

Structure of the *Plasmodium* 6-cysteine s48/45 domain

Silvia A. Arredondo^{a,1}, Mengli Cai^{b,1}, Yuki Takayama^b, Nicholas J. MacDonald^c, D. Eric Anderson^d, L. Aravind^{e,2}, G. Marius Clore^{b,2}, and Louis H. Miller^{f,2}

^aLaboratory of Immunogenetics, ^cLaboratory of Malaria Immunology and Vaccinology, and ^fLaboratory of Malaria Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852; ^bLaboratory of Chemical Physics and ^dMass Spectrometry Facility, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892; and ^eNational 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 aconoidasidans. 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 Pf12 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 metazoan 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 s48/45 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 aconoidasidan (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, P12p, 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, 18–20). 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

Author contributions: S.A.A., M.C., G.M.C., and L.H.M. designed research; S.A.A., M.C., Y.T., N.J.M., D.E.A., G.M.C., and L.A. performed research; S.A.A., M.C., L.A., G.M.C., and L.H.M. analyzed data; and S.A.A., L.A., G.M.C., and L.H.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The coordinates and experimental restraints reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2LOE), and the chemical shift assignments have been deposited in the Biological Magnetic Resonance Bank, www.bmrb.wisc.edu (BMRB ID 18210).

¹S.A.A. and M.C. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: aravind@mail.nih.gov, marius@intra.niddk.nih.gov, or lmiller@niaid.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1204363109/-DCSupplemental.

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- d_3 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 residual 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 ($\beta 2$, $\beta 1$, $\beta 4$, $\beta 9$, $\beta 8$) has a 1x, -2x, -2x, 1 topology; the four-stranded β -sheet B ($\beta 3$, $\beta 11$, $\beta 10$, $\beta 5$) has a 3x, -1, -1 topology; and the two-stranded β -sheet C ($\beta 6$, $\beta 7$) 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 $\beta 1$ to $\beta 3$, and $\beta 4$ to $\beta 10$, respectively; disulfide bond C (Cys210–Cys258) bridges the loop connecting strands $\beta 5$ and $\beta 6$ to strand $\beta 10$ 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 $\beta 3$ is disordered (see, for example, the structure superpositions in Fig. 1A, and the low values of the $^1\text{H}\{-^{15}\text{N}\}$ heteronuclear NOEs in Fig. 1C).

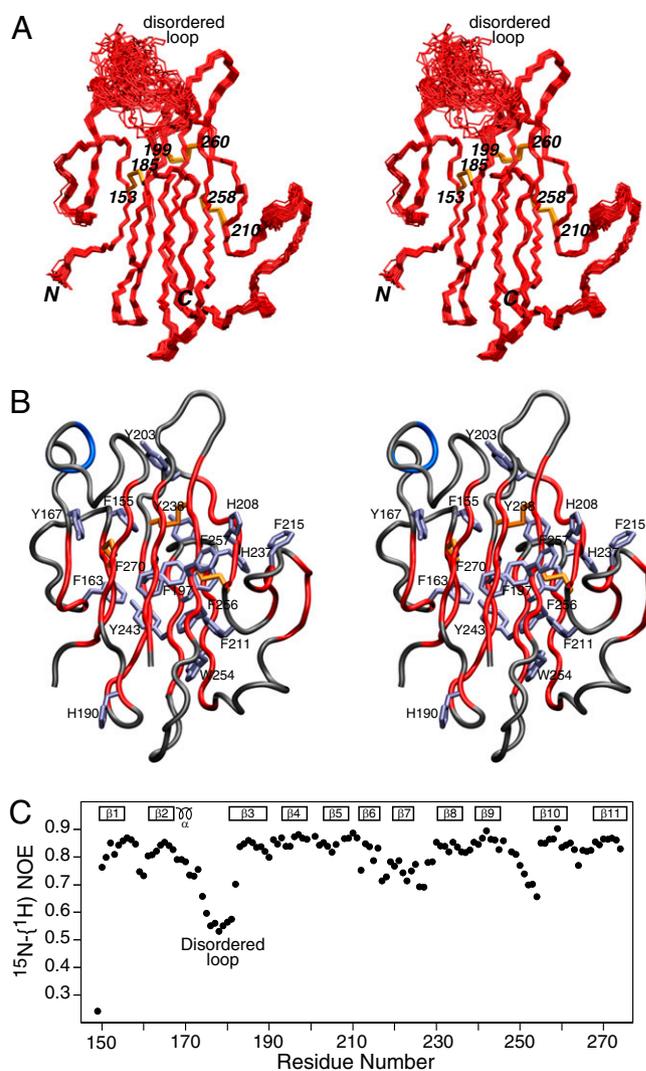


Fig. 1. NMR structure of Pf12 D2. (A) Stereoview of the backbone superposition of the 100 final simulated annealing structures. The backbone atoms (N, C α , C) are in red, and the three disulfide bonds (between residues 153–185, 199–260, and 210–258) are in yellow. (B) Stereoview of a tube representation of the backbone (with strands colored in red, and the helix in blue) showing the large number of aromatic residues (ice blue). (C) Heteronuclear $^1\text{H}\{-^{15}\text{N}\}$ -NOE profile as a function of residue number, showing that the loop connecting the single helix with strand $\beta 3$ is highly mobile and disordered.

