REVIEWS

Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families

Matthew K. Higgins[1]* and Mark Carrington[2]*

[1]Department of Biochemistry, University of Oxford, OX1 3QU, Oxford
[2]Department of Biochemistry, University of Cambridge, CB2 1QW, Cambridge

Abstract: *Trypanosoma* and *Plasmodium* species are unicellular, eukaryotic pathogens that have evolved the capacity to survive and proliferate within a human host, causing sleeping sickness and malaria, respectively. They have very different survival strategies. African trypanosomes divide in blood and extracellular spaces, whereas *Plasmodium* species invade and proliferate within host cells. Interaction with host macromolecules is central to establishment and maintenance of an infection by both parasites. Proteins that mediate these interactions are under selection pressure to bind host ligands without compromising immune avoidance strategies. In both parasites, the expansion of genes encoding a small number of protein folds has established large protein families. This has permitted both diversification to form novel ligand binding sites and variation in sequence that contributes to avoidance of immune recognition. In this review we consider two such parasite surface protein families, one from each species. In each case, known structures demonstrate how extensive sequence variation around a conserved molecular architecture provides an adaptable protein scaffold that the parasites can mobilise to mediate interactions with their hosts.

Keywords: malaria; trypanosomes; surface proteins; PfEMP1; VSG; antigenic variation

Introduction

Surface proteins lie at the heart of the interactions between parasites and their hosts. As a result they are highly adapted to mediate diverse functions.

They bind to host ligands, allowing them to invade host cells or interact with host tissues. They bind to host macromolecules that can be used as nutrients. However, they also avoid interactions with molecules of the host immune system, both innate and adaptive, to avoid clearance of the infection.

Plasmodium parasites invade host cells, including hepatocytes and erythrocytes, allowing them to access an environment sheltered from the host immune system in which they can multiply. Two families of parasite proteins, the "erythrocyte-

binding like" and "reticulocyte-binding like" (EBL and RBL) proteins, are critical in the early stages of erythrocyte invasion, though interactions with human receptors.[1,2] Of these, the EBLs, including EBA-175 from *P. falciparum* and the Duffy-binding protein (DBP) from *P. vivax*, are best understood, and contain the *Plasmodium*-specific Duffy-binding like (DBL) domain as a ligand-binding module.[3,4]

In the most deadly forms of malaria, caused by *Plasmodium falciparum*, adhesive PfEMP1 proteins are present on the surface of infected erythrocytes.[5] These DBL domain containing proteins mediate attachment to tissues and endothelial surfaces, holding the parasites away from splenic clearance and allowing them to divide and develop. Important symptoms of malaria result from this adhesion and acquired immunity to severe and pregnancy-associated malaria correlates with the presence of antibodies that target these proteins.[6,7]

In contrast, African trypanosomes are predominantly extracellular parasites, living free in the host blood and tissue spaces. The absence of a requirement for host cell invasion does not remove the need for surface protein families. Indeed, to evade population clearance despite constant exposure to the host adaptive immune system, they have evolved a unique surface. This includes a layer of the variant surface glycoprotein (VSG) that coats the entire cell surface providing protection to other proteins.[8] Within this coat operate important nutrient receptors and surface proteins with a role in avoiding the toxic effects of innate immune factors.[9]

While *Plasmodium* and African trypanosome species have very different life cycles, their surface proteins have many similar requirements. Both parasites operate in the context of the adaptive immune system, and the VSGs and PfEMP1s have therefore diversified into large and complex families, allowing parasites to switch expression through antigenic variation to avoid immune detection. However, members of both families must also interact with unvarying host receptors and nutrient molecules, leading to a requirement for conserved binding faces. Here, we consider the protein folds at the heart of these two surface protein families, the VSG-fold and related three-helical bundle, and the DBL domains, reviewing what we know about these adaptable architectures and how they are used by the parasites that express them.

The trypanosome surface and the three-helical bundle fold

The surface coat of an African trypanosome is highly adapted and unique. It is packed with proteins, is under constant flux through rapid membrane recycling and is regularly remodelled by antigenic variation. The primary protein component, with some 5×10^6 dimers per cell, equivalent to ~10% of total

cellular protein, is the variant surface glycoprotein (VSG). The VSG is the key component in a population survival strategy. While many hundreds of VSG genes are present in the *Trypanosoma brucei* genome, only one is expressed in any individual cell. When the immunoglobulin titre against this VSG is sufficient, the population expressing it will be killed. The parasite population then survives through a low frequency stochastic switch to expression of a different VSG. If the VSG is novel in the host, clonal expansion allows the new population to expand, until it in turn is recognized. Iterations of switching and clonal expansion produce an infection that can last for decades.[8,10]

Structural studies, conducted some 20 years ago, demonstrated that, despite a high degree of sequence diversification, VSGs share a common fold.[11] More recent studies have shown that this fold, and the related three-helical bundle architecture, can diversify further, generating ligand-binding functions essential for trypanosome survival and human infectivity.[12]

The variant surface glycoprotein

Structural studies showed that the VSGs form elongated dimers.[11,13,14] In *T. brucei,* each monomer contains a large N-terminal domain of some 350–400 residues, and a smaller, 40–80 residue C-terminal domain, both of which can be classified based on their patterns of disulphide bonds.[15] In some other African trypanosome species, including *T. congolense* and *T. vivax,* the VSGs lack C-terminal domains.

The VSG N-terminal domains show as little as 16–20% sequence identity [Fig. 1(A)], and yet structures of two such domains revealed a remarkable conservation of molecular architecture [Fig. 1(B)].[11,13,14] Each monomer is dominated by a long α-helical hairpin, spanning nearly the full ~100 Å length of the domain. The hairpins are twisted, with kinks towards the centre of each helix. Each dimer broadens at both the "tip" and the "base" to 40–60 Å in diameter. The "tip" lies furthest from the C-terminal membrane attachment site and is formed from the end of the helical hairpin and a small three-stranded β-sheet, both decorated with a series of loop insertions and stabilized by two disulphide bonds. A third strand snakes down the hairpin towards the C-terminus, partially forming a third helix, before reaching the base. This "base" is made from a series of conserved α-helices. In *T. brucei* this is linked to a small, compact, disulphide bond stabilized C-terminal domain[16] or didomain[17] with a GPI membrane anchor.

