University of Oxford

Thesis submitted for the degree

Doctor of Philosophy in Biochemistry

Structural studies of the PIR and RIFIN families from malaria-infected erythrocytes



Thomas Edward Harrison

Merton College

Professor Matthew Higgins

Abstract

Malaria, caused by the *Plasmodium* parasite, is responsible for the deaths of close to half a million people a year. During the course of infection, the parasite invades and remodels the host red blood cell, often requiring the export of a number of proteins to the erythrocyte surface. These include the RIFIN family of *Plasmodium falciparum*, which have been shown to directly modulate the host immune response through interactions with inhibitory immune receptors such as LILRB1 and LAIR1. Additionally, the PIR family of a number of human- and rodent-infective *Plasmodium* species have been implicated in the determination of parasite virulence and the outcome of infection, along with cytoadhesion of infected erythrocytes to the vasculature. Here, I present the structure of the variable region of a RIFIN in complex with LILRB1. This is the first structure of any member of the RIFIN family and reveals that the RIFIN utilises a conserved mode of binding, which can be blocked through a single mutation in all RIFINs tested, and which mimics the interaction of MHC class I, the native ligand of LILRB1. The RIFIN is also able to mimic the effect of MHC binding both on signalling through LILRB1, and on inhibition of NK cell activation at the immunological synapse. I also present the structure of the extracellular domain of a PIR protein from *Plasmodium chabaudi*, the first for a member of this family, revealing a novel α-helical fold. Sequence analysis and homology modelling show conservation of this fold across PIRs from 4 human- and 3 rodent-infective *Plasmodium* species, as well as excluding the RIFINs and STEVORs from membership of the PIR family. These findings provide both a molecular mechanism for RIFIN function, along with a structural definition of the PIRs, giving valuable insight into 2 major families of *Plasmodium* variant surface antigens.

Acknowledgements

I would first like to thank my supervisor, Matt Higgins, for his guidance, support, and advice across my time in the lab. His help from harvesting crystals and planning experiments to writing papers has proved invaluable, and I am very grateful for everything he has contributed over this D.Phil. I also thank the other members of the lab, who have made my time there so enjoyable, while also providing constant advice throughout my D.Phil, with a particular mention for Ivan, Kate, and Natalie for their help and support over the last few years.

I would also like to thank our many collaborators who have played an instrumental part in these projects. Members of Hisashi Arase's group (Department of Immunochemistry, University of Osaka), particularly Akihito Sakoguchi, provided the KIR-binding RIFIN sequences, and carried out both the LILRB1-signalling reporter assay and transgenic parasite experiments for the RIFIN project. Alex Morch and James Felce of Mike Dustin's group (Kennedy Institute of Rheumatology, University of Oxford) carried out the microscopy for the same project, which included the use of protein produced by Edward Jenkins of Simon Davis's group (MRC Weatherall Institute of Molecular Medicine, University of Oxford). Members of Jean Langhorne's group (Francis Crick Institute) provided the constructs used for expression of the CIRs, and Adam Reid (Wellcome Sanger Institute) provided valuable advice and assistance in carrying out the bioinformatic analysis for both the RIFIN and CIR projects. Additionally I would like to thank Ed Lowe for his help with crystallography, David Staunton for help with biophysics experiments, and Elspeth Garman for her comments and suggestions on improving this thesis.

Finally, I would like to thank my friends and family who have been there for me throughout, and whose encouragement and support has made this possible.

Contents

Abstract i

Acknowledgements ii

Contents iii

List of Figures x

List of Tables xv

Abbreviations xvii

1 Introduction 1

1.1 Malaria . . . 1

1.1.1 History . . . 1

1.1.2 Global Disease Burden . . . 2

1.1.3 The *Plasmodium* Parasite . . . 3

1.1.4 The *Plasmodium* Life Cycle . . . 4

1.1.4.1 The Exo-Erythrocytic Stage . . . 4

1.1.4.2 The Erythrocytic Stage . . . 5

1.1.4.3 Gametocytogenesis and Sporogony . . . 6

1.1.5 Disease . . . 7

1.1.6 Disease Prevention and Treatment . . . 8

1.1.6.1 Vector Control . . . 8

1.1.6.2 Antimalarial Drugs . . . 9

1.1.6.3 Vaccine Development . . . 10

1.2 Variant Surface Antigens . . . 12

1.2.1 Overview . . . 12

1.2.2 Variant Surface Antigens of *Plasmodium falciparum* . . . 14

1.2.2.1 Expression 14
1.2.2.2 Transport 15
1.2.2.3 PfEMP1 17
1.2.2.4 STEVORs 19
1.3 RIFINs 20
1.3.1 Immunogenenicity 22
1.3.2 Interactions 23
1.4 PIRs 24
1.4.1 PIR Function 27
1.5 Inhibitory Immune Receptors 29
1.5.1 LILR Family 30
1.5.1.1 LILRB1 31
1.5.2 LAIR1 33
1.5.3 KIR Family 34
1.6 Project Aims 36

2 Materials and Methods **38**

2.1 Cloning 38
2.1.1 Vectors 38
2.1.2 Polymerase Chain Reaction 38
2.1.3 Restriction Digestion 40
2.1.4 Agarose Gel Electrophoresis 40
2.1.5 Ligation 41
2.1.6 Transformation and Sequencing 41
2.1.7 Site Directed Mutagenesis 42
2.2 Protein Expression and Purification 42
2.2.1 *E. coli* Expression 42
2.2.1.1 Expression Trials 42

2.2.1.2 Large Scale Expression 43
2.2.1.3 Sonication 43
2.2.1.4 Nickel-NTA Affinity Purification 44
2.2.1.5 Size Exclusion Chromatography 44
2.2.1.6 Insoluble Protein Purification and Refolding 44
2.2.1.7 Tangential Flow Filtration 45
2.2.2 HEK293F Cell Expression 46
2.2.2.1 Cell Maintenance 46
2.2.2.2 Transfection and Protein Purification 46
2.2.2.3 Selenomethionine Incorporation 47
2.2.3 Expi293F Cell Expression 47
2.2.3.1 Cell Maintenance 48
2.2.3.2 Transfection 48
2.2.4 S2 Cell Expression 48
2.2.4.1 Cell Maintenance 48
2.2.4.2 Transfection 49
2.2.5 Polyacrylamide Gel Electrophoresis 49
2.2.6 Thrombin Cleavage 50
2.2.7 Protein Deglycosylation 50
2.2.8 Protein Methylation 50
2.3 Biophysical Methods 51
2.3.1 Circular Dichroism 51
2.3.2 Surface Plasmon Resonance 51
2.4 X-Ray Crystallography 53
2.4.1 Crystallisation 53
2.4.2 Crystal Harvesting 54
2.4.3 Data Collection 54
2.4.4 Data Processing 55

2.4.5 Phase Determination 56
2.4.5.1 Molecular Replacement 56
2.4.5.2 Experimental Phasing 57
2.4.6 Model Building and Refinement 58
2.5 Bioinformatics 59

3 Production and Biophysical Characterisation of RIFINs 61

3.1 Summary 61
3.2 RIFIN Ectodomain Expression 62
3.2.1 Design of Ectodomain Constructs 62
3.2.2 Expression Trials of RIFIN Ectodomains in HEK Cells, *E. coli*, and S2 Cells 65
3.2.2.1 HEK293F Cell Trials 65
3.2.2.2 Expi293F Cell Trials 66
3.2.2.3 *E. coli* Trials 67
3.2.2.4 S2 Cell Trials 70
3.3 Characterisation of a LILRB1-Binding RIFIN Variable Region 71
3.3.1 Expression and Purification 72
3.3.1.1 Expression Trials in Expi293F Cells 72
3.3.1.2 Expression Trials in *E. coli* 72
3.3.2 Assessment of the Secondary Structure of the Variable Region 74
3.3.3 Expression and Purification of LILRB1 76
3.3.3.1 Expression in HEK293F Cells 76
3.3.3.2 Refolding of LILRB1 D1D2 from Insoluble Expression in *E. coli* 77
3.3.4 Binding of the Variable Region to LILRB1 79
3.4 Characterisation of a RIFIN Constant Region 81
3.4.1 Expression and Purification 82

3.4.1.1 Expression of the Constant Region of a LILRB1-Binding RIFIN 82
3.4.1.2 Assessment of the Secondary Structure of the Constant Region 84
3.4.1.3 Binding of the Constant Region to LILRB1 85
3.4.2 Expression of 4 RIFIN Constant Regions 85
3.4.3 Crystallisation Trials 87
3.5 Characterisation of a KIR-binding RIFIN Variable Region 88
3.5.1 Expression and Purification 89
3.5.2 Expression and Purification of KIR2DL1 and KIR2DS1 91
3.5.3 Binding of the Variable Region to KIR2DL1 and KIR2DS1 . . 92
3.5.4 Crystallisation Trials 94
3.6 Discussion 95

4 Structural Basis for RIFIN Binding to LILRB1 98

4.1 Summary 98
4.2 Crystallisation, Data Collection, and Structure Solution of the RIFIN Variable Region:LILRB1 Complex 99
4.2.1 Crystallisation 99
4.2.2 Data Collection and Processing 100
4.3 The Structure of the RIFIN Variable Region 106
4.3.1 Conservation of the Variable Region 107
4.4 Analysis of the Interaction Between the RIFIN and LILRB1 112
4.4.1 Confirmation of the Binding Interface 115
4.5 The LILRB1-Binding Mode of Other RIFINs 121
4.5.1 Expression and Purification of the Wild-type and Mutant Variable Regions 122

4.5.2 Effect of Predicted Binding-Site Mutations on the Interaction of RIFIN Variable Regions with LILRB1 124
4.5.2.1 Crystallisation of the Pf3D7_0100400 Variable Region:LILRB1 Complex 127
4.6 Comparison with the interaction between MHC Class I and LILRB1 . 129
4.6.1 Analysis of the NK Cell Immunological Synapse using TIRF Microscopy 133
4.6.1.1 Effect of RIFIN Binding on RIFIN and LILRB1 Accumulation at the Immunological Synapse 134
4.6.1.2 Effect of RIFIN Binding on Perforin Release 137
4.7 Discussion 139

5 The Structure of a CIR 143

5.1 Summary 143
5.2 Expression and Purification of CIRs 144
5.3 Assessment of the Fold of CIRs 147
5.4 CIR Crystallisation, Data Collection, and Data Processing 147
5.4.1 Crystallisation of the CIR, PCHAS_1200500 147
5.4.2 Data Collection and Processing 149
5.4.3 Solving the CIR Structure 149
5.4.3.1 Heavy Atom Derivatives 151
5.4.3.2 Selenomethionine Incorporation 152
5.4.3.3 Sulphur-SAD 153
5.5 The Structure of a CIR 157
5.6 Conservation Among CIRs 160
5.6.1 Sequence Conservation 160
5.6.2 Structure-Based Threading 163
5.7 Conservation Among Other Putative PIR Families 166

5.7.1 Is the CIR Protein Fold Conserved Across Rodent-Infective *Plasmodium* Species? . . . 167
5.7.2 Is the CIR Protein Fold Conserved Across Human-Infective *Plasmodium* Species? . . . 170
5.7.3 Is the CIR Protein Fold Conserved Across the RIFINs and STEVORs? . . . 175
5.7.4 Is the CIR Protein Fold Conserved Across the *Laverania*? . . . 177
5.7.5 Is the CIR Protein Fold Conserved Across Avian-Infective *Plasmodium* Species? . . . 177
5.8 Discussion . . . 178

6 Conclusions and Future Directions 181
6.1 Characterisation of the Interaction Between RIFINs and LILRB1 . . 182
6.1.1 The RIFIN Variable Region Structure . . . 182
6.1.2 The RIFIN:LILRB1 Interaction . . . 183
6.2 Characterisation of a PIR Protein . . . 185
6.2.1 The CIR Structure . . . 185
6.2.2 Conservation of the PIRs . . . 186
6.3 Summary of the Findings of this Thesis . . . 187

A Constructs and Primers 188

B Kinetic Values for KIR-Binding RIFINs 193

C CIR Modelling Data 194

D Publications Arising from this Thesis 198

Bibliography 199

List of Figures

1.1 Map Showing the Death Rate from Malaria 2

1.2 Life Cycle of the *Plasmodium* Parasite 5

1.3 Structures of PfEMP1 Domains Bound to 3 Receptors 19

1.4 Conserved Protein Architecture of the RIFINs 21

1.5 Sequence Alignment of Orthologous PIRs 27

1.6 A Structure of HLA-A2 Bound to LILRB1 32

1.7 A Structure of the Ectodomain of LAIR1 33

1.8 A Structure of HLA-Cw4 Bound to KIR2DL1 35

1.9 Comparison of LILRB1 and KIR2DL1 Binding to MHC Class I . . . 36

2.1 pET15b Vector Map 39

2.2 pHLsec Vector Map 39

2.3 pExpreS^2-1 Vector Map 40

3.1 Phylogenetic Tree for All Known LILRB1-Binding RIFINs 63

3.2 Sequence Alignment of LILRB1-Binding RIFINs 64

3.3 Expression Trial of RIFIN Ectodomain Constructs in HEK293F Cells 66

3.4 Expression Trial of RIFIN Ectodomain Constructs in Expi293F cells . 67

3.5 Expression Trial of RIFIN Ectodomain Constructs in *E. coli* 69

3.6 Expression Trial of RIFIN Ectodomain Constructs in S2 Cells 71

3.7 Expression Trial of a LILRB1-Binding RIFIN Variable Region in Expi293F Cells 73

3.8 Expression Trial of a LILRB1-Binding RIFIN Variable Region in *E. coli* 73

3.9 Refolding of a LILRB1-Binding RIFIN Variable Region from *E. coli* Inclusion Bodies 75

3.10 Circular Dichroism Spectrum for a LILRB1-Binding RIFIN Variable Region 75

3.11 Expression Trial of LILRB1 Ectodomain Constructs in HEK293F Cells 77
3.12 Purification of LILRB1 Full-Length and D3D4 Constructs from HEK293F Cells 78
3.13 Refolding of LILRB1 D1D2 from *E. coli* Inclusion Bodies 79
3.14 Analytical Gel Filtration Showing a RIFIN Variable Region Binding to LILRB1 80
3.15 Surface Plasmon Resonance Showing a RIFIN Variable Region Binding to 3 LILRB1 Constructs 81
3.16 Expression Trial of the Constant Region of a LILRB1-Binding RIFIN in HEK293F Cells 83
3.17 Purification of the Constant Region of a LILRB1-Binding RIFIN from HEK293F Cells 84
3.18 Circular Dichroism Spectrum for the Constant Region of a LILRB1-Binding RIFIN 84
3.19 Surface Plasmon Resonance for the Interaction Between a RIFIN Constant Region and LILRB1 85
3.20 Expression Trials for RIFIN Constant Region Constructs in HEK293F and Expi293F Cells 86
3.21 Purification of 3 RIFIN Constant Region Constructs from HEK293F Cells 87
3.22 Expression Trial of KIR-Binding RIFIN Constructs in HEK293F and Expi293F Cells 90
3.23 Purification of a KIR2DL1/DS1-Binding RIFIN Variable Region from HEK293F Cells 91
3.24 Purification of the KIR2DL1 and KIR2DS1 Ectodomains from HEK293F Cells 92
3.25 Binding of a RIFIN Variable Region and HLA-Cw4 to the KIR2DL1 and KIR2DS1 Ectodomains 93

