

UNIVERSITY OF OXFORD

DOCTORATE OF PHILOSOPHY

**Structural studies of *Pf*EMP1 bound to human
Cluster of Differentiation 36 and of an unusual
human antibody**

Fu-Lien Hsieh



A thesis submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy

Department of Biochemistry

St. John's College

Trinity Term 2017

Abstract

Structural studies of *Pf*EMP1 bound to human Cluster of Differentiation 36 and of an unusual human antibody

Fu-Lien Hsieh

Doctor of Philosophy

St. John's College, University of Oxford

Trinity Term 2017

In *Plasmodium falciparum*-infected humans, parasites escape from splenic clearance and cause severe or chronic symptoms partly through the action of the variable surface antigens expressed on the surfaces of infected erythrocytes, called *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1). *Pf*EMP1 that interact with human Cluster of Differentiation 36 (CD36) receptor mediate the most common adhesion phenotype found in infected erythrocytes. Here the first crystal structure of CD36 in complex with a *Pf*EMP1 CIDR domain is presented. Together with sequence analysis and biophysical data, it shows how parasite *Pf*EMP1 proteins are able to diversify hugely while retaining a chemically conserved CD36 binding surface. This reveals insight into the structure and function of CD36 and shows how the parasite targets a region of CD36 that is critical for its physiological role, avoiding mutations in host receptors that would prevent parasite binding.

Repetitive interspersed family proteins (RIFINs) are the most abundant variant surface antigens in *P. falciparum*. However, unlike the well-studied *Pf*EMP1, the function of RIFIN remains unclear. Recently, a subset of human antibodies was identified from malaria-endemic areas that can recognise varied RIFINs. These antibodies have the unique feature of the insertion of a human receptor, LAIR1 (CD305), into the third complementarity-determining region of their heavy chain. Here, the first structure of one of these unusual LAIR1-containing antibodies is presented. It reveals that the CDR loops are mostly occluded by the insert and form a surface that positions LAIR1 for antigen binding. The variable regions of the antibody therefore no longer perform their classical role in antigen recognition. Instead they act as an adaptor between a human LAIR1 insert with antigen binding properties and the immune-modulatory antibody constant regions.

Acknowledgements

I would like to thank my supervisor Prof. Matt Higgins for his guidance and encouragement over the course of my doctoral studies. I am very lucky to have a mentor like him, understanding my shortcomings and always having faith in me.

The project was completed in collaboration with Dr. Thomas Lavstsen (University of Copenhagen), Dr. Louise Turner (University of Copenhagen), Dr. Jani Bolla (University of Oxford) and Prof. Carol Robinson (University of Oxford). Thank you for the discussion, advice and assistance. I would also like to express my thanks to everyone in Higgins' lab for making this place so enjoyable and I am grateful to have this opportunity to work with them. Ray and Jon are such kind colleagues and I feel lucky to have them sharing my emotion always. Thanks are reserved for Dr. Edward Lowe and Dr. David Staunton (Biochemistry Department) for keeping crystallography and biophysical facilities working well. I would like to extend my thanks to Ms. Margaret Jones in STRUBI for all her help to purchase a key reagent in my projects. In particular, I wish to acknowledge the help provided by Dr. Tao-Hsin Chang. This thesis would not have been successful without him. I am also very grateful to the funding source provided by Taiwan Bio-Development Foundation.

Finally, I would like to offer my special thanks to my family and friends, especially my husband Tao-Hsin, for their love, support and encouragement.

Table of contents

Abstract	I
Acknowledgements	II
Table of contents	III
List of figures	VI
List of tables	VIII
List of abbreviations	IX
 Chapter 1 Introduction	 1
1.1 Malaria	2
1.1.1 An ancient disease with a global distribution	2
1.1.2 Malaria is caused by different species of <i>Plasmodium</i> parasites	3
1.1.3 Life cycle	4
1.1.4 Symptomatic and asymptomatic malaria	7
1.2 Multi-gene family in malaria: variant surface antigens	11
1.3 <i>P. falciparum</i> erythrocyte membrane protein1 (<i>PfEMP1</i>)	14
1.3.1 Structure of CIDR and DBL domains	17
1.3.2 The endothelial receptors of <i>PfEMP1</i>	20
1.3.2.1 The interaction of <i>PfEMP1</i> with ICAM-1 and EPCR	21
1.3.2.2 The interaction of <i>PfEMP1</i> with CD36	22
1.3.2.3 Human cluster of differentiation 36 (CD36)	23
1.3.2.4 Natural ligands of CD36	24
1.3.2.5 CD36: a friend or enemy in malaria	26
1.4 Repetitive interspersed family proteins (RIFINs)	29
1.5 Antibody structure and function	31
1.5.1 Basic antibody structure	31
1.5.2 The antigen-binding site and its diversity	33
1.5.3 A LAIR1-containing human antibody	36
1.6 Leukocyte-associated immunoglobulin-like receptor-1 (LAIR1)	37
1.7 Aims	39
 Chapter 2 Materials and methods	 40
2.1 Construct design and molecular cloning	41
2.1.1 Constructs for CD36 and MCvar1 CIDR α 2.8	42
2.1.2 Constructs for Fab-MGD21	42
2.1.3 Constructs for RIFIN, LAIR1 and LAIR1m	43
2.2 Protein expression and purification	44
2.2.1 HEK cells maintenance and large-scale expression	44

2.2.2 Expression and purification of CD36 and MCvar1 CIDR α 2.8	45
2.2.3 Expression and purification of Fab-MGD21, RIFIN, LAIR1 and LAIR1m	46
2.3 Crystallisation and cryo-protection	47
2.3.1 Crystallisation of CD36:MCvar1 CIDR α 2.8	47
2.3.2 Crystallisation of Fab-MGD21	48
2.4 X-ray data collection and structure determination	49
2.4.1 Data collection and structure determination for CD36:MCvar1 CIDR α 2.8	49
2.4.2 Data collection and structure determination of Fab-MGD21	50
2.5 Surface plasmon resonance (SPR)	51
2.5.1 SPR analysis of the interaction of CD36 and CIDR domains	51
2.5.2 SPR assay of the interaction of RIFIN with LAIR1-containing ligands	52
2.6 Size-elution chromatography multi-angle light scattering (SEC-MALS)	53
2.7 Analysis of <i>PfEMP1</i> sequences	54
2.8 Determination of unknown fatty acids using GC/MS	55
Chapter 3 Structural and functional characterisation of the interaction of <i>PfEMP1</i> with CD36	56
3.1 Project aim and summary	57
3.1.1 Declaration	58
3.2 Results	59
3.2.1 Protein production and characterisation of CD36 and CIDR α complex	59
3.2.2 Crystallisation and structure determination of CD36 bound to a CIDR α domain	62
3.2.3 The structural and functional analysis of CD36 binding by the <i>PfEMP1</i> CIDR α 2.8 domain	70
3.2.4 Structural comparison of CD36 with LIMP-2 and analysis of the binding of CD36 to long-chain fatty acid	74
3.2.5 Conserved structural features of CIDR α domains allow CD36 binding	81
3.2.6 The binding of CIDR α domain affects the natural function of CD36	90
3.3 Conclusion	95
Chapter 4 A LAIR1-containing human antibody structure	97
4.1 Project aim and summary	98
4.2 Results	100
4.2.1 Expression and purification of the MGD21 Fab fragment	100
4.2.2 Crystal structure of Fab-MGD21	102

4.2.3 Structural comparison of germ line LAIR1 and antibody-inserted LAIR1m	116
4.2.4 Novel CDR H3 analysis	120
4.3 Biophysical characterisation of RIFIN-ligand binding	122
4.4 Conclusion	125
Chapter 5 Discussion and perspectives	126
5.1 The <i>PfEMP1</i> CIDR domains	127
5.1.1 The opened and closed form of the MCvar1 CIDR α 2.8 domain	127
5.1.2 CIDR α domains for human CD36 or EPCR	131
5.1.3 A clever trick of the <i>PfEMP1</i> CIDR α domains	134
5.2 Insights from LAIR1m-containing antibody structure	137
5.2.1 Unusual antibody structures	137
5.2.2 CDR H3 comparison	140
5.2.3 The antibody engineering	142
5.3 Further questions and perspective	144
Appendices	146
Appendix 1. Polylinkers of pHLsec-based vector variants	147
Appendix 2. Primers used in the thesis	150
Appendix 3. CIDR domain sequences for recombinant proteins	152
Appendix 4. Publications arising from the thesis	155
References	156

List of figures

1.1	Population risk of malaria	3
1.2	The generalised life cycle of <i>Plasmodium</i> parasites	5
1.3	The complications of severe malaria in children and adults	9
1.4	Schematic of the <i>var</i> gene locus	14
1.5	<i>var</i> gene organisation and PfEMP1 proteins structure	16
1.6	The DBL3X domain of var2csa	18
1.7	Structural alignment of CIDR domains	19
1.8	Schematic diagrams of PfEMP1 and MC179 structure	20
1.9	Schematic representation of CD36	24
1.10	Schematic representation of RIFIN proteins	30
1.11	Schematic diagram of immunoglobulin structure	32
1.12	The generation of antibody diversity	34
1.13	LAIR-1 structure	38
3.1	Protein expression, purification and biophysical analysis	60
3.2	Crystal structure of the CD36:CIDR α 2.8 complex	66
3.3	Analysis of crystal packing of CD36:CIDR α 2.8 complex	68
3.4	The structural basis for CD36 binding by CIDR α 2.8 domain	71
3.5	Structural details of binding sites in the CD36:CIDR α 2.8 complex	72
3.6	Superimposition of CD36 and LIMP-2	76
3.7	Sequence analysis of CD36 and its homologues	77
3.8	CD36 and its binding of fatty acids	79
3.9	CD36 tunnels for fatty acids	80
3.10	Extensive diversity in CD36 binding CIDR α domains	85
3.11	Conservation across the CD36-binding CIDR α 2-6 domains	86
3.12	Limited chemical conservation of CIDR α domain allows CD36 binding	88
3.13	The effect of the F153A mutation of CD36 on the binding of CIDR α 2-6 variants	89
3.14	The use of SPR experiments to study competition binding	92
3.15	PfEMP1 CIDR α domains prevent oxidised low-density lipoprotein (oxLDL), but not thrombospondin 1(TSP1), from binding to CD36	93
3.16	A model for the inhibition of oxLDL binding by PfEMP1	95
4.1	Secreted Fab-MGD21 protein expression and purification	101
4.2	A view of the electron density of two Fab-MGD21 molecules in the asymmetric unit	104
4.3	Structure of a LAIR1m-containing antibody Fab-MGD21 fragment	105
4.4	Protein sequence of antibody Fab-MGD21 and its alignment to	106

	germ line LAIR1	
4.5	Crystal packing and order	109
4.6	Structural features of Fab-MGD21	110
4.7	The organisation of the CDRs of Fab-MGD21	115
4.8	Structure and polymorphism in the LAIR1 insertion	118
4.9	Comparison of the LAIR1m-containing antibody with other unusual antibodies	121
4.10	SPR experiments for the RIFIN binding assays	124
5.1	Structural comparison of CIDR domains	128
5.2	Sequence alignment of the CIDR domains	129
5.3	Structural comparison of CIDR α 2.8 domains at different pH values	130
5.4	A comparison of the features that allow CIDR α domains to bind to EPCR or CD36	133
5.5	Schematic representation of different antibodies	139
5.6	Structural comparison of cow antibody (BLV5B8, PDB ID: 4K3E) and human LAIR1m-containing antibody (PDB ID: 3KGR)	141

List of tables

3.1	Data collection and refinement statistics	65
3.2	The interactions between CD36 and CIDR α 2.8	73
4.1	Data collection and refinement statistics	103
4.2	A list of interactions between the LAIR1m insert and linkers that occupies the heavy chain CDR3 loop and the other five CDR loops of Fab-MGD21	114

List of abbreviations

ATS	acid terminal segment
BirA	biotin ligase
CD36	cluster of differentiation 36
CDR	complementarity determining region
CIDR	cysteine-rich interdomain region
DBL	duffy binding ligand
DC	dendritic cell
DMEM	Dulbecco's Modified Eagle Medium
Endo-F ₁	endoglycosidase-F ₁
EPCR	endothelial protein C receptor
FAT	fatty acid translocase
FBS	fetal bovine serum
GC/MS	gas chromatography coupled with mass spectrometry
HEK	human embryonic kidney
ICAM-1	Intercellular adhesion molecule-1
IE	Infected erythrocytes
Ig	immunoglobulin
IMAC	immobilised metal ion affinity chromatography
iRBC	infected red blood cells
ITIM	immunoreceptor tyrosine-based inhibitory motif
L-Glu	L-glutamine
LAIR1	leukocyte-associated immunoglobulin-like receptor-1
LB	Luria-Bertani
LIMP-2	lysosomal integral membrane protein 2
MR	molecular replacement
NEAA	non-essential amino acids
NTS	N-terminal segment
oxLDL	oxidised low-density lipoprotein
PBS	phosphate-buffered saline
PCR	polymerase chain reactions
PCT	pre-crystallisation test
PE	parasitized erythrocytes
PEG	polyethylene glycol
PEI	polyethylenimine
PEXEL	Plasmodium export element
<i>PfEMP1</i>	<i>Plasmodium falciparum</i> erythrocyte membrane protein 1
pir	<i>Plasmodium</i> interspersed repeat

RBC	red blood cell
RIFIN	repetitive interspersed family proteins
RU	Response Unit
SEC	size-exclusion chromatography
SEC-MALS	size-exclusion chromatography multi-angle light scattering
SH	somatic hypermutation
SICA	schizont-infected cell agglutination
SP	signal peptide
SPR	surface plasmon resonance
SR-BI	scavenger receptor class B member 1
TM	transmembrane
TSP	thrombospondin
UPS	upstream promoter sequence
VSA	variant surface antigens
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Malaria

1.1.1 An ancient disease with a global distribution

Malaria is a disease that has afflicted human beings for thousands of years. The name “malaria” is derived from Italian words meaning “bad air”, because it was believed that sickening “miasmas” carried illness and death in the air. The records of malaria-like symptoms go back as far as 2700 BC in China; but it wasn't until the late 19th century that the aetiology of malaria was identified (Cox, 2010) with the discovery that malaria is caused by *Plasmodium* parasites.

The association between the human race and malaria parasites has a very long history, with malaria reported in all continents except Antarctica (Hay et al., 2004), and it is human efforts that markedly control its distribution in space and time. Today, the disease has been geographically restricted to areas of poverty and to underdeveloped countries with climates suitable for transmission (**Figure 1.1**). Most reported cases occurred in sub-Saharan Africa, followed by South-East Asia and the Eastern Mediterranean. In 2015, the World Health Organisation (WHO) estimated that there were 212 million cases with approximately 429,000 deaths worldwide. However, some distortion in estimates of global malaria burden and cases could have resulted from inherent uncertainties in surveillance systems (Snow et al., 2005).

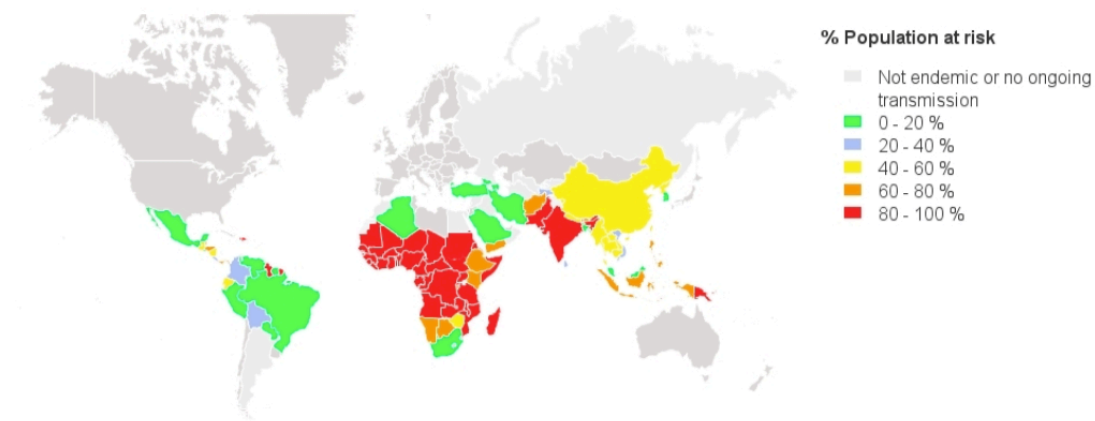


Figure 1.1. Population risk of malaria. Map created by the Medicines for Malaria Venture and the WHO Global Malaria Programme. Figure is adapted from <http://www.worldmalaria-report.org/node/68>

1.1.2 Malaria is caused by different species of *Plasmodium* parasites

Malaria is caused by infection with *Plasmodium* protozoans. After transfer through the bite of an infected mosquito, *Plasmodium* species can reproduce in various vertebrate hosts including humans, primates, birds, rodents, bats and so on. There are more than 100 species that comprise the *Plasmodium* family, of which five are relevant for human infection and these are *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*.

P. falciparum is the most deadly human-infective malaria species. It develops the highest complications and is responsible for most malaria deaths, especially of children under five, in countries in sub-Saharan Africa (WHO, 2016). *P. vivax* is less virulent than *P. falciparum*, but is the most geographically widespread species outside Africa and accounts for the major distribution of malaria in the world (Genton et al., 2008; Price et al., 2007;

Tjitra et al., 2008). *P. malariae* and *P. ovale* have a more benign status and are often found together with *P. falciparum* and *P. vivax* (Mueller et al., 2007). *P. Knowlesi* is traditionally considered a non-human parasite, but it has recently emerged a potentially life-threatening zoonotic form for human infection (Cox-Singh et al., 2008).

1.1.3 Life cycle

The malaria parasite has a very complex, multi-stage life cycle involving sexual and asexual stages within both its female Anopheles mosquito host and its human host (**Figure 1.2**). In general, the parasite species share a similar life cycle but important differences in their life cycles affects virulence and clinical outcomes.

Malaria (*Plasmodium* spp.)

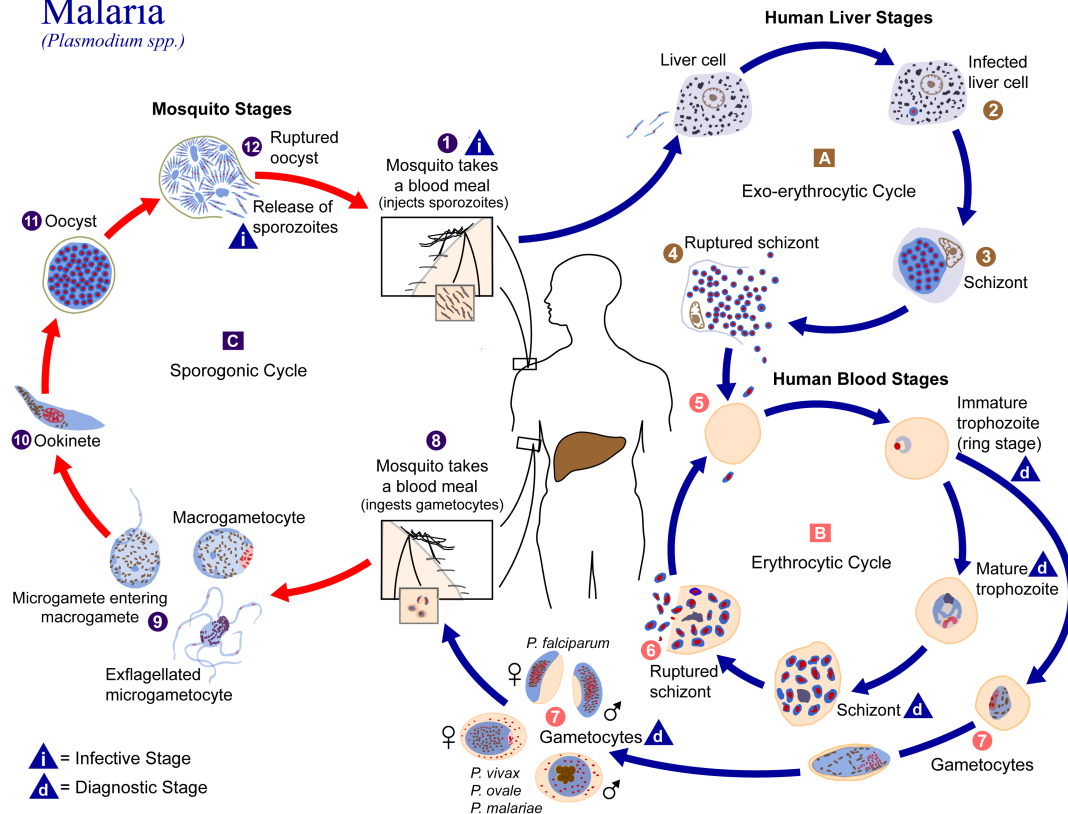


Figure 1.2. The generalised life cycle of *Plasmodium* parasites. The life cycle can be divided into human stages (A and B) and mosquito stage (C). By the bite of a mosquito, parasites enter the human bloodstream and migrate to the liver, where they replicate into merozoites. The merozoites infect red blood cells and produce more merozoites. Sexual forms of the parasites are also produced and are taken into a mosquito to start the life cycle in the insect. Figure is adapted from the Centers for Disease Control and Prevention (CDC).

When a malaria-infected female *Anopheles* mosquito takes a blood meal, she injects a particular stage of *Plasmodium*, the sporozoites, together with her saliva into the human bloodstream. A small proportion of the sporozoites then move in the circulatory system and make their way to target cells in the liver (Amino et al., 2006; Menard et al., 2013; Pradel et al., 2002). Once inside a hepatic cell, the sporozoites undergo rounds of replication and form mature

schizonts in the hepatic cell. When the schizonts ultimately rupture, they release thousands of the new stage of the *Plasmodium* parasite: the merozoites. In *P.vivax* and *P.ovale*, a few sporozoites lie dormant in their host hepatocytes forming what are known as hypnozoites for months to years. This “sleeping stage” is the cause of relapse malaria.

The merozoites are released into the circulating blood and rapidly invade red blood cells (RBCs) (Cowman and Crabb, 2006; Sturm et al., 2006), attaching and entering the cells by way of specific receptors on the erythrocyte surface. Each *Plasmodium* species has distinct predilections for invasion of different types of RBCs: *P.vivax* and *P.ovale* invade the youngest reticulocytes, *P. malariae* in mature RBCs, while *P. falciparum* in all RBC age classes (Kerlin and Gatton, 2013; Mons, 1990).

Once inside the erythrocyte, the parasites begin to differentiate into the feeding state, the trophozoites. At this point haemoglobin is digested and replication begins to create new intra-erythrocytic merozoites, enlarging the cell and forming another schizont (Francis et al., 1997; Marti et al., 2004). When the schizont ruptures, the new batch of merozoites are released from the burst erythrocytes; each then invades a new host erythrocyte and repeats the process again and again (Blackman, 2008).

After a number of replicative cycles, instead of progressing to trophozoites, a fraction of merozoites follow a different developmental route to become gametocytes (Talman et al., 2004). These gametocytes are the sexual stage

of the parasite and will perform no further function in the human host. Merozoites and gametocytes are transmitted from humans' blood into mosquitos. Once the gametocytes are taken up into the mosquito gut, the next stage of sexual reproduction takes place, which is known as the sporogonic cycle.

The gametocytes develop into female macrogametocytes (oocytes) and male microgametocytes (spermatocytes) (Billker et al., 1998; Vlachou et al., 2006), and then fuse to form the zygote. In the next 12-14 hours, the zygotes will elongate and change form into the ookinetes, which then invade the mosquito midgut epithelium and finally develop into the oocysts (Vlachou et al., 2004). Thousands of sporozoites can develop inside each oocyst, and egress from the mature oocyst. The sporozoites then migrate to the salivary glands of the mosquito where they are ready to begin another new human host infection (Akaki and Dvorak, 2005; Aly and Matuschewski, 2005; Rodriguez and Hernandez-Hernandez Fde, 2004). This completes the general *Plasmodium* life cycle.

1.1.4 Symptomatic and asymptomatic malaria

The symptoms of malaria first appear 6-14 days after infection, depending on the parasite species. When the merozoites are released from infected erythrocytes (IE), cellular debris and other pyrogenic toxins are released synchronously. This causes a significant decline in red blood cell numbers and elicits an immune response, resulting in development of acute clinical symptoms. Together with flu-like syndrome (chills, followed by fever and

sweating), a wide variety of clinical symptoms include headache, body aches, cramps, coughing, varying levels of mental disorientation, convulsions and coma.

Severe malaria is mostly caused by *P. falciparum*, although cases have also been reported that are due to *P. vivax* and *P. knowlesi* infections (TMIH, 2014). The criteria for severe malaria is established by WHO and the major complications include cerebral malaria, metabolic acidosis, hypoglycaemia, severe anaemia and renal impairment, involving nervous, respiratory, renal and/or hematopoietic systems (**Figure 1.3**) (WHO, 2012). In clinical practice, these complications can develop rapidly and progress to death within hours or days (Pasvol, 2005), and are particularly dangerous for children with naïve immune systems. Pregnant women with malaria also present adverse outcomes. The combination of compromised immune systems (mother) and placental infection leads to a high risk for low birth weight, pre-term delivery and infant mortality (Rogerson et al., 2007).

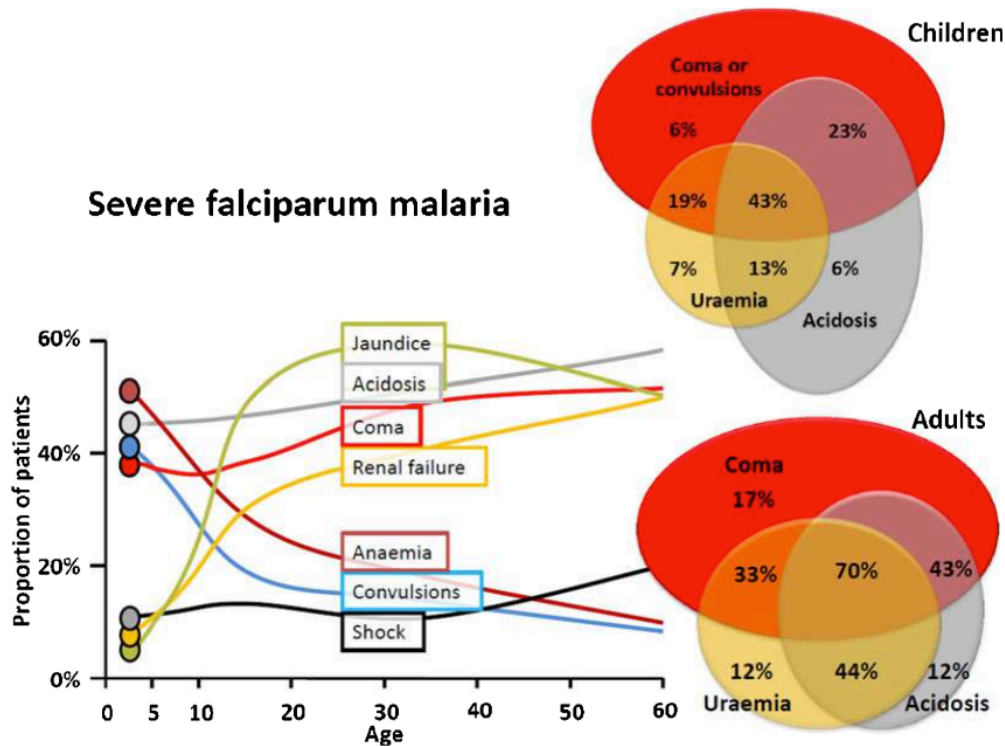


Figure 1.3. The complications of severe malaria in children and adults.

The left panel shows the prevalence of different features of severe falciparum malaria by age. Venn diagrams on the right show the mortality in children and adults associated with manifestations of cerebral and renal impairment and metabolic acidosis alone or in combination. The sizes of the ovals are proportional to number of cases. Figure is adapted from (TMIH, 2014).

In many malaria-endemic areas, more than 75% of infections are asymptomatic. Some of these can shift to symptomatic malaria within days or weeks while most remain asymptomatic for months to years (Bousema et al., 2014; Chen et al., 2016; Nzobo et al., 2015). The diagnosis and identification of cases of asymptomatic parasitemia is a major obstacle in its study. In general, individuals before receiving antimalarial treatment have detectable malarial parasitemia but in the absence of fever or other acute symptoms can

be categorised as having asymptomatic malaria (Laishram et al., 2012; Lindblade et al., 2013). Decades of studies shows that two types of immunity may result in asymptomatic malaria (Daubersies et al., 1996; Day and Marsh, 1991; Filipe et al., 2007; Trape et al., 1994): 1) Anti-disease immunity allows an individual to gain protection without symptoms. In areas with moderate to high levels of malaria transmission, adults and most children develop immunity to decrease symptomatic illness and reduce the possibility of severe malaria. 2) Anti-parasite immunity allows an individual to suppress parasite loads. With maturation of the immune system, infected individuals can limit parasite densities at very low levels following infection and therefore exhibit no clinical symptoms. Interestingly, asymptomatic malaria has been reported in both high and low transmission areas (Cerutti et al., 2007; Cucunuba et al., 2008; Harris et al., 2010).

Diagnostic clinical manifestations have been established in severe and complicated malaria and therefore form an integrated approach to reduce morbidity and mortality (WHO, 2012). On the other hand, for individuals with mild, uncomplicated or asymptomatic malaria, a lack of diagnostic criteria has become a worry for malaria control and elimination. It is not clear how much impact this may make, but it is worth putting efforts to investigate proper ways to improve the diagnosis of mild, uncomplicated and asymptomatic malaria.

1.2 Multi-gene family in malaria: variant surface antigens

In the vertebrate host, *Plasmodium* infection can last for long periods, increasing the possibility of transmission to a new host. In order to maintain prolonged chronic infection, the parasite requires a mechanism that facilitates interaction with the host for entry into the cell, allowing it to evade immune destruction. Antigenic variation in *Plasmodium* species was first suggested by Cox in the relapse studies of mice infected with *P. berghei* (Cox, 1959). Antigenic variation is a process by which an infectious agent (like a protozoan, bacterium or virus) changes its surface proteins, resulting in display of a new surface protein and in evasion of detection by host antibodies against the previously expressed surface protein to evade immune surveillance. A few years later, Brown and Brown demonstrated the first *Plasmodium* gene family, named *SICAvar* (*schizont-infected cell agglutination*), undergoing antigenic variation in *P. knowlesi*-infected monkeys (Brown and Brown, 1965). Subsequent discoveries of antigenic variation in *Plasmodium* species infecting humans (Biggs et al., 1991; Hommel et al., 1983), rodents (McLean et al., 1982) and other primates (Handunnetti et al., 1987) showed how switching variant surface antigens (VSAs) on the surface of parasite infected erythrocyte facilitates the evasion of immune response and the establishment of a chronic infection.

Plasmodium parasites possess an extensive and often species-specific array of VSAs throughout their complex life cycle. The major multi-gene family known to date is the *Plasmodium interspersed repeat (pir)* gene family, which has been identified in many *Plasmodium* species: *P. vivax* (vir, ~300 genes)

(del Portillo et al., 2001), *P. knowlesi* (*kir*, 68 genes) (Janssen et al., 2004; Pain et al., 2008), *P. yoelii* (*yir*, ~800 genes) (Carlton et al., 2002), *P. berghei* (*bir*, ~250 genes) (Hall et al., 2005), and *P. chabaudi* (*cir*, ~200 genes) (Janssen et al., 2002). Most members of this gene family are found in sub-telomeric locations with 6-90 % sequence identity at the amino acid level. The predicted structure suggests that PIR proteins comprise a long variable amino terminus followed by a trans-membrane region (Janssen et al., 2004). To date no known *pir* homologs have been identified in *P. falciparum*.

In *P. falciparum*, these multi-gene families include the *rif* (~150 genes), the *stevor* (~35 genes), the *var* (60 genes), the *surfin* (10 genes) and the *Pfmc-2TM* (12 genes) (Cheng et al., 1998; Gardner et al., 2002; Sam-Yellowe et al., 2004; Smith et al., 2001; Winter et al., 2005). Summaries of the features of the *var* and *rif* families are presented in section 1.2.2 and 1.2.3. *SICAvar* (28 genes) is another *P. knowlesi*-specific variant antigen gene family (al-Khedery et al., 1999). It has been shown that the *var* gene family in *P. falciparum* and the *SICAvar* family in *P. knowlesi* share many fundamental biological characteristics (Galinski and Corredor, 2004), despite these two species being phylogenetically distant.

Due to their large number and complexity, biological studies of VSAs are very challenging. In addition to immune-evasion, VSAs likely have other functions in iRBCs such as affecting host vesicular protein trafficking and altering cellular signalling pathway (Di Girolamo et al., 2008) and are associated with

clinical symptoms (Abdi et al., 2015; Fried and Duffy, 1996; Storm and Craig, 2014).

1.3 *P. falciparum* erythrocyte membrane protein1 (*PfEMP1*)

The best studied VSA gene family is the clinically important *var* gene family of *P. falciparum*, with approximately 60 *var* genes in each parasite genome (Gardner et al., 2002). These encode erythrocyte surface-expressed proteins known as *P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*) (Baruch et al., 1995; Leech et al., 1984; Smith et al., 1995; Su et al., 1995). In most cases the parasite only expresses one *PfEMP1* at a time to facilitate evasion of the host immune response. Each protein-coding gene is of the same basic two-exon structure: Exon 1 encodes the extracellular part of the protein and is composed of an N-terminal segment (NTS), varying combinations of Duffy binding ligand (DBL) and cysteine-rich inter-domain region (CIDR) domains, and a single transmembrane (TM) domain; Exon 2 encodes highly conserved cytoplasmic acid terminal segment (ATS) domain (**Figure 1.4**).



Figure 1.4. Schematic of the *var* gene locus. Exon 1 codes for the variable extracellular domain and transmembrane region, while exon 2 codes for a highly conserved intracellular region. UTR: untranslated regions.

The majority of *var* genes can be divided into groups according to their chromosomal location, upstream promoter sequence (UPS) and direction of transcription. Group A, Group B, and Group B/A *var* genes are found in the subtelomeric chromosome regions and transcribed in opposite directions. In contrast, Group C and Group B/C are typically found in internal regions

(**Figure 1.5**) (Gardner et al., 2002; Rask et al., 2010). Although the very extensive nucleotide diversity in exon1 makes the primary structure (the amino acid sequence) of the extracellular part of *PfEMP1* proteins differ in size and number, nearly all proteins have a modular architecture and contain a tandem DBL-CIDR domain, called the semi-conserved head structure, at the N-terminus. However, the sequence identity of the head structures is less than 50% within a given *var* repertoire (Kraemer et al., 2007). In small *PfEMP1*s (Group B, B/C and C), this head structure is followed by a second DBL-CIDR tandem domain to form a four-domain extracellular unit. In large *PfEMP1*s (Group A and B/A), there are additional DBL domains inserted before or after the C-terminal tandem (**Figure. 1.5**). With an extensive sequence analysis from several *P. falciparum* strains, *PfEMP1* domain families have been classified into DBL (α , β , γ , δ , ϵ , and ζ) and CIDR (α , β , γ) within different subgroups (Lavstsen et al., 2003; Smith et al., 2000b).

Recently, specific conserved architectural motifs known as domain cassettes were identified from the seven genomes analysis (Rask et al., 2010). Each domain cassette contains different combination of DBL and CIDR domains. These different domain cassettes are often retained as a unit during recombination and contain binding capacity for different ligands.

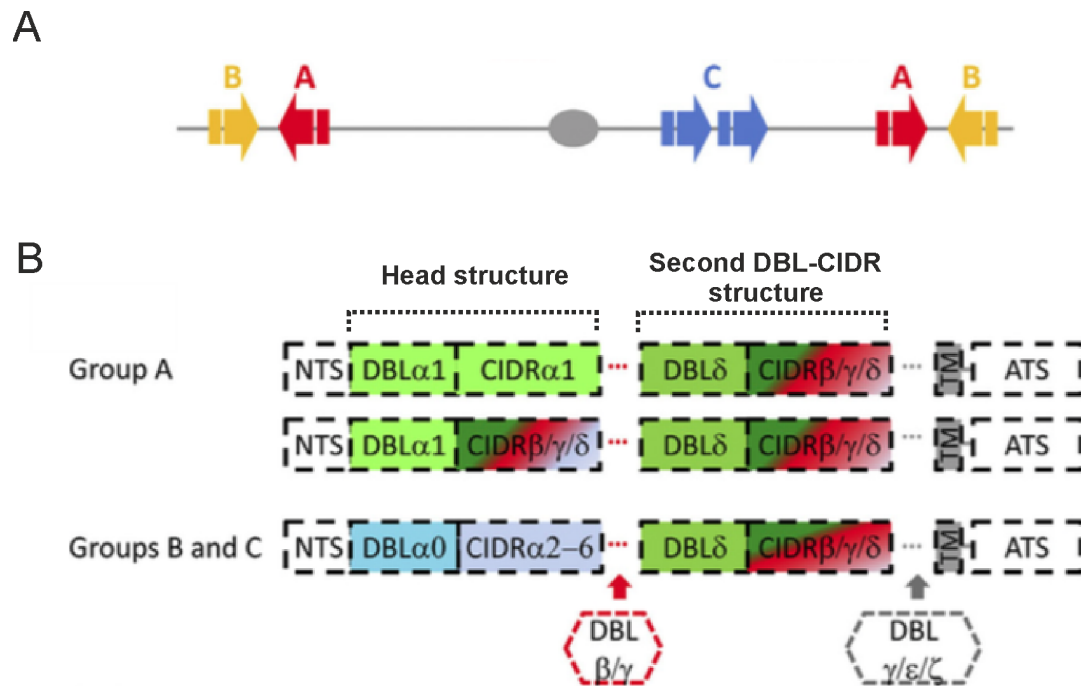


Figure 1.5. *var* gene organisation and *Pf*EMP1 proteins structure. (A) Chromosomal organisation of *var* genes. Group A and B *var* genes are located in subtelomeric regions, but transcribed in opposite directions. Group A genes lie towards the telomere, Group B and Group B/A *var* genes lie towards the centromere. Group C are typically found in central chromosomal regions. **(B)** *Pf*EMP1 proteins are built of different subtypes DBL and CIDR domains. Small *Pf*EMP1 proteins predominantly have a four-domain structure, whereas larger *Pf*EMP1 proteins have additional DBL domains after the head structure or the second DBL-CIDR structure. Figure is modified from (Hviid and Jensen, 2015).

1.3.1 Structure of CIDR and DBL domains

To date, there is no crystal structure of any full-length *PfEMP1* ectodomain, with only low-resolution data from small-angle X-ray spectrometry analysis in solution generating structural models for the arrangement of DBL and CIDR domains (Brown et al., 2013; Clausen et al., 2012; Srivastava et al., 2010; Stevenson et al., 2015).

The DBL domains are present in erythrocyte-binding proteins and *PfEMP1* proteins. Despite the remarkably low sequence identity of DBL domains, the crystal structures of DBL domains from invasion receptors (Batchelor et al., 2011; Hodder et al., 2012; Lin et al., 2012; Singh et al., 2006; Tolia et al., 2005) and *PfEMP1* proteins (Higgins, 2008; Khunrae et al., 2009; Vigan-Womas et al., 2012) present the same basic scaffold with a core helical architecture sharing a core set of conserved residues (**Figure 1.6**). These conserved residues are located in different subdomains (SD), stabilising interactions between helices and maintaining the core architecture (Higgins and Carrington, 2014).

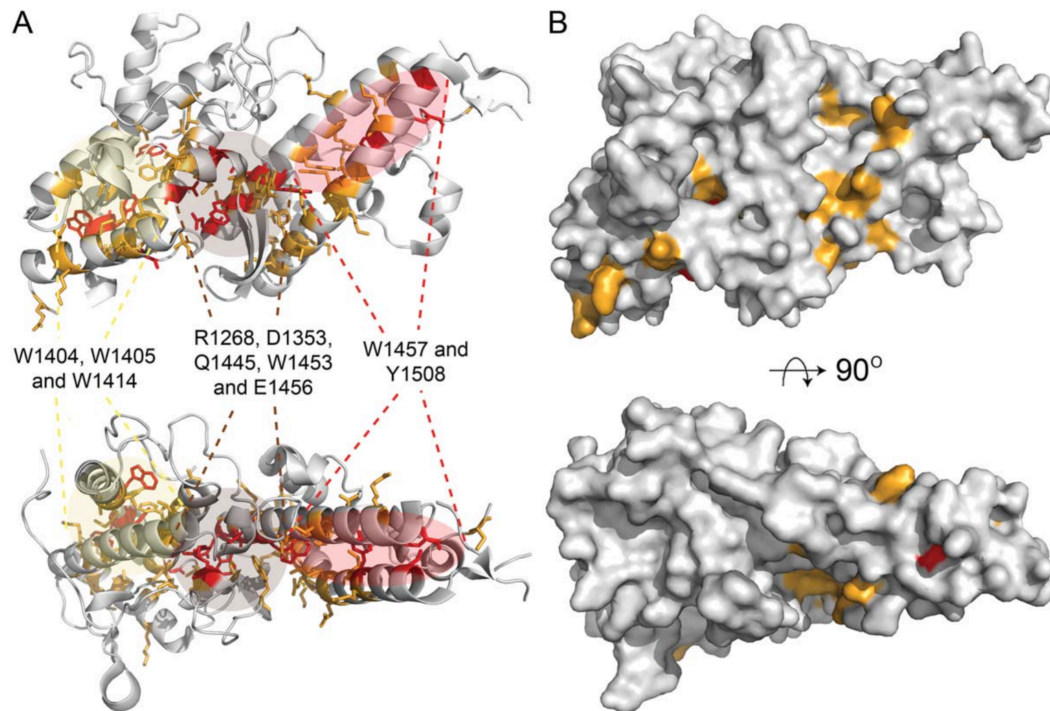


Figure 1.6. The DBL3X domain of var2csa. (A) Conserved residues are showing in red sticks and similar residues in yellow. **(B)** A surface representation revealed very few of these conserved residues are found on the protein surface. Figure is adapted from (Higgins and Carrington, 2014).

The CIDR domains are smaller and have been divided into CIDR α 1, CIDR α 2-6 and CIDR $\beta/\gamma/\delta$. Each CIDR domain is built up from a core three α -helical bundle with an insertion between the second and third core helices, that is perpendicular to the core bundle (Hsieh et al., 2016; Klein et al., 2008; Lau et al., 2015; Ng, 2012; Vigan-Womas et al., 2012). The sequence composition and length of the inserts are different and form part of a “homology block” used to classify CIDR domains into subgroups (**Figure 1.7**).

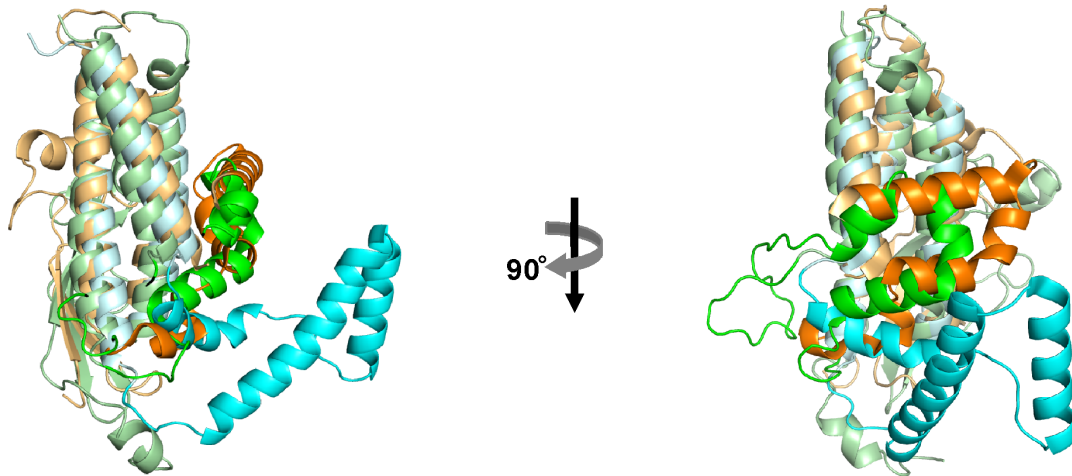


Figure 1.7. Structural alignment of CIDR domains. Superimposition of the CIDR domains from MCvar1 CIDR α 2.8 (cyan, PDB ID: 3C64), HB3var03 CIDR α 1 (orange; PDB ID: 4V3D) and *Pf*EMP1varO CIDR γ (green; PDB ID: 2YK0). Domains share core three α -helical bundles, and the homology blocks that emerge from between the second and third of these core helices are highlighted.

The CIDR α 2.8 domain of MCvar1 can be divided into three subdomains: M1, M2 and M3 (**Figure 1.8**) (Klein et al., 2008). M1 and M3 are the respective N- and C-terminal portions, while Klein *et al.*, determined the crystal structure of M2 domain, known as MC179, revealing an “opened” V-shaped architecture (**Figure 1.7 and 1.8B**). In contrast, the crystal structures of CIDR α 1 and CIDR γ domains revealed “closed” compact packing between the core three α -helical bundle and homology block (**Figure 1.7**) (Lau et al., 2015; Vigan-Womas et al., 2012). A further detailed comparison of CIDR domains is discussed in **Chapter 5**.

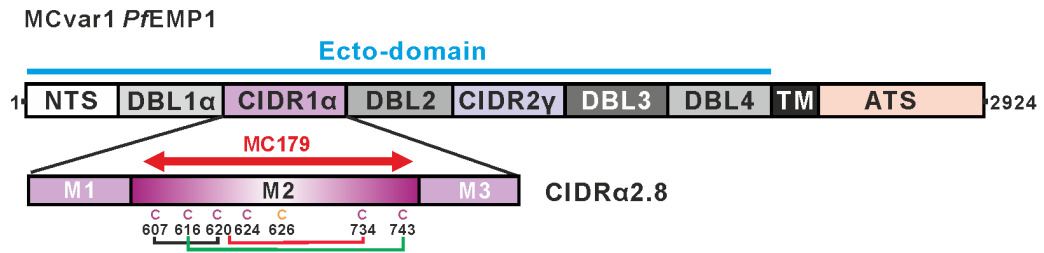
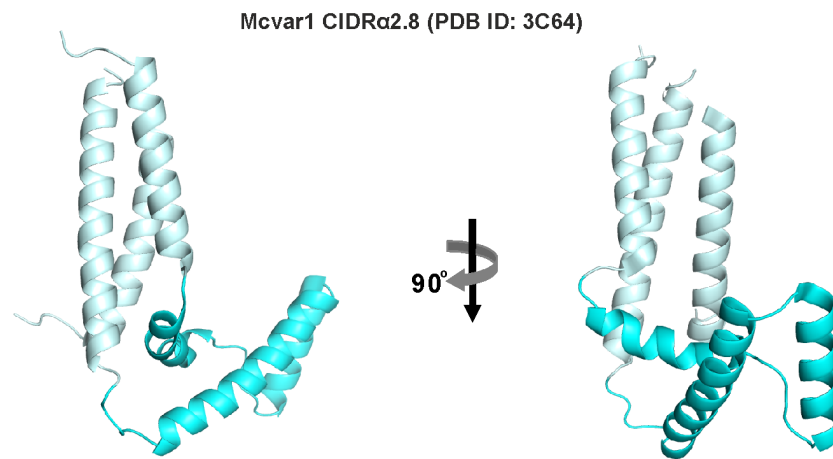
A**B**

Figure 1.8. Schematic diagrams of PfEMP1 and MC179 structure. (A) The MCvar1 PfEMP1 protein consists of four DBL and two CIDR domains. MC179 contains about two thirds of the CIDR1 α domain and is a minimum domain for CD36 binding. **(B)** The crystal structure of MC179 (PDB ID: 3C64).

1.3.2 The endothelial receptors of *PfEMP1*

In addition to varying to evade immune recognition, *PfEMP1* has a distinctive feature that contributes to its important role in malarial disease called cytoadherence, which causes infected erythrocytes to sequester in the vasculatures. Sequestration allows parasites to avoid spleen-dependent killing mechanisms, increasing parasitaemia and aiding transmission to mosquitoes (David et al., 1983; Hayward et al., 1999; Hommel et al., 1983). Previous studies show that *PfEMP1* act as adhesion ligands for both uninfected erythrocytes and receptors on endothelial cells and indicate that binding specificity is related to domain content. The ability of *PfEMP1* to bind adhesion receptors on the surfaces of uninfected erythrocytes leads to the formation of rosettes, which are clumps of uninfected erythrocytes bound to an infected erythrocyte (Adams et al., 2014). The function of rosetting is unclear but it has been associated with severe malaria in some studies (Carlson et al., 1990; Kun et al., 1998; Rowe et al., 1995).

1.3.2.1 The interaction of *PfEMP1* with ICAM-1 and EPCR

PfEMP1 variants mediate cytoadhesion to different receptors. Intercellular adhesion molecule-1 (ICAM-1) and endothelial protein C receptor (EPCR) are two endothelial receptors associated with cerebral and/or severe malaria (Baruch et al., 1996; Smith et al., 2000a; Turner et al., 2013). ICAM-1 is an endothelial- and leukocyte-associated glycoprotein, which is important in cell signalling (Dustin et al., 1986). It was shown that DBL β domains from *PfEMP1* proteins adhered specifically to ICAM-1; and ICAM-1 binding by group A *PfEMP1* proteins is associated with severe disease while that mediated by

group B and C *PfEMP1* is associated with uncomplicated malaria (Bengtsson et al., 2013; Howell et al., 2008; Janes et al., 2011).

EPCR is a glycoprotein known for its essential role in the activation of protein C (Laszik et al., 1997). In recent studies, domain cassettes 8 and domain cassettes 13 from *PfEMP1* variants are found to contain EPCR-binding CIDR α 1 domains that are associated with severe malaria (Turner et al., 2013). More recently, the *PfEMP1* containing both EPCR-binding CIDR α 1 and ICAM-1 binding DBL β domains were found to be expressed more frequently in cases of cerebral malaria (Avril et al., 2016; Francis et al., 2017).

1.3.2.2 The interaction of *PfEMP1* with CD36

In contrast to ICAM-1 and EPCR, *PfEMP1* variants binding to cluster of differentiation 36 (CD36), also known as platelet glycoprotein 4, fatty acid translocase (FAT), scavenger receptor class B member 3 (SCARB3) glycoprotein 88 (GP88), IIIb (GPIIb), or IV (GPIV), are associated with uncomplicated malaria. The scavenger receptor, CD36 is the host receptor most commonly found to interact with parasite isolates from patients (Heddini et al., 2001; Newbold et al., 1997; Rogerson et al., 1999; Udomsangpetch et al., 1997) and with laboratory-adapted parasite strains (Janes et al., 2011). In the case of the domain families, many domains have been tested and only CIDR α 2-6 domains bind CD36 (Robinson et al., 2003; Smith et al., 1998). The *PfEMP1* domain region that binds to CD36 has been determined from the Malayan Camp (MC) strain of *P. falciparum* (Baruch et al., 1997). Experiments identified a 179-residue region in the MCvar1 CIDR α domain

(MC179, residue 676-753) that is the minimal region for CD36 binding (**Figure 1.8**) (Baruch et al., 1997). Binding to CD36 is associated with uncomplicated malaria and antibodies against CD36 can disrupt the interaction between MC179 and CD36 to prevent early infection (Baruch et al., 1997; Gowda et al., 2013).

1.3.2.3 Human cluster of differentiation 36 (CD36)

CD36 is a broadly expressed protein presented on various types of cells of the cardiovascular system (Silverstein and Febbraio, 2009) and a type II membrane protein containing 472 amino acids with two transmembrane helices and both N- and C-termini in the cytoplasm (**Figure 1.9**). The N-terminus is only seven residues long, and the C-terminal tail is thought to be the signal transduction site (Rahaman et al., 2006). The large extracellular domain is heavily glycosylated with 10 putative N-linked glycosylation sites, which are shown to be necessary for trafficking onto the cell membrane (Hoosdally et al., 2009). Six of the ten cysteine residues lie in the extracellular domain and form three disulphide bonds, which are required for the CD36 transport from endoplasmic reticulum to Golgi (Gruarin et al., 1997). Crystal structures of the CD36 homolog, LIMP-2 (Lysosomal integral membrane protein II) at different pHs, enabled modeling of the CD36 ectodomain (Dang et al., 2014; Neculai et al., 2013; Zhao et al., 2014). In **Chapter 3**, the detailed crystal structure of CD36 is revealed.

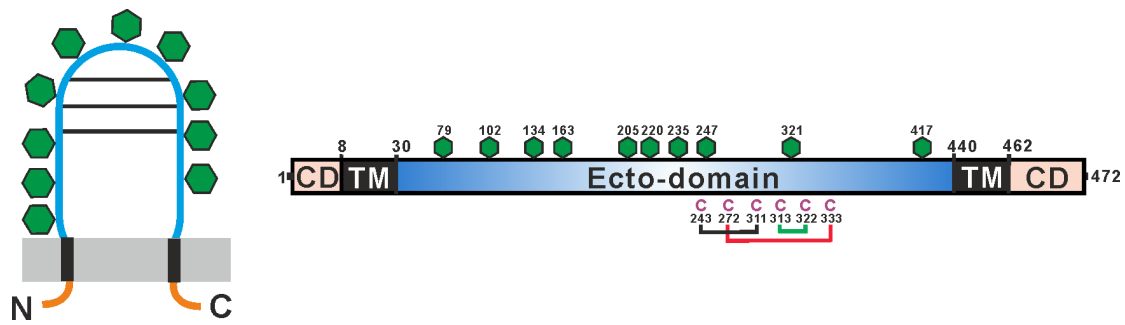


Figure 1.9. Schematic representation of CD36. CD36 has short intracellular domains at both termini, two transmembrane helices and a large extracellular domain with 10 putative N-linked glycosylation sites (green hexagon). The disulphide bonds are shown as black, red and green lines.

1.3.2.4 Natural ligands of CD36

CD36 is a multi-ligand receptor involved in many physiological processes including the immune response, fatty acid metabolism, and *Plasmodium* interaction. Currently known ligands include collagen, thrombospondin (TSP), hexarelin, oxidised phospholipid, oxidised low-density lipoprotein (oxLDL) and long-chain fatty acid. Despite decades of investigation, CD36 remains a mysterious receptor with distinct proposed roles in diverse cell types. In this thesis the binding of CD36 to two physiological ligands (TSP and oxLDL) has been investigated.

TSPs are a family of five secreted glycoproteins (TSP homolog 1 to 5) with multiple functions, including angiogenesis, tissue organisation and synaptogenesis (Adams and Lawler, 2011; Baenziger et al., 1971). At the molecular level, TSP homolog 1 (TSP1) interacts with CD36 and plays a critical role in the regulation of angiogenesis. Residues 93-120 of CD36, called the CLESH domain, were been shown to interact with residues 447-452

of TSP1 (Asch et al., 1992; Dawson et al., 1997; Frieda et al., 1995). TSP1 binding to CD36 inhibits angiogenesis by triggering apoptosis. Pharmacologic and genetic knockout experiments have demonstrated that the inhibition of angiogenesis requires CD36 activation by inhibiting growth factor-induced proangiogenic signals, instead of generating anti-angiogenic signals (Rege et al., 2009; Volpert et al., 2002). TSP1 has also been found to play an essential role in platelet activation and aggregation. In the blood vessel, CD36 contributes to the Ca^{2+} -dependent adhesion of platelets to TSP1 and can induce platelet activation (Agbanyo et al., 1993; Bonnefoy et al., 2001; Roberts et al., 2010).

