Structure of the Plasmodium 6-cysteine s48/45 domain

Silvia A. Arredondo ah1, Mengli Caih b, Yuki Takayama b, Nicholas J. MacDonald d, D. Eric Anderson d, L. Aravind eb, G. Marius Clore b, 2, and Louis H. Miller e, 2

Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; Laboratory of Malaria Immunology and Vaccinology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; Laboratory of Chemical Physics and Mass Spectrometry Facility, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892; Laboratory of Malaria Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Contributed by Louis H. Miller, March 15, 2012 (sent for review February 10, 2012)

The s48/45 domain was first noted in Plasmodium proteins more than 15 y ago. Previously believed to be unique to Plasmodium, the s48/45 domain is present in other aconoidasians. In Plasmodium, members of the s48/45 family of proteins are localized on the surface of the parasite in different stages, mostly by glycosylphosphatidylinositol-anchoring. Members such as P52 and P36 seem to play a role in invasion of hepatocytes, and PfS230 and PfS48/45 are involved in fertilization in the sexual stages and have been consistently studied as targets of transmission-blocking vaccines for years. In this report, we present the molecular structure for the s48/45 domain corresponding to the C-terminal domain of the blood-stage protein P12 from Plasmodium falciparum, obtained by NMR. Our results indicate that this domain is a β-sandwich formed by two sheets with a mixture of parallel and antiparallel strands. Of the six conserved cysteines, two pairs link the β-sheets by two disulfide bonds, and the third pair forms a bond outside the core. The structure of the s48/45 domain conforms well to the previously defined surface antigen 1 (SAG1)-related-sequence (SRS) fold observed in the SAG family of surface antigens found in Toxoplasma gondii. Despite extreme sequence divergence, remarkable spatial conservation of one of the disulfide bonds is observed, supporting the hypothesis that the domains have evolved from a common ancestor. Furthermore, a homologous domain is present in ephrins, raising the possibility that the precursor of the s48/45 and SRS domains emerged from an ancient transfer to Apicomplexa from metazoaan hosts.

Plasmodium domain evolution | Escherichia coli periplasmic expression | solution conformation

Malaria is caused by Plasmodium parasites one of which, Plasmodium falciparum, claims the lives of nearly a million children each year in Africa alone. Efforts to develop a malaria vaccine would benefit considerably from a more complete understanding of the structure and function of Plasmodium proteins that play key roles in infection. Plasmodium is transmitted to humans when infected mosquitoes inoculate sporozoites into the skin while taking a blood meal. The sporozoites travel from the skin to the liver, where they invade hepatocytes and replicate tens–of-thousands-fold, differentiating into merozoites that exit the liver and invade erythrocytes. After several rounds of replication in erythrocytes, some parasites differentiate into male and female gametocytes and are subsequently taken up by an Anopheles mosquito in a blood meal. Within the mosquito, fertilization takes place and the resulting ookinetes invade the midgut epithelium, resulting in sporozoites that travel to the salivary glands and are transmitted to the human host, completing the cycle.

The 6-cysteine Plasmodium gamete-surface homology 45 domain originally identified by Williamson et al. (1) (partially defined by Pfam model PF07422) is found in P. falciparum proteins that are expressed in all stages of the parasite life cycle in both the human and mosquito hosts (Table S1) (1–3). The domain might occur in 1–14 copies per protein, either as the sole globular domain or in combination with β-helix–forming hexapeptide repeats, as is the case of sequestrin. Previously believed to be exclusive to Plasmodium (4–6), the domain also exists in proteins found in all members of the aconoidasian (hematozoan) clade of Apicomplexa, which unites the haemosporidians (Plasmodium) and piroplasms. In Plasmodium, the s48/45 family has undergone a lineage-specific expansion, with 12 distinct members (PfS230, PfS48/45, PfS230p, PfS47, P52, P36, Pf41, Pf38, Pf12, Pf12p, Pf92 and sequestrin (not previously included) (Table S1)) encoded by the genome of P. falciparum (2, 7). Conserved homologs are present in the genomes of other Plasmodium species. Most of the proteins containing the s48/45 domain are expressed on the surface of the parasite (8–11). PfS48/45 and PfS230 are current targets of malaria transmission-blocking vaccine development (12, 13).