Comparison of the two available N-terminal domain structures revealed significant architectural similarity, with 60% of the residues aligning with an rmsd of just 1.8 Å despite just 16% sequence identity [Fig. 1(B)]. Structure-based alignment of ten A type VSGs based on these two structures revealed only
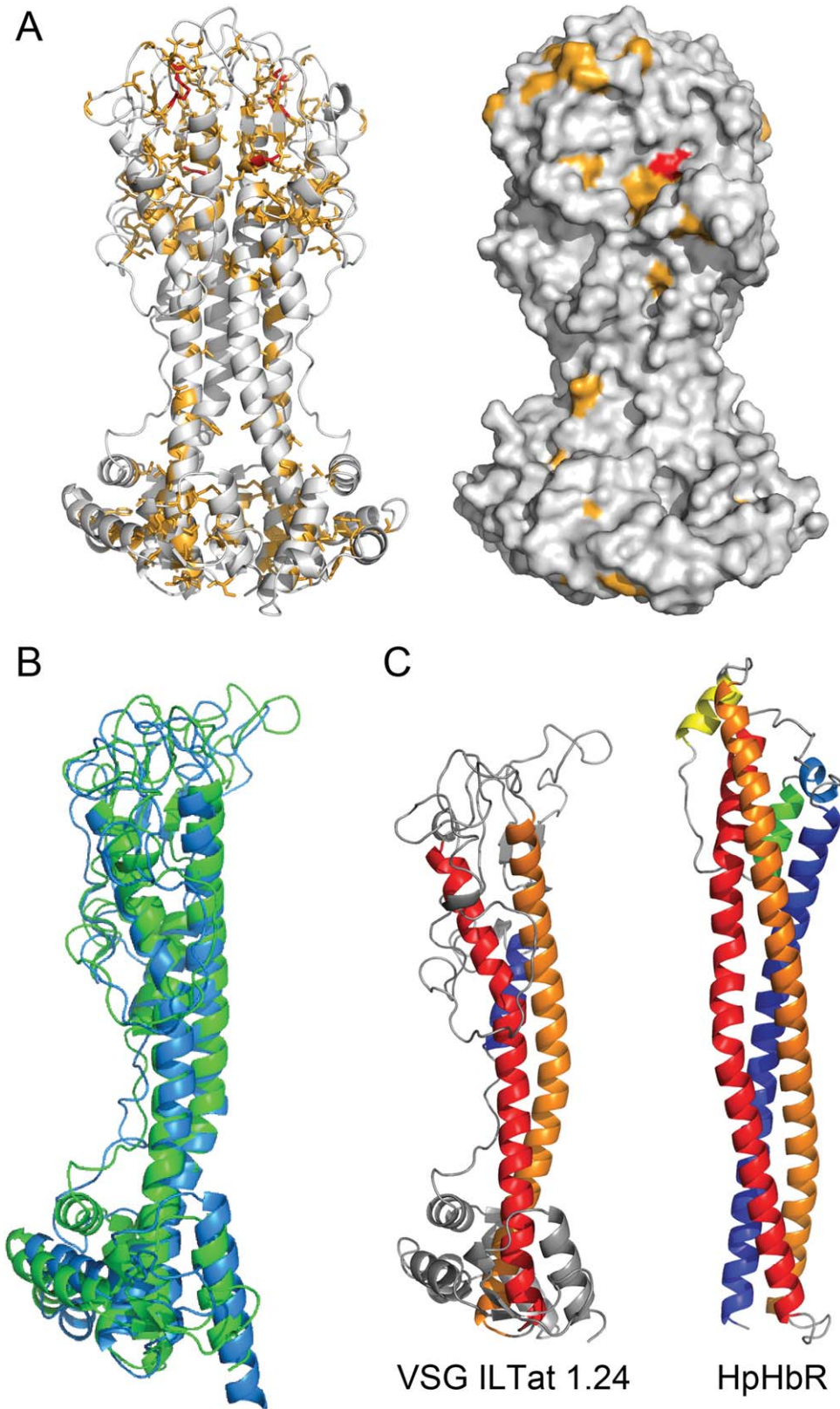
Figure 1. The trypanosome surface proteins. A: The conservation of the variant surface glycoproteins (VSGs). Residues that are conserved in the A type VSGs are plotted onto the structure of ILTat 1.24 with absolutely conserved residues in red and similar residues in yellow. Conserved residues are predominantly buried in the core of the protein, as seen in the surface representation. B: Alignment of the structures of the VSGs ILTat 1.24 (green) and MITat 1.2 (blue) showing the conserved architecture. C: A comparison of the structures of the VSG ILTat 1.24 and the *T. congolense* haptoglobin-hemoglobin receptor (HpHbR). Equivalent helices are shown in the same color, indicating the conserved basic architecture of the three helical bundles.

Parasite Surface Protein Structure and Function

five totally conserved residues (four cysteines and a glycine) and 81 residues with conserved chemical properties [Fig. 1(A)]. These conserved residues are predominantly buried and are important in determining the protein fold.[11] They include heptad repeats with hydrophobic faces that seal together the α-helices of the hairpin, glycine, or proline residues at the kinks in these α-helices and a conserved glycine that allows the two monomers to approach closely at the dimer interface. These features are also observed in the other VSG classes, demonstrating the likelihood of a conserved fold across the entire protein family.[18]

The VSGs therefore show a remarkable capacity to diverge in sequence while maintaining a conserved architecture through retention of a core set of structurally important residues. Through this, they maintain the integrity of the parasite surface, while allowing surface sequence variation to permit evasion of immune detection.

Development of ligand binding by the VSGs

While VSGs have no specific ligand, members of the wider VSG family provide the trypanosomes with important ligand-binding properties. The first identified of these was the transferrin receptor (TfR). This heterodimer is encoded by the ESAG6 and ESAG7 genes, which have evolved from an A-type VSG.[19,20] These two subunits share the essential features of the VSG fold. These include the heptad repeats, the conserved disulphides and the conserved glycine that is indicative of the dimer interface. The receptor is therefore predicted to adopt a classical dimeric VSG fold. It has been proposed that the absence of the C-terminal domains shorten the long axis of the TfR and allows the longer VSG to confer some protection.[21] However, this model is dependent on assumptions about the relative degree of extension of unstructured parts of both the VSG and TfR. Four blocks of sequence which map to the loops at the membrane distal tip of TfR are predicted to form the binding site, and mutations in these loops alter transferrin binding, suggesting that sequence variation of exposed surface loops resulted in a VSG with ligand binding properties.[22]

In addition, two distinct genes, both derived from different VSG genes, encode proteins that play a critical role in determining whether trypanosomes can survive the onslaught of trypanolytic factors (TLFs), innate immune factors present in human serum. TLFs are large lipoprotein complexes that contain the pore forming toxin, ApoLI.[23] If taken up into the lysosome, ApoLI causes lysosomal swelling and cell death.[24] Two trypanosome subspecies, *T. b. rhodesiense* and *T. b. gambiense* are able to infect humans by resisting the effects of TLFs. In *T. b. rhodesiense*, the serum resistance associated protein (SRA) allows trypanosomes to bind to, and detoxify ApoLI.[25] In *T.*

b. gambiense, the presence of TgsGP is necessary but not sufficient for protection from TLFs through a not fully substantiated mechanism.[26,27]

Both SRA and TgsGP are derived from VSGs. TgsGP has the features of a classical VSG N-terminal domain, with greatest similarity to a B-type VSG, but lacks a C-terminal domain.[28] In contrast, SRA contains a C-terminal didomain, but has a truncated N-terminal domain, marked by an internal 126-residue deletion. The features required to generate the α-helical hairpin, the heptad repeats and conserved disulphide bonds, are present, but the deletion is predicted to remove the loops found at the membrane distal tip, leading to a structure which will be as long, but narrower, than a VSG.[29]

Therefore, a variety of modifications to the VSG fold, including loss of C-terminal domains and truncations of membrane distal loops, combined with sequence variation, have generated atypical variants with altered functions. This allowed the development of a novel class of receptors that can function in the context of a VSG coat, permitting parasites to take up nutrients and to evade innate immunity.

The simplified architecture of the three-helical bundles

In addition, trypanosomes express a variety of other GPI-anchored membrane proteins that show no sequence similarity to VSGs. These include the haptoglobin-hemglobin receptor (HpHbR), which plays a dual role in the uptake of HpHb as a source of haem and in the uptake of TLF particles.[30] A recent structure of HpHbR from *T. congolense*[12] revealed an elongated, monomeric molecule, with a total length of 112 Å [Fig. 1(C)]. The receptor is built primarily from a three-helical bundle, with the helices spanning nearly its whole length. At the membrane distal side, it broadens into a compact head structure containing an additional three short helices. This is stabilised by the only disulphide bond present in the molecule and contains a patch of conserved residues essential for binding to HpHb.[12] This architecture is not unique, but is also observed in a trypanosome protein of unknown function, GARP,[31] and is likely to be shared by a family of other GPI-anchored surface molecules.

Although HpHbR is significantly simpler in structure than a VSG, it has a similar molecular architecture [Fig. 1(C)]. The N-terminal two helices of HpHbR share a path with the helical hairpin of the VSGs, including the kink in the first helix. The third helix of HpHbR also shares the path of the helical part of the third strand of the VSGs. It therefore seems highly likely that the VSGs, and members of the HpHbR-related protein family have evolved from a common ancestor built on a three-helical bundle architecture. In HpHbR, this fold has remained simple, allowing evolution of a ligand binding surface

patch, but without significant further elaboration. In the VSGs, the selection pressure to diversify for immune evasion has led to development of further structural complexity, with breaking of the third helix present on the most exposed side of the molecule into a less regular strand and the generation of complex loops at the membrane distal tip. Through these means, the simple helical bundle architecture has developed into an array of proteins that can play diverse roles in the same membrane system.

