Insights into Duffy Binding-Like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from *P. falciparum*

Anthony N. Hodder¹,², Peter E. Czabotar¹,², Alessandro D. Uboldi¹, Oliver B. Clarke¹,², Clara S. Lin¹,², Julie Healer¹, Brian J. Smith¹,²,³ and Alan F. Cowman¹,²

¹The Walter and Eliza Hall Institute of Medical Research, Melbourne 3052, Australia.
²Department of Medical Biology, University of Melbourne, Melbourne 3052, Australia.
³La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086, Australia.

Running head: Structure and function of DBL domains

To whom correspondence should be addressed: Alan F. Cowman or Tony Hodder, The Walter and Eliza Hall Institute of Medical Research. 1G Royal Parade, Parkville 3052, Melbourne, Australia.
Telephone: 61-3-93452446. Facsimile: 61-3-93470852. Email: cowman@wehi.edu.au or hodder@wehi.edu.au

These authors contributed equally

Keywords: malaria; structure; merozoite surface protein

**Background:** *Plasmodium falciparum* invades red blood cells by binding of specific parasite ligands to host cell receptors.

**Results:** We have solved the 3-dimensional structure of the erythrocyte-binding Duffy binding-like domain of an important red blood cell binding protein.

**Conclusion:** The architecture of this protein is distinct from other related proteins.

**Significance:** This has provided insights into the function of this protein and enabled us to map a receptor-binding site.

**SUMMARY**

Invasion of human red blood cells by *Plasmodium falciparum* involves interaction of the merozoite form through proteins on the surface coat. The Erythrocyte Binding-Like (EBL) protein family functions after initial merozoite interaction by binding via the Duffy Binding-Like (DBL) domain to receptors on the host red blood cell. The Merozoite Surface Proteins DBL 1 and 2 (*Pf*MSPDBL1 and *Pf*MSPDBL2) (PF10_0348 and PF10_0355) are extrinsically associated with the merozoite, and both have a DBL domain in each protein. We expressed and refolded recombinant DBL domains for *Pf*MSPDBL1 and 2 and show they are functional. The red cell binding characteristics of these domains was shown to be similar to full-length forms of these proteins isolated from parasite cultures. Furthermore, metal co-factors were found to enhance the binding of both the DBL domains and the parasite-derived full-length proteins to erythrocytes, which has implications for receptor binding of other DBL containing proteins in *Plasmodium spp*. We solved the structure of the erythrocyte-binding DBL domain of *Pf*MSPDBL2 to 2.09 Å resolution and modelled that of *Pf*MSPDBL1, revealing a canonical DBL fold consisting of a boomerang shaped α-helical core formed from three sub-domains. *Pf*MSPDBL2 is highly polymorphic and mapping of these mutations shows they are on the surface, predominantly in the first two domains. For both *Pf*MSPDBL proteins polymorphic variation spares the cleft separating domains 1 and 2 from domain 3, and the groove between the two major helices of domain 3 extending beyond the cleft, indicating these regions are functionally important and likely to be associated with the binding of a receptor on the red blood cell.

The recognition of host receptors by specific parasite ligands is a critical step in the pathogenic cycle of invasion and growth of *Plasmodium* parasites. Proteins belonging to the Erythrocyte Binding-Like (EBL) superfamily are known to play key roles in the complex series of interactions required for merozoite invasion of erythrocytes (reviewed in (1)). The defining feature that classifies members of this superfamily is the Duffy-binding-like (DBL) domain, so called from its identification in *P. knowlesi* and *P. vivax* as the ligand that binds the Duffy antigen receptor for chemokines (DARC) (2,3). DBL domains are encoded in many genes within the *P. falciparum* genome and are found in both EBL and var genes (4). These domains are responsible for diverse pathogenic phenotypes including sequestration of infected erythrocytes in various tissues (5), and rosetting
Structure and function of DBL domains

(6), as well as erythrocyte invasion (reviewed in (1)).

For invasion-related proteins, two types of EBLs have been identified in the genus, having either a single or double, tandemly-repeated DBL domain (7). In *P. knowlesi* and *P. vivax*, a single DBL-encoding gene exists (*Pk*DBP and *Pv*DBP), possessing a single DBL domain, which binds to the DARC receptor as a dimerized form (3,8). The *P. falciparum* EBL ligands, which bind to sialic acid residues of distinct erythrocyte surface proteins, are Erythrocyte-Binding Antigen-175 (EBA-175), which binds to glycophorin A (9), EBA-140 (also known as Baebl), which binds to glycophorin C (10-12), and EBA-181 (also known as Jsebl) whose receptor is unknown (13). In each of these *P. falciparum* proteins, two tandem DBL domains are present.

Two hypothetical proteins containing single DBL domains have been identified in *P. falciparum* (PF10_0348 and PF10_0355). Both proteins have been localized to the merozoite surface (14,15) and functional studies have showed that the full length form of the PF10_0348 protein has erythrocyte binding activity when expressed on the surface of COS cells (14). Furthermore, PF10_0355 was recently identified in a genome-wide association study designed to highlight putative genes involved in anti-malarial drug-resistance, and has been associated with resistance to halofantrine (16). Population genomic analysis has revealed a large number of polymorphisms in both genes, particularly within the *dbl* region, and that both genes are subject to strong balancing selection, an indicator that they are likely to be under intense immune pressure, consistent with their presence on the surface of the merozoite (17).

These DBL-containing merozoite surface proteins (MSP) are of interest as potential malaria vaccine candidates, and as such, it will be useful to obtain a better understanding of their functional role during parasite invasion. Given the very polymorphic nature of these proteins and the functional importance of the DBL domain, a structural analysis would be informative to understand how they bind to their erythrocyte receptors. The crystal structures for both single and tandemly-arranged DBL domains of *Pv*DBP, *Pk*DBP and EBA-175, respectively, have been solved (3,8,18). In both cases, a dimeric organization of the domain appears to be critical in receptor binding.

In this study, we characterize two closely related merozoite surface proteins, *P*MSDBL1 (PF10_0348) and 2 (PF10_0355), and show that erythrocyte binding occurs through the DBL domains of these proteins in a manner enhanced by the presence of specific metal-ion cofactors. We have also derived the crystal structure of the erythrocyte-binding DBL domain of *P*MSDBL2 to 2.09 Å resolution.

**EXPERIMENTAL PROCEDURES**

**Cloning, protein production and purification:** Synthetic genes, codon optimized for expression in *E. coli*, were produced corresponding to residues K161 to L457 in *PfMSPDBL2* and residues K143 to D443 in *PfMSPDBL1* (Geneart). Each gene was cut from the supplier’s vector using *Bam* HI and *Xho* I then ligated into the pProExHTb vector (Life Technologies), which added a cleavable N-terminal Hexa-Histidine fusion tag. The vectors were then transformed into *E. coli* strain BL21 (DE3) for expression of the recombinant DBL domains. Bacteria were grown in super broth at 37°C to an OD600nm of 0.5. Expression was induced with 1 mM IPTG and bacteria were grown for a further 3 hr at 37°C. Both proteins were deposited as insoluble inclusion bodies. Cells were lysed by sonication and the inclusion bodies solubilized by the addition of 6 M guanidine-HCl, pH 8.0. The solubilized proteins were purified by metal-chelating chromatography using a NiNTA column (Qiagen) and eluted in an 8 M urea pH 8.0 solution containing 1 M imidazole. Each protein was refolded at room temperature for two days in 2 M urea/100 mM NaCl/20 mMTris pH 8.0 with 1 mM reduced and 1 mM oxidized glutathione to facilitate disulfide bond formation.

The refolded *P*MSDBL2 DBL domain was further purified using anion-exchange chromatography. Buffer A was comprised of 20 mM Tris pH8.0/250 mM urea and buffer B was comprised of 20 mM Tris pH8.0/250 mM urea/1 M NaCl. Bound proteins were eluted from a 1 ml Hitrap Q column (GE Healthcare) using a linear gradient from 1-50%B and the relevant fractions
pooled and buffer-exchanged into 20 mM Bis-Tris pH6.5/20 mM NaCl then concentrated to 3 mg/ml for crystallization. The PfMSPDBL1 DBL domain was further purified using cation-exchange chromatography. Buffer A was comprised of 20 mM MES pH6.5/20 mM NaCl and buffer B was comprised of 20 mM MES pH8.0/250 mM urea. Bound proteins were eluted from a 1 ml Hitrap S column (GE Healthcare) using a linear gradient from 1-50%B and the relevant fractions pooled and buffer-exchanged into 20 mM MES pH6.5/20 mM NaCl then concentrated to 3 mg/ml for crystallization trials. The majority of the N-terminal fusion tag was removed from the expressed proteins using TEV protease (Life Technologies). However, the five most C-terminal residues from the vector (GAMGS) on the P’ side of the TEV cleavage remained attached to the DBL domains at their N-terminus.

Proteins were electrophoresed using 4-12% Bis-Tris precast gels (Invitrogen) in the presence of DTT-reduced or non-reduced sample buffer. Proteins were electrophoretically transferred onto PVDF membrane using an iBlot apparatus (Invitrogen). Protein G purified pooled hyperimmune sera, from individuals living in malaria endemic regions in PNG, were used to detect proteins on immunoblots.

RP-HPLC analyses of the refolded or denatured forms of PfMSPDBL2 and PfMSPDBL1 DBL domains were performed using an Agilent 1100 LC system. Buffer A comprised 0.05% (v/v) TFA (HPLC/Spectro grade, Pierce, Rockford, Ill. USA) in Milli-Q grade water (Millipore, Bedford, MA. USA); buffer B comprised 0.05% (v/v) TFA in acetonitrile (ChromAR HPLC grade, Malinckrodt, Paris, Kentucky, USA). Samples were loaded onto a C8 column (2.1 mm I.D. x 100 mm (Brownlee columns, Perkin-Elmer Instruments, Norwalk, CT, USA) in the presence of buffer A. Bound proteins were eluted using a linear gradient of 0-100% B over 12 min at a flow rate of 0.5 ml/min.

