Exploring the transcriptome of the malaria sporozoite stage

Stefan H. I. Kappe*†, Malcolm J. Gardner‡, Stuart M. Brown§, Jessica Ross*, Kai Matuschewski*, Jose M. Ribeiro¶, John H. Adams‖, John Quackenbush‡, Jennifer Cho‡, Daniel J. Carucci, Stephen L. Hoffman††, and Victor Nussenzweig***

*Michael Heidelberger Division, Department of Pathology, Kaplan Cancer Center, New York University School of Medicine, New York, NY 10016; ‡The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; §Research Computing Resource, New York University Medical Center, New York, NY 10016; ¶Medical Entomology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-0425; ‖Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; **Malaria Program, Naval Medical Research Center, Silver Spring, MD 20910; and ††Celera Genomics, 45 West Gude Drive, Rockville, MD 20850

Edited by Louis H. Miller, National Institutes of Health, Bethesda, MD, and approved June 19, 2001 (received for review April 13, 2001)

Most studies of gene expression in *Plasmodium* have been concerned with asexual and/or sexual erythrocytic stages. Identification and cloning of genes expressed in the preerythrocytic stages lag far behind. We have constructed a high quality cDNA library of the *Plasmodium* sporozoite stage by using the rodent malaria parasite *P. yoelii*, an important model for malaria vaccine development. The technical obstacles associated with limited amounts of RNA material were overcome by PCR-amplifying the transcriptome before cloning. Contamination with mosquito RNA was negligible. Generation of 1,972 expressed sequence tags (EST) resulted in a total of 1,547 unique sequences, allowing insight into sporozoite gene expression. The circumsporozoite protein (CS) and the sporozoite surface protein 2 (SSP2) are well represented in the data set. A BLASTX search with all tags of the nonredundant protein database gave only 161 unique significant matches ($P(N) \leq 10^{-4}$), whereas 1,386 of the unique sequences represented novel sporozoite-expressed genes. We identified ESTs for three proteins that may be involved in host cell invasion and documented their expression in sporozoites. These data should facilitate our understanding of the preerythrocytic *Plasmodium* life cycle stages and the development of preerythrocytic vaccines.

Plasmodium yoelii yoelii | expressed sequence tag

Protozoan parasites of the genus *Plasmodium* are the causative agents of malaria, the most devastating parasitic disease in humans. The parasites occur in distinct morphological and antigenic stages as they progress through a complex life cycle, thwarting decades of efforts to develop an effective malaria vaccine. *Plasmodium* is transmitted via the bite of an infected *Anopheles* mosquito, which releases the sporozoite stage into the skin. Sporozoites enter the bloodstream and, on reaching the liver, invade hepatocytes and develop into exo-erythrocytic forms (EEF). After multiple cycles of DNA replication, the EEF contains thousands of merozoites (liver schizont) that are released into the blood stream and initiate the erythrocytic cycle (asexual blood stage) that causes the disease malaria. Changes in life cycle stages are accompanied by major changes in gene expression and therefore by major changes in antigenic composition. The form of the parasite best studied is the asexual blood stage, mainly because of its comparatively easy experimental accessibility. Therefore, most *Plasmodium* proteins that have been well characterized are expressed during the erythrocytic cycle, among them some major erythrocytic-stage vaccine candidates such as merozoite surface protein-1 (MSP-1) and apical membrane antigen-1 (AMA-1; ref. 1). Erythrocytic-stage vaccines are aimed at inducing an immune response that suppresses or eradicates parasite load in the blood. In contrast, preerythrocytic vaccines are aimed at eliciting an immune response that destroys the sporozoites and the EEF, thereby preventing progression of the parasite to the blood stage. The feasibility of a preerythrocytic vaccine is demonstrated by the fact that immunization with radiation-attenuated sporozoites leads to protective, sterile immunity (2, 3). The effector mechanisms are antibodies (4), cytotoxic T lymphocytes (CTL; ref. 4), and lymphokines (5, 6). Hence, it is desirable to systematically identify proteins synthesized by sporozoites and EEF to select new potential vaccine candidates. Antibodies against surface-exposed sporozoite proteins block hepatocyte entry (7). In addition, sporozoite proteins can be carried over into the invaded hepatocyte and become a target for CTL (8). By using mixtures of these proteins, it might be possible to formulate a vaccine that mimics the sterile immunity achieved by immunization with irradiated sporozoites. Sporozoite proteins could also be the target of transmission-blocking strategies. Past efforts to prepare cDNA libraries of sporozoites and identify new sporozoite antigens were hindered by difficulties in obtaining adequate numbers of purified parasites. Thus far, few sporozoite-expressed proteins have been identified. The best characterized of these proteins are the circumsporozoite protein (CS; ref. 2) and the sporozoite surface protein 2 (SSP2), also called thrombospondin-related anonymous protein (TRAP; refs. 9–11). CS and SSP2/TRAP are involved in the invasion of hepatocytes and are detected in the hepatocyte after sporozoite invasion. Both proteins are found in all *Plasmodium* species examined. A few other sporozoite antigens have been identified in *P. falciparum* (12, 13), but their function is unknown.

To facilitate the identification of genes that are expressed in the sporozoite stage, we have constructed a cDNA library from salivary gland sporozoites of the rodent malaria parasite *Plasmodium yoelii* and generated 1,972 expressed sequence tags (ESTs). We document the quality of the library by the presence of CS and SSP2/TRAP transcripts and the absence of erythrocytic stage-specific transcripts. The sequence data provide insight into sporozoite gene expression. We show sporozoite expression of MAEBL (14), a protein previously thought to be present only in erythrocytic stages. In addition, we identify two putative sporozoite adhesion ligands. Transcripts of a key enzyme of the shikimate pathway (15) are present in the data set, indicating that this pathway is likely to be operational in sporozoites and liver stages.