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,

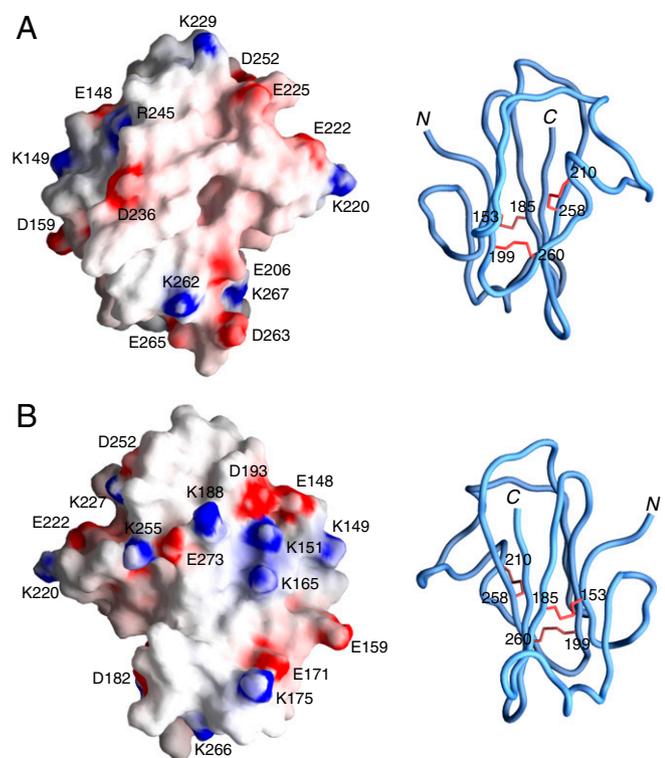


Fig. 3. Molecular surface of Pf12 D2. Two views (*A* and *B*) related by a 180° rotation displaying the molecular surface of Pf12 D2 with the electrostatic potential mapped onto the surface (*Left*). The corresponding panels (*Right*) are backbone tube representations in the same orientation as the surfaces to guide the eye. The electrostatic potential was created by the program GRASP (65).

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 $C\alpha$ 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

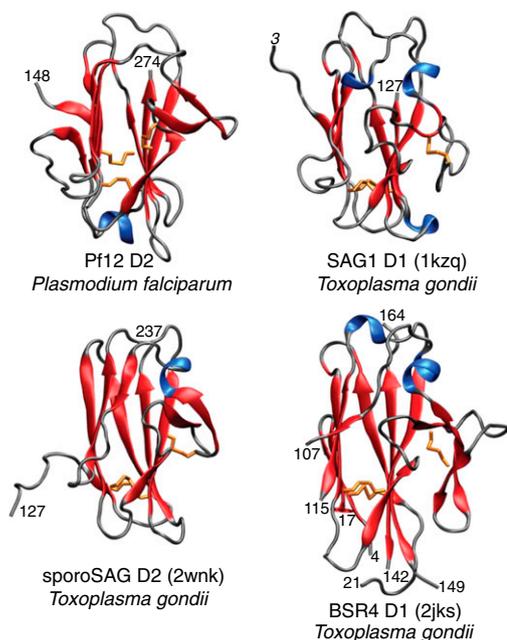


Fig. 4. Comparison of the fold of Pf12 D2 with the folds of three surface antigens from *T. gondii* (SAG D1, sporoSAG D2, and BSR4 D1) belonging to the SRS superfamily. The ribbon diagrams show the four domains in identical orientations.

D2, but distinct characteristics are also observed in the periphery of the core (Fig. 2) [i.e., a double-stranded β -sheet (β_6 , β_7), 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 - $\{^{15}\text{N}\}$ 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 Pfs230, Pfs48/45, 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 Pfs48/45, 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 Pfs230, Pfs48/45, Pfs47, 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, 26); however, no 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, β -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 only known from apicomplexan clades of coccidia and aconoidasida. These members are apparently absent in *Cryptosporidium*, which belongs to a more basal lineage of apicomplexans (Fig. S9), suggesting that these domains emerged in the common ancestor of the coccidians and aconoidasidans. This pattern of evolution is congruent with that of several other surface molecules [for example, the AMA1/MAEBL adhesins (44)]. Given that profile-profile searches indicate the ephrins as the next best hit after the SRS superfamily (Table S3), it is conceivable that the precursor of the s48/45 and SRS domains was derived from an ephrin-like precursor originally acquired from their metazoan hosts. Subsequent to their divergence, the SRS domain has undergone independent explosive radiations in different coccidians, but the s48/45 domains have undergone an expansion only in the *Plasmodium* lineage.

Despite the early recognition of Pfs230 and Pfs48/45 as transmission-blocking targets (9, 19) and subsequent identification of other members of the initially named 6-cys family (1, 5, 6), technical difficulties in recombinant protein expression have been a major hurdle in the successful production of full-length s48/45 proteins in soluble and stable conditions. The one exception has been Pfs48/45 (12). In the case of Pf12, after multiple failed attempts for the production of recombinant full length Pf12, we opted for engineering the protein for expression as separate domains with success in the production of the soluble

D2 domain in the periplasm of *E. coli*. The conservation of this structural component is important in *Plasmodium* biology. Thus, further functional and structural characterization of s48/45 proteins is indispensable.

Methods

Cloning, Expression, and Purification. Recombinant soluble Pf12 D2, the C-terminal domain of the *P. falciparum* protein Pf12 (PFF0615c), was produced in *E. coli*. Details on cloning, expression, labeling and purification are provided in *SI Methods*.

