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 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 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 ^{15}N- and ^{13}C-labeled by growing in minimal medium containing $^{15}NH_4Cl$ and $^{13}C_6$-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 His_6 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-d_3 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 crystalized 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-idiotype antibodies. *Proc Natl Acad Sci USA* 88:3175–3179.

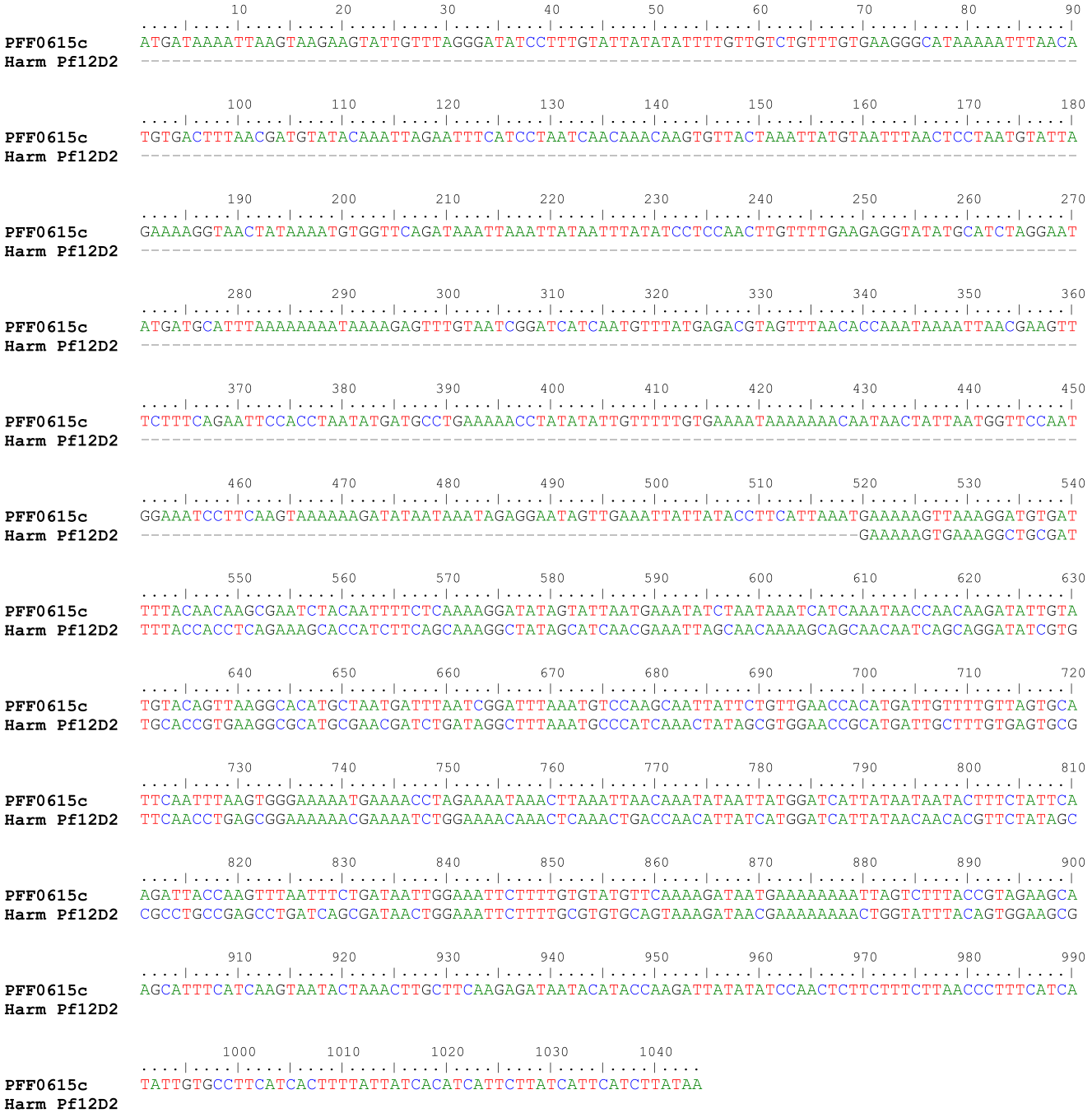

Fig. S1. Harmonized Pf12 D2 DNA sequence (bp 520–900) compared to the full sequence of native Pf12 (PFF0615c) from the 3D7 *P. falciparum* strain.