MICROBIOLOGY

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CS, circumsporozoite protein; SSP2, sporozoite surface protein 2; TRAP, thrombospondin-related anonymous protein; EST, expressed sequence tag; EEF exo-erythrocytic form; MSP-1, merozoite surface protein-1; MyoA, myosin A; TSR, thrombospondin type 1 repeat; SPATR, secreted protein with altered thrombospondin repeat.

Data deposition: The EST sequences reported in this paper have been deposited in the GenBank dbEST database (accession nos. BG601070–BG603042). Complete gene sequences have been deposited in the GenBank database (accession nos. AF390551–AF390553).

†To whom reprint requests should be addressed. E-mail: kappes01@popmail.med.nyu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Materials and Methods

Parasite Preparation. Two million *P. yoelii* (17XNL) sporozoites were obtained in a salivary gland homogenate from dissection of 100 infected *Anopheles stephensi* mosquitos. The crude salivary gland homogenate was passed over a DEAE cellulose column to remove contaminating mosquito tissue. Sporozoites (4×10^5) were recovered after purification. The preparation was almost free of mosquito contaminants as judged by microscopic inspection. Sporozoites were immediately subjected to poly(A)$^+$ RNA extraction.

RNA Extraction and cDNA Synthesis. Poly(A)$^+$ RNA was directly isolated from the sporozoites by using the MicroFastTrack procedure (Invitrogen) and was resuspended in a final volume of 10 μl elution buffer (10 mM Tris, pH 7.5). The obtained poly(A)$^+$ RNA was treated with Dnase I (Life Technologies, Rockville, MD) to remove possible genomic DNA contamination. RNA quantification was not possible because of the minute amounts obtained. The RNA was reverse-transcribed by using Superscript II (Life Technologies), a modified oligo(dT) oligonucleotide for first strand priming (5′-AAGCAGTGGTAACAACGCAGAGTACT$_{30}$VN-3′; V = A/C/G, N = A/C/G/T) and a primer called cap switch oligonucleotide (5′-AAGCAGTGGTAACAACGCAGAGTACGCGGG-3′) that allows extension of the template at the 5′ end (CLONTECH). Second strand synthesis and subsequent PCR amplification was done with an oligonucleotide that anneals to both the modified oligo(dT) oligonucleotide and the cap switch oligonucleotide.

cDNA Cloning and Sequencing. The cDNA was size selected on a CHROMA-SPIN 400 column (CLONTECH) that resulted in a cutoff at ≥300 bp and was ligated into vector pCR4 (Invitrogen). Ligations were transformed into *Escherichia coli* TOP10-competent cells. Template preparation and sequencing were done as described (16). Sequencing was performed in both directions.

Assemblies and Database Searches. All obtained sequences were subjected to vector sequence removal and screened for overlaps, and matching sequences were then assembled by using the TIGR assembler program. The nonredundant (NR) sequence database at the National Center for Biotechnology Information (NCBI) was searched with the complete data set, consisting of the assembled sequences and singletons, by using the Basic Local Alignment Search Tool X (BLASTX) algorithm.

Sources of Sequence Data. Sequence data were obtained from the TIGR *P. yoelii* genome project (www.tigr.org) and the *Plasmodium* genome consortium PlasmoDB (http://PlasmoDB.org).

cDNA Blots. cDNA was separated on agarose gels and transferred to nylon membranes (Roche). Gene-specific probes were prepared by using the digoxigenin (DIG) High Prime Labeling system (Roche). cDNA blots were incubated and washed according to the manufacturer's instructions (Roche).

Reverse Transcription–PCR. Poly(A)$^+$ RNA was reverse-transcribed by using Superscript II. Gene-specific PCR was done by using oligonucleotide primers specific for *P. yoelii MSP-1* (L22551; sense, 5′-GGTAAAAGCTGGCGTCATTGATCC-3′; antisense, 5′-GTCTAATTCAAAATCATCGGCAGG-3′) or *P. yoelii MAEBL* (AF031886; sense, 5′-ATGCTGCTCAATATCAGATTATTGC-3′; antisense, 5′-AACAATTTCATCAAAAGCAACTTCC-3′).

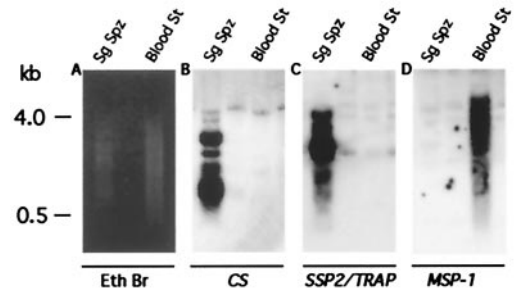

Fig. 1. Quality assessment of the generated cDNA populations. cDNA blot hybridization with stage-specific probes demonstrates that stage-specific transcript representation is not altered by cDNA amplification. (*A*) Ethidium bromide-stained agarose gel of cDNA amplified from salivary gland sporozoites (Sg Spz) or mixed blood stages (Blood St). Note the distinct bands visible in the sporozoite preparation. (*B*) Hybridization to a *CS* probe. (*C*) Hybridization to an *SSP2/TRAP* probe. (*D*) Hybridization to an *MSP-1* probe. Sizes are given in kb.

Indirect Immunofluorescence Assay. Salivary gland sporozoites and midgut sporozoites were incubated in 3% BSA/RPMI medium 1640 on BSA-covered glass-slides for 30 min, fixed, and permeabilized with 0.05% saponin. MAEBL was detected with the polyclonal antisera against the M2 domain or the 3′-carboxyl cysteine-rich region (1:200; ref. 14) and FITC-conjugated goat anti-rabbit IgG (1:100; Kirkegaard & Perry Laboratories).