CD36 has also been shown to bind to oxLDL (Endemann et al., 1993). In CD36-null mice, macrophage binding and oxLDL uptake are significantly reduced (Febbraio et al., 1999). The CD36-null mouse showed less atherosclerotic lesion formation, suggesting that CD36 is likely to be a key determinant of macrophage foam cell formation and atherosclerosis (Febbraio et al., 1999). The mechanism of macrophage expression of CD36 and foam cell formation may be perpetuated by a cycle in which internalised oxLDL within macrophages provides specific oxidised lipids as CD36 ligands to activate the transcription factor, PPAR- γ (peroxisome proliferator activated receptor- γ), and up-regulates CD36 expression by releasing an “eat me signal” (Nagy et al., 1998; Tontonoz et al., 1998). This induces further uptake of oxLDL via CD36. PPAR γ is a member of a nuclear hormone superfamily that heterodimerize with the retinoid X receptor (RXR). These proteins are transcriptional regulators of genes encoding proteins involved in adipogenesis

and lipid metabolism (Tontonoz et al., 1995). This oxLDL neutralisation also induces cytokines secretion and establishes atherosclerotic diseases (Jiang et al., 2012).

1.3.2.5 CD36: a friend or enemy in malaria

Most malaria-infected individuals show asymptomatic or uncomplicated symptoms, with only 1-2% developing severe disease. It is therefore reasonable to speculate that host immunity towards either protection or pathology might be in balance to permit parasite propagation while also maintaining host survival, thereby prolonging infection and increasing the likelihood of transmission of the parasite through mosquito bite. CD36 is abundantly expressed in skin, muscle and adipose tissue and is a major receptor for adhesion by most parasitised erythrocytes (PEs), which could help to create an environment for parasite replication and result in iRBCs sequestration for parasite survival by avoiding splenic clearance. Meanwhile, CD36 expressed on monocytes, macrophages and dendritic cells (DCs) causes a stress-like effect, enhancing defensive responses within the immune system.

CD36 expressed on various immune cells has been shown to bind infected erythrocytes. Monocytes are derived from haematopoietic stem cells and can differentiate into either macrophages or DCs (Dalton et al., 2014). In malaria-infected hosts, monocyte and macrophage-mediated phagocytosis contribute to the early control of parasite replication, especially in non-immune children. Before antigen-specific immune responses occur, CD36 plays an important factor by functioning as a pattern recognition receptor (PRR) expressed onto

monocytes and macrophages to identify PEs for innate clearance (McGilvray et al., 2000; Smith et al., 2003). Phagocytic uptake of PEs was strongly inhibited by cleavage of the CD36-binding PfEMP1 on the surface of PEs or by functional blockage of CD36, while macrophages lacking CD36 were shown to phagocytose significantly fewer PEs (McGilvray et al., 2000; Smith et al., 2003). On the other hand, haemozoin (a by-product degraded from haemoglobin by the parasite in the iRBC) accumulating within macrophages that have ingested PEs leads to macrophage dysfunction and reduced phagocytosis (Schwarzer et al., 2003), subsequently decreasing the efficiency of innate immunity.

PEs also interact with DCs which are responsible for the initiation of adaptive immune and the amplification of innate immune responses. Immature DCs internalise PEs, which induces DC maturation. This is characterised by antigen presentation, naïve T cell activation, cytokine production and enhancement of antibody production (Liu, 2001). The capability of innate and adaptive immunity is then stimulated to reduce numbers of infected parasites. Meanwhile, by interacting with CD36 on DCs, *P. falciparum*-iRBCs inhibit DC maturation and subsequently delay the development of adaptive immunity (Urban et al., 1999; Urban et al., 2001), although the modulation of DCs is not necessarily applicable to other *Plasmodium* species (Seixas et al., 2001).

The apparently varied roles of CD36 involving in interacting with cells of the immune system are very complex (Cabrera et al., 2014). What is addressed above is a drop in the ocean. Understanding the mechanisms by which CD36

affects these diverse functions is a challenge and important topic for therapeutic strategies. However, what is known suggests that CD36 binding might allow the parasite to be maintained at a level which allows parasite survival and transmission without putting the host at risk of severe malaria.

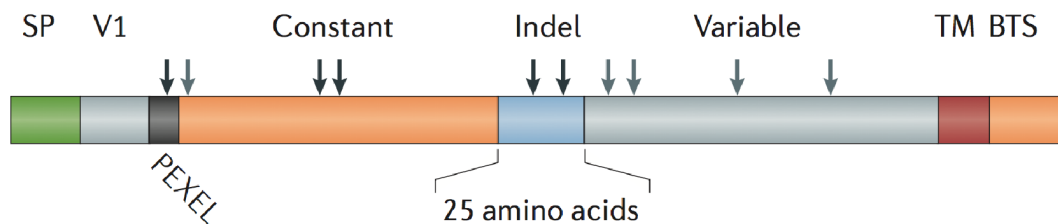
1.4 Repetitive interspersed family proteins (RIFINs)

The *rif* genes constitute the most abundant multi-gene family in *P. falciparum* with ~150 genes in the 3D7 genome (Gardner et al., 2002). They are present at the subtelomeric regions and encode a group of variant RIFIN proteins of about 27-45 KDa in size, which are expressed on the surface of iRBC (Fernandez et al., 1999; Kyes et al., 1999). In contrast to *PfEMP1*, multiple RIFINs can be expressed on iRBC surfaces simultaneously. Sequence analysis of RIFIN proteins revealed two major subgroups in this family: RIF_A and RIF_B (Joannin et al., 2008; Pizzi and Frontali, 2001). Both share a similar architecture: a signal peptide at the N-terminus, a *Plasmodium* export element (PEXEL) motif which is necessary for targeting protein outside cell (Horrocks and Muhia, 2005) and one transmembrane helix. The main difference is due to a 25-amino acid (25 AA) insert downstream of the PEXEL motif only in RIF_As (**Figure 1.10**) (Goel et al., 2015). Interestingly, recent evidence demonstrated that RIF_As and RIF_Bs have different sub-cellular localisations (Bultrini et al., 2009; Petter et al., 2007). RIF_As are trafficked to the iRBC membrane, while RIF_Bs remain confined inside the parasite.

Recent evidence also revealed that RIFINs are also expressed in other development stages (merozoites, sporozoites and gametocytes) (Bachmann et al., 2012; Florens et al., 2002; Mwakalinga et al., 2012; Petter et al., 2007; Wang et al., 2010); which suggests that antigenic variation is not restricted to the IE surface and RIFINs may have multiple roles in different parasite life cycle stages. Recent studies showed that RIF_As are able to mediate rosetting by the binding of the blood group A antigen on RBCs and the RIFIN-

mediated rosettes appear to protect *PfEMP1* from antibody recognition (Goel et al., 2015; Moll et al., 2015).

RIF_A



RIF_B

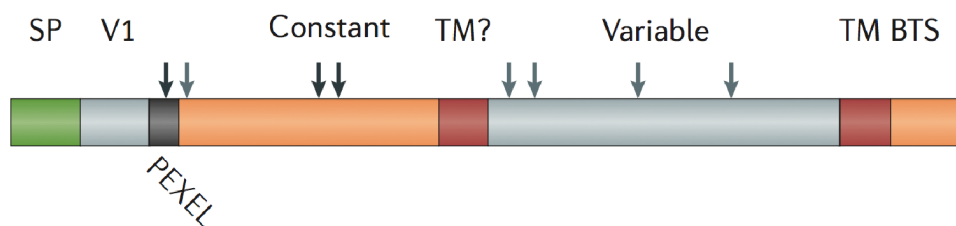


Figure 1.10. Schematic representation of RIFIN proteins. Both RIFIN subgroups share a similar secondary structure: A signal peptide (SP) at N-terminus, a PEXEL motif and a transmembrane (TM) region. There is an additional insertion of 25 amino acid (25 AA) in RIF_A. A second TM may be present in some RIF_B. Common conserved cysteine residues are shown in grey arrows and subtype-specific cysteine residues are in black arrows. V1: first variable domain. BTS: basic terminal sequence. Figure is adapted from (Wahlgren et al., 2017).

1.5 Antibody structure and function

Antibodies are a globular protein, also known as immunoglobulins (Ig), which are used to neutralise pathogens in the adaptive immune system. Antibodies are presented either as a membrane-bound form attached to the surface of B cells or as a soluble form secreted to be free in the blood plasma. All antibodies share certain structural features, search and neutralise antigen for elimination, and participate in a limited number of effector functions.

1.5.1 Basic antibody structure

A natural antibody is made up of four peptide chains: two heavy chains (H) and two light chains (L). Each light chain is connected to a heavy chain by a disulphide bond, and the two heavy and light chains (H-L) are linked by disulphide bonds and noncovalent interaction to form a basic “ Y-shaped ” molecule, the dimer $(H-L)_2$ (**Figure 1.11**). The light chain is a small polypeptide of about 25,000 Daltons molecular weight, composed of one variable domain (V_L) and one constant domain (C_L). There are two types of constant domain: lambda (λ) and kappa (κ); and only one of them is present in a single antibody molecule. The heavy chain is a larger polypeptide having a molecular weight of 50,000 Dalton or more, composed of one variable domain (V_H) and three constant domains (C_{H1} , C_{H2} , C_{H3}) subgrouped into five distinct classes: alpha (α), gamma (γ), delta (δ), epsilon (ϵ) and mu (μ). Antibodies can be classified by their isotypes, determined by the constant domain of the heavy chain and are grouped into IgA (α), IgG (γ), IgD (δ), IgE (ϵ) and IgM (μ).

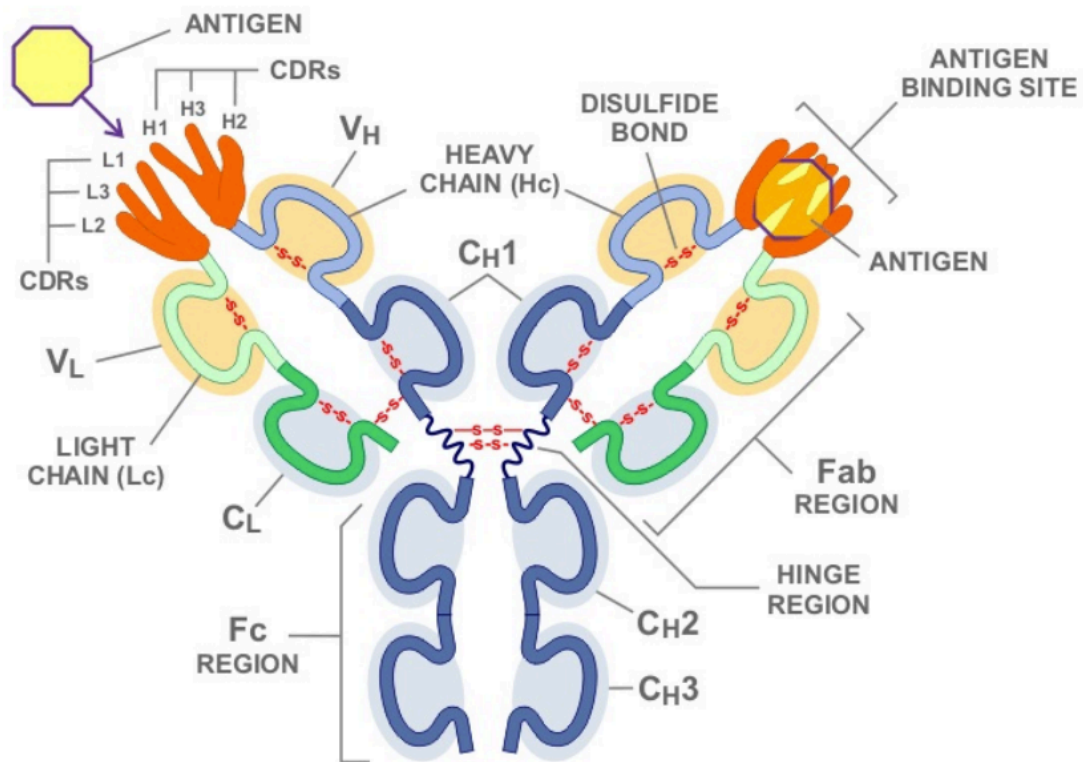


Figure 1.11. Schematic diagram of immunoglobulin structure. Each heavy and light chain contains variable (V_H for heavy chain, V_L for light chain) and constant regions (C_H for heavy chain, C_L for light chain). The antigen binding sites are found on the six CDRs (L1, L2 and L3 on light chain; H1, H2, and H3 on heavy chain) at the tip of variable regions of heavy and light chains.

Figure is adapted from <https://teaching.ncl.ac.uk/bms/wiki/index.php/Antibody>

1.5.2 The antigen-binding site and its diversity

The tip of the antibody “Y” contains an antigen-binding site that is composed of one constant and one variable domain of each of the heavy and the light chain, called the Fab fragment (fragment, antigen binding). The stem of the “Y”, called the Fc fragment (fragment, crystallisation) is able to interact with cell surface receptors of the immune system (**Figure 1.11**). The Fab fragment has unique antigen binding specificity, and only a very small part of the variable domain (V_H and V_L) is actually involving in antigen binding. The majority of the affinity and specificity of antibodies is mediated by the tips of six hypervariable regions (H1, H2, H3 from the heavy chain and L1, L2, L3 from the light chain). Since these are complementary to the surface of the antigen that the antibody will bind, they are also called complementarity-determining regions (CDRs).

Antibody diversity is mediated by two types of DNA altering mechanism: V(D)J recombination and somatic hypermutation (SH) (Fugmann et al., 2000; Kato et al., 2012; Schatz and Swanson, 2011; Smider and Chu, 1997). V(D)J recombination, named for the V (variable), D (diversity) and J (joining) coding gene segments, occurs through assembly of the germline DNA due to site-specific cleavage and end-joining reactions which generate the V-(D)-J-C coding unit for gene translation (**Figure 1.12**). After exposure to new antigens, the immune system then undergoes “somatic hypermutation” with mutations occurring in the variable regions. Within the six CDR loops, CDR H3 is normally the most variable in both sequence and length, and therefore plays a crucial role in the antigen binding.

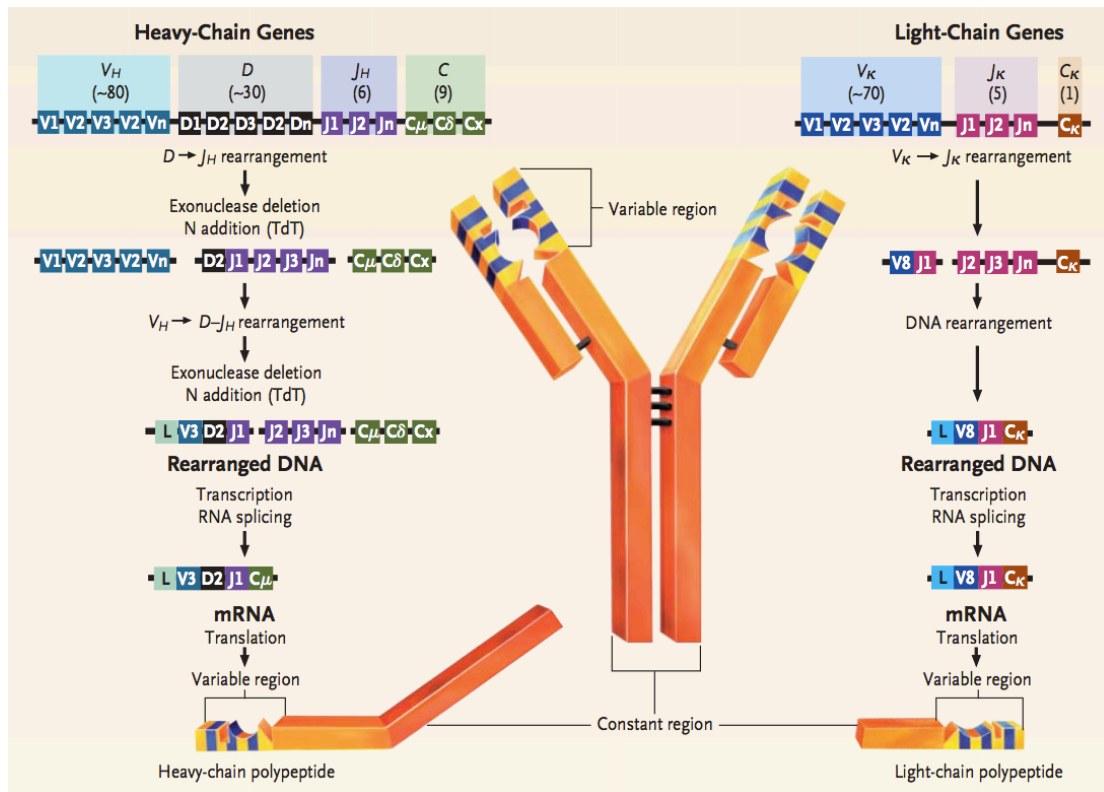


Figure 1.12. The generation of antibody diversity. V(D)J recombination in B-cell development. Germline DNA contains many different variable (V), diversity (D) (in the case of the heavy chain only) and joining (J) gene segments. The process begins by the random joining of one D segment to a J segments (D-J unit) followed by addition of a V gene to the D-J unit. After rearrangement, the composite V-D-J trio is then brought together with the constant gene (C) to form the V-D-J-C coding unit of the heavy chain. Similar, random rearrangement of V-J followed by joining to a constant gene forms the V-J-C coding unit of the light chain. The antigen receptors of T cells use a similar process to generate diversity. Figure is adopted from (Schwartz, 2003).

Despite this genetic diversity, the typical antibody structure is maintained within the Ig fold, and the associated CDRs come together to form a relatively flat binding surface. Today, high-resolution X-ray crystallographic studies provide valuable insight into the structure of CDRs and the interactions

between antigens and antibodies. In general, CDRs are supported by the conserved scaffold of the variable domains and most or all six CDRs usually contact the antigen. Although CDRs vary in sequence and size, five of them (all except CDR H3) have one of a small number of main chain conformation termed as “canonical structure”, which is determined by the length of the loop and the residues present at key sites (Al-Lazikani et al., 1997; Chothia et al., 1989). CDR H3, on the other hand, has the largest variety in its length, sequence and structure. CDR H3 therefore contributes more to antigen recognition and sometimes changes its conformation significantly upon antigen binding (Stanfield and Wilson, 1994). Interesting, a subset of antibodies in cow (~10% of the total antibody pool) was found to have “exceptionally long” CDR H3 regions, with length up to nearly 70 residues (Saini et al., 1999). Crystal structures of these ultra long CDR H3 show that they fold into a long stalk on top of which sits a knob domain that protrudes from the antibody surface, forming a very unusual architecture that can accommodate significant structural variation (Wang et al., 2013). Mutating residues in the knob greatly influences antigen binding and removal of the knob completely abolishes binding. It is clear that the knob domain is essential for antigen binding but it is not clear what role the stalk and the other five CDRs play as there is no structure of such an antibody bound to antigen.

1.5.3 A LAIR1-containing antibody

In malaria endemic areas, anti-RIFIN antibodies were detected in the majority of the adult population whereas the prevalence of such antibodies was much lower in children (Abdel-Latif et al., 2004; Abdel-Latif et al., 2003; Abdel-Latif et al., 2002). It was shown that anti-RIFIN antibodies could cross-react with several RIFIN proteins in many parasite isolates, suggesting that the generation of the anti-RIFIN antibodies repertoire is likely to be stable over time. In addition, elevated levels of anti-RIFIN antibodies in children were associated with rapid parasite clearance. Recently, a subset of unusual human antibodies was identified in malaria-endemic areas (Tan et al., 2016). These antibodies adopted a human receptor protein, leukocyte-associated immunoglobulin-like receptor-1 (LAIR1), into the antigen-binding site to generate broad reactivity. This kind of immune response appears to provide a certain degree of protection against malaria, and is detected in milder/asymptomatic infectious individuals. In **Chapter 4**, the crystal structure of the Fab-MGD21 fragment containing a LAIR1 insert in the CDR H3 region is presented; which reveals an even more remarkable architecture of the CDR H3 regions, providing new ideas in terms of antibody engineering and vaccine development.

1.6 Leukocyte-associated immunoglobulin-like receptor-1 (LAIR1)

LAIR1 is a member of the immunoglobulin (Ig) superfamily receptor that plays a role in controlling immune cells to prevent improper activation or over-activation. In the immune inhibitory receptors, immune-receptor tyrosine-based inhibitory motifs (ITIMs) play an important role in regulating the immune system. Phosphorylation of tyrosine residues in the ITIMs of LAIR1 can deliver a negative signal to inhibit cellular functions of natural killer (NK) cells, effector T cells, B cells and DC precursors. Unlike most immune inhibitory receptors, which are restricted in their expression, LAIR1 is expressed on most immune cells including NK cells, T cells, B cells, monocytes, and CD34⁺ hematopoietic progenitor cells (Meyaard, 2008; Meyaard et al., 1997; Meyaard et al., 1999; Ouyang et al., 2003; Poggi et al., 1998).

Collagen is the most abundant protein in humans, important in maintenance of tissue integrity and hemostasis. LAIR1 recognises a common collagen motif, the Gly-Pro-Hyp repeats, with a nano-molar affinity, and this peptide was shown to inhibit immune cell activation *in vitro* (Lebbink et al., 2006). In addition, LAIR1-collagen interactions also occurs at different differentiation or activation stages, implying that collagens act as functional ligands that modulate the role of LAIR1 in immune regulation (Jansen et al., 2007; Lebbink et al., 2006; Lebbink et al., 2007).

LAIR1 is a type-I transmembrane protein of 287 amino acids containing an extracellular C2-type Ig-like domain followed by a transmembrane domain and

two cytoplasmic ITIMs that relay the inhibitory signal. As shown in **Figure 1.11**, the crystal structure of human LAIR-1 collagen-binding domain (hLAIR1-CBD, residue 22-165, PDB ID: 3KGR) together with NMR mapping and site-directed mutagenesis, identified key binding residues that are important for collagen binding (Brondijk et al., 2010). Residues R59, E61 and W109 form a conserved surface patch on LAIR1 that is essential for collagen binding, while mutants R59A, E61A and R65A showed a strongly reduced collagen binding (**Figure 1.13**).

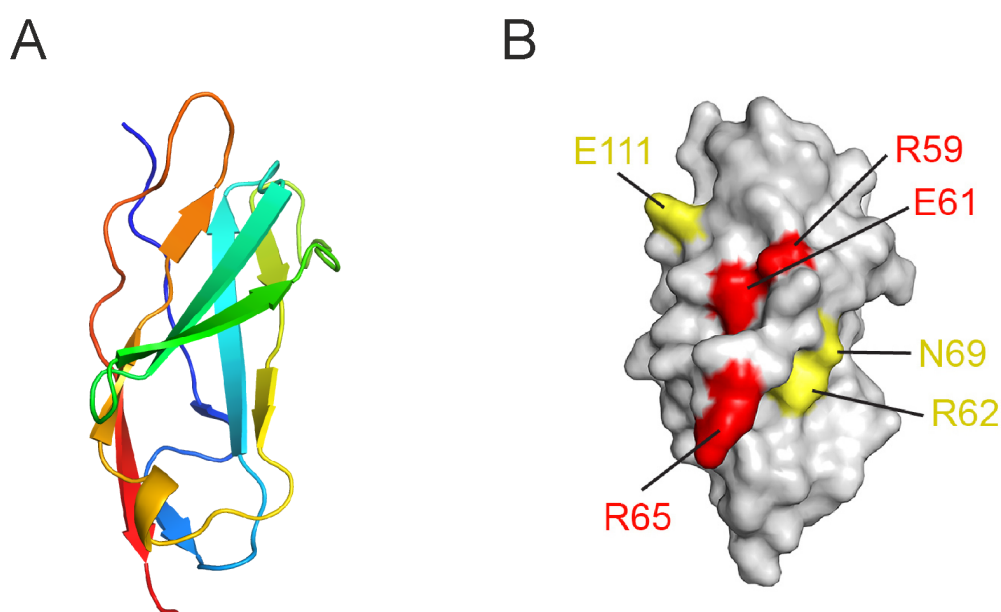


Figure 1.13. LAIR-1 structure. (A) hLAIR1-CBD (PDB ID: 3KGR) is drawn in ribbons and coloured in rainbow colours from N-terminus (blue) to C-terminus (red). (B) A surface representation of the structure of LAIR1 (grey) with residues whose mutation has a major (red) or minor (yellow) effect on collagen binding highlighted.

1.7 Aims

The first aim of this thesis is to determine how *Pf*EMP1 CIDR α domains bind to CD36 by using X-ray crystallography. Together with biophysical methods as well as *in vivo* assays, I will reveal the molecular determinants in *Pf*EMP1 proteins that allow the interaction with CD36 and provide insight into the molecular basis of this interaction. I also want to understand how CD36-binding is so prevalent and why the host cannot avoid this binding. The second aim of this thesis is to determine a high-resolution crystal structure of one of the unusual human LAIR-1-containing antibody structures. The novel antibody structure reveals a molecular detail that can provide inspiration not only in the malaria field but also in its biomedical application.

Chapter 2

Materials and methods

2.1 Construct design and molecular cloning

The mammalian expression vectors, pHLsec and pHLsec-BAP, are gifts from Dr. A. R. Aricescu (University of Cambridge, UK). The C-terminal 8xHis tag modified vector, pHLsec-8H, was derived from pHLsec and kindly provided by Dr. T.-H. Chang (University of Oxford, UK) (**Appendix 1**). Vectors pOPIN VH and pOPIN VL for Fab protein expression are gifts from Dr. Ray Owens (Oxford Protein Production Facility-UK). All synthetic complementary DNA clones are codon-optimised for expression in mammalian cells and obtained from GeneArt (ThermoFisher). All primers used in the thesis are described in **Appendix 2**.

Polymerase chain reactions (PCR) were carried out using KOD HotStart polymerase (Merck Millipore) to amplify DNA fragments on an MJ Mini Personal Thermal Cycler (Bio-Rad, USA), according to the manufacturer's protocol. Successful amplification of DNA fragments was verified by agarose gel electrophoresis followed by DNA gel extraction and purification by QIAquick Kit (QIAGEN). Purified PCR products and vectors were digested with the corresponding restriction enzymes (New England BioLabs) using the recommended buffer and temperatures. Ligations were performed with insert and vector in approximately a 3:1 ratio using the Quick Ligation Kit (New England BioLabs). Ligation mixtures were transformed into DH5 α library efficient competent cells (ThermoFisher) according to the manufacturer's protocol. Transformed cells were spread on Luria-Bertani (LB) medium-agar plates containing 50 μ g/ml ampicillin and were incubated at 37°C for 16 hrs. Colonies for each construct were picked, grown in 10 ml of LB media

overnight at 37°C and used to prepare plasmid DNA using QIAprep Spin Miniprep Kit (QIAGEN) for DNA sequencing verification (Source Bioscience).

2.1.1 Constructs for CD36 and MCvar1 CIDR α 2.8

A gene encoding the human CD36 ectodomain (UniprotKB/Swiss-prot ID: P16671; residues 35-439) was inserted into the pHLsec-8H vector. A synthetic complementary DNA clone of MCvar1 CIDR α 2.8 (UniprotKB/Swiss-prot ID: Q25733; residues 576-745), with Cys626 mutated into Ser to remove an unpaired cysteine, was cloned into pHLsec-8H. CD36 constructs used for surface plasmon resonance (SPR) experiments were created in the pHLsec-BAP vector, containing a biotin ligase recognition site at the C-terminus of the insert prior to a 6xHis tag (**Appendix 1**). Mutants were generated using two-step overlapping PCR experiments (Stemmer et al., 1995). The amplified PCR fragments were verified by agarose gel electrophoresis and used for the molecular cloning following the procedures described above. Each mutant construct was verified by DNA sequencing (Source Bioscience).

2.1.2 Constructs for Fab-MGD21

Based on the sequences of human monoclonal antibody MGD21 (Tan et al., 2016), two synthetic complementary DNA clones were obtained from GeneArt (ThermoFisher). The heavy and light chain variable regions were cloned into pOPINVH (heavy chain variable region) and pOPINVL (light chain variable region) by In-Fusion cloning (Clontech) (Nettleship et al., 2008).

2.1.3 Constructs for RIFIN, LAIR1 and LAIR1m

Synthetic complementary DNA clones of RIFIN (residues 1-378; UniprotKB/Swiss-prot ID: Q8IJ06) and LAIR1 ectodomain (residues 22-122; UniprotKB/Swiss-prot ID: Q6GTX8) were obtained from GeneArt (ThermoFisher). Residue 23-334 of RIFIN were inserted into pHLsec-BAP vector for SPR assays. LAIR1 fragment was constructed in the pHLsec-8H vector. LAIR1m fragment was amplified from Fab-MGD21 heavy chain by PCR and cloned into pHLsec-8H.

2.2 Protein expression and purification

2.2.1 HEK cell maintenance and large-scale expression

Human embryonic kidney (HEK) cell line 293T (ATCC CRL-11268) was the main expression cell line used for protein production. HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich) supplemented with 5 µg/ml L-Glutamine (L-Glu, Sigma-Aldrich), 5 µg/ml non-essential amino acids (NEAA, Invitrogen) and 10% (v/v) Fetal Bovine Serum (FBS, Invitrogen) in a humidified 37 °C incubator with 5% CO₂.

For large-scale protein production, 90 % confluent HEK293T cells in expanded surface roller bottles (2125 cm², Greiner Bio-One) were transfected with a cocktail of 2 mg of purified DNA (Endotoxin-Free Plasmid Giga Kit, Qiagen) and 4 mg polyethylenimine (PEI, Sigma-Aldrich) per liter of cell media. The roller bottles contained media with 2 % (v/v) FBS concentration and were incubated at 37 °C following transfection. Five to seven days post transfection, media was clarified by centrifugation at 4000g for 30 min and filtered using Steritop™ Filter units (Merck Millipore).

For proteins containing N-linked glycosylation sites, high diversity in glycan composition is common for HEK293T cell produced proteins and can interfere with protein crystallisation. To produce a highly homogeneous population of N-linked glycans, 5 µM kifunensine (Toronto Research Chemicals) was added into the culture media immediately after transfection. Kifunensine is a class I α-mannosidase inhibitor used to halt the N-linked glycosylation pathway in the *cis*-Golgi at the Man₉GlcNAc₂ stage. (Chang et al., 2007). *Flavobacterium*

meningosepticum endoglycosidase-F₁ (Endo-F₁) is then able to hydrolyse the N-linked oligomannose-type glycans, leaving a single N-acetylglucosamine at each N-linked glycosylation site, while hybrid and complex N-linked glycans are resistant to this enzyme.

2.2.2 Expression and purification of CD36 and MCvar1 CIDR α 2.8

CD36 and MCvar1 CIDR α 2.8 were co-expressed in HEK293T cells in the presence of kifunensine. The conditioned media was dialysed against phosphate-buffered saline (PBS) and the protein complex was purified by immobilised metal ion affinity chromatography (IMAC) using TALON resin (Clontech). The TALON resin was washed with buffer containing 25 mM Tris pH 7.5, 0.15 M NaCl, 10 mM imidazole, 10% (w/v) glycerol and eluted with buffer containing 25 mM Tris pH 7.5, 0.15 M NaCl, 0.5 M imidazole. The isolated CD36:CIDR α 2.8 complex was further purified by size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare Life Sciences) and buffer containing 10 mM Bis-Tris pH 6.5 and 0.3 M NaCl.

For size-exclusion chromatography multi-angle light scattering (SEC-MALS) and crystallisation experiments, proteins were expressed and purified as above. For SPR experiments, CD36 and MCvar1 CIDR α 2.8 proteins and their mutants were expressed individually in HEK293T cells without kifunensine treatment, to ensure that the proteins bore their native, complex N-linked glycans, as these might be important to give natural binding affinity. Proteins were purified as above before buffer exchange into 10 mM HEPES pH 7.5,

0.15 M NaCl.

Recombinant CIDRa2-6 domains for SPR analysis were expressed and purified by colleagues at the University of Copenhagen as described (Lau et al., 2015). In short, domain sequences with boundaries as defined in (Rask et al., 2010) were used to design domain-encoding DNA. These were codon optimised for expression in *Drosophila* Sf9 cells and were inserted into the baculovirus-insect cell expression vector pAcGP67-A GeneArt (ThermoFisher). All proteins were expressed with C-terminal V5 and His-tags, and were purified by nickel affinity chromatography. Sequences are given in **Appendix 3**.

2.2.3 Expression and purification of Fab-MGD21, RIFIN, LAIR1 and LAIR1m

DNA constructs expressing the heavy and light chains of Fab-MGD21 were mixed in a 1 to 1 ratio and co-transfected into HEK293T cells in the presence of 5 μ M kifunensine. The conditioned media was dialysed against PBS and purified by IMAC using TALON resin (Clontech). The Fab heterodimer was eluted with buffer containing 25 mM Tris pH 7.5, 0.15 M NaCl, 0.5 M imidazole and further purified by SEC using a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare Life Sciences) with buffer containing 10 mM HEPES pH 7.5 and 0.15 M NaCl. For SPR assays, RIFIN, LAIR1 and LAIR1m proteins were prepared the same as described in section **2.2.2**.

2.3 Crystallisation and cryo-protection

A pre-crystallisation test (PCT, Hampton Research) was performed to determine appropriate protein concentration prior to setting up crystallisation trials. The protein samples were then subjected to sitting drop vapour diffusion crystallisation trials in SwisSci 96-well plates by mixing 100 nl protein with 100 nl reservoir solution. For initial screening experiments, the following commercial screen kits from Molecular Dimensions were used: JCSG-plus, PACT, Morphenus, ProPlex, and PGA. Crystallisation plates were kept at 18 °C to allow crystals growth.

Cryoprotectants such as glycerol, ethylene glycol, and low molecular weight polyethylene glycol (PEG) were used to protect crystals from the damaging effects of ice formation, unless there was a cryoprotectant present in the crystallisation condition and the crystals could be mounted directly. A single protein crystal was fished into each cryoloop and moved to a cryoprotectant solution before being flash-cooled in liquid nitrogen for storage and data collection.

2.3.1 Crystallisation of CD36:MCvar1 CIDR α 2.8

The purified CD36:CIDR α 2.8 complex was concentrated to 10 mg/ml in the presence of 1% (v/v) Endo-F₁ and carboxypeptidase Y (Sigma-Aldrich) for *in situ* de-glycosylation and proteolysis (Chang et al., 2007; Dong et al., 2007). The CD36:CIDR α 2.8 complex crystallised with a well solution of 0.2 M NaCl, 20% (w/v) PEG6000, 0.1 M Tris pH 8.0 at 18 °C . It was cryoprotected using well solution supplemented with 25% (v/v) glycerol.

2.3.2 Crystallisation of Fab-MGD21

Concentrated Fab-MGD21 protein (10 mg/ml) was incubated with 1% (v/v) Endo-F₁ for *in situ* deglycosylation. The protein samples were then subjected to crystallisation trials described as above. Protein crystals were obtained with a well solution of 20% (w/v) PEG4000, 0.1 M sodium citrate pH 4.5 at 18°C. The cryoprotectant for these crystals was well solution with the addition of 25% (v/v) glycerol and crystals were flash frozen as described above.

2.4 X-ray data collection and structure determination

X-ray diffraction data were collected on beamlines IO2 and IO4-1 at Diamond Light Source, UK. At the synchrotron, the data collection strategy functions in EDNA and Mosflm were used to determine the optimum orientation of the crystal at which to start data collection as well as the exposure time, oscillation degree and crystal rotation for the complete data set.

Diffraction data were indexed and integrated using XIA2 (Winter et al., 2013) coupled with XDS (Kabsch, 2010), and scaled and merged using Aimless (Evans and Murshudov, 2013). A subset of 5% of randomly selected diffraction data was used for the calculation of R_{free} (Brunger, 1993). PHASER (McCoy et al., 2007) was used to determine molecular replacement (MR) solutions. The molecular models were completed by manual building in COOT (Emsley et al., 2010) and refined in autoBuster (Smart et al., 2012). All final refined models were validated with MolProbity (Chen et al., 2010) for the assessment of stereochemical quality and were deposited into Protein Data Bank (<http://www.rcsb.org/>) together with diffraction data.

2.4.1 Data collection and structure determination for CD36:MCvar1

CIDR α 2.8

Data were collected at beamline IO2 with space group C_12_1 and a resolution of 2.07 Å. The structure of the CD36:CIDR α 2.8 complex was determined using the structure of the homolog LIMP-2 (PDB ID: 4F7B; (Neculai et al., 2013)) as a search model for MR in PHASER (McCoy et al., 2007) to obtain the initial phases. The additional electron density corresponding to CIDR α 2.8

was clearly identified after initial refinement in autoBuster (Smart et al., 2012). Subsequently, the complex structure was determined by searching for CIDR α 2.8 used the core three α -helical bundle of MC179 (residues 580-645 and 719-737) (PDB ID: 3C64) in PHASER (McCoy et al., 2007) while the first solution of CD36 was fixed in place. An interpretable electron density generated by autoBuster (Smart et al., 2012) was used for manual model building in COOT to complete the model (Emsley et al., 2010). The data collection and refinement statistics for CD36:CIDR α 2.8 complex is listed in **Table 3.1**.

2.4.2 Data collection and structure determination of Fab-MGD21

Data were collected at beamline IO4-1 with the space group of C_{121} and a resolution of 2.52 Å. PHASER (McCoy et al., 2007) was used to determine a MR solution, using the known structures of LAIR1 (PDB ID: 3KGR; (Brondijk et al., 2010)) and a human monoclonal antibody Fab fragment, OX117 (PDB ID: 3DIF; (Nettleship et al., 2008)) with the hypervariable loops truncated, as search models. This identified two copies of the LAIR1m-containing Fab fragment in the asymmetric unit. Refinement and manual model rebuilding was completed using autoBuster (Smart et al., 2012) and COOT (Emsley et al., 2010), respectively. The data collection and refinement statistics for Fab-MGD21 are listed in **Table 4.1**.

2.5 Surface plasmon resonance (SPR)

Constructs used for SPR binding experiments were cloned into the pHLsec-BAP vector carrying a C-terminal biotin ligase (BirA) recognition site and a 6xHis tag (**Appendix 1**). After IMAC and SEC purification, 1 mg of purified protein was incubated with 20 μ g of enzyme BirA (Sigma-Aldrich) in buffer containing 10 mM HEPES pH7.5, 0.15 M NaCl, 5 mM ATP, 0.3 mM Biotin, and 5 mM MgCl₂ for *in vitro* biotinylation at 25°C for 15 hours. This buffer was then exchanged into SPR buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.005% (w/v) Tween20) using a PD10 column (GE Healthcare Life Sciences). Analysed proteins were cloned into pHLsec-8H, expressed and purified as in section 2.2. SPR binding measurements were performed using a BIAcore T200 instrument (GE Healthcare Life Science) at 25 °C in appropriate buffer solutions.

2.5.1 SPR analysis of the interactions of CD36 and CIDR domains

SPR binding measurements were performed at 25 °C in 20 mM HEPES pH 7.5, 0.15 M NaCl, 0.005% (w/v) Tween20. CD36 wild-type and mutants with BAP tags were expressed in HEK293T cells, incubated with BirA (Sigma-Aldrich) for *in vitro* biotinylation and immobilised onto a sensor CAPture chip (GE Healthcare Life Sciences) to about 700 Response Units (RU). MCvar1 CIDR α 2.8 domain and mutants used as analytes were expressed in HEK293T cells and purified as described in **2.2.2**. Dr. Louise Turner in the University of Copenhagen provided different CIDR proteins for binding studies described in section 3.2.5. In all binding experiments, analytes were injected at 30 μ l min⁻¹ for 240 sec with a 300 sec dissociation time before chip regeneration.

Affinities (dissociation constant, K_d) were estimated using equilibrium binding measurements and data were fitted with the BIAevaluation software (GE Healthcare Life Sciences) using a one site-binding model.

For competition experiments, different CIDR α proteins were injected onto the chip with 500 sec association time, followed by either 400 nM recombinant human Thrombospondin-1 (TSP-1; R&D SYSTEMS) or 10 nM native human oxidised low-density lipoprotein (oxLDL; AbD Serotec). All buffers were supplemented with 72.5 mM oleic acid (MP Biochemicals Australasia Pty Limited) for oxLDL binding experiments.

2.5.2 SPR assay of the interaction of RIFIN with LAIR1-containing ligands

SPR binding experiments were performed at 25 °C in DMEM (Sigma-Aldrich) and 0.05% (w/v) Tween20. RIFIN with a C-terminal BAP tag was immobilised onto a sensor SA chip to about 800 RU (GE Healthcare Life Sciences) by coupling through the C-terminal biotin. Fab-MGD21, LAIR1 and LAIR1m used as analytes were expressed in HEK293T cells and purified as described in section 2.2.3. In all binding experiments, a concentration series of analyte was injected at 30 $\mu\text{l min}^{-1}$ for 240 sec with a 480 sec dissociation time after injecting 10 mg/ml bovine serum albumin (Sigma-Aldrich) over the chip at 30 $\mu\text{l min}^{-1}$ for 60 sec. The chip surface was then regenerated with 300 sec of SPR buffer at 100 $\mu\text{l min}^{-1}$. Data were analysed as above.

2.6 Size-elution chromatography multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were performed using an analytical Superdex S200 10/300 GL column (GE Healthcare Life Sciences) connected to online static light-scattering (DAWN HELEOS 8, Wyatt Technology) and differential refractive index (Optilabr EX, Wyatt Technology). Purified sample (MCvar1 CIDRa2.8 at 1.5 mg/ml and CD36:MCvar1 CIDRa2.8 complex at 1 mg/ml) were injected into a column equilibrated with 10 mM HEPES pH 7.5, 150 mM NaCl. Molecular mass determination was performed using an adapted RI increment value (dn/dc standard value; 0.186 ml/g) to account for glycosylation. The theoretical molecular weight was predicated from amino acid sequence plus 1865 Da per N-linked glycosylation site for recombinant protein produced from HEK293T cells in the presence of kifunensine with limited glycosylation. Data were analysed using the ASTRA 6.1 software package (Wyatt Technology).

2.7 Analysis of *PfEMP1* sequences

Sequence analysis was conducted by Dr. Thomas Lavstsen in the University of Copenhagen. A total of 263 CIDRa2-6 domains from seven previously annotated whole genome sequenced parasites (Rask et al., 2010) were used as described in (Lau et al., 2015), to blastp extract CIDRa sequences from assemblies of Illumina whole genome sequencing data from 226 parasite samples collected in both Africa and Asia (Miotto et al., 2013) available through the MalariaGEN community. Only full length domain sequences were kept (domain boundaries as defined in (Rask et al., 2010)). This resulted in a total of 2386 CIDRa2-6 sequences, including 11 sequences from *P. reichenowi* CIDRa2-6 domains, which were used for the analysis of sequence diversity. A hand-corrected MUSCLE alignment was used to generate sequence logos by WebLogo 3 (Crooks et al., 2004) and neighbour joining sequence distance trees by MEGA (Tamura et al., 2013). No major sequence clusters could be verified by bootstrapping, but sequences distributed as previously observed (Rask et al., 2010).

2.8 Determination of unknown fatty acids using GC/MS

Mass spectrometry experiments were conducted and analysed by Dr. Jani Reddy Bolla and Professor Carol V. Robinson in the University of Oxford. To identify the nature of the bound ligand in crystals of the CD36:MCvar1 CIDR α 2.8 complex, gas chromatography coupled with mass spectrometry (GC/MS) was used. Purified CD36:MCvar1 CIDR α 2.8 protein was incubated at 100 °C for 5 min, and chloroform was added to a final concentration of 80% (v/v) to denature the protein and extract the ligand. The sample was derivatized with N,O-Bis(trimethylsilyl)trifluoroacetamide. The derivatized sample was then analysed by GC/MS on a Waters GCT Classic instrument. Ligand identification was performed by comparison of mass spectrum of the analyte with the NIST library database.

Chapter 3

Structural and functional characterisation of the interaction of *Pf*EMP1 with CD36

3.1 Project aim and summary

The role of CD36 binding in malaria has been studied extensively. Approximately 85% of *PfEMP1* proteins contain domains predicted to bind to CD36, making this the most common adhesion phenotype (Smith et al., 2013). A peptide-based study identified residues 146-164 of CD36 as forming the binding site for *PfEMP1* proteins (Baruch et al., 1999). Although the crystal structures of MCvar1 CIDR α 2.8 (Klein et al., 2008) and the CD36 homolog, lysosomal integral membrane protein 2 (LIMP-2) (Dang et al., 2014; Neculai et al., 2013; Zhao et al., 2014), have been solved, significant sequence diversity in CD36-binding CIDR domains hampered efforts to understand the molecular determinants that are retained to allow interaction with CD36. To understand what makes CD36-binding such a successful interaction for the parasite, I have therefore combined structural and biophysical studies with analysis of thousands of sequences to show how CIDR α domains interact with CD36. It is found that, despite extensive sequence diversity, the CD36-binding CIDR α domains retain a conserved hydrophobic pocket that accepts a phenylalanine residue found at the membrane distal tip of CD36. This phenylalanine is also required for CD36 to bind to oxidised low-density lipoprotein particles. In addition, the diverse *PfEMP1* CIDR proteins retain a chemically conserved surface feature that allows them to interact with an invariant region of their receptor, maintaining the ability to tether to the endothelium and to avoid clearance by the spleen.

3.1.1 Declaration

Mass spectrometry experiment was conducted and analysed by Dr. Jani Reddy Bolla and Dr. Carol V. Robinson in the University of Oxford. Sequence distance trees and logos were conducted by Dr. Thomas Lavstsen in the University of Copenhagen and variant CIDR proteins used for SPR binding assays were generated by Dr. Louise Turner in the University of Copenhagen.

3.2 Results

3.2.1 Protein production and characterisation of CD36 and CIDR α complex

To determine the structure of CD36 in complex with a CIDR α domain, the CIDR α 2.8 domain was selected from the MCvar1 *PfEMP1*, as the structure of this isolated domain was already known, providing structure-based insight into the domain boundaries (Klein et al., 2008). As shown in **Figure 3.1A**, the optimised constructs of CD36 (residues 35-439) and MCvar1 CIDR α 2.8 domain (residues 576-745 with a C626S mutation to remove an unpaired cysteine) were co-transfected at the DNA ratio 1:3 into HEK293T cells in the presence of the kifunensine. The secreted complex of CD36:CIDR α 2.8 was purified from dialysed conditioned media by IMAC and further isolated by SEC (**Figure 3.1B**). The final CD36:CIDR α 2.8 complex protein yield is approximately 0.1 mg from each liter of conditioned media.

To investigate the stoichiometry of complex formation, peak fractions containing the complex were subjected to SEC-MALS measurement, demonstrating the formation of a 1:1 complex, with no higher-order protein assemblies (**Figure 3.1C**). SEC-MALS also showed that CIDR α 2.8 proteins exist as monomers (**Figure 3.1C**).

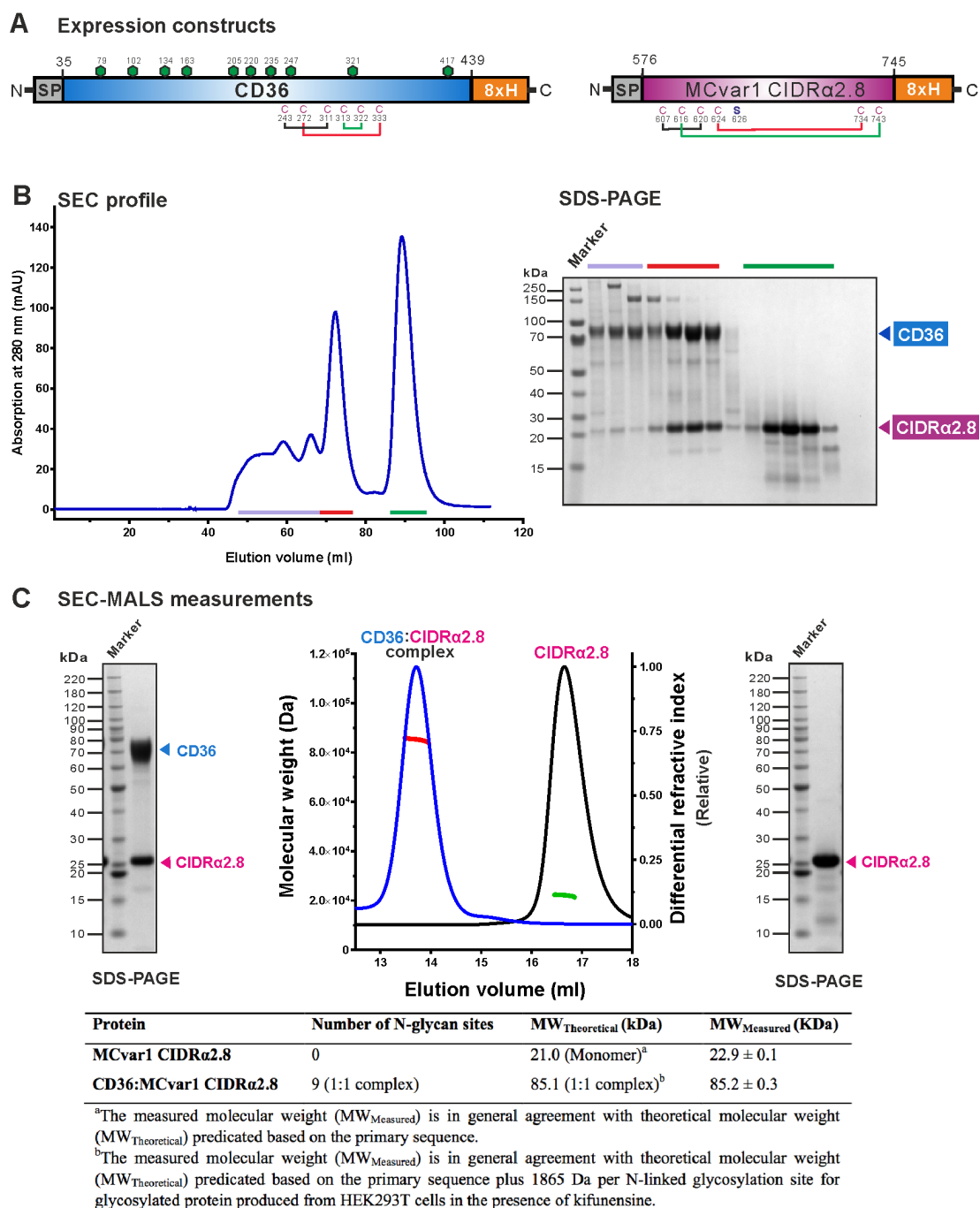


Figure 3.1. Protein expression, purification and biophysical analysis. (A) Schematic diagrams of expression constructs (SP, signal peptide from pHLsec vector; 8xH, eight histidine residues at the C-terminus). The disulphide bonds are numbered and drawn as lines. The N-linked glycosylation sites of CD36 are marked as green hexagons. **(B)** SEC elution profile and fractions analysed by SDS-PAGE under reducing conditions are marked with a red line for the CD36:CIDRα2.8 complex and a green line for CIDRα2.8 alone. **(C)** SEC-MALS analyses reveal the formation of a 1:1

complex of CD36 bound to the CIDR α 2.8 and monomer of MCvar1 CIDR α 2.8 alone. SDS-PAGE analysis of the CD36:CIDR α 2.8 complex and MCvar1 CIDR α 2.8 studied by SEC-MALS measurements are shown either side of the SEC-MALS traces. The SEC-MALS profiles of molecular weight (right ordinate axis) and relative differential refractive index (left ordinate axis) are shown as red and blue lines for CD36:CIDR α 2.8 complex and black and green lines for CIDR α 2.8 alone.

3.2.2 Crystallisation and structure determination of CD36 bound to a CIDR α domain

After purification, the complex was treated with Endo-F₁ to remove labile N-linked glycan chains and was concentrated to 10 mg/ml prior to set up of crystallisation trials. However initial crystallisation trials failed to generate crystals. The CD36:CIDR α 2.8 complex finally grew crystals with *in situ* deglycosylation with Endo-F₁ treatment and proteolysis with carboxypeptidase Y cleavage, suggesting that glycans and/or a disordered region at the C-terminus were preventing crystal formation. The CD36:CIDR α 2.8 complex crystallised in the C₁2₁ space group and a complete data set was collected to 2.07Å resolution (**Table 3.1**).

Molecular replacement attempts in which the MCvar1 CIDR α 2.8 domain (PDB 3C64) (Klein et al., 2008) was used as a search model failed, while using LIMP-2 (PDB 4F7B) (Neculai et al., 2013) as a search model yielded a solution. Subsequent analysis of the electron density map revealed that the MCvar1 CIDR α 2.8 domain in the CD36:CIDR α 2.8 complex had adopted a very different conformation from that seen in crystal of the domain alone (PDB 3C64; the detailed structural comparison is described in **Chapter 5**). The molecular replacement solution obtained using LIMP-2 was therefore used as the starting point for an iterative cycle of model building and refinement to generate a structure of the CD36:CIDR α 2.8 complex (**Table 3.1** and **Figure 3.2**). This structure was refined to an R_{work} of 21.2% and R_{free} of 25.2%, with all residues located in the favoured and allowed regions of a Ramachandran plot (**Table 3.1**).

As shown in **Figure 3.2A**, the CD36 ectodomain adopts a bean-shaped fold, consisting of antiparallel β -barrel core around short α -helical segments and stabilised by three disulphide bridges (Cys243-Cys311, Cys272-Cys333, and Cys313-Cys322). The architecture of CD36 is similar to LIMP-2, with an elongated structure, and is decorated with N-linked glycans distributed across the protein surface (**Figure 3.2A**). Nine of the ten predicted N-linked glycosylation sites of Asn79, Asn102, Asn134, Asn205, Asn220, Asn235, Asn247, Asn321, and Asn417 have well-defined electron density for the glycan attachment (**Figure 3.2A**), but the electron density is too diffuse to allow modelling of the glycan attached to Asn163. The membrane-distal head of CD36 is free from the decoration of N-linked glycans and consists of a three-helix bundle, required for binding to CIDR α 2.8 domain (**Figure 3.2A** and **3.2B**).

As expected from analysis of the stoichiometry of the CD36:CIDR α 2.8 complex (**Figure 3.1C**), the crystals contained an equal number of CD36 and CIDR α 2.8 molecules in the asymmetric unit. However, two CIDR α 2.8 domains make significant contacts with each CD36 ectodomain (**Figure 3.3A**). To determine which of these represents the physiological interaction and which is due to crystal packing, I assessed the binding of mutant proteins by surface plasmon resonance (SPR) experiments (**Figure 3.3B**). Glycosylation sites were introduced into the CIDR α 2.8 domain to disrupt each of the two interfaces (**Figure 3.3A**). The insertion of a glycosylation site into site 1 was partially successful (**Figure 3.3B**), but still disrupted the binding of the CIDR α 2.8 domain to CD36 (**Figure 3.3D**). In contrast, a mutation inserting a

glycan into site 2 had no effect on the interaction (**Figure 3.3E**), showing the same binding affinity as wild-type CDR α 2.8 domain (**Figure 3.3C**). Site 1 is therefore identified as the authentic binding interface for the CDR α 2.8 domain (**Figure 3.2A** and **3.3A**).

Table 3.1. Data collection and refinement statistics.

CD36:CIDR α 2.8	
Data collection	
Space group	C_{121}
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.8, 40.7, 138.8
α , β , γ (°)	90.0, 114.9, 90.0
Wavelength	0.97949
Resolution (Å)	63.00 - 2.07 (2.12 - 2.07)
R_{pim} (%)	4.6 (53.1)
$I / \sigma I$	8.2 (1.5)
Completeness (%)	99.8 (99.5)
Multiplicity	3.2 (3.3)
Refinement	
Number of reflections	40525 (3984)
$R_{\text{work}} / R_{\text{free}}$	21.2 / 25.2
Number of atoms	
Protein	4490
Ligands	233
Water	152
<i>B</i> -factors	
Protein	47.6
Ligand/ion	64.4
Water	44.6
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.19
Ramachandran plot	
Favored (%)	97.0
Allowed (%)	3.0

All structures were determined from one crystal.
Values in parentheses are for highest-resolution shell.
The structure is deposited with pdb code **5LGD**.

