



Structure of the *Plasmodium*-interspersed repeat proteins of the malaria parasite

Thomas E. Harrison^a, Adam J. Reid^b, Deirdre Cunningham^c, Jean Langhorne^c, and Matthew K. Higgins^{a,1}

^aDepartment of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; ^bParasite Genomics, Wellcome Sanger Institute, Cambridge CB10 1SA, United Kingdom; and ^cThe Francis Crick Institute, London NW1 1AT, United Kingdom

Edited by Louis H. Miller, National Institute of Allergy and Infectious Diseases, Rockville, MD, and approved October 16, 2020 (received for review August 7, 2020)

The deadly symptoms of malaria occur as *Plasmodium* parasites replicate within blood cells. Members of several variant surface protein families are expressed on infected blood cell surfaces. Of these, the largest and most ubiquitous are the *Plasmodium*-interspersed repeat (PIR) proteins, with more than 1,000 variants in some genomes. Their functions are mysterious, but differential *pir* gene expression associates with acute or chronic infection in a mouse malaria model. The membership of the PIR superfamily, and whether the family includes *Plasmodium falciparum* variant surface proteins, such as RIFINs and STEVORs, is controversial. Here we reveal the structure of the extracellular domain of a PIR from *Plasmodium chabaudi*. We use structure-guided sequence analysis and molecular modeling to show that this fold is found across PIR proteins from mouse- and human-infective malaria parasites. Moreover, we show that RIFINs and STEVORs are not PIRs. This study provides a structure-guided definition of the PIRs and a molecular framework to understand their evolution.

The symptoms of malaria occur as *Plasmodium* parasites replicate within blood. This is a rich environment, replete with the nutrients required for growth and providing the opportunity for transmission by blood-sucking insects. However, blood also contains much of the machinery of the host immune defense. To survive under immune attack, *Plasmodium* parasites have evolved to replicate while hidden within host cells. Only a few parasite proteins are exposed on host cell surfaces, and these have mostly diversified into large protein families, allowing a population survival strategy based on antigenic variation (1, 2).

The best understood of the infected-erythrocyte surface protein families is the PfEMP1, members of which interact with human endothelial receptors, causing infected erythrocytes to adhere within the vasculature away from splenic clearance (1, 3, 4). However, PfEMP1 are found only in *Plasmodium falciparum* and the closely related *Laverania*. More ubiquitous across *Plasmodium* species are families of small variant surface antigens (VSAs) (1). These include the CIRs of *Plasmodium chabaudi* (5) and the VIRs of *Plasmodium vivax* (6), often known as the “*Plasmodium*-interspersed repeats” (PIRs) (7). The PIRs can be very abundant, with thousands of members in some genomes (8). However, whether some families of small VSAs in the *Laverania* and more distantly related *Plasmodium* species are part of the PIR superfamily is unclear. For example, the RIFINs and STEVORs of *P. falciparum* (9–11) were proposed to be PIRs due to their sizes, cellular locations, and the presence of shared sequence elements within intron regions of their genes (12). However, differences in gene structure and a low protein sequence identity make this assignment uncertain (7). Are the small VSAs part of a larger superfamily with related functions, or are they different protein families with different roles?

Also uncertain is whether the small VSAs are universally found on infected erythrocyte surfaces. Studies have located PIRs on or close to the surfaces of blood cells infected with *P. vivax* (6), *Plasmodium yoelii* (13), and *Plasmodium berghei* (14). Similarly, RIFINs and STEVORs have been located to surfaces of *P. falciparum*-infected erythrocytes (10, 15–17). Indeed,

natural infection with *P. vivax* induces antibodies that target VIRs (18), while unusual RIFIN-targeting antibodies found in adults in malaria endemic regions of Africa also recognize infected erythrocytes (19–21). These studies suggest that the small VSAs are molecules of infected blood cell surfaces and indicate that family expansion and diversification have occurred to allow them to undergo antigenic variation. However, other studies have cast doubt on the universality of this model, indicating an intracellular location for some small VSAs (22–24) or showing them be expressed in other life cycle stages of the parasite, including merozoites or gametocytes (23, 25–27). Has diversification of these small VSA families led to their use at different life cycle stages and different cellular locations during infection?

A number of recent studies indicate that small VSAs have important functions. First, *P. chabaudi* introduced into mice through mosquito bite are less virulent than those introduced by direct injection of infected blood (28). The major differences in gene expression in these parasites are in the *cir* gene repertoires, with a broader range of *cir* genes expressed on mosquito transmission (28). Different *cir* genes are also transcribed during the chronic phase of a mosquito-transmitted *P. chabaudi* infection compared with those expressed during the acute phase (29). When passaged in naïve mice, these chronic parasites are more virulent than those from acute stages of infection. Indeed, a more virulent *P. chabaudi* strain, PcCB, expresses more of the *pir* genes associated with chronic infection than a less virulent strain, PcAS (30). These findings combine to suggest an as-yet unknown

Significance

The *Plasmodium* parasites that cause malaria replicate within blood cells of an infected host. These parasites send a small number of proteins to infected blood cell surfaces, allowing them to bind host molecules but also risking their detection by the host immune system. These proteins have diversified into large families, allowing the parasite to avoid detection by using antigenic variation. The most ubiquitous of these families is the *Plasmodium*-interspersed repeat (PIR) protein family. Here we present the structure of a PIR protein, revealing the architecture of its ectodomain and showing how it has diversified. Finally, we use structure-guided methods to understand which small variant surface antigen families are PIRs and to understand their evolution across malaria parasites.

Author contributions: T.E.H., A.J.R., J.L., and M.K.H. designed research; T.E.H. and A.J.R. performed research; T.E.H., D.C., and J.L. contributed new reagents/analytic tools; T.E.H. analyzed data; and T.E.H. and M.K.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: matthew.higgins@bioch.ox.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2016775117/-DCSupplemental>.