Results

Quality Assessment of the cDNA Library. The amplified sporozoite cDNAs showed a visible size distribution between 300 and 4,000 bp on ethidium bromide-stained agarose gels, with highest density between 500 and 3,000 bp (Fig. 1*A*). No amplification was detected when the reverse transcription step was omitted (data not shown). To assess the quality of the sporozoite cDNA population, we performed cDNA blot analysis with probes for the sporozoite-expressed *SSP2/TRAP* and *CS*. cDNAs for both proteins were found to be abundant in salivary gland sporozoite preparations but absent in blood stage parasite preparations (Fig. 1 *B* and *C*). Conversely, cDNAs for the blood stage-expressed *MSP-1* were detected in blood stage parasite preparations but absent in sporozoites (Fig. 1*D*). The cDNA blot analysis documented the presence of cDNAs of the approximate full-length size of each transcript. In addition, smaller sized cDNA fragments were present for each transcript, resulting in multiple signals from distinctly sized cDNAs (Fig. 1). To assure that no trace amounts of genomic DNA were amplified, we analyzed the sporozoite cDNA for the presence of introns by using the transcript of myosin A (MyoA), a myosin that is expressed in the sporozoite stage (17). *MyoA* contains two introns, and neither was detected in the sporozoite cDNA preparation (data not shown). Sequencing of 100 clones confirmed the cDNA fragmentation, which was mainly due to internal priming by the modified oligo(dT) oligonucleotide. It annealed to homo-polymeric runs of adenine in the untranslated regions (UTR) and the coding sequences of this AT-rich organism. We took advantage of the AT-richness of the *P. yoelii* genome to differentiate between cDNAs of parasite origin and cDNAs amplified from contaminating mosquito RNA. Based on the total number of cDNA clones of mosquito origin, contamination was estimated to be ≈1%.

Characteristics of the EST Data Set. We obtained a final number of 1,972 sequence reads of sufficient quality to be subjected to further analysis (Table 1). The average length of EST sequence was 377 bp. Six hundred forty-eight of the sequence reads could be assembled into 223 consensus sequences (input files), and

Table 1. General characteristics of the *P. yoelii* sporozoite EST project

ESTs submitted to NCBI	1,972
ESTs in input files	648
Input files	223
Singletons	1,324
Total number of unique sequences	1,547
BLASTX matches	286
Unique BLASTX matches	161
Matches with proteins of unknown function	75
BLASTX matches with *Plasmodium* proteins	70
ESTs for CS	33
ESTs for SSP2/TRAP	13
ESTs for MAEBL	10
ESTs for HSP-70	10

1,324 sequences did not match another sequence in the data set sufficiently to allow assembly (singletons). This analysis gave a total of 1,547 unique sequences. A BLASTN comparison between the 1,547 unique sequences and the incomplete *P. yoelii* genome (2× coverage) database resulted in 1,135 matches. A BLASTX search of the predicted proteins from *the P. falciparum* genome (translated ORFs of >100 bases) resulted in only 356 matches, with a smallest sum probability of $P(N) \leq 10^{-4}$. A BLASTX search of the NR sequence database at NCBI resulted in only 286 matches, with a smallest sum probability of $P(N) \leq 10^{-4}$. Of those, 70 were matches with known *Plasmodium* proteins. The matches were grouped in functional categories shown in Fig. 2 (see Table 2, which is published as supplemental data on the PNAS web site, www.pnas.org, for a complete list of all BLASTX matches). All ESTs have been deposited in the GenBank dbEST database (accession nos. BG601070–BG603042). In addition, data are made available through the *P. yoelii* gene index (http://www.tigr.org/tdb/pygi/).

Functional Groups of ESTs. Ribosomal proteins were not very abundant, with only 7 of the estimated 80 components of the ribosome represented. Only 4 ESTs gave matches with other proteins involved in translation. This low representation of proteins of the translation machinery contrasts with the relative abundance of ribosomal proteins found in EST sequencing projects for *Toxoplasma* tachyzoites (12% of all ESTs; refs. 18 and 19) and *Cryptosporidium* sporozoites (8% of all ESTs; ref. 20). However, a *P. falciparum* blood stage parasite EST project found that proteins involved in translation were also underrepresented (21). There were 18 ESTs in the transcription category, 7 matching a *P. falciparum* RNA recognition motif binding protein and two matching a human zinc finger protein potentially involved in transcription.

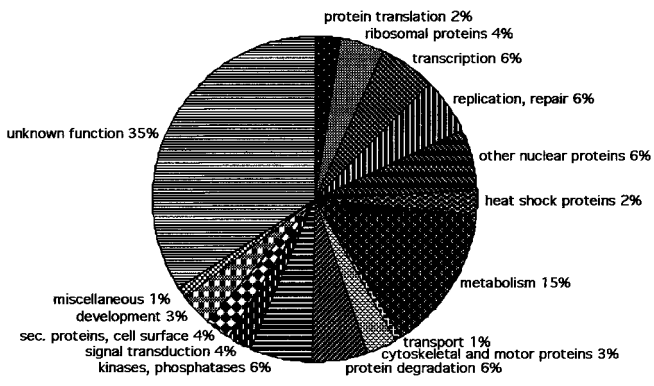

Fig. 2. Functional classification of *P. yoelii* sporozoite ESTs. One hundred sixty-one unique BLASTX matches were classified according to their putative biological function. Refer to Table 2 for a complete list of all BLASTX matches.