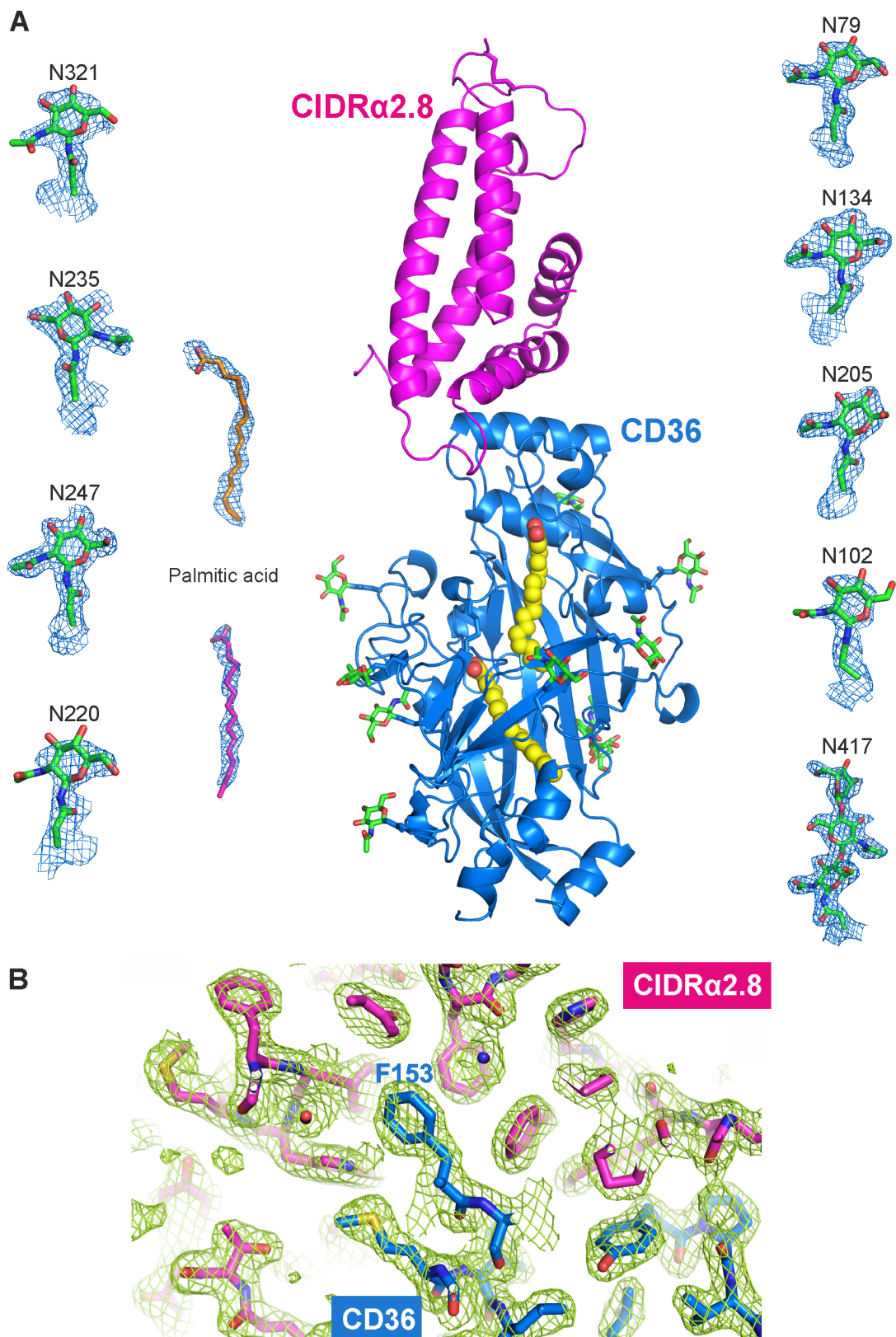


Figure 3.2. Crystal structure of the CD36:CIDR α 2.8 complex. (A) Cartoon representation of CD36 (blue) in a 1:1 complex with CIDR α 2.8 (magenta). The nine N-linked glycosylation sites and associated glycans are green while two

palmitic acids are shown as yellow. Close-up view of $2|F_o|-|F_c|$ electron density maps for N-linked glycans and palmitic acids are contoured at 0.9σ and 0.8σ as blue meshes, respectively. **(B)** The $2|F_o|-|F_c|$ electron density maps for the binding interface of CD36:CIDR α 2.8 complex are contoured at 1σ as green meshes. This is representative of the quality of electron density throughout the complex and contains the important CD36 residue, F153, which lies at the binding interface with the CIDR α 2.8 domain.

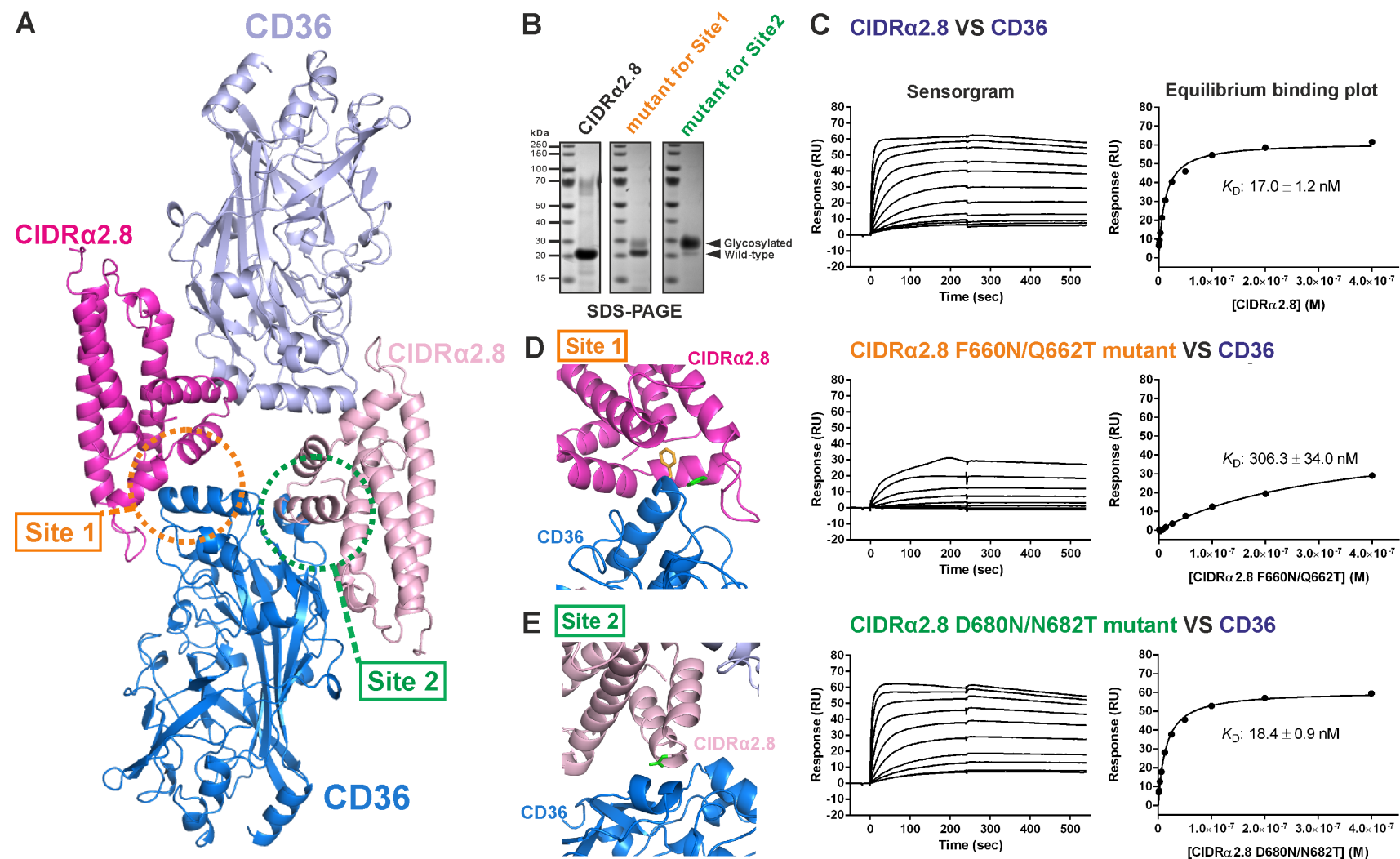


Figure 3.3. Analysis of crystal packing of CD36:CIDRα2.8 complex. (continued)

Figure 3.3. (continued)

(A) In crystals of CD36 bound to the MCvar1 CIDR α 2.8, two CIDR α 2.8 domains (magenta and light pink) interact with each CD36 (blue and light blue). Two distinct binding interfaces are marked as site 1 (an orange dotted circle) and site 2 (a green dotted circle). **(B)** The expression of wild and mutants forms of the CIDR α 2.8 domains with N-linked glycosylation sites introduced in site 1 (F660N/Q662T) and site 2 (D680N/N682T). A small fraction of the site 1 mutant was glycosylated while the majority of the site 2 mutant was glycosylated. **(C)-(E)** SPR sensorgrams (left panels) and equilibrium binding plots (right panels) against a series of concentrations of CIDR α 2.8 wild-type and mutant proteins are presented. **(C)** SPR experiments show that CD36 binds to CIDR α 2.8 with a binding constant (K_d) of 17 nM. **(D)** and **(E)** SPR showing binding of CD36 and CIDR α 2.8 mutants. Introduction of a partial glycan in site 1 disrupted binding, while a glycan inserted in site 2 did not affect binding. This confirms that site 1 represents the interaction that occurs in solution while site 2 represents a crystal contact.

3.2.3 The structural and functional analysis of CD36 binding by the *PfEMP1* CIDR α 2.8 domain

A small hydrophobic cavity on the surface of the *PfEMP1* CIDR α 2.8 domain therefore binds to the α -helical bundle at the membrane-distal tip of CD36 (**Figure 3.2B and 3.4**), in the region of the CIDR domain identified previously as containing the CD36 binding site (Baruch et al., 1999). The majority of the interaction is mediated by F153 from CD36, which fits into a hydrophobic pocket on the CIDR α 2.8 domain surface. The detailed interactions between CD36 and CIDR α 2.8 are described in **Table 3.2**. To dissect the relative importance of the binding site residues in the CD36:CIDR α 2.8 complex, interface residues were selected for mutagenesis and SPR experiments were used to assess binding affinities. Mutation of either F153 of CD36 , or of hydrophobic residues of the CIDR α 2.8 domain that line this pocket, abolishes the interaction (**Figure 3.5**). Secondary interactions are made by a smaller α -helix (residues 668-678) from the CIDR α 2.8 domain (mediating hydrogen bonding to CD36 through E672) and a loop (residues 647-656; again hydrogen bonding, this time through Q648 and D650). Neither of these interactions are critical in the MCvar1 CIDR α 2.8 domain, as disrupting hydrogen bonds by mutagenesis led only to small reductions in affinity (**Figure 3.5 and Table 3.2**).

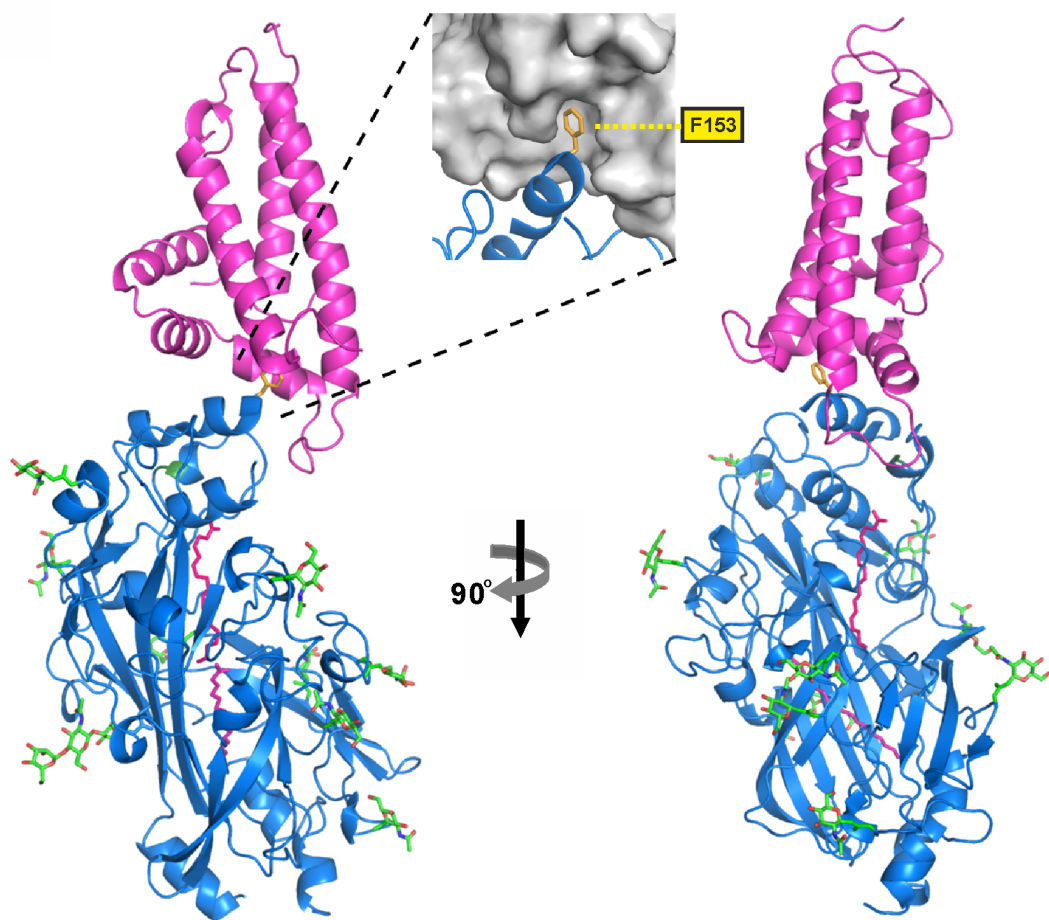


Figure 3.4. The structural basis for CD36 binding by CIDRα2.8 domain. The structure of CD36 (blue) bound to the CIDRα2.8 domain of the MCvar1 *PfEMP1* protein (magenta). N-linked glycans and palmitic acids on CD36 are shown in green and pink. The inset shows a surface representation of the CIDRα2.8 domain in gray with CD36 as a blue cartoon and F153 of CD36 in orange.

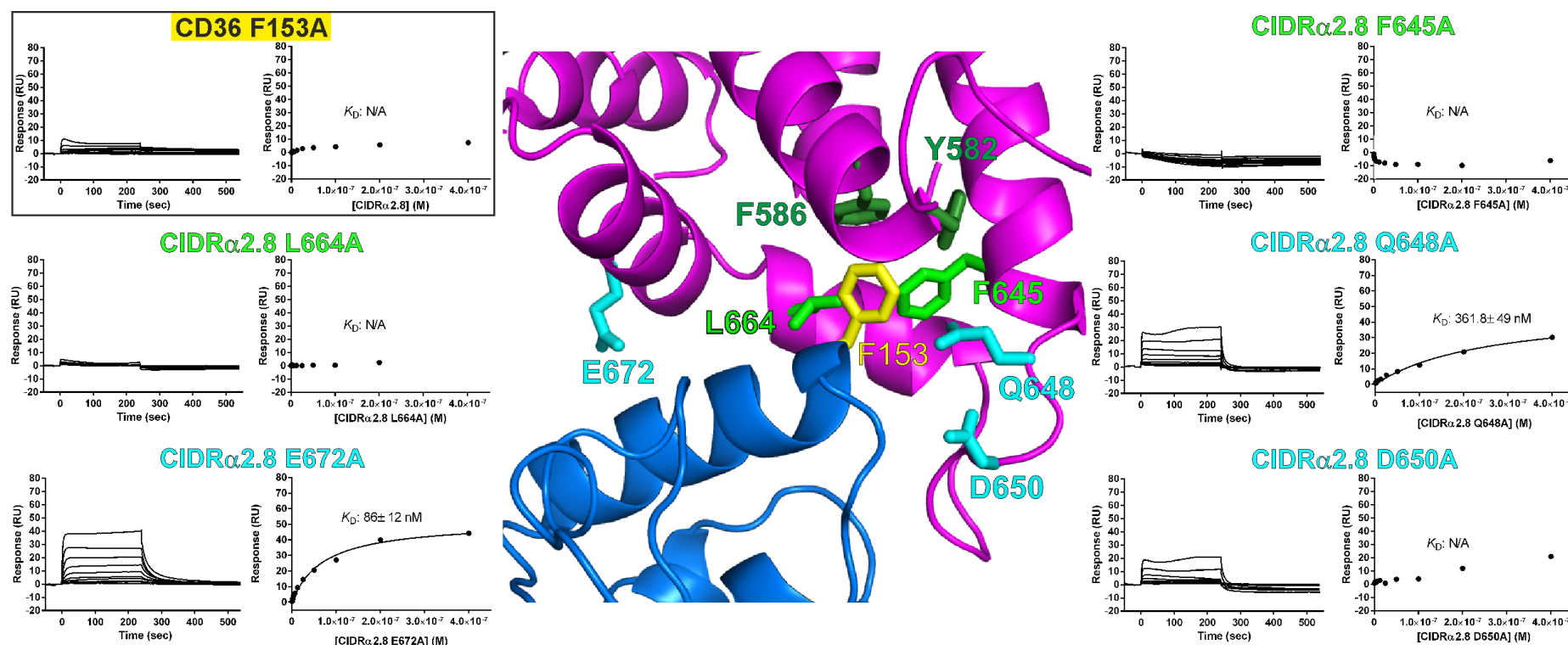


Figure 3.5. Structural details of binding sites in the CD36:CIDRα2.8 complex. A close-up view of the interface between CD36 (blue) and the CIDRα2.8 domain (magenta) with the key interacting residues labeled. Also shown is the effect of mutagenesis of key interacting residues. SPR sensorgrams and equilibrium binding plots are presented. The most critical interactions are mediated by F153 of CD36, which fits into a hydrophobic pocket lined by residues including F645 and L664 of the CIDRα2.8 domain.

Table 3.2. The interactions between CD36 and CIDR α 2.8.

CD36		MCvar1 CIDR α 2		Interaction
Residue	Group	Residue	Group	
TYR149	side chain	PRO652	side chain	Hydrophobic
ASN151	side chain NH2	GLN648	side chain CO	Hydrogen bond
ASN151	side chain NH2	ASP650	side chain	Hydrogen bond
GLN152	side chain NH2	ASN583	side chain CO	Hydrogen bond
PHE153	side chain	TRY582	side chain	Hydrophobic
PHE153	side chain	ASN583	back bone	Hydrophobic
PHE153	side chain	PHE586	side chain	Hydrophobic
PHE153	side chain	TRP587	side chain	Hydrophobic
PHE153	side chain	PHE645	side chain	Hydrophobic
PHE153	side chain	PHE660	side chain	Hydrophobic
PHE153	side chain	LEU664	side chain	Hydrophobic
VAL154	side chain	PRO652	side chain	Hydrophobic
VAL154	side chain	PHE660	side chain	Hydrophobic
MET156	side chain	TRP587	side chain	Hydrophobic
MET156	side chain	ILE664	side chain	Hydrophobic
MET156	side chain	ILE673	side chain	Hydrophobic
ILE157	side chain	TRP655	side chain	Hydrophobic
ILE157	side chain	PHE660	side chain	Hydrophobic
ILE157	side chain	THR663	side chain	Hydrophobic
ILE157	side chain	LEU664	side chain	Hydrophobic
ILE157	side chain	LEU669	side chain	Hydrophobic
ASN159	side chain NH2	ASP676	side chain	Hydrogen bond
SER160	side chain	GLU672	side chain	Hydrogen bond
PRO191	backbone O	TRP655	side chain NH	Hydrogen bond
TYR192	side chain	VAL154	side chain	Hydrophobic
TYR192	side chain	TRP655	side chain	Hydrophobic
TYR192	side chain	PRO652	side chain	Hydrophobic
PRO193	side chain	TRP655	back bone	Hydrophobic
PRO193	side chain	ASP654	back bone	Hydrophobic
LYS398	side chain NH3	ASN583	side chain CO	Hydrogen bond
GLN400	side chain NH2	ASP676	side chain	Hydrogen bond

Note: Interface analysis is based on PISA (<http://www.ebi.ac.uk/pdbe/pisa/>) followed a detailed check by PyMOL.

3.2.4 Structural comparison of CD36 with LIMP-2 and analysis of the binding of CD36 to long-chain fatty acid

The CD36 ectodomain adopts architecture similar to its homologue LIMP-2, with an elongated structure (**Figure 3.6A** and **3.7**) (Neculai et al., 2013). While LIMP-2 undergoes a pH-dependent conformational change, with an 'open' conformation at pH 5.5 and a 'closed' conformation at pH 7.5 (Dang et al., 2014; Neculai et al., 2013; Zhao et al., 2014), CD36, which was crystallised at pH 8.0, adopts an open conformation (**Figure 3.6A**). Indeed the pH-sensor residue, H150 of LIMP-2, which stabilises its closed conformation (Zhao et al., 2014) has been replaced by phenylalanine residue (F153) in CD36, and CD36 will therefore not undergo the pH-driven conformational change observed in LIMP-2, but will remain open and poised to accept fatty acids at neutral pH (**Figure 3.6B**).

A distinctive feature of LIMP-2 is a cavity that spans most of the length of the molecule (Neculai et al., 2013). This is also present in CD36 and is lined predominantly by hydrophobic side chains. Unlike in LIMP-2, the cavity of CD36 contains two electron density features that resemble the extended hydrocarbon chain of fatty acids (**Figure 3.2A**). Working with Dr. Jani Reddy Bolla and Dr. Carol V. Robinson in the University of Oxford, we identified these fatty acids by chloroform extraction of ligands from the CD36:CIDR α 2.8 complex used for crystallisation, followed by analysis of the extracted material by gas chromatography coupled with mass spectrometry (GC/MS). This revealed that palmitic acid and stearic acid were the most prevalent of a mixture of fatty acids present in the structure (**Figure 3.8**). Indeed CD36 has

been shown to play a role in the uptake of these fatty acids (Ibrahimi et al., 1996) and they have most likely bound to the protein during expression in human culture cells. Palmitic acid molecules were therefore modeled into these electron densities and I observed a close fit (**Figure 3.2A**). Both fatty acids are oriented with their carboxylic acid moieties at the membrane distal end. The majority of their interactions with CD36 are mediated by hydrophobic contacts as the fatty acids lie along the length of the cavity. However, the carboxylic acid group of the membrane distal fatty acid also makes water-mediated hydrogen-bonding interactions with T195.

The entrance for the fatty acid translocation pathway has previously been proposed to lie close to K164 (Neculai et al., 2013; Pepino et al., 2014). As shown in **Figure 3.9**, the central cavity indeed has an opening close to this residue (Entrance 1) that is a likely entry point for the membrane distal fatty acid and is the entrance that opens and closes in response to pH changes in LIMP-2. In addition, a second opening (Entrance 2) is also found at the membrane distal side of the CD36 ectodomain and the second fatty acid occupying the cavity lies in a tunnel that proceeds from this entrance. The CD36 structure therefore supports a model in which fatty acids move through the cavity, identifying a second possible entrance and further delineating favourable positions through which translocating ligands move as they pass towards the membrane proximal exit and the plasma membrane.

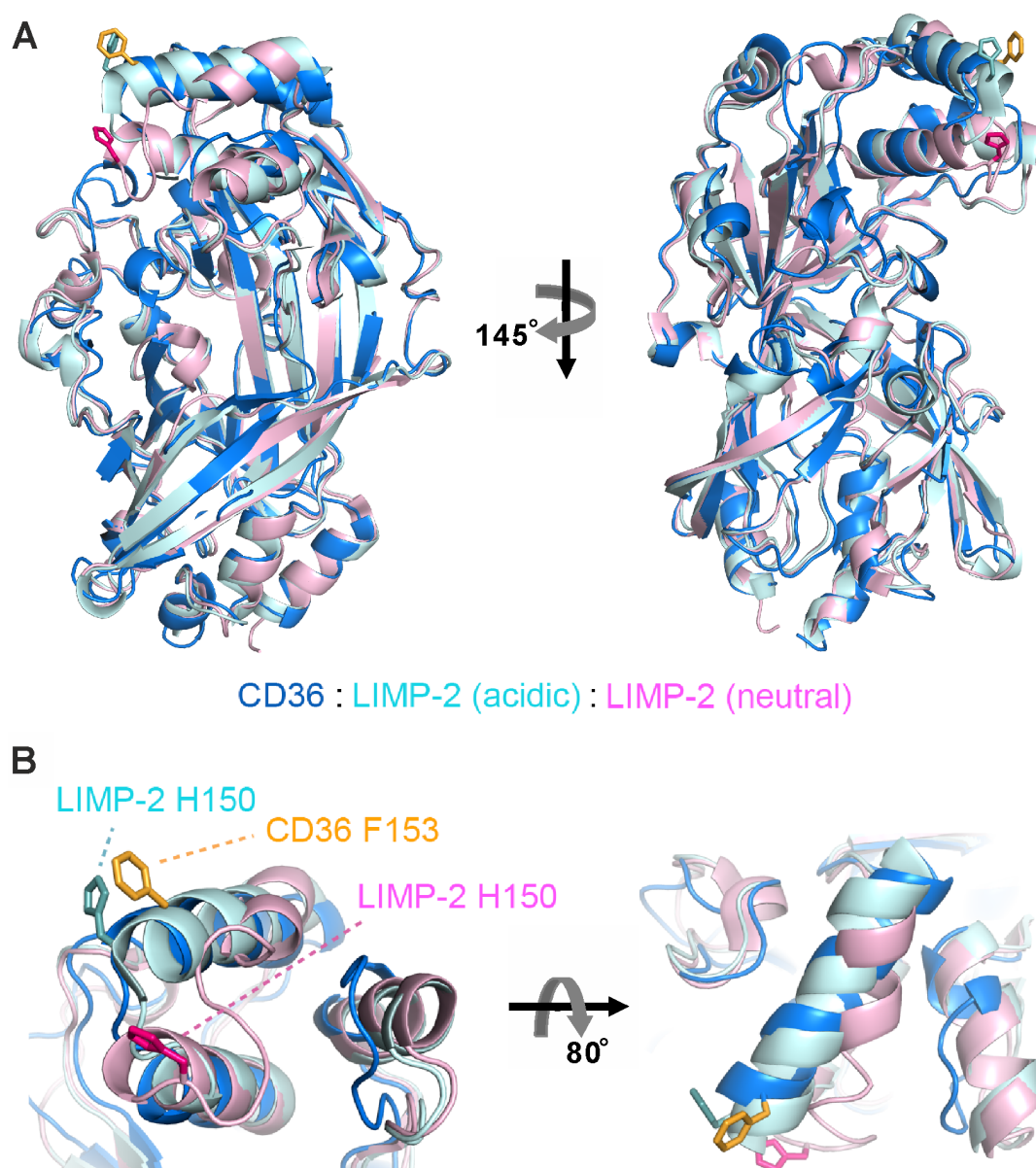


Figure 3.6. Superimposition of CD36 and LIMP-2 (A) A structural alignment of CD36 (blue) with the structures of LIMP-2 at acidic (cyan) and neutral (pink) pH. **(B)** Residues F153 from CD36 (orange) and H150 from LIMP-2 at acidic (light blue) and neutral (pink) pH are highlighted.

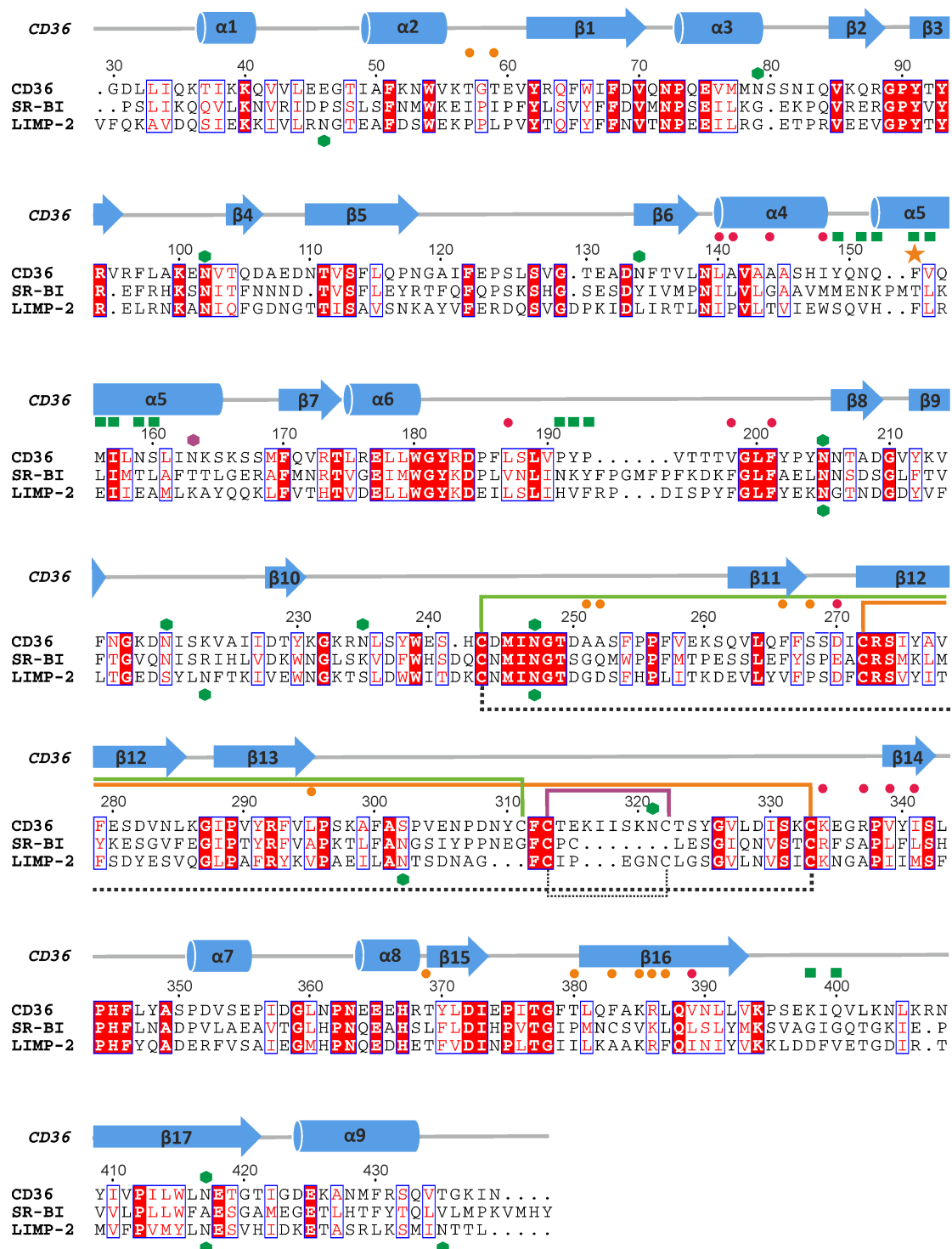


Figure 3.7. Sequence analysis of CD36 and its homologues. An alignment of the sequence of CD36 with those of its homologues scavenger receptor class B member 1 (SR-B1) and LIMP-2. Above the alignment is the secondary structure of CD36, derived from the crystal structure. Orange and red circles show residues that interact with the two fatty acids while green squares show those that interact with the MCvar1 CIDRα2.8 domain. Orange, green and

purple lines show the positions of the disulphide bonds of CD36; and dotted lines show the disulphide bonds of LIMP-2. Green hexagons show the positions of N-linked glycans on CD36 (above the sequence) and LIMP-2 (below the sequence) while the purple hexagon shows a position predicted to be glycosylated on CD36 but without a glycan observed in the electron density. CD36 F153 is highlighted with a yellow star.

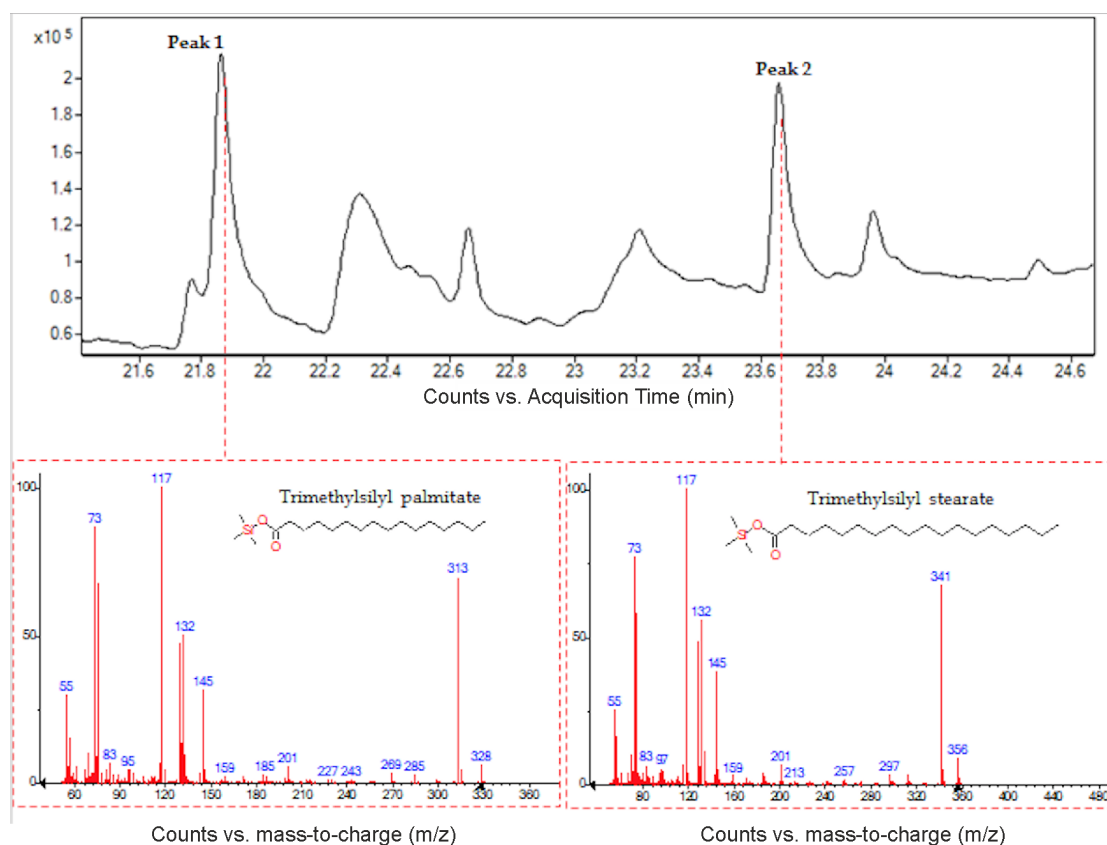


Figure 3.8. CD36 and its binding of fatty acids. Fatty acid identification by gas chromatography coupled with mass spectrometry (GC/MS). Fatty acids were chloroform extracted from the protein complex used for crystallisation and were converted to their corresponding trimethylsilyl esters and analysed by GC/MS. Masses were consistent with the identification of the most prevalent species as palmitic acid and stearic acid.

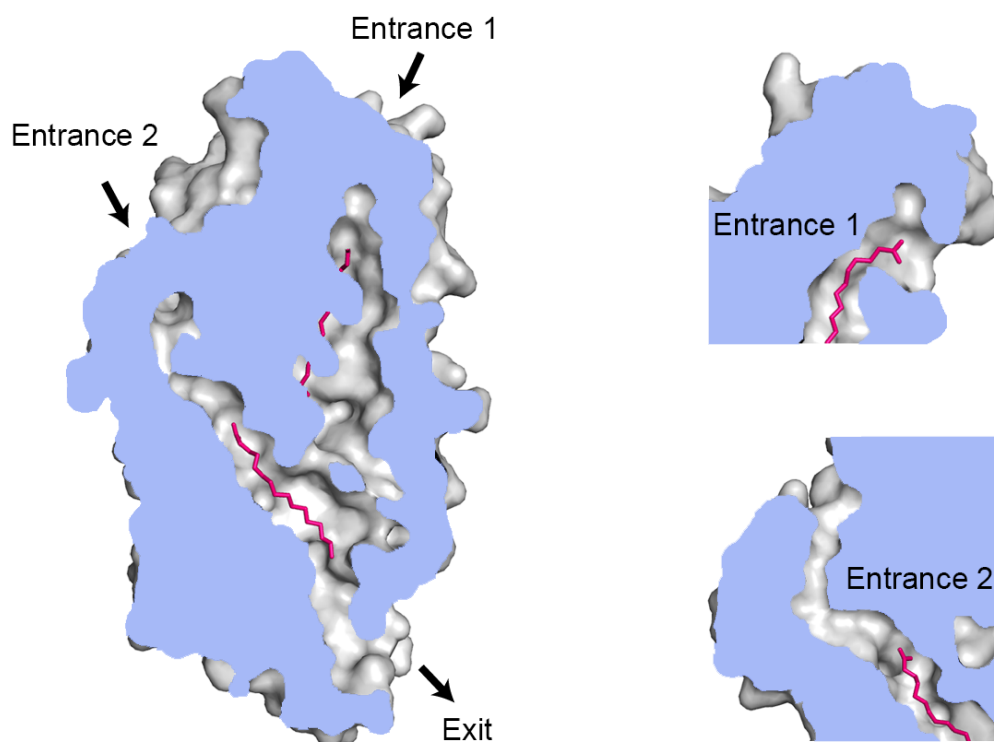


Figure 3.9. CD36 tunnels for fatty acids. A section through a surface representation of CD36 showing the central core cavity occupied by two palmitic acids (magenta). Insets show two putative entrances to this central cavity.

3.2.5 Conserved structural features of CIDR α domains allow CD36 binding

CD36 binding is an extremely common phenotype, with approximately 85% of *PfEMP1* proteins containing the CIDR α 2-6 domains that are predicted to bind this receptor (Robinson et al., 2003; Smith et al., 1998; Smith et al., 2013). The genome sequence of the 3D7 strain of *Plasmodium falciparum* allowed a systematic analysis of *PfEMP1* to identify those that contain CIDR domains that interact with CD36 (Robinson et al., 2003). Three classes of CIDR domains were identified, CIDR α , β , and γ . All tested CIDR α domains, with the exception of those from the CIDR α 1 subgroup, bound to CD36 while CIDR α 1 domains bind to EPCR (Turner et al., 2013). However the CIDR α 2-6 domains are highly variable with less than 33% pairwise sequence similarity. In order to understand how the CD36-binding CIDR α 2-6 binding domains diversify, while retaining ligand binding, I collaborated with Dr. Thomas Lavstsen and Dr. Louise Turner in the University of Copenhagen to conduct sequence analysis and set up biophysical assays.

With many *PfEMP1* sequences emerging from recent parasite genome sequencing projects, a more comprehensive analysis than those previously conducted was possible. As shown in **Figure 3.10**, the sequence distance trees was built by using 2386 CIDR α 2-6 domain sequences extracted from 233 parasite genomes that had been selected to be representative of global diversity (Crooks et al., 2004). No major sequence clusters could be verified by bootstrapping, but the sequences distributed into small subgroups as previously described (**Figure 3.10**) (Rask et al., 2010). The alignment

included 11 CIDR α domains from *Plasmodium reichenowi* that distributed across the tree of *Plasmodium falciparum* CIDR α 2-6 domains, as seen for other surface protein families (Turner et al., 2013).

To test the ability of CIDR to bind to CD36, a panel of CIDR α 2-6 domains were randomly selected from across this sequence diversity tree. I used SPR experiments to measure the affinity of these CIDR α domains for CD36. The SPR results showed that all CIDR α 2-6 domains adopt nanomolar interactions with slow off-rates (**Figure 3.10**). Together with previously identified CD36 binders (Janes et al., 2011; Robinson et al., 2003; Turner et al., 2013), this confirmed that all tested members of the CIDR α 2-6 classes bind to CD36.

As shown in **Figure 3.11A**, a sequence logo for all members of these domain subclasses was generated to understand sequence conservation. The majority of the domain sequences could be aligned into a single logo. However, the loop containing residues 673-683 was divergent in the CIDR α 3.1 subclass (**Figure 3.11A** and **3.11B**). As found before in the DBL and CIDR domains, the majority of conserved residues were cysteines or aromatics (Higgins and Carrington, 2014). A sequence logo for residues that make direct interactions with CD36 (**Figure 3.11C**) showed limited sequence conservation, but with hydrophobic residues exchanged for other hydrophobic residues and hydrogen bonding capacity conserved. Three of the most conserved residues directly contact CD36 (MCvar1 CIDR α 2.8 Y582, F586 and F645) and contribute to the hydrophobic pocket that accepts F153 (**Figure 3.5** and **Table 3.2**).

To understand the distribution of conservation across the CD36-binding regions of *Pf*EMP1 CIDR α domains, sequence entropy for each residue (as a measure of chemical conservation) was plotted onto the structure of the MCvar1 CIDR α 2.8 domain (**Figure 3.12A**). As seen previously for the CIDR α 1 domains (Lau et al., 2015), the significant majority of conserved residues were found in the core of the domain, stabilising the protein fold. The major patch of surface exposed conservation was in, and immediately surrounding, the conserved hydrophobic patch that binds CD36, with substitutions in this region leading to conservation of the hydrophobicity of this pocket, maintaining the ability to accept the conserved F153 residue from CD36 (**Figure 3.12B**). Other contacting residues lacked direct conservation, but retained the ability to form hydrogen bonds. This suggests a shared binding mode for CD36 across the CIDR α 2-6 domains.

To support this conclusion, I further tested the effect of the F153A mutation of CD36 on the binding of a sequence diverse set of ten CIDR α 2-6 domains by SPR (**Figure 3.13A**). This showed that the binding of each of these domains to CD36 was significantly reduced in the presence of the F153A mutations, with reduced overall binding levels and significantly faster off rates (**Figure 3.13A**). As shown in **Figure 3.13B**, the F153A mutation did not have an equivalent effect on all of the CIDR α domains, suggesting that the importance of additional interactions differ across the family. However, conservation of chemical property in the F153 binding pocket, combined with a significant effect of F153 mutation on all of the interactions supports the idea of a

conserved binding mode across the domain family, despite significant sequence diversification.

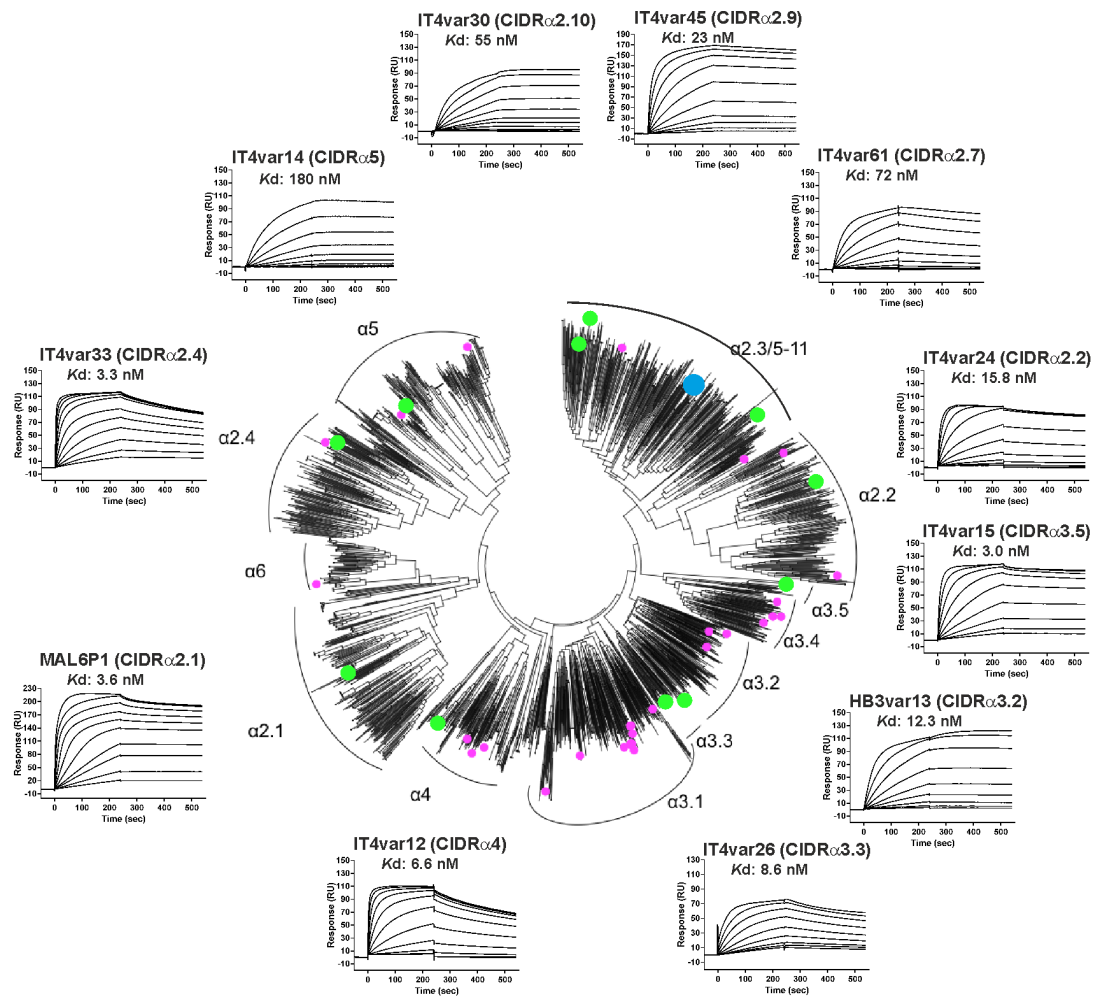


Figure 3.10. Extensive diversity in CD36 binding CIDR α domains. Sequence distance tree of 2386 full-length CIDR α 2-6 domains. A blue circle marks the sequence of the crystallised CIDR α 2.8 domain. Green circles mark sequences of recombinant CIDR α domains for which the affinity for CD36 binding were tested. Pink circles mark sequences of CIDR domains previously demonstrated to bind CD36. Annotated clusters contained previously defined CIDR α 2-6 subclasses. Each set of SPR sensorgrams shows a series of concentrations of CIDR α domains and presents K_d values. All tested domains from the CIDR α 2-6 subclasses bind to CD36.

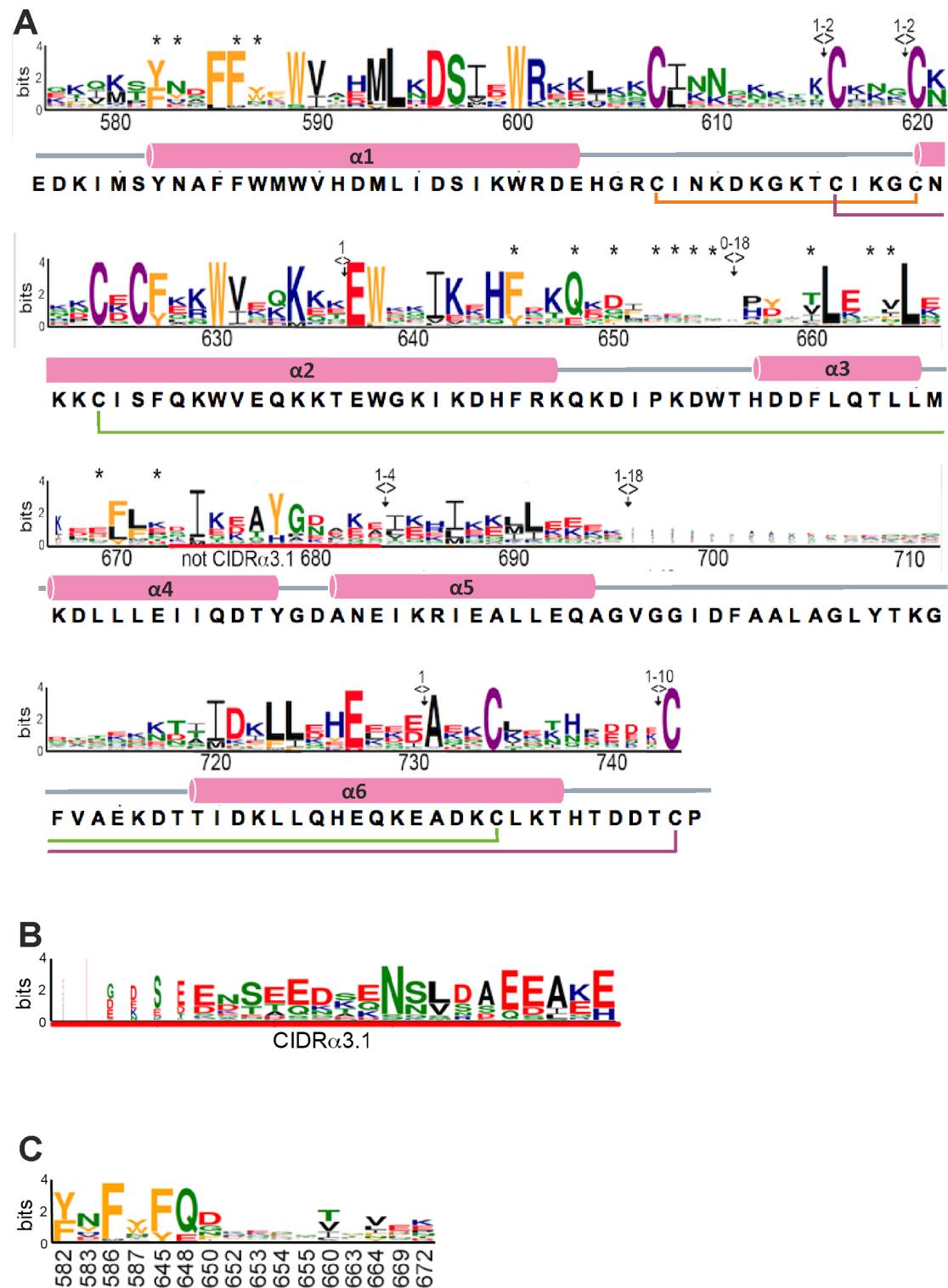


Figure 3.11. Conservation across the CD36-binding CIDRα2-6 domains. (A) Sequences of 2386 CIDRα2-6 domains were aligned and a sequence logo generated of residues equivalent to those found in the MCvar1 CIDRα2.8 domain. Deletions (><) and insertions (<>) are indicated, and underneath the sequence is a representation of the secondary structure of the MCvar1 CIDRα2.8. The green, purple and orange lines show the positions of

disulphide bonds. The region underlined by a red line is found in all CIDR α 2-6 domains except for CIDR α 3.1. Residues labeled with * make direct contact to CD36. **(B)** The region found in CIDR α 3.1 domains that replaces that underlined in red in (A). **(C)** A sequence logo for the residues that make direct contacts to CD36.

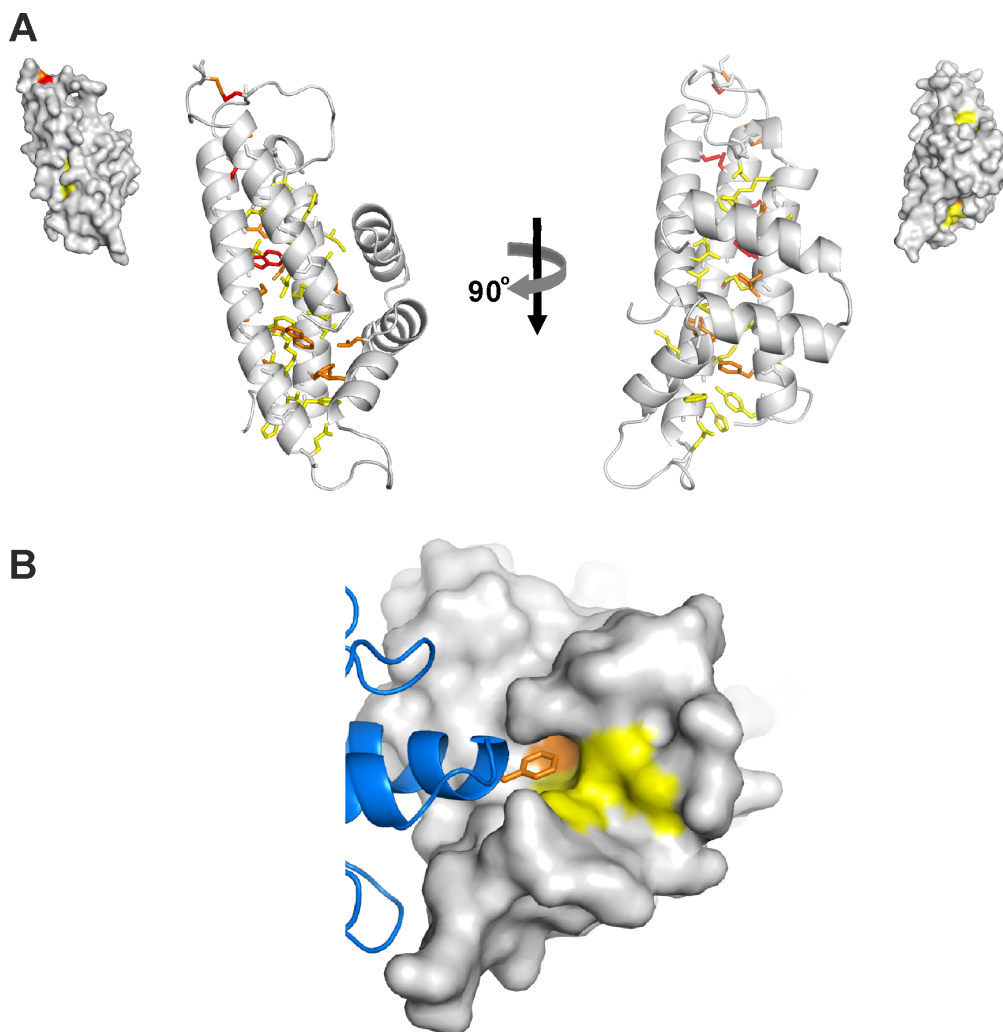


Figure 3.12. Limited chemical conservation of CIDR α domain allows CD36 binding. **(A)** Conservation in the CD36-binding CIDR α domain is plotted onto the structure of the MCvar1 CIDR α 2.8 domain. Absolutely conserved residues are shown as red sticks. Residues with a property entropy score of less than 0.1 (but not totally conserved) are orange and those with scores of 0.1-0.3 are yellow. The inset shows a surface representation in the same orientation and colours, showing that conserved residues cluster predominantly in core of the domain, stabilising its structure. **(B)** A surface representation of the CIDR α domain coloured as in **Figure 2.12A**, with CD36 in blue. This shows that residues in the hydrophobic pocket of the CIDR α domain are the most chemically conserved feature on the CIDR α domain surface.

