Structural characterisation of the interaction between *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) and their human ligands in severe malaria

Hong Kin Ng

This thesis is submitted for a degree of
Masters in Science by research



Merton College

August 2012

Declaration

The work described in this dissertation was performed in the Department of Biochemistry, University of Oxford, between October 2011 and August 2012. I declare that the work described in this dissertation is the result of my own work and includes nothing which is the outcome of the work done in collaboration unless specifically indicated in the text. No part of this dissertation has already been or is concurrently being submitted to any other degree, diploma or qualification at other academic institutes. The length of this dissertation does not exceed the word limits.

Hong Kin Ng

University of Oxford

August 2012

Abstract

Structural characterisation of the interaction between *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) and their human ligands in severe malaria

Sequestration of infected erythrocytes during *P. falciparum* infection can lead to severe malaria. PfEMP1 proteins are central to this process. Multimodular PfEMP1 proteins can bind to a variety of human ligands due to the presence of CIDR and DBL domains. CIDR and DBL domains are classified to three and six families respectively, with each thought to bind to unique human ligands. The interaction between PfEMP1 proteins and human ligands protects the infected erythrocytes against spleen-mediated destruction. Additionally, these PfEMP1 proteins undergo antigenic variation to evade immune surveillance. They also elicit immune suppression by binding to immunoglobulins within the circulatory system.

The work described here focuses on two interactions between PfEMP1 on the infected erythrocytes and their human ligands, the first involving CD36 and CIDR1α domains and the second involving human IgM and DBL domains. I have expressed and purified recombinant proteins involved in both interactions.

In this thesis, I have confirmed using a variety of biophysical methods that CIDR1α domains are able to bind CD36 in a 1:1 binding stoichiometry, and this accounts for the binding affinity of the full length protein. This interaction was not disrupted after performing deglycosylation procedures on CD36, allowing me to obtain stable complexes of these interacting partners for structural characterisation by crystallography. I have also confirmed the interaction between DBL domains and human IgM Fc region using similar biophysical methods. These experiments are important for paving the way towards further structural characterisation of these interactions that leads to sequestration.

Hong Kin Ng
Thesis submitted for the degree of Masters in Science by research
Merton College
Trinity 2012

Contents

Declaration i
Abstract ii
Contents iii
List of Figures vi
List of Tables viii
List of Abbreviations ix
Acknowledgements xi

1. Introduction 1
1.1 Introduction to Malaria 1
1.1.1. Plasmodium parasite: Cause of malaria *1*
1.1.2. Evolution and genomics of Plasmodium *2*
1.1.3. Life cycle of human-adapted Plasmodium parasites *2*
1.2 Plasmodium falciparum and severe malaria 5
1.2.1. Invasive properties of Plasmodium falciparum *5*
1.2.2. PfEMP1 and its role in antigenic variation and immune suppression *7*
1.3 Sequestration and PfEMP1 – the link between P. falciparum and severe malaria 10
1.3.1. Molecular mechanism of sequestration *10*
1.3.2. Var genes *11*
1.3.3. PfEMP1 proteins and domain classification *14*
1.3.4. The domains involved in cytoadherence in PfEMP1 proteins: DBL and CIDR domains. *17*
1.3.5. Higher order arrangement of PfEMP1 *25*
1.3.6. Implications of domain and var gene classification in studies of malaria *28*
1.4 General aims 30

2. Materials and methods **31**
2.1 Gene cloning 31
2.1.1 Polymerase Chain Reaction *31*
2.1.2 Restriction Enzyme Digests *31*
2.1.3 DNA Ligations *32*
2.1.4 Transformation of DH5α cells with Plasmid DNA *32*
2.1.5 Plasmid Isolation *32*
2.1.6 Gel Electrophoresis *33*
2.1.7 DNA Sequencing *33*
2.1.8 DNA plasmids for Escherichia Coli expression *33*
2.2 Protein expression 35
2.2.1 Bacterial expression in Escherichia coli cells *35*
2.2.2 Baculovirus-mediated insect cell expression *37*
2.2.3 Mammalian cell expression system *39*
2.3 Protein purification 41
2.3.1 Sonication of bacterial cells *41*

2.3.2 Buffer exchange of cell supernatant 42
2.3.3 Dialysis of cell supernatant 42
2.3.4 Buffer exchange using PD-10 Desalting columns 43
2.3.5 Ni-NTA affinity chromatography for His_6-tag soluble proteins 43
2.3.6 IgM affinity chromatography 43
2.3.7 Refolding on Ni-NTA affinity column 44
2.3.8 TEV cleavage 45
2.3.9 Deglycosylation of glycoproteins from baculovirus-infected insect cell expression 45
2.3.10 Preparative size-exclusion chromatography 45
2.4 Protein detection and characterisation techniques 46
2.4.1 SDS Polyacrylamide gel electrophoresis (SDS-PAGE) 46
2.4.2 Detection using Western blotting 46
2.4.3 Circular dichroism 47
2.4.4 Fluorescence-based thermal shift assays 48
2.5 Biophysical methods for interaction studies 49
2.5.1 Ni-NTA Pull down assay 49
2.5.2 Surface Plasmon resonance 49
2.5.3 Analytical size-exclusion chromatography 52
2.5.4 Analytical ultracentrifugation 52
2.6 Co-crystallisation of hCD36ED-MC179 complex 54
2.6.1 Preparation of hCD36ED-MC179 complex. 54
2.6.2 Crystallisation trials 54

3. Biophysical characterisation of the interaction between human CD36 and PfEMP1 proteins 55
3.1. CD36 and severe malaria 55
3.1.1. Properties and features of human CD36 56
3.1.2. The pathogenic role of CD36 in malaria 59
3.1.3. Project aims 64
3.2. Characterisation of the interaction between CD36 and CIDR1α domains 66
3.2.1. Expression and purification of CD36 using Baculovirus expression system 66
3.2.2. Cloning, expression and purification of recombinant PfEMP1 CIDR1α domains using E coli expression system 71
3.2.3. Do CIDR1α-M2 domains interact with hCD36ED? 78
3.2.4. What is the stoichiometry of the complex formed between CIDR1α-M2 domains and hCD36ED? 83
3.3. What is the minimum CD36-binding region of a PfEMP1? 95
3.3.1. Expression and purification of the PF11_0007 series 95
3.3.2. Does the context of CIDR1α affect the interaction between hCD36ED and PfEMP1? 100
3.4. CD36 deglycosylation and affirmation of its interaction with CIDR1α domain 105
3.4.1. CD36 deglycosylation 105
3.4.2. Biophysical characterisation of the function and structure of deglycosylated hCD36ED 111
3.5. Crystallisation trials 118
3.5.1. Can lipid molecules block CIDR1α interaction with CD36? 118

3.5.2. Fluorescence-based thermal shift stability assay 119
3.5.3. Crystallisation screens 125
3.6. Consolidation of binding interaction studies between hCD36ED and PfEMP1 127

4. Biophysical characterisation of the interaction between IgM and PfEMP1 129
4.1. IgM and severe malaria 129
4.1.1. Human IgM and its role in immune surveillance 129
4.1.2. The pathogenic role of IgM in malaria 134
4.1.3. Project aims 139
4.2. Interaction studies involving recombinant human IgM Fc 140
4.2.1. Expression and purification of recombinant human IgM Fc using HEK cells 140
4.2.2. Cloning, expression and purification of recombinant DBL domains using E coli expression system 147
4.2.3. Interaction studies using recombinant IgM and DBL domains 152
4.2.4. Consolidation of interaction studies of hIgM-Fc and DBL domains 163

5. Perspectives and Discussion 164
5.1. PfEMP1 and CD36 164
5.1.1. Binding interactions between CIDR1α domains and CD36 164
5.1.2. Crystallisation trials 165
5.1.3. Future work 166
5.2. DBL and IgM domains 167
5.2.1. Interaction between DBL domains and IgM 167
5.2.2. Generation of a heterogenous human IgM Fc core 167
5.2.3. Future work 168
5.3. Conclusion 169

Bibliography 170
Appendix 194

List of Figures

1.1 Life cycle of *P. falciparum*. 3
1.2. Erythrocyte invasion by *P. falciparum*. 6
1.3. The two exon structure of v*ar* genes. 12
1.4. Graphical representation of nuclear compartmentalisation of *var* genes. 13
1.5. The domain architecture of a hypothetical PfEMP1. 15
1.6. Graphical representation of knob like formations on the membrane. 16
1.7. NTS region of the NTS-DBLα1 structure from VarO variant (PDB accession 2XU0). 17
1.8. Architecture and structure of DBL domains. 21
1.9. Architecture and structure of CIDRα domain. 23
1.10. Crystal structure of NTS-DBL1α1-CIDRγ domains of VarO variant (PDB accession 2YK0). 25
2.1. Converted pET15bcon vector. 34
3.1 Graphical representation of the human CD36 protein structure from studies on the primary sequence. 57
3.2. Analysis of the human CD36 sequence. 67
3.3 Graphical representation of the expression process of hCD36ED using baculovirus expression system. 69
3.4. Purification of soluble hCD36ED. 71
3.5. Sequence alignment of the CIDR1α-M2 regions from different *P. falciparum* strains with the sequence of the crystal structure of MC179 from MCvar1. 72
3.6. Agarose gel electrophoresis for the cloning of PF11CIDR and MC179. 74
3.7. Western blot for the expression trials of MC179 and PF11CIDR. 75
3.8. Purification of MC179. 77
3.9. Response curves from the surface plasmon resonance measurements for a concentration series of CIDR1, CIDR2, CIDR5 and MC179 for their interaction with glycosylated hCD36ED. 80
3.10. Data fitting for surface plasmon resonance measurements of CIDR1α-M2 domains. 82
3.11. Analytical size-exclusion chromatography of MC179 and hCD36ED. 86
3.12. Sedimentation velocity data for MC179. 89
3.13. Sedimentation velocity data for hCD36ED. 90
3.14. Sedimentation velocity data for the complex between hCD36ED and MC179. 92
3.15. Sedimentation equilibrium data for the complex between hCD36ED and MC179. 94
3.16. Graphical representation of the series of PF11 constructs. 97
3.17. Agarose gel electrophoresis for the cloning of PF11CIDR series. 98
3.18. Purification of $PF11CIDR_{EXT}$ and $PF11CIDR_{TRUN}$. 99
3.19. Surface plasmon resonance measurements for the PF11_0007 series. 101
3.20. Surface plasmon resonance measurements of the NTS-DBL1-CIDR1α constructs. 103
3.21. Graphical representation of the N-glycosylation species found in expression in insect cells. 106
3.22. Illustration of the outcome of treatment of N-glycosylation with PNGaseF and endoglycosidase H_f or Endo F_3. 108
3.23. Mass spectra of hCD36ED after endoglycosidase treatment using endoglycosidase H_f and F_3. 109
3.24. Deglycosylation of hCD36ED. 110
3.25. Circular dichroism spectroscopy of glycosylated and deglycosylated hCD36ED. 112
3.26. Surface Plasmon Resonance measurements of CIDR1α constructs for their interaction with

deglycosylated hCD36ED. 114
3.27. Analytical size-exclusion chromatography of MC179 and deglycosylated hCD36ED. 117
3.28. The size-exclusion chromatography trace of the co-complex between deglycosylated hCD36ED and MC179 in the presence of oleic acid (OA). 119
3.29. Example of a complex melting curve of hCD36ED in Tris-based buffers 121
3.30. Thermal shift assays of hCD36ED in the presence of oleic acid. 124
3.31. Purification process of the co-crystallisation complex between deglycosylated hCD36ED and MC179. 126
4.1. Graphical representation of the monomeric and pentameric human IgM. 130
4.2. Expression trials of hIgM-Fc. 142
4.3. Western blots showing the purification of hIgM-Fc expressed in SFM. 143
4.4. Purification trial of hIgM-Fc in different buffers. 144
4.5. Purification of hIgM-Fc. 146
4.6. Agarose gel electrophoresis of ligated MSPDBL2 samples after restriction enzyme digestion. 149
4.7. Purification of MSPDBL2. 150
4.8. Circular dichroism of refolded MSPDBL2. 151
4.9. Western blot of Ni-NTA pull down assay of hIgM-Fc using His-tagged DBL6ε and MSPDBL2 154
4.10. The response curves from the surface plasmon resonance measurements for a concentration series of MSPDBL2 and DBL6ε for their interaction with hIgM-Fc. 155
4.11. Analytical size-exclusion chromatography of hIgM-Fc and MSPDBL2 complex. 157
4.12. Sedimentation velocity data for MSPDBL2. 160
4.13. Sedimentation velocity data for hIgM-Fc and hIgM-Fc:MSPDBL2 complex. 162

List of Tables

1.1. *Var* gene classification of the 3D7 strain according to 5′ promoter class, genomic location and transcription orientation. 14
1.2. Table of PfEMP1 DBL and CIDR domains and their best characterised human ligands. 28
3.1. Annotations for the different CIDR1α constructs. 73
3.2. Affinities of CIDR1α-M2 domains for glycosylated hCD36ED. 83
3.3. Affinities of PF11_0007 series for glycosylated hCD36ED. 102
3.4. Affinities of NTS-DBL1-CIDR1α constructs for glycosylated hCD36ED. 104
3.5. Buffer screen in a 96-well format for hCD36ED stability thermal shift assay. 120
3.6. Lipid binding experiment in a 96-well format for hCD36ED stability thermal shift assay. 122

List of Abbreviations

$\Delta G_{folding}$	Gibbs free energy of unfolding
ACNPV	Autographa californica nuclear polyhedrosis virus
ATS	Acidic terminal sequences
AUC	Analytical ultracentrifugation
CD	Circular Dichroism
CD36	Cluster of differentiation 36
CHO	Chinese hamster ovary
CIDR	Cysteine interdomain rich domain
CLA-1	CD36 LIMPII analogous-1
CPSG	Chondroitin sulphate proteoglycan
CR1	Complement receptor 1
Cryo-AFM	Cryo-atomic force microscopy
CSA	Chondroitin sulfate A
DBL	Duffy binding like
DMEM	Dulbecco's Modified Eagle Medium
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Endo	Endoglycosidase
f/f_0	Weight average frictional ratio
Fab	Fragment antigen binding
FBS	Fetal bovine serum
Fc	Fragment crystallisable region
FcR	Fc receptors
hCD36ED	Human CD36 ectodomain
HEK	Human embryonic kidney
High Five	*Trichoplusia ni* insect cells
hIgM-Fc	Human IgM Fc core
ICAM1	Intercellular adhesion molecule 1
IFA	Immunofluorescence assay
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Dissociation Constant
LB	Lysogeny broth
LDL	Low density lipoproteins
LIMPII	Lysosomal integral membrane protein II
MALDI-TOF-MS	Matrix-assisted-laser-desorption ionisation/time of flight mass spectrometry
MC179	Minimal binding region of MC*var*1
MCS	Multiple Cloning site
MRE	Mean residue ellipticity
MRW	Mean residue weight
MSP3	Merozoite surface protein 3
MSPDBL2	DBL domain of MSP3.8
MWCO	Molecular weight cutoff
NTA	Nitrilotriacetic acid
NTS	N-terminal segment
OA	Oleic Acid

OD_{600}	Optical density at wavelength 600 nm
PBS	Phosphate buffered saline
PCR	Polymerase Chain reaction
PDB	Protein data bank
PECAM	Platelet endothelial cell adhesion molecule 1
PEI	Linearised polyethlenimine
PfEMP1	*P. falciparum* erythrocytic membrane protein 1
PFU	Plaque forming unit
PNGase	Peptide-N-Glycosidase
R_{eq}	Response at equilibrium
R_{max}	Maximum response
rmsd	Root mean square deviation
RPM	Rotations per minute
RU	Resonance response units
S200	Superdex 200 gel filtration column
S75	Superdex 75 gel filtration column
SAM	Serum activated Dulbecco's Modified Eagle Medium
SAXS	Small angle x-ray scattering
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SF9	*Spodoptera frugiperda* insect cells
SFM	Serum free Dulbecco's Modified Eagle Medium
SPAM	Secreted polymorphic antigen associated with merozoites domains
SPR	Surface plasmon resonance
TEV	Tobacco etch virus
T_m	Melting Temperature
TSP1	Thrombospondin 1
Vbar	Protein partial specific volume
VCAM	Vascular cell adhesion molecule
KAHRP	Knob-associated histidine-rich protein

Acknowledgements

I would like to firstly thank Dr Matthew Higgins for the supervision and support I have received during the course of this project. The encouragement, advice and helpful discussion were very important during my time in the lab and in the completion of this thesis.

I am also thankful towards Katherine Wright, Livia Lai, Ailsa Powell and other members of the neighbouring labs for their advice, company and their essential support towards the completion of this project. I am also grateful to the help I received from Dr David Staunton for his help in analytical ultracentrifugation.

Last but not least, I would like to thank my family and Charmaine for the understanding and encouragement they have given me during my time in the UK.

Chapter 1
Introduction

1.1 Introduction to Malaria

Malaria is the most fatal parasitic disease in the world, with around 3.3 billion people living in areas with high risk of contracting malaria, resulting in an estimated 2.16 billion incidences and 655 thousand deaths worldwide in 2010[1]. Despite a reduction in overall global deaths through improved diagnostics and malaria vector control in the past decade, advances in preventive and curative intervention remain elusive.

Infection by *Plasmodium falciparum* causes the most severe cases of malaria[2]. A major cause of its pathogenicity is the ability of *P. falciparum* infected erythrocytes to adhere to human ligands leading to sequestration. *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) protein is central to this property.

1.1.1. Plasmodium parasite: Cause of malaria

Malaria is a vector-borne infectious disease caused by pathogenic parasites of the *Plasmodium* genus, in the *Apicomplexan* family. *Plasmodium* parasites are able to infect primates, humans and other vertebrates such as livestocks and birds. Malaria in human hosts is caused by five known host-adapted species, namely *P. falciparum, Plasmodium vivax*[3]*, Plasmodium ovale*[4]*, Plasmodium malariae*[5] *and Plasmodium knowlesi*[6]. Nearly 91% of the global confirmed malarial diagnosis is due to *P. falciparum*[1].

1.1.2. Evolution and genomics of Plasmodium

Genomic techniques have allowed us to study in detail the full genome of various *Plasmodium* strains[7–9]. The genomes of human host-adapted *Plasmodium* species are fairly similar in terms of size, number of chromosomes, and the number and structure of genes. Other than having 14 linear chromosomes constituting the nuclear genome, the *Plasmodium* species have an additional mitochondrial genome, as well as a 35-kb plastid in the apicoplast. 77% of the genes in sequenced *Plasmodium* species (*P. falciparum, P. vivax, Plasmodium yoelii yoelii)* are orthologous, including important proteins participating in key metabolic pathways, housekeeping processes and membrane transporters[7]. Through comparative genomics, it has been postulated that divergent expansion of recognised gene families has allowed for development of specific biology for each of the different *Plasmodium* lineages[8]. Although most stages in the life cycle of these different *Plasmodium* species remain similar, the unique proteomic identity at specific stages of parasite development of each *Plasmodium* parasite has profound implications on the biology and hence severity of pathology in malaria caused by different species of *Plasmodium*[2].

1.1.3. Life cycle of human-adapted Plasmodium parasites

The life cycle of all human infecting *Plasmodium* parasites involve two separate hosts. An exogenous sexual phase involves the *Anopheles* mosquito, and an endogenous asexual phase takes place in the vertebrate human host (Figure 1.1).

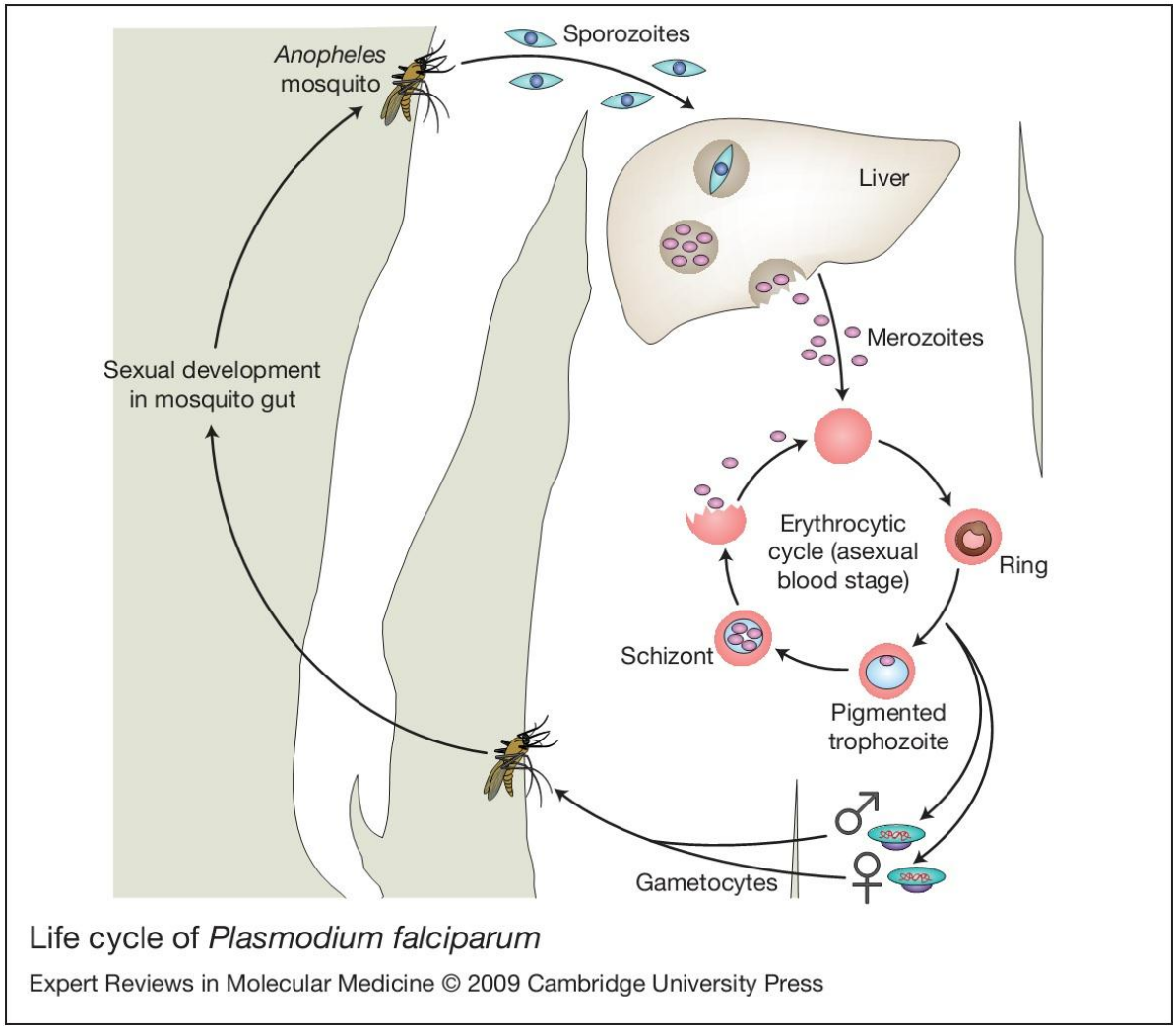


Figure 1.1 Life cycle of *P. falciparum*. Reproduced from Rowe *et al*., 2009[10].

The exogenous sexual phase (sporogeny) begins with the ingestion of matured male and female gametocytes during a blood meal by a suitable species of female *Anopheles* mosquito. Differences in *Anopheles* vector population and dynamics, in addition to differences in climates, geographical locations, and human populations are important factors in the global distribution of various *Plasmodium* species. When gametocytes reach the midgut of the insect, female macrogamete and male microgametes are released from the erythrocytes, allowing fertilization. Fusion of the two nuclei results in the formation of a zygote, which evolves into a motile ookinete. The ookinete can penetrate the epithelial cells of the midgut, which is the site for ookinete development into a matured oocyst containing

thousands of sporozoites. Epithelial cell ruptures to release mature sporozoites, which migrate to the acinal cells of the salivary glands of the mosquito. These highly infective sporozoites are ejected into the bloodstream of the human host during dermal penetration by the mosquito's proboscis in search of blood vessels.

Inoculated sporozoites then begin the endogenous asexual phase of their life cycle in the human host. This asexual phase is comprised of two separate phases– an exo-erythrocytic phase, which involves invasion of liver cells, and an erythrocytic phase which involves invasion of erythrocytes.

The exo-erythrocytic phase is characterised by the invasion of hepatocytes in the liver by sporozoites. Here, sporozoites mature into schizonts, in process that can take about 6 (*P. falciparum)* to 16 (*P. malariae)* days. Matured schizonts rupture to release merozoites into the bloodstream, to begin the erythrocytic phase. Interestingly, *P. vivax*[11,12] and *P. ovale*[4,13] are able to remain dormant during tissue phase in the liver cells through the development of hypnozoites, leading to recurring relapse of malaria.

Pathogenicity of malaria largely occurs during the erythrocytic phase of the *Plasmodium* life cycle[2]. Merozoites invade the erythrocytes through species-specific pathways. Parasites develop within infected erythrocytes into ring stage trophozoites, which then further mature into blood stage schizonts, each containing merozoites that are released when the erythrocytes rupture, to infect neighbouring red blood cells. Again, the process and time required for maturation is species-specific.

During trophozoite maturation, some parasites undergo gametocytogenesis to form macrogametocytes and microgametocytes, completing the transmission cycle between human hosts via an *Anopheles* mosquito vector. The amount of time taken for gametocytes to be detected differs largely between species, thus affecting relative transmission efficiency between infected and uninfected hosts. It takes only 3 days for gametocytes to be detected in *P. vivax* infection[14], while it might take up to 7-10 days for mature gametes to be found circulating in a *P. falciparum* infection[15]. These can then be ingested by a female *Anopheles* mosquito to restart the exogenous sporogenic cycle in the mosquito vector.

1.2 Plasmodium falciparum and severe malaria

In addition to causing 91% of all diagnosed cases of malaria in 2010[1], there are three reasons why *P. falciparum* is more deadly than other *Plasmodium* species. The first reason lies in *P. falciparum's* highly adaptive invasive properties. Secondly, *P. falciparum* is able to undergo antigenic variation to evade immune surveillance, leading to increased parasitemia and hence chronic infections in patients. Thirdly, the unique adherent properties of *P. falciparum* infected erythrocytes allow adhesion to human ligands on endothelial membrane and uninfected erythrocytes. This adherence is the key process leading to exclusive clinical manifestations of cerebral and placental malaria, as well as increasing the likelihood of severe malaria in *P. falciparum* infected patients[16,17].

1.2.1. Invasive properties of Plasmodium falciparum

Invasion of erythrocytes by *Plasmodium* species involves a series of sequential steps (Figure 1.2). Invasion begins with recognition of specific surface receptors on the erythrocyte, with

which receptors on the merozoite associates weakly[18]. The attached merozoite then tumbles, and in the process reorientates its apical tip into contact with the erythrocytic membrane[19,20]. Apical reorientation is followed by the formation of a moving junction[21–23], between malarial microneme protein PfAMA1 and PfRON4, the latter of which is a member of the RON (Rhoptry neck) complex that is inserted into the cytosolic face of the erythrocytic membrane during rhoptory discharge[24,25]. The junction then inserts into the erythrocytic membrane, in a process requiring an actomyosin motor system[21,26]. The process finally ends with sealing of the invaded merozoite to form a parasitophorous vacuole within an infected erythrocyte[27].

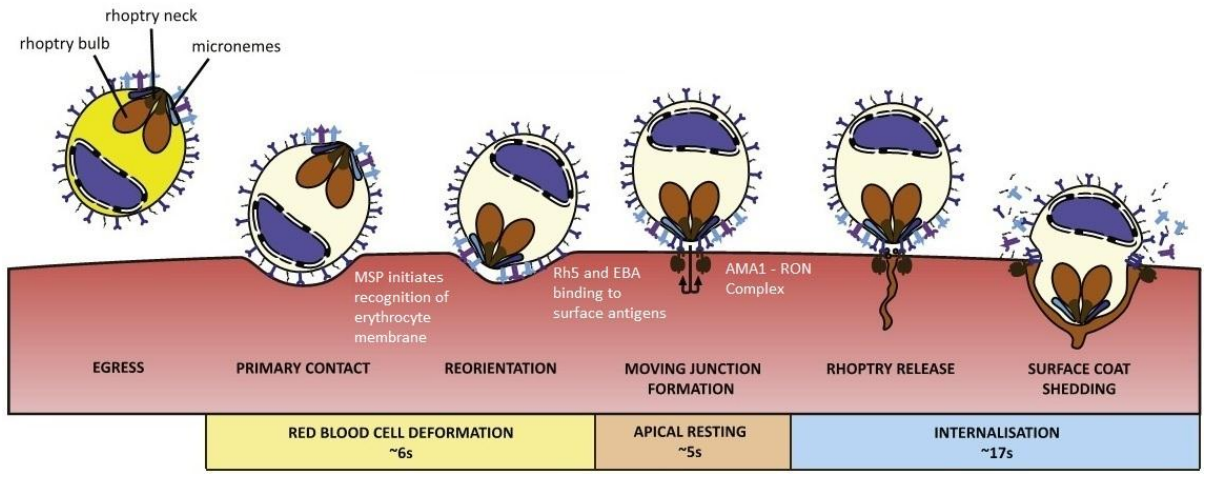


Figure 1.2. Erythrocyte invasion by *P. falciparum*. Reproduced and edited from Harvey *et al.,* 2010[28].

Two unique invasive properties make *P. falciparum* a more dangerous pathogen than other *Plasmodium* species. Firstly, it can invade erythrocytes at various levels of maturity[29], while *P. vivax* can only infect reticulocytes. Polymorphisms in the Duffy blood group result in immunity against *P. vivax* infection. This is illustrated in the selection and inheritance of the Duffy-negative trait by the native population in malaria endemic West Africa[30], where *P. vivax* infections has been eradicated, now causing less than 1% out of all diagnosed malaria infections[3]. Additionally, *P. falciparum* which causes severe malaria can invade most types of erythrocytes while *P. falciparum* involved in uncomplicated malaria can only invade a

subpopulation of erythrocytes during episodes, suggesting that the ability to invade a repertoire of cells correlates to severe clinical manifestations[31]. This makes the general population who are naïve to malaria more susceptible to *P. falciparum* due to its non-selective invasive properties.

Secondly, *P. falciparum* is able to make use of redundant pathways for invasion. *P. falciparum* merozoites are able to recognise a variety of receptors during initial recognition of host specific erythrocytes[29,32–34], and invade via recognition of different receptors on the erythrocyte membrane[28]. Switching between various pathways occurs in the presence of different selection pressures, and this switching mechanism occurs reversibly, such as between a sialic acid-dependent and -independent pathway[35]. In contrast, *P. vivax* and *P. knowlesi* can only invade erythrocytes by recognition of the Duffy blood group antigen[36,37].

The flexibility in invasive properties of *P. falciparum* allows for greater survivability by being able to invade different subtypes of erythrocytes, hence increasing severity of malaria as a result of increased parasitic burden.

1.2.2. PfEMP1 and its role in antigenic variation and immune suppression

PfEMP1: The link between cytoadherence and antigenic variation

Antigenic variation is a mechanism by which the pathogen is able to evade host immune clearance by temporal coordination of variant surface molecules. In the case of *P. falciparum*, this leads to the establishment of chronic recurring infections, characterised by fluctuating levels of parasitemia[38].

P. falciparum erythrocytic membrane protein 1 (PfEMP1) proteins, encoded by *var* genes, are expressed on the infected erythrocytes. These proteins mediate cytoadherence, a crucial part of the survival strategy of *P. falciparum* infected erythrocytes to allow avoidance of splenic clearance[39–41]. Complicated malaria such as in the case of placental[42,43] and cerebral malaria[44,45] can arise as a result of this increase in parasite burden. Several studies have linked PfEMP1 to antigenic variation[46–49].

PfEMP1 protein is a target antigen of the adaptive immune system[50,51]. A classic example is the reduced risk of repeat cases of placental malaria for pregnant women who had been previously infected[52–54]. However, recognition of these PfEMP1 proteins seems to be variant specific[50]. Antigenic variation by changing the exposed PfEMP1 protein of the infected erythrocyte at different time points during the entire clinical malaria episode protects the infected erythrocyte from being recognised by antibodies, thus leading to persistent parasitemia.

Molecular mechanism of antigenic variation

Many members of the *var* gene family are located within the sub telomeric regions of the chromosomes[7,55,56]. This allows for antigenic variation to occur by three mechanisms. Firstly, heterogenetic silencing can define the domain boundaries of the gene loci at telomeres which are accessible for transcription, within a general environment which is otherwise silenced[57–59]. Secondly, *var* genes at telomeric region localise to the nuclear periphery where genetic recombination can occur at recombination hotspots to contribute to the overall repertoire of PfEMP1[60–63]. Lastly, *var* genes are subjected to epigenetic regulation leading to changes in the chromosomal status and form the molecular basis for

which the parasite can switch between single *var* genes for monoallelic expression of a single clonal PfEMP1 variant during the course of infection[64,65].

Temporal regulation of antigenic variation

Global coordination of antigenic switching between singly expressed *var* genes of the *P. falciparum* population during infection is dependent on a few factors. The first *in vivo* study on human volunteers suggests that *var* gene expression switches with the developmental stage of the parasite during different stages of infection[66]. Although one might postulate that switching events are due to selective pressure to avoid recognition by the immune system[67], other studies suggests that these events can be driven by the selection for strong cytoadherent properties, allowing escape from splenic destruction[47,49]. Mathematical models[68,69] and studies using genome wide transcriptome profiling[70] supported the presence of a balance between intrinsic switching mechanisms elicited by the parasite, in addition to immune-mediated selection. This can account for the presence of specific locations of v*ar* genes and the non-random combinations of domains on some PfEMP1 proteins that have occurred as a result of selective recombination[71].

1.3 Sequestration and PfEMP1 – the link between P. falciparum and severe malaria

A major factor that contributes to the severity of malaria in a *P. falciparum* infection is the modification of infected erythrocyte membrane to allow adhesion to ligands on the surface of host cells. This is known as sequestration. A function of the spleen is to mediate destruction of abnormal and infected erythrocytes within the bloodstream[72]. This occurs by reducing the velocity of blood flow in the spleen for the T and B lymphocytes to sample the blood for microorganisms and abnormal erythrocytes. The parasite has evolved to circumvent spleen-mediated destruction by sequestration. Sequestration leads to chronic infection as infected erythrocytes reside in the microvasculature for a longer period of time. This survival strategy increases the parasite burden during a *P. falciparum* infection, leading to severe malaria as well as organ specific (cerebral and placental) malaria [2]. PfEMP1s are central to cytoadhesion[73].

1.3.1. Molecular mechanism of sequestration

Three mechanisms can lead to sequestration. The best studied involves adhesion of infected erythrocytes to host receptors on endothelial cells[74]. The second process involves rosetting of infected erythrocytes to uninfected erythrocytes to form clumps of erythrocytes in the microvasculature[75–77]. The third process involves clumping of platelets together with infected erythrocytes, though the precise contribution of this process in severe malaria remain unclear[78–80]. Despite differences in mechanistic details between these processes, there are certain similarities which are important in the defining the role of sequestration in severe malaria.

Sequestration, via rosetting or cytoadherence, leads to similar pathology. An increase in parasite burden occurs as infected erythrocytes are not cleared by splenic mediated destruction[2], and infected erythrocytes obstruct the microvasculature as a result of clumping to endothelial cells and uninfected red blood cells, such as during organ-specific malaria[81]. It may be that a combination of both is required for manifestation of severe malaria[10,76].

The main proteins involved in both cytoadherence[39–41] and rosetting[77,82] are PfEMP1s. In both processes, individual DBL and CIDR domains have been identified as interaction partners to the various human ligands on their respective targets[10,73].

Both genetic and protein-based approaches have been used to identify the contribution of specific adhesion phenotypes to pathogenesis leading to severe malaria. The genetic approach uses genomic tools such as comparative genomics and transcriptome profiling to link adhesion phenotypes of *P. falciparum* infected erythrocytes from field isolates to respective *var* genes. The protein approach involves studying binding of *P. falciparum* infected erythrocytes to receptors by using static or flow binding assays. The binding capacities of members from the same class of PfEMP1-encoding *var* gene are usually similar[83]. It is therefore important to understand how PfEMP1 encoding *var* genes are classified in addition to the molecular and structural functions of individual PfEMP1 in order to draw links between an observed adhesive phenotype and clinical manifestations.

1.3.2. Var genes

Var genes encode variable surface antigens that are involved in cytoadherence[39–41]. Each *P. falciparum* strain carries around 60 *var* genes, each encoding a different PfEMP1 protein.

These are expressed in a mono-allelic fashion[84]. The 3D7 strain carries 59 members of the *var* gene family and 1 *var* pseudogene[7].

1.3.2.1 Structure of a var gene.

Each *var* gene is made up of two exons (Figure 1.3). The first measures 3.5 to 9.0 kb and encodes the extracellular, polymorphic, functional region of the PfEMP1. The second exon is much shorter, measuring 1.0 – 1.6 kb. It encodes a conserved cytoplasmic terminus, consisting of a single acidic terminal sequence (ATS).

The intervening introns are postulated to have an important regulatory role in the expression status of the gene loci[85–87]. These intron sequences can vary from 170 base pairs to 1.2 kb in length, and are A/T rich[7].

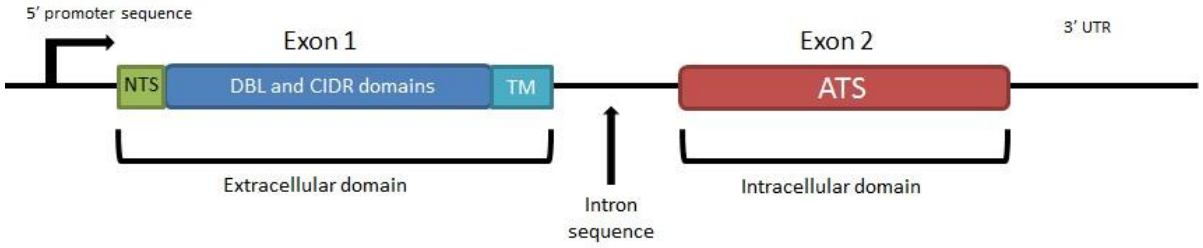


Figure 1.3. The two exon structure of v*ar* genes.

1.3.2.2 Classification of var genes

Classification of *var* genes has a profound impact on the understanding of the cytoadherent properties of PfEMP1. Different groups of *var* genes have a unique gene location, orientation, and flanking sequences (Figure 1.4, Table 1.1). Each group of *var* genes encodes for a distinct functional combination of domains on PfEMP1 which are preserved by limiting recombination events between different groups[71,88,89].

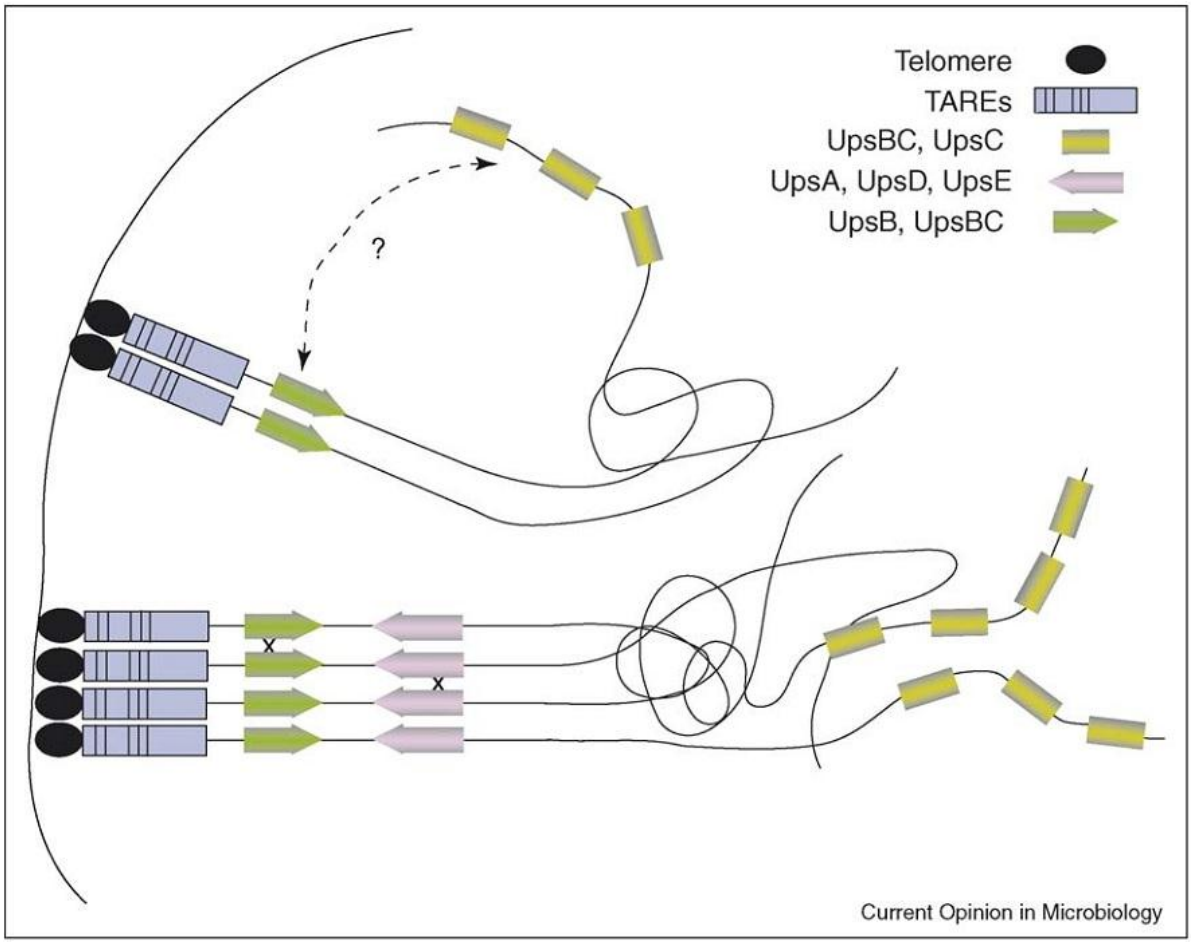


Figure 1.4. Graphical representation of nuclear compartmentalisation of *var* genes. Telomeric and subtelomeric groups A and B *var* genes locate to the nuclear periphery, while group C var genes are central in location. Recombination (x) only occurs within *var* genes within the same group. Suspected recombination between *var* genes of group B and C represented by (?) sign. Reproduced from Kraemer *et al*., 2006[73].

Var genes are classified into subfamilies according to their position along the chromosome and their alignment of 5′ sequences 1.6 to 2 kb upstream of the *var* genes into five upstream sequence classes (Ups) A, B, C, D and E[7,62,71,73,88].

These subfamilies can be further categorised into three broad groups, according to their gene location, their 5′ promoter subfamily as well as their domain composition[88] (Table 1.1). Group A consists of subtelomeric *var* genes flanked by UpsA sequence and is transcribed towards the telomere. Group B consists of subtelomeric *var* genes flanked by UpsB sequences and is transcribed towards the centromeres, while group C harbours all of the *var*

genes residing in the central regions of the chromosomes flanked by UpsC sequence. Additionally, there are two other intermediate groups for which the domain composition and gene location differs from the predicted Ups-linked nomenclature, termed the groups B/C and B/A for their mixed promoter sequence[62,73,88]. There are also three strain-transcendent variants (Table 1.1), namely *var1CSA* (UpsD-linked)[90], *var2CSA* (UpsE-linked)[91], and *var*3 (UpsA-linked)[92] which carries genes which encode for strain specific PfEMP1 that are encoded by a set of 23 conserved domain cassettes[62].

var gene group	Number of members	5' flanking promoter class	Genomic Location	Transcription orientation
A	9	UpsA	Subtelomeric	Towards telomere
B/A	4	Ups B/A	Subtelomeric	Towards Centromeres
B	22	Ups B	Subtelomeric	Towards Centromeres
B/C	9	Ups B/C	Centromeric	Towards Telomeres
C	13	UpsC	Centromeric	towards telomeres
Strain-transcendent *var*				
*var*1CSA	1	UpsD	Telomeric	Telomere
*var*2CSA	1	UpsE	Subtelomeric	Telomere
*var*3	1	UpsA	Subtelomeric	Telomere
Total	59 var genes, 1 pseudo gene (*var*1CSA)			

Table 1.1. *Var* gene classification of the 3D7 strain according to 5′ promoter class, genomic location and transcription orientation[73].

1.3.3. PfEMP1 proteins and domain classification

Sequence classification has been conducted based on the primary sequence of PfEMP1 proteins in conjunction with studies on their corresponding *var* gene to better understand their sequence diversity and binding properties[89,93].

PfEMP1 proteins are highly modular in structure (Figure 1.5). They can be largely separated into extracellular and cytoplasmic portions. The extracellular framework of the PfEMP1 protein is made up of two main domains: Duffy binding like (DBL) and cysteine interdomain rich (CIDR) domains. This is followed by a hydrophobic transmembrane segment, which is

linked to the ATS domain, accounting for the cytoplasmic facing segment of PfEMP1 protein.

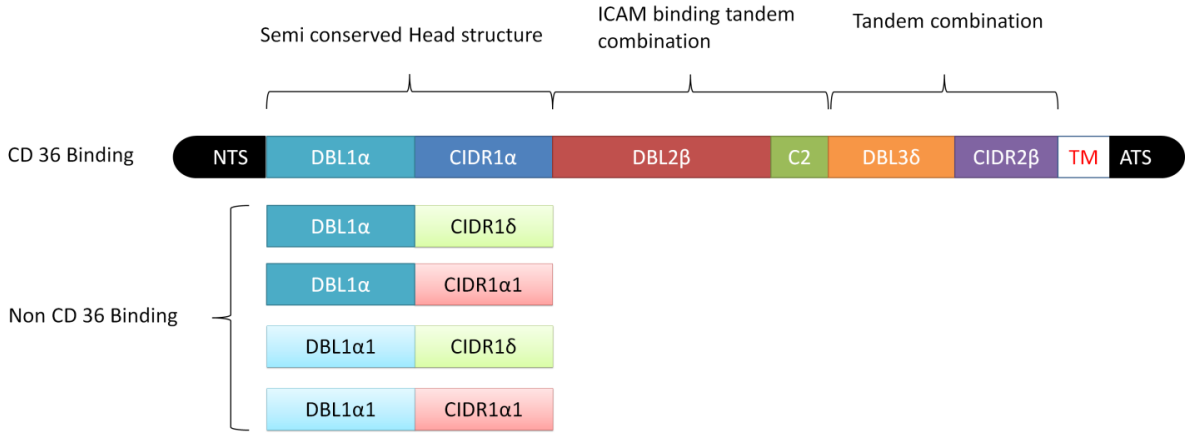


Figure 1.5. The domain architecture of a hypothetical PfEMP1. This PfEMP1 consists of different DBL and CIDR domains, as well as a C2, ATS and NTS domain. DBL and CIDR domains are identified by the number at which the domain is position in the PfEMP1 relative to the amino terminal, followed by the domain class[88]. There are three conserved tandem combinations (DBL1α-CIDR1α, DBLβ-C2 and DBLδ-CIDRβ) that are shown in the figure[71]. The head structures of non CD36-binding PfEMP1 are shown in a different colour combination below the canonical CD36-binding head structure.

1.3.3.1 ATS

The ATS domain is located on the cytoplasmic-facing carboxy terminus of PfEMP1[94]. Despite their cytoplasmic orientation, ATS domains have important functions in cytoadherence. The ATS domain has homology blocks that can act as a signal peptide that allow transportation of PfEMP1 to the erythrocytic membrane[62]. At the erythrocytic membrane, the ATS domain interacts with Knob-associated histidine-rich protein (KAHRP) proteins[95,96] to organise PfEMP1 in clusters on knob-like deformations on the membranes of the infected erythrocytes[39,97,98] (Figure 1.6). PfEMP1 clustering enhance sequestration as KAHRP knockout abolishes knob formation and eradicates the binding properties of infected erythrocytes to CD36, a glycoprotein involved in sequestration, in a flow assay[99]. ATS sequences can classified into ATSA, ATSB, ATSpam, ATSvar1 and ATSvar3[62].

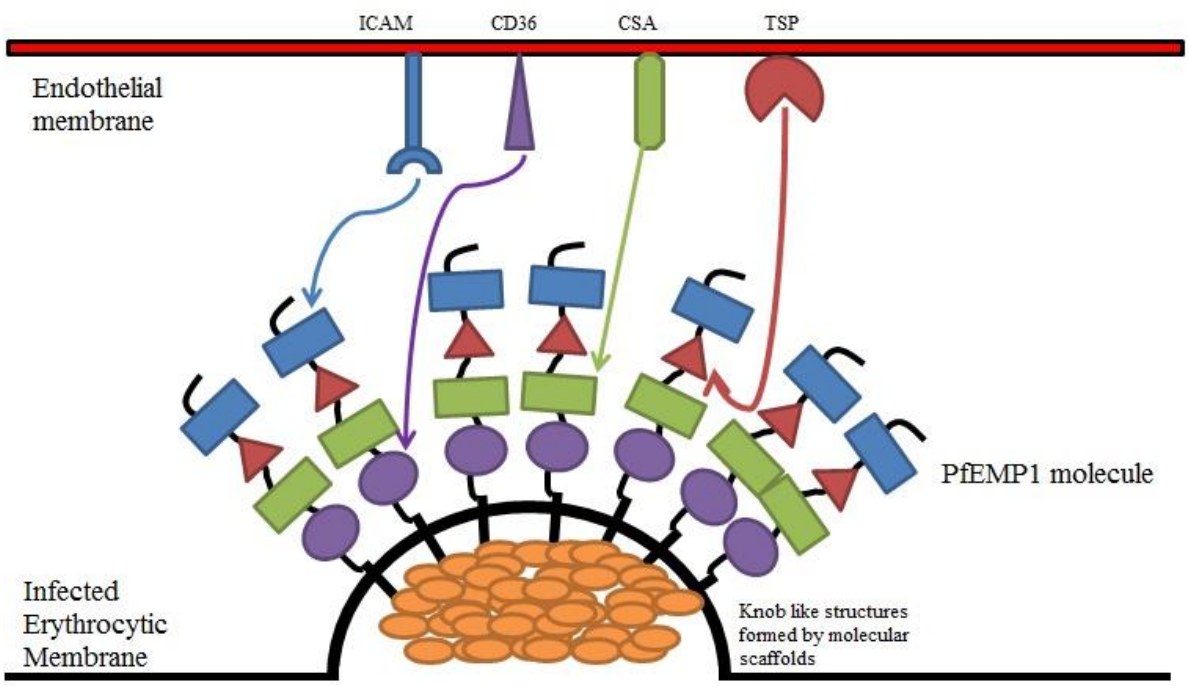


Figure 1.6. Graphical representation of knob like formations on the membrane. Modular PfEMP1 is orientated in clusters by interaction between ATS domain and KAHRP proteins. Binding domains on PfEMP1 molecules are then able to interact with receptors on the endothelial membrane.

1.3.3.2 N-terminal segment (NTS)

The NTS at the amino terminus of PfEMP1 protein ranges from 75 to 107 amino acids in length. The NTS may serve as a possible signal for transportation of PfEMP1 proteins to the membrane of the infected erythrocyte[100].

A crystal structure of an NTS has been solved as part of a DBLα1 domain of a PfEMP1 protein that form rosettes by interacting with tri-saccharides of ABO blood group antigen[101,102]. This NTS contains a central α helix (αH2) flanked by two other short α helices (αH1 and αH3) (Figure 1.7). The NTS sits in a cleft in DBLα1 and makes extensive contacts with the DBLα1 domain which are difficult to disrupt, thus suggesting that it forms an integral portion of the DBLα domain structure. Indeed, NTS domains always occurs N-terminal to DBLα domains and in no other DBL domains[7,101].

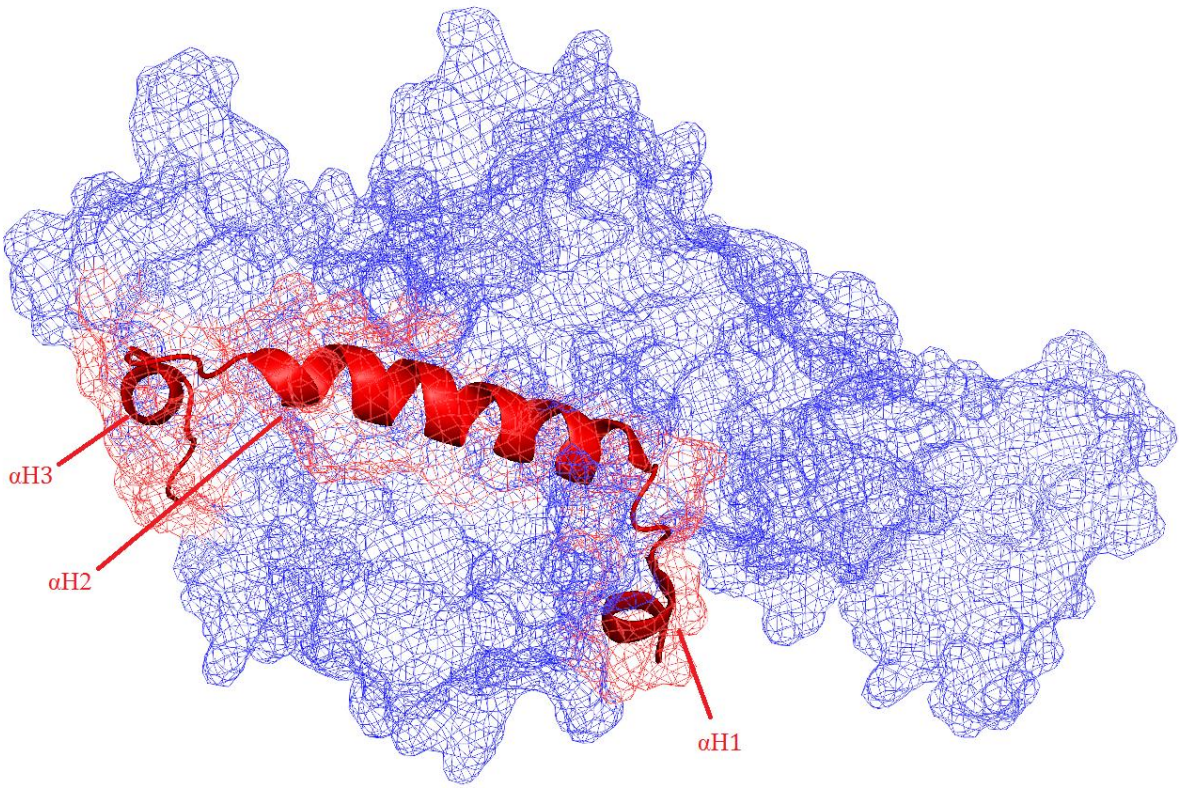


Figure 1.7. NTS region of the NTS-DBLα1 structure from VarO variant (PDB accession 2XU0). NTS region is coloured red, against a blue electron mesh of the DBL1α1 domain. The NTS structure is made up of a central α-helix (αH2) with two short spanning α helices (αH1 and αH3).

1.3.4. The domains involved in cytoadherence in PfEMP1 proteins: DBL and CIDR domains.

Despite the lack of sequence similarity between the various DBL and CIDR domains, it is possible to classify them to different groups via amino acid sequence alignment as well as the presence of conserved motifs and folds in their tertiary structure. As such, DBL domains are grouped into six sub-classes (α, β, γ, δ, ε, ζ) while CIDR domains are grouped into five sub-classes (α, β, γ, δ and pam)[62,88,89].

1.3.4.1 DBL domains – Consensus motif and structure

Consensus motif

Unlike CIDR domains which are unique to *P. falciparum,* DBL domains of PfEMP1 are similar to the DBL domains found in erythrocyte binding antigens (EBA) proteins involved in invasion of erythrocytes[103]. Given that erythrocyte invasion is a necessary property of all malaria causing *Plasmodium* species while cytoadherence is unique to *P. falciparum*, it is probable that DBL domains in PfEMP1 evolved much later from ancestors of these EBA proteins. Indeed, PfEMP1 DBL domains are significantly divergent from the EBA DBL domains, supporting the above view[89]. This divergence into six distinct sequence classes may be attributed to functional and immunosuppressive pressures, resulting in a plethora of domains for the modular PfEMP1[88].

Consensus sequences on DBL domains of PfEMP1 proteins have been elucidated using multiple sequence alignment. DBL domains are classified by the presence of homology blocks[62,73,89]. Alignment of these main homology blocks which consist of different conserved residues allow us to classify DBL domains of PfEMP1 into six sub-classes as described above. Each of these different homology blocks are predicted to have secondary structure[62], supported by mapping them into DBL domain structures[43,104–106]. Using alignment analysis, Rask and colleagues described five homology blocks out of the ten identified[89] to contain the most strongly conserved structural features shared by nearly all DBL domains in PfEMP1 (Figure 1.8a). DBL domains of PfEMP1 also share 10 evenly distributed invariant cysteine residues[93]. These cysteine residues are important in maintaining disulphide bonds, which serve to keep the various subdomains of DBL domains intact[104,106,107].

Divergence between various DBL classes is concentrated in 10 hypervariable blocks which are interspersed between homology blocks of DBL domains[89]. Hypervariation in DBL domains can result either from variation in length or amino acid sequence. Unlike EBA DBL domains which have an invariant length, DBL domains are observed to manifest frequent length variation. Variations in lengths may have arisen by two different mechanisms. DBL domains of different lengths have been identified to possess similar binding partners, an example being DBL3X and DBL6ε of *var*2CSA binding to CSA which has been linked to placental malaria. Both DBL6ε (302 residues) and DBL3X (359 reisudes) are able to interact with CSA using different surfaces despite differences in length[43]. This strongly suggests that the cytoadherent function of the DBL domains may be a result of conserved charged patches present in different surface locations for different DBL domains. This could be due to the different lengths of the DBL domains, especially in the flexible linker regions between α helices[108]. Variation in lengths can also account for the capacity for to express different PfEMP1 proteins for immune evasion. This dual selection under which DBL evolves allows for diversification in amino acid sequence and hence leads to conserved hypervariable blocks within DBL domains of PfEMP1 not found in EBA-DBL[62,71,88,109].