Three of the s48/45 proteins (PfS48/45, PfS230, and PfS47) are involved in male/female gamete fusion in the mosquito midgut. PfS48/45 is predicted to be glycosylphosphatidylinositol (GPI)-anchored to the gamete surface (14). PfS230 is a soluble protein that associates with the gamete membrane by binding to PfS48/45 (15). Deletion of either PfS48/45 or PfS230 greatly reduces zygote formation (16, 17). It is unknown if the role of PfS48/45 is only to bind to PfS230, which functions directly in gamete fusion, or if both are directly involved in facilitating fusion. PfS48/45 has three s48/45 domains, whereas PfS230 has 14 s48/45 domains (2, 4). Antibodies specific for either protein block transmission to the mosquito, and complement enhances the effect of antibodies specific for PfS230 (9, 18). PfS47, a duplicated and contiguous paralogue of PfS48/45, is found on female gametes; its deletion in rodent Plasmodium berghei, but not in P. falciparum, greatly reduces fertilization in the mosquito (11, 21). The presence of multiple species-specific paralogues of this family might result in backup functions in different lineages within the genus Plasmodium.

Sporozoites express two members of the s48/45 family, P36 and P52, from contiguous, duplicated genes. The functions of these two proteins appear to be different because the deletion of either gene influences the ability of the parasite to efficiently infect hepatocytes, with no effect on their ability to invade salivary gland cells in the mosquito (22, 23). In P. berghei there is also evidence that these genes may be required for invasion of hepatocytes and to promote normal development in the liver. Deletion of Pbs36 and Pbs36p results in sporozoites that are unable to develop beyond a small round body, reflecting a defect in normal invasion (8, 24).

The asexual-stage parasites that replicate within erythrocytes express four members of the s48/45 family (Pf12, Pf38, Pf41, and...
Pf92), all located on the surface of merozoites. Each member has a signal sequence and three of them are GPI-anchored to the merozoite surface (10, 14). Because Pf41 has no GPI anchor signal or transmembrane domain for attachment to the membrane, it is likely to bind another merozoite surface protein. Pf41, Pf38, and Pf12 are strongly recognized by immune sera from naturally infected patients, and Pf38 and Pf92 are under balancing selection (10, 25–27). Attempts to delete Pf92 have been reportedly unsuccessful; however, no information for the deletion of the other genes is available (28). Although a specific function has not been described for any of these proteins, the possibility exists for them to have a role in red blood cell binding.

To date, no experimentally determined molecular structure for any s48/45 domain has been reported, mainly because of protein-expression challenges (29–34). However, considering the importance of structural information in the effective design of subunit vaccine candidates, we present the solution three-dimensional structure of the s48/45 domain obtained by NMR spectroscopy of the C-terminal domain of Pf12 expressed in Escherichia coli.

Results

Production of Soluble Pf12 D2. The C-terminal s48/45 domain of Pf12 (Pf12 D2) was produced in E. coli as a soluble histidine-tagged protein by initially harmonizing the DNA sequence of Pf12 for expression in E. coli based on the procedure published by Angov et al. (35), and then targeting the protein for expression in the bacterial periplasm (Figs. S1 and S2) (35, 36). Only the protein expressed as a soluble fraction, extracted through the osmotic shock procedure as described, was used for our studies. The use of PBS as a buffer was avoided because it led to protein aggregation. Soluble expression was assessed by Western blot analysis probing with α-His antibody (Fig. S3). Following purification by gradient immobilized metal affinity chromatography, size-exclusion FPLC on a Superdex75 column resulted in the elution of a single protein peak, indicating that Pf12 D2 is a monomer in solution (Fig. S4). The protein was more than 95% pure, as judged by Gelcode blue-stained SDS/PAGE electrophoresis (Fig. S5); the expected molecular weight of cleaved Pf12 D2 was 14.5 kDa. The identity of the protein was confirmed by N-terminal sequencing and mass spectrometry. For NMR spectroscopy, the final isotopically labeled product was found to be most stable at concentrations no higher than 7 mg/mL in 25 mM sodium acetate-d3 pH 5.0.