```
          10         20         30         40         50         60         70         80         90
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TCTAGATTAAAGAGGAGAAAGGTACCCATGAAGAAAGGTTTTATGTTGTTTACTTTGTTAGCGGCGTTTTCAGGCTTTGCTCAGGCTCAC
                                  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
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CATCACCACCATCATAGCAGCGGCCTGGTGCCGCGCGGCAGCGAAAAAGTGAAAGGCTGCGATTTTACCACCTCAGAAAGCACCATCTTC
 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
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AGCAAAGGCTATAGCATCAACGAAATTTAGCAACAAAAGCAGCAACAATCAGCAGGATATCGTGTGCACCGTGAAGGCGCATGCGAACGAT
 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
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CTGATAGGCTTTAAATGCCCATCAAACTATAGCGTGGAACCGCATGATTGCTTTGTGAGTGCGTTCAACCTGAGCGGAAAAAACGAAAAT
 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
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CTGGAAAACAAACTCAAACTGACCAACATTATCATGGATCATTATAACAACACGTTCTATAGCCGCCTGCCGAGCCTGATCAGCGATAAC
 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
....|....|....|....|....|....|....|....|....|....|....|....|....|..
TGGAAATTCTTTTGCGTGTGCAGTAAAGATAACGAAAAAAAACTGGTATTTACAGTGGAAGCGTAACTCGAG
 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*.

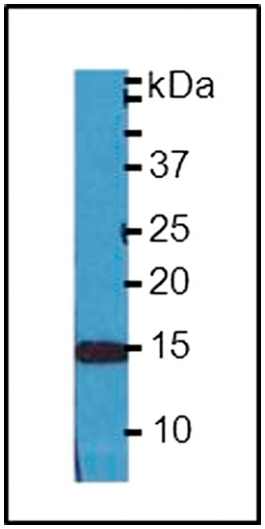

Fig. S3. Western blot analysis of the soluble fraction obtained by osmotic shock from the expression of Pf12 D2 in the *E. coli* periplasm. The membrane was probed with monoclonal antipolyhistidine HRP-conjugated antibody (Sigma).

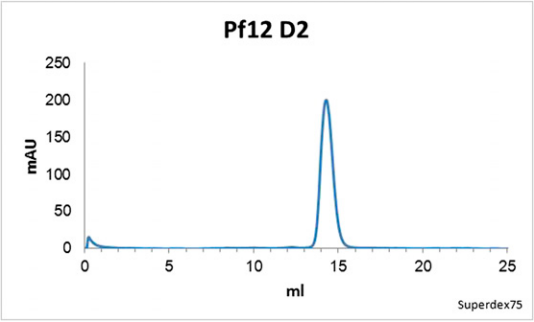

Fig. S4. Gel filtration FPLC of Pf12 D2. The purified protein was run on a Superdex75 column.

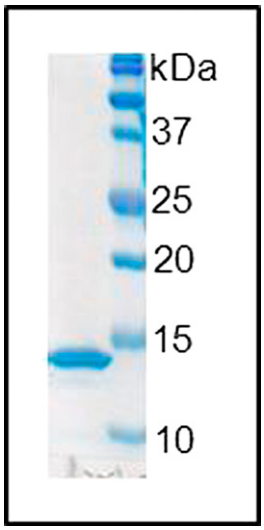

Fig. S5. SDS-gel electrophoresis of purified Pf12 D2 stained with Gelcode blue reagent (Thermo Fisher Scientific).