Structure

The DBL domain structure was described as a "boomerang" shaped module made up of an α helical core[104–106,110] (Figure 1.8b,c,d). Despite divergence in sequence and length, the overall structure of EBA175 DBL domain is similar to that of DBL domains found in PfEMP1[105,110] (Figure 1.8b,c,d), containing three main subdomains[43,101,104,110]. Subdomain 1 (S1) contains a coiled structure with interspersed helical structures. This structure is held by two internal disulphide bonds in DBL3x of var2CSA[104,110], but these disulphide bonds are absent in a similar subdomain in DBL6ε domain from the same PfEMP1 protein. Subdomain

2 (S2) is formed by four α-helices held by four loop insertions. Again, internal disulphide bonds coordinate and stabilise the DBL folds in both DBL3x[104,110] and Pkα-DBL[105], but are once again absent from DBL6ε[43]. Thus, this suggests that the intra- and inter-strand disulphide bonds in S1 and S2 are not essential in maintaining the DBL fold.

Subdomain 3 (S3) consists of two elongated antiparallel α-helices. These form a bundle with a third strand, that is held together by inter-strand disulphide bonds[43,104] (two disulphide bonds in DBL6ε and one in DBL3X). A conserved feature seen in all the crystal structures of solved DBL domains is the stabilisation of the distal end of the bundle by disulphide bonds (Figure 1.8b,c,d). This stabilises the most distal loop structure. In comparison with the two other subdomains, the S3 region is better conserved within DBL domains. Two out of the five crucial homology blocks present in all DBL domains are mapped to the two alpha helices of S3. Homology block 1 (HB1), which consists of the largest number of residues among the five homology blocks is mapped almost entirely onto one of the elongated alpha helix of the S3[62]. Conserved residues in HB1 also span the entire inward facing surface of the alpha helix[62], thus highlighting the importance of S3 in the structure of DBL domains. Indeed, the S3 region of DBL3x has been found to be the minimal region required to bind to CSA using structure-based mutagenesis and NMR spectroscopy[111], suggesting the importance of the S3 region in cytoadhesion.

Structures of NTS-DBLα1-CIDRγ and NTS-DBLα1 tandem domains showed the importance of NTS in maintaining the structure of DBLα1[101,102] (Figure 1.7, 1.10). Mutational studies on the NTS-DBLα1 domain also highlighted the importance of the NTS-DBLα1 hinge region in coordinating the binding pocket of the DBLα1 domain for ABO blood group antigen[102]. Although the NTS is exclusive to the DBLα domains[7], the presence

of linker regions outside the canonical DBL domain boundaries may prove to be important in coordinating DBL domain-mediated binding in a full length PfEMP1 protein.

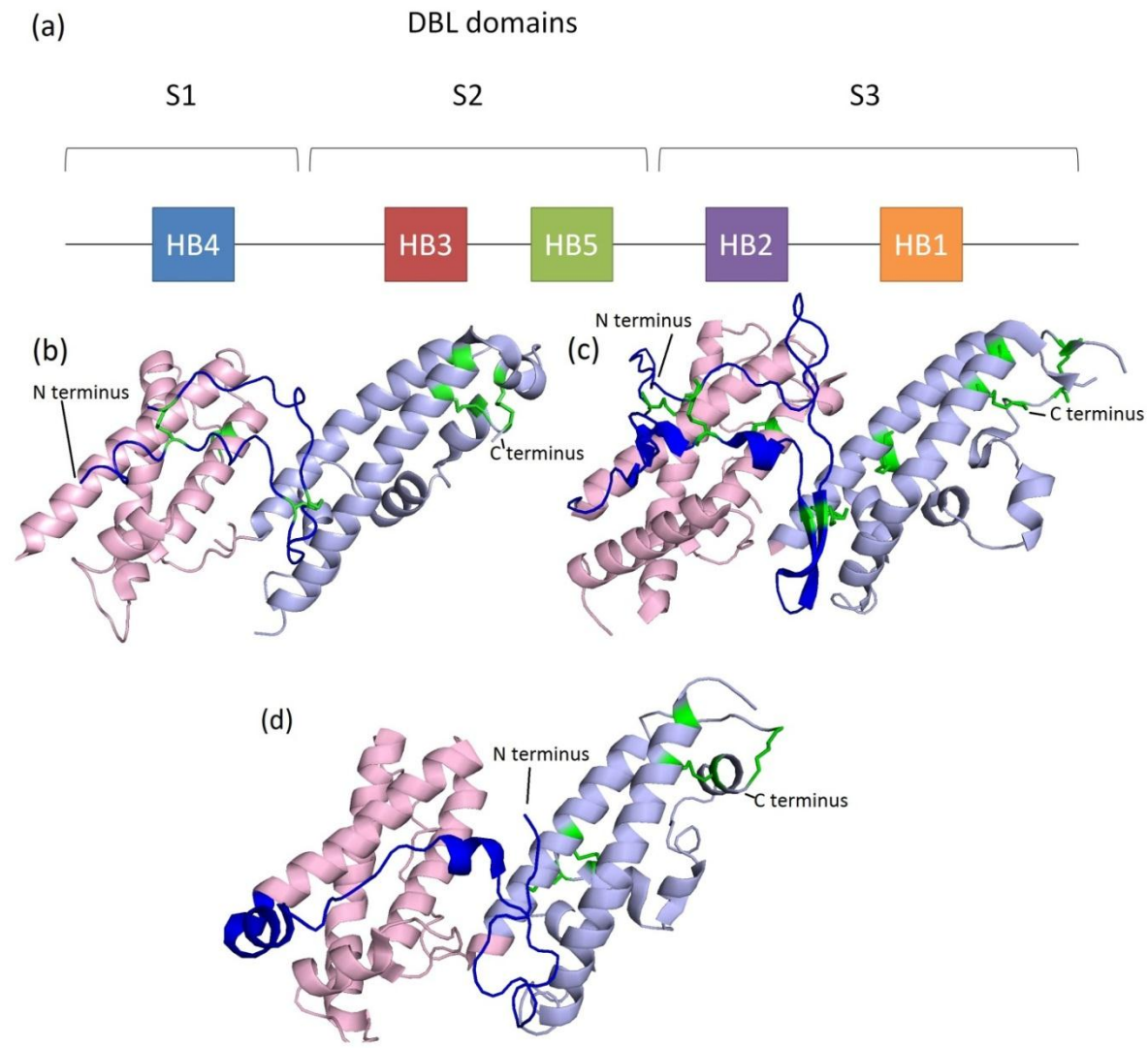


Figure 1.8. Architecture and structure of DBL domains. **(a)** The five most prevalent homology blocks on DBL subdomains 1, 2 and 3 respectively. **(b)** Crystal structure of *P. knowlesi* Pkα-DBL domain[105]. (PDB accession 2C6J). **(c)** Crystal structure of DBL3x domain from *P. falciparum var*2CSA[104] (PDB accession 3BQK). **(d)** Crystal structure of DBL6ε domain from *P. falciparum var*2CSA[43](PDB accession 2WAU). The Pkα-DBL domain involved in erythrocyte invasion has a similar shape to the DBL3x and DBL6ε domains involved in adhesion to CSA with the disordered S1 (blue) and ordered S2 (light pink) and S3 (violet) are coordinated in a similar boomerang shape. Conserved disulphide bridges (green) found at the most distal end of S3 are present in all three structures.

1.3.4.2 CIDR domains – Consensus motifs and structure

Consensus motif

CIDR domains are separated into three main subclasses[89,93] (α, β, γ). A recent study identified another two CIDR domain subclasses, termed as CIDRδ and CIDRpam[62]. CIDRpam domains are found in the *var*1 and *var*2csa genes that are divergent from the rest of the CIDR domains in PfEMP1.

CIDR domains can be classified according to their function, as demonstrated by CD36-binding activity of MC*var*1[112], into three separate portions, namely the M1, M2 and M3 from N to C terminal. M2 is the minimal CD36-binding domain as affirmed by binding assays[112,113]. The conserved core of CIDR domains can again be mapped by alignment of homology blocks. Within the M1 region, there exist six invariant cysteine residues that are common to all CIDR sub groups. In addition, the domain boundary of CIDR is marked by highly conserved FL residues[89]. The most conserved homology blocks (HB1, 8 and 10) align in the M2 region[62] (Figure 1.9a). Like DBL domains, these homology blocks are mapped to structural features that are highly conserved between CIDR domains and maintain the CIDR fold. CIDR domains are also rich in cysteine residues, as their name suggests. The consensus cysteine rich motif ($C_{7-12}X_{3-5}CX_3CX_{1-2}CX_3CWX_{7-8}W$) in the M2 region is characteristic of CIDR domains[39,41,89]. The M3 region however is less conserved. Rask and colleagues have suggested separating the M3 portion from CIDR domains for greater accuracy in CIDR domain classification[62].

Structure

The structure of the minimal CIDRα-M2 region of MC*var*1 (MC179) required for CD36-binding has been solved[114]. It resembles a V-shaped module made up of a three helix bundle (H1, H2 and H3) linked together by 3 short α helices (helix **a**, **b**, **c**) (Figure 1.9b). In a manner that is similar to the structure of the DBL domain, the distal end of the three helix bundle is held together by three disulphide bonds (two of which are inferred and not observed in the electron density map).

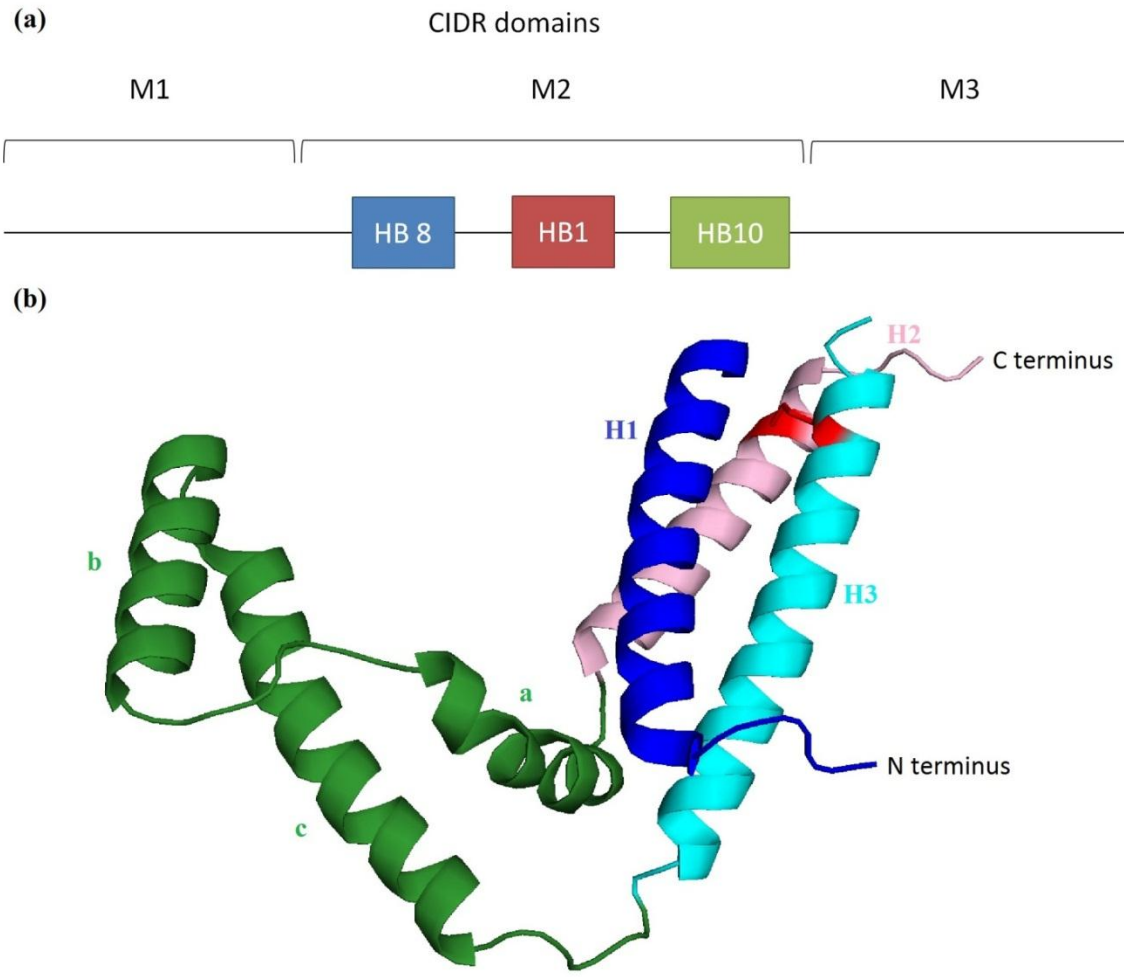


Figure 1.9. Architecture and structure of CIDRα domain. **(a)** The three most prominent homology blocks are found in M2 region of CIDR domains. **(b)** Crystal structure of MC179 (M2 region of CIDRα) from *P. falciparum* MCvar1[114] (PDB accession 3C64). The α-helices in the three helical bundle is termed H1 (blue), H2 (light pink) and H3 (turquoise) respectively. The three additional helices are termed a, b, and c (green). Disulphide bond detected during diffraction is coloured red (inferred disulphide bonds are not modelled).

A structure of CIDRγ domain in the presence of a DBLα domain of the VarO variant shows that this CIDR domain forms a compact structure in the presence of DBLα domains[102] (Figure 1.10). Here, the N-terminal region of the CIDR1γ, made up of four antiparallel β-sheets with three short interspersed α-helices, interacts with the neighbouring DBL1α1 domain extensively via several charged and hydrophobic interactions (Figure 1.10). This allows CIDRγ to assume a more compact structure in comparison to the structure of MC179. The NTS-DBL1α1-CIDR1γ structure also confirmed previous observations that the helices involved in the three helix bundle of CIDRα domains overlap with the conserved homologous α-helices of DBL domains[114]. The similarity between DBL and CIDR domains suggests that a full length PfEMP1 will be a series of helical bundles linked together by flexible loops and linkers.

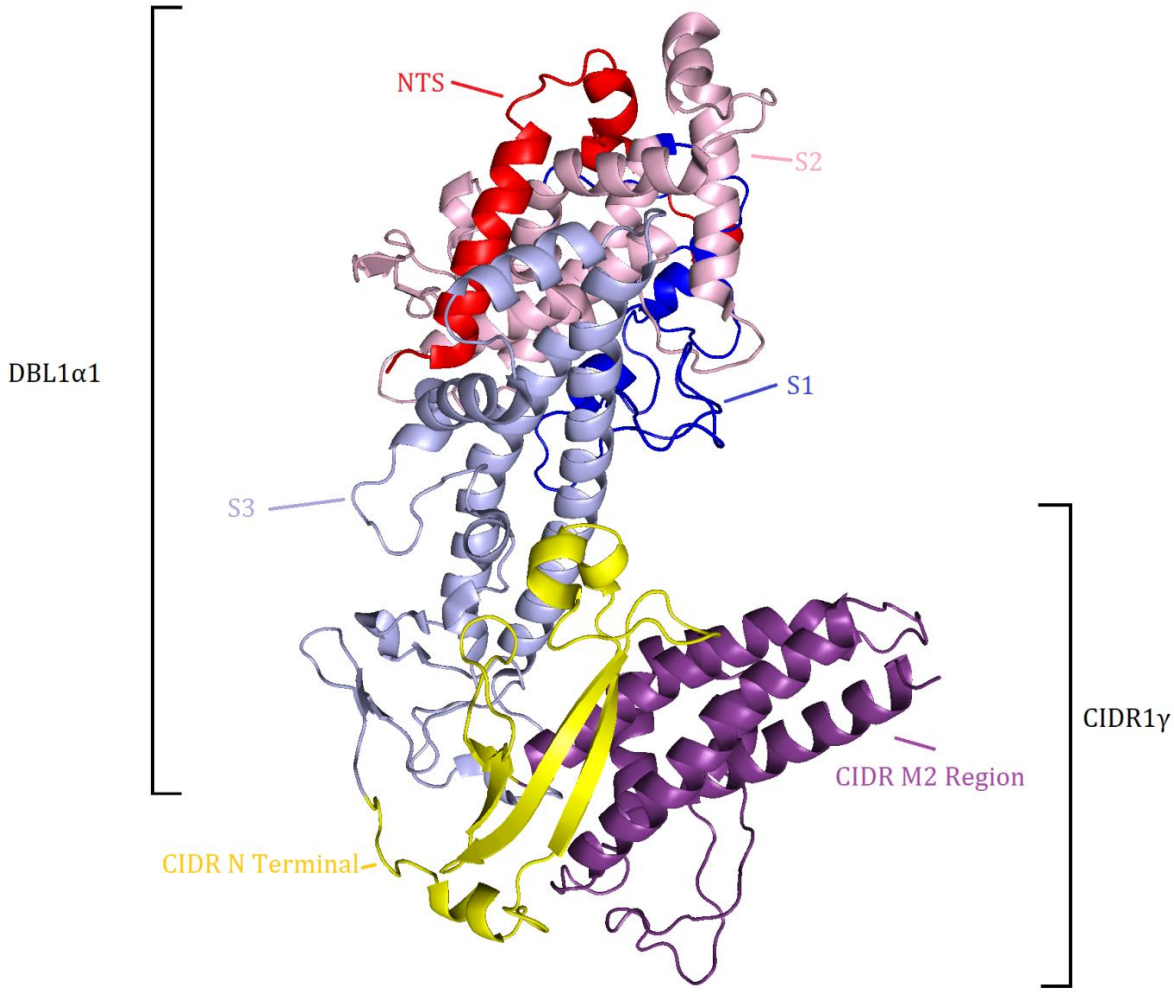


Figure 1.10. Crystal structure of NTS-DBL1α1-CIDRγ domains of VarO variant (PDB accession 2YK0). The NTS (red) is coordinated deep into a cleft with extensive contact with S2 (light pink) and S3 (violet). The CIDR N-terminal region (yellow) connects the CIDR1γ M2 region and allow for the CIDR M2 region to fold back to give a compact structure. S1 of the DBL1α1 is shown in blue.

1.3.5. Higher order arrangement of PfEMP1

Given that PfEMP1 is a modular protein made up of variable combinations of DBL and CIDR domains of different classes, the number of possible PfEMP1 proteins can be large. However PfEMP1 has a semi-conserved head structure[39]. Deeper analysis of PfEMP proteins of various strains of *P. falciparum* elucidated a few combinations of tandem domains that appear to evolve as a functional combination[62,88,89]. Furthermore, the overall arrangement of PfEMP1 can be further classified according to their size, as they can vary between having 2

– 7 DBL domains and 1 – 2 CIDR domains[83]. Full length PfEMP1 proteins can be subdivided into small canonical four domain structures (DBL-CIDR-DBL-CIDR) or larger proteins with complex composition of domains[83,88].

The first domain combination that occurs at a high frequency is the semi-conserved head structure, which consists of a DBL1α in tandem with a CIDR1 of variable class (Figure 1.4)[89]. DBL1α-CIDR1α pair is found in 51 out of 59 of the PfEMP *var* genes in the 3D7 genome[88] and is the canonical head structure of CD36-binding PfEMP1 proteins[73,83]. This conserved feature suggests that N-terminal portion of the PfEMP1, which is most distal from the membrane, is commonly involved in the adhesion trait that is displayed by infected erythrocytes[39].

Another combination of domains described in early studies is the DBLβ and C2 domains (Figure 1.4). C2 domains associate exclusively with DBLβ domains[89]. It has been shown that the DBLβ-C2 tandem domain from group B and C *var* genes is the minimal functional unit required for PfEMP1 binding to ICAM1[109,115]. Interestingly, mutagenesis studies using alanine scanning identified some residues in the first helix of C2 as part of the glycan binding pocket. By sequence alignment, it was observed that the first helix of C2 can act as the third stabilising helix in the three helix bundle for DBLβ as described in the S3 of DBL domains[109]. Indeed C2 is not a discrete domain and instead the DBLβ-C2 tandem is a DBLβ domain[62].

The last combination of domains involves DBL2δ and CIDR2γ or β. This combination is found frequently after the transmembrane region of canonical short four-domain PfEMP1[89,91]. Although a function has not yet been found, the lack of variation in this region

may either suggest a strong selective pressure for, or a lack of selective pressure against, the occurrence of this combination in most groups B and C PfEMP1.

Occurrence of these tandem combinations raises the question of the role of selection pressure for adhesion in retaining repeated features in the presence of immunity mediated selection pressure. Given that PfEMP1 are more likely to have evolved under a combination of different pressures, the non-random combinations intuitively suggest a greater functional significance even in the presence of immune-mediated selection pressure. Moreover, the first two domain combinations described (DBL1α-CIDR1α and DBL2β-C2) are involved in two of the most well defined interactions with human ligands during sequestration. However, it is uncertain whether higher order arrangements of binding modules in the full length PfEMP1 contribute to binding. Although ligand binding might be mapped to a certain minimal region of a domain, it is difficult to quantitatively compare the affinity of PfEMP1 in different domain context to a human ligand. Thus, it is interesting to ask whether higher order arrangements of full length PfEMP1 affect its affinity for human ligands, and whether the suggested minimal region or domain to which the interaction is mapped can account for the whole affinity for their human ligand partner. Binding experiments have shown that the binding affinity between an intact extracellular portion of *var*2CSA and chondroitin sulphate proteoglycan (CPSG) cannot be fully accounted for by single recombinant DBL domains. This suggests that multiple domains appear to be important for coordinating the binding of *var*2CSA to CPSG[53].

1.3.6. Implications of domain and var gene classification in studies of malaria

The classification of *var* genes and various domains of PfEMP1 have implications for the ability to predict PfEMP1 binding attributes to human ligands from the primary sequence of PfEMP1. A wide range of human ligands have been shown to bind PfEMP1 proteins. These include CD36[116], ICAM1[117], CSA[118], Thrombospondin 1 (TSP1)[119,120] and thrombomodulin[121], Platelet endothelial cell adhesion molecule 1 (PECAM/CD31)[122], vascular cell adhesion molecule (VCAM)[122], hyaluronic acid[123], heparin sulphate[124], human natural IgG[125] and IgM[126–129], complement receptor 1 (CR1)[77] and AB blood group antigen[130](Table 1.2).

Domain	Associated Human Ligand	Severe malaria	Mechanism of pathogenesis
CIDRα	CD36	Area specific	Cytoadherence / Rosetting
	IgM	x	Rosetting
	PECAM	x	Cytoadherence
DBLα	Heparan sulphate	x	Cytoadherence / rosetting leading to placental malaria
	CR1	x	Rosetting
	AB antigen	x	Rosetting
DBLβ-C2	ICAM1	✓	Cytoadherence / rosetting leading to cerebral malaria
DBLγ	CSA	x	Cytoadherence leading to placental malaria
DBLδ	PECAM	x	Cytoadherence
DBL6ε	IgM	x	Rosetting
	CSA	✓	Cytoadherence leading to placental malaria
DBL2x	CSA	✓	Cytoadherence leading to placental malaria
DBL3x	CSA	✓	Cytoadherence leading to placental malaria

Table 1.2. Table of PfEMP1 DBL and CIDR domains and their best characterised human ligands. Correlation to severe malaria and intended mechanism of pathogenesis was collated using evidences cited by Rowe *et al.*, 2009[10]. Cerebral and placental malaria is highlighted in red for the combination of PfEMP1 and human ligands related to the respective diseases.

A recent study by Jane *et al*[83] suggests linking adhesion phenotype by sequence and domain classification. Binding assays were performed using a panel of parasite lines that expressed single *var* genes, testing their binding to host receptors to investigate the predictability of adhesion phenotypes from *var* gene and PfEMP1 classification. In their study, interactions with CD36 and ICAM1 could be predicted by domain classification. Other receptors that were tested either failed to bind or rarely bound. The distinction between group A and groups B/C *var* genes was demonstrated by these binding studies as PfEMP1 encoded by group A failed to bind to CD36 as opposed to PfEMP1 encoded by group B or C, hence affirming previous observations[131].

In essence, classification of PfEMP1 and *var* genes has allowed us to do three things. Firstly, we are now able to appreciate the evolution of different PfEMP1 subfamilies as a result of recombination hierarchy, leading to diversity in function amongst them. Secondly, it is possible to locate conserved residues and important structural features that are common to PfEMP1 within similar ancestral families. Thirdly, being able to predict adhesion phenotype and function from these classifications allows us to identify possible vaccine targets. These have opened up opportunities for further functional studies and vaccine development for cross reactive antibody in the bid to reduce malarial occurrence globally.

1.4 General aims

The ability of *P. falciparum* infected erythrocytes to evade splenic destruction by sequestration is an important cause of pathology in severe malaria. PfEMP1 is central to the process of sequestration, through interactions with human receptors. This exposes *Plasmodium* proteins to the immune system, making evasion of immune recognition by antigenic variation necessary. The use of genomic tools and protein classification has allowed us to understand the important domain architecture of PfEMP1. We now understand that certain domains or combination of domains are conserved to bind to different human targets and presumably conserved features are required for interaction with specific ligands. One of best understood ligands is CD36, which has been described as one of the most common targets for PfEMP1 during cytoadherence. However, deeper understanding of the molecular basis of these interactions can be further expanded on by biophysical means.

The long-term aim of my project is to characterise the conserved features on PfEMP1 that are important for interaction with human ligands through structural methods. This will allow us to provide molecular insight to understand the models of binding postulated by genetic and functional studies. The two sets of interactions I am particularly interested in are between CIDR1α domains and CD36, and between DBL domains and human IgM. The work I have described in the following chapters was performed to characterise these interactions biophysically, in an attempt to crystallise these components.

Chapter 2

Materials and methods

2.1 Gene cloning

2.1.1 Polymerase Chain Reaction

Oligonucleotides were synthesised by Sigma-Aldrich and solubilised in water to a concentration of 1 μg/μl. Polymerase Chain reaction (PCR) was conducted using KOD Hotstart DNA polymerase (Novagen). This begins with three minutes denaturation at 95 °C. The PCR cycle then involves denaturation at 95°C for 1 minute, annealing at 50°C for 1 minute, and elongation at 68°C for 1 minute, for 24 cycles, before a final elongation of 68°C for 5 minutes. PCR purification was performed using a QIAquick PCR purification kit (Qiagen) according to manufacturer's protocol.

2.1.2 Restriction Enzyme Digests

DNA restriction digests were conducted using New England Biolabs (NEB) enzymes and buffers at suggested concentrations. Reactions were incubated at 37°C for 3 hours or overnight. 5′ terminal phosphate residues were removed from cleaved vectors via alkaline phosphatase treatment using calf intestinal alkaline phosphastase (Invitrogen), incubated at 37°C for an hour.

2.1.3 *DNA Ligations*

DNA ligations were conducted using T4 DNA ligase (NEB). 1 µl of T4 DNA ligase and 1 µl of T4 DNA ligase reaction buffer were added to make up a 10 µl ligation reaction containing variable ratios of appropriately digested inserts and vectors. Ligation experiments were conducted at 16°C for 4 hours. Ligation was followed by transformation into subcloning competent DH5α cells (NEB).

2.1.4 Transformation of DH5α cells with Plasmid DNA

1 µl of ligated DNA was incubated on ice with 50 µl of subcloning competent DH5α (NEB) cells for 30 minutes. The cells were then heat-shocked for 45 seconds at 42°C. 250µl of Lysogeny broth (LB) medium was then added to the cells and they were incubated for 1 hour in a 37°C shaker incubator before plating on ampicillin selective agar plates (100 µg/ml of ampicillin) and incubating overnight at 37°C.

2.1.5 Plasmid Isolation

DH5α cells were grown in 5 ml of antibiotic selective LB media (ampicillin 100 µg/ml) in a 37°C shaking incubator overnight. Cells were harvested by spinning at 13,000g for 2 minutes, before DNA plasmid was extracted using a QIAprep Spin Miniprep kit (Qiagen), according to the manufacturer's protocol.

2.1.6 Gel Electrophoresis

DNA samples were visualized by gel electrophoresis on a 1% agarose gel in 1 x TBE buffer with SYBR® Safe (Invitrogen, 10 000X). Samples were loaded with Qiagen Gelpilot loading buffer at a 5:1 v/v ratio. Both 1 kb and 100 kb molecular markers (NEB) were used to determine fragment size. Gels were run at 100 V for 30 minutes and DNA was visualized with 302 nm UV-B light.

Gel extraction was performed using a QIAquick Gel extraction kit (Qiagen) according to manufacturer's protocol.

2.1.7 DNA Sequencing

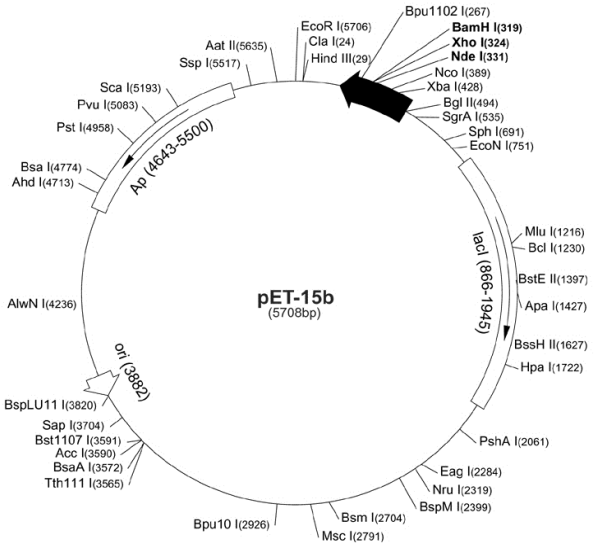
DNA sequencing was performed using the Sanger Sequencing service by Source Bioscience using their T7F primers.

2.1.8 DNA plasmids for Escherichia Coli expression

A converted pET15b vector (pET15bcon) has been used for the expression of recombinant proteins in an *Escherichia coli* expression system. Dr Matthew Higgins had converted the pET15bcon vector to add a His_6-tag and TEV site into the NcoI-BamHI site, together with a multiple cloning site containing BamHI, NheI, XhoI, KpnI, NdeI sites (Figure 2.1). Cloning into the NheI-BamHI site in frame gives a fusion with the His_6-tag and TEV site at the N terminus.

The sequences of the recombinant proteins cloned into pET15bcon for bacterial expression are shown in Appendix A.

CATGGGC*CACCATCACCATCACCAT***GAGAACCTGTATTTTCAGGGC**
GGATCCGCTAGCTCGAGGTACCATATGC



Vector:	vector base	pEt15b
	insertion site	NcoI-BamHI
	resistance	amp

Figure 2.1. Converted pET15bcon vector.

2.2 Protein expression

2.2.1 Bacterial expression in Escherichia coli cells

2.2.1.1 Transformation of Escherichia coli cells with Plasmid DNA

Escherichia coli cells were innoculated from their respective glycerol stock into 5 ml of LB media at 37°C for a few hours to a OD_{600} of 1.0. One millilitre aliquots were centrifuged (13,000g, 1 min, 25°C) using a benchtop microfuge (Eppendorf). LB media was aspirated, and the pellet was resuspended in ice cold 1 ml of $CaCl_2$. This was centrifuged (13,000g, 1 min, 25°C) and the resulting supernatant was aspirated. The pellet containing competent *E. coli* cells was resuspended with 50 µl of ice cold $CaCl_2$ for transformation.

1 µl of ligated DNA was incubated on ice with 50 µl of competent *E. coli* cells for 1 hour. The cells were then heat-shocked for 45 seconds at 42°C. 250µl of Lysogeny broth (LB) medium was then added to the cells and they were incubated for 1 hour in a 37°C shaker incubator before plating on antibiotic selective agar plates (100 µg/ml of Ampicillin and 35 µg/ml of Chloramphenicol) and incubating overnight at 37°C.

2.2.1.2 Expression trials

Competent *E. coli* cells were transformed with the pET15bcon vectors containing various constructs and grown on an ampicillin selective plate (100 µg/ml of Ampicillin and 35 µg/ml of Chloramphenicol) overnight at 37°C. A single colony was picked and grown in 10 ml of LB media in a 37°C shaker to an OD_{600} of 1.0. Expression was then induced by addition of IPTG to a final concentration of 1 mM. Induced cells were transferred to shaking incubators

at various temperatures (18°C, 25°C and 30°C). 1 ml of cells for each condition was collected after a 3 hour interval and after overnight induction.

Harvested cells were centrifuged (13,000g, 1 min, 25°C) in a benchtop microfuge (Eppedorf). The supernatant was aspirated and the cell pellet resuspended in 1 ml of phosphate buffered saline (PBS) (0.01 M Phosphate pH 7.4, 0.0027 M KCl, 0.137 M NaCl)(Fisher Scientific). The cells were sonicated (5 pulses at output 4.0 for 2 seconds each with 5 second intervals between each pulse) using a Misonix S4000 (Qsonica). A 50 µl sample was taken of whole cell extract before centrifugation (13,000g for 1 minute) and of the soluble fraction after centrifugation. The samples for each condition were subjected to Western blot detection for the presence of His_6-tagged recombinant CIDRα protein.

2.2.1.3 Expression of recombinant proteins in Escherichia coli

Transformed cells were plated on antibiotic selective agar plates (100 µg/ml of Ampicillin and 35 µg/ml of Chloramphenicol) and incubated at 37°C overnight. One colony was picked and inoculated in 100 ml of antibiotic selective LB media (100 µg/ml of Ampicillin and 35 µg/ml of Chloramphenicol) at 37°C overnight. 10 ml of overnight culture was inoculated into each 1 L of antibiotic selective LB media (100 µg/ml of Ampicillin and 35 µg/ml of Chloramphenicol). These were left in a 37°C shaker incubator (220 RPM). IPTG was added to the cell culture to a final concentration of 1 mM for induction of protein expression when the OD600 of the cells reaches 1.0. The cell cultures were then incubated at various temperatures (25°C or 30°C) in a shaking incubator (220 RPM). Cells were harvested after either three hours or overnight by centrifugation (6000g, 20 minutes, 4°C). Cell pellets were stored at -80°C or used for protein purification.

2.2.2 Baculovirus-mediated insect cell expression

2.2.2.1 Co - Transfection of DNA.

DNA encoding the recombinant protein of interest was previously cloned into a converted transfer vector pAcGP67A by Dr Matthew Higgins. The converted pAcGP67A contains an N-terminal His_6-tag, TEV cleavage site and a multiple cloning site (MCS). Co-transfection of sterilized recombinant DNA plasmid into SF9 (*Spodoptera frugiperda*) cells (Invitrogen) was performed using a BD Baculogold transfection Kit (BD Biosciences), according to manufacturer's protocol. Supernatant consisting of viral particles were used for plaque assays.

All steps were performed in sterile conditions.

2.2.2.2 Plaque assays

1.5 x 10^6 SF9 cells from a log phase culture were seeded into each sterile 60 mm tissue culture dish with 2 ml of Sf-900 II SFM media (Invitrogen). These were left to incubate at room temperature for 15 minutes. Serial ten-fold dilutions (10^{-3} to 10^{-5}) with volume of 1 ml were made from a viral supernatant. These were added into each tissue culture dish and incubated for 60 minutes at room temperature. A 1% agarose mixture was prepared using 4% Plaque assay agarose (BD Biosciences) with 1.3X SF900II SFM media (Gibco) and was incubated in a 37°C water bath. After 60 minutes incubation, the viral inoculated medium was aspirated and replaced with the agarose mixture. Agarose plates were allowed to set and incubated in a 27°C incubator for 6 days. Plaques were identified and picked under a Nikon Diapshot Microscope (Nikon). An agarose plug directly above the plaque was removed using a pipette tip and placed in 1 ml of Sf-900 II SFM media and incubated at 4°C overnight.

2.2.2.3 Amplification of baculoviruses with DNA insert of recombinant protein

Sequential amplification of the DNA insert of the protein of interest was performed to obtain a high-titre viral stock.

P1 viral stocks were prepared with 1.5 x 10^6 cells in 5 ml of Sf-900 II SFM media using a sterile T25cm cell culture flask (Corning). The cells were inoculated with 1 ml of viral stock from the plaque assays and incubated in a 27°C incubator for 6 days. Viral supernatants were harvested by centrifugation at 1000g for 10 minutes at room temperature.

P2 viral stocks were prepared with 5 x 10^6 cells in 10 ml of Sf-900 II SFM media using a sterile T75cm cell culture flask (Corning). The cells were inoculated with 0.25 ml of each P1 viral stock and incubated in a 27°C incubator for 6 days. Viral supernatant was harvested by centrifugation at 1000g for 10 minutes at room temperature.

P3 viral stocks were prepared using 5 X 10^7 cells in 500 ml of Sf-900 II SFM media in 2 L cell culture roller bottles (Corning). Cells were inoculated with 1 ml of P2 viral stocks (about 0.1-0.2 pfu/cell) and incubated in a 27°C shaking incubator for 6 days. Cell supernatant was harvested by centrifugation at 1000g for 10 minutes at room temperature. Viral supernatants were stored at 4°C.

2.2.2.4 Expression Trials of recombinant protein expression from viral stock

Expression trials were performed by pipetting 100 µl of High-Five (*Trichoplusia ni*) cells in logarithmic growth phase (from a stock of concentration 10^6 cells/ml) (Invitrogen) into required number of wells on a 96 well plate. Cells were inoculated with 50 µl of each viral stock. Subsequently, 50 µl of Express Five SFM media (Invitrogen) was added to each well.

The 96 well plate was incubated in a 27°C incubator for 4 days. Cell supernatants were then harvested by microcentrifugation (13,000g, 1 minute) using a benchtop microfuge (Eppendorf) before a Western blot was conducted to detect the His_6-tag recombinant protein.

2.2.2.5 Expression of recombinant protein of interest in High-Five cells

High throughput expression of soluble recombinant protein was performed in High-Five (*Trichopulsia ni*) cells using P3 viral stocks. Cells were inoculated with P3 viral stock at a ratio of 0.2 pfu/cell (50 ml of P3 viral supernatant with 500 ml of High-Five cells at 10^6 cells/ml). These were incubated in a 27°C shaking incubator for 4 days. Cell supernatant containing soluble protein was harvested by centrifugation (6000g, 20 minutes, 4°C).

2.2.3 Mammalian cell expression system

2.2.3.1 Preparation of DNA of interest for mammalian cell transfection

The gene of the protein of interest was previously cloned into a pFUSE vetor. This was transformed into DH5α and selected on a low salt agar plate containing 25 μg/ml zeocin during a 37°C overnight incubation. A 100 ml culture of transformed DH5α in zeocin selection low salt LB media was incubated overnight in a 37°C shaking incubator (220 RPM). 10 ml of the overnight culture was used to inoculate each litre of zeocin containing low salt LB media, and grown overnight at 37°C in a shaking incubator (220 RPM). Cells were harvested by centrifugation (6000g, 15 min, 4°C). DNA purification was carried out using a QIAGEN Plasmid Giga Kit (Qiagen) according to manufacturer's protocol. Ethanol precipitation of the DNA was performed under sterile conditions in the fumehood. The DNA was then resuspended in endotoxin-free water to a concentration of ~2.5 mg/ml.

2.2.3.2 Preparation of media for HEK293T cells expression

Serum activated Dulbecco's Modified Eagle Medium (DMEM) was used to passage HEK293T cells. Serum activated DMEM (SAM) was prepared by adding 55 ml of Hyclone fetal bovine serum (FBS) (Thermo scientific) and 5 ml of Penicillin-Streptomycin 100X mixture (Sigma-Aldrich) to 500 ml of high glucose DMEM with L glutamine and sodium pyruvate (PAA). Serum free DMEM (SFM) was prepared in the same way, without addition of FBS. All HEK293T cells were grown at 5% CO_2 in a 37°C incubator (Sanyo).

2.2.3.3 Expression trials for recombinant protein in HEK293T cells

Expression trials in 6 well dishes (Corning) were conducted before doing large scale roller bottle expression. 1 x 10^5 HEK293T cells were plated in a single well of the dish and resuspended in 2 ml of SAM. Plated cells were incubated at 37°C overnight to allow the cells to reach 50% – 80% confuency before transfection using FuGENE® 6 Transfection Reagent (Promega). The DNA transfection mixture was made using 3 µl of FuGENE® 6 Transfection Reagent, 1 µl of pFUSE vector with DNA insert of interest (1 µg/µl) in 97 µl of SF900II SFM media. The mixture was added dropwise to cells after a 15 minute incubation at room temperature. Cells were then incubated at 37°C and the 50 µl of the supernatant were collected at different time points for Western blot detection for expression.

For expression trials involving SFM, the SAM was aspirated by pipetting and replaced with 2 ml of SFM immediately before transfection.

2.2.3.4 Roller bottle expression of recombinant proteins in HEK293T cells

Protein was expressed at large scale in roller bottles. Adherent HEK293T cells were prepared in T75 cell culture flask (corning). When cells in a T75 flask reach 80% confluence (7.5 x

10^5 cells/ml) after incubation at 37°C, the cells were trypsinized using 5 ml of Trypsin-EDTA solution (PAA) in Dulbecco's phosphate-buffered saline (Invitrogen) and transferred into a roller bottle in 200 ml of fresh SAM. The roller bottles were incubated in a 37°C rolling incubator (2 RPM)(Wheaton) for 3 days before transfection with sterile DNA.

For transfection of ten bottles, 5 mg of sterile DNA (0.5 mg per roller bottle) was added to 500 ml of high glucose DMEM without penicillin-streptomycin mixture or FBS. 10 ml of linear polyethlenimine (PEI) at a concentration of 1 mg/ml was added as a transfection reagent [132] (1 mg of linear PEI required per roller bottle). This was thoroughly vortexed and incubated at room temperature for 10 minutes.

Media in the roller bottles was replaced with 200 ml of SFM before transfection. 50 ml of the DNA mixture was added to each roller bottle. The roller bottles were incubated in a 37°C rolling incubator (2 RPM) for 5 days before harvesting the supernatant by centrifugation.

2.3 Protein purification

2.3.1 Sonication of bacterial cells

Cell pellets were first resuspended in 30 ml of Ni-NTA binding buffer (20 mM Tris pH 8.0, 0.3 m NaCl, 0.5% Triton-X 100, 10 mM Imidazole). Cell suspension was sonicated (Output 5.0 V, 4 cycles of 15 seconds pulse with 45 seconds of rest between each cycle) using a Misonix S4000 (Qsonica). Cell debris containing inclusion bodies was separated by centrifugation (45,000g, 20 min, 4°C) and cell supernatant containing soluble protein of interest was collected. Insoluble proteins in the inclusion bodies were purified by a refolding

on Ni-NTA column protocol while soluble protein of interest was purified using a normal Ni-NTA affinity chromatography.

2.3.2 Buffer exchange of cell supernatant

Cell supernatants from insect cell or HEK293 cell culture containing soluble protein were harvested by centrifugation (10,000g, 20 min, 4°C). The collected supernatants were filtered using a Steritop-GP 0.22 μM filter unit (Millipore). Cell supernatant was buffer exchanged and concentrated using Masterflex® I/P® Precision Brushless Drive tangential flow machine (Cole Parmer). The Centramate™ PE Lab tangential flow systems fitted with the Omega™ 5 kDa PES membrane (Pall) was first equilibrated with 2 litres required buffer, before pumping the supernatant through at ~5ml/min. Protein samples were collected after a 0.5 L wash with required buffer.

2.3.3 Dialysis of cell supernatant

Dialysis of cell supernatant is performed using a SnakeSkin dialysis tubing (10 kDa MWCO) (Thermoscientific). This was performed usually against 2 – 4 L of buffer for a variable amount of time (2 hours to overnight) at either room temperature or 4°C. All buffers were stirred using a magnetic stirrer.

2.3.4 Buffer exchange using PD-10 Desalting columns

PD-10 Desalting columns (GE-Healthcare) were used to buffer exchange small volumes of proteins. Columns were first equilibrated with 25 ml of the buffer of interest. 2.5 ml of protein sample is then applied are then eluted using 3.5 ml of buffer.

2.3.5 *Ni-NTA affinity chromatography for His_6-tag soluble proteins*

Bacterial expression system

Ni-NTA agarose beads (Qiagen) were first equilibrated with 3 columns Ni-NTA binding buffer. Harvested cell supernatant containing His_6-tag soluble protein was then incubated with Ni-NTA agarose beads (Qiagen) for 10 minutes. Ni-NTA agarose beads were washed with 1 L of Ni-NTA binding buffer, followed by 1 L of Ni-NTA wash buffer (20 mM Tris pH 8.0, 0.5 M mM NaCl, 10 mM imidazole) before eluting with Ni-NTA elution buffer (20 mM Tris pH 8.0, 300 mM NaCl, 200 mM imidazole) into 10 ml elution fractions.

Baculovirus insect cell expression system

Buffer exchanged cell supernatant (20 mM Tris pH 8.0, 0.3 M NaCl) was incubated with Ni-NTA agarose beads (Qiagen) overnight at 4°C. This is followed by the same wash and elution protocol as above.

2.3.6 IgM affinity chromatography

Supernatants containing hIgM-Fc were buffer exchanged into PBS (Fisher Scientific). Protein sample in PBS was then applied to an affinity column containing Anti-Human IgM

(μ-chain specific) − Agarose antibody (Sigma-Aldrich). The agarose beads were equilibrated with 3 column volumes of IgM binding buffer (0.01 M sodium phosphate buffer pH 7.2, 0.5 M NaCl) before applying the protein samples. The flow through was reapplied twice. Beads were then washed with 0.5 L of IgM binding buffer. hIgM-Fc was then eluted using an IgM elution buffer (0.5 M acetic acid pH 2.4, 0.15 M NaCl) into 10 ml fractions. 1 ml of 1 M Tris pH 8.0 buffer was immediately added to each of the fractions and the pH was corrected to pH 8.0 using 5 M NaOH.

2.3.7 Refolding on Ni-NTA affinity column

Inclusion bodies containing insoluble proteins were resolubilised in a Ni-NTA refolding buffer (20 mM Tris pH 8.0, 6M Guanidine HCl, 0.3 M NaCl, 10 mM β-mercaptoethanol) by dounce homogenization and sonication (Output 5.0 V, 4 cycles of 15 seconds pulse with 45 seconds of rest between each cycle). Resuspended inclusion bodies were subjected to centrifugation (45,000g, 20 min, 4°C). Supernatant containing denatured protein was incubated with Ni-NTA agarose beads (Qiagen) for 10 minutes at room temperature. Refolding was performed on the Ni-NTA column using a Ni-NTA refolding wash buffer 1 (20 mM Tris pH 8.0, 6 M Guanidine HCl, 0.3 M NaCl, 10 mM β-mercaptoethanol, 0.3 mM oxidized gluthathione, 3 mM reduced gluthathione). The amount of guanindine HCl in the wash buffer was gradually decreased with subsequent washes. A concentration series of wash buffers containing 6M, 4.8M, 3.6M, 2.4M and 1.2M of guanidine HCL was obtained by diluting Ni-NTA refolding wash buffer 1 with Ni-NTA refolding wash buffer 2 that contains no guanidine (20 mM Tris pH 8.0, 0.3 M NaCl, 10 mM β-mercaptoethanol, 0.3 mM oxidized gluthathione, 3 mM reduced gluthathione). The final wash step was performed with 100 ml of Ni-NTA refolding wash buffer 2, before eluting with Ni-NTA elution buffer.

2.3.8 TEV cleavage

Elution fractions were buffer exchanged into PBS using PD-10 Desalting columns (GE Healthcare). 0.3 mM oxidized gluthathione and 3 mM reduced gluthathione were added. TEV enzyme (OD_{280} = 1.0) was added at a ratio of 1:25 (TEV enzyme:recombinant protein) w/w and incubated overnight at 4°C. This solution was subjected to Ni-NTA affinity chromatography to remove the His_6-tagged TEV enzymes (elute) and uncleaved protein from the cleaved protein of interest (flowthrough).

2.3.9 Deglycosylation of glycoproteins from baculovirus-infected insect cell expression

Deglycosylation of proteins expressed in baculovirus-infected insect cell expression system was performed using endoglycosidase (Endo) H_f (NEB) and $EndoF_3$ (made by Dr Matthew Higgins). Protein samples were first adjusted to pH 5.6 using 1 M Sodium Acetate pH 4.0 before addition of 5000 units (1 000 000 units/ml) of $EndoH_f$ and $EndoF_3$ (1.2 mg/ml) at 1:100 (w/w ratio) at 37°C. Sample was dialysed overnight in 4 L of 20 mM Tris pH 8.0, 0.15 M NaCl at 4°C.

2.3.10 Preparative size-exclusion chromatography

Samples were first concentrated to 2 ml using Amicon Ultra-15 Centrifugal Filter Unit (Millipore) with an appropriate molecular weight cut off. These were then centrifuged (13,000g , 5 min, 25°C) to remove aggregates. The samples were then subjected to preparative size exclusion chromatography using Superdex 75 HR 16/60 Gel Filtration column (GE Healthcare) or Superdex 200 HR 16/60 Gel Filtration column (GE Healthcare)

using an ÄKTApurifier (GE healthcare) FC 950. The 2 ml samples were injected and the column was run at 1 ml/min using 20 mM Tris pH 8.0, 0.15 M NaCl.

2.4 Protein detection and characterisation techniques

2.4.1 SDS Polyacrylamide gel electrophoresis (SDS-PAGE)

Samples from protein purifications were mixed with 5x protein loading buffer (50 mM Tris pH 6.8, 10% Glycerol, 2% SDS, 5% ß-mercaptoethanol, 0.02% bromophenol blue) and run on a 12% SDS polyacrylamide gel using SDS-PAGE running buffer (30 mM Tris pH7.5, 192 mM glycine, 0.1% SDS). Proteins were stained with commassie stain (25% Ethanol, 10% acetic acid, 0.5% Brilliant blue) and destained with destain buffer (6% Ethanol, 6% acetic acid). Protein sizes were compared with the markers found in PageRuler Prestained Protein Ladder (Fermantas).

2.4.2 Detection using Western blotting

Proteins with His_6-tag were subjected to SDS-PAGE as above. Proteins were transferred onto a PVDF Immobilon-P membrane (Millipore) using semi-dry blotting buffer (25 mM Tris pH 7.5, 192 mM glycine, 20% Ethanol, 0.1% SDS) and a semi-dry blotting machine (Bio-Rad) run at a constant current (50 mA) for 1 hour. Membranes were blocked for 1 hour at room temperature using either 20 ml of 5% (w/v) milk in wash buffer (20 mM Tris pH 7.5, 0.5 M NaCl, 0.05% Tween-20, 0.2% Triton X-100), or using 20 ml of commercial blocking reagent (Qiagen), on a rocking table. Membranes were washed 3 times for 10 minutes each with

wash buffer before incubating with HRP-conjugated antibodies for 1 hour at room temperature. Next, the membranes were washed 3 times for 10 minutes each with wash buffer. Chemisluminescent detection was performed using ECL Plus Western blotting Detection Reagents (Amersham) and Hyperfilm ECL (Amersham), according to manufacturer's protocol and required exposure timing in a dark room.

Antibodies required for:

1) IgM Detection:

Anti-Human IgM (μ-chain specific) −Peroxidase antibody (Sigma)

Concentration: 1/10 000 made up in wash buffer (20 mM Tris pH 7.5, 0.5 M NaCl, 0.05% Tween-20, 0.2% Triton X-100).

2) His_6-tagged protein detection:

Penta-His HRP Conjugate (Qiagen)

Concentration: 1/2000 made up in blocking buffer made according to manufacturer protocol (Qiagen)

2.4.3 Circular dichroism

300 μl of proteins at a concentration of 0.1 mg/ml were prepared by buffer exchange into CD buffer (20 mM Tris pH 8.0, 0.15 M NaF) using a PD-10 desalting column. The samples were pipetted into a Starna Scientific quartz cuvette with a light path of 1 mm and studied using a Jasco J-720 CD Spectrometer. The ellipicity (mDeg) of the samples were measured from 195 – 260 nm wavelength at 25°C. Data was taken as an average of ten readings with a 30 second interval between each cycle. Averaged data was buffer corrected by subtraction of

averaged measurement of a buffer blank and analysed by the K2D3 algorithm[133] (http://www.ogic.ca/projects/k2d3/) to predict the proportion of secondary structures within the sample by comparison with known structures in the PDB. The mean residue ellipticity (MRE) for the spectrums was calculated according to the equations that takes into account the mean residue weight (MRW) of the protein sample[134],

$$[\theta]_{mrw_1\lambda} = \frac{MRW \times \theta_\lambda}{10 \times d \times c}$$

where θ_λ is the measured ellipticity (deg) at wavelength λ, *d* is the pathlength (cm), *c* is the concentration of the sample (g/ml), and $MRW = \frac{M}{(N-1)}$, where M is the molecular mass (Da) and (N – 1) is the number of peptide bond within the protein.

2.4.4 Fluorescence-based thermal shift assays

Thermal shift assays were conducted with a Strategene MX3005P quantitative PCR machine (Agilent Technologies) in a 96 well plate format. Proteins were at 10 μM concentration in 20 mM Tris pH 8.0, 150 mM NaCl. SYPRO® Orange (Invitrogen) was added from a 50X stock solution (Invitrogen). The emission fluorescence wavelength of SYPRO® Orange at 610 nm was measured over a range of temperatures from 25°C to 80°C, with a temperature ramp rate of 1°C per 90 seconds. Buffer blanks were subtracted during data processing using GraphPad Prism 5.0.

2.5 Biophysical methods for interaction studies

2.5.1 Ni-NTA Pull down assay

Cell pellets containing soluble His_6-tagged proteins were harvested after centrifugation (10,000g, 4°C, 20 min) and resuspended in 5 ml of Ni-NTA binding buffer. The cell pellets were sonicated (5 pulses at output 4.0 for 2 second each with 5 second interval in between each pulse) using a Misonix S4000 (Qsonica). The resulting mixture was centrifuged (13,000g, 2 min) using a benchtop microfuge (Eppendorf).

100 µl of Ni-NTA beads were washed three times with binding buffer. Supernatants containing His_6-tagged DBL domains were incubated with Ni-NTA beads for 30 minutes at room temperature. After 30 minute incubation, the beads were washed three times with Ni-NTA wash buffer. The binding partner was then applied to the Ni-NTA mixture and incubated for 30 minutes at room temperature. The beads were washed three times with Ni-NTA wash buffer before being resuspended in 50 µl of protein loading buffer and detected by Western blot. Positive and negative controls were performed according to descriptions in their individual chapters.

2.5.2 *Surface Plasmon resonance*

Surface plasmon resonance (SPR) was performed using a Biacore T100 system (GE Healthcare). All experiments were performed on a Series S sensor ship CM5 (GE Healthcare).

2.5.2.1 Ligand immobilisation

The CM5 sensor chip has four flow cells. Flow cell 1 was used as a reference without any immobilised ligands. All ligand samples were buffer exchanged into the Biacore buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 50 µM ethylenediaminetetraacetic acid, 0.05 % Tween 20 (v/v)) using PD-10 Desalting columns (GE Healthcare). The follow steps were all performed at a flow rate of 5 µl/min.

The Biacore T100 was first equilibrated by priming with Biacore buffer before immobilising the ligand on the CM5 chip using amine coupling as suggested by the manufacturer. All flow cells were subjected to 1:1 mixture of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide and N-hydroxysuccinimide (Biacore) for 420 seconds. Ligands were immobilised on the chip surface using 10 mM Sodium Acetate pH 5.0 until the response increased by 5000 RU. The reactive groups were then blocked by treating with ethanolamine (1 M Ethanolamine – HCl pH 8.5) for 420 s.

2.5.2.2 Continuous flow experiment using analyte

The analyte proteins were equilibrated using PD-10 desalting columns (GE Healthcare) into the biacore buffer before being concentrated for analysis. Ten 200 µl two-fold serial dilutions were made of each protein from their top concentration to form a concentration series. Top concentrations of different proteins were altered accordingly in order to obtain a saturated response according to their affinity.

Each concentration series were subjected to a continuous flow experiment using the following method. Firstly, 120 µl of each analyte at the lowest concentration were injected over each of the flow cells at a continuous flow (30 µl/min) at 25°C in the flow cell chamber

for the association phase. This was followed by a dissociation phase, in which 120 µl of Biacore buffer was injected to allow dissociation of the analyte from the ligand. The chip was then regenerated by injecting 30 µl of 4M NaCl over flow cells to recover the ligand binding sites for reproducible binding during subsequent cycles. Ten such cycles were performed to study the response obtained from a two-fold dilution series. A buffer injection was performed at the end of each concentration series to obtain a blank. The resulting response curves from the specific interaction between the analyte and immobilised ligand was calculated by a subtraction of the measured reference response (flow cell 1) from the measured experimental response (flow cells 2, 3, 4).

2.5.2.3 Data analysis

Response data was analysed using the BIAevaluation software (Biacore). Response data was first reference subtracted using data from flow cell 1. Any buffer effects were corrected by blank subtraction. The BIAevaluation software was used to obtain the maximum response for each concentration at steady state. These maximum responses were plotted against protein concentration to give a saturated curve. Curves were then fitted with a steady state 1:1 binding affinity model.

$$R_{eq} = \frac{CR_{max}}{K_D + C} + RI$$

This equilibrium binding model is defined by a plot of R_{eq} (response at equilibrium) against time. At high concentrations of ligands where $R_{eq} = R_{max}$ (maximum response), the binding sites are assumed to be fully saturated in a 1:1 equilibrium binding model. The dissociation constant (K_D) is defined by the concentration of free analyte at which 50% of the total amount of immobilised ligand is in complex with the analyte. Thus, in a 1:1 binding model, the K_D is the analyte concentration at which $R_{eq} = 0.5\ R_{max}$. The K_D values presented in the

analysis below were fitted using this model in the Biacore evaluation software. RI was defined as the bulk refractive index contribution which was derived during the fitting process. Further association and dissociation kinetics were analysed using the respective models explained in further in their respective sections via the BIAevaluation software.

2.5.3 Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was performed using 50 µl of protein samples at a concentration of 1 mg/ml. Protein samples were loaded onto Superdex 200 10/300 GL (GE Healthcare) or Superose 6 PC 3.2/30 columns (GE healthcare) at 0.25 ml/min and 0.10 ml/min respectively at 25°C using an ÄKTApurifier (GE healthcare) FC 950. Protein levels were estimated by UV measurements at a wavelength of 280 nm.