Structural Characterization of Pf12 D2. The solution NMR structure of Pf12 D2 was determined on the basis of 2,361 experimental NMR restraints, including 1,477 NOE-derived interproton distance restraints and 192 residue dipolar couplings. The latter provide long-range orientational information in the form of bond vector orientations relative to an external alignment tensor (37). A summary of the structural statistics is provided in Table S2, and a best-fit superposition of the final simulated annealing structures is shown in Fig. 1A. The three disulfide bonds are uniquely determined from the NMR structure and comprise Cys153-Cys185, Cys199-Cys260, and Cys210-Cys258. The structure comprises three β-sheets and a short helix (residues 168–173) (Fig. 2); the five-stranded β-sheet A (β2, β1, β4, β9, β8) has a 1x, –2x, –2x, 1 topology; the four-stranded β-sheet B (β3, β11, β10, β5) has a 3x, –1, –1, 1 topology; and the two-stranded β-sheet C (β6, β7) comprises a simple antiparallel β-sheet connected by a type I turn. Disulfide bonds A (Cys153-Cys185) and B (Cys199-Cys260) bridge β-sheets A and B connecting strands β1 to β3, and β4 to β10, respectively; disulfide bond C (Cys210-Cys258) bridges the loop connecting strands β5 and β6 to strand β10 in sheet B (Fig. 2B). The core of the protein is packed by an unusually large preponderance of aromatic residues (Fig. 1B). The loop connecting the single helix with strand β3 is disordered (see, for example, the structure superpositions in Fig. 1A, and the low values of the 1H-(15)N heteronuclear NOEs in Fig. 1C).

The electrostatic potential mapped on the molecular surface of Pf12 D2 is shown in Fig. 3. Although there are certainly patches of charged residues on the surface, it can also be seen that most of the protein surface is hydrophobic. In the absence of the D1 domain and knowledge of the interactions between the D1 and D2 domain, it is difficult to speculate as to the potential ligand binding site. Nevertheless, certain features of the surface of Pf12 D2 are worth noting. On the front surface in the view of Fig. 3A, the largely hydrophobic surface is surrounded by a ring of four negative charges comprising Glu206, Glu222, Glu225, and Asp236. In the center of the front face of the protein is a deep hydrophobic pocket, at the bottom of which lies Phe197 and the disulfide bridge formed by Cys210 and Cys258, bounded by Val205, Pro207, Val212, Met235, and His237. It is possible that this pocket may constitute a ligand binding site. The backside of the protein, displayed in Fig. 3B, is mildly concave and comprises a central hydrophobic region surrounded by a partial ring of charged residues comprising Lys255, Glu273, Lys188, Asp193, Glu148, Lys151,
Lys165, Glu171, and Lys175. This type of surface is characteristic of many protein-interaction surfaces (38) and may therefore represent the site of interaction with another domain or protein. The surface of Pf12 D2, however, does not possess the equivalent of the basic groove or acid cap seen in the structurally related surface antigen 1 (SAG1) (39) and SporoSAG (40) proteins (see below), respectively, from Toxoplasma gondii.

**Structural Neighbors of the s48/45 Domain and Its Relationship to the SAG1-Related Sequence Superfamily.** Using fold-prediction methods, it was previously shown that the s48/45 domain is likely to adopt the same fold as the SAG1-related sequence (SRS) domain, a β-sandwich found in the lineage-specifically expanded group of proteins from T. gondii, such as SAG1, SporoSAG, and BSR4 (2, 39). We confirmed this hypothesis using profile-profile comparisons with the HHpred program (41): a comparison of the PSI-BLAST–derived profile of the s48/45 family with that of the SRS superfamily yielded highly significant hits (P = 10⁻¹¹; probability 94%) (Table S3). This finding was further validated by a Dali (42) search of the Protein Data Bank using the structure of Pf12 s48/45 determined in this study, which recovers the SRS domains with z > 5 (Fig. 4). Indeed, 96 residues of SAG D1, 92 residues of SporoSAG D2, and 91 residues of BSR4 D1 can be superimposed onto Pf12 D2 with Cα atomic rms differences of 2.6, 2.9, and 2.7 Å, respectively. However, the percentage sequence identities for the corresponding structurally superimposed regions are only 16%, 13%, and 12%, suggesting rather extensive sequence divergence between the SRS and s48/45 domains, despite their sharing a common structural scaffold (Fig. 4). Structure similarity searches using the DaliLite (43) program indicate that the β-sandwich fold shared by s48/45 and the SRS superfamily is further related to comparable β-sandwich domains found in a diverse array of cell-surface molecules from various organisms (z > 4 in reciprocal structure similarity searches) (44). These molecules include the cysteine protease inhibitors amoebiasin-2 (3m86) from Entamoeba histolytica and chagasin (2nnr) from Trypanosoma cruzi, ephrins (2wo3), which are animal neural signaling molecules, and the copper-binding plastocyanins (2q5b). All these β-sandwich domains are united by the pattern of crossover between the two sheets of the sandwich and the zone of variability, which is restricted to one edge of the β-sandwich (Figs. S6 and S7). The ephrins, s48/45, and SRS display a cysteine at the end of strand 3 of their respective β-sandwich domains (equivalent to the first cysteine participating in disulfide bridge B in the latter two domains) (Fig. S6). However, the partner cysteine on the opposite sheet in the ephrins is not equivalent to that seen in the s48/45 and SRS domains (Figs. S6 and S7). This structurally equivalent, unique disulfide bond B shared by s48/45 and SRS distinguishes them from the rest of the related β-sandwich domains. Disulfide bond A bridges strands β1 and β3 of the sandwich in the s48/45 domain, but in the case of SRS it bridges strand β1 to the final strand β11, which lies adjacent to the second strand in the structure (Figs. 2 and 4). Thus, one of the cysteine pairs forming disulfide bond A appears to have diverged between the s48/45 and
SRS; yet, each disulfide bond mediates a comparable cross-sheet contact in both of these domains.

