Genes for Glycosylphosphatidylinositol Toxin Biosynthesis in *Plasmodium falciparum*

Mauro Delorenzi, Adrienne Sexton, Hosam Shams-Eldin, Ralph T. Schwarz, Terry Speed, and Louis Schofield

The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia, and Medizinisches Zentrum fur Hygiene und Medizinische Philips Universita¨t Mikrobiologie, Institut fur Virologie, Marburg 35037, Germany

Received 25 January 2002/Returned for modification 21 March 2002/Accepted 16 April 2002

About 2.5 million people die of *Plasmodium falciparum* malaria every year. Fatalities are associated with systemic and organ-specific inflammation initiated by a parasite toxin. Recent studies show that glycosylphosphatidylinositol (GPI) functions as the dominant parasite toxin in the context of infection. GPIs also serve as membrane anchors for several of the most important surface antigens of parasitic invasive stages. GPI anchoring is a complex posttranslational modification produced through the coordinated action of a multi-component biosynthetic pathway. Here we present eight new genes of *P. falciparum* selected for encoding homologs of proteins essential for GPI synthesis: PIG-A, PIG-B, PIG-M, PIG-O, GPI1, GPI8, GAA-1, and DFM1. We describe the experimentally verified mRNA and predicted amino acid sequences and in situ localization of the gene products to the parasite endoplasmic reticulum. Moreover, we show preliminary evidence for the PIG-L and PIG-C genes. The biosynthetic pathway of the malaria parasite GPI offers potential targets for drug development and may be useful for studying parasite cell biology and the molecular basis for the pathophysiology of parasitic diseases.

*Plasmodium falciparum* malaria ranks along with human immunodeficiency virus disease and tuberculosis as one of the most serious infectious diseases of humanity (81). It infects 5 to 10% of the global population and kills approximately 2.5 million people annually. Malarial fatalities are strongly associated with an exacerbated systemic or organ-specific inflammatory cascade, which is initiated by a parasite toxin. This toxin induces the expression of proinflammatory cytokines tumor necrosis factor alpha and interleukin-1 from macrophages and may directly activate the vascular endothelium. (56). Furthermore, both cytokines and malarial toxins can each directly induce the expression of proinflammatory loci such as that for inducible nitric oxide synthase (66), thereby raising levels of nitric oxide, which may be a regulator of pathogenesis (8). The syndromes seen in severe *P. falciparum* malaria in African children and in nonimmune adults may consist of organ-specific as well as multiorgan and system disturbances, including fever, metabolic acidosis, hypoglycemia, shock, and jaundice as well as renal failure, pulmonary edema, and cerebral involvement, including seizures and coma. Despite their diversity, these various signs, symptoms, and syndromes are thought to represent in part manifestations of an underlying inflammatory cascade driven by a parasite toxin.

Recent studies suggest that the glycosylphosphatidylinositol (GPI) of parasite origin functions as the dominant malarial toxin in the context of infection (56, 57, 66, 67, 68). GPIs of *Trypanosoma brucei* have the same function (68). These findings have been confirmed by others (42) and extended to *Trypanosoma cruzi* (2). These data support the view (57) that GPIs of the parasitic protozoa are the dominant proinflammatory agents in this class of eukaryotic pathogen.

GPIs are ubiquitous among eukaryotes, having been described for *Trypanosoma* (1, 15, 23), *Plasmodium* (18, 19), *Leishmania* (43, 47, 55), and *Toxoplasma* (73), as well as yeast (13), fish, plants, and numerous mammalian sources (27, 53, 82). Structurally related to phosphatidylglycerol (PI), a membrane phospholipid, they consist of a conserved core glycan (Manα1→2Manβ1→6H9251) linked to position 6 of the inositol ring of PI. The non-N-acetylated glucosamine is a characteristic feature of GPIs. The synthesis consists of sequential additions to PI, and the mannose (Man) units are numbered accordingly. An ethanolamine phosphate (EtNP) is added at position 6 on Man3 and can be used for attachment to a polypeptide.