4.1 Purification of the RIFIN Variable Region:LILRB1 Complex 99
4.2 Crystals of the RIFIN Variable Region:LILRB1 Complex 101
4.3 Final RIFIN Variable Region:LILRB1 Electron Density Map 104
4.4 Ramachandran Plot for the RIFIN Variable Region:LILRB1 Structure 105
4.5 The Structure of a RIFIN Variable Region 107
4.6 Sequence Logos for RIFIN Variable Regions 109
4.7 Conservation of the RIFIN Variable Region 110
4.8 The Structure of the RIFIN Variable Region:LILRB1 Complex 112
4.9 Change in Angle Between LILRB1 Domains Upon Ligand Binding . . 115
4.10 Purification and Biophysical Characterisation of the RIFIN C223S Mutant 116
4.11 Location of the P184 and G234 Residues of the RIFIN Within the Binding Interface 117
4.12 Purification of Binding Site Mutants for the RIFIN Variable Region . 118
4.13 Surface Plasmon Resonance Showing Binding of RIFIN Variable Region Mutants to LILRB1 119
4.14 Circular Dichroism Spectrum for the G234R Mutant of the RIFIN Variable Region 120
4.15 NFAT-GFP Reporter Assay Showing the Effect of the G234R Mutation on Signalling Through LILRB1 121
4.16 A Sequence Logo for Residues from LILRB1-Binding RIFINs corresponding to Residues 220-255 of Pf3D7_1254800 122
4.17 Purification of Wild Type and Binding Site Mutants for 3 LILRB1-Binding RIFIN Variable Regions 123
4.18 Circular Dichroism for the Wild Type and L264R Mutant of the Pf3D7_0100400 Variable Region 124
4.19 Surface Plasmon Resonance for the Binding of the Wild Type and L264R Mutant of the Pf3D7_0100400 Variable Region to LILRB1 . . 125

4.20 Surface Plasmon Resonance for the Binding of the Pf3D7_0100400 Variable Region to 3 LILRB1 Constructs 125
4.21 Flow Cytometry Analysis of the Binding of LILRB1-Fc to Erythrocytes Infected with Transgenic *P. falciparum* Expressing Specific RIFINs . 127
4.22 Purification of the Pf3D7_0100400 Variable Region:LILRB1 Complex 128
4.23 Crystals of the Pf3D7_0100400 Variable Region:LILRB1 Complex . . 129
4.24 A Structural Comparison of Binding to LILRB1 by a RIFIN Variable Region and an MHC Class I Complex 130
4.25 Supported Lipid Bilayer Setup for TIRF Microscopy 133
4.26 Effect of RH5 and anti-RH5 Antibody on NK Cell Activation 135
4.27 Images of RIFIN, MHC, and LILRB1 Localisation at the Immunological Synapse Formed Between NK Cells and a Supported Lipid Bilayer 135
4.28 Analysis of Quantity and Co-Localisation of the RIFIN Variable Region, MHC, and LILRB1 at the NK Cell Immunological Synapse . 137
4.29 Images of Perforin Accumulation at the Immunological Synapse Formed Between NK Cells and a Supported Lipid Bilayer 138
4.30 Analysis of the Quantity of Perforin at the NK Cell Immunological Synapse in the Presence of the RIFIN Variable Region and MHC . . 139

5.1 Expression Trial of CIR Ectodomain Constructs in HEK293F Cells . 145
5.2 Purification of 3 CIRs from HEK293F Cells 146
5.3 Circular Dichorism Spectra for 3 CIR Ectodomains 147
5.4 A Crystal of the CIR, PCHAS_1200500 148
5.5 Selenomethionine Incorporation into the CIR, PCHAS_1200500 . . . 153
5.6 Final CIR Electron Density Map 155
5.7 Ramachandran Plot for the Structure of the CIR, PCHAS_1200500 . 157
5.8 The Structure of a CIR 158

5.9 Flexibility within the CIR Structure 159
5.10 Sequence Logo for the Ectodomain of CIRs 161
5.11 Conservation of the Ectodomain of CIRs 162
5.12 Disorder Predictions for Representatives of Each CIR Familiy 165
5.13 Conserved Protein Architecture of the CIRs 166
5.14 Sequence Analysis of Putative PIR Families 168
5.15 Phyre2 Threading Confidence Scores for Putative PIR Families 170
5.16 Sequence Logo for the Ectodomain of Putative PIRs 172
5.17 Comparison of the CIR and RIFIN Variable Region Structures 175
5.18 Evolutionary Tree of *Plasmodium* Species with Those Expressing PIR Proteins in Red 178

List of Tables

3.1 Kinetic Values for Binding of the Pf3D7_1254800 Variable Region to LILRB1 Full-Length and D1D2 81

3.2 Kinetic Values for Binding of the KIR2DL1/DS1-Binding RIFIN Variable Region and HLA-Cw4 to KIR2DL1 and KIR2DS1 93

4.1 Data Processing Statistics for the RIFIN Variable Region:LILRB1 Structure 102

4.2 Refinement Statistics for the RIFIN Variable Region:LILRB1 Structure 106

4.3 Structure-Based Threading of LILRB1-Binding RIFINs 111

4.4 Table of Interactions between the RIFIN Variable Region and LILRB1 114

4.5 Kinetic Values for Binding of the Pf3D7_1254800 C223S Mutant to LILRB1 116

4.6 Kinetic Values for Binding of the Pf3D7_0100400 Variable Region to LILRB1 Full-Length and D3D4 126

4.7 Table of Interactions between HLA-A2 and LILRB1 132

5.1 Data Processing Statistics for the CIR, PCHAS_1200500 150

5.2 Sulphur-SAD Data Processing Statistics for the CIR, PCHAS_1200500 154

5.3 Refinement Statistics for the CIR, PCHAS_1200500 156

5.4 Variation in Loop Lengths Among CIRs 163

5.5 Variation in Lengths of the Linker Between the CIR Fold and the Transmembrane Helix Among CIR Families 164

5.6 Structure-Based Threading of PIRs 169

5.7 Structure-Based Threading of the RIFIN Constant Region 176

A.1 Summary of Protein Constructs and Primers 192

B.1 Kinetic Values for Binding of the KIR2DL1/DS1-Binding RIFIN Variable Region and HLA-Cw4 to KIR2DL1 and KIR2DS1 193

C.1 Structure-Based Threading of Putative PIRs 196

C.2 Structure-Based Threading of RIFINs 197

C.3 Structure-Based Threading of Putative PIRs from *P. reichenowi* onto the RIFIN Variable Region Structure 197

Abbreviations

2xYT	2 x Yeast tryptone broth
β2M	Beta-2-microglobulin
ABC	Dimethylamine borane complex
ACT	Artemisinin-combination therapies
ADCC	Antibody-dependent cell-mediated cytotoxicity
AMA1	Apical membrane antigen 1
ATS	Acidic terminal segment
CD	Circular dichroism
CDR	Complimentarity-determining region
CIDR	Cysteine-rich interdomain region
CSP	Circumsporozoite protein
DBL	Duffy-binding-like
DBP	Duffy binding protein
DDT	Dichorodiphenyltrichloroethane
DTT	Dithiothreitol
EBL	Erythrocyte binding-like
EDTA	Ethylenediaminetetraacetic acid
EPIC	Exported protein-interacting complex
ER	Endoplasmic reticulum
FCS	Fetal calf serum
GFP	Green fluorescent protein
GPI	Glycosylphosphatidylinositol
HEK	Human embryonic kidney
HLA	Human leukocyte antigen
Ig	Immunoglobulin

IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITAM	Immunoreceptor tyrosine-based activation motif
ITIM	Immunoreceptor tyrosine-based inhibition motif
KIR	Killer-cell immunoglobulin-like receptor
LAIR-1	Leukocyte-associated immunoglobulin-like receptor 1
LB	Lysogeny broth
LILR	Leukocyte immunoglobulin-like receptor
MHC	Major histocompatibility complex
MPD	2-Methyl-2,4-propanediol
MSP1	Merozoite surface protein 1
NFAT	Nuclear factor of activated T cells
Ni-NTA	Nickel-nitrilotriacetic acid
NK	Natural killer
OD	Optical density
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethylenimine
PEXEL	*Plasmodium* export element
PfCRT	*Plasmodium falciparum* chloroquine resistance transporter
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PfHRP II	*Plasmodium falciparum* histidine-rich protein II
PfRH	*Plasmodium falciparum* reticulocyte binding protein homologue
PHIST	*Plasmodium* helical interspersed subtelomeric
PI(3)P	Phosphatidylinositol-3-phosphate
PI3K	Phosphatidylinositol-4-phosphate 3-kinase
PIR	Plasmodium interspersed repeat
PMSF	Phenylmethylsulfonyl fluoride

PNEP	PEXEL-negative exported protein
PPM	Parasite plasma membrane
PTEX	*Plasmodium* translocon of exported proteins
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
RIFIN	Repetitive interspered family of polypeptides
RON2	Rhoptry neck protein 2
S2	Schneider 2
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SICAvar	Schizont-infected cell agglutination antigens
SLB	Supported lipid bilayer
SPR	Surface plasmon resonance
STEVOR	Subtelomeric open reading frame
SURFIN	Surface-associated interspersed proteins
TEV	Tobacco etch virus
TFF	Tangential flow filtration
TIRF	Total internal reflection fluorescence
TRX	Thioredoxin
VSA	Varient surface antigen

Chapter 1

Introduction

1.1 Malaria

1.1.1 History

References to malaria data back several thousand years, with Chinese documentation from 2700 BC, Mesopotamian tablets from 2000 BC, and Egyptian papyri from 1570 BC all describing symptoms such as fevers and enlarged spleens, typical of malaria infections in those living around marshes and swamps[1].

These symptoms were believed to be caused by miasma, a poisonous mist filled with particles from decomposed matter, a hypothesis which remained until the 19^{th} century, when Charles Louis Alphonse Laveran discovered the presence of parasites in the blood of malaria patients in 1880[2], and Ronald Ross identified the mosquito as the vector for avian malaria in 1897[3]. This was shortly followed by the confirmation of the mosquito as the vector for human-infective malaria parasites between 1898 and 1900 by a number of Italian scientists[2].

This was followed by significant research into the treatment and prevention of malaria infection throughout the 20^{th} century, and the establishment of a Global Malaria Eradication Programme in 1955 by the World Health Organisation, which succeeded in eliminating malaria from a number of regions around the world, including Europe and North America[4]. While strong progress has also since been made globally, with

the impact of malaria in terms of both cases and deaths significantly reduced, the disease still presents a substantial challenge to this day.

1.1.2 Global Disease Burden

Malaria remains a significant global health problem, infecting and killing vast numbers of people around the world every year. In 2018 alone there were an estimated 228 million cases of the disease. These were primarily found across Africa, but large numbers of cases were also seen in India and Southeast Asia, the Mediterranean, and the Americas[5].

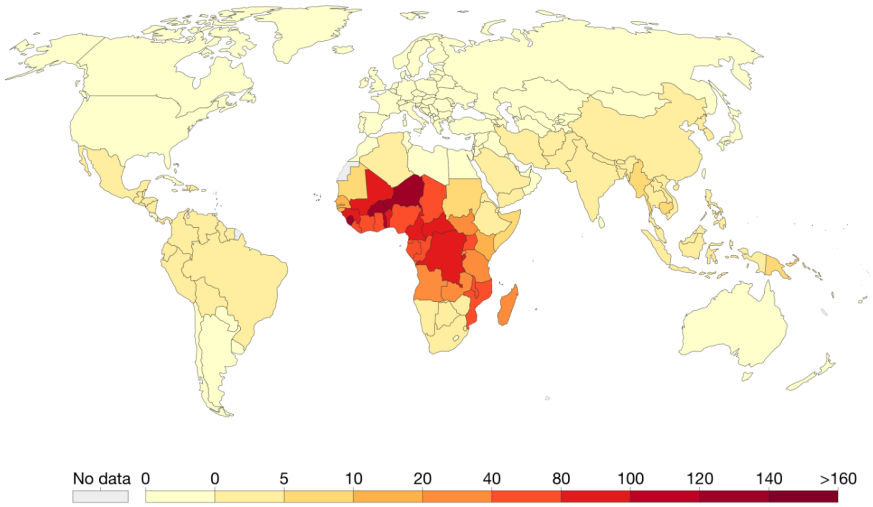


Figure 1.1: Map Showing the Number of Deaths from Malaria per 100000 People by Country in 2017 Figure adapted from https://ourworldindata.org/

These cases lead to huge numbers of deaths every year, with an estimated 405000 people killed by the disease in 2018[5]. The majority of these deaths, around 67%, were children under the age of 5, and almost all, around 94%, were in Africa **(Figure 1.1)**.

Significant progress has been made across recent history in decreasing both the incidence and death rates of malaria, with the number of cases falling from an estimated 262 million in 2000. Even more progress has been made in decreasing

mortality rates, with the number of deaths falling by over half since 2000, down from 839000 to 405000 in 2018[6], primarily due to the use of bed nets, indoor residual spraying and artemisinin-based therapies[7]. This progress is stalling[8], however, and so further understanding of the disease, and the parasite which causes it, are needed to help decrease the still vast numbers of malaria cases and deaths occurring each year.

1.1.3 The *Plasmodium* Parasite

The *Plasmodium* parasite which is responsible for malaria is a member of the *Apicomplexa*[9], a phylum named for their distinctive apical organellar complexes, utilised for host cell invasion[10].

There are 6 known human-infective species of *Plasmodium*[11]. *P. falciparum* is responsible for both the majority of cases and almost all malaria deaths. This species is predominantly found in sub-Saharan Africa, where it is responsible for the vast majority of cases, but is also responsible for around 50% of cases in Southeast Asia, and 71% of those in the Mediterranean region[5].

P. vivax is the next most prevalent species of the parasite. This is most common across India, where it acounts for 47% of cases, and the Americas where around 75% of cases are caused by the species[5].