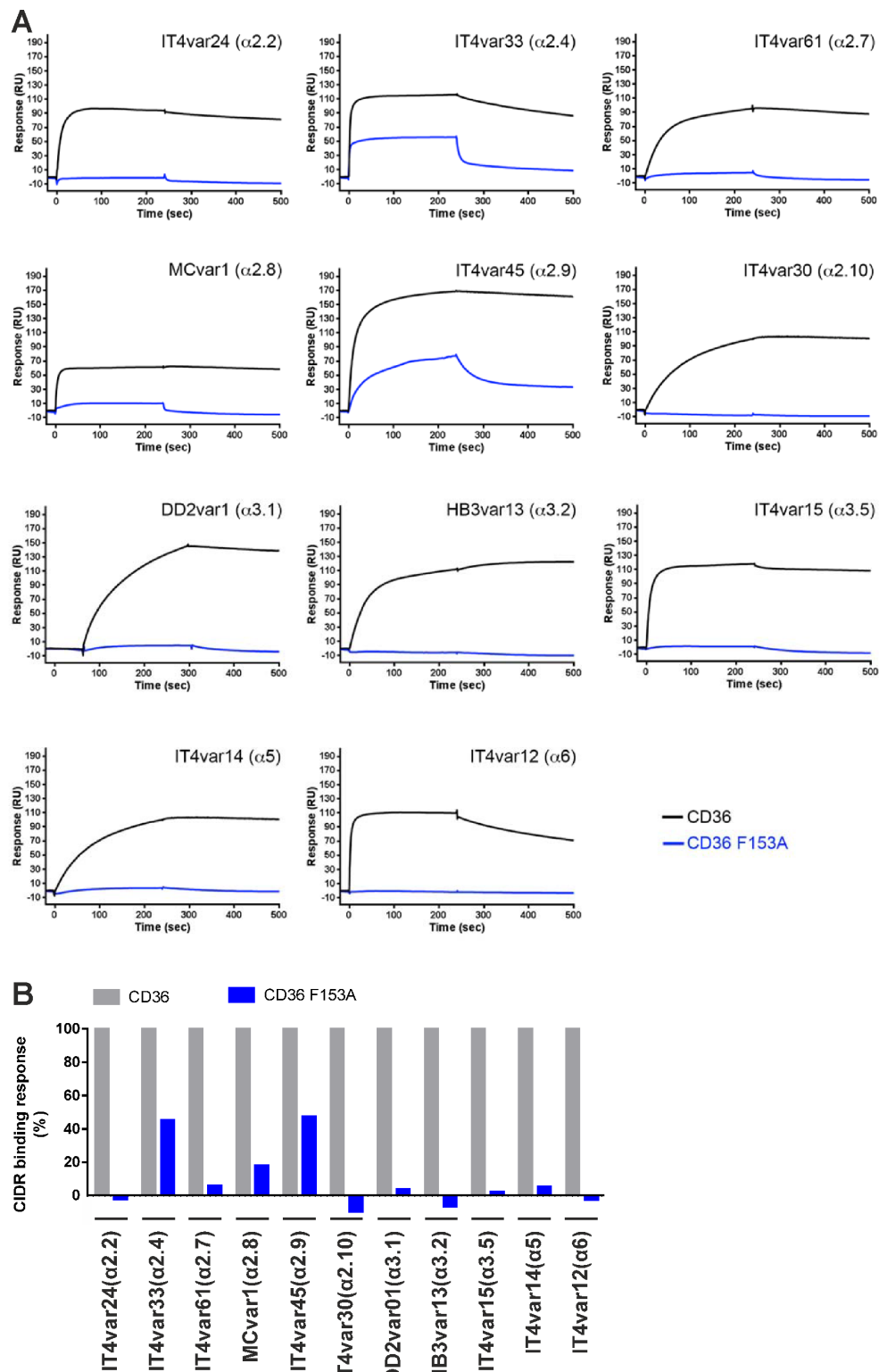


Figure 3.13. The effect of the F153A mutation of CD36 on the binding of CIDR α 2-6 variants. SPR analysis of the binding of equivalent concentrations of different CIDR α 2-6 subclasses to CD36 (black) and the F153A mutant of CD36 (blue). In each case, the concentration of CIDR α domain was 400 nM.

3.2.6 The binding of CIDR α domains affects the natural function of CD36

CD36 has a number of functions in mammalian physiology, including lipid metabolism, vascularisation, and immune response (Silverstein and Febbraio, 2009). It has been known that CD36 has the ability to bind to oxidised low-density lipoprotein (oxLDL) particles to facilitate fatty acid uptake (Neculai et al., 2013). It also interacts with thrombospondin (TSP) for the regulation of angiogenesis (Asch et al., 1987; Silverstein and Febbraio, 2009). As the CD36 ectodomain is decorated with nine N-linked glycans (**Figure 3.2A**), there is only a limited region on the surface of CD36 ectodomain available for protein-protein interactions, making it possible that the binding of *Pf*EMP1 CIDR α domain to CD36 may prevent it from interacting with physiologically relevant binding partners e.g. oxLDL and TSP. To test the hypothesis, an SPR experiment was designed to test for competition (**Figure 3.14**). As shown in **Figure 3.15A**, SPR showed that the binding of CD36 to TSP1 (TSP homolog 1 used as representation) was not altered by pre-incubation with CIDR α 2.9 domain. In contrast, the presence of CIDR α 2.9 domain abolished the interaction of CD36 with oxLDL particles (**Figure 3.15B**). The same was seen with five other CIDR α domains, selected to represent sequence diversity, suggesting this to be a conserved feature (**Figure 3.15D**). Furthermore, the CD36 F153A mutation, which disrupts the interaction of CD36 with CIDR α 2-6 (**Figure 3.13**) also abolishes binding of CD36 to oxLDL particles (**Figure 2.15C**), showing that this region of the protein is essential to interact both with oxLDL particles and *Pf*EMP1 CIDR α 2-6 domain (**Figure 2.15D**). Therefore both parasite proteins and oxLDL bind to a similar area on CD36 (**Figure 3.16**). The parasite therefore targets a part of CD36 that is essential for its

physiological role in fatty acid uptake, reducing the likelihood that the human host can escape from *Pf*EMP1 binding by polymorphisms in CD36 and these are also likely to alter the binding to its natural ligands.

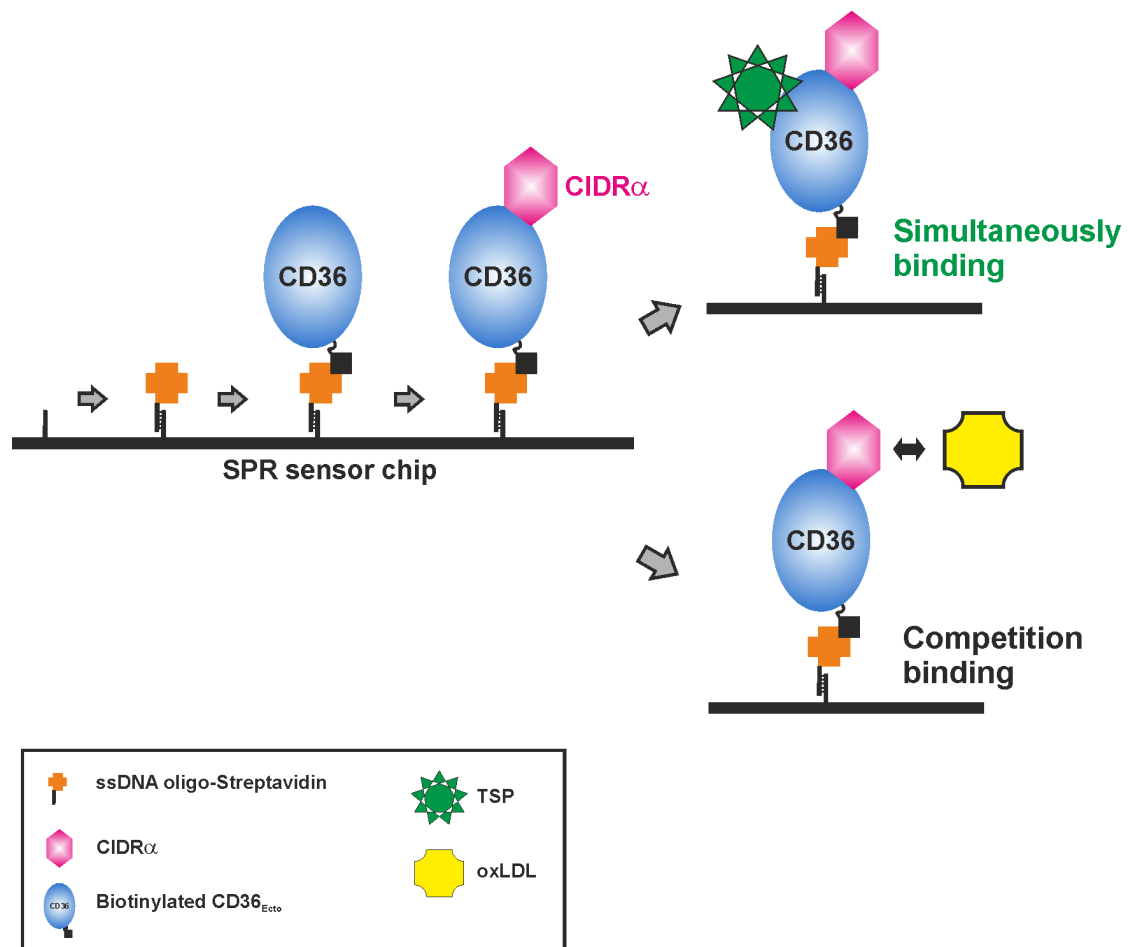


Figure 3.14. The use of SPR experiments to study competition binding. A streptavidin-coated CAP SPR sensor chip was used to capture biotinylated CD36 ectodomain. After CIDR α domain was bound to the immobilised CD36 ectodomain, CD36 ligands (e.g. oxLDL and TSP) were injected to assess whether they can bind simultaneously or in competition.

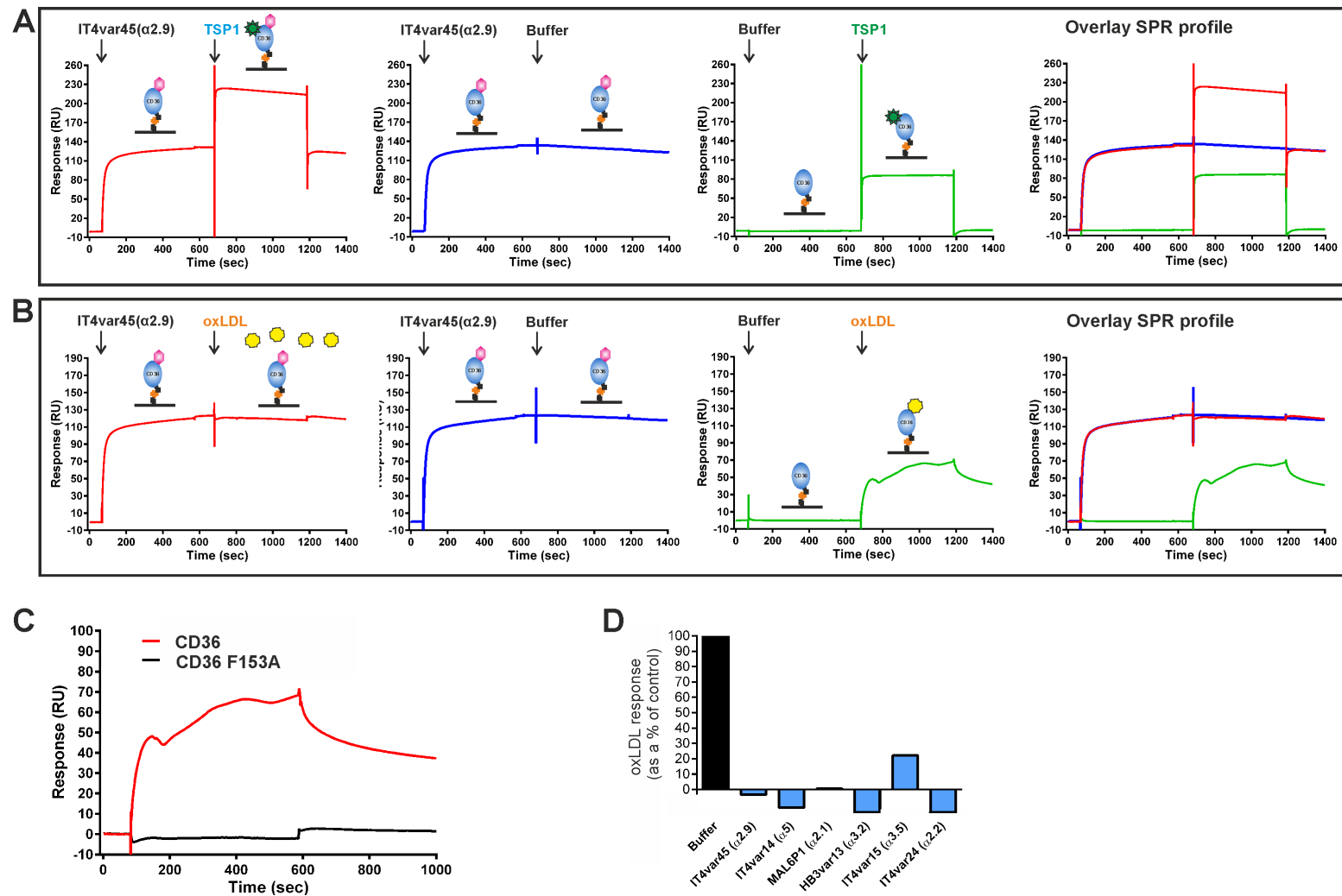


Figure 3.15. *Pf*EMP1 CIDR α domains prevent oxidised low-density lipoprotein (oxLDL), but not thrombospondin 1 (TSP1), from binding to CD36. (continued)

Figure 3.15. (continued)

(A) and (B) SPR sensorgrams showing that the preincubation of CD36 with IT4var45 CIDR α 2.9 does not affect the binding of TSP1 to CD36. But the preincubation of CD36 with IT4var45 CIDR α 2.9 prevents the binding of oxLDL to CD36. (C) Demonstration by SPR sensorgrams that the CD36 F153A mutation blocks the binding of oxLDL to CD36. (D) SPR data showing the inhibition of oxLDL binding to CD36 by a panel of diverse CIDR α 2-6 domains.

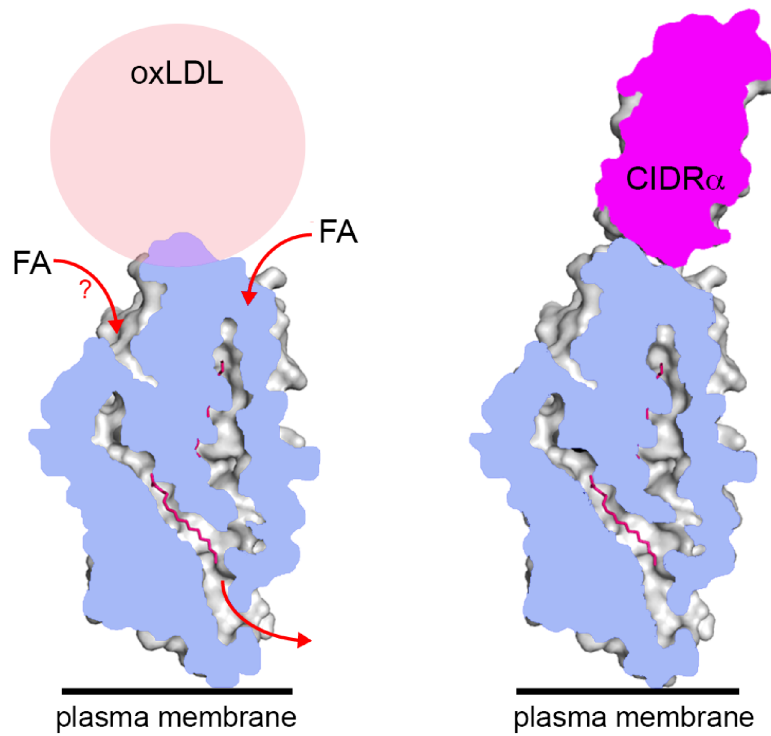


Figure 3.16. A model for the inhibition of oxLDL binding by *PfEMP1*. Fatty acids can be incorporated from oxLDL particles or other transport systems and pass through the central cavity of CD36 to the membrane. The *PfEMP1* CIDR α 2-6 domains interact with the same surface of CD36 as oxLDL and compete for binding.

3.3 Conclusion

In **Chapter 3**, I set up a reliable system for protein expression and purification of CD36:CIDR α 2.8 complex and CIDR α 2.8 protein in HEK293 cells. This allowed a high-resolution crystal structure of the CD36:CIDR α 2.8 complex to be determined. Structural and biophysical analyses together with sequence conservation analysis identified the molecular features that allow CIDR α domains to bind CD36. The structure of the CD36:CIDR α 2.8 complex bound to a long-chain fatty acid also allowed me to propose a molecular model for fatty acid translocation, in agreement with the previous identification of CD36 as a lipid transporter. Finally, I investigated whether CIDR α domain binding prevents CD36 from interacting with its natural ligands (oxLDL and TSP). Further discussion of how the malaria parasite takes advantage of the binding of the CIDR α domain to the oxLDL binding site of CD36 to maintain its capacity to cytoadhere and to protect it from splenic clearance is presented in **Chapter 5**.

Chapter 4

A LAIR1-containing human antibody structure

4.1 Project aim and summary

The general architecture of human antibodies is very conserved, with the variable domains of the heavy and light chain each providing three hypervariable loops at one tip, to generate diverse antigen binding sites. The most variable parts are the complementarity-determining regions (CDRs). While five of the six CDRs (including three CDRs of light chain, L1, L2, and L3, and the first and second CDRs of heavy chain, H1 and H2) can be classified with similar lengths and architectures (Goldsby et al., 2000), the third CDR of the heavy chain (H3) is more structurally diverse and variable in sequence, typically with a length of 8-16 residues (Goldsby et al., 2000; Gu et al., 1991). However, a recent study identified a group of unusual human antibodies in malaria endemic regions of Africa (Tan et al., 2016). These antibodies were discovered through their capacity to agglutinate human erythrocytes infected with different strains of *Plasmodium falciparum* and they bind to a subset of RIFINs. These RIFINs are displayed by the parasite on infected erythrocyte surfaces and are of unknown function (Chan et al., 2014; Gardner et al., 1998; Kyes et al., 1999). These antibodies show a remarkable adaptation with the 98-residue collagen-binding domain of human LAIR1 (also known as CD305) inserted into the third CDR loop of the antibody heavy chain (Tan et al., 2016). Mutation of residues in the LAIR1 domain, generates an antibody inserted form of LAIR1, “LAIR1m”, reducing binding to the natural ligand of LAIR1, collagen, but proposed to increase binding to infected erythrocytes via RIFINs (Tan et al., 2016). However, structural insight into the architecture of a LAIR1m-containing novel human antibody and understanding of how the mutations in LAIR1m change the ligand binding specificity, were

lacking. In the **Chapter 4**, the structure of the Fab fragment of one of these antibodies is revealed, showing that its antigen-binding site has adopted architecture that positions LAIR1m, while itself being occluded. This therefore represents a novel means of antigen recognition, in which the Fab fragment of an antibody acts as an adaptor, linking a human protein insert with antigen binding potential to the constant antibody regions which mediate immune cell recruitment.

4.2 Results

4.2.1 Expression and purification of the MGD21 Fab fragment

To design constructs for the expression of a LAIR1m-containing human monoclonal antibody, I cloned the variable regions of the heavy and light chains of antibody MGD21 (Tan et al., 2016) into vectors pOPINVH (containing a Cy1 from a mouse IgG1) and pOPINVL (containing a Ck from a mouse IgG1), respectively (Nettleship et al., 2008), as shown in **Figure 4.1**. This generated a chimeric His-tagged Fab fragment (fragment, antigen binding) with the variable domains of antibody MGD21. This was co-expressed in HEK293T cells in the presence of the class I α -mannosidase inhibitor, kifunensine (Chang et al., 2007). The secreted Fab-MGD21 fragment was purified from conditioned medium by IMAC and further isolated by SEC (**Figure 4.1C**) using Superdex200 and buffer containing 10 mM HEPES, pH 7.5, 0.15 M NaCl. The protein yield of Fab-MGD21 fragment is approximately 1 mg from each liter of conditioned media. The purified Fab-MGD21 fragment was concentrated to 10 mg/ml for extensive crystallisation trials.

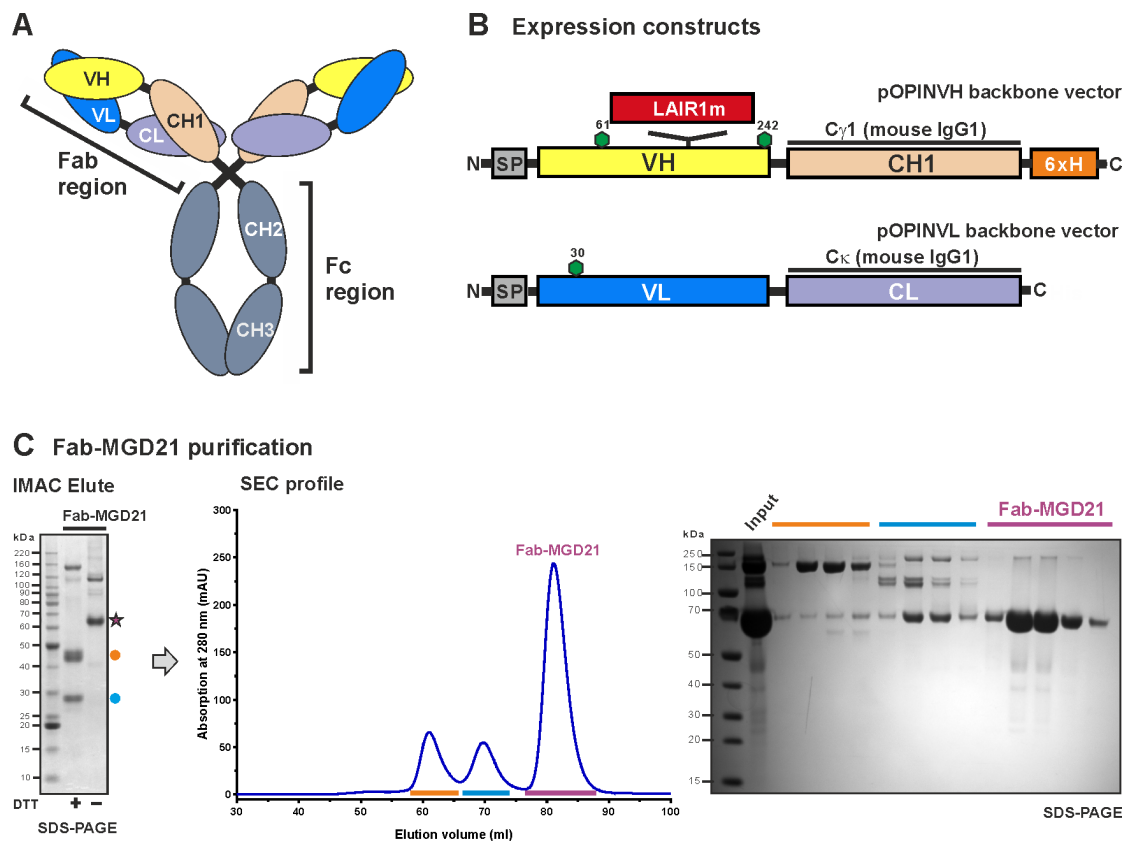


Figure 4.1. Secreted Fab-MGD21 protein expression and purification. (A) Schematic representation of a conventional intact antibody, including two light chains (composed of variable and constant regions, VL and CL) and two heavy chains (each chain contains one variable region, VH, and three constant regions, CH1, CH2, and CH3). (B) Schematic diagrams of expression constructs (SP, signal peptide from pHLsec vector; 6xH, six histidine residues at the C-terminus). The region of the LAIR1m domain is marked as a red bar. The putative N-linked glycosylation sites are marked as green hexagons. (C) His-tagged Fab-MGD21 proteins was purified by IMAC and analysed by SDS-PAGE under either non-reducing or reducing conditions. Orange and cyan circles represent a LAIR1m-containing heavy chain and a light chain of Fab-MGD21, respectively. The purple star represents a disulfide-linked Fab-MGD21. SEC elution profile and fractions analysed by SDS-PAGE under non-reducing conditions are marked with a purple line for the Fab-MGD21 fragments.

4.2.2 Crystal structure of Fab-MGD21

The Fab-MGD21 fragment was crystallised in the C_{121} space group with *in situ* deglycosylation with Endo-F₁ treatment, allowing a complete dataset to be collected to 2.52Å resolution (**Table 4.1**). The structure was determined by molecular replacement using the ectodomain of LAIR1 (PDB ID: 3KGR, (Brondijk et al., 2010)) and the Fab fragment of antibody OX117 (PDB ID: 3DIF (Nettleship et al., 2008)) as search models. This identified two copies of the Fab-MGD21 fragment in the asymmetric unit of the crystal (**Figure 4.2**). A model was built for residues 2-211 of the light chain and 1-351 (with residues 214-219 and 264-270 disordered) of the heavy chain (**Figure 4.3**). The detailed protein sequence information of the Fab-MGD21 fragment was shown in **Figure 4.4**.

Table 4.1. Data collection and refinement statistics

	Fab-MGD21
Data collection	
Space group	C_{121}
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	169.8, 86.5, 104.0
α , β , γ (°)	90.0, 126.7, 90.0
Wavelength	0.92819
Resolution (Å)	81.90 - 2.52 (2.56 - 2.52)
<i>R</i> _{pim} (%)	5.4 (67.8)
<i>I</i> / σ (<i>I</i>)	7.4 (1.0)
Completeness (%)	99.8 (98.3)
Multiplicity	3.2 (2.7)
Refinement	
Number of reflections	40946
<i>R</i> _{work} / <i>R</i> _{free}	21.9 / 26.7
Number of residues	
Protein	1076
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.25
All Atom clash score	5
B factors	
All atoms	71.53
Solvent	63.12
Variable domains	65.17
Constant domains	74.29
LAIR1 insert	73.70
Linkers	94.71
Ramachandran plot	
Favored (%)	95.2%
Allowed (%)	4.8%

The structure was determined from a single crystal.
Values in parentheses are for highest-resolution shell.
The structure is deposited with pdb code **5NST**.

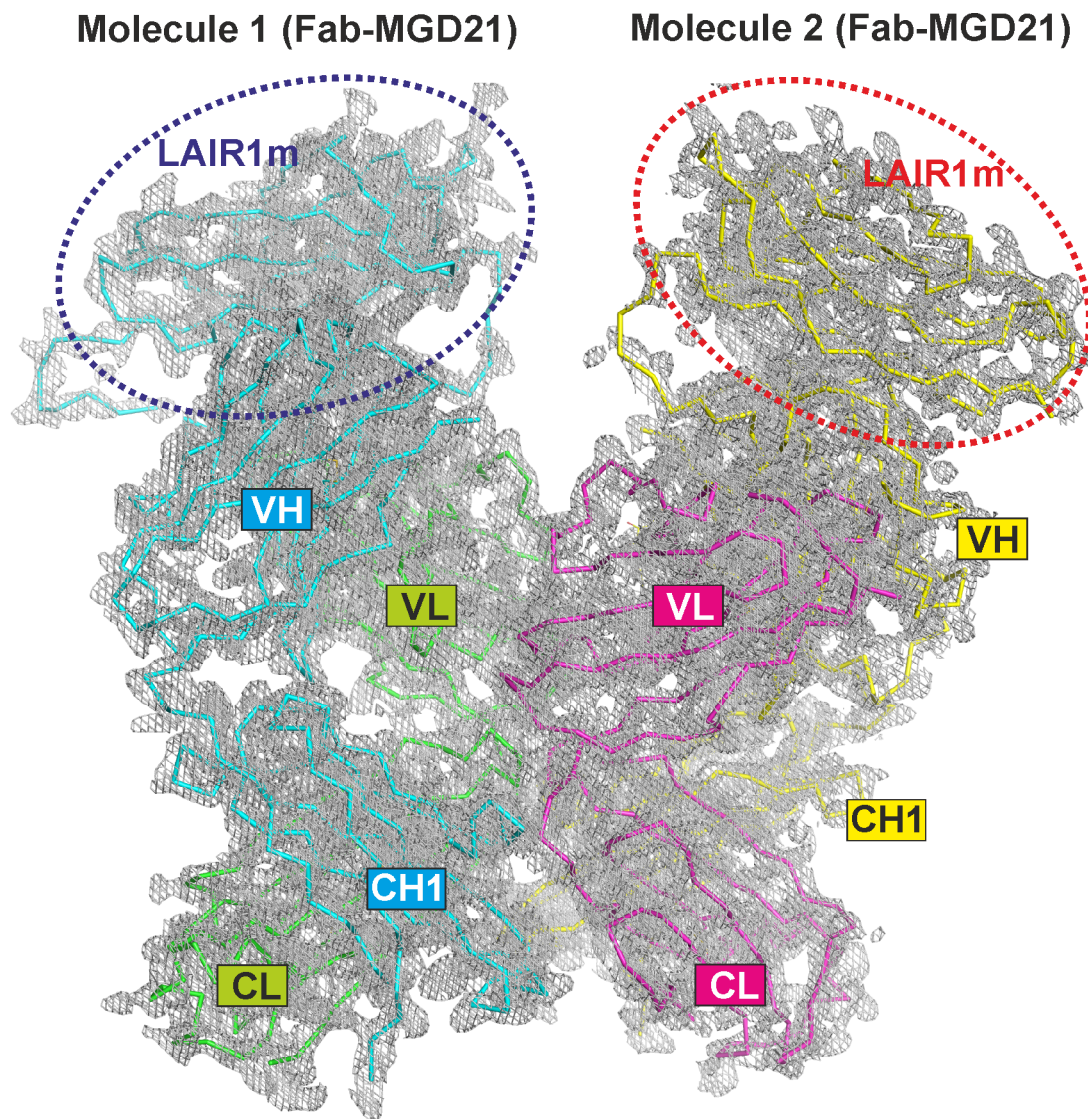


Figure 4.2. A view of the electron density of two Fab-MGD21 molecules in the asymmetric unit. The $2|F_o|-|F_c|$ electron density map in grey is contoured at 1 σ . The domains of Fab-MGD21 fragment are marked as LAIR1m, VH, CH1, VL, and CL.

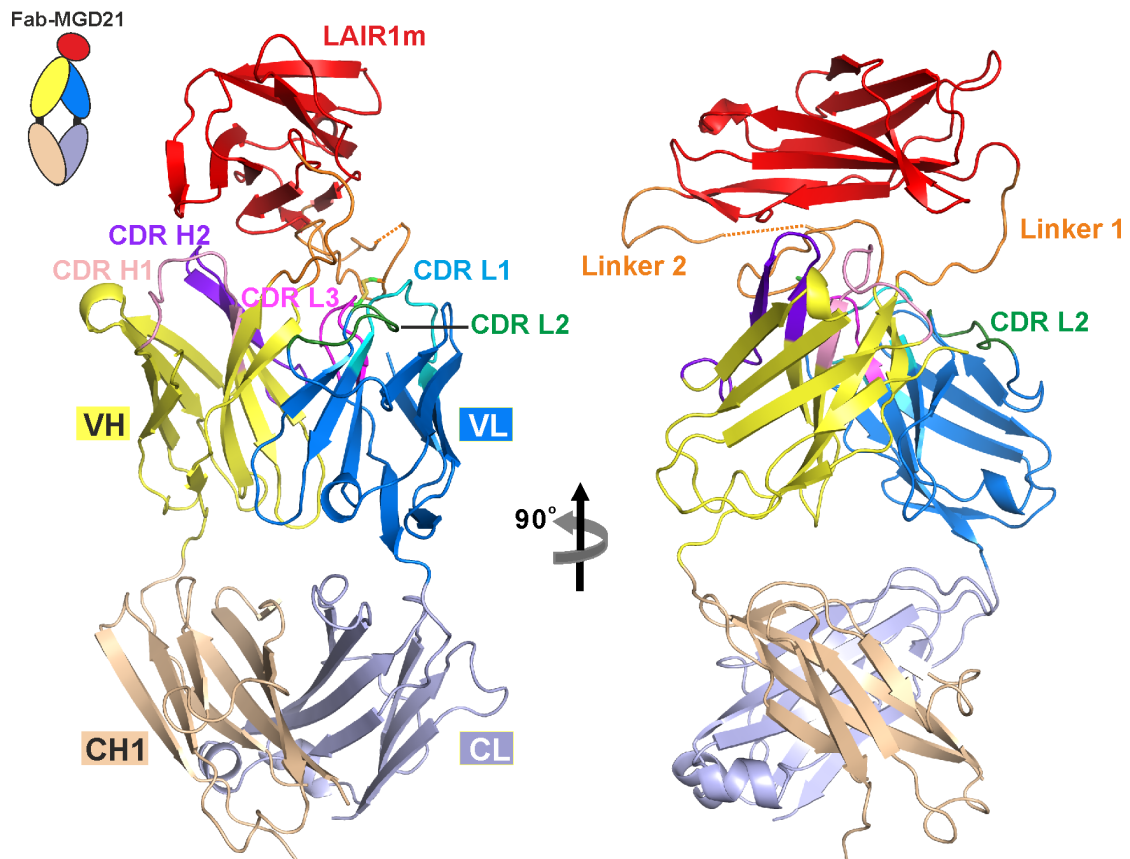


Figure 4.3. Structure of a LAIR1m-containing antibody Fab-MGD21 fragment. Cartoon representation of the Fab-MGD21 fragment. LAIR1m (red) is inserted into the third CDR loop of the heavy chain variable region (VH; yellow) through two extended linkers (Linker 1 and 2; orange). The VL and CL of light chain are shown as blue and grey blue, respectively. The dashed orange line represents protein disordered in the structure. Each of the CDR loops and its corresponding label is shown in different colour in the annotated sequences in **Figure 4.4**.

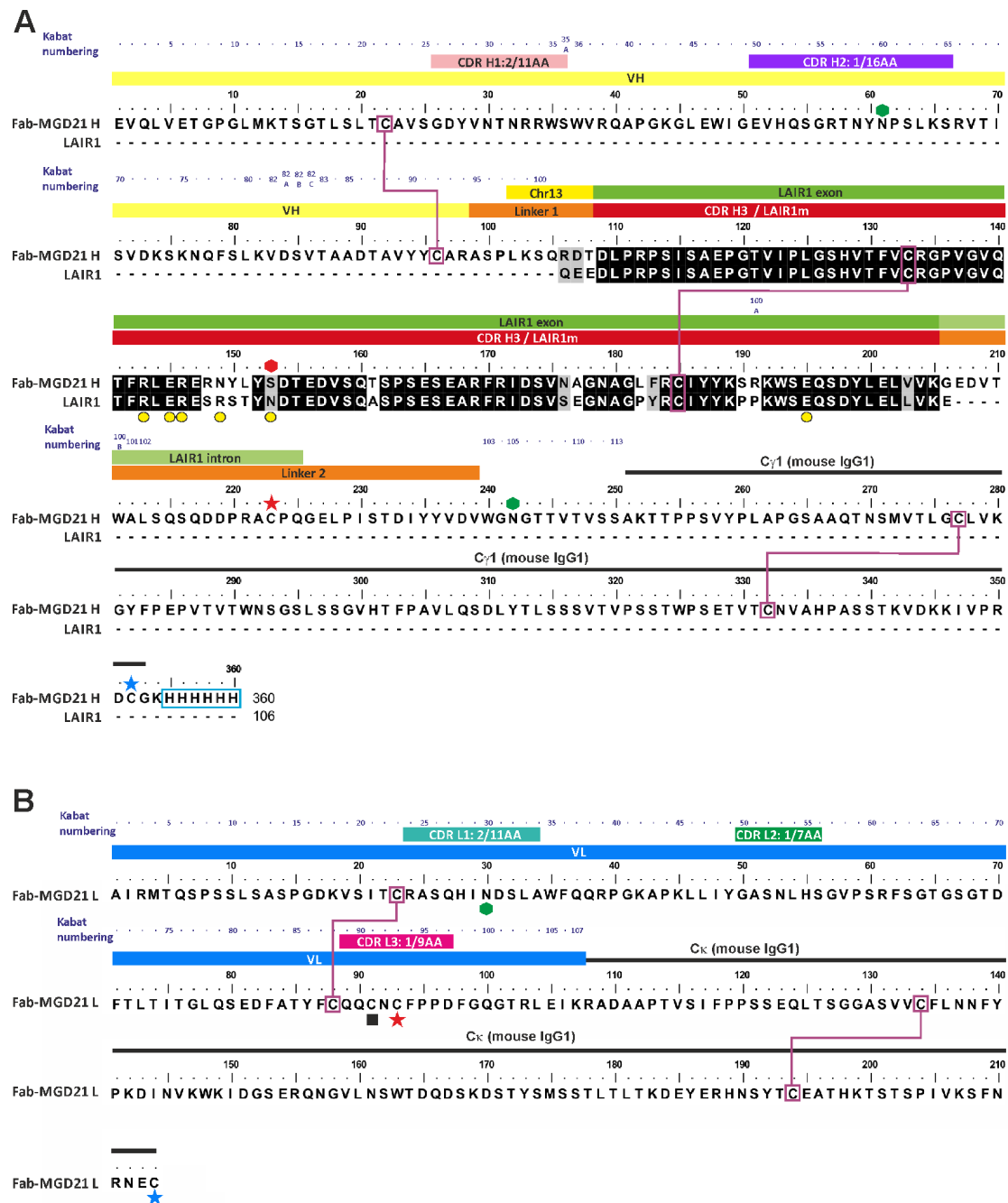


Figure 4.4. Protein sequence of antibody Fab-MGD21 and its alignment to germ line LAIR1 (A). The sequences of the heavy chain of Fab-MGD21 aligned to that of germ line LAIR1. The domain regions of VH, linkers, LAIR1m insertion, and Cγ1 are highlighted according to **Figure 4.3**. CDR H1 and H2 are shown in light pink and purple. LAIR1 segments of exon and intron are shown in green and light green. Chromosome 13 (Chr13) sequences are shown in yellow. Yellow circles mark sites in germ line LAIR1 known to affect collagen binding. **(B).** The sequences of the light chain of Fab-MGD21 with the CDR loops indicated by cyan, green and magenta. The domain regions of

VL and Cκ are highlighted according to **Figure 4.3**. In **(A)** and **(B)**, hexagons represent putative N-linked glycosylation sites with the red hexagon representing a site lost in the mutation of LAIR1m insertion. The intramolecular disulphide bonds are drawn as purple lines. Stars in red and blue represent cysteine residues that make a cross-chain disulphide bonds. One free cysteine residue of the light chain is marked with a black square. Kabat numbers (Dunbar and Deane, 2016) of residues are given above the sequences. In addition, all the CDRs, with the exception of the third CDR of the heavy chain, are labeled with their canonical class (Martin and Thornton, 1996)

The architecture of Fab-MGD21 adopts a canonical Fab fragment structure, comprising heavy (VH and Cy1) and light chains (VL and Cκ), except for the unique LAIR1m insertion into CDR3 of VH (**Figure 4.3**). The two Fab fragments from the asymmetric unit adopt the same structure with a root mean square deviation of 0.26Å (calculated over 475 C α atoms) suggesting a highly ordered linkage between the variable domains of the antibody and LAIR1m (**Figure 4.5**). As shown in **Figure 4.6**, Fab-MGD21 contains five intramolecular disulphide binds; Cys23-Cys88 and Cys134-Cys194 within light chain; Cys22-Cys96 and Cys277-Cys332 within heavy chain; Cys133-Cys185 within LAIR1m. There are also two intermolecular disulphide bridges; light chain Cys93-heavy chain Cys223 and light chain Cys214-heavy chain Cys352, the later of which is in a disordered region of Fab-MGD21 and therefore missing from the model. The antibody sequence of Fab-MGD21 has three putative N-linked glycosylation sites, but of these (light chain Asn30; heavy chain Asn61 and Asn242) only Asn242 shows a distinguishable electron density corresponding to an Asn-linked N-acetyl glucosamine, in a position distant from the LAIR1m insert (**Figure 4.4** and **4.6**).

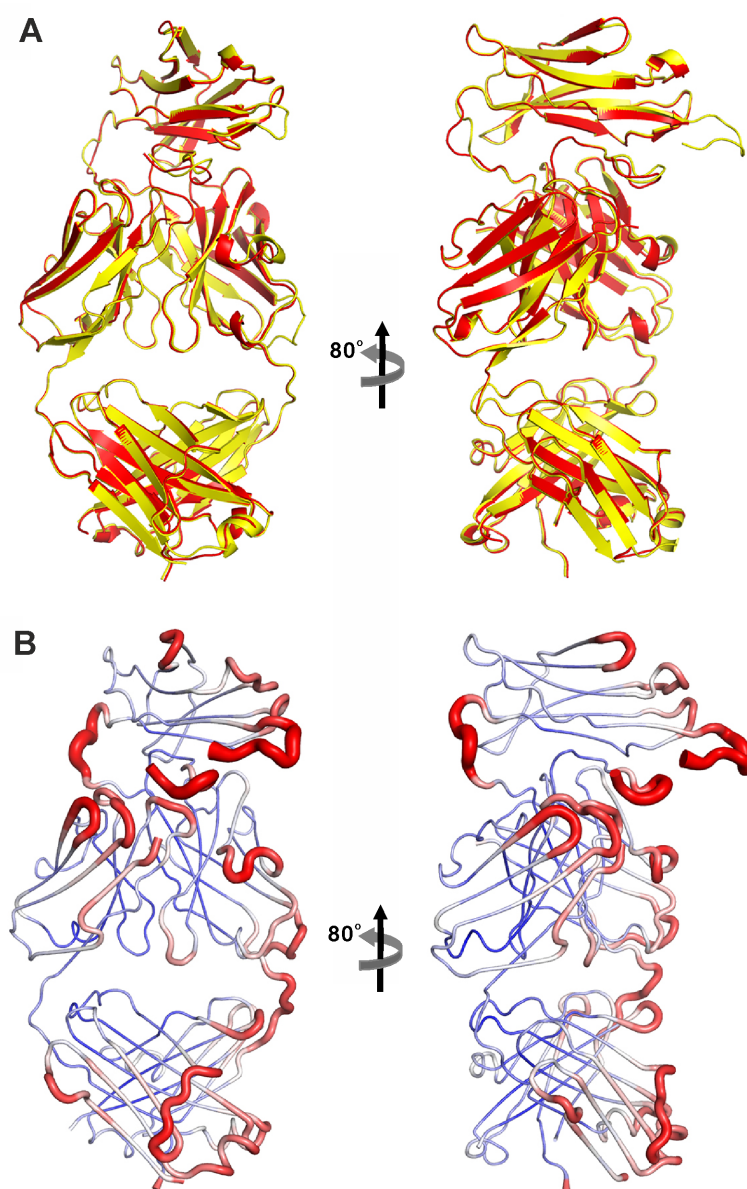


Figure 4.5. Crystal packing and order. (A). A superimposition of the two Fab-MGD21 molecules in the asymmetric unit, with chains A and B in yellow and C and D in red. (B). The structure of chains A and B shown in putty representation with putty thickness determined by B factor. The colour scale for blue to red is from B factors of 30 to 100.

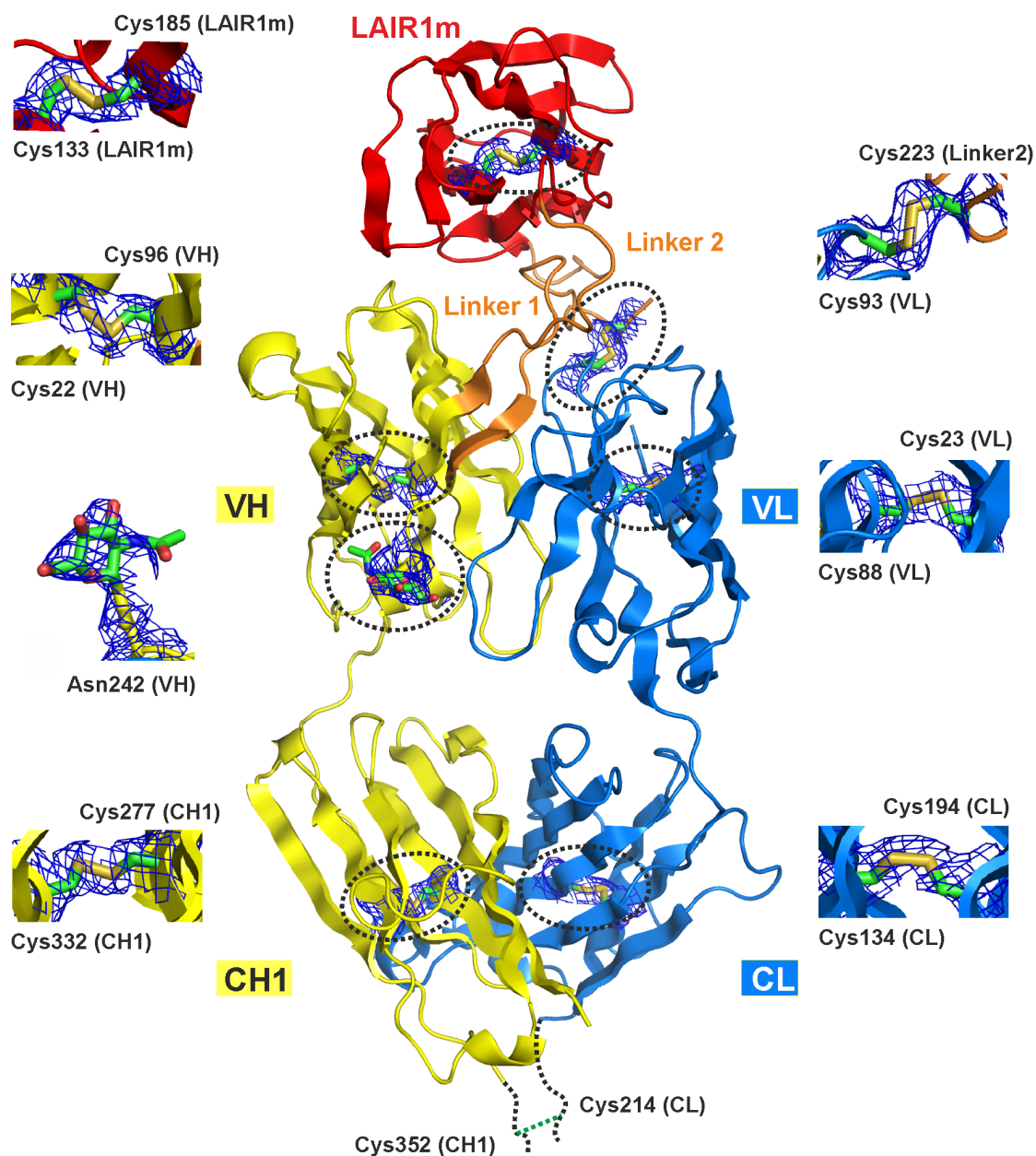


Figure 4.6. Structural features of Fab-MGD21. Cartoon representation of Fab-MGD21, highlighting LAIR1m (red), linkers (orange), heavy (yellow) and light (blue) chains. Close-up view of $2|F_o| - |F_c|$ electron density maps for disulphide bonds and a N-linked glycan on Asn242 are contoured at 1σ as blue meshes. An intermolecular disulphide bond (Cys214 and Cys352) disordered in the structure is drawn as dotted lines.

The structure shows LAIR1m emerging from the CDR3 loop of the heavy chain and lying across the antigen-binding surface of the variable domains of the Fab fragment (**Figure 4.3**). The long axis of the LAIR1m is positioned with the β -strands aligned approximately perpendicular to the groove between the heavy and light chain CDRs and the insertion and linkers interacts with, and occlude all five of the remaining CDR loops (**Figure 4.3**). The N- and C-termini of LAIR1m lie at opposite ends of its structure, necessitating long linkers between the sites from which CDR3 emerges from the antibody heavy chain and each terminus of the LAIR1m (**Figure 4.3**). The N-terminal linker (Linker 1) is 10 residues long and adopts a simple loop structure that joins the antibody variable domain to the N-terminus of the LAIR1m. The C-terminal linker (Linker 2) is longer at 34 residues and is more complex in structure. It extends out from the C-terminus of the LAIR1m before zigzagging back towards the insertion site in the heavy chain variable domain. Linker 2 is stabilised by hydrogen bonds to the LAIR1m and to the antibody heavy chain as well as by a disulphide bond of Cys223 on the linker 2 to Cys93 on the variable region of light chain (**Table 4.2** and **Figure 4.6**).

The linkers of the LAIR1m-containing antibodies sequenced to date are variable both in length and content, involving different parts of the intronic regions of the LAIR1 gene, or intergenic sequences of chromosome 13 (Tan et al., 2016). The arrangement of these linkers, which radiate away from the remainder of the antibody, will in theory accommodate almost limitless variation in both length and sequence without disturbing the packing of LAIR1 against the variable domains of the antibody.

The five CDR loops lacking the LAIR1m are representatives of previously identified canonical classes (**Figure 4.4**) (Al-Lazikani et al., 1997; Chothia et al., 1989). The most unusual residue is Cys93, which is the residue that forms a disulphide bond with linker 2, Cys223 (**Figure 4.6**). The heavy chain CDR H3 loop has a base that adopts the ‘kinked’ conformation (Shirai et al., 1999), with the loop rapidly spreading to the two termini of LAIR1m.

In previous structures of antibodies with extended heavy chain CDR3 loops, the remaining five CDRs of the antibody are exposed, with the potential to engage in antigen binding (McLellan et al., 2011; Stanfield et al., 2016; Wang et al., 2013). One of the remarkable features of the LAIR1m-containing antibody is therefore the occlusion of large parts of each of the remaining five CDRs, with these loops each interacting directly with the LAIR1m and/or linkers (**Figure 4.7** and **Table 4.2**). Both the first and second CDRs of the heavy chain directly contact the LAIR1m (**Table 4.2**).

The structure of Fab-MGD21 argues for a rigid association of LAIR1m with the remainder of the antibody. Firstly, the structures of the two molecules of the antibody superimpose closely (**Figure 4.5**). It is unlikely that this is due solely to constraints from crystal packing as LAIR1m is anchored to the variable domains of the antibody through three fixed positions: the attachment sites of the two linkers, and the disulphide bond between light chain Cys93 and Linker 2 Cys223 (**Figure 4.6**). In addition, each of the five CDR loops not containing the LAIR1m insert makes direct interactions with either LAIR1m or the linkers, through contacts found in both copies of the antibody in the asymmetric unit of

the crystal. This will stabilise a tight association between LAIR1m and the antibody variable domains. As these antibodies can include multiple different light chains, and very different linkers (Tan et al., 2016), these interaction will not be replicated precisely across the antibody family, but some variant of interaction between light chain CDRs and the linkers is likely.

Table 4.2. A list of interactions between the LAIR1m insert and linkers that occupies the heavy chain CDR3 loop and the other five CDR loops of Fab-MGD21.

CDR loop	Residue	Group	LAIR1 region	Residue	Group	Interaction
Light chain CDR1	Q27	Side chain	Linker 2	A222	Main chain	Hydrogen Bond
Light chain CDR2	Y49	Side chain	Linker 1	L102	Side chain	Hydrophobic Packing
Light chain CDR2	N53	Side chain	Linker 1	S104	Side chain	Hydrogen Bond
Light chain CDR3	C93	Side chain	Linker 2	C223	Side chain	Disulphide Bond
Light chain CDR3	F94	Main chain	Linker 2	E227	Side chain	Hydrogen Bond
Heavy chain CDR1	N32	Side chain	LAIR1 m	R134	Side chain	Hydrogen Bond
Heavy chain CDR2	R57	Side chain	LAIR1m	P109	Main chain	Hydrogen Bond

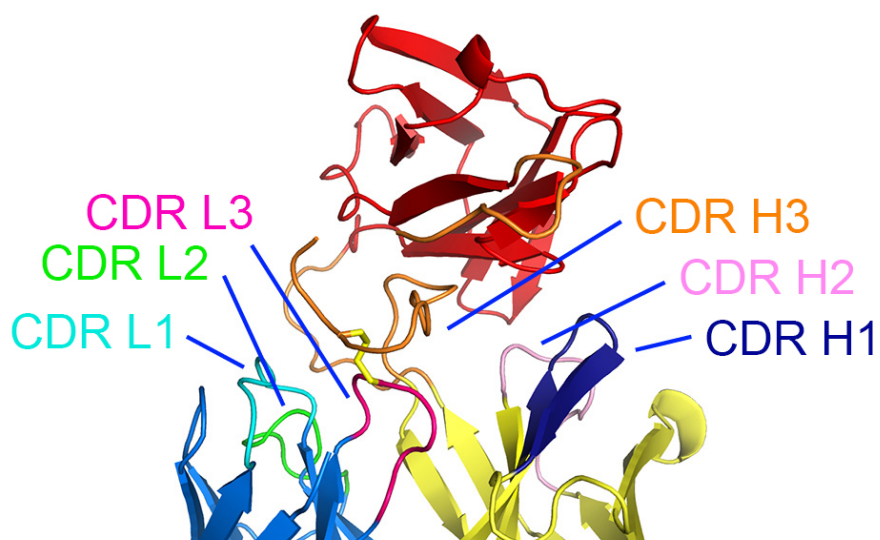


Figure 4.7. The organisation of the CDRs of Fab-MGD21. The three CDR loops of the light chain and remaining two CDR loops of the heavy chain directly contact the LAIR1m insert or the linkers. See also **Table 4.2**. Each of the CDR loops and its corresponding label is shown in a different colour.

4.2.3 Structural comparison of germ line LAIR1 and antibody-inserted LAIR1m

A structural comparison of the LAIR1m insertion in antibody MGD21 with that of the chromosomal copy of LAIR1 (Brondijk et al., 2010) reveals that no global structural changes have taken place (root mean square deviation 0.43Å for the 82 C α residues) (**Figure 4.8A**). Indeed, the LAIR1m insertion in the MDG21 antibody differs in only 13 positions relative to the LAIR1 sequence (**Figure 4.8B**). Mapping these sites onto the structure reveals that they do not alter residues through which the LAIR1m interacts with the rest of the antibody (**Figure 4.8B**). Presumably the stable interaction between LAIR1m and the antibody has therefore come instead from adaptations to the CDR loops. In contrast, polymorphisms are mostly located on the surface of LAIR1m distal to the rest of the antibody and have the potential to alter its interaction with its original ligand, collagen, and with the RIFIN proteins which are the target of these antibodies.

The normal function of LAIR1 is to interact with collagen (Meyaard, 2008). The structure of germ line LAIR1, together with NMR analysis and mutagenesis, allowed the mapping of residues critical for the collagen interaction onto a germ line LAIR1 crystal structure (Brondijk et al., 2010). In particular, mutations in residues R59, E61 and R65 have a significant impact on collagen binding (Brondijk et al., 2010). These residues map onto the surface of LAIR1 (**Figure 4.8C**) that is most exposed in the context of the antibody (**Figure 4.8D**). Indeed, mapping of the polymorphisms found in the 27 LAIR1m-containing antibodies sequenced to date shows that large parts of this surface

are mutable (**Figure 4.8E**). As shown in **Figure 4.8F**, the polymorphisms in LAIR1m include R149N, which is in the position equivalent to the residue R65 in germ line LAIR1 and this change may impact collagen binding. A second polymorphism, found in 7/27 of the antibodies (including MGD21) alters the N-linked glycosylation site at residue N69 of germ line LAIR1 (Wollscheid et al., 2009), which may alter collagen binding and/or increase RIFIN binding, but is not conserved across the antibody family. Indeed 11 of the 27 sequenced antibodies have mutations in at least one of the residues implicated in collagen binding, or other polymorphisms that reduce the interaction (Tan et al., 2016).