Especially significant among the ESTs giving BLASTX matches with proteins involved in metabolic pathways is chorismate synthase, the final enzyme of the shikimate pathway. This pathway generates the aromatic precursor chorismate, which is used for aromatic amino acid biosynthesis. The shikimate pathway is present in plants, fungi, and Apicomplexa (15) but is not found in vertebrates.

The salivary gland sporozoite is highly motile, and its main function is the invasion of the vertebrate hepatocyte. Of relevance to motility and invasion are tags for two apicomplexan unconventional class XIV myosins, MyoA and MyoB. MyoA localized under the plasma membrane within all invasive stages of *Plasmodium* (sporozoite, merozoite, and ookinete; refs. 17, 22, and 23), and a homologous protein was expressed in the *Toxoplasma* tachyzoite (24, 25). This myosin is currently the best candidate for the motor protein that drives Apicomplexan motility and host cell penetration.

Kinases and phosphatases are likely to be involved in the regulation of motility and host cell invasion (26), and we find 10 different input files and singletons in this category. Recently it was shown that a calmodulin-domain kinase, represented with one EST in the data set, played a crucial role in *Toxoplasma* tachyzoite motility and host cell invasion (27). Phospholipase A_2 is represented with one EST. Involvement of secreted phospholipase A_2 in the invasion process was shown in *Toxoplasma* tachyzoites (28). It will be of interest to find out whether this *Plasmodium* homologue has a role in hepatocyte invasion and/or plays a role in the migration of sporozoites through cells before establishing an infection (29).

The group of predicted secreted proteins and proteins that have a membrane anchor are of special interest, because they may be involved in host cell recognition and/or invasion. Within this group is the CS protein, most likely glycosylphasphatidylinositol-anchored, and SSP2/TRAP, a type one transmembrane protein. CS had one of the highest representations in the EST set with 33 matches, and TRAP was represented with 13 matches (Table 1).

Identification of Three Potential Sporozoite Invasion Ligands. Unexpectedly, we found that MAEBL was represented with 10 ESTs (Table 1). It was reported previously that MAEBL is expressed in *P. yoelii* and *P. berghei* merozoites, where it localized to the rhoptry organelles (14, 30). MAEBL is a type one transmembrane protein with a chimeric structure. It shares similarity with apical membrane antigen-1 (AMA-1) in the N-terminal portion, and similarity with the erythrocyte binding protein (EBP) family in the C-terminal portion (31). To ensure that the representation of a merozoite rhoptry protein in our EST library was not an artifact, we hybridized a salivary gland and midgut sporozoite cDNA blot to a *MAEBL*-specific probe, resulting in strong signals for both populations (Fig. 3*A*). In addition, reverse transcription–PCR with gene-specific primers resulted in *MAEBL* amplification from salivary gland sporozoite poly(A)$^+$ RNA and from blood stage poly(A)$^+$ RNA. In contrast, *MSP-1* expression was detected only in blood stages (Fig. 3*B*). A polyclonal antiserum against the carboxyl cysteine-rich region of *P. yoelii* MAEBL strongly reacted with permeabilized *P. yoelii* salivary gland sporozoites and midgut sporozoites in indirect immunofluorescence assay (IFA), indicating that this protein is indeed expressed in the sporozoite stages (Fig. 3 *C* and *D*). MAEBL localization was heterogeneous but was frequently more pronounced in one end of the sporozoites. Similar staining was obtained with a polyclonal antiserum against the M2 domain of MAEBL (data not shown).

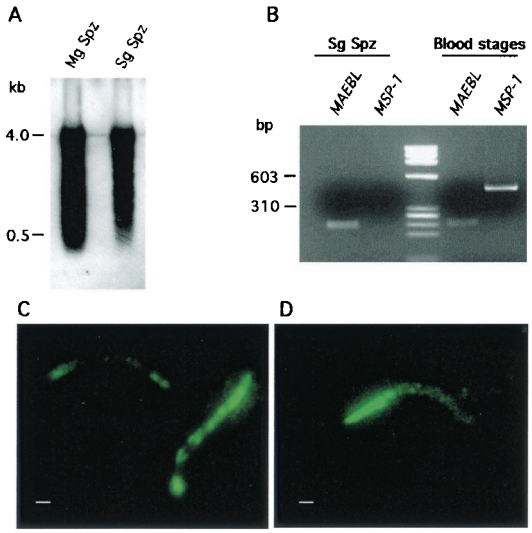

Fig. 3. Sporozoite expression of MAEBL. (*A*) cDNA blot showing *MAEBL* expression in midgut sporozoites (Mg Spz) and salivary gland sporozoites (Sg Spz). (*B*) Reverse transcription–PCR confirming *MAEBL* expression in salivary gland sporozoites. *MAEBL* expression is also detected in blood stages. Amplification with *MSP-1*-specific primers shows *MSP-1* expression in blood stages. *MSP-1* expression is not detected in salivary gland sporozoites. Sizes are given in base pairs (bp). (*C*) Localization of MAEBL by indirect immunofluorescence assay in *P. yoelii* salivary gland sporozoites with antisera against the carboxyl cysteine-rich region. (*D*) Localization of MAEBL by indirect immunofluorescence in *P. yoelii* midgut sporozoites with antisera against the carboxyl cysteine-rich region. Scale bar for *C* and *D* = 1 μm.