Host parasite interactions in malaria

Malaria is caused by *Plasmodium* parasites. These unicellular organisms invade host cells, including human hepatocytes and erythrocytes, in which they can divide and proliferate away from detection by the immune system. Host cell invasion is a complex process, requiring intricate molecular machinery that includes parasite surface proteins that interact with human erythrocyte surface ligands.[32] These parasite proteins come predominantly from two protein families, the EBL and RBL proteins.[1,2]

Despite the advantages of total seclusion from the immune system within a host cell, *Plasmodium falciparum*, the parasite that causes the most deadly forms of malaria, also exports proteins, including the PfEMP1 family, to the infected erythrocyte surface. PfEMP1s are surface exposed, and mediate interactions with a variety of human ligands, including ICAM-1, CD36 and EPCR.[33–35] This aids parasite survival as it tethers infected erythrocytes within the microvasculature, preventing them from being filtered from the blood by the spleen. It also leads to some of the most deadly symptoms of the disease, with infected erythrocyte accumulation in the brain resulting in inflammation during cerebral malaria and accumulation within the placenta resulting in pregnancy associated malaria. As a result, naturally acquired immunity to severe malaria correlates with immunoglobulins that bind to PfEMP1s.[6,7]

The architecture of the DBL domain

Plasmodium species have evolved a number of protein folds for molecular recognition. One of the most common, and the best understood, is the DBL domain. This ~40 kDa domain is present as a receptor recognition module in the EBL invasion proteins and is the predominant domain found in the PfEMP1s. In recent years, a number of crystal structures have been solved for DBL domains, and studies have started to show how multiple domains can be combined to produce binding proteins with diverse ligands.

Structures of ten different DBL domains are available from invasion receptors (EBA-175, EBA-140, DBP, and MSPDBL2)[3,4,36–38] and PfEMP1s.[39–41] Structure-based sequence alignment of these domains reveals a remarkably low level of sequence

identity with just 14 residues (about 4%) totally conserved in these 10 domains, while around 15% of residues are similar (Fig. 2). Despite this, DBL domain structures are built on a conserved scaffold, with a core helical architecture present in all domains. All of the conserved residues lie buried within the DBL domain fold and play structurally important roles [Fig. 3(A,B)].

The DBL domains have been described as being composed of three subdomains (SD1, SD2, and SD3)[4] [Fig. 3(C)]. The first two subdomains fold together, with SD2 containing a four helical bundle present in all DBL domains and SD1 lacking conserved secondary structure and wrapping around SD2. Three of the conserved residues, all tryptophans (W1404, W1405, and W1413 in DBL3X of var2csa), lie on the forth helix of SD2, stabilizing interactions with other helices and maintaining the fold. This core architecture is decorated with a wide variety of loops and helices in different domains. These loops are often longer and more complex in DBL domains from PfEMP1 proteins than in DBL domains from proteins used in invasion processes. Indeed, the DBL1 domain of var0 is the most decorated DBL domain structure visualised to date[41] [Fig. 3(D)]. The DBL1 domains of PfEMP1s protrude furthest from the infected erythrocyte and are therefore the most exposed to the immune system. The greater structural complexity of DBL1 domains may be driven by this greater accessibility to immunoglobulin binding and provides greater potential for sequence diversity, facilitating immune evasion and development of novel binding functions.

SD3 also has a conserved core architecture, containing a long two α-helical hairpin, together with a third partially helical strand, that snakes back along the bundle. Two conserved residues (W1457 and Y1508) stabilize the interaction between these two helices while a third (G1360) allows a tight turn between helices two and three. In addition, in nearly all cases, the distal end of SD3 is stabilised by the presence of three disulphide bonds. SD2 and SD3 are linked together through a rigid interface, in which the remaining conserved residues on both SD2 (R1268, D1353) and SD3 (Q1445, W1453, and E1456) form a series of salt bridges.