2.5.4 Analytical ultracentrifugation

Analytical ultracentrifugation was performed using a Beckman Optima XL-I analytical ultracentrifuge. A 4-hole rotor was used, with one position used to house the calibration cell.

2.5.4.1 Sedimentation Velocity

For sedimentation velocity experiments, 400 µl of each protein sample (0.2 mg/ml) in 20 mM Tris pH 8.0, 150 mM NaCl was loaded into the sample cell of an aluminium cell housing with sapphire mirrors. Buffer was loaded into the reference cell.

For sedimentation velocity readings, a method scan using the Proteomelab software was performed at a speed of 40,000 RPM at 20°C. 200 scans were made with 3 min in between each scan. Both absorbance and interference readings were measured. The absorbance reading with an *Rmin* of 6 and *Rmax* of 7.2 was measured at 280 nm. Velocity interference data was also measured with 22 pixels per fringe with the internal radii of the rotor being at 5.992 inches.

Data was analysed using the Sedfit program. Protein partial specific volume, buffer density and viscosity were estimated using Sednterp. Data from sedimentation velocity were analysed using the Lamm equation distributions *c(s)* on Sedfit[135–137].

2.5.4.2 Sedimentation Equilibrium

For sedimentation equilibrium experiments, 100 µl of each protein sample (0.2 mg/ml) in 20 mM Tris pH 8.0, 150 mM NaCl were loaded into the sample cell of an aluminium cell housing with sapphire mirrors. Buffer was loaded into the reference cell.

A method scan using Proteomelab software was performed with measurements made at each of 10,000, 12,000, 14,000, 16,000 and 18,000 RPM. Both interference and absorbance reading were made with the same parameters as for sedimentation velocity experiments. Two scans were made at each speed, with one measurement after 18 hours at each speed, and another after 2 hours. Each scan was performed at a 0.0001 cm/s increment. Analysis was performed using Origin 6.0 (Optima XL-A/XL-I data analysis software, Beckman).

2.6 Co-crystallisation of hCD36ED-MC179 complex

2.6.1 Preparation of hCD36ED-MC179 complex.

The hCD36ED-MC179 complex was prepared by first incubating 1 mg of deglycosylated hCD36ED with excess MC179 for 30 minutes at room temperature. The sample subjected to preparative size-exclusion chromatography using a Superdex 200 HR 16/60 Gel Filtration column (GE Healthcare). The elution fractions were concentrated to 8.79 mg/ml using Vivaspin® 500 Centrifugal Concentrator (Sartorius stedim). The sample was then centrifuged (13,000g, 10 min, 4°C) on a benchtop centrifuge (Eppendorf) to remove aggregation before putting them on the crystal plates.

2.6.2 Crystallisation trials

Crystallisation trials were conducted by sitting drop crystallisation using reagents from ProPlex™ (Molecular Dimensions Ltd.), Morpheus™ (Molecular Dimensions Ltd.), JCSG-plus™ (Molecular Dimensions Ltd.), and PACT premier™ (Molecular Dimensions Ltd.). 50 µl of each of the 96 crystallisation reagent for a single screen was pipetted onto single reservoir on at MRC 96 well Crystallization Plate ™ UV (Molecular Dimensions Ltd.). 100 nl of the protein samples were dispensed by the Mosquito® (TTP Labtech) robot in crystallisation well A of the same MRC 96 well Crystallization Plate ™ UV. 100 nl of each crystallisation reagent present in each reservoir was dispensed by the Mosquito® robot onto each of the protein sample. Crystallisation trays were stored at 4°C and monitored electronically over 2 weeks using the CrystalMation (Rigaku) crystallisation system.

Chapter 3

Biophysical characterisation of the interaction between human CD36 and PfEMP1 proteins

3.1. CD36 and severe malaria

CD36 is a human scavenger receptor first found as a glycoprotein on platelets that interact with TSP1[138,139]. CD36 has also been found on a variety of cells such as macrophages[140], endothelial cells[141] and smooth muscle cells[142]. It interacts with many different partners, including long chain fatty acids[142,143], low density lipoproteins[140] (LDL), phospholipids[144], collagen[78] and apoptotic cells. A myriad of functions have been ascribed to CD36, the best studied being its role in glucose and lipid metabolism[145,146], and its role as a scavenger receptor on macrophages in the recognition of apoptotic cells[140,147]. Other normal functions of CD36 includes roles in modulating homeostasis, immunosurveillance, inflammatory responses and maintaining retinal integrity[148].

Intriguingly, knockouts in animal models has linked CD36 abnormalities to diseases including vascular atherosclerosis, diabetes, angiogenesis, hypertension and cardiomyopathy[149–152]. CD36 has also been identified as a receptor for adhesion to *P. falciparum* infected erythrocytes and has been widely linked to pathogenesis during development of severe *P. falciparum* malaria[116,153].

3.1.1. Properties and features of human CD36

3.1.1.1. Gene structure of CD36

The gene encoding human CD36 is found on chromosome 7[154] and contains 15 exons[155]. Mature CD36 is encoded from exon 3 up to exon 14 from N to C terminus[156]. Interestingly many isoforms of CD36 have been found in different cell types, and each of the different isoforms have a different estimated molecular weight[141]. The presence of various functional forms of CD36 has been attributed to the fact that the pre-mRNA of the gene is subjected to alternative splicing[157–159].

3.1.1.2. Protein structure predictions from primary sequence of human CD36

Despite functional characterisation of CD36, a crystal structure of CD36 has been elusive. A cartoon structure of human CD36 (Figure 3.1) has been modelled based on the amino acid sequence of placental CD36[141,150] (SwissProt: P16671). CD36 contains 471 amino acids, with two predicted transmembrane helices. A central extracellular ectodomain lies between these helices, leaving two short intracellular cytoplasmic amino acid sequences at each end[141,160] (Figure 3.1). CD36 is a member of a family of proteins with conserved sequence, structure and function that include lysosomal integral membrane protein II[161] (LIMPII), murine CD36, and CD36 LIMPII analogous-1[159] (CLA-1). Several patches of hydrophobicity are postulated by sequence alignment to be membrane associated in both human and rat CD36[141,162]. CD36 is therefore predicted to be a ditopic protein with two embedded regions at the amino and carboxy ends[163].

Human CD36 contains ten cysteine residues with six in the ectodomain. These extracellular cysteine residues are located in a proline-rich region located in the N-terminal half of the CD36 ectodomain. The remaining four cysteine residues are distributed evenly between the two cytoplasmic fragments[141,150]. Rasmussen and colleagues have mapped disulphide bridges between distant pairs of cysteine residues (C271-C332, C312-C321, C242-C310) in the extracellular domain of the human CD36 by performing matrix-assisted-laser-desorption ionisation/time of flight mass spectrometry (MALDI-TOF MS) using the conserved bovine CD36 as a model[164]. The intracellular cysteine residues are palmitoylated[165].

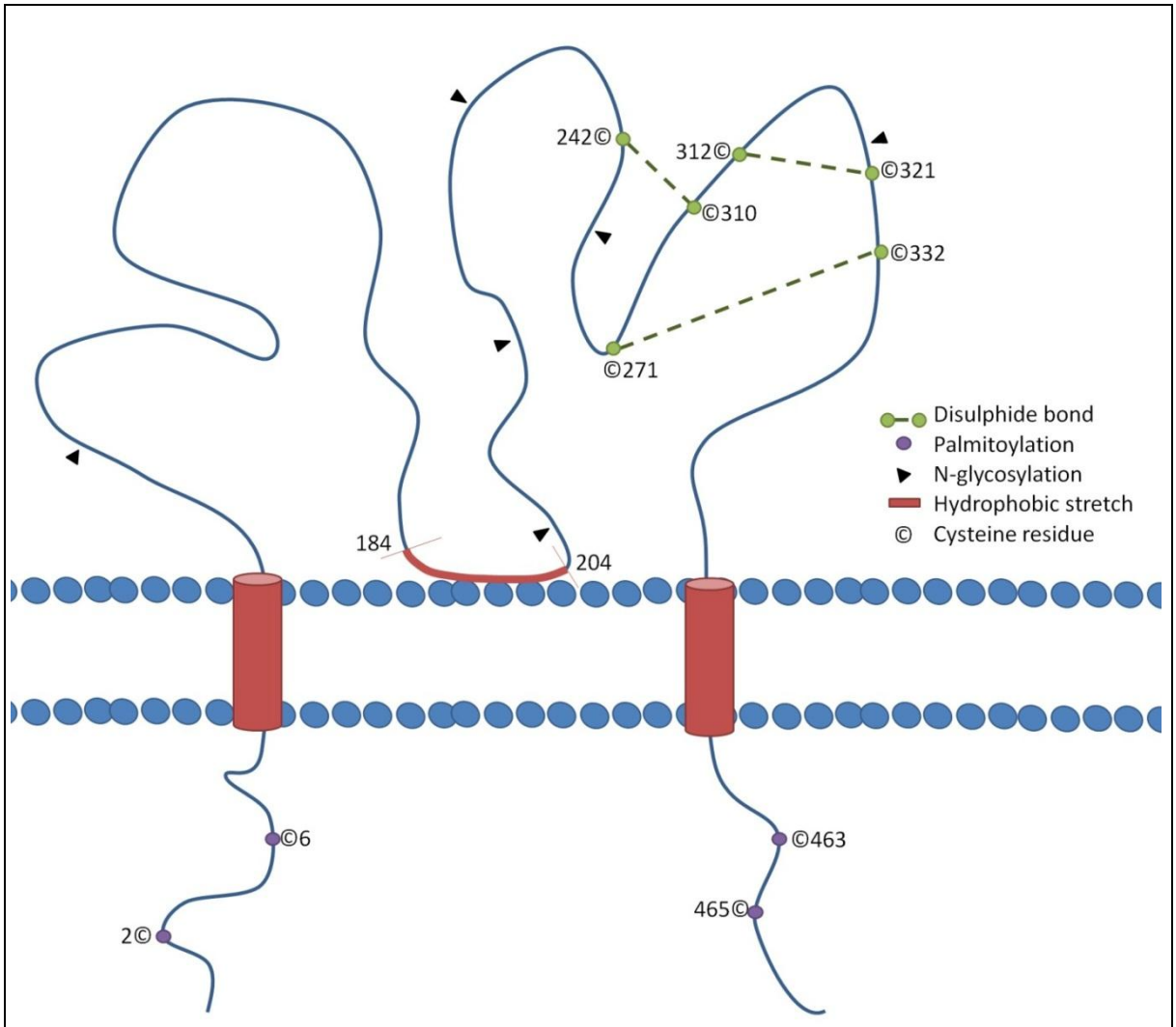


Figure 3.1 Graphical representation of the human CD36 protein structure from studies on the primary sequence.

3.1.1.3. Glycosylation patterns on CD36

The predicted theoretical molecular weight of human CD36 estimated by ProtParam[166] is ~53 kDa. However, purified CD36 from human platelets mammary epithelial cells and erythroblasts have a mass of 88 and 78 kDa respectively[78,167,168]. Treatment of purified CD36 from different cell types with glycosidases[141,167,168] resulted in a smaller protein close to the theoretical molecular weight suggesting extensive glycosylation (~55 kDa). The difference in weight observed in different tissues might be due to two factors. Firstly, some exons with potential N-glycosylation sites might have been spliced out in different isoforms of CD36 in different cell types[159]. Secondly, glycans attached to CD36 may undergo substantial but variable post-translational modification in different cells[141]. As the glycosylation status of an extracellular protein plays an important role in determining correct folding, membrane trafficking, as well as binding and function, understanding the pattern of glycosylation may provide insights to the functional role and binding properties of CD36 in different cell types or disease.

The human CD36 has ten predicted N-glycosylation sites[141]. Using mutagenesis and mass spectrometry in conjunction with a membrane trafficking assay, N-glycosylation has been observed in nine out of ten sites[169]. This is in agreement with conserved glycosylation sites on bovine CD36[170]. A mutant lacking all ten N-glycosylation sites wasn't surface-expressed in HEK293T cells despite effective translation. This suggests that the folding and trafficking of CD36 to the cell membrane is disrupted when N-glycosylation is disrupted[169]. Given the large number of glycosylation sites, finding a minimal number of N-glycosylation sites required for surface expression is complex. However, it was shown that if either N120 or N205 was restored with N247, N321 and N417, there was a near 50% increase in surface

expression of CD36[169]. This suggests that multiple N-glycosylation sites are necessary for membrane trafficking. Indeed, the surface expression of folded CD36 could be rescued synergistically by restoring the glycosylation sites systematically in both the carboxy and amino terminus of the ectodomain[169].

Interestingly, mature glycosylation patterns are not necessary for trafficking of soluble folded CD36 onto the mammalian cell membrane or for its interaction with oxidized LDL[169]. This was performed by transient expression of both wild type and N-glycosylation site mutated CD36 in HEK293T cells. Endoglycosidase H is only able to cleave oligomannose glycans and not matured complex glycans. Thus, the level of glycan maturity on the surface-expressed CD36 was determined by its sensitivity to endoglycosidase H. It has been shown that a subset of surface-expressed CD36 was sensitive to endoglycosidase H amongst others that were not. This suggests that complex glycosylation might not be required for the folding, trafficking, and binding function of native CD36[169]. This information is important for the production of soluble CD36 in insect cells.

3.1.2. The pathogenic role of CD36 in malaria

3.1.2.1. Molecular mechanism behind CD36-binding in sequestration

<u>CD36 and its role in sequestration</u>

CD36 plays multiple roles as an adhesion receptor during the process of sequestration. One of them involves being a cytoadherent target for parasite infected erythrocytes. The role of CD36 in cytoadherence was first identified when *P. falciparum* parasitised erythrocytes were shown to interact with CD36-expressing COS7 cells[116]. This interaction was further

determined using infected erythrocytes from field isolates with severe malaria. These infected erythrocytes were able to bind to both purified CD36 and CD36 expressing C32 melanoma cells[171]. Indeed, *in vitro* flow binding assays show that the infected erythrocytes form a static interaction with CD36 as opposed to other receptors, such as ICAM1 which binds to PfEMP1 in a rolling manner[119]. This is likened to a process similar to leucocyte adhesion. In this model, rolling adhesion in ICAM1 and static adhesion in CD36 to infected erythrocytes can act synergistically to result in greater adhesion and hence enhances sequestration in a progressive multistep process[172–174].

CD36 is also found to play a role in the forming of rosettes[175] as well as platelet-mediated clumping[80]. However, there is little molecular evidence for the role of CD36 in these processes.

Evidence for the interaction between CIDR1α domains and CD36

Blocking studies using antibodies[176] and CD36 fragments[177] were performed to map the region on CD36 that is responsible for the binding interaction with infected erythrocytes to residues 139 – 184[177].

The minimum M2 region in CIDR1α was identified as the domain responsible for binding to CD36 by counting the number of CD36 expressing mammalian cells (Chinese hamster ovary (CHO) cells) bound to immobilised GST-tagged recombinant PfEMP1 fragments in comparison to both normal CHO cells and ICAM1-CHO cells[112]. The CIDR1α domain is part of a semi-conserved head group structure containing the NTS-DBL1α-CIDR1α combination[39,88,89,131]. The DBLα domain has been shown systematically not to be involved in this interaction by immunofluorescent detection of COS7 cells expressing recombinant

PfEMP1 domains (DBL or CIDR) interacting with immobilized dynal magnetic beads coated with antibodies binding to recombinant human CD36[178].

The minimal region on MC*var*1 required for CD36-binding is a stretch of 179 amino acids within the M2 region (MC179) of CIDR1α[112,179] (Figure 1.7b). Structure-based sequence alignment of MC179 with other known CD36-binding CIDR1α domains highlighted a conserved surface not found in non CD36-binding CIDR domains[114]. However, point mutations of these conserved residues did not affect CD36-binding[113,114]. This suggests that the helical bundle of the MC179 interacts with CD36 as a whole domain, and surface interaction and structural fit is more important than specific residue mediated interaction.

A three amino acid motif differs between CIDR domains from CSA-binding and CD36-binding PfEMP1. CIDR domains from a CSA-binding PfEMP1 (CIDR1α1) cannot bind to CD36[93]. Comparison of the CD36-binding ItG2 strain and a CSA-binding ItG2-CS2 strain[180] revealed a three amino acid substitution. The CD36-binding CIDR1α domain contained an Asp-Ile-Glu (DIE) motif instead of a Gly-His-Arg (GHR) motif found on the CSA-binding CIDR1α1[113]. When DIE was replaced by GHR, CD36-binding ItG2 strain was abolished. Mutation of Glu-Ile-Lys (EIK) motif in the CD36-binding MC179 to GHR disrupted CD36-binding, thus signifying the importance of this three residue motif in maintaining CD36-binding[113].

In a subsequent experiment, Gamain *et al* showed that combining a CD36-binding CIDR1α domain with a CSA-binding DBL domain abolish the CSA-binding ability of the DBL domain[181]. This effect is once again negated by a modification of the DIE motif (CD36-binding) to a GHR motif (non CD36-binding) in the CIDR1α domain. However, the CSA-

binding ability of this CSA-binding DBL domain was retained when paired with a CIDR domain from a non-CD36-binding PfEMP1 or a DBLβ-C2 from an ICAM1-binding species. These experiments highlight the importance of higher order arrangement in PfEMP1's role in cytoadherence. It might indeed be too simplistic to see PfEMP1 proteins as multimodular proteins with individual binding domains. Studies on the importance of cooperation between binding domains have been highlighted in studies of the CSA-binding DBL domains in *var*2CSA[53].

In a genome wide characterisation of the CD36-binding properties from a range of PfEMP1, Robinson and colleagues found that the CD36-binding PfEMP1s are encoded by *var* genes of groups B and C[131]. This was confirmed by a fluorescent flow cytometry experiment conducted by Jane *et al*[83]. A series of erythrocytes infected by a panel of parasite strains were generated by limited dilution cloning. PfEMP1s expressed on each infected erythrocyte were determined by RT PCR, before performing flow cytometry using immobilised CD36 expressing mammalian cells (CHO). The main advantage of this method is the use of infected erythrocytes that are expressing a single *var* gene. The second advantage is use of native full length PfEMP1 instead of immobilised recombinant proteins on the infected erythrocytes in the binding assays.

The first conclusion from these studies is that CD36-binding occurs most frequently compared to other binding phenotypes. Secondly, the data agrees with previous work[131] that the CIDR1α1 domain on the PfEMP1 from CSA-binding strains encoded by *var* genes from group A does not bind to CD36 cross referenced to the CIDR1α domains from PfEMP1 encoded by *var* genes from groups B and C.

3.1.2.2. Epidemiological studies linking CD36 and severe malaria

Sequestration has long been suspected to be an important determinant of severe malaria[2]. However, unlike ICAM1[45,120,171,182] and CSA[42,53,183,184] which have been linked to the occurrence of cerebral and placental malaria respectively, the clinical outcomes of sequestration due to adhesion to CD36 remains inconclusive.

A stable interaction formed between CD36 and PfEMP1 is common in different strains of *P. falciparum* that cause severe malaria[83,131]. However, epidemiological evidence for the role of the interaction between CD36 and PfEMP1 in causing severe malaria is conflicting. Studies using clinical isolates have demonstrated that most of the adhesive phenotypes detected in severe malaria involves binding to CD36, over other suspected receptors[174,182]. Intriguingly, although CD36 adhesion is identified in most *P. falciparum* malarial cases, it had no obvious correlation with the manifestation of severe malaria[44,182,185,186]. This was not the case for severe malaria in Asia where epidemiological studies do correlate CD36-binding to severe malaria[171,187].

Polymorphisms in CD36 within a population in a malaria endemic region are also thought to correlate with development of severe malaria. Initial studies by Aitman *et al* suggested that that mutations in CD36 gene causing CD36 deficiencies in populations in certain portions of Africa unexpectedly correlate with susceptibility to severe malaria[188]. Pain and colleagues, however suggested that the T188G mutation is associated with CD36 deficiencies resulting in protection against severe malaria through decreased parasite sequestration[189]. These views remain controversial, with alternative evidences supporting[190,191] and contradicting[192–194] the idea that CD36 polymorphism, and hence CD36 deficiency, renders protection from malaria.

Thus, there is still no conclusion as to whether CD36 plays a crucial role in the clinical manifestation of severe malaria.

3.1.3. Project aims

The focus of my project will be to characterise the interaction between CD36 and CIDR domain of the PfEMP1 using biophysical methods, leading towards the determination of a co-crystal structure of a CIDR-CD36 complex.

The interaction between CD36 and PfEMP1 has been characterised mainly by fluorescence microscopy and cytometry. Given that sequestration remains central to the development of severe malaria, structural characterisation of the interactions of infected erythrocytes with human receptors will provide insight to future therapeutic endeavours. The work I describe hopes to answer a few key questions:

1) Do CIDR1α domains bind to hCD36ED?
2) What is the stoichiometry of the resulting complex?
3) How does the binding affinity of a single CIDR1α domain compare to that of a full length PfEMP1?
4) Does the glycosylation status affect the binding affinity of CIDR1α domains for hCD36ED?

I have expressed and purified a number of CIDR1α domains, as well as CD36. Here I present the first characterisation of their interaction by biophysical methods and determine the stoichiometry of the complex formed between CIDR1α domains and CD36. This will be

useful for future comparison of affinities between PfEMP1 of different strains and will allow further dissection of the importance of CD36-binding properties of PfEMP1.

Next I tested whether the context of the CIDRα domain, within the conserved DBLα-CIDRα head structure, affects the interaction with CD36. To do this I produced various truncations of a single CD36-binding PfEMP1 protein and investigated their binding to CD36, showing that the CIDR1α domain alone can account for the interaction with CD36.

Finally, CD36 is highly glycosylated in many different cell types. Crystallising glycoproteins is difficult[195]. Therefore I characterised the binding of deglycosylated CD36 to CIDR domains, to assess whether glycans can be removed to facilitate future structural studies of these proteins. This is important in our eventual aim of obtaining a crystal structure of either CD36 on its own, or in complex with either PfEMP1 or its native ligands.

3.2. Characterisation of the interaction between CD36 and CIDR1α domains

Recombinant CD36 and CIDR1α domains have been produced using the baculovirus insect cell system and using *Escherichia coli* expression respectively.

3.2.1. Expression and purification of CD36 using Baculovirus expression system

3.2.1.1. A soluble CD36 ectodomain construct

The primary sequence of the full human CD36 (472 residues) is shown in Figure 3.2a. Prediction of the hydrophobic transmembrane region using an algorithm at the Transmembrane Hidden Markov Model (TMHMM) server[196] highlighted two plausible transmembrane domains, from residues 7 – 29 and 441 to 463 (Figure 3.2a). Although one other hydrophobic domain stretching from residues 185 – 202 was predicted, it thought to be part of the extracellular ectodomain[163].

Soluble murine CD36 ectodomain has been previously expressed using the baculovirus expression system[197]. Alignment of the human CD36 ectodomain with murine CD36 ectodomain using the lalign algorithm[198] gave a sequence identity of 84% with an overlap in 409 residues. The PONDR® (Predictor of Naturally Disordered Regions) algorithms predicted that human CD36 is generally ordered, with two stretches of disordered regions from residues 106 – 130 and 353 – 373 detected (Figure 3.2b)[199–201]. Structural prediction using the I-TASSER platform suggested a series of α helices and β sheets for the human

CD36[202]. From this, a 410 residue sequence corresponding to the extracellular ectodomain (hCD36ED) was selected for expression, including residues G30 to N439 (Figure 3.2a).

This was cloned by PCR from human CD36 cDNA (IMAGE clone 4244251) with flanking NheI and BamHI restriction sites and inserted into a converted baculovirus transfer vector pAcGP67A, which has an additional N-terminal Histidine tag (His_6-tag), TEV cleavage site and a multiple cloning site (MCS).

a) MGCDRNCGLIAGAVIGAVLAVFGGILMPVGDLLIQKTIKKQVVLEEGTIAFKNWVKTGTEVYRQF
WIFDVQNPQEVMMNSSNIQVKQRGPYTYRVRFLAKENVTQDAEDNTVSFLQPNGAIFEPSLSVGT
EADNFTVLNLAVAAASHIYQNQFVQMILNSLINKSKSSMFQVRTLRELLWGYRDPFLSLVPYPVTT
TVGLFYPYNNTADGVYKVFNGKDNISKVAIIDTYKGKRNLSYWESHCDMINGTDAASFPPFVEKS
QVLQFFSSDICRSIYAVFESDVNLKGIPVYRFVLPSKAFASPVENPDNYCFCTEKIISKNCTSYGVLDI
SKCKEGRPVYISLPHFLYASPDVSEPIDGLNPNEEEHRTYLDIEPITGFTLQFAKRLQVNLLVKPSEK
IQVLKNLKRNYIVPILWLNETGTIGDEKANMFRSQVTGKINLLGLIEMILLSVGVVMFVAFMISYC
ACRSKTIK

b)

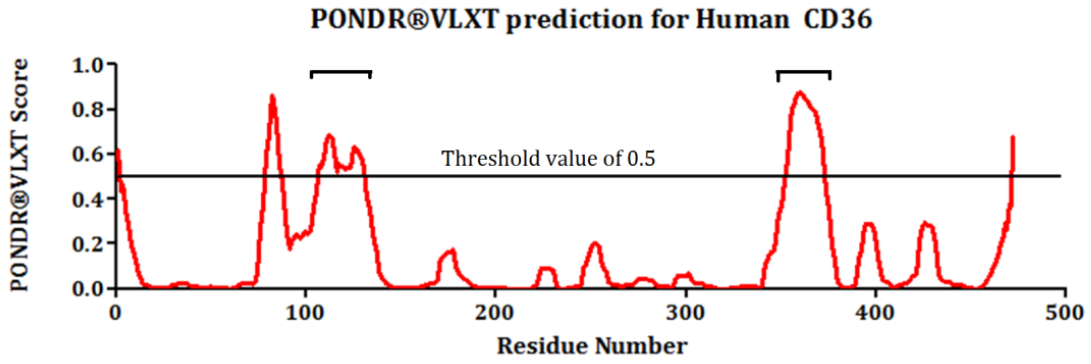


Figure 3.2. Analysis of the human CD36 sequence **(a)** The 472 amino acid residues of the human CD36. The two putative transmembrane domains from residues 7 – 29 and 441 – 463 are underlined in red. The predicted membrane associated hydrophobic patch (185 – 202) is coloured red and the ectodomain highlighted in yellow. The ten extracellular Asn residues which are predicted to be N-glycosylated are highlighted in black. The 6 extracellular cysteine residues coloured and underlined in blue are involved in disulphide bond formation. The 410 residues (highlighted in yellow) which makes up the CD36 ectodomain stretch from G30 to N439. These were chosen for expression. **(b)** The PONDR®VLXT prediction for disordered regions for human CD36. Sequences with a PONDR score above the threshold value of 0.5 are predicted to be disordered. The two main disordered regions are marked with an overhead bracket.

3.2.1.2. Baculovirus expression system

The baculovirus expression system was employed for the expression of soluble hCD36ED. This allowed production of glycosylated protein, a feature which might be important for folding and secretion of the native soluble protein[169]. As discussed above, it is suggested that complex glycosylation is not required for surface expression of CD36 with a subset of CD36 trafficked to the membrane with N-glycosylation patterns that are endoglycosidase H sensitive[169]. Proteins with these N-glycosylation patterns are likely to be similar to those produced by insect cell expression system.

A summary of the baculovirus expression system is shown in Figure 3.3. This uses a modified *Autographa californica* nuclear polyhedrosis virus (AcNPV), with an incomplete linearised genome that only becomes viable when recombined with a transfer vector. The pAcGP67A transfer vector with an hCD36ED insertion allows the incorporation of the hCD36ED sequence into a viable viral genome, allowing expression of soluble hCD36ED under a strong polyhedrin promoter (Figure 3.3). These viable viruses were plaque purified and were amplified via repeated infection of *Spodoptera frugiperda* (SF9) insect cells. They were used for high levels of protein expression using *Trichoplusia ni* (High-Five) insect cells. Expressed hCD36ED is in a soluble form. Expression trials were performed at various stages of the purification process to check for the viability of the viral stock. Detection for the expression of hCD36ED was confirmed by Western blot detection of the N-terminal His_6-tag (Figure 3.4a).

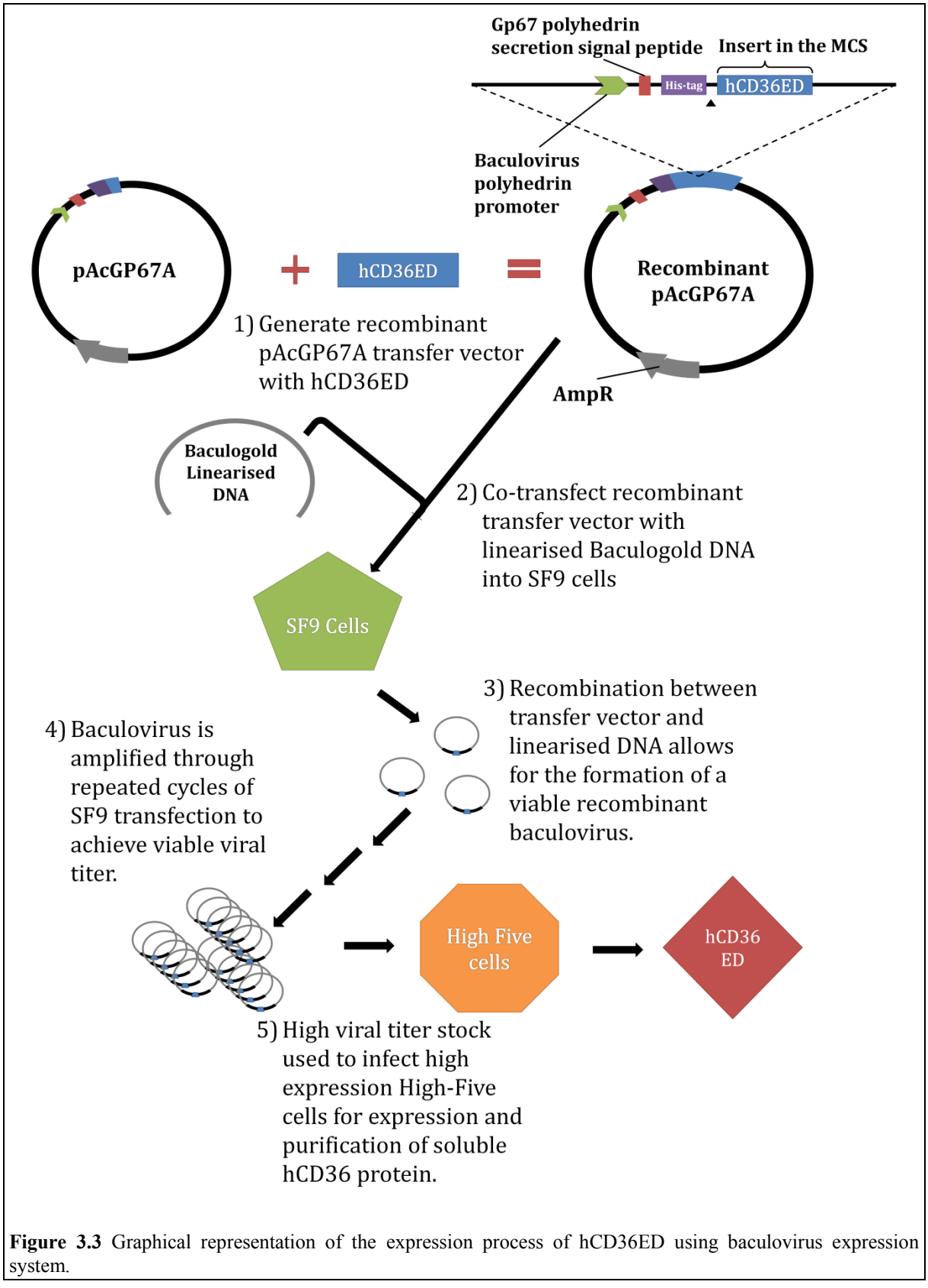


Figure 3.3 Graphical representation of the expression process of hCD36ED using baculovirus expression system.

3.2.1.3. CD36 purification

Ni-NTA affinity chromatography was used to purify the hCD36ED. The supernatant was buffer exchanged and dialysed into 3 litres of 20 mM Tris pH 8.0, 300 mM NaCl for 3 hours at 4°C, before incubating overnight with the Ni-NTA beads. This resulted in the majority of the hCD36ED binding to the beads (Figure 3.4 b). hCD36ED was then subjected to TEV cleavage of the N-terminal His_6-tag before a second Ni-NTA affinity chromatography step to remove uncleaved material and His_6-tagged TEV. This was followed by further purification by preparative size-exclusion chromatography using a Superdex 200 HR 16/60 Gel Filtration column (S200) (Figure 3.4b,c). About 0.5 mg of proteins was purified from every 2 L of High-Five cultures.

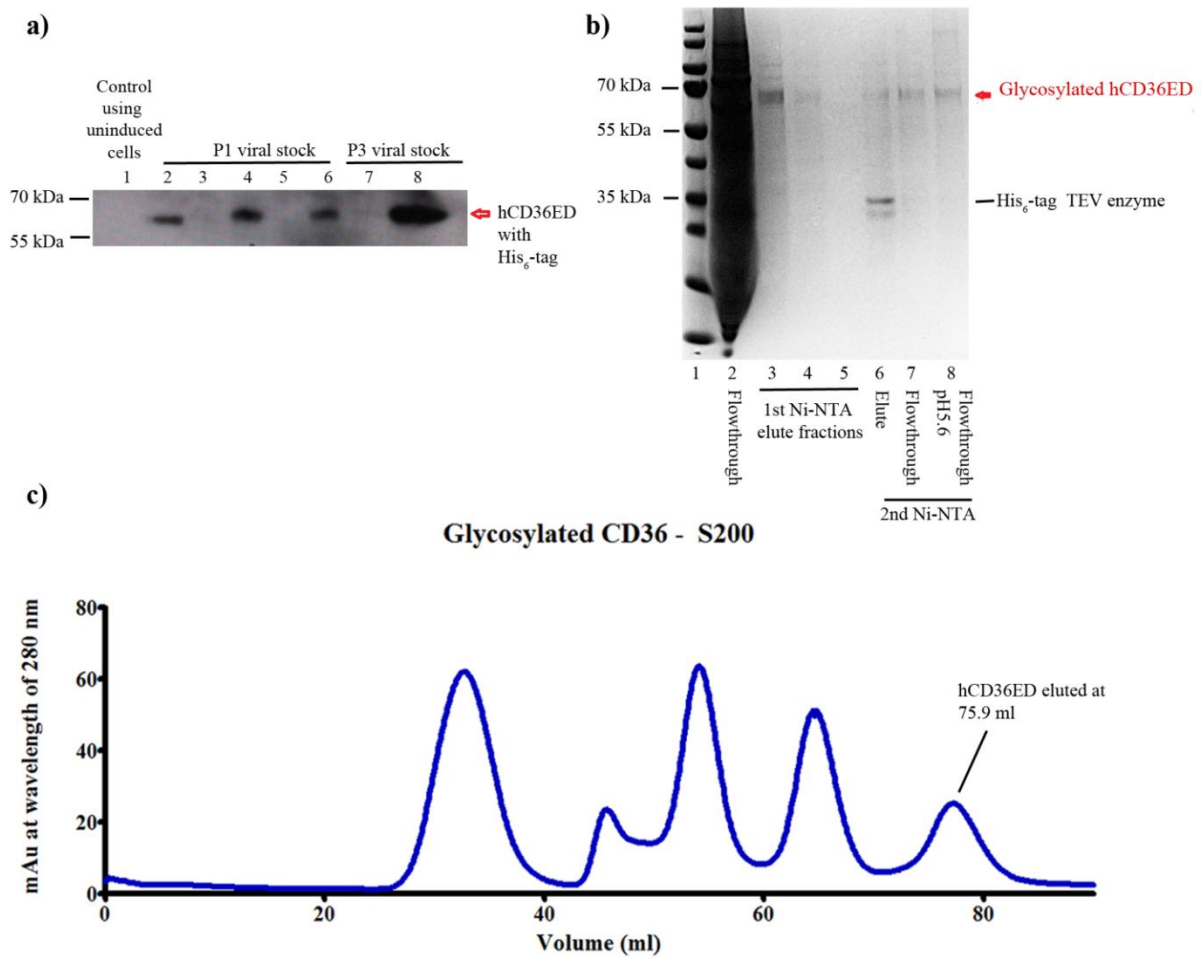


Figure 3.4. Purification of soluble hCD36ED. **(a)** Western blot of an expression trial done with P1 and P3 viral stocks. Only viral stocks of P1 in lanes 2, 4, and 6 were able to lead to hCD36ED expression. Only one P3 viral stock, in lane 8, caused hCD36ED expression. A control experiment was done with uninfected High-Five cells. **(b)** SDS-PAGE of the different fractions collected from the Ni-NTA chromatography, TEV cleavage, deglycosylation and size-exclusion chromatography procedures. **(c)** The size-exclusion chromatography trace for purified glycosylated hCD36ED after TEV cleavage. Glycosylated hCD36ED eluted at 75.9 ml on an S200 column. The other peaks contain contaminants.

3.2.2. Cloning, expression and purification of recombinant PfEMP1 CIDR1α domains using E coli expression system

Nearly 30 PfEMP1s with a similar CIDR1α binding signature have been predicted to be able to bind to CD36[83,131]. In this section, I will describe the process for the selection, cloning and purification of the various CIDR1α constructs studied in this thesis.

3.2.2.1. CIDR1α-M2 construct design

A series of M2 regions of CIDR1α domains from PfEMP1s from different *P. falciparum* strains (3D7 and Malayan Camp (MC)) were selected. The boundaries of the CIDR1α-M2 domains were defined using the fold recognition algorithm FUGUE[203] and sequence alignment using Clustal Omega algorithm[204]. The alignment of the CIDR1α-M2 regions of different PfEMP1s studied here is shown below (Figure 3.5a). There are six conserved cysteine residues in the CIDR1α-M2 constructs. These six conserved cysteine residues are well aligned and thought to form inter-helical disulphide bridges that stabilises the distal end of the alpha helical bundle of the CIDR1α-M2 although not all of them are observed in the crystal structures (Figure 1.8b). In addition, there is a conserved hydrophobic $FFX_2WVX_3LX_5W$ motif observed at the start of each of the CIDR1α-M2 constructs (Figure 3.5).

```
PFL0020w          ---IMKFYEFFIFWVTYMLNDCIDWKKK-ITKCINNGTKWR---CKNKCKNNCKCFEKWV
PFL1955w          ---QKTFHDFFYFWIRRFLNDSMYWRGK-VNSCINKSKREK---CKEECKKECECFQKWI
PFL2665C          ----MPYHPFFWKWVHDMLHDSVEWKTE-LSKCINNKKGSN---CIGGCKKTCECFQKWV
PF11_0007         ---QKTFNPFFYYWVVHMLKDSIHWRTKKLDKCINNSNESKACKNNKKCNKECECFQRWI
MCvar1            EDKIMSYNAFFWMWVHDMLIDSIKWRDE-HGRCINKDKGKT---CIKGCNKKCICFQKWV

PFL0020w          KKKQVEWNTMKEQYEKQPDLVNNRHFT-------TLEWFLE-GQFLKTIEKAYGNEDAIE
PFL1955w          DKKEKEWGQIKIHFNTQDFGSKGGLLGGVMTADFVLKTVLNIEDLFKDIKDGYGNAKELK
PFL2665C          VQKGKEWDEIKVHFNKQEDIVQPGSFIE-FSPYGVLDLVLKGGNLLQNIKDVHGDTEDIK
PF11_0007         TQKQQEWEQIKKHFRKQKNIDNVGGFFK-LSHDDVLQQVLDKDLLLKSLQEAYGKPEDIK
MCvar1            EQKKTEWGKIKDHFRKQKDIPKDWTHDD------FLQTLLMKDLLLEIIQDTYGDANEIK

PFL0020w          KIQEFLQKKSKQE-----------DDEIKDKRDIIDILLEHELDEAQECKDNNREEKNCS
PFL1955w          GIENMLEKEKQKNKE------EEAGASGGEDNTTIDKLLQHEGDEAGECLDTHKETCPDE
PFL2665C          YIKDLLNDEEAAG--------VLVVGSGGENKTTIDKLIEHEKEQAEQCKKCEEAQKP--
PF11_0007         HIKDLLEKEAAAV----------AVVLGGKDNTTIDKLLNHEDKDAKDCLKKQEECKEQ-
MCvar1            RIEALLEQAGVGGIDFAALAGLYTKGFVAEKDTTIDKLLQHEQKEADKCLKTHTDDTCPP

PFL0020w          KEPHD----
PFL1955w          GLG------
PFL2665C          ---------
PF11_0007         ---------
MCvar1            QEDRSVARS
```

Figure 3.5. Sequence alignment of the CIDR1α-M2 regions from different *P. falciparum* strains with the sequence of the crystal structure of MC179 from MCvar1. Highlighted in yellow are the conserved cysteine residues. The conserved hydrophobic residues are highlighted in red. The conserved three residue motifs are highlighted in blue.

In order not to cause confusion between various CIDR1α-M2 constructs, I will employ the annotations in Table 3.1 to address the CIDR1α-M2 regions studied here.

Var gene	Annotation of various CIDR1α constructs	Residue number chosen for expression
PFL0020w	CIDR1	576 - 735
PFL1955w	CIDR2	636 - 806
PFL2665C	CIDR5	610 - 771
PF11_0007	PF11CIDR	643 - 807
	$PF11CIDR_{TRUN}$	643 - 802
	$PF11CIDR_{EXT}$	496 - 802
	$PF11CIDR_{NTS}$	1 - 802
MCvar1	MC179	576 - 755

Table 3.1. Annotations for the different CIDR1α constructs. In addition to the CIDR1α-M2 constructs, three other PF11CIDR constructs from chapter 3.3.1.1 (Figure 3.16) are included.

3.2.2.2. CIDR1α-M2 constructs – PF11CIDR and MC179

3.2.2.2.1. Cloning

The PF11CIDR and MC179 constructs were cloned into pET15bcon using NheI and BamHI restriction enzymes. The figure below shows the PCR products and the restriction digest using NheI and BamHI to check that the inserts were ligated correctly the pET15bcon vector (Figure 3.6b,c). Constructs were checked by sequencing.

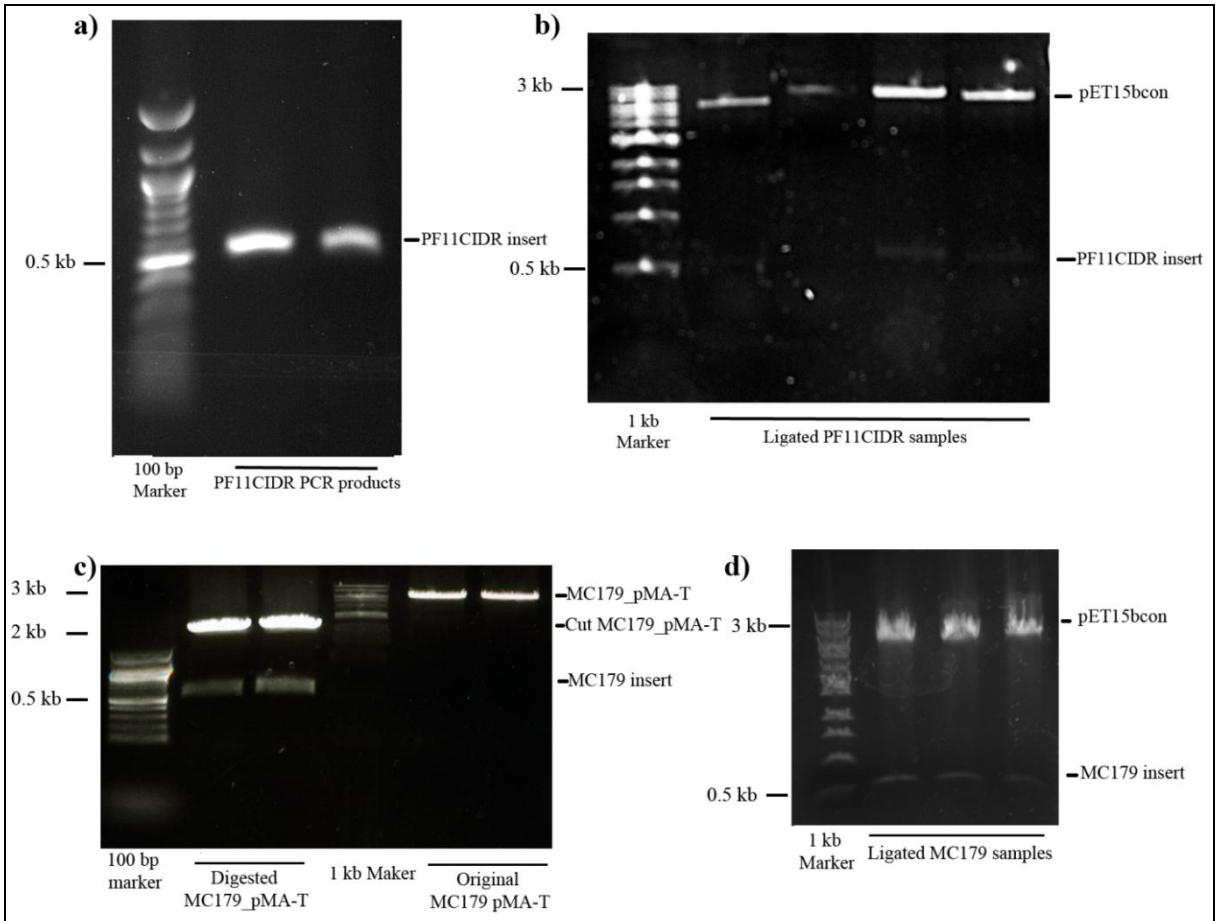


Figure 3.6. Agarose gel electrophoresis for the cloning of PF11CIDR **(a, b)** and MC179 **(c,d)** at different stages. The sizes of PF11CIDR and MC179 are 495 bp and 537 bp. Restriction digest were performed with NheI and BamHI according to manufacturer's protocol. The presence of an insert after ligation suggests successful cloning. These clones were then sent for sequencing to confirm their identity.

3.2.2.2.2. Expression trials for PF11CIDR and MC179

Expression trials were performed to optimise the expression of soluble protein from each of the constructs. These were performed using various *E. coli* strains that can express soluble proteins with disulphide bonds. Origami B cells that have an additional pRIG vector[104,205] to provide an additional three tRNA (Arg, Ile, Gly) to recognise rare codons on *P. falciparum* genes. Origami B cells are also optimised for making soluble proteins with disulphide bridges by the deletion of thioredoxin reductase and glutathione reductase genes (Novagen). BL21(DE3) cells with additional pMJS9 (BL21(pMJS9)) or pFH255 (BL21(pFH255))

vectors[206,207] allow for expression of soluble eukaryotic proteins with disulphide bonds by co-expression of Erv1p, a sulfhydryl oxidase and disulfide isomerase.

Competent *E. coli* cells were transformed with pET15bcon containing the various constructs and were grown at different expression temperatures (18°C, 25°C, 30°C) after induction with IPTG. Samples were collected at three hours or overnight after induction. Western blotting using a Penta-His HRP Conjugate (Qiagen) was employed to detect the recombinant His_6-tagged CIDR1α domains (Figure 3.7 a,b). Only the blots of the cell type for which the various constructs were expressed at the highest level are shown. Thus, the conditions for the CIDR1α-M2 domains were the following. Soluble MC179 was expressed in Origami (pRIG) cells and harvested after an overnight incubation after induction at 25°C. Soluble PF11CIDR was expressed in BL21(pMJS9) cells and harvested after an overnight incubation after induction at 25°C.

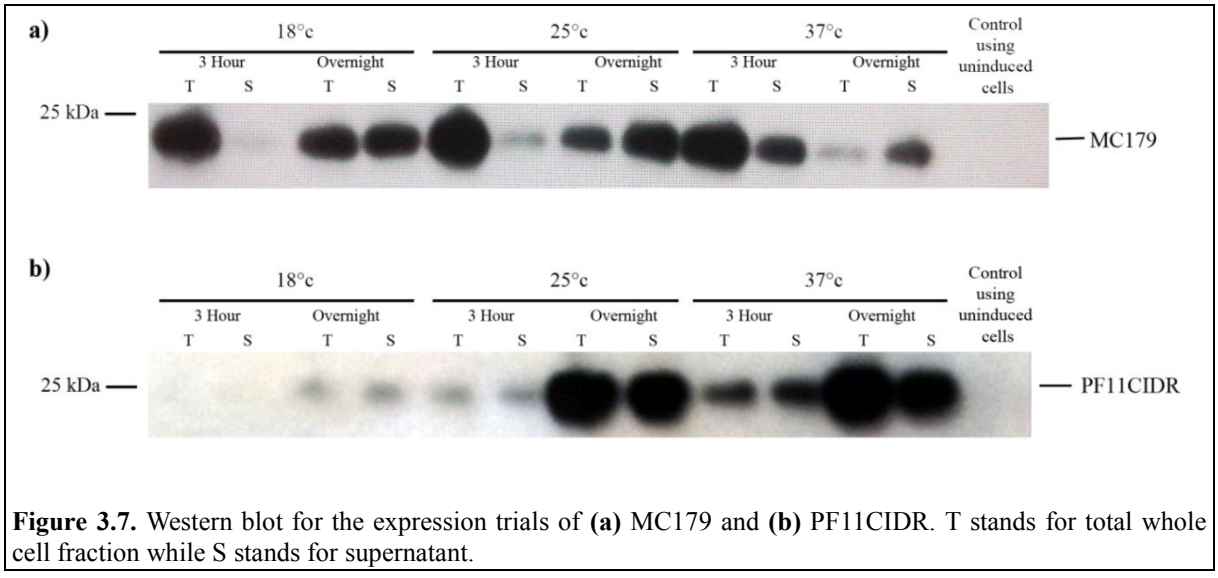


Figure 3.7. Western blot for the expression trials of **(a)** MC179 and **(b)** PF11CIDR. T stands for total whole cell fraction while S stands for supernatant.

3.2.2.2.3. Purification strategy

Ni-NTA affinity chromatography was used to purify the recombinant His_6-tagged CIDR1α domains. Cells were sonicated and subjected to centrifugation to obtain supernatants containing soluble recombinant proteins. These readily attached to Ni-NTA agarose beads and were detected in elution fractions (Figure 3.8a). These fractions were subjected to TEV cleavage before a second Ni-NTA affinity chromatography step to remove uncleaved protein and His_6-tagged TEV enzyme. Finally, protein samples were subjected to preparative size-exclusion chromatography using a Superdex 75 HR 16/60 column (S75). The figure below shows the SDS-PAGE and size-exclusion chromatography from the entire purification process of MC179 (Figure 3.8). The SDS-PAGE and size-exclusion chromatography trace were similar for PF11CIDR (data not shown).

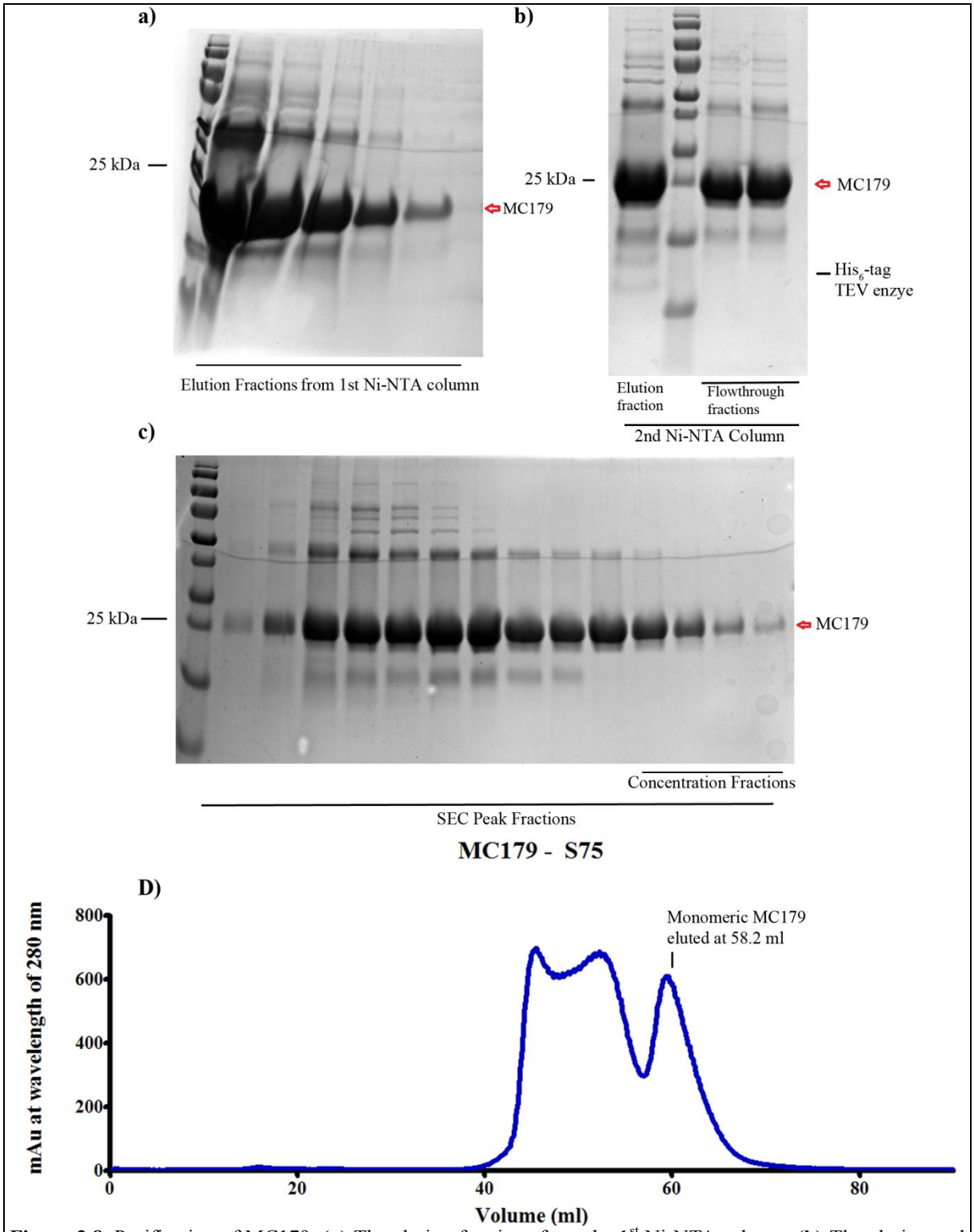


Figure 3.8. Purification of MC179. **(a)** The elution fractions from the 1st Ni-NTA column. **(b)** The elution and flowthrough fractions from the 2nd Ni-NTA column containing TEV enzyme and MC179 respectively after TEV cleavage. **(c)** The size-exclusion chromatography peak fractions **(d)** The size-exclusion chromatography trace MC179 after TEV cleavage. Monomeric MC179 eluted at 58.2 ml on a preparatory S75 column.

3.2.3. Do CIDR1α-M2 domains interact with hCD36ED?

3.2.3.1. Surface plasmon resonance

Surface plasmon resonance measurement using a Biacore T100 was used to characterise the binding affinity of CIDR1α domains for hCD36ED. hCD36ED was immobilised on a sensor CM5 chip using amine coupling. This has a matrix containing carboxymethylated dextrans covalently attached on a gold surface. Amine coupling makes use of the $–NH_2$ amino groups on lysine residues on the ligand (hCD36ED) for covalent immobilisation on the carboxymethylated dextran. The analyte (CIDR1α domains) was then injected over the flow cell chambers under continuous flow conditions. Complex formation as a result of an interaction between the analyte and ligand accumulates proteins on the chip matrix, resulting in a change in refractive index during absorption of the emitted plasmon waves. The optical detector is able to detect changes in the resonance conditions of the reflected surface plasmon waves as a result of the absorption on the molecules that are attached on the metallic surface of the chip, and results is plotted as resonance response units (RU) over time (s)[208].

Experimental procedure

The CM5 chip has four flow cells, allowing attachment of different ligands or a ligand at different surface densities. In my experiments, flow cell 1 was used as a control flow cell with no ligands attached. Flow cell 2 was coupled with glycosylated hCD36ED to a response of 5000 RU.

For equilibrium analysis, a concentration series of each of the recombinant CIDR1α-M2 domains was flowed over immobilised hCD36ED, starting with the lowest concentration. A buffer injection was added at the end of each series to account for any change in signal due to buffer flow. All SPR binding measurements conducted in Biacore buffer.

3.2.3.1.1. Interaction between hCD36ED and CIDR1α-M2 domains

SPR measurements were made to characterise the association and dissociation of PfEMP1 CIDR1α-M2 domains to hCD36ED. All the CIDR1α-M2 domains were able to bind to hCD36ED. The response obtained from binding hCD36ED was greatest for MC179 compared to the other CIDR1α-M2 domains (Figure 3.9). Initial experiments with CIDR1 and CIDR2 were previously performed with a 90 s association and 500 s dissociation phases. Subsequent experiments were performed using 240 s association and 240 s dissociation time. Despite the shorter association time, the interaction of CIDR1 and CIDR2 with the immobilised hCD36ED reached stabilisation.