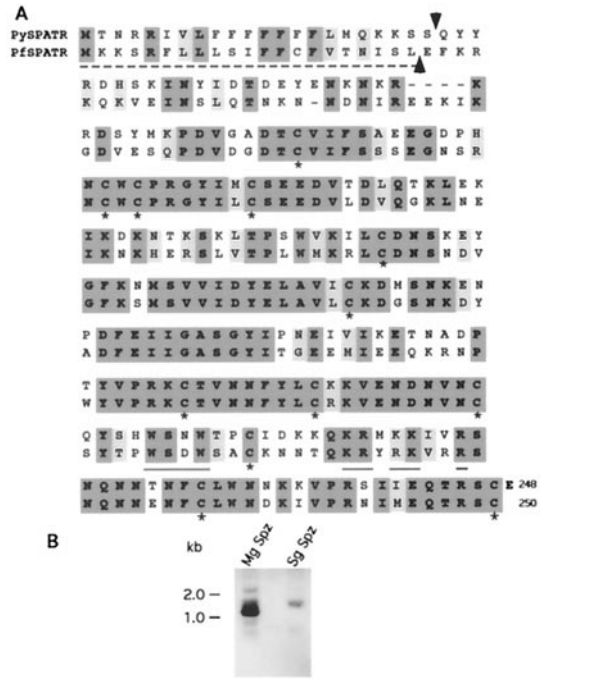

Fig. 4. Alignment of SPATR and expression in sporozoites. (*A*) Comparison of the deduced amino acid sequences of the *P. yoelii* SPATR with the homologue in *P. falciparum* (accession no. C71611). The conserved residues of the altered TSR are underlined with a solid line. The putative signal peptides are underlined with a dashed line. Putative signal peptide cleavage sites are marked with arrowheads (▲, ▼). Conserved cysteine residues are marked with an asterisk (*). Identical residues are shaded dark gray. Conserved amino acid changes are shaded light gray, and radical changes are not shaded. (*B*) cDNA blot demonstrating *SPATR* expression in midgut sporozoites (Mg Spz) and salivary gland sporozoites (Sg Spz). Sizes are given in kb.

One EST in the data set identified another potential sporozoite invasion ligand, matching a hypothetical ORF on chromosome 2 of *P. falciparum* (PFB0570w; ref. 16). We determined the complete ORF for this *P. yoelii* EST. The predicted protein has a putative cleavable signal peptide predicting that it is secreted (Fig. 4*A*). Significantly, the protein carries a motif with similarity to the thrombospondin type 1 repeat (TSR) (32). We therefore named it SPATR (secreted protein with altered thrombospondin repeat). The most conserved motif of the TSR is present (WSXW), followed by a stretch of basic residues. The central CSXTCG that follows the WSXW motif in a number of the TSR superfamily members (33) is not present in SPATR. Interestingly, this motif is present in the TSR of CS but it is not important for CS binding to the hepatocyte surface (34). The *P. yoelii* and *P. falciparum* SPATR proteins share 63% amino acid sequence identity, including 12 conserved cysteine residues (Fig. 4*A*). The N-terminal intron of *SPATR* is conserved in both species (data not shown). This overall similarity suggests that the proteins are homologous. To confirm *SPATR* transcription, we hybridized a salivary gland and midgut sporozoite cDNA blot to a *SPATR*-specific probe. *SPATR* cDNA seemed more abundant in the midgut sporozoite preparations (Fig. 4*B*).

One EST showed weak similarity with Pbs48/45, a member of the six-cysteine (6-cys) superfamily (35). A *P. yoelii* contig from the *P. yoelii* genome project that matched this EST showed a single ORF of 1,440 bp coding for a predicted mature 52-kDa protein. Search of the *P. falciparum* genome database identified a putative homologue that shared 40% amino acid sequence identity with the *P. yoelii* protein (Fig. 5*A*). Both predicted proteins have consensus amino terminal cleavable signal peptides followed by two tandem 6-cys domains. A carboxyl-terminal hydrophobic domain indicated that the proteins could be membrane-anchored by a glycosylphasphatidylinositol linkage. The presence of the 6-cys domain and the overall structure clearly identified the proteins as new members of the 6-cys superfamily. According to the nomenclature of this superfamily by predicted molecular mass of the mature protein, we named the proteins Py52 and Pf52. To confirm *Py52* expression, we hybridized a salivary gland and midgut sporozoite cDNA blot to a *Py52* specific probe. *Py52* cDNA seemed more abundant in the midgut sporozoite preparations (Fig. 5*B*).

Finally, it is noteworthy that none of our ESTs resulted in significant matches with sporozoite-threonine asparagine-rich protein and liver stage antigen-3, proteins that have been described in *P. falciparum* sporozoites (12, 13).

Discussion

The nearly complete genome sequence of *P. falciparum* is now available, and its annotation will be concluded in the near future (36). It has been estimated that the 25–30 megabase genome harbors about 6,000 expressed genes. In addition, a 2× sequence coverage of the *P. yoelii* genome has very recently been completed and made publicly available (www.tigr.org). Malaria parasites occur in a number of different life cycle stages, making it a challenging task to determine which subset of the 6,000 genes is represented in the transcriptome of each stage. Microarrays will be the method of choice for expression analysis in asexual and sexual blood stage parasites where the acquisition of sufficient RNA is not a limitation. Although whole genome microarrays are not yet available, partial arrays from mung bean genomic libraries (37) or blood stage cDNA libraries (38) have been used successfully to study gene expression in blood stages. However, microarray analysis of gene expression in ookinetes, early oocysts, sporozoites, and EEF of mammalian Plasmodia will be difficult because large quantities of these stages are not available.