First published November 30, 2020.

role for PIRs in modulating the virulence of infection. One possible mechanism in some *Plasmodium* species might be through causing adhesion of infected cells within the vasculature, allowing avoidance of splenic clearance. Indeed, VIRs from *P. vivax* are proposed to cause infected-reticulocytes to adhere to endothelial receptors including ICAM-1 (22, 31), while RIFINs and STEVORs can cause infected erythrocytes to adhere to uninfected erythrocytes, by interacting with blood group antigens or glycoprotein C (16, 32). Alternatively, subgroups of RIFINs have been shown to interact with human inhibitory immune receptors, such as LILRB1 and LAIR1 (17). The LILRB1-binding RIFINs mimic the natural ligand of LILRB1, MHC class I, allowing the RIFIN to inhibit markers of natural killer cell activation, most likely reducing parasite clearance (33). These studies imply wide-ranging roles for the small VSAs in both mammalian and insect hosts.

The important roles emerging for different small VSAs highlights the need to understand these mysterious parasite protein families in greater detail. In particular, are they members of the same *Plasmodium* superfamily, evolved to perform similar functions during infection, or do the small VSAs represent different protein families with different roles? To explore this question, we determined the structure of the extracellular domain of a CIR protein from *P. chabaudi*. We compared this structure with that of the variable domain of a LILRB1-binding RIFIN (33) and used structure-guided sequence analysis to predict which small VSAs are part of the PIR superfamily. This provides a framework for understanding PIR protein function and evolution.

Results

The Structure of a *P. chabaudi* PIR Protein Ectodomain. To determine the structure of a member of the PIR family, we focused on proteins from *P. chabaudi*, often known as CIRs. These consist of an N-terminal extracellular domain ranging in size from 28 kDa (236 residues) to 133 kDa (1,331 residues) in the AS strain. This is followed by a predicted transmembrane helix and small intracellular peptide. We assessed the expression of a panel of seven of the smaller CIR ectodomains (*SI Appendix, Fig. S1*). Five of these were expressed in HEK293F cells and were subjected to crystallization trials. Crystals formed for PCHAS_1200500 and contained three copies in the asymmetric unit. A complete dataset was collected to 2.15 Å, and the structure was determined by Sulfur-SAD phasing, using anomalous scattering from the sulfur atoms found in the five disulfide bonds and six methionine residues in each monomer (Fig. 1A and *SI Appendix, Table S1*). A complete model was built for residues 5 to 242. Another 20 residues, which connect the ectodomain to the predicted transmembrane helix, are missing from the C terminus, suggesting that this domain is connected to the membrane through a flexible, disordered linker. Indeed, residues 242 to 258 are predicted to be disordered (Fig. 1B).

The CIR ectodomain adopts a compact α -helical structure, arranged into two distinct clusters of helices (Fig. 1A). It does not adopt the same fold as known *Plasmodium* surface protein modules, such as the DBL and CIDR domains found in the PfEMP1 proteins, and the DALI server did not identify a similar protein of known structure. The N-terminal half of the ectodomain is formed from five tightly packed helices connected by long loops largely lacking in secondary structure. This packs against the C-terminal portion of the ectodomain, which is folded as a cluster of three parallel helices, connected by short loops. The C-terminal end of the fold consists of a long stretch of residues lacking in secondary structure, running alongside the final two helices. Five disulfide bonds stabilize the fold, supporting the hypothesis that the CIRs are surface proteins, because disulfide bonds are characteristic of proteins located in the oxidizing extracellular environment. Two of the disulfide bonds

lie within the loop between the first and second helices, and a third bond stabilizes the loop between helices 2 and 3. The final two disulfide bonds link helices within the core of the domain, with one linking helices 1 and 6 and another linking helices 7 and 8.

The CIR ectodomain is predicted to be tethered to membranes through a C-terminal transmembrane helix. The surface opposite this attachment site is dominated by a long loop that links helices 1 and 2. This is one of the most flexible regions of the domain, containing the residues with the highest crystallographic B factors (Fig. 1C). In addition, two copies of the domain in the asymmetric unit of the crystal have missing density in this region, with chain B missing residues 45 to 48 and chain C missing residues 25 to 29, again indicating flexibility. Other regions of higher B factors include the loops linking helices 4 and 5 and helices 6 and 7, both lying on the same side of the domain. For any CIR proteins found on infected erythrocytes, this surface will be the most exposed to the immune system. With disorder reducing the likelihood of antibody attachment, this suggests that flexible and disordered sequences have evolved on this exposed surface to decrease the likelihood of immune detection.

Diversity and Conservation within the CIR Proteins. To understand the degree to which CIR proteins diversify, we analyzed all 198 CIR sequences from the AS strain of *P. chabaudi* by aligning them to the structure of PCHAS_1200500. A conservation LOGO for sequences aligned in this way identified the majority of conserved residues as cysteines or aromatic residues, found predominantly on helices (Fig. 2A). The Shannon sequence entropy, which takes into account the chemical properties of amino acid side chains in assessing chemical conservation, was calculated for each residue, and these scores were plotted on to the PCHAS_1200500 structure (Fig. 2B). The majority of chemically well-conserved residues, with low entropy, were found in the core of the protein, stabilizing the packing of the helices and indicating conservation of the protein fold. Two of the disulfide bonds, corresponding to Cys12-Cys165 and Cys42-Cys52, were also well conserved. In contrast, the surface of the protein shows no conservation, as would be expected for a family of antigenically distinct proteins under immune pressure. This lack of conservation is particularly stark in both the sequence and the length of the surface loops that link the helices (Fig. 2). This is reminiscent of the DBL and CIDR domains of *Plasmodium* parasites, which also have conserved core aromatics and disulfide bonds to stabilize their fold, while showing great diversification on the surface (34–37).