Four other species of *Plasmodium* are responsible for the remainder of the global malaria burden, although in far smaller numbers. *P. knowlesi* is found predominantly across Southeast Asia[12], and is the cause of the majority of severe disease in some regions, such as Malaysia[13,14]. *P. malariae* is present mostly across sub-Saharan Africa, where it is seen in between 15 and 30% of cases, commonly during co-infection with *P. falciparum*, and also to a lesser extent across Asia and South America, in around 1% of cases[15]. Two closely related species, *P. ovale curtisi* and *P. ovale wallikeri*[16] are the cause of the remainder of malaria cases. These species are found

mostly in Western Africa, causing around 10% of cases in the region, but also cause between 3 and 5% of cases in India and Southeast Asia[15].

In addition to this, there are more than 50 other species of *Plasmodium* which infect a wide range of hosts, such as *P. reichenowi* which infects non-human primates[17], *P. chabaudi* which infects rodents[18], and *P. gallinaceum* which infects birds[19].

1.1.4 The *Plasmodium* Life Cycle

1.1.4.1 The Exo-Erythrocytic Stage

The *Plasmodium* parasite has a complex life cycle consisting of distinct stages in both the mosquito and human hosts **(Figure 1.2)**. The Human-infective species of the *Plasmodium* parasite are transmitted by the *Anopheles* mosquito. When the mosquito takes a blood meal, it deposits saliva into the skin to prevent the blood from coagulating[20]. Elongated and motile sporozoites are ejected into the skin, typically a few hundred[21]. These move by gliding until they enter the blood circulation either through penetration of a blood vessel, or via the lymphatic system[22,23].

The parasite then travels to the liver, where the sporozoite invades a hepatocyte[24]. This occurs through localisation to the liver sinusoid, mediated by binding of circumsporozoite protein (CSP) to heparan sulfate proteoglycans in the extracellular matrix[25]. This is followed by movement through sinusoidal endothelial cells and liver macrophages, termed Kupffer cells, until the parasite enters the liver parenchyma[26]. The sporozoites migrate through a number of hepatocytes, before invasion of a final hepatocyte. This migration damages the hepatocytes, causing rupture of the membrane, inducing secretion of hepatocyte growth factor which increase susceptibility to infection[27].

Following hepatocyte invasion, the parasitophorous vacuole (PV) is formed, within which the sporozoite replicates asexually, and develops into schizonts containing thou-

sands of merozoites. Merozoites are then released within merosomes, formed from the host cell membrane, into the bloodstream[28]. Upon reaching the pulmonary vasculature the merozoites are released from the merosomes into the bloodstream[29].

During an infection with *P. vivax* or *P. ovale*, development of sporozoites into schizonts can be delayed by lengthy periods of time. In this case they are termed hypnozoites, and this latency can allow for relapse of infections from these *Plasmodium* species[30].

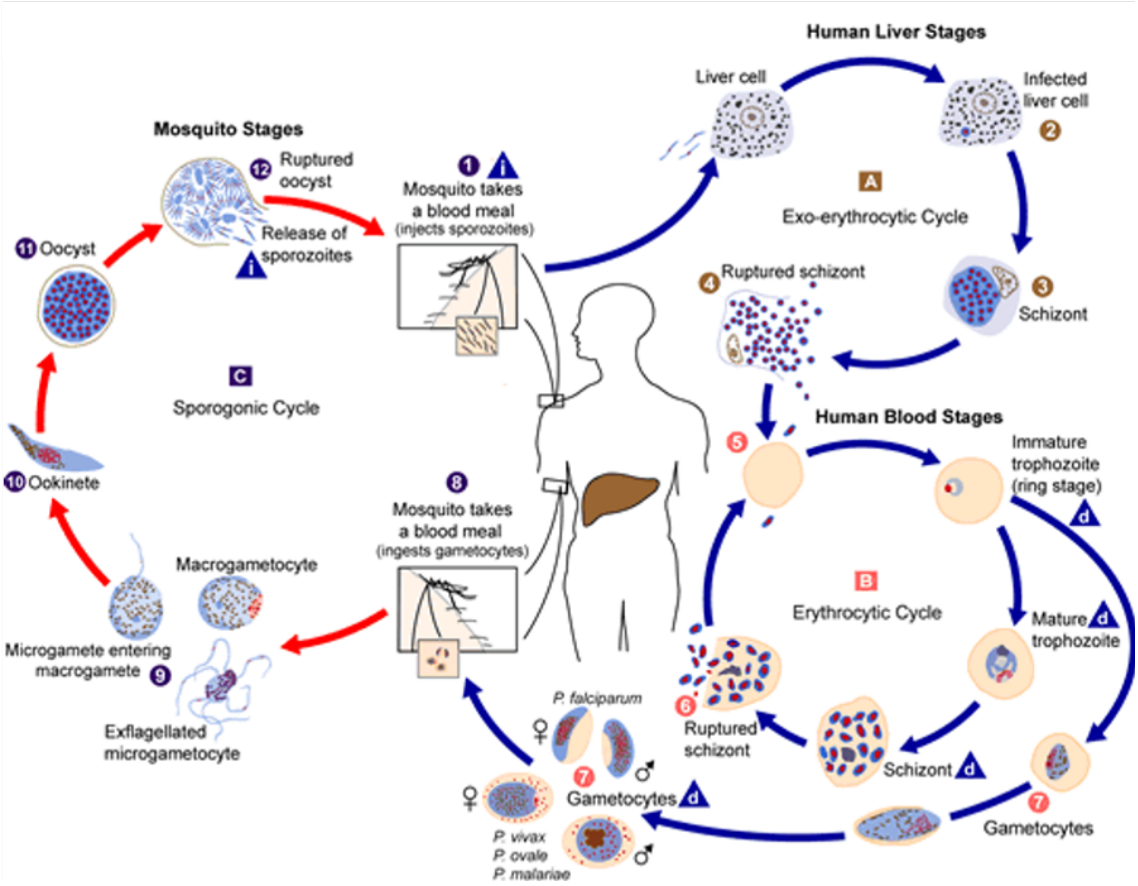


Figure 1.2: Life Cycle of the *Plasmodium* Parasite Figure adapted from https://www.cdc.gov/malaria/about/biology/index.html

1.1.4.2 The Erythrocytic Stage

Once released into the blood, merozoites rapidly attach to and invade red blood cells. The initial attachment of the parasite to the red blood cells is mediated by low-affinity interactions between ligands on the merozoite surface with host receptors[31].

These ligands are primarily glycosylphosphatidylinositol (GPI) anchored membrane proteins, such as merozoite surface protein 1 (MSP1) which is believed to interact with band 3 on the surface of the red blood cell[32].

This is followed by reorientation of the parasite to place its apical end against the erythrocyte membrane, allowing for the interaction of invasion proteins with host receptors. These differ between *Plasmodium* species, with, for example, *P. vivax* utilising the interaction between Duffy binding protein (DBP) with Duffy antigen on the host membrane[33], and *P. falciparum* relying on interactions formed by the *P. falciparum* reticulocyte binding protein homolog (PfRH) and erythrocyte binding-like (EBL) families[34].

These interactions lead to the formation of a tight junction, through interactions formed by the RON-AMA1 complex[35], and calcium mediated signalling between the parasite and host cell. This results in the release of proteins and lipids from the rhoptries of the merozoite, and the formation of a parasitophorous vacuole membrane (PVM) within which the parasite is enveloped[36].

Following invasion, the parasite enters the ring stage, also termed the early-trophozoite stage, which is followed by the mid-trophozoite stage. This is then followed by repetitive nuclear division and maturation into schizonts containing between 16 and 20 nuclei. These divide to form merozoites, which burst from the PVM and red blood cell membrane into the blood, where they can invade new erythrocytes and repeat the cycle[37]. Whilst inside the red blood cell, the parasite extensively modifies the host cell through secretion of proteins to the host cell membrane. This will be discussed in detail in section 1.2.

1.1.4.3 Gametocytogenesis and Sporogony

An alternative developmental pathway for merozoites also exists. Following invasion, instead of forming ring-stage parasites, a small minority of merozoites undergo

gametocytogenesis to form male and female gametocytes, termed micro- and macro-gametocytes respectively[38]. This primarily occurs in the spleen and bone marrow[39]. When taken up in a mosquito blood meal, these gametocytes undergo further development into gametes, a process triggered by xanthurenic acid, a mosquito-derived metabolite. These gametes then fuse to form a zygote, which matures into a motile ookinete[40]. This is followed by traversal of the ookinete across the midgut, and encystment beneath the midgut basal lamina, where development into oocysts occurs[41]. After multiple rounds of cellular division and differentiation, sporozoites form and are released. The sporozoites then invade the mosquito salivary glands, from where they can be injected into a new host.

1.1.5 Disease

Initial symptoms of malaria are non-specific, and typically include headache, fatigue, muscle aches and abdominal discomfort. This is often accompanied by a fever, along with nausea, vomiting, and orthostatic hypertension[42].

Most patients will have what is termed an uncomplicated infection, which may progress to include mild anaemia, caused by the destruction of large numbers of red blood cells, but is typically not fatal, with a mortality rate of around 0.1% when treated with antimalarial drugs[42].

A small proportion, around 1%, of cases progress to severe malaria. Manifestations of severe disease vary depending on age. Severe anaemia and hypoglycemia are common in children, while jaundice, kidney injury, and acute pulmonary oedema are more common in adults. Acidosis, and impaired consciousness or coma occur independent of age[43].

The vast majority of severe malaria is caused by *P. falciparum*, and severe disease is believed to be caused by both a high parasite load and adherence of infected red blood cells to either the vasculature[44], other infected erythrocytes through platelet-mediated

agglutination[45], or to uninfected erythrocytes, termed rosetting[46]. Sequestration of infected erythrocytes leads to both restricted blood flow and inflammation caused by an innate immune response against the parasites. Cytoadherence of infected red blood cells has the additional effect of allowing the parasite to avoid both immune detection and clearance in the spleen[47].

The most severe malaria cases occur when sequestration of infected erythrocytes takes place within the vasculature of the brain, leading to cerebral malaria. This is characterised by impaired consciousness and coma, and even with appropriate treatment has a mortality rate of 15 to 20%[48].

The adherence of infected erythrocytes is mediated by proteins exported to the erythrocyte surface by the parasite. In an infection with *P. falciparum*, three major families of proteins are exported, the PfEMP1, RIFIN, and STEVOR families, which have been implicated in both the adherent properties of infected erythrocytes and interactions with the host immune system to aid evasion by the parasite. These proteins will be discussed in more detail in section 1.2.

1.1.6 Disease Prevention and Treatment

The main ways in which malaria is currently managed are through attempting to prevent infection by targeting the *Anopholes* mosquito which spreads the *Plasmodium* parasite, and through the use of anti-malarial drugs to treat the disease.

1.1.6.1 Vector Control

The sole means by which the *Plasmodium* parasite can be transmitted to humans in a natural setting is through the *Anopholes* mosquito, and so attempts to reduce the spread of disease have largely focused on targeting the mosquito population.

The use of insecticides in the way of both indoor residual spraying and insecticide treated bed-nets in malaria endemic regions resulted in up to 50 fold decreases in

the local mosquito population, leading to significant declines in the prevalence of malaria infections in those areas[49].

However, the main insecticides used are dichlorodiphenyltrichloroethane (DDT) and pyrethroid insecticides which target a voltage-gated sodium channel. Mutations within this channel conferring resistance have since developed, and so insecticide resistance in mosquitoes is increasing[50].

Alternative mechanisms of vector control have been suggested. These include the use of symbiotic microorganisms which can infect the mosquito and reduce their capacity to transmit the *Plasmodium* parasite. Infection of the *Anopheles* mosquito with the bacteria *Wolbachia* significantly reduces *Plasmodium* levels within the insect[51]. Similarly, mosquitoes infected with the fungal parasite *Microsporidia MB* have been shown to have undetectable levels of sporozoites following infection with *P. falciparum*[52]. This type of control could prove invaluable in the future as a means of decreasing the spread of the *Plasmodium* parasite.

1.1.6.2 Antimalarial Drugs

Historically the most commonly used antimalarial drugs have been quinoline-containing drugs, primarily chloroquine. This was the standard treatment for malaria in the 1950s, functioning by inhibiting the polymerisation of haem, which is released during the proteolysis of haemoglobin by the parasite. This haem is usually polymerised to form harmless haemozoin crystals. Quinoline forms a complex with haem, which is incorporated into the polymer to terminate chain extension, resulting in a buildup of toxic haem[53].

Resistance to chloroquine is now widespread, due to mutations in the chloroquine resistance transporter (PfCRT). This is an integral membrane protein found in the internal digestive vacuole of the parasite, and mutations have emerged which reduce the accumulation of chloroquine by the parasite by directly transporting the drug

out of the vacuole[54].

Artemisinin is now used instead as the main anti-malarial drug. This also acts by targeting the haem polymerisation pathway. The endoperoxide bridge within the molecule is cleaved through binding of free iron formed during haemoglobin digestion, producing radicals which damage a range of cellular components through alkylation, includng the membrane of the digestive vacuole of the parasite. In addition to this, haem-artemisinin adducts can interact with the *P. falciparum* histidine-rich protein II (PfHRP II), which is a haem polymerase. This interaction displaces the haem from PfHRP II, inhibiting polymerisation, and leading to a buildup of toxic haem[55]. Artemisinin has a very short half-life, only around 1 hour, and so is typically combined with a longer-lasting drug with a different mode of action, such as lumefantrine, in artemisinin-combination therapies (ACT). This acts both to improve the efficacy of treatment, as well as to reduce the risk of the development of resistance.

Despite this, resistance appears to be developing to these treatments. In western Cambodia, for example, a rise in resistance is being seen which leads to slower parasite clearance when treated with ACT[56]. This worrying development highlights the need for further efforts to both modify existing drugs to attempt to reduce resistance, as well as to produce new anti-malarial treatments.

1.1.6.3 Vaccine Development

A more permanent solution to reducing the numbers of both infections and deaths from malaria would be the development of a vaccine, so that infection can be prevented prior to the need for drug intervention. Due to the complex life cycle of the parasite, there are a number of potential vaccine strategies in development, primarily focused on three stages of the life cycle.

Pre-erythrocytic, or liver stage, vaccines, attempt to target the parasite during the sporozoite stage. This is before any symptoms develop, and when the parasite load

is relatively low. The current leading malaria vaccine candidate, RTS,S, acts at this stage. This consists of fragments of the circumsporozoite protein (CSP) fused to a hepatitis B surface antigen, and has progressed as far as phase 3 trials, where the "R" and "T" refer to the central repeat region and T-cell epitopes of the CSP respectively, and the "S" for the hepatitis B surface antigen[57]. This vaccine has showed limited efficacy, reducing the number of cases by 39%, and the number of severe cases by just 29% over a 4 year period after a total of 4 doses[58].