NMR Spectroscopy. All NMR spectra were collected on 0.4-mM protein samples in 15 mM sodium d_3 -acetate at pH 5.0, and were recorded at 35 °C on Bruker DRX600, DRX800, and Avance 900 spectrometers, equipped with z-gradient cryoprobes. Spectra were processed using the program NMRPipe (52), and analyzed using the programs PIPP and CAPP (53). Sequential assignment of 1H , ^{15}N , and ^{13}C resonances was achieved by means of through-bond heteronuclear scalar correlations along the protein backbone and side chains (54) (55) using 3D HNCO, CBCACONH, HNCACB, (H)C(CO)NH TOCSY, and H(CCO)NH-TOCSY experiments, and a 30-ms mixing time ^{13}C -separated NOE experiment. Interproton distance restraints were derived from 3D ^{15}N - and ^{13}C -separated NOE experiments at a mixing time of 120 ms. Side chain torsion angle restraints were derived from $^3J_{NC}$ (aromatic, methyl, and methylene) and $^3J_{CC}$ (aromatic, methyl, and methylene) scalar couplings measured by quantitative J correlation spectroscopy (56), in combination with data from a short mixing time (30 ms) 3D ^{13}C -separated NOE spectrum recorded in H_2O .

Residual $^1D_{NH}$, $^1D_{\alpha H}$, $^1D_{NC}$, and $^2D_{HNC}$ dipolar couplings (57) were measured in a liquid crystalline medium of phage pf1 (11 mg/mL) (58). The magnitudes of the axial (D_a^{NH}) and rhombic (η) components of the alignment tensor D^{NH} were obtained by examining the distribution of the normalized residual dipolar couplings (59). Heteronuclear ^{15}N - $\{^1H\}$ NOE data were collected and analyzed as described previously (60), on a 0.3 mM ^{15}N -labeled sample in 25 mM sodium acetate, pH 5.0.

Structure Calculations. Interproton distance restraints were classified into four distance ranges: 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5.0 Å and 1.8–6.0 Å, corresponding to strong, medium, weak, and very weak NOEs, respectively (54). Next, 0.5 Å was added to the upper distance limits for NOEs involving methyl groups. Distances involving ambiguous NOEs, nonstereospecifically assigned methylene and methyl groups, and the H_δ and H_ϵ protons of Tyr and Phe were represented as a $\Sigma(r^{-6})^{-1/6}$ sum. Backbone ϕ/ψ torsion angle restraints were obtained on the basis of backbone ^{15}N , 1H_N , $^{13}C\alpha$, $^1H\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts using the program TALOS+ (61). Structures were calculated by simulated annealing in torsion angle space (62) using the program Xplor-NIH (63). The target function comprised terms for the experimental NMR restraints (NOE-derived interproton distances, backbone, and side-chain torsion angles, $^{13}C\alpha/^{13}C\beta$ chemical shifts, and residual dipolar couplings), terms to maintain idealized covalent geometry, a quartic van der Waals repulsion term to prevent atomic overlap, a multidimensional torsion-angle database potential of mean force, and a hydrogen bonding database potential of mean force. Structure figures were created using the programs VMD-XPLOR (64) and GRASP (65). The coordinates and experimental restraints have been deposited in the Protein Data Bank (PDB ID 2LOE) and the chemical-shift assignments in the Biological Magnetic Resonance Bank (BMRB ID 18210).

ACKNOWLEDGMENTS. We thank Susan Pierce (Laboratory of Immunogenetics) and Kim Williamson and Prakash Srinivasan (Laboratory of Malaria Vector Research) for very insightful discussions. This work was supported by the intramural funds from the National Library of Medicine (to L.A.), the National Institute of Allergy and Infectious Diseases (L.H.M.), and the National Institute of Diabetes and Digestive and Kidney Diseases (G.M.C.) of the National Institutes of Health.

- Williamson KC, Criscio MD, Kaslow DC (1993) Cloning and expression of the gene for *Plasmodium falciparum* transmission-blocking target antigen, Pfs230. *Mol Biochem Parasitol* 58:355–358.
- Gerloff DL, Creasey A, Maslau S, Carter R (2005) Structural models for the protein family characterized by gamete surface protein Pfs230 of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 102:13598–13603.
- Punta M, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue, D1):D290–D301.

- Carter R, Coulson A, Bhatti S, Taylor BJ, Elliott JF (1995) Predicted disulfide-bonded structures for three uniquely related proteins of *Plasmodium falciparum*, Pfs230, Pfs48/45 and Pf12. *Mol Biochem Parasitol* 71:203–210.
- Templeton TJ, Kaslow DC (1999) Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol Biochem Parasitol* 101:223–227.
- Thompson J, Janse CJ, Waters AP (2001) Comparative genomics in *Plasmodium*: A tool for the identification of genes and functional analysis. *Mol Biochem Parasitol* 118:147–154.