Therefore, as in the VSGs, the few conserved residues in the DBL domains are buried, stabilizing the domain structure. Around this simple conserved helical architecture the domain can diversify, both through changes to the exposed faces of helices and through decoration with loops of highly varying length and sequence. These loops emerge from the domain in all directions, ensuring different surface shapes and chemical properties. Diversification of these surfaces, built on a versatile protein fold, has allowed the domain to be used for binding functions, and to vary to allow evasion of immune detection.
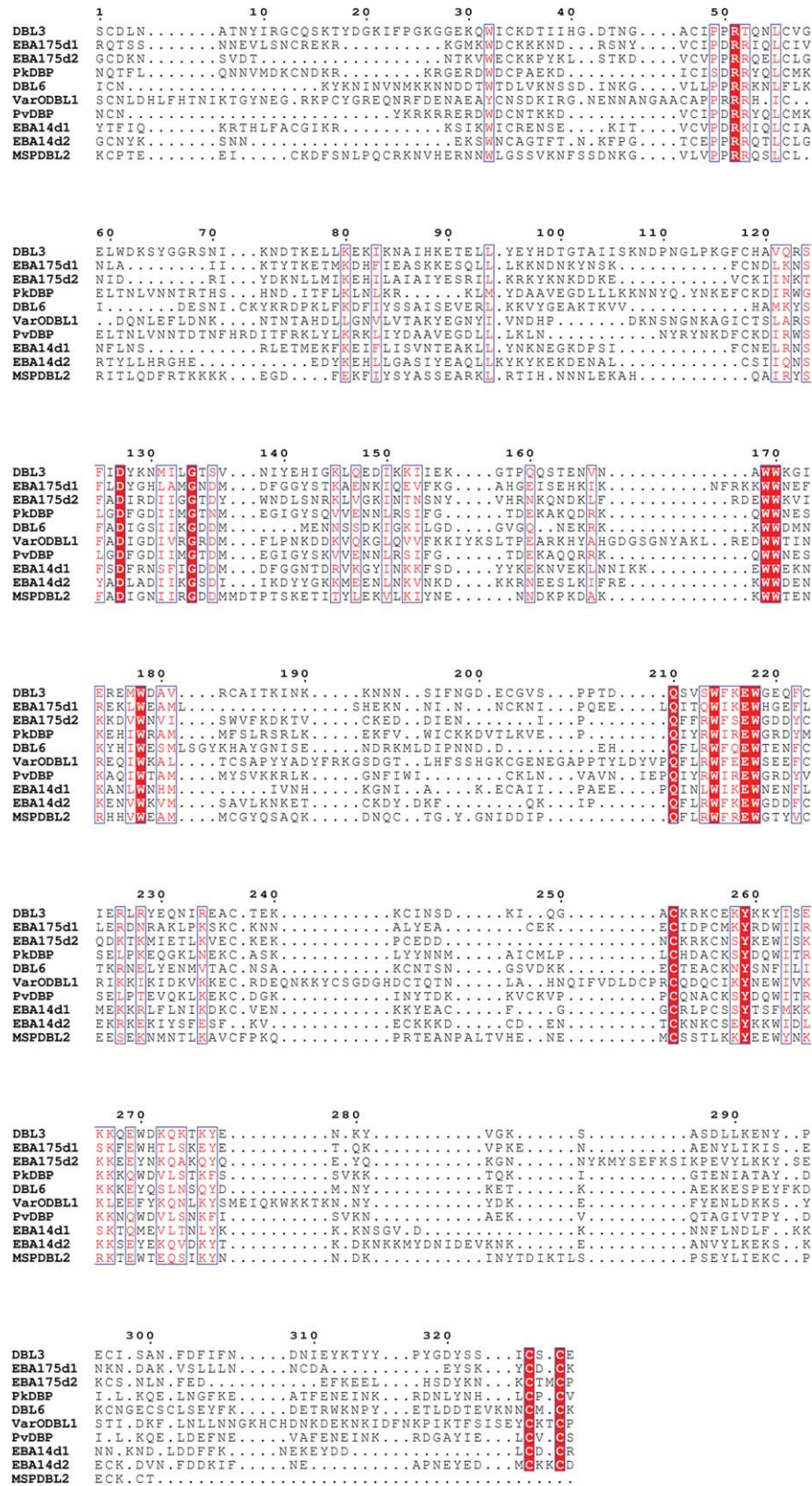
```
            1         10        20        30        40        50
DBL3        SCDLN.......ATNYIRGCQSKTYDGKIFPGKGGEKQWICKDTIIHG.DTNG....ACIFPRTQNLCVG
EBA175d1    RQTSS.......NNEVLSNCREKR.........KGMKWDCKKKND...RSNY....VCIPDRRIQLCIV
EBA175d2    GCDKN.......SVDT.............NTKVWECKKPYKL..STKD...VCVPPRRQELCLG
PkDBP       NQTFL.......QNNVMDKCNDKR.........KRGERDWDCPAEKD........ICISDRRYQLCMK
DBL6        ICN.........KYKNINVNMKKNNSSDDTWTDLVKNSSD.INKG...VLLPPRRKNLFLK
VarODBL1    SCNLDHLFHTNIKTGYNEG.RKPCYGREQNRFDENAEAYCNSDKIRG.NENNANGAACAPPRRRH.IC..
PvDBP       NCN.........YKRKRRERDWDCNTKKD...VCIPDRRYQLCMK
EBA14d1     YTFIQ.......KRTHLFACGIKR.........KSIKWICRENSE...KIT...VCVPDRRIQLCIA
EBA14d2     GCNYK.......SNN.............EKSWNCAGTFT.N.KFPG....TCEFPRRQTLCLG
MSPDBL2     KCPTE.......EI....CKDFSNLPQCRKNVHERNNWLGSSVKNFSSDNKG....VLVPPRRQSLCL.

            60        70        80        90        100       110       120
DBL3        ELWDKSYGGRSNI...KNDTKELLKEKIKNAIHKETELL.YEYHDTGTAIISKNDPNGLPKGFCHAVQRS
EBA175d1    NLA.......II...KTYTKETMKDHFIEASKKESQLL.LKKNDNKYNSK..........FCNDLKNS
EBA175d2    NID.......RI...YDKNLLMIKEHILAIAIYESRIL.KRKYKNKDDKE..........VCKIINKT
PkDBP       ELTNLVNNTRTHS...HND.ITFLKLNLKR......KLM.YDAAVEGDLLLKKNNYQ.YNKEFCKDIRWG
DBL6        I.......DESNI.CKYKRDPKLFKDFIYSSAISEVERL.KKVYGEAKTKVV..........HAMKYS
VarODBL1    ..DQNLEFLDNK....NTNTAHDLLGNYLVTAKYEGNYI.VNDHP.......DKNSNGNKAGICTSLARS
PvDBP       ELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLL.LKLN..........NYRYNKDFCKDIRWS
EBA14d1     NFLNS.........RLETMEKFKEIFLISVNTEAKLL.YNKNEGKDPSI..........FCNELRNS
EBA14d2     RTYLLHRGHE..........EDYKEHLLGASIYEAQLLKYKYKEKDENAL..........CSIIQNS
MSPDBL2     RITLQDFRTKKKK...EGD....FEKFIYSYASSEARKL.RTIH.NNNLEKAH..........QAIRYS

            130       140       150       160       170
DBL3        FIDYKNMILGTSV...NIYEHIGKLQEDIKKIIEK.....GTPQQSTENVN..........AWWKGI
EBA175d1    FLDYGHLAMGNDM...DFGGYSTKAENKIQEVFKG.....AHGEISEHKIK..........NFRKKWWNEF
EBA175d2    FADIRDIIGGTDY...WNDLSNRKLVGKINTNSNY....VHRNKQDKLF..........RDEWWKVI
PkDBP       LGDFGDIIMGTNM...EGIGYSQVVENNLRSIFG.....TDEKAKQDRK..........QWWNES
DBL6        FADIGSIIKGDDM......MENNSSDKIGKILGD...GVGQ.NEKRK..........KWWDMN
VarODBL1    FADIGDIVRGRDM...FLPNKDDKVQKGLQVVFKKIYKSLTPEARKHYAHGDGSGNYAKL.REDWWTIN
PvDBP       LGDFGDIIMGTDM...EGIGYSKVVENNLRSIFG.....TDEKAQQRRK..........QWWNES
EBA14d1     FSDFRNSFIGDDM...DFGGNTDRVKGYINKKFSD....YYKENVKLLNNIKK......EWWEKN
EBA14d2     YADLADIIKGSDI...IKDYYGKKMEENLNKVNKD....KKRNEESLKIFRE......KWWDEN
MSPDBL2     FADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNE......NNDKPKDAK..........KWWTEN

            180       190       200       210       220
DBL3        EREMWDAV....RCAITKINK.....KNNN..SIFNGD.ECGVS....PPTD.....QSVSWFKEWGEQFC
EBA175d1    REKLWEAML................SHEKN..NI.N...NCKNI..PQEE.....LQITQWIKEWHGEFL
EBA175d2    KKDVWNVI......SWVFKDKIV...CKED..DIEN.....I....P.....QFFRWFSEWGDDYC
PkDBP       KEHIWRAM....MFSLRSRLK.....EKFV..WICKKDVTLKVE...P.....QIYRWIREWGRDYM
DBL6        KYHLWESMLSGYKHAYGNISE.....NDRKMLDIPNND.D.........EH.....QFLRWFQEWTENFC
VarODBL1    REQIWKAL....TCSAPYYADYFRKGSDGT..LHFSSHGKCGENEGAPPTYLDYVPQFLRWFEEWSEEFC
PvDBP       KAQIWTAM....MYSVKKRLK.....GNFIWI......CLN..VAVN..IEPQIYRWIREWGRDYV
EBA14d1     KANLWNHM......IVNH.....KGNI.A....K.ECAII..PAEE.....PQINLWIKEWNENFL
EBA14d2     KENVWKVM....SAVLKNKET....CKDY.DKF.......QK..IP.....QFLRWFKEWGDDFC
MSPDBL2     RHHVWEAM....MCGYQSAQK.....DNQC..TG.Y.GNIDDIP.......QFLRWFREWGTYVC

            230       240       250       260
DBL3        IERLRYEQNIREAC.TEK..........KCINSD.....KI..QG......ACKRKCEKYKKYISE
EBA175d1    LERDNRAKLPKSKC.KNN..........ALYEA.....CEK..........ECIDPCMYYRDWIIR
EBA175d2    QDKTKMIETLKVEC.KEK..........PCEDD.....I...........NCKRKCNSYKEWISK
PkDBP       SELPKEQGKLNEKC.ASK..........LYYNNM...AICMLP........LCHDACKSYDQWITR
DBL6        TKRNELYENMVTAC.NSA..........KCNTSN....GSVDKK.......ECTEACKNYSNFILI
VarODBL1    RIKKIKIDKVKKEC.RDEQNKKYCSGDGHDCTQTN....LA..HNQIFVDLDCPRCQDQCIKYNEWIVK
PvDBP       SELPIEVQKLKEKC.DGK..........INYTDK....KVCKVP.......PCQNACKSYDQWITR
EBA14d1     MEKKRLFLNIKDKC.VEN..........KKYEAC....F....G.......GCRLPCSSYTSFMKK
EBA14d2     EKRKEKIYSFESF..KV...........ECKKKD....CD..EN.......TCKNKCSEYKKWIDL
MSPDBL2     EESEKNMNTLKAVCFPKQ..........PRTEANPALTVHE..NE........MCSSTLKKYEEWYNK

            270       280       290
DBL3        KKQEWDKQKTKYE...........N.KY........VGK.......S..........ASDLLKENY..P
EBA175d1    SKFEWHTLSKEYE...........T.QK........VPKE......N..........AENYLIKIS..E
EBA175d2    KKEEYNKQAKQYQ...........E.YQ........KGN....NYKMYSEFKSIKPEVYLKKY.SE
PkDBP       KKKQWDVLSTKFS...........SVKK........TQK......I..........GTENIATAY..D
DBL6        KKKEYQSLNSQYD...........M.NY........KET......K..........AEKKESPEYFKD
VarODBL1    KLEEFYKQNLKYSMEIQKWKKTKN.NY........YDK......K..........FYENLDKKS..Y
PvDBP       KKNQWDVLSNKFI...........SVKN........AEK......V..........QTAGIVTPY..D
EBA14d1     SKTQMEVLTNLYK...........K.KNSGV.D........K..........NNFLNDLF..KK
EBA14d2     KKSEYEKQVDKYT...........K.DKNKKMYDNIDEVKNK......E..........ANVYLKEKS..K
MSPDBL2     RKTEWTEQSIKYN...........N.DK........INYTDIKTLS..........PSEYLIEKC..P

            300       310       320
DBL3        ECI.SAN.FDFIFN.....DNIEYKTYY...PYGDYSS...ICS.CE
EBA175d1    NKN.DAK.VSLLLN.....NCDA..........EYSK...YCD.CK
EBA175d2    KCS.NLN.FED..........EFKEEL....HSDYKN...KCTMCP
PkDBP       I.L.KQE.LNGFKE.....ATFENEINK...RDNLYNH...LCP.CV
DBL6        KCNGECSCLSEYFK.....DETRWKNPY...ETLDDTEVKNNCM.CK
VarODBL1    STI.DKF.LNLLNNGKHCHDNKDEKNKIDFNKPIKTFSISEYCKTCP
PvDBP       I.L.KQE.LDEFNE.....VAFENEINK...RDGAYIE...LCV.CS
EBA14d1     NN.KND.LDDFFK.....NEKEYDD................LCD.CR
EBA14d2     ECK.DVN.FDDKIF.....NE..........APNEYED...MCKKCD
MSPDBL2     ECK.CT.................................
```