GPIs are built up in the endoplasmic reticulum (ER) by the sequential addition of sugar residues to PI by the action of glycosyltransferases (65). At some stage during this process, the maturing GPI is translocated across the membrane from the cytoplasmic to the luminal side of the ER by an undefined mechanism (75, 76). After the GPI glycolipid is completed, it may be exported to the cell surface, free or in covalent association with proteins. GPI can be coupled to a protein via an amide bond between the terminal EtNP of the GPI and the protein C terminus formed by the removal of the terminal amino acids of the protein chain, the GPI signal sequence (74). The tetrasaccharide core glycan may be further substituted with sugars, phosphates, and ethanolamine groups in a species- and tissue-specific manner. GPI fatty acid moieties can be either saturated or unsaturated diacylglycerols, alkylacylglycerols, monoalkylglycerols, or ceramides, with additional acyl modifications to the inositol ring, variously C14:0, C16:0, C18:0.
C16:1 and C18:2. The overall picture is of a closely related family of glycolipids sharing certain core features but with a high level of variation in fatty acid composition and side chain modifications to the conserved core glycan. The significance of the structural differences among GPIs is not yet clear, although structure-activity studies on GPI toxicity demonstrate important functional differences among GPIs with different fatty acid compositions (2, 7, 67).

The structure of the GPI toxin of *P. falciparum* has been elucidated and shown to consist of EtNP-6(Mano1-2)Mano1-2Mano1-6Mano1-4GlcNol-1(6-acyl-2)myoInsl-1-P-(sn1,2 diacyl)-glycerol (18). The acyl components of GPIs are probably variable; in *P. falciparum* a preference for palmitoyl at the glycerol and myristoyl at the inositol was described (19). Some differences between GPIs of mammals and *Plasmodium* have been identified, as follows. (1) *Plasmodium* GPIs lack any modification of Man2 and Man3, while in mammals and yeast GPIs always carry an EtNP side chain at the position 2 on Man1 and sometimes also carry this side chain at position 6 on Man2. (ii) The Man4 is present in most *Plasmodium* anchors but on only a minority of mammalian GPIs. (iii) The lipid moiety is predominantly 1-alkyl, 2-acyl glycerol in mammalian erythrocytes but is invariably a diacylglycerol in intraerythrocytic *Plasmodium*. However, diacylglycerol-containing mammalian GPIs from kidney and spleen have been described.

Among the GPI-anchored proteins in *Plasmodium*, the most important are the circumsporozoite protein, which coats the sporozoites (49), and the merozoite surface proteins MSP-1 and MSP-2, two leading vaccine candidates that are thought to be of major importance for the infection of red blood cells (25, 26, 62). Only a four-mannose structure was detected for the GPI moieties of MSP-1 and MSP-2 which accumulate during schizogony (19), but a form lacking the fourth mannose might be of major importance for the infection of red blood cells (25,80) (at http://au.expasy.org/cgi-bin/protscale.pl) and a lack of cysteines. We selected oligopeptides with good predicted antigenicity on the basis of hydrophilicity (30,52,80) (at http://au.expasy.org/cgi-bin/protscal.ep) and a lack of cysteines. We chose the following sequences for peptide synthesis: anti-GA1, YNTNRIG KKHRSST, anti-PIG-O, DKKLKLKTNLNEEN, anti-PIG-A, GKVKENVKNILOTGH; anti-PIG-B, NEDNKRKEDDENNGN; and anti-DPM1, HKYIYIFKIKOREKN. Peptides were coupled to diphtheria toxoid, and 50 µg was emulsified in Freund’s complete adjuvant and used to immunize mice, followed at 4-week intervals with two boosts of equal doses in Freund’s incomplete adjuvant. Sera were collected into heparin and screened for reactivity to *P. falciparum* by an indirect fluorescent-antibody test. Thin films of parasites at mature stages were fixed in cold acetone and incubated with 1:80 dilutions of antibody. Localization to the parasite ER was determined by counterstaining with rabbit polyclonal antibodies to *P. falciparum* ERC1 (endoplasmic reticulum-located, calcium-binding protein [37]). The slides were washed extensively in phosphate-buffered saline and incubated with an appropriate mixture of fluorescein-conjugated anti-mouse and rhodamine-conjugated anti-rabbit antibodies.
Preimmune sera served as negative controls. Slides were photographed under appropriate illumination for fluorescein isothiocyanate and rhodamine.