Erythrocytic, or blood stage, vaccine targets are typically merozoite surface proteins which are key to red blood cell invasion. Leading candidates for these include the apical membrane antigen 1 (AMA1)-rhoptry neck protein 2 (RON2) complex, and PfRH5. Immunisation of *Aotus* monkeys with the AMA1-RON2 complex resulted in complete protection from infection in 4 of 8 monkeys, and delayed infection in 3 of the remaining 4, demonstrating significant potential for use as a vaccine[59]. Efforts are also ongoing to develop a vaccine targeting the conserved and essential invasion protein, PfRH5. Antibodies targeting this protein are able to block erythrocyte invasion[60], and immunisation of *Aotus* monkeys resulted in protection from infection[61]. Human trials testing the efficacy of PfRH5 as a vaccine component are ongoing.

Transmission-blocking vaccines target antigens of the parasite following gametocytogenesis, with an aim to induce antibodies during the human-stage of the life cycle. These act once the parasite is in the mosquito midgut to ultimately prevent oocyst formation, either through preventing fertilisation, or by destroying gametes or ookinetes. Potential candidates here include the gamete surface antigens Pfs48/45 and Pfs230[62], and the ookinete surface protein Pfs25[63]. Antibodies targeting both Pfs48/45[64] and Pfs25[65] have been shown to have transmission-blocking activity, demonstrating the potential of these candidates.

While progress is being made on a number of fronts in the development of a malaria vaccine, there is still a long way to go, with the best candidates so far providing

relatively limited protection. Further development is clearly needed, therefore, along with a better understanding of the interaction between the immune system and the parasite.

1.2 Variant Surface Antigens

1.2.1 Overview

Following red blood cell invasion, a number of modifications are made to the infected erythrocyte which facilitate evasion of the host immune response. This occurs through the export of parasite proteins to the red blood cell surface, which are able to interact with a number of host receptors on a wide range of host cells, including endothelial cells, erythrocytes, and leukocytes, allowing immune evasion both through cytoadhesion to avoid the splenic filter[66] and through direct modulation of the host immune response[67].

Antigenic variation in human-infective malaria was first seen in *P. knowlesi*[68], where it was demonstrated that sera isolated from monkeys infected with a specific parasite strain were ineffective at targeting infection with another strain. More specific characterisation was then achieved in 1984, when a parasite derived antigen was identified on the surface of a *P. falciparum*-infected erythrocyte, which was termed *P. falciparum* erythrocyte membrane protein 1 (PfEMP1)[69]. The *var* gene family encoding the PfEMP1 proteins was then characterised in 1995 as the first known variant surface antigen (VSA) family produced by *P. falciparum*[70], with the family implicated in both cytoadherence[71] and rosetting[72].

Genomic analysis revealed the existence of a second variable multigene family, the *rif* (repetitive interspersed family) genes[73], and the encoded RIFIN proteins were soon shown to also be present on the surface of the infected erythrocyte[74]. The function of the RIFINs remained a mystery until recently, when they were implicated in both

rosetting[75] and direct immune modulation[67].

A third major VSA family produced by *P. falciparum* was identified at a similar time, the STEVOR (subtelomeric variable open reading frame) protein family[76]. This was later confirmed to be exported to the surface of a *P. falciparum*-infected red blood cell[77], and while the function of this family is less clear, they have been implicated in rosetting[78].

Two much smaller families, the surface-associated interspersed proteins, SURFINs, and the *P. falciparum* Maurer's clefts 2 transmembrane, *Pf*MC-2TM, have also been identified on the surface of the infected erythrocyte[79,80], although little is known about the function of members of either family.

While the VSA families of *P. falciparum*, and in particular the PfEMP1 family, are the most well studied, VSA families have also been identified for a number of other *Plasmodium* species. The Schizont-infected cell agglutination antigens (SICAvar) are found exclusively in *P. knowlesi*[81], and while again little is known about the function of these proteins, both genomic[82] and proteomic[83] analysis suggests an evolutionary relationship with the PfEMP1 family.

In addition to this, a much larger gene superfamily has been proposed, containing multigene families of putative variant surface antigens from a number of both human- and rodent-infective *Plasmodium* species. The first members of the *Plasmodium* interspersed repeat gene family, *pir*, were identified as the *vir* genes in the *P. vivax* genome[84]. This was followed by the discovery of homologous gene families in three rodent-infective species of *Plasmodium*, the *cir*, *bir*, and *yir* families from *P. chabaudi*, *P. berghei*, and *P. yoelii* respectively[85]. The *kir* family of *P. knowlesi*[86], and putative *pir* gene families in *P. malariae*, *P. ovale curtisi*, and *P. ovale wallikeri* have since also been identified through genomic analysis[87].

A function of the CIR proteins in modulation of the virulence of infection has been

suggested[88,89], and a member of the VIR family has been implicated in endothelial adhesion of infected erythrocytes[90], providing a possible mechanism of action. This adhesion has only been seen for one member of the VIR family, and so it is unclear whether there is a universal mechanism by which the PIR family functions.

While no clearly homologous family to the PIR proteins exists for *P. falciparum*, it has been suggested that the RIFINs may be part of the PIR family, due to the presence of unique DNA sequence motifs, along with similarity of gene structure, genome location, expression pattern, and gene size[86].

1.2.2 Variant Surface Antigens of *Plasmodium falciparum*

The largest VSA families expressed by *P. falciparum*, and the best understood, are the PfEMP1, RIFIN, and STEVOR families. These families show several similarities, including in their genomic organisation, expression, localisation, and function. All three gene families are typically found within subtelomeric repeat sequences[91], and following expression, the proteins are trafficked to the erythrocyte membrane via the *Plasmodium* translocon of exported proteins (PTEX) and through the Maurer's clefts[92].

1.2.2.1 Expression

All three families are expressed during the erythrocytic phase of the *P. falciparum* life cycle, though they do vary in the timing of expression. *Var* are transcribed during the ring stage, with *rif* genes transcribed slightly later in the early trophozoite phase[93]. *Stevor* transcription occurs later still, during the mid-trophozoite phase[94].

Typically, only a single PfEMP1 protein is produced and exported to the infected erythocyte surface at any one time[95], with *var* gene expression believed to be controlled by both regulatory regions upstream of each gene, and epigenetic modifications which

allow for mutually exclusive *var* gene expression. This does not appear to be true for either the *rif*[96] or *stevor*[97] genes, where multiple genes are transcriptionally active at a time.

In addition to this, while they have all been implicated in modulating the adherent properties of infected erythrocytes, the mechanism by which this is achieved varies, and other functions, such as the ability of some RIFIN proteins to directly modulate the activity of elements of the immune response, are, so far, unique to a particular family.

The PfEMP1 and STEVOR families are introduced in this section, while RIFINs are discussed in more detail in section 1.3.

1.2.2.2 Transport

During the blood stage of infection, the parasite grows within the parasitophorous vacuole (PV), and so all VSAs exported to the host cell membrane must pass through the parasite plasma membrane (PPM), and the PV membrane (PVM), before then being trafficked to the host cell membrane.

Transport of RIFINs and STEVORs is dictated by the presence of a conserved export motif termed the *Plasmodium* export element, or PEXEL[54]. This is a 5 residue motif, with the sequence RxLxE/Q/D. While the exact mechanism of export remains unclear, it is known that the PEXEL motif is bound to by phosphatidylinositol-3-phosphate (PI(3)P)[98] in the parasite endoplasmic reticulum (ER), followed by cleavage by the aspartic protease plasmepsin V[99] C-terminal of the conserved leucine residue. The remaining N-terminal residue is then acetylated[100]. Plasmempsin V interacts with the signal peptidase complex 25, the Sec61 ER translocon, and an ancillary protein PfSec62, which allow the processed PEXEL proteins to enter the ER. They are subsequently trafficked to the PV by an as yet unknown mechanism, potentially either through insertion into the PPM, which would then require extraction for

further trafficking, or by trafficking as soluble chaperoned complexes[92].

Unlike the RIFINs and STEVORs, PfEMP1 proteins lack the PEXEL motif[101], and so are referred to as PEXEL-negative exported proteins (PNEPs). These proteins instead appear to have an N-terminal hydrophobic region which acts as an export signal in lieu of the conserved motif[102].

Following export to the PV, the proteins need to be trafficked again across the PVM, with this process carried out by the *Plasmodium* translocon of exported proteins (PTEX)[103]. This is a complex with five currently identified components which work to transport the protein in an unfolded state[104] across the PVM. Three of these components are essential. The first, HSP101, is believed to unfold translocating proteins and drive translocation[103] through the channel formed by EXP2[105]. The final essential component is PTEX150, which is believed to provide structural support to the complex[106]. In addition to this, two auxiliary proteins complete the complex. The first, TRX2, is a thioredoxin which is believed to aid protein unfolding[107]. The role of the second auxiliary protein, PTEX88. remains unclear, although it has been shown to interact with a second protein complex, the exported protein-interacting complex (EPIC)[108]. This complex has been proposed to assist in transport of proteins across the PV to the PTEX following secretion from the ER[109].

Following translocation by the PTEX, the exported proteins are moved through the host cytoplasm as chaperone-associated transport complexes[110], either directly to the host cell membrane, or via the Maurer's clefts[111]. These are membranous structures established within the host cytoplasm by the parasite, which are believed to aid in sorting and trafficking of proteins to the erythrocyte membrane[112]. Movement of proteins from the Maurer's clefts to the erythrocyte membrane is again not well understood, although there is evidence that both chaperone-[113] and vesicle-mediated[114] transport occurs.

1.2.2.3 PfEMP1

Around 50 to 100 *var* genes are present in the genome of each parasite strain[115], encoding PfEMP1 proteins which are typically between 200 to 350 kDa in size[116], all of which share a common overall architecture despite only around 40% sequence identity between members of the family[117]. A small cytoplasmic domain termed the acidic terminal segment, ATS, is largely flexible and highly conserved[118]. This is connected to a single trans-membrane helix, and a much larger extracellular region. The ectodomain is formed predominantly from two types of domain, the Duffy-binding like (DBL) domain, which bears strong similarity to a number of *Plasmodium* invasion ligands such as DBP from *P. vivax*, and the cysteine-rich interdomain region (CIDR)[119]. The DBL domains can be further divided into 5 major families, and the CIDR domains into 4. The number and organisation of these domains varies between sequences, with extracellular regions ranging from 4 to 9 domains in length[119].

Only a single PfEMP1 protein is typically presented at any point during infection, due to mutually exclusive expression of *var* genes[95], although a parasite line which co-expresses two different PfEMP1 antigens has been observed[120]. This allows for antigenic variation through the successive expression of different members of the family with variable protein surfaces, and so little or no immunological cross-reactivity[91].

PfEMP1s are exported to regions on the surface of the erythrocyte, seen as electron-dense by electron microscopy, termed knobs[121,122], where they interact with a number of proteins which appear to boost their ability to mediate cytoadherence. Knobs are formed mostly from KAHRP, knob-associated histidine rich protein. Knockouts of this protein lead to the diffuse spreading of PfEMP1 across the membrane, which severely diminishes cytoadherence of infected red blood cells[123]. In addition to this, the PHIST, *Plasmodium* helical interspersed subtelomeric, domain proteins have been shown to localise to the knob complex, which are believed to link PfEMP1 to the host

cytoskeleton to again provide the mechanical strength needed for cytoadherence[124], through interactions with the cytoskeletal component Band 4.1[125].

PfEMP1 have been shown to interact with a variety of receptors to modulate both adhesion to vasculature endothelial cells and uninfected erythrocytes during rosetting, with these interactions known to contribute to both the course and severity of infection. Known binding partners include ICAM-1[126], commonly expressed on endothelial cells and leukocytes[127], CD36[128], expressed on a wide range of cells including endothelial cells, erythrocytes, and leukocytes[129], and EPCR[130], primarily found on endothelial cells of large blood vessels[131]. Additionally, PfEMP1 binding has been demonstrated to chondroitin sulfate A (CSA)[132], a sulfated glycosaminoglycan commonly found on proteoglycans in the placental syncytium[133].

Structural characterisation of the interactions with ICAM-1, CD36, and EPCR **(Figure 1.3)**, have shown a number of similarities. While ICAM-1 binding is mediated by interactions with DBL domains, in contrast to CD36 and EPCR binding which occurs through interactions with CIDR domains, all interactions have a strong hydrophobic component. A hydrophobic patch forms the core of all structurally-characterised interactions, with surrounding residues contributing hydrogen bonds which stabilise the interaction[134–136]. While the shape and chemistry of binding sites are typically well conserved across PfEMP1 which bind the same ligands, there is typically a high degree of sequence diversity, even in residues directly involved in forming interactions[134–136].

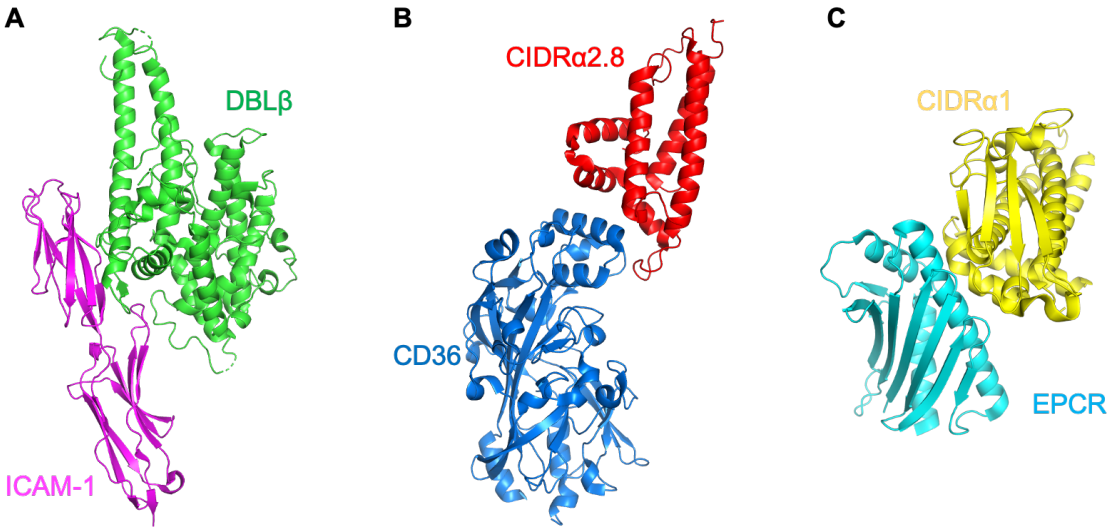


Figure 1.3: Structures of PfEMP1 Domains Bound to 3 Receptors A) A 3.67 Å structure of the DBLβ domain from IT4var13 PfEMP1 (green) bound to ICAM-1 (magenta) (PDB Code: 6S8U[134]). B) A 2.07 Å structure of the CIDRα2.8 domain from MCvar1 PfEMP1 (red) vound to CD36 (blue) (PDB Code: 5LGD[136]). C) A 2.65 Å structure of the CIDRα1 domain from HB3var03 PfEMP1 (yellow) bound to EPCR (cyan) (PDB Code: 4V3D[135]).