Discussion

The structure of the C-terminal domain of the protein Pf12 from P. falciparum solved by NMR defines the s48/45 domain, previously considered exclusive to the Plasmodium species. As shown in Fig. 2, this domain features an SRS fold, characterized by a sandwich formed by β-sheets resulting from a mixture of parallel and antiparallel strands (39) that positions the s48/45 domain as a homolog of the SRS family. This family of surface antigens from T. gondii is defined by the presence of an N-terminal secretion signal, a GPI anchor, and a set of conserved amino acids, among which are six cysteines involved in the formation of three disulfide bonds important in maintaining the structure of the SRS fold (40).

To date, more than 160 DNA sequences from T. gondii have been identified as belonging to the SRS superfamily (45), but only the structures for SAG1, SporoSAG, and BSR4 have been solved (39, 40, 46, 47). Although there is some structural divergence among these three proteins, and within each protein the N-terminal (D1) and C-terminal (D2) domains are also structurally distinct, the SRS fold is observed as a common feature in the domains of all three (40) (Fig. 4). Overall, comparison of the structure of Pf12 D2 to the D1 domain of SAG1, D2 of SporoSAG, and D1 of BSR4 (Fig. 4) indicates that despite having a low identity for both the overall sequence (4–9%) and the structurally matched regions (12–16%), the structure is indeed superimposable on the SRS structures with Co rms deviations ranging from 2.6 to 2.9 Å, not a significant difference compared with the deviation reported between the superimposition of the D1 and D2 domains from the crystal structure of SAG1 (2.7 Å) (39). As seen in Fig. 4, the core structure of the SRS fold is evidently conserved in Pf12 D2, but distinct characteristics are also observed in the periphery of the core (Fig. 2) [i.e., a double-stranded β-sheet (β6, β7), a short α-helix, and a disordered loop (residues 174–180)] (Fig. 1). This last feature, indicative of extensive mobility (as demonstrated by the low 1H-{15N} NOE values), could in principle be present in the T. gondii SRS proteins for which the structures are available. However, because these features were derived from crystals, at the moment it is not clear whether this is a particular feature of Pf12, the s48/45 domain, or the SRS fold in general.

A common characteristic of both the SRS and the s48/45 domains is their tendency to occur as tandem paralogous pairs. The two tandem domains forming a pair in s48/45 proteins were respectively labeled as “A-type” and “B-type” domains in the original work defining the s48/45 domains (2.4). The structure of Pf12 D2 reported here corresponds to the B-type or second domain in the pair.

The remarkable conservation of one of the disulfide bonds in all four structures, Pf12 D2, SAG1, SporoSAG, and BSR4 (Fig. 4 and Fig. S8), corresponding to disulfide B (Cys199–Cys260) in Pf12 D2 (Fig. 2), suggests that this structural component is essential in the stability of the SRS fold. A second bridge, disulfide A (Cys153–Cys185), although not spatially matched to the disulfide found in the three structures from T. gondii at the same position, seems to serve the equivalent purpose of keeping the main β-sheets of the sandwich together (Figs. 2 and 4, and Fig. S8). Despite structural differences, three disulfide bonds are positionally conserved within both domains of SAG1, BSR4, and SporoSAG; however, analysis of other SRS sequences predicts that the third disulfide bond might not be strictly conserved, or is even absent in some members of the family (40). In the case of Pf12 D2, the third disulfide bond (Cys210–Cys258) is positioned outside the main fold, bringing the third β-sheet closer to the core (Fig. 2); similar to the T. gondii family, not all of the s48/45 domains may require the presence of this disulfide. Analysis of the sequences from proteins Pf6230, Pf64/45, and Pf12 by Carter et al. (4), leading to the original prediction of the disulfide bond connectivity in the s48/45 domain, showed that with the exception of the first domain of Pf64/45, all predicted domains have an even number of cysteines, most of them anticipated to form three or...
two disulfide bonds, but a few would only form one bond within each domain.