The core PIR protein domain starts within five residues of the N terminus of each of these 198 CIR proteins. However, there are significant differences in the length and nature of the linker between this domain and the transmembrane helix, which ranges from 18 residues to 1,117 residues. The CIRs have previously been divided into the S family and the L family, with the length of the linker the primary difference between these groups (Fig. 2D and *SI Appendix, Fig. S2*). For example, in the S7 CIR subfamily, which contains PCHAS_1200500 and is associated with acute infection (29), this linker ranges from 20 to 68 residues, with an average of just 37. In contrast, in the L1 family, whose expression is associated with chronic infection, it is typically much longer, ranging from 59 to 1117 residues, with an average of 533. Analysis using five different predictors of protein disorder suggests that this portion of the ectodomain is largely disordered in all CIR protein classes (Fig. 2D and *SI Appendix, Fig. S2*). This indicates that the ectodomains of all CIRs share a similar architecture, with a structurally conserved yet sequence-diverse N-terminal fold connected to a transmembrane helix through a flexible linker of highly variable length.

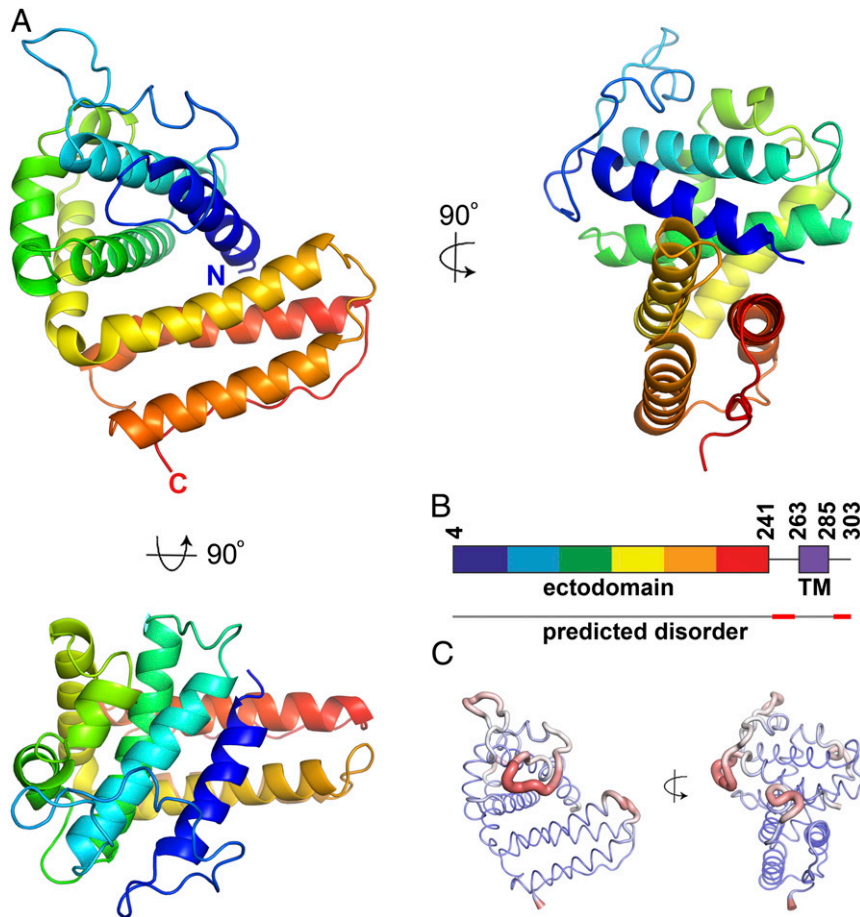


Fig. 1. The structure of a *P. chabaudi* PIR protein. (A) The structure of the CIR protein ectodomain from PCHAS_1200500 is shown in rainbow representation, with the N terminus in blue and the C terminus in red. (B) A schematic of the full CIR protein PCHAS_1200500 showing the ectodomain (rainbow) and the transmembrane helix (TM; purple). Below the schematic is a line representing the predicted disorder across the domain, with disordered regions in red. (C) The structure of the CIR domain represented as a wire, with the thickness scaled according to the crystallographic B factor of this region of the model. Broad, red regions of wire indicate more disordered regions of the protein domain.

Which Other Plasmodium Protein Families Share the PIR Protein Fold?

We next asked whether the CIR protein fold is conserved across the small VSAs, using structural insight to define the PIR superfamily. Comparisons of the CIR structure with our structure of the variable domain of a RIFIN (PF3D7_1254800) (33) revealed very different architectures (Fig. 3A). The RIFINs are mostly small proteins of 30 to 50 kDa with an N-terminal semi-conserved domain and a C-terminal variable domain, followed by a putative transmembrane helix (32, 38). A hydrophobic region lying between the variable and constant domains has been suggested to act as a second transmembrane helix, but recent studies show that it is more likely part of the ectodomain (32, 38, 39). The RIFIN variable domain consists of a three α -helix core decorated with complex loops, which include the LILRB1-binding site. This structure most closely resembles the three C-terminal helices of the CIR protein; however, the topology and arrangement of these helices do not match. It is also unlikely that the N-terminal conserved domain of the RIFIN is structurally similar to the CIR protein structure, owing to the size difference. While the CIR protein domain of PCHAS_1200500 is 241 residues long, the N-terminal domain of RIFIN_1254800 is just 124 residues long. Therefore, the CIR ectodomain fold is too large to be accommodated within the RIFIN conserved domain. These findings suggest that RIFINs and CIRs are not structurally equivalent but rather are two distinct classes of *Plasmodium* surface proteins.