Figure 2. Structure-based alignment of the sequences of the 10 structurally characterized DBL domains. The sequences of the ten DBL domains with known structures were aligned using structural information, revealing very few conserved residues. DBL3 is the DBL3X domain of var2csa (pdb:3BQI); EBA175D1 and D2 are the two DBL domains from EBA-175 (pdb:1ZRL); PkDBP is from the Duffy-binding protein from *P. knowlesi* (pdb:2CGJ); DBL6 is the DBL6ε domain of var2csa (pdb:2WAU); VarODBL1 is the DBL1α domain from the var0 PfEMP1 (pdb:2YK0); PvDBP is from the Duffy-binding protein from *P. vivax* (pdb:3RRC); EBA140D1 and D2 are the DBL domains from EBA-140 (pdb:4JN0); MSPDBL2 is the DBL domain from merozoite surface protein MSPDBL2 (pdb:2VUU). Alignments were performed in fugue and visualized using Esprit.
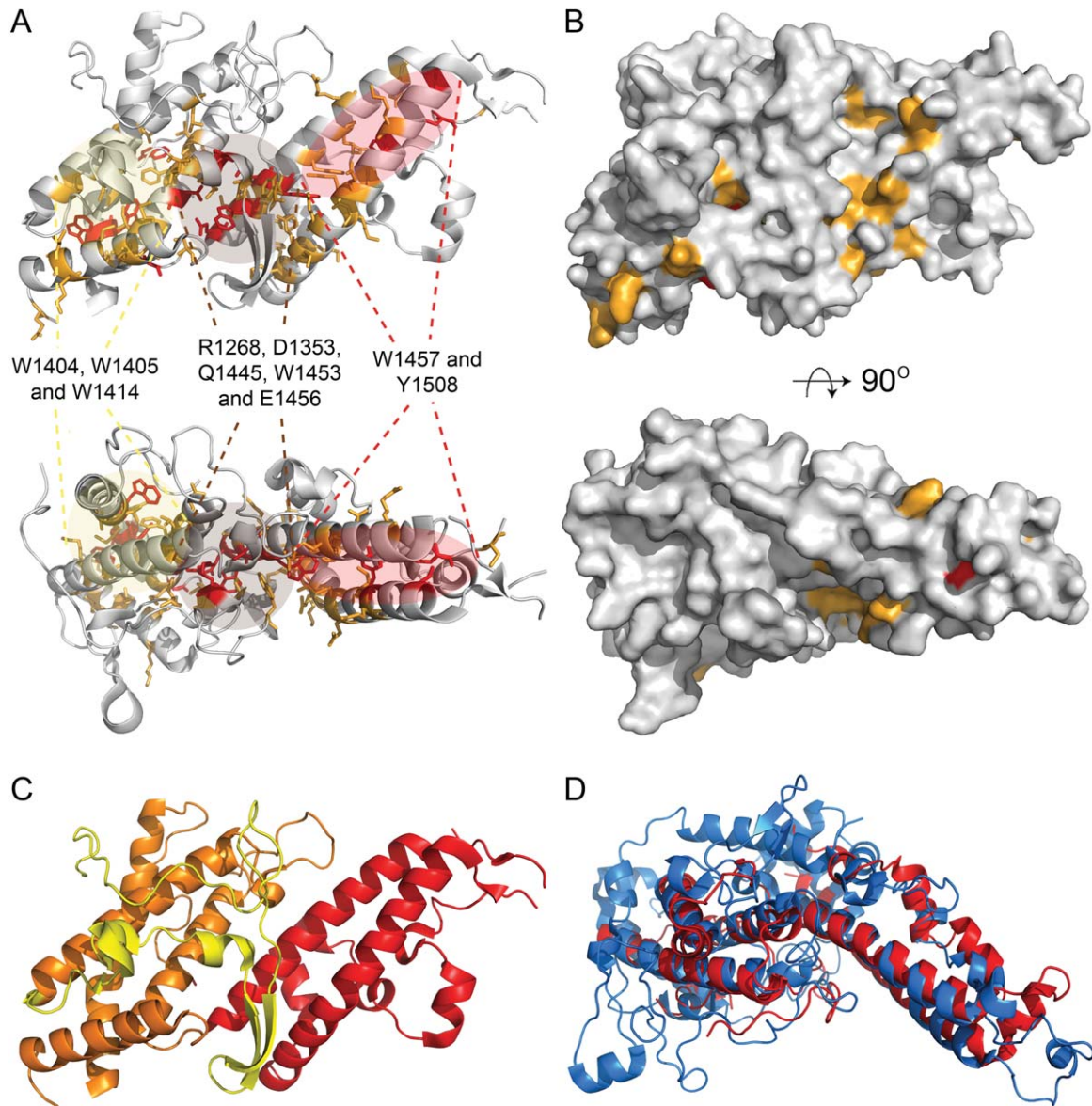
Figure 3. The structure of the DBL domain. A: Conservation of the DBL fold. Conserved residues are plotted onto the structure of the DBL3X domain of var2csa, with absolutely conserved residues in red and similar residues in yellow. The conserved residues include three main patches: one stabilizing subdomain 2 (W1404, W1405, and W1414), one stabilizing subdomain 3 (W1457 and Y1508) and one stabilizing the interface between subdomains 2 and 3 (R1268, D1353, Q1445, W1453, and E1456). B: Very few of these conserved residues are found on the surface of the protein, as illustrated by the surface representations colored as in A. C: The division of the DBL domain into three subdomains: subdomain 1 in yellow, subdomain 2 in orange, and subdomain 3 in red. D: Comparison of the DBL1 domain of EBA175 (red) with the DBL1 domain of the var0 PfEMP1 proteins (blue), showing the greater complexity of the more exposed PfEMP1 domain.

Numerous *Plasmodium* proteins are therefore built from this fold, with domains linked together in tandem to generate complex protein ectodomains.

The invasion proteins—DBL tandems and dimerization

Although the modular nature of DBL domain containing proteins suggested a mix-and-match arrangement of distinct binding modules, the DBL domains in invasion proteins appear to not work alone. Structural studies of both *P. falciparum* EBA-175 and *P. vivax*

DBP suggest that their DBL domains dimerise on ligand engagement, with ligand-binding sites formed in clefts at the dimerization interface.

EBA-175 is EBL protein that binds to the sialic acid modified erythrocyte surface protein, glycophorin A, with both the protein chain and the carbohydrate modification required for high-affinity binding. The structure reveals an ectodomain containing two DBL domains linked by a three helical bundle to form a rigid, elongated structure.[3] In the crystal, the ectodomains are arranged as an anti-parallel dimer and
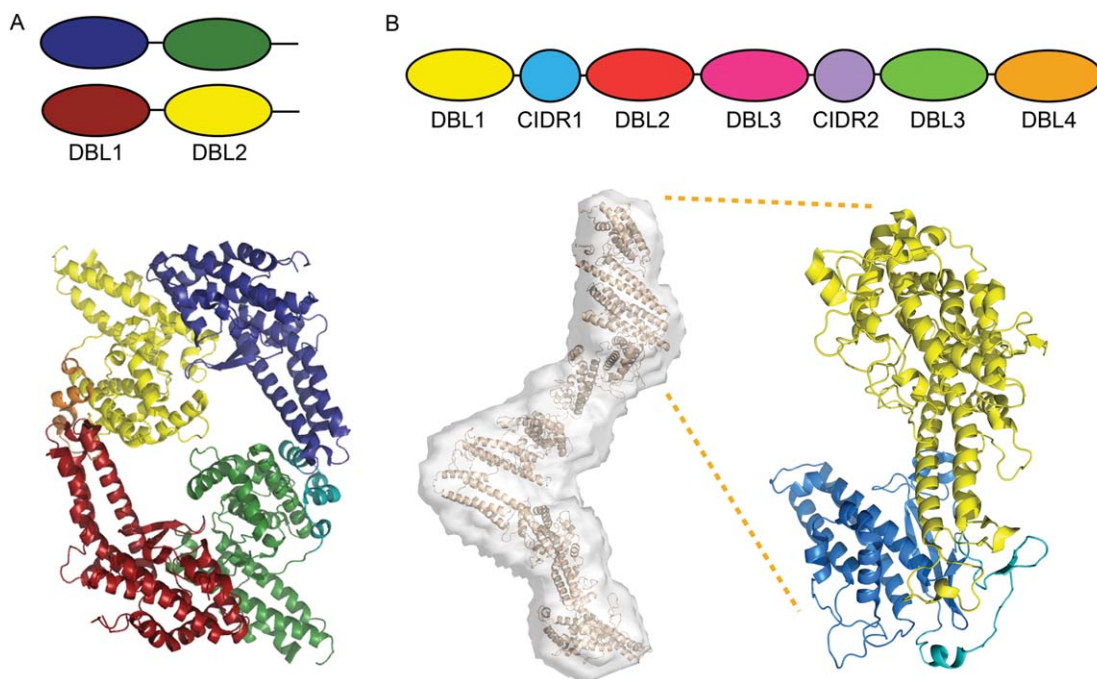
Figure 4. Higher order organization in DBL domain containing proteins. A: EBA-175 contains two DBL domains and forms a homodimer in the crystal. Sialic acid binding sites are located at the dimerization interface, suggesting the dimer to be the functional unit. B: An envelope derived from small angle x-ray scattering data for the IT4var13 PfEMP1 reveals an elongated but rigid structure. The structure of the DBL1-CIDR1 didomain of the var0 PfEMP1 protein also shows rigid organization of the two domains into a single structural unit.