7. Sanders PR, et al. (2007) Identification of protein complexes in detergent-resistant membranes of *Plasmodium falciparum* schizonts. *Mol Biochem Parasitol* 154:148–157.
8. Ishino T, Chinzei Y, Yuda M (2005) Two proteins with 6-cys motifs are required for malarial parasites to commit to infection of the hepatocyte. *Mol Microbiol* 58:1264–1275.
9. Quakyi IA, et al. (1987) The 230-kDa gamete surface protein of *Plasmodium falciparum* is also a target for transmission-blocking antibodies. *J Immunol* 139:4213–4217.
10. Sanders PR, et al. (2005) Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *J Biol Chem* 280:40169–40176.
11. van Schaijk BC, et al. (2006) Pfs47, paralog of the male fertility factor Pfs48/45, is a female specific surface protein in *Plasmodium falciparum*. *Mol Biochem Parasitol* 149:216–222.
12. Chowdhury DR, Angov E, Kariuki T, Kumar N (2009) A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PLoS ONE* 4:e6352.
13. Williamson KC (2003) Pfs230: From malaria transmission-blocking vaccine candidate toward function. *Parasite Immunol* 25:351–359.
14. Gilson PR, et al. (2006) Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 5:1286–1299.
15. Kumar N (1987) Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex. *Parasite Immunol* 9:321–335.
16. Eksi S, et al. (2006) Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production. *Mol Microbiol* 61:991–998.
17. van Dijk MR, et al. (2001) A central role for P48/45 in malaria parasite male gamete fertility. *Cell* 104:153–164.
18. Healer J, et al. (1997) Complement-mediated lysis of *Plasmodium falciparum* gametes by malaria-immune human sera is associated with antibodies to the gamete surface antigen Pfs230. *Infect Immun* 65:3017–3023.
19. Vermeulen AN, et al. (1985) Sequential expression of antigens on sexual stages of *Plasmodium falciparum* accessible to transmission-blocking antibodies in the mosquito. *J Exp Med* 162:1460–1476.
20. Vermeulen AN, et al. (1985) *Plasmodium falciparum* transmission blocking monoclonal antibodies recognize monovalently expressed epitopes. *Dev Biol Stand* 62:91–97.
21. van Dijk MR, et al. (2010) Three members of the 6-cys protein family of *Plasmodium* play a role in gamete fertility. *PLoS Pathog* 6:e1000853.
22. van Schaijk BC, et al. (2008) Gene disruption of *Plasmodium falciparum* p52 results in attenuation of malaria liver stage development in cultured primary human hepatocytes. *PLoS ONE* 3:e3549.
23. VanBuskirk KM, et al. (2009) Preerythrocytic, live-attenuated *Plasmodium falciparum* vaccine candidates by design. *Proc Natl Acad Sci USA* 106:13004–13009.
24. van Dijk MR, et al. (2005) Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc Natl Acad Sci USA* 102:12194–12199.
25. Elliott JF, et al. (1990) Genes for *Plasmodium falciparum* surface antigens cloned by expression in COS cells. *Proc Natl Acad Sci USA* 87:6363–6367.
26. Reeder JC, Wapling J, Mueller I, Siba PM, Barry AE (2011) Population genetic analysis of the *Plasmodium falciparum* 6-cys protein Pf38 in Papua New Guinea reveals domain-specific balancing selection. *Malar J* 10:126.
27. Tetteh KK, et al. (2009) Prospective identification of malaria parasite genes under balancing selection. *PLoS ONE* 4:e5568.
28. Cowman AF, Crabb BS (2006) Invasion of red blood cells by malaria parasites. *Cell* 124:755–766.
29. Bustamante PJ, et al. (2000) Differential ability of specific regions of *Plasmodium falciparum* sexual-stage antigen, Pfs230, to induce malaria transmission-blocking immunity. *Parasite Immunol* 22:373–380.
30. Farrance CE, et al. (2011) A plant-produced Pfs230 vaccine candidate blocks transmission of *Plasmodium falciparum*. *Clin Vaccine Immunol* 18:1351–1357.
31. Milek RL, Stunnenberg HG, Konings RN (2000) Assembly and expression of a synthetic gene encoding the antigen Pfs48/45 of the human malaria parasite *Plasmodium falciparum* in yeast. *Vaccine* 18:1402–1411.
32. Outchkourou NS, et al. (2008) Correctly folded Pfs48/45 protein of *Plasmodium falciparum* elicits malaria transmission-blocking immunity in mice. *Proc Natl Acad Sci USA* 105:4301–4305.
33. Tachibana M, et al. (2011) N-terminal prodomain of Pfs230 synthesized using a cell-free system is sufficient to induce complement-dependent malaria transmission-blocking activity. *Clin Vaccine Immunol* 18:1343–1350.
34. Williamson KC, Keister DB, Muratova O, Kaslow DC (1995) Recombinant Pfs230, a *Plasmodium falciparum* gametocyte protein, induces antisera that reduce the infectivity of *Plasmodium falciparum* to mosquitoes. *Mol Biochem Parasitol* 75:33–42.
35. Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* 3:e2189.
36. Thanaraj TA, Argos P (1996) Ribosome-mediated translational pause and protein domain organization. *Protein Sci* 5:1594–1612.
37. Bax A, Grishaev A (2005) Weak alignment NMR: A hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* 15:563–570.
38. Williams DC, Jr., Cai M, Suh JY, Peterkofsky A, Clore GM (2005) Solution NMR structure of the 48-kDa IIAmannose-HPr complex of the *Escherichia coli* mannose phosphotransferase system. *J Biol Chem* 280:20775–20784.
39. He XL, Grigg ME, Boothroyd JC, Garcia KC (2002) Structure of the immunodominant surface antigen from the *Toxoplasma gondii* SRS superfamily. *Nat Struct Biol* 9:606–611.
40. Crawford J, et al. (2010) Structural and functional characterization of Sporozoite Surface Antigen (SAG2)-related surface antigen from *Toxoplasma gondii*. *J Biol Chem* 285:12063–12070.
41. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
42. Holm L, Rosenstrom P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–W549.
43. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567.
44. Anantharaman V, Iyer LM, Balaji S, Aravind L (2007) Adhesion molecules and other secreted host-interaction determinants in Apicomplexa: Insights from comparative genomics. *Int Rev Cytol* 262:1–74.
45. Jung C, Lee CY, Grigg ME (2004) The SRS superfamily of *Toxoplasma* surface proteins. *Int J Parasitol* 34:285–296.
46. Crawford J, et al. (2009) Structural characterization of the bradyzoite surface antigen (BSR4) from *Toxoplasma gondii*, a unique addition to the surface antigen glycoprotein 1-related superfamily. *J Biol Chem* 284:9192–9198.
47. Graille M, et al. (2005) Crystal structure of the complex between the monomeric form of *Toxoplasma gondii* surface antigen 1 (SAG1) and a monoclonal antibody that mimics the human immune response. *J Mol Biol* 354:447–458.
48. Carruthers V, Boothroyd JC (2007) Pulling together: An integrated model of *Toxoplasma* cell invasion. *Curr Opin Microbiol* 10:83–89.
49. Lekutis C, Ferguson DJ, Grigg ME, Camps M, Boothroyd JC (2001) Surface antigens of *Toxoplasma gondii*: Variations on a theme. *Int J Parasitol* 31:1285–1292.
50. Mineo JR, Kasper LH (1994) Attachment of *Toxoplasma gondii* to host cells involves major surface protein, SAG-1 (P30). *Exp Parasitol* 79:11–20.
51. Mineo JR, et al. (1993) Antibodies to *Toxoplasma gondii* major surface protein (SAG-1, P30) inhibit infection of host cells and are produced in murine intestine after peroral infection. *J Immunol* 150:3951–3964.
52. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
53. Garrett DS, Powers R, Gronenborn AM, Clore GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* 95:214–220.
54. Clore GM, Gronenborn AM (1991) Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252:1390–1399.
55. Bax A, Grzesiek S (1993) Methodological advances in protein NMR. *Acc Chem Res* 26:131–138.
56. Bax A, et al. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol* 239:79–105.
57. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114.
58. Clore GM, Starich MR, Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120:10571–10572.
59. Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* 133:216–221.
60. Grzesiek S, Bax A (1993) The importance of not saturating H2O in protein NMR—Application to sensitivity enhancement and NOE measurements. *J Am Chem Soc* 115:12593–12594.
61. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223.
62. Schwieters CD, Clore GM (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 152:288–302.
63. Schwieters CD, Kuszewski J, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* 48:47–62.
64. Schwieters CD, Clore GM (2001) The VMD-XPLOR visualization package for NMR structure refinement. *J Magn Reson* 149:239–244.
65. Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11:281–296.