We next aimed to determine whether other small VSAs, which have not yet been structurally characterized, adopt the same fold as either a CIR or a RIFIN. We studied diverse small VSA sequences from the rodent-infective *Vinckeia* clade, the monkey- and human-infecting *Plasmodium* clade, ape-infecting *Laverania* species, and rodent- and bird-infective *Plasmodium* species, aiming to survey across the *Plasmodium* evolutionary tree (*SI Appendix, Fig. S3*). For each protein family, we selected 10 sequences, chosen to represent a range of sequence diversity (*SI Appendix, Fig. S3*). These were analyzed by structure-based threading using Phyre2 (40), generating homology models and providing a score that represents confidence in the model generated (Fig. 3B and *SI Appendix, Fig. S3*). Threading 10 sequence-diverse CIRs onto the PCHAS_1200500 structure generated models with a Phyre2 confidence score of 100%, supporting the conservation of this fold among PIRs from *P. chabaudi*. In contrast, threading 10 sequence-diverse RIFINs onto the PCHAS_1200500 structure gave an average confidence score of only 1% (range, 0.1 to 2.6%), as the RIFINs do not share the CIR protein fold.

To determine whether the fold is conserved in small VSAs from other rodent-infecting *Plasmodium* species, we selected proteins from *P. berghei* and *P. yoelii*. In each case, Phyre2 generated molecular models, based on threading these sequences onto the PCHAS_1200500 structure, with confidence values of 98.94 to 100%, strongly indicative of a conserved fold (Fig. 3B

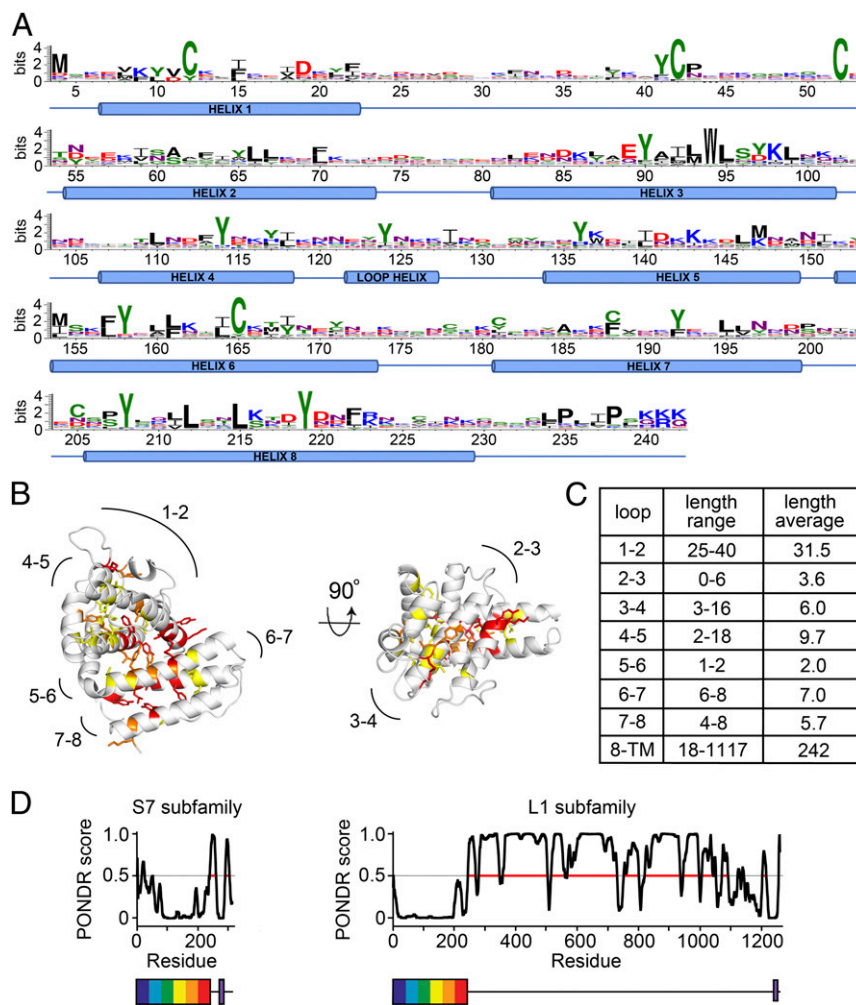


Fig. 2. Conservation and diversity in the CIR proteins. (A) All CIR sequences from the AS strain of *P. chabaudi* were aligned against the structure of PCHAS_1200500, and a sequence logo was generated, numbered according to PCHAS_1200500. (B) Residues most chemically conserved across the CIR proteins of the AS strain are shown as sticks. Residues are color-coded by Shannon sequence entropy: 0.75 to 1.0 in yellow, 0.5 to 0.75 in orange, and <0.5 in red. Lower sequence entropy indicates greater conservation of side chain chemical properties. (C) A table of the lengths of the loops and of the linker between the PIR domain fold and the transmembrane helix. (D) Prediction of disorder (PONDNR score) for members of the S7 and L1 CIR protein subfamilies, in each case representing the protein closest in sequence to the sequence logo for that protein family, PCHAS_0500200 for the S7 family and PCHAS_0601000 for the L1 family. Prediction of disorder, determined using PONDNR, is plotted against residue number. Below the plots are representations of the two proteins, showing the PIR protein domain as a rainbow box and the transmembrane helix in purple.

and *SI Appendix, Fig. S3*). To reveal the degree of conservation across the *P. berghei* and *P. yoelii* VSAs, we also generated sequence LOGOs for 135 *P. berghei* proteins and 1,011 *P. yoelii* proteins, allowing their comparison with sequence LOGOs for the CIRs. Plotting these onto the homology models showed that the *P. berghei* and *P. yoelii* VSAs share conserved sequence features with the CIRs, including conservation of two disulfide bonds, equivalent to conserved disulfide bonds C12-C165 and C42-C52 of the CIRs, as well as conservation of hydrophobic residues found in the domain core (Fig. 3C and *SI Appendix, Fig. S4*). These data show that it is highly likely that the *P. berghei* and *P. yoelii* VSAs adopt the same global structure as the *P. chabaudi* CIRs and are part of the same PIR protein family.