RESULTS AND DISCUSSION

The aim of the project was to identify the full set of genes involved in GPI biosynthesis in \textit{Plasmodium}. We summarize the present state of the evidence that we gathered for these genes. Kinoshita and Inoue (34) have subdivided mammalian GPI biosynthesis into 11 steps, and we have adapted their diagram for \textit{P. falciparum} GPI biosynthesis (Fig. 1); based on the known structure of \textit{P. falciparum} GPI, addition of EtNP to Man1 and Man2 is not expected, while an extra step for the addition of Man4 is included, resulting in 10 steps. Data suggest that steps 8 and 9 can occur in either order in \textit{P. falciparum}. In this paper we present the sequences of proteins needed for steps 1 (PIG-A and GPI1), 5 (PIG-M), 7 (PIG-B), 9 (PIG-O), and 10 (GPI8 and GAA-1) and for the generation of the mannose donor dolichol-phosphate-mannose (Dol-P-Man) (DPM1). A candidate ortholog for step 2 (deacetylation of N-glucosamine by human PIG-L or yeast GPI12) was found very recently and has not yet been characterized. For steps 3 (inositol acylation), 4 (transport), and 6 (second mannosylation), no genes have been discovered in any species. No candidate gene for the addition of Man4 in step 8 (performed in yeast by Smp3) could be identified.

The protein and nucleotide sequences derived from our analysis are available at our website (http://www.wehi.edu.au/bioweb/Mauro/GPI). Table 1 gives information on the evolutionary conservation and lengths of the proteins, as well as the chromosome localization and structures of eight key genes. The proteins have evolved rather slowly, with a 40 to 70\% similarity, which allows recognition by sequence, although non-conserved regions cannot be identified by similarity alone and required the exact determination of the position of the introns. PIG-A and DPM1 are strongly conserved proteins; GAA-1 is the most variable. Typically the similarities between the human and yeast sequences are, in accordance with the phylogenetic tree of the species, marginally higher than those with \textit{Plasmodium}, except for DPM1 (Table 1). The \textit{PIG-O} gene probably belongs to chromosome 12 but is also present in the BLOB file (chromosomes 6 to 8 as designated by the Malaria Genome Consortium).

The sequences of the experimentally determined intron-exon boundaries and their flanking sequences are shown in Table 2. The predictions based on the multiple alignment were almost perfectly correct. The localization of the start codon is
resolved and are referred to as the Blob.

At least 51% of the aligned sequences; otherwise, positions are indicated by asterisks.

CT-AC second intron in the PIG-O gene. This is possibly the first time a CT-AC splice donor has been observed. In general, there is a high frequency of thymine in the last 40 bases of the introns (Table 2, splice donor).

Identity and similarity were computed for the optimal global pairwise alignment driven by scoring matrix BLOSUM45 with gap penalties 4 and 12. Similarity is defined by a positive score. The percentage is calculated with respect to the length of the shorter sequence. Accession numbers are provided in the figure legends.