The disulfide bond pattern for the s48/45 domain, where cysteines 1–2, 3–6, and 4–5 are connected, previously proposed independently by Carter et al. through visual inspection and scoring of the sequences mentioned above (4), and by Gerloff et al. through comparative structural homology models of Pf230, Pf48/45, Pf54-7, and Pf12 based on SAG1 as a template (2), is now validated by the disulfide bonds observed between Cys153–Cys185, Cys199–Cys260, and Cys210–Cys258 in the structure of Pf12 D2 (Fig. 2).

The members of the SRS superfamily, developmentally expressed on the surface of T. gondii, are generally considered adhesins involved in host-cell attachment and invasion (45, 48). In particular, SAG1 is a highly immunogenic prototypic member of the family expressed on the invasive tachyzoite, predicted to bind polyanionic ligands and presumed to serve as an immune decoy facilitating chronic infection; anti-SAG1 antibodies reduce parasite invasion in vitro, and SAG1-null parasites showed less virulence in mice experiments (39, 49–51).

In Plasmodium, so far there are four proteins (Pf12, Pf38, Pf41, Pf92) from the s48/45 family that have been validated as merozoite surface proteins. Pf12, Pf38, and Pf41 are readily recognized by immune sera from infected patients from endemic areas (10, 25, 46); however, Pf92 function has been ascribed to any of them yet. The first step in invasion of merozoites into red blood cells is the binding of any part of the merozoite to the red blood cell surface. It is tempting to speculate that one of the s48/45 proteins could possibly be the long-anticipated initiator of the invasion cascade.

From an evolutionary perspective, b-sandwich domains with a structure comparable to the s48/45 and SRS are found in a wide range of cell-surface proteins from bacteria to eukaryotes, including versions from other parasites, such as Entamoeba and trypanosomes. This finding indicates that such domains have an ancient history of being used in extracellular interactions. However, classic members of the s48/45 and SRS family are currently ancient history of being used in extracellular interactions. How-


Supporting Information

**Arredondo et al. 10.1073/pnas.1204363109**

**SI Methods**

The sequence of gene PFF0615c corresponding to the C-terminal domain (bp 520–900, encoding residues 148–274) of the *Plasmodium falciparum* protein Pt12 was codon harmonized for *Escherichia coli* expression following the protocol of Angov et al. (1) (Fig. S1). For expression and purification purposes, the synthetic gene was preceded by a sequence containing a ribosomal binding site and sequences coding for the signal peptide of DsbC for translocation into the periplasm, a hexahistidine tag, and a thrombin cleavage site (Fig. S2). Following digestion with XbaI and XhoI, the construct was ligated into pET-28a and transformed into *E. coli* BL21 Star(DE3) (Invitrogen).

Cells were grown in Luria-Bertani (LB) medium with 50 μg/mL kanamycin at 37 °C and induced at an OD_{600} ~0.6–0.7 by addition of isopropyl β-D-1-thiogalactopyranoside (at a final concentration of 0.01 mM). After 4 h, the soluble periplasmic fraction was extracted by cold osmotic shock, where cell pellets were resuspended in 30 mM Tris, 0.5 mM EDTA, 20% (wt/vol) sucrose, pH 8.5, and incubated shaking at room temperature for 10 min; after centrifugation at 10,460 g for 15 min at 4 °C, cold pellets were resuspended in ice-cold buffer (5 mM Tris, 0.5 mM EDTA, pH 8.5), incubated on ice for 20 min while shaking, and centrifuged at 18,590 × g for 30 min at 4 °C. The soluble protein in the supernatant was further purified by gradient immobilized metal affinity chromatography (IMAC) (25–500 mM imidazole) using a HisTrap FF column (GE Healthcare) and by size-exclusion chromatography (SEC) using a Superdex75 column (GE Healthcare), both steps in 20 mM Tris, 200 mM NaCl, pH 7.5.

For NMR studies, the protein was expressed and purified as above with the following exceptions: cells were uniformly 15N- and 13C-labeled by growing in minimal medium containing 15NH_4Cl and 13C_6-glucose as the sole nitrogen and carbon sources, respectively (2), and expression was induced at an OD_{600} ~0.9 and continued for 7 h. After IMAC purification as described above, the His6 tag was removed by thrombin cleavage and Ni-NTA agarose passage. Thrombin was removed by passage over benzamidine Sepharose. Following size-exclusion chromatography in 25 mM sodium acetate pH 5.0, the protein was concentrated to ~6–7 mg/mL in 25 mM sodium acetate-d3 pH 5.0.