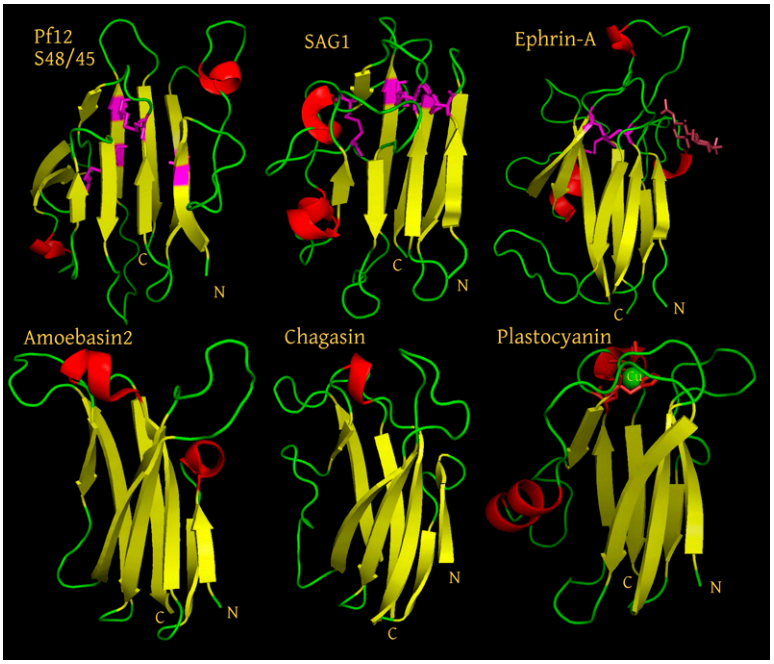

Fig. S6. A collection of β-sandwich domains related to the s48/45 domain. The domains are arranged so the shared, structurally equivalent core of the β-sandwich element in each domain is aligned to their respective right edge where their N termini are located. The domains differ from each other primarily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

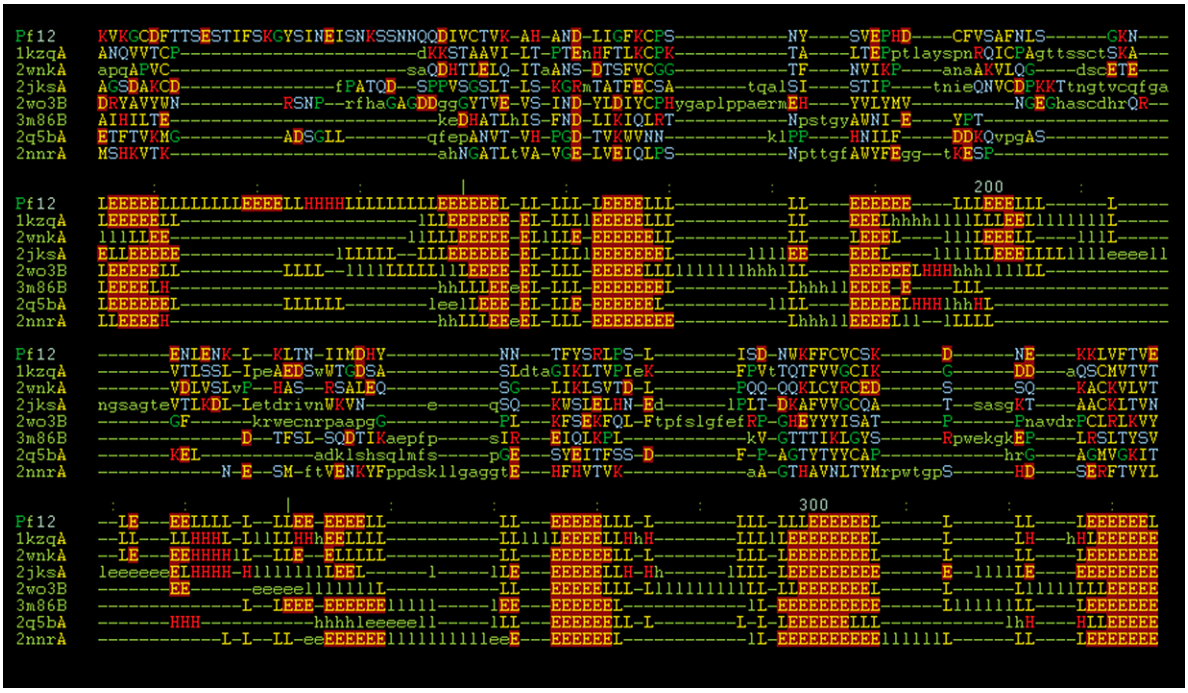

Fig. S7. A structure-based multiple alignment of the selected β-sandwich domains compared with Pf12 D2. 1kzq-A MAJOR SURFACE ANTIGEN P30, SAG1 D1; 2wnk A SPOROZOITE-SPECIFIC SAG PROTEIN, SporoSAG D1; 2jks-A BRADYZOITE SURFACE ANTIGEN, BSR4 D1; 2wo3-B EPHRIN TYPE-A RECEPTOR; 3m86-B AMOEBIASIN-2; 2q5b-A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

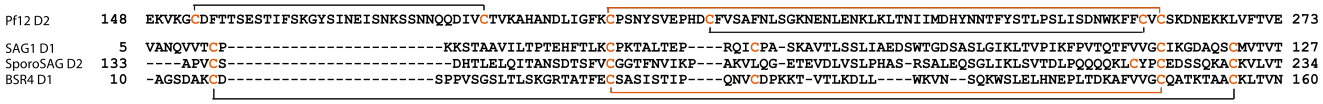

Fig. S8. Structure-based sequence alignment of Pf12 D2 and the three surface antigens from *Toxoplasma gondii* (SAG D1, SporoSAG D2, and BSR4 D1) belonging to the SRS superfamily.

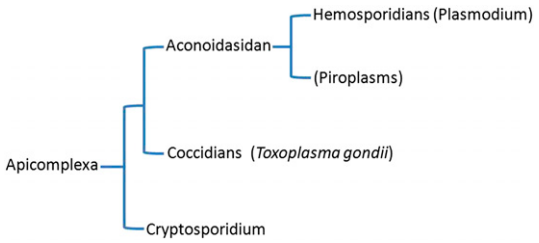

Fig. S9. Phylogeny of relevant apicomplexan organisms [for a more comprehensive version refer to Anantharaman et al. (1) or Templeton et al. (2)].

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		GPI-anchor (3)	Role	Additional references
			mRNA	Protein (2)			
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 (*Plasmodium berghei*)	(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.

1. 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.
2. Aurrecoechea C, et al. (2009) PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res* 37(Database issue):D539–D543.
3. 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.