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 Pf12 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 OD₆₀₀ ~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 with 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 ¹⁵NH₄Cl and ¹³C₆-glucose as the sole nitrogen and carbon sources, respectively (2), and expression was induced at an OD₆₀₀ ~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.

1. Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* 3:e2189.
2. Cai M, et al. (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*. *J Biomol NMR* 11:97–102.
3. Gerloff DL, Creasey A, Maslau S, Carter R (2005) Structural models for the protein family characterized by gamete surface protein Pfs230 of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 102:13598–13603.
4. He XL, Grigg ME, Boothroyd JC, Garcia KC (2002) Structure of the immunodominant surface antigen from the *Toxoplasma gondii* SRS superfamily. *Nat Struct Biol* 9:606–611.

5. Crawford J, et al. (2010) Structural and functional characterization of SporoSAG: A SAG2-related surface antigen from *Toxoplasma gondii*. *J Biol Chem* 285:12063–12070.
6. Jacquet A, et al. (2001) The surface antigen SAG3 mediates the attachment of *Toxoplasma gondii* to cell-surface proteoglycans. *Mol Biochem Parasitol* 116:35–44.
7. Toth J, et al. (2001) Crystal structure of an ephrin ectodomain. *Dev Cell* 1:83–92.
8. Ockenhouse CF, Klotz FW, Tandon NN, Jamieson GA (1991) Sequestrin, a CD36 recognition protein on *Plasmodium falciparum* malaria-infected erythrocytes identified by anti-idiotypic antibodies. *Proc Natl Acad Sci USA* 88:3175–3179.

10 20 30 40 50 60 70 80 90
 TCTAGATTAAAGAGGAGAAAGGTACCCATGAAGAAAGGTTTATGTTGTTTACTTGTAGCGGCGTTTCAGGCTTTCAGGCTCAC
 M K K G F M L F T L L A A F S G F A Q A H

100 110 120 130 140 150 160 170 180
 CATCACCACCATCATAGCAGCGGCTGGTGCAGCGGCGAGCAAAAAGTGAAGGCTGCGATTTTACCACCTCAGAAAGCACCATCTTC
 H H H H H S S G L V P R G S E K V K G C D F T T S E S T I F

190 200 210 220 230 240 250 260 270
 AGCAAAGGCTATAGCATCAACGAAATTAGCAACAAAAGCAGCAACAATCAGCAGGATATCGTGTGCACCGTGAAGGCGCATGCGAACGAT
 S K G Y S I N E I S N K S S N N Q Q D I V C T V K A H A N D

280 290 300 310 320 330 340 350 360
 CTGATAGGCTTTAAATGCCCATCAAACCTATAGCGTGAACCCGCATGATTGCTTTGTGAGTGCGTTCAACCTGAGCGGAAAAAACGAAAT
 L I G F K C P S N Y S V E P H D C F V S A F N L S G K N E N

370 380 390 400 410 420 430 440 450
 CTGGAAAACAAACTCAAACCTGACCAACATTTATCATGGATCATTTATAACAACACGTTCTATAGCCGCTGCCGAGCCTGATCAGCGATAAC
 L E N K L K L T N I I M D H Y N N T F Y S R L P S L I S D N

460 470 480 490 500 510 520
 TGGAAATTCCTTTGCGTGTGCAGTAAAGATAACGAAAAAAACTGGTATTTACAGTGGAAAGCGTAACTCGAG
 W K F F C V C S K D N E K K L V F T V E A * L E

Fig. S2. DNA sequence (and amino acid translation) of the complete construct used for periplasmic expression of Pf12 D2 in *E. coli*.