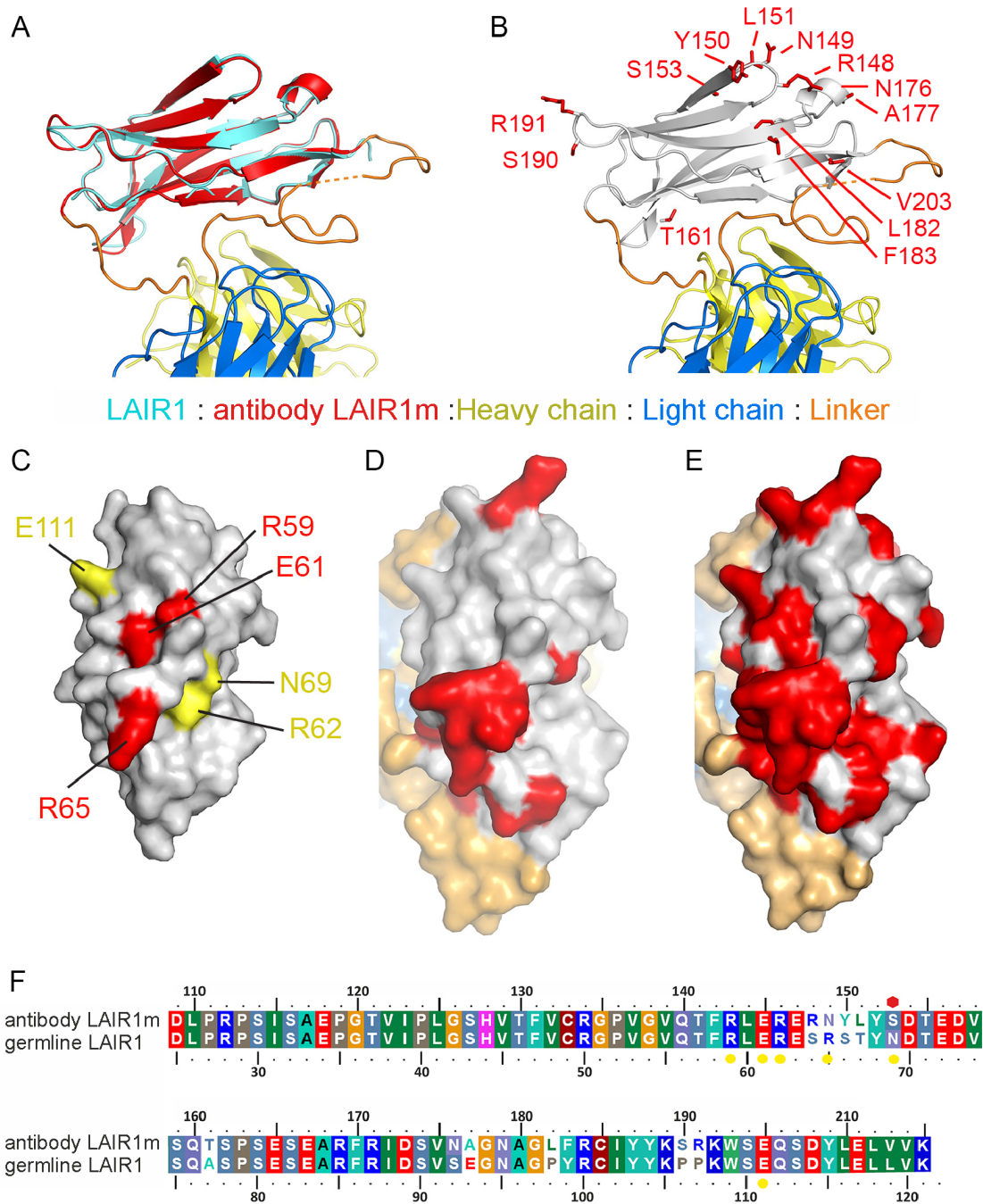


Figure 4.8. Structure and polymorphism in the LAIR1 insertion. (A). A structural alignment of germ line LAIR1 (cyan) with the LAIR1m insert of the Fab-MGD21 (red). (B). The residues that differ between the LAIR1m insert of Fab-MGD21 and germ line LAIR1 are shown as red sticks. (C). A surface representation of the structure of LAIR1 (grey) with residues whose mutation has a major (red) or minor (yellow) effect on collagen binding highlighted (Brondijk et al., 2010). (D). A surface view of the LAIR1m insert of Fab-MGD21 (grey) with residues that differ from germ line LAIR1 highlighted (red).

(E). A surface view of the LAIR1m insert (grey) with residues that differ from germ line LAIR1 in all 27 antibodies tested to date (Tan et al., 2016) highlighted (red). **(F).** A sequence alignment of germ line LAIR1 and the LAIR1m insert in the MGD21 antibody. Yellow circles are residues shown to play a role in collagen binding while a red hexagon represents a potential N-linked glycosylation site mutated in the LAIR1m insert.

4.2.4 Novel CDR H3 analysis

The LAIR1m-containing antibodies are a remarkable variant of the standard immunoglobulin fold. While the majority of mammalian antibodies have predictable and short CDRs, the third CDR of the heavy chain can accommodate unusual diversity (**Figure 4.9**) (McLellan et al., 2011; Wang et al., 2013; Weitzner et al., 2015). This is seen in the elongated CDR3 of the broadly neutralising antibodies that interact with human immunodeficiency virus (HIV) surface proteins and in the insertion of a β -hairpin and disulphide-rich domain in a fraction of bovine antibodies. However, in both of these cases, only the heavy chain CDR3 is altered and the remaining CDR loops remain exposed for antigen binding. The LAIR1m-containing antibodies are an exception to this, with the LAIR1m insert interacting with, and partly occluding, all five of the remaining CDR loops. In many ways, the structure resembles an antibody with CDR loops adapted for LAIR1m binding, into which LAIR1m has also been inserted.

This occlusion of large parts of the CDR loops by the LAIR1m insert has major consequences for its role in antigen recognition, as the majority of the antigen-binding surface is contributed by LAIR1m. Indeed it has been shown that the LAIR1m insert alone can bind to infected erythrocytes, as can a LAIR1m-containing antibody with the heavy and light chain regions exchanged (Tan et al., 2016).

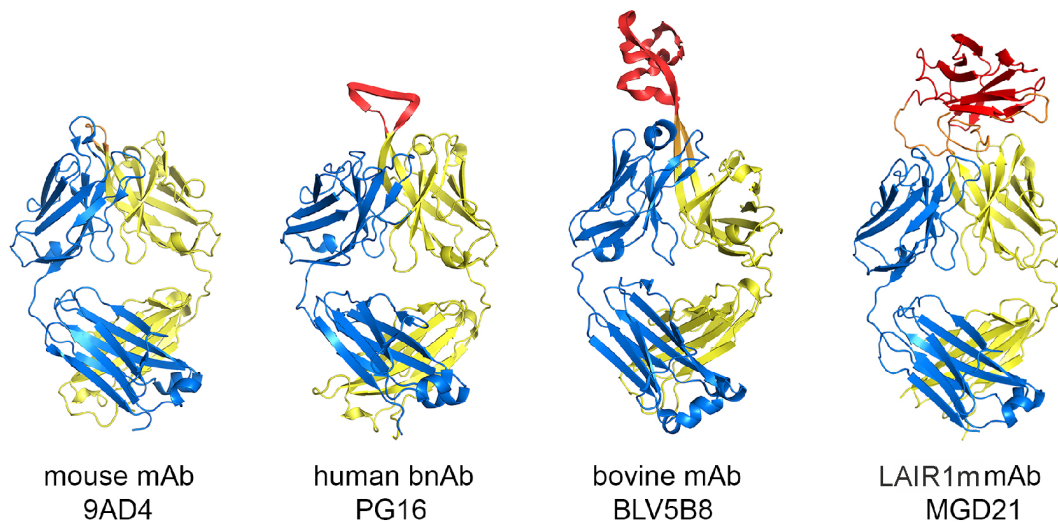


Figure 4.9. Comparison of the LAIR1m-containing antibody with other unusual antibodies. The structure of the Fab-MGD21 is compared with a classical mouse monoclonal antibody (9AD4; PDB code: 4U0R), a human monoclonal antibody with broadly neutralizing potential against HIV (PG16; PDB code: 4DQ0) and a bovine monoclonal antibody (BLV5B8; PDB code: 4K3E). In each case, the light chain is blue and the two immunoglobulin domains of the heavy chain are yellow. Inserted domains are shown in red with linker regions in orange.

4.3 Biophysical characterisation of RIFIN-ligand binding

Tan *et al*, stated that an antibody in which the LAIR1m insert has been deleted or exchanged for the germ line LAIR1 did not bind to erythrocytes which express RIFINs on their cell surface (Tan et al., 2016). However the binding affinities of LAIR1-containing antibodies for RIFINs were not determined. Furthermore, whether RIFINs bind to germ line LAIR1 alone was not reported. To address these questions, I used SPR experiments to measure the binding affinity of RIFIN for Fab-MGD21, LAIR1m, and LAIR1 (**Figure 4.10A**).

The ectodomain of a RIFIN previously identified to bind to the LAIR1m-containing antibodies (UniprotKB/Swiss-prot ID: Q8IJ06; residues 23-334) was expressed in HEK293T cells with C-terminal BAP and 6xHis tags and was purified from conditioned media by IMAC. The biotinylated RIFIN was immobilised on a SPR SA sensor chip coating with streptavidin for the binding experiments (**Figure 4.10A**). In the initial SPR experiment, I found that Fab-MGD21 showed a significant background binding to the matrix of sensor chip, particularly with the sensor CAP chip, which was used for assessment of CD36 and CIDR interaction, having a notably higher non-specific binding than sensor SA chip. Bovine serum albumin and salmon DNA were included to block the non-specific binding of chip matrix to Fab-MGD21 and bovine serum albumin reduced the background binding. The final optimised SPR method, including buffer and SPR flow-rate, is described in **Chapter 2** (section 2.5.2). However, attempts to determine the affinity binding of Fab-MGD21 to RIFIN by SPR accurately were challenging, as methods to reduce non-specific

background binding were ineffective. As shown in **Figure 4.10B**, RIFIN interacts with antibody Fab-MGD21 with a binding constant (K_d) of 12 μM .

As the LAIR1m insert was suggested to be the binding site to RIFIN, I tested whether that LAIR1m insert has the same affinity as the MGD21 Fab for RIFIN. I hence expressed the LAIR1m insert and germ line LAIR1 in HEK293T cells and purified proteins for SPR experiments to measure the binding constant. SPR sensorgrams showed that the LAIR1m insert has less non-specific binding in contrast to antibody Fab-MGD21 (**Figure 4.10B and C**). This allowed the determination of binding affinities, (**Figure 4.10C and D**) showing that the RIFIN has stronger affinity for the LAIR1m insert (K_d value of 6 μM) than for the germ line LAIR1 insert (K_d value of 17 μM) showing about a 3-fold difference in binding affinity.

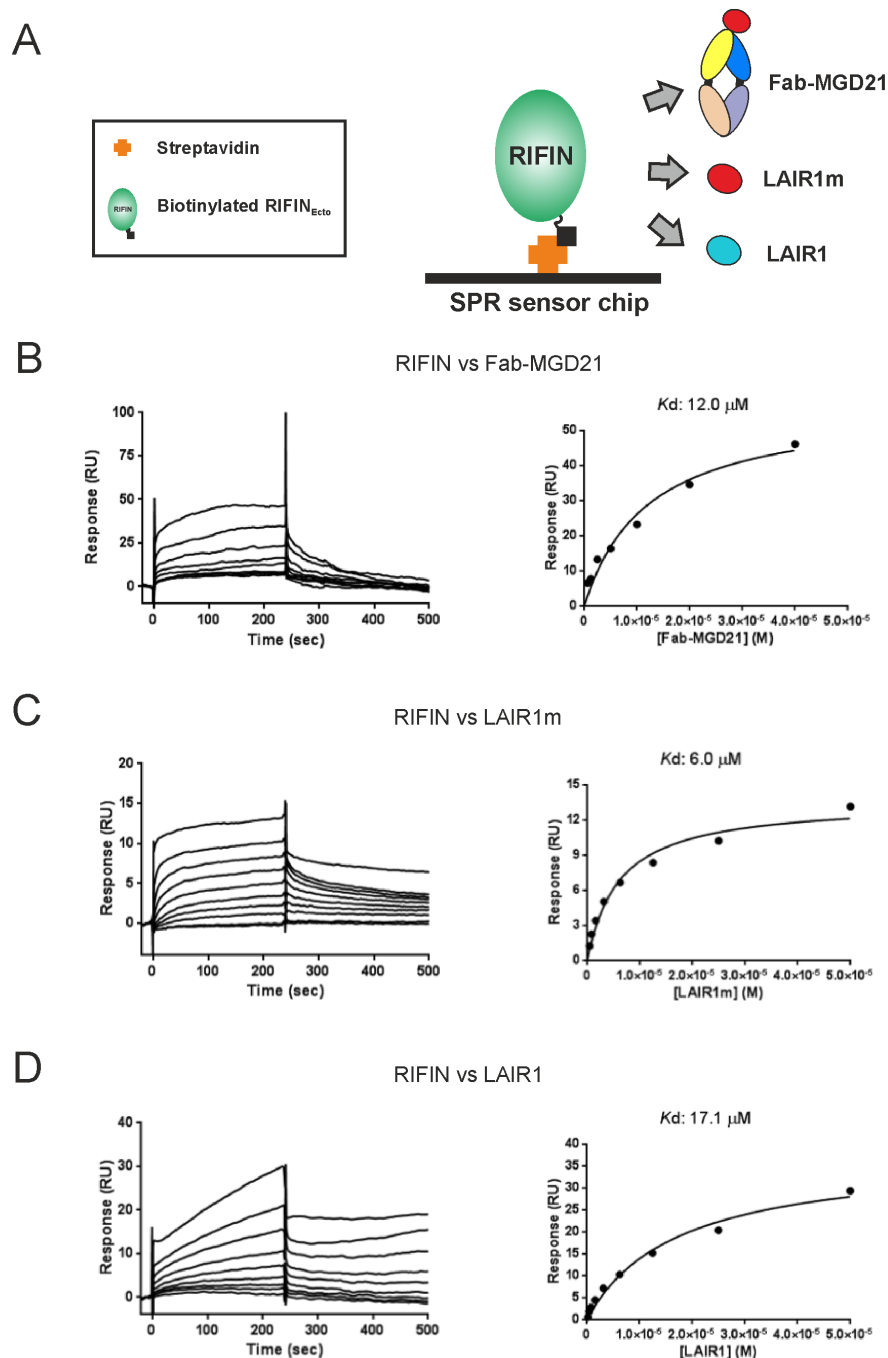


Figure 4.10. SPR experiments for the RIFIN binding assays. (A). An SPR sensor chip SA immobilised with streptavidin was used to capture biotinylated RIFIN ectodomain on the sensor chip surface. Analytes, Fab-MGD21, LAIR1 and LAIR1m, were then injected over this surface to assess their binding affinities for the RIFIN. SPR sensorgrams with reference and blank subtraction (left panels) and equilibrium binding plots (right panels) against a series of concentrations of analytes Fab-MGD21 **(B)**, LAIR1m **(C)**, and germ line LAIR1 **(D)**.

4.4 Conclusion

LAIR1, also named as CD305, is an inhibitory receptor expressed on most cells of the immune system, and its interaction with collagen suggests that it plays an important role in controlling immune cell activation (Lebbink et al., 2006). The SPR experiment (**Figure 4.10**) suggests that germ line LAIR1 may be a physiological ligand of the group of RIFINs identified as binding to LAIR1m-containing antibodies, although the binding affinity of LAIR1 to RIFINs is in the micromolar range. This moderate binding of germ line LAIR1 to RIFIN appears to then be affinity matured when LAIR1 inserted into the Fab fragment of an antibody to increase its capacity for immune recognition and recruitment of immune cells.

The structure of the LAIR1m-containing antibody shows how it uses the classical hypervariable loops for a novel function: to position an inserted auxillary domain for antigen recognition. The classical Fab fragment therefore now acts as a link between a ligand for a pathogen surface receptor and the Fc region of the antibody with immune recruitment capability. It will be interesting to investigate the biological role of these broadly reactive antibodies in the immune system. It is also fascinating to see if this is a paradigm that is repeated in other novel antibodies, as yet undiscovered.

Chapter 5

Discussion and perspectives

5.1 The *Pf*EMP1 CIDR domains

5.1.1 The opened and closed form of the MCvar1 CIDR α 2.8 domain

The CIDR domains have been divided into three groups, CIDR α 1, CIDR α 2-6 and CIDR $\beta/\gamma/\delta$; and domains structures have been determined for each of these domain classes (Hsieh et al., 2016; Klein et al., 2008; Lau et al., 2015; Ng, 2012). In each case, the CIDR domain contains a core three α -helical bundle with an insertion between the second and third core helices (**Figure 5.1 and 5.2**). This folds to make smaller helices that lie approximately perpendicular to the core helices, and the insert has been identified as a “homology block” as part of the classification of the CIDR domains (Rask et al., 2010). Three of the current CIDR domain structures solved before the structures described in this thesis adopt a similar conformation of core helices and homology block, while the MCvar1 CIDR α 2.8 domain (PDB ID: 3C64) adopts a different conformation when crystallised at pH 4.2 (**Figure 5.1 B**). This raised the question of whether this different conformation is genuine, or is the result of a pH-induced change in structure. **Chapter 3** described the structure of the MCvar1 CIDR α 2.8 domain in complex with CD36 (PDB ID: 5LGD), crystallised at pH 8.0. While the homology block of MCvar1 CIDR α 2.8 domain (PDB ID: 3C64) at acidic condition has an opened conformation, in the structure of the CD36:CIDR α 2.8 complex (PDB ID: 5LGD) it adopts the typical compact conformation found in other CIDR domains (PDB ID: 4V3D and 2YK0) crystallised around physiological pH (**Figure 5.1**).

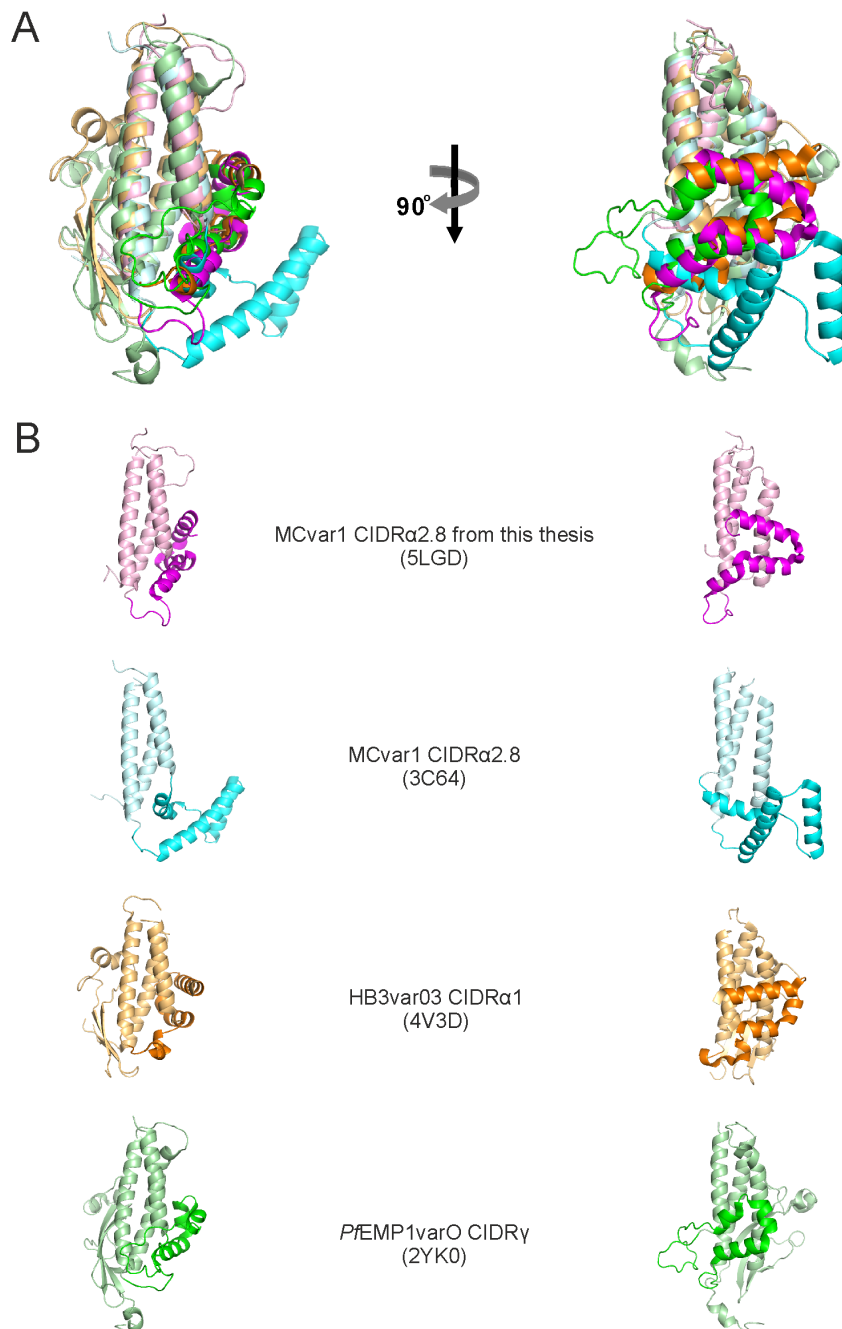


Figure 5.1. Structural comparison of CIDR domains (A) Superimposition of the CIDR domains from MCvar1 CIDRα2.8 (pink, PDB ID: 5LGD; cyan, PDB ID: 3C64), HB3var03 CIDRα1 (orange; PDB ID: 4V3D) and *PfEMP1varO* CIDRγ (green; PDB ID: 2YK0). Domains share core three α-helical bundles, and the homology blocks that emerge from between the second and third of these core helices are highlighted. **(B)** Structures show that the core helices lie perpendicular to the insertion.

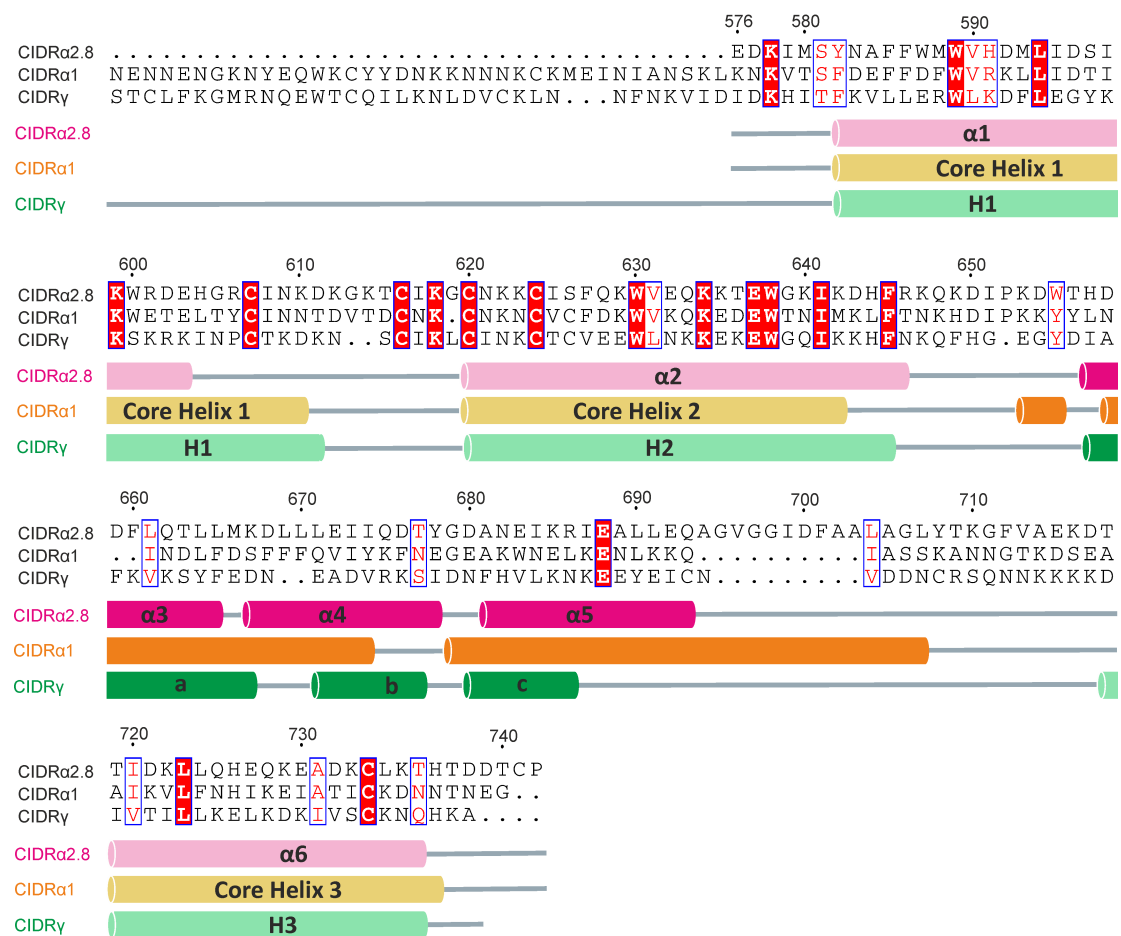


Figure 5.2. Sequence alignment of the CIDR domains. Above the alignment is the residue number from MC179 *PfEMP1*. Underneath is indicated the secondary structure (Hsieh et al., 2016; Lau et al., 2015; Vigan-Womas et al., 2012), shown in the same colours as in **Figure 5.1**.

To understand the reasons for the structural discrepancy amongst MCvar1 CIDR α 2.8 domains (**Figure 5.3**), I expressed MCvar1 CIDR α 2.8 in HEK293T cells and obtained crystals. When comparing this protein, crystallised at pH 4.5, with the equivalent bacterial expressed protein crystallised at pH 4.2 (Klein et al., 2008), the structures have the same space group ($P6_522$) and close to identical cell dimension ($a=b=95.9\text{\AA}$, $c=91.9\text{\AA}$), showing that the HEK293T-expressed protein adopts the same open structure when crystallised alone at pH 4.5. The opened conformation is therefore likely to be an artifact of crystallising the isolated CIDR α 2.8 domain in acidic conditions, and the domain adopts a more normal compact architecture under physiological conditions.

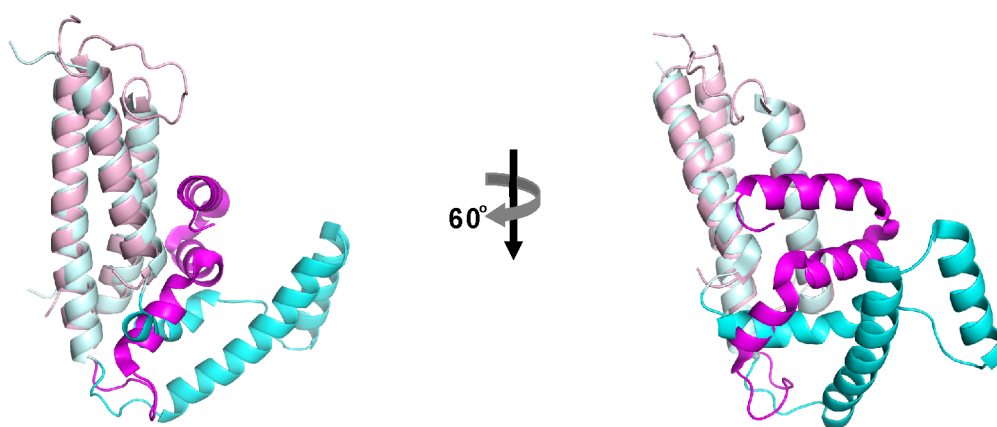


Figure 5.3. Structural comparison of CIDR α 2.8 domains at different pH values. The structure of the MCvar1 CIDR α 2.8 crystallised at pH 4.2 (cyan; PDB ID: 3C64) in an opened conformation, while that in the complex structure (pink; PDB ID: 5LGD) crystallised at pH 8.0 has a compact conformation.

5.1.2 CIDR α domains for human CD36 or EPCR

The scavenger receptor, CD36, binds to CIDR α 2-6 domains of the *Pf*EMP1 proteins and is the host receptor most commonly found to interact with parasites isolated from patients and with laboratory-adapted parasite strains. On the other hand, most CIDR α 1 domains bind to endothelial protein C receptor (EPCR). As the structures of both human receptors bound to CIDR α domains have been revealed (Hsieh et al., 2016; Lau et al., 2015), this allows comparison of how the homology blocks fold to contribute to the formation of ligand-binding sites.

A comparison of the molecular features used by CIDR α domains to bind to CD36 and EPCR shows striking similarities, but with a structural inversion of the mechanism of binding (**Figure 5.4**). In the CIDR α 1 domains, the insertion is ~ 60 residues in length and folds to form a helix and kinked-helix, with a critical phenylalanine residue (F656 in HB3var03) protruding from the kink, to provide the majority of the EPCR-binding surface (Lau et al., 2015). This hydrophobic protrusion forms the centre of the binding site, occupying the hydrophobic groove of EPCR. It is surrounded by hydrogen bonding residues that further stabilise the interaction. In contrast, the CIDR α 2-6 domains, with a ~ 70 residue insertion, adopts a subtly different structure. The helix and kinked-helix of CIDR α 1 domains are replaced by three shorter helices (α 3, α 4 and α 5 in **Figure 5.2**) in the CIDR α 2-6 domains. Together with residues emerging from the core three α -helical bundle, these form an open, hydrophobic cavity containing conserved aromatic residues, which is missing in the CIDR α 1 domains. It is this cavity that provides the majority of the

interaction with CD36, accepting a protruding phenylalanine residue (F153) from CD36. Here too, the hydrophobic core of the interaction surface is surrounded by hydrophilic regions that can make hydrogen bonds. Therefore in both cases, the domains form an extremely stable interaction using a contact surface built around a small hydrophobic core, surrounded by hydrogen bond donors and acceptors. The insertion of different homology blocks into the core three α -helical bundles of the CIDR α domains allows them to do this in different ways, presenting ligand-binding surfaces with different architectures to interact with differently structured ligands.

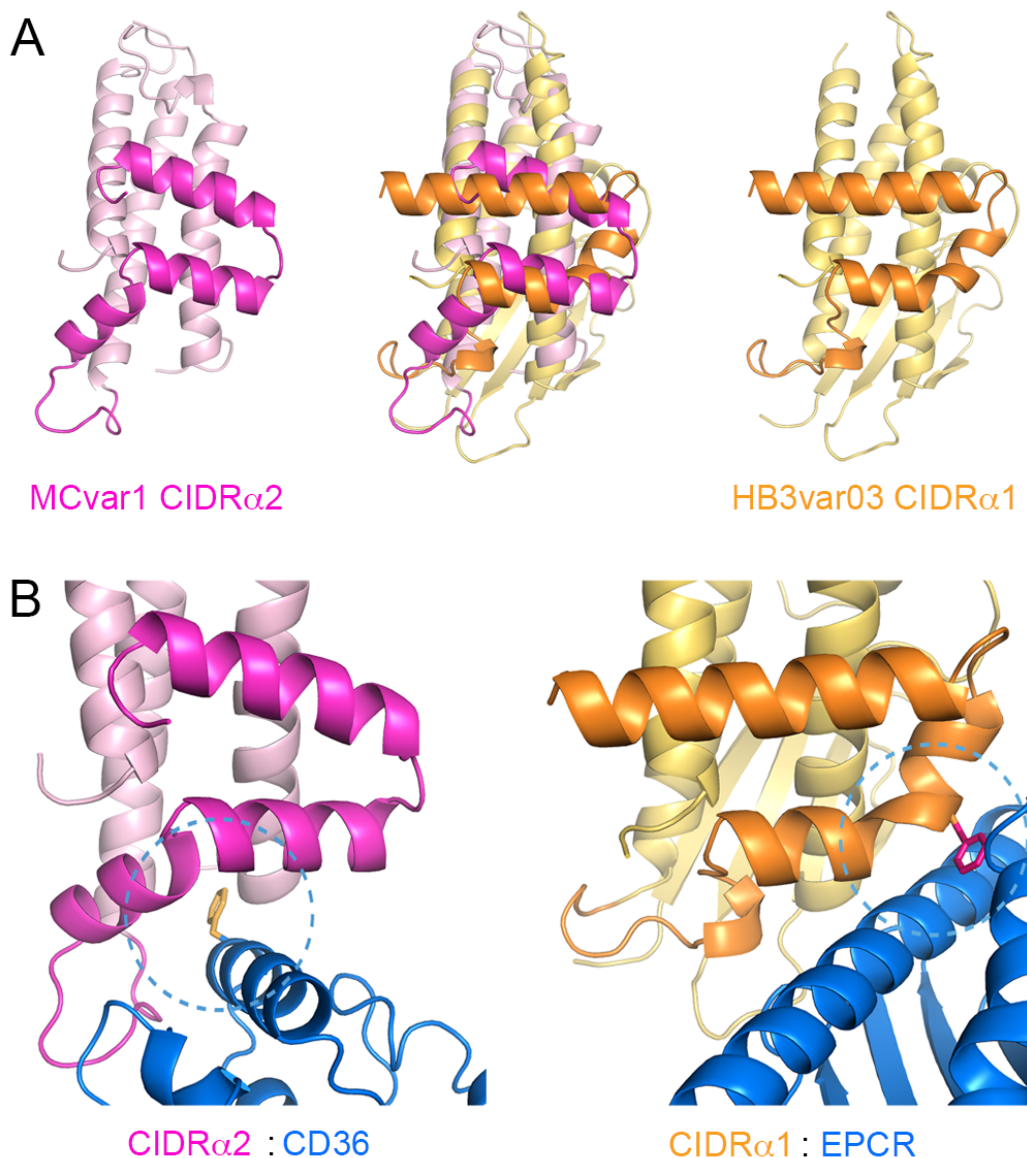


Figure 5.4. A comparison of the features that allow CIDR α domains to bind to EPCR or CD36. (A) A comparison of the structure of a CD36-binding CIDR α 2 domain (pink) with that of an EPCR-binding CIDR α 1 domain (orange). Both domains share a core three α -helical bundle. The insertion that emerges between the second and third of these core helices forms a docking platform for ligands. **(B)** A close up of the binding interfaces that mediates the CIDR α 2:CD36 and CIDR α 1:EPCR interactions. The CIDR α 1 domains have a phenylalanine residue (F656) on a convex surface of the domain that protrudes into the hydrophobic groove of EPCR. In contrast, the CIDR α 2 domains have a hydrophobic pocket that binds to a protruding phenylalanine residue (F153) from CD36.

5.1.3 A clever trick of the *PfEMP1* CIDR α domains

The *PfEMP1* proteins have evolved under conflicting evolutionary pressures. On one hand, they are constantly exposed to the immune system and are therefore under constant pressure to diversify, allowing immune evasion. On the other hand, they must retain functionally important features, maintaining the capacity to interact with specific endothelial ligands, allowing them to perform their role in endothelial adhesion. In this thesis (**Chapter 3**), I have combined structural studies of a MCvar1 CIDR α 2.8 domain in complex with CD36 with the analysis of the sequences that represent much of the global diversity of CD36 binding *PfEMP1* domains, in order to identify the molecular features that are required to mediate CD36 binding.

The CD36-binding domains present a concave hydrophobic pocket, into which docks a phenylalanine F153 from CD36 (**Figure 3.4** and **5.4**). A loop from the CIDR α domain then makes hydrogen bonds with other regions of CD36. This combination of a small hydrophobic core, together with a larger, complementary hydrophilic interface, generates a high affinity and stable binding site, allowing the domains to interact with CD36 with a slow off rate, stabilising cytoadhesion of infected erythrocytes against the strong forces of blood flow. In particular, the CD36 binding surface is surrounded by a sequence diverse protein surface containing a flexible loop (residues 647-657) that might make antibody-mediated recognition less likely. Perhaps this is why the CD36 interaction is such a successful one from the perspective of the parasite, with the binding site reducing the exposure of chemically conserved determinants from detection by acquired immunity, allowing a diverse set of

CIDR α 2-6 domains and a system of antigenic variation to provide the changes necessary to maintain the capacity to cytoadhere and to protect the parasite from splenic clearance.

A second strategy used by both EPCR- and CD36-binding CIDR domains is to interact with functionally important sites in their human receptors to avoid the evolution of escape mutants in the human genome. The CIDR α 1 domains bind to the region of EPCR that is required for its interaction with its ligand, protein C. The conserved phenylalanine of the CIDR α 1 domains occupies a pocket that is also occupied by a phenylalanine from protein C. The parasite has evolved a similar trick in its CIDR α 2-6 domains (**Figure 3.16 and 5.4**). Here the CIDR domains interact with a phenylalanine residue that protrudes from the surface of CD36. But why are polymorphisms in this residue not favoured in malaria endemic regions, as these would be expected to allow escape from binding? It is shown here that this residue is also important for binding to oxidised LDL particles, one of the physiological binding partners of CD36, and that the CIDR α domains prevent oxidised LDL from binding (**Figure 3.15**). Once again, the *Pf*EMP1 proteins target a functionally important site, reducing the likelihood of a host escape mutant that would prevent infected erythrocytes from sequestering away from splenic destruction.

In addition, the region of CD36 on which this phenylalanine is displayed is the site of recognition for long chain fatty acids (**Figure 3.9**). Fatty acids can be cross-linked to K164 of CD36 and are then thought to move across to enter

the central CD36 cavity and pass through the CD36 ectodomain to the membrane surface. The CIDR α 2.8 domain lies directly in between K164 and the cavity entrance, potentially blocking lipid uptake (**Figure 3.16**). With fatty acid uptake and oxidised LDL binding linked to fatty acid metabolism and atherosclerosis, it will be interesting to see whether CIDR α 2-6 domains can prevent fatty acid uptake as a proof of principle of targeting this region of CD36.

In summary, the characterisation of the structural basis for CD36-binding by *Pf*EMP1 has shown how the complex *Pf*EMP1 protein family balances the need to adhere tightly to unchanging human receptors while diversifying to evade antibody-mediated recognition. Functionally important regions of human receptors are targeted, reducing the likelihood of mutations that disrupt binding, but also, inadvertently blocking the physiological roles of these receptors, with the potential to lead to disease outcomes. While human receptors are restricted from evolving away from the interaction, the *Pf*EMP1 proteins are hugely diverse, retaining their structure and the chemical features of their binding site while altering nearly all of their surface residues. In this way, the parasite has evolved a surface protein family that can maintain the capacity to tether it away from splenic clearance, while avoiding clearance by the acquired immune system.

5.2 Insights from LAIR1m-containing antibody structure

5.2.1 Unusual antibody structures

Compared with conventional antibodies, some species have unique antibody (Ab) architectures (**Figure 5.5**). Camels can generate unusual Abs with only one heavy chain variable region, devoid of pairing with light chain, and these are termed heavy-chain antibodies (HCAbs) (**Figure 5.5**). Thus the Fab domains found in conventional antibodies are replaced by these smaller domains referred to as VHH or nanobodies in camels (Ge and Elghetany, 2005; Muyldermans et al., 1994). Today, engineered camel nanobodies are exploited in the pharmaceutical industry for clinical development. On the other hand, the immunoglobulin new antigen receptor (IgNAR) in the shark has a long-Fc region (**Figure 5.5**). IgNAR has five constant domains and one variable domain in each chain and forms dimers linked by disulphide bonds (Greenberg et al., 1995; Nuttall et al., 2001; Roux et al., 1998).

While camels and sharks use unusual subunit compositions to generate unique examples of antibodies, cow antibodies have ultra long CDR H3 regions, revealing a novel mechanism of genetic and structural diversity (**Figure 4.9**) (Wang et al., 2013). In fact, it appears that the CDR H3 of around 10% of cow antibodies have lengths of up to 69 residues, while in human they are up to 35 residues long and in camel, up to 24 residues (Griffin et al., 2014; Zemlin et al., 2008). The structure of these cow antibodies, with their ultra long CDR H3, revealed that their CDR H3 are formed from a β -ribbon stalk topped by a knob that leads to diversity in antigen recognition (Stanfield et al., 2016; Wang et al., 2013). Structural analysis along with sequence data

suggests that the stalk can be of variable length and the knob can have different disulphide patterns as well as differences in surface shape and binding potential. In **Chapter 4**, the structure of the LAIR1m-containing human antibody revealed an even larger CDR H3 region, where the conventional CDR H3 region was replaced by a human receptor LAIR1 insert, for antigen binding (**Figure 4.9**). Now the structures of both antibodies with unusual CDR H3 regions are revealed, it can be used to compare the differences.

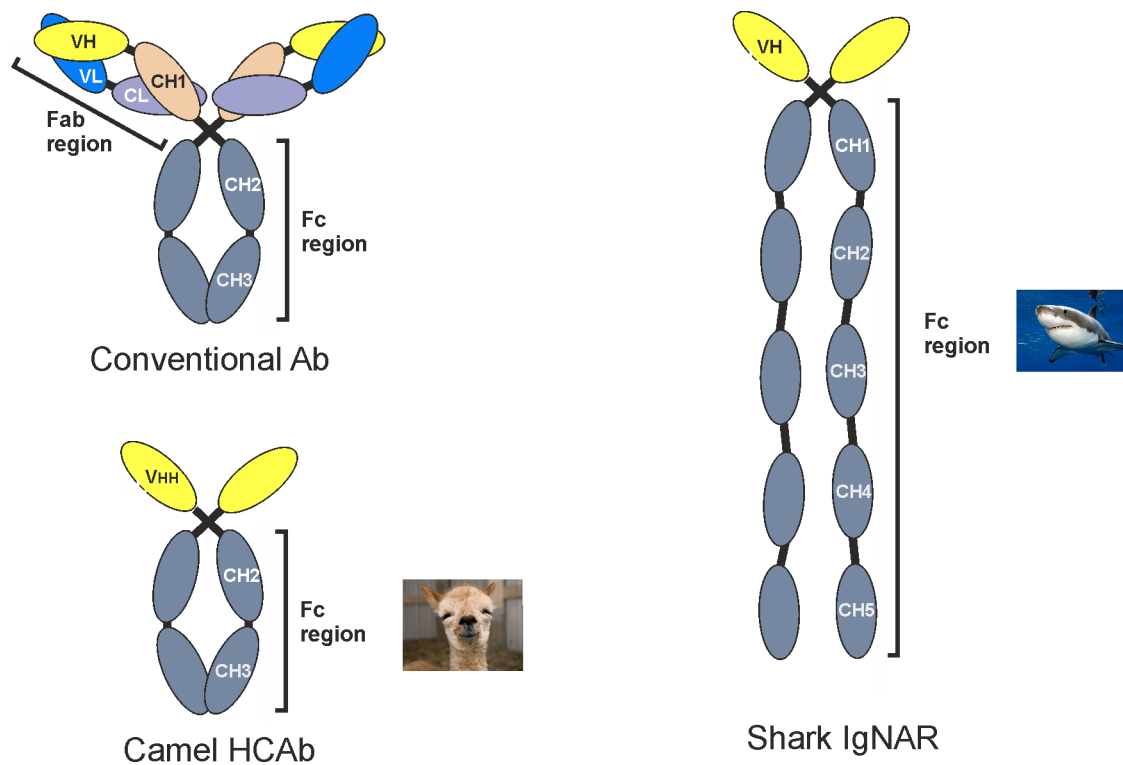


Figure 5.5. Schematic representation of different antibodies. The conventional antibody (top left) includes two light chains (variable and constant regions, VL and CL) and two heavy chains (variable region, VH, and three constant regions, CH1, CH2, and CH3) forming a functional molecule. The camel antibody (lower left) consists of heavy chains only in the antigen binding region. These heavy-chain only antibodies contain a single variable domain (VHH) and two constant domains (CH2, CH3). IgNAR antibody from shark (right) has five constant domains and one variable domain in each chain, forming a dimer for antigen binding.

5.2.2 CDR H3 comparison

Superposition of variable domains of the cow antibody and human LAIR1m-containing antibody reveals a similar structure within VH, apart from the CDR H3 loop, with a root mean square deviation of 0.48 Å (calculated over 83 C_α atoms) (**Figure 5.6**). Five CDRs (L1, L2, L3, H1 and H2) adopt canonical forms, while the CDR H3 use different scaffold to accommodate diverse structures. The ascending and descending β-strands aligned with linkers of variable length serve as a support to position CDR H3 on the top. The linkers in the LAIR1m-containing antibody are stabilised by other CDRs and a disulphide bond between the antibody light chain and linker, while the stalk in the cow antibody is stabilised by other CDRs (**Figure 4.6** and **5.6**). In both structures, antigen-binding sites are thought to be mediated through the top domains (LAIR1m insert of human antibody and knob of cow antibody). Interestingly, cows perform V(D)J recombination for antibody generation, similar to that of other species, but with limited potential for combinatorial diversity. The repertoire of an ultra long CDR H3 derives from a single germ line rearrangement, with diversity arising mainly at the V-D and D-J junctions and somatic hypermutation (SH) acting to generate further diversity prior to antigen exposure. In human LAIR1m-containing antibody, on the other hand, a novel, and little characterised mechanism inserted germ line LAIR1 between V and DJ segments and carried a high load of somatic mutations after stimulation by malaria antigens to contribute antibody diversification (Tan et al., 2016). The transposition of LAIR1 sequences into V-DJ gene segments is the first example in which a novel binding domain is inserted and modified to generate functional antibodies. It is not clear how often this novel mechanism

may happen and whether other fundamental human domains can serve similar roles as LAIR1 insertion. It will be very interesting to find out more similar but unusual antibodies in malaria-endemic areas that may provide new ways to eliminate malaria. Meanwhile, this novel mechanism of generating antibodies raises the question of whether this is a specific case in malaria infection or a general mechanism in other infectious diseases.

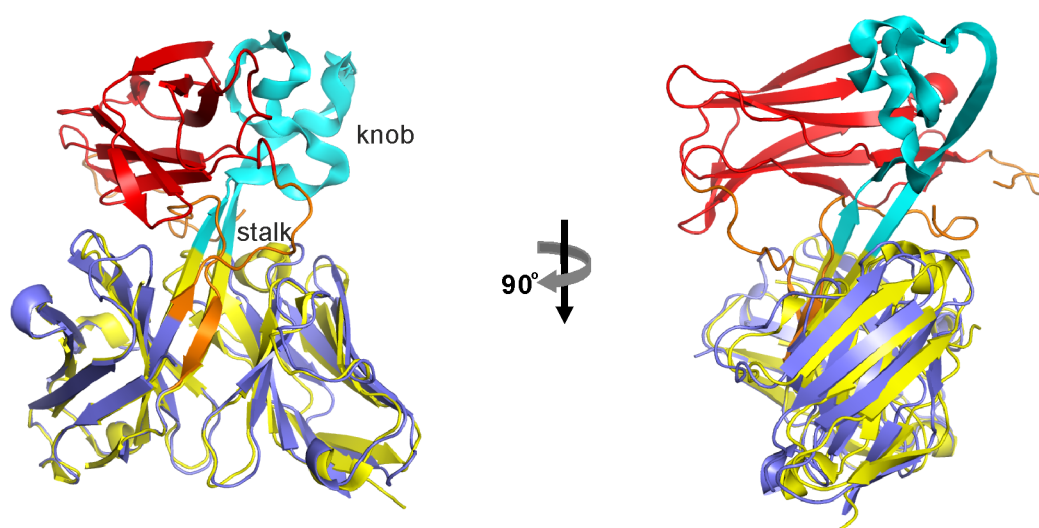


Figure 5.6. Structural comparison of cow antibody (BLV5B8, PDB ID: 4K3E) and human LAIR1m-containing antibody (PDB ID: 3KGR). Structural alignment reveals both antibodies hold similar architectures in variable regions apart from CDR H3 (human variable region shown in purple, cow variable region shown in yellow). The cow CDR H3 is shown in cyan while the LAIR1m insert in red. The linkers connecting the LAIR1m insert to VH are in orange.

5.2.3 The antibody engineering

Malaria is one of the major global diseases and affects hundreds of millions of people. The global efforts of improved quality of health systems and disease management have reduced the number of cases of infection and of malaria-related deaths over the past decade. While the mortality rates are gradually decreasing, mild or uncomplicated malaria, leading to persistent low-grade parasitaemia is a potential threat for malaria control and elimination. Individuals with asymptomatic malaria can generate broadly neutralisation by their innate or adaptive immune mechanisms to avoid clinical symptoms. It is therefore clear that these antibodies can serve as potential therapeutic tools to reduce parasites from the blood, but the main challenge is how to generate complete sterilising immunity.

RTS,S (MosquirixTM) is the only approved vaccine so far to show some protective effect in Phase III clinical trials (Heppner et al., 2005; Rts, 2015). It is a recombinant protein-based malaria vaccine against *P. falciparum*-infected malaria that works by preventing parasite infection at pre-erythrocytic stage in liver. However, a relatively low efficacy (25-50%) leaves much grounds for improvement. Although it is very challenging to transform naturally occurring antibody responses to parasite proteins into high neutralization in immunized patients, engineered LAIR1m-containing antibodies may be able to offer potential therapeutic strategies to malaria treatment. As shown in **Figure 4.10**, antibody Fab-MGD21 and RIFIN bind together relatively weakly. This weak-binding affinity can be strengthened by either engineering of the LAIR1m insert or by creating bi-specific antibodies, with one Fab fragment targeting

the antigen and the other fragment targeting an alternative antigen or an effector to enhance the binding. The engineered product could then be used as a potential therapeutic agent for malaria infection, a bio-marker or bait for RIFINs detection. Second, the structural architecture of Fab-MGD21 suggests that many human proteins might be inserted into the CDR H3 region to create novel chimera antibodies, which may lead to new discoveries and applications in biomedical research. For example, if the CDR H3 region is replaced by engineered CD36/ICAM/EPCR or other human proteins, will this manipulation be sufficient to generate broadly reactive antibodies as well? If yes, maybe these chimeric antibody cocktails can serve as sensor-makers to diagnosis malaria infection even in cases of low-parasitaemia.

5.3 Further questions and perspective

However, bear in mind that the SPR experiments which show RIFINs binding to LAIR1m-containing antibodies, presented in **Chapter 4**, are not yet completed and a few technical challenges need to be addressed. First, LAIR1m-containing antibodies show significant background binding to the surface matrix of the SPR sensor chip, making SPR measurements challenging. As shown in **Figure 4.10**, RIFIN has a weak binding ($K_d = 12 \mu\text{M}$) to Fab-MGD21 in an *in vitro* SPR binding assay. On one hand, this weak binding may be an underestimate because the LAIRm insert shows less non-specific background binding to the chip matrix and stronger binding ($K_d = 6 \mu\text{M}$) to RIFIN. Furthermore, it appears to be possible that the LAIR1m-containing antibody has a tight binding to RIFIN when RIFIN is expressed on infected erythrocyte surfaces as the binding affinity could be strengthened if RIFINs form multimers, leading to binding avidity.

The effort of malaria control over the past decades has constrained its spread in tropical and subtropical regions. However, the ambitious goal of malaria elimination still has a long way to go. In order to reduce the clinical impact of malaria, most current efforts are focusing on severe malaria. However, it is important to also focus on mild, asymptomatic malaria if aiming to eliminate malaria infection. Meanwhile, it is possible that the persistent low-grade parasitaemia would progress to severe malaria by unknown mechanisms; and new strain infection would also let the malaria genetic pool generate new severe malaria via mutagenesis. In the thesis, the structure of CD36 in complex with *Pf*EMP1 CIDR domain provides atomic level information to

explain how parasites take advantage and human hosts make balance to survival. Like CD36, which has multiple functions in the immune system, LAIR1, an inhibitory receptor found on natural killer cells, T cells and B cells, also plays roles in the immune response. To date, only one study addresses the connection between LAIR1 and malaria. It will be fascinating to understand what roles LAIR1 plays in parasite-induced immune responses. It will be interesting to understand the structural detail of the LAIR1m-containing antibody and RIFIN interaction. Perhaps, combining structural information to design a magic chimera antibody will be an exciting and novel strategy against malaria infection.

Appendices

Appendix 1. Polylinkers of pHLsec-based vector variants

1.1 The pHLsec vector

The N-terminal Kozak sequence is highlighted in green and chick receptor protein tyrosine phosphatase sigma signal peptide in cyan, the C-terminal 6xHis tag (KHHHHHH) is shown in red and stop codons in black. Constructs can be cloned into the vector by using AgeI (pink) and KpnI (purple).

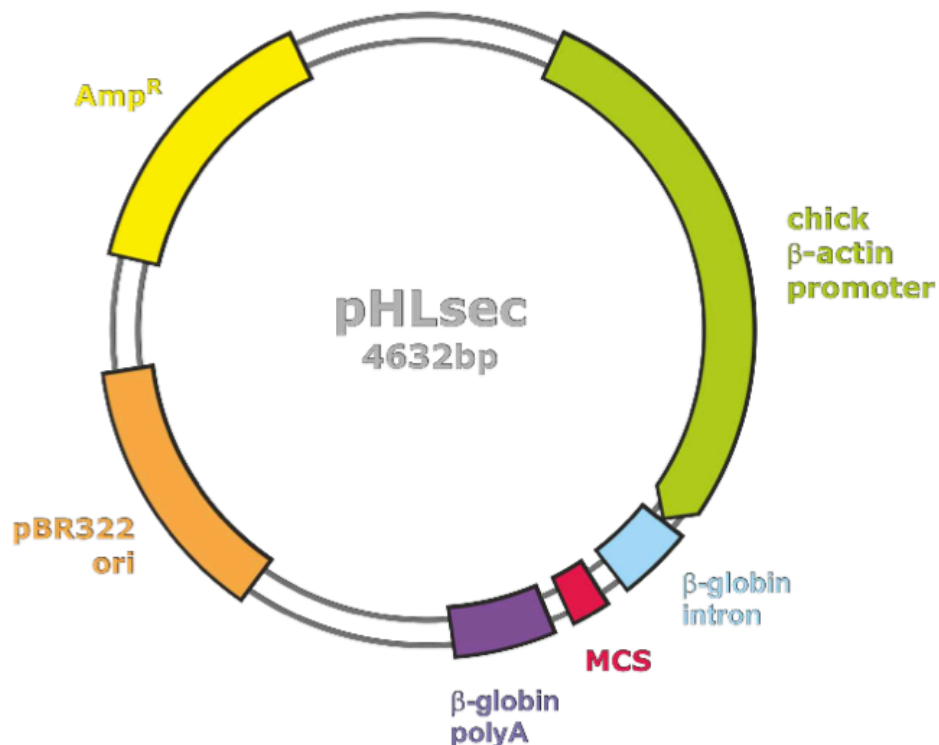
```

EcoRI HindIII Kozak M G I L P S P G M P A L L S
GAATTCAAGCTT GCCACCATGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCC

  L V S L L S V L L M G C V A E T G
CTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGTAGCTGAAACCGGT...insert...

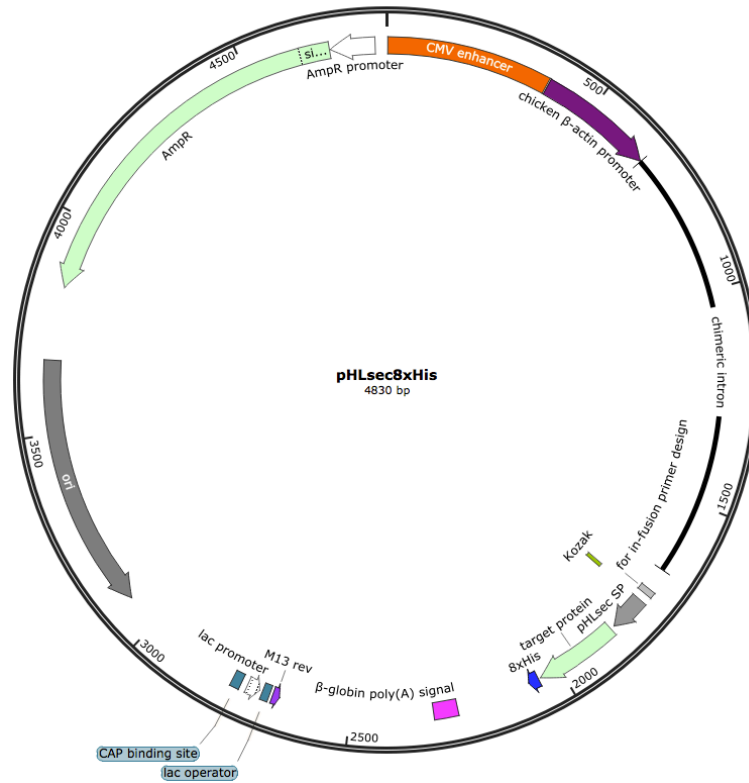
  G T K H H H H H H * * XhoI
GGTACCAAGCACCACCATCACCATCACTAATGATCACTCGAG
  
```

The vector map of pHLsec is shown as follows. Figure provided by Dr. A. R. Aricescu and based on reference (Aricescu et al., 2006).



1.2 The pHLsec-8H vector

The vector is modified from pHLsec vector, adding two more His tag in the C-terminal end. The vector map is drawn by SnapGene.