Reduced and alkylated forms of the DBL domain for PfMSPDBL1 and PfMSPDBL2 were produced by incubating with 50 mM DTT for 1 hr at room temperature then adding iodoacetic acid to a final concentration of 250 mM and leaving the solution for 1 hr in the dark at room temperature. Proteins were then dialysed into 2 M Urea/150 mM NaCl/20 mM Tris pH8.0.

**Generation of HA-tagged PfMSPDBL1 and PfMSPDBL2 parasite lines:** To attach a triple HA tag (3xHA) to the 3’ end of the Pfmspdbl1 (MSPDBL1) gene, an 834 bp fragment of Pfmspdbl1 was amplified from 3D7 genomic DNA using the primers 5’-ATATCCCGCGAATGTGATTGTAAATATAAAG-3’ and 5’-GAAACCTCGAGTTTTGAAATAATCTGTC-3’ (Sac II and Xho I restriction sites underlined). Similarly, to attach a 3xHA tag to the 3’ end of the Pfmspdbl2 (Pfmspdbl2) gene, a 922 bp fragment of PfMSPDBL1 was amplified from 3D7 genomic DNA using the primers 5’-ATTCCCGGAAAAAGCTTTATTAGAAAAG-3’ and 5’-ATTACTCGAGATTTTTAAAAATTTGTAATATC-3’. The DNA fragments were digested with Sac II and Xho I, and cloned into pHAST, a derivative of pGEM-3Z containing a 3xHA tag and a Strep II tag in tandem. Parasites were transfected as described previously (19). Successful integration of the 3xHA tag was determined by Western and Southern blotting.

**Red blood cell binding assay using HA-tagged PfMSPDBL1 and PfMSPDBL2 derived from post-invasion culture supernatants:** For PfMSPDBL1-HA and PfMSPDBL2-HA RBC binding assays, 200 µl of red blood cells (type O+) from three donors were washed twice at room temperature (RT) with 10 ml of RPMI-Hepes containing 0.2% NaHCO₃ (RHN buffer), followed by centrifugation at 1500 rpm for 5 min. Red blood cells were a gift from the Red Cross Blood Service (Melbourne, Australia). For enzyme-treated red blood cells, erythrocytes were incubated with trypsin (bovine pancreas-TPCK treated, 1.5 mg/ml in RHN buffer) and neuraminidase (Vibrio cholera 0.06 U/ml (final) in RHN buffer) for 1 hr at room temperature to digest red blood cell surface proteins or sialic acid residues, respectively. Cells were then washed for 15 min at room temperature with RHN+0.5 mg/ml soybean trypsin-inhibitor or RHN only buffers (for trypsin-treated or neuraminidase treated cells, respectively). The untreated and enzyme-treated red blood cells were then resuspended in 10 ml of post-invasion
culture supernatant containing 1 mM Ca\(^{2+}\) and either \( Pf\)MSPDBL1-HA, \( Pf\)MSPDBL2-HA or control supernatant lacking HA-tagged \( Pf\)MSPDBL proteins, and incubated with gentle shaking for 2 hr at room temperature. The culture supernatant was removed by centrifugation and the red blood cells were washed with 1 ml of RHN buffer. The RHN buffer was discarded and the red blood cells were resuspended in 200 \( \mu l \) of elution buffer (50 mM Tris- HCl, pH 7.6; 1.5 M NaCl) and incubated at room temperature for 10 min with gentle shaking before centrifuging at 10,000 rpm for 30 sec. The eluate was subjected to an additional centrifugation step at maximum speed for 3 min to pellet any red blood cells present. The supernatant was then precipitated using a mixture of methanol-chloroform-water (20), resuspended in SDS-PAGE sample buffer and subjected to western blotting. RBC-binding assays for \( Pf\)MSPDBL1-HA and \( Pf\)MSPDBL2-HA were carried out with or without 1 mM CaCl\(_2\) added to the RHN buffer.

**RBC-binding assay using recombinant DBL domains for \( Pf\)MSPDBL1 and \( Pf\)MSPDBL2:** Assays using recombinant DBL domains for \( Pf\)MSPDBL1 and \( Pf\)MSPDBL2 were carried out as described above with several modifications. Recombinant proteins (10 \( \mu g/assay \)) were made up to 300 \( \mu l \) with the filtrate that was used for their concentration during purification and combined with 300 \( \mu l \) of RHN buffer to give a total volume of 600 \( \mu l \). This volume was incubated with 200 \( \mu l \) red blood cells to allow binding to occur. For some assays, 5 mM Ca\(^{2+}\) or other metal ions (5 mM, final concentration) were included in the assay. It was not necessary to concentrate the protein in the final high salt eluate for analysis of binding by Western blot.

**FACS-based RBC-binding assay using the recombinant DBL domain for \( Pf\)MSPDBL1:** This assay was modified from a previously described protocol (21,22). Trypsin- and neuraminidase-treated red blood cells were prepared as described above. In addition, chymotrypsin-treated red blood cells were produced in a similar manner to that of trypsin-treated red blood cells, using chymotrypsin (TLCK-treated) at 1.5 mg/ml final concentration. Erythrocytes were washed three times with RHN then diluted to a final volume of 1 \( \times 10^7 \) cells/mL. For binding, the recombinant DBL domain of MSPDBL1 was incubated at 0.15 mg/mL for 1 hr at room temperature in RHN buffer containing 1 mM Ca\(^{2+}\) (CaCl\(_2\)). Post binding, erythrocytes were washed twice with 200 \( \mu l \) PBS/1% BSA then incubated with rabbit serum containing \( \alpha \) MSPDBL1-DBL domain antibodies diluted 1:100 in PBS/1% BSA buffer. After an hour incubation, the erythrocytes were washed two times with 200 \( \mu l \) PBS/1% BSA and incubated with goat \( \alpha \) rabbit Alexa Fluor 488 conjugated secondary antibodies (diluted 1:200 in PBS/1% BSA, Molecular Probes) for 1 hr. Samples were washed four times and resuspended in 400 \( \mu l \) PBS/0.1% BSA. 50 000 events were acquired per sample using CellQuest Software (BD Bioscience) on the FACs Calibur flow cytometer (BD Bioscience). The data was analyzed using Flowjo 8.8.7 (Treestar) where the Alexa Fluor 488 fluorescently labeled proteins bound to erythrocytes were calculated as a percentage over unbound erythrocytes. Recombinant PfRh4.9 protein was a generous gift from Dr Wai-Hong Tham (WEHI). This protein was used as an additional control as its red blood cell binding behavior is sensitive to trypsin- but not neuraminidase-treatment of erythrocytes. Recombinant PfRh4.9, in RHN buffer, was incubated with red blood cells at 0.1mg/ml for 1 hr at room temperature then assayed as per the DBL domain of MSPDBL1. PfRh4.9 binding to erythrocytes was detected using an anti-PfRh4 IgG antibody, followed by a secondary anti-rabbit Alexa Fluor 488 antibody (21).

To demonstrate the specific effect of Ca\(^{2+}\) on the binding of the DBL domain of MSPDBL1 to red blood cells, RHN buffer containing protein and 1 mM CaCl\(_2\) was supplemented with increasing amounts of ethyleneglycol-bis-(\( \beta \)-aminoethyl ether) N, N', N',-tetraacetic acid (EGTA) over a range from 0 to 60 mM.

**Crystallization Diffraction Data and structure determination:** Crystals obtained in initial sparse matrix screens with His-tagged protein diffracted relatively poorly, and diffraction was often...
irreproducible. Removal was effected by adding TEV protease (1 to 50 volume ratio) to the protein sample for 3 hr prior to setting up drops, leaving five residues (GAMGS) from the vector attached to the N-terminus of the DBL domain. The protein crystallized out of a variety of conditions containing medium weight PEGs and salt - the crystal used for data collection was crystallized at a protein concentration of 2 mg/mL from 20 % PEG3350 0.2 M sodium thiocyanate. In an attempt at derivatisation, 0.8 µL of well solution and 0.2 µL 10mM KAuCl4 was added to the crystals 24 hr prior to mounting and cryocooling, however, we found no evidence for derivatisation in either the anomalous signal nor in the final maps. The crystal used for data collection was cryoprotected by quickly soaking in 25% PEG3350, 15% ethylene glycol, 0.2M NaCl and cryocooled by plunging into liquid nitrogen.

X-ray datasets were collected at the Australian Synchrotron on beam line MX2 at 100 K. Data were processed with HKL2000 (23) and the structure was solved by molecular replacement with PHASER (24) using a BALBES (25) model based on the structure of PfEMP1 VAR2CSA DBL 6e (2WAU). Initial maps revealed little density beyond the molecular replacement model and multiple attempts at building and refinement neither enhanced maps nor reduced R_work/R_free. Multi-crystal averaging was subsequently performed using DMMULTI from the CCP4 suite (26) incorporating data from a second lower resolution crystal form (P21212, a = 61.6, b = 89.6, c = 60.8 Å; resolution 2.7 Å) grown in crystal drops lacking TEV protease and thus retaining vector derived sequence. The subsequent maps were significantly improved permitting manual model building in COOT (27) and refinement in PHENIX (28). The final refinement statistics are shown in Table 1.

Crystals of the structure determined for the DBL domain of PfMSPDBL2 in the presence of Zn2+ were obtained from a protein preparation of a larger refolded recombinant fragment of PfMSPDBL2 (unpublished data). During crystallization trials this protein was processed by an unknown protease leaving only the DBL domain of PfMSPDBL2. Crystals grew from this preparation in the same crystallization conditions described above for DBL domain and were of the same space group. Crystals were soaked for 5 min in mother liquor supplemented with 50 mM ZnCl2, then washed in mother liquor prior to cryopreservation. X-ray datasets were collected at the Australian Synchrotron on beam line MX2 at 100 K and at a wavelength near the Zinc absorption edge (1.2781 Å). The structure was solved by molecular replacement using PHASER (24) with the model of the DBL domain of PfMSPDBL2 as a search model. Model building was performed in COOT (27) with refinement performed, and anomalous difference maps generated, with PHENIX (28).

Structural comparison: Structural alignments of the DBL domains from PF10_0355, P/EBA 175 (F1 and F2) (1ZRO), Pk DBP (2C6j), Pv DBP (3RRC), PfEMP1-NTS-DBL1-VarO (2XU0), PfEMP1 VAR2 CSA DBL3X (3BOi/3CMl), PfEMP1 VAR2 CSA DBL6 (2WAU), and the modelled structure for P/332 DBL domain (29) were performed using the Sequioa program (30).