Binding of PfEMP1 to ICAM-1 has been shown to be associated with the more severe cerebral malaria, while CD36-binding, by contrast, was associated with uncomplicated infection[137]. PfEMP1 binding to CSA found on placental syncytiotrophoblasts leads to the severe form of disease termed pregnancy associated malaria, caused by the sequestration of infected erythrocytes in the placenta[133]. The export of PfEMP1 to the red blood cell surface, therefore, is important in both the outcome of infection, through interactions with different receptors, as well as immune evasion, through both sequestration of infected cells, as well as antigenic variation and switching of var gene expression.

1.2.2.4 STEVORs

The *stevor* genes form a slightly smaller family of VSAs in the *P. falciparum* genome, with typically around 40 genes seen in most isolates[138]. All *stevor* genes share a common 2 exon architecture, with a short first exon encoding a start codon and

putative signal peptide, followed by a longer second exon encoding the bulk of the around 30 kDa protein[139]. This gives rise to a similarly conserved protein architecture, where the signal peptide is followed by a PEXEL motif utilised for protein trafficking. There is then a semiconserved region, followed by a variable region, with the two separated by a hydrophobic stretch. This comprises the ectodomain, which is followed by a single transmembrane region and a highly conserved C terminus which forms the cytoplasmic domain[138]. While the hydrophobic stretch separating the semiconserved and variable regions has previously been suggested to form a second transmembrane domain[139], it has since been confirmed that both regions are exposed on the erythrocyte surface, dispelling this theory[80].

While less is known of the STEVORs relative to the PfEMP1 family, the proteins have been implicated in a few different processes, suggesting potential functions of the family. Interactions between STEVORs and Glycophorin C on the red blood cell surface have been shown to mediate rosette formation, indicating a potential function in immune evasion[78], and they have also been shown to contribute to the rigidity of the infected erythrocyte[140], which facilitates blockage of microcapillaries, and therefore sequestration of infected cells[141], again assisting in the evasion of the immune response, as well as helping to prevent splenic clearance.

In addition to this, a secondary localisation of STEVORs to the apical tip of the merozoite has been demonstrated[142], with antibodies targeting these proteins able to inhibit erythrocyte invasion, suggesting another role as an invasion protein, although the exact function here is still unknown[138].

1.3 RIFINs

The RIFINs form the largest family of VSAs expressed by *P. falciparum*, with typically between 150 and 200 *rif* genes present in the subtelomeric regions of the

genome of each strain of the parasite[143]. The protein products of these genes are typically between 30 and 45 kDa[144], with a common gene structure encoding a conserved protein architecture. A signal peptide is encoded by the short first exon, with the remainder of the protein encoded within a second, longer exon[145]. Following the signal peptide is a conserved PEXEL motif, followed by a semiconserved domain, a short hydrophobic stretch, and a variable domain. There is then a conserved transmembrane region and C-terminal domain[146] **(Figure 1.4)**.



Figure 1.4: Conserved Protein Architecture of the RIFINs

It has previously been suggested that the hydrophobic region separating the conserved and variable domains formed a second transmembrane helix, leading to a model whereby the variable region formed an exposed loop on the surface of the erythrocyte, with the conserved N- and C-terminal regions remaining within the cell[145]. More recent studies have challenged this model, however. Trypsin treatment of infected erythrocytes demonstrated that both the semi-conserved N-terminal region of protein and the variable region were exposed on the red blood cell surface[80]. This is supported by *in vitro* glycosylation experiments, where the RIFIN was purified within ER-derived microsomes, and only the region of protein exposed to the luminal side of the membrane could be glycosylated. This demonstrated that both the semiconserved and variable regions could be glycosylated, and so were both exposed on the same side of the membrane[147]. These studies suggest that the hydrophobic region separating the semiconserved and variable regions is not a second transmembrane helix, and so these two regions combined form the ectodomain of the RIFIN.

In addition to this, it has been confirmed that the PEXEL motif of the RIFINs is cleaved by Plasmepsin V during processing and export of the proteins[148], and so the signal peptide upstream of this does not form part of the mature ectodomain

following export to the erythrocyte membrane.

The sequence identity between members of the RIFIN family is low, at around 45%, which has made grouping the proteins very difficult. They have, however, been separated into two groups, A and B, based on the presence or absence respectively of a 25 residue stretch within the semiconserved region of the ectodomain[146]. While there is little known about the functional consequences of this difference, it has been suggested that only the group A RIFINs are exported to the red blood cell surface, while the group B RIFINs are retained within the parasite[149]. The localisation of group A RIFINs to the erythrocyte membrane has been confirmed through the use of immunofluorescence using multiple members of the group[80,150]. RIFINs have also been seen within the merozoites, with the group A RIFINs found at the apical tip[150], and group B RIFINs in the cytosol[149], although immunofluorescence data have also shown group B RIFINs on the merozoite surface[151]. The functional consequences of RIFIN export to the merozoite surface are currently, however, unknown.

1.3.1 Immunogenenicity

As expected for proteins exposed on the red blood cell surface, RIFINs are targets for the host immune system, with those in malaria endemic regions developing antibodies particularly to the semiconserved domain[152]. The fact that the antibody response against the semiconserved region was stronger is expected, due to the relative lack of variation in this region compared to the variable region of the ectodomain. This antibody response has been shown to be associated with faster parasite clearance, although does not appear to be strongly protective[153,154].

Antibodies targeting RIFINs have, however, been shown to be sufficient to induce antibody-dependent cell-mediated cytotoxicity by NK, natural killer, cells, thereby inhibiting parasite growth[155], highlighting the risk to the parasite of exporting proteins to the host cell surface.

In addition to this, more unusual antibody responses have been seen which target this protein family. Broadly reactive anti-RIFIN antibodies have been isolated from patients in malaria-endemic regions where the collagen-binding domain of the inhibitory immune receptor LAIR1 has been inserted into the third CDR loop[156]. This receptor is introduced in more detail in section 1.5.2. Structural studies of this LAIR1-containing antibody revealed that the collagen-binding residues lie on the exposed surface of the domain[157], suggesting that some RIFINs are able to bind LAIR1 at or near its collagen-binding site, and this has been utilised by the immune response as a means of targeting the RIFINs.

1.3.2 Interactions

The emergence of antibodies containing the LAIR1 ectodomain which were able to bind RIFINs revealed a potential function of the proteins through binding to inhibitory immune receptors. Three RIFINs have been identified which have the capability to bind LAIR1[67,156]. However, RIFIN binding to LAIR1 was not able to induce inhibitory signalling in a GFP-reporter system[67], suggesting that RIFINs are not able to directly modulate the immune response through this receptor.

A number of RIFINs have also been identified which were able to bind to another inhibitory immune receptor, LILRB1[67]. The receptor itself is introduced in more detail in section 1.5.1.1. Eleven LILRB1-binding RIFINs were identified, and in contrast to those interacting with LAIR1, RIFIN binding was sufficient to induce signalling through the receptor in a GFP-reporter assay[67].

In addition to this, erythrocytes infected with parasites expressing LILRB1-binding RIFINs were able to inhibit the production of IgM by primary human B cells, and the response of NK cells was similarly inhibited in the presence of LILRB1-binding RIFINs[67]. This strongly indicates that a key function of RIFINs is to downregulate the host immune response through interactions with inhibitory immune

receptors.

The LILRB1-binding capability of infected erythrocytes from Tanzanian patients with malaria was also characterised. This showed a significantly higher proportion of cells capable of binding the receptor in patients with severe illness compared to those with an uncomplicated infection, suggesting a contribution of LILRB1-binding to severe disease. However, this study was carried out with a particularly small sample of patients, just 9 with severe disease and 30 with non-severe malaria, and so only limited conclusions can be drawn from this[67].

RIFINs have also been implicated in rosetting. One RIFIN was shown to interact with the blood group A antigen, and expression of this RIFIN by parasites significantly increased the rate of rosetting in culture[75]. In addition to this, it was shown that sequestration of red blood cells infected with parasites expressing the blood group A antigen-binding RIFIN in the microvasculature was significantly increased[75]. These results combine to suggest another potential role for the RIFINs in immune evasion through mediation of rosetting and sequestration of infected erythrocytes.

1.4 PIRs

The largest multigene family known across a range of *Plasmodium* species is the *Plasmodium* interspersed repeat, or *pir*, gene family. Members of this family are seen within the genomes of *Plasmodia* infecting a variety of hosts, including human-infective and rodent-infective species of the parasite.

For the best characterised species, the *pir* family has been named according to the species, termed *vir* for those in *P. vivax*, *kir* for *P. knowlesi*, *cir* for *P. chabaudi*, *bir* for *P. berghei*, *yir* for *P. yoelii*. *Pir* genes have also been identified more recently for *P. malariae*, *P. ovale curtisi*, and *P. ovale wallikeri*[87].

The number of *pir* genes varies highly between species, from just 70 genes in the *P.*

knowlesi strain H genome[158], to nearly 1500 in the *P. ovale curtisi* GH01 genome[159]. These genes typically share a common architecture, with a short first exon, followed by a longer second exon usually ending in a putative transmembrane helix, and a highly conserved third exon predicted to contain the cytoplasmic domain[85]. Much like the RIFINs, these genes are found in the sub-telomeric regions of chromosomes for most species, although in *P. yoelii* some members are found in other regions of the chromosomes[160], and in *P. knowlesi* they are found randomly across the chromosomes[158].

The organisation and structure of the genes is one of the main reasons for the suggestion that the *rif* family may form part of the *pir* superfamily. The second intron of the rodent *pir* genes shows significant sequence homology to the *rif* intron[86]. Similarity of localisation within the genome, gene size, and predicted localisation of both PIR and RIFIN proteins to the infected red blood cell surface has added to these comparisons. Homology between the RIFINs and the rodent PIRs in particular would provide a highly valuable animal model for the study of RIFIN function, and so is a particularly attractive prospect. However, the lack of amino acid sequence similarity has made comparisons difficult, making it challenging to confidently place the RIFINs in the PIR superfamily.

In addition to this, while RIFINs have been shown to localise primarily to the infected red blood cell surface, along with the merozoite surface[80,149,150], localisation of PIR proteins is less clear. Members of the YIR family have been shown to localise to the erythrocyte membrane[161], as have some VIRs[84,90], although some members of the VIR family appear to remain within the merozoite[90]. Similarly, some BIRs have been shown to be exported to the host cell cytosol, but not displayed on the erythrocyte membrane[162], while others appear to be found at the host cell surface[163]. Localisation is particularly unclear for the CIR proteins. While some CIRs have been observed at the infected red blood cell membrane, the majority show alternate

localisation, including remaining in the PV, localising to the merozoite surface, or export to the host cell cytosol[164].

Despite this variety of protein localisation, the PIRs have a number of similarities. The majority of PIRs are predicted to be transmembrane proteins, with over 95% of PIRs from all species other than *P. vivax* predicted to contain a TM helix[165]. The VIRs, however, stand out in seeing predicted TM helices in only around half of the family[166].

The defining feature across all PIRs is that they contain a conserved N-terminal domain. This is a region of around 200 residues in length predicted to be highly alpha-helical in structure, with secondary structure predictions suggesting around 8 α-helices of around 12 to 20 residues in length. This is consistently followed by a far more variable stretch of sequence separating the conserved domain from the predicted TM helix[86].

In addition to this, sequence analysis of PIRs from the 3 rodent-infective *Plasmodium* species along with *P. vivax* and *P. knowlesi* revealed the existence of an orthologous group, with one member in each species genome, of PIR sequences with significantly higher conservation than the rest of the family **(Figure 1.5)**[82]. These sequences have been suggested to act as the ancestral PIRs, from which the rest of the family may have diverged. No homologue of these sequences was found in the *P. falciparum* genome, suggesting the absence of PIRs from this species.

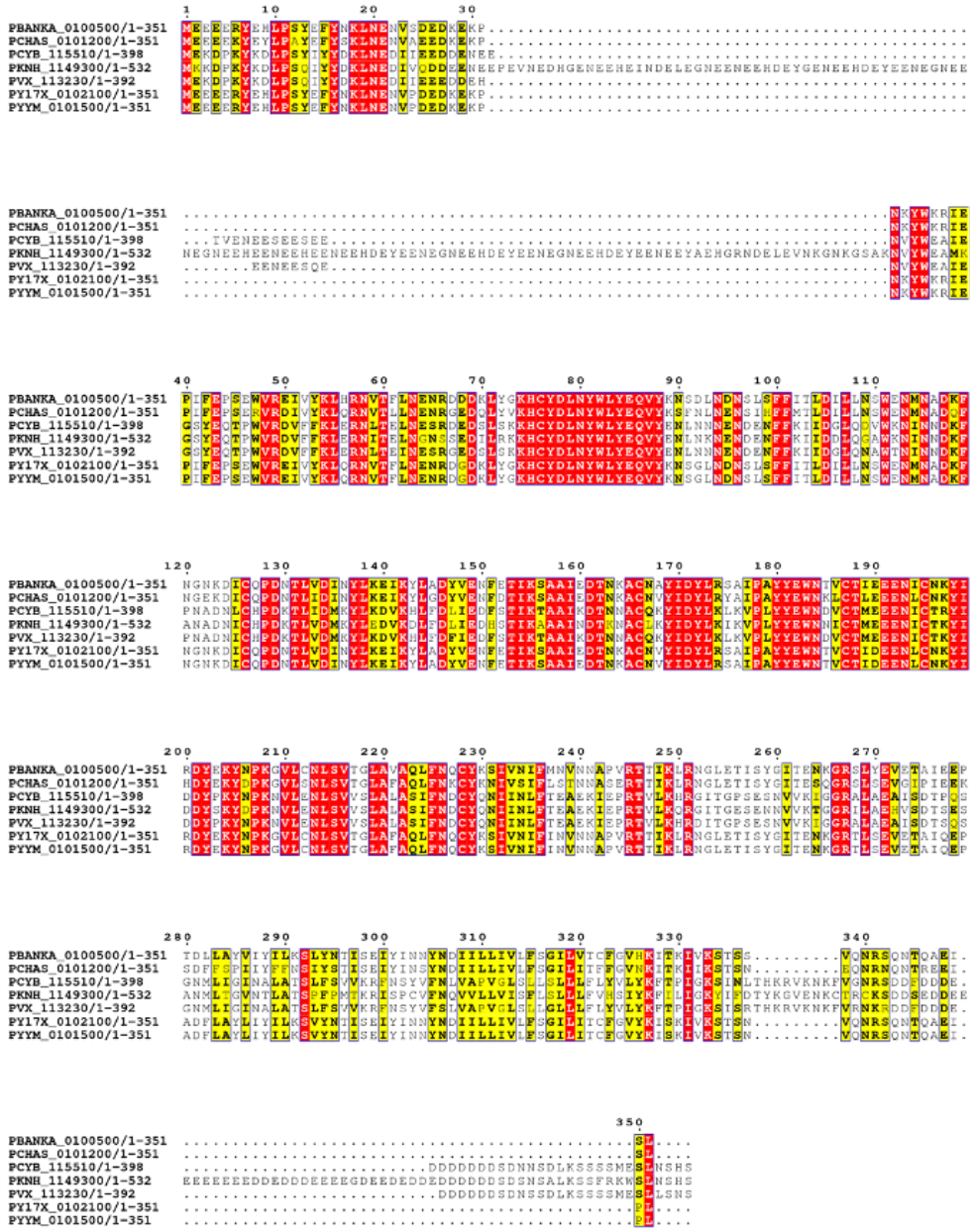


Figure 1.5: Sequence Alignment of Orthologous PIRs A red background for a column indicates a fully conserved residue, while a yellow background indicates conservation of residues with similar chemical properties. Figure prepared using ESPript 3.0[167].

