

Data analysis

Peak lists containing the precursor masses and the corresponding MS/MS fragment masses were generated from the original data file and searched in the annotated *P. falciparum* database (Sanger/TIGR) combined with the human IPI database (European Bioinformatics Institute) using the Mascot program (Matrix Science). Identified peptides that were not unique or had a score less than 20 were removed. Proteins identified with a combined peptide score of higher than 60 were considered significant, and lower scoring proteins were manually verified or rejected. Iterative calibration algorithms on the basis of identified peptides were used to achieve a final average absolute mass accuracy of better than 20 p.p.m. in both the precursor and fragment ions. Relative protein abundance between the malaria stages was based on the total number of unique peptides identified for each protein and was further supported by extracted ion chromatograms (XIC) for individual peptides within an elution time window of 3 min (see also Supplementary Fig. A).

Quantitative real-time RT–PCR

Relative mRNA abundance was quantified by real-time PCR with the GeneAmp 5700 Sequence Detection System (Applied Biosystems) using the SYBR Green PCR Master Mix kit (Applied Biosystems). Total RNA was isolated from two stages: asexual blood stage (trophozoites and schizonts) and the sexual blood stage (gametocytes), both prepared as described above. RNA was isolated from 4×10^7 parasites using TRIzol reagent (Gibco BRL). A total of 2 μg of RNA was used for complementary DNA synthesis. RNA was treated with DNaseI (Pharmacia) after which cDNA was synthesized with random hexamers and Superscript II enzyme (Gibco BRL), according to standard protocols. cDNA was dissolved in 50 μl diethyl pyrocarbonate (DEPC)-treated H_2O. A total of 3 μl of × 10 and × 100 diluted cDNA was used for real-time PCR. Primers were designed by Primer Express software (Applied Biosystems). The mRNA levels for each gene were normalized against heat shock protein 86. For each gene the relative mRNA abundance was determined by calculating the ratio of asexual:gametocyte stage and gametocyte:asexual stage.

Received 31 July; accepted 9 September 2002; doi:10.1038/nature01111.

1. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405,** 837–846 (2000).
2. Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17,** 676–682 (1999).
3. Griffin, T. J. & Aebersold, R. Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* **276,** 45497–45500 (2001).
4. Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100,** 191–200 (1990).
5. Richie, T. L. & Saul, A. Progress and challenges for malaria vaccines. *Nature* **415,** 694–701 (2002).
6. Kocken, C. H. *et al.* Cloning and expression of the gene coding for the transmission blocking target antigen Pfs48/45 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **61,** 59–68 (1993).
7. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567 (1999).
8. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419,** 498–511 (2002).
9. van Dijk, M. R. *et al.* A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104,** 153–164 (2001).
10. Alano, P. *et al.* COS cell expression cloning of Pfg377, a *Plasmodium falciparum* gametocyte antigen associated with osmiophilic bodies. *Mol. Biochem. Parasitol.* **74,** 143–156 (1995).
11. Wickham, M. E. *et al.* Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J.* **20,** 5636–5649 (2001).

12. Cowman, A. F. *et al.* Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. *FEBS Lett.* **476,** 84–88 (2000).
13. Rahlfs, S., Fischer, M. & Becker, K. *Plasmodium falciparum* possesses a classical glutaredoxin and a second, glutaredoxin-like protein with a PICOT homology domain. *J. Biol. Chem.* **276,** 37133–37140 (2001).
14. Fowler, R. E. *et al.* Microtubule associated motor proteins of *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* **117,** 187–200 (2001).
15. Templeton, T. J., Keister, D. B., Muratova, O., Procter, J. L. & Kaslow, D. C. Adherence of erythrocytes during exflagellation of *Plasmodium falciparum* microgametes is dependent on erythrocyte surface sialic acid and glycophorins. *J. Exp. Med.* **187,** 1599–1609 (1998).
16. Delrieu, I. *et al.* PSLAP, a protein with multiple adhesive motifs, is expressed in *Plasmodium falciparum* gametocytes. *Mol. Biochem. Parasitol.* **121,** 11–20 (2002).
17. Kuster, B., Mortensen, P., Andersen, J. S. & Mann, M. Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics* **1,** 641–650 (2001).
18. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16,** 944–945 (2000).
19. Brooks, S. R. & Williamson, K. C. Proteolysis of *Plasmodium falciparum* surface antigen, Pfs230, during gametogenesis. *Mol. Biochem. Parasitol.* **106,** 77–82 (2000).
20. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858 (1996).
21. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12,** 1231–1245 (2002).
22. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95,** 5857–5864 (1998).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We thank our colleagues for help and discussions. This work was supported by the Danish National Research Foundation, the Dutch Science Foundation (NWO), the European Union and the World Health Organization (WHO) Special Program for Research and Training in Tropical Diseases.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.G.S. (e-mail: h.stunnenberg@ncmls.kun.nl) or M.M. (e-mail: mann@bmb.sdu.dk). Sequence data for the genes newly annotated according to the present study can be found at http://www.sanger.ac.uk/Projects/P_falciparum.