Modeling of the PfMSPDBL1 structure: The structure of the PfMSPDBL2 DBL domain was used as a template in comparative modelling of the corresponding DBL domain region in PfMSPDBL1 using the MODELLER (9v7) program (31) based on the sequence alignment presented in Fig. S1. From 25 initial models, the model with the lowest Modeller Objective Function was used for further loop modelling. Two loops missing the electron density of the structure of PfMSPDBL2 (residues 172–185, and 375–387) were refined using the loop modelling utility, a part of the MODELLER package; the final model yielded the lowest Modeller Objective Function from 25 loop refinement models.

Molecular electrostatics were calculated using the MEAD program (32) using PARSE atomic charges and radii (33) an internal dielectric of 4.0, external dielectric of 78, and an ionic strength of 0.10. The molecular surface was calculated using the MSMS program (34). The molecular electrostatic surface of PfMSPDBL2 was calculated on a model in which the loops missing in the X-ray crystal structural were included using the MODELLER program.

RESULTS
PfMSPDBL2 is a red cell binding protein on the merozoite surface that contains DBL and SPAM domains. The merozoite surface protein PfMSPDBL1 (PF10_0348), which binds the erythrocyte surface, consists of DBL and SPAM domains (residues 542-696 in PfMSPDBL1, and residues 571-672 in PfMSPDBL2) (14) – the SPAM domain is characterised by an acidic rich region followed by a leucine zipper-like (LZL) region (Fig. 1). In all members of the MSP3-like family and in both PfMSPDBLs there is also a conserved sequence motif of NLR[K/N][A/G/N] near the N-terminus whose significance is not known. Since it lacks a transmembrane region or a glycosyl phosphatidyl inositol (GPI)-anchoring motif, PfMSPDBL1 is extrinsically associated with the merozoite surface (Fig. 1). A second gene (PF10_0355) that is closely linked on the same chromosome, encodes a protein of similar structure, also consisting of a single DBL and SPAM domain. We henceforth refer to this latter protein as PfMSPDBL2 and the former (PF10_0348) as PfMSPDBL1 (Fig. 1). The sequence identity and similarity between the DBL domains of PfMSPDBL2 and PfMSPDBL1 is 38% and 62%, respectively (Fig. S1), whilst for the SPAM domain is 30% and 53%, respectively. The similar domain structure and significant sequence homology between PfMSPDBL1 and PfMSPDBL2 suggest that they are both involved in the interaction of the merozoite with the red blood cell (Fig. 1).

To further analyse PfMSPDBL2 and PfMSPDBL1 we transfected the 3D7 strain of P. falciparum with a plasmid that when inserted by homologous recombination would place a sequence encoding a haemagglutinin (HA) tag at the 3’ end of the pfmspdlb1 and pfmspdlb2 genes (3D7MSPDBL1-HA and 3D7MSPDBL2-HA). Immunoblots with anti-HA antibodies confirmed that both 3D7MSPDBL1-HA and 3D7MSPDBL2-HA modified parasite lines expressed full-length forms of HA-tagged PfMSPDBL1 and PfMSPDBL2, respectively (Fig. 2A). Interestingly, parasite-derived PfMSPDBL1 and 2 harvested from invasion supernatants was largely observed as a processed form that was approximately 10-15 kDa smaller than the larger forms. This processing must occur at the N-terminus as the HA tag used for recognition remained attached to the C-terminus of these proteins that were expressed in transfected parasites. Similar processing of parasite-derived PfMSPDBL1 has also been observed previously, however, the orientation of the cleavage event could not be determined by the antibodies used for detection of this protein (14). PfSUB1 cleavage sites have been predicted near to the N-terminus of PfMSPDBL1 and 2, and cleavage at these sites would result in processed forms of similar size to those observed in Fig 2A (35). Similar processing was also observed in wild type parasites (Fig. S2). Saponin-lysis of late-stage schizonts revealed the majority of the PfMSPDBL proteins were found in a soluble and unprocessed form. However, the processed forms of these proteins were found only in the insoluble pellet fraction, suggesting that processing is required before they can be incorporated onto the merozoite surface (Fig. S2).

To confirm erythrocyte-binding activity of native parasite protein PfMSPDBL1 and to determine if PfMSPDBL2 was also able to bind human erythrocytes, C-terminal HA-tagged protein was harvested from culture supernatants and tested in red blood cell binding assays (Fig. 2A). Both PfMSPDBL1 and PfMSPDBL2 parasite-derived proteins bound red blood cells. Binding of both ligands was consistently resistant to trypsin and neuraminidase treatment of the red blood cells (Fig. 2B). This was in contrast to binding of EBA-175 on the same enzyme-treated erythrocytes, which was sensitive to both enzyme treatments as described previously (36). The binding phenotype observed for parasite-derived PfMSPDBL1 differs to that previously reported for this protein expressed on the surface of COS-7 cells (14). However, it was clear that PfMSPDBL2 is also a P. falciparum invasion ligand that binds to human erythrocytes suggesting that it plays a direct role in merozoite invasion. PfMSPDBL1 and 2 were also found to have identical red blood cell binding phenotypes under the conditions used.

The location of PfMSPDBL2 on the merozoite surface was determined using immuno-fluorescence assays (IFA) with anti-PfMSPDBL2 antibodies. It showed strong co-localisation with the known merozoite surface protein Merozoite Surface Protein 1 (MSP1) in both schizont and free merozoite stages (Fig. 2C, first and second row). PfMSPDBL2 co-localised
with *Pf*MSPDBL1 in mature schizont stages both giving a ‘bunch of grapes’ pattern, consistent with their presence on the merozoite surface (Fig. 2C, first and third row). They were also co-localised on the surface of free merozoites although *Pf*MSPDBL2 appeared to be more concentrated at the apical end of the cell as indicated by its close apposition to RON4, a protein located in the neck of the rhoptries (Fig 2C, bottom row) (37). This confirmed previous work showing that *Pf*MSPDBL1 was located on the merozoite surface and also demonstrated that *Pf*MSPDBL2 had a similar localisation consistent with them having similar functions (14).

**Recombinant DBL domains of *Pf*MSPDBL1 and *Pf*MSPDBL2 bind human red blood cells.** To further analyse the binding of *Pf*MSPDBL1 and *Pf*MSPDBL2 to red blood cells we produced recombinant forms of the DBL domains for both proteins in *E. coli*. These expressed proteins were found in inclusion bodies which were separated and dissolved in 6 M guanidinium hydrochloride, purified under denaturing conditions by NiNTA immobilised metal affinity chromatography (IMAC), oxidatively refolded, and further purified by ion-exchange chromatography (Fig. 3A, B, C and D). Differential migration was observed for both refolded proteins on SDS-PAGE under reducing and non-reducing conditions indicating an influence of disulfide bond architecture on the shape and/or SDS-binding ability of these domain fragments (Fig. 3C and D). Attainment of native conformational epitopes in the *in vitro* refolded DBL domains was demonstrated using immunoblots. The DBL domains of MSPDBL1 and 2 were electrophoresed in sample buffer with or without DTT, transferred onto PVDF membrane and the immunoblots probed with antibodies from pooled hyperimmune sera obtained from adult individuals living in malaria endemic regions of Papua New Guinea (Fig. 3E and F). The ability of these antibodies to recognize the DBL domains was largely dependent on the existence of epitopes with a conformation stabilized by disulfide bonds. Antibodies targeting the *Pf*MSPDBL1 DBL domain reacted exclusively with disulfide bond dependent epitopes (Fig. 3E), while those targeting the *Pf*MSPDBL2 DBL domain largely recognized conformational epitopes but additionally linear epitopes were also recognized in the DTT-reduced sample (Fig. 3F). RP-HPLC was used to demonstrate a decrease in hydrophobicity for each of the DBL domains had occurred as a result of the oxidative *in vitro* refolding process (Fig. 3G and 3H). The refolded material eluted significantly earlier than the denatured starting material consistent with internalization of hydrophobic residues upon refolding (Fig. 3G and 3H). The monomeric forms of these DBL domains were found to be stable for extended periods at 4°C, further indicating no reactive surface accessible Cys residues were present in the final products.

Recombinant *Pf*MSPDBL1 and *Pf*MSPDBL2 DBL domains were functional in traditional red blood cell binding assays in which bound proteins are eluted from the cells, after passage through oil, and detected on immunoblots. Both DBL domains demonstrated a binding phenotype that was insensitive to treatment of erythrocytes with either trypsin or neuraminidase (Fig. 4A). Furthermore, the addition of Ca²⁺ to these assays was found to enhance the level of binding for both the parasite-derived, full-length (Fig. 2B, 4D) and recombinant DBL domains of *Pf*MSPDBL1 and *Pf*MSPDBL2 (Fig. 4A) while the effect could also be reversed in the presence of EGTA (Fig. 4D). However this increased level of binding was only induced by the presence of specific metal ions, with Ca²⁺, Cu²⁺ and Ni²⁺ having the greatest effect for *Pf*MSPDBL1, and Cu²⁺, Co²⁺, Zn²⁺ and Ca²⁺ for *Pf*MSPDBL2 (Fig. 4B). In addition, binding was found to be dependent on a disulfide-bond stabilised conformation within the DBL domains of MSPDBL1 and 2, as reduced and alkylated forms of these proteins could not bind red blood cells (Fig. 4C).