Table 1: Lengths numbers of introns, and chromosomal locations of P. falciparum genes involved in GPI biosynthesis and sequence similarities to homologs in human and yeast

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosomea</th>
<th>No. of introns</th>
<th>Product length (amino acids)</th>
<th>% Amino acid identity (similarity)b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P. falciparum vs. H. sapiens</td>
<td>P. falciparum vs. S. cerevisiae</td>
</tr>
<tr>
<td>PIG-A</td>
<td>10</td>
<td>6</td>
<td>461</td>
<td>46 (66)</td>
</tr>
<tr>
<td>GPI1</td>
<td>Blob</td>
<td>8</td>
<td>669</td>
<td>20 (46)</td>
</tr>
<tr>
<td>PIG-M</td>
<td>12</td>
<td>7</td>
<td>441</td>
<td>37 (62)</td>
</tr>
<tr>
<td>PIG-O</td>
<td>13</td>
<td>0</td>
<td>786</td>
<td>25 (50)</td>
</tr>
<tr>
<td>PIG-O</td>
<td>12</td>
<td>2</td>
<td>1,238</td>
<td>24 (47)</td>
</tr>
<tr>
<td>GAA1</td>
<td>13</td>
<td>3</td>
<td>700</td>
<td>19 (40)</td>
</tr>
<tr>
<td>GPI8</td>
<td>11</td>
<td>0</td>
<td>493</td>
<td>30 (54)</td>
</tr>
<tr>
<td>DPM1</td>
<td>11</td>
<td>0</td>
<td>259</td>
<td>49 (70)</td>
</tr>
</tbody>
</table>

a Chromosome assignments are based on database information, using the system provided by the Malaria Genome Consortium. Chromosomes 6 to 8 are not yet resolved and are referred to as the Blob.

b Identity and similarity were computed for the optimal global pairwise alignment driven by scoring matrix BLOSUM45 with gap penalties 4 and 12. Similarity is defined by a positive score. The percentage is calculated with respect to the length of the shorter sequence. Accession numbers are provided in the figure legends.

Table 2: Sequences of intron-exon boundaries of the P. falciparum genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Intron Length (bp)</th>
<th>Splice donor</th>
<th>Sequencec</th>
<th>Splice acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAA-1</td>
<td>172</td>
<td>gattgtaaag</td>
<td>gtaatcata</td>
<td>ccaacatga</td>
</tr>
<tr>
<td></td>
<td>318</td>
<td>gtaacaaaaa</td>
<td>aaatagat</td>
<td>tatattttact</td>
</tr>
<tr>
<td></td>
<td>219</td>
<td>gtaatgtaacc</td>
<td>cttaaataaa</td>
<td>tatatttttttaa</td>
</tr>
<tr>
<td>PIG-O</td>
<td>160</td>
<td>gtaatacaaat</td>
<td>ccacacatta</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>gtaagtaacc</td>
<td>ccacacatta</td>
<td>tataatttata</td>
</tr>
<tr>
<td>PIG-A</td>
<td>209</td>
<td>gtaatacaaat</td>
<td>ccacacatta</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>123</td>
<td>gtaacagcaac</td>
<td>aaatataaa</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>gtttatcatac</td>
<td>gtttatcata</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>gaaagacagcc</td>
<td>aataataataa</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>86</td>
<td>aataataataa</td>
<td>aataataataa</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>147</td>
<td>gaaagacagcc</td>
<td>aataataataa</td>
<td>tataatttata</td>
</tr>
<tr>
<td>PIG-M</td>
<td>150</td>
<td>gtaagtattc</td>
<td>aataatatgt</td>
<td>tttaaaaaa</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>gtaatcagcaag</td>
<td>gagaataaa</td>
<td>tataataataa</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>gtttttacaag</td>
<td>gtttatcatac</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>179</td>
<td>gtttatcatac</td>
<td>gtttatcatac</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>169</td>
<td>gtttttacaa</td>
<td>gtttatcatac</td>
<td>tataatttata</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>gtttttacaa</td>
<td>gtttatcatac</td>
<td>tataatttata</td>
</tr>
</tbody>
</table>

Consensus ********AG gtaaa*aaaa aa*aa*aaa ttt**t**tt ttttttttttt ttttttttttt ttttttttttt ttttttttttt
provided by PIG-A, whose yeast ortholog (GPI3) has been shown to bind the substrate UDP-GlcNAc (36). Its sequence reveals motifs shared with a large family of glycosyltransferases, which transfer activated sugars from different nucleotide carriers to a variety of substrates, including bacterial lipopolysaccharides. These motifs are also encoded in the P. falciparum gene that we have identified. The best conserved block, possibly encompassing the active site, is shown in Fig. 3A.