We next conducted the same analysis with VSAs from human-infecting malaria parasites, *P. vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium knowlesi*. Ten of each of these proteins were threaded onto the PCHAS_1200500 structure in Phyre2, and scores were obtained. In each case, the average score exceeded the 95% cutoff for a high-confidence model; 97.9% for *P. vivax*, 97.7% for *P. malariae*, 98.3% for *P. ovale*, and

98.7% for *P. knowlesi* (Fig. 3 and *SI Appendix, Fig. S3*). These findings indicate that small VSAs from these rodent-infecting and monkey-infecting clades are part of the PIR superfamily and that they share a common overall architecture.

We next studied small VSAs from *P. falciparum*. As described above, threading 10 diverse *P. falciparum* RIFIN sequences onto the CIR structure generated an average score of 1%, suggesting no structural similarity (Fig. 3B), agreeing with our comparison of the CIR structure with that of the variable domain of the RIFIN (Fig. 3A). However, no structure is available for the constant domain of a RIFIN. As this is smaller than the CIR ectodomain, we assessed whether the constant domain of the RIFIN shares structural homology with either the N-terminal or C-terminal lobes of the CIR. To do so, we threaded 10 sequence-diverse RIFIN constant domain sequences onto either the N-terminal (helices 1 to 5) or the C-terminal (helices 6 to 8) lobes of the CIR. In both cases, threading RIFIN sequences onto CIR structures gave confidence scores of <1%. A similar analysis of the *P. falciparum* STEVOR gave an average confidence score of 83.0% when we threaded the STEVOR sequences on the CIR

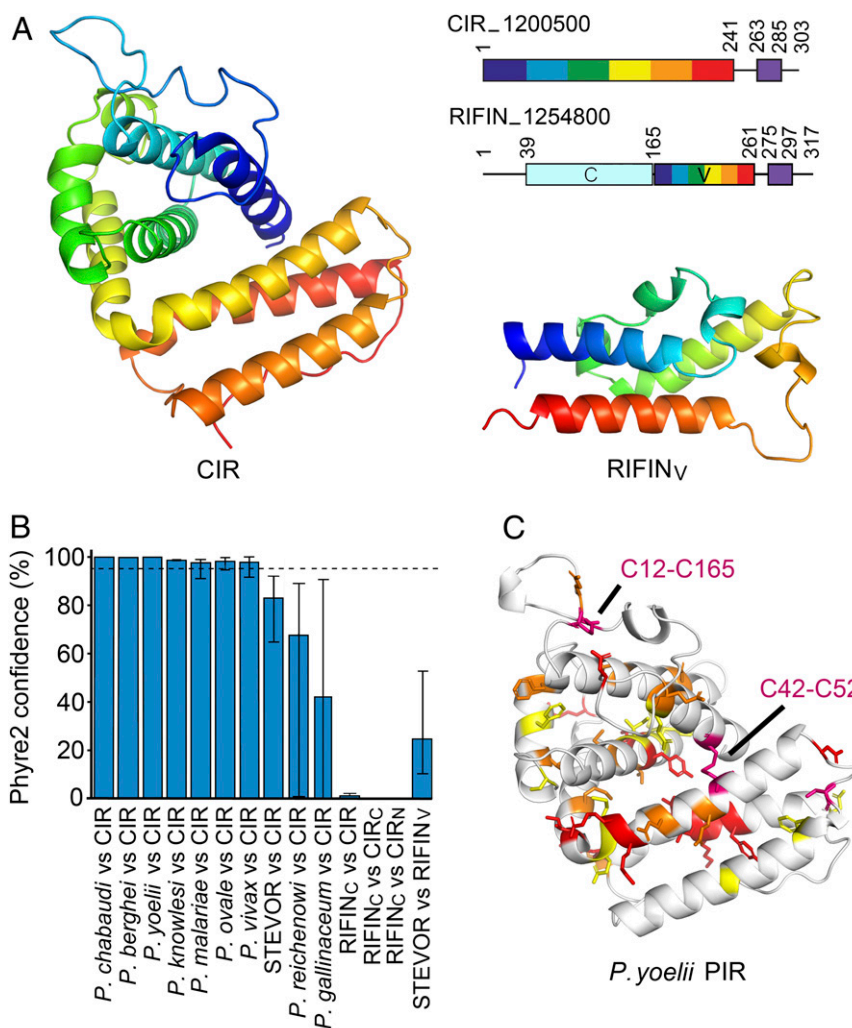


Fig. 3. Structure-guided determination of membership of the PIR protein superfamily. (A) Comparison of the structure of the CIR protein ectodomain with that of the variable domain of the RIFIN (RIFIN_V). Both are shown as rainbow representation from N terminus (blue) to C terminus (red). The top right panel shows schematics for these two proteins, with numbers referring to the position in the protein sequence. Both have a transmembrane helix, represented as a purple box, close to the C terminus. The RIFIN has a variable domain (V; rainbow representation) and a constant domain (C; pale blue). (B) The Phyre2 confidence scores for modeling of small VSAs from different *Plasmodium* species using the known structures of the CIR protein, its C-terminal lobe (CIR_C; residues 152 to 241), its N-terminal lobe (CIR_N; residues 4 to 151), or against the RIFIN_V domain. The dashed line represents the high confidence cutoff limit of 95%, and the error bars show the range of scores obtained from the 10 examples of each PIR protein modeled. (C) A model of a *P. yoelii* PIR protein extracellular domain, PY17X_094500, based on the CIR structure. Residues with a property entropy of 0.75 to 1.0 are in yellow, those with 0.5 to 0.75 are in orange, and those with <0.5 are in red. Two disulfide bonds shared with the CIR protein structure are shown in purple, with the relevant cysteine residues labeled.

structure and 24.7% when we threaded the STEVEOR sequences onto the structure of the RIFIN variable domain. In neither case do these scores exceed the 95% threshold for high confidence. This compares with a confidence score of 94.3% when threading nine RIFIN structures onto the RIFIN variable domain structure (excluding one outlier).