solution studies at higher concentrations also show the presence of dimer [Fig. 4(A)]. Crystallization in the presence of a sialic acid derivative revealed six binding sites per dimer, all of which are found at dimer interfaces. Indeed mutation of residues involved in dimerization, and those that form any of the three sialic acid binding sites, reduced EBA-175 binding to erythrocytes.[3] In addition, a recent structure of EBA-175 in complex with a Fab fragment from an inhibitory monoclonal antibody shows the antibody to interact with residues in both the sialic acid binding and dimerization sites, holding the molecule in a monomeric conformation and preventing ligand binding.[42]

A distinct invasion receptor is involved in host cell invasion by *Plasmodium vivax*. This parasite infects reticulocytes rather than erythrocytes, through binding of a single DBL domain containing protein, DBP to the tyrosine sulphated ectodomain of the reticulocyte receptor DARC. The structure of DBP from *P. knowlesi*, revealed a single DBL domain.[4] A more recent structure, from *P. vivax* showed a very similar molecular architecture for the DBL monomer. However, this domain packed as a dimer in the crystal with the dimerization interface forming a positively charged groove that contains two sulphate-binding sites.[37] Mutagenesis, small angle x-ray scattering and analytical ultracentriguation all suggested that *P. vivax* DBP is predominantly monomeric in solution and the presence of DARC induces dimerization.[37]

Although dimerization was proposed to be "conserved in DBL-domain receptor engagement,"[37] the glycophorin C binding EBL protein, EBA-140 shows no evidence of dimerization, either in the crystal or in solution.[36] A structure of EBA-140 bound to a short carbohydrate also reveals a monomeric molecule with two sialic acid binding sites in different places from those found in EBA-175.[43] An integrated set of cellular and molecular studies are therefore needed to determine whether invasion protein dimerization is required on ligand binding *in vivo*, to see how universal it is, and to assess its functional consequences.

The multidomain PfEMP1 proteins

Around 60 genes encode PfEMP1s in each *P. falciparum* genome.[44,45] Each PfEMP1 contains a large extracellular ectodomain, linked through a single transmembrane helix to a C-terminal cytoplasmic region [Fig. 4(B)]. These ectodomains are built from individual domains of two types: the DBL and CIDR domains. A DBL-CIDR di-domain is present at the N-terminus of nearly all of the ectodomains, followed by different combinations of DBLs and CIDRs. Computational studies have suggested that the "modules" from which the PfEMP1 proteins are built are often tandem domain combinations that are maintained through evolution, rather than single domains.[45] In the seven genomes analyzed in this study, 22 such "domain cassettes" were identified,

containing different combinations of DBL and CIDR domains and recombination that generates PfEMP1 diversity often retains these domain cassette units.

The DBL domains present in the PfEMP1s have the same basic architecture as those from invasion receptors. The CIDR domains are smaller, with the two available crystal structures showing significant differences (Fig. 5). In both cases, the domain is built upon a core three α-helical bundle with similarity to subdomain 3 of a DBL domain.[41,46] However, the MC179 CIDR domain adopts an "open" V-shaped architecture with the loop between the second and third helices folding into three short α-helices that lie at approximately 90° to the main bundle.[46] In contrast, in the structure of the N-terminal DBL1-CIDR1 didomain of var0, the CIDR domain forms a compact α-helical bundle. This packs tightly against the DBL domain through an additional region containing a four-stranded anti-parallel β-sheet.[41] Whether different CIDR domains adopt different architectures, or whether truncation of the MC179 CIDR domain or the crystallization conditions have caused this structure to spread apart, remains to be seen and will require the determination of more CIDR structures, alone and in complex with their ligands.

Higher order organization in the PfEMP1s

The modular nature of the PfEMP1s naturally led to the suggestion that they operate as strings of distinct ligand binding modules. Indeed, binding properties were ascribed to a variety of single domains, with, for example, ICAM-1 shown to bind to DBLβ domains and the CD36-interaction mapped to a subset of CIDRα domains.[47,48] A number of recent studies have therefore used biophysical tools to compare the ligand-binding affinities of individual PfEMP1 domains with the binding behaviour of intact ectodomains and to assess whether individual domains do indeed mediate binding. The first such study investigated the intact var2csa ectodomain, showing it to have a greater affinity and specificity for chondoitin sulphate A (CSA) than individual component "CSA-binding" domains.[40] This led to the suggestion of higher-order architecture in this protein, with multiple domains coming together to generate specific binding site. Indeed, small angle x-ray scattering (SAXS) studies showed that the domains of var2csa fold together to adopt a compact architecture.[49,50] More recent studies reveal that CSA binding is mediated predominantly by the DBL2 domain, together with flanking sequences, and that this domain lies at the tip of this folded ectodomain structure.[49]

In contrast, affinities of the entire IT4var13 and IT4var20 PfEMP1 ectodomains for ICAM-1[51] and EPCR,[35] respectively, are extremely similar to those of individual domains (CIDRα to ECPR and DBLβ to ICAM-1), suggesting that these PfEMP1s are
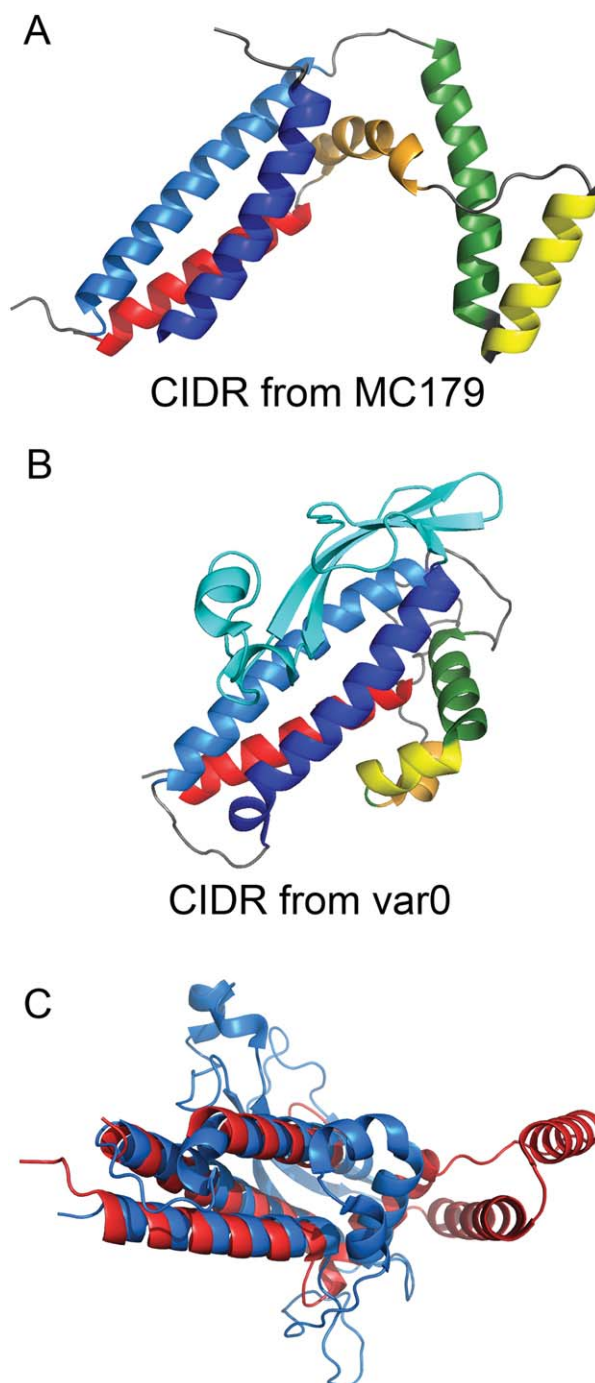