The GPI1 gene of P. falciparum was identified by its ability to complement a GPI1-defective yeast strain (cDNA clone with GenBank accession number AJ249657) (58). A comparison between the cDNA and the genomic sequence reveals eight introns and the apparent existence of a microexon 11 nucleotides in length (not shown).

With the exception of what might be a fragment of PIG-C (Fig. 3C), no homolog for the other mammalian genes implicated in this step (PIG-H, PIG-P, and DPM2) could be found. The roles of these genes are not yet understood. The genes themselves are not vertebrate specific, as we found convincing sequence homologs for all of them in nonvertebrate and non-animal species (not shown), such as Schizosaccharomyces pombe. DPM2 was first identified as a gene required for assisting the transfer of mannose units from dolichol phosphate by the catalytic DPM1 (40), suggesting that it actually plays an indirect accessory role in both reactions. As a homolog gene has not been identified in the completed sequence of the genome of Saccharomyces cerevisiae, DPM2 appears to be dispensable for GPI synthesis in some species, so it is plausible that Plasmodium may not require the PIG-H, PIG-P, and DPM2 genes, although a final judgment is not yet possible.

(ii) Step 2: deacetylation to GlcN-PI. The second reaction also takes place on the cytoplasmic side of the ER, and the catalyzing enzyme, N-acetylgalactosaminylphosphatidylinositol de-N-acetylase, is encoded by the PIG-L gene in humans (79). When we started our work we found only a short expressed
sequence tag sequence of *Plasmodium berghei* that potentially encodes 60 amino acids with significant similarity to yeast and human *PIG-L*, particularly in a dodecamer motif, **AHPDDEXMFFXP** (Fig. 4). Recently, genomic sequences that contain this presumptive *PIG-L* gene have been made available in databases for *P. yoelii*, *P. falciparum*, and *Plasmodium knowlesi*. The best local alignment extends the one shown in Fig. 4 by about 50 amino acids (across an intron) and contains a second motif, **YGVSGHPNHIS**, which is invariant in the three *Plasmodium* species (not shown).

(iii) Steps 5, 7, and 8: addition of mannoses. The three mannose units in the GPI core are linked by **1,4**, **1,6**, and **1,2** bonds. The three enzymes are not expected to be closely related. The fourth mannose is added in **1,2** linkage like the third one. While the human Man3 transferase (PIG-B) has been known for some time (69), the Man1 and Man4 transferases, called PIG-M and Smp3 (yeast), respectively, have just recently been discovered (21, 41). However, no transferase for Man2 has been described so far. In *P. falciparum* we found genes encoding putative Man1 and Man3 transferases (*PIG-B*).
PIG-M) (Fig. 5), which align very well to their counterparts, but not an ortholog of Smp3 (the most similar gene is PIG-B). Very close relatives of Smp3 can easily be found for various species (S. pombe, Drosophila melanogaster, and Homo sapiens). Assuming that a similar degree of conservation extends to Plasmodium, the failure to identify an Smp3 homolog could be due to its absence in the database to date. The P. falciparum genome sequence is thought to be almost complete, but substantial fragments that are difficult to clone could still be missing, at least for chromosomes 6 to 8 (the "Blob"). Alternatively, a different gene might perform this function in Plasmodium.