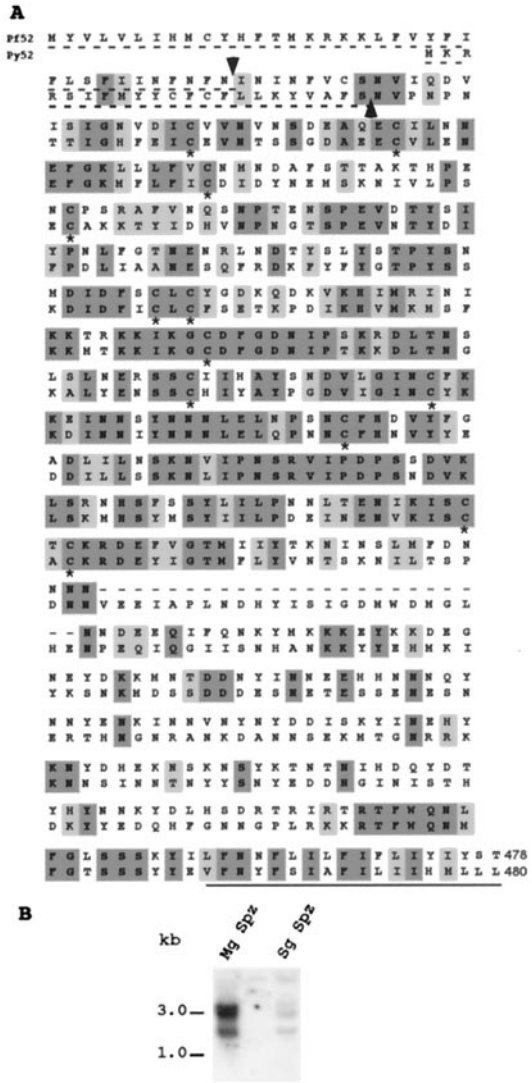

Fig. 5. Alignment of P52 and expression in sporozoites. (*A*) Comparison of the deduced amino acid sequences of the *P. yoelii,* Py52, with the homologue in *P. falciparum*, Pf52. The putative signal peptides are underlined with a dashed line. Putative signal peptide cleavage sites are marked with arrowheads (▲, ▼). Conserved cysteine residues of the tandem 6-cys motifs are marked with an asterisk (*). The carboxyl-terminal hydrophobic putative membrane anchor is underlined with a solid line. Identical residues are shaded dark gray. Conserved amino acid changes are shaded light gray, and radical changes are not shaded. (*B*) cDNA blot demonstrating *Py52* expression in midgut sporozoites (Mg Spz) and salivary gland sporozoites (Sg Spz). Sizes are given in kb.

Herein, we have described a survey of genes expressed in the infectious *Plasmodium* salivary gland sporozoite. We have demonstrated that, with a PCR-based amplification of the transcriptome, it is possible to obtain enough cDNA to construct a library for EST sequence acquisition. CS and SSP2/TRAP are highly expressed in the salivary gland sporozoites. On the basis of Western blot analysis of salivary gland sporozoites, CS is more abundant than SSP2/TRAP (data not shown), and this result is in agreement with the number of ESTs for CS (33 ESTs) and SSP2/TRAP (13 ESTs). We do not know whether the low number of ribosomal protein ESTs in the cDNA data set reflects true abundance of transcripts for those proteins in the sporozoite. PCR amplification of cDNA before cloning and sequencing could have biased the representation. Yet, it is possible that the bulk of proteins of the translation machinery are synthesized in the developing oocyst or in midgut sporozoites. The EST data set gives unprecedented insight into sporozoite gene expression, opening up new avenues of exploration. Expression of chorismate synthase in sporozoites is one example. The shikimate pathway was shown to be functional in blood stage *Plasmodium*, and the herbicide glyphosate had a clear inhibitory effect on parasite growth (15). If the shikimate pathway is also operational in sporozoites and EEF, inhibitory drugs (39) could be used to eliminate the preerythrocytic stages, avoiding progression to the blood stage and therefore disease.

The presence of MAEBL in the sporozoite stage raises interesting questions about its function. Binding of MAEBL to erythrocytes suggested that it had a role in merozoite red blood cell invasion (14). It will be worthwhile to investigate whether MAEBL also has a role in mosquito salivary gland and hepatocyte invasion, and therefore acts as a multifunctional parasite ligand in the merozoite and sporozoite stages. Regardless, its dual expression could make MAEBL the target of an inhibitory immune response against erythrocytic and preerythrocytic stages.

We show here that sporozoites express *SPATR*, coding for a putative secreted protein with a degenerate TSR. The CS protein and SSP2/TRAP each carry a TSR, and both proteins have demonstrated roles in sporozoite motility, host cell attachment, and invasion (34, 40–42). TSRs are also present in CS/TRAP-related protein (43), a protein essential for ookinete motility and host cell invasion (44–46).

The 6-cys motif defines a superfamily of proteins that seems to be restricted to the genus *Plasmodium* (35). Where studied, expression of members of this family was restricted to sexual erythrocytic stages. Recently, targeted gene disruption of *P48/45* identified the protein as a male gamete fertility factor (47). We have identified *Py52* and *Pf52* as genes coding for new members of the 6-cys family. *Py52* is expressed in sporozoites, and, like *SPATR*, *Py52* was expressed at higher level in midgut sporozoites than in salivary gland sporozoites. These expression patterns contrast with expression patterns of *SSP2/TRAP* and *CS*, which appeared equally abundant in both sporozoite stages (data not shown). Although we have not yet analyzed SPATR and Py52 protein expression, it is tempting to speculate, based on transcript level, that both proteins may have a role in sporozoite invasion of the mosquito salivary glands.

We have presented and discussed here only an initial analysis of the EST data set and further characterized a few selected examples with emphasis on putative sporozoite ligands for host cell attachment and invasion. A detailed analysis of all ESTs is beyond the scope of this first description. The amount of redundancy present in the EST data set is relatively low. It is therefore likely that the generation of more sequence data will identify novel sporozoite-expressed genes. However, many ESTs do not have significant database matches, and a number of ESTs produce matches with proteins of unknown function. A comprehensive expression analysis will determine which subset of the identified genes is exclusively expressed in the sporozoite stages. Sporozoite-specific genes are amenable to functional genetic analysis because loss-of-function mutants can be isolated and analyzed (48), a tool not yet available for genes essential in the asexual erythrocytic cycle (49). All told, we can now generate more of the urgently needed information about the sporozoite stage, a stage of the complex malaria life cycle that has so far eluded comprehensive experimental study.