Finally, we modeled the small VSAs from *Plasmodium reichenowi* and *Plasmodium gallinaceum*. *P. reichenowi*, like *P. falciparum*, is a member of the *Laverania* group of parasites. Here, threading 10 sequences onto the structure of the CIR generated an average score of 66.4%, while threading these sequences onto the RIFIN variable domain generated a score of 71.5%, both of which are below the confidence cutoff. These findings support the placement of the RIFINs and STEVEORs of *P. falciparum* and the putative PIR proteins of *P. reichenowi* outside the PIR protein superfamily. Similarly, threading 10 sequences of small VSAs from bird-infective *P. gallinaceum* onto the CIR structure

gave an average confidence score of 42.5%, while threading onto the RIFIN variable domain gave an average of 7.2%. This analysis suggests that the emergence of the PIR superfamily occurred after the separation of the rodent and simian-infective *Plasmodium* species from the *Laverania* and the bird-infective species, and indicates that the small VSAs found in the latter are not PIR proteins.

Discussion

The structure of the ectodomain of a *P. chabaudi* CIR protein has allowed us to explore the fold which underlies this protein family and to determine the degree to which it has diversified. We find the CIRs to adopt a novel structure, consisting of eight α -helices, divided into two lobes. Despite their different architectures, the CIRs share properties with other proteins from the surfaces of *Plasmodium*-infected erythrocytes. Like the CIDR and DBL domains of *P. falciparum* PfEMP1 proteins, CIRs are

built on a core α -helical scaffold. The most conserved residues are hydrophobic and aromatic or are disulfide bond-forming cysteine residues. These are all internal, stabilizing the structure. As disulfide bonds form only in the oxidizing extracellular environment, the presence of five such bonds in the structure of PCHAS_1200500, and the conservation of two of these across the rodent-infective malaria species, are indicative of a cell surface role for the PIRs.

Similar too, when comparing the CIR proteins with the PfEMP1, is that each has its own conserved structural fold, which is decorated with flexible surface loops. This has allowed extensive surface diversification in both sequence and loop length. This again is indicative of a surface location for the PIR proteins, with the development of antigenically distinct molecules to allow sequential deployment during an infection a potential driving force for diversification. In the CIRs, these domains are attached to the cell surface through a linker of variable length, perhaps acting as a protein–protein interaction module swinging on a flexible chain.

Families of small VSAs from across the *Plasmodium* genus have been proposed to be part of the PIR superfamily. However, the true extent of this family—and thus the origins, common or otherwise, of genes from different species—has remained unclear. As structures are more conserved than sequences, the structure of a PIR protein family member from *P. chabaudi* has allowed us to explore which other *Plasmodium* small VSAs are part of the PIR protein superfamily. This analysis predicts that the majority of other proteins examined, including those from the rodent-infective parasites *P. berghei* and *P. yoelii* and from human-infective species *P. ovale*, *P. vivax*, *P. knowlesi*, and *P. malariae*, share the PIR protein fold and can be placed within the PIR superfamily. In contrast, analysis of RIFIN and STEVOR proteins of *P. falciparum* and the small VSAs of *P. reichenowi* and *P. gallinaceum* suggests that they do not share this fold, challenging their membership of the PIR superfamily. Examination of the *Plasmodium* family tree therefore suggests that the genesis of the PIR superfamily might have occurred after the divergence of the *Laverania* and bird-infective *Plasmodium* species from the rodent and simian-infective species, and reveals that the small VSAs from the *Plasmodium* species cannot be considered a homogeneous whole.

The placement of the RIFINs outside the PIR protein superfamily has functional consequences. The expression of different *P. chabaudi* pirs has recently been associated with acute or chronic *P. chabaudi* infection (28, 29). The growing evidence that RIFINs of *P. falciparum* bind to inhibitory immune receptors, such as LILRB1 and LAIR1 (17), and dampen immune signaling (33), is thus tantalizing. Could PIR proteins also interact with inhibitory immune receptors, thereby promoting chronic disease? It is still possible that they do, but the placement of the PIRs and the RIFINs in different protein families suggests that if this is indeed the case, it is the outcome of convergent evolution.

Materials and Methods

Protein Expression and Purification. Synthetic genes encoding a panel of CIR ectodomains were cloned into a pTT3 vector, giving a C-terminal His6 tag. In the case of PCHAS_1200500, this contained five changes (S5A, T125A, S152A, S176A, and T179A) intended to remove putative N-linked glycosylation sites. These were transfected into HEK293F cells (Thermo Fisher Scientific) using polyethylenimine, and after 5 d harvested by centrifugation at $5,000 \times g$. The supernatant was buffer-exchanged into 20 mM Hepes pH 7.5, 150 mM NaCl, and 20 mM imidazole by tangential flow filtration, and the protein was purified by immobilized metal affinity chromatography using Ni²⁺-NTA resin, followed by size exclusion chromatography using a Superdex 75 10/300 column (GE Healthcare Life Sciences).

To methylate the protein, 20 mM dimethylamine-borane complex (ABC) and 40 mM formaldehyde were added, and the mixture was incubated for 2 h at room temperature. After incubation, this was repeated two more times, followed by the addition of 10 mM ABC and incubation of the

reaction overnight. The protein was then purified by size exclusion chromatography as above.