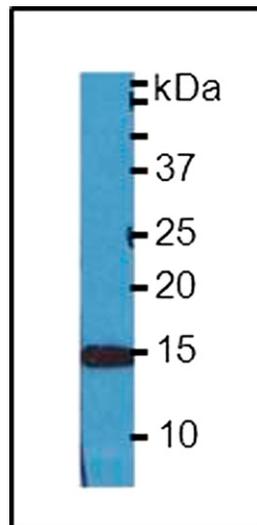


Fig. S3. Western blot analysis of the soluble fraction obtained by osmotic shock from the expression of Pf12 D2 in the *E. coli* periplasm. The membrane was probed with monoclonal antipolyhistidine HRP-conjugated antibody (Sigma).

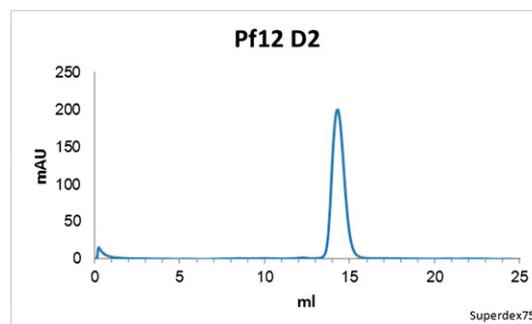


Fig. S4. Gel filtration FPLC of Pf12 D2. The purified protein was run on a Superdex75 column.

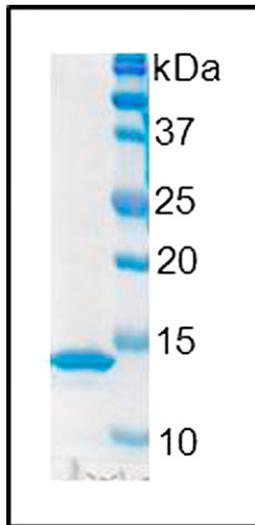


Fig. S5. SDS-gel electrophoresis of purified Pf12 D2 stained with Gelcode blue reagent (Thermo Fisher Scientific).

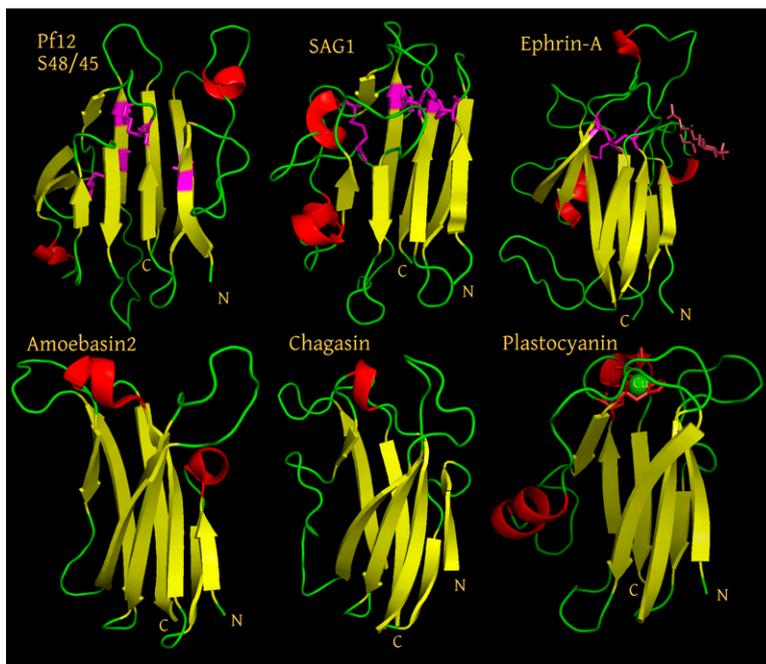


Fig. S6. A collection of β -sandwich domains related to the s48/45 domain. The domains are arranged so the shared, structurally equivalent core of the β -sandwich element in each domain is aligned to their respective right edge where their N termini are located. The domains differ from each other primarily 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 amoebasin-2, 2nnr chagasin, and 2q5b plastocyanin.