Note Added in Proof. Recently, 1,117 additional ESTs were generated. These ESTs are not included in the analysis presented here. The additional ESTs have been deposited in the GenBank dbEST database (accession nos. BG603043–BG604160) and are also available through the *P. yoelii* gene index (http://www.tigr.org/tdb/pygi/).

We thank Tirza Doniger at the New York University School of Medicine Research Computing Resource for bioinformatics support. This work was supported by National Institutes of Health Grant AI-47102, the United Nations Development Program/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases (TDR), the Naval Medical Research Center Work Units 61102AA0101BFX and 611102A0101BCX, and a U.S. Army Medical Research and Material Command Contract (DAMD17-98-2-8005). S.H.I.K. is a recipient of the B. Levine fellowship in malaria vaccinology. We thank the scientists and funding agencies comprising the international Malaria Genome Project for making sequence data from the genome of *P. falciparum* (3D7) public prior to publication of the completed sequence. The Sanger Centre (Hinxton, U.K.) provided sequence for chromosomes 1, 3-9, and 13, with financial support from the Wellcome Trust. A consortium composed of the Institute for Genome Research, along with the Naval Medical Research Center (Silver Spring, MD) sequenced chromosomes 2, 10, 11, and 14, with support from the National Institute of Allergy and Infectious Diseases/National Institutes of Health, the Burroughs Wellcome Fund, and the Department of Defense. The Stanford Genome Technology Center sequenced chromosome 12, with support from the Burroughs Wellcome Fund. The Plasmodium Genome Database is a collaborative effort of investigators at the University of Pennsylvania and Monash University (Melbourne, Australia) supported by the Burroughs Wellcome Fund.