In order to further validate the red blood cell binding behaviour of the DBL domains of MSPDBL1 and 2 we used FACS-based assays to assess their binding to a suspension of cells. The antibody used for the detection of rDBL1 on immunoblots (Fig. 4A, B and C) was suitable for use in the FACS-based assay. However, several different antibodies, including the mouse anti-HexaHis monoclonal antibody could not detect rDBL2, recombinant or parasite-derived forms of full-length MSPDBL1 and 2 or parasite-derived EBA175. We assume this is due to steric...
hinderance arising from the interaction between the red blood cell receptor(s) and these proteins. However, the FACS-based erythrocyte binding data obtained for rDBL1 (Fig 4E, F and G) was found to complement the observations determined for this protein in binding assays where rDBL1 was eluted from erythrocytes and detected on immunoblots (Fig 4 A, B and C). The binding of rDBL1 was found to be insensitive to treatment of erythrocytes with either trypsin, neuraminidase or chymotrypsin (Fig. 4E). Furthermore, the binding of rDBL1 to erythrocytes was dependent on the disulfide-bond stabilised conformation of this domain (Fig. 4F). The effect of Ca\(^{2+}\) on binding was demonstrated by titrating EGTA (a metal-ion chelating agent) over a concentration range from 0 to 60 mM. This effect was specific for rDBL1 as the binding of recombinant PfRh4.9 to erythrocytes was unaffected by the level of EGTA under similar assay conditions (Fig. 4G).

Combined these results show that the recombinant DBL domains of PfMSPDBL1 and PfMSPDBL2 have the same functional properties as the full-length mature forms of the PfMSPDBL proteins derived from parasite culture supernatants (Fig. 2A), and suggest that the affinity for receptor binding of both PfMSPDBL1 and PfMSPDBL2 is influenced by the presence of specific metal ions.

Structure of the PfMSPDBL2 DBL domain. We determined the crystal structure of the PfMSPDBL2 DBL domain (residues 161 to 457) to 2.09 Å resolution revealing a canonical DBL fold consisting of a boomerang shaped α-helical core formed from three subdomains (Table 1, Fig. 5). Electron density was apparent for residues 161 to 171, 186 to 374 and 388 to 457. Patches of poorly defined density connected residues 171 to 186 but was of inadequate quality for confident model building. However, the disulfide bond that was found nested within the Cys177-Cys212 linkage (C2-C3 discussed below) was not present in subdomain 1. Additionally, unlike the canonical structure, disulfide linkages occur between Cys162-Cys334 and Cys168-Cys323 in PfMSPDBL2 (Fig. 7). These linkages tether a small helical region in subdomain 1 (helix 1a), between residues 161 to 171, to the C-terminal end of helix 5 and the loop region between helix 5 and helix 6 in subdomain 2 (Fig. 5C, Fig. 6).

PfMSPDBL2 subdomain 3 (344–460) is a helical bundle comprised of two long α-helices (helix 6 and 7) and two smaller α-helices (helix 8 and 9) (Fig. 6, Fig. 5D). The two larger helices are tethered to each other via a single disulfide bond (Cys372-Cys393) at the distal end of the bundle. A Cys358-Cys446 disulfide linkage anchors the small loop region between helices 8 and 9 to the mid-region of helix 6. A disulfide linkage between Cys441 and Cys444 constrains this same loop region and brings helices 8 and 9 in close proximity to each other in an anti-parallel orientation (Fig. 7). The loop between helices 6 and 7 (residues 374 to 388) was disordered (Fig. 6), as has been observed for other DBL structures (eg. PfEMP1 VAR2CSA DBL 6ε, 2WAU) (38). Structures of DBL domains in which this region is visible, for example the PfEBA175 F1 DBL (1ZRU F1) (18), contain C-terminal extensions that interact with this loop and thus stabilize the region. Residues corresponding to this C-terminal
extension were not included in the construct for PfMSPDBL2 used in these studies.

For a second structure of PfMSPDBL2, crystals were soaked with ZnCl₂ and data collected near the Zinc absorption edge. Anomalous difference maps revealed the location of four sites of zinc coordination, none of which conform to the classical zinc functional types (catalytic, structural, co-catalytic, or interfacial) (39) (Fig. 8). At three sites the zinc ion coordinates a single histidine and either a glutamate or tyrosine. Water molecules fill the remaining coordination valency at all three sites. At the fourth site a single aspartate coordinates the metal ion. While the data for this structure was nominally at lower resolution than that collected for the native, electron density unambiguously extended beyond residue 171 whilst in the native dataset only patchy density had been observed. This permitted additional model building from residue Ser172 to Lys179, including the cysteine residue at position 177 – this additional density also confirmed the disulfide connectivity between Cys177 and Cys212.

A model of the DBL domain for PfMSPDBL1. The DBL domain of PfMSPDBL1 would not crystallize under the myriad of conditions used. The high sequence similarity between PfMSPDBL1 and PfMSPDBL2 permitted the comparative modeling of the DBL domain of PfMSPDBL1 using PfMSPDBL2 as a template; in particular, the disulfide bond connectivity pattern was predicted to be similar in both domains (Fig. S1). Notably, however, an additional Cys residue (Cys398) remains unpaired in helix 7 of subdomain 3 of PfMSPDBL1. This Cys residue differs in location to other unpaired Cys residues found in 2WAU (C2385 in a small helix found in subdomain 1) and 1ZRO(F1) (C273 in helix 9 of subdomain 3) (Fig. 6). Ligand interaction sites in DBL domains other than PfMSPDBL1 and PfMSPDBL2 are associated with clusters of positively charged basic residues. The electrostatic potential surface of PfMSPDBL1 and PfMSPDBL2 (Fig. 9B and 9D, respectively) show large patches of positive potential, particularly at the C-terminus of helix 2a. The cleft between subdomains 2 and 3 of PfMSPDBL1 shows a large negatively charged region, which was not observed on the surface of PfMSPDBL2.

The DBL domains of PfMSPDBL2 and PfMSPDBL1 are highly polymorphic. Both PfMSPDBL2 and PfMSPDBL1 proteins display some of the highest levels of sequence diversity so far found within the P. falciparum genome, with the majority of the polymorphisms located in the DBL domains of these proteins (17). Many of the mutations are dimorphic in nature, with approximately 10-20% of the mutated amino acids involving a switch in polarity (Table 2). Placement of the naturally occurring polymorphisms onto a surface structure for the DBL domain of PfMSPDBL2 showed good coverage around the surface of the molecule (Fig. 10A and B), except for a cleft region that lies between helices 1 and 6, and the groove between helices 6 and 7 that extends beyond the cleft (Fig. 10A). The composition of the residues associated with this region between subdomains 2 and 3 is conserved and restricted from mutational change, indicating that this region of the DBL domain is functionally important and may be associated with the interaction of a receptor on the surface of the red blood cell. Significantly fewer polymorphisms are known for PfMSPDBL1 than for PfMSPDBL2. Mapping of the known polymorphisms of PfMSPDBL1 onto the surface of the modeled structure reveals that most are distributed equally across the surface (Fig. 10D) and, like PfMSPDBL2, the region between subdomains 2 and 3 remains free of polymorphisms.

**DISCUSSION**

Key host/parasite molecular interactions that occur during the invasion process remain poorly understood. Here we have characterized two unusual members of the PfMSP3 family that interact with red blood cells via a single DBL domain located near each N-terminus. We have determined the 3-D structure of the DBL domain from PfMSPDBL2 and using this information modeled PfMSPDBL1 to show that they both have a canonical DBL fold consisting of a boomerang shaped α-helical core formed from three subdomains. We have used this to map the sequence polymorphisms associated with these proteins in the P. falciparum population and this has highlighted a region most likely corresponding to the receptor-binding domain.
The similar domain structures for P/MSPDBL1 and 2 and their shared subcellular localization on the merozoite surface suggest that they both play a similar role in merozoite invasion, most likely in the initial interaction of the parasite cell with the host.

P/MSPDBL1 and 2 have no apparent transmembrane spanning domains or GPI-anchoring motifs and they associate peripherally with other parasite proteins to form complexes at the merozoite surface. We have found that processing of these proteins was linked with their association on the merozoite surface (Fig. S2). Both P/MSPDBL1 and 2 have potential P/SUB1 cleavage sites located near to their N-terminus and processing by this protease would produce fragments of similar size to those observed for the processed forms of these proteins. A similar P/SUB1 processing event is required for incorporation of MSP6, another member of the MSP3 family, into the MSP1 complex (40,41). It is possible that unprocessed forms of P/MSPDBL1 and 2, which are not incorporated into the merozoite surface (Fig S2), may act as immuno-decoys as these proteins will be presented to the host immune system in a different form to the P/MSPDBL proteins which form part of a complex on the merozoite surface.

We believe the DBL domain from each P/MSPDBL protein is the region of the molecule that interacts with the receptor(s) on the surface of the red blood cell for the following reasons: (i) the red blood cell binding properties of the recombinant DBL domains and the mature full-length forms of the P/MSPDBL proteins, obtained from parasite culture supernatants, are similar and greatly enhanced in the presence of specific metal ions, such as Ca$^{2+}$; (ii) other members of the MSP3 family which do not contain DBL domains yet retain the C-terminal SPAM domain, as found in the P/MSPDBL proteins, do not bind red blood cells in identical binding assays (Unpublished data) and (iii) the interaction between the DBL domains of other merozoite surface proteins involved in invasion, such as members of the EBL family, is well documented (18, (18)).

The structure of the DBL domain for P/MSPDBL2 was also determined in the presence of Zn$^{2+}$, which was one of the metal ions identified to enhance the binding of this DBL domain to red blood cells. Four distinct Zn$^{2+}$ binding sites were identified within the structure. Functional zinc binding sites generally employ at least three side-chain ligands to coordinate the metal ion (42). In three of the binding sites identified two side-chain ligands are observed, and in the fourth only a single side-chain ligand is observed. Enhanced binding of P/MSPDBL1 and P/MSPDBL2 to red blood cells is also observed for calcium ions and these do not generally coordinate histidine. Therefore, it is unlikely that calcium could function in at least three of these sites in the same manner as zinc, indicating other metal ion binding sites may exist within these DBL domains. However, the enhanced binding of these parasite ligands to the red blood cell was found to be specific to only a few metal ions, as not all cations tested in the red blood cell binding assay resulted in elevated levels of binding for each DBL domain. Some members of the lectin family rely on the presence of metal ions, such as Ca$^{2+}$, to enable the formation of high affinity interactions with carbohydrate moieties in their binding partners (43). The Ca$^{2+}$ concentration in the parasitophorous vacuole, that contains the dividing merozoites, is approximately 100 nM and 1 mM in serum and therefore it is likely that some of the Ca$^{2+}$ binding sites in P/MSPDBL1 and P/MSPDBL2 are occupied thus enabling enhanced affinities of these surface proteins for their specific receptors (44). It will be interesting to determine if DBL domains in other EBL family members are able to bind Ca$^{2+}$ or metal ions and if this plays a role in affinity of binding to their receptor.