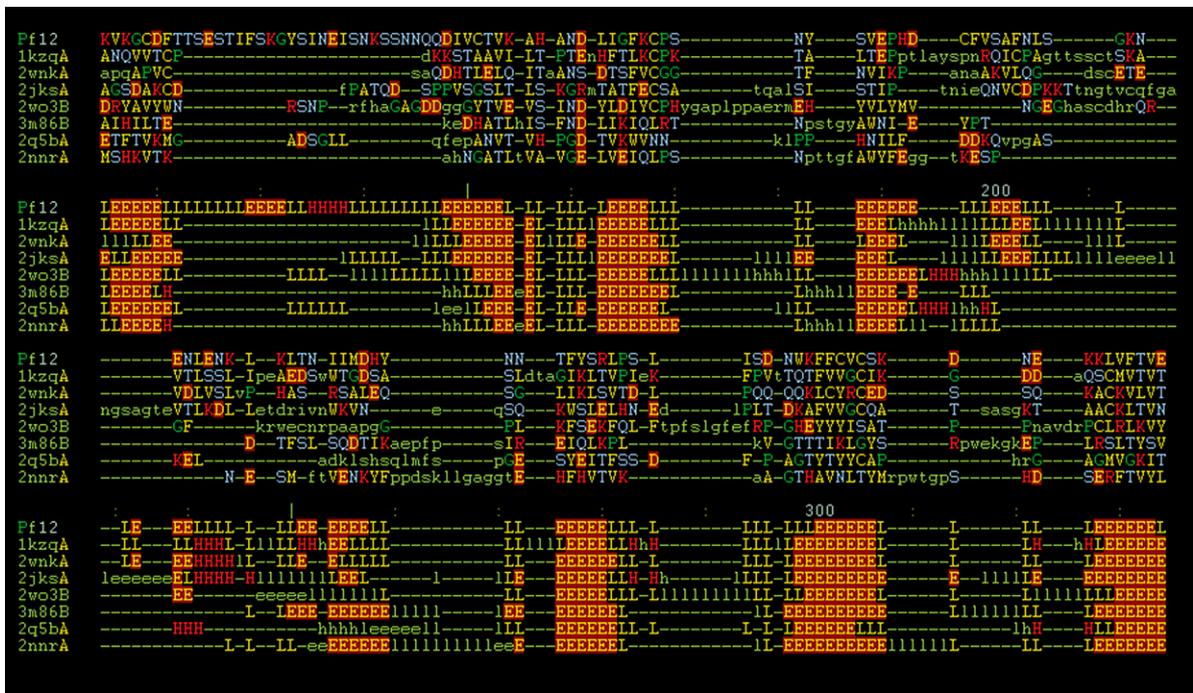


Fig. S7. A structure-based multiple alignment of the selected β -sandwich domains compared with Pf12 D2. 1kzq-A MAJOR SURFACE ANTIGEN P30, SAG1 D1; 2wnk A SPOOROZOITE-SPECIFIC SAG PROTEIN, SporoSAG D1; 2jks-A BRADYZOITE SURFACE ANTIGEN, BSR4 D1; 2wo3-B EPHRIN TYPE-A RECEPTOR; 3m86-B AMOEBIASIN-2; 2q5b-A PLASTOCYANIN; 2nr-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.

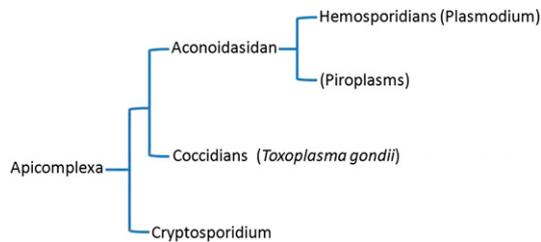


Fig. S9. Phylogeny of relevant apicomplexan organisms [for a more comprehensive version refer to Anantharaman et al. (1) or Templeton et al. (2)].

1. Anantharaman V, Iyer LM, Balaji S, Aravind L (2007) Adhesion molecules and other secreted host-interaction determinants in Apicomplexa: Insights from comparative genomics. *Int Rev Cytol* 262:1-74.
2. Templeton TJ, et al. (2010) A genome-sequence survey for *Ascogregarina taiwanensis* supports evolutionary affiliation but metabolic diversity between a Gregarine and Cryptosporidium. *Mol Biol Evol* 27:235-248.

Table S1. Summary of the *P. falciparum* proteins containing s48/45 domains

Protein	PlasmoDB Gene ID	No. of s48/45 domains (1)	Expression			Role	Additional references
			mRNA	Protein (2)	GPI-anchor (3)		
Pfs230	PFB0405w	14		Gm	No	RBC binding, oocyst development	(4–8)
Pfs48/45	PF13_0247	3	Gm (3)	Gm	Predicted	Male gamete fertility	(9–12)
Pfs230p	PFB0400w	14	Gm (13)	Gm	No	ND	(13)
Pfs47	PF13_0248	3	Gm (3)	Gm	Predicted	ND/ female fertility (<i>Plasmodium berghei</i>)	(14–16)
P36	PFD0210c	2		Sp	No	Hepatocyte infection	(14, 17–19)
P52	PFD0215c	2	Sp (3)	Sp	Predicted	Hepatocyte infection	(17, 18, 20, 21)
Pf41	PFD0240c	2	A (22)	A	No	ND	(14, 22)
Pf38	PFE0395c	2	A/Sp/Gm (22)	A/Gm	Yes	ND	(20, 22–24)
Pf12	PFF0615c	2	A/Sp/Gm (22)	A	Yes	ND	(22, 25)
P12p	PFF0620c	2		A (3)	Predicted	ND	(1)
Pf92	PF13_0338	1*	A (22)	A	Yes	ND	(22, 24, 26)
Sequestrin	PFD0260c	2*		A (27)	No	CD36 binding	(27)

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

*As determined by L.A.