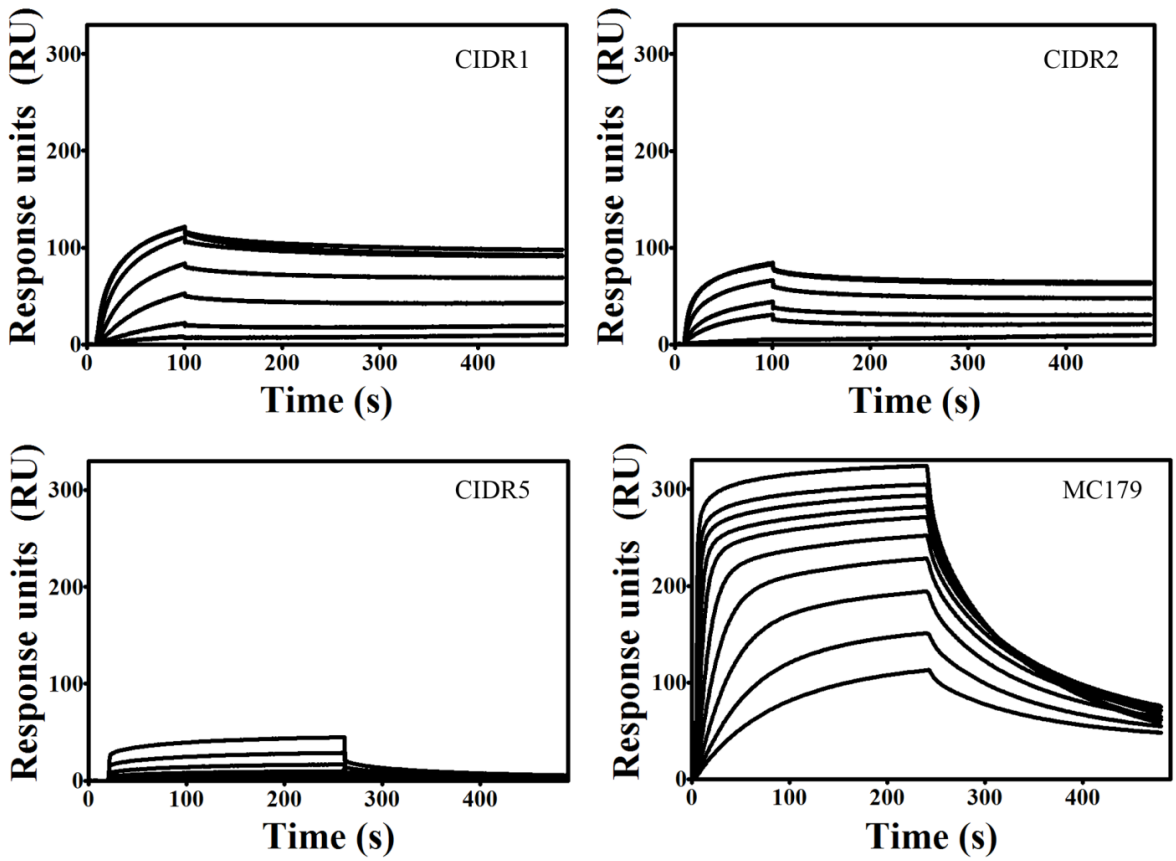


Figure 3.9. Response curves from the surface plasmon resonance measurements for a concentration series of CIDR1, CIDR2, CIDR5 and MC179 for their interaction with glycosylated hCD36ED. The association and dissociation time for CIDR1 and 2 was 90 seconds and 500 seconds, where it was 240 seconds each for CIDR5 and MC179, both at a flow rate of 30 µl/min. The data was analysed using the BIAevaluation software and constructed using GraphPad Prism 5.

The binding affinity of different CIDR1α-M2 domains for CD36 was estimated by equilibrium analysis of the saturated response during complex formation when different concentrations of CIDR1α-M2 domains were flowed over the immobilised hCD36ED chip surface. This was performed by fitting the data to a 1:1 equilibrium binding model as described previously (Figure 3.10a). The dissociation constant (K_D) measured could be used as a measurement for binding affinity.

The data did not fit cleanly to a simple 1:1 binding kinetics model, suggesting heterogeneity in the protein samples. This was consistent with the CIDR1α-M2 domains being present in

both monomeric and dimeric forms, as observed during purification and subsequent AUC experiments (Figure 3.8d, 3.12c). Indeed, a better fit was observed when the data was fitted to a heterogenous analyte model. Using MC179 as an example, fitting was performed by taking into account the relative proportions of monomers and dimers determined by AUC experiment (80% monomer, 20% dimer) during kinetics analysis using the heterogenous analyte model. The fitted curves adhere more accurately to the association and dissociation of the original SPR curves (Figure 3.10b). This gave an estimate of a K_D value close to that estimated using the steady state affinity model. Although the curve gave a more accurate fit, the inability to estimate perfectly the relative amount of different species within the samples of all the different CIDR1α domains limits the use of this method to determine precise kinetic parameters, but allows comparative measurement of binding affinities to CD36 across the different constructs.

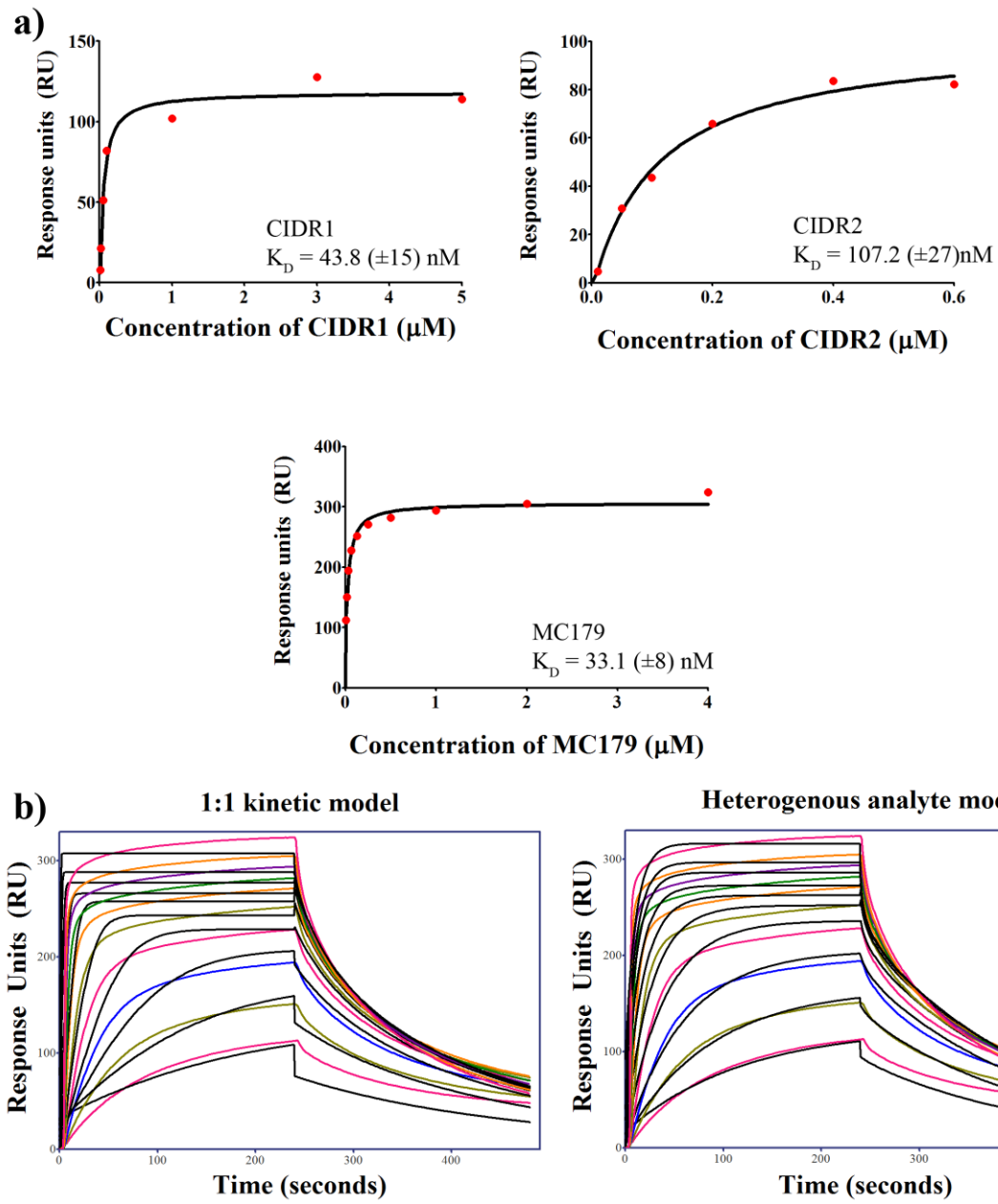


Figure 3.10. Data fitting for surface plasmon resonance measurements of CIDR1α-M2 domains. **(a)** The saturated binding curves for concentration series of CIDR1, CIDR2 and MC179 binding to glycosylated hCD36ED. The curves were fitted using the BIAevaluation programme using a steady state affinity 1:1 binding model. K_D values were estimated to be the concentration of the analyte at which 50% of the ligand were bound in a saturated binding curve. **(b)** An example of the SPR date being fitted using a 1:1 kinetic model and a heterogenous analyte model.

The binding affinity of the interaction between a series of CIDR1α-M2 domains and hCD36ED was examined using SPR at concentrations according to the table below (Table 3.2). The K_D values that were obtained for MC179, CIDR1 and CIDR2 were in the mid-

nanomolar range. MC179 had the highest binding affinity (K_D = 33.1 (± 8) nM). CIDR5 did not saturate at a top concentration of 16 μM, suggesting that the binding affinity to hCD36ED is significantly lower than that of other CIDR1α-M2 domains (data not shown) while the other CIDR1α-M2 domains interacted with hCD36ED under flow conditions with high affinity.

CIDRα-M2 domains	Top concentration (μM)	K_D (nM) to hCD36ED	Chi^2 (RU^2) of fit
CIDR1	4	43.8 (±15)	23.1
CIDR2	4	107.2 (±27)	14.7
CIDR5	32	Not determined	Not determined
MC179	4	33.1 (± 8)	103.0

Table 3.2. Affinities of CIDR1α-M2 domains for glycosylated hCD36ED. The top concentration was the maximum concentration at which the protein interaction was measured. The K_D (nM) was estimated by the BIAevaluation software by fitting the saturation curve via a steady state affinity 1:1 binding model. The errors presented are from BIAevaluation software. The goodness of the fit was determined by the Chi^2 value. CIDR5 did not bind sufficiently strongly to allow KD determination.

3.2.4. What is the stoichiometry of the complex formed between CIDR1α-M2 domains and hCD36ED?

Following the characterisation of the interaction between hCD36ED and CIDR1α-M2 domains, I used analytical size-exclusion chromatography and AUC determining the stoichiometry of complexes between CIDR1α-M2 domains and hCD36ED.

3.2.4.1. Analytical size-exclusion chromatography

Analytical size-exclusion chromatography uses the chromatographic techniques separation of particles based on their hydrodynamic volume. It can be used to study self-association of monomers as well as protein-protein interactions[209,210]. Complex formation results in

changes of the combined hydrodynamic volume of the interacting protein partners. Hence, the interaction between hCD36ED and CIDR1α-M2 domains could be observed as a shift in the elution volume compared to when they are in their respective monomeric form. The analytical GL 10/300 version of the Superdex 200 (S200) were used in all these experiments, given the smaller bed volume, higher sensitivity, and smaller sample requirements.

Analytical size-exclusion chromatography was used not only to affirm the interaction between CIDR1α-M2 domain and CD36, but also to estimate the stoichiometry of the interaction by observing the change in hydrodynamic properties of the complex on the analytical size-exclusion chromatography column by varying the concentrations of various components.

3.2.4.1.1. Complex formation between hCD36ED and MC179

100 µl of complex at 0.2 mg/ml was injected onto an analytical S200 column with a flow rate of 0.25 ml/min at room temperature. MC179 was used in this experiment because of the large quantities at which it can be produced and the relatively high binding affinity for CD36 as manifested in the SPR data.

For the formation of complex, hCD36ED was incubated with MC179 at a molar ratio of 2:1 (hCD36ED:MC179) at room temperature for 30 minutes before analytical size-exclusion chromatography. The hCD36ED is the bigger component (theoretical molecular weight is 46.6 kDa) of the complex relative to MC179 (theoretical molecular weight is 24.8 kDa). A 1:1 complex should have a theoretical molecular weight of ~75 kDa including glycosylation. A peak corresponding to excess hCD36ED at 46.6 kDA should be observed.

Figure 3.11a shows the elution volumes of MC179, hCD36ED, and the MC179:CD36 complex on an analytical S200. Glycosylated hCD36ED and MC179 elute at 15.05 ml and 16.13 ml respectively while the MC179:CD36 complex co-elutes at 13.63 ml. Column calibration with standards indicates that a peak at 13.63 ml is appropriate in mass for a 1:1 MC179:hCD36ED complex (Appendix D). The disappearance of the MC179 peak at ~16ml in the size-exclusion chromatography of the complex also suggested that all of the MC179 is incorporated into the complex, which is shown in the SDS-PAGE (Figure 3.11b). A peak corresponding to excess hCD36ED was also observed.

From the analytical size-exclusion chromatography trace, a shift in elution volume by ~1.5 ml when MC179 was added to hCD36ED and disappearance of the MC179 peak as a result strongly suggested formation of a stable complex of MC179 and hCD36ED.

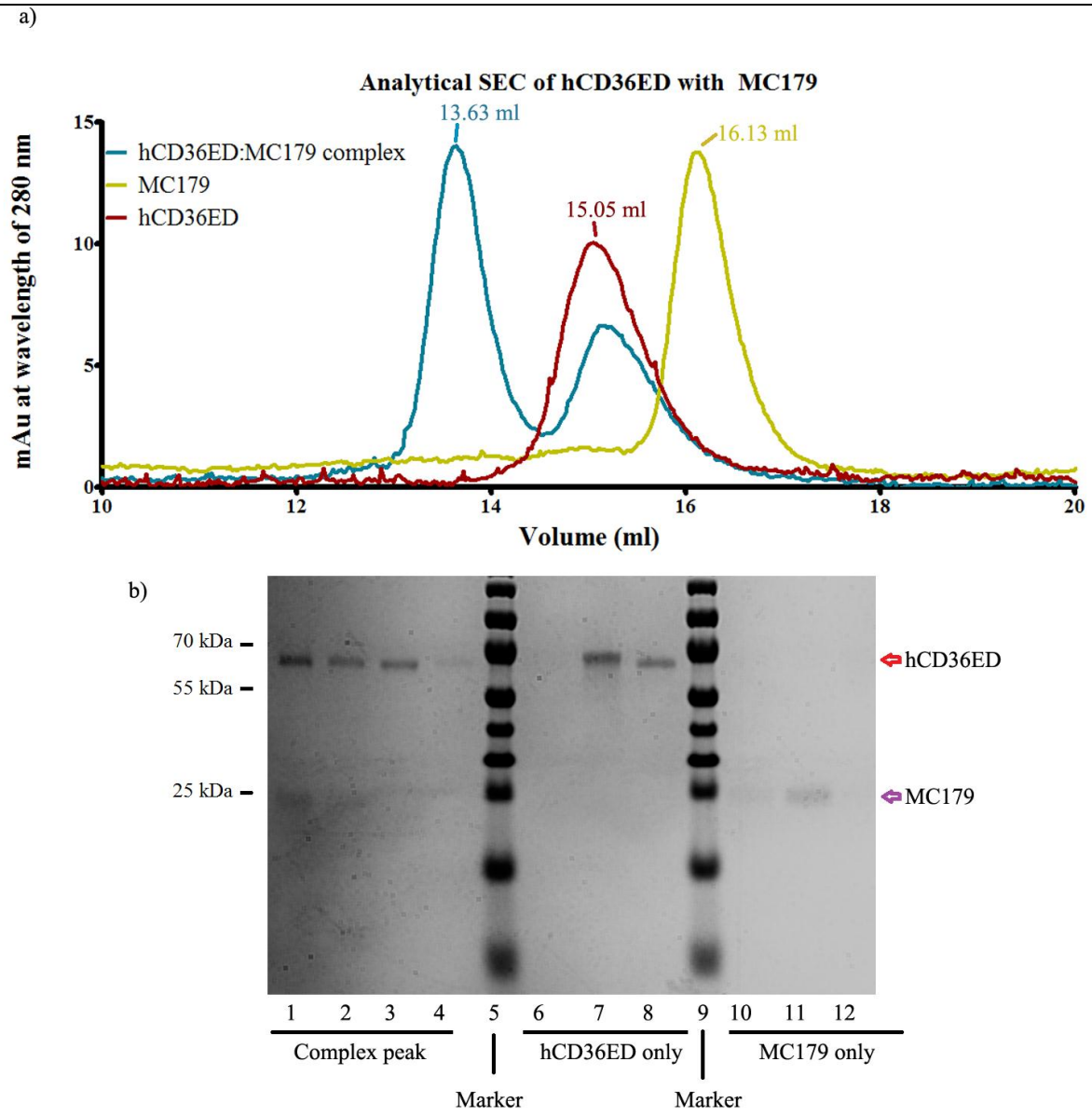


Figure 3.11. Analytical size-exclusion chromatography of MC179 and hCD36ED. **(a)** The traces for MC179, hcD36ED and the co-complex (hCD36ED:MC179) were overlaid on the same axes. MC179, hCD36ED and the co-complex eluted at 16.13 ml, 15.05 ml and 13.63 ml respectively on an analytical S200 column. **(b)** An SDS-PAGE of the different peaks from the analytical column. The complex peak was ran in lanes 1 – 4. MC179 could be seen to have co-eluted with hCD36ED in lanes 1 and 2. Lane 3 consists of the excess hCD36ED. The other lanes consist of hCD36ED and MC179 alone. MC179 could be seen in lane 11 and hCD36ED could be seen in lanes 7 and 8.

3.2.4.2. Analytical ultracentrifugation

The final method used to assess the interaction and stoichiometry of the complex formed between hCD36ED and CIDR1α-M2 domain was analytical ultracentrifugation (AUC). There are two techniques of specific interest in my project; sedimentation velocity and sedimentation equilibrium experiments.

3.2.4.2.1. Sedimentation velocity

Sedimentation velocity experiments were conducted to better understand the stoichiometry and hydrodynamic behaviour of the complex. This follows the sedimentation of proteins when subjected to ultracentrifugal conditions. Sedimentation depends a balance of two factors; the centripetal force (dependent on molecular weight of the species and the speed of the ultracentrifuge) and the frictional force (dependent by the hydrodynamic volume and shape of the protein)[135–137,211]. Hence, large proteins sediment more quickly than small proteins. Compact globular proteins also sediment faster than asymmetric or unfolded proteins. The sedimentation coefficient (S) of the protein, determined by AUC at a constant velocity, hence incorporates both the molecular mass and hydrodynamic shape assumed by the protein in solution[135–137,211].

For sedimentation velocity experiments, all the samples were prepared at a concentration of 0.2 mg/ml. The complex sample was prepared by incubating MC179 with hCD36ED for 30 minutes in a 1:1 molar ratio to a combined concentration of 0.2 mg/ml. Both absorbance and interference data were measured to track the sedimentation properties of the samples in solution when being subjected to centrifugation at 40,000 RPM at 20°C.

Absorbance data was used for data analysis using SEDFIT. The protein partial specific volume (vbar) of each protein and buffer viscosity (1.0002) and density (1.0049) were estimated using Sednterp. Data were then fitted to distribution curves that satisfy the Lamm equation c(s) distribution[135–137]. The weight averaged frictional ratio (f/f_0) and molecular weight estimated by the programme also indicate the properties of the individual components as well as the complex that is formed in solution.

MC179

The vbar of MC179 was estimated to be 0.7372. Data were fitted to the distribution curves and satisfied the Lamm equation with an rmsd of 0.006 (Figure 3.12c). The f/f_0 ratio estimated by the analysis was 1.32, consistent with a compact globular structure[135]. There were two peaks at sedimentation coefficient 2.2S (71.8%) and 3.8S (15.2%) respectively, suggesting MC179 to be present in solution in both monomeric and dimeric forms. The estimated molecular weight from the sedimentation coefficient was ~28.2 kDa, comparable with a theoretical molecular weight of 24.8 kDa.

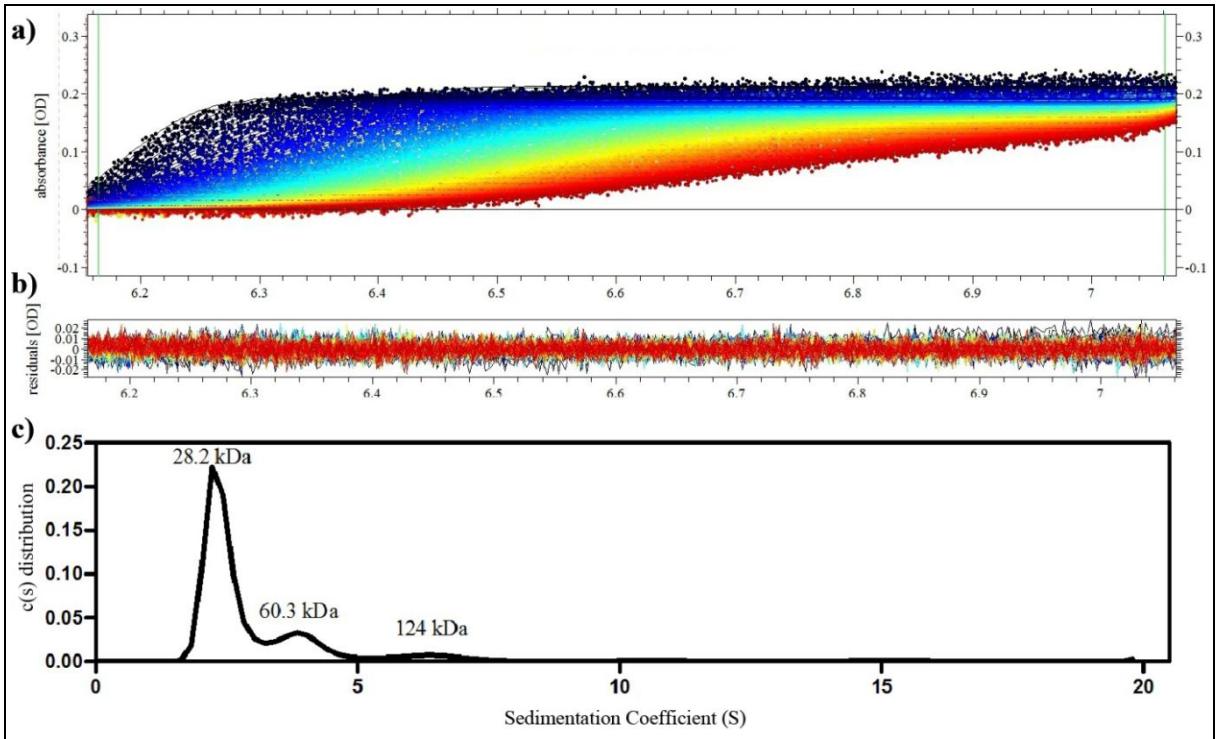


Figure 3.12. Sedimentation velocity data for MC179. **(a)** Theoretical curves that were fitted with AUC data to satisfy the Lamm equation with an rmsd of 0.006. **(b)** The residuals showing the error of the fit for each data set. **(c)** The normal c(s) distribution of the species from the AUC data set, with 71.8% of the species in the major peak at 2.2S and 15.2% of the species residing at the peak at 3.8S.

Glycosylated hCD36ED

The vbar for hCD36ED was estimated to be 0.7413. Data were fitted to distribution curves and satisfied the Lamm equation with an rmsd of 0.006 (Figure 3.13c). The f/f_0 ratio estimated by the analysis was 1.09, and was consistent with a compact globular structure. A single peak was observed at sedimentation coefficient 4.0S (88.2%). This was broader than would be expected for a single species. This may be due to heterogeneity in mass as a result of variation in the glycosylation state of hCD36ED. The estimated molecular weight of ~50.8 kDa correlates well with that of the theoretical molecular weight of 46.6 kDa.

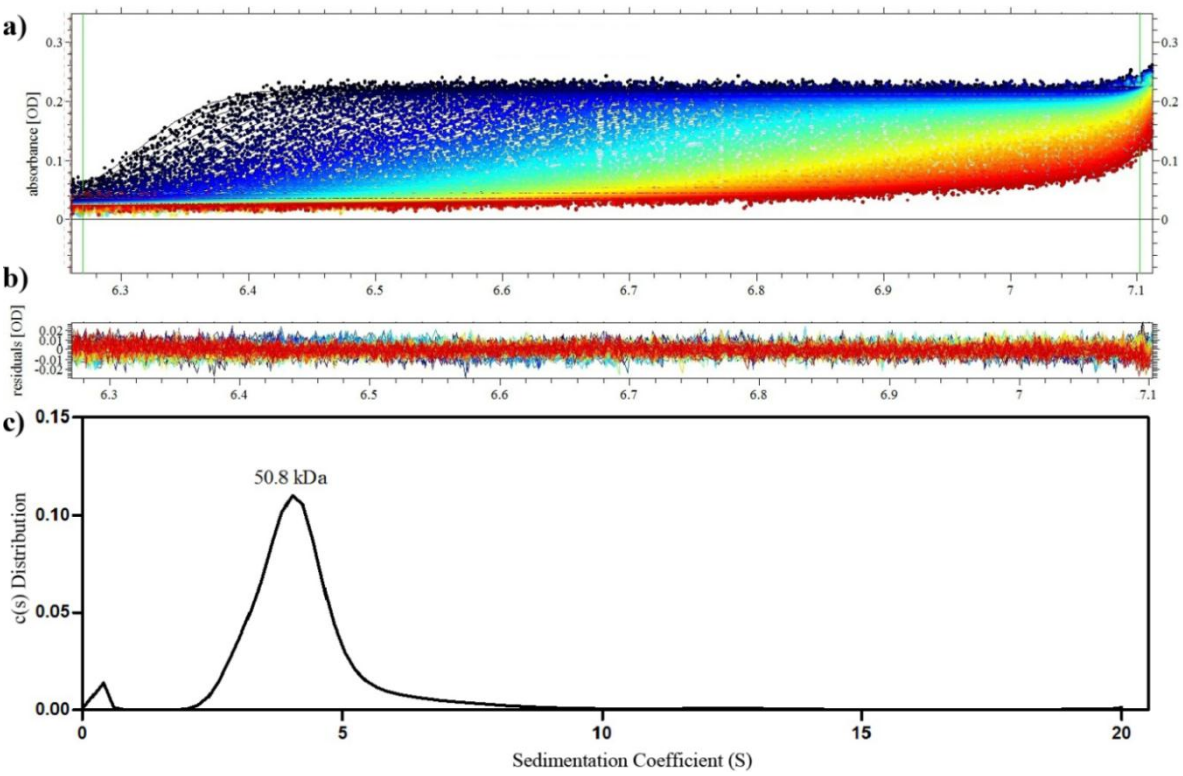


Figure 3.13. Sedimentation velocity data for hCD36ED. **(a)** Theoretical curves were fitted with AUC data to satisfy the Lamm equation with an rmsd of 0.006. **(b)** The residuals showing the error of the fit for each data set. **(c)** The normal c(s) distribution of the species from the AUC data set, with 88.2% of the species residing at the 4.0S peak.

hCD36ED-MC179 Complex

The vbar of the complex was estimated to be 0.740. Data were fitted to distribution curves and satisfied the Lamm equation with an rmsd of 0.006 (Figure 3.14c). The f/f_o ratio estimated by the analysis was 1.31, which suggested a change in hydrodynamic shape from that of the hCD36ED (1.09) when MC179 was added. This f/f_o value predicted for the complex is consistent for a protein with compact globular structure.

Three significant peaks were observed with 75.2% of the sample distributed under the biggest peak at a sedimentation coefficient of 4.5S. This correlates with a complex with a molecular weight of ~70.1 kDa, which was near the theoretical molecular weight for a 1:1 complex (71.4 kDa). In addition, 3.1% and 6.3% of the sample was distributed under two

peaks at sedimentation coefficient 2.8S and 6.9S respectively. The first peak at 2.8S could be attributed to monomeric MC179. The second peak at 6.9S can correspond to a dimer of MC179:hCD36ED complex. The formation of these heavier species might be nucleated by dimeric MC179 found in the sample. The rest of the sample was found distributed at lower sedimentation coefficients.

The individual AUC traces are overlaid in Figure 3.14d. This shows a similar trend as that found in analytical size-exclusion chromatography; there was a shift in size of the protein complex resulting from an interaction between MC179 and hCD36ED in a 1:1 stoichiometry. Hence, the sedimentation velocity experiments were able to affirm the above findings using SPR and analytical size-exclusion chromatography that hCD36ED forms a stable complex with MC179 in solution. The heterogenous properties of MC179 were also determined clearly by the experiment. This enabled fitting of the SPR measurements to a heterogenous analyte model when the ratio of heterogenous species of MC179 was taken into account.

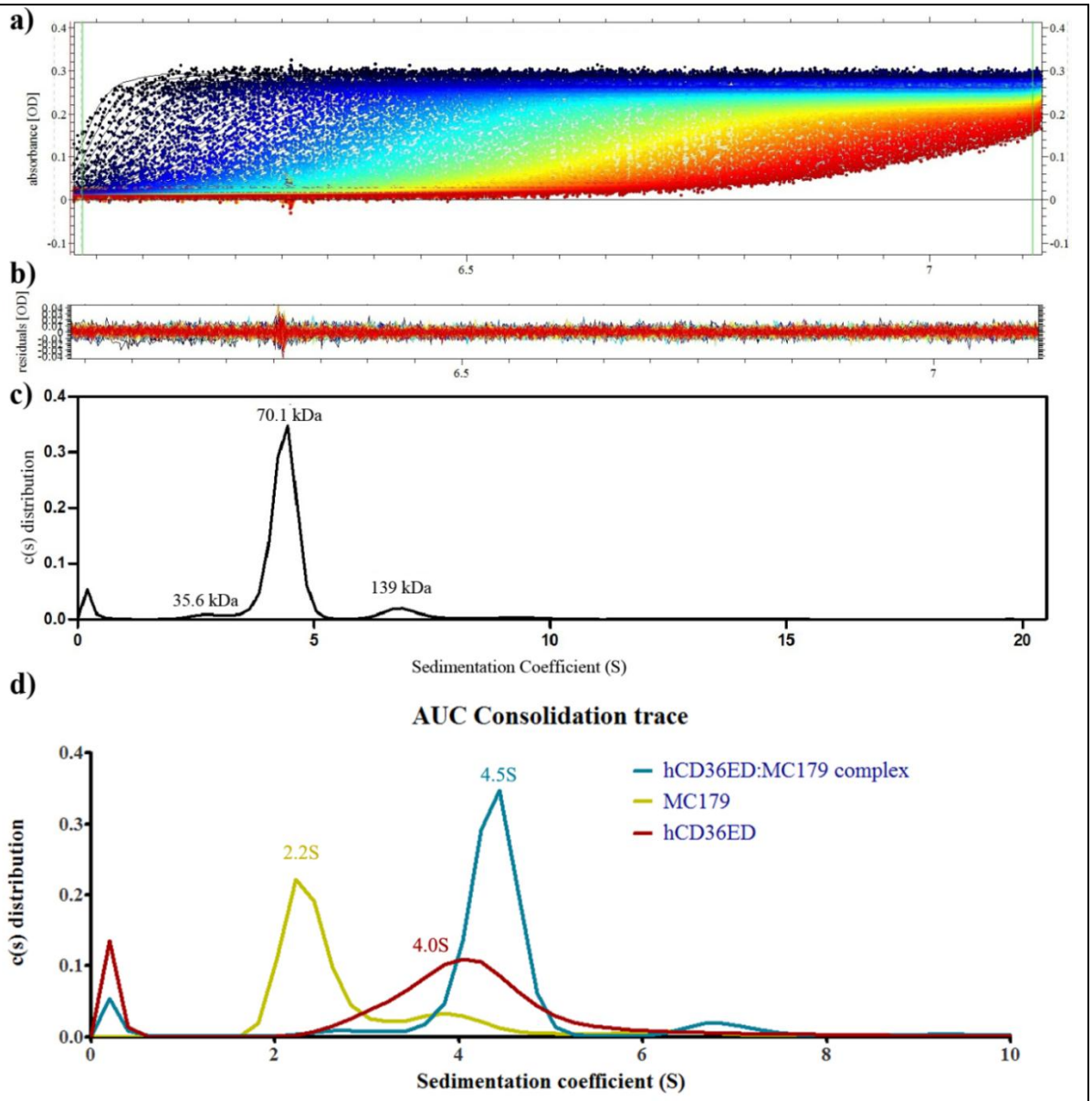


Figure 3.14. Sedimentation velocity data for the complex between hCD36ED and MC179. **(a)** Theoretical curves that were fitted with AUC data to satisfy the Lamm equation with an rmsd of 0.006. **(b)** The residuals showing the error of the fit for each data set. **(c)** The normal c(s) distribution of the species from the AUC data set, with 75.2% of the species residing at the major 4.5S peak. Two smaller peaks at 3.8S (3.1%) and 6.9S (6.3%) could also be seen. **(d)** A consolidation of the c(s) distribution curves of the three AUC experiments.

3.2.4.2.2. Sedimentation equilibrium

In contrast to sedimentation velocity experiments, sedimentation equilibrium measurements give measurements of mass unbiased by shape. Centrifugal force is required to sediment proteins during velocity experiments. However, at low speeds, centrifugal force creates a

concentration distribution of proteins against an opposing diffusional flux. When the sedimentation flux is balanced out by the diffusional flux, an equilibrium concentration distribution of the protein is reached. At equilibrium, the measurements are independent of the shape of the protein, and hence equilibrium AUC is an accurate way to measure mass[135,212,213].

Equilibrium AUC measurements were made for a 1:1 complex of glycosylated hCD36ED with MC179 at a molar ratio of 1:1. A 100 µl sample at 0.2 mg/ml was incubated for 30 minutes at room temperature. The samples were spun at speeds of 10,000 RPM to 18,000 RPM, with a 2000 RPM increase per cycle. A delay time of 18 hours was given before a scan is made at each speed to allow equilibrium to be reached. A second scan was made with a delay of 2 hours after the first scan was made to reaffirm the equilibrium position for each speed. Both absorbance and interference data were collected.

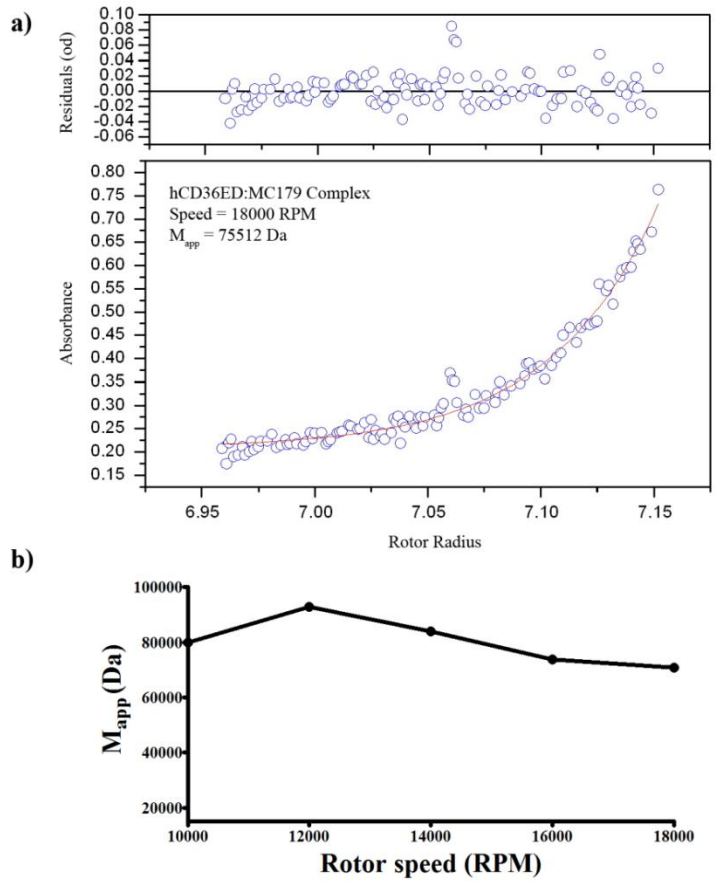


Figure 3.15. Sedimentation equilibrium data for the complex between hCD36ED and MC179. **(a)** An example of a single fitted curve at 18000 RPM that satisfies the self-associating (heterogeneous) model. The M_{app} suggested is via a global fit of all the data points for all five speeds (10 000, 12 000, 14 000, 16 000 and 18 000 RPM). The residuals displayed are showing the error of the fit for each data set. **(b)** The plot of the M_{app} suggested for each speed when they were individually fitted. This was used to validate the error of the suggested M_{app} between each speed.

Data was analysed using the absorbance data on the Origin 6.0 using the self (-association) model for heterogeneous species. Individual curves at different speeds were fitted in order to estimate the apparent masses (M_{app}). From Figure 3.15b, we can see that the estimated M_{app} at different speeds was about ~80 kDa, with a fluctuation of less than 10%. This suggests that an equilibrium concentration distribution has been reached at the speeds from 10 000 RPM onwards. Thus, these curves were then globally fitted to get a more robust M_{app} from multiple data points from the scan at equilibrium. An example of the global fit for a single speed at 18 000 RPM was shown in Figure 3.15a. The M_{app} for the complex between glycosylated hCD36ED and MC179 was 75.512 kDa. The M_{app} was very close to the theoretical molecular weight (71.4 kDa) of a 1:1 complex (hCD36ED:MC179).

3.3. What is the minimum CD36-binding region of a PfEMP1?

In the previous section, we confirmed the interaction between CIDR1α domains and hCD36ED. Next I investigated whether CIDR1α domains account for the full binding affinity of PfEMP1 to CD36 as this is not the case for *var*2CSA[53]. A series of longer CIDR1α constructs were therefore made for further biophysical investigation.

3.3.1. Expression and purification of the PF11_0007 series

3.3.1.1. The PF11 series – CIDR1α in different PfEMP1 contexts

Three constructs of PF11_0007 were designed in order to investigate the importance PfEMP1 context on CD36-binding by CIDR1α. Full length PF11_0007 protein (PF11VAR with the entire extracellular framework from residues 1 – 1702) and a DNA clone were kindly provided to us by Dr Louise Turner (University of Copenhagen).

Design of constructs

PF11_0007 is a PfEMP1 encoded by a group B *var* gene. The ecotodomain has a canonical four domain structure, NTS-DBL1α-CIDR1α-DBL2δ-CIDR2β (Figure 3.16). In a recent crystal structure of a DBLα1-CIDRγ head structure, it was observed that the N-terminal domain in the M1 region of the CIDRγ made multiple contacts with the flexible linkers that were important in coordinating the adjacent DLB1α1 domain[102]. Additionally, the CIDRγ assumes a more compact structure than that of a previous CIDR1α structure[114], suggesting

that the extended N-terminal region of the CIDR domain including the flexible linkers might be important in positioning the CIDR domain in the full length PfEMP1.

A few factors contributed to the design of these constructs. Firstly, the binding affinity of the original PF11CIDR for CD36 differed considerably from that of PF11VAR as discussed below. This suggested that other regions of PF11VAR might contribute to binding. An alternative reason for reduced affinity of PF11CIDR might be due to the propensity to form dimers, which might block the CD36-binding interface. Indeed a free cysteine near the C-terminal of the protein may form disulphide-induced dimers. Thus, $PF11CIDR_{TRUN}$ was designed as a truncated version of the PF11CIDR, lacking five residues (ECKEQ) at the carboxy terminus to remove this additional Cys. The subsequent constructs were all designed with a five residue truncation at carboxy terminus.

Two longer constructs were also designed (Figure 3.16). $PF11CIDR_{NTS}$ includes NTS, DBL1α and CIDR1α of the PF11_0007 (residues 1 – 802). $PF11CIDR_{EXT}$ starts from an extended region that is N-terminal to the CIDR1α domain immediately after the predicted end of the DBL1α domain (residues 496 to 802). Prediction of order using PONDR®VLXT algorithm suggested that the linker between the DBL1α and CIDR1α domains is generally a disordered stretch of residues (data not shown).

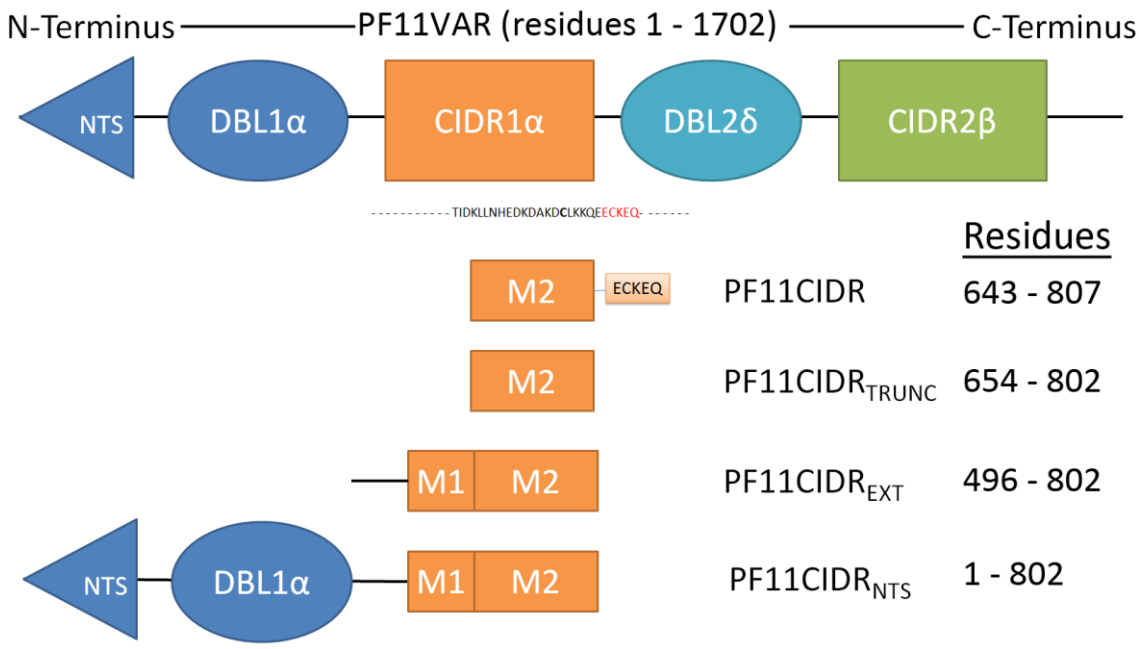


Figure 3.16. Graphical representation of the series of PF11 constructs.

3.3.1.2. Cloning, expression and purification of PF11 series

The three new PF11CIDR constructs were cloned into pET15bcon (Figure 3.17). Expression trials were done as in section 3.2.2.2.2. $PF11CIDR_{TRUN}$ and $PF11CIDR_{EXT}$ expressed well in Origami (pRIG) cells (Figure 3.18 a,b). $PF11CIDR_{NTS}$, however, was not expressed to a suitable level (data not shown). The lack of expression could largely be due to the relatively larger size of the $PF11CIDR_{NTS}$ (~90 kDa) in comparison to the rest of the other CIDR1α constructs, or due to proteolysis of the large protein into smaller domains (data not shown). Both $PF11CIDR_{TRUN}$ and $PF11CIDR_{EXT}$ were expressed in Origami (pRIG) cells and harvested after an overnight incubation after induction at 25°C.

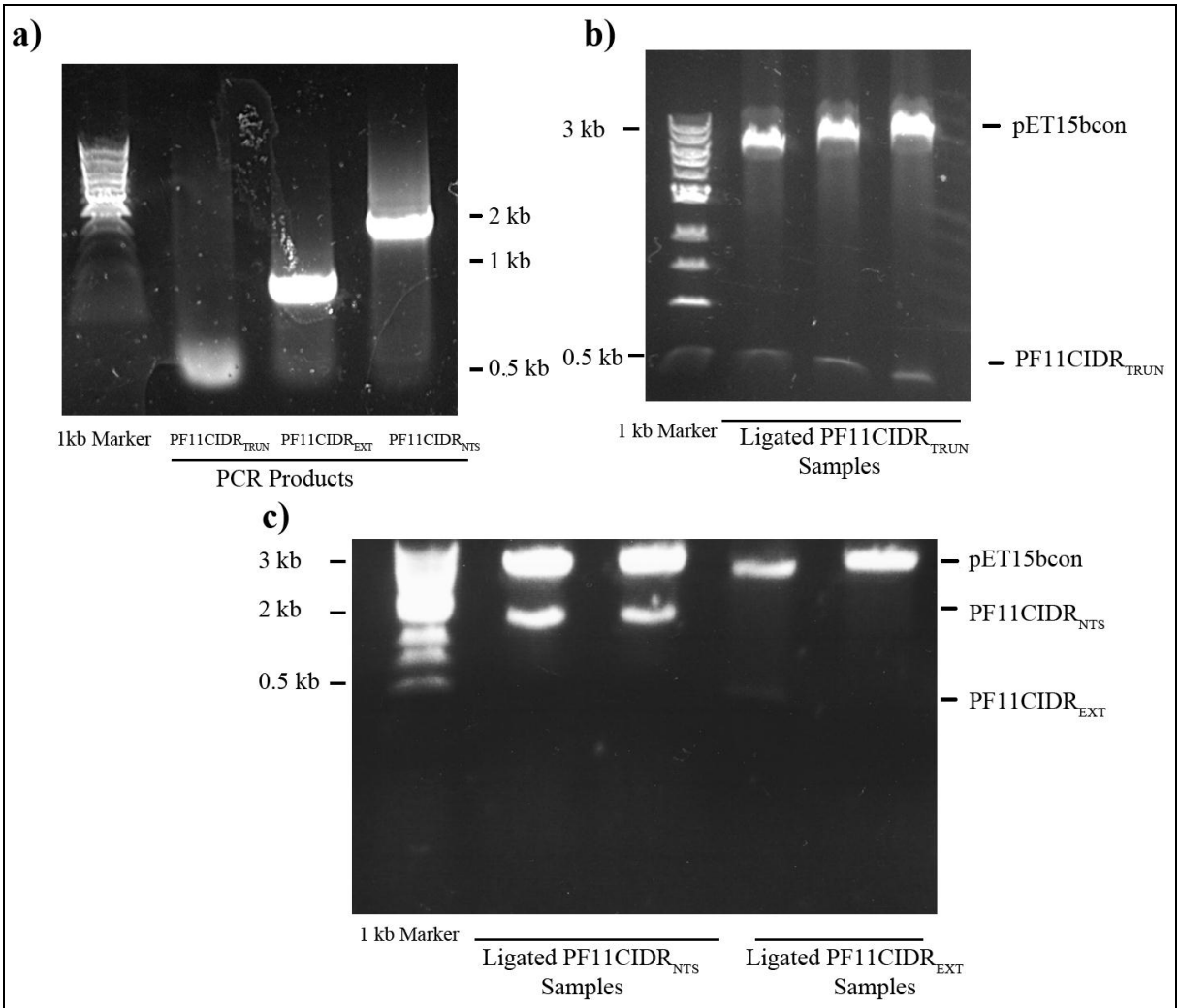


Figure 3.17. Agarose gel electrophoresis for the cloning of PF11CIDR series. **(a)** PCR products of the PF11CIDR series including PF11CIDR$_{TRUN}$ (480 bp), PF11CIDR$_{EXT}$ (921 bp), PF11CIDR$_{NTS}$ (1566 bp). **(b,c)** Restriction digests of PF11CIDR$_{TRUN}$, PF11CIDR$_{EXT}$ and PF11CIDR$_{NTS}$. Restriction digest were performed with NheI and BamHI enzymes according to manufacturer's protocol. The presence of an insert after ligation suggests successful cloning. These clones were then sent for sequencing to confirm their identity.

Soluble versions of both PF11CIDR$_{TRUN}$ and PF11CIDR$_{EXT}$ were purified as for other CIDR1α constructs (Chapter 3.2.2.2.3) using Ni-NTA affinity chromatography followed by TEV cleavage, and finally preparative size-exclusion chromatography (Figure 3.18 c – f).

Although the free cysteine has been removed from PF11CIDR$_{TRUN}$, it still forms dimers. This correlates well with dimerisation seen during the purification of MC179[114] (Figure 3.8d) and other CIDR1α-M2 constructs (data not shown). In contrast, the longer PF11CIDR$_{EXT}$

remains monodispersed during purification. This suggests that the N-terminal region of the extended CIDR1α construct stabilises the monodispersed version of the CIDR construct during expression and purification.

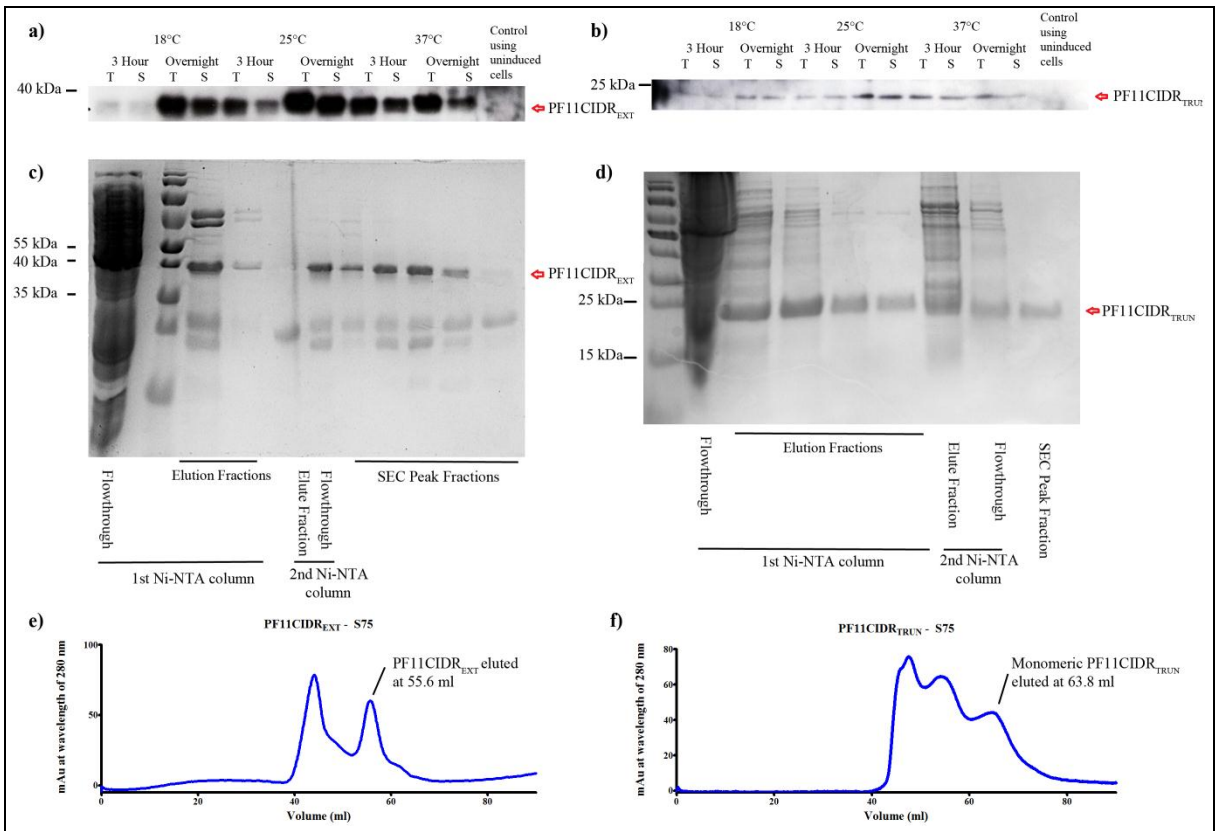


Figure 3.18. Purification of PF11CIDR$_{EXT}$ and PF11CIDR$_{TRUN}$. **(a,b)** Western blots of the expression trials for PF11CIDR$_{EXT}$ and PF11CIDR$_{TRUN}$ using Origami (pRIG) cells. T stands for total whole cell fraction while S stands for supernatant. **(c,d)** SDS-PAGE of fractions collected from the Ni-NTA chromatography, TEV cleavage, deglycosylation and size-exclusion chromatography procedures for both PF11CIDR$_{EXT}$ and PF11CIDR$_{TRUN}$. **(e,f)** The size-exclusion chromatography trace for PF11CIDR$_{EXT}$ and PF11CIDR$_{TRUN}$ after TEV cleavage. PF11CIDR$_{EXT}$ and monomeric PF11CIDR$_{TRUN}$ eluted at 55.6 ml and 63.8 ml respectively on a preparatory S75.

3.3.2. Does the context of CIDR1α affect the interaction between hCD36ED and PfEMP1?

Binding of the PF11 CIDR1α constructs to hCD36ED was investigated as described for the SPR experiment in chapter 3.2.3.1. SPR was used to measure and compare the binding affinities of single CIDR1α domains in their interaction with hCD36ED to that of the full length PfEMP1 using the PF11 series of constructs.

The PF11CIDR$_{EXT}$ construct bound more strongly to the hCD36ED under flow conditions than the other PF11 constructs (Figure 3.19a). The binding affinity of PF11CIDR$_{EXT}$ (K_D value of 710.2 (±69) nM) was similar to that of PF11VAR (K_D = 798.9 (±130) nM) (Table 3.3). The binding affinity of PF11CIDR for hCD36ED was not determined because of the poor saturation at its top concentration (32 μM) (Figure 3.19a). The data obtained for PF11CID$_{TRUN}$ had a distorted decreasing response, and is not presented.

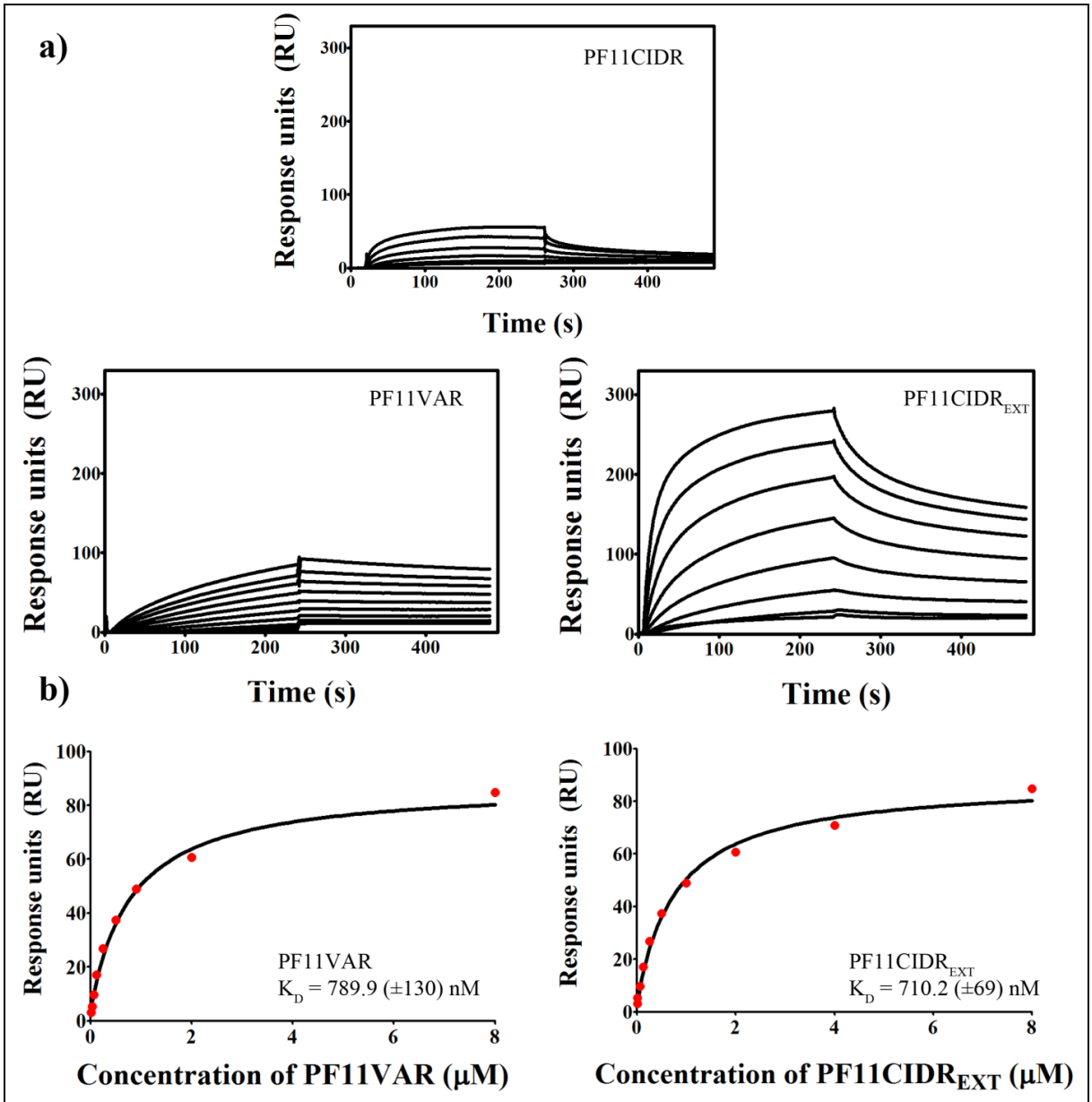


Figure 3.19. Surface plasmon resonance measurements for the PF11_0007 series**. (a)** Response curves from surface plasmon resonance measurements for concentration series of PF11CIDR$_{TRUN}$, PF11CIDR$_{EXT}$ and PF11VAR for their interaction with glycosylated hCD36ED. **(b)** The saturated binding curves for the concentration series of PF11CIDR$_{EXT}$ and PF11VAR for their interaction with glycosylated hCD36ED. The curves were fitted using the BIAevaluation programme using a steady state affinity 1:1 binding model. K_D values were estimated to be the concentration of the analyte at which 50% of the ligand were bound in a saturated binding curve.

PF11 series	Top concentration (μM)	K_D (nM) to hCD36ED	Chi^2 (RU^2) of fit
PF11CIDR	32	Not determined	Not determined
$PF11CIDR_{TRUN}$	32	Not determined	Not determined
$PF11CIDR_{EXT}$	4	710.2 (±69)	29.8
PF11VAR	8	798.9 (±130)	9.2

Table 3.3. Affinities of PF11_0007 series for glycosylated hCD36ED. The top concentration was the maximum concentration at which the protein was measured. The K_D (nM) was estimated by the BIAevaluation software by fitting the saturation curve to a steady state affinity 1:1 binding model. Errors were derived by the same programme during fitting. The goodness of fit was determined by the Chi^2 value. The K_D of some of the constructs were not determined either because their curves were not saturated at the various top concentrations, or their binding curves were too distorted for fitting.