The binding of the DBL domains, from the P/MSPDBL proteins, to their receptors on the red blood cell surface was dependent on a conformation stabilized by disulfide bonds, as the reduced and alkylated forms of these proteins were unable to bind the host cell. Furthermore, both proteins had a similar red blood cell binding phenotype that was both trypsin and neuraminidase resistant. This binding phenotype was also observed for the parasite-derived proteins but differed from that previously reported for P/MSPDBL1 expressed on the surface of COS-7 cells (14). It is not clear why the pattern of binding differs between these two studies, but may possibly be due to misfolding of the protein expressed in COS-7 cells. It will be
of interest to identify the specific receptor(s) for these merozoite surface proteins.

The SPAM domain, which is present in most other members of the MSP3 family, consists of a glutamic acid rich region and a leucine zipper–like region – the alanine-rich heptad repeat region of the parent SPAM protein (45) is absent from all other members of the MSP3 family in *P. falciparum* (15). The leucine zipper-like region has been shown to be responsible for oligomerisation of MSP3. Analysis of the physical properties of MSP3 has suggested that it forms a highly extended and oligomeric structure on the merozoite surface that would allow it to interact with the red blood cell at relatively long distances (46) – notably, dimerization of MSP3 is solely through the leucine zipper-like domain. As *Pf*MSPDBL1 and 2 bind red blood cells through the DBL domains and share with MSP3 a SPAM domain at the C-terminus, they are also likely to be long extended molecules able to bind to their specific receptors. The function of these proteins is likely to be associated with the initial interaction of the merozoite with the red blood cell and as such important for the invasion process.

Although the overall structures for the various DBL domains found in *Plasmodium* spp. are relatively conserved, there is great variation in the host receptors with which they interact (reviewed in (1,47)). The specificity of the DBL domain structure is poorly understood in relation to function/receptor binding. An alignment of the sequences of Duffy binding-like (DBL) domains shows that the size and location of the major helices (i.e. helices 1-9) is highly conserved in the structure of known DBL domains (Fig. 6). However, the extent of the connecting loops between these helices varies with each structure. The DBL fold is classically divided into three sub-domains. Subdomain 1 contains little conserved secondary structure except for a short helical segment (helix 1) between Cys3 and Cys4 conserved in all structures except 2C6J. However, this region of 2C6J includes a sequence anomaly (Val39Asp) in the solved structure. Subdomain 2 is composed of 4 structurally conserved helices (helices 2-5), with helix 5 bent mid-way in all structures. Subdomain 3 is formed by two long anti-parallel helices (6 and 7) and several smaller helices across one face of this helix bundle; in the canonical DBL structure this region includes two helices, 8 and 9, however there are 5 helices found in the structure 2WAU.

Despite a high degree of sequence similarity, few residues are strictly conserved across the seven DBL structures determined to date. An arginine (Arg207) in the short helix between Cys3 and Cys4 forms salt bridges with a conserved aspartate (Asp266) on helix 2 and glutamate (Glu352) on helix 6. A conserved glutamate in helix 2 forms a hydrogen bond with the backbone amide nitrogen immediately prior to Cys4. The indole nitrogen of a conserved tryptophan in helix 6 (W349 in *Pf*MSPDBL2) forms a hydrogen bond with the carboxylate side-chain of a conserved aspartate in helix 3 (D266 in *Pf*MSPDBL2) and packs against a conserved proline at the beginning of the short helix between Cys 3 and Cys4 (P205 in *Pf*MSPDBL2; notably, in 2C6J this residue is a serine and the short helix 1 is not correctly folded). The N-terminus of helix 1 is capped by another conserved tryptophan two residues C-terminal of Cys2 (W187 in *Pf*MSPDBL2). The conserved aspartate in helix 3 (D266 in *Pf*MSPDBL2) also forms hydrogen bonds with a conserved glutamine (Q345 in *Pf*MSPDBL2) at the beginning of helix 6. Thus, residue conservation in these DBL domains is spatially focused about the junction between helices 1 and 6.

Although the DBL domains from different protein families share similar overall folds, different DBL subgroups demonstrate distinct disulphide linkage patterns (Fig. 7). The canonical disulfide linkage pattern is based on that observed in the F2 DBL domain of *Pf*EBA175, which is a member of the EBL family that binds glycoporphin A (18). Members of the EBL family generally; (1) do not have disulfide-connectivity between any of the subdomains, (2) conserve the C1-C4 and C2-C3 linkages in subdomain 1, (3) conserve the C5-C6 linkage in subdomain 2, and (4) conserve the C8-C10, C9-C14 and C11-C13 linkages in subdomain 3. In contrast, *Pf*EMP1 family members have a modified linking pattern across subdomain 3. In particular, C14 links to a cysteine not observed in the EBL family (C9a) located in an insertion between helices 6 and 7,
and C9 is either absent or links with an additional cysteine (C9b) also within this insertion. Additionally, \( P/\text{EMP1-NTS-DBL1} \) and \( P/\text{EMP1 VAR2CSA DBL3X} \) contain an extra disulfide link connecting subdomains 1 and 2 (C1a-C6a). Whilst disulfide connectivity for \( P/\text{EMP1 VAR2CSA DBL 6e} \) is similar to \( P/\text{EMP1-NTS-DBL1} \) and \( P/\text{EMP1 VAR2CSA DBL3X} \) within subdomain 3, it is unique in that it has lost all disulfide links within subdomains 1 and 2. Additionally, it has acquired an extra disulfide linkage (C12b-C12a) arising from cysteine residues present in a small insertion at the C-terminus of alpha 8. Amongst all the structures determined to date, there occurs only two instances where a cysteine residue is not found in a disulfide bond, Cys273 in 1ZRO(F1), and Cys2385 in 2WAU. Both are surface exposed, and could potentially form a disulfide bond with partner cysteine residues.

The disulfide bond architecture observed in the DBL domain of \( P/\text{MSPDBL2} \) is similar to that of \( P/\text{EMP1-NTS-DBL1} \) and \( P/\text{EMP1 VAR2CSA DBL3X} \) in that it contains disulfide links connecting subdomains 1 and 2 (C1a-C6a). However, unlike other DBL domains, C6 of \( P/\text{MSPDBL2} \) links to a novel N-terminal cysteine (C1b) providing an extra covalent tether between these domains. Additionally, \( P/\text{MSPDBL2} \) has lost the C2-C3 linkage in subdomain 1, and the C9-C14 and C11-C13 linkages in subdomain 3. Subdomain 3 contains the small insertion at the end of alpha 8 containing the short hairpin disulfide bond between C12b and C12a observed in \( P/\text{EMP1 VAR2CSA DBL 6e} \). These unique characteristics place the \( P/\text{MSPDBL2} \) proteins in their own subgroup in regards to their disulfide architecture.

Previous structures of DBL domains show almost no similarity in the location of proposed substrate binding sites (Fig. S3). Additionally, the majority of residues implicated in ligand interaction are not conserved across DBL domains (Fig. 6). A heparin-binding site on \( P/\text{EMP1-NTS-DBL1} \) (2XU0) maps to the N-terminal extension to subdomain 1 and the flanking C-terminal region of subdomain 3 (48). In contrast, the primary binding site for chondroitin sulfate proteoglycans on VAR2CSA DBL6e (2WAU) lies at the N-terminus of helix 2, on the 'distal side' of subdomain 2 (49). The ligand-binding site on \( P/\text{EMP1 DBL (2C6J)} \), which binds to Duffy antigen receptor for chemokines (DARC), comprises residues from the N-terminus of helix 3 and the loop region between subdomains 2 and 3 (3). The proposed binding site for chondroitin sulfate A (CSA) on VAR2CSA DBL3x (3BQI, 3CML) is located in the region formed at the junction of all three domains. Whilst disparate in location, most sites identified contain lysine residues. Likewise, \( P/\text{MSPDBL2} \) contains lysine residues in similar regions indicating that it has the capacity to form similar types of interaction.

It has also been proposed for both EBA-175 (1ZRO) and \( P/\text{DBP (3RCC)} \) that dimerization is required for receptor binding, and that residues at the dimer interfaces are involved in these interactions. Again, however, two disparate dimerization models have been identified with different ligand binding sites. EBA-175 contains two tandem DBL domains joined through a three-turn helical linker (18). Dimerization of these domains produces a "handshake" orientation between two RII regions and results in the formation of a channel. The glycan (sialic acid) binding sites occur at dimer interfaces, on loops between strands in subdomain 1 and the base of subdomain 3. However, as \( P/\text{MSPDBL2} \) contains only one DBL domain, and the ligand is not neuraminidase sensitive, it is unlikely that it utilizes the same binding mechanism. Unlike EBA-175, the DARC binding site of \( P/\text{DBP DBL (3RCC)} \) occurs at the interface of a pair of dimers involving subdomain 2 (Fig. S3). The \( P/\text{MSPDBL2} \) structure is incompatible with this dimer due to steric clashes at the interface involving helix 1c and helix 3. Thus, it is also unlikely that \( P/\text{MSPDBL2} \) utilizes this dimerization surface for ligand interaction.

Whilst \( P/\text{MSPDBL2} \) is unlikely to adopt the dimerization interfaces described for either EBA-175 (18) or \( P/\text{DBP DBL (3)} \), it remains possible that dimerization is still required for ligand recognition. In the full-length protein dimerization is probably facilitated by the C-terminal SPAM domain present in both \( P/\text{MSPDBL1} \) and 2, as is the case for MSP3 (46). Dimerization via this motif would likely orientate two DBL domains in a parallel fashion,
however, flexibility within the linker region between the DBL and SPAM domains may allow for a variety of DBL dimer interfaces. The structure of full length PfMSPDBL1 or 2, or of the DBL domains in complex with receptor, will likely inform the nature of the receptor recognition complex.