1.4.1 PIR Function

While the function of the PIR family has not been conclusively determined, recent studies have implicated the proteins in the modulation of virulence of infection.

Evidence for this came from the study of differential gene expression between *P. chabaudi* parasites introduced in mice either through serial blood passage or through mosquito transmission[168]. Parasite virulence is increased through serial blood passage[169], which suggests that mosquito transmission acts to regulate parasite virulence within the host. Analysis of gene expression of *P. chabaudi* before and after mosquito transmission revealed that the largest changes in expression were within the *cir* genes, where a much broader range of *cirs* were expressed following mosquito transmission[168].

Further evidence for the role of PIRs in parasite virulence came from study of gene expression during chronic and acute infections with *P. chabaudi*[88]. The *cir* genes were again the most differentially expressed gene family, with more than half of the genes differentially expressed between the acute and chronic phases of infection. The bulk of the *cirs* showed higher expression during an acute infection, although some were more highly expressed in the chronic phase.

Further analysis showed that the genes were often arranged as clusters in the subtelomeres, and that genes within clusters were co-expressed during infection. The CIRs had been previously classified into short and long form, termed the S and L families respectively[170], and clusters containing mostly genes within the short S7 family were highly expressed during an acute infection, while clusters enriched with the long L1 family were dominant during a chronic infection[88], again suggesting that the CIRs are playing a role in determining parasite virulence, while also implying that the length of the protein may be a determining factor in its function.

The mechanism by which this occurs is not yet known, although the PIRs have been implicated in rosetting and vascular adhesion, which could provide an explanation. CHO cells transfected with some CIRs were able to bind uninfected red blood cells, and this could be partially disrupted using anti-CIR antibodies. This was confirmed by demonstrating binding of recombinant CIRs to red blood cells by flow

cytometry[164].

Additionally, *P. vivax*-infected erythrocytes have been shown to cytoadhere to endothelial cells, and this could be significantly inhibited through the addition of polyclonal VIR antisera[171], which could be explained by the demonstration that a member of the VIR family is able to bind to the endothelial receptor ICAM-1[90]. However, this interaction has so far only been confirmed for a single VIR, and so this cannot yet be used to form a general model for the mode of function of the whole family.

1.5 Inhibitory Immune Receptors

A key mechanism for regulation of the human immune response is through the balance of activating or inhibiting signals allowing for the rapid activation of components of the immune system upon recognition of a foreign body, or to prevent a response against self antigens[172].

One way in which this is achieved is through the pairing of activatory and inhibitory receptors with closely related ectodomains capable of binding to similar ligands, but with key differences in cytoplasmic domains which allow for modulation of the immune response[173]. Such receptors include the leukocyte immunoglobulin-like receptors (LILR)[174], expressed on a wide range of leukocytes including B cells, T cells, NK cells, and dendritic cells[175], and the NK cell-specific killer cell immunoglobulin-like receptors (KIR)[176].

Activatory receptors within these families contain immunoreceptor tyrosine-based activation motifs (ITAM), pairs of YxxL/I motifs separated by 6 to 8 residues[177], which are phosphorylated by Src family kinases upon receptor activation. This results in the recruitment of the tyrosine kinases ZAP-70 and Syk through SH2 domain-based interactions, leading to a cascade of phosphorylation events[178], which

ultimately lead to activation of the presenting cells.

Inhibitory receptors, by contrast, contain immunoreceptor tyrosine-based inhibition motifs (ITIM) within their cytoplasmic tails, which have the sequence V/I/LxYxxL/V[179]. These receptors cluster at immune synapses, leading to tyrosine phosphorylation by Src family kinases. This leads to recruitment of the tyrosine phosphatase SHP-1 and SHP-2, and the inositol phosphatase SHIP. While the downstream targets of these phosphatases have not been clearly defined, they are known to include both Syk and ZAP-70, the kinases involved in ITAM signalling, as well as other downstream components of activatory signalling such as phosphatidylinositol-4-phosphate 3-kinase (PI3K)[180]. This inhibitory response is typically dominant over the ITAM-mediated activatory response, as activatory signals at the immunological synapse are blocked[181].

1.5.1 LILR Family

One such family of immune receptors are the LILRs, which are members of the immunoglobulin (Ig) superfamily[174]. This is a family of 11 receptors, with 5 activatory and 5 inhibitor members of the family, along with a single soluble protein[182]. The activatory members of the family are termed LILRAs, and the inhibitory members LILRBs.

Each member contains either 2 or 4 Ig-like domains. For all other than the soluble LILRA3, these form the extracellular portion of the protein, which is connected via a single transmembrane helix to the cytoplasmic domain which contains the characteristic ITAM or ITIM motifs defining whether they are activatory or inhibitory receptors respectively.

Members of the family can be found on the surface of a broad range of cell types, including B, T, and NK cells, along with monocytes, dendritic cells, granulocytes, macrophages, and neutrophils[182]. The ligands for these receptors are typically both

classical and non-classical class I major histocompatibility complexes (MHC)[183]. Interactions have also been seen, however, with a range of other proteins including bone marrow stromal antigen 2, BST2[184], angiopoietin-like proteins[185], a damage-associated molecular pattern protein, S100A9[186], myelin inhibitors[187], and β-amyloids[188]. For some members of the family, such as LILRA2 and LILRB3, ligands are still unknown[189].

1.5.1.1 LILRB1

LILRB1, also known as LIR-1 or ILT-2, is an inhibitory member of the LILR family, and consists of four Ig-like domains, followed by a single transmembrane helix, and a cytoplasmic tail containing 4 ITIM motifs[174].

LILRB1 is widely expressed on most leukocytes[189], and has been shown to bind to a wide spectrum of both classical and non-classical MHC class I complexes[190], Previous structural characterisation has demonstrated that this is largely through interactions between the conserved β-2-microglobulin (β_2M) chain of the complexes and the hinge region between the two membrane-distal Ig-like domains of the receptor **(Figure 1.6)**[191–193].

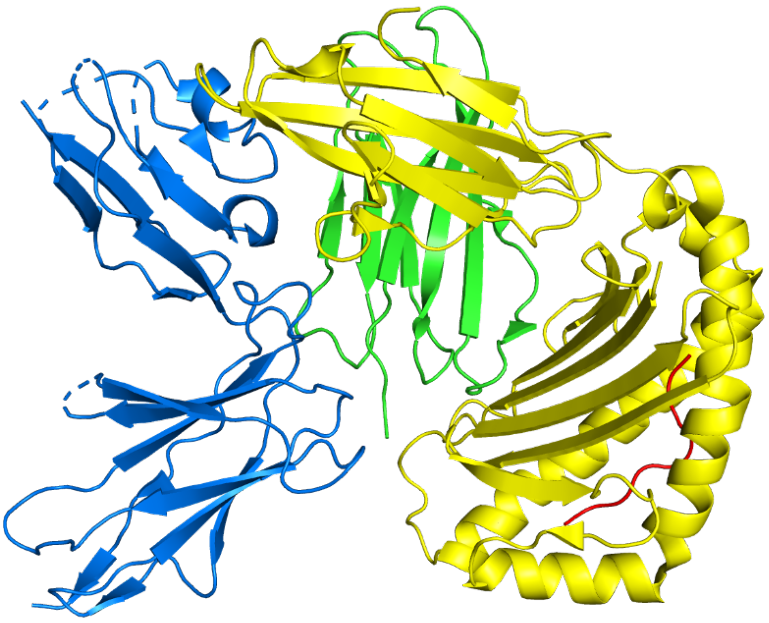


Figure 1.6: A Structure of HLA-A2 Bound to LILRB1 A 2.39 Å structure of the complex between the 2 membrane distal domains of LILRB1 (blue), with HLA-A2, where the α-chain is in yellow, the β_2M chain in green, and the bound peptide is in red (PDB Code: 6EWA[194]).

In addition to this, LILRB1 has been previously shown to be the target of multiple viruses as a means of inhibiting the immune response during an infection. Human cytomegalovirus produces the UL18 protein which has been shown to bind LILRB1[195], and in doing so is capable of reducing NK cell cytotoxicity, thereby inhibiting the immune response against the virus[196]. Structural studies have demonstrated this is achieved through closely mimicking the structure of the alpha-chain of an MHC class I molecule and associating with host β_2M[197].

Similarly, LILRB1 has been shown to be bound during infection with dengue virus, which downregulated antiviral responses triggered during antibody-dependent infection with the virus, again aiding avoidance of a strong immune response[198].

1.5.2 LAIR1

LAIR1, leukocyte-associated immunoglobulin-like receptor-1, is a similar inhibitory immune receptor, found on a wide range of peripheral mononuclear cells, including monocytes and dendritic cells, along with B cells, T cells, and NK cells[199].

The extracellular domain of the receptor contains a single Ig-like domain **(Figure 1.7)**, followed by a single transmembrane domain, then a cytoplasmic tail containing 2 ITIM motifs[200].

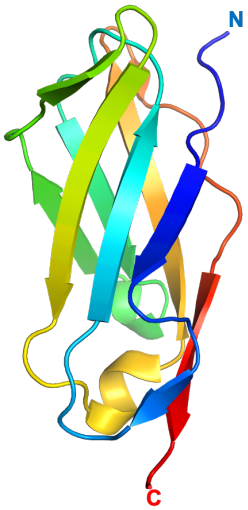


Figure 1.7: A Structure of the Ectodomain of LAIR1 A 1.80 Å structure of the ectodomain of LAIR1, shown in a rainbow representation where the N-terminus is in blue, and C-terminus in red (PDB Code: 3KGR[201]).

Unlike the LILR family of receptors, LAIR1 does not bind to MHC complexes, and instead has been shown to interact with collagen through conserved glycine-proline-hydroxyproline repeats[202]. This is the only known example of an inhibitory receptor that utilises ITIM signalling which is triggered by a component of the extracellular matrix[202]. Collagen binding to LAIR1 has been shown to inhibit the activity of a wide range of immune cells, including T cells, B cells, and NK cells[203].

1.5.3 KIR Family

The KIR, killer-cell immunoglobulin-like receptor, family is another member of the Ig superfamily, expressed on both the NK cell surface[204], and a small subset of T cells[205]. 14 genes encode members of the family[206], 7 of which are activatory, and 7 inhibitory. Each member contains either 2 or 3 Ig-like domains in the extracellular region of protein, with a single transmembrane domain connecting this to either a long cytoplasmic tail containing 1 or 2 inhibitory ITIMs, or a short cytoplasmic tail indicative of the activatory members of the family[204]. These do not directly contain ITAMs, but instead a conserved lysine residue allows for association with the adaptor protein DAP12, which itself contains 2 ITAMs[207].

Members of the family are named to indicate both the length of the extracellular domain, with either "2D" or "3D" used to denote 2 or 3 domains respectively, and whether they contain the long, inhibitory, cytoplasmic tail, or the short, activatory alternative, with L or S used respectively.

While the ligands for some members of the family, such as KIR2DL5 and KIR2DS3 are still unknown, the majority have been shown to recognise a wide range of MHC class I molecules[208]. Structural studies have revealed that, much like the interaction between MHC complexes and LILRB1, binding to KIR receptors is at the hinge region between two Ig-like domains **(Figure 1.8)**[209,210]. For those receptors with 3 Ig-like domains, this binding is at the hinge between the two membrane proximal domains[211].

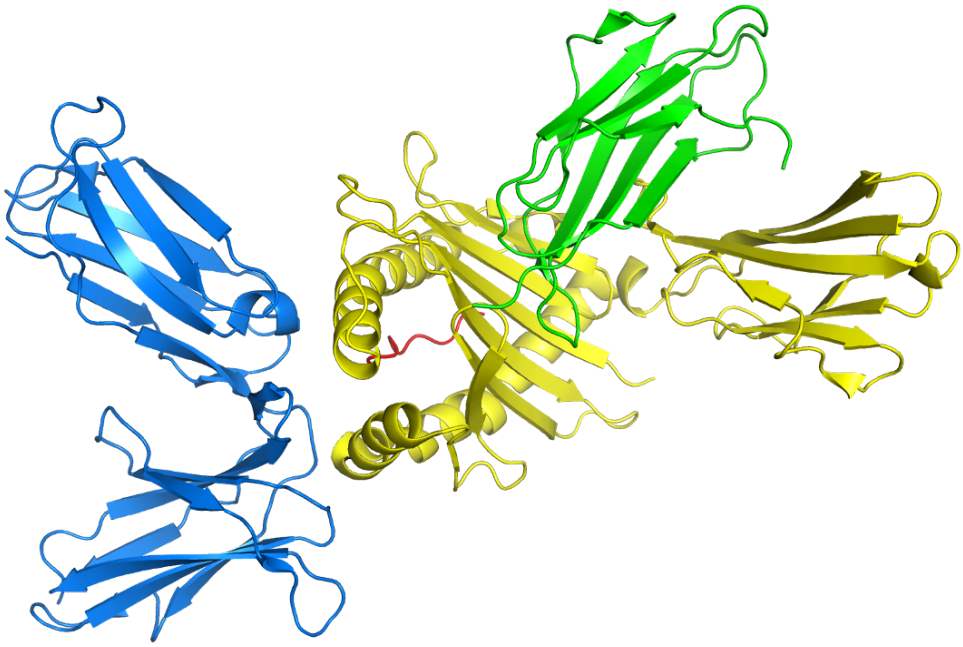


Figure 1.8: A Structure of HLA-Cw4 Bound to KIR2DL1 A 2.80 Å structure of the complex between the 2 extracellular domains of KIR2DL1 (blue), with HLA-Cw4, where the α-chain is in yellow, the β_2M chain in green, and the bound peptide is in red (PDB Code: 1IM9[210]).