Crystallization, Data Collection, and Structure Determination. For crystallization, methylated PCHAS_1200500 (2–264) was concentrated to 34 mg/mL, with trials carried out using vapor diffusion in sitting drops with a mixture of 100 nL of protein solution and 100 nL of well solution. Crystals were obtained after 40 d at 4 °C using a well solution of 0.1 M Hepes pH 6.5 and 45% wt/vol poly(acrylic acid sodium salt) 2100, and then cryocooled for data collection in liquid nitrogen.

Sulfur-SAD data were collected at the I23 long-wavelength beamline at the Diamond Light Source, using a wavelength of 2.75520 Å. Data were indexed and scaled using XDS, giving a resolution of 3 Å. Phases were obtained using SHELXD, and an initial model was then built using Buccaneer, with three molecules in the asymmetric unit of the crystal.

Native data were collected at the ID23-1 beamline at the European Synchrotron Radiation Facility and indexed and scaled using GRENDES fastproc, giving a resolution of 2.15 Å. The initial model from the Sulfur-SAD data was used as a search model, allowing it to be located within the higher-resolution dataset using Phaser MR. The structure was then built and refined using cycles of Coot and Buster.

Circular Dichroism Analysis. Circular dichroism experiments were conducted using a Jasco J815 CD spectrophotometer. The CIR was desalted into 20 mM sodium phosphate pH 7.5 and 150 mM NaF buffer using PD-10 columns (GE Healthcare) and diluted to a concentration of 0.26 mg/mL. A spectrum was obtained at 20 °C between 260 nm and 190 nm using a 1-mm path length, with measurements taken every 0.5 nm, subtracting a baseline determined using buffer. Ten equivalent spectra were averaged together.

Sequence Analysis. Small VSA sequences were extracted from *Plasmodium* genomes, including 198 amino acid sequences from the *P. chabaudi* AS v3 genome assembly (29), 135 sequences from the *P. berghei* ANKA genome assembly (25), 1,011 sequences from the *P. yoelii* 17X v3 genome assembly (41), 70 sequences from the *P. knowlesi* strain H genome assembly (42), 136 sequences from the *P. malariae* UG01 genome assembly (43), 1,495 sequences from the *P. ovale* curtisi GH01 genome assembly (43), 185 RIFIN sequences and 31 STEVOR sequences from the *P. falciparum* 3D7 genome assembly (44), 1,086 sequences from the *P. vivax* P01 genome assembly (8), 487 sequences from the *P. reichenowi* G01 genome assembly (45) and 20 sequences from the *P. gallinaceum* 8A genome assembly. All sequences were retrieved from PlasmoDB (46).

Sequences were aligned using MUSCLE v3.8.31 with default parameters. Sequence LOGOs were generated using <http://weblogo.threeplusone.com>. Homology models were produced using the Phyre2 server (40).

Phylogenetic Analysis. Evolutionary analyses were conducted in MEGA X (47, 48). The evolutionary history was inferred using the maximum likelihood method and the Whelan and Goldman frequency model (49). The tree with the highest log-likelihood value is shown. Initial trees for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a Jones–Taylor–Thornton model and then selecting the topology with the superior log-likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

Data Availability. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID code 6ZYV). All other data and protein constructs are available from the authors on request.

ACKNOWLEDGMENTS. We thank Ed Lowe and the staff of the Promixa1 beamline at Soleil for help with crystallographic data collection and Ramona Duman and Kamel El Omari of beamline I23 at Diamond Light Source for assistance with collection of long-wavelength crystallographic data. M.K.H. is a Wellcome Trust Investigator (101020/Z/13/Z) and is funded by the Medical Research Council (MR/T000368/1). T.E.H. is funded by the Peter J. Braam graduate scholarship for Global Wellbeing at Merton College, Oxford. A.J.R. is funded by the Wellcome Trust (Grant 206194/Z/17/Z) and a Medical Research Council Programme grant (MR/M003906/1). J.L. is a Wellcome Trust senior investigator (104777/Z/14/Z). J.L. and D.C. are supported by the Francis Crick Institute, which receives its core funding from the Wellcome Trust, the UK Medical Research Council, and Cancer Research UK (FC001101).