1. Holder, A. A. (1996) in *Malaria Vaccine Development: A Multi-Immune Response Approach*, ed. Hoffman, S. L. (Am. Soc. Microbiol., Washington, DC), pp. 35–75.
2. Nussenzweig, V. & Nussenzweig, R. S. (1989) *Adv. Immunol.* **45,** 283–334.
3. Nussenzweig, R. S. & Nussenzweig, V. (1989) *Rev. Infect. Dis.* **11,** S579–S585.
4. Schofield, L., Villaquiran, J., Ferreira, A., Schellekens, H., Nussenzweig, R. S. & Nussenzweig, V. (1987) *Nature (London)* **330,** 664–666.
5. Schofield, L., Ferreira, A., Altszuler, R., Nussenzweig, V. & Nussenzweig, R. S. (1987) *J. Immunol.* **139,** 2020–2025.
6. Ferreira, A., Schofield, L., Enea, V., Schellekens, H., van der Meide, P., Collins, W. E., Nussenzweig, R. S. & Nussenzweig, V. (1986) *Science* **232,** 881–884.
7. Sinnis, P. & Nussenzweig, V. (1996) in *Malaria Vaccine Development: A Multi-Immune Response Approach*, ed. Hoffman, S. L. (Am. Soc. Microbiol., Washington, DC), pp. 15–33.
8. Hoffman, S. L., Franke, E. D., Hollingdale, M. R. & Druilhe, P. (1996) in *Malaria Vaccine Development: A Multi-Immune Response Approach*, ed. Hoffman, S. L. (Am. Soc. Microbiol., Washington, DC), pp. 35–75.
9. Charoenvit, Y., Leef, M. F., Yuan, L. F., Sedegah, M. & Beaudoin, R. L. (1987) *Infect. Immun.* **55,** 604–608.
10. Rogers, W. O., Malik, A., Mellouk, S., Nakamura, K., Rogers, M. D., Szarfman, A., Gordon, D. M., Nussler, A. K., Aikawa, M. & Hoffman, S. L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9176–9180.
11. Robson, K. J., Hall, J. R., Jennings, M. W., Harris, T. J., Marsh, K., Newbold, C. I., Tate, V. E. & Weatherall, D. J. (1988) *Nature (London)* **335,** 79–82.
12. Fidock, D. A., Bottius, E., Brahimi, K., Moelans, I. M. D., Aikawa, M., Konings, R. N., Certa, U., Olafsson, P., Kaidoh, T., Asavanich, A., *et al*. (1994) *Mol. Biochem. Parasitol.* **64,** 219–232.
13. Daubersies, P., Thomas, A. W., Millet, P., Brahimi, K., Langermans, J. A. M., Ollomo, B., Mohamed, L. B., Slierendregt, B., Eling, W., Van Belkum, A., *et al*. (2000) *Nat. Med.* **6,** 1258–1263.
14. Kappe, S. H. I., Noe, A. R., Fraser, T. S., Blair, P. L. & Adams, J. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1230–1235.
15. Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tzipori, S., Ferguson, D. J. P., Chakrabarti, D. & McLeod, R. (1998) *Nature (London)* **393,** 801–805.
16. Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., *et al*. (1998) *Science* **282,** 1126–1132.
17. Matuschewski, K., Mota, M. M., Pinder, J. C., Nussenzweig, V. & Kappe, S. H. I. (2001) *Mol. Biochem. Parasitol.* **112,** 157–161.
18. Wan, K. L., Blackwell, J. M. & Ajioka, J. W. (1996) *Mol. Biochem. Parasitol.* **75,** 179–186.
19. Ajioka, J. W., Boothroyd, J. C., Brunk, B. P., Hehl, A., Hillier, L., Manger, I. D., Marra, M., Overton, G. C., Roos, D. S., Wan, K. L., *et al*. (1998) *Genome Res.* **8,** 18–28.
20. Strong, W. B. & Nelson, R. G. (2000) *Mol. Biochem. Parasitol.* **107,** 1–32.
21. Chakrabarti, D., Reddy, G. R., Dame, J. B., Almira, E. C., Laipis, P. J., Ferl, R. J., Yang, T. P., Rowe, T. C. & Schuster, S. M. (1994) *Mol. Biochem. Parasitol.* **66,** 97–104.
22. Pinder, J. C., Fowler, R. E., Dluzewski, A. R., Bannister, L. H., Lavin, F. M., Mitchell, G. H., Wilson, R. J. & Gratzer, W. B. (1998) *J. Cell. Sci.* **111,** 1831–1839.
23. Margos, G., Siden-Kiamos, I., Fowler, R. E., Gillman, T. R., Spaccapelo, R., Lycett, G., Vlachou, D., Papagiannakis, G., Eling, W. M., Mitchell, G. H. & Louis, C. (2000) *Mol. Biochem. Parasitol.* **111,** 465–469.
24. Heintzelman, M. B. & Schwartzman, J. D. (1997) *J. Mol. Biol.* **271,** 139–146.
25. Heintzelman, M. B. & Schwartzman, J. D. (1999) *Cell Motil. Cytoskeleton* **44,** 58–67.
26. Bonhomme, A., Bouchot, A., Pezzella, N., Gomez, J., Le Moal, H. & Pinon, J. M. (1999) *FEMS Microbiol. Rev.* **23,** 551–561.
27. Kieschnick, H., Wakefield, T., Narducci, C. A. & Beckers C. (2001) *J. Biol. Chem.* **276,** 12369–12377.
28. Cassaing, S., Fauvel, J., Bessieres, M. H., Guy, S., Seguela, J. P. & Chap, H. (2000) *Int. J. Parasitol.* **30,** 1137–1142.
29. Mota, M. M., Pradel, G., Vanderberg, J. P., Hafalla, J. C. R., Frevert, U., Nussenzweig, R. S., Nussenzweig, V. & Rodriguez, A. (2001) *Science* **291,** 141–144.
30. Kappe, S. H. I., Curley, G. P., Noe, A. R., Dalton, J. P. & Adams, J. H. (1997) *Mol. Biochem. Parasitol.* **89,** 137–148.
31. Adams, J. H., Sim, B. K. L., Dolan, S. A., Fang, X., Kaslow, D. C. & Miller, L. H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7085–7089.
32. Lawler, J. & Hynes, R. O. (1986) *J. Cell Biol.* **103,** 1635–1648.
33. Adams, J. C. & Tucker, R. P. (2000) *Dev. Dyn.* **218,** 280–299.
34. Gantt, S. M., Clavijo, P., Bai, X., Esko, J. D. & Sinnis, P. (1997) *J. Biol. Chem.* **272,** 19205–19213.
35. Templeton, T. J. & Kaslow, D. C. (1999) *Mol. Biochem. Parasitol.* **101,** 223–227.
36. Carucci, D. J. & Hoffman, S. L. (2000) *Nat. Med.* **6,** 1–6.
37. Hayward, R. E., Derisi, J. L., Alfadhli, S., Kaslow, D. C., Brown, P. O. & Rathod, P. K. (2000) *Mol. Microbiol.* **35,** 6–14.
38. Mamoun, C. B., Gluzman, I. Y., Hott, C., MacMillan, S. K., Amarakone, A. S., Anderson, D. L., Carlton, J. M.-R., Dame, J. B., Chakrabarti, D., Martin, R. K., *et al*. (2001) *Mol. Microbiol.* **39,** 26–36.
39. McConkey, G. A. (1999) *Antimicrob. Agents Chemother.* **43,** 175–177.
40. Sinnis, P. (1996) *Infect. Agents Dis.* **5,** 182–189.
41. Sultan, A. A., Thathy, V., Frevert, U., Robson, K. J., Crisanti, A., Nussenzweig, V., Nussenzweig, R. S. & Ménard, R. (1997) *Cell* **90,** 511–522.
42. Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V. & Ménard, R. (1999) *J. Cell Biol.* **147,** 937–944.
43. Trottein, F., Triglia, T. & Cowman, A. F. (1995) *Mol. Biochem. Parasitol.* **74,** 129–141.
44. Dessens, J. T., Beetsma, A. L., Dimopoulos, G., Wengelnik, K., Crisanti, A., Kafatos, F. C. & Sinden, R. E. (1999) *EMBO J.* **18,** 6221–6227.
45. Yuda, M., Sakaida, H. & Chinzei, Y. (1999) *J. Exp. Med.* **190,** 1711–1716.
46. Templeton, T. J., Kaslow, D. C. & Fidock, D. A. (2000) *Mol. Microbiol.* **36,** 1–9.
47. van Dijk, M. R., Janse, C. J., Thompson, J., Waters, A. P., Braks, J. A. M., Dodemont, H. J., Stunnenberg, H. G., Van Gemert, G.-J., Sauerwein, R. W. & Eling, W. (2001) *Cell* **104,** 153–164.
48. Ménard, R. & Janse, C. (1997) in *Methods: A Companion to Methods in Enzymology—Analysis of Apicomplexan Parasites* (Academic, Orlando, FL), Vol. 13, pp. 148–157.
49. De Koning-Ward, T. F., Janse, C. J. & Waters, A. P. (2000) *Annu. Rev. Microbiol.* **54,** 157–185.