In contrast to binding to LILRB1, KIR-binding by MHC class I molecules is not mediated by the β_2M, and instead the receptor binds to one end of the peptide-binding groove of the α-chain of the complex, forming interactions with both α1- and α2-domain helices**(Figure 1.9)**[210].

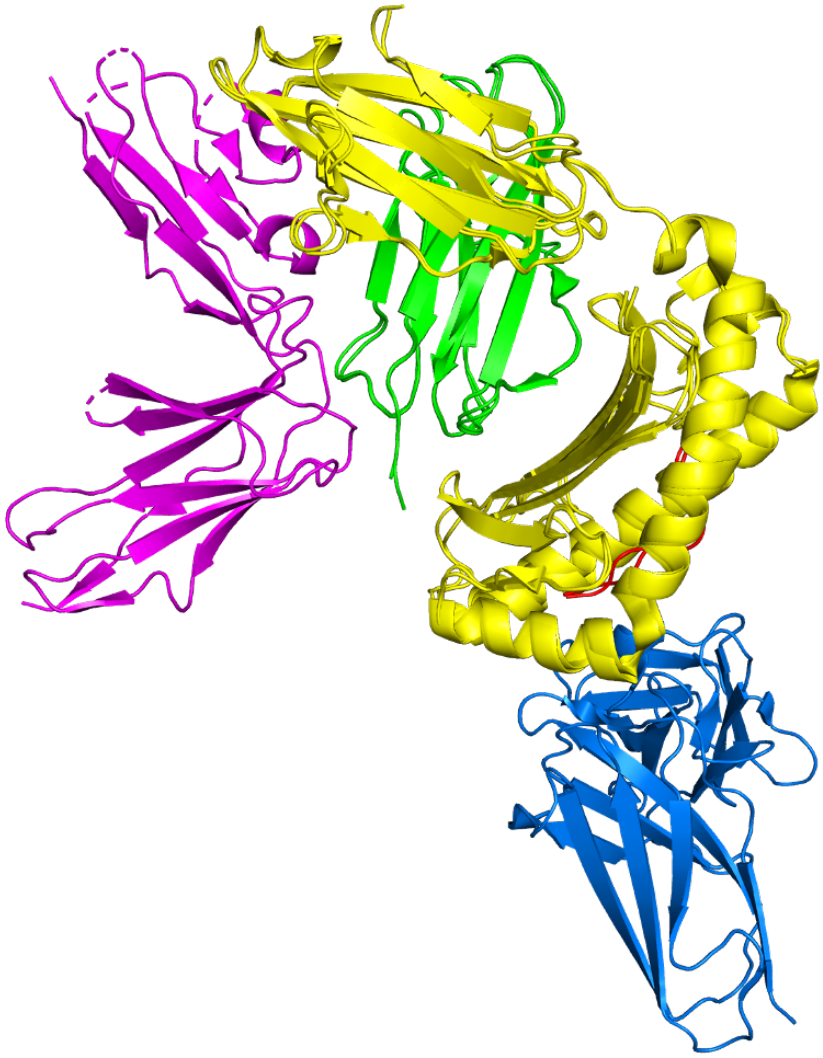


Figure 1.9: Comparison of LILRB1 and KIR2DL1 Binding to MHC Class I Structures of LILRB1 bound to HLA-A2 (PDB Code: 6EWA[194]) and KIR2DL1 bound to HLA-Cw4 (PDB Code: 1IM9[210]), aligned using the α-chain of each HLA complex. The α-chain is in yellow, the β_2M chain in green, and the bound peptide is in red for each HLA complex, LILRB1 is in magenta, and KIR2DL1 is in blue.

1.6 Project Aims

The RIFINs form the largest family of variant surface antigens expressed by *P. falciparum*, the parasite responsible for the most common and most deadly form of malaria, yet relatively little is known about how they function. There is no structural information available for the family, and little is known about how exactly they are

able to modulate the immune response through interactions with inhibitory immune receptors such as LILRB1 or LAIR1.

The first aim of this D.Phil. project, therefore, was to determine the structure of a RIFIN in complex with an inhibitory immune receptor. This would give insight both into structural conservation across the RIFIN family, through revealing the fold of the protein, as well as providing a molecular mechanism of binding to the receptors, which would significantly enhance our understanding of how the family functions, and open up more avenues for further study in the future.

In addition to this, little is known again about the PIR family, the largest multigene family across a wide range of *Plasmodium* species, including even about the membership of the family. The second aim of this project was to determine the structure of a PIR protein from *P. chabaudi*. This could then be used to assess structural conservation across putative PIRs from a range of species, and determine whether the VSA families from *P. falciparum*, including the RIFINs, could be considered as PIR proteins. This would also validate whether the rodent-infective *Plasmodium* species could be used as an animal model for studying the PIRs from human-infective *Plasmodium* species, including the VIRs from *P. vivax*, and potentially the RIFINs from *P. falciparum*.

Chapter 2

Materials and Methods

2.1 Cloning

2.1.1 Vectors

All constructs were purchased as GeneArt DNA strings (ThermoFisher). For expression in bacteria, these were cloned into the pET15b vector[212] **(Figure 2.1)**, giving an N-terminal hexahistidine-tag (His-tag) which can be removed by cleavage using thrombin. For expression in mammalian cell lines, constructs were cloned into the pHLsec vector[213] **(Figure 2.2)**, giving a C-terminal His-tag, and for expression in S2 cells, they were cloned into the pExpreS^2-1 vector (ExpreS^{2}ion Biotechnologies) **(Figure 2.3)**, giving an N-terminal His-tag.

2.1.2 Polymerase Chain Reaction

All polymerase chain reactions (PCR) for cloning were carried out using Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher), using an MJ Mini Personal Thermal Cycler (Bio-Rad, USA). Reactions were carried out per the manufacturers instructions, typically consisting of an initial 10 s 98°C denaturation step, followed by 30 cycles of denaturation for 1 s at 98°C, primer annealing at around 60°C, and extension at 72°C for 15 s/kb. A final extension step of 72°C for 1 minute was then carried out. All primers used for cloning are described in **Table A.1**.



Figure 2.1: pET15b Vector Map

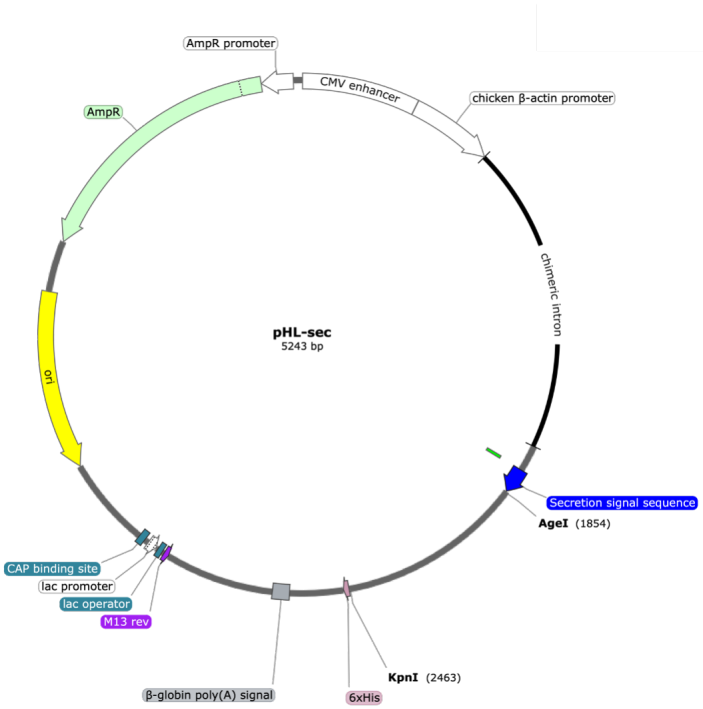


Figure 2.2: pHLsec Vector Map

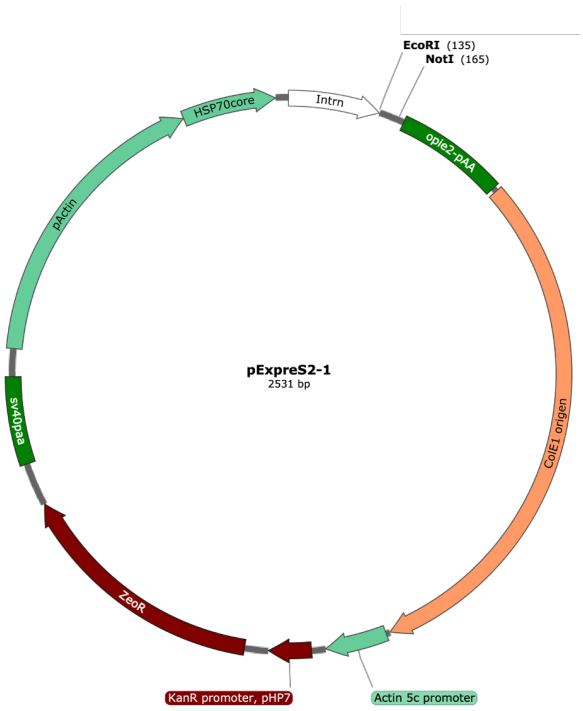


Figure 2.3: pExpreS^2-1 Vector Map

2.1.3 Restriction Digestion

All restriction digestion reactions were performed using enzymes from New England Biolabs. For cloning into pET15b, the enzymes NdeI and BamHI were typically used. For pHLsec, NdeI and KpnI were used, and for pExpreS^2-1, EcoRI and NotI were used. For each reaction, 1 μL of enzyme was added to 17 μL of DNA, along with 2 μL of CutSmart Buffer (NEB). Reactions were incubated for a minimum of 1 hour, at 37°C. Digested DNA strings were then purified using a PCR Purification Kit (QIAGEN), while digested vector was purified by agarose gel electrophoresis.

2.1.4 Agarose Gel Electrophoresis

1% w/v agarose gels were made using TAE buffer (40 mM Tris pH 8.0, 20 mM Acetic Acid, 1 mM ethylenediaminetetraacetic acid (EDTA)), with SYBR Safe (Invitrogen) added at a 1 in 10000 dilution. Gel Loading Dye, Purple (6X) (NEB) was added to samples prior to loading, and gels were run for 1 hour at 100 V. Resulting bands were

visualised using a G:Box Chemi-XRQ imaging system (Syngene). To isolate DNA, bands were cut and subsequently purified using a Monarch DNA Gel Extraction Kit (NEB).

2.1.5 Ligation

Following restriction digestion, constructs were ligated into digested vector using T4 DNA Ligase (NEB). For each reaction, 1 μL of ligase and 2 μL of 10x DNA T4 Ligase Buffer (NEB) were added to 10 μL of vector and 7 μL of insert. Reactions were incubated at 25°C for at least 1 hour, followed by a 10 minute heat inactivation at 65°C.

2.1.6 Transformation and Sequencing

50 μL of cells were used for each transformation. For cloning, Subcloning Efficiency DH5 Competent Cells (Invitrogen) were used, and for protein expression SHuffle T7 Competent Cells (NEB) were typically used. After the cells were thawed on ice, 2 μL of ligated DNA was added. After a 30 minute incubation, the cells were heat shocked for 45 seconds at 42°C, and then incubated on ice for 2 minutes. 150 μL of Lysogeny broth (LB) medium was added to the cells, which were then incubated for 45 minutes in a shaking incubator at 37°C. The cells were then plated on an agar plate containing either ampicillin at 100 μg/mL for constructs using the pET15b or pHLsec vectors, or zeocin at 10 μg/mL for constructs using the pExpreS^2-1 vector. Plates were incubated overnight in a static incubator at 37°C. Individual bacterial colonies were picked and added to 10 mL of LB medium containing the appropriate antibiotic, and grown overnight in a shaking incubator at 37°C. DNA was extracted using a QIAprep Spin Miniprep kit (QIAGEN), and Sanger sequencing carried out by Source Bioscience.

2.1.7 Site Directed Mutagenesis

Site directed mutagenesis was carried out using complementary primers which flanked the residues to be mutated by 15 bases in each direction. The vector was amplified by using these primers, as described in section 2.1.2. The PCR product was then treated with DpnI (NEB) overnight at 16°C to digest the methylated template DNA, and then transformed into XL-10 Gold Ultracompetent cells (Agilent Technologies). Following an overnight incubation at 37°C, colonies were picked and grown as described in section 2.1.6. DNA was then isolated using a QIAprep Spin Miniprep kit (QIAGEN), and Sanger sequencing carried out by Source Bioscience.

2.2 Protein Expression and Purification

2.2.1 *E. coli* Expression

2.2.1.1 Expression Trials

For expression in *E. coli*, constructs were transformed into an expression strain of the bacteria, typically SHuffle T7 Competent *E. coli* (NEB). For expression trials, 10 mL of LB medium containing 100 μg/mL Ampicillin was inoculated with the transformed bacteria, and incubated at 37°C overnight in a shaking incubator. 100 μL of this culture was then added to 10 mL of fresh LB, and this culture incubated until an optical density at 280 nm (OD_{280}) of 0.6 was reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a concentration of 1 mM, and the colony incubated either for 3 hours or overnight, typically at either 25°C or 16°C. Cell pellets were then harvested by centrifugation at 5000 x g for 10 minutes. These pellets were resuspended in 1 mL of a lysis buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.05% v/v TRITON X-100, 10% v/v glycerol, and the cells lysed by sonication. Samples were then taken for analysis by sodium

dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), representing total protein expression, both soluble and insoluble. The lysed cells were then centrifuged at 17000 x *g* for 10 minutes, and samples of the supernatants taken for SDS-PAGE analysis, representing soluble protein expression.