A large positively charged region on the surface of PfMSPDBL2 is reminiscent of similarly charged regions on the surfaces of other DBL domains, regions that are associated with binding of receptors or protein partners. However, this region on the surface of PfMSPDBL2 is highly polymorphic and thus it is doubtful that the positive potential has any significance for receptor binding. The significance of the acidic region at the interface of subdomains 2 and 3 of PfMSPDBL1 is less clear since the characterization of polymorphisms is less complete, although this region is resistant to substitution in PfMSPDBL2 also. This region coincides with the few residues conserved across all DBL domains (which tether helix 1), raising the prospect that this region also plays a role in receptor binding.

In summary, we have characterized both PfMSPDBL1 and PfMSPDBL2 and shown that the DBL domains of these proteins are responsible for red blood cell binding. This binding, although not sensitive to trypsin, chymotrypsin or neuraminidase treatment, is enhanced in the presence of divalent cations, specifically Ca$^{2+}$, Co$^{2+}$, Cu$^{2+}$ or Zn$^{2+}$. We have determined the structure of PfMSPDBL2 and find that it displays a typical DBL fold. We propose that the disulfide bonding architecture of PfMSPDBL1 and PfMSPDBL2 is distinct from both the PfEMP1 family and the EBL family differentiating DBL proteins into three subgroups. We also find that the structure, and domain architecture, of PfMSPDBL2 is inconsistent with dimerization models proposed for other DBL domains of erythrocyte-binding ligands.

REFERENCES

1. Cowman, A. F., and Crabb, B. S. (2006) Cell 124, 755-766
2. Chitnis, C. E., and Miller, L. H. (1994) J. Exp. Med. 180, 497-506
3. Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E., and Sharma, A. (2006) Nature 439, 741-744
4. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shalmon, S. J., Suh, B., Peterson, J., Angiuli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002) Nature 419, 498-511.
5. Howell, D. P., Levin, E. A., Springer, A. L., Kraemer, S. M., Phippard, D. J., Schief, W. R., and Smith, J. D. (2008) Mol Microbiol 67, 78-87
6. Rowe, J. A., Moulds, J. M., Newbold, C. I., and Miller, L. H. (1997) Nature 388, 292-295
7. Adams, J. H., Blair, P. L., Kaneko, O., and Peterson, D. S. (2001) Trends Parasitol. 17, 297-299
8. Batchelor, J. D., Zahm, J. A., and Tolia, N. H. (2011) Nat. Struct. Mol. Biol. 18, 908-914
9. Sim, B. K. L., Chitnis, C. E., Wasniowska, K., Hadley, T. J., and Miller, L. H. (1994) Science 264, 1941-1944
10. Mayer, D. C., Kaneko, O., Hudson-Taylor, D. E., Reid, M. E., and Miller, L. H. (2001) Proc. Natl. Acad. Sci. USA 98, 5222-5227
11. Maier, A. G., Duraisingham, M. T., Reeder, J. C., Patel, S. S., Kazura, J. W., Zimmerman, P. A., and Cowman, A. F. (2003) Nat. Med. 9, 87-92.
12. Lobo, C. A., Rodriguez, M., Reid, M., and Lustigman, S. (2003) Blood 101, 4628-4631
13. Gilberger, T. W., Thompson, J. K., Triglia, T., Good, R. T., Duraisingham, M. T., and Cowman, A. F. (2003) J. Biol. Chem. 278, 14480-14486
14. Wickramarachchi, T., Cabrera, A. L., Sinha, D., Dhawan, S., Chandran, T., Devi, Y. S., Kono, M., Spielmann, T., Gilberger, T. W., Chauhan, V. S., and Mohammed, A. (2009) *Int. J. Parasitol.* 39, 763-773

15. Singh, S., Soe, S., Weisman, S., Barnwell, J. W., Perignon, J. L., and Druihle, P. (2009) PLoS One 4, e5410

16. Van Tyne, D., Park, D. J., Schaffner, S. F., Neafsey, D. E., Angelino, E., Cortese, J. F., Barnes, K. G., Rosen, D. M., Lukens, A. K., Daniels, R. F., Milner, D. A., Jr., Johnson, C. A., Shylakhter, I., Grossman, S. R., Becker, J. S., Yamins, D., Karlsson, E. K., Ndiaye, D., Sarr, O., Mboup, S., Happi, C., Furlotte, N. A., Eskin, E., Kang, H. M., Hartl, D. L., Birren, B. W., Wiegand, R. C., Lander, E. S., Wirth, D. F., Volkman, S. K., and Sabeti, P. C. (2011) PLoS Genet 7, e1001383

17. Ochola, L. I., Tetteh, K. K., Stewart, L. B., Ritho, V., Marsh, K., and Conway, D. J. (2010) Mol Biol Evol 27, 2344-2351

18. Tolia, N. H., Enemark, E. J., Sim, B. K., and Joshua-Tor, L. (2005) Cell 122, 183-193

19. Crabb, B. S., Cooke, B. M., Reeder, J. C., Waller, R. F., Caruana, S. R., Davern, K. M., Wickham, M. E., Brown, G. V., Coppell, R. L., and Cowman, A. F. (1997) Cell 89, 287-296

20. Wessel, D., and Flugge, U. I. (1984) Anal. Biochem. 138, 141-143

21. Tham, W. H., Wilson, D. W., Lopaticki, S., Schmidt, C. Q., Tetteh-Quarcoo, P. B., Barlow, P. N., Richard, D., Corbin, J. E., Beeson, J. G., and Cowman, A. F. (2010) Proc Natl Acad Sci U S A 107, 17327-17332

22. Tham, W. H., Schmidt, C. Q., Hauhart, R. E., Guariento, M., Tetteh-Quarcoo, P. B., Lopaticki, S., Atkinson, J. P., Barlow, P. N., and Cowman, A. F. (2011) Blood

23. Otwinowski, Z., and Minor, W. (1997) Methods in Enzymology 276, 307-326

24. McCoy, A. J., Grosske-Kunstele, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) J. Appl. Cryst. 40, 658-674

25. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) Acta Crystallogr. D. Biol. Crystallogr. 64, 125-132

26. Collaborative Computational Project, N. (1994) Acta Cryst. D. Biol. Crystallogr. 50, 760-763

27. Emsley, P., and Cowtan, K. (2004) Acta Cryst. D. Biol. Crystallogr. 60, 2126-2132

28. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosske-Kunstele, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) Acta Cryst. D. Biol. Crystallogr. 66, 213-221

29. Hodder, A. N., Maior, A. G., Rug, M., Brown, M., Hommel, M., Pantic, I., Puig-de-Morales-Marinkovic, M., Smith, B., Triglia, T., Beeson, J., and Cowman, A. F. (2009) Mol Microbiol 71, 48-65

30. Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B., and Tainer, J. A. (1999) J. Mol. Biol. 288, 427-439

31. Fiser, A., and Sali, A. (2003) Methods Enzymol. 374, 461-491

32. Bashford, D. (1997) Lecture Notes in Computer Science 1343, 233-240

33. Sitkoff, D. S., KA, Honig, B. (1994) J. Phys. Chem. 98, 1978-1988

34. Sanner M, O. A., Spehner JC. (1996) Biopolymers 38, 305-320

35. Silmon de Monerri, N. C., Flynn, H. R., Campos, M. G., Hackett, F., Koussis, K., Withers-Martinez, C., Skehel, J. M., and Blackman, M. J. (2011) Infect Immun 79, 1086-1097

36. Camus, D., and Hadley, T. J. (1985) Science 230, 553-556

37. Richard, D., MacRaid, C. A., Riglar, D. T., Chan, J. A., Foley, M., Baum, J., Ralph, S. A., Norton, R. S., and Cowman, A. F. (2010) J Biol Chem 285, 14815-14822

38. Holm, L., and Rosenstrom, P. (2010) Nucleic Acids. Res. 38, W545-549

39. Auld, D. S. (2009) Biometals 22, 141-148

40. Kauth, C. W., Epp, C., Bujard, H., and Lutz, R. (2003) J Biol Chem 278, 22257-22264

41. Kauth, C. W., Woehlbi, U., Kern, M., Mekonnen, Z., Lutz, R., Mucke, N., Langowski, J., and Bujard, H. (2006) J Biol Chem 281, 31517-31527
42. Auld, D. S. (2001) *Biometals* 14, 271-313
43. Zelensky, A. N., and Gready, J. E. (2005) *FEBS J* 272, 6179-6217
44. Gazarini, M. L., Thomas, A. P., Pozzan, T., and Garcia, C. R. (2003) *J. Cell Biol.* 161, 103-110
45. McColl, D. J., Silva, A., Foley, M., Kun, J. F. J., Favaloro, J. M., Thompson, J. K., Marshall, V. M., Coppol, R. L., Kemp, D. J., and Anders, R. F. (1994) *Mol. Biochem. Parasitol.* 68, 53-67
46. Burgess, B. R., Schuck, P., and Garboczi, D. N. (2005) *J. Biol. Chem.* 280, 37236-37245
47. Tham, W. H., Healer, J., and Cowman, A. F. (2012) *Trends Parasitol.* 28, 23-30
48. Juillerat, A., Lewit-Bentley, A., Guillote, M., Gangnard, S., Hessel, A., Baron, B., Vigan-Womas, I., England, P., Mercereau-Puijalon, O., and Bentley, G. A. (2011) *Proc Natl Acad Sci U S A* 108, 5243-5248
49. Khunrae, P., Philip, J. M., Bull, D. R., and Higgins, M. K. (2009) *J Mol Biol* 393, 202-213
50. Hodder, A. N., Crewther, P. E., Matthew, M. L., Reid, G. E., Moritz, R. L., Simpson, R. J., and Anders, R. F. (1996) *Journal of Biological Chemistry* 271, 29446-29452
51. Reed, M. B., Caruana, S. R., Batchelor, A. H., Thompson, J. K., Crabb, B. S., and Cowman, A. F. (2000) *Proc. Natl. Acad. Sci. USA* 97, 7509-7514

**FOOTNOTES**

We thank the Red Cross Blood Service (Melbourne, Australia) for supply of red blood cells and serum. Crystallization trials were performed at the Bio21 Collaborative Crystallization Centre, Victoria, Australia. Diffraction data were collected at the Australian Synchrotron and we thank staff from the macromolecular beam lines for their assistance. This work was made possible through Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRISS.