Additional SPR experiments were performed using longer recombinant NTS-DBL1-CIDR1α constructs (ITvar34, ITvar44 and ITvar67) from the IT strain of *P. falciparum* kindly provided to us by Professor Alexandra Rowe (University of Edinburgh). ITvar34 and ITvar44 both bound more strongly than ITvar67. Fitting of the SPR data of ITvar34 and ITvar44 to a steady state affinity binding model (1:1) gave a K_D of 1670.0 (±140) nM and 753.6 (±41) nM respectively (Figure 3.20b, Table 3.4). ITvar67 was not fitted because its binding was not saturated at the top concentration of 2 μM. Although it is not possible to directly compare these data with that of PF11 series, the binding affinities of these constructs were similar to that of the PF11VAR and $PF11CIDR_{EXT}$.

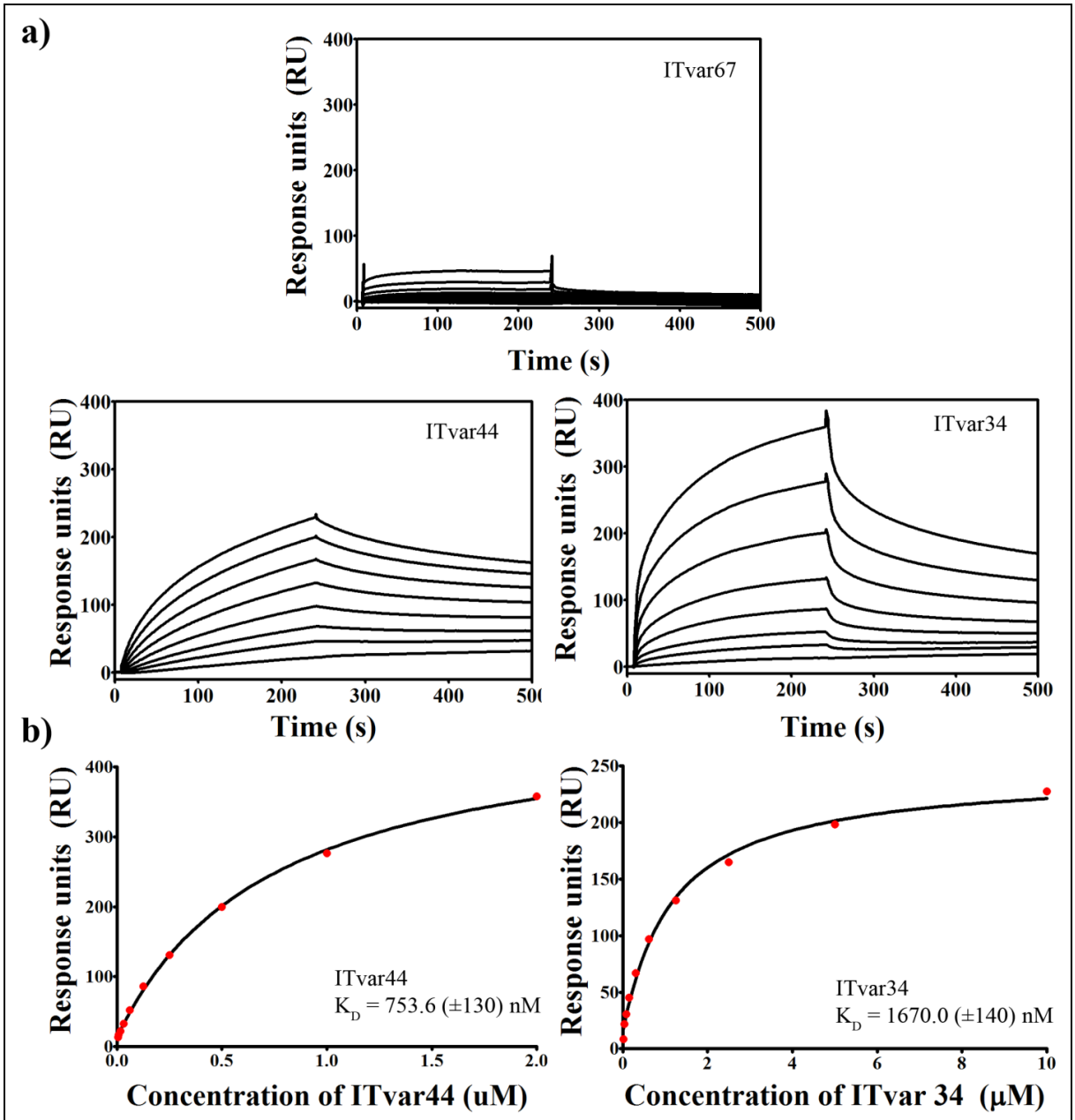


Figure 3.20. Surface plasmon resonance measurements of the NTS-DBL1-CIDR1α constructs. **(a)** Response curves from surface plasmon resonance measurements for a concentration series of ITvar34, ITvar44 and ITvar67 for their interaction with glycosylated hCD36ED **(b)** The saturated binding curves for the concentration series of ITvar34 and ITvar44 for their interaction with glycosylated hCD36ED. Curves were fitted using the BIAevaluation programme using a steady state affinity 1:1 binding model. K_D values were estimated to be the concentration of the analyte at which 50% of the ligand were bound in a saturated binding curve.

NTS constructs	Top concentration (μM)	K_D (nM) to hCD36ED	Chi^2 (RU^2) of fit
ITvar34	10	1670.0 (±140)	36.2
ITvar44	2	753.6 (±41)	13.1
ITvar67	2	Not determined	Not determined

Table 3.4. Affinities of NTS-DBL1-CIDR1α constructs for glycosylated hCD36ED. The top concentration was the maximum concentration at which the protein was measured. The K_D (nM) was estimated by the BIAevaluation software by fitting the saturation curve via a steady state affinity 1:1 binding model. The errors presented were derived by the same programme during the fitting process, in nM. The goodness of the fit was determined by the Chi^2 value during fitting by the programme. The K_D of some of the constructs were not determined either because their curves were not saturated at the various top concentrations, or their binding curves were too distorted for fitting.

The experiments performed with PF11 suggest that the extended CIDR1α domain alone contains the entire CD36-binding site. In essence, these data suggest that the binding affinity of PF11_0007 can be accounted for by a single CIDR1α domain. Interestingly, the N-terminal M1 region of the CIDR1α domain that was included in the $PF11CIDR_{EXT}$ construct seems to contribute to hCD36ED binding. Additionally, the NTS constructs provided by Dr. Alexandra Rowe also bound to hCD36ED with similar high nanomolar affinities to those of full length PF11VAR.

However, this might not be true for all CD36-binding PfEMP1. Some of the CIDR1α-M2 regions that were investigated, such as MC179, had low nanomolar binding affinities to CD36 (Table 3.2). Given the sequence variability of the CIDR domains even within the same class, it will be interesting to compare the binding affinities of other full length PfEMP1 proteins with that of their individual CIDR1α domains. This will give further understanding on the role of CIDR1α domains in the context of a full length PfEMP1 during interaction.

In conclusion, SPR affinity measurements determined the binding affinities of different PfEMP1 constructs for CD36. Firstly, different CIDR1α domains were shown to have a

binding affinity in the nanomolar range. In addition, the SPR measurements were able to be used as a direct comparison between the binding affinities of a single CIDR1α domain, and a full length PfEMP1 for CD36. In this protein it is clear that the CIDR1α alone contains the determinants for CD36-binding in the context full length PfEMP1. This is in contrast to observations of var2csa.

3.4. CD36 deglycosylation and affirmation of its interaction with CIDR1α domain

The human CD36 ectodomain is N-glycosylated at nine out of ten potential sites. Mutation of all ten of the N-glycosylation sites obliterated surface expression of CD36 in HEK293T cells[169]. This suggested that glycosylation was necessary for the folding and trafficking of native CD36 to the cell surface. Crystallisation of macromolecules is hindered by the presence of these flexible post-translational modifications[195]. This requires that hCD36ED is produced and purified as a soluble folded glycoprotein from insect cells, and is subsequently degycosylated. A deglycosylation protocol was hence developed for hCD36ED and biophysical methods were used to affirm the effect of CD36-binding.

3.4.1. CD36 deglycosylation

3.4.1.1. Glycosylation pattern on Trichoplusia ni (High-Five)

Glycans present on proteins produced by High-Five cells are mainly composed of high and pauci-mannose type glycosylation[214–218] (Figure 3.21). Unlike mammalian cells, insect cells

have a limited capability to produce complex glycosylation patterns[195,215] due to the presence of a N-acetylhexosaminidase, an enzyme which hydrolyses the terminal β1-2 N-acetyl-glucosamine (GluNAc), resulting in the formation of fucosylated paucimannose structures[219]. Thus, complex oligosaccharide formation is actively prevented in insect cells, although this has been altered in certain strains through genetic engineering[215,216].

Studies on glycosylation patterns have affirmed that High-Five cells expresses glycoproteins with characteristic core fucosylated $Man_3GlcNac_2$ structure with variable paucimannose structures[195,218,220–222]. This has been affirmed previously in the laboratory by Dr Max Crispin and Dr Matthew Higgins by MALDI-TOF MS for hCD36ED (Figure 3.23). The glycans that were detected for hCD36ED included oligomannose types such as Man_5 – Man_9 species, as well as paucimannose types such as α1-6 fucosylated $Man_3GlcNAc_2$, $Glc_1Man_3GlcNAc_2$, and the $Man_3GlcNAc_2$.

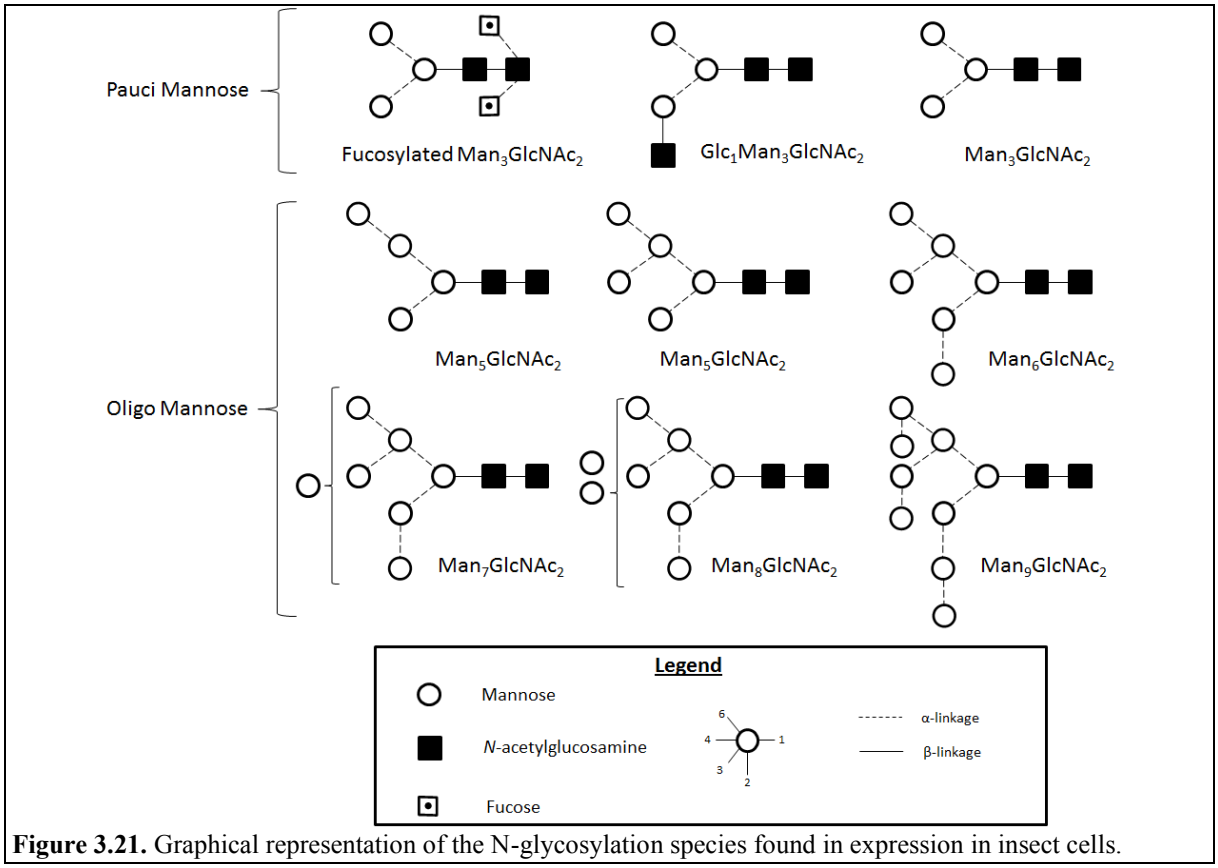


Figure 3.21. Graphical representation of the N-glycosylation species found in expression in insect cells.

3.4.1.2. A deglycosylation strategy

Deglycosylation of CD36 purified from both insect and mammalian cells has often been performed with peptide-N-glycosidases[169,197](PNGase). PNGase F is most commonly used for the complete removal of complex glycans on glycoproteins purified from mammalian cells by hydrolysing the amide bond between the N-linked oligosaccharide from the Asn residue of the N-glycosylation site. However, PNGase F is unable to cleave glycans that have been core fucosylated[197](α1-3 fucose), which is observed frequently in glycoproteins purified from insect cells[214,222]. Moreover, there is a greater propensity for glycoproteins that have been completely deglycosylated by PNGase F to aggregate. This is due to the introduction of a point charge at the cleaved Asn residue which often occurs within hydrophobic patches, resulting in an overall destabilisation of the protein[195,223,224].

A few strategies can generate soluble monodispersed glycoproteins that are deglycosylated for structural analysis[195]. One is to produce soluble glycoproteins that are endoglycosidase sensitive. Unlike PNGase F, endoglycosidases act on the β1-4 linkage di-N-acetylchitobiose bond between the two glucosamine residues that forms the core glycan structure during N-glycosylation. This method of N-glycan cleavage leaves a single cleaved glucosamine linked to the Asn of the N-glycosylation site (Figure 3.22), which may confer stability by coordinating the glycan-protein interaction within the hydrophobic regions on the protein surface[224]. This strategy was successful in crystallising several glycoproteins[225,226].

<u>Endoglycosidase treatment</u>

Glycans added by High-Five cells are dominated by the presence of α1-6 core fucosylated paucimannose species[220], with some instances of α1-3 core fucosylation[222]. Treatment of

these soluble glycoproteins with specific combinations of endoglycosidases can remove the oligomannose and paucimannose structures for structural studies[225,226]. Endoglycosidases are specific for a subset of glycan. For example, Endo H (*Streptococcus plicatus*) cleaves oligomannose and hybrid-type structures[227]. The Endo F series of endoglycosidase (*Flavobacterium meningosepticum*) can cleave oligomannose (Endo F_1), mono- and biantennary glycans (Endo F_2), as well as α1-6 core fucosylated mono-, bi- and triantennary glycans (Endo F_3)[228].

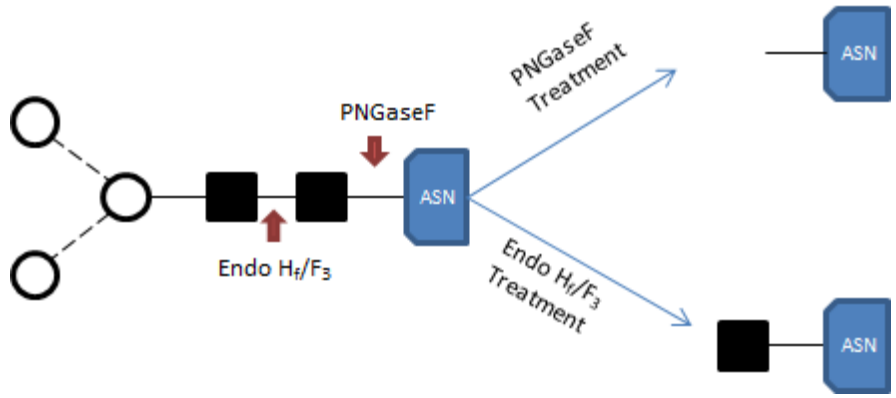


Figure 3.22. Illustration of the outcome of treatment of N-glycosylation with PNGaseF and endoglycosidase H_f or Endo F_3. A single glucosamine residue is left on the asparagine when treated with endoglycosidases but not when treated with PNGase F.

3.4.1.3. Deglycosylation of hCD36ED

In order to produce stable deglycosylated form of hCD36E for structural analysis, the soluble hCD36ED purified from High-Five cells was treated with endoglycosidases. A combination of Endo H_f and F_3 were selected to deglycosylate hCD36ED based on the oligomannose and core fucosylated paucimannose identified in the mass spectrum (Figure 3.23). hCD36ED was treated with Endo H_f and F_3 separately and subjected to MALDI-TOF-MS to identify the species of glycans removed by the respective endoglycosidases (Figure 3.23). The MALDI-TOF-MS experiment was performed by Dr Max Crispin and Dr. Matthew Higgins.

Treatment of hCD36ED with Endo H_f resulted in the decrease in the mass spectrum peaks corresponding to GlcNAc-Man_3-$GlcNAc_2$ and higher mannose species. However, α1-6 fucosylated $Man_3GlcNAc_2$ was not removed by Endo H_f treatment, corresponding to the peak at 1078.7 m/z. On the other hand, treatment of hCD36ED with Endo F_3 resulted in the removal of α1-6 fucosylated paucimannoses (fucosylated $Man_3GlcNAc_2$ (1078.7 m/z) and $Glc_1Man_3GlcNAc_2$ (1281.6 m/z)), but not all the high mannose species. Given that two different subsets of glycans were removed with treatment with the two different endoglycosidases, complete treatment of the hCD36ED with both endoglycosidases will result in complete deglycosylation, leaving behind a single cleaved glucosamine on each glycosylation site.

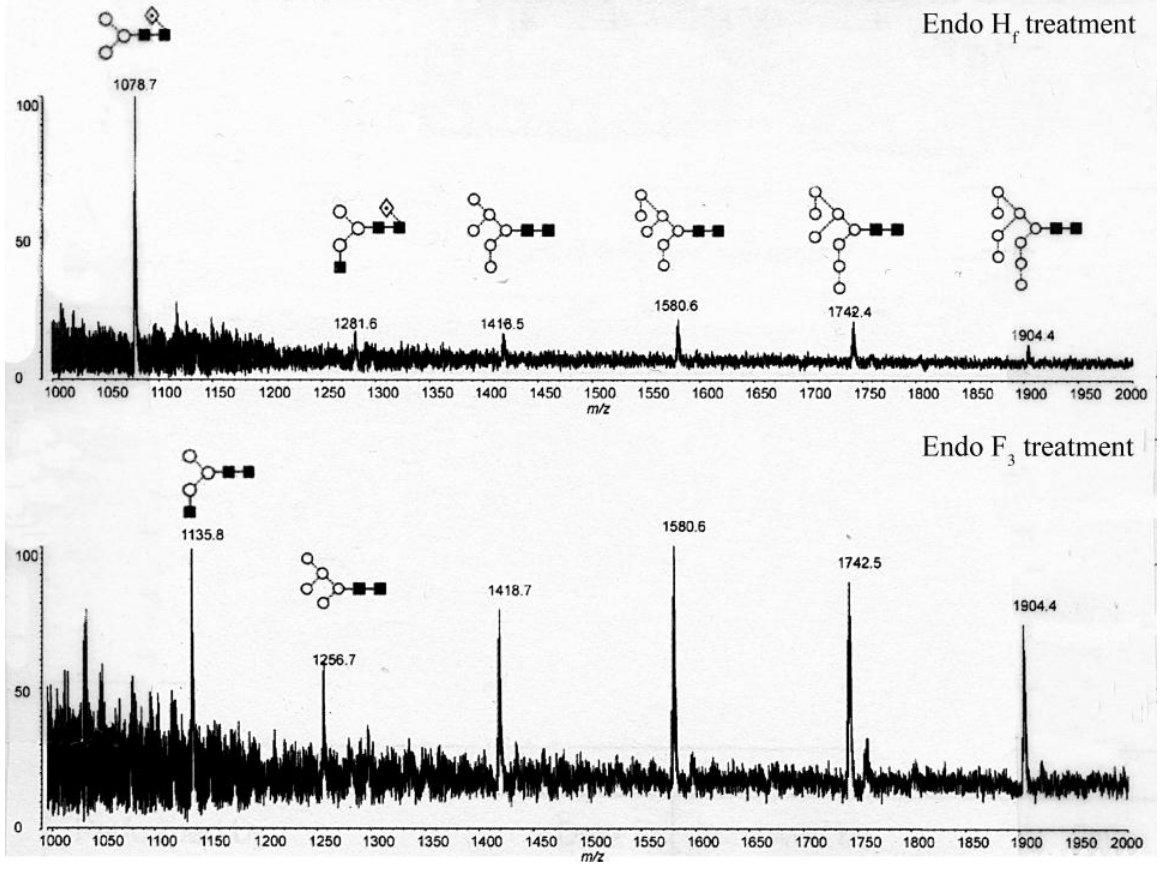


Figure 3.23. Mass spectra of hCD36ED after endoglycosidase treatment using endoglycosidase H_f and F_3. Mass spectrum data was collected by MALDI-TOF Mass spectrometry conducted by Dr. Max Crispin.

Glycosylated hCD36ED was therefore treated for 90 and 180 minutes with 5000 units (1 000 000 units/ml) of Endo H_f, and Endo F_3 at 1:100 (w/w ratio) at 37°C (Figure 3.24a). A shift in size for deglycosylated hCD36ED has occurred after 90 minutes of treatment with the endoglycosidases and no additional shift was seen after 180 minutes. Further purification of hCD36ED post deglycosylation on a preparatory S200 did not register a shift in elution volume although there was a shift in size on the SDS-PAGE (Figure 3.24a,c). This was consistent with the results from analytical size-exclusion chromatography results discussed below.

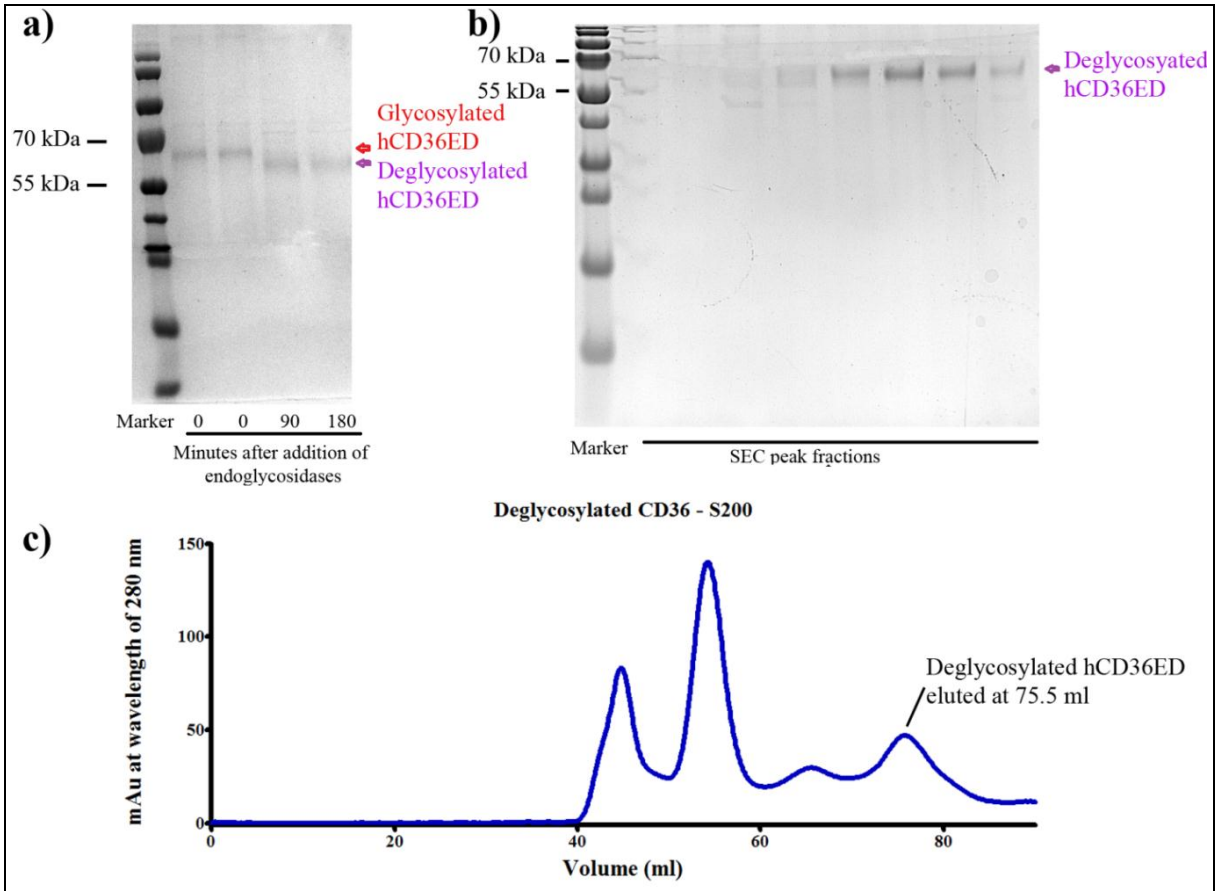


Figure 3.24. Deglycosylation of hCD36ED. **(a)** SDS-PAGE of the deglycosylated material after 90 mins and 180 mins of treatment with 5000 units of both endoglycosidase H_f and F_3. **(b)** SDS-PAGE of the size-exclusion chromatography peak fractions of deglycosylated hCD36ED. **(c)** The size-exclusion chromatography trace for deglycosylated hCD36ED. Deglycosylated hCD36ED eluted at 75.5 ml on a preparatory S200.

3.4.2. Biophysical characterisation of the function and structure of deglycosylated hCD36ED

3.4.2.1. Circular Dichroism spectroscopy

Circular dichroism (CD) spectroscopy was employed to compare the secondary structure of both glycosylated and deglycosylated forms of hCD36ED in solution. This was to affirm that the hCD36ED maintained its fold after deglycosylation.

Proteins are chiral and are able to absorb both left- and right-hand circularly polarised light to different degrees. This is caused by chromophores on the protein such as amide bonds which absorb light of wavelength in the far UV range[134,229,230](180 – 250 nm). A CD spectrum obtained in this range will give insight into the local secondary structures surrounding these backbone amide groups. Both glycosylated and deglycosylated forms of hCD36ED were studied by CD spectroscopy to compare the fold of hCD36ED before and after glycosylation. This also assessed secondary structure elements present in the purified hCD36ED.

To reduce background noise caused by chloride ions, protein samples were buffer exchanged into a CD buffer which has minimal optical activity (20 mM Tris pH 8.0, 0.15 M NaF). The ellipticity (mdeg) was measured from wavelengths 195 nm to 260 nm. An average from ten consecutive runs was taken for each sample. The data were processed by blank subtraction using an averaged spectrum of a buffer blank. The mean residue ellipticity (MRE) of the measured CD spectrums was calculated using the protein concentration.

The complete spectrum of both glycosylated and deglycosylated forms of hCD36ED are overlaid in Figure 3.25. The spectra showed elements of both α-helical (indicated by negative ellipticity in the 208 nm and 218 nm regions) and β-sheet structures (indicated by negative ellipticity in the 222 nm regions). Both spectra were predicted to have 70% α-helix with 5% β-sheet character using the K2D3 algorithm[133]. This suggests that hCD36ED assumes a globular structure with predominantly α-helical character, and that this structure is retained after deglycosylation.

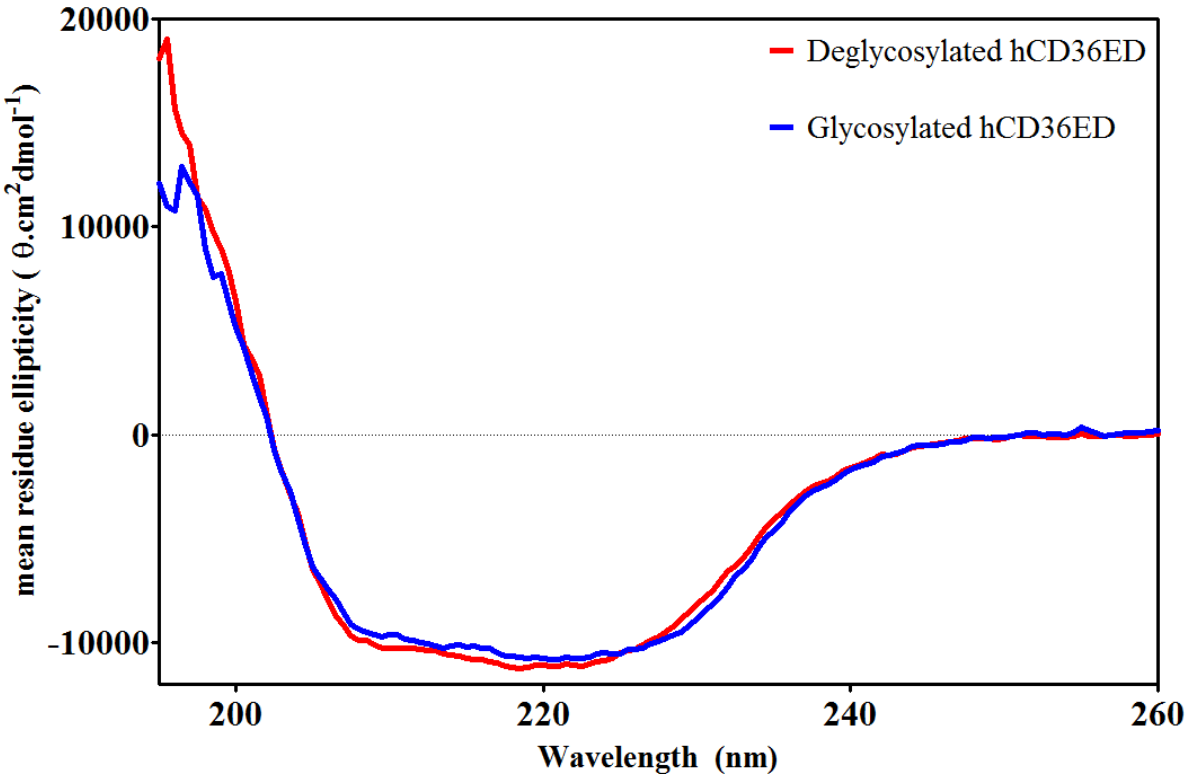


Figure 3.25. Circular dichroism spectroscopy of glycosylated and deglycosylated hCD36ED. A far UV spectrum over the range of wavelength from 195-260 nm is shows the presence of α-helical and β-sheet character found in the proteins. The spectra show that the folded properties did not change for hcD36ED after deglycosylation.

3.4.2.2. Interaction between deglycosylated hCD36ED and CIDR1α domains

The interaction of deglycosylated hCD36ED with MC179 was examined using SPR to affirm the functionality of hCD36ED after deglycosylation. Deglycosylated hCD36ED was

immobilised in flow cell 3 of the same CM5 chip surface to a similar level of glycosylated hCD36ED (5000 RU) as explained in the SPR experiment in chapter 3.2.3.1. Samples injected over flow cell 2 (glycosylated hCD36ED) passed through flow cell 3 (deglycosylated hCD36ED) on the same injection run on the same chip. This allowed for the direct comparison of the response data using the same samples at the same concentration for both deglycosylated and glycosylated hCD36ED. MC179 and PF11CIDR$_{EXT}$ were used to examine the binding function of deglycosylated hCD36ED.

The response for the binding curves were diminished for deglycosylated hCD36ED (Figure 3.26a,b), even though deglycosylated hCD36ED and glycosylated hCD36ED were immobilised to the same level of response on their respective flow cells (5000 RU). This observation suggests that less CIDR1α domains binds to deglycosylated hCD36ED than glycosylated hCD36ED. The K_D observed for MC179 for glycosylated hCD36ED was 33.1(±8) nM but that for deglycosylated hCD36ED was 333.6 (±100) nM. A reduction in affinity for deglycosylated hCD36ED (K_D = 1234.0 (±130) nM) compared to glycosylated hCD36ED (K_D = 710.2.0 (±69) nM) was also observed for PF11CIDR$_{EXT}$.

The reason for the lower response of the CIDR1α constructs to deglycosylated hCD36ED could be that the amount of CIDR1α was reduced through interaction with hCD36ED after passing over flow cell 2, such that the concentration of CIDR1α was reduced before being flowed over flow cell 3. Another possible reason may be that the overall number of available binding sites is reduced on immobilised deglycosylated hCD36ED chip, even though the same amount of protein was present on the chip surface at the start. This could occur if the deglycosylation of hCD36ED alters its ability to interact with the chip surface, resulting in an immobilised orientation that does not accommodate interaction with CIDR1α domains.

Thus, deglycosylated hCD36ED was still able to interact with both CIDR1α domains albeit with reduced affinity. This suggests that the deglycosylated hCD36ED can still be used for co-crystallisation with its CIDR1α binding partner, but indicated a potential role of glycans in the interaction.

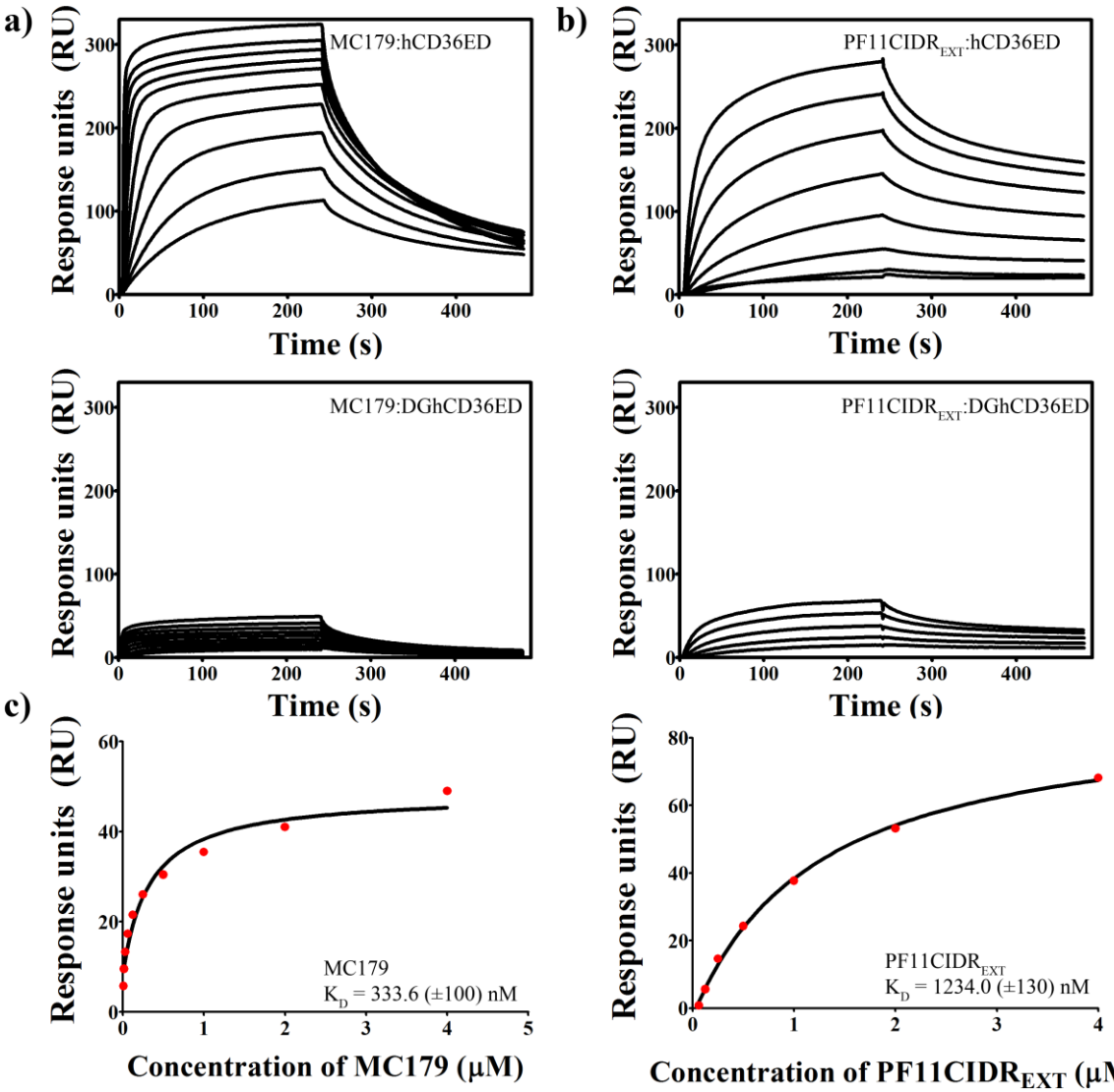


Figure 3.26. Surface Plasmon Resonance measurements of CIDR1α constructs for their interaction with deglycosylated hCD36ED. **(a)** The response curves from the surface plasmon resonance measurements for a concentration series of MC179 for its interaction with glycosylated hCD36ED and deglycosylated DGhCD36ED. **(b)** The response curves from the surface plasmon resonance measurements for a concentration series of PF11CIDR$_{EXT}$ for its interaction with glycosylated hCD36ED and deglycosylated DGhCD36ED. **(c)** The saturated binding curves for the concentration series of MC179 and PF11CIDR$_{EXT}$ for their interaction with deglycosylated DGhCD36ED. The curves were fitted using the BIAevaluation programme using a steady state affinity 1:1 binding model. K_D values were estimated to be the concentration of the analyte at which 50% of the ligand were bound in a saturated binding curve.

3.4.2.3. Analytical size-exclusion chromatography of deglycosylated CD36 and MC179

Analytical size-exclusion chromatography was used to further characterise the complex formed between deglycosylated hCD36ED and CIDR1α-M2 domains as described in chapter 3.2.4.1.

In the previous experiment, an excess of glycosylated hCD36ED compared to MC179 was used to visualise the shift in size. In this experiment with deglycosylated hCD36ED, MC179 was added in excess compared to hCD36ED at a molar ratio of 3:1. There are two advantages to this. Firstly, it provides a reference point to validate the shift in elution volume for the complex. More importantly, it allows us to confirm the stoichiometry of the complex in reference with the previous experiment. If the complex formed between MC179 and deglycosylated hCD36ED assumed a stoichiometry more than a 1:1 (MC179:hCD36ED) ratio, the co-elution volume of the bigger complex will result in a bigger shift towards the left with a bigger signal (mAu at wavelength 280 nm) from binding more MC179, given that the hydrodynamic volume of the overall complex has changed. Excess CIDR1α domains should appear as a peak at elution volume of ~16 ml if it forms a 1:1 complex with hCD36ED.

The analytical size-exclusion chromatography trace for deglycosylated hCD36ED alone, and the complex between deglycosylated CD36 and MC179 are shown in Figure 3.27a. Deglycosylated hCD36ED eluted at 14.90 ml, which is similar to where glycosylated hCD36ED eluted (Figure 3.27b). Additionally, the complex between deglycosylated hCD36ED and MC179 co-eluted at 13.69 ml, suggesting that hCD36ED can indeed form a stable complex with MC179 in the absence of glycosylation. The elution volume of the

complex between deglycosylated hCD36ED and MC179 (13.69 ml) was nearly the same as that of the complex between glycosylated hCD36ED and MC179 (13.63 ml), even in the presence excess MC179. In addition, the excess MC179 was clearly seen at the elution volume of 16.05ml. Therefore the complex formed between a CIDR1α domain and CD36 is in a 1:1 stoichiometry.

In essence, the analytical size-exclusion chromatography experiments confirmed that deglycosylated hCD36ED and MC179 were able to form a complex on an analytical size-exclusion chromatography column albeit the lower affinity discovered through SPR. Additionally, the SDS-PAGE of the co-eluted complex showed only two species. This suggests that the interaction was specific and stable on a size-exclusion chromatography column, which is an important feature for producing clean co-crystallisation complex using preparative size-exclusion chromatography.

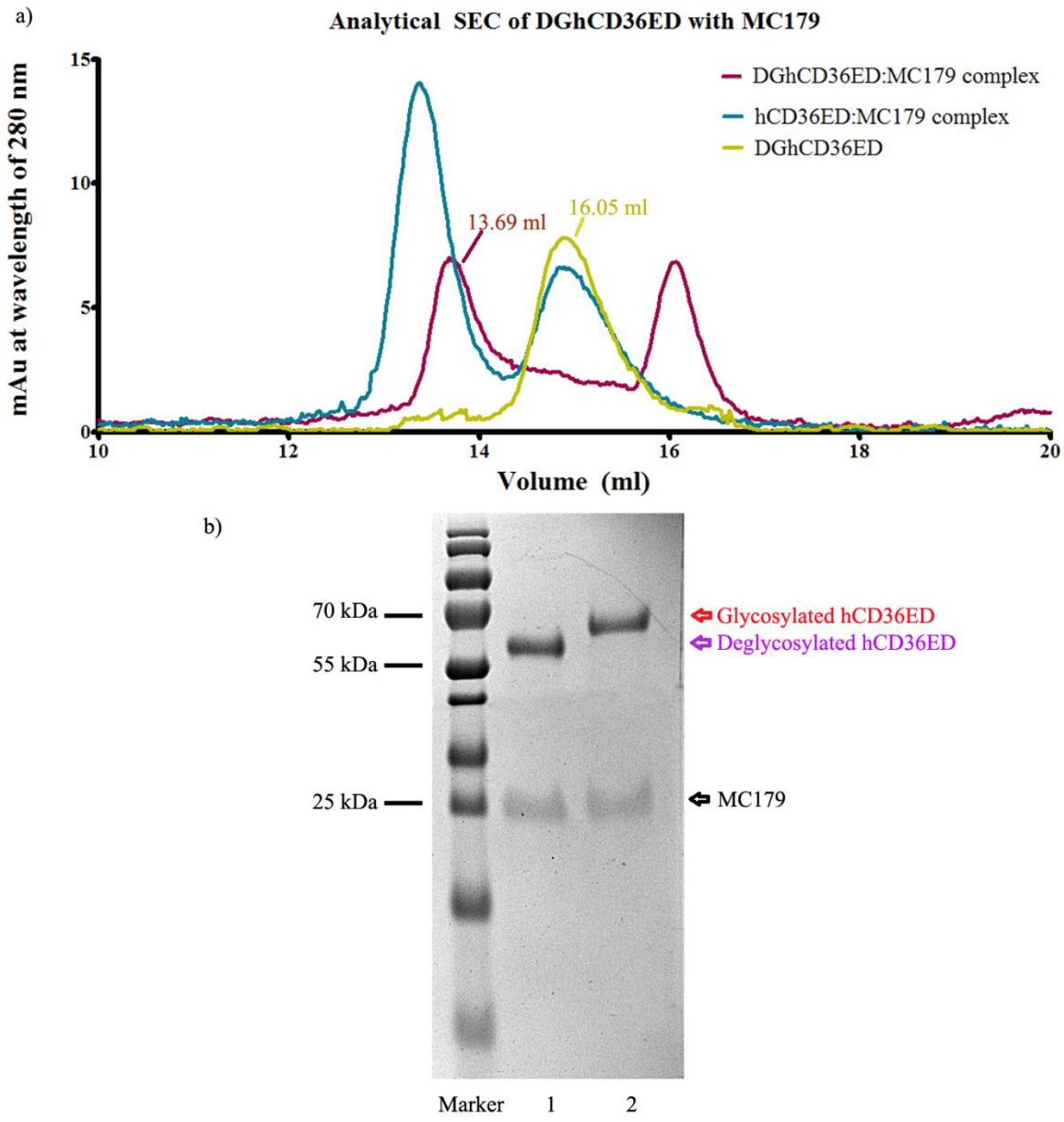


Figure 3.27. Analytical size-exclusion chromatography of MC179 and deglycosylated hCD36ED. **(a)** The size-exclusion chromatography traces of the co-complex (hCD36ED:MC179), the co-complex with deglycosylated hCD36ED (DGhCD36ED:MC179, and deglycosylated hCD36ED (DGhCD36ED) was overlaid on the same axes. The deglycosylated hCD36ED alone and in complex with MC179 eluted at 16.05 ml and 13.69 ml respectively on an analytical S200 column. **(b)** An SDS-PAGE of the elute fraction of the complex involving either deglycosylated hCD36ED (lane 1) or glycosylated hCD36ED (lane 2) with MC179.

3.5. Crystallisation trials

Following biophysical characterisation of the hCD36ED:CIDR1α domain complex, the next step is the determination of a high resolution structure of a complex. Crystallisation trials of the complex between deglycosylated hCD36ED and MC179 were therefore conducted using a variety of crystal screens for complex formation. Additionally, the possibility of crystallising hCD36ED in the presence of oleic acid has also been explored by both analytical size-exclusion chromatography and fluorescence-based thermal shift assays.

3.5.1. Can lipid molecules block CIDR1α interaction with CD36?

Oleic acid (OA) has been shown to be one of the native long chain fatty acids that is transported by CD36[143]. A competition experiment was designed using analytical size-exclusion chromatography to understand whether native ligand can disrupt the interaction between CD36 and CIDR1α domains. Deglycosylated hCD36ED was incubated with OA at a 1:100 (hCD36ED:OA) molar ratio for 1 hour. Oleic acid was dissolved in PBS with 1% ethanol. MC179 was then added to this mixture and incubated for another 30 minutes at a 1:1 molar ratio (MC179:hCD36ED). The resulting mixture was injected on an analytical S200 column.

The resulting size-exclusion chromatography trace is featured in Figure 3.28. Deglycosylated hCD36ED was able to form a complex with that MC179 in the presence of OA. The elution peak at 13.52 ml was also close to the elution peak estimated for the complex between deglycosylated hCD36ED and MC179 from previous experiments (Figure 3.28). This

suggested that deglycosylated hCD36ED was able to interact with MC179 in the presence of OA.

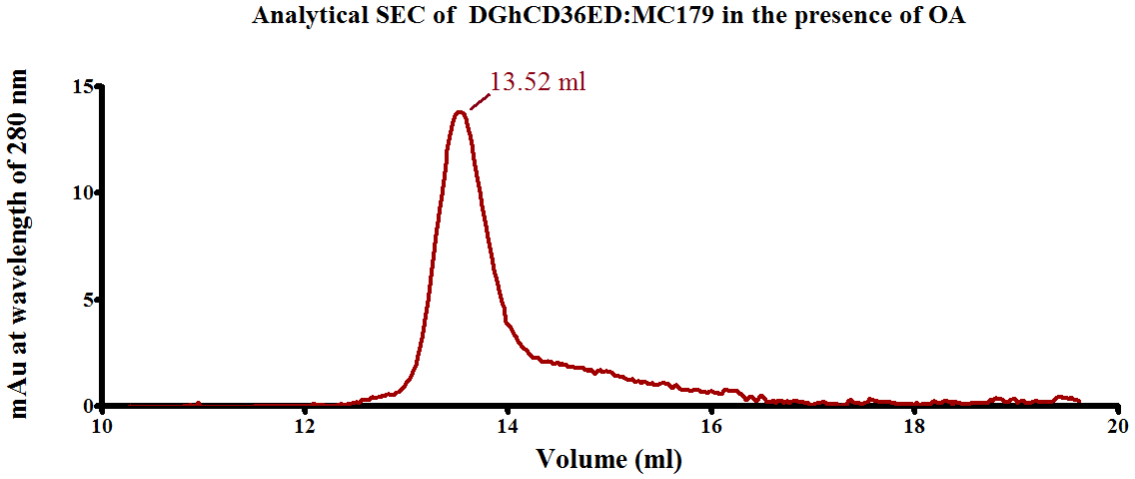


Figure 3.28. The size-exclusion chromatography trace of the co-complex between deglycosylated hCD36ED and MC179 in the presence of oleic acid (OA). The deglycosylated hCD36ED eluted in complex with MC179 at 13.52 ml on an analytical S200 column.

3.5.2. Fluorescence-based thermal shift stability assay

3.5.2.1. Investigation of hCD36ED stability using buffer screen

A thermal shift assay was first conducted to ascertain the stability and melting temperature (T_m) of hCD36ED in different buffers, ionic conditions and pH according to the Table 3.5.

The unfolding of hCD36ED was monitored by a scanning fluorimetry on a 96 well plate. As temperature increases, the amount of exposed hydrophobic surface increases as a result of protein unfolding. A fluorescent dye (SYPRO® Orange) that increases fluorescence emission in non-polar conditions can bind to these surfaces. Hence, fluorescent emission increases with protein unfolding. The excitation wavelength of SYPRO® Orange is 492 nm, and the emission wavelength detected is 610 nm. The T_m was determined by the level of

fluorescent emission (dR) as a probe for unfolding over a temperature range. T_m is defined as the temperature at which the concentration of unfolding protein is the same as that of folded proteins i.e, $\Delta G_{folding} = 0$.

Experiments were performed in a 96 well format. The top volume for each well was 50 µl. In each well, there was 1 µl of protein (10 µM) in 47 µl of each buffer (Table 3.5). 2 µl of SYPRO® Orange 50X stock solution was added to each well and pipetted to ensure proper mixing. For control experiments, the buffer screen was repeated on the same 96 well plate with 48 µl of buffer and 2 µl of SYPRO® Orange 50X stock solution without any proteins. A quantitative PCR machine was used to measure the emission wavelength of SYPRO® Orange at 610 nm for the temperature range from 25°C to 80°C.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20 mM MES pH 5.2 0.05 M NaCl	20 mM MES pH 5.2 0.15 M NaCl	20 mM MES pH 5.2 0.3 M NaCl	20 mM MES pH 5.7 0.05 M NaCl	20 mM MES pH 5.7 0.15 M NaCl	20 mM MES pH 5.7 0.3 M NaCl	20 mM MES pH 6.2 0.05 M NaCl	20 mM MES pH 6.2 0.15 M NaCl	20 mM MES pH 6.2 0.3 M NaCl	20 mM MES pH 6.5 0.05 M NaCl	20 mM MES pH 6.5 0.15 M NaCl	20 mM MES pH 6.5 0.3 M NaCl
B	20 mM HEPES pH 7 0.05 M NaCl	20 mM HEPES pH 7 0.15 M NaCl	20 mM HEPES pH 7 0.3 M NaCl	20 mM HEPES pH 7.5 0.05 M NaCl	20 mM HEPES pH 7.5 0.15 M NaCl	20 mM HEPES pH 7.5 0.3 M NaCl	20 mM HEPES pH 8.0 0.05 M NaCl	20 mM HEPES pH 8.0 0.15 M NaCl	20 mM HEPES pH 8.0 0.3 M NaCl	20 mM HEPES pH 8.5 0.05 M NaCl	20 mM HEPES pH 8.5 0.15 M NaCl	20 mM HEPES pH 8.5 0.3 M NaCl
C	20 mM TRIS pH 7 0.05 M NaCl	20 mM TRIS pH 7 0.15 M NaCl	20 mM TRIS pH 7 0.3 M NaCl	20 mM TRIS pH 7.5 0.05 M NaCl	20 mM TRIS pH 7.5 0.15 M NaCl	20 mM TRIS pH 7.5 0.3 M NaCl	20 mM TRIS pH 8.0 0.05 M NaCl	20 mM TRIS pH 8.0 0.15 M NaCl	20 mM TRIS pH 8.0 0.3 M NaCl	20 mM TRIS pH 8.5 0.05 M NaCl	20 mM TRIS pH 8.5 0.15 M NaCl	20 mM TRIS pH 8.5 0.3 M NaCl

Table 3.5. Buffer screen in a 96-well format for hCD36ED stability thermal shift assay.

A striking observation from the thermal shift assays was the complex shape of the melting curves. An example of a series of melting curves for hCD36ED in Tris buffer (Table 3.5) is shown in Figure 3.29 (data for MES and HEPEs buffer not shown). The shape of the melting curves obtained suggested complex unfolding properties for hCD36ED. The complexity of

the curves also made it difficult to account for shifts in T_m in the various buffers although the phase in the 50-60°C region most likely represents the unfolding transition.

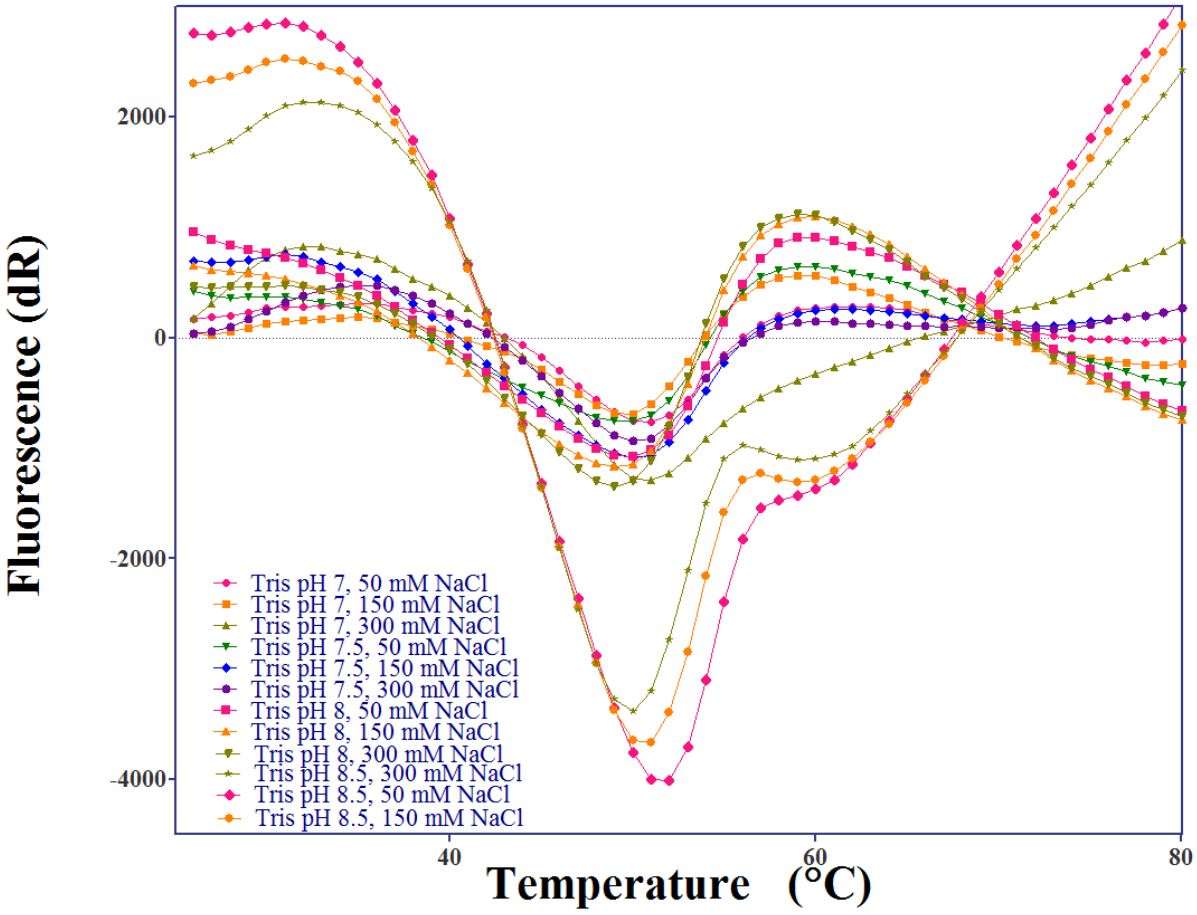


Figure 3.29. Example of a complex melting curve of hCD36ED in Tris-based buffers (according to table 3.5)

3.5.2.2. Thermal shift of hCD36ED in the presence of oleic acid

An interesting feature observed from Figure 3.29 was the tilted shape of the melting curves. The initial downward sloping curve at low temperatures suggested that the protein might be interacting with the fluorescent dye even before unfolding has taken place, since the amount of signal due to SYPRO® Orange emission was decreasing. Given that hCD36ED was involved in transportation of lipids across membranes[231], one possibility is that there might

be a hydrophobic cleft to which the lipids could bind. Thus, it is possible that this accessible hydrophobic cleft can bind to SYPRO® Orange even when the protein is folded at low temperatures.

Oleic acid (OA) was added to hCD36ED at three buffer conditions (Table 3.6) to investigate if the presence of OA could stabilise the protein unfolding curve. If the postulation that hCD36ED indeed has a hydrophobic cleft that could potentially interact with OA is true, addition of OA would reduce the upward sloping trend at low temperatures observed for the apoprotein. The experimental procedure described above was employed. OA mixture was prepared to a concentration of 1 mM in PBS with ethanol (10 mM Phosphate pH 7.4, 27 mM KCl, 0.137 M NaCl, 1% Ethanol). OA was added to hCD36ED at 1:100 (hCD36ED:OA) molar ratio according to Table 3.6. Buffer blanks were performed by replacing OA and protein using the same volume of their respective buffers. The 96 well plate was left to incubate for 30 minutes are room temperature before starting the run. Data was analysed after subtraction of the controls from the buffer run and OA alone in buffer.

	20 mM Tris pH 7.5, 0.15 M NaCl			20 mM Tris pH 8.5, 0.15 M NaCl			20 mM HEPES pH 8.5, 0.3 M NaCl		
	hCD36ED only	Oleic acid only	OA + hCD36ED	hCD36ED only	Oleic acid only	OA + hCD36ED	hCD36ED only	Oleic acid only	OA + hCD36ED
Buffer	46.5	45.5	45.5	46.5	45.5	45.5	46.5	45.5	45.5
100% Ethanol	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Protein (10µM)	1	1	1	1	1	1	1	1	1
1 mM Oleic acid mixture (PBS, 1% ethanol)	0	1	1	0	1	1	0	1	1
50X Sypro orange solution	2	2	2	2	2	2	2	2	2
TOTAL	50	50	50	50	50	50	50	50	50

Table 3.6. Lipid binding experiment in a 96-well format for hCD36ED stability thermal shift assay.

OA was shown not to bind SYPRO® Orange on its own in buffer (Figure 3.30b). However, when OA was added to hCD36ED at 1:100 (hCD36ED:OA) ratio, the downward sloping curve of hCD36ED observed without OA was replaced by a flat fluorescence signal (Figure 3.30a). This suggested that the presence of oleic acid was able eliminate the binding of SYPRO® Orange to hCD36ED at low temperatures below the Tm. Hence, the postulation that the hCD36ED has a hydrophobic cleft might be true. More experiments would have to be performed to investigate the lipid binding properties of hCD36ED. The last observation was that there was no apparent shift in T_m of hCD36ED even in the presence of OA (Figure 3.30a).