The mannosyltransferases form a superfamily. Interestingly, the general organization of the PIG-M protein is similar to that of PIG-B and Smp3, which is presumably a sign of a common origin, although conservation at the amino acid level has been lost completely over evolutionary time. The PIG-B, Smp3, and PIG-M proteins all have multiple potential transmembrane domains interspersed with short hydrophilic loops, most of which are very short. The first transmembrane domain is followed by the longest hydrophilic region (about 30 amino acids). As an exception, the P. falciparum PIG-B has two additional long hydrophilic insertions. This tendency to have additional hydrophilic stretches has been observed in other proteins of P. falciparum, and we have verified that these are not intronic sequences. The first transmembrane domain contains a characteristic arginine in the middle, which is conserved in 17 of 18 PIG-B/PIG-M/Smp-3 homologs that we have aligned (not shown). In each of the three families, other hydrophobic stretches also show one conserved charged or hydrophilic residue. These residues might be engaged in strong inter- or intramolecular contacts in the membrane. The presence of one or several hydrophilic residues within a transmembrane domain is a common feature of ER-resident proteins that contributes to their localization, and the efficacy is stronger for amino acids D and R and when the hydrophilic residues are positioned in the middle of a transmembrane domain (6, 38, 39).

There are structural similarities between the α-1,2-mannosyltransferases. Database analysis suggests that the Smp3 gene is very conserved, with five or six well-defined motifs (not shown). Two of them correspond to the only two blocks of strong similarity in the PIG-B genes (Fig. 5). Among many others, three strongly related proteins, with National Center for Biotechnology Information identifiers 1302525 (YNR030w, S. cerevisiae), 3738170 (S. pombe), and 12804615 (H. sapiens), whose function seems to be unknown, also have a similar motif and a general similarity to the PIG-B sequences and therefore probably form another family of related but yet-uncharacterized glycosyltransferases.

In the PIG-M α-1,4-mannosyltransferases, a few sequence blocks are very well conserved and likely encode a critical

FIG. 5. Two well-conserved sequence blocks characteristic of the α-1,2-mannosyltransferase PIG-B. The top line represents the consensus sequence (CS). The species are A. thaliana (At), H. sapiens (Hs), S. cerevisiae (Sc), S. pombe (Sp), T. brucei (Tb), Zymomonas mobilis (Z), and P. falciparum (Pf). The Zymomonas protein is a bacterial sequence homolog of unproven function. It serves to tentatively identify the best-conserved positions, which are more likely to be essential for function. (A) The best-conserved sequence block in PIG-B follows the first transmembrane domain and is one of the few hydrophilic segments. Of the 40 positions, only 8 are invariant (12 without Zymomonas), but the consensus can be defined for most positions (27). The bottom line shows that a homologous block of sequence is also present in the Smp3 glycosyltransferase of yeast, with most residues similar and 11 identical to the consensus sequence. (B) The second well-conserved block of PIG-B and Smp3 is also present in other α-1,2-mannosyltransferases, for example, in yeast ALG9 (accession number NP 014180). Accession numbers for PIG-B are CAC01884 (At), BAA07709 (Hs), CAA96854 (Sc), CAB53078 (Sp), BAA94863 (Tb), and AAD53921 (Z); that for yeast Smp3 is NP 014792. Symbols and figure preparation are as described in the legend to Fig. 3.
function. The PIG-M proteins are slightly shorter than PIG-B and Smp3. Again, the first transmembrane domain, with the conserved R, is followed by one of the best conserved motifs (Fig. 6), embedded in a relatively hydrophilic short loop. In contrast to the case for PIG-B, there is also some degree of conservation in the terminal parts of the PIG-M proteins. Most notably, all of six presumptive PIG-M sequences (those in Fig. 6 and one from Caenorhabditis elegans), have a lysine as the third-to-last amino acid (not shown). Lysines near the carboxy terminus and particularly at position \( H11002 \) have been implicated in the ER retention of proteins, especially a dilysine motif in type Ia ER membrane proteins (33). The lysines are preferably located at the third- and fourth-to-last positions (KKXX-COOH motif), but variations in position and some replacements with arginine are not uncommon (60, 72). A second positively charged K or R residue is indeed present at position \( H11002 \) in many (but not all) PIG-M proteins, reinforcing the hypothesis of a function as a retention signal. In contrast to the similarities at the protein level, the gene structures of \( P. falciparum \) PIG-B and PIG-M differ considerably in that the first gene has no introns while the second has seven.