**SI Discussion**

Studies on the s48/45 family suggest the majority of proteins, including sequestrin, which was not reported in the original study (3), contain at least one tandem pair of these domains. This pattern suggests that the two tandem β-sandwich domains of surface antigen 1 (SAG1) protein-related sequence (SRS) and the s48/45 families are likely to function as a pair. Although members of the SRS family have crystallized as both monomers and dimers, it is conceivable that under native conditions the pair of β-sandwich domains facilitates dimerization in a fashion similar to the SAG1 protein (4, 5). It is thus possible that also the domain pair in the s48/45 family forms an intertwined structure, with a central binding pocket formed at the dimer interface. The localization of polymorphism in the s48/45 (3) tends to favor this model, suggesting that the SRS and s48/45 domains might share a binding pocket located at a spatially similar position across the dimer. However, the extreme sequence divergence between the SRS and s48/45 suggests that they are likely to have different binding specificities. Indeed, their independent expansions in different apicomplexan (Fig. S9) lineages point out their potential versatility in mediating a diverse set of extracellular interactions. The SRS domains of the SAG1-related clade have been proposed to bind sulfated proteoglycans found on host cell surfaces (4, 6). The more distantly related ephrins are known to bind carbohydrates (7). At least one protein with s48/45, namely sequestrin, has been proposed to interact with a heavily glycosylated host protein (8). In light of these observations, it would be of interest to determine if at least a subset of the s48/45 might have carbohydrate-binding capacity.

Fig. S1. Harmonized Pf12 D2 DNA sequence (bp 520–900) compared to the full sequence of native Pf12 (PFF0615c) from the 3D7 P. falciparum strain.

Arredondo et al. www.pnas.org/cgi/content/short/1204363109
Fig. S2. DNA sequence (and amino acid translation) of the complete construct used for periplasmic expression of Pf12 D2 in E. coli.

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCTAGATTTAAGGAGGAAAGGTTACCAATAAATGTTGTTACTTTTTGTTGCCGTTTGCACAGGCCTACAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
</tr>
<tr>
<td>CACCAAGCCAATGAGCCGCGCCGCCGAGGCAAGAGGTGACGTTACACCCCAAGCAGCAGCTTCTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
<td>240</td>
<td>250</td>
<td>260</td>
<td>270</td>
</tr>
<tr>
<td>SSKYSHSGLVPRGSEKVKGCDFTT肄E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>290</td>
<td>300</td>
<td>310</td>
<td>320</td>
<td>330</td>
<td>340</td>
<td>350</td>
<td>360</td>
</tr>
<tr>
<td>CTGATAGCCTATAGCTGACCGAAATGACCGCTGGTCAGGAGCCATGCATGTTGTCAGCCTGAACCTAGGAGAAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>380</td>
<td>390</td>
<td>400</td>
<td>410</td>
<td>420</td>
<td>430</td>
<td>440</td>
<td>450</td>
</tr>
<tr>
<td>LIGFPCFCPSNYSEVFDCPVSAPNLSGKHNEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. S3. Western blot analysis of the soluble fraction obtained by osmotic shock from the expression of Pf12 D2 in the E. coli periplasm. The membrane was probed with monoclonal antipolyhistidine HRP-conjugated antibody (Sigma).

Fig. S4. Gel filtration FPLC of Pf12 D2. The purified protein was run on a Superdex75 column.
Fig. S5. SDS-gel electrophoresis of purified Pf12 D2 stained with Gelcode blue reagent (Thermo Fisher Scientific).

Fig. S6. A collection of β-sandwich domains related to the s48/45 domain. The domains are arranged so the shared, structurally equivalent core of the β-sandwich element in each domain is aligned to their respective right edge where their N termini are located. The domains differ from each other primarily in the variable region that is positioned to their respective left edge in the figure. The domains depicted in the figure are: Pf12 D2, 1kzq SAG1 D1, 2wo3-B ephrin A, 3m86 amoebiasin-2, 2nnr chagasin, and 2q5b plastocyanin.


Fig. S7. A structure-based multiple alignment of the selected β-sandwich domains compared with Pf12 D2. 1kzq-A MAJOR SURFACE ANTIGEN P30, SAG1 D1; 2wnk A SPOROZOITE-SPECIFIC SAG PROTEIN, SporoSAG D1; 2jks-A BRADYZOITE SURFACE ANTIGEN, BSR4 D1; 2wo3-B EPHRIN TYPE-A RECEPTOR; 3m86-B AMOEBIASIN-2; 2qs5-A PLASTOCYANIN; 2nnr-A CHAGASIN.