- Gerloff DL, Creasey A, Maslau S, Carter R (2005) Structural models for the protein family characterized by gamete surface protein Pfs230 of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 102:13598–13603.
- Aurrecochea C, et al. (2009) PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res* 37(Database issue):D539–D543.
- Gilson PR, et al. (2006) Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 5:1286–1299.
- Williamson KC, Criscio MD, Kaslow DC (1993) Cloning and expression of the gene for *Plasmodium falciparum* transmission-blocking target antigen, Pfs230. *Mol Biochem Parasitol* 58:355–358.
- Quakyi IA, et al. (1987) The 230-kDa gamete surface protein of *Plasmodium falciparum* is also a target for transmission-blocking antibodies. *J Immunol* 139:4213–4217.
- Williamson KC (2003) Pfs230: From malaria transmission-blocking vaccine candidate toward function. *Parasite Immunol* 25:351–359.
- Kumar N (1987) Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex. *Parasite Immunol* 9:321–335.
- Eksi S, et al. (2006) Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production. *Mol Microbiol* 61:991–998.
- Chowdhury DR, Angov E, Kariuki T, Kumar N (2009) A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PLoS ONE* 4:e6352.
- van Dijk MR, et al. (2001) A central role for P48/45 in malaria parasite male gamete fertility. *Cell* 104:153–164.
- Vermeulen AN, et al. (1985) Sequential expression of antigens on sexual stages of *Plasmodium falciparum* accessible to transmission-blocking antibodies in the mosquito. *J Exp Med* 162:1460–1476.
- Vermeulen AN, et al. (1985) *Plasmodium falciparum* transmission blocking monoclonal antibodies recognize monovalently expressed epitopes. *Dev Biol Stand* 62:91–97.
- Eksi S, Williamson KC (2002) Male-specific expression of the paralog of malaria transmission-blocking target antigen Pfs230, Pfb0400w. *Mol Biochem Parasitol* 122:127–130.
- Templeton TJ, Kaslow DC (1999) Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol Biochem Parasitol* 101:223–227.
- van Schaijk BC, et al. (2006) Pfs47, paralog of the male fertility factor Pfs48/45, is a female specific surface protein in *Plasmodium falciparum*. *Mol Biochem Parasitol* 149:216–222.
- van Dijk MR, et al. (2010) Three members of the 6-cys protein family of *Plasmodium* play a role in gamete fertility. *PLoS Pathog* 6:e1000853.
- Ishino T, Chinzei Y, Yuda M (2005) Two proteins with 6-cys motifs are required for malarial parasites to commit to infection of the hepatocyte. *Mol Microbiol* 58:1264–1275.
- VanBuskirk KM, et al. (2009) Preerythrocytic, live-attenuated *Plasmodium falciparum* vaccine candidates by design. *Proc Natl Acad Sci USA* 106:13004–13009.
- van Dijk MR, et al. (2005) Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc Natl Acad Sci USA* 102:12194–12199.
- Thompson J, Janse CJ, Waters AP (2001) Comparative genomics in *Plasmodium*: A tool for the identification of genes and functional analysis. *Mol Biochem Parasitol* 118:147–154.
- van Schaijk BC, et al. (2008) Gene disruption of *Plasmodium falciparum* p52 results in attenuation of malaria liver stage development in cultured primary human hepatocytes. *PLoS ONE* 3:e3549.
- Sanders PR, et al. (2005) Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *J Biol Chem* 280:40169–40176.
- Reeder JC, Wapling J, Mueller I, Siba PM, Barry AE (2011) Population genetic analysis of the *Plasmodium falciparum* 6-cys protein Pf38 in Papua New Guinea reveals domain-specific balancing selection. *Malar J* 10:126.
- Tetteh KK, et al. (2009) Prospective identification of malaria parasite genes under balancing selection. *PLoS ONE* 4:e5568.
- Elliott JF, et al. (1990) Genes for *Plasmodium falciparum* surface antigens cloned by expression in COS cells. *Proc Natl Acad Sci USA* 87:6363–6367.
- Cowman AF, Crabb BS (2006) Invasion of red blood cells by malaria parasites. *Cell* 124:755–766.
- Ockenhouse CF, Klotz FW, Tandon NN, Jamieson GA (1991) Sequestrin, a CD36 recognition protein on *Plasmodium falciparum* malaria-infected erythrocytes identified by anti-idiotypic antibodies. *Proc Natl Acad Sci USA* 88:3175–3179.

Table S2. Structural statistics

Experimental restraints	<SA>
Rmsd from experimental restraints*	
Distances (Å) (1,543)	0.012 ± 0.001
Torsion angles (deg) (380)	0.46 ± 0.05
¹³ C α / ¹³ C β shifts (ppm) (125/121)	1.09 ± 0.03/1.32 ± 0.03
Dipolar coupling R-factors (%) [†]	
¹ D _{NH} (88)	2.5 ± 0.3
¹ D _{NC} (51)	24.3 ± 1.2
¹ D _{HNC} (53)	22.2 ± 2.0
Deviations from idealized geometry	
Bonds (Å)	0.002 ± 0
Angles (deg)	0.34 ± 0.01
Impropers (deg)	0.40 ± 0.02
Nonbonded contacts	
E _{repel} (kcal·mol ⁻¹)	27.5 ± 6.0
E _{LJ} (kcal·mol ⁻¹) [‡]	-419 ± 16
Percentage residues in most favorable region of Ramachandran map ^{§,¶}	86.0 ± 1.7
Coordinate precision (Å) [¶]	
Backbone	0.44
All atoms	1.22

*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 ϕ , 144 ψ , and 144 χ 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 η are the magnitude of the alignment tensor and the rhombicity, respectively. The values of D_a and η 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).

[¶]Excludes residues 174–180 comprising the disordered mobile loop.

1. Clore GM, Garrett DS (1991) R-factor, free R and complete cross-validation for dipolar coupling refinement of NMR structures. *J Am Chem Soc* 113:9008–9012.
2. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK—A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.

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

No.	Profile Hit to s48/45 profile	Prob	P value	Score	SS	Cols	Query HMM	Template	HMM
1	SRS domain Major surface antigen	94.8	1.8e-11	51.8	6.8	81	30–126	155–242	(289)
2	SRS domain Sporosag	86.0	1.2e-9	48.5	5.4	74	35–126	150–225	(238)
3	SRS domain BSR4, bradyzoite antigen	80.7	1.3e-7	33.0	5.5	82	30–126	188–279	(315)
4	Ephrin-A, neural molecule	49.2	0.00016	23.8	3.3	28	30–57	3–48	(132)
5	Ephrin-B, neural molecule	39.4	0.00034	22.2	3.4	30	30–59	3–52	(140)

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.