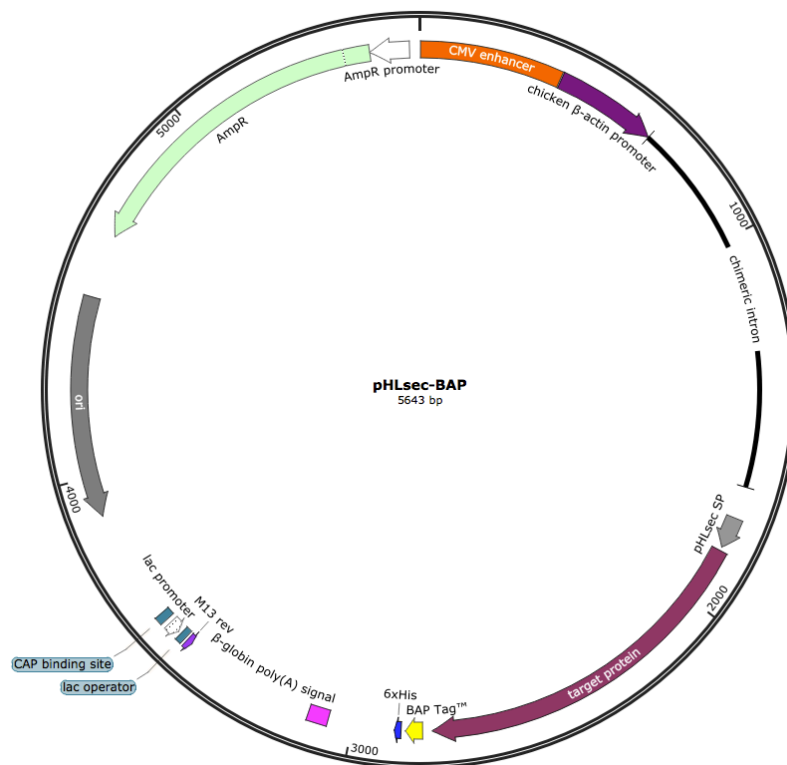
1.3 The pHLsec-BAP vector

The vector contains GGS linker regions lighted in orange, followed by the biotin ligase (BirA) recognition site in blue and finally 6xHis tag (KHHHHHH) in red (Aricescu et al., 2006). Constructs can be cloned into the vector by using EcoRI (pink) and KpnI (purple).

EcoRI
GAATTCinsert.....

KpnI
G T G G S G G S G L N D I F E A Q K I E
GGTACC

W H E G R T K H H H H H H * * XhoI
TGGCATGAAGGACGTACCAAGCACCACCATCACCATCACTAATGATCACTCGAG



Appendix 2. Primers used in the thesis

2.1 Primers for cloning CD36 and MCvar1 CIDR2.8

CD36 ectodomain(35-439)	
Primer 1	ctccaccggtcagaagacaattaaaaagcaag
Primer 2	gtggggtaccctattagttattttccagttacttgacttc
MCvar1 CIDR α 2.8	
Primer 1	ctccaccggtgaggacaagattatgagctac
Primer 2	gtggggtacctggtgggcatgtgtc

2.2 Primers of mutagenesis in CD36 and MCvar1 CIDR α 2.8

CD36_F153A	
Primer 1	catatctatcaaaatcaagccgttcaaattgac
Primer 2	gatcattgaacggcttgatttgatagatatg
MCvar1 CIDR α 2.8_F645A	
Primer 1	aagatcaaggaccacgccagaaagcagaaggac
Primer 2	gtccttctgctttctggcgtggctctgatctt
MCvar1 CIDR α 2.8_Q648A	
Primer 1	gaccactcagaaaggccaaggacatccccaag
Primer 2	cttgggggatgtccttggcctttctgaagtggtc
MCvar1 CIDR α 2.8_D650A	
Primer 1	ttcagaaagcagaaggccatccccaaggactgg
Primer 2	ccagtccttggggatggccttctgctttctgaa
MCvar1 CIDR α 2.8_L664A	
Primer 1	gactttctgcagaccgccctgatgaaggacctg
Primer 2	caggtcctcatcaggcggtctgcagaaagtc
MCvar1 CIDR α 2.8_E672A	
Primer 1	aaggacctgctgctggccatcatccaggacacc
Primer 2	ggtgtcctggatgatggccagcagcaggctcctt
MCvar1 CIDR α 2.8_F660N:Q662T	
Primer 1	tggaccacgacgacaacctgaccacctgctgatgaag
Primer 2	cttcatcagcagggtggcaggtgtcgtcgtgggtcca
MCvar1 CIDR α 2.8_D680N:N682T	
Primer 1	caggacacctacggcaacgccaccgagatcaagcggatc
Primer 2	gatccgcttgatctcgggtggcgttgccgtaggtgtcctg

2.3 Primers for Fab-MGD21 fragment

Heavy chain_variable region	
Primer 1	gatgggtgcgtagctgaagtgcagctgggtgaaacaggc
Primer 2	gggtgtcgtttggcgctagacactgtcacgggtggtgcc
Light chain_variable region	
primer 1	gatgggtgcgtagctgccatcagaatgacctcagagcccc
primer 2	gtgcagcatcagcccgttgatttcagccgggtgccc

2.4 Primers for RIFIN fragment

RIFIN	
Primer 1	ctccaccggtcgggtgtacaacaagcggaaccacaag
Primer 2	gtggggtaccggtctggtgctcatgaagatgg

2.5 Primers for LAIR1 fragment

LAIR1	
Primer 1	ctccaccggtcaagaagaggacctgcctagacctagc
Primer 2	gtggggtaccctctttgaccagcagttccaggtagtc

2.6 Primers for LAIR1M fragment

LAIR1m	
Primer 1	ctccaccggtcccagcgggataccgacctg
Primer 2	gtggggtacccttcacgaccagttccaggtagtcg

Appendix 3. CIDR domain sequences for recombinant proteins

> CIDR α 2.8 MCvar1

EDKIMSYNAFFWMVHMLIDSIKWRDEHGRCINKDKGKTCIKGCNKKCICFQKWVE
QKKTEWGGIKDHFQKQKDI PKDWTHTDDFLQTLLMKDLLLEIIQD TYGDANEIKR IEA
LLEQAGVGGIDFAALAGLYTKGFVAEKD TTIDKLLQHEQKEADKCLKTH TDDT C PPQ
EDRSVARS

>CIDR α 6 IT4var12

PWCGLKKQADGTWKRLYENDPQCPIKPKYEPPKGEEPT EIDVLYTGKENKDIIVKLR
EFCKTDGNTGFKNEEWNCYYQVGNDKCVLENGEELGGEKKVKDYDNFLMFVVAHMLK
DSIEWRSKLSNCLKSDKKT CIKKCNDNCKCYEKWIGKKKVEWTQIKKHFDKQTD FQG
WGRYFVLETVLEGDQFFTDITKAYGDAREIVHIQEMLQKKKEQVLHEDASNMKTIID
ELLDHELKEAKQCIVNHKDNNC

>CIDR α 5 IT4var14

PWCGIEEQKDGGKWK RINDHSACKEEELYTPKENAKYTKINVLTS GEGHEDIAKRLKE
FCTKTQNGGGGSDDCGGNSDSSLCEPWQCYQPDQLEKVG GGEVDDKLGAGGLCIFE
KMKGEKKVKKQKTFNNFFNFVVAHV LKDSIDWRTQLTKCLSEDK LKKCEKGCKSNCE
CFKKWIEKKEKEWIKVKDQFNKQTD FLEWKHYLVLETILENYYFENIQKAYGDLKSI
QEMKKMIKENKQKNRTKDDEDALDVLF DHEKEEAEDCLDIHE

>CIDR α 2.2 IT4var24

PLCGVNGPKGNWTDIEDKECTLVEKKKYDSNNTTNIDILTADKSQKNILQKYNKFCK
NGANDEKSATPTANGGGQIKNWQCY YDESKE SGQNNNCILGKWEDFTGNEDVTSYNV
FFYNSIIEMLNDSIEWKDKLNICINNKTGKCRKVCKNPCECYKRWIEKKKTELEKIK
DHFRKQKDIGDAAQREMTLNITLNNN FLNDIKDAYPVKQQLQKIEERLKD KMQEDFI
FERTKTSIDKFLQEEEQFAEKCIEKHKC

>CIDR α 2.10 IT4var30

PWCGIEEQKVDGKWKAKNNGECAKVENKIYDPKKT TTIEILTGDNEKFGIYEKYNKF
CANGKNGVTSGAPDTAPGKNGNQIVTWQCY YDKKKRSGQNNNDNCVEGTWEKFTGKQT
VKSYNVFFWDWVYHMLHDSLEWRERLNSCINNAKSQNCKNNKCNRECGCF AKWVVKK

KDEWDKIKDHFNKQENIGQNVGFIEFNHYGVLEGVLDKDELLKNIKDTHADAKDIER
IDKMLQQAGVVGGGRGGENNTIIDKFLQEEERFAKDCQOKQEEC

>CIDRq2.4 IT4var33

PLCGLGMQEPPWNPKKDTCEDRGIKTFDDRNSTPINLLVKDVTGTSIVEKLGGLCG
NGAKKNIQTWKCRFESSQNYCYVLQNDKKNTPQQEIESFNSLFWHWITEMLKDSIDW
RKEHENCINNNNTCKKGCKSKCECFEKWVKRMKEEWKQVLEVYDKQPDFKEVFTPYF
TLGYLLKEYFTKIKAPYEEVESVQEFIKEMEQIIDENSNNINATKENNSITKFLQHE
EGIATECKKTHNEEK

>CIDRq2.9 IT4var45

PWCGVNGTKGNWTPKNDDCKPGNEYKNYENTQIPILTGDKTKSEIVERYRKFCNN
GKNGANGREGGVGGSENGAASNSDNATTGycGTNNNDKDPslCEKWTCYYKKKEKND
GKKAINFCVLQDDKVGKSQEKSMHYNSFFWDWVYHMLHDSLDWRKQLGRCINKKEET
KCIGSCNKKCECFAKWVGHKQQEWEQIKKHFLKQDDIGQETNCDPMVTLEILLDIDE
LLKNIKDTHANADDIDRIQNMLQQAGFDGGVAALVGRCTEGFVAEKDTTIDKFLEKE
LQEAekCLETHKEK

>CIDRq2.7 IT4var61

PWCGVKNRGGGWKAKSDGECakeKKTYKKKNITEIPVLTPDKEKHNILQKYENFCKN
SDGNNGDQIKNWQCYEEKDESNDGDSNICVLQENENIGKKEEKSMpyHPFFWKWVT
EMLIDSMYWRKELKRCINKETKACKNGCKNNCDcyKRWVEEKKKEWGQIKTHFNTQE
DMREDIGENTDPGIILAALLNIEDLFENIKDtyGDVKEIKDINQMLEKENEENEGTA
GADSKKKNTIDLminHEQKDAQKCVTNNPDkDC

>CIDRq3.2 HB3var13

PLCGVKKEKGTWVRKDSMNDCPRIKLYKPINDKVGTPINFLYSGDGQTEIAEKLKkf
CRTENGSDGSSSARANGASGDKNGGSGSQELYQYWKCYQIGDLQKVREGEDDEDGQ
YDQEEVENAGGLCILQKTNGKENVNKQKTSHEIQKTFNPFFYYWVVHMLKDSIHWRTE
KIKSCINNTNELKACKNNKKCNSDCDCFKRWVKQKGKEWGQIKVHFktQDIRGKVVN
GNTVVSFFLDHDELLEGVLDKGLLLESQEAyGNAEDIKHikLLQETDVVGGGEHK
TTIDFLLKEELNEAEELKKQKDNEC

>CIDRq3.5 IT4var15

PGCGVELIGNEWKEKNKGECKGGKRYNIPKGTKHNVIPVLSFGDEHKEIIEKIEQFC
AESNSDSSKLTEQWKCYGDKEYEVCTLENRNKSEEDPEEIQKTFHNFFYFWIRHLL
NDSIEWRDKINNCIEKAKEGKCKNECKTDCGCFQRWIGKKKEEWGEIKKHFKTQDGF
SIFGNNDYDFVLNVLNIDELFQDITEAYGNSQKIQGIKDTLAKKKTQAADDATEQKN
TIDLLFEYDSEEA EKCKKIQEEC

>CIDRα2.1 MAL6P1.252

PWCGVKPGGPPWKDNDIDSCGKKEISFSDKDTT DISILSTDRAKKNILQKLENFCRD
TEHINHDIWKCHYENTDNDNCILQNENTGSEKQKIMPFDAFFLWLTQMLDDSI EWR
KKLKT CINNEKPTNCIRGCKKPCECFERWVEQKEE EWISIEKHFDKQRDI SEEERYI
TLEYILNEFFMDKIEKAYGIEKSKELKEKLKSNKGHGIIRDTEHSQDAIKILLEHEL
EDAKKCTETHNDEKC

>CIDRα3.1 DD2var01

PYCGVKKVNNGGSSNEWEEKNNGKCKSGKLYEPKPDKEGTTITILKSGKGHDDIEEK
LNKFCDEKNGDTINSGSGTGSGGGNSGRQELYEEWKCYKGEDVVKVGHDEDEED
YENVKNAGGLCILKNQKKNKEEGGNTSEKEPDEIQKTFNPFY YWVAHMLKDSIHWK
KKLQRC LQNGNRIKCGNNKCNNDCECFKRWITQKKDEWGKIVQHFKTQNIKGRGGS D
NTAELIPFDHDYVLQYNLQEEFLKGDSEDA SEEKS ENSLDAEEAEELKHLREIIESE
DNNQEASVGGGVTEQKNIMDKLLNYEKDEADLCLEIH

Appendix 4. Publications arising from the thesis

Hsieh, F.L., Turner, L., Bolla, J.R., Robinson, C.V., Lavstsen, T., and Higgins, M.K. (2016). The structural basis for CD36 binding by the malaria parasite. **Nature communications** 7, 12837.

Hsieh, F.L., and Higgins, M.K. (2017). The structure of a LAIR1-containing human antibody reveals a novel mechanism of antigen recognition. **Elife** 6, e27311

References

Abdel-Latif, M.S., Cabrera, G., Kohler, C., Kremsner, P.G., Luty, A.J., and Team, C.S. (2004). Antibodies to rifin: a component of naturally acquired responses to *Plasmodium falciparum* variant surface antigens on infected erythrocytes. *The American journal of tropical medicine and hygiene* 71, 179-186.

Abdel-Latif, M.S., Dietz, K., Issifou, S., Kremsner, P.G., and Klinkert, M.Q. (2003). Antibodies to *Plasmodium falciparum* rifin proteins are associated with rapid parasite clearance and asymptomatic infections. *Infection and immunity* 71, 6229-6233.

Abdel-Latif, M.S., Khattab, A., Lindenthal, C., Kremsner, P.G., and Klinkert, M.Q. (2002). Recognition of variant Rifin antigens by human antibodies induced during natural *Plasmodium falciparum* infections. *Infection and immunity* 70, 7013-7021.

Abdi, A.I., Kariuki, S.M., Muthui, M.K., Kivisi, C.A., Fegan, G., Gitau, E., Newton, C.R., and Bull, P.C. (2015). Differential *Plasmodium falciparum* surface antigen expression among children with Malarial Retinopathy. *Sci Rep* 5, 18034.

Adams, J.C., and Lawler, J. (2011). The thrombospondins. *Cold Spring Harb Perspect Biol* 3, a009712.

Adams, Y., Kuhnrae, P., Higgins, M.K., Ghumra, A., and Rowe, J.A. (2014). Rosetting *Plasmodium falciparum*-infected erythrocytes bind to human brain microvascular endothelial cells in vitro, demonstrating a dual adhesion phenotype mediated by distinct *P. falciparum* erythrocyte membrane protein 1 domains. *Infection and immunity* 82, 949-959.

Agbanyo, F.R., Sixma, J.J., de Groot, P.G., Languino, L.R., and Plow, E.F. (1993). Thrombospondin-platelet interactions. Role of divalent cations, wall shear rate, and platelet membrane glycoproteins. *The Journal of clinical investigation* 92, 288-296.

Akaki, M., and Dvorak, J.A. (2005). A chemotactic response facilitates mosquito salivary gland infection by malaria sporozoites. *J Exp Biol* 208, 3211-3218.

al-Khedery, B., Barnwell, J.W., and Galinski, M.R. (1999). Antigenic variation in malaria: a 3' genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Mol Cell* 3, 131-141.

Al-Lazikani, B., Lesk, A.M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273, 927-948.

Aly, A.S., and Matuschewski, K. (2005). A malarial cysteine protease is necessary for *Plasmodium* sporozoite egress from oocysts. *The Journal of experimental medicine* 202, 225-230.

Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., and Menard, R. (2006). Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nature medicine* 12, 220-224.

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica Section D, Biological crystallography* 62, 1243-1250.

Asch, A.S., Barnwell, J., Silverstein, R.L., and Nachman, R.L. (1987). Isolation of the thrombospondin membrane receptor. *The Journal of clinical investigation* 79, 1054-1061.

Asch, A.S., Silbiger, S., Heimer, E., and Nachman, R.L. (1992). Thrombospondin sequence motif (CSVTCG) is responsible for CD36 binding. *Biochem Biophys Res Commun* 182, 1208-1217.

Avril, M., Bernabeu, M., Benjamin, M., Brazier, A.J., and Smith, J.D. (2016). Interaction between Endothelial Protein C Receptor and Intercellular Adhesion Molecule 1 to Mediate Binding of Plasmodium falciparum-Infected Erythrocytes to Endothelial Cells. *MBio* 7.

Bachmann, A., Petter, M., Tilly, A.-K., Biller, L., Uliczka, K.A., Duffy, M.F., Tannich, E., and Bruchhaus, I. (2012). Temporal Expression and Localization Patterns of Variant Surface Antigens in Clinical Plasmodium falciparum Isolates during Erythrocyte Schizogony. *PLOS ONE* 7, e49540.

Baenziger, N.L., Brodie, G.N., and Majerus, P.W. (1971). A thrombin-sensitive protein of human platelet membranes. *Proceedings of the National Academy of Sciences of the United States of America* 68, 240-243.

Baruch, D.I., Gormely, J.A., Ma, C., Howard, R.J., and Pasloske, B.L. (1996). Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proceedings of the National Academy of Sciences of the United States of America* 93, 3497-3502.

Baruch, D.I., Ma, X.C., Pasloske, B., Howard, R.J., and Miller, L.H. (1999). CD36 peptides that block cytoadherence define the CD36 binding region for Plasmodium falciparum-infected erythrocytes. *Blood* 94, 2121-2127.

Baruch, D.I., Ma, X.C., Singh, H.B., Bi, X., Pasloske, B.L., and Howard, R.J. (1997). Identification of a region of PfEMP1 that mediates adherence of Plasmodium falciparum infected erythrocytes to CD36: conserved function with variant sequence. *Blood* 90, 3766-3775.

Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X., Ma, X.C., Feldman, M., Taraschi, T.F., and Howard, R.J. (1995). Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* 82, 77-87.

Batchelor, J.D., Zahm, J.A., and Tolia, N.H. (2011). Dimerization of Plasmodium vivax DBP is induced upon receptor binding and drives recognition of DARC. *Nat Struct Mol Biol* 18, 908-914.

Bengtsson, A., Joergensen, L., Rask, T.S., Olsen, R.W., Andersen, M.A., Turner, L., Theander, T.G., Hviid, L., Higgins, M.K., Craig, A., *et al.* (2013). A novel domain cassette identifies Plasmodium falciparum PfEMP1 proteins binding ICAM-1 and is a target of cross-reactive, adhesion-inhibitory antibodies. *J Immunol* 190, 240-249.

Biggs, B.A., Gooze, L., Wycherley, K., Wollish, W., Southwell, B., Leech, J.H., and Brown, G.V. (1991). Antigenic variation in Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America* 88, 9171-9174.

Billker, O., Lindo, V., Panico, M., Etienne, A.E., Paxton, T., Dell, A., Rogers, M., Sinden, R.E., and Morris, H.R. (1998). Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* 392, 289-292.

Blackman, M.J. (2008). Malarial proteases and host cell egress: an 'emerging' cascade. *Cellular microbiology* 10, 1925-1934.

Bonnefoy, A., Hantgan, R., Legrand, C., and Frojmovic, M.M. (2001). A model of platelet aggregation involving multiple interactions of thrombospondin-1, fibrinogen, and GPIIb/IIIa receptor. *J Biol Chem* 276, 5605-5612.

Bousema, T., Okell, L., Felger, I., and Drakeley, C. (2014). Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat Rev Microbiol* 12, 833-840.

Brondijk, T.H., de Ruiter, T., Ballering, J., Wienk, H., Lebbink, R.J., van Ingen, H., Boelens, R., Farndale, R.W., Meyaard, L., and Huizinga, E.G. (2010). Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood* 115, 1364-1373.

Brown, A., Turner, L., Christoffersen, S., Andrews, K.A., Szeszak, T., Zhao, Y., Larsen, S., Craig, A.G., and Higgins, M.K. (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.

Brown, K.N., and Brown, I.N. (1965). Immunity to malaria: antigenic variation in chronic infections of *Plasmodium knowlesi*. *Nature* 208, 1286-1288.

Brunger, A.T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta crystallographica Section D, Biological crystallography* 49, 24-36.

Bultrini, E., Brick, K., Mukherjee, S., Zhang, Y., Silvestrini, F., Alano, P., and Pizzi, E. (2009). Revisiting the *Plasmodium falciparum* RIFIN family: from comparative genomics to 3D-model prediction. *BMC Genomics* 10, 445.

Cabrera, A., Neculai, D., and Kain, K.C. (2014). CD36 and malaria: friends or foes? A decade of data provides some answers. *Trends in parasitology* 30, 436-444.

Carlson, J., Helmby, H., Hill, A.V., Brewster, D., Greenwood, B.M., and Wahlgren, M. (1990). Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet* 336, 1457-1460.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Perte, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., *et al.* (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512-519.

Cerutti, C., Jr., Boulos, M., Coutinho, A.F., Hatab Mdo, C., Falqueto, A., Rezende, H.R., Duarte, A.M., Collins, W., and Malafronte, R.S. (2007). Epidemiologic aspects of the malaria transmission cycle in an area of very low incidence in Brazil. *Malar J* 6, 33.

Chan, J.A., Fowkes, F.J., and Beeson, J.G. (2014). Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell Mol Life Sci* 71, 3633-3657.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., *et al.* (2007). Glycoprotein structural genomics: solving the glycosylation problem. *Structure* 15, 267-273.

Chen, I., Clarke, S.E., Gosling, R., Hamainza, B., Killeen, G., Magill, A., O'Meara, W., Price, R.N., and Riley, E.M. (2016). "Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated. *PLoS Med* 13, e1001942.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* 66, 12-21.

Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M., and Saul, A. (1998). *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which

potentially encode variant antigens. *Molecular and biochemical parasitology* 97, 161-176.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., *et al.* (1989). Conformations of immunoglobulin hypervariable regions. *Nature* 342, 877-883.

Clausen, T.M., Christoffersen, S., Dahlback, M., Langkilde, A.E., Jensen, K.E., Resende, M., Agerbaek, M.O., Andersen, D., Berisha, B., Ditlev, S.B., *et al.* (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.

Cowman, A.F., and Crabb, B.S. (2006). Invasion of red blood cells by malaria parasites. *Cell* 124, 755-766.

Cox, F.E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasit Vectors* 3, 5.

Cox, H.W. (1959). A study of relapse *Plasmodium berghei* infections isolated from white mice. *J Immunol* 82, 209-214.

Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., Ratnam, S., Rahman, H.A., Conway, D.J., and Singh, B. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis* 46, 165-171.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* 14, 1188-1190.

Cucunuba, Z.M., Guerra, A.P., Rahirant, S.J., Rivera, J.A., Cortes, L.J., and Nicholls, R.S. (2008). Asymptomatic *Plasmodium* spp. infection in Tierralta, Colombia. *Mem Inst Oswaldo Cruz* 103, 668-673.

Dalton, H.J., Armaiz-Pena, G.N., Gonzalez-Villasana, V., Lopez-Berestein, G., Bar-Eli, M., and Sood, A.K. (2014). Monocyte subpopulations in angiogenesis. *Cancer Res* 74, 1287-1293.

Dang, M., Wang, X., Wang, Q., Wang, Y., Lin, J., Sun, Y., Li, X., Zhang, L., Lou, Z., Wang, J., *et al.* (2014). Molecular mechanism of SCARB2-mediated attachment and uncoating of EV71. *Protein & cell* 5, 692-703.

Daubersies, P., Sallenave-Sales, S., Magne, S., Trape, J.F., Contamin, H., Fandeur, T., Rogier, C., Mercereau-Puijalon, O., and Druilhe, P. (1996). Rapid turnover of *Plasmodium falciparum* populations in asymptomatic individuals living in a high transmission area. *The American journal of tropical medicine and hygiene* 54, 18-26.

David, P.H., Hommel, M., Miller, L.H., Udeinya, I.J., and Oligino, L.D. (1983). Parasite sequestration in *Plasmodium falciparum* malaria: spleen and antibody modulation of cytoadherence of infected erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America* 80, 5075-5079.

Dawson, D.W., Pearce, S.F., Zhong, R., Silverstein, R.L., Frazier, W.A., and Bouck, N.P. (1997). CD36 mediates the *In vitro* inhibitory effects of thrombospondin-1 on endothelial cells. *J Cell Biol* 138, 707-717.

Day, K.P., and Marsh, K. (1991). Naturally acquired immunity to *Plasmodium falciparum*. *Immunol Today* 12, A68-71.

del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., *et al.* (2001). A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410, 839-842.

Di Girolamo, F., Raggi, C., Birago, C., Pizzi, E., Lalle, M., Picci, L., Pace, T., Bachi, A., de Jong, J., Janse, C.J., *et al.* (2008). 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.

Dong, A., Xu, X., Edwards, A.M., Midwest Center for Structural, G., Structural Genomics, C., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., *et al.* (2007). In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 4, 1019-1021.

Dunbar, J., and Deane, C.M. (2016). ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* 32, 298-300.

Dustin, M.L., Rothlein, R., Bhan, A.K., Dinarello, C.A., and Springer, T.A. (1986). Induction by IL 1 and interferon-gamma: tissue distribution, biochemistry, and function of a natural adherence molecule (ICAM-1). *J Immunol* 137, 245-254.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta crystallographica Section D, Biological crystallography* 66, 486-501.

Endemann, G., Stanton, L.W., Madden, K.S., Bryant, C.M., White, R.T., and Protter, A.A. (1993). CD36 is a receptor for oxidized low density lipoprotein. *J Biol Chem* 268, 11811-11816.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography* 69, 1204-1214.

Febbraio, M., Abumrad, N.A., Hajjar, D.P., Sharma, K., Cheng, W., Pearce, S.F., and Silverstein, R.L. (1999). A null mutation in murine CD36 reveals an important role in fatty acid and lipoprotein metabolism. *J Biol Chem* 274, 19055-19062.

Fernandez, V., Hommel, M., Chen, Q., Hagblom, P., and Wahlgren, M. (1999). 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. *The Journal of experimental medicine* 190, 1393-1404.

Filipe, J.A., Riley, E.M., Drakeley, C.J., Sutherland, C.J., and Ghani, A.C. (2007). Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Comput Biol* 3, e255.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., *et al.* (2002). A proteomic view of the Plasmodium falciparum life cycle. *Nature* 419, 520-526.

Francis, N.K., Foster, J.D., and Salib, E. (2017). Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery: common mistakes and methodological issues. *Tech Coloproctol* 21, 403.

Francis, S.E., Sullivan, D.J., Jr., and Goldberg, D.E. (1997). Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annu Rev Microbiol* 51, 97-123.

Fried, M., and Duffy, P.E. (1996). Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science* 272, 1502-1504.

Frieda, S., Pearce, A., Wu, J., and Silverstein, R.L. (1995). Recombinant GST/CD36 fusion proteins define a thrombospondin binding domain. Evidence for a single calcium-dependent binding site on CD36. *J Biol Chem* 270, 2981-2986.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., and Schatz, D.G. (2000). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* 18, 495-527.

Galinski, M.R., and Corredor, V. (2004). Variant antigen expression in malaria infections: posttranscriptional gene silencing, virulence and severe pathology. *Molecular and biochemical parasitology* 134, 17-25.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., *et al.* (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498-511.

Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., *et al.* (1998). Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* 282, 1126-1132.

Ge, Y., and Elghetany, M.T. (2005). CD36: a multiligand molecule. *Lab Hematol* 11, 31-37.

Genton, B., D'Acremont, V., Rare, L., Baea, K., Reeder, J.C., Alpers, M.P., and Muller, I. (2008). *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. *PLoS Med* 5, e127.

Goel, S., Palmkvist, M., Moll, K., Joannin, N., Lara, P., Akhouri, R.R., Moradi, N., Ojemalm, K., Westman, M., Angeletti, D., *et al.* (2015). RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nature medicine* 21, 314-317.

Goldsby, R.A., Kindt, T.J., Osborne, B.A., and Kuby, J. (2000). *Kuby immunology*, 4th edn (New York: W.H. Freeman).

Gowda, N.M., Wu, X., Kumar, S., Febbraio, M., and Gowda, D.C. (2013). CD36 contributes to malaria parasite-induced pro-inflammatory cytokine production and NK and T cell activation by dendritic cells. *PLoS One* 8, e77604.

Greenberg, A.S., Avila, D., Hughes, M., Hughes, A., McKinney, E.C., and Flajnik, M.F. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* 374, 168-173.

Griffin, L.M., Snowden, J.R., Lawson, A.D., Wernery, U., Kinne, J., and Baker, T.S. (2014). Analysis of heavy and light chain sequences of conventional camelid antibodies from *Camelus dromedarius* and *Camelus bactrianus* species. *J Immunol Methods* 405, 35-46.

Gruarin, P., Sitia, R., and Alessio, M. (1997). Formation of one or more intrachain disulphide bonds is required for the intracellular processing and transport of CD36. *Biochem J* 328 (Pt 2), 635-642.

Gu, H., Kitamura, D., and Rajewsky, K. (1991). B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound D mu protein and selection of DH element reading frames. *Cell* 65, 47-54.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., *et al.* (2005). A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307, 82-86.

Handunnetti, S.M., Mendis, K.N., and David, P.H. (1987). Antigenic variation of cloned *Plasmodium fragile* in its natural host *Macaca sinica*. Sequential appearance of successive variant antigenic types. *The Journal of experimental medicine* 165, 1269-1283.

Harris, I., Sharrock, W.W., Bain, L.M., Gray, K.A., Bobogare, A., Boaz, L., Lilley, K., Krause, D., Vallely, A., Johnson, M.L., *et al.* (2010). A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. *Malar J* 9, 254.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., and Snow, R.W. (2004). The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 4, 327-336.

Hayward, R.E., Tiwari, B., Piper, K.P., Baruch, D.I., and Day, K.P. (1999). Virulence and transmission success of the malarial parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4563-4568.

Heddi, A., Pettersson, F., Kai, O., Shafi, J., Obiero, J., Chen, Q., Barragan, A., Wahlgren, M., and Marsh, K. (2001). Fresh isolates from children with severe *Plasmodium falciparum* malaria bind to multiple receptors. *Infection and immunity* 69, 5849-5856.

Heppner, D.G., Jr., Kester, K.E., Ockenhouse, C.F., Tornieporth, N., Ofori, O., Lyon, J.A., Stewart, V.A., Dubois, P., Lanar, D.E., Krzych, U., *et al.* (2005). Towards an RTS,S-based, multi-stage, multi-antigen vaccine against *falciparum* malaria: progress at the Walter Reed Army Institute of Research. *Vaccine* 23, 2243-2250.

Higgins, M.K. (2008). The structure of a chondroitin sulfate-binding domain important in placental malaria. *J Biol Chem* 283, 21842-21846.

Higgins, M.K., and Carrington, M. (2014). Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein Sci* 23, 354-365.

Hodder, A.N., Czabotar, P.E., Uboldi, A.D., Clarke, O.B., Lin, C.S., Healer, J., Smith, B.J., and Cowman, A.F. (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.

Hommel, M., David, P.H., and Oligino, L.D. (1983). Surface alterations of erythrocytes in *Plasmodium falciparum* malaria. Antigenic variation, antigenic diversity, and the role of the spleen. *The Journal of experimental medicine* 157, 1137-1148.

Hoosdally, S.J., Andress, E.J., Wooding, C., Martin, C.A., and Linton, K.J. (2009). The Human Scavenger Receptor CD36: glycosylation status and its role in trafficking and function. *J Biol Chem* 284, 16277-16288.

Horrocks, P., and Muhia, D. (2005). Pexel/VTS: a protein-export motif in erythrocytes infected with malaria parasites. *Trends in parasitology* 21, 396-399.

Howell, D.P., Levin, E.A., Springer, A.L., Kraemer, S.M., Phippard, D.J., Schief, W.R., and Smith, J.D. (2008). Mapping a common interaction site used by *Plasmodium falciparum* Duffy binding-like domains to bind diverse host receptors. *Mol Microbiol* 67, 78-87.

Hsieh, F.L., Turner, L., Bolla, J.R., Robinson, C.V., Lavstsen, T., and Higgins, M.K. (2016). The structural basis for CD36 binding by the malaria parasite. *Nature communications* 7, 12837.

Hviid, L., and Jensen, A.T. (2015). PfEMP1 - A Parasite Protein Family of Key Importance in *Plasmodium falciparum* Malaria Immunity and Pathogenesis. *Adv Parasitol* 88, 51-84.

Ibrahimi, A., Sfeir, Z., Magharaie, H., Amri, E.Z., Grimaldi, P., and Abumrad, N.A. (1996). Expression of the CD36 homolog (FAT) in fibroblast cells: effects on fatty acid transport. *Proceedings of the National Academy of Sciences of the United States of America* 93, 2646-2651.

Janes, J.H., Wang, C.P., Levin-Edens, E., Vigan-Womas, I., Guillotte, M., Melcher, M., Mercereau-Puijalon, O., and Smith, J.D. (2011). Investigating the host binding signature on the *Plasmodium falciparum* PfEMP1 protein family. *PLoS pathogens* 7, e1002032.

Jansen, C.A., Cruijssen, C.W., de Ruiter, T., Nanlohy, N., Willems, N., Janssens-Korpela, P.L., and Meyaard, L. (2007). Regulated expression of the inhibitory receptor LAIR-1 on human peripheral T cells during T cell activation and differentiation. *Eur J Immunol* 37, 914-924.

Janssen, C.S., Barrett, M.P., Turner, C.M., and Phillips, R.S. (2002). A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proc Biol Sci* 269, 431-436.

Janssen, C.S., Phillips, R.S., Turner, C.M., and Barrett, M.P. (2004). *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res* 32, 5712-5720.

Jiang, Y., Wang, M., Huang, K., Zhang, Z., Shao, N., Zhang, Y., Wang, W., and Wang, S. (2012). Oxidized low-density lipoprotein induces secretion of interleukin-1 β by macrophages via reactive oxygen species-dependent NLRP3 inflammasome activation. *Biochem Biophys Res Commun* 425, 121-126.

Joannin, N., Abhiman, S., Sonnhhammer, E.L., and Wahlgren, M. (2008). Subgrouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics* 9, 19.

Kabsch, W. (2010). Xds. *Acta crystallographica Section D, Biological crystallography* 66, 125-132.

Kato, L., Stanlie, A., Begum, N.A., Kobayashi, M., Aida, M., and Honjo, T. (2012). An evolutionary view of the mechanism for immune and genome diversity. *J Immunol* 188, 3559-3566.

Kerlin, D.H., and Gatton, M.L. (2013). Preferential invasion by *Plasmodium* merozoites and the self-regulation of parasite burden. *PLoS One* 8, e57434.

Khunrae, P., Philip, J.M., Bull, D.R., and Higgins, M.K. (2009). Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. *J Mol Biol* 393, 202-213.

Klein, M.M., Gittis, A.G., Su, H.P., Makobongo, M.O., Moore, J.M., Singh, S., Miller, L.H., and Garboczi, D.N. (2008). The cysteine-rich interdomain region from the highly variable *plasmodium falciparum* erythrocyte membrane protein-1 exhibits a conserved structure. *PLoS pathogens* 4, e1000147.

Kraemer, S.M., Kyes, S.A., Aggarwal, G., Springer, A.L., Nelson, S.O., Christodoulou, Z., Smith, L.M., Wang, W., Levin, E., Newbold, C.I., *et al.* (2007). Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics* 8, 45.

Kun, J.F., Schmidt-Ott, R.J., Lehman, L.G., Lell, B., Luckner, D., Greve, B., Matousek, P., and Kremsner, P.G. (1998). Merozoite surface antigen 1 and 2 genotypes and rosetting of *Plasmodium falciparum* in severe and mild malaria in Lambarene, Gabon. *Trans R Soc Trop Med Hyg* 92, 110-114.

Kyes, S.A., Rowe, J.A., Kriek, N., and Newbold, C.I. (1999). Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with

Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America 96, 9333-9338.

Laishram, D.D., Sutton, P.L., Nanda, N., Sharma, V.L., Sobti, R.C., Carlton, J.M., and Joshi, H. (2012). The complexities of malaria disease manifestations with a focus on asymptomatic malaria. *Malar J* 11, 29.

Laszik, Z., Mitro, A., Taylor, F.B., Jr., Ferrell, G., and Esmon, C.T. (1997). Human protein C receptor is present primarily on endothelium of large blood vessels: implications for the control of the protein C pathway. *Circulation* 96, 3633-3640.

Lau, C.K., Turner, L., Jespersen, J.S., Lowe, E.D., Petersen, B., Wang, C.W., Petersen, J.E., Lusingu, J., Theander, T.G., Lavstsen, T., *et al.* (2015). Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. *Cell host & microbe* 17, 118-129.

Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E., and Theander, T.G. (2003). Subgrouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J* 2, 27.

Lebbink, R.J., de Ruiter, T., Adelmeijer, J., Brenkman, A.B., van Helvoort, J.M., Koch, M., Farndale, R.W., Lisman, T., Sonnenberg, A., Lenting, P.J., *et al.* (2006). Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *The Journal of experimental medicine* 203, 1419-1425.

Lebbink, R.J., de Ruiter, T., Kaptijn, G.J., Bihan, D.G., Jansen, C.A., Lenting, P.J., and Meyaard, L. (2007). Mouse leukocyte-associated Ig-like receptor-1 (mLAIR-1) functions as an inhibitory collagen-binding receptor on immune cells. *Int Immunol* 19, 1011-1019.

Leech, J.H., Barnwell, J.W., Miller, L.H., and Howard, R.J. (1984). Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *The Journal of experimental medicine* 159, 1567-1575.

Lin, D.H., Malpede, B.M., Batchelor, J.D., and Tolia, N.H. (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.

Lindblade, K.A., Steinhart, L., Samuels, A., Kachur, S.P., and Slutsker, L. (2013). The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev Anti Infect Ther* 11, 623-639.

Liu, Y.J. (2001). Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity. *Cell* 106, 259-262.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* 306, 1930-1933.

Martin, A.C., and Thornton, J.M. (1996). Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. *J Mol Biol* 263, 800-815.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

McGilvray, I.D., Serghides, L., Kapus, A., Rotstein, O.D., and Kain, K.C. (2000). Nonopsonic monocyte/macrophage phagocytosis of *Plasmodium falciparum*-parasitized erythrocytes: a role for CD36 in malarial clearance. *Blood* 96, 3231-3240.

McLean, S.A., Pearson, C.D., and Phillips, R.S. (1982). *Plasmodium chabaudi*: antigenic variation during recrudescence parasitaemias in mice. *Experimental parasitology* 54, 296-302.

McLellan, J.S., Pancera, M., Carrico, C., Gorman, J., Julien, J.P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K., *et al.* (2011). Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 480, 336-343.

Menard, R., Tavares, J., Cockburn, I., Markus, M., Zavala, F., and Amino, R. (2013). Looking under the skin: the first steps in malarial infection and immunity. *Nat Rev Microbiol* 11, 701-712.

Meyaard, L. (2008). The inhibitory collagen receptor LAIR-1 (CD305). *J Leukoc Biol* 83, 799-803.

Meyaard, L., Adema, G.J., Chang, C., Woollatt, E., Sutherland, G.R., Lanier, L.L., and Phillips, J.H. (1997). LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes. *Immunity* 7, 283-290.

Meyaard, L., Hurenkamp, J., Clevers, H., Lanier, L.L., and Phillips, J.H. (1999). Leukocyte-associated Ig-like receptor-1 functions as an inhibitory receptor on cytotoxic T cells. *J Immunol* 162, 5800-5804.

Miotto, O., Almagro-Garcia, J., Manske, M., Macinnis, B., Campino, S., Rockett, K.A., Amaratunga, C., Lim, P., Suon, S., Sreng, S., *et al.* (2013). Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet* 45, 648-655.

Moll, K., Palmkvist, M., Ch'ng, J., Kiwuwa, M.S., and Wahlgren, M. (2015). Evasion of Immunity to *Plasmodium falciparum*: Rosettes of Blood Group A Impair Recognition of PfEMP1. *PLoS One* 10, e0145120.

Mons, B. (1990). Preferential invasion of malarial merozoites into young red blood cells. *Blood Cells* 16, 299-312.

Mueller, I., Zimmerman, P.A., and Reeder, J.C. (2007). *Plasmodium malariae* and *Plasmodium ovale*--the "bashful" malaria parasites. *Trends in parasitology* 23, 278-283.

Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J.A., and Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng* 7, 1129-1135.

Mwakalinga, S.B., Wang, C.W., Bengtsson, D.C., Turner, L., Dinko, B., Lusingu, J.P., Arnot, D.E., Sutherland, C.J., Theander, T.G., and Lavstsen, T. (2012). Expression of a type B RIFIN in *Plasmodium falciparum* merozoites and gametes. *Malar J* 11, 429.

Nagy, L., Tontono, P., Alvarez, J.G., Chen, H., and Evans, R.M. (1998). Oxidized LDL regulates macrophage gene expression through ligand activation of PPARgamma. *Cell* 93, 229-240.

Neculai, D., Schwake, M., Ravichandran, M., Zunke, F., Collins, R.F., Peters, J., Neculai, M., Plumb, J., Loppnau, P., Pizarro, J.C., *et al.* (2013). Structure of LIMP-2 provides functional insights with implications for SR-BI and CD36. *Nature* 504, 172-176.

Nettleship, J.E., Ren, J., Rahman, N., Berrow, N.S., Hatherley, D., Barclay, A.N., and Owens, R.J. (2008). A pipeline for the production of antibody fragments for structural studies using transient expression in HEK 293T cells. *Protein Expr Purif* 62, 83-89.

Newbold, C., Warn, P., Black, G., Berendt, A., Craig, A., Snow, B., Msobo, M., Peshu, N., and Marsh, K. (1997). Receptor-specific adhesion and clinical disease in

Plasmodium falciparum. The American journal of tropical medicine and hygiene 57, 389-398.

Ng, H.K. (2012). Structural characterisation of the interaction between *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) and their human ligands in severe malaria. In Department of Biochemistry (University of Oxford).

Nuttall, S.D., Krishnan, U.V., Hattarki, M., De Gori, R., Irving, R.A., and Hudson, P.J. (2001). Isolation of the new antigen receptor from wobbegong sharks, and use as a scaffold for the display of protein loop libraries. *Mol Immunol* 38, 313-326.

Nzobo, B.J., Ngasala, B.E., and Kihamia, C.M. (2015). Prevalence of asymptomatic malaria infection and use of different malaria control measures among primary school children in Morogoro Municipality, Tanzania. *Malar J* 14, 491.

Ouyang, W., Ma, D., Lin, D., Sun, Y., Liu, X., Li, Q., Jia, W., Cao, Y., Zhu, Y., and Jin, B. (2003). 9.1C3 is identical to LAIR-1, which is expressed on hematopoietic progenitors. *Biochem Biophys Res Commun* 310, 1236-1240.

Pain, A., Bohme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Aslett, M.A., *et al.* (2008). The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 455, 799-803.

Pasvol, G. (2005). The treatment of complicated and severe malaria. *Br Med Bull* 75-76, 29-47.

Pepino, M.Y., Kuda, O., Samovski, D., and Abumrad, N.A. (2014). Structure-function of CD36 and importance of fatty acid signal transduction in fat metabolism. *Annual review of nutrition* 34, 281-303.

Petter, M., Haeggstrom, M., Khattab, A., Fernandez, V., Klinkert, M.Q., and Wahlgren, M. (2007). Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Molecular and biochemical parasitology* 156, 51-61.

Pizzi, E., and Frontali, C. (2001). Fine structure of *Plasmodium falciparum* subtelomeric sequences. *Molecular and biochemical parasitology* 118, 253-258.

Poggi, A., Tomasello, E., Ferrero, E., Zocchi, M.R., and Moretta, L. (1998). p40/LAIR-1 regulates the differentiation of peripheral blood precursors to dendritic cells induced by granulocyte-monocyte colony-stimulating factor. *Eur J Immunol* 28, 2086-2091.

Pradel, G., Garapaty, S., and Frevert, U. (2002). Proteoglycans mediate malaria sporozoite targeting to the liver. *Mol Microbiol* 45, 637-651.

Price, R.N., Tjitra, E., Guerra, C.A., Yeung, S., White, N.J., and Anstey, N.M. (2007). Vivax malaria: neglected and not benign. *The American journal of tropical medicine and hygiene* 77, 79-87.

Rahaman, S.O., Lennon, D.J., Febbraio, M., Podrez, E.A., Hazen, S.L., and Silverstein, R.L. (2006). A CD36-dependent signaling cascade is necessary for macrophage foam cell formation. *Cell Metab* 4, 211-221.

Rask, T.S., Hansen, D.A., Theander, T.G., Gorm Pedersen, A., and Lavstsen, T. (2010). *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput Biol* 6.

Rege, T.A., Stewart, J., Jr., Dranka, B., Benveniste, E.N., Silverstein, R.L., and Gladson, C.L. (2009). Thrombospondin-1-induced apoptosis of brain microvascular endothelial cells can be mediated by TNF-R1. *J Cell Physiol* 218, 94-103.

Roberts, W., Magwenzi, S., Aburima, A., and Naseem, K.M. (2010). Thrombospondin-1 induces platelet activation through CD36-dependent inhibition of the cAMP/protein kinase A signaling cascade. *Blood* 116, 4297-4306.

Robinson, B.A., Welch, T.L., and Smith, J.D. (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.

Rodriguez, M.H., and Hernandez-Hernandez Fde, L. (2004). Insect-malaria parasites interactions: the salivary gland. *Insect Biochem Mol Biol* 34, 615-624.

Rogerson, S.J., Hviid, L., Duffy, P.E., Leke, R.F., and Taylor, D.W. (2007). Malaria in pregnancy: pathogenesis and immunity. *Lancet Infect Dis* 7, 105-117.

Rogerson, S.J., Tembenu, R., Dobano, C., Plitt, S., Taylor, T.E., and Molyneux, M.E. (1999). Cytoadherence characteristics of *Plasmodium falciparum*-infected erythrocytes from Malawian children with severe and uncomplicated malaria. *The American journal of tropical medicine and hygiene* 61, 467-472.

Roux, K.H., Greenberg, A.S., Greene, L., Strelets, L., Avila, D., McKinney, E.C., and Flajnik, M.F. (1998). Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. *Proceedings of the National Academy of Sciences of the United States of America* 95, 11804-11809.

Rowe, A., Obeiro, J., Newbold, C.I., and Marsh, K. (1995). *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. *Infection and immunity* 63, 2323-2326.

Rts, S.C.T.P. (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet* 386, 31-45.

Saini, S.S., Allore, B., Jacobs, R.M., and Kaushik, A. (1999). Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur J Immunol* 29, 2420-2426.

Sam-Yellowe, T.Y., Florens, L., Johnson, J.R., Wang, T., Drazba, J.A., Le Roch, K.G., Zhou, Y., Batalov, S., Carucci, D.J., Winzeler, E.A., *et al.* (2004). A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. *Genome Res* 14, 1052-1059.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* 45, 167-202.

Schwartz, R.S. (2003). Shattuck lecture: Diversity of the immune repertoire and immunoregulation. *N Engl J Med* 348, 1017-1026.

Schwarzer, E., Kuhn, H., Valente, E., and Arese, P. (2003). Malaria-parasitized erythrocytes and hemozoin nonenzymatically generate large amounts of hydroxy fatty acids that inhibit monocyte functions. *Blood* 101, 722-728.

Seixas, E., Cross, C., Quin, S., and Langhorne, J. (2001). Direct activation of dendritic cells by the malaria parasite, *Plasmodium chabaudi chabaudi*. *Eur J Immunol* 31, 2970-2978.

Shirai, H., Kidera, A., and Nakamura, H. (1999). H3-rules: identification of CDR-H3 structures in antibodies. *FEBS Lett* 455, 188-197.

Silverstein, R.L., and Febbraio, M. (2009). CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Sci Signal* 2, re3.

Singh, S.K., Hora, R., Belrhali, H., Chitnis, C.E., and Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* 439, 741-744.

Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonnrhein, C., and Bricogne, G. (2012). Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta crystallographica Section D, Biological crystallography* 68, 368-380.

Smider, V., and Chu, G. (1997). The end-joining reaction in V(D)J recombination. *Semin Immunol* 9, 189-197.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., and Miller, L.H. (1995). Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82, 101-110.

Smith, J.D., Craig, A.G., Kriek, N., Hudson-Taylor, D., Kyes, S., Fagan, T., Pinches, R., Baruch, D.I., Newbold, C.I., and Miller, L.H. (2000a). Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proceedings of the National Academy of Sciences of the United States of America* 97, 1766-1771.

Smith, J.D., Gamain, B., Baruch, D.I., and Kyes, S. (2001). Decoding the language of var genes and *Plasmodium falciparum* sequestration. *Trends in parasitology* 17, 538-545.

Smith, J.D., Kyes, S., Craig, A.G., Fagan, T., Hudson-Taylor, D., Miller, L.H., Baruch, D.I., and Newbold, C.I. (1998). Analysis of adhesive domains from the A4VAR *Plasmodium falciparum* erythrocyte membrane protein-1 identifies a CD36 binding domain. *Molecular and biochemical parasitology* 97, 133-148.

Smith, J.D., Rowe, J.A., Higgins, M.K., and Lavstsen, T. (2013). Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cellular microbiology* 15, 1976-1983.

Smith, J.D., Subramanian, G., Gamain, B., Baruch, D.I., and Miller, L.H. (2000b). Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. *Molecular and biochemical parasitology* 110, 293-310.

Smith, T.G., Serghides, L., Patel, S.N., Febbraio, M., Silverstein, R.L., and Kain, K.C. (2003). CD36-mediated nonopsonic phagocytosis of erythrocytes infected with stage I and IIA gametocytes of *Plasmodium falciparum*. *Infection and immunity* 71, 393-400.

Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., and Hay, S.I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 434, 214-217.

Srivastava, A., Gangnard, S., Round, A., Dechavanne, S., Juillerat, A., Raynal, B., Faure, G., Baron, B., Ramboarina, S., Singh, S.K., *et al.* (2010). Full-length extracellular region of the var2CSA variant of PfEMP1 is required for specific, high-affinity binding to CSA. *Proceedings of the National Academy of Sciences of the United States of America* 107, 4884-4889.

Stanfield, R.L., and Wilson, I.A. (1994). Antigen-induced conformational changes in antibodies: a problem for structural prediction and design. *Trends Biotechnol* 12, 275-279.

Stanfield, R.L., Wilson, I.A., and Smider, V.V. (2016). Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Sci Immunol* 1.

Stemmer, W.P., Cramer, A., Ha, K.D., Brennan, T.M., and Heyneker, H.L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164, 49-53.

Stevenson, L., Huda, P., Jeppesen, A., Laursen, E., Rowe, J.A., Craig, A., Streicher, W., Barfod, L., and Hviid, L. (2015). Investigating the function of Fc-specific binding of IgM to *Plasmodium falciparum* erythrocyte membrane protein 1 mediating erythrocyte rosetting. *Cellular microbiology* 17, 819-831.

Storm, J., and Craig, A.G. (2014). Pathogenesis of cerebral malaria--inflammation and cytoadherence. *Front Cell Infect Microbiol* 4, 100.

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.M., Menard, R., and Heussler, V.T. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science* 313, 1287-1290.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A., and Wellems, T.E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82, 89-100.

Talman, A.M., Domarle, O., McKenzie, F.E., Arie, F., and Robert, V. (2004). Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malar J* 3, 24.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725-2729.

Tan, J., Pieper, K., Piccoli, L., Abdi, A., Foglierini, M., Geiger, R., Tully, C.M., Jarrossay, D., Ndungu, F.M., Wambua, J., *et al.* (2016). A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* 529, 105-109.

Tjitra, E., Anstey, N.M., Sugiarto, P., Warikar, N., Kenangalem, E., Karyana, M., Lampah, D.A., and Price, R.N. (2008). Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Med* 5, e128.

TMIH (2014). Severe malaria. *Trop Med Int Health* 19 Suppl 1, 7-131.

Tolia, N.H., Enemark, E.J., Sim, B.K., and Joshua-Tor, L. (2005). Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* 122, 183-193.

Tontonoz, P., Hu, E., and Spiegelman, B.M. (1995). Regulation of adipocyte gene expression and differentiation by peroxisome proliferator activated receptor gamma. *Curr Opin Genet Dev* 5, 571-576.

Tontonoz, P., Nagy, L., Alvarez, J.G., Thomazy, V.A., and Evans, R.M. (1998). PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. *Cell* 93, 241-252.

Trape, J.F., Rogier, C., Konate, L., Diagne, N., Bouganali, H., Canque, B., Legros, F., Badji, A., Ndiaye, G., Ndiaye, P., *et al.* (1994). The Dielmo project: a longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of Senegal. *The American journal of tropical medicine and hygiene* 51, 123-137.

Turner, L., Lavstsen, T., Berger, S.S., Wang, C.W., Petersen, J.E., Avril, M., Brazier, A.J., Freeth, J., Jespersen, J.S., Nielsen, M.A., *et al.* (2013). Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* **498**, 502-505.

Udomsangpetch, R., Reinhardt, P.H., Schollaardt, T., Elliott, J.F., Kubes, P., and Ho, M. (1997). Promiscuity of clinical *Plasmodium falciparum* isolates for multiple adhesion molecules under flow conditions. *J Immunol* **158**, 4358-4364.

Urban, B.C., Ferguson, D.J., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M., and Roberts, D.J. (1999). *Plasmodium falciparum*-infected erythrocytes modulate the maturation of dendritic cells. *Nature* **400**, 73-77.

Urban, B.C., Willcox, N., and Roberts, D.J. (2001). A role for CD36 in the regulation of dendritic cell function. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8750-8755.

Vigan-Womas, I., Guillotte, M., Juillerat, A., Hessel, A., Raynal, B., England, P., Cohen, J.H., Bertrand, O., Peyrard, T., Bentley, G.A., *et al.* (2012). Structural basis for the ABO blood-group dependence of *Plasmodium falciparum* rosetting. *PLoS pathogens* **8**, e1002781.

Vlachou, D., Schlegelmilch, T., Runn, E., Mendes, A., and Kafatos, F.C. (2006). The developmental migration of *Plasmodium* in mosquitoes. *Curr Opin Genet Dev* **16**, 384-391.

Vlachou, D., Zimmermann, T., Cantera, R., Janse, C.J., Waters, A.P., and Kafatos, F.C. (2004). Real-time, in vivo analysis of malaria ookinete locomotion and mosquito midgut invasion. *Cellular microbiology* **6**, 671-685.

Volpert, O.V., Zaichuk, T., Zhou, W., Reiher, F., Ferguson, T.A., Stuart, P.M., Amin, M., and Bouck, N.P. (2002). Inducer-stimulated Fas targets activated endothelium for destruction by anti-angiogenic thrombospondin-1 and pigment epithelium-derived factor. *Nature medicine* **8**, 349-357.

Wahlgren, M., Goel, S., and Akhouri, R.R. (2017). Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat Rev Microbiol*.

Wang, C.W., Mwakalinga, S.B., Sutherland, C.J., Schwank, S., Sharp, S., Hermsen, C.C., Sauerwein, R.W., Theander, T.G., and Lavstsen, T. (2010). Identification of a major rif transcript common to gametocytes and sporozoites of *Plasmodium falciparum*. *Malar J* **9**, 147.

Wang, F., Ekiert, D.C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., Torkamani, A., Raudsepp, T., Mwangi, W., Criscitiello, M.F., *et al.* (2013). Reshaping antibody diversity. *Cell* **153**, 1379-1393.