Fig. S8. Structure-based sequence alignment of Pf12 D2 and the three surface antigens from Toxoplasma gondii (SAG D1, SporoSAG D2, and BSR4 D1) belonging to the SRS superfamily.

Fig. S9. Phylogeny of relevant apicomplexan organisms [for a more comprehensive version refer to Anantharaman et al. (1) or Templeton et al. (2)].


<table>
<thead>
<tr>
<th>Protein</th>
<th>PlasmoDB Gene ID</th>
<th>No. of s48/45 domains (1)</th>
<th>mRNA</th>
<th>Protein (2)</th>
<th>GPI-anchor (3)</th>
<th>Role</th>
<th>Additional references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pf230</td>
<td>PFB0405w</td>
<td>14</td>
<td>Gm</td>
<td>No</td>
<td>RBC binding, oocyst development</td>
<td>(4–8)</td>
<td></td>
</tr>
<tr>
<td>Pf48/45</td>
<td>PF13_0247</td>
<td>3</td>
<td>Gm (3)</td>
<td>Predicted</td>
<td>Male gamete fertility</td>
<td>(9–12)</td>
<td></td>
</tr>
<tr>
<td>Pf230p</td>
<td>PFB0400w</td>
<td>14</td>
<td>Gm (13)</td>
<td>No</td>
<td>ND</td>
<td>(13)</td>
<td></td>
</tr>
<tr>
<td>Pf47</td>
<td>PF13_0248</td>
<td>3</td>
<td>Gm (3)</td>
<td>Predicted</td>
<td>ND/ female fertility (Plasmodium berghei)</td>
<td>(14–16)</td>
<td></td>
</tr>
<tr>
<td>P36</td>
<td>PFD0210c</td>
<td>2</td>
<td>Sp</td>
<td>No</td>
<td>Hepatocyte infection</td>
<td>(14, 17–19)</td>
<td></td>
</tr>
<tr>
<td>P52</td>
<td>PFD0215c</td>
<td>2</td>
<td>Sp (3)</td>
<td>Predicted</td>
<td>Hepatocyte infection</td>
<td>(17, 18, 20, 21)</td>
<td></td>
</tr>
<tr>
<td>P41</td>
<td>PFD0240c</td>
<td>2</td>
<td>A (22)</td>
<td>A</td>
<td>ND</td>
<td>(14, 22)</td>
<td></td>
</tr>
<tr>
<td>P38</td>
<td>PFE0395c</td>
<td>2</td>
<td>A/Sp/Gm (22)</td>
<td>A/Gm</td>
<td>Yes</td>
<td>ND</td>
<td>(20, 22–24)</td>
</tr>
<tr>
<td>P12</td>
<td>PFD0615c</td>
<td>2</td>
<td>A/Sp/Gm (22)</td>
<td>A</td>
<td>Yes</td>
<td>ND</td>
<td>(22, 25)</td>
</tr>
<tr>
<td>P12p</td>
<td>PFF0620c</td>
<td>2</td>
<td>A (3)</td>
<td>Predicted</td>
<td>ND</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>P92</td>
<td>PF13_0338</td>
<td>1*</td>
<td>A (22)</td>
<td>A</td>
<td>Yes</td>
<td>ND</td>
<td>(22, 24, 26)</td>
</tr>
<tr>
<td>Sequestrin</td>
<td>PFD0260c</td>
<td>2*</td>
<td>A (27)</td>
<td>No</td>
<td>CD36 binding</td>
<td>(27)</td>
<td></td>
</tr>
</tbody>
</table>

A, asexual blood stage; Gm, gametocyte; GPI, glycosylphosphatidylinositol; ND, not determined; Sp, sporozoite.

*As determined by L.A.