Thus, these thermal shift assays suggest a model whereby the presence of OA might stabilise the hCD36ED at low temperatures by binding to an otherwise exposed hydrophobic cleft.

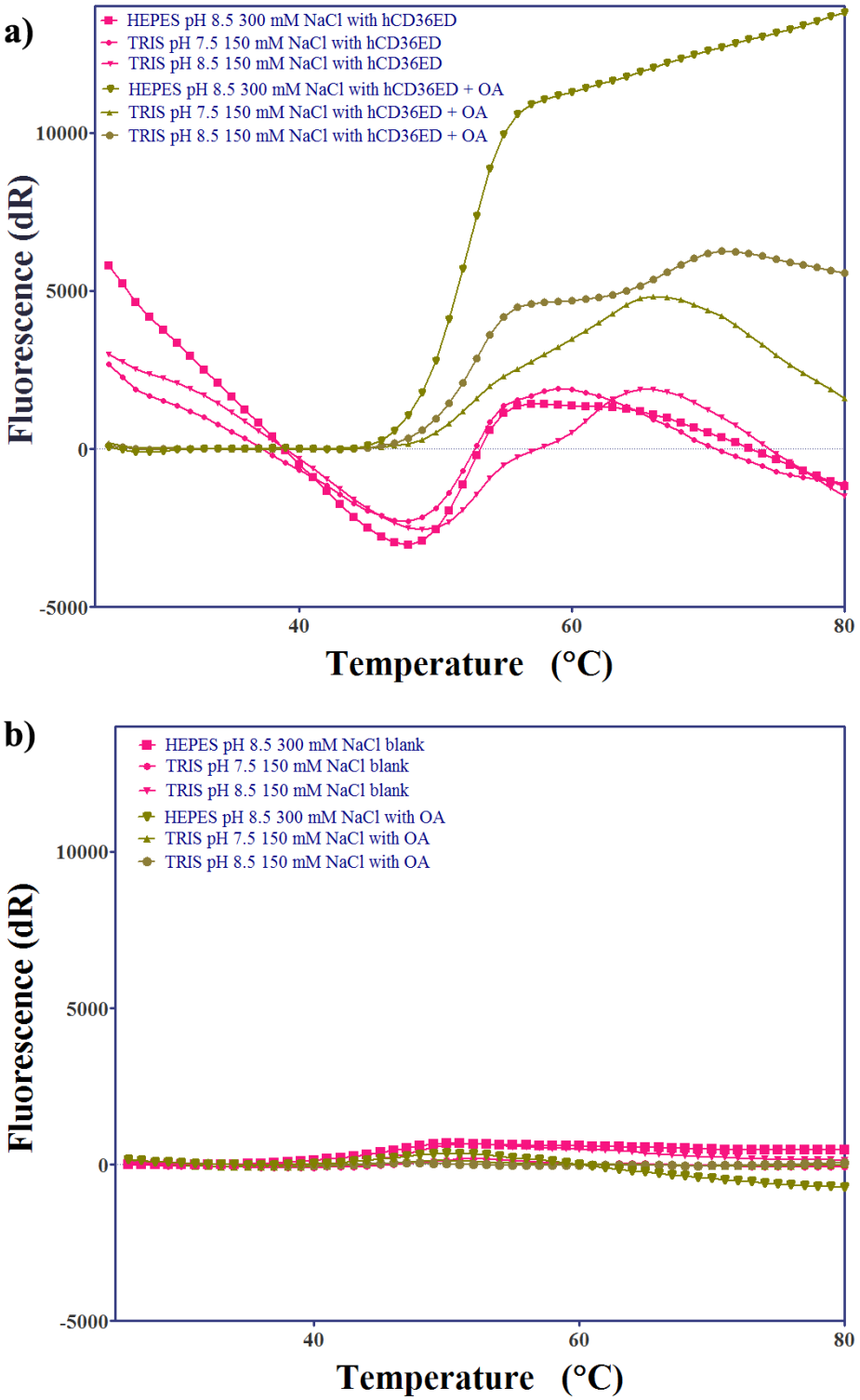


Figure 3.30. Thermal shift assays of hCD36ED in the presence of oleic acid **(a)** The thermal shift experiments involving hCD36ED with (olive) and without (pink) oleic acid. There is a flattening of the initial downward sloping curve in the presence of oleic acid. **(b)** The thermal shift assay of buffer blank (pink) and buffer with oleic acid (olive). This shows that oleic acid does not bind to SYPRO® Orange.

3.5.3. Crystallisation screens

3.5.3.1. Preparation of the material for crystallisation

The co-complex between deglycosylated hCD36ED and MC179 were prepared using size-exclusion chromatography on a preparative S200 column (Figure 3.31b). The elution peak fraction of the complex (69.3 ml) was visibly shifted compared to that of the deglycosylated hCD36ED on its own (75.9 ml) (Figure 3.4c). A peak corresponding to excess MC179 could also be seen on the size-exclusion chromatography trace (84.8 ml). The identity of the complex was confirmed by SDS-PAGE.

3.5.3.2. Crystallisation trials

The chosen crystal screens included sparse matrix (ProPlex™ and JCSG-plus™), ionic screening kits (PACT premier™), and reagents based on successful crystallisation conditions found from PDB (Morpheus™). Although heavy precipitation could be seen in some of the wells, there were no observed crystals. Further crystallisation trials of deglycosylated hCD36ED on its own and with oleic acid were being conducted at the point of thesis submission to observe if the presence of oleic acid can result in stabilisation and crystallisation of hCD36ED at low temperatures.

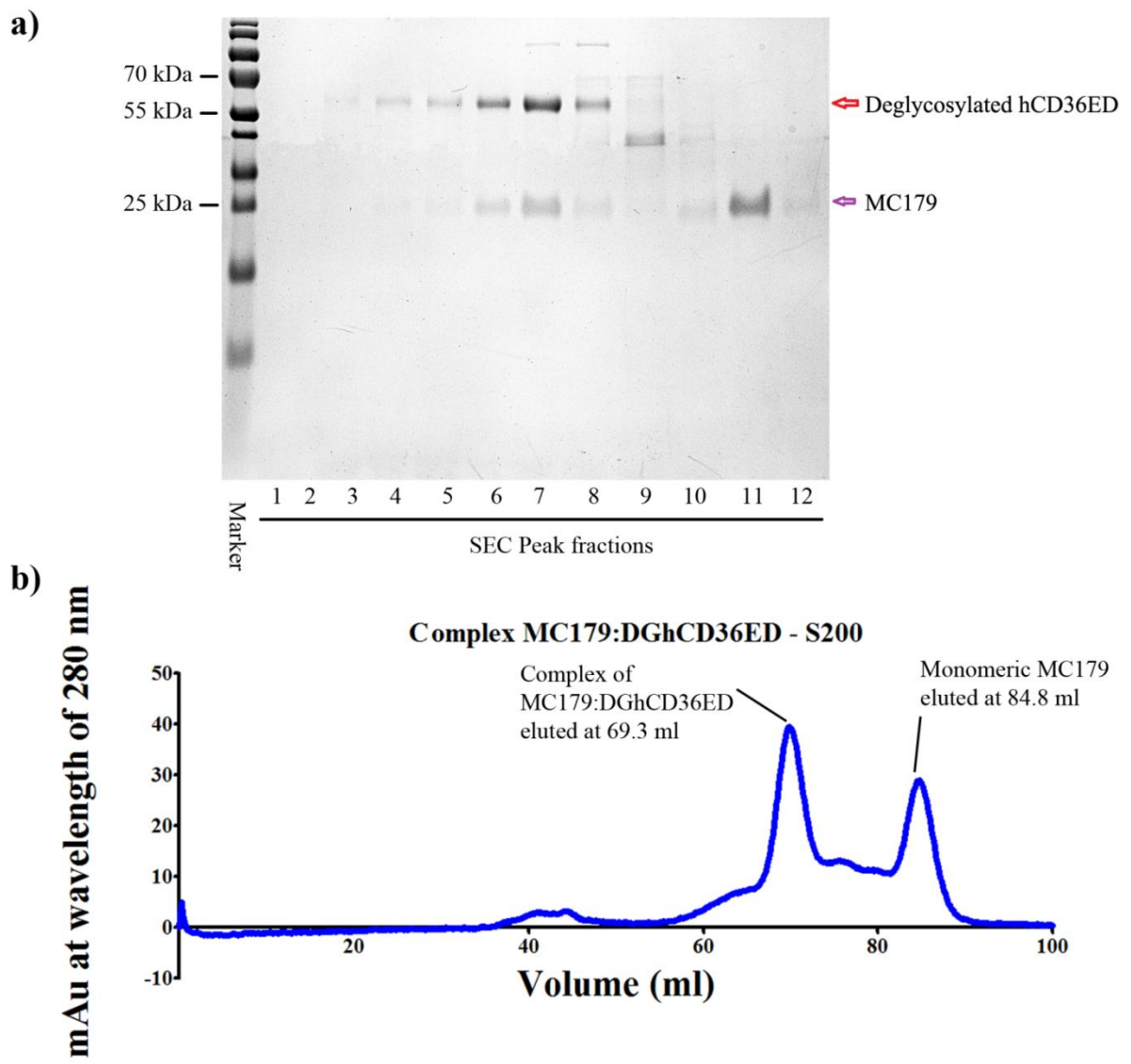


Figure 3.31. Purification process of the co-crystallisation complex between deglycosylated hCD36ED and MC179. **(a)** SDS-PAGE of the peak fractions from preparative S200 the complex between hCD36ED and MC179 could be clearly seen in lanes 6 – 8. Excess MC179 was detected in lanes 10 – 12. **(b)** The size-exclusion chromatography trace for purified between deglycosylated hCD36ED and MC179. The complex eluted at a volume of 69.3ml on a preparative S200 column.

3.6. Consolidation of binding interaction studies between hCD36ED and PfEMP1

The interaction between PfEMP1 and CD36 has been characterised by a variety of biophysical methods. These data have allowed us to conclude the following.

Firstly, CIDR1α binds to CD36 with an affinity ranging from low to high nanomolar. This specific binding interaction was observed with a variety of constructs from different PfEMP1s that were suggested to bind to CD36 using SPR measurements. These results were also observed using analytical size-exclusion chromatography and using AUC.

Secondly, results from analytical size-exclusion chromatography and AUC show that hCD36ED interacts with MC179 in a 1:1 stoichiometry. Given that every PfEMP1 protein has only one CIDR1α domain, this information can be extrapolated for the full length PfEMP1 constructs as well. Further validation of this stoichiometry can be confirmed through structural studies.

Thirdly, in the case of PF11_0007 the binding interaction of a single CIDR1α domain was similar to that of a full length PfEMP1. This is in contrast to var2csa, for which the intact extracellular domain was able to bind to CSA with higher affinity than any individual CSA-binding DBL domain. Although more experiments will need to be performed on other PfEMP1 domains to validate these findings, these biophysical data substantiate the central role of CIDR1α in mediating the adhesion of PfEMP1 infected erythrocytes to CD36 of endothelial tissues.

Lastly, SPR and analytical size-exclusion chromatography experiments showed clearly that hCD36ED was able to interact with MC179 after deglycosylation. The secondary structures of hCD36ED were confirmed by circular dichroism to unaffected by deglycosylation. Furthermore, complex containing glycosylated or deglycosylated CD36 co-eluted on a size-exclusion chromatography column to produce a co-crystallisation complex between deglycoslyated hCD36ED and MC179. Thermal shift assays also suggested the possibility of producing co-crystallisation complexes between deglycosylated hCD36ED and its native ligand oleic acid.

Chapter 4

Biophysical characterisation of the interaction between IgM and PfEMP1

4.1. IgM and severe malaria

4.1.1. Human IgM and its role in immune surveillance

Natural human serum IgM is a polyreactive immunoglobulin produced by peritoneal B-1 lyphocytes as an important initial defence against infection by bacteria and viruses by acting as a conduit between innate and adaptive immune system[232–234]. IgM is especially potent in neutralising blood pathogens, preventing them from transferring from an infected erythrocyte to an uninfected erythrocyte[235]. Recognition of antigens by IgM involves the fragment antigen binding (Fab) domains of IgM. The cellular consequences include ligand neutralisation, phagocytosis and as well as complement activation and result in the interaction between the fragment crystallisable (Fc) region of the IgM and their respective Fc receptors (FcR) on cells involved in immunity.

4.1.1.1. Structure of IgM

Human IgM can be made in two forms; a transmembrane monomeric IgM expressed on B cells, as well as multimeric secreted IgM.

Structure of the IgM monomer

Monomeric IgM is composed of a homodimer of two heavy chains, each attached to a light chain. Immunoglobulin classes are defined by their heavy chain.

The heavy chain of IgM (μ-chain) is made up of five domains. The first domain is a variable μ domain (Vμ). This is followed by a constant region of the μ-chain made up of four domains, denoted Cμ1 – Cμ4[236] (Figure 4.1a). Monomeric IgM is coordinated by the formation of both inter- and intra-chain disulphide bridges. The inter-chain disulphide bridge is formed between C140 of the Cμ1 and the light chain[237]. The intra-chain disulphide bridge is formed between C337 of two adjacent μ-chains[238]. Studies on the structure of individual portions of IgM is conducted by IgM fragmentation either through trypsin[239] or papain[240].

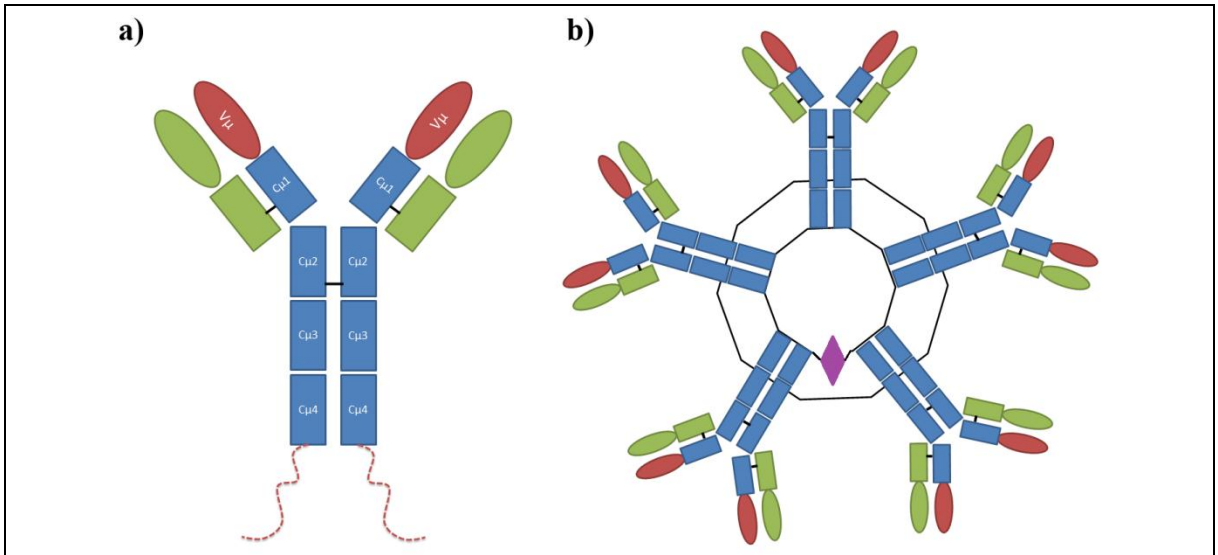


Figure 4.1. Graphical representation of the monomeric **(a)** and pentameric **(b)** human IgM. The variable μ-domains (Vμ) are denoted as red ovals. Constant μ (Cμ) domains are denoted as blue boxes. Light chains are coloured in green. Red dashed lines are the transmembrane portions only found in monomeric IgM. Disulphide bonds are shown by black solid lines. The J chain is denoted as the purple diamond in pentameric IgM.

Human IgM is N-glycosylated at five different sites, with three of them carrying complex glycans and another two carrying oligomannose glycans. The glycosylation of human IgM is important for both B-lymphocyte antigen presentation and formation of secretory IgM[241].

Monomeric IgM is only present as a transmembrane receptor on B-lymphocytes, attached by a transmembrane polypeptide C-terminal to Cμ4[242]. In secreted pentameric IgM, this transmembrane region is not present.

Structure of multimeric IgM.

The structure of human IgM has been studied by a few methods. The earliest evidence that supports the pentameric form of IgM came from electron microscopy[243]. Structural studies using small angle x-ray scattering (SAXS) also suggested that IgM forms a star shaped pentamer in solution[240]. The μ-chain has 86% sequence similarity as the ε-chain of IgE. Structural modelling of human IgM using IgE-Fc crystal structure postulated that the human IgM pentamer is a flexible, non-planar complex[244]. This model is supported by the use of cryo-atomic force microscopy (cryo-AFM), which reproduced the uneven surface contours of the individual complex. Additionally, cryo-AFM studies also showed that most of the IgM complex exists in pentameric form. Observations of complexes with more than five Fab arms are likely to be due to the separation of Fab domains in the complex.

Pentameric secreted IgM contains five 7S subunits and a single J chain[245]. It has 10 antigen binding sites and has a combined sedimentation coefficient of 19S[246]. The J chain is a 15 kDa protein involved in promoting the pentameric conformation of secreted IgM, which is coordinated by two sets of disulphide bonds with two different cysteine residues in the Cμ region (Figure 4.1b). The first set of disulphide bonds are associated with the C575 of a

"tailpiece" region C terminal to the Cμ4 domain. Disulphide bond formation between inter-subunit μ-chains at C575 allow for the formation of a multimeric IgM complex. J chain is able to insert in between two neighbouring IgM subunit and form a disulphide bonds each with both of them, thus coordinating the formation of a pentameric IgM with a J chain[247,248]. The second set of disulphide bonds are formed between C414 of the Cμ3 chains of each subunit[237,238]. The J chain is not involved in the second set of disulphide bonds.

Despite the individual studies suggesting that IgM complex exist as pentamers, certain studies have identified IgM hexamers[249–253] which lack J chains[250,251,254]. Although hexameric IgM complexes initiate the complement cascade with greater efficiency than pentamers[249,251,252,255], the physiological function of these hexameric species are not well established[253].

4.1.1.2. FcR and the IgM effector mechanism

The FcR for IgM (FcμR) is ectopically expressed on human cells. FcμR is expressed on a defined population of human B, T, natural killer (NK) and phagocytic cells[256]. The FcμR cDNA in humans suggests that FcμR is a transmembrane sialoglycoprotein which binds to $Fc_5\mu$ fragments of IgM specifically[256]. Binding of IgM to the FcμR during an infection when the host is challenged antigenically results in enhanced humoral as well as innate response.

Fcα/μR is a human receptor for both IgM and IgA[257,258] but is better defined in its cellular action in both bacterial and viral infections[257]. It is found largely on a certain subset of B cells[259] and follicular dendritic cells[260] and is involved in antigenic presentation and B cell selection during the onset of humoral response[256,261]. Sequence comparison of the

immunoglobulin domain between the Fcα/μR and FcμR suggests distinct modes of interaction with natural human IgM[256].

Binding of the IgM Fc region and FcR on various cell types initiates effector mechanisms that are involved in IgM elicited immunity. The protective roles of IgM are diverse. It can act in various stages of innate and adaptive immunity. On top of recognising bacteria and virus infected cells, evidence for IgM in complement mediated phagocytosis of via recognition of self-apoptotic cells are well established[262]. Murine models of IgM protection also showed that serum IgM participates in anti-inflammatory pathways during innate immunity and similar mechanisms have been identified in protecting humans against several diseases such as atherosclerosis[263].

FcR inhibition – an evolved mechanism of immune evasion

FcR inhibition has evolved as an important approach for the evasion of Fc-mediated destruction during bacterial and viral infections. The secretion of proteins that sequester FcR interferes with the interaction between FcR and pathogen recognising antibodies, and hence dampens that effector mechanism of antibodies. This mechanism of immune evasion is well documented for IgA and IgG in various bacterial and viral infection systems[264,265]. Likewise, IgM binding proteins have been described for several species of parasites, including *Toxoplasma gondii*, *Trypanosomatidae*, as well as for *Plasmodium* species[266].

4.1.2. The pathogenic role of IgM in malaria

IgM has been shown to be able to bind to *P. falciparum* infected erythrocytes[126,127,129,267,268]. Characterisation of the interaction has identified PfEMP1 as the main binding partner to IgM.

4.1.2.1. Interaction between IgM and PfEMP1

IgM interaction with DBL domains

Numerous domains on PfEMP1 have been identified to interact with IgM. These includes both DBL[125–127,267,268] and CIDR domains[269]. These interactions are largely characterised by immunofluorescence assay (IFA) and enzyme-linked immunosorbent assay (ELISA) methods.

Structural modelling suggests that the IgM monomer-monomer crevice interacts with loops in the S2 region of the DBL domain[127,270]. Inspection of the modelled interaction surface revealed important charged residues conserved only within IgM binding DBL that interact with complementary charged residues in the IgM binding site. These residues on IgM are in the same two conserved loops which disrupted PfEMP1 binding when mutated explained in details below[127].

Another interesting feature is that the DBL domains interacting with IgM are proximal to the infected erythrocyte membrane within their respective PfEMP1 architectures. This suggests why IgM binding does not compromise the intrinsic cytoadherent properties of PfEMP1, which is often mediated by domains at the membrane-distal end of the protein. This was

suggested for the CSA binding DBL5ε domain of *var*2CSA, which can bind to IgM even in the presence of free soluble CSA. Although this suggests that IgM binding might occur simultaneously with sequestration, this has yet to be validated *in vivo*.

IgM binding site for PfEMP1

The binding site on the IgM for PfEMP1 is mapped to the Cμ4 domain[127]. Similar overlapping regions are also implicated in binding to Fc-binding proteins for IgG and IgA in bacterial and viral systems[127,264,266,271].

This Cμ4 domain found on polymeric IgM has been shown to be a common binding site for a series of domains from PfEMP1 of different strains by binding studies involving IFA and ELISA[127]. This suggests the interaction of PfEMP1 with IgM is shared between PfEMP1 with different cytoadherent properties (E.g. CSA binding or rosetting). Additionally, soluble CSA does not inhibit the binding of IgM to CSA binding DBL domains[270]. This suggests that PfEMP1 might have evolved to interact with IgM under a separate pressure from that for their adherent properties towards other ligands.

Residues on two exposed loops (PNRV (Residues 444 – 447) and PLSP (394 – 397)) of the Cμ4 region are involved in this binding interaction[127]. Sequence comparison show a difference in these two loops between human IgM, and both bovine and mouse IgM. Mutation of these loops to mimic mouse IgM disrupted binding of the PfEMP1 by both ELISA and IFA[127]. Interestingly, the PNRV loop has been mapped to homologous regions on the IgG and IgA that interacts with Fc-binding proteins secreted by bacteria[264].

The interaction between IgM and PfEMP1 requires the polymeric nature of IgM. PfEMP1 do not bind to monomeric IgM, or to domain-swapped antibodies that consist of IgG molecule with a IgM Fc tailpiece[127]. Thus, PfEMP1 recognises the structural features on the polymeric IgM in addition to the presence of conserved residues on the exposed loops. Structural modelling of the binding between DBL4β domain and pentameric IgM suggested that the DBL4β is docked in the crevice between a monomer-monomer interface of the IgM. This might hence explain for the importance of multivalency in IgM for PfEMP1 binding.

4.1.2.2. Functional implications of the interaction between IgM and PfEMP1

There are a number of models that describe the functional contributions of PfEMP1 as an IgM Fcμ sequestering protein during severe malaria. In this section, I will briefly discuss how this interaction could contribute to severe malaria via various processes.

Interference with effector mechanisms

Sequestration of IgM Fc can result in interference of immune surveillance by various B lymphocytes, T lymphocytes and NK cells carrying the effector FcR[256]. IgM is required for the formation of an immunological synapse between FcμR expressing cytotoxic T cells and B cells that provide immunity against parasites, disruption of which will interfere with cytotoxicity. In addition, IgM activation of Fcα/μR is involved in antigen presentation and B cell maturation in follicular dendritic cells, as well as T –independent humoral response. The binding site of IgM for Fcα/μR and PfEMP1 proteins are shown to be closely overlapping[261]. IgM binding to Fcα/μR could therefore be blocked by PfEMP1.

Severe malaria is associated with polyclonal B-cell activation leading to hypergamaglobulinemia[128]. It has been previously shown that Cμ4 binding ligands are required in mediating B cell clustering via the monomeric IgM like B cell receptors for B cell activation[272]. This unimpeded activation of B-cells might be linked to the Cμ4 binding ability of PfEMP1 which causes pathology during severe malaria.

Prevention of antigenic recognition

IgM is able to trigger the classical complement cascade after antigenic recognition of surface antigens on infected cells[273]. This is due to the binding of complement factor C1q to the Cμ3 domain, which is adjacent to the Cμ4 domain that interacts with PfEMP1[244]. One postulation is that by binding PfEMP1, the Cμ3 binding site is rendered inaccessible by the C1q factors and therefore disrupts complement activation and cell lysis.

The binding of IgM to the surface of infected erythrocytes might also prevent opsonisation, which is the recognition of surface antigens by molecules that enhances immune response. This might hence directly prevent phagocytosis of infected erythrocytes. Immunisation has been shown to be effective in increasing immunity against *P. falciparum* infection[274]. Recognition of surface antigens by immunoglobulin and subsequent effector mechanism through the FcR such as FcγRI are important in rendering protection against malaria[275]. The binding of large multivalent IgM molecules on the surface might therefore prevent further opsonisation and hence allow masking of critical surface antigens from being recognised by the humoral immune response[276].

Rosetting and placental malaria

The ability of infected erythrocytes to form rosettes correlates with the development of severe malaria[77,82,126]. IgM has been detected in field isolates that contains rosettes[277–280]. Some models suggests that the polymeric IgM may act as a bridging agent that stabilise the formation of rosettes between infected and uninfected erythrocytes[278,279]. One example of a receptor that IgM could bind to during rosette formation is complement receptor 1 (CR1) on uninfected erythrocytes[77,126]. Taken together, a possible model for pathogenesis during severe malaria would be through the formation of rosettes via the interaction of PfEMP1 to IgM[126].

Additionally, high levels of IgM has also been found in field isolates derived from placental malaria[129,268]. Single domains on PfEMP1 encoded by *var* genes from CSA binding strains (*var*2CSA and *var*1CSA) have been implicated in IgM binding[267]. Despite studies having demonstrated strong relationship between rosette formation, IgM binding and placental malaria[126,129], causal links between rosette formation and placental malaria are not well established[281]. Thus, the role of IgM binding in placental malaria remains to be discovered.

4.1.3. Project aims

The work described here investigates the plausibility of characterising the interaction between PfEMP1 and human IgM structurally. In order to do that, preparation of a single homogenous, monodispersed species of IgM Fc core is of paramount importance.

Crystallisation of IgM is difficult for several reasons. Firstly, crystallisation of a flexible multimeric IgM complex is challenging because of its large size and flexibility. Secondly, it is known that IgM is able to form different multimeric species[252,253]. An IgM complex that can form either hexamers or pentamers will be a challenge for structural characterisation.

Producing the human IgM Fc core as a single species is essential to my aim to crystallise the human IgM Fc core or to characterise the interaction between the IgM Fc core and DBL domains using biophysical methods. Also important is to be able to regulate the degree of glycosylation for IgM preparation. For this reason, rather than starting IgM preparations from native sources, I tested an IgM expression, which will allow regulation of glycosylation through a combination of methods including glycosylation processing inhibitors and endoglycosidases.

4.2. Interaction studies involving recombinant human IgM Fc

4.2.1. Expression and purification of recombinant human IgM Fc using HEK cells

4.2.1.1. Human IgM Fc construct

The human IgM Fc (hIgM-Fc) construct was kindly provided to us by Professor Richard Pleass (Liverpool School of Tropical Medicine). The construct consisted of 349 residues stretching from Cμ2 to Cμ4 regions of the IgM monomer Fc domain. Sequence analysis highlighted a few difference between this construct from the sequence of the Fc region from human IgM genomic DNA (Genbank X14940). The first is the additional N-terminal RS residues that were remaining from the BglII cleavage site. There are also two point mutations; a G84S substitution, and an insertion of a valine residue at position 308. The construct was inserted into the MCS of a pFUSE vector. This pFUSE-hIgM-Fc construct can be readily transfected into mammalian cells to produce secreted soluble hIgM-Fc.

4.2.1.2. Expression of hIgM-Fc in HEK293T cells

HEK293T cells were used for the expression of soluble hIgM-Fc. Sterile pFUSE- hIgM-Fc DNA was transfected into HEK293T cells using linearised PEI before purification.

Optimisation of expression

Initial expression of hIgM-Fc in HEK293T cells were performed in media with serum. Given the similarity in size of BSA and IgM monomer, the presence of BSA covered hIgM-Fc on a Western blot (Figure 4.2a). Initial purification of hIgM-Fc using affinity columns also

resulted in poor yields (data not shown). Thus, a series of steps were taken to optimise the expression of hIgM-Fc in HEK293T cells.

Firstly, an expression trial was performed to observe the difference in expression of hIgM-Fc with and without serum in the media. This was done using FuGENE6 transfection kit. Serum was removed immediately before transfection by replacing SAM with SFM. Adherent HEK293T cells were grown on a 6 well culture dish to 80% confluence before performing transfection. Cells were incubated at 37°C after transfection. The supernatant was collected three days after transfection for detection of hIgM-Fc by Western blot. Control experiments were done with cells that were not transfected and hence should not produce hIgM-Fc.

The removal of serum resulted in reduced BSA detection and hence the IgM band could be seen more clearly (Figure 4.2a). This has further consequences in the purification process, discussed below.

Additionally, an expression time trial over 7 days was performed to determine the amount of time required for maximum expression of soluble hIgM-Fc. 100 µl of the cell supernatant was collected from day 3 onwards. In addition to the time trial, I also tested the importance of pH correction for the expression of hIgM-Fc. It was observed that there was a change in the colour of the pH indicator (phenol red) in the media after transfection of the cells. A change in pH might affect the viability of the adherent cells and hence expression. Therefore, a trial for pH correction was included in parallel with the expression time trial. Cell cultures were pH corrected by adding SFM to the solution until the pH indicator returned to red (alkaline) from yellow (acidic).

It was observed from the results of this expression time trial that most hIgM-Fc could be detected on day 5 (Figure 4.2b). There were also no significant advantages to performing pH correction. Thus, the hIgM-Fc is optimally expressed HEK293T cells after 5 day incubation at 37°C after transfection in SFM without pH correction.

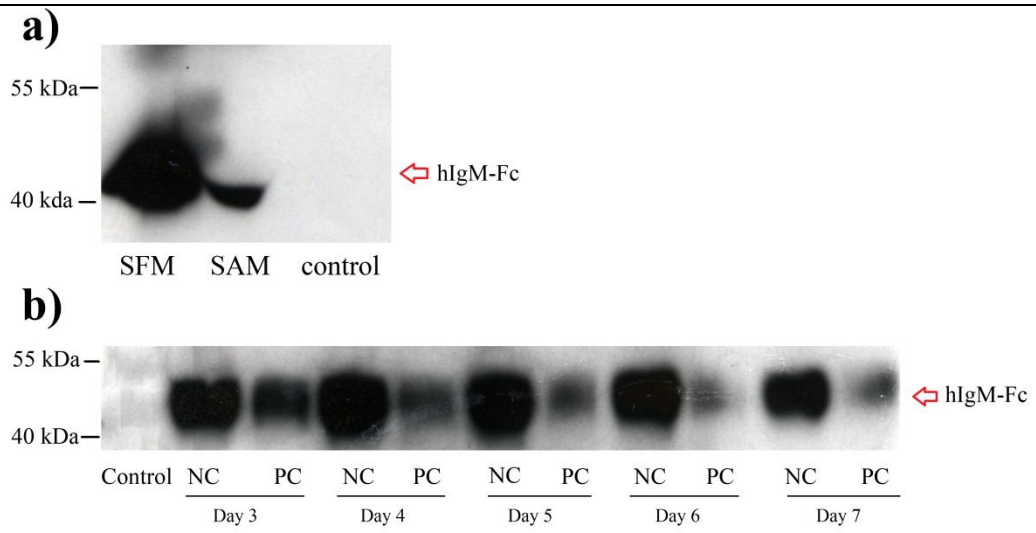


Figure 4.2. Expression trials of hIgM-Fc. **(a)** A Western blot for the expression trial for the determination of expression with and without serum in the media. **(b)** A Western blot for time trial to determine the number of days required for optimal expression of hIgM-Fc. The lanes denoted "NC" not pH corrected. The lanes denoted "PC" were samples for which the pH had been corrected by adding fresh media. All controls were performed with just SDS-PAGE loading buffer without hIgM-Fc. Western blots were performed using Anti-Human IgM (μ-chain specific) −Peroxidase antibody. Control experiments were performed with cells that were not transfected.

4.2.1.3. Purification

hIgM-Fc was purified using ion exchange chromatography (5 ml HiTrap Q HP column), IgM affinity (1 ml HiTrap IgM HP column) and Anti-Human IgM (μ-chain specific) agarose beads.

Optimisation of the purification process.

The first purification trial of the serum containing supernatant resulted in high amount of BSA being detected regardless of the usage of any of the methods described above (data not shown). Subsequent removal of BSA from expression media increased the level of binding of hIgM-Fc to the HiTrap IgM HP column, but hIgM-Fc was still undetectable by SDS-PAGE in the elution fractions despite being detected by Western blotting (Figure 4.3). Instead, high level of hIgM-Fc was detected in the flow through (Figure 4.3). This suggested that the HiTrap IgM column was unable to bind to hIgM-Fc efficiently using suggested binding buffers. The situation was similar in the use of the ion exchange. The levels of non-specific binding remained quite high in both chromatographic methods (data not shown).

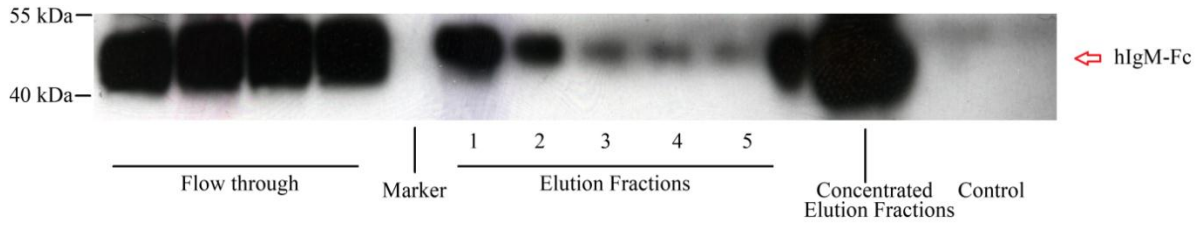


Figure 4.3. Western blots showing the purification of hIgM-Fc expressed in SFM. Purification was performed using HiTrap IgM HP column. No occlusion of the bands by BSA was seen. More IgM were found in the elution fractions. However, most of the hIgM-Fc was found in the flowthrough fractions. Western blots were performed using Anti-Human IgM (μ-chain specific) −Peroxidase antibody.

The use of Anti-Human IgM (μ-chain specific) agarose beads (IgM affinity beads) for the purification of hIgM-Fc initially registered lower levels of nonspecific binding. Collected serum free supernatant was buffer exchanged into either 20 mM Tris pH 8.0, 0.3 M NaCl or into PBS (10 mM Phosphate pH 7.4, 2.7 mM KCl, 137 mM NaCl) to see whether a change in buffer conditions can increase the affinity of hIgM-Fc for the IgM affinity beads. The elution method involved the use of a low pH elution buffer (0.5 M acetic acid pH 2.4, 0.15 M NaCl) in relation to the theoretical isoelectric point of the protein (pI = 6.72) suggested by ProtParam[166]. Eluted fractions were immediately pH corrected to prevent any protein aggregation at low pH.

The trial suggested that purification of the hIgM-Fc containing supernatant that were buffer exchanged into PBS allowed hIgM-Fc to bind to the affinity beads with a much higher efficiency compared to the use of the Tris based binding buffer (Figure 4.4). Purified hIgM-Fc in the elution fractions were now detectable by both SDS-PAGE and Western blot (Figure 4.4a,b).

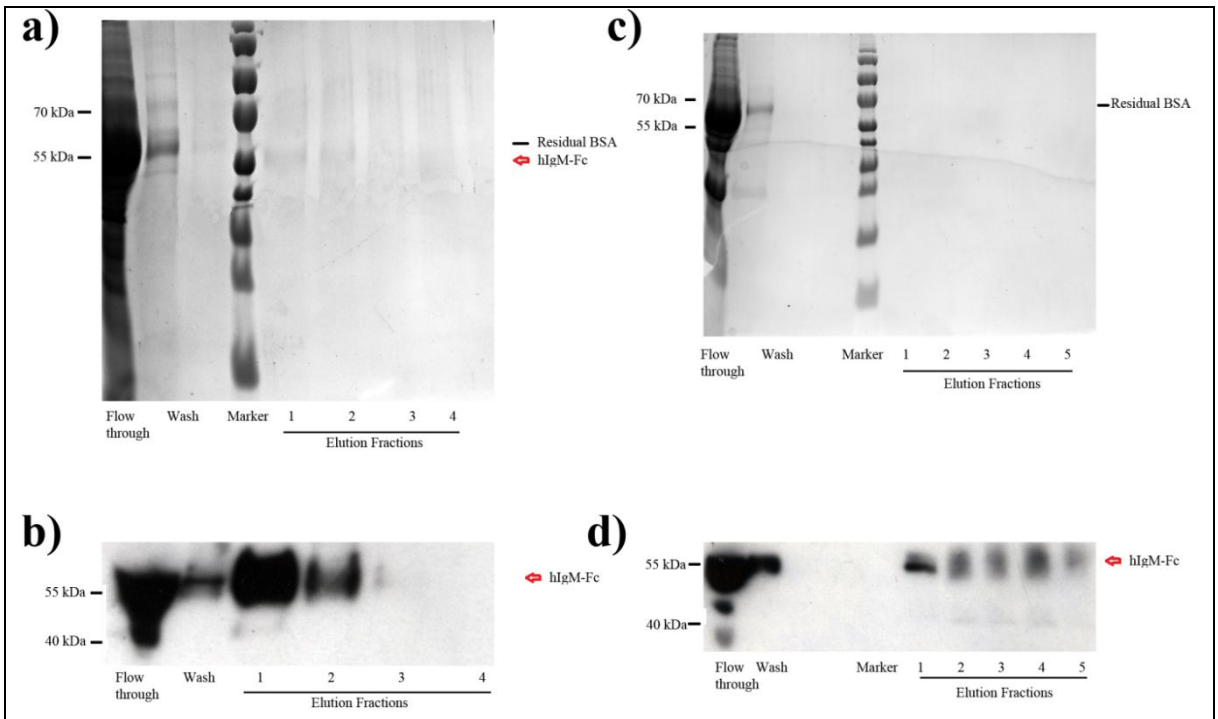


Figure 4.4. Purification trial of hIgM-Fc in different buffers. **(a)** SDS-PAGE and **(b)** Western blot of a purification trial performed after dialysing the supernatant in PBS by affinity beads. hIgM-Fc could be detected in both SDS page and elution fractions. **(c)** SDS Page and **(d)** Western blot of the purification trial performed after dialysing the supernatant into 20 mM Tris pH 8.0, 0.3 M NaCl. No hIgM-Fc could be detected in the SDS Page. Less hIgM-Fc was also detected by Western blot compared to the western blot shown in (b). Western blots were performed using Anti-Human IgM (μ-chain specific) −Peroxidase antibody.

Purification of hIgM-Fc

Purification of hIgM-Fc was consolidated to a two-step chromatographic protocol after optimisation of the method.

Firstly, cell supernatant containing hIgM-Fc was buffer exchanged into PBS. Collected samples containing hIgM-Fc were subjected to affinity chromatography using IgM affinity

beads. The flow through was reapplied to the same agarose beads several times. This increased the amount of hIgM-Fc bound, as we can see by the reduced amount of hIgM-Fc in the flow through (Figure 4.5a). Elution was performed in a low pH elution buffer suggested above. The eluted fractions were immediately pH corrected to pH 8.0 using NaOH after elution. Presence of hIgM-Fc in the elution fractions could be detected by both SDS-PAGE (data not shown) and Western blot (Figure 4.5a). The elution fractions were then subjected to a preparative S200 size-exclusion chromatography column. The resulting multipeak size-exclusion chromatography trace, with all peaks containing hIgM-Fc suggested the presence of various polymeric forms of hIgM-Fc (Figure 4.5b,c). The peak fractions at 49.31 ml corresponding to the estimated size of a pentameric hIgM-Fc were collected. The amount of hIgM-Fc purified from this method was ~ 2 mg/L of supernatant.

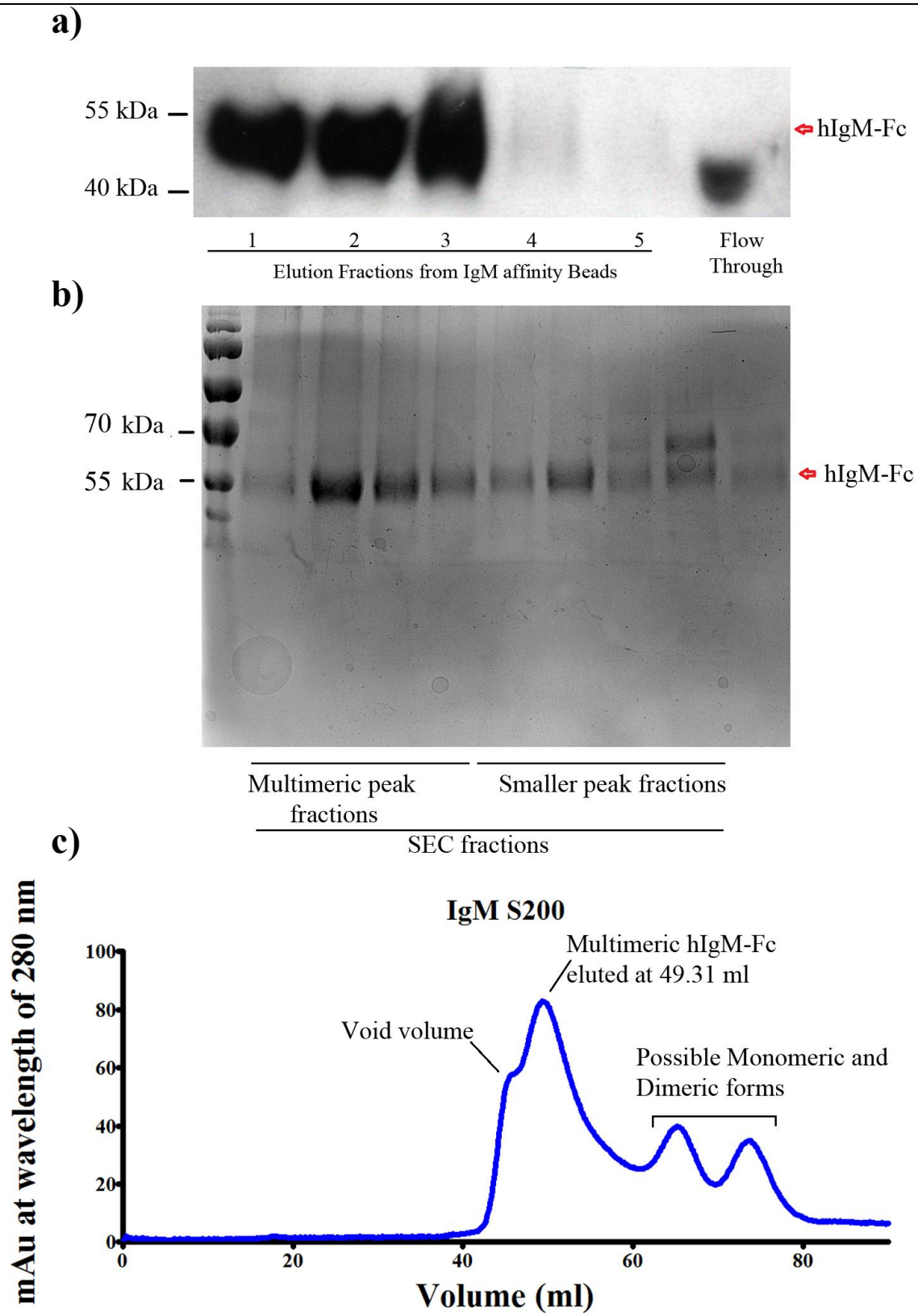


Figure 4.5. Purification of hIgM-Fc. **(a)** Western blot of the elution fractions from IgM affinity beads. Trace amounts of hIgM-Fc is detected in the flow through compared to Figure 4.3. **(b)** The SDS-PAGE of size-exclusion chromatography peak fractions showed that hIgM-Fc was present in multiple peaks. This suggested that hIgM-Fc is in multivalent forms during purification. **(c)** The size-exclusion chromatography trace of hIgM-Fc using a prep-grade S200 column. Western blots were performed using Anti-Human IgM (μ-chain specific) −Peroxidase antibody.

4.2.2. Cloning, expression and purification of recombinant DBL domains using E coli expression system

4.2.2.1. DBL6ε

DBL6ε from *var*2CSA has been shown to be able to interact with IgM[267].

Cloning, expression and purification of DBL6ε

The DBL6ε domain construct was previously made in the laboratory for structure determination by Dr. Pongsak Khunrae[43]. This construct consist of residues 2333 – 2634 of *var*2CSA (Genbank AE014844.1). Expression and purification of the DBL6ε was followed as documented from previous work in the lab in Origami (pRIG) cells[282] using a purification protocol identical to that used for the purification of CIDR domains in the previous chapter.

4.2.2.2. MSPDBL2 domain

Merozoite surface protein 3 (MSP3) is important in the erythrocyte invasion process[283]. It has also been identified as a target antigen by the human immunity system[284,285]. MSP3 is predicted to contain a series of elongated α helical coil domains that are made from blocks of heptad repeats[286,287]. Analysis of the sequences of MSP3 genes from various strains suggested that these are polymorphic and are clustered in the N-terminal region. Conversely, the C terminal region consist of a leucine zipper[288] and is highly conserved[289]. These leucine zippers allow MSP3 proteins to form dimers and long chain elongated oligomers in solution that are able to bind to erythrocytes[286]. The use of antibody-dependent cellular inhibition

assays has established MSP3 as a viable target antigen for vaccine development[290–292]. Initial vaccine trials in primates has given positive results for malaria resistance[293,294]

Two members of the MSP3 family, MSP3.4 (Genbank AE014185.2) and MSP3.8 (Genbank AAN35552.1), contain DBL domains[295,296]. In addition to the DBL domains, they contain a secreted polymorphic antigen associated with merozoites (SPAM) domain. The DBL containing MSP proteins are able to bind to erythrocytes[296]. The structures of the DBL domains in the MSP proteins are similar to the previous DBL domains from EBL and PfEMP1 proteins[297]. Antibodies raised against these MSP proteins disrupt erythrocyte invasion[298]. MSP3.8 has also been shown to be able to bind to IgM (Gavin Wright, unpublished data). This DBL domain from MSP3.8 has been annotated as MSPDBL2 in the literature.

4.2.2.2.1. Cloning of MSPDBL2

The template construct for the MSP3.8 which displayed IgM binding properties was kindly provided to us by Dr Gavin Wright (Wellcome Trust, Sanger institute). This contained a construct for which the two predicted N-glycosylation sites had been mutated by alanine substitutions (S196A and T426A).

The domain boundaries of the DBL domain in MSP3.8 (MSPDBL2) were defined by structural alignment to a few DBL domains from PfEMP1 using the FUGUE algorithm[203]. The construct contained residues Q176 – T447, corresponding to a DBL domain. There are ten cysteine residues within the construct with an arrangement that is similar to that found in DBL domains found in *var* and *ebl* gene families. This was amplified by PCR and inserted

into a pET15bcon vector using restriction enzymes NheI and BamHI (Figure 4.6). The identities of the clones were checked by sequencing.

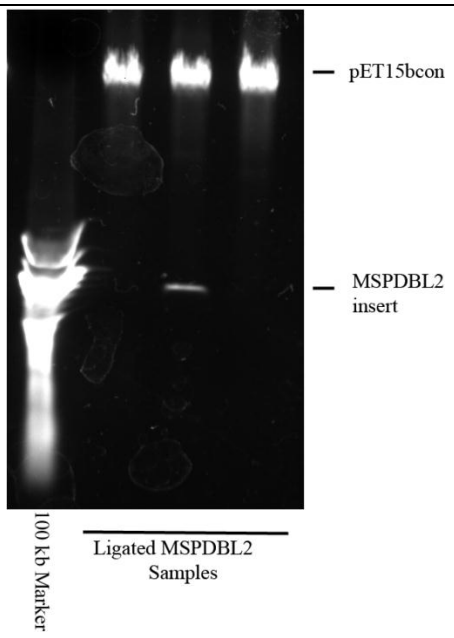


Figure 4.6. Agarose gel electrophoresis of ligated MSPDBL2 samples after restriction enzyme digestion. Restriction digest of the ligated sample using NheI and BamHI restriction enzymes allow visualisation of the MSPDBL2 insert. Sample 2 was sent for sequencing to confirm the identity of MSPDBL2.

4.2.2.2.2. Expression and purification of MSPDBL2

An expression trial of MSPDBL2 was conducted as described in chapter 2.2.1.2 (Figure 4.7a). MSPDBL2 was expressed in the inclusion bodies in Origami (pRIG) cells[205] incubated overnight at 25°C after IPTG induction.

Purification of insoluble His_6-tagged MSPDBL2 therefore required a refolding procedure. Inclusion bodies were first lysed by sonication in a buffer containing guanidine HCl. Guanidine HCl is a chaotropic agent that denatures insoluble protein by disruption of disulphide bonds and hydrogen bonds, thus rendering the protein unfolded. Refolding of proteins from guanidine HCl has been used for purification of many small domains from

inclusion bodies[299,300]. A protocol involving refolding on a Ni-NTA column using decreasing concentration of guanidine was performed as described in chapter 2.3.7. The refolded protein was subsequently TEV cleaved and introduced onto a preparative S200 for size-exclusion chromatography (Figure 4.7b,c,d).

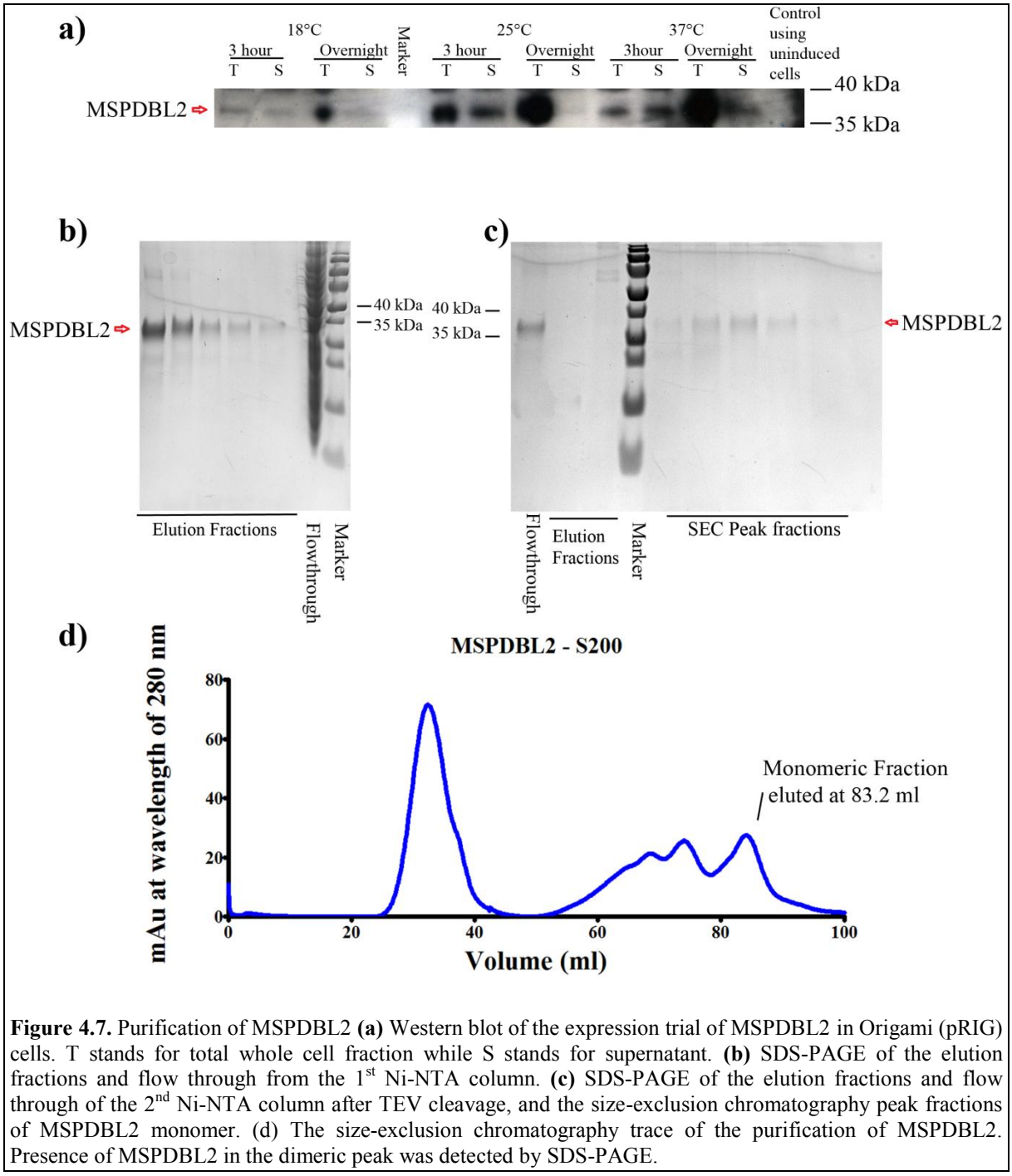


Figure 4.7. Purification of MSPDBL2 **(a)** Western blot of the expression trial of MSPDBL2 in Origami (pRIG) cells. T stands for total whole cell fraction while S stands for supernatant. **(b)** SDS-PAGE of the elution fractions and flow through from the 1st Ni-NTA column. **(c)** SDS-PAGE of the elution fractions and flow through of the 2nd Ni-NTA column after TEV cleavage, and the size-exclusion chromatography peak fractions of MSPDBL2 monomer. (d) The size-exclusion chromatography trace of the purification of MSPDBL2. Presence of MSPDBL2 in the dimeric peak was detected by SDS-PAGE.

4.2.2.2.3. Circular dichroism

CD spectroscopy was performed to investigate the presence of secondary structure to ensure proper folding of MSPDBL2 during the purification process. The purified refolded MSPDBL2 was buffer exchanged into a CD buffer. The ellipicity (mDeg) of MSPDBL2 was measured from wavelength 195 nm to 260 nm (Figure 4.8). An averaged spectrum of a buffer blank was subtracted and mean residue ellipticity (MRE) of the measured CD spectrums was calculated.

The CD spectrum of the purified MSPDBL2 is shown below. The spectrum showed mostly α-helical content, with negative ellipticity in the 208 nm and 218 nm regions. The spectrum was then analysed by using the K2D3 algorithm to predict secondary structure content[133]. This suggested a helical content of 81% which agrees well with the crystal structure of MSPDBL2[297].

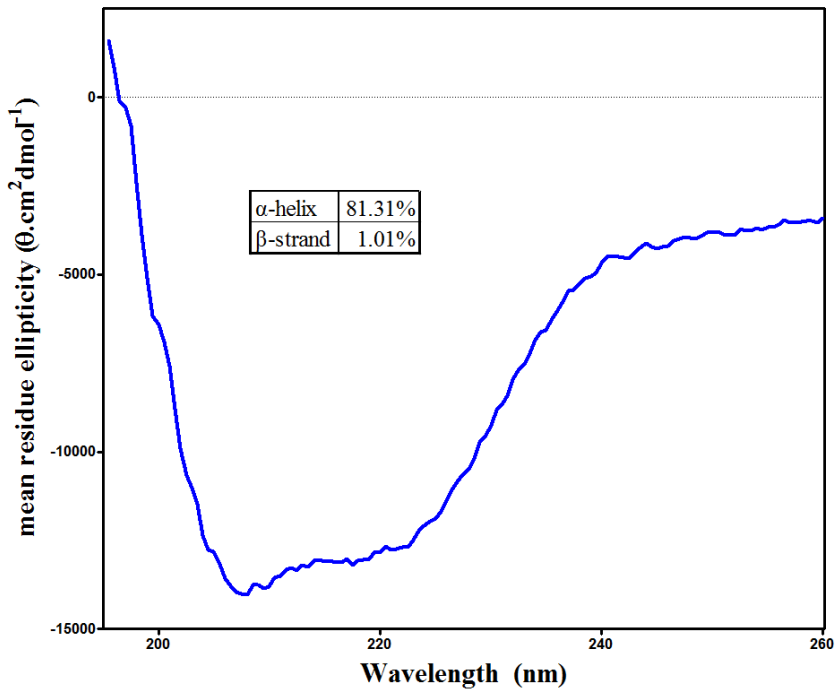


Figure 4.8. Circular dichroism of refolded MSPDBL2. Percentages of α-helices and β-strands were determined by K2D3 algorithm[133].

4.2.3. Interaction studies using recombinant IgM and DBL domains

Interaction studies were performed to characterise the binding of purified recombinant hIgM-Fc to the various recombinant DBL domains. This was done by a variety of biophysical and biochemical methods.