**FIGURE LEGENDS**

Fig. 1. The structure of *Pf*MSPDBL1, 2, MSP3 and EBL proteins. Schematic showing the important structural characteristics for *Pf*MSPDBL1 and 2 in relation to the MSP3 and EBL protein families.

Fig. 2. Expression, red cell binding and surface localisation of parasite-derived *Pf*MSPDBL1 and 2. (A) HA tagging of endogenous *Pf*MSPDBL1 and *Pf*MSPDBL2 proteins. Triple HA tags (3xHA) were attached to the 3’ end of the *Pf*MSPDBL1 and *Pf*MSPDBL2 genes as described in the experimental procedures section. Full-length forms of the HA-tagged proteins were detected on immunoblots using a mouse anti-HA antibody. (B) Parasite derived HA-tagged *Pf*MSPDBL1 and *Pf*MSPDBL2 proteins bind to red blood cells in an enzyme-independent manner. Post-invasion culture supernatants from parasites expressing endogenously HA-tagged *Pf*MSPDBL proteins or from wild type (WT) parasites, were incubated with untreated red blood cells (-) or cells treated with either trypsin (T) or neuraminidase (N) as described in the experimental procedures section. Bound proteins were eluted and detected on immunoblots using mouse anti-HA antibodies. Rabbit anti-EBA175 antibodies were used to confirm that the enzyme treatments had effectively removed selective red blood cell surface receptors. (C) *Pf*MSPDBL1 and 2 localise to the merozoite surface. Immunofluorescence assays of late stage schizonts and merozoites prepared from *P. falciparum* (3D7 line) parasites probed with either rabbit or mouse polyclonal antibodies made to *Pf*MSPDBL2 and marker proteins MSP119 (Rows 1 and 2), *Pf*MSPDBL1 (Rows 3 and 4) or RON4 (Row 5). Column 1=phase view, column 2=DAPI stained parasites, column 3=localisation of *Pf*MSPDBL2 (green), column 4=localisation of marker protein (red), column 5=the merge of columns 3 and 4 and column 6=the merge of columns 2, 3 and 4.
Fig. 3. Expression and purification of the DBL domains from *P*/MSPDBL2 and *P*/MSPDBL1. Panels (A) and (B) show the elution profiles for *P*/MSPDBL1 and *P*/MSPDBL2, respectively, from NiNTA resin under denaturing conditions. Panels (C) and (D) show SDS-PAGE analyses for the refolded DBL domains of *P*/MSPDBL1 and *P*/MSPDBL2, respectively, after ion-exchange purification. Samples were electrophoresed in sample buffer with (RD) or without (NR) reducing agent. Panels (E) and (F) show immunoblots for *P*/MSPDBL1 and *P*/MSPDBL2 DBL domains respectively, probed with IgG obtained from pooled hyperimmune sera from malaria endemic regions of PNG. Proteins were electrophoresed with (RD) or without (NR) reducing agent in the sample buffer prior to transfer onto PVDF membrane. Panels (G) and (H) show RP-HPLC profiles for the denatured (SM) and refolded (50) DBL domains for *P*/MSPDBL1 and *P*/MSPDBL2. See experimental procedures for additional information.

Fig. 4. Red blood cell binding of recombinant DBL domains from *P*/MSPDBL1 and 2. (A) Recombinant DBL domains from *P*/MSPDBL1 and 2 proteins (rDBL1 and rDBL2, respectively) bind to red blood cells in an enzyme-independent manner with binding enhanced in the presence of Ca^{2+}. Recombinant DBL domains were incubated with untreated red blood cells (+) or cells treated with either trypsin (T) or neuraminidase (N) in the presence or absence of 5 mM Ca^{2+}, as described in the experimental procedures section. After passage through oil the bound proteins were eluted with high salt, subjected to western transfer then detected on immunoblots with either rabbit anti-*P*/MSPDBL1 or a mouse monoclonal antibody directed against the HexaHis tag of the recombinant *P*/MSPDBL2 DBL domain. For rDBL1 (top blot), lower bands were additionally shown (arrow) in the immunoblot as an indication of sample loading. These bands were of red blood cell protein origin and found to react only with the secondary sheep anti rabbit IgG HRP conjugated antibody and not the primary anti-*P*/MSPDBL1 antibody. Sheep anti mouse IgG HRP antibodies used to assist in visualising the rDBL2 (middle blot) did not cross-react with red blood cell proteins. EBA-175, obtained from parasite culture supernatants, was used to demonstrate effective trypsin and neuraminidase treatment of red blood cells. The EBA-175 binding phenotype is sensitive to treatment with either of these enzymes. An antibody targeting the C-terminal cysteine rich region (CR) or region 6 of EBA175 was used for detection (51) (lower blot). (B). The binding of the recombinant DBL domains of MSPDBL1 and 2 to erythrocytes is enhanced in the presence of specific metal ions. Red blood cells were incubated with recombinant DBL domains in the presence of a selection of di- and mono-valent cations (5 mM final concentration) as described in the experimental procedures section. Only metal ions that enhanced the binding of rDBL1 or rDBL2 to erythrocytes resulted in detectable levels of these proteins on immunoblots. Assays were performed using untreated erythrocytes and carried out as described in (A) and the experimental procedures section. Lanes labelled rDBL1 or rDBL2 show recombinant protein not used in the binding assay. Other lanes show outcome of binding assays performed with different metal ions in conjunction with rDBL1 (upper blot) or rDBL2 (lower blot). (C) The binding of recombinant DBL domains of *P*/MSPDBL1 and 2 to red blood cells is dependent upon their disulfide bond stabilised conformation. Red blood cells were incubated with refolded rDBL domain (rDBL1 or rDBL2) or reduced and alkylated rDBL domains (rDBL1-R/A or rDBL2-R/A) in the presence of 5 mM Ca^{2+} as described in (A) and the experimental procedures section. Lanes 1 and 2 of each immunoblot contain recombinant rDBL1 or rDBL2 and rDBL1-R/A or rDBL2-R/A not used in binding assays as an identification control. Lanes 4 and 5 show the recovery of the same proteins when used in binding assays. Note that the rDBL1-R/A or rDBL2-R/A proteins could not be recovered from these assays. (D) The effect of Ca^{2+} on the binding of parasite-derived, full length *P*/MSPDBL2 to red blood cells can be reversed in the presence of EGTA. HA-tagged MSPDBL2 (i.e. full-length MSPDBL2-HA) derived from parasite cultures (see Fig. 2(A)) was used in these binding assays. Lane 1= parasite culture supernatant only, not used in the binding assay, Lane 2= no Ca^{2+} added to the culture supernatant used in the binding assay, Lane 3= 1 mM Ca^{2+} added to the culture supernatant and binding assay, Lane 4= 10 mM EGTA in culture supernatant and binding assay, Lane 5= 1 mM Ca^{2+} + 1mM EGTA in culture supernatant and binding assay, Lane 6= 1 mM Ca^{2+} + 2 mM EGTA.
in culture supernatant and binding assay, Lane 7 = 1 mM Ca\(^{2+}\) + 5 mM EGTA in culture supernatant and binding assay, Lane 8 = 1 mM Ca\(^{2+}\) + 10 mM EGTA in culture supernatant and binding assay and Lane 9 = a wild type parasite culture supernatant control, in which MSPDBL2 does not have the C-terminal HA tag + 1 mM Ca\(^{2+}\) used in the binding assay. EBA-175, obtained from the same parasite culture supernatants as a source of PfMSPDBL2, was used as a binding control under identical assay conditions. The rabbit polyclonal antibody used to detect EBA-175 targeted regions 3-5 of the molecule (R35) (51). (E) Histograms and dot blots obtained from FACS analyses of the binding of rDBL1 to untreated and enzyme-treated erythrocytes. Enzyme treatments and assay conditions are discussed further in the experimental procedures section. Error bars on histograms indicate standard error of the mean (SEM) obtained from three independent assays. Recombinant PfRh4.9 was used as a binding control for enzyme treated erythrocytes. This protein can bind to neuraminidase-treated erythrocytes but not erythrocytes treated with either trypsin or chymotrypsin. Representative dot blots are shown for each of the rDBL1 assay conditions. Numbers outside of the purple boxes refer to the percentage of erythrocytes with bound rDBL1 relative to the erythrocyte population. The various enzyme-treated erythrocytes used in the assays are indicated by the following symbols: (-) = no treatment, (T) = trypsin treatment, (N) = neuraminidase treatment and (C) = chymotrypsin treatment. (F) Histograms and dot blots obtained from FACS analyses of the binding of refolded and reduced and alkylated forms of rDBL1 to erythrocytes. Binding was performed on untreated erythrocytes in the presence of 1 mM Ca\(^{2+}\). See part (E) and the experimental procedures section for further details. A student’s t test was performed on the mean values obtained for the % binding of the refolded (rDBL1) and reduced and alkylated (rDBL-R/A) forms of rDBL1. A p<0.001 is considered significant. (G) Histograms and dot blots obtained from FACS analyses of the binding of rDBL1 in the presence of increasing concentrations of EGTA. Binding was performed using untreated erythrocytes with 1 mM Ca\(^{2+}\) present in the binding buffer. The EGTA concentration within the binding assays ranged from 0 to 60 mM as indicated. Recombinant PfRh4.9 was used as a binding control over the EGTA concentration range used in the binding assays. See part (A) and the experimental procedures section for further details.

**Fig. 5.** The structure of the DBL domain from PfMSPDBL2. (A) Ribbon diagram displaying the structure of the PfMSPDBL2 DBL domain. Disulphide bridges are highlighted in yellow. A 180° rotation of the structure is shown and helices are labelled as depicted in Fig. 6. Regions corresponding to subdomains 1, 2 and 3 are coloured magenta, blue and green respectively. Only electron density corresponding to Cys177 was observed in the region between residues Ser172 to Asn185 and no electron density was observed for the loop (Gln375-Val387) between helices 5 and 6. These regions were modelled using the LoopModel utility, part of the MODELLER package (31) and coloured in black (B) Overlay of PfMSPDBL2-DBL (Blue) with the DBL 6e domain of PfEMP1 VAR2CSA (red). (C) Schematic showing the disulfide linkages (yellow) between subdomains 1 (magenta) and 2 (blue) in the DBL domain of PfMSPDBL2. The depicted view is obtained by a 90° anticlockwise rotation of the first schematic in Fig. 5A.