Weitzner, B.D., Dunbrack, R.L., Jr., and Gray, J.J. (2015). The origin of CDR H3 structural diversity. *Structure* **23**, 302-311.

WHO (2012). Management of severe malaria.

WHO (2016). World malaria report 2016.

Winter, G., Kawai, S., Haeggstrom, M., Kaneko, O., von Euler, A., Kawazu, S., Palm, D., Fernandez, V., and Wahlgren, M. (2005). SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *The Journal of experimental medicine* **201**, 1853-1863.

Winter, G., Lobley, C.M., and Prince, S.M. (2013). Decision making in xia2. *Acta crystallographica Section D, Biological crystallography* **69**, 1260-1273.

Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., and Watts, J.D. (2009). Mass-spectrometric identification and

relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* 27, 378-386.

Zemlin, M., Schelonka, R.L., Ippolito, G.C., Zemlin, C., Zhuang, Y., Gartland, G.L., Nitschke, L., Pelkonen, J., Rajewsky, K., and Schroeder, H.W., Jr. (2008). Regulation of repertoire development through genetic control of DH reading frame preference. *J Immunol* 181, 8416-8424.

Zhao, Y., Ren, J., Padilla-Parra, S., Fry, E.E., and Stuart, D.I. (2014). Lysosome sorting of beta-glucocerebrosidase by LIMP-2 is targeted by the mannose 6-phosphate receptor. *Nature communications* 5, 4321.

ARTICLE

Received 5 May 2016 | Accepted 4 Aug 2016 | Published 26 Sep 2016

DOI: 10.1038/ncomms12837

OPEN

The structural basis for CD36 binding by the malaria parasite

Fu-Lien Hsieh¹, Louise Turner², Jani Reddy Bolla³, Carol V. Robinson³, Thomas Lavstsen² & Matthew K. Higgins¹

CD36 is a scavenger receptor involved in fatty acid metabolism, innate immunity and angiogenesis. It interacts with lipoprotein particles and facilitates uptake of long chain fatty acids. It is also the most common target of the PfEMP1 proteins of the malaria parasite, *Plasmodium falciparum*, tethering parasite-infected erythrocytes to endothelial receptors. This prevents their destruction by splenic clearance and allows increased parasitaemia. Here we describe the structure of CD36 in complex with long chain fatty acids and a CD36-binding PfEMP1 protein domain. A conserved hydrophobic pocket allows the hugely diverse PfEMP1 protein family to bind to a conserved phenylalanine residue at the membrane distal tip of CD36. This phenylalanine is also required for CD36 to interact with lipoprotein particles. By targeting a site on CD36 that is required for its physiological function, PfEMP1 proteins maintain the ability to tether to the endothelium and avoid splenic clearance.

¹Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ²Centre for Medical Parasitology, Department of International Health, Immunology & Microbiology, University of Copenhagen and Department of Infectious Diseases, Rigshospitalet, Copenhagen 1017, Denmark.

³Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks, Oxford OX1 3QZ, UK. Correspondence and requests for materials should be addressed to M.K.H. (email: matthew.higgins@bioch.ox.ac.uk).

The majority of human deaths caused by malaria result from infection by *Plasmodium falciparum*¹. Unlike other human-infective *Plasmodium* species, this parasite contains a set of adhesive proteins, the PfEMP1, which are expressed on the surfaces of infected red blood cells, causing them to bind to blood vessels and tissues². This protects the parasite within from clearance by the spleen, allowing the development of increased parasite load^{3,4}. It also leads to specific pathology, with PfEMP1 causing the accumulation of parasite-infected erythrocytes in the placenta during pregnancy-associated malaria^{5,6} and in the brain during cerebral malaria^{7,8}.

The PfEMP1 are among the few parasite proteins constantly exposed to the host immune system during the blood phase of the parasite life cycle. They are therefore under conflicting selection pressure, experiencing pressure to diversify to aid immune evasion while retaining the capacity to interact with invariant human receptors to fulfil their cytoadhesive role. As a result they have expanded into a complex protein family, with around 60 members in each parasite genome and a system of antigenic variation that allows parasites to switch which PfEMP1 they expose to the immune system^{2,9}. Antibodies that bind PfEMP1 are found in people from malaria endemic regions and correlate with immunity from severe disease^{6,10–12}.

To mediate endothelial binding, PfEMP1 proteins contain large extracellular ectodomains consisting of 2–10 copies of two domain types, the Duffy-binding like (DBL) and cysteine-rich interdomain region (CIDR) domains^{9,13}. While these domains are extremely variable in sequence, they share a conserved architecture¹⁴ and can be classified into different subgroups based on sequence¹⁵. The vast majority of PfEMP1 proteins contain a DBL domain at the membrane-distal end, followed by a CIDR domain², with a variable number of succeeding domains. PfEMP1 are frequently modular, with individual domains maintaining the capacity to bind to specific receptors¹⁴. The domains closest to the membrane-distal end interact with endothelial receptors^{16–18}, while serum components such as IgM^{19,20} and α 2-macroglobulin²¹ interact with the membrane-proximal domains.

While a variety of ligands have been implicated in binding to PfEMP1, there are three major human endothelial proteins that have been shown to bind in a consistent manner that is predictable based on the sequences of their constituent domains². The scavenger receptor, CD36 binds to many classes of CIDR α domain (CIDR α 2–6)^{22,23} and is the host receptor most commonly found to interact with parasite isolates from patients^{24–26} and with laboratory-adapted parasite strains²⁷. ICAM-1 is also a common partner, with binding capacity retained in a subset of the DBL β domains^{16,28,29}. Finally, most CIDR α 1 domains bind to endothelial protein C receptor (EPCR) with 5–7 such domains predicted in the genome of each parasite isolate^{2,17}.

Approximately 84% of PfEMP1 proteins contain domains predicted to bind to CD36, making this the most common adhesion phenotype². CD36 is a cell surface scavenger receptor that plays a role in fatty acid uptake, angiogenesis and phagocytosis³⁰. It contains two transmembrane helices, between which lies a ~47 kDa extracellular domain. The structure of LIMP-2, a homologue of CD36, revealed an oval-shaped ectodomain with a hydrophobic cavity that traverses the whole length of the molecule. This cavity is thought to be the site through which fatty acids pass as they move towards the plasma membrane, although no fatty acids are present in the existing structures³¹. An α -helical ‘platform’ at the membrane distal tip of CD36 has been identified as a recognition site for oxidized low density lipoprotein (LDL) and fatty acids, with a lysine residue (K164) on this platform cross-linking to fatty acids and important for oxidized LDL binding^{31,32}. The structure of CD36 and its binding surface for

PfEMP1 were unknown, although a peptide-based study identified residues 146–164 of CD36 as forming the binding site³³.

CD36 is present in many tissues, allowing adhesion of infected erythrocytes in different sites within the vascular endothelium¹⁸. It is also present in the cells of the innate and acquired immune systems, modulating the binding of infected erythrocytes to macrophages and dendritic cells^{34,35}. This leads to a nuanced interaction between parasite and host³⁶. On one hand, the frequency of CD36 binding shows that it is advantageous to the parasite, allowing the avoidance of splenic clearance. CD36 binding may also benefit the parasite by providing a mechanism to reduce dendritic cell-mediated T-cell activation, hampering the capacity of the host immune system to clear the infection^{34,37}. On the other hand, CD36 binding increases the likelihood of infected erythrocytes being cleared by macrophages, perhaps reducing the chances of developing high parasitaemia^{35,38}. CD36 binding is therefore thought to allow the parasite to sustain a viable infection, while not putting the host at significant danger from elevated parasite levels³⁶.

The role of CD36 binding in malaria has been studied extensively. However, significant sequence diversity in CD36-binding CIDR domains has hampered efforts to understand the molecular determinants that are retained to allow interaction with CD36. To understand what makes CD36-binding such a successful interaction for the parasite, we have therefore combined structural and biophysical studies with analysis of thousands of sequences to show how CIDR α domains interact with CD36. We find that, despite extensive sequence diversity, the CD36-binding CIDR α domains retain a conserved hydrophobic pocket that accepts a phenylalanine residue found at the membrane distal tip of CD36. This phenylalanine is also required for CD36 to bind to oxidized LDL particles. We therefore show that the diverse PfEMP1 proteins retain a chemically conserved surface feature that allows them to interact with an invariant region of their receptor, maintaining the ability to tether to the endothelium and to avoid clearance.

Results

Determining the structure of CD36 bound to a CIDR α domain.

To determine the structure of CD36 in complex with a CIDR α domain, we selected the CIDR α 2.8 domain from the MCvar1 PfEMP1, as the structure of this isolated domain was already known³⁹. We coexpressed the ectodomain of CD36 (residues 35–439) together with the MCvar1 CIDR α 2.8 domain (residues 576–745 with a C626S mutation to remove an unpaired cysteine). These were expressed in human embryonic kidney (HEK) 293 cells in the presence of the class I α -mannosidase inhibitor, kifunensine and the secreted complex was purified. We used size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) to demonstrate the formation of a 1:1 complex, with no higher-order protein assemblies (Supplementary Fig. 1).

The CD36:CIDR α 2.8 complex was subjected to crystallization trials with *in situ* deglycosylation and proteolysis, allowing a complete data set to be collected to 2.07 Å resolution. Molecular replacement attempts in which the MCvar1 CIDR α 2.8 domain (PDB 3C64)³⁹ was used as a search model failed, while using LIMP-2 (PDB 4F7B)³¹ as a search model yielded a solution (Supplementary Fig. 2). Subsequent analysis of the electron density revealed that the MCvar1 CIDR α 2.8 domain structure had adopted a very different conformation from that seen in crystals of the domain alone (Supplementary Fig. 3). The molecular replacement solution obtained using LIMP-2 was therefore used as the starting point for an iterative cycle of model building and refinement to generate a structure of the CD36:CIDR α 2.8 complex (Table 1).

The structure of CD36 bound to fatty acids. The CD36 ectodomain adopts an architecture similar to LIMP-2, with an elongated structure, and is decorated with nine *N*-linked glycans distributed across the protein surface (Fig. 1a, Supplementary Fig. 4)³¹. While LIMP-2 undergoes a pH-dependent conformational change, with an ‘open’ conformation at pH 5.5 and a ‘closed’ conformation at pH 7.5 (refs 31,40,41), CD36, which was crystallized at pH 8.0, adopts an open conformation (Fig. 1b). Indeed the pH-sensor residue, H150 of LIMP-2, which stabilizes its closed conformation⁴¹ has been replaced by phenylalanine residue (F153) in CD36, and CD36 will therefore not undergo the pH-driven conformational change observed in LIMP-2, but will remain open and poised to accept fatty acids at neutral pH.

A distinctive feature of LIMP-2 is a cavity that spans most of the length of the molecule³¹. This is also present in CD36 and is lined predominantly by hydrophobic side chains. Unlike in LIMP-2, the cavity of CD36 contains two electron density features that resemble the extended hydrocarbon chain of a fatty acid. We identified these fatty acids by chloroform extraction of ligands from the crystallized protein, followed by GC/MS of the extracted material. This revealed that palmitic acid and stearic acid were most prevalent of a mixture of fatty acids present in the structure (Supplementary Fig. 5). Indeed CD36 has been shown to play a role in the uptake of these fatty acids⁴² and they have most likely bound to the protein during expression in human culture cells. We therefore modelled palmitic acid into these electron densities and observed a close fit. Both fatty acids are oriented with their carboxylic acid moieties at the membrane distal end. The majority of their interactions with CD36 are mediated by hydrophobic contacts as the fatty acids lie along the length of the cavity. However, the carboxylic acid group of the

membrane distal fatty acid also makes water-mediated hydrogen-bonding interactions with T195.

The entrance for the fatty acid translocation pathway has previously been proposed to lie close to K164 (refs 31,32). Indeed the central cavity has an opening close to this residue (entrance 1) that is a likely entry point for the membrane distal fatty acid and is the entrance that opens and closes in response to pH changes in LIMP-2 (Fig. 1c). In addition, a second opening (entrance 2) is also found at the membrane distal side of the CD36 ectodomain (Fig. 1c) and the second fatty acid occupying the cavity lies in a tunnel that proceeds from this entrance. Our CD36 structure therefore supports a model in which fatty acids move through the cavity, identifying a second possible entrance and further delineating favourable positions through which translocating

Table 1 Data collection and refinement statistics.	
CD36:CIDRα2.8	
Data collection	
Space group	C121
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.8, 40.7, 138.8
α , β , γ (°)	90.0, 114.9, 90.0
Wavelength	0.97949
Resolution (Å)	63.00–2.07 (2.12–2.07)
<i>R</i> _{pim} (%)	4.6 (53.1)
<i>I</i> / σ <i>I</i>	8.2 (1.5)
Completeness (%)	99.8 (99.5)
Multiplicity	3.2 (3.3)
Refinement	
Number of reflections	40,525 (3,984)
<i>R</i> _{work} / <i>R</i> _{free}	21.2/25.2
Number of atoms	
Protein	4,490
Ligands	233
Water	152
<i>B</i> -factors	
Protein	47.6
Ligand/ion	64.4
Water	44.6
<i>R.m.s</i> deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.19
Ramachandran plot	
Favoured (%)	97.0
Allowed (%)	3.0
All structures were determined from one crystal. Values in parentheses are for highest-resolution shell.	

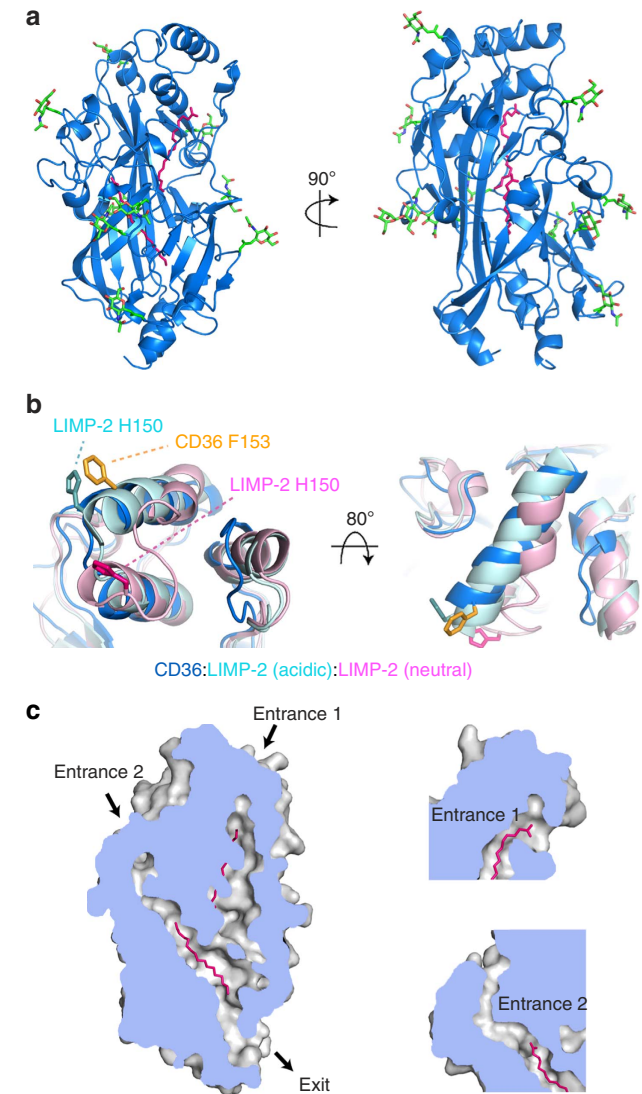


Figure 1 | The structure of CD36 and its binding of fatty acids. (a) The structure of CD36, shown in blue. The nine *N*-linked glycosylation sites and associated sugars are green while two palmitic acids are shown as pink sticks. (b) An alignment of CD36 (blue) with the structures of LIMP-2 at acidic (cyan) and neutral (pink) pH. Residues F153 from CD36 (orange) and H150 from LIMP-2 at acidic (light blue) and neutral (pink) pH are highlighted. (c) A section through a surface representation of CD36 showing the central core cavity occupied by two palmitic acids (pink). Insets show two putative entrances to this central cavity.

ligands move as they pass towards the membrane proximal exit and the plasma membrane.

The structural basis for CD36 binding by PfEMP1. As expected from characterizing the stoichiometry of the CD36:CIDR α 2.8 complex (Supplementary Fig. 1), the crystals contained an equal number of CD36 and CIDR α 2.8 molecules. However, two CIDR α 2.8 domains make significant contacts with each CD36 ectodomain. To determine which of these represents the physiological interaction, and which is due to crystal packing we assessed the binding of mutant proteins by surface plasmon resonance. Glycosylation sites were introduced into the CIDR α 2.8 domain to disrupt each of the two interfaces. In addition, a key interaction at site 1 is mediated by the insertion of a phenylalanine residue (F153) from CD36 into a hydrophobic pocket on the CIDR α 2.8 domain and we therefore produced the F153A mutant of CD36. The insertion of a glycosylation site into site 1 and the F153A mutation both disrupted the binding of the CIDR α 2.8 domain to CD36 (Supplementary Fig. 6). In contrast, a mutation inserting a glycan into site 2 had no effect on the interaction, identifying site 1 as the authentic binding surface.

The PfEMP1 CIDR α 2.8 domain therefore binds to the α -helical bundle at the membrane-distal tip of CD36, close to the major entrance (entrance 1) to the hydrophobic cavity (Fig. 2a) and in the region identified previously³³. The majority of the interaction is mediated by F153 from CD36, which fits into a hydrophobic pocket on the CIDR α 2.8 domain surface. Mutation of either F153, or of hydrophobic residues of the CIDR α 2.8 domain that line this pocket, abolishes the interaction (Fig. 2b, Supplementary Fig. 7). Secondary interactions are made by a smaller α -helix (residues 668–678) from the CIDR α 2.8 domain (mediating hydrogen bonding to CD36 through E672) and a loop (residues 647–656; again hydrogen bonding, this time through D650). Neither of these interactions is critical in the MCvar1 CIDR α 2.8 domain, as disrupting hydrogen bonds by mutagenesis led only to small reductions in affinity (Fig. 2b, Supplementary Fig. 7, Supplementary Table 1).

Previous studies have shown reduced binding of infected erythrocytes to CD36 in response to phosphorylation of T92 (ref. 43). We see no evidence of phosphorylation of T92 in the electron density maps and this side chain is >40 Å from the CIDR α 2.8 domain, suggesting that T92 phosphorylation will not directly block the interaction.

Four structures of CIDR α domains have been solved to date^{39,44,45}. Of these, three adopt a similar compact conformation^{44,45}, while the MCvar1 CIDR α 2.8 domain adopts an open conformation when crystallized at pH 4.2 (ref. 39) (Supplementary Figs. 3 and 8). In the structure presented here, the MCvar1 CIDR α 2.8 domain (crystallized at pH 8.0) adopts the typical compact configuration found in other CIDR domains crystallized around physiological pH, and aligns to the CIDR α 1 and γ domains with rmsd of 3.4 and 4.3 Å respectively (Supplementary Fig. 3). To understand the reasons for this structural discrepancy, we crystallized our HEK293 cell expressed MCvar1 CIDR α 2.8 domain and obtained crystals at pH 4.5 with the same space group and close to identical cell dimensions ($p6_32_2$; $a = b = 95.9$ Å, $c = 91.9$ Å) to those reported for bacterially expressed protein crystallized at pH 4.2. We therefore propose that the open conformation is an artefact of crystallizing the isolated CIDR α 2.8 domain in acidic conditions and that the domains adopt a conserved compact architecture under physiological conditions.

A comparison of the molecular features used by CIDR α domains to bind to CD36 and EPCR shows striking similarities, but with a structural inversion of the mechanism of binding (Fig. 3). In both cases, the CIDR α domains contain a

core three α -helical bundle with an insertion between the second and third core helices, α_2 and α_6 (Supplementary Fig. 8). This folds to make smaller helices that lie approximately perpendicular to the core helices. This insert has been identified as a 'homology block' and CIDR domains can be classified into three groups, the CIDR α 1 domains, the CIDR α 2–6 domains and the CIDR $\beta/\gamma/\delta$ domains, which each have a different homology block in this region¹⁵. Now that we have structures for members of each of these domain classes (Supplementary Fig. 3) we can see how these homology blocks fold to contribute to the formation of ligand-binding sites.

In the CIDR α 1 domains, the insert folds to form a helix and a kinked helix, with a critical phenylalanine residue protruding from the kink. This hydrophobic protrusion forms the centre of the binding site, occupying the hydrophobic groove of EPCR. It is surrounded by hydrogen bonding residues that further stabilize the interaction. In contrast, CIDR α 2–6 domains lack the kinked helix and the protruding phenylalanine. Instead the equivalent region of the protein contributes, together with residues from the core three-helical bundle, to a hydrophobic cavity (Fig. 3), which accepts a protruding phenylalanine residue from the ligand (F153 from CD36). Here too, the hydrophobic core of the interaction surface is surrounded by hydrophilic regions that can make hydrogen bonds. Residues previously identified to distinguish between CD36-binding and non-binding CIDR domains lie within this homology block, but do not directly contact CD36 (ref. 46). Therefore in both cases the domains form an extremely stable interaction using a contact surface built around a small hydrophobic core, surrounded by hydrogen bond donors and acceptors. The insertion of different homology blocks into the core helical bundle of the CIDR α domains allows them to do this in different ways, presenting ligand-binding surfaces with different architectures to interact with differently structured ligands.

Conserved structural features allow CD36 binding. We next aimed to understand how the CD36-binding CIDR α 2–6 binding domains diversify, while retaining ligand binding. The genome sequence of the 3D7 strain of *Plasmodium falciparum* allowed a previous systematic analysis of PfEMP1 to identify those that contain CIDR domains that interact with CD36 (ref. 23). Three classes of CIDR domains were identified, CIDR α , β and γ . All tested CIDR α domains, with the exception of those from the CIDR α 1 subgroup, bound to CD36 (ref. 23) while CIDR α 1 domains bind to EPCR (ref. 17). This suggested that CD36 binding is an extremely common phenotype, with $\sim 85\%$ of PfEMP1 proteins containing the CIDR α 2–6 domains predicted to bind. However the CIDR α 2–6 domains are highly variable with less than 33% pairwise sequence similarity².

With many PfEMP1 sequences emerging from recent parasite genome sequencing projects, we were able to conduct a more comprehensive analysis. We built sequence distance trees using 2386 CIDR α 2–6 domain sequences extracted from 233 parasite genomes that had been selected to be representative of global diversity¹⁵. No major sequence clusters could be verified by bootstrapping, but the sequences distributed into small subgroups as previously described (Fig. 4a)¹⁵. The alignment included 11 CIDR α domains from *Plasmodium reichenowi* that distributed across the tree of *Plasmodium falciparum* CIDR α 2–6 domains, as seen for other surface protein families⁴⁷. We used surface plasmon resonance to test a panel of CIDR α 2–6 domains randomly selected from across this sequence diversity tree for the ability to bind to CD36, showing them all to adopt nanomolar interactions with slow off-rates (Fig. 4a, Supplementary Fig. 9). Together with previously identified CD36 binders^{17,23,27}, this confirmed that all tested members of the CIDR α 2–6 classes bind to CD36.

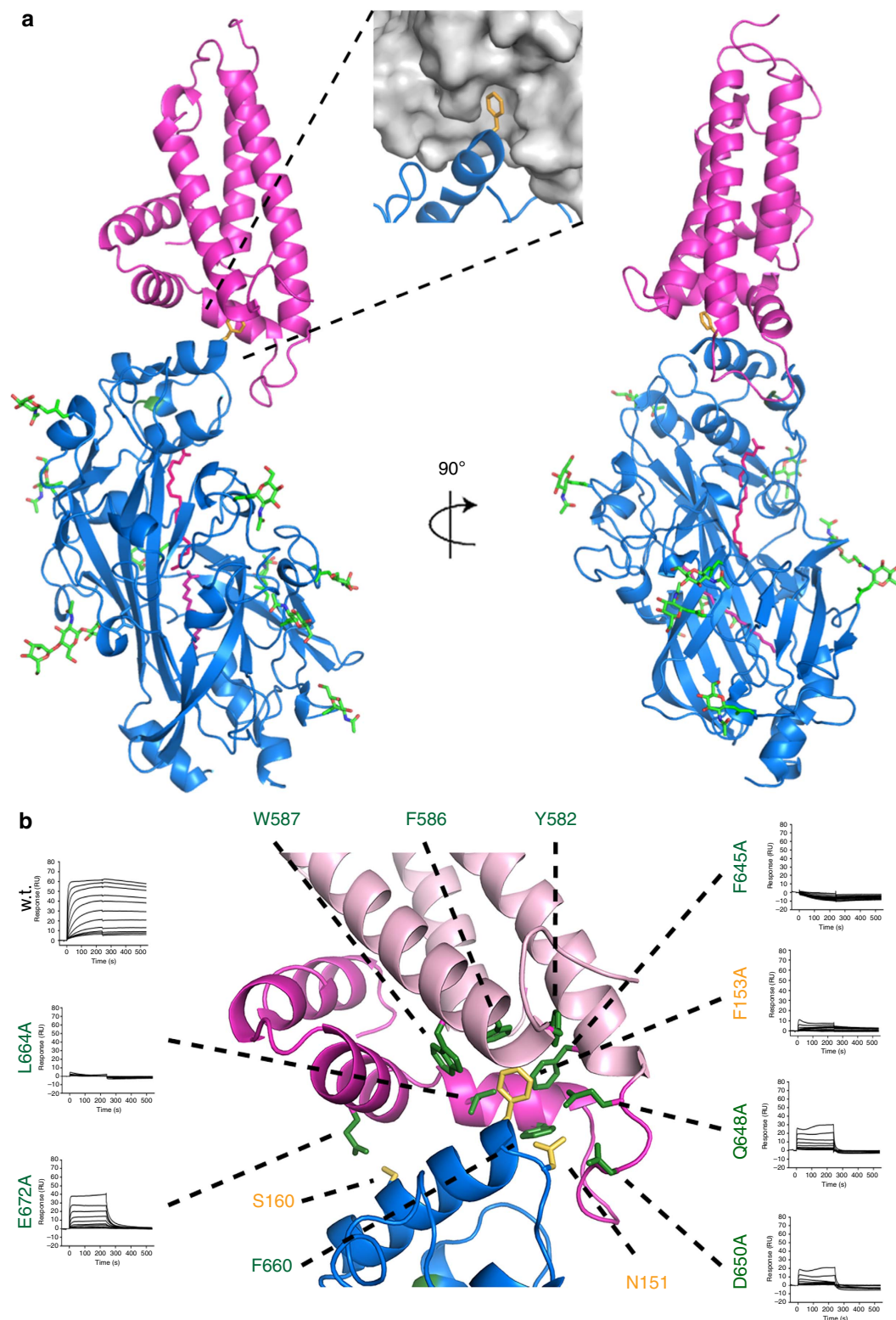


Figure 2 | The structural basis for CD36 binding by PfEMP1 proteins. (a) The structure of the complex of CD36 (blue) with the CIDR α 2.8 domain of the MCvar1 PfEMP1 protein (pink). N-linked glycans are shown in green. The inset shows a surface representation of the CIDR α domain with CD36 as a blue cartoon and F153 of CD36 in orange. (b) A close-up view of the interface between CD36 and the CIDR α domain with the key interacting residues labelled. The three-core α -helices of the CIDR α domain (α 1, α 2 and α 6) are coloured in light pink while the insert (α 3– α 5) is dark pink. Also shown is the effect of mutagenesis of key interacting residues as determined by surface plasmon resonance. The most critical interactions are mediated by F153 of CD36, which fits into a hydrophobic pocket lined by residues including F645 and L664 of the CIDR α domain.

We next produced a sequence logo for all members of these domain subclasses (Fig. 4b,c). The majority of the domain sequences could be aligned into a single logo. However, the loop

containing residues 673–683 was divergent in the CIDR α 3.1 subclass. As found before in the DBL and CIDR domains, the majority of conserved residues were cysteines or aromatics^{14,44}.

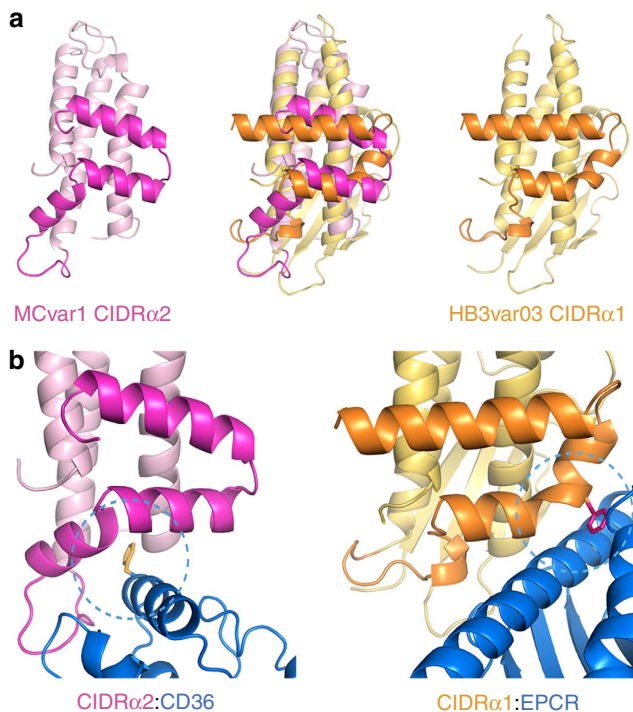


Figure 3 | A comparison of the features that allow CIDR α domains to bind to EPCR or CD36. (a) A comparison of the structure of a CD36-binding CIDR α 2 domain (pink) with that of an EPCR-binding CIDR α 1 domain (orange). Both domains share a core three α -helical bundle. The insertion that emerges between the second and third of these core helices forms a docking platform for ligands. **(b)** A close up of the binding interfaces that mediates the CIDR α 2:CD36 and CIDR α 1:EPCR interactions. The CIDR α 1 domains have a phenylalanine residue (F656) on a convex surface of the domain that protrudes into the hydrophobic groove of EPCR. In contrast, the CIDR α 2 domains have a hydrophobic pocket that binds to a protruding phenylalanine residue (F153) from CD36.

A sequence logo for residues that make direct interactions with CD36 (Fig. 4d) showed limited sequence conservation, but with hydrophobic residues exchanged for other hydrophobic residues and hydrogen bonding capacity conserved. Three of the most conserved residues that directly contact CD36 (Y582, F586 and F645) contribute to the hydrophobic pocket that accepts F153 (Fig. 2b).

To understand the distribution of conservation across the CD36-binding domains, we plotted sequence entropy, as a measure of chemical conservation, onto the structure of the MCvar1 CIDR α 2.8 domain (Fig. 5a,b). As seen previously for the CIDR α 1 domains, the significant majority of conserved residues were found in the core of the domain, stabilizing the protein fold. The major patch of surface exposed conservation was in, and immediately surrounding, the conserved hydrophobic patch, with substitutions in this region leading to conservation of the hydrophobicity of this pocket, maintaining the ability to accept the conserved F153 residue from CD36. Other contacting residues lacked direct conservation, but retained the ability to form hydrogen bonds. This suggests a shared binding mode for CD36 across the CIDR α 2-6 domains.

To support this conclusion, we tested the effect of the F153A mutation of CD36 on the binding of a sequence diverse set of ten CIDR α 2-6 domains. We found that the binding of each of these domains to CD36 was significantly reduced in the presence of the F153A mutations, with reduced overall binding levels and significantly faster off rates (Fig. 5c, Supplementary Fig. 10).

The F153A mutation did not have an equivalent effect on all of the CIDR α domains, suggesting that the importance of additional interactions differ across the family. However, conservation of chemical property in the F153 binding pocket, combined with a significant effect of F153 mutation on all of the interactions supports the idea of a conserved binding mode across the domain family, despite significant sequence diversification.

The effect of CIDR domains on the natural function of CD36. CD36 has a number of functions in mammalian physiology. It has the ability to bind oxidized lipoprotein particles³¹ and to interact with thrombospondin (TSP) to regulate angiogenesis^{30,48}. We therefore investigated whether the binding of PfEMP1 to CD36 can prevent it from interacting with physiological binding partners. Surface plasmon resonance experiments showed that the binding of CD36 to TSP was not altered by pre-incubation with CIDR α 2-6 domains (Fig. 6a). In contrast, the presence of CIDR α 2-6 domains abolished the interaction of CD36 with oxidized LDL particles (Fig. 6b,c). Furthermore, the F153 mutation, which disrupts the interaction of CD36 with CIDR α 2-6 also abolishes binding of CD36 to oxidized LDL particles, showing that this region of the protein is essential to interact both with lipoprotein particles and PfEMP1 (Fig. 6d). The parasite therefore targets a part of CD36 that is essential for its physiological role in fatty acid uptake, reducing the likelihood that the human host can escape from PfEMP1 binding by altering its CD36.

Discussion

The PfEMP1 proteins of *Plasmodium falciparum* have evolved under conflicting evolutionary pressures. On one hand, they are constantly exposed to the immune system and are therefore under constant pressure to diversify, allowing immune evasion. On the other hand, they must retain functionally important features, maintaining the capacity to interact with specific endothelial ligands, allowing them to perform their role in endothelial adhesion. In this study we have combined structural studies of a MCvar1 CIDR α 2.8 domain in complex with CD36 with the analysis of the sequences that represent much of the global diversity of CD36 binding PfEMP1 domains, in order to identify the molecular features that are required to mediate CD36 binding.

The CIDR domains provide a versatile molecular scaffold for the evolution of different binding phenotypes. Their structure is built upon a core three-helical bundle and between the second and third core helices emerges an insertion of ~ 60 residues. Much of this insertion has been identified as a ‘homology-block’, and is a module that differs between the main CIDR domain classes¹⁵. In the CIDR α 1 domains the insertion is ~ 60 residues in length and folds to form the helix and kinked-helix that provide the majority of the EPCR-binding surface⁴⁴. In the CIDR α 2-6 domains, a different insertion in this region, of ~ 70 residues, adopts a subtly different structure. The helix and kinked-helix of CIDR α 1 domains are replaced in CIDR α 2-6 domains by two shorter helices ($\alpha 4$ and $\alpha 5$ in Supplementary Fig. 8). Together with residues emerging from the core three-helical bundle, these form an open, hydrophobic cavity containing conserved aromatic residues, which is missing in the CIDR α 1 domains. It is this cavity that forms the majority of the interaction site for CD36. Therefore, the equivalent regions of CIDR α domains adopt very different architectures, providing binding sites for very different ligands.

Although the binding sites for EPCR and CD36 are very different in shape, both interactions are mediated by similar chemical properties. In the case of EPCR binding, the CIDR α 1 domains present a convex surface formed from two hydrophobic

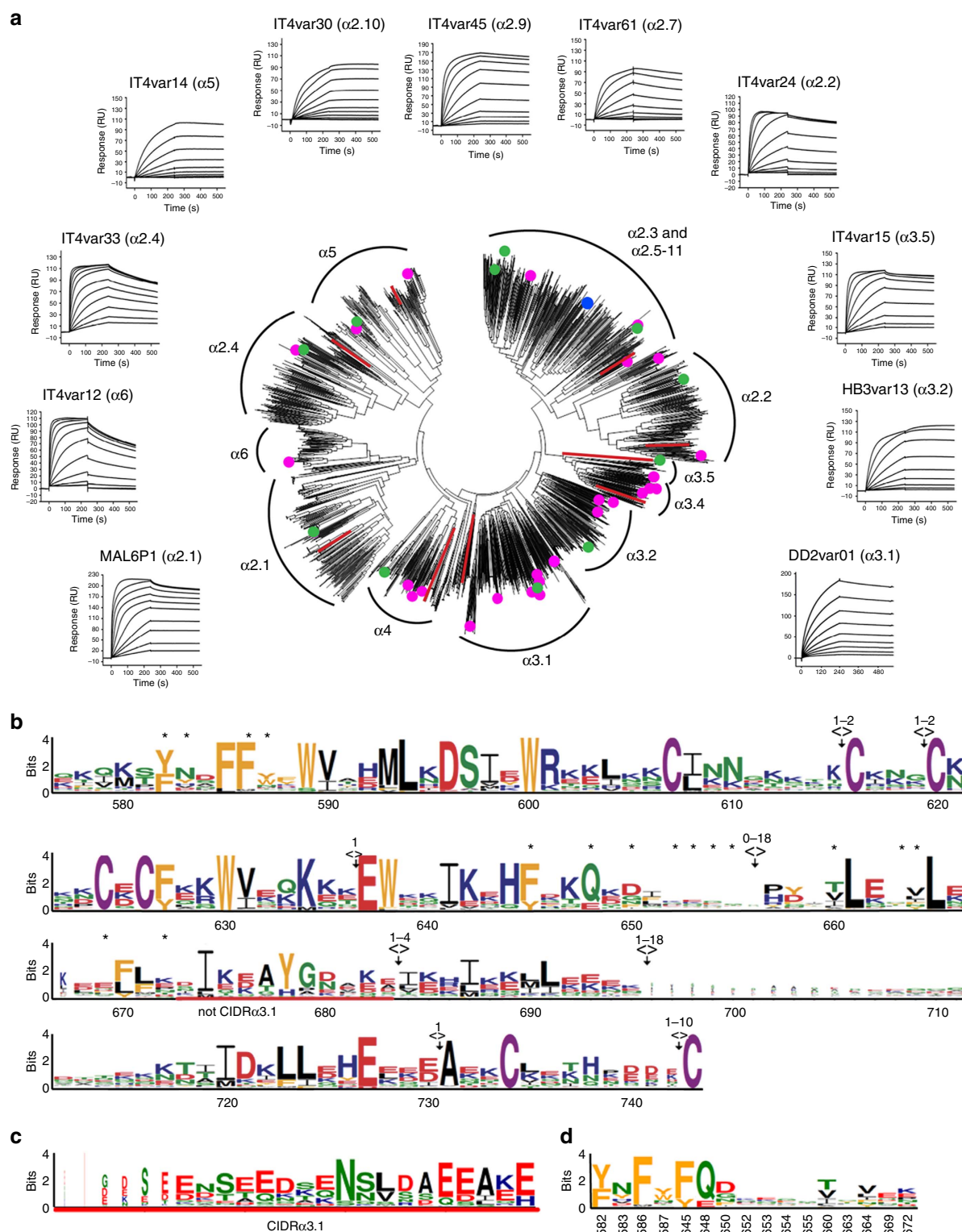


Figure 4 | Extensive diversity in CD36 binding CDR α domains. (a) Sequence distance tree of 2386 full-length CDR $\alpha 2-6$ domains. Red lines represent sequences from *Plasmodium reichenowi*. A blue circle marks the sequence of the crystallized CDR $\alpha 2.8$ domain (also see Supplementary Fig. 8). Green circles mark sequences of recombinant CDR α domains for which the affinity for CD36 binding were tested (corresponding SPR traces given). Pink circles mark sequences of CDR domains previously demonstrated to bind CD36. Annotated clusters contained previously defined CDR $\alpha 2-6$ subclasses. All tested domains from the CDR $\alpha 2-6$ subclasses bind to CD36. (b) Sequences of the 2386 CDR α domains were aligned and a sequence logo generated of residues equivalent to those found in the MCvar1 CDR α domain (numbered as in MCvar1). Deletions (>) and insertions (< >) are indicated. The region underlined by a red line is found in all CDR $\alpha 2-6$ domains except for CDR $\alpha 3.1$. Residues labelled with * make direct contact to CD36. (c) The region found in CDR $\alpha 3.1$ domains that replaces that underlined in red. (d) A sequence logo for the residues that make direct contacts to CD36.

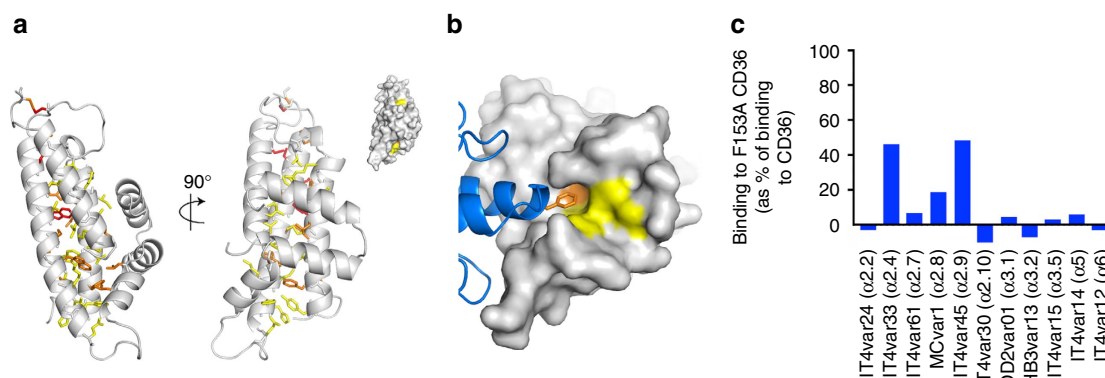


Figure 5 | Limited chemical conservation allows CD36 binding. (a) Conservation in the CD36-binding CIDR α domain is plotted onto the structure of the MCvar1 CIDR α 2.8 domain. Absolutely conserved residues are shown as red sticks. Residues with property entropy score of less than 0.1 (but not totally conserved) are orange and those with scores of 0.1–0.3 are yellow. The inset shows a surface representation in the same orientation and colours, showing that conserved residues cluster predominantly in core of the domain, stabilizing its structure. (b) A surface representation of the CIDR α domain coloured as in A, with CD36 in blue. This shows that residues in the hydrophobic pocket of the CIDR α domain are the most chemically conserved feature on the CIDR α domain surface. (c) The effect of the F153A mutant of CD36 on the binding of a diverse panel of CIDR α 2–6 domains shows that the interaction mediated by F153 of CD36 plays an important role in binding across the CIDR α 2–6 domain family.

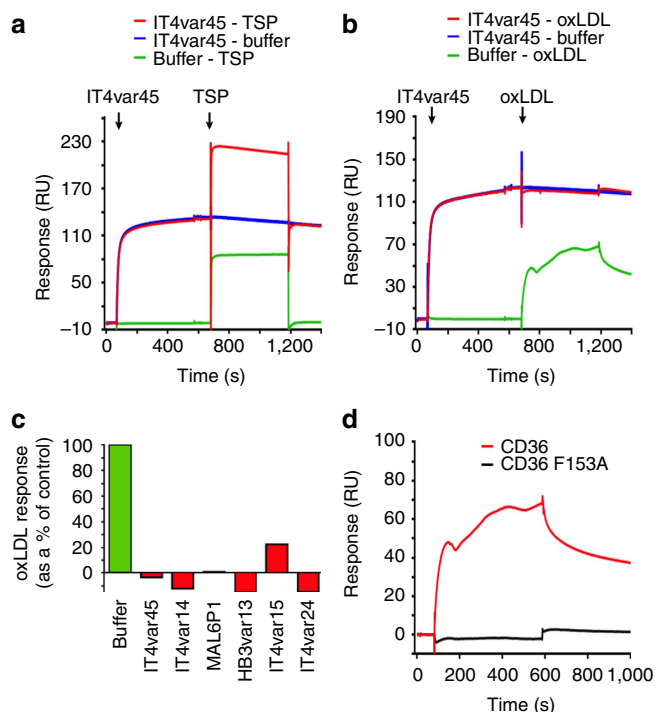


Figure 6 | PfEMP1 proteins prevent oxidized LDL from binding to CD36.

(a) Surface plasmon resonance analysis showing that the preincubation of CD36 with IT4var45 CIDR α 2.9 does not affect the binding of thrombospondin (TSP) to CD36. (b) Surface plasmon resonance analysis showing that the preincubation of CD36 with IT4var45 CIDR α 2.9 prevents the binding of oxidized LDL particles (oxLDL) to CD36. (c) SPR data showing the inhibition of oxLDL binding to CD36 by a panel of CIDR α 2–6 domains. (d) Demonstration by surface plasmon resonance that the F153A mutation blocks the binding of oxLDL to CD36.

phenylalanine residues. These residues, and in particular F656 in the HB3var03 CIDR α 1 domain, protrude into the hydrophobic groove of EPCR, while hydrophilic residues that surround this hydrophobic cluster make hydrogen-bonding interactions. The CD36-binding domains present a binding site that is conceptually

the inverse of this. Instead of a hydrophobic protrusion, the CD36-binding surface is predominantly formed from a concave hydrophobic pocket into which docks a phenylalanine, F153 from CD36. A loop from the CIDR α domain then makes hydrogen bonds with other regions of CD36. In both cases, this combination of a small hydrophobic core, together with a larger, complementary hydrophilic interface, generates a high affinity and stable binding site, allowing the domains to interact with their ligand with a slow off rate, stabilizing cytoadhesion of infected erythrocytes against the strong forces of blood flow.

In both cases, these binding surfaces have diversified significantly, with little sequence conservation in the residues that made direct contacts with EPCR or CD36. However, in both cases, residues involved in stabilizing the fold are conserved and sequence variation in residues that interact directly with the ligands is conservative in chemistry, maintaining the capacity to bind. This raises the possibility of generating antibodies that recognize these conserved chemical features and show broadly inhibitory potential. In particular, the EPCR binding surface of the CIDR α 1 domains protrudes from the domain, presenting a surface that appears compatible with recognition by an antibody. In contrast, the CD36 binding site is concave, with its conserved hydrophobic residues concealed in a pocket that may be less readily recognized. This binding surface is surrounded by a sequence diverse protein surface containing a flexible loop (residues 647–657) that might make antibody-mediated recognition less likely. Perhaps this is why the CD36 interaction is such a successful one from the perspective of the parasite, with the binding site reducing the exposure of chemically conserved determinants from detection by acquired immunity, allowing a diverse set of CIDR α 2–6 domains and a system of antigenic variation to provide the changes necessary to maintain the capacity to cytoadhere and to protect the parasite from splenic clearance.

A second strategy used by both EPCR- and CD36-binding CIDR domains is to interact with functionally important sites in their human receptors to avoid the ready evolution of escape mutants in the human genome. The CIDR α 1 domains bind to the region of EPCR that is required for its interaction with its ligand, protein C. Indeed the conserved phenylalanine of the CIDR α 1 domains occupies a pocket that is also occupied by a phenylalanine from protein C. The parasite has evolved a similar trick in its CIDR α 2–6 domains (Fig. 7). Here the CIDR domains

interact with a phenylalanine residue that protrudes from the surface of CD36. But why are polymorphisms in this residue not favoured in malaria endemic regions, as these would be expected to allow escape from binding? We show here that this residue is also important for binding to oxidized LDL particles, one of the physiological binding partners of CD36, and that the CIDR α domains prevent oxidized LDL from binding. Once again, the PfEMP1 proteins target a functionally important site, reducing the likelihood of a host escape mutant that would prevent infected erythrocytes from sequestering away from splenic destruction.

In addition, the region of CD36 on which this phenylalanine is displayed is the site of recognition for long chain fatty acids. Fatty acids can be cross-linked to K164 of CD36 and are then thought to move across to enter the central CD36 cavity and pass through the CD36 ectodomain to the membrane surface. The CIDR α 2.8 domain lies directly in between K164 and the cavity entrance, potentially blocking lipid uptake. With fatty acid uptake and oxidized LDL binding linked to fatty acid metabolism and atherosclerosis, it will be interesting to see whether CIDR α 2-6 domains can prevent fatty acid uptake as a proof of principle of targeting this region of CD36.

In summary, our characterization of the structural basis for CD36-binding by PfEMP1, coupled with our previous analysis of EPCR-binding has shown how the complex PfEMP1 protein family balances the need to adhere tightly to unchanging human receptors while diversifying to evade antibody-mediated recognition. Functionally important regions of human receptors are targeted, reducing the likelihood of mutations that disrupt binding, but also, inadvertently blocking the physiological roles of these receptors, with the potential to lead to disease outcomes. While human receptors are restricted from evolving away from the interaction, the PfEMP1 proteins are hugely diverse, retaining their structure and the chemical features of their binding site while altering nearly all of their surface residues. The CD36-binding CIDR domains also tuck their ligand-binding site in a small pocket, reducing the likelihood of antibody-mediated recognition. In this way, the parasite has evolved a surface protein family that

can maintain the capacity to tether it away from splenic clearance, while avoiding clearance by the acquired immune system.

Methods

Protein expression and purification. A gene encoding the human CD36 ectodomain (residues 35–439) was inserted into the pHLsec vector⁴⁹, modified with a C-terminal 8xHis tag (pHLsec-8H). A synthetic complementary DNA clone of MCvar1 CIDR α 2.8 with Cys626 mutated into Ser to remove an unpaired cysteine (codon-optimized for expression in mammalian cells) was obtained from GeneArt (Invitrogen), and residues 576–745 were cloned into pHLsec-8H. CD36 constructs used for SPR experiments had an additional C-terminal BAP tag that was biotinylated *in vitro* by incubation with BirA ligase (Sigma-Aldrich). Mutants were generated using two-step overlapping PCR experiments and verified by DNA sequencing (Source Bioscience) and cloning primers are given in Supplementary Table 2.

CD36 and MCvar1 CIDR α 2.8 were coexpressed in HEK293 (ATCC CRL-11268) cells in the presence of 5 μ M kifunensine. The conditioned media was dialysed against phosphate-buffered saline with the addition of 0.5 M NaCl and the protein complex was purified by immobilized metal ion affinity chromatography using TALON resin (Clontech). The CD36:CIDR α 2.8 complex was further purified by size-exclusion chromatography using a Superdex 200 16/600 column (GE Healthcare Life Sciences) and 10 mM Bis-Tris, pH 6.5 and 0.3 M NaCl. For SEC-MALS experiments, proteins were expressed and purified as above. For surface plasmon resonance analysis, CD36 and MCvar1 CIDR α 2.8 proteins and their mutants were expressed individually in HEK293 cells, and were purified as above, before buffer exchange into 10 mM HEPES, pH 7.5, 0.15 M NaCl.

Recombinant CIDR α 2-6 domains were expressed and purified as previously described⁴⁴. In short, domain sequences with boundaries as defined in Rask *et al.*¹⁵ were used to design domain-encoding DNA. These were codon optimized for expression in *Drosophila* Sf9 cells and were delivered already inserted into the baculovirus-insect cell expression vector pAcGP67-A (Geneart, Regensburg, Germany). All proteins were expressed with C-terminal V5 and His-tags, and were purified by nickel affinity chromatography. Sequences are given in Supplementary Table 3.

Crystallization of a complex of CD36 and MCvar1 CIDR α 2.8. The purified CD36:CIDR α 2.8 complex was concentrated to 10 mg/ml in the presence of 1% (v/v) *Flavobacterium meningosepticum* endoglycosidase-F1 and carboxypeptidase Y (Sigma-Aldrich) for *in situ* de-glycosylation and proteolysis. The protein samples were then subjected to sitting drop vapour diffusion crystallization trials in SwisSci 96-well plates by mixing 100 nl protein with 100 nl reservoir solution. The CD36:CIDR α 2.8 complex crystallized in 0.2 M NaCl, 20% (w/v) PEG6000, 0.1 M Tris, pH 8.0 at 18°. For cryo-protection, crystals were transferred into mother liquor supplemented with 25% (w/v) glycerol and then cryo-cooled in liquid nitrogen for storage and data collection.

Data collection and structure determination. Data were collected at beamline IO2 (Diamond Light Source UK) with a Pilatus 6M detector. Diffraction data of the CD36:CIDR α 2.8 complex were indexed and integrated using XIA2 (ref. 50) coupled with XDS⁵¹, and scaled and merged using Aimless⁵² to a resolution of 2.07 Å. A subset of 5% of randomly selected diffraction data were used for the calculation of R_{free} .

The structure of the CD36:CIDR α 2.8 complex was determined by molecular replacement in PHASER⁵³ with the structure of LIMP-2 (PDB 4F7B) used as a search model to obtain the initial phases. The models were completed by manual building in COOT⁵⁴ and refinement in autoBuster⁵⁵.

Surface plasmon resonance. SPR binding measurements were performed using a Biacore T200 instrument (GE Healthcare Life Science) at 25 °C in 20 mM HEPES, pH 7.5, 0.15 M NaCl, 0.005% (w/v) Tween20. CD36 wild-type and mutants with BAP tags were expressed in HEK293T cells, incubated with BirA for *in vitro* biotinylation and immobilized onto sensor CAPture chip (GE Healthcare Life Sciences) to ~700 RU. MCvar1 CIDR α 2.8 domain and mutants used as analytes were expressed in HEK293 cells and purified as described above. In all binding experiments, analytes were injected for 240 s with a 300 s dissociation time before chip regeneration. Affinities were estimated using equilibrium binding measurements and data were fitted with the BIAevaluation software using a one site-binding model. For competition experiments, different CIDR α proteins were injected onto the chip with 500 s association time, followed by injecting either 400 nM Recombinant Human Thrombospondin-1 (TSP-1; R&D SYSTEMS) or 10 nM Native Human low density lipoprotein (oxidized)(oxLDL; AbD Serotec). All buffers were supplemented with 72.5 μ M Oleic acid (MP Biochemicals Australasia Pty Limited) for oxLDL binding experiments.

Size-exclusion chromatography multi-angle light scattering. SEC-MALS experiments were performed using an analytical Superdex S200 10/300 GL column (GE Healthcare Life Sciences) connected to online static light-scattering (DAWN HELEOS 8, Wyatt Technology) and differential refractive index (Optilab EX, Wyatt Technology). Purified sample (MCvar1 CIDR α 2.8 at 1.5 mg/ml and CD36:MCvar1 CIDR α 2.8 complex at 1 mg/ml) were injected into a column

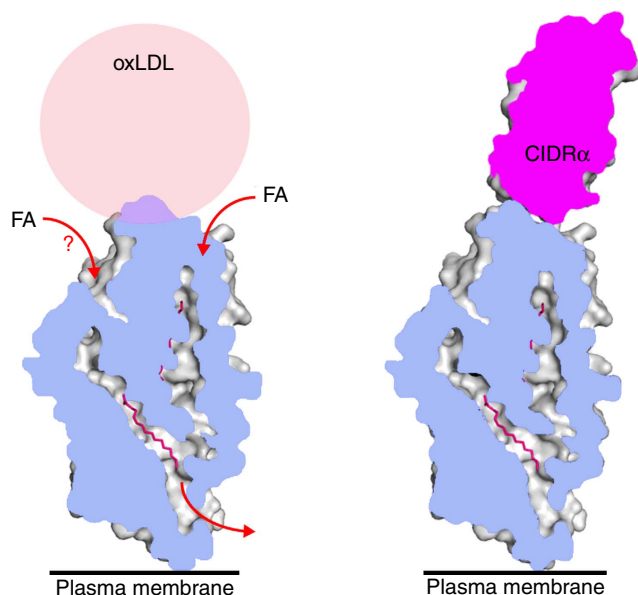


Figure 7 | A model for the inhibition of oxLDL binding by PfEMP1. One of the physiological roles of CD36 is to interact with oxidized LDL (oxLDL). Fatty acids can be incorporated from oxLDL particles, or other transport systems and pass through the central cavity of CD36 to the membrane. The PfEMP1 CIDR α 2-6 domains interact with the same surface of CD36 as oxLDL and compete for binding.

equilibrated with 10 mM HEPES pH 7.5, 0.15 M NaCl. Molecular mass determination was performed using an adapted RI increment value (dn/dc standard value; 0.186 ml/g) to account for glycosylation. The theoretical molecular weight was predicated from amino acid sequence plus 1865 Da per N-linked glycosylation site for recombinant protein produced from HEK293T cells in the presence of kifunensine with limited glycosylation. Data were analysed using the ASTRA 6.1 software package (Wyatt Technology).

Analysis of PfEMP1 sequences. A total of 263 CIDR α 2-6 domains from seven previously annotated whole genome sequenced parasites¹⁵ were used as previously described⁴⁴, to blastp extract CIDR α sequences from assemblies of Illumina whole genome sequencing data from 226 parasite samples collected in both Africa and Asia⁵⁶ (Study number ERP000190) available through the MalariaGEN community. Only full length domain sequences were kept (domain boundaries as defined in¹⁵). This resulted in a total of 2386 CIDR α 2-6 sequences, including 11 sequences from *P. reichenowi* CIDR α 2-6 domains, which were used for the analysis of sequence diversity. A hand-corrected MUSCLE alignment was used to generate sequence logos by WebLogo 3 (ref. 57) and neighbour joining sequence distance trees by MEGA⁵⁸. No major sequence clusters could be verified by bootstrapping, but sequences distributed as previously observed¹⁵.