4. 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.
5. 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.
6. Williamson KC (2003) Pfs230: From malaria transmission-blocking vaccine candidate toward function. *Parasite Immunol* 25:351–359.
7. Kumar N (1987) Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex. *Parasite Immunol* 9:321–335.
8. 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.
9. 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.
10. van Dijk MR, et al. (2001) A central role for P48/45 in malaria parasite male gamete fertility. *Cell* 104:153–164.
11. 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.
12. Vermeulen AN, et al. (1985) *Plasmodium falciparum* transmission blocking monoclonal antibodies recognize monovalently expressed epitopes. *Dev Biol Stand* 62:91–97.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. VanBuskirk KM, et al. (2009) Preerythrocytic, live-attenuated *Plasmodium falciparum* vaccine candidates by design. *Proc Natl Acad Sci USA* 106:13004–13009.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. Tetteh KK, et al. (2009) Prospective identification of malaria parasite genes under balancing selection. *PLoS ONE* 4:e5568.
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. Cowman AF, Crabb BS (2006) Invasion of red blood cells by malaria parasites. *Cell* 124:755–766.
27. Ockenhouse CF, Klotz FW, Tandon NN, Jamieson GA (1991) Sequestrin, a CD36 recognition protein on *Plasmodium falciparum* malaria-infected erythrocytes identified by anti-idiotype 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
$^{13}C\alpha/^{13}C\beta$ 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* 121:9008–9012.
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.