Figure 5. The structure of the CIDR domain. The structures of CIDR domains from A. MC179 and B. var0. The three main helices of the core helical bundle are colored dark blue, light blue, and red. Three additional helices form in the loop that emerges from this core helical bundle and are colored green, yellow, and beige. In var0 the N-terminal β-strands are shown in teal. C: Alignment of the two CIDR domains with MC179 in red and var0 in blue. This reveals the conserved organization of the three-helical bundle and the divergent organization of the loop helices.

modular in nature. SAXS analysis of one such ectodomain [Fig. 4(B)], IT4var13, in the presence of ICAM-1, confirmed this view, revealing a single

protrusion for bound ICAM-1 emerging from the elongated ectodomain structure.[51] Nevertheless, this ectodomain is not flexible, adopting a very similar architecture in the presence and absence of ligand.[51] This rigid architecture could be advantageous to the parasite as it will reduce the surface area of the protein exposed to immune recognition and also ensure presentation of the N-terminal ligand binding domains away from the membrane surface.

While only two ectodomains have been structurally characterized by SAXS, it seems likely that var2csa is an outlier, with the need to bind to a flexible and complex carbohydrate substrate leading to an unusual architecture. Most PfEMP1s are likely to adopt an elongated architecture, presenting individual binding sites for protein ligands on individual domains.

Finding binding sites in variable ectodomains

The PfEMP1s exist under conflicting selection pressure. They diversify to enable immune evasion, and yet maintain the ability to bind human receptors, presumably by retaining a binding surface with conserved chemical properties. However, the identification of such a surface in the extremely divergent domains has proven a significant challenge. Even after a comprehensive classification of CIDR domains into CD36 binders and non-binders, mutation of three of the residues that characterised the binders only reduced CD36 binding to 67% of wildtype.[52] For ICAM-1 binding PfEMP1s, mutagenesis has been more successful, with alteration of multiple residues located at the convex side on the domain, contributed from both subdomains 2 and 3, abolishing binding.[53] Future studies will require crystal structures of PfEMP1 domains in complex with these protein receptors, providing a more detailed understanding of the nature of these conserved binding sites and revealing the degree to which sequence conservation is required.

Conclusions

Surface proteins lie at the heart of the conflict between parasite and host. They are essential for ligand binding processes in host cell invasion, avoidance of innate immunity and nutrient uptake, requiring conservation of key binding surfaces. In contrast, their exposure to the host adaptive immune system provides a pressure towards antigenic variation to avoid immunoglobulin binding. In this review we have considered two such protein families, the three-helical bundles of the trypanosome surface and the DBL domains used by *Plasmodium*, showing some of the ways in which diversification of a basic protein fold can allow the development of different properties to facilitate parasite survival.

Structural studies of the trypanosome VSGs led to the conclusion that "the African trypanosomal antigens accomplish antigenic variation through the variation of sequence and limited conformational modification and not by gross alteration of structure."[11] More recent studies have shown further diversification of this basic fold to allow the development of ligand binding in proteins that bind nutrients or innate immune factors. Furthermore, a more simple monomeric, and probably ancestral, three-helical bundle architecture is found in receptors such as that for haptoglobin-haemoglobin.

Similarly, the DBL domains from *Plasmodium* have remarkably conserved folds, despite significant sequence diversity. These have been used in a variety of different ways to develop novel binding function. They can be linked together in extended arrays, as seen in many PfEMP1s, can act as individual binding modules as in DBP, or can fold into a compact architecture as in var2csa. They can also dimerise around ligands, as seen in some of the invasion proteins, not only creating a binding surface, but potentially providing the opportunity to contribute to cellular signaling.

Certain themes are common to both families. Both proteins are built on an α-helical architecture, with the versatility of this structural element allowing extensive sequence variation without disruption of protein fold. Both folds also show an increased pressure towards complexity for proteins that are under increased exposure to the immune system. Therefore, VSGs and PfEMP1s, proteins that are constantly exposed to the host immunoglobulins, have more sequence variation and loop complexity than proteins that are transiently exposed such as the *Plasmodium* invasion receptors and the trypanosome receptors.

Many questions remain. In *Plasmodium*, structural studies are required to demonstrate how highly divergent PfEMP1 sequences can generate binding sites that interact with the same essential human ligands. In trypanosomes, experiments are needed to show how receptors interact with their ligands, and how these binding sites are arranged in the context of the VSG layer. Answers to these questions will help us to understand the molecular details of host-parasite interactions and will guide the development of therapeutics to target the conserved protein surfaces and to tackle parasitic disease.

References

1. Iyer J, Grüner AC, Rénia L, Snounou G, Preiser PR (2007) Invasion of host cells by malaria parasites: a tale of two protein families. Mol Microbiol 65:231–249.
2. Gunalan K, Gao X, Yap SS, Huang X, Preiser PR (2013) The role of the reticulocyte-binding-like protein homologues of Plasmodium in erythrocyte sensing and invasion. Cell Microbiol 15:35–44.
3. Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell 122:183–193.

4. Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A (2006) Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. Nature 439:741–744.

5. Smith JD, Rowe JA, Higgins MK, Lavstsen T (2013) Malaria's deadly grip: cytoadhesion of Plasmodium falciparum-infected erythrocytes. Cell Microbiol 15: 1976–1983.

6. Chan JA, Howell KB, Reiling L, Ataide R, Mackintosh CL, Fowkes FJ, Petter M, Chesson JM, Langer C, Warimwe GM, Duffy MF, Rogerson SJ, Bull PC, Cowman AF, Marsh K, Beeson JG (2012) Targets of antibodies against Plasmodium falciparum-infected erythrocytes in malaria immunity. J Clin Invest 122: 3227–3238.

7. Fried M, Nosten F, Brockman A, Brabin BJ, Duffy PE (1998) Maternal antibodies block malaria. Nature 395: 851–852.

8. Schwede A, Carrington M (2010) Bloodstream form trypanosome plasma membrane proteins: antigenic variation and invariant antigens. Parasitology 137:2029–2039.

9. Pays E, Vanhollebeke B (2009) Human innate immunity against African trypanosomes. Curr Opin Immunol 21:493–498.

10. MacGregor P, Szoor B, Savill NJ, Matthews KR (2012) Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. Nat Rev Microbiol 10:431–438.

11. Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC (1993) A structural motif in the variant surface glycoproteins of Trypanosoma brucei. Nature 362:603–609.

12. Higgins MK, Tkachenko O, Brown A, Reed J, Raper J, Carrington M (2013) Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. Proc Natl Acad Sci USA 110:1905–1910.

13. Freymann D, Down J, Carrington M, Roditi I, Turner M, Wiley D (1990) 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. J Mol Biol 216:141–160.

14. Metcalf P, Blum M, Freymann D, Turner M, Wiley DC (1987) Two variant surface glycoproteins of Trypanosoma brucei of different sequence classes have similar 6 A resolution X-ray structures. Nature 325:84–86.

15. Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M (1991) Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. J Mol Biol 221:823–835.

16. Chattopadhyay A, Jones NG, Nietlispach D, Nielsen PR, Voorheis HP, Mott HR, Carrington M (2005) Structure of the C-terminal domain from Trypanosoma brucei variant surface glycoprotein MITat1.2. J Biol Chem 280:7228–7235.

17. Jones NG, Nietlispach D, Sharma R, Burke DF, Eyres I, Mues M, Mott HR, Carrington M (2008) Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein. J Biol Chem 283:3584–3593.

18. Bussler H, Linder M, Linder D, Reinwald E (1998). Determination of the disulfide bonds within a B domain variant surface glycoprotein from Trypanosoma congolense. J Biol Chem 273:32582–32586.