Determination of unknown ligand(s) using GC/MS. To identify the nature of the bound ligand in crystals of the CD36:CIDR α complex, we used gas chromatography coupled with mass spectrometry (GC/MS). Briefly, purified CD36:CIDR α protein was incubated at 100 °C for 5 min, and then chloroform was added to a final concentration of 80% (v/v) to denature the protein and extract the ligand. The sample was derivatized with N,O-Bis(trimethylsilyl)trifluoroacetamide. The derivatized sample was then analysed by GC/MS on a Waters GCT Classic instrument. Ligand identification was performed by comparison of mass spectrum of the analyte with the NIST library database.

Data availability. The structure is deposited with pdb code 5LGD. All other data, including CIDR α domain sequences, are available from the authors on request.

References

- World Health Organisation. *World Malaria Report 2014* (World Health Organisation, 2014).
- Smith, J. D., Rowe, J. A., Higgins, M. K. & Lavstsen, T. Malaria's deadly grip: cytoadhesion of Plasmodium falciparum-infected erythrocytes. *Cell Microbiol.* **15**, 1976–1983 (2013).
- Bachmann, A. *et al.* Absence of erythrocyte sequestration and lack of multicopy gene family expression in Plasmodium falciparum from a splenectomized malaria patient. *PLoS ONE* **4**, e7459 (2009).
- David, P. H., Hommel, M., Miller, L. H., Udeinya, I. J. & Oligino, L. D. Parasite sequestration in Plasmodium falciparum malaria: spleen and antibody modulation of cytoadherence of infected erythrocytes. *Proc. Natl Acad. Sci. USA* **80**, 5075–5079 (1983).
- Fried, M. & Duffy, P. E. Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science* **272**, 1502–1504 (1996).
- Fried, M. & Duffy, P. E. Designing a VAR2CSA-based vaccine to prevent placental malaria. *Vaccine* **33**, 7483–7488 (2015).
- Storm, J. & Craig, A. G. Pathogenesis of cerebral malaria— inflammation and cytoadherence. *Front. Cell Infect. Microbiol.* **4**, 100 (2014).
- MacPherson, G. G., Warrell, M. J., White, N. J., Looareesuwan, S. & Warrell, D. A. Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. *Am. J. Pathol.* **119**, 385–401 (1985).
- Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).
- Fried, M., Nosten, F., Brockman, A., Brabin, B. J. & Duffy, P. E. Maternal antibodies block malaria. *Nature* **395**, 851–852 (1998).
- Bull, P. C. *et al.* Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat. Med.* **4**, 358–360 (1998).
- Chan, J. A. *et al.* Targets of antibodies against Plasmodium falciparum-infected erythrocytes in malaria immunity. *J. Clin. Invest.* **122**, 3227–3238 (2012).
- Su, X. Z. *et al.* The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell* **82**, 89–100 (1995).
- Higgins, M. K. & Carrington, M. Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein Sci.* **23**, 354–365 (2014).
- Rask, T. S., Hansen, D. A., Theander, T. G., Gorm Pedersen, A. & Lavstsen, T. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes—divide and conquer. *PLoS Comput. Biol.* **6** (2010).
- Berendt, A. R., Simmons, D. L., Tansey, J., Newbold, C. I. & Marsh, K. Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum. *Nature* **341**, 57–59 (1989).
- Turner, L. *et al.* Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* **498**, 502–505 (2013).
- Ockenhouse, C. F., Tandon, N. N., Magowan, C., Jamieson, G. A. & Chulay, J. D. Identification of a platelet membrane glycoprotein as a falciparum malaria sequestration receptor. *Science* **243**, 1469–1471 (1989).
- Semblat, J. P., Raza, A., Kyes, S. A. & Rowe, J. A. Identification of Plasmodium falciparum var1CSA and var2CSA domains that bind IgM natural antibodies. *Mol. Biochem. Parasitol.* **146**, 192–197 (2006).
- Stevenson, L. *et al.* Investigating the function of Fc-specific binding of IgM to Plasmodium falciparum erythrocyte membrane protein 1 mediating erythrocyte rosetting. *Cell Microbiol.* **17**, 819–831 (2015).
- Stevenson, L. *et al.* alpha2-macroglobulin can crosslink multiple plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) molecules and may facilitate adhesion of parasitized erythrocytes. *PLoS Pathog.* **11**, e1005022 (2015).
- Smith, J. D. *et al.* Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. *Mol. Biochem. Parasitol.* **97**, 133–148 (1998).
- Robinson, B. A., Welch, T. L. & Smith, J. D. 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 (2003).
- Heddi, A. *et al.* Fresh isolates from children with severe Plasmodium falciparum malaria bind to multiple receptors. *Infect. Immun.* **69**, 5849–5856 (2001).
- Rogerson, S. J. *et al.* Cytoadherence characteristics of Plasmodium falciparum-infected erythrocytes from Malawian children with severe and uncomplicated malaria. *Am. J. Trop. Med. Hyg.* **61**, 467–472 (1999).
- Udomsangpet, R. *et al.* Promiscuity of clinical Plasmodium falciparum isolates for multiple adhesion molecules under flow conditions. *J. Immunol.* **158**, 4358–4364 (1997).
- Janes, J. H. *et al.* Investigating the host binding signature on the Plasmodium falciparum PfEMP1 protein family. *PLoS Pathog.* **7**, e1002032 (2011).
- Smith, J. D. *et al.* 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 (2000).
- Bengtsson, A. *et al.* A novel domain cassette identifies Plasmodium falciparum PfEMP1 proteins binding ICAM-1 and is a target of cross-reactive, adhesion-inhibitory antibodies. *J. Immunol.* **190**, 240–249 (2013).
- Silverstein, R. L. & Febbraio, M. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Sci. Signal.* **2**, re3 (2009).
- Neculai, D. *et al.* Structure of LIMP-2 provides functional insights with implications for SR-BI and CD36. *Nature* **504**, 172–176 (2013).
- Pepino, M. Y., Kuda, O., Samovski, D. & Abumrad, N. A. Structure-function of CD36 and importance of fatty acid signal transduction in fat metabolism. *Annu. Rev. Nutr.* **34**, 281–303 (2014).
- Baruch, D. I., Ma, X. C., Pasloske, B., Howard, R. J. & Miller, L. H. CD36 peptides that block cytoadherence define the CD36 binding region for Plasmodium falciparum-infected erythrocytes. *Blood* **94**, 2121–2127 (1999).
- Urban, B. C. *et al.* Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature* **400**, 73–77 (1999).
- McGilvray, I. D., Serghides, L., Kapus, A., Rotstein, O. D. & Kain, K. C. Nonopsonic monocyte/macrophage phagocytosis of Plasmodium falciparum-parasitized erythrocytes: a role for CD36 in malarial clearance. *Blood* **96**, 3231–3240 (2000).
- Serghides, L., Smith, T. G., Patel, S. N. & Kain, K. C. CD36 and malaria: friends or foes? *Trends Parasitol.* **19**, 461–469 (2003).
- Elliott, S. R. *et al.* Inhibition of dendritic cell maturation by malaria is dose dependent and does not require Plasmodium falciparum erythrocyte membrane protein 1. *Infect. Immun.* **75**, 3621–3632 (2007).
- Patel, S. N. *et al.* CD36 mediates the phagocytosis of Plasmodium falciparum-infected erythrocytes by rodent macrophages. *J. Infect. Dis.* **189**, 204–213 (2004).
- Klein, M. M. *et al.* The cysteine-rich interdomain region from the highly variable plasmodium falciparum erythrocyte membrane protein-1 exhibits a conserved structure. *PLoS Pathog.* **4**, e1000147 (2008).
- Dang, M. *et al.* Molecular mechanism of SCARB2-mediated attachment and uncoating of EV71. *Protein Cell* **5**, 692–703 (2014).
- Zhao, Y., Ren, J., Padilla-Parra, S., Fry, E. E. & Stuart, D. I. Lysosome sorting of beta-glucocerebrosidase by LIMP-2 is targeted by the mannose 6-phosphate receptor. *Nat. Commun.* **5**, 4321 (2014).
- Ibrahimi, A. *et al.* Expression of the CD36 homolog (FAT) in fibroblast cells: effects on fatty acid transport. *Proc. Natl Acad. Sci. USA* **93**, 2646–2651 (1996).

43. Ho, M. *et al.* Ectophosphorylation of CD36 regulates cytoadherence of *Plasmodium falciparum* to microvascular endothelium under flow conditions. *Infect. Immun.* **73**, 8179–8187 (2005).
44. Lau, C. K. *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).
45. Vigan-Womas, I. *et al.* Structural basis for the ABO blood-group dependence of *Plasmodium falciparum* rosetting. *PLoS Pathog.* **8**, e1002781 (2012).
46. Gamain, B., Smith, J. D. & Miller, L. H. & Baruch, D.I. Modifications in the CD36 binding domain of the *Plasmodium falciparum* variant antigen are responsible for the inability of chondroitin sulfate A adherent parasites to bind CD36. *Blood* **97**, 3268–3274 (2001).
47. Otto, T. D. *et al.* Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* **5**, 4754 (2014).
48. Asch, A. S., Barnwell, J., Silverstein, R. L. & Nachman, R. L. Isolation of the thrombospondin membrane receptor. *J. Clin. Invest.* **79**, 1054–1061 (1987).
49. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1243–1250 (2006).
50. Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
51. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
52. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
55. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380 (2012).
56. Miotto, O. *et al.* Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat. Genet.* **45**, 648–655 (2013).
57. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
58. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

Acknowledgements

This work was supported by a Wellcome Trust Investigator Award to MKH. F.-L.H. is supported by a fellowship from the Taiwan Bio-Development Foundation. L.T. and T.L. received support from Danish Council for Independent Research, Sapere Aude program DFF-4004-00624B and The Lundbeck Foundation. We thank David Staunton for help with biophysical methods, the beamline scientists at Diamond Light Source for assistance with data collection, Jakob Jespersen for assistance with sequence analysis, Frank Lennartz for valuable discussion and comments on the manuscript and the mass spectrometry research facility at the University of Oxford. We are grateful for use of data from the MalariaGEN community project on *Plasmodium falciparum* population genomics as described⁵⁶.

Author contributions

F.-L.H. conducted protein production, crystallization and biophysical experiments. F.-L.H. and M.K.H. determined crystal structures. J.R.B. and C.V.R. conducted and analysed mass spectrometry experiments. L.T. produced variant proteins and conducted binding studies. T.L. conducted sequence analysis. F.-L.H., T.L. and M.K.H. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Hsieh F.-L. *et al.* The structural basis for CD36 binding by the malaria parasite. *Nat. Commun.* **7**:12837 doi: 10.1038/ncomms12837 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

The structure of a LAIR1-containing human antibody reveals a novel mechanism of antigen recognition

Fu-Lien Hsieh, Matthew K Higgins*

Department of Biochemistry, University of Oxford, Oxford, United Kingdom

Abstract Antibodies are critical components of the human adaptive immune system, providing versatile scaffolds to display diverse antigen-binding surfaces. Nevertheless, most antibodies have similar architectures, with the variable immunoglobulin domains of the heavy and light chain each providing three hypervariable loops, which are varied to generate diversity. The recent identification of a novel class of antibody in humans from malaria endemic regions of Africa was therefore surprising as one hypervariable loop contains the entire collagen-binding domain of human LAIR1. Here, we present the structure of the Fab fragment of such an antibody. We show that its antigen-binding site has adopted an architecture that positions LAIR1, while itself being occluded. This therefore represents a novel means of antigen recognition, in which the Fab fragment of an antibody acts as an adaptor, linking a human protein insert with antigen-binding potential to the constant antibody regions which mediate immune cell recruitment.

DOI: [10.7554/eLife.27311.001](https://doi.org/10.7554/eLife.27311.001)

*For correspondence: matthew.higgins@bioch.ox.ac.uk

Competing interests: The authors declare that no competing interests exist.

Funding: See page 9

Received: 29 March 2017

Accepted: 08 May 2017

Published: 20 May 2017

Reviewing editor: Pamela J Bjorkman, California Institute of Technology, United States

© Copyright Hsieh and Higgins. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

The antigen-binding sites of human antibodies commonly adopt similar structures, with the light and heavy chains each providing three hypervariable loops that combine to form a surface that is complementary to the epitope. While the sequences of these complementarity determining regions (CDRs) are highly variable, five of the six CDRs (L1, L2, L3, H1 and H2) can be classified into a number of relatively small sets, with similar lengths and architectures, and their structures are predictable from sequence (Chothia et al., 1989; North et al., 2011). In contrast, the third CDR loop of the heavy chain (CDR H3) is more structurally diverse, most likely due to its location close to the V(D)J recombination site (Weitzner et al., 2015). Human antibodies typically have CDR H3 lengths of 8–16 residues (Zemlin et al., 2008) while mouse antibodies have CDR H3 lengths of 5–26 residues (Zemlin et al., 2003).

However, recent years have seen the discovery of antibodies with major differences from the norm, in particular due to changes in the length of the third CDR of the heavy chain. A set of antibodies with broadly neutralizing potential against HIV is one such example. Here, the third CDR loop of the heavy chain is elongated, allowing it to reach through the glycan shield that surrounds the gp120 protein to bind an otherwise concealed epitope (McLellan et al., 2011; Pancera et al., 2013; Pejchal et al., 2010). Such antibodies are rare, making the induction of a broadly inhibitory response against HIV a major challenge (Corti and Lanzavecchia, 2013).

In a more extreme example, while the majority of bovine antibodies have CDR H3 loops of around 23 residues, around 10% contain a highly elongated third CDR loop of up to 69 residues, containing a small disulphide rich domain (Saini et al., 1999; Wang et al., 2013). These domains adopt a conserved β -sheet structure that displays variable loops and are each presented on an elongated, but rigid β -hairpin (Stanfield et al., 2016; Wang et al., 2013). While it is clear that the

eLife digest When bacteria, viruses or parasites invade the human body, the immune system responds by producing proteins called antibodies. Antibodies recognize and bind to molecules (known as antigens) on the surface of the invaders. This binding can either neutralize the invader directly or trigger signals that cause other parts of the immune system to destroy it.

Our blood contains a huge range of different antibody molecules that each bind to a different antigen. This is despite most human antibodies having the same basic shape and structure. Six loops, known as complementarity determining regions (CDRs), emerge from the surface of the antibody to form the surface that recognizes the antigen. However, variations in the structure of the loops alter this surface enough to allow different antibodies to recognize completely different molecules.

In 2016, a new class of antibodies was identified. Unlike previously identified antibodies, these molecules had an entire human protein, called LAIR1, inserted into one of their CDR loops. Members of this group of antibodies bind to a molecule, known as a RIFIN, that is found on the surface of human red blood cells that are infected with the parasite that causes malaria.

How do LAIR1-containing antibodies bind to their RIFIN targets? Hsieh and Higgins investigated this question by using a technique called X-ray crystallography to determine the structure of the antibody. This revealed that instead of binding directly to an antigen, all of the six CDR loops in the LAIR1-containing antibody bind to the LAIR1 insert. By doing so, LAIR1 is oriented in a manner that enables it to bind to the RIFIN molecule from the parasite.

This is the first known example of an antibody that recruits another protein to bind to an antigen rather than binding directly to the pathogen itself. A future challenge will be to see if other antibodies exist that use this mechanism and whether it can be employed to design new therapeutic antibodies.

DOI: [10.7554/eLife.27311.002](https://doi.org/10.7554/eLife.27311.002)

additional domains play an important role in ligand binding, the remaining five CDR loops are also exposed and further studies are needed to see the contribution that they make (Wang et al., 2013).

A recent study identified a group of even more unusual human antibodies in malaria endemic regions of Africa (Tan et al., 2016). These antibodies were discovered through their capacity to agglutinate human erythrocytes infected with different strains of *Plasmodium falciparum*, and they bind to a subset of RIFIN proteins. These RIFINs are displayed by the parasite on infected erythrocyte surfaces and are of uncertain function (Chan et al., 2014; Gardner et al., 1998; Kyes et al., 1999). The antibodies show a remarkable adaptation with an intact 96 residue protein, LAIR1, inserted into the third CDR loop of the antibody heavy chain. Indeed, LAIR1 was shown to be essential for the antibody to interact with RIFINs (Tan et al., 2016). In this study, we reveal the structure of the Fab fragment of one of these antibodies, showing how LAIR1 is presented on the antibody surface and drawing conclusions about how this class of antibody can recognize its ligand.

Results

We expressed the two chains that make up the Fab fragment of antibody MGD21 (Tan et al., 2016) in a secreted form from HEK293 cells. This antibody has a kappa light chain (VK1-8/JK5) and a heavy chain in which LAIR1 has been inserted into CDR H3. This fragment was purified and crystallised, allowing a dataset to be collected to 2.52 Å resolution. The structure was determined by molecular replacement using LAIR1 (Brondijk et al., 2010) and the Fab fragment of antibody OX117 (Nettleship et al., 2008) as search models. This identified two copies of the MGD21 Fab fragment in the asymmetric unit of the crystal. A model was built for residues 2–211 of the light chain and 1–351 (with 214–219 and 264–270 disordered) of the heavy chain (Figure 1, Figure 1—figure supplement 1, Figure 1—figure supplement 2, Table 1). The two Fab fragments adopt the same structure with a root mean square deviation of 0.26 Å (calculated over 475 Cα atoms) suggesting a highly ordered linkage between the variable domains of the antibody and the LAIR1 insert (Figure 1—figure supplement 3). The antibody sequence has three putative N-linked glycosylation sites, but of

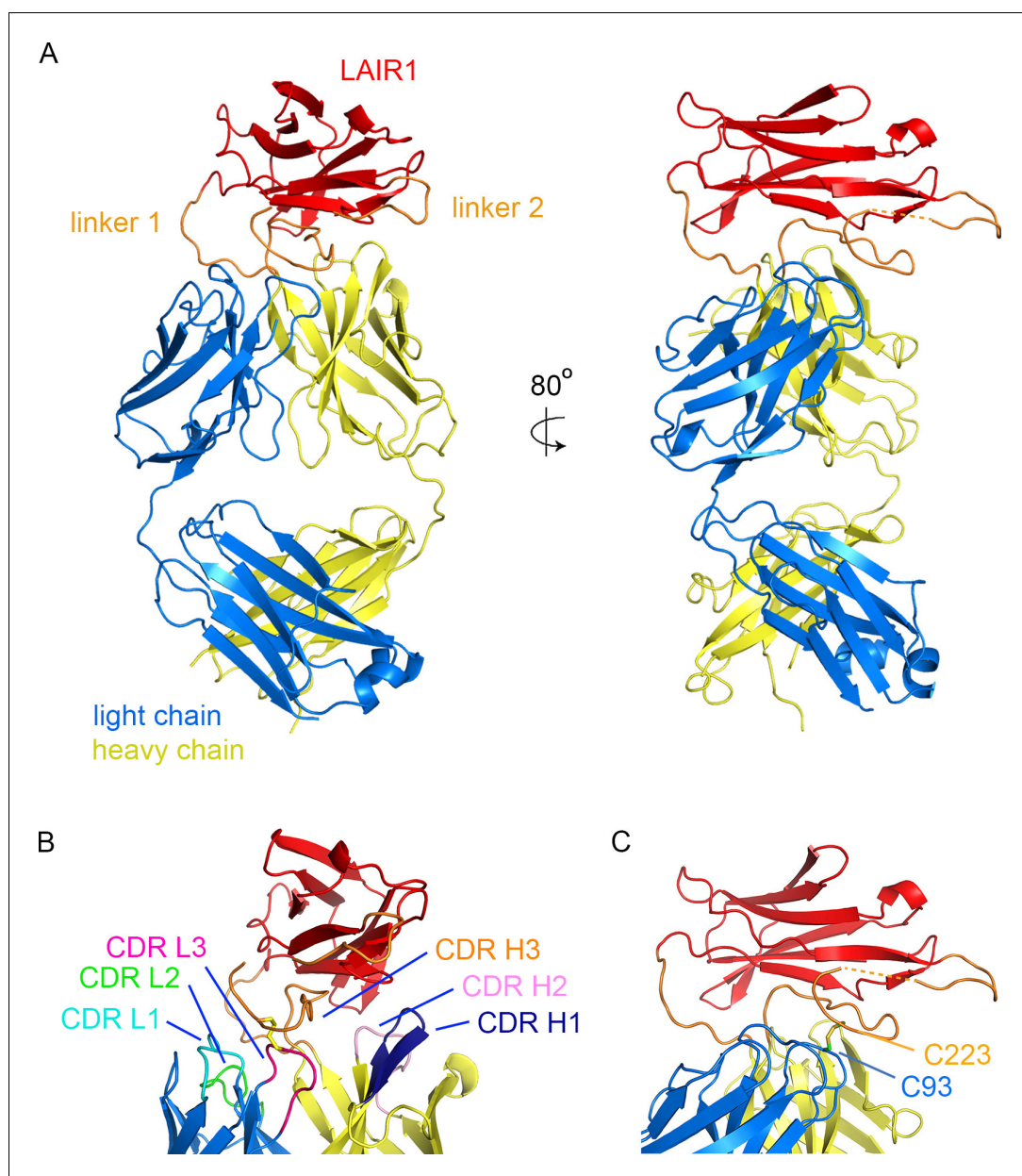


Figure 1. Structure of a LAIR1-containing antibody Fab fragment. (A) The structure of the Fab fragment. LAIR1 (red) is inserted into the third CDR loop of the heavy chain (yellow) through two extended linkers (orange). The light chain is blue. The dashed orange link represents protein disordered in the structure. (B) The organization of the CDRs. The three CDR loops of the light chain and remaining two CDR loops of the heavy chain directly contact the LAIR1 insert or the linkers. Each of the CDR loops and its corresponding label is shown in a different colour. (C) A disulphide bond between C93 of the light chain and C223 of the heavy chain stabilizes the interface (cysteine residues are shown as sticks).

DOI: [10.7554/eLife.27311.003](https://doi.org/10.7554/eLife.27311.003)

The following figure supplements are available for figure 1:

Figure supplement 1. Annotated sequence of antibody MGD21 and its alignment to germ line LAIR1.

DOI: [10.7554/eLife.27311.004](https://doi.org/10.7554/eLife.27311.004)

Figure supplement 2. Electron density.

DOI: [10.7554/eLife.27311.005](https://doi.org/10.7554/eLife.27311.005)

Figure supplement 3. Crystal packing and order.

DOI: [10.7554/eLife.27311.006](https://doi.org/10.7554/eLife.27311.006)

Table 1. Data collection and refinement statistics. The structure was determined from a single crystal. Values in parentheses are for highest-resolution shell. R_{free} was determined using 1968 reflections (4.8%) The structure is deposited with pdb code 5NST.

	Fab-MGD21
Data collection	
Space group	C121
Cell dimensions	
a, b, c (Å)	169.8, 86.5, 104.0
$\alpha \beta \gamma$ (°)	90.0, 126.7, 90.0
Wavelength	0.92819
Resolution (Å)	81.90–2.52 (2.56–2.52)
Total Observations	131833 (5451)
Total Unique	40946 (2031)
R_{pim} (%)	5.4 (67.8)
R_{merge} (%)	8.3 (88.5)
R_{meas} (%)	9.9 (112.1)
$CC_{1/2}$	0.992 (0.571)
$I/\sigma(I)$	7.4 (1.0)
Completeness (%)	99.8 (98.3)
Multiplicity	3.2 (2.7)
Wilson B factor	55.216
Refinement	
Number of reflections	40946
$R_{\text{work}} / R_{\text{free}}$	21.9/26.7
Number of residues	
Protein	1076
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.25
All Atom clash score	5
B factors	
All atoms	71.53
Solvent	63.12
Variable domains	65.17
Constant domains	74.29
LAIR1 insert	73.70
Linkers	94.71
Ramachandran plot	
Favored (%)	95.2%
Allowed (%)	4.8%
Disallowed (%)	0.0%

DOI: [10.7554/eLife.27311.007](https://doi.org/10.7554/eLife.27311.007)

these (light chain N30; heavy chain N61 and N242) only N242 shows electron density corresponding to an Asn-linked N-acetyl glucosamine, in a position distant from the LAIR1 insert.

The structure shows LAIR1 emerging from the CDR3 loop of the heavy chain and lying across the antigen-binding surface of the variable domains of the Fab fragment (**Figure 1A**). The long axis of the LAIR1 insert is positioned with the β -strands aligned approximately perpendicular to the groove between the heavy and light chain CDRs and the insertion and linkers interact with, and occlude all five of the remaining CDR loops. The N- and C-termini of LAIR1 lie at opposite ends of its structure, necessitating long linkers between the sites from which CDR3 emerges from the antibody heavy chain and each terminus of the LAIR1 insert (**Figure 1A**). The N-terminal linker (linker 1) is 10 residues long and adopts a simple loop structure that joins the antibody variable domain to the N-terminus of the LAIR1 insert. The C-terminal linker (linker 2) is longer at 34 residues and is more complex in structure. It extends out from the C-terminus of the LAIR1 insert before zigzagging back towards the insertion site in the heavy chain variable domain. It is stabilized by hydrogen bonds to the LAIR1 insert and to the antibody heavy chain as well as by a disulphide bond to C93 of the antibody light chain. The linkers of the LAIR1-containing antibodies sequenced to date are variable both in length and content, involving different parts of the intronic regions of the LAIR1 gene, or intergenic sequences of chromosome 13 (**Tan et al., 2016**). The arrangement of these linkers, which radiate away from the remainder of the antibody, will in theory accommodate almost limitless variation in both length and sequence without disturbing the packing of LAIR1 against the variable domains of the antibody.

The five CDR loops lacking the LAIR1 insertion are representatives of previously identified canonical classes (**Figure 1—figure supplement 2**) (**Martin and Thornton, 1996**). However, a search using the Abcheck server (**Martin, 1996**) identified seven unusual residues within the antibody structure; C91, C93, D97 and I106 from the light chain and Y28, R34 and Q54 from the heavy chain, all within the CDR loops. In particular, C91, C93 and D97 all lie in CDR3 of the light chain, perhaps facilitating its interaction with linker 2. Indeed, the most unusual residue is C93, which is found in only 0.096% of light chains, and is the residue that forms a disulphide bond with linker 2 (**Figure 1B**). The heavy chain CDR H3 loop has a base that adopts the ‘kinked’ conformation (**Shirai et al., 1999**), with the loop rapidly spreading to the two termini of LAIR1.

In previous structures of antibodies with extended heavy chain CDR3 loops, the remaining five CDRs of the antibody are exposed, with the potential to engage in antigen binding (**McLellan et al., 2011; Stanfield et al., 2016, Wang et al., 2013**). One of the remarkable features of the LAIR1-containing antibody is therefore the occlusion of large parts of each of the remaining five CDRs, with these loops each interacting directly with the LAIR1 insert and/or linkers (**Figure 1B, Table 2**). The degree of occlusion of the CDRs by LAIR1 was determined by accessibility to a 1.4 Å probe in the presence and absence of LAIR1 and the linkers. Each of these five CDR loops was partly occluded by the presence of LAIR1 or the linkers (occluding 12.7% of the accessible surface area of CDR L1, 18.3% of CDR L2, 47.3% of CDR L3, 34.7% of CDR H1 and 16.0% of CDR H2). Indeed, both the first and second CDRs of the heavy chain directly contact the LAIR1 insert (**Table 2**). In addition, each of the three CDR loops of the light chain interacts with one of the two linkers, with interactions

Table 2. A list of interactions between the LAIR1 insert and linkers that occupies the heavy chain CDR3 loop and the other five CDR loops of the antibody.

CDR loop	Residue	Group	LAIR1 region	Residue	Group	Interaction
Light chain CDR1	Q27	Side chain	Linker 2	A222	Main Chain	Hydrogen Bond
Light chain CDR2	Y49	Side chain	Linker 1	L102	Side chain	Hydrophobic Packing
Light chain CDR2	N53	Side chain	Linker 1	S104	Side chain	Hydrogen Bond
Light chain CDR3	C93	Side chain	Linker 2	C223	Side chain	Disulphide Bond
Light chain CDR3	F94	Main Chain	Linker 2	E227	Side Chain	Hydrogen Bond
Heavy chain CDR1	N32	Side chain	LAIR1	R134	Side chain	Hydrogen Bond
Heavy chain CDR2	R57	Side Chain	LAIR1	P109	Main Chain	Hydrogen Bond

DOI: 10.7554/eLife.27311.008

including a disulphide bond between C93 of the light chain and C223 of linker 2 (**Figure 1C**, **Table 2**). These interactions are replicated in both copies of the molecule in the asymmetric unit of the crystal.

The structure of MGD21 argues for a rigid association of the LAIR1 insert with the remainder of the antibody. Firstly, the structures of the two molecules of the antibody in the asymmetric unit of the crystal superimpose closely (**Figure 1—figure supplement 2**). It is unlikely that this is due solely to constraints from crystal packing as LAIR1 is anchored to the variable domains of the antibody through three fixed positions: the attachment sites of the two linkers, and the disulphide bond between light chain C93 and heavy chain C223 (**Figure 1C**). In addition, each of the five CDR loops not bearing a LAIR1 insertion makes direct interactions with either LAIR1 or the linker, through contacts found in both copies of the antibody in the asymmetric unit of the crystal. This will stabilize a tight association between LAIR1 and the antibody variable domains. As these antibodies can include multiple different light chains, and very different linkers (**Tan et al., 2016**), these interaction will not be replicated precisely across the antibody family, but some variant of interaction between light chain CDRs and the linkers is likely.

A comparison of the LAIR1 insert with that of the chromosomal copy of LAIR1 (referred to below as germ line) (**Brondijk et al., 2010**) reveals that no global structural changes have taken place (root mean square deviation 0.43 Å for the 82 Cα residues) (**Figure 2A,F**). Indeed, the LAIR1 insertion in the MGD21 antibody differs in only 13 positions relative to the germ line sequence. Mapping these sites onto the structure reveals that they do not alter residues through which the LAIR1 insert interacts with the rest of the antibody (**Figure 2B**). Presumably, the stable interaction between LAIR1 and the antibody has therefore come instead from adaptations to the CDR loops. In contrast, polymorphisms are mostly located on the surface of LAIR1 distal to the rest of the antibody and have the potential to alter its interaction with its original ligand, collagen, and with the RIFIN proteins which are the target of these antibodies.

The normal function of LAIR1 is to interact with collagen (**Meyaard, 2008**). The structure of germ line LAIR1, together with NMR analysis and mutagenesis, allowed the mapping of residues critical for the collagen interaction onto a LAIR1 crystal structure (**Brondijk et al., 2010**). In particular, mutations in residues R59, E61 and R65 have a significant impact on collagen binding (**Brondijk et al., 2010**). These residues map onto the surface of LAIR1 (**Figure 2C**) that is most exposed in the context of the antibody (**Figure 2D**). Indeed, mapping of the polymorphisms found in the 27 LAIR1-containing antibodies sequenced to date shows that large parts of this surface are mutable (**Figure 2E**). The polymorphisms in LAIR1 include R149N, which is in the position equivalent to R65 in germ line LAIR1 and this change may impact collagen binding. A second polymorphism, found in 7/27 of the antibodies (including MGD21) alters the N-linked glycosylation site at residue 69 of LAIR1 (**Wollscheid et al., 2009**), which may alter collagen binding and/or increase RIFIN binding, but is not conserved across the antibody family. Indeed 11 of the 27 sequenced antibodies have mutations in at least one of the residues implicated in collagen binding, or other polymorphisms that reduce the interaction (**Tan et al., 2016**).

Discussion

The LAIR1-containing antibodies are a remarkable variant of the standard immunoglobulin fold. While the majority of mammalian antibodies have predictable and short CDRs, the third CDR of the heavy chain can accommodate usual diversity (**Figure 3**) (**McLellan et al., 2011; Wang et al., 2013; Weitzner et al., 2015**). This is seen in the elongated CDR3 of the broadly neutralizing antibodies that interact with HIV surface proteins and in the insertion of a β-hairpin and disulphide-rich domain in a fraction of bovine antibodies. However, in both of these cases, only the heavy chain CDR3 is altered and the remaining CDR loops remain exposed for antigen binding. The LAIR1-containing antibodies are an exception to this, with the LAIR1-insert interacting with, and partly occluding, all five of the remaining CDR loops. In many ways, the structure resembles an antibody with CDR loops adapted for LAIR1 binding, into which LAIR1 has also been inserted.

This occlusion of large parts of the CDR loops by the LAIR1 insert has major consequences for its role in antigen recognition, as the majority of the antigen-binding surface will be contributed by LAIR1. Indeed, it has been shown that the LAIR1 insert alone can bind to infected erythrocytes, as can a LAIR1-containing antibody with the heavy and light chain regions exchanged (**Tan et al.,**

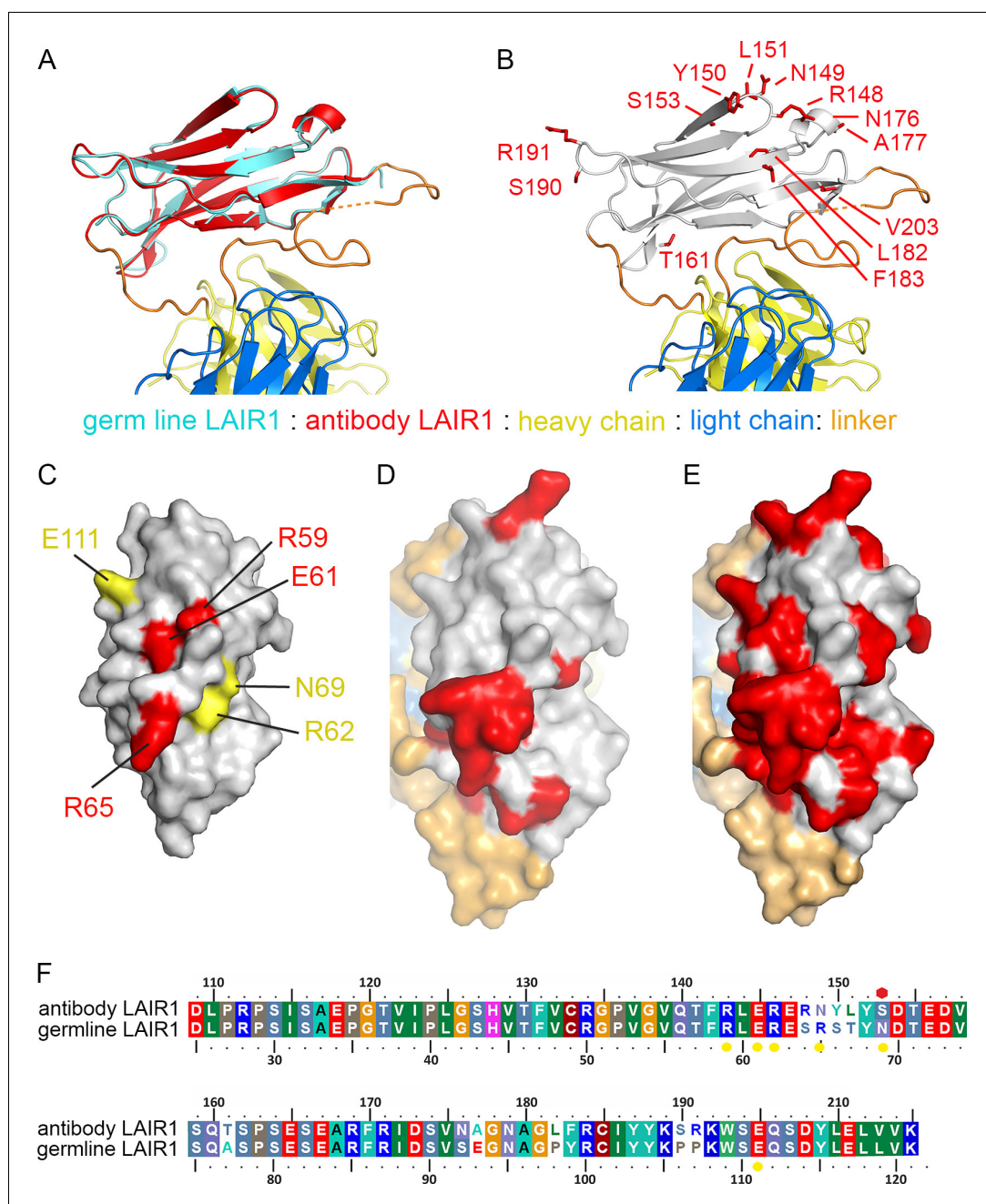


Figure 2. Structure and polymorphism in the LAIR1 insertion. (A) An alignment of germ line LAIR1 (cyan) with the antibody LAIR1 insertion (red). (B) The residues that differ between the LAIR1 insertion in antibody MGD21 and germ line LAIR1 are shown as red sticks. (C) A surface representation of the structure of LAIR1 (grey) with residues whose mutation has a major (red) or minor (yellow) effect on collagen binding highlighted (Brondijk et al., 2010). (D) A surface view of the LAIR1 insert in antibody MGD21 (grey) with residues that differ from germ line LAIR1 highlighted (red). (E) A surface view of the LAIR1 insert (grey) with residues that differ from germ line LAIR1 in all 27 antibodies tested to date (Tan et al., 2016) highlighted (red). (F) A sequence alignment of germ line LAIR1 and the LAIR1 insert in the MGD21 antibody. Yellow circles are sites residues shown to play a role in collagen binding while a red hexagon represents a potential N-linked glycosylation site mutated in the LAIR1 insert.

DOI: [10.7554/eLife.27311.009](https://doi.org/10.7554/eLife.27311.009)

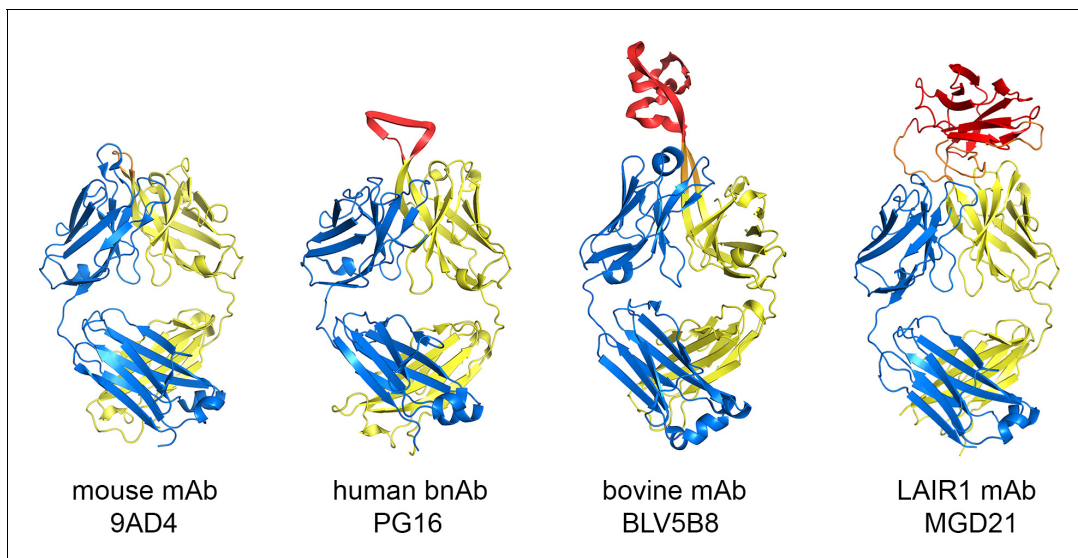


Figure 3. Comparison of the LAIR1-containing antibody with other unusual antibodies. The structure of the LAIR1-containing monoclonal antibody is compared with a classical mouse monoclonal antibody (9AD4; PDB code 4U0R), a human monoclonal antibody with broadly neutralizing potential against HIV (PG16; PDB code 4DQ0) and a bovine monoclonal antibody (BLV5B8; PDB code 4K3E). In each case, the light chain is blue and the two immunoglobulin domains of the heavy chain are yellow. Inserted domains are shown in red with linker regions in orange.

DOI: [10.7554/eLife.27311.010](https://doi.org/10.7554/eLife.27311.010)

2016). Surprisingly an antibody in which the LAIR1 insert has been exchanged for the unaltered germ line LAIR1 did not bind to erythrocytes, although the folding of this chimera was not tested (**Tan et al., 2016**). In addition, the capacity of RIFINs to bind to unaltered LAIR1 alone has not yet been reported. Indeed, it seems most likely that LAIR1, or a highly related homologue, is the physiological ligand of the group of RIFINs that are recognized by these antibodies and that its insertion into the Fab fragment of an antibody allows it to be affinity matured to mobilise it for immune recognition and recruitment of immune cells. This remarkable LAIR1-containing antibody therefore uses the classical hypervariable loops for a novel function: to position an inserted auxillary domain for antigen recognition. The classical Fab fragment therefore now acts as a link between a ligand for a pathogen surface receptor and the Fc region of the antibody with its immune recruitment capability. It will be fascinating to see if this is a paradigm that is repeated in other novel antibodies, as yet undiscovered.

Materials and methods

Construction, protein expression and purification

Two synthetic complementary DNA clones based on MGD21 (**Tan et al., 2016**) were obtained from GeneArt (ThermoFisher, UK). The heavy chain variable region was amplified using primers VH-F: 5'-GATGGGTTGCGTAGCTGAAGTGCAGCTGGTGGAAACAGGC-3' and VH-R: 5'-GGGTGTCGTTTGGCGCTAGACACTGTCACGGTGGTGCC-3'. The light chain variable region was amplified using primers VL-F: 5'-GATGGGTTGCGTAGCTGCCATCAGAATGACCCAGAGCCCC-3' and VL-R: 5'-GTGCAGCATCAGCCCGCTTGATTCCAGCCGGGTGCCC-3'. The resulting PCR products were cloned into pOPIN VH (heavy chain variable region) and pOPIN VL (light chain variable region) by In-Fusion cloning (Clontech, Mountain View, CA) (**Nettleship et al., 2008**). Therefore the variable domains from MGD21 were fused to the constant domains derived from the pOPIN VH and pOPIN VL vectors.

DNA constructs expressing heavy and light chains were mixed into a 1 to 1 ratio and co-transfected in HEK293T cells (ThermoFisher Scientific, UK) with polyethyleneimine in the presence of 5 μ M kifunensine (**Aricescu et al., 2006**). After five days, conditioned media was dialysed against phosphate-buffered saline and purified by immobilised metal ion affinity chromatography using

TALON resin (Clontech, Mountain View, CA). The Fab heterodimer was further purified by size-exclusion chromatography using a Superdex 200 16/600 column (GE Healthcare Life Sciences) in 10 mM HEPES, pH 7.5 and 150 mM NaCl.

Crystallisation, data collection and structure determination

Concentrated protein (10 mg/ml) was incubated with *Flavobacterium meningosepticum* endoglycosidase-F1 for in situ deglycosylation (Hsieh et al., 2016). The protein samples were then subjected to sitting drop vapour diffusion crystallisation trials in SwisSci 96-well plates by mixing 100 nl protein with 100 nl reservoir solution. The protein crystals were obtained in 20% (w/v) PEG4000, 0.1 M sodium citrate, pH 4.5 at 18°C. Crystals were transferred into mother liquor containing 25% (w/v) glycerol and were then cryo-cooled in liquid nitrogen for storage and data collection. Data were collected on beamline I04-1 at the Diamond Light Source and were indexed and scaled using XDS (Kabsch, 2010). Phaser (McCoy et al., 2007) was used to determine a molecular replacement model, using the known structures of LAIR1 (pdb: 3KGR (Brondijk et al., 2010)) and a human monoclonal antibody Fab fragment similar to MGD21 (pdb: 3DIF, (Nettleship et al., 2008)) separated into two files containing the variable and the constant regions, as search models. This identified two copies of the LAIR1-containing Fab fragment in the asymmetric unit of the antibody. Refinement and rebuilding was completed using Buster (Blanc et al., 2004) and Coot (Emsley et al., 2010) respectively. To determine the effect of the LAIR1 insert on the accessible surface area of the CDR loops, we used AREAIMOL from the CCP4 suite (Winn et al., 2011) to determine the accessible surface area of each CDR loop both in the presence and absence of LAIR1 and the linkers.

Acknowledgements

We thank Dr. Ray Owens at the Oxford Protein Production Facility for the provision of vectors pOPINVH and pOPINVL and the beamline scientists at Diamond Light Source for assistance with data collection.

Additional information

Funding

Funder	Grant reference number	Author
Wellcome	101020/Z/13/Z	Fu-Lien Hsieh Matthew K Higgins
Taiwan Bio-Development Foundation		Fu-Lien Hsieh

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

F-LH, Conceptualization, Formal analysis, Investigation, Writing—original draft; MKH, Conceptualization, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing—original draft, Project administration

Author ORCIDs

Matthew K Higgins,  <http://orcid.org/0000-0002-2870-1955>

References

- Aricescu AR, Lu W, Jones EY. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D Biological Crystallography* **62**:1243–1250. doi: [10.1107/S0907444906029799](https://doi.org/10.1107/S0907444906029799), PMID: [17001101](https://pubmed.ncbi.nlm.nih.gov/17001101/)
- Blanc E, Roversi P, Vornrhein C, Flensburg C, Lea SM, Bricogne G. 2004. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D Biological Crystallography* **60**:2210–2221. doi: [10.1107/S0907444904016427](https://doi.org/10.1107/S0907444904016427), PMID: [15572774](https://pubmed.ncbi.nlm.nih.gov/15572774/)

- Brondijk TH**, de Ruiter T, Ballering J, Wienk H, Lebbink RJ, van Ingen H, Boelens R, Farndale RW, Meyaard L, Huizinga EG. 2010. Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood* **115**:1364–1373. doi: [10.1182/blood-2009-10-246322](https://doi.org/10.1182/blood-2009-10-246322), PMID: [20007810](https://pubmed.ncbi.nlm.nih.gov/20007810/)
- Chan JA**, Fowkes FJ, Beeson JG. 2014. Surface antigens of plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. *Cellular and Molecular Life Sciences* **71**:3633–3657. doi: [10.1007/s00018-014-1614-3](https://doi.org/10.1007/s00018-014-1614-3), PMID: [24691798](https://pubmed.ncbi.nlm.nih.gov/24691798/)
- Chothia C**, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR. 1989. Conformations of immunoglobulin hypervariable regions. *Nature* **342**:877–883. doi: [10.1038/342877a0](https://doi.org/10.1038/342877a0), PMID: [2687698](https://pubmed.ncbi.nlm.nih.gov/2687698/)
- Corti D**, Lanzavecchia A. 2013. Broadly neutralizing antiviral antibodies. *Annual Review of Immunology* **31**:705–742. doi: [10.1146/annurev-immunol-032712-095916](https://doi.org/10.1146/annurev-immunol-032712-095916), PMID: [23330954](https://pubmed.ncbi.nlm.nih.gov/23330954/)
- Dunbar J**, Deane CM. 2016. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**: 298–300. doi: [10.1093/bioinformatics/btv552](https://doi.org/10.1093/bioinformatics/btv552), PMID: [26424857](https://pubmed.ncbi.nlm.nih.gov/26424857/)
- Emsley P**, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. *Acta Crystallographica Section D Biological Crystallography* **66**:486–501. doi: [10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493), PMID: [20383002](https://pubmed.ncbi.nlm.nih.gov/20383002/)
- Gardner MJ**, Tettelin H, Carucci DJ, Cummings LM, Aravind L, Koonin EV, Shallom S, Mason T, Yu K, Fujii C, Pederson J, Shen K, Jing J, Aston C, Lai Z, Schwartz DC, Perteau M, Salzberg S, Zhou L, Sutton GG, et al. 1998. Chromosome 2 sequence of the human malaria parasite plasmodium falciparum. *Science* **282**:1126–1132. doi: [10.1126/science.282.5391.1126](https://doi.org/10.1126/science.282.5391.1126), PMID: [9804551](https://pubmed.ncbi.nlm.nih.gov/9804551/)
- Hsieh FL**, Turner L, Bolla JR, Robinson CV, Lavstsen T, Higgins MK. 2016. The structural basis for CD36 binding by the malaria parasite. *Nature Communications* **7**:12837. doi: [10.1038/ncomms12837](https://doi.org/10.1038/ncomms12837), PMID: [27667267](https://pubmed.ncbi.nlm.nih.gov/27667267/)
- Kabsch W**. 2010. XDS. *Acta Crystallographica. Section D, Biological Crystallography* **66**:125–132. doi: [10.1107/S0907444909047337](https://doi.org/10.1107/S0907444909047337), PMID: [20124692](https://pubmed.ncbi.nlm.nih.gov/20124692/)
- Kyes SA**, Rowe JA, Kriek N, Newbold CI. 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with plasmodium falciparum. *PNAS* **96**:9333–9338. doi: [10.1073/pnas.96.16.9333](https://doi.org/10.1073/pnas.96.16.9333), PMID: [10430943](https://pubmed.ncbi.nlm.nih.gov/10430943/)
- Martin AC**. 1996. Accessing the kabat antibody sequence database by computer. *Proteins: Structure, Function, and Genetics* **25**:130–133. doi: [10.1002/\(SICI\)1097-0134\(199605\)25:1<130::AID-PROT11>3.3.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0134(199605)25:1<130::AID-PROT11>3.3.CO;2-Y), PMID: [8727325](https://pubmed.ncbi.nlm.nih.gov/8727325/)
- Martin AC**, Thornton JM. 1996. Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. *Journal of Molecular Biology* **263**:800–815. doi: [10.1006/jmbi.1996.0617](https://doi.org/10.1006/jmbi.1996.0617), PMID: [8947577](https://pubmed.ncbi.nlm.nih.gov/8947577/)
- McCoy AJ**, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674. doi: [10.1107/S0021889807021206](https://doi.org/10.1107/S0021889807021206), PMID: [19461840](https://pubmed.ncbi.nlm.nih.gov/19461840/)
- McLellan JS**, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, et al. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**:336–343. doi: [10.1038/nature10696](https://doi.org/10.1038/nature10696), PMID: [22113616](https://pubmed.ncbi.nlm.nih.gov/22113616/)
- Meyaard L**. 2008. The inhibitory collagen receptor LAIR-1 (CD305). *Journal of Leukocyte Biology* **83**:799–803. doi: [10.1189/jlb.0907609](https://doi.org/10.1189/jlb.0907609), PMID: [18063695](https://pubmed.ncbi.nlm.nih.gov/18063695/)
- Nettleship JE**, Ren J, Rahman N, Berrow NS, Hatherley D, Barclay AN, Owens RJ. 2008. A pipeline for the production of antibody fragments for structural studies using transient expression in HEK 293T cells. *Protein Expression and Purification* **62**:83–89. doi: [10.1016/j.pep.2008.06.017](https://doi.org/10.1016/j.pep.2008.06.017), PMID: [18662785](https://pubmed.ncbi.nlm.nih.gov/18662785/)
- North B**, Lehmann A, Dunbrack RL. 2011. A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology* **406**:228–256. doi: [10.1016/j.jmb.2010.10.030](https://doi.org/10.1016/j.jmb.2010.10.030), PMID: [21035459](https://pubmed.ncbi.nlm.nih.gov/21035459/)
- Pancera M**, Shahzad-ul-Hussan S, Doria-Rose NA, McLellan JS, Bailer RT, Dai K, Loesgen S, Louder MK, Staupe RP, Yang Y, Zhang B, Parks R, Eudailey J, Lloyd KE, Blinn J, Alam SM, Haynes BF, Amin MN, Wang LX, Burton DR, et al. 2013. Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nature Structural & Molecular Biology* **20**:804–813. doi: [10.1038/nsmb.2600](https://doi.org/10.1038/nsmb.2600), PMID: [23708607](https://pubmed.ncbi.nlm.nih.gov/23708607/)
- Pejchal R**, Walker LM, Stanfield RL, Phogat SK, Koff WC, Pognard P, Burton DR, Wilson IA. 2010. Structure and function of broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent neutralization of HIV-1. *PNAS* **107**:11483–11488. doi: [10.1073/pnas.1004600107](https://doi.org/10.1073/pnas.1004600107), PMID: [20534513](https://pubmed.ncbi.nlm.nih.gov/20534513/)
- Saini SS**, Allore B, Jacobs RM, Kaushik A. 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *European Journal of Immunology* **29**:2420–2426. doi: [10.1002/\(SICI\)1521-4141\(199908\)29:08<2420::AID-IMMU2420>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A), PMID: [10458755](https://pubmed.ncbi.nlm.nih.gov/10458755/)
- Shirai H**, Kidera A, Nakamura H. 1999. H3-rules: identification of CDR-H3 structures in antibodies. *FEBS Letters* **455**:188–197. doi: [10.1016/S0014-5793\(99\)00821-2](https://doi.org/10.1016/S0014-5793(99)00821-2), PMID: [10428499](https://pubmed.ncbi.nlm.nih.gov/10428499/)
- Stanfield RL**, Wilson IA, Smider VV. 2016. Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Science Immunology* **1**:aaf7962. doi: [10.1126/sciimmunol.aaf7962](https://doi.org/10.1126/sciimmunol.aaf7962), PMID: [27574710](https://pubmed.ncbi.nlm.nih.gov/27574710/)
- Tan J**, Pieper K, Piccoli L, Abdi A, Foglierini M, Geiger R, Tully CM, Jarrossay D, Ndungu FM, Wambua J, Bejon P, Fregni CS, Fernandez-Rodriguez B, Barbieri S, Bianchi S, Marsh K, Thathy V, Corti D, Sallusto F, Bull P, et al. 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* **529**: 105–109. doi: [10.1038/nature16450](https://doi.org/10.1038/nature16450), PMID: [26700814](https://pubmed.ncbi.nlm.nih.gov/26700814/)

- Wang F**, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, Torkamani A, Raudsepp T, Mwangi W, Criscitiello MF, Wilson IA, Schultz PG, Smider VV. 2013. Reshaping antibody diversity. *Cell* **153**:1379–1393. doi: [10.1016/j.cell.2013.04.049](https://doi.org/10.1016/j.cell.2013.04.049), PMID: [23746848](https://pubmed.ncbi.nlm.nih.gov/23746848/)
- Weitzner BD**, Dunbrack RL, Gray JJ. 2015. The origin of CDR H3 structural diversity. *Structure* **23**:302–311. doi: [10.1016/j.str.2014.11.010](https://doi.org/10.1016/j.str.2014.11.010), PMID: [25579815](https://pubmed.ncbi.nlm.nih.gov/25579815/)
- Winn MD**, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**:235–242. doi: [10.1107/S0907444910045749](https://doi.org/10.1107/S0907444910045749), PMID: [21460441](https://pubmed.ncbi.nlm.nih.gov/21460441/)
- Wollscheid B**, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, Schiess R, Aebersold R, Watts JD. 2009. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology* **27**:378–386. doi: [10.1038/nbt.1532](https://doi.org/10.1038/nbt.1532), PMID: [19349973](https://pubmed.ncbi.nlm.nih.gov/19349973/)
- Zemlin M**, Klinger M, Link J, Zemlin C, Bauer K, Engler JA, Schroeder HW, Kirkham PM. 2003. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *Journal of Molecular Biology* **334**:733–749. doi: [10.1016/j.jmb.2003.10.007](https://doi.org/10.1016/j.jmb.2003.10.007), PMID: [14636599](https://pubmed.ncbi.nlm.nih.gov/14636599/)
- Zemlin M**, Schelonka RL, Ippolito GC, Zemlin C, Zhuang Y, Gartland GL, Nitschke L, Pelkonen J, Rajewsky K, Schroeder HW. 2008. Regulation of repertoire development through genetic control of DH reading frame preference. *The Journal of Immunology* **181**:8416–8424. doi: [10.4049/jimmunol.181.12.8416](https://doi.org/10.4049/jimmunol.181.12.8416), PMID: [19050259](https://pubmed.ncbi.nlm.nih.gov/19050259/)