1. M. Wahlgren, S. Goel, R. R. Akhouri, Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat. Rev. Microbiol.* **15**, 479–491 (2017).
2. K. W. Deitsch, R. Dzikiowski, Variant gene expression and antigenic variation by malaria parasites. *Annu. Rev. Microbiol.* **71**, 625–641 (2017).
3. L. Hviid, A. T. Jensen, PfEMP1–A parasite protein family of key importance in *Plasmodium falciparum* malaria immunity and pathogenesis. *Adv. Parasitol.* **88**, 51–84 (2015).
4. J. D. Smith, J. A. Rowe, M. K. Higgins, T. Lavstsen, Malaria's deadly grip: Cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cell. Microbiol.* **15**, 1976–1983 (2013).
5. K. Fischer *et al.*, Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of *Plasmodium chabaudi*, a malaria species that undergoes antigenic variation in the laboratory mouse. *Mol. Microbiol.* **48**, 1209–1223 (2003).
6. H. A. del Portillo *et al.*, A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842 (2001).
7. D. Cunningham, J. Lawton, W. Jarra, P. Preiser, J. Langhorne, The *pir* multigene family of *Plasmodium*: Antigenic variation and beyond. *Mol. Biochem. Parasitol.* **170**, 65–73 (2010).
8. S. Auburn *et al.*, A new *Plasmodium vivax* reference sequence with improved assembly of the subtelomeres reveals an abundance of *pir* genes. *Wellcome Open Res.* **1**, 4 (2016).
9. Q. Cheng *et al.*, *Stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97**, 161–176 (1998).
10. S. A. Kyes, J. A. Rowe, N. Kriek, C. I. Newbold, Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9333–9338 (1999).
11. V. Fernandez, M. Hommel, Q. Chen, P. Hagblom, M. Wahlgren, Small, clonally variant antigens expressed on the surface of the *Plasmodium falciparum*-infected erythrocyte are encoded by the *rif* gene family and are the target of human immune responses. *J. Exp. Med.* **190**, 1393–1404 (1999).
12. C. S. Janssen, R. S. Phillips, C. M. Turner, M. P. Barrett, *Plasmodium* interspersed repeats: The major multigene superfamily of malaria parasites. *Nucleic Acids Res.* **32**, 5712–5720 (2004).
13. D. A. Cunningham *et al.*, Host immunity modulates transcriptional changes in a multigene family (*yir*) of rodent malaria. *Mol. Microbiol.* **58**, 636–647 (2005).
14. F. Di Girolamo *et al.*, *Plasmodium* lipid rafts contain proteins implicated in vesicular trafficking and signalling as well as members of the PIR superfamily, potentially implicated in host immune system interactions. *Proteomics* **8**, 2500–2513 (2008).
15. M. Niang, X. Yan Yam, P. R. Preiser, The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog.* **5**, e1000307 (2009).
16. M. Niang *et al.*, STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* **16**, 81–93 (2014).
17. F. Saito *et al.*, Immune evasion of *Plasmodium falciparum* by RIFIN via inhibitory receptors. *Nature* **552**, 101–105 (2017).
18. P. Requena *et al.*, *Plasmodium vivax* VIR proteins are targets of naturally-acquired antibody and T cell immune responses to malaria in pregnant women. *PLoS Negl. Trop. Dis.* **10**, e0005009 (2016).
19. J. Tan *et al.*, A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* **529**, 105–109 (2016).
20. K. Pieper *et al.*, Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. *Nature* **548**, 597–601 (2017).
21. F. L. Hsieh, M. K. Higgins, The structure of a LAIR1-containing human antibody reveals a novel mechanism of antigen recognition. *eLife* **6**, e27311 (2017).
22. M. Bernabeu *et al.*, Functional analysis of *Plasmodium vivax* VIR proteins reveals different subcellular localizations and cytoadherence to the ICAM-1 endothelial receptor. *Cell. Microbiol.* **14**, 386–400 (2012).
23. X. Y. Yam *et al.*, Characterization of the *Plasmodium* Interspersed Repeats (PIR) proteins of *Plasmodium chabaudi* indicates functional diversity. *Sci. Rep.* **6**, 23449 (2016).
24. M. Petter *et al.*, Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol. Biochem. Parasitol.* **156**, 51–61 (2007).
25. A. Fougère *et al.*, Variant exported blood-stage proteins encoded by *Plasmodium* multigene families are expressed in liver stages where they are exported into the parasitophorous vacuole. *PLoS Pathog.* **12**, e1005917 (2016).
26. A. J. Reid *et al.*, Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *eLife* **7**, e33105 (2018).
27. S. B. Mwakalinga *et al.*, Expression of a type B RIFIN in *Plasmodium falciparum* merozoites and gametes. *Malar. J.* **11**, 429 (2012).
28. P. J. Spence *et al.*, Vector transmission regulates immune control of *Plasmodium* virulence. *Nature* **498**, 228–231 (2013).
29. T. Brugat *et al.*, Antibody-independent mechanisms regulate the establishment of chronic *Plasmodium* infection. *Nat. Microbiol.* **2**, 16276 (2017).
30. J. W. Lin *et al.*, Genomic and transcriptomic comparisons of closely related malaria parasites differing in virulence and sequestration pattern. *Wellcome Open Res.* **3**, 142 (2018).
31. B. O. Carvalho *et al.*, On the cytoadhesion of *Plasmodium vivax*-infected erythrocytes. *J. Infect. Dis.* **202**, 638–647 (2010).
32. S. Goel *et al.*, RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nat. Med.* **21**, 314–317 (2015).
33. T. E. Harrison *et al.*, Structural basis for RIFIN-mediated activation of LILRB1 in malaria. *Nature* **587**, 309–312, 10.1038/s41586-020-2530-3 (2020).
34. M. K. Higgins, M. Carrington, Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein Sci.* **23**, 354–365 (2014).
35. C. K. Lau *et al.*, Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. *Cell Host Microbe* **17**, 118–129 (2015).
36. F. L. Hsieh *et al.*, The structural basis for CD36 binding by the malaria parasite. *Nat. Commun.* **7**, 12837 (2016).
37. F. Lennartz, C. Smith, A. G. Craig, M. K. Higgins, Structural insights into diverse modes of ICAM-1 binding by *Plasmodium falciparum*-infected erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20124–20134 (2019).
38. A. Bachmann *et al.*, A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar. J.* **14**, 274 (2015).
39. A. Andersson *et al.*, Membrane integration and topology of RIFIN and STEVOR proteins of the *Plasmodium falciparum* parasite. *FEBS J.* **287**, 2744–2762 (2019).
40. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
41. T. D. Otto *et al.*, A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol.* **12**, 86 (2014).
42. A. Pain *et al.*, The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799–803 (2008).
43. G. G. Rutledge *et al.*, *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
44. U. Böhme, T. D. Otto, M. Sanders, C. I. Newbold, M. Berriman, Progression of the canonical reference malaria parasite genome from 2002–2019. *Wellcome Open Res.* **4**, 58 (2019).
45. T. D. Otto *et al.*, Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* **5**, 4754 (2014).
46. C. Aurrecochea *et al.*, PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009).
47. G. Stecher, K. Tamura, S. Kumar, Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol. Biol. Evol.* **37**, 1237–1239 (2020).
48. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
49. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).