(iv) Step 8: addition of EtNP. Preliminary evidence produced in different laboratories indicates that the three enzymes that add the EtNP groups are PIG-N/MCD4 (EtNP addition to Man1), Gpi7 (Man2), and PIG-O/GPI13 (Man3) (4, 16, 17, 28, 70, 83). In \( P. falciparum \) only the terminal EtNP that serves as a bridge to the protein is present. In addition, mammals also require PIG-F (31), which has an unknown biochemical function. In a phylogenetic tree with sequences from \( H. sapiens \), \( S. cerevisiae \), \( S. pombe \), C. elegans, and D. melanogaster, the phosphoesterases cluster according to orthologous genes rather than taxonomic group (not shown). This suggests that development of the ability to add EtNP to all three mannoses preceded the evolutionary splitting of yeasts and animals, since convergent evolution seems implausible. In \( P. falciparum \) we identified, as expected, exactly one gene of the family. It clusters with the PIG-O subfamily. This reinforces the view that PIG-O is the Man3 phosphoesterase. It also suggests that the three genes existed before the splitting of Opisthokonta, Apicomplexa, and Plantae in early eukaryote evolution (3) and that two genes were lost in the evolutionary history of \( P. falciparum \). A loss could be explained either as an adaptation to parasitic life and rapid growth or as a more specific selection of some biochemical property with functional importance.

The three phosphoesterases are all large proteins of approximately 100 to 120 kDa and 800 to 1100 amino acids, and they share a common organization, with a short hydrophobic segment (probably a signal sequence) near the N-terminal end, a large hydrophilic N-terminal half, and a hydrophobic second half that includes many potential transmembrane domains. Short conserved sequence motifs have been identified in the hydrophilic half (4, 16). The order of the three motifs is con-
served, but in *P. falciparum* the highest similarity to the third one is shifted by about 100 amino acids due to a hydrophilic insertion. Motifs 1 and 2 (Fig. 7) are part of pattern 01663 in the Pfam motif library (63) derived from type I phosphodiesterases and nucleotide pyrophosphatases that catalyze the cleavage of phosphodiester bonds in NAD, deoxynucleotides, and nucleotide sugars. One might expect that motif 1 is involved in binding the EtNP and that motif 2 is involved in catalysis. Motif 3 is more specific for the phosphodiesterases of GPI synthesis (Fig. 7C). A partially similar block can be found in some other enzymes, i.e., phosphohexose mutase, phosphoglucomutase, and nucleoside diphosphate kinase (not shown). The meaning of this observation is unclear, but it might suggest a role in binding the GPI substrate. The second of many conserved hydrophobic stretches in the PIG-O gene products has very strongly conserved D and R residues with consensus O6-D-(GA)-L-R-O-D-O3, where O represents a hydrophilic amino acid. As noted above for the mannosyltransferases, this might be a transmembrane domain with charged residues engaged in protein interactions and/or involved in ER retention.

(v) **Step 10: covalent linking to the protein (transamidation).** The transamidation reaction requires at least two proteins, GAA-1 and GPI8 (5, 22, 84), which form a complex in mammalian cells (50). For both proteins we have identified the *P. falciparum* ortholog. GPI8 is most probably the catalytic subunit, as it associates with substrate proteins (64, 77). It has homologies to proteinases (5, 12), and it is related to the caspase family of cysteine proteases. Mutational analysis has identified a cysteine and a histidine as being likely components of the active site and essential for the transamidation reaction (44, 50). Another common feature is the potential for a transmembrane domain near the C terminus. A cDNA for GPI8 which complements a GPI8 mutant yeast strain was obtained (H. Shams-Eldin et al., unpublished data). It has the expected amino acid sequence features (Fig. 8). A comparison with genomic DNA reveals that the gene has no introns.