**Fig. 6.** Structure based sequence alignment for various DBL domains found in *Plasmodium spp.* Residues participating in helices are coloured red, while those in beta strands are coloured blue. Two cysteine residues not engaged in disulfide bond formation are underlined. Residues that were either engineered mutations or that are anomalous are presented in italics. Cysteine residues engaged in disulfide bonds and strictly conserved residues are highlighted. The location of the canonical helices is shown. Residues in lower-case for PfMSPDBL2, are not observed in the X-ray structure. Residues presented in bold font and highlighted have been implicated in the binding of substrate. Magenta, blue and green underline represents regions defined as subdomains 1, 2 and 3, respectively. Protein identification is based on the PDB identifier code with 1ZRO=PfEBA175 DBL, 3RRC=Pv DBP DBL, 2C6J=Pk DBP DBL, 2XUO=PfEMP1-NTS-DBL1a1–VarO, 3BQI/3CML=PfEMP1 VAR2 CSA DBL 3X and 2WAU=PfEMP1 VAR CSA DBL 6e.
Fig. 7. A comparison of the disulfide bond architectures found in various DBL domains of *Plasmodium* spp. Numbers 1 to 14 represents the relative position of cysteine residues conserved in the structure of the EBA-175 F2 DBL domain. Additional numbers and letters represent a shift in the position of cysteine residues found in other DBL domains relative to those in the EBA-175 F2 DBL domain. A red cross through numbers 1 to 14 represents the loss of a cysteine from a position found in EBA-175 F2 DBL domain. The identity of individual Cys residues involved in disulfide bonds within the DBL domain of *PfMSPDBL2* is shown. Magenta, blue and green lines indicate the location of Cys residues in subdomain 1, 2 or 3 respectively. PDB identifier codes for each structure are shown in brackets where known. The disulfide architecture for *Pf332* is shown as described (29).

Fig. 8. Structure of *PfMSPDBL2* and the zinc binding sites. The four zinc binding sites are shown S1 - H249, Y356; S2 - E254, H257; S3 - H316, E319; S4 - D278, with the zinc cation represented as a green sphere.

Fig. 9. Modelled structures of the *PfMSPDBL1* and 2 DBL domains. (A) Ribbon schematic of the modelled *PfMSPDBL1* DBL domain. Disulfide bridges are displayed in yellow. (B) Electrostatic surface potential diagrams, with 180° rotation, for the DBL domain of *PfMSPDBL1*. (C) Ribbon schematic of the modelled DBL domain for *PfMSPDBL2*. Disulfide bridges are displayed in yellow. (D) Electrostatic surface potential diagrams, with 180° rotation, for the DBL domain of *PfMSPDBL2*. Electrostatic surface diagrams in B and D are based on the modelled structures for the DBL domains of *PfMSPDBL1* and 2, including loops missing in the X-ray structure of *PfMSPDBL2*. Arrows indicate the location of the cleft that occurs between subdomains 2 and 3 in these DBL domains.

Fig. 10. Naturally occurring polymorphisms of the *PfMSPDBL2* and *PfMSPDBL1* DBL domains. The structures are depicted in surface format with polymorphisms coloured red. The orientation in (A) is as that displayed in Figure 5A. The orientation in (B) is rotated 180° to that displayed in (A). The orientation in (C) is as that displayed in Figure 9A. The orientation in (D) is rotated 180° to that displayed in (C). Arrows shown in A and C indicate the location of the cleft between subdomains 2 and 3 in the DBL domains of *PfMSPDBL2* and 1, respectively. Colouring of the electrostatic potential is from red (-15 kJ mol⁻¹), through white (0 kJ mol⁻¹) to blue (+15 kJ mol⁻¹).
### Table 1. Crystallographic statistics.

| **Data Collection** | **Native** | **Zn\(^{2+}\)** |
|---------------------|------------|------------------|
| **Space Group**     | P2\(_1\)   | P2\(_1\)         |
| **Cell Dimensions** | a = 36.6 Å, b = 101.5 Å, c = 38.4 Å, \(\beta = 101.6^\circ\) | a = 36.7 Å, b = 102.1 Å, c = 38.2 Å, \(\beta = 101.8^\circ\) |
| **Wavelength (Å)**  | 1.0163     | 1.2718           |
| **Resolution Range (Å)** | 50.0 - 2.09 (2.16 - 1.09)* | 50.0 - 2.11 (2.24 - 2.11) |
| **R\(_{merge}\)**   | 0.074 (0.489) | 0.085 (0.683)    |
| **I/\(\sigma_I\)**  | 19.23 (2.4)  | 12.33 (2.15)     |
| **Completeness (%)** | 99.7 (97.5)  | 99.1 (95.7)      |
| **Redundancy**      | 4.6 (4.0)   | 3.7 (3.6)        |

**Refinement**

| **Resolution**       | 35.9 – 2.09 (2.25 – 2.09) | 37.4 – 2.11 (2.25 – 2.11) |
| **No. reflections R\(_{work}\)** | 15167 (2552) | 14918 (2434) |
| **No. reflections R\(_{free}\)**  | 765 (133) | 791 (125)    |
| **R\(_{work}\)**     | 0.18 (0.24)  | 0.19 (0.27)   |
| **R\(_{free}\)**     | 0.24 (0.31)  | 0.24 (0.35)   |
| **No. atoms**        | Protein 2233 | 2287             |
|                      | Water/ions 55 | 74               |

**R.M.S. deviations**

| **Bond lengths (Å)** | 0.008 | 0.018 |
| **Bond angles (°)**  | 1.0   | 1.0   |

**Ramachandran**

| **Favoured (%)**   | 96.2 | 97.5 |
| **Allowed (%)**    | 3.8  | 2.5  |
| **Rotamer outliers (%)** | 0.0  | 2.0  |

*Numbers in brackets represent statistics for highest resolution shell.
Table 2: Location of polymorphisms within the DBL domains of (A) PfMSPDBL2 and (B) PfMSPDBL1.

(A)

| Residue | Strain   |
|---------|----------|
|         | 164 165 166 167 169 170 171 172 174 176 180 188 189 191 193 194 196 197 198 200 200 201 200 210 210 210 213 213 216 216 217 217 218 218 219 219 221 222 223 226 227 229 232 235 237 239 |
|         | J07      |
|         | Ghana1   |
|         | HB3      |
|         | D10      |
|         | IT       |
|         | ROA33    |
| D6      | A K      |
| K1      | A K      |
| V1-S    | D K      |
| D02     | A K      |

(B)

| Residue | Strain   |
|---------|----------|
|         | 242 244 246 247 248 250 253 258 260 261 263 271 272 274 277 283 284 286 289 291 294 295 296 297 298 300 301 302 308 311 312 314 315 316 319 327 |
|         | J07      |
|         | Ghana1   |
|         | HB3      |
|         | D10      |
|         | IT       |
|         | ROA33    |
| D6      | A K      |
| K1      | V1-S     |
| D02     | V1-S     |

| Residue | Strain   |
|---------|----------|
|         | 328 329 333 336 341 355 363 366 377 281 383 385 399 409 420 426 426 435 439 448 452 457 |
|         | J07      |
|         | Ghana1   |
|         | HB3      |
|         | D10      |
|         | IT       |
|         | ROA33    |
| D6      | A K      |
| K1      | V1-S     |
| D02     | V1-S     |

Note: The location of amino acid polymorphisms within the DBL domains of PfMSPDBL1 and 2 in various strains of P. falciparum were obtained from PlasmoDB.org. Additional polymorphisms for the DBL domain of PfMSPDBL1 were obtained from [14].
Figure 1

MSPDBL1

MSPDBL2

MSP3

MSP6

PkαDBP

PfEBA140

PfEBA175

Signal sequence

NLR[K/N][A/G/N]

motif

SPAM domain

DBL domain

Transmembrane domain

MSP3-like family

EBP family
Structure and function of DBL domains

Figure 3
Figure 4

A: Structure and function of DBL domains

B: Overview of DBL domains

C: RBC binding assays

D: MSPDBL2-HA assay

E: Binding kinetics analysis

F: Effect of Ca^2+ on binding

G: EGTA concentration impact
Structure and function of DBL domains

Figure 5
Figure 7

| Subdomain 1 | Subdomain 2 | Subdomain 3 |
|-------------|-------------|-------------|
| **PyBEA 175 F2 DBL (1ZRO)** | 1 2 3 4 | 5 6 | 7 8 9 10 11 12 13 14 |
| **PyBEA 175 F1 DBL (1ZRO)** | 1 2 3 4 | 5 6 | 8 9 10 11 12 13 14 |
| **Pv DBP DBL (1RRC)** | 1 2 3 4 | 5 6 | 8 9 10 11 12 13 14 |
| **Pv DBP DBL (2C6I)** | 1 2 3 4 | 5 6 | 8 9 10 11 13 14 |
| **Pv332 DBL** | 2 3 | 5a 5 6a 6 | 7 8 10 11 12 13 14 |
| **MSPDBL2 DBL** | 1b 1a 1 | 4 | 6a 6 | 7 8 10 12b 12a 12 15 16 |
| **PcEMPI NTS DBL3_A-VarO (2XU0)** | 1a 1 2 3 4 | 5 6a 6 | 7 8 9a 9 10 11 12 13 14 15 16 |
| **PcEMPI VAR2 CSA DBL3X (3BQG3CM)** | 1a 12 3 4 | 5 6a 6 | 7 8 9a 10 11 12 13 14 |
| **PcEMPI VAR2 CSA DBL6c (2WAU)** | 5b | | 7 8 9a 10 11 12 13 14 |
Figure 8
Figure 9
Figure 10
Insights into duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from P. falciparum
Anthony N. Hodder, Peter E. Czabotar, Alessandro D. Uboldi, Oliver B. Clarke, Clara S. Lin, Julie Healer, Brian J. Smith and Alan F. Cowman

J. Biol. Chem. published online July 26, 2012

Access the most updated version of this article at doi: 10.1074/jbc.M112.350504

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2012/07/26/M112.350504.DC1