19. Steverding D, Overath P (1996) Trypanosoma brucei with an active metacyclic variant surface gene expression site expresses a transferrin receptor derived from esag6 and esag7. Mol Biochem Parasitol 78:285–288.

20. Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, Brown R, Browne H, Corton N, Hauser H, Gamble J, Gilderthorp R, Marcello L, McQuillan J, Otto TD, Quail MA, Sanders MJ, van Tonder A, Ginger ML, Field MC, Barry JD, Hertz-Fowler C, Berriman M (2012) Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. Proc Natl Acad Sci USA 109: 3416–3421.

21. Mehlert A, Wormald MR, Ferguson MA (2012) Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of Trypanosoma brucei. PLoS Pathog 8:e1002618.

22. Salmon D, Hanocq-Quertier J, Paturiaux-Hanocq F, Pays A, Tebabi P, Nolan DP, Michel A, Pays E (1997) Characterization of the ligand-binding site of the transferrin receptor in Trypanosoma brucei demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. EMBO J 16:7272–7278.

23. Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, Nolan DP, Lins L, Van Den Abbeele J, Pays A, Tebabi P, Van Xong H, Jacquet A, Moguilevsky N, Dieu M, Kane JP, De Baetselier P, Brasseur R, Pays E (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature 422:83–87.

24. Pérez-Morga D, Vanhollebeke B, Paturiaux-Hanocq F, Nolan DP, Lins L, Homblé F, Vanhamme L, Tebabi P, Pays A, Poelvoorde P, Jacquet A, Brasseur R, Pays E (2005) Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. Science 309:469–472.

25. Vanhollebeke B, Pays E (2010) The trypanolytic factor of human serum: many ways to enter the parasite, a single way to kill. Mol Microbiol 76:806–814.

26. Capewell P, Clucas C, Dejesus E, Kieft R, Hajduk S, Veitch N, Steketee PC, Cooper A, Weir W, Macleod A (2013) The TgsGP gene is essential for resistance to human serum in Trypanosoma brucei gambiense. PLoS Pathog 9:e1003686.

27. Uzureau P, Uzureau S, Lecordier L, Fontaine F, Tebabi P, Homblé F, Grélard A, Zhendre V, Nolan DP, Lins L, Crowet JM, Pays A, Felu C, Poelvoorde P, Vanhollebeke B, Moestrup SK, Lyngsø J, Pedersen JS, Mottram JC, Dufourc EJ, Pérez-Morga D, Pays E (2013) Mechanism of Trypanosoma brucei gambiense resistance to human serum. Nature 501:430–434.

28. Berberof M, Pérez-Morga D, Pays E (2001) A receptor-like flagellar pocket glycoprotein specific to Trypanosoma brucei gambiense. Mol Biochem Parasitol 113: 127–138.

29. Campillo N, Carrington M (2003) The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from Trypanosoma brucei rhodesiense. Mol Biochem Parasitol 127: 79–84.

30. Vanhollebeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, Raes M, Moestrup SK, Pays E (2008) A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. Science 320:677–681.

31. Loveless BC, Mason JW, Sakurai T, Inoue N, Razavi M, Pearson TW, Boulanger MJ (2011) Structural characterization and epitope mapping of the glutamic acid/alanine-rich protein from Trypanosoma congolense: defining assembly on the parasite cell surface. J Biol Chem 286:20658–20665.

32. Cowman AF, Berry D, Baum J (2012) The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol 198:961–971.

Parasite Surface Protein Structure and Function

33. Berendt AR, Simmons DL, Tansey J, Newbold CI, Marsh K (1989) Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmdium falciparum. Nature 341:57–59.

34. Oquendo P, Hundt E, Lawler J, Seed B (1989) CD36 directly mediates cytoadherence of Plasmodium falciparum parasitized erythrocytes. Cell 58:95–101.

35. Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE, Avril M, Brazier AJ, Freeth J, Jespersen JS, Nielsen MA, Magistrado P, Lusingu J, Smith JD, Higgins MK, Theander TG (2013) Severe malaria is associated with parasite binding to endothelial protein C receptor. Nature 498:502–505.

36. Lin DH, Malpede BM, Batchelor JD, Tolia NH (2012) Crystal and solution structures of Plasmodium falciparum erythrocyte-binding antigen 140 reveal determinants of receptor specificity during erythrocyte invasion. J Biol Chem 287:36830–36836.

37. Batchelor JD, Zahm JA, Tolia NH (2011) Dimerization of Plasmodium vivax DBP is induced upon receptor binding and drives recognition of DARC. Nat Struct Mol Biol 18:908–914.

38. Hodder AN, Czabotar PE, Uboldi AD, Clarke OB, Lin CS, Healer J, Smith BJ, Cowman AF (2012) Insights into Duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from Plasmodium falciparum. J Biol Chem 287:32922–32939.

39. Higgins MK (2008) The structure of a chondroitin sulfate-binding domain important in placental malaria. J Biol Chem 283:21842–21846.

40. Khunrae P, Philip JM, Bull DR, Higgins MK (2009) Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. J Mol Biol 393:202–213.

41. Vigan-Womas I, Guillotte M, Juillerat A, Hessel A, Raynal B, England P, Cohen JH, Bertrand O, Peyrard T, Bentley GA, Lewit-Bentley A, Mercereau-Puijalon O (2012) Structural basis for the ABO blood-group dependence of Plasmodium falciparum rosetting. PLoS Pathog 8:e1002781.

42. Chen E, Paing MM, Salinas N, Sim BK, Tolia NH (2013) Structural and functional basis for inhibition of erythrocyte invasion by antibodies that target Plasmodium falciparum EBA-175. PLoS Pathog 9:e1003390.

43. Malpede BM, Lin DH, Tolia NH (2013) Molecular basis for sialic acid-dependent receptor recognition by the Plasmodium falciparum invasion protein Erythrocyte-binding Antigen-140/BAEBL. J Biol Chem 288:12406–12415.

44. Gardner MJ, et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511.

45. Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T (2010) Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes—divide and conquer. PLoS Comput Biol 6:e1000933.

46. Klein MM, Gittis AG, Su HP, Makobongo MO, Moore JM, Singh S, Miller LH, Garboczi DN (2008) The cysteine-rich interdomain region from the highly variable plasmodium falciparum erythrocyte membrane protein-1 exhibits a conserved structure. PLoS Pathog 4:e1000147.

47. Smith JD, Kyes S, Craig AG, Fagan T, Hudson-Taylor D, Miller LH, Baruch DI, Newbold CI (1998) Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol Biochem Parasitol 97:133–148.

48. Smith JD, Craig AG, Kriek N, Hudson-Taylor D, Kyes S, Fagan T, Pinches R, Baruch DI, Newbold CI, Miller LH (2000) Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci USA 97:1766–1771.

49. Clausen TM, Christoffersen S, Dahlbäck M, Langkilde AE, Jensen KE, Resende M, Agerbæk MØ, Andersen D, Berisha B, Ditlev SB, Pinto VV, Nielsen MA, Theander TG, Larsen S, Salanti A (2012) Structural and functional insight into how the Plasmodium falciparum VAR2CSA protein mediates binding to chondroitin sulfate A in placental malaria. J Biol Chem 287: 23332–23345.

50. Srivastava A, Gangnard S, Round A, Dechavanne S, Juillerat A, Raynal B, Faure G, Baron B, Ramboarina S, Singh SK, Belrhali H, England P, Lewit-Bentley A, Scherf A, Bentley GA, Gamain B (2010) Full-length extracellular region of the var2CSA variant of PfEMP1 is required for specific, high-affinity binding to CSA. Proc Natl Acad Sci USA 107:4884–4889.

51. Brown A, Turner L, Christoffersen S, Andrews KA, Szestak T, Zhao Y, Larsen S, Craig AG, Higgins MK (2013) Molecular architecture of a complex between an adhesion protein from the malaria parasite and intracellular adhesion molecule 1. J Biol Chem 288:5992–6003.

52. Robinson BA, Welch TL, Smith JD (2003) Widespread functional specialization of Plasmodium falciparum erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol 47:1265–1278.

53. Howell DP, Levin EA, Springer AL, Kraemer SM, Phippard DJ, Schief WR, Smith JD (2008) Mapping a common interaction site used by Plasmodium falciparum Duffy binding-like domains to bind diverse host receptors. Mol Microbiol 67:78–87.