4.2.3.1. hIgM-Fc pull down assay

A Ni-NTA agarose pull down assay was first used to test for interaction between hIgM-Fc and DBL domains (MSPDBL2 and DBL6ε) (Figure 4.9). For the experiment, the cell supernatant containing His_6-tagged DBL domains were first incubated with Ni-NTA agarose beads for 30 minutes at room temperature. Beads were washing with three columns of Ni-NTA wash buffer. Subsequently, hIgM-Fc (50 µl of 0.2 mg/ml in 20 mM Tris pH 8.0, 0.3M NaCl) was added in excess of the Ni-NTA bound DBL domains and incubated for 30 minutes at room temperature. This was once again washed with three columns of wash buffer. SDS-PAGE loading dye was added directly to the beads and heated at 100°C before subjecting to SDS-PAGE. Western blot was performed using antibodies that detect either His_6-tag or IgM-Fc.

hIgM-Fc was loaded straight to lane 8 for comparison, generating a band adjacent to the 55 kDa marker. Positive controls for His_6-tagged DBL6ε and MSPDBL2 in lanes 1 and 5 respectively were performed by adding buffer instead of hIgM-Fc to Ni-NTA bound His_6-tagged DBL domains. A band corresponding to ~35 kDa was observed in both these lanes. The negative control for binding was performed by applying wash buffer instead of cell

supernatant containing His_6-tagged DBL domains to the Ni-NTA agarose beads. This was done for lanes 2 and 6.

The first experimental observation from the pull down assay was that hIgM-Fc can interact with both DBL6ε and MSPDBL2. hIgM-Fc was detected in the presence of His_6-tagged DBL domains in lanes 3 and 7. No bands corresponding to hIgM-Fc was observed in the absence of His_6-tagged DBL domains in lanes 2 and 6 (Figure 4.9). This suggests that hIgM-Fc was unable to interact with Ni-NTA beads in the absence of His_6-tagged DBL domains.

Secondly, the pull down assay showed that the recombinant DBL6ε pulled down hIgM-Fc to a significantly lesser extent then MSPDBL2, as seen by the relative intensity of the bands between hIgM-Fc and the His_6-tagged DBL domains.

Therefore, the pull down assay showed that the DBL domains are able to interact with the recombinant hIgM-Fc core, and suggests that the hIgM-Fc core was in a correct multimeric conformation to allow interaction with DBL domains.

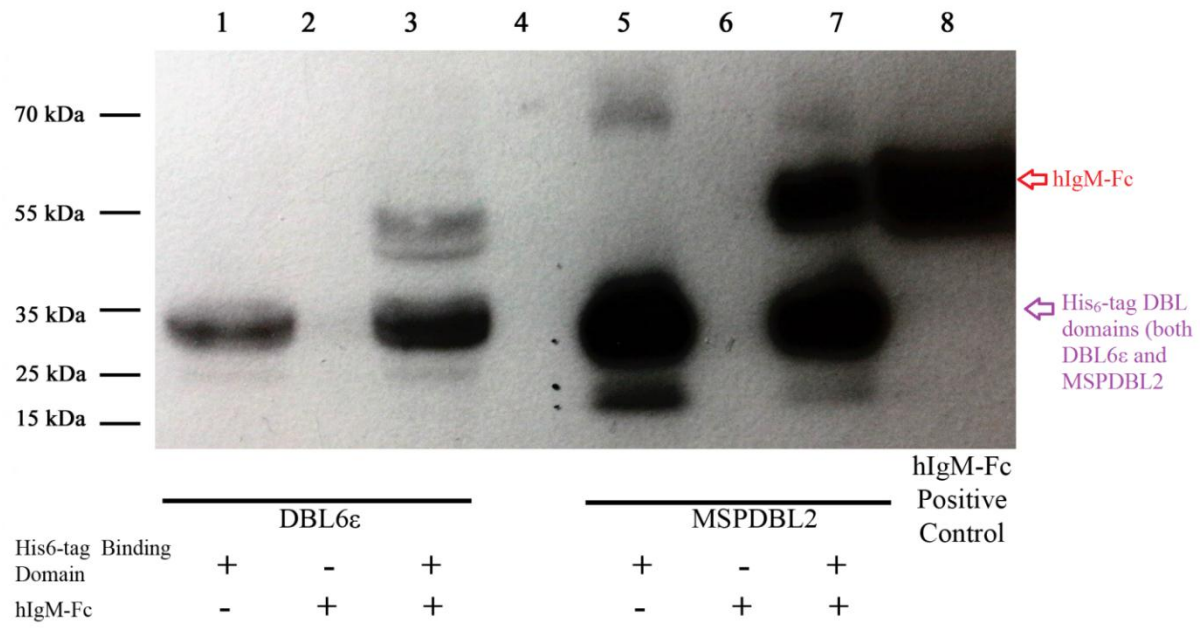


Figure 4.9. Western blot of Ni-NTA column pull down assay of hIgM-Fc using His-tagged DBL6ε and MSPDBL2. Western blotting was performed using both Anti-Human IgM (μ-chain specific)−Peroxidase antibody and Penta-His HRP Conjugate. Lanes 1 and 5 were positive controls for the His_6-tagged DBL domains. Lanes 2 and 6 were the negative controls for binding. Lane 8 was the positive control for hIgM-Fc. Lane 4 was used for running the markers. Lanes 3 and 7 was used for the experiment samples with both hIgM-Fc and His_6-tagged DBL domains.

4.2.3.2. Surface Plasmon Resonance measurements

SPR was further conducted to characterise the interaction between hIgM-Fc and the DBL domains. In these experiments, the hIgM-Fc was immobilised on the CM5 sensor chip surface of flow cell 2 to 5000 RU. Flow cell 1 was used as a reference cell. A concentration series of both DBL domains were flowed over the chip surface at a continuous flow rate of 30 μl/min. The top concentration used for both MSPDBL2 and DBL6ε was 32 μM. A buffer injection was added at the end of each series to account for any change in signal due to buffer flow. All SPR binding measurements were measured with proteins in Biacore buffer at 25°C in the flow cell chamber.

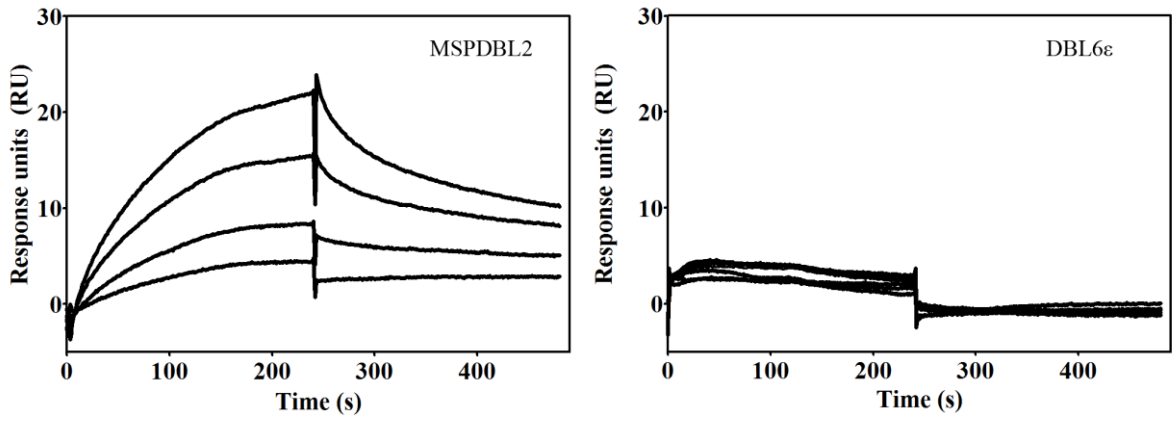


Figure 4.10. The response curves from the surface plasmon resonance measurements for a concentration series of MSPDBL2 and DBL6ε for their interaction with hIgM-Fc. The data was analysed using the BIAevaluation software and constructed using GraphPad Prism 5.

The first observation was that MSPDBL2 was able to bind to hIgM-Fc (Figure 4.10). This was in agreement with the results from the Ni-NTA pull down assay. However, titration of DBL6ε at a top concentration of 32 μM did not show significant binding (Figure 4.10). These results were not in agreement with previous studies regarding IgM binding domains on *var*2CSA[267]. Despite the saturated response curve, the measured SPR data for MSPDBL2 did not fit the various binding models to satisfaction (data not shown). Although the binding curves estimated a low μM affinity for the interaction between MSPDBL2 and hIgM-Fc, an exact binding affinity cannot be derived accurately.

Also, the change in response units that was measured due to the interaction of DBL domains to hIgM-Fc was small (25 RU) relative to the amount of hIgM-Fc that was immobilised on the surface (Fig 4.10a). There might be a few reasons for this observation.

One reason for this observation might be due to the immobilisation of a relatively larger multimeric hIgM-Fc (~250 kDa) compared to the bound DBL domains (Theoretical molecular weight for MSPDBL2 and DBL6ε are 32.16 kDa and 35.58 kDa respectively).

The change in response measured will depend on the number of attached DBL domains for each immobilised hIgM-Fc multimer. Thus, in the hypothetical scenario that the multimeric hIgM-Fc only has one binding site for DBL domains, the change in response measured might be small relative to the total mass immobilised on the chip surface.

Furthermore, the number of binding sites of the immobilised hIgM-Fc available on the surface depends largely on the homogeneity of the hIgM-Fc sample. The homogeneity of hIgM-Fc was further investigated using sedimentation velocity experiments, which will be discussed further below.

These observations were consistent across two independent SPR experiments using different batches of protein. Thus, the SPR data suggests that MSPDBL2 was able to bind to hIgM-Fc. This was further studied using analytical size-exclusion chromatography.

4.2.3.3. Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was used to confirm the interaction between MSPDBL2 and hIgM-Fc. DBL6ε was not included in view of its apparent poor affinity to hIgM-Fc. A 2.4 ml analytical grade Superose 6 column was used for these studies. 50 µl of each protein sample (0.5 mg/ml) was injected onto the analytical column at a flow rate of 0.10 ml/min using 20 mM Tris pH 8.0, 0.15 M NaCl. For the formation of the complex peak, MSPDBL2 was incubated with hIgM-Fc in a 1:1 (monomer:monomer) weight ratio for 30 minutes at room temperature before being injected into the column.

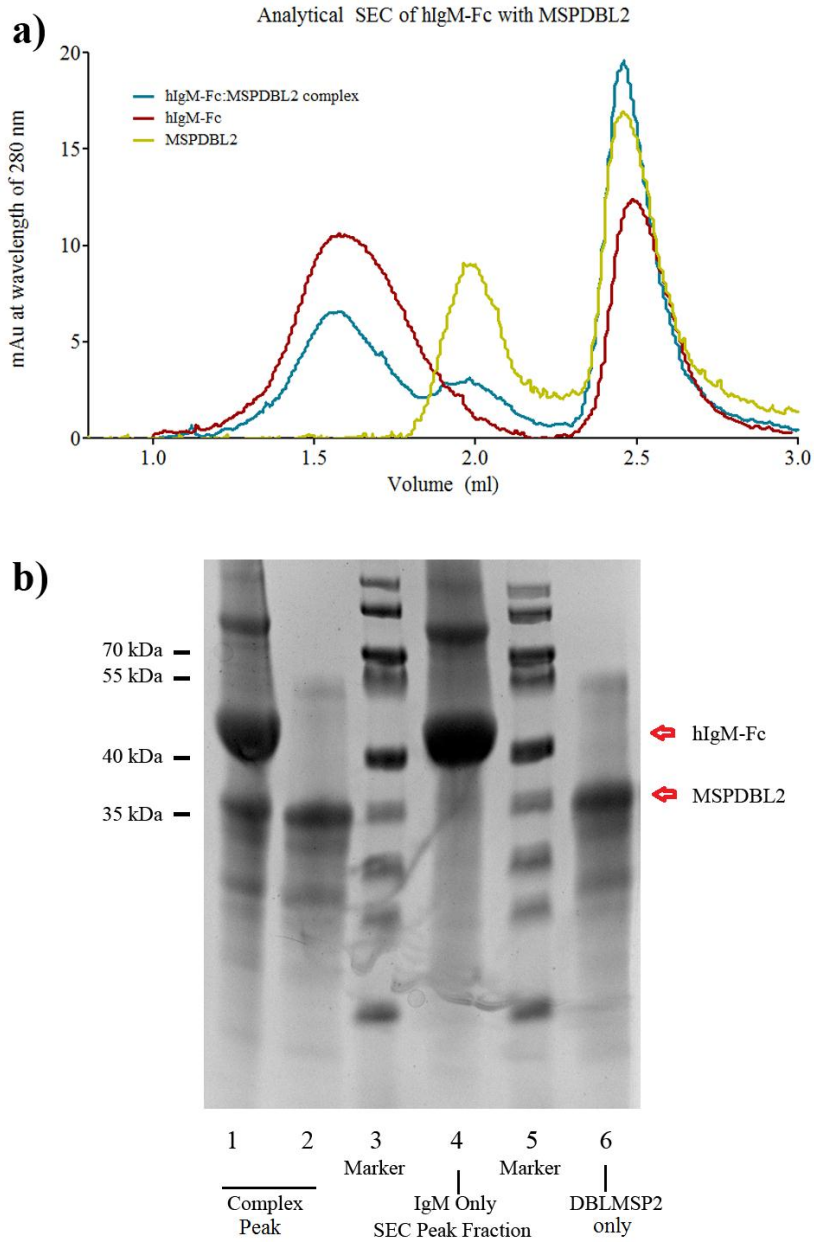


Figure 4.11. Analytical size-exclusion chromatography of hIgM-Fc and MSPDBL2 complex. **(a)** Comparative curves of the analytical size-exclusion chromatography traces of hIgM-Fc (red), MSPDBL2 (gold), and the hIgM-Fc:MSPDBL2 complex (blue). **(b)** SDS-PAGE of the peak fraction at the respective elution volumes for hIgM-Fc (1.58 ml), MSPDBL2 (1.98 ml), and hIgM-Fc:MSPDBL2 complex (1.61 ml). Lane 1 and 2 corresponds to the fractions containing the hIgM-Fc:MSPDBL2 complex and excess MSPDBL2 respectively from the complex run. Lane 4 corresponds to the peak fraction from the size-exclusion chromatography run with only hIgM-Fc. Lane 6 corresponds to the peak fraction from the size-exclusion chromatography run with only MSPDBL2.

MSPDBL2 and multimeric hIgM-Fc eluted at 1.98 ml and 1.61 ml respectively on the analytical Superose 6 column. When MSPDBL2 was added to hIgM-Fc, the peak at 1.98 ml corresponding to monomeric MSPDBL2 had diminished (Figure 4.11a). This suggested that

the MSPDBL2 has been incorporated into a complex. This was confirmed by the presence of both hIgM-Fc and MSPDBL2 bands on SDS-PAGE in lane 1 (Figure 4.11b). Thus, this clearly suggests that hIgM-Fc and MSPDBL2 were able to form a complex that co-eluted on an analytical size-exclusion chromatography column.

The mobility in the elution volume of the complex (1.58 ml) in comparison to that of the hIgM-Fc multimer (1.61 ml) was minimal. This might be due to a few reasons. Firstly, the resolution of the analytical Superose 6 column was unable to separate the change in hydrodynamic volume of the hIgM-Fc multimer from the complex.

Secondly, the change in size of the complex might not be large enough to see a change. Excess MSPDBL2 was detected both at wavelength of 280 nm (1.94 ml of the complex size-exclusion chromatography trace) and by SDS page (lane 2) despite being incubated with hIgM-Fc at a 1:1 (monomer:monomer) molar ratio. (Figure 4.11a,b). This suggests that not all the MSPDBL2 were involved in an interaction with hIgM-Fc.

In consolidation, the analytical size-exclusion chromatography results were able to confirm the binding of MSPDBL2 to hIgM-Fc. This was consistent with the results from the Ni-NTA pulldown assay as well as from the SPR measurements. Through the three methods described above, I was able to confirm the specific interaction between MSPDBL2 and hIgM-Fc. However, what was also consistent between the analytical size-exclusion chromatography and SPR results was the ambiguity in the stoichiometry of the interaction between hIgM-Fc and MSPDBL2. In order to understand more about the stoichiometry, an AUC experiment was conducted.

4.2.3.4. Analytical ultracentrifugation

Sedimentation velocity experiment was carried out to further characterise the nature of the interaction between hIgM-Fc and MSPDBL2. The AUC data should report more accurately the stoichiometry of the complex and reveal whether the expressed recombinant hIgM-Fc is homogenous in solution.

Experimental procedure

All the samples (0.2 mg/ml) were prepared in 20mM Tris pH 8.0, 0.15 M NaCl. The complex sample was prepared by incubating MSPDBL2 with hIgM-Fc for 30 minutes in a 1:1 (monomer to monomer) molar ratio to make up to a combined concentration of 0.2 mg/ml. The samples and a buffer blank were injected into the sample and reference cell respectively. Both absorbance and interference data were measured to track the sedimentation properties of the samples in solution when being subjected to centrifugation at 40,000 RPM at 20°C.

The absorbance data was analysed on SEDFIT. The protein partial specific volume (vbar) of each protein and buffer viscosity (1.0002) and density (1.0049) were estimated using Sednterp. The data were then fitted to distribution curves that satisfy the Lamm equation c(s) distribution[135–137]. The weight averaged frictional ratio (f/f_0) and molecular weight estimated by the programme also gives us an idea about the properties of the individual components as well as the complex that is formed in solution.

4.2.3.4.1. AUC for MSPDBL2

The vbar of MSPDBL2 was estimated to be 0.7255. The data which were fitted to the distribution curves satisfied the Lamm equation with an rmsd of 0.006 (Figure 4.12b). The f/f_0 ratio estimated by the programme came up to 1.20, which was close to the frictional ratio estimated for a compact structure. Two species were observed at sedimentation 2.6S (81.3%) and 3.8S (11.3%) (Figure 4.12c). This suggested that MSPDBL2 might be present in both monomeric and dimeric forms, which was consistent with the data from the purification process. The estimated molecular weight of the peak at 2.6S was 32.7 kDa, which was close to the theoretical molecular weight (32.2 kDa).

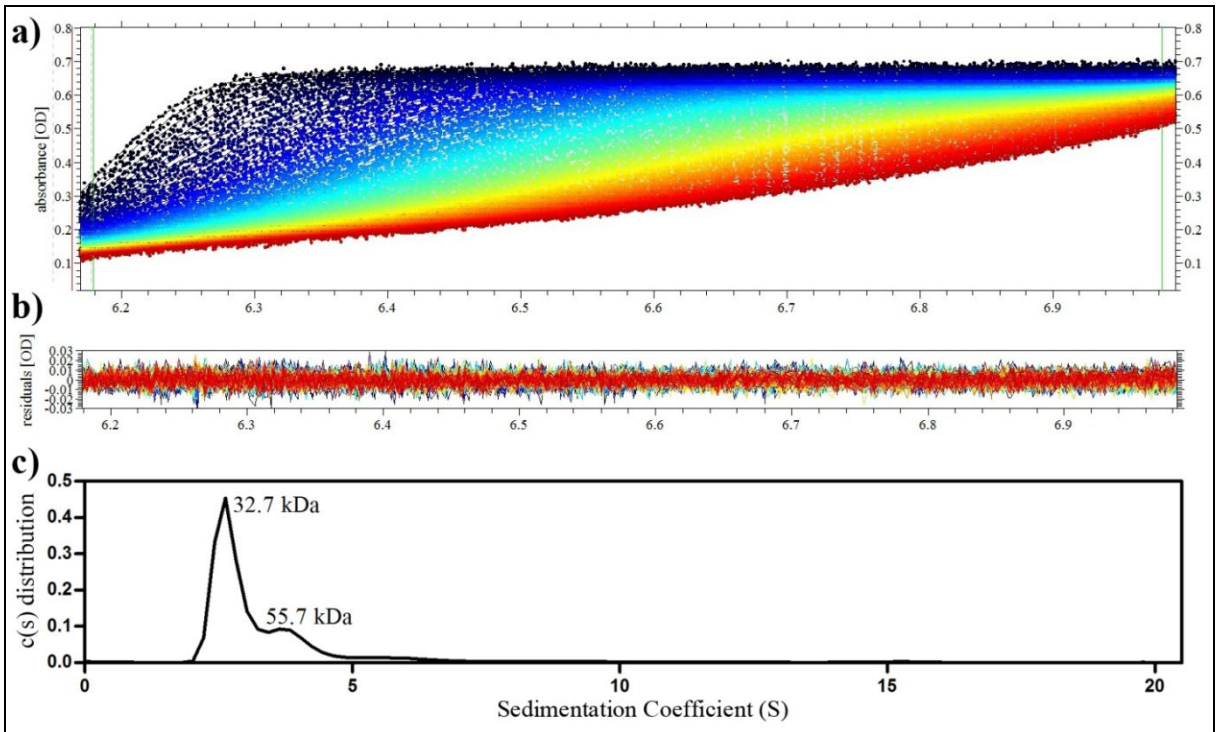


Figure 4.12. Sedimentation velocity data for MSPDBL2. **(a)** The theoretical curves that were fitted with AUC data to satisfy the Lamm equation with an rmsd of 0.006. **(b)** The residuals showing the error of the fit for each data set. **(c)** This is the normal c(s) distribution of the species from the AUC data set, with 81.3% of the species in the major peak at 2.6S and 11.3% of the species residing at the peak at 3.8S.

4.2.3.4.2. The heterogenous nature of hIgM-Fc

The AUC trace for recombinant hIgM-Fc showed that the protein exist in multiple multimeric forms. 85.9% of the hIgM-Fc sample resides in at least three different multimeric states with peaks at sedimentation coefficient of 8.1S (~91.8 kDa), 10.9S (~170 kDa) and 13.7S (~277 kDa) (Figure 4.13a). This suggests that the purified recombinant hIgM-Fc was in multiple multivalent conformations. This result was consistent with what was observed in the results from the other interaction studies.

This multipeak trend was observed even in the presence of MSPDBL2 at a 1:1 (monomer:monomer) molar ratio (Figure 4.13b). Because of the complexity of the multivalent hIgM-Fc in the background of the AUC trace for the hIgM-FC:MSPDBL2 complex, it was not possible to identify a peak that corresponds to the complex. Hence, it is difficult to determine which multimeric forms of hIgM-Fc were able to interact with MSPDBL2.

Additionally, a peak could be clearly seen at sedimentation coefficient of 2.6S (22.4%) in the AUC trace for the complex (Figure 4.13b). This corresponds with what would be left of the MSPDBL2 monomers that were not involved in complex formation (Figure 4.13b). Given that MSPDBL2 (Theoretical molecular weight = 32.2 kDa) was added to hIgM-Fc (Theoretical molecular weight = 38.5 kDa) in a 1:1 (monomer:monomer) molar ratio according to the theoretical weight of their monomeric species, 45.5% of the complex sample should be comprised of MSPDBL2. Thus, this suggests that only half of the MSPDBL2 was involved in an interaction with multimeric hIgM-Fc. This observation was consistent with the results from the analytical size-exclusion chromatography.

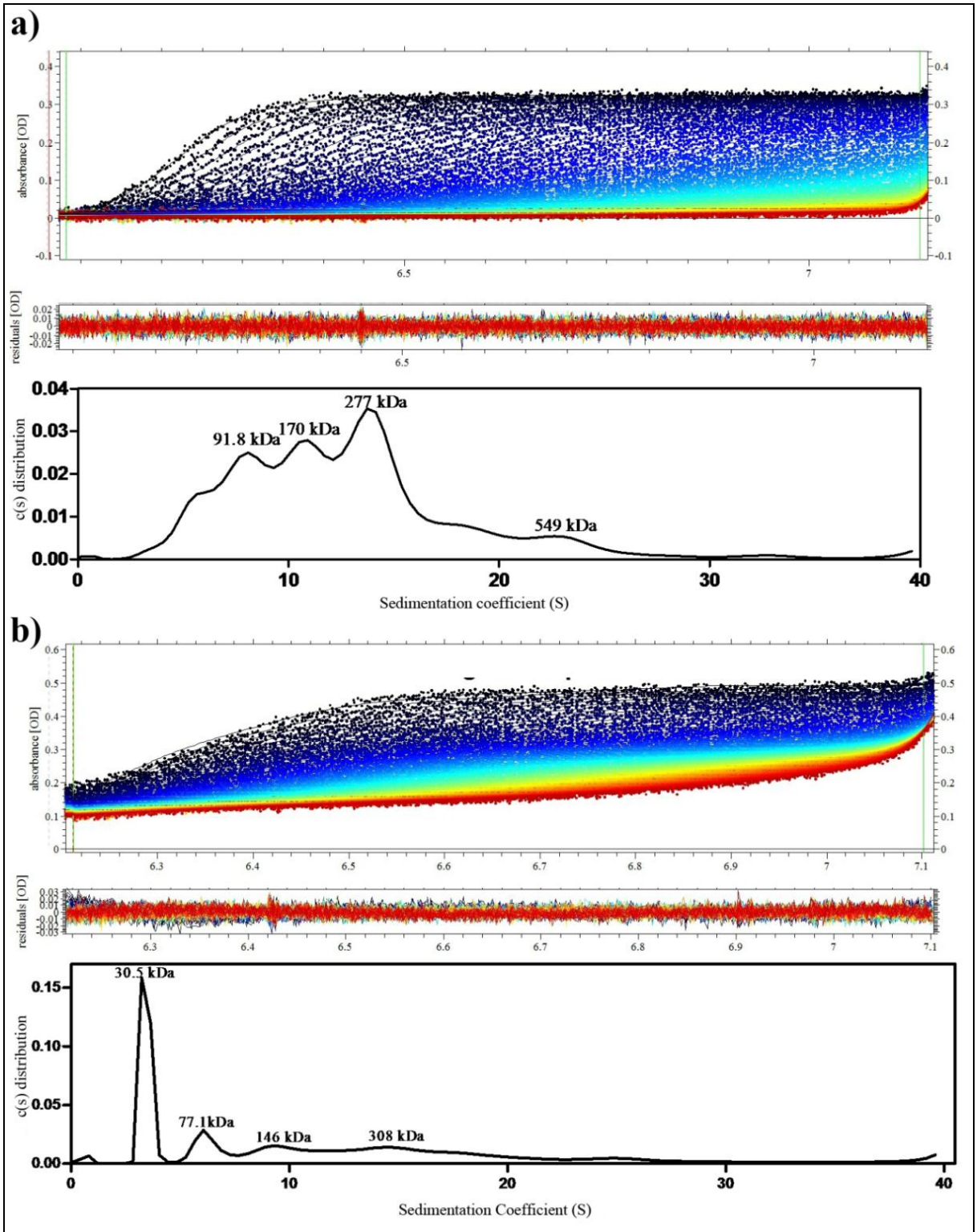


Figure 4.13. Sedimentation velocity data for hIgM-Fc and hIgM-Fc:MSPDBL2 complex. For each data set, there is a theoretical curve that was fitted with AUC data to satisfy the Lamm equation with an rmsd of 0.006, and a residual plot showing the error of the fit for each data set. **(a)** The c(s) distribution of the AUC from the hIgM-Fc data set suggests that 85.9% of the hIgM-Fc are in a multipeak of sedimentation coefficients 8.1S, 10.9S and 13.7S. **(b)** The c(s) distribution of the AUC from the hIgM-Fc:MSPDBL2 complex data set showed that 22.4% of the species are at a peak of sedimentation coefficient of 2.6S. The rest of the species were spread out in across multiple peaks ranging from 7S to 20S.

4.2.4. Consolidation of interaction studies of hIgM-Fc and DBL domains

In essence, I was able to confirm the interaction of MSPDBL2 with hIgM-Fc with a variety of biophysical methods. This interaction was consistent with the data that suggested that MSP3.8 was able to interact with IgM (Dr Gavin Wright, unpublished data). However, another interesting finding was the lack of interaction between hIgM-Fc and DBL6ε as shown by two separate methods. This was surprising because it has been shown previously to be able to bind to hIgM-Fc by live cell IFA using DBL6ε expressing COS-7 cells[267].

Further structural characterisation of the interaction between hIgM-Fc and DBL domains was difficult because of lack of homogeneity of hIgM-Fc. This is especially so for crystallisation, where homogenous hIgM-Fc samples are needed. The complex multimeric properties of expressed recombinant hIgM-Fc might be due to a few reasons. One question would be whether there is a critical concentration of hIgM-Fc monomers required for formation of multimers of a certain form. Secondly, the formation of hexameric natural human IgM has been linked to lack of the J chain[250–253]. This could mean that the J chain might be important in stabilising the pentameric form of the expressed hIgM-Fc core *in vitro*.

Chapter 5

Perspectives and discussion

5.1. PfEMP1 and CD36

5.1.1. Binding interactions between CIDR1α domains and CD36

The work described here presents the first biophysical characterisation of the interaction between a series of CIDR1α domains and CD36. The data confirmed the previous findings that sequence analysis can predict of which CIDRs can interact with CD36[83]. The stoichiometry of the hCD36ED and CIDR1α complex was determined to be 1:1 by different biophysical methods. The binding affinities of these interactions with CD36 has been determined by SPR measurements using the same protocols, hence allowing us to perform comparative studies between different CIDR1α domains from different strains.

The comparison of the binding affinities between a CIDR1α domain and the full length protein of PF11_0007 suggested that an elongated CIDR1α containing the M1 region could account for the full binding affinity of the full length PfEMP1 to CD36. The binding affinity of this full length PF11VAR was similar in comparison to that of other NTS-DBLα-CIDR1α constructs from PfEMP1 proteins from the IT strain of *P. falciparum*. Although an extrapolation of these results is not possible, an attractive model of binding involving the CIDR1α as the central binding module of PfEMP1 in its interaction with CD36 can be suggested by taking into account these two pieces of evidences.

One interesting observation was also that the binding affinities of the different CIDR1α containing constructs to CD36 ranging from mid nanomolar to low micromolar affinity. Previous studies have classified CD36-binding PfEMP1 proteins against non CD36-binding PfEMP1 proteins. This led to suggestion that only PfEMP1 from *var* gene groups B and C can bind to CD36[62,83,131]. However, this observed trend of strong CD36-binding PfEMP1 against weak binding CD36-binding PfEMP1 by comparing the binding affinities of different CIDR1α constructs to CD36 suggests it might be important to discern between the differences in binding affinities even for CD36-binding PfEMP1. This is important because epidemiological studies linking severe malaria with infected erythrocytes displaying CD36-binding abilities are not consistent in views. A very primitive postulation would be that the PfEMP1 involved in these studies might all be able to bind to CD36, but with significant differences in affinities to CD36, which affect the binding of infected erythrocytes to CD36 on the microvasculature under flow conditions[44,171,182,185–187]. Thus, it will be interesting to test this postulation further.

5.1.2. Crystallisation trials

Recombinant hCD36ED has also been successfully purified from baculovirus insect cell expression system. This hCD36ED has also been deglycosylated. The deglycosylated version has also been structurally characterised using circular dichroism to show that the overall secondary structure profile was nearly identical before and after deglycosylation. Deglycosylated hCD36ED has also been functionally tested by various biophysical interaction studies, revealing the formation of a 1:1 complex with CD36. Deglycosylated hCD36ED has also been made in complex with MC179 for crystallisation studies. Even though initial crystallisation trials did not give positive results, other crystallisation trials on-

going co-crystallisation complex between deglycosylated hCD36ED and oleic acid might prove to be fruitful.

5.1.3. Future work

5.1.3.1. Further characterisation of the context of CIDR1α binding domains

An important thing to do would be to confirm the central role of CIDR1α binding domain in mediating the PfEMP1 binding to CD36. This could be performed by making longer constructs of the CIDR1α-M2 domains that were tested to bind. This will give us a better answer as to whether CIDR1α domains can really account for the entire binding toward hCD36ED.

It will also be interesting to test binding affinities of the different CIDR1α constructs studied in different epidemiological studies both linking and not linking CD36-binding phenotype to severe malaria.

5.1.3.2. Crystallisation of hCD36ED

The most direct way of characterising the interaction between hCD36ED to PfEMP1 is to crystallise the co-complex. Even though initial crystallisation trials did not give positive results, crystallisation trials on co-crystallisation complex between deglycosylated hCD36ED and other CIDR1α domains should be done. The crystallisation of hCD36ED on its own or in complex with its native fatty acid ligand (oleic acid), which might stabilise the CD36 structure can also be tried in parallel. Indeed, I show here that oleic acid does not disrupt the fomation of the CD36:CIDR complex. The crystal structure of hCD36ED alone

will be important in answering questions not just within studies on malaria given the role of CD36 in different diseases.

5.2. DBL and IgM domains

5.2.1. Interaction between DBL domains and IgM

The interaction between DBL domains and IgM has been confirmed by biophysical studies. The interaction of MSPDBL2 with IgM has been shown using Ni-NTA affinity pull down assay, analytical size-exclusion chromatography and SPR response measurements. However, an intriguing finding from my studies was that recombinant DBL6ε was unable to bind to hIgM-Fc as opposed to previous studies[267,268]. His_6-tagged DBL6ε domain was only able to pull down hIgM-Fc to a small extent in a Ni-NTA affinity pull down assay. Interestingly, the DBL6ε also did not exhibit any binding to hIgM-Fc during SPR measurements. The experiments were repeated using different batches of each component and the results were the same. Moreover, the functionality of this DBL6ε construct has been determined both by SPR and structural studies with CSA previously in the laboratory[43]. Further examination of binding to hIgM-Fc can be performed with other DBL domains to confirm that the interaction was specific between DBL and hIgM-Fc.

5.2.2. Generation of a heterogeneous human IgM Fc core

Despite being able to express and purify hIgM-Fc from mammalian cells, AUC data suggests that hIgM-Fc is heterogonous in solution. This could account for the small response

measurements during the SPR movement. DBL domains are thought to only interact with multimeric IgM[127,270]. It is possible that only one species within the heterogeneous hIgM-Fc could interact with DBL domains. In essence, the multivalency of the recombinant hIgM-Fc prevented further biophysical and structural characterisation of the interaction with DBL domains.

5.2.3. Future work

Firstly, commercial human IgM from serum can be purchased for protease fragmentation to release the pentameric human IgM Fc core[239,240]. Indeed I conducted initial trials to attempt this. In our construct, the J chain that is thought to be important for pentameric IgM formation was not included. One advantage of performing human IgM fragmentation is the preservation of the J chain, which is thought to stabilise the formation of pentameric IgM in solution[248,250]. The homogeneity of the samples can be tested by AUC before conducting other experiments. If homogenous multimeric species could be obtained from cleavage of native IgM, this will suggest that the J chains are important in stabilising a certain multimeric conformation. Further binding interaction studies can then be performed.

However, it is difficult to assess and control the glycosylation patterns on the commercial material. Thus, it may not be straightforward processing them for crystallisation. We could therefore co-express J chains with the IgM core to allow formation of these complexes in HEK cells, allowing control over glycosylation.

5.3. Conclusion

I therefore present the production and first biophysical characterisation of CD36-CIDR1α and IgM-DBL complexes in this thesis, paving the way towards further structural characterisation in future. These studies will be important in further defining and understanding the structural basis for the interaction between PfEMP1 and their human partners in severe malaria. This will provide a platform for which useful drugs and vaccines can be developed to perturb these interactions during sequestration to reduce the incidences of severe malaria as a result of *P. falciparum* infection.

Bibliography

1. World Health Organisation World Malaria Report 2011. *WHO Press, Geneva, Switzerland.* (2011).

2. Miller, L. H., Baruch, D. I., Marsh, K. & Doumbo, O. K. The pathogenic basis of malaria. *Nature* **415**, 673–9 (2002).

3. Mendis, K., Sina, B. J., Marchesini, P. & Carter, R. The neglected burden of Plasmodium vivax malaria. *The American journal of tropical medicine and hygiene* **64**, 97–106 (2001).

4. Collins, W. E. & Jeffery, G. M. Plasmodium ovale: parasite and disease. *Clinical microbiology reviews* **18**, 570–81 (2005).

5. Collins, W. E. & Jeffery, G. M. Plasmodium malariae: parasite and disease. *Clinical microbiology reviews* **20**, 579–92 (2007).

6. Cox-Singh, J. *et al.* Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 165–71 (2008).

7. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).

8. Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature* **455**, 757–63 (2008).

9. Pain, a *et al.* The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* **455**, 799–803 (2008).

10. Rowe, J. A., Claessens, A., Corrigan, R. a & Arman, M. Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert reviews in molecular medicine* **11**, e16 (2009).

11. Krotoski, W. A. Discovery of the hypnozoite and a new theory of malarial relapse. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **79**, 1–11 (1985).

12. White, N. J. Determinants of relapse periodicity in Plasmodium vivax malaria. *Malaria journal* **10**, 297 (2011).

13. Garnham, P. C. . *et al.* The pre-erythrocytic stage of Plasmodium ovals. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **49**, 158–167 (1955).

14. Bousema, T. & Drakeley, C. Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. *Clinical microbiology reviews* **24**, 377–410 (2011).

15. Paul, R. E. L., Brey, P. T. & Robert, V. Plasmodium sex determination and transmission to mosquitoes. *Trends in parasitology* **18**, 32–8 (2002).

16. Mackintosh, C. L., Beeson, J. G. & Marsh, K. Clinical features and pathogenesis of severe malaria. *Trends in parasitology* **20**, 597–603 (2004).

17. Bartoloni, A. & Zammarchi, L. Clinical aspects of uncomplicated and severe malaria. *Mediterranean journal of hematology and infectious diseases* **4**, e2012026 (2012).

18. Miller, L. H., Aikawa, M., Johnson, J. G. & Shiroishi, T. Interaction between cytochalasin B-treated malarial parasites and erythrocytes. Attachment and junction formation. *The Journal of experimental medicine* **149**, 172–84 (1979).

19. Dvorak, J. A., Miller, L. H., Whitehouse, W. C. & Shiroishi, T. Invasion of erythrocytes by malaria merozoites. *Science (New York, N.Y.)* **187**, 748–50 (1975).

20. Gilson, P. R. & Crabb, B. S. Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. *International journal for parasitology* **39**, 91–6 (2009).

21. Aikawa, M., Miller, L. H., Johnson, J. & Rabbege, J. Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. *The Journal of cell biology* **77**, 72–82 (1978).

22. Richard, D. *et al.* Interaction between Plasmodium falciparum apical membrane antigen 1 and the rhoptry neck protein complex defines a key step in the erythrocyte invasion process of malaria parasites. *The Journal of biological chemistry* **285**, 14815–22 (2010).

23. Collins, C. R., Withers-Martinez, C., Hackett, F. & Blackman, M. J. An inhibitory antibody blocks interactions between components of the malarial invasion machinery. *PLoS pathogens* **5**, e1000273 (2009).

24. Bradley, P. J. *et al.* Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii. *The Journal of biological chemistry* **280**, 34245–58 (2005).

25. Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J.-F. & Lebrun, M. Export of a Toxoplasma gondii rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS pathogens* **5**, e1000309 (2009).

26. Baum, J., Papenfuss, A. T., Baum, B., Speed, T. P. & Cowman, A. F. Regulation of apicomplexan actin-based motility. *Nature reviews. Microbiology* **4**, 621–8 (2006).

27. Lauer, S. *et al.* Vacuolar uptake of host components, and a role for cholesterol and sphingomyelin in malarial infection. *The EMBO journal* **19**, 3556–64 (2000).

28. Harvey, K. L., Gilson, P. R. & Crabb, B. S. A model for the progression of receptor–ligand interactions during erythrocyte invasion by Plasmodium falciparum. *International Journal for Parasitology* **42**, 567–573 (2012).

29. Mitchell, G. H., Hadley, T. J., McGinniss, M. H., Klotz, F. W. & Miller, L. H. Invasion of erythrocytes by Plasmodium falciparum malaria parasites: evidence for receptor heterogeneity and two receptors. *Blood* **67**, 1519–21 (1986).

30. Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *The New England journal of medicine* **295**, 302–4 (1976).

31. Chotivanich, K. *et al.* Parasite multiplication potential and the severity of Falciparum malaria. *The Journal of infectious diseases* **181**, 1206–9 (2000).

32. Perkins, M. E. & Holt, E. H. Erythrocyte receptor recognition varies in Plasmodium falciparum isolates. *Molecular and biochemical parasitology* **27**, 23–34 (1988).

33. Narum, D. L. *et al.* Antibodies against the Plasmodium falciparum receptor binding domain of EBA-175 block invasion pathways that do not involve sialic acids. *Infection and immunity* **68**, 1964–6 (2000).

34. Adams, J. H., Blair, P. L., Kaneko, O. & Peterson, D. S. An expanding ebl family of Plasmodium falciparum. *Trends in parasitology* **17**, 297–9 (2001).

35. Stubbs, J. *et al.* Molecular mechanism for switching of P. falciparum invasion pathways into human erythrocytes. *Science (New York, N.Y.)* **309**, 1384–7 (2005).

36. Galinski, M. R., Medina, C. C., Ingravallo, P. & Barnwell, J. W. A reticulocyte-binding protein complex of Plasmodium vivax merozoites. *Cell* **69**, 1213–26 (1992).

37. Barnwell, J. W., Nichols, M. E. & Rubinstein, P. In vitro evaluation of the role of the Duffy blood group in erythrocyte invasion by Plasmodium vivax. *The Journal of experimental medicine* **169**, 1795–802 (1989).

38. Miller, L. H., Good, M. F. & Milon, G. Malaria pathogenesis. *Science (New York, N.Y.)* **264**, 1878–83 (1994).

39. 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).

40. Smith, J. D. *et al.* Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–10 (1995).

41. Baruch, D. I. *et al.* 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 (1995).

42. Fried, M. & Duffy, P. E. Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science (New York, N.Y.)* **272**, 1502–4 (1996).

43. Khunrae, P., Philip, J. M. D., Bull, D. R. & Higgins, M. K. Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. *Journal of molecular biology* **393**, 202–13 (2009).

44. Kyriacou, H. M. *et al.* Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. *Molecular and biochemical parasitology* **150**, 211–8 (2006).

45. Smith, J. D. *et al.* 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**, 1766–71 (2000).

46. Leech, J. H. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. *Journal of Experimental Medicine* **159**, 1567–1575 (1984).

47. Roberts, D. J. *et al.* Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357**, 689–92 (1992).

48. Biggs, B. a *et al.* Antigenic variation in Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9171–4 (1991).

49. Biggs, B. A. *et al.* Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of Plasmodium falciparum. *Journal of immunology (Baltimore, Md. : 1950)* **149**, 2047–54 (1992).

50. Bull, P. C. *et al.* Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature medicine* **4**, 358–60 (1998).

51. Gupta, S., Snow, R. W., Donnelly, C. a, Marsh, K. & Newbold, C. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nature medicine* **5**, 340–3 (1999).

52. Fried, M., Nosten, F., Brockman, a, Brabin, B. J. & Duffy, P. E. Maternal antibodies block malaria. *Nature* **395**, 851–2 (1998).

53. Khunrae, P. *et al.* Full-length recombinant Plasmodium falciparum VAR2CSA binds specifically to CSPG and induces potent parasite adhesion-blocking antibodies. *Journal of molecular biology* **397**, 826–34 (2010).

54. Staalsoe, T. *et al.* Variant surface antigen-specific IgG and protection against clinical consequences of pregnancy-associated Plasmodium falciparum malaria. *Lancet* **363**, 283–9 (2004).

55. Hernandez-Rivas, R. *et al.* Expressed var genes are found in Plasmodium falciparum subtelomeric regions. *Molecular and cellular biology* **17**, 604–11 (1997).

56. Rubio, J. P., Thompson, J. K. & Cowman, a F. The var genes of Plasmodium falciparum are located in the subtelomeric region of most chromosomes. *The EMBO journal* **15**, 4069–77 (1996).

57. Duraisingh, M. T. *et al.* Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell* **121**, 13–24 (2005).

58. Marty, A. J. *et al.* Evidence that Plasmodium falciparum chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. *Molecular microbiology* **62**, 72–83 (2006).

59. Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J.-C. & Scherf, A. A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. *The EMBO journal* **21**, 815–24 (2002).

60. Freitas-Junior, L. H. *et al.* Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature* **407**, 1018–22 (2000).

61. Taylor, H. M., Kyes, S. a & Newbold, C. I. Var gene diversity in Plasmodium falciparum is generated by frequent recombination events. *Molecular and biochemical parasitology* **110**, 391–7 (2000).

62. 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 computational biology* **6**, (2010).

63. Mu, J. *et al.* Recombination hotspots and population structure in Plasmodium falciparum. *PLoS biology* **3**, e335 (2005).

64. Scherf, a *et al.* Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. *The EMBO journal* **17**, 5418–26 (1998).

65. Freitas-Junior, L. H. *et al.* Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* **121**, 25–36 (2005).

66. Peters, J. *et al.* High diversity and rapid changeover of expressed var genes during the acute phase of Plasmodium falciparum infections in human volunteers. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10689–94 (2002).

67. Recker, M. *et al.* Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature* **429**, 555–8 (2004).

68. Molineaux, L. *et al.* Plasmodium falciparum parasitaemia described by a new mathematical model. *Parasitology* **122**, 379–91 (2001).

69. Paget-McNicol, S., Gatton, M., Hastings, I. & Saul, a The Plasmodium falciparum var gene switching rate, switching mechanism and patterns of parasite recrudescence described by mathematical modelling. *Parasitology* **124**, 225–35 (2002).

70. Recker, M. *et al.* Antigenic variation in Plasmodium falciparum malaria involves a highly structured switching pattern. *PLoS pathogens* **7**, e1001306 (2011).

71. Kraemer, S. M. & Smith, J. D. Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. *Molecular Microbiology* **50**, 1527–1538 (2003).

72. Moghimi, S. M. Mechanisms of splenic clearance of blood cells and particles: towards development of new splenotropic agents. *Advanced Drug Delivery Reviews* **17**, 103–115 (1995).

73. Kraemer, S. M. & Smith, J. D. A family affair: var genes, PfEMP1 binding, and malaria disease. *Current opinion in microbiology* **9**, 374–80 (2006).

74. Ho, M. & White, N. J. Molecular mechanisms of cytoadherence in malaria. *The American journal of physiology* **276**, C1231–42 (1999).

75. Udomsangpetch, R. *et al.* Plasmodium falciparum-infected erythrocytes form spontaneous erythrocyte rosettes. *The Journal of experimental medicine* **169**, 1835–40 (1989).

76. Kaul, D. K., Roth, E. F., Nagel, R. L., Howard, R. J. & Handunnetti, S. M. Rosetting of Plasmodium falciparum-infected red blood cells with uninfected red blood cells enhances microvascular obstruction under flow conditions. *Blood* **78**, 812–9 (1991).

77. Rowe, J. A., Moulds, J. M., Newbold, C. I. & Miller, L. H. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* **388**, 292–5 (1997).

78. Tandon, N. N., Kralisz, U. & Jamieson, G. a Identification of glycoprotein IV (CD36) as a primary receptor for platelet-collagen adhesion. *The Journal of biological chemistry* **264**, 7576–83 (1989).

79. Heddini, a *et al.* Binding of Plasmodium falciparum-infected erythrocytes to soluble platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31): frequent recognition by clinical isolates. *The American journal of tropical medicine and hygiene* **65**, 47–51 (2001).

80. Pain, a *et al.* Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1805–10 (2001).

81. Rasti, N., Wahlgren, M. & Chen, Q. Molecular aspects of malaria pathogenesis. *FEMS immunology and medical microbiology* **41**, 9–26 (2004).

82. Chen, Q. *et al.* Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum. *The Journal of experimental medicine* **187**, 15–23 (1998).

83. Janes, J. H. *et al.* Investigating the host binding signature on the Plasmodium falciparum PfEMP1 protein family. *PLoS pathogens* **7**, e1002032 (2011).

84. Frank, M. & Deitsch, K. Activation, silencing and mutually exclusive expression within the var gene family of Plasmodium falciparum. *International journal for parasitology* **36**, 975–85 (2006).

85. Deitsch, K. W., Calderwood, M. S. & Wellems, T. E. Malaria. Cooperative silencing elements in var genes. *Nature* **412**, 875–6 (2001).

86. Calderwood, M. S., Gannoun-Zaki, L., Wellems, T. E. & Deitsch, K. W. Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *The Journal of biological chemistry* **278**, 34125–32 (2003).

87. Dzikowski, R. *et al.* Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO reports* **8**, 959–65 (2007).

88. Lavstsen, T., Salanti, A., Jensen, A. T. R., Arnot, D. E. & Theander, T. G. Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malaria journal* **2**, 27 (2003).

89. Smith, J. D., Subramanian, G., Gamain, B., Baruch, D. I. & Miller, L. H. Classification of adhesive domains in the Plasmodium falciparum Erythrocyte Membrane Protein 1 family. *Molecular and Biochemical Parasitology* **110**, 293–310 (2000).

90. Salanti, A. *et al.* A sub-family of common and highly conserved Plasmodium falciparum var genes. *Molecular and biochemical parasitology* **122**, 111–5 (2002).

91. Salanti, A. *et al.* Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. *Molecular Microbiology* **49**, 179–191 (2003).

92. Trimnell, A. R. *et al.* Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. *Molecular and biochemical parasitology* **148**, 169–80 (2006).

93. Smith, J. D., Gamain, B., Baruch, D. I. & Kyes, S. Decoding the language of var genes and Plasmodium falciparum sequestration. *Trends in parasitology* **17**, 538–45 (2001).

94. Kriek, N. *et al.* Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes. *Molecular Microbiology* **50**, 1215–1227 (2003).

95. Oh, S. S. *et al.* Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. *Molecular and biochemical parasitology* **108**, 237–47 (2000).

96. Waller, K. L., Cooke, B. M., Nunomura, W., Mohandas, N. & Coppel, R. L. Mapping the binding domains involved in the interaction between the Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and the cytoadherence ligand P. falciparum erythrocyte membrane protein 1 (PfEMP1). *The Journal of biological chemistry* **274**, 23808–13 (1999).

97. Luse, S. A. & Miller, L. H. Plasmodium falciparum malaria. Ultrastructure of parasitized erythrocytes in cardiac vessels. *The American journal of tropical medicine and hygiene* **20**, 655–60 (1971).

98. Maier, A. G., Cooke, B. M., Cowman, A. F. & Tilley, L. Malaria parasite proteins that remodel the host erythrocyte. *Nature reviews. Microbiology* **7**, 341–54 (2009).

99. Crabb, B. S. *et al.* Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* **89**, 287–96 (1997).

100. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science (New York, N.Y.)* **306**, 1930–3 (2004).

101. Juillerat, A. *et al.* Structure of a Plasmodium falciparum PfEMP1 rosetting domain reveals a role for the N-terminal segment in heparin-mediated rosette inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5243–8 (2011).

102. Vigan-Womas, I. *et al.* Structural Basis for the ABO Blood-Group Dependence of Plasmodium falciparum Rosetting. *PLoS Pathogens* **8**, e1002781 (2012).

103. Chitnis, C. E. & Miller, L. H. Identification of the erythrocyte binding domains of Plasmodium vivax and Plasmodium knowlesi proteins involved in erythrocyte invasion. *The Journal of experimental medicine* **180**, 497–506 (1994).

104. Higgins, M. K. The structure of a chondroitin sulfate-binding domain important in placental malaria. *The Journal of biological chemistry* **283**, 21842–6 (2008).

105. Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E. & Sharma, A. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* **439**, 741–4 (2006).

106. Tolia, N. H., Enemark, E. J., Sim, B. K. L. & Joshua-Tor, L. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. *Cell* **122**, 183–93 (2005).

107. Andersen, P. *et al.* Structural insight into epitopes in the pregnancy-associated malaria protein VAR2CSA. *PLoS pathogens* **4**, e42 (2008).

108. Dahlbäck, M. *et al.* Epitope mapping and topographic analysis of VAR2CSA DBL3X involved in P. falciparum placental sequestration. *PLoS pathogens* **2**, e124 (2006).

109. Howell, D. P.-G. *et al.* Mapping a common interaction site used by Plasmodium falciparum Duffy binding-like domains to bind diverse host receptors. *Molecular microbiology* **67**, 78–87 (2008).

110. Singh, K. *et al.* Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A. *Nature structural & molecular biology* **15**, 932–8 (2008).

111. Singh, K. *et al.* Subdomain 3 of Plasmodium falciparum VAR2CSA DBL3x is identified as a minimal chondroitin sulfate A-binding region. *The Journal of biological chemistry* **285**, 24855–62 (2010).

112. Baruch, D. I. *et al.* Identification of a region of PfEMP1 that mediates adherence of Plasmodium falciparum infected erythrocytes to CD36: conserved function with variant sequence. *Blood* **90**, 3766–75 (1997).

113. 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**, 3268–74 (2001).

114. 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 pathogens* **4**, e1000147 (2008).

115. Springer, A. L., Smith, L. M., Mackay, D. Q., Nelson, S. O. & Smith, J. D. Functional interdependence of the DBLbeta domain and c2 region for binding of the Plasmodium falciparum variant antigen to ICAM-1. *Molecular and biochemical parasitology* **137**, 55–64 (2004).

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

117. 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**, 57–9 (1989).

118. Rogerson, S. J., Chaiyaroj, S. C., Ng, K., Reeder, J. C. & Brown, G. V. Chondroitin sulfate A is a cell surface receptor for Plasmodium falciparum-infected erythrocytes. *The Journal of experimental medicine* **182**, 15–20 (1995).

119. Cooke, B. M. *et al.* Rolling and stationary cytoadhesion of red blood cells parasitized by Plasmodium falciparum: separate roles for ICAM-1, CD36 and thrombospondin. *British journal of haematology* **87**, 162–70 (1994).

120. Baruch, D. I., Gormely, J. a, Ma, C., Howard, R. J. & Pasloske, B. L. 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**, 3497–502 (1996).

121. Rogerson, S. J., Novakovic, S., Cooke, B. M. & Brown, G. V. Plasmodium falciparum-infected erythrocytes adhere to the proteoglycan thrombomodulin in static and flow-based systems. *Experimental parasitology* **86**, 8–18 (1997).

122. Ockenhouse, C. F. *et al.* Human vascular endothelial cell adhesion receptors for Plasmodium falciparum-infected erythrocytes: roles for endothelial leukocyte adhesion molecule 1 and vascular cell adhesion molecule 1. *The Journal of experimental medicine* **176**, 1183–9 (1992).

123. Beeson, J. G. *et al.* Adhesion of Plasmodium falciparum-infected erythrocytes to hyaluronic acid in placental malaria. *Nature medicine* **6**, 86–90 (2000).

124. Vogt, A. M. *et al.* Heparan sulfate on endothelial cells mediates the binding of Plasmodium falciparum-infected erythrocytes via the DBL1alpha domain of PfEMP1. *Blood* **101**, 2405–11 (2003).

125. Flick, K. *et al.* Role of nonimmune IgG bound to PfEMP1 in placental malaria. *Science (New York, N.Y.)* **293**, 2098–100 (2001).

126. Rowe, J. A., Shafi, J., Kai, O. K., Marsh, K. & Raza, A. Nonimmune IgM, but not IgG binds to the surface of Plasmodium falciparum-infected erythrocytes and correlates with rosetting and severe malaria. *The American journal of tropical medicine and hygiene* **66**, 692–9 (2002).

127. Ghumra, A. *et al.* Identification of residues in the Cmu4 domain of polymeric IgM essential for interaction with Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1). *Journal of immunology (Baltimore, Md. : 1950)* **181**, 1988–2000 (2008).

128. Donati, D. *et al.* Identification of a polyclonal B-cell activator in Plasmodium falciparum. *Infection and immunity* **72**, 5412–8 (2004).

129. Creasey, A., Staalsoe, T. & Raza, A. Nonspecific immunoglobulin M binding and chondroitin sulfate A binding are linked phenotypes of Plasmodium falciparum isolates implicated in malaria during. *Infection and* **71**, 4767–4771 (2003).

130. Barragan, A., Kremsner, P. G., Wahlgren, M. & Carlson, J. Blood group A antigen is a coreceptor in Plasmodium falciparum rosetting. *Infection and immunity* **68**, 2971–5 (2000).

131. 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. *Molecular microbiology* **47**, 1265–78 (2003).

132. Reed, S. E., Staley, E. M., Mayginnes, J. P., Pintel, D. J. & Tullis, G. E. Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. *Journal of virological methods* **138**, 85–98 (2006).

133. Louis-Jeune, C., Andrade-Navarro, M. a & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* 374–381 (2011).doi:10.1002/prot.23188

134. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochimica et biophysica acta* **1751**, 119–39 (2005).

135. Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science : a publication of the Protein Society* **11**, 2067–79 (2002).

136. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **78**, 1606–19 (2000).

137. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical Biochemistry* **320**, 104–124 (2003).

138. Okumura, T. & Jamieson, G. A. Platelet glycocalicin. I. Orientation of glycoproteins of the human platelet surface. *The Journal of biological chemistry* **251**, 5944–9 (1976).

139. Podolsak, B. Effects of thrombin, chymotrypsin and aggregated gamma-globulins on the proteins of the human platelet membrane. *Thrombosis and haemostasis* **37**, 396–406 (1977).

140. Endemann, G. *et al.* CD36 is a receptor for oxidized low density lipoprotein. *The Journal of biological chemistry* **268**, 11811–6 (1993).

141. Greenwalt, D. E. *et al.* Membrane glycoprotein CD36: a review of its roles in adherence, signal transduction, and transfusion medicine. *Blood* **80**, 1105–15 (1992).

142. Harmon, C. M. & Abumrad, N. A. Binding of sulfosuccinimidyl fatty acids to adipocyte membrane proteins: isolation and amino-terminal sequence of an 88-kD protein implicated in transport of long-chain fatty acids. *The Journal of membrane biology* **133**, 43–9 (1993).

143. Coburn, C. T. Defective Uptake and Utilization of Long Chain Fatty Acids in Muscle and Adipose Tissues of CD36 Knockout Mice. *Journal of Biological Chemistry* **275**, 32523–32529 (2000).

144. Podrez, E. a *et al.* Identification of a novel family of oxidized phospholipids that serve as ligands for the macrophage scavenger receptor CD36. *The Journal of biological chemistry* **277**, 38503–16 (2002).

145. Febbraio, M. *et al.* A null mutation in murine CD36 reveals an important role in fatty acid and lipoprotein metabolism. *The Journal of biological chemistry* **274**, 19055–62 (1999).

146. Hajri, T., Han, X. X., Bonen, A. & Abumrad, N. A. Defective fatty acid uptake modulates insulin responsiveness and metabolic responses to diet in CD36-null mice. *The Journal of clinical investigation* **109**, 1381–9 (2002).

147. Platt, N., da Silva, R. P. & Gordon, S. Recognizing death: the phagocytosis of apoptotic cells. *Trends in cell biology* **8**, 365–72 (1998).