*P. falciparum* GAA1 is a gene with three introns that encodes a protein of 700 amino acids. GAA1 proteins are probably required for correct localization of GPI8 and are very hydrophobic. While this property is conserved, their sequence has diverged considerably, indicating a rather unspecific biochem-
ical role. In a multiple alignment, only one or two short conserved blocks could be identified (Fig 8).

Two additional components of the transamidase complex were recently identified in humans and yeast: PIG-S/GPI17 and PIG-T/GPI16 (50). Homologous sequences can easily be found for animal and yeast species. The sequence of PIG-T appears to be under strong selection pressure, as it is highly conserved. A coding region with similarity to human PIG-T can be found in plasmodia. It is strongly conserved between P. falciparum, P. knowlesi, and P. yoelii, but the level of similarity to the other sequences is too low to identify it as a likely PIG-T ortholog (data not shown).

(vi) Auxiliary step: synthesis of the mannose donor Dol-P-Man. DPM1 is the Dol-P-Man synthase that catalyzes the production of Dol-P-Man (51). The protein has been well characterized for many species. There is a P. falciparum sequence with strong similarity (Fig. 9). A very good alignment with almost no gaps is obtained with a conceptual translation of the genomic sequence, strongly suggesting that the gene has no introns. An uncertainty remains as to the transcriptional start site, as there are two AUG codons in an appropriate position, the first of which has been used for Fig. 9. As noted before (9), a phylogenetic tree curiously splits the proteins in two clusters (not shown), with those with a presumptive C-terminal transmembrane domain (Saccharomyces, Ustilago, Leishmania, and Trypanosoma) on one side and those without it (Homo, Caenorhabditis, Schizosaccharomyces, Arabidopsis, and Plasmodium) on the other side, while each species seems to have only one gene. This pattern is not easy to interpret, as Plasmodium is evolutionarily closer to Kinetoplastida and Schizosaccharomyces would be expected to cluster with other yeasts and fungi. It suggests convergent evolution in different evolutionary lines.

Conclusions. To identify systematically the genes involved in the GPI pathway in Plasmodium, we used critical evaluation of amino acid sequence similarity data, exon prediction, confirmation by PCR, and protein localization. With the genome sequences at hand, it is thus possible rapidly to identify many candidates for further research. Except for Smp3, it seems that the genes for which no ortholog was identified are those likely to have a role more in stabilization or regulation than in catalysis. Either their sequences have diverged to the point of

FIG. 8. (A and B) Alignment of the two best-conserved blocks characteristic of GAA-1. The protein evolved considerable divergence at the amino acid level, and only seven positions are invariant in the two aligned blocks (taken together) with the species A. thaliana (At), H. sapiens (Hs), S. cerevisiae (Sc), and P. falciparum (Pf). Accession numbers are CAA55944 (Sc), NP_197414 (At), and AAIH3071 (Hs). (C and D). The histidine (C) and cysteine (D) in the boxed motifs of GPI8 are essential for the transamidation reaction (see text) in other species and are conserved in P. falciparum. Accession numbers are AAB81597 (Hs), P49018 (Sc), T40853 (S. pombe [Sp]), and AJ401202 (Pf). CS, consensus sequence. Symbols and figure preparation are as described in the legend to Fig. 3.

FIG. 9. Global alignment of the DPM1 protein sequences from A. thaliana (At), H. sapiens (Hs), and P. falciparum (Pf). The alignment shows a high degree of overall conservation. The only gaps are a short deletion in the Arabidopsis gene near the presumptive start codon and a one-amino-acid deletion at position 24 in the Plasmodium protein. The second aspartate in the boxed IDDGS motif and the first aspartate in the boxed MDAD motif are seen in a range of N-acetyltransferases that transfer a single sugar unit and that have been suggested to participate in catalysis (53). Also boxed is a conserved GTRY motif, where the R is invariant in a broad family that includes sequences of unknown function annotated as DPM1-like proteins in databases. Accession numbers are AAIH7073 (Hs) and NP_173481 (At). CS, consensus sequence. Symbols and figure preparation are as described in the legend to Fig. 3.