Table S2. Structural statistics

Experimental restraints

<table>
<thead>
<tr>
<th>Rmsd from experimental restraints*</th>
<th>&lt;SA&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distances (Å) (1,543)</td>
<td>0.012 ± 0.001</td>
</tr>
<tr>
<td>Torsion angles (deg) (380)</td>
<td>0.46 ± 0.05</td>
</tr>
<tr>
<td>$^{13}$C/($^{13}$C) shifts (ppm) (125/121)</td>
<td>1.09 ± 0.03/1.32 ± 0.03</td>
</tr>
</tbody>
</table>

Dipolar coupling R-factors (%)†

<table>
<thead>
<tr>
<th>R-factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{1}$DNH (88)</td>
<td>2.5 ± 0.3</td>
</tr>
<tr>
<td>$^{1}$DNC' (51)</td>
<td>24.3 ± 1.2</td>
</tr>
<tr>
<td>$^{1}$DHNC' (53)</td>
<td>22.2 ± 2.0</td>
</tr>
</tbody>
</table>

Deviations from idealized geometry

<table>
<thead>
<tr>
<th>Deviation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds (Å)</td>
<td>0.002 ± 0</td>
</tr>
<tr>
<td>Angles (deg)</td>
<td>0.34 ± 0.01</td>
</tr>
<tr>
<td>Impropers (deg)</td>
<td>0.40 ± 0.02</td>
</tr>
</tbody>
</table>

Nonbonded contacts

<table>
<thead>
<tr>
<th>Contact Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erepel (kcal·mol$^{-1}$)</td>
<td>27.5 ± 6.0</td>
</tr>
<tr>
<td>ELJ (kcal·mol$^{-1}$)$^{‡}$</td>
<td>$-$419 ± 16</td>
</tr>
<tr>
<td>Percentage residues in most favorable region of Ramachandran map§</td>
<td>86.0 ± 1.7</td>
</tr>
</tbody>
</table>

Coordinate precision (Å)

<table>
<thead>
<tr>
<th>Precision</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone</td>
<td>0.44</td>
</tr>
<tr>
<td>All atoms</td>
<td>1.22</td>
</tr>
</tbody>
</table>

*The distance restraints comprise 1,477 NOE-derived interproton distances (353 intraresidue, 424 sequential, 160 medium range, and 540 long-range), and 66 distances for 33 backbone hydrogen bonds added during the final stages of iterative refinement. The torsion angles comprise 122 $\phi$, 144 $\psi$, and 144 $\chi$ angles. There are no NOE or torsion angle violations greater than 0.5 Å or 5°, respectively.

†The dipolar coupling R-factor is defined as the ratio of the rmsd between observed and calculated values and the expected rmsd for a random distribution of vectors. The latter is given by $[2D_{a}^{2}(4+3\eta^{2})/5]^{1/2}$, where $D_{a}$ and $\eta$ are the magnitude of the alignment tensor and the rhombicity, respectively. The values of $D_{a}$ and $\eta$ are 14.0 Hz and 0.15, respectively. The R-factor scales are between 0 and 100% (1).

‡Calculated using the CHARMM 22 Lennard-Jones energy but not used in the simulated annealing calculations.

§Calculated using PROCHECK (2).

Table S3. Statistical support for the sequence relationship between the s48/45 and SRS domains

<table>
<thead>
<tr>
<th>No.</th>
<th>Profile Hit to s48/45 profile</th>
<th>Prob</th>
<th>P value</th>
<th>Score</th>
<th>SS</th>
<th>Cols</th>
<th>Query HMM</th>
<th>Template HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SRS domain Major surface antigen</td>
<td>94.8</td>
<td>1.8e-11</td>
<td>51.8</td>
<td>6.8</td>
<td>81</td>
<td>30–126</td>
<td>155–242 (289)</td>
</tr>
<tr>
<td>2</td>
<td>SRS domain Sporosag</td>
<td>86.0</td>
<td>1.2e-9</td>
<td>48.5</td>
<td>5.4</td>
<td>74</td>
<td>35–126</td>
<td>150–225 (238)</td>
</tr>
<tr>
<td>3</td>
<td>SRS domain BSR4, bradyzoite antigen</td>
<td>80.7</td>
<td>1.3e-7</td>
<td>33.0</td>
<td>5.5</td>
<td>82</td>
<td>30–126</td>
<td>188–279 (315)</td>
</tr>
<tr>
<td>4</td>
<td>Ephrin-A, neural molecule</td>
<td>49.2</td>
<td>0.00034</td>
<td>23.8</td>
<td>3.3</td>
<td>28</td>
<td>30–57</td>
<td>3–48 (132)</td>
</tr>
<tr>
<td>5</td>
<td>Ephrin-B, neural molecule</td>
<td>59.4</td>
<td>0.00064</td>
<td>22.2</td>
<td>3.4</td>
<td>30</td>
<td>30–59</td>
<td>3–52 (140)</td>
</tr>
</tbody>
</table>

Cols, column-wise match; HMM, Hidden Markov model; Prob, percentage probability; Query HMM, extent of match on query HMM; SS, secondary structure match; Template, extent of match on template HMM.