148. Silverstein, R. L. & Febbraio, M. CD36 and atherosclerosis. *Current opinion in lipidology* **11**, 483–91 (2000).

149. Febbraio, M., Hajjar, D. P. & Silverstein, R. L. CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. *The Journal of clinical investigation* **108**, 785–91 (2001).

150. Rać, M. E., Safranow, K. & Poncyljusz, W. Molecular basis of human CD36 gene mutations. *Molecular medicine (Cambridge, Mass.)* **13**, 288–96 (2007).

151. Gautam, S. & Banerjee, M. The macrophage Ox-LDL receptor, CD36 and its association with type II diabetes mellitus. *Molecular genetics and metabolism* **102**, 389–98 (2011).

152. Febbraio, M. *et al.* Targeted disruption of the class B scavenger receptor CD36 protects against atherosclerotic lesion development in mice. *The Journal of clinical investigation* **105**, 1049–56 (2000).

153. Barnwell, J. W. *et al.* A human 88-kD membrane glycoprotein (CD36) functions in vitro as a receptor for a cytoadherence ligand on Plasmodium falciparum-infected erythrocytes. *The Journal of clinical investigation* **84**, 765–72 (1989).

154. Fernández-Ruiz, E., Armesilla, A. L., Sánchez-Madrid, F. & Vega, M. A. Gene encoding the collagen type I and thrombospondin receptor CD36 is located on chromosome 7q11.2. *Genomics* **17**, 759–61 (1993).

155. Armesilla, a L. & Vega, M. a Structural organization of the gene for human CD36 glycoprotein. *The Journal of biological chemistry* **269**, 18985–91 (1994).

156. Armesilla, A. L., Calvo, D. & Vega, M. A. Structural and functional characterization of the human CD36 gene promoter: identification of a proximal PEBP2/CBF site. *The Journal of biological chemistry* **271**, 7781–7 (1996).

157. Taylor, K. T., Tang, Y., Sobieski, D. A. & Lipsky, R. H. Characterization of two alternatively spliced 5'-untranslated exons of the human CD36 gene in different cell types. *Gene* **133**, 205–12 (1993).

158. Noguchi, K. *et al.* cDNA expression cloning of the 85-kDa protein overexpressed in adriamycin-resistant cells. *Biochemical and biophysical research communications* **192**, 88–95 (1993).

159. Tang, Y., Taylor, K. T., Sobieski, D. A., Medved, E. S. & Lipsky, R. H. Identification of a human CD36 isoform produced by exon skipping. Conservation of exon organization and pre-mRNA splicing patterns with a CD36 gene family member, CLA-1. *The Journal of biological chemistry* **269**, 6011–5 (1994).

160. Asch, A. S. *et al.* Analysis of CD36 binding domains: ligand specificity controlled by dephosphorylation of an ectodomain. *Science (New York, N.Y.)* **262**, 1436–40 (1993).

161. Vega, M. a *et al.* Cloning, sequencing, and expression of a cDNA encoding rat LIMP II, a novel 74-kDa lysosomal membrane protein related to the surface adhesion protein CD36. *The Journal of biological chemistry* **266**, 16818–24 (1991).

162. Abumrad, N. a, el-Maghrabi, M. R., Amri, E. Z., Lopez, E. & Grimaldi, P. a Cloning of a rat adipocyte membrane protein implicated in binding or transport of long-chain fatty acids that is induced during preadipocyte differentiation. Homology with human CD36. *The Journal of biological chemistry* **268**, 17665–8 (1993).

163. Gruarin, P. *et al.* CD36 is a ditopic glycoprotein with the N-terminal domain implicated in intracellular transport. *Biochemical and biophysical research communications* **275**, 446–54 (2000).

164. Rasmussen, J. T., Berglund, L., Rasmussen, M. S. & Petersen, T. E. Assignment of disulfide bridges in bovine CD36. *European journal of biochemistry / FEBS* **257**, 488–94 (1998).

165. Tao, N., Wagner, S. J. & Lublin, D. M. CD36 is palmitoylated on both N- and C-terminal cytoplasmic tails. *The Journal of biological chemistry* **271**, 22315–20 (1996).

166. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry* **182**, 319–26 (1989).

167. Kieffer, N. *et al.* Developmentally regulated expression of a 78 kDa erythroblast membrane glycoprotein immunologically related to the platelet thrombospondin receptor. *The Biochemical journal* **262**, 835–42 (1989).

168. Greenwalt, D. E., Watt, K. W., So, O. Y. & Jiwani, N. PAS IV, an integral membrane protein of mammary epithelial cells, is related to platelet and endothelial cell CD36 (GP IV). *Biochemistry* **29**, 7054–9 (1990).

169. Hoosdally, S. J., Andress, E. J., Wooding, C., Martin, C. a & Linton, K. J. The Human Scavenger Receptor CD36: glycosylation status and its role in trafficking and function. *The Journal of biological chemistry* **284**, 16277–88 (2009).

170. Berglund, L., Petersen, T. E. & Rasmussen, J. T. Structural characterization of bovine CD36 from the milk fat globule membrane. *Biochimica et biophysica acta* **1309**, 63–8 (1996).

171. Ockenhouse, C. F. *et al.* Molecular basis of sequestration in severe and uncomplicated Plasmodium falciparum malaria: differential adhesion of infected erythrocytes to CD36 and ICAM-1. *The Journal of infectious diseases* **164**, 163–9 (1991).

172. Ho, M., Hickey, M. J., Murray, a G., Andonegui, G. & Kubes, P. Visualization of Plasmodium falciparum-endothelium interactions in human microvasculature: mimicry of leukocyte recruitment. *The Journal of experimental medicine* **192**, 1205–11 (2000).

173. Yipp, B. G. *et al.* Differential roles of CD36, ICAM-1, and P-selectin in Plasmodium falciparum cytoadherence in vivo. *Microcirculation (New York, N.Y. : 1994)* **14**, 593–602 (2007).

174. Udomsangpetch, R. *et al.* Receptor specificity of clinical Plasmodium falciparum isolates: nonadherence to cell-bound E-selectin and vascular cell adhesion molecule-1. *Blood* **88**, 2754–60 (1996).

175. Handunnetti, S. M. *et al.* Involvement of CD36 on erythrocytes as a rosetting receptor for Plasmodium falciparum-infected erythrocytes. *Blood* **80**, 2097–104 (1992).

176. Daviet, L. *et al.* Characterization of two vaccinia CD36 recombinant-virus-generated monoclonal antibodies (10/5, 13/10): effects on malarial cytoadherence and platelet functions. *European journal of biochemistry / FEBS* **243**, 344–9 (1997).

177. 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**, 2121–7 (1999).

178. Smith, J. D. *et al.* Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. *Molecular and biochemical parasitology* **97**, 133–48 (1998).

179. Miller, L. H., Hudson-Taylor, D., Gamain, B. & Saul, A. J. Definition of the minimal domain of CIDR1alpha of Plasmodium falciparum PfEMP1 for binding CD36. *Molecular and biochemical parasitology* **120**, 321–3 (2002).

180. Reeder, J. C. *et al.* The adhesion of Plasmodium falciparum-infected erythrocytes to chondroitin sulfate A is mediated by P. falciparum erythrocyte membrane protein 1. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5198–202 (1999).

181. Gamain, B., Gratepanche, S., Miller, L. H. & Baruch, D. I. Molecular basis for the dichotomy in Plasmodium falciparum adhesion to CD36 and chondroitin sulfate A. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10020–4 (2002).

182. Newbold, C. *et al.* Receptor-specific adhesion and clinical disease in Plasmodium falciparum. *The American journal of tropical medicine and hygiene* **57**, 389–98 (1997).

183. Beeson, J. G. *et al.* Plasmodium falciparum isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties. *The Journal of infectious diseases* **180**, 464–72 (1999).

184. Buffet, P. a *et al.* Plasmodium falciparum domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 12743–8 (1999).

185. Rogerson, S. J. *et al.* Cytoadherence characteristics of Plasmodium falciparum-infected erythrocytes from Malawian children with severe and uncomplicated malaria. *The American journal of tropical medicine and hygiene* **61**, 467–72 (1999).

186. Heddini, A. *et al.* Fresh isolates from children with severe Plasmodium falciparum malaria bind to multiple receptors. *Infection and immunity* **69**, 5849–56 (2001).

187. Ho, M. *et al.* Clinical correlates of in vitro Plasmodium falciparum cytoadherence. *Infection and immunity* **59**, 873–8 (1991).

188. Aitman, T. J. *et al.* Malaria susceptibility and CD36 mutation. *Nature* **405**, 1015–6 (2000).

189. Pain, A. *et al.* A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet* **357**, 1502–3 (2001).

190. Omi, K., Ohashi, J. & Patarapotikul, J. < i> CD36</i> Polymorphism Is Associated with Protection from Cerebral Malaria. *The American Journal of* **36**, 364–374 (2003).

191. Chilongola, J., Balthazary, S., Mpina, M., Mhando, M. & Mbugi, E. CD36 deficiency protects against malarial anaemia in children by reducing Plasmodium falciparum-infected red blood cell adherence to vascular endothelium. *Tropical medicine & international health : TM & IH* **14**, 810–6 (2009).

192. Amodu, O. K. *et al.* Plasmodium falciparum malaria in south-west Nigerian children: is the polymorphism of ICAM-1 and E-selectin genes contributing to the clinical severity of malaria? *Acta tropica* **95**, 248–55 (2005).

193. Ayodo, G. *et al.* Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants. *American journal of human genetics* **81**, 234–42 (2007).

194. Fry, A. E. *et al.* Positive selection of a CD36 nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. *Human molecular genetics* **18**, 2683–92 (2009).

195. Davis, S. J. & Crispin, M. *Functional and Structural Proteomics of Glycoproteins*. (Springer Netherlands: Dordrecht, 2011).doi:10.1007/978-90-481-9355-4

196. Krogh, a, Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567–80 (2001).

197. Jimenez-Dalmaroni, M. J. *et al.* Soluble CD36 ectodomain binds negatively charged diacylglycerol ligands and acts as a co-receptor for TLR2. *PloS one* **4**, e7411 (2009).

198. Pearson, W. R. Empirical statistical estimates for sequence similarity searches. *Journal of molecular biology* **276**, 71–84 (1998).

199. Romero, Obradovic & Dunker, K. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics. Workshop on Genome Informatics* **8**, 110–124 (1997).

200. Li, X., Romero, P., Rani, M., Dunker, A. & Obradovic, Z. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics* **10**, 30–40 (1999).

201. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).

202. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725–38 (2010).

203. Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of molecular biology* **310**, 243–57 (2001).

204. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).

205. Baca, A. M. & Hol, W. G. Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. *International journal for parasitology* **30**, 113–8 (2000).

206. Veggiani, G. & de Marco, A. Improved quantitative and qualitative production of single-domain intrabodies mediated by the co-expression of Erv1p sulfhydryl oxidase. *Protein expression and purification* **79**, 111–4 (2011).

207. Nguyen, V. D. *et al.* Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microbial cell factories* **10**, 1 (2011).

208. Pattnaik, P. Surface plasmon resonance: applications in understanding receptor-ligand interaction. *Applied biochemistry and biotechnology* **126**, 79–92 (2005).

209. Winzor, D. J. Analytical exclusion chromatography. *Journal of Biochemical and Biophysical Methods* **56**, 15–52 (2003).

210. Hagel, L. Gel filtration: size exclusion chromatography. *Methods of biochemical analysis* **54**, 51–91 (2011).

211. McRorie, D. K. & Voelker, P. J. Self-Associating Systems in the Analytical Ultracentrifuge. *Beckman*

212. Cole, J. L., Lary, J. W., P Moody, T. & Laue, T. M. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in cell biology* **84**, 143–79 (2008).

213. Howlett, G. J., Minton, A. P. & Rivas, G. Analytical ultracentrifugation for the study of protein association and assembly. *Current opinion in chemical biology* **10**, 430–6 (2006).

214. Altmann, F., Staudacher, E., Wilson, I. B. & März, L. Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconjugate journal* **16**, 109–23 (1999).

215. Jarvis, D. L. *Baculovirus-insect cell expression systems. Methods in enzymology* **463**, 191–222 (Elsevier Inc.: 2009).

216. Jarvis, D. L., Kawar, Z. S. & Hollister, J. R. Engineering N-glycosylation pathways in the baculovirus-insect cell system. *Current opinion in biotechnology* **9**, 528–33 (1998).

217. Davis, T. R., Trotter, K. M., Granados, R. R. & Wood, H. A. Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production, glycosylation and secretion. *Bio/technology (Nature Publishing Company)* **10**, 1148–50 (1992).

218. Rendic, D., Wilson, I. B. H. & Paschinger, K. ChemInform Abstract: The Glycosylation Capacity of Insect Cells. *ChemInform* **39**, 7–21 (2008).

219. Tomiya, N. *et al.* Purification, characterization, and cloning of a Spodoptera frugiperda Sf9 beta-N-acetylhexosaminidase that hydrolyzes terminal N-acetylglucosamine on the N-glycan core. *The Journal of biological chemistry* **281**, 19545–60 (2006).

220. Hsu, T. a *et al.* Differential N-glycan patterns of secreted and intracellular IgG produced in Trichoplusia ni cells. *The Journal of biological chemistry* **272**, 9062–70 (1997).

221. Kulakosky, P. C., Hughes, P. R. & Wood, H. a N-Linked glycosylation of a baculovirus-expressed recombinant glycoprotein in insect larvae and tissue culture cells. *Glycobiology* **8**, 741–5 (1998).

222. Hancock, K. *et al.* False positive reactivity of recombinant, diagnostic, glycoproteins produced in High Five insect cells: effect of glycosylation. *Journal of immunological methods* **330**, 130–6 (2008).

223. Davis, S. J. *et al.* Ligand binding by the immunoglobulin superfamily recognition molecule CD2 is glycosylation-independent. *The Journal of biological chemistry* **270**, 369–75 (1995).

224. Petrescu, A.-J., Milac, A.-L., Petrescu, S. M., Dwek, R. a & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103–14 (2004).

225. Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure (London, England : 1993)* **15**, 267–73 (2007).

226. Kwong, P. D. *et al.* Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *The Journal of biological chemistry* **274**, 4115–23 (1999).

227. Robbins, P. W., Wirth, D. F. & Hering, C. Expression of the Streptomyces enzyme endoglycosidase H in Escherichia coli. *The Journal of biological chemistry* **256**, 10640–4 (1981).

228. Tarentino, A. L., Quinones, G., Changchien, L. M. & Plummer, T. H. Multiple endoglycosidase F activities expressed by Flavobacterium meningosepticum endoglycosidases F2 and F3. Molecular cloning, primary sequence, and enzyme expression. *The Journal of biological chemistry* **268**, 9702–8 (1993).

229. Greenfield, N. J. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Analytical biochemistry* **235**, 1–10 (1996).

230. Johnson, W. C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**, 307–12 (1999).

231. Noushmehr, H. *et al.* Fatty acid translocase (FAT/CD36) is localized on insulin-containing granules in human pancreatic beta-cells and mediates fatty acid effects on insulin secretion. *Diabetes* **54**, 472–81 (2005).

232. Boes, M., Prodeus, A. P., Schmidt, T., Carroll, M. C. & Chen, J. A critical role of natural immunoglobulin M in immediate defense against systemic bacterial infection. *The Journal of experimental medicine* **188**, 2381–6 (1998).

233. Ochsenbein, a. F. Control of Early Viral and Bacterial Distribution and Disease by Natural Antibodies. *Science* **286**, 2156–2159 (1999).

234. Zhou, Z., Tzioufas, A. G. & Notkins, A. L. Properties and function of polyreactive antibodies and polyreactive antigen-binding B cells. *Journal of autoimmunity* **29**, 219–28 (2007).

235. Racine, R. & Winslow, G. M. IgM in microbial infections: taken for granted? *Immunology letters* **125**, 79–85 (2009).

236. Janeway, C. A., Travers, P., Walport, M. & Shlomchik., M. J. *Immunobiology: The Immune System in Health and Disease. 5th edition.* 27106 (Garland Science: 2001).

237. Beale, D. & Buttress, N. Studies on a human 19-S immunoglobulin M. The arrangement of inter-chain disulphide bridges and carbohydrate sites. *Biochimica et biophysica acta* **181**, 250–67 (1969).

238. Beale, D. & Feinstein, A. Evidence for C-terminal intra-subunit disulphide bridges between immunoglobulin-M heavy chains. *FEBS letters* **7**, 175–176 (1970).

239. Plaut Immunoglobulin M: pentameric Fcμ fragments released by trypsin at higher temperatures. *Proceedings of the National* (1970).

240. Volkov, V. V. *et al.* Low-resolution structure of immunoglobulins IgG1, IgM and rheumatoid factor IgM-RF from solution X-ray scattering data. *Journal of Applied Crystallography* **36**, 503–508 (2003).

241. Sitia, R., Rubartelli, a & Hämmerling, U. The role of glycosylation in secretion and membrane expression of immunoglobulins M and A. *Molecular immunology* **21**, 709–19 (1984).

242. Kehry, M. *et al.* The immunoglobulin mu chains of membrane-bound and secreted IgM molecules differ in their C-terminal segments. *Cell* **21**, 393–406 (1980).

243. Feinstein, A. & Munn, E. A. Conformation of the free and antigen-bound IgM antibody molecules. *Nature* **224**, 1307–9 (1969).

244. Czajkowsky, D. M. & Shao, Z. The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14960–5 (2009).

245. Klimovich, V. B. IgM and its receptors: structural and functional aspects. *Biochemistry. Biokhimiia* **76**, 534–49 (2011).

246. Dorrington, K. J. & Mihaesco, C. Subunit structure o human gamma-M-globulins. *Immunochemistry* **7**, 651–60 (1970).

247. Mestecky, J. & Schrohenloher, R. E. Site of attachment of J chain to human immunoglobulin M. *Nature* **249**, 650–2 (1974).

248. Koshland, M. E. Structure and function of the J chain. *Advances in immunology* **20**, 41–69 (1975).

249. Davis, a C., Roux, K. H. & Shulman, M. J. On the structure of polymeric IgM. *European journal of immunology* **18**, 1001–8 (1988).

250. Randall, T. D., Parkhouse, R. M. & Corley, R. B. J chain synthesis and secretion of hexameric IgM is differentially regulated by lipopolysaccharide and interleukin 5. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 962–6 (1992).

251. Randall, T. D., King, L. B. & Corley, R. B. The biological effects of IgM hexamer formation. *European journal of immunology* **20**, 1971–9 (1990).

252. Hughey, C. T., Brewer, J. W., Colosia, A. D., Rosse, W. F. & Corley, R. B. Production of IgM hexamers by normal and autoimmune B cells: implications for the physiologic role of hexameric IgM. *Journal of immunology (Baltimore, Md. : 1950)* **161**, 4091–7 (1998).

253. Petrušić, V. *et al.* Hexameric immunoglobulin M in humans: desired or unwanted? *Medical hypotheses* **77**, 959–61 (2011).

254. Brewer, J. W., Randall, T. D., Parkhouse, R. M. & Corley, R. B. IgM hexamers? *Immunology today* **15**, 165–8 (1994).

255. Collins, C., Tsui, F. W. L. & Shulman, M. J. Differential activation of human and guinea pig complement by pentameric and hexameric IgM. *European journal of immunology* **32**, 1802–10 (2002).

256. Kubagawa, H. *et al.* Identity of the elusive IgM Fc receptor (FcmuR) in humans. *The Journal of experimental medicine* **206**, 2779–93 (2009).

257. Shibuya, A. & Honda, S.-I. Molecular and functional characteristics of the Fcalpha/muR, a novel Fc receptor for IgM and IgA. *Springer seminars in immunopathology* **28**, 377–82 (2006).

258. Shibuya, a *et al.* Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. *Nature immunology* **1**, 441–6 (2000).

259. Honda, S. *et al.* Enhanced humoral immune responses against T-independent antigens in Fc alpha/muR-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11230–5 (2009).

260. Kikuno, K. *et al.* Unusual biochemical features and follicular dendritic cell expression of human Fcalpha/mu receptor. *European journal of immunology* **37**, 3540–50 (2007).

261. Ghumra, A. *et al.* Structural requirements for the interaction of human IgM and IgA with the human Fcalpha/mu receptor. *European journal of immunology* **39**, 1147–56 (2009).

262. Chen, Y., Park, Y.-B., Patel, E. & Silverman, G. J. IgM antibodies to apoptosis-associated determinants recruit C1q and enhance dendritic cell phagocytosis of apoptotic cells. *Journal of immunology (Baltimore, Md. : 1950)* **182**, 6031–43 (2009).

263. Grönwall, C., Vas, J. & Silverman, G. J. Protective Roles of IgM Antibodies. *Frontiers in immunology* **3**, 66 (2012).

264. Pleass, R. J., Areschoug, T., Lindahl, G. & Woof, J. M. Streptococcal IgA-binding proteins bind in the Calpha 2-Calpha 3 interdomain region and inhibit binding of IgA to human CD89. *The Journal of biological chemistry* **276**, 8197–204 (2001).

265. Nagashunmugam, T. *et al.* In vivo immune evasion mediated by the herpes simplex virus type 1 immunoglobulin G Fc receptor. *Journal of virology* **72**, 5351–9 (1998).

266. Pleass, R. J. & Woof, J. M. Fc receptors and immunity to parasites. *Trends in parasitology* **17**, 545–51 (2001).

267. Semblat, J.-P., Raza, A., Kyes, S. A. & Rowe, J. A. Identification of Plasmodium falciparum var1CSA and var2CSA domains that bind IgM natural antibodies. *Molecular and biochemical parasitology* **146**, 192–7 (2006).

268. Rasti, N. *et al.* Nonimmune immunoglobulin binding and multiple adhesion characterize Plasmodium falciparum-infected erythrocytes of placental origin. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13795–800 (2006).

269. Chen, Q. *et al.* The semiconserved head structure of Plasmodium falciparum erythrocyte membrane protein 1 mediates binding to multiple independent host receptors. *The Journal of experimental medicine* **192**, 1–10 (2000).

270. Czajkowsky, D. M. *et al.* IgM, Fc mu Rs, and malarial immune evasion. *Journal of immunology (Baltimore, Md. : 1950)* **184**, 4597–603 (2010).

271. Pleass, R. J., Dunlop, J. I., Anderson, C. M. & Woof, J. M. Identification of residues in the CH2/CH3 domain interface of IgA essential for interaction with the human fcalpha receptor (FcalphaR) CD89. *The Journal of biological chemistry* **274**, 23508–14 (1999).

272. Mongini, P. K., Blessinger, C. a, Chiorazzi, N., Rajaram, N. & Rudich, S. M. A monovalent C mu 4-specific ligand enhances the activation of human B cells by membrane IgM cross-linking ligands. *International immunology* **7**, 317–30 (1995).

273. Carroll, M. C. The complement system in regulation of adaptive immunity. *Nature immunology* **5**, 981–6 (2004).

274. Good, M. F., Kaslow, D. C. & Miller, L. H. Pathways and strategies for developing a malaria blood-stage vaccine. *Annual review of immunology* **16**, 57–87 (1998).

275. McIntosh, R. S. *et al.* The importance of human FcgammaRI in mediating protection to malaria. *PLoS pathogens* **3**, e72 (2007).

276. Barfod, L. *et al.* Evasion of immunity to Plasmodium falciparum malaria by IgM masking of protective IgG epitopes in infected erythrocyte surface-exposed PfEMP1. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12485–90 (2011).

277. Clough, B., Atilola, F. A., Black, J. & Pasvol, G. Plasmodium falciparum: the importance of IgM in the rosetting of parasite-infected erythrocytes. *Experimental parasitology* **89**, 129–32 (1998).

278. Treutiger, C. J. *et al.* Rouleaux-forming serum proteins are involved in the rosetting of Plasmodium falciparum-infected erythrocytes. *Experimental parasitology* **93**, 215–24 (1999).

279. Scholander, C., Treutiger, C. J., Hultenby, K. & Wahlgren, M. Novel fibrillar structure confers adhesive property to malaria-infected erythrocytes. *Nature medicine* **2**, 204–8 (1996).

280. Somner, E. A., Black, J. & Pasvol, G. Multiple human serum components act as bridging molecules in rosette formation by Plasmodium falciparum-infected erythrocytes. *Blood* **95**, 674–82 (2000).

281. Rogerson, S. J., Beeson, J. G., Mhango, C. G., Dzinjalamala, F. K. & Molyneux, M. E. Plasmodium falciparum Rosette Formation Is Uncommon in Isolates from Pregnant Women. *Infection and Immunity* **68**, 391–393 (2000).

282. Higgins, M. K. Overproduction, purification and crystallization of a chondroitin sulfate A-binding DBL domain from a Plasmodium falciparum var2csa-encoded PfEMP1 protein. *Acta crystallographica. Section F, Structural biology and crystallization communications* **64**, 221–3 (2008).

283. Rodríguez, L. E. *et al.* Identifying Plasmodium falciparum merozoite surface antigen 3 (MSP3) protein peptides that bind specifically to erythrocytes and inhibit merozoite invasion. *Protein science : a publication of the Protein Society* **14**, 1778–86 (2005).

284. Polley, S. D. *et al.* Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *The Journal of infectious diseases* **195**, 279–87 (2007).

285. Roussilhon, C. *et al.* Long-term clinical protection from falciparum malaria is strongly associated with IgG3 antibodies to merozoite surface protein 3. *PLoS medicine* **4**, e320 (2007).

286. Burgess, B. R., Schuck, P. & Garboczi, D. N. Dissection of merozoite surface protein 3, a representative of a family of Plasmodium falciparum surface proteins, reveals an oligomeric and highly elongated molecule. *The Journal of biological chemistry* **280**, 37236–45 (2005).

287. Mulhern, T. D. *et al.* Solution structure of a polypeptide containing four heptad repeat units from a merozoite surface antigen of Plasmodium falciparum. *Biochemistry* **34**, 3479–91 (1995).

288. McColl, D. J. & Anders, R. F. Conservation of structural motifs and antigenic diversity in the Plasmodium falciparum merozoite surface protein-3 (MSP-3). *Molecular and biochemical parasitology* **90**, 21–31 (1997).

289. Huber, W. *et al.* Limited sequence polymorphism in the Plasmodium falciparum merozoite surface protein 3. *Molecular and biochemical parasitology* **87**, 231–4 (1997).

290. Oeuvray, C. *et al.* Merozoite surface protein-3: a malaria protein inducing antibodies that promote Plasmodium falciparum killing by cooperation with blood monocytes. *Blood* **84**, 1594–602 (1994).

291. Singh, S. *et al.* Identification of a conserved region of Plasmodium falciparum MSP3 targeted by biologically active antibodies to improve vaccine design. *The Journal of infectious diseases* **190**, 1010–8 (2004).

292. Demanga, C. G. *et al.* Toward the rational design of a malaria vaccine construct using the MSP3 family as an example: contribution of antigenicity studies in humans. *Infection and immunity* **78**, 486–94 (2010).

293. Hisaeda, H. *et al.* Merozoite surface protein 3 and protection against malaria in Aotus nancymai monkeys. *The Journal of infectious diseases* **185**, 657–64 (2002).

294. Carvalho, L. J. M. *et al.* Immunization of Saimiri sciureus monkeys with Plasmodium falciparum merozoite surface protein-3 and glutamate-rich protein suggests that protection is related to antibody levels. *Scandinavian journal of immunology* **59**, 363–72 (2004).

295. Singh, S. *et al.* A conserved multi-gene family induces cross-reactive antibodies effective in defense against Plasmodium falciparum. *PloS one* **4**, e5410 (2009).

296. Wickramarachchi, T. *et al.* A novel Plasmodium falciparum erythrocyte binding protein associated with the merozoite surface, PfDBLMSP. *International Journal for Parasitology* **39**, 763–773 (2009).

297. Hodder, A. N. *et al.* Insights into duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from P. falciparum. *The Journal of biological chemistry* (2012).doi:10.1074/jbc.M112.350504

298. Sakamoto, H. *et al.* Antibodies against a Plasmodium falciparum antigen PfMSPDBL1 inhibit merozoite invasion into human erythrocytes. *Vaccine* **30**, 1972–80 (2012).

299. Maxwell, K. L., Bona, D., Liu, C., Arrowsmith, C. H. & Edwards, A. M. Refolding out of guanidine hydrochloride is an effective approach for high-throughput structural

studies of small proteins. *Protein science : a publication of the Protein Society* **12**, 2073–80 (2003).

300. Hagihara, Y., Aimoto, S., Fink, A. L. & Goto, Y. Guanidine hydrochloride-induced folding of proteins. *Journal of molecular biology* **231**, 180–4 (1993).

Appendix

A. Sequences of all constructs cloned into pET15bcon for bacterial expression

MC179

Protein sequence

EDKIMSYNAFFWMWVHDMLIDSIKWRDEHGRCINKDKGKTCIKGCNKKCICFQKW
VEQKKTEWGKIKDHFRKQKDIPKDWTHDDFLQTLLMKDLLLEIIQDTYGDANEIKRI
EALLEQAGVGGIDFAALAGLYTKGFVAEKDTTIDKLLQHEQKEADKCLKTHTDDTC
PPQEDRSVARS

DNA sequence

GAAGATAAAATTATGAGCTATAACGCGTTTTTTTGGATGTGGGTGCATGATATGC
TGATTGATAGCATTAAATGGCGCGATGAACATGGCCGCTGCATTAACAAAGATA
AAGGCAAAACCTGCATTAAAGGCTGCAACAAAAAATGCATTTGCTTTCAGAAAT
GGGTGGAACAGAAAAAAACCGAATGGGGCAAAATTAAAGATCATTTTCGCAAA
CAGAAAGATATTCCGAAAGATTGGACCCATGATGATTTTCTGCAGACCCTGCTG
ATGAAAGATCTGCTGCTGGAAATTATTCAGGATACCTATGGCGATGCGAACGAA
ATTAAACGCATTGAAGCGCTGCTGGAACAGGCGGGCGTGGGCGGCATTGATTTT
GCGGCGCTGGCGGGCCTGTATACCAAAGGCTTTGTGGCGGAAAAAGATACCACC
ATTGATAAACTGCTGCAGCATGAACAGAAAGAAGCGGATAAATGCCTGAAAAC
CCATACCGATGATACCTGCCCGCCGCAGGAAGATCGCAGCGTGGCGCGCAGC

PF11CIDR

Protein sequence

QKTFNPFFYYWVVHMLKDSIHWRTKKLDKCINNSNESKACKNNKKCNKECECFQR
WITQKQQEWEQIKKHFRKQKNIDNVGGFFKLSHDDVLQQVLDKDLLLKSLQEAYG
KPEDIKHIKDLLEKEAAAVAVVLGGKDNTTIDKLLNHEDKDAKDCLKKQEECKEQ

DNA sequence

CAAAAGACATTCAATCCTTTTTTTTACTATTGGGTGGTACATATGTTAAAAGATT
CCATACACTGGAGAACAAAAAAACTTGATAAGTGTATAAATAACAGTAATGAAT
CAAAAGCATGTAAAAACAATAAGAAATGTAATAAAGAATGTGAATGCTTCCAA
AGATGGATTACACAAAAACAACAAGAATGGGAACAAATAAAAAAGCATTTTCG
CAAACAAAAAAATATTGACAATGTAGGAGGTTTTTTCAAACTTAGCCACGATGA
TGTTCTTCAACAAGTTTTGGACAAAGATCTACTTTTGAAAAGTCTTCAAGAGGCT
TATGGGAAACCAGAGGACATAAAACACATTAAGGATCTGTTGGAAAAGGAAGC
AGCAGCAGTAGCAGTTGTTCTTGGTGGCAAGGACAATACCACAATTGATAAATT
GCTCAATCACGAAGATAAAGATGCCAAAGACTGCCTAAAAAAACAGGAAGAAT
GCAAAGAACAA

Primers used for cloning

Forward : 5′ – AAGGATCCCAAAAGACATTCAATCCTTTTTTTTAC – 3′

Reverse: 5′ – TTGCTAGCCTATTGTTCTTTGCATTCTTCCTG – 3′

PF11CIDR$_{TRUN}$

Protein sequence

QKTFNPFFYYWVVHMLKDSIHWRTKKLDKCINNSNESKACKNNKKCNKECECFQR
WITQKQQEWEQIKKHFRKQKNIDNVGGFFKLSHDDVLQQVLDKDLLLKSLQEAYG
KPEDIKHIKDLLEKEAAAVAVVLGGKDNTTIDKLLNHEDKDAKDCLKKQE

DNA sequence

CAAAAGACATTCAATCCTTTTTTTTACTATTGGGTGGTACATATGTTAAAAGATT
CCATACACTGGAGAACAAAAAAACTTGATAAGTGTATAAATAACAGTAATGAAT
CAAAAGCATGTAAAAACAATAAGAAATGTAATAAAGAATGTGAATGCTTCCAA
AGATGGATTACACAAAAACAACAAGAATGGGAACAAATAAAAAAGCATTTTCG
CAAACAAAAAAATATTGACAATGTAGGAGGTTTTTTCAAACTTAGCCACGATGA
TGTTCTTCAACAAGTTTTGGACAAAGATCTACTTTTGAAAAGTCTTCAAGAGGCT
TATGGGAAACCAGAGGACATAAAACACATTAAGGATCTGTTGGAAAAGGAAGC
AGCAGCAGTAGCAGTTGTTCTTGGTGGCAAGGACAATACCACAATTGATAAATT
GCTCAATCACGAAGATAAAGATGCCAAAGACTGCCTAAAAAAACAGGAA

Primers used for cloning

Forward : 5′ – AAGGATCCCAAAAGACATTCAATCCTTTTTTTTAC – 3′

Reverse: 5′ – TTGCTAGCCTATTCCTGTTTTTTTAGGCA – 3′

PF11CIDR$_{EXT}$

Protein sequence

KESNGGGGNTKWKRKEDMDKCPPINLYKPIDDKGGITINFLYSGDEPTEIGKKLNAF
CLTHSGNSVVSSSSARGNGASRDKNGGSDSQDLYQKWTCYEIDELEKDEKEDGVD
DPQYEKDVKTGGGLCILEKTNGKEKVNKQKTSHEMQKTFNPFFYYWVVHMLKDSI
HWRTKKLDKCINNSNESKACKNNKKCNKECECFQRWITQKQQEWEQIKKHFRKQK
NIDNVGGFFKLSHDDVLQQVLDKDLLLKSLQEAYGKPEDIKHIKDLLEKEAAAVAV
VLGGKDNTTIDKLLNHEDKDAKDCLKKQE

DNA sequence

AAGGAAAGTAATGGTGGTGGTGGTAATACCAAATGGAAAAGGAAAGAGGATAT
GGATAAATGTCCGCCTATAAACCTTTATAAGCCTATAGATGACAAAGGTGGTAT
TACTATTAATTTCCTGTACAGTGGTGATGAACCAACAGAAATTGGAAAAAAATT

AAACGCATTTTGCTTAACACACAGTGGTAATAGTGTTGTTAGTAGTAGTAGTGCT
CGTGGTAACGGTGCTAGTCGTGATAAAAATGGTGGTAGTGATAGTCAGGATTTG
TATCAAAAATGGACATGTTATGAAATTGACGAGCTTGAGAAAGATGAAAAAGA
AGACGGTGTGGATGACCCACAGTATGAAAAGGATGTAAAAACTGGAGGCGGAT
TATGTATATTGGAAAAGACGAACGGCAAAGAAAAAGTGAACAAACAAAAGACA
TCTCATGAAATGCAAAAGACATTCAATCCTTTTTTTTACTATTGGGTGGTACATA
TGTTAAAAGATTCCATACACTGGAGAACAAAAAAACTTGATAAGTGTATAAATA
ACAGTAATGAATCAAAAGCATGTAAAAACAATAAGAAATGTAATAAAGAATGT
GAATGCTTCCAAAGATGGATTACACAAAAACAACAAGAATGGGAACAAATAAA
AAAGCATTTTCGCAAACAAAAAAATATTGACAATGTAGGAGGTTTTTTCAAACT
TAGCCACGATGATGTTCTTCAACAAGTTTTGGACAAAGATCTACTTTTGAAAAGT
CTTCAAGAGGCTTATGGGAAACCAGAGGACATAAAACACATTAAGGATCTGTTG
GAAAAGGAAGCAGCAGCAGTAGCAGTTGTTCTTGGTGGCAAGGACAATACCAC
AATTGATAAATTGCTCAATCACGAAGATAAAGATGCCAAAGACTGCCTAAAAAA
ACAGGAA

Primers used for cloning

Forward : 5′ – AAGGATCCAAGGAAAGTAATGGTGGTGGTGGTAATACC – 3′

Reverse: 5′ – TTGCTAGCCTATTCCTGTTTTTTTAGGCA – 3′

PF11CIDR$_{NTS}$

Protein sequence

MEPHGGSGGGGDVIDHQSAKHLLDSIGKIVHDQVEKEAADYRSSLQGRLKDATYSK
KPDNQEKPSDPCELKHEYRTNATNGTSYPCRAGKEERFSDTLGGQCTDQQIEGNDR
NNGGACAPYRRLHLCNKNMVKMDTNNDDSSKAKHNLLLDVCMAAKYEGYSINTH
YPKHQEKYKDTGTASQLCTVLARSFADIGDIVRGRDLFRGNKKKSENKREKEKLEE
NFKKYFQQIHEDVTSTSGKNRKALRDRYHKDGPDYYQLREDWWNNNRKMVWIA
MTCEAGGSQYFRRTCSNGISETNNKCRCTIETVPTYFDYVPQYLRWFEEWAEDFCR
KKKHKLKDAIDKCRGQYQDADRYCDLNGYDCEKTKRGINMYRWDHKCTGCFLSC
SHFRTWIDNQKEQFHKQKKIYDKEITRGGSDHKGYEKKFYGKLKGTDYKDVGKFL
EKLSDEDVCTKFREDEGIIDFKNVNSSSGSGGGEGSNKTFSHTEYCQACPLCGVEKE
SNGGGGNTKWKRKEDMDKCPPINLYKPIDDKGGITINFLYSGDEPTEIGKKLNAFCL
THSGNSVVSSSSARGNGASRDKNGGSDSQDLYQKWTCYEIDELEKDEKEDGVDDP
QYEKDVKTGGGLCILEKTNGKEKVNKQKTSHEMQKTFNPFFYYWVVHMLKDSIH
WRTKKLDKCINNSNESKACKNNKKCNKECECFQRWITQKQQEWEQIKKHFRKQKN
IDNVGGFFKLSHDDVLQQVLDKDLLLKSLQEAYGKPEDIKHIKDLLEKEAAAVAVV
LGGKDNTTIDKLLNHEDKDAKDCLKKQE

DNA sequence

ATGGAGCCGCATGGAGGTAGCGGTGGTGGCGGGGATGTTATTGACCACCAAAGT
GCCAAACATTTATTGGATAGCATAGGGAAAATAGTGCACGACCAAGTGGAAAA
GGAAGCTGCAGATTATAGAAGTTCTTTACAAGGACGTTTGAAAGACGCAACATA

TTCAAAAAAACCAGATAATCAAGAAAAACCTTCTGATCCATGCGAACTTAAACA
TGAATATCGTACTAATGCTACTAACGGTACAAGTTATCCATGTAGAGCTGGAAA
AGAAGAACGTTTTTCCGATACACTTGGTGGCCAATGCACTGATCAACAAATAGA
AGGTAATGATCGTAATAATGGTGGCGCCTGTGCGCCATATAGACGATTACATTT
ATGCAACAAAAATATGGTAAAAATGGACACAAATAATGATGATAGTAGTAAAG
CTAAACATAATTTATTGTTAGACGTGTGTATGGCAGCCAAATATGAAGGGTACT
CAATAAATACACATTATCCAAAACATCAAGAAAAATATAAGGATACTGGTACTG
CTTCCCAATTATGTACTGTATTAGCACGCAGTTTTGCAGATATTGGAGATATTGT
AAGAGGAAGAGATCTATTTCGTGGTAATAAAAAAAAAAGTGAAAATAAAAGAG
AGAAAGAAAAATTAGAAGAAAATTTTAAAAAATATTTCCAGCAAATACATGAG
GACGTGACGTCGACGAGCGGGAAGAATAGGAAGGCGCTAAGAGATCGCTACCA
TAAGGACGGTCCAGATTATTATCAATTAAGAGAAGATTGGTGGAATAATAATAG
AAAAATGGTATGGATTGCTATGACGTGCGAAGCTGGGGGTTCTCAATATTTTCG
ACGAACATGTTCTAATGGAATAAGTGAAACTAATAACAAATGCCGATGCACTAT
TGAAACTGTTCCCACATATTTTGACTACGTGCCACAATATTTGAGTGGTTCGAGG
AATGGGCAGAAGATTTTTGTAGGAAAAAAAAACACAAATTAAAGGATGCTATA
GACAAATGTCGTGGCCAGTACCAGGATGCGGATCGATATTGTGATCTTAATGGC
TACGATTGCGAAAAAACCAAAAGAGGAATAAACATGTATCGTTGGGATCACAA
ATGTACTGGCTGCTTCCTTTCTTGCTCTCATTTTAGAACATGGATAGATAACCAA
AAAGAACAATTTCACAAACAAAAAAAAATATATGATAAAGAAATAACACGTGG
TGGTAGTGATCATAAAGGATATGAAAAAAAATTTTATGGAAAACTGAAAGGAA
CTGACTATAAAGATGTTGGTAAATTTTTGGAAAAATTAAGTGATGAAGATGTAT
GCACAAAGTTTAGAGAAGACGAAGGAATAATTGATTTCAAAAACGTTAATAGTA
GTAGTGGTAGTGGTGGTGGTGAGGGTAGTAATAAAACATTTTCTCATACGGAAT
ATTGCCAAGCATGTCCTCTATGTGGGGTCGAAAAGGAAAGTAATGGTGGTGGTG
GTAATACCAAATGGAAAAGGAAAGAGGATATGGATAAATGTCCGCCTATAAAC
CTTTATAAGCCTATAGATGACAAAGGTGGTATTACTATTAATTTCCTGTACAGTG
GTGATGAACCAACAGAAATTGGAAAAAAATTAAACGCATTTTGCTTAACACACA
GTGGTAATAGTGTTGTTAGTAGTAGTAGTGCTCGTGGTAACGGTGCTAGTCGTGA
TAAAAATGGTGGTAGTGATAGTCAGGATTTGTATCAAAAATGGACATGTTATGA
AATTGACGAGCTTGAGAAAGATGAAAAAGAAGACGGTGTGGATGACCCACAGT
ATGAAAAGGATGTAAAAACTGGAGGCGGATTATGTATATTGGAAAAGACGAAC
GGCAAAGAAAAAGTGAACAAACAAAAGACATCTCATGAAATGCAAAAGACATT
CAATCCTTTTTTTTACTATTGGGTGGTACATATGTTAAAAGATTCCATACACTGG
AGAACAAAAAAACTTGATAAGTGTATAAATAACAGTAATGAATCAAAAGCATG
TAAAAACAATAAGAAATGTAATAAAGAATGTGAATGCTTCCAAAGATGGATTAC
ACAAAAACAACAAGAATGGGAACAAATAAAAAAGCATTTTCGCAAACAAAAAA
ATATTGACAATGTAGGAGGTTTTTTCAAACTTAGCCACGATGATGTTCTTCAACA
AGTTTTGGACAAAGATCTACTTTTGAAAAGTCTTCAAGAGGCTTATGGGAAACC
AGAGGACATAAAACACATTAAGGATCTGTTGGAAAAGGAAGCAGCAGCAGTAG
CAGTTGTTCTTGGTGGCAAGGACAATACCACAATTGATAAATTGCTCAATCACG
AAGATAAAGATGCCAAAGACTGCCTAAAAAAACAGGAA

Primers used for cloning

Forward : 5′ – AAGGATCCATGGAGCCGCATGGAGGTAG – 3′

Reverse: 5′ – TTGCTAGCCTATTCCTGTTTTTTTAGGCA – 3′

MSPDBL2

Protein sequence
QCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFE KFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYL EKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDD IPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLK KYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCT
DNA sequence
CAGTGCCGGAAGAACGTGCACGAGCGGAACAACTGGCTGGGCAGCAGCGTGAA GAACTTCGCCAGCGACAACAAGGGCGTGCTGGTGCCCCCCAGACGGCAGAGCCT GTGCCTGAGAATCACCCTGCAGGACTTCCGGACCAAGAAGAAGAAAGAGGGCG ACTTCGAGAAGTTCATCTACAGCTACGCCAGCAGCGAGGCCCGGAAGCTGCGGA CCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCCATCCGGTACAGCTTCG CCGACATCGGCAACATCATCCGGGGCGACGACATGATGGACACCCCCACCAGCA AAGAGACAATCACCTATCTGGAAAAGGTGCTGAAGATCTACAATGAGAACAAC GACAAGCCCAAGGACGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTG GGAGGCCATGATGTGCGGCTACCAGAGCGCCCAGAAAGACAACCAGTGCACCG GCTACGGCAACATCGACGATATCCCCCAGTTCCTGCGGTGGTTCCGCGAGTGGG GCACCTACGTGTGCGAAGAGTCCGAGAAGAACATGAATACCCTGAAGGCCGTGT GCTTCCCCAAGCAGCCCAGAACCGAGGCCAACCCTGCCCTGACCGTGCACGAGA ACGAGATGTGCAGCAGCACCCTGAAGAAATACGAGGAATGGTACAACAAGCGC AAGACCGAGTGGACCGAGCAGAGCATCAAGTATAACAACGATAAGATCAACTA CGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCGAGAAGTGCCCCGA GTGCAAGTGCACC
Primers used for cloning
Forward : 5′ – AAGGATCCCAGTGCCGGAAGAACGTGCACG – 3′ Reverse: 5′ – TTGCTAGCCTAGGTGCACTTGCACTCGGGGCA – 3′

B. Sequence of recombinant hCD36ED in pAcGP67A for baculovirus insect cell expression

Protein sequence
GDLLIQKTIKKQVVLEEGTIAFKNWVKTGTEVYRQFWIFDVQNPQEVMMNSSNIQV KQRGPYTYRVRFLAKENVTQDAEDNTVSFLQPNGAIFEPSLSVGTEADNFTVLNLA VAAASHIYQNQFVQMILNSLINKSKSSMFQVRTLRELLWGYRDPFLSLVPYPVTTTV GLFYPYNNTADGVYKVFNGKDNISKVAIIDTYKGKRNLSYWESHCDMINGTDAASF PPFVEKSQVLQFFSSDICRSIYAVFESDVNLKGIPVYRFVLPSKAFASPVENPDNYCFC

TEKIISKNCTSYGVLDISKCKEGRPVYISLPHFLYASPDVSEPIDGLNPNEEEHRTYLDI
EPITGFTLQFAKRLQVNLLVKPSEKIQVLKNLKRNYIVPILWLNETGTIGDEKANMFR
SQVTGKIN

DNA sequence

GGAGACCTGCTTATCCAGAAGACAATTAAAAAGCAAGTTGTCCTCGAAGAAGGT
ACAATTGCTTTTAAAAATTGGGTTAAAACAGGCACAGAAGTTTACAGACAGTTT
TGGATCTTTGATGTGCAAAATCCACAGGAAGTGATGATGAACAGCAGCAACATT
CAAGTTAAGCAAAGAGGTCCTTATACGTACAGAGTTCGTTTTCTAGCCAAGGAA
AATGTAACCCAGGACGCTGAGGACAACACAGTCTCTTTCCTGCAGCCCAATGGT
GCCATCTTCGAACCTTCACTATCAGTTGGAACAGAGGCTGACAACTTCACAGTTC
TCAATCTGGCTGTGGCAGCTGCATCCCATATCTATCAAAATCAATTTGTTCAAAT
GATCCTCAATTCACTTATTAACAAGTCAAAATCTTCTATGTTCCAAGTCAGAACT
TTGAGAGAACTGTTATGGGGCTATAGGGATCCATTTTTGAGTTTGGTTCCATACC
CTGTTACTACCACAGTTGGTCTGTTTTATCCTTACAACAATACTGCAGATGGAGT
TTATAAAGTTTTCAATGGAAAAGATAACATAAGTAAAGTTGCCATAATCGACAC
ATATAAAGGTAAAAGGAATCTGTCCTATTGGGAAAGTCACTGCGACATGATTAA
TGGTACAGATGCAGCCTCATTTCCACCTTTTGTTGAGAAAAGCCAGGTATTGCAG
TTCTTTTCTTCTGATATTTGCAGGTCAATCTATGCTGTATTTGAATCCGACGTTAA
TCTGAAAGGAATCCCTGTGTATAGATTTGTTCTTCCATCCAAGGCCTTTGCCTCT
CCAGTTGAAAACCCAGACAACTATTGTTTCTGCACAGAAAAAATTATCTCAAAA
AATTGTACATCATATGGTGTGCTAGACATCAGCAAATGCAAAGAAGGGAGACCT
GTGTACATTTCACTTCCTCATTTTCTGTATGCAAGTCCTGATGTTTCAGAACCTAT
TGATGGATTAAACCCAAATGAAGAAGAACATAGGACATACTTGGATATTGAACC
TATAACTGGATTCACTTTACAATTTGCAAAACGGCTGCAGGTCAACCTATTGGTC
AAGCCATCAGAAAAAATTCAAGTATTAAAGAATCTGAAGAGGAACTATATTGTG
CCTATTCTTTGGCTTAATGAGACTGGGACCATTGGTGATGAGAAGGCAAACATG
TTCAGAAGTCAAGTAACTGGAAAAATAAAC

C. Sequence of recombinant hIgM-Fc in pFUSE vector of mammalian cell expression

Protein sequence

RSIAELPPKVSVFVPPRDGFFGNPRKSKLICQATGFSPRQIQVSWLREGKQVGSGVTT
DQVQAEAKESGPTTYKVTSTLTIKESDWLSQSMFTCRVDHRGLTFQQNASSMCVPQ
DTAIRVFAIPPSFASIFLTKSTKLTCLVTDLTTYDSVTISWTRQNGEAVKTHTNISESH
PNATFSAVGEASICEDDWNSGERFTCTVTHTDLPSPLKQTISRPKGVALHRPDVYLL
PPAREQLNLRESATITCLVTGFSPADVFVQWMQRGQPLSPEKYVTSAPMPEPQAPGR
YFAHSILTVSEEEWNTGETYTCVVAHEALPNRVTERTVDKSTGKPTLYNVSLVMSD
TAGTCY

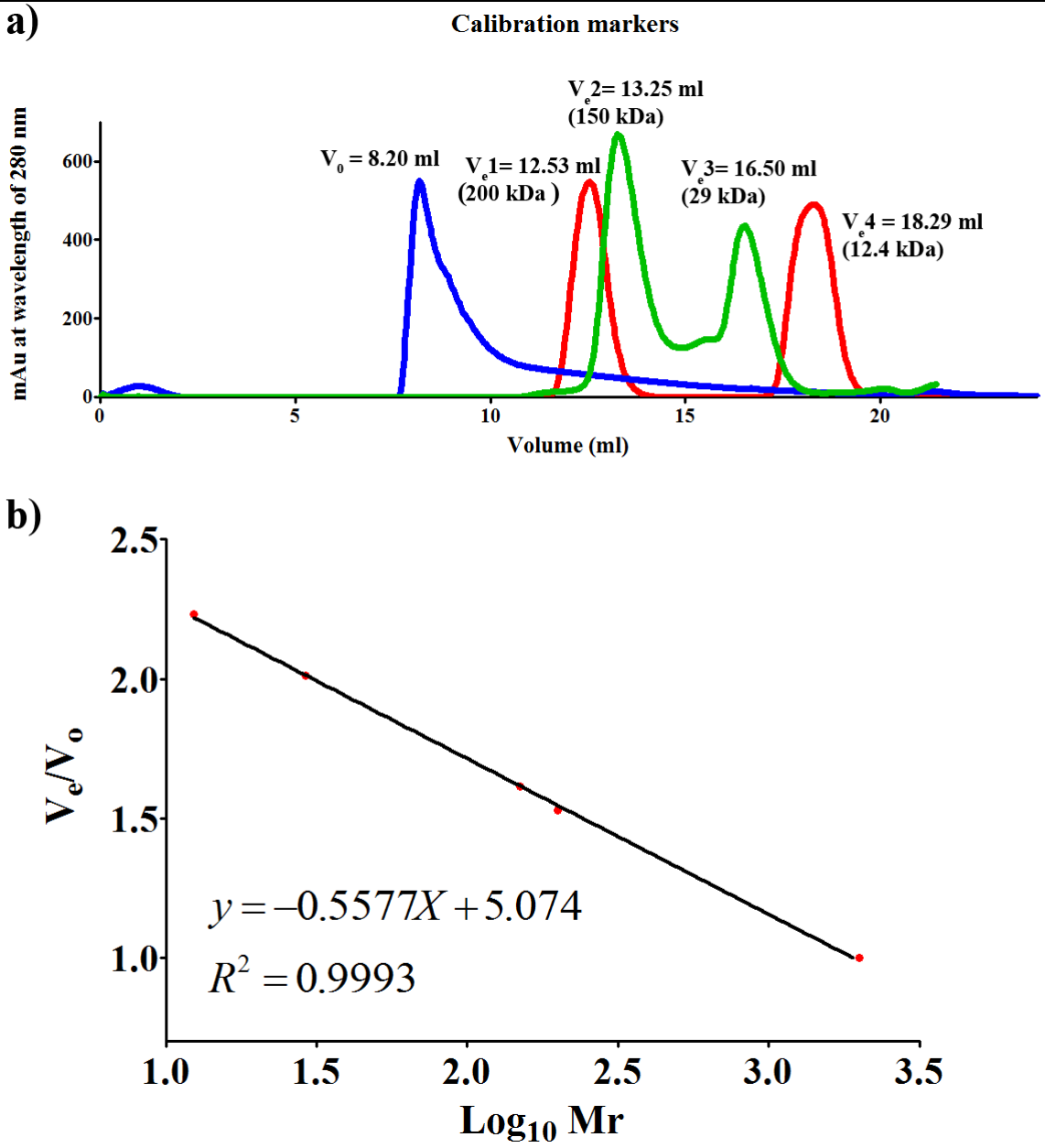
DNA sequence

ATTGCCGAGCTGCCTCCCAAAGTGAGCGTCTTCGTCCCACCCCGCGACGGCTTCT
TCGGCAACCCCCGCAAGTCCAAGCTCATCTGCCAGGCCACGGGTTTCAGTCCCC

GGCAGATTCAGGTGTCCTGGCTGCGCGAGGGGAAGCAGGTGGGGTCTGGCGTCA
CCACGGACCAGGTGCAGGCTGAGGCCAAAGAGTCTGGGCCCACGACCTACAAG
GTGACCAGCACACTGACCATCAAAGAGAGCGACTGGCTCAGCCAGAGCATGTTC
ACCTGCCGCGTGGATCACAGGGGCCTGACCTTCCAGCAGAATGCGTCCTCCATG
TGTGTCCCCGGTGAGTGACCTGTCCCCAGGGGCAGCACCCACCGACACACAGGG
GTCCACTCGGGTCTGGCATTCGCCACCCCGGATGCAGCCATCTACTCCCTGAGCC
TTGGCTTCCCAGAGCGGCCAAGGGCAGGGGCTCGGGCGGCAGGACCCCTGGGCT
CGGCAGAGGCAGTTGCTACTCTTTGGGTGGGAACCATGCCTCCGCCCACATCCA
CACCTGCCCCACCTCTGACTCCCTTCTCTTGACTCCAGATCAAGACACAGCCATC
CGGGTCTTCGCCATCCCCCCATCCTTTGCCAGCATCTTCCTCACCAAGTCCACCA
AGTTGACCTGCCTGGTCACAGACCTGACCACCTATGACAGCGTGACCATCTCCT
GGACCCGCCAGAATGGCGAAGCTGTGAAAACCCACACCAACATCTCCGAGAGC
CACCCCAATGCCACTTTCAGCGCCGTGGGTGAGGCCAGCATCTGCGAGGATGAC
TGGAATTCCGGGGAGAGGTTCACGTGCACCGTGAC

CCACACAGACCTGCCCTCGCCACTGAAGCAGACCATCTCCCGGCCCAAGGGTAG
GCCCCACTCTTGCCCCTCTTCCTGCACTCCCTGGGACCTCCCTTGGCCTCTGGGG
CATGGTGGAAAGCACCCCTCACTCCCCCGTTGTCTGGGCAACTGGGGAAAAGGG
GACTCAACCCCAGCCCACAGGCTGGTCCCCCCACTGCCCCGCCCTCACCACCAT
CTCTGTTCACAGGGGTGGCCCTGCACAGGCCCGATGTCTACTTGCTGCCACCAGC
CCGGGAGCAGCTGAACCTGCGGGAGTCGGCCACCATCACGTGCCTGGTGACGGG
CTTCTCTCCCGCGGACGTCTTCGTGCAGTGGATGCAGAGGGGGCAGCCCTTGTCC
CCGGAGAAGTATGTGACCAGCGCCCCAATGCCTGAGCCCCAGGCCCCAGGCCGG
TACTTCGCCCACAGCATCCTGACCGTGTCCGAAGAGGAATGGAACACGGGGGAG
ACCTACACCTGCGTGGTGGCCCATGAGGCCCTGCCCAACAGGGTCACCGAGAGG
ACCGTGGACAAGTCCACCGGTAAACCCACCCTGTACAACGTGTCCCTGGTCATG
TCCGACACAGCTGGCACCTGCTAC

D. Calibration curve of Superdex 200 10/300 GL (GE Healthcare)



Appendix D. Calibration curve for analytical Superdex 200 10/300 GL (GE Healthcare). **(a)** Calibration markers (Sigma-aldrich) were run down the Superdex 200 10/300 GL at a speed of 0.25 ml/min. Elution volume (V_e) and void volume (V_0) were determined for each size. **(b)** A calibration curve of (Ve/V_0) against Log_{10}(Molecular Mass). Hence, assuming the protein has a perfectly spherical hydrodynamic volume, the molecular mass of a protein will be given by

$$Mr = 10^{\left(\frac{\left[\frac{V_e}{8.2}\right]-5.074}{-0.5577}\right)}$$

Where V_e is the elution volume of the protein in question.