2.2.1.2 Large Scale Expression

For large scale protein purification from *E. coli*, 240 mL of LB medium containing 100 μg/mL Ampicillin was inoculated with the transformed bacteria, and incubated overnight at 37°C. 20 mL of this culture was then added to 1 L of 2 x Yeast-Tryptone Broth (2xYT, Melford), with typically either 6 or 12 L of culture used. This culture was incubated at 37°C until an OD_{280} of 0.6 was reached, and IPTG was then added to 1 mM. After incubation, usually overnight at 25°C, the cells were centrifuged for 15 minutes at 4500 x *g*. Cell pellets were resuspended to a total volume of either 80 mL for 6 L of culture, or 160 mL for 12 L of culture, using lysis buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.05% v/v TRITON X-100, 10% v/v glycerol. The cells were then lysed by sonication, and centrifuged for 30 minutes at 80000 x *g*. For soluble proteins, the protein could then be purified from the supernatant by Ni-NTA affinity purification, followed by size exclusion chromatography. For insoluble protein, an alternate protocol was used, described in section 2.2.1.6, whereby the protein was purified and refolded from the pellet.

2.2.1.3 Sonication

Bacterial cell lysis was carried out using a Misonix S4000 (Qsonica) sonicator. For small samples, such as in expression trials, resuspended cells were lysed at 10 W, in 1 second bursts, for a total of 5 seconds. For large scale protein purification, cells were lysed at 70 W in 3 second bursts, for a total of 3 minutes, with 7 second increments between each burst.

2.2.1.4 Nickel-NTA Affinity Purification

His-tagged proteins were purified using immobilised metal affinity chromatography, using nickel-nitrilotriacetic acid (Ni-NTA) agarose (QIAGEN). The beads were first equilibrated with 20 column volumes of buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole. The protein was then flowed over the beads, which were then washed with at least 20 column volumes of the equilibration buffer. The protein was then eluted using the same buffer with a higher imidazole concentration of 300 mM.

2.2.1.5 Size Exclusion Chromatography

Following Ni-NTA purification, proteins were purified further by size exclusion chromatography (SEC). The eluted protein was concentrated to 0.5 mL using Amicon Ultra-15 Centrifugal Filters (Millipore), using either a 3 kDa or a 10 kDa molecular weight cut-off depending on the size of the protein. The protein was then purified using a Superdex 75 Increase 10/300 GL or a HiLoad 16/600 Superdex 75 prep grade (GE Healthcare), depending on the expected yield of protein, on either an *Ä*kta Pure or *Ä*kta Purifier system (GE Healthcare), using buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl. 0.25 mL fractions were collected if a Superdex 75 Increase 10/300 GL was used, and 1 mL fractions if a HiLoad 16/600 Superdex 75 prep grade was used. Fractions within a peak of absorbance at 280 nm (A_{280}) were analysed by SDS-PAGE, and those containing pure protein were collected and pooled.

2.2.1.6 Insoluble Protein Purification and Refolding

Some proteins were not folded correctly during expression, and instead aggregated into insoluble inclusion bodies, which are found in the pellet after cell lysis and centrifugation. To purify these proteins, the pellets were first resuspended into buffer containing 6 M guanidine, 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole.

This was then incubated at room temperature for 2 hours while stirring rapidly. Dithiothreitol (DTT), was then added to a concentration of 25 mM, followed by another 1 hour incubation. The suspension was then homogenised using a Dounce homogeniser, and centrifuged for 30 minutes at 80000 x *g*. The unfolded protein was then purified from the supernatant by Ni-NTA affinity purification, where the beads were equiibrated with 20 column volumes of buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, 6 M guanidine, and the protein eluted with 20 ml of buffer containing 20 mM Hepes pH 7.5, 500 mM NaCl, 1 M imidazole, 6 M guanidine.

The protein was then refolded by flash dilution, with the Ni-NTA elution added dropwise to 2 L of refolding buffer (1 M arginine, 100 mM Tris pH 8.5, 2 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 3 mM reduced glutathione, 0.3 mM oxidized glutathione). This was then incubated at 4°C for 48 hours, followed by concentration and buffer exchange into 20 mM Hepes pH 7.5, 150 mM NaCl by tangential flow filtration. The protein was then further purified by SEC, as previously described. For the purification of the PF3D7_0100400 variable region, both wild-type and the L264R mutant, glycerol was added to the flash dilution buffer to a concentration of 15% v/v.

2.2.1.7 Tangential Flow Filtration

Flash diluted protein during refolding, or large volumes of supernatant from mammalian cell expression, were concentrated and buffer exchanged using a Centramate tangential flow filtration (TFF) system (Pall Corp., USA), with a 5 kDa cut-off membrane. After equilibration with 2 L of buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, up to 2L of the solution containing the protein of interest was loaded into the TFF machine at a rate of up to around 20 mL/minute. This was then washed with at least 1 L of the equilibration buffer, and the protein

then eluted in a volume of around 150 mL.

2.2.2 HEK293F Cell Expression

2.2.2.1 Cell Maintenance

Human embryonic kidney 293F (HEK293F) cells were grown and maintained in Freestyle F17 Expression Medium (Gibco), supplemented with 2 mM glutamine (Gibco), 100 U penicillin and 0.1 mg/mL streptomycin (Sigma), and 1% v/v MEM Non-Essential Amino Acids Solution (Gibco). Cells were grown at a density of between 0.5 and 2.5 million cells/mL, using Corning Erlenmeyer Shaker Flasks at 37°C, 8% CO2, and 70% humidity, shaking at 125 rpm.

2.2.2.2 Transfection and Protein Purification

The HEK293F cells were transiently transfected using linear 25kDa polyethylenimine (PEI, Polysciences). A stock solution of PEI was prepared at a concentration of 1 mg/mL in buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, and this was diluted to 0.5 mg/mL with media immediately prior to use. Cells were grown to a density of 2.5 million cells/mL, at which point they were centrifuged at 300 x g for 10 minutes, and resuspended in fresh media to the same density. DNA was then added, with 3 μg of DNA added per mL of cells. This was then incubated in a shaking incubator for 5 minutes. This was followed by the addition of PEI to a concentration of 9 μg/mL of cells. Following an overnight incubation, fresh media was added to double the total culture volume, and the cells were then incubated for 4 to 6 days. Following this, the cells were centrifuged at 5000 x g for 20 minutes, and the protein purified from the supernatant by Ni-NTA affinity purification, followed by SEC.

2.2.2.3 Selenomethionine Incorporation

To incorporate selenomethionine into protein produced by HEK293F cells, the transfection protocol was modified. The cells were grown as normal to a density of 2.5 million cells/mL, and centrifuged at 300 x *g*. The cells were then resuspended in RPMI 1640 media without methionine (Gibco), that was supplemented with 25 mM Hepes pH 7.1, 2 mM glutamine (Gibco), 100 U penicillin and 0.1 mg/mL streptomycin (Sigma), 1% v/v MEM Non-Essential Amino Acids Solution (Gibco), and 1% v/v Pluronic F68 (Gibco).

Two different protocols were then trialled. In the first, the cells were immediately transfected as previously described, but following the overnight incubation, the cells were again centrifuged at 300 x *g* for 10 minutes, and resuspended in the fresh RPMI media, which had been additionally supplemented with 60 mg/L L-selenomethionine. The cells were then incubated for 4 to 6 days, and the protein purified as previously described. The overnight incubation here serves to deplete cellular methionine stores, and protein produced after this point should contain the supplemented selenomethionine.

A second protocol was trialled, where after the initial centrifugation, the cells were resuspended in the RPMI media lacking methionine, and were incubated for 12 hours prior to transfection in order to deplete cellular methionine stores. The cells were then centrifuged and resuspended in RPMI supplemented with selenomethionine, followed by transfection and protein harvesting.

2.2.3 Expi293F Cell Expression

In addition to HEK293F cells, the Expi293F cell system (ThermoFisher) was used. These cells are modified HEK293F cells, which are able to grow more quickly, and to higher densities than the standard cells. Additionally the Expi293F system commonly

gives higher protein yields when compared to HEK293F cells.

2.2.3.1 Cell Maintenance

The Expi293F cells were grown and maintained in Expi293 Expression Medium (ThermoFisher), at a density between 0.5 and 5 million cells/mL, using Corning Erlenmeyer Shaker Flasks at 37°C, 8% CO2, and 70% humidity, shaking at 125 rpm.

2.2.3.2 Transfection

These cells were transiently transfected using the ExpiFectamine 293 Transfection Kit (ThermoFisher) after reaching a density of between 3 and 4 million cells/mL. For every 100 mL of cells to be transfected, 100 μg of DNA was diluted into 5 ml of Opti-MEM I Reduced Serum Medium (Gibco), and 267 μL of ExpiFectamine 293 Reagent was diluted separately into 5 ml of the same media. After a 5 minute incubation, the DNA and ExpiFectamine solutions were mixed, and incubated for a further 20 minutes at room temperature. This mixture was then added to the cells, which were then incubated for 16 to 18 hours. Following this, 500 μL of ExpiFectamine 293 Transfection Enhancer 1 and 5 ml of ExpiFectamine 293 Transfection Enhancer 2 were added to each 100 ml cell culture. The cells were then incubated for 4 to 6 days, followed by centrifugation at 5000 x g for 20 minutes, with the protein then purified from the supernatant using Ni-NTA affinity purification and SEC.

2.2.4 S2 Cell Expression

2.2.4.1 Cell Maintenance

Schneider 2 (S2) cells, derived from *Drosophila melanogaster*, were grown and maintained in EX-CELL 420 Serum-Free Medium for Insect Cells (Sigma-Aldrich), supplemented with 10% v/v fetal calf serum (FCS, Gibco), 100 U penicillin and 0.1

mg/mL streptomycin (Sigma), at a density between 5 and 50 million cells/mL, using Corning Erlenmeyer Shaker Flasks at 25°C. Following transfection, stable cell lines were maintained in medium lacking FCS and split to 8 million cells/mL.

2.2.4.2 Transfection

These cells were stably transfected using ExpreS^2 Insect-TRx5 Transfection Reagent (ExpreS^{2}ion Biotechnologies). The cells were first centrifuged at 300 x *g* for 3 minutes, and then resuspended at a density of 2 million cells/mL in serum-free EX-CELL 420 media. 50 μL of the transfection reagent was then added to 5 ml of cells, followed by 12.5 μg of DNA. The cells were then transferred to 25 cm^2 canted neck cell culture flasks (Corning), and incubated for 4 hours. FCS was then added to a final concentration of 12.5% v/v, and after a further 24 hour incubation zeocin was added at 1.5 mg/mL. The cells were maintained for 3 weeks, with 2.5 ml of fresh media containing zeocin added twice a week, to select for a polyclonal cell line.

2.2.5 Polyacrylamide Gel Electrophoresis

Protein samples were first mixed with a 2x loading dye (100 mM Tris pH 6.8, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), and then incubated at 95°C for 1 minute. 7 μL of each sample was loaded onto either 12% or 15% w/v polyacrylamide gels, along with 7 μL of PageRuler Prestained Protein Ladder (ThermoFisher). Gels were run using a Bio-Rad Mini- PROTEAN gel system, at 200 V for 45 minutes, with a running buffer containing 30 mM Tris pH 7.5, 192 mM glycine, and 0.1% w/v SDS. Protein bands were visualised by staining with InstantBlue Protein Gel Stain (Expedeon).

2.2.6 Thrombin Cleavage

The N-terminal His-tag on protein expressed in *E. coli* can be removed by cleavage using thrombin protease (GE Healthcare). 1 unit of thrombin was added per 100 μg of protein, and this was then incubated overnight at room temperature. The cleaved protein was then passed over Ni-NTA beads, which will bind the cleaved His-tag and any remaining uncleaved protein, and the tag-free protein was collected in the flow-through.

2.2.7 Protein Deglycosylation

Protein purified from mammalian or insect cell systems is often modified by glycosylation, which can interfere with both the function of the protein, as well as crystallisation. To remove this glycosylation, EndoHf (NEB) was added to protein produced in S2 cells at a molar ratio of 1:100, and incubated overnight at room temperature.

For protein produced in HEK293F or Expi293F systems, in order to facilitate deglycosylation, the mannosidase inhibitor kifunensine was added during transfection to a concentration of 2.2 μM. This results in high-mannose glycosylation, which can be removed by adding EndoHF (NEB) at a molar ratio of 1:100, and incubated overnight at room temperature.

2.2.8 Protein Methylation

Flexible lysine side chains can often interfere with the formation of crystal contacts, so to improve the likelihood of crystallisation these lysines can be methylated which reduces flexibility[214]. For methylation, 20 μL of 1 M dimethylamine borane complex (ABC, Sigma Aldrich) and 40 μL of 1 M formaldehyde was added to 1 mL of protein at a concentration of at least 1 mg/mL, followed by incubation in the dark at room

temperature for 2 hours. This was repeated twice, and after the third incubation, a further 10 μL of ABC was added, and the protein then incubated overnight. The methylated protein was then centrifuged at 17000 x g for 10 minutes to remove any precipitate, and then purified by SEC.

2.3 Biophysical Methods

2.3.1 Circular Dichroism

Circular dichroism (CD) measures differential absorbance of left-handed and right-handed circularly polarized light. Different protein secondary structures give characteristic spectra within the range between 190 and 260 nm, with, for example, α-helical proteins typically giving negative peaks at 208 and 222 nm, and β-sheets resulting in a negative peak at 218 nm[215].

Experiments were carried out using a Jasco J815 CD Spectrophotometer. The proteins were first desalted into 20 mM sodium phosphate pH 7.5, 150 mM NaF buffer using PD-10 columns (GE Healthcare), and were concentrated or diluted to 0.25 mg/mL. Spectra were taken between 190 and 260 nm, using a 1 mm path length, with measuremen ts taken every 0.5 nm. 10 spectra were taken for each protein and averaged, and a baseline taken using buffer alone was subtracted from this.

2.3.2 Surface Plasmon Resonance

Surface plasmon resonsance (SPR) allows for the determination of binding kinetics for interactions between proteins, through coupling one of the binding partners to a gold-coated chip. The other component of the interaction is then flowed over this in solution, and the change in mass on the chip upon binding is measured through observation of the change in critical angle where plasmons are produced, which is dependent on the local environment of the chip surface, and so changes upon protein

binding[216].

SPR experiments were carried out using a Biacore T200 instrument (GE Healthcare). For LILRB1-binding experiments, the receptor was coupled onto a CM5 sensor chip (GE Healthcare) via amine coupling, to a total response of around 300 RU for the full-length receptor, or a correspondingly lower response for smaller LILRB1 constructs with lower molecular weights. Initial binding tests were carried out where a RIFIN construct, typically at 8 μM was injected over the chip for 240 s with a flow rate of 30 μL/min, followed by a 240 s dissociation time, and regeneration by injection of 10 mM Glycine pH 2.5 for 10 s. Dilution series were then injected over the chip, using successive two-fold dilutions of the RIFIN.

For determination of binding kinetics for the KIR-binding RIFIN, the RIFIN was coupled onto a CM5 sensor chip via amine coupling to a total response of around 150 RU. Two-fold dilution series of KIR2DL1 and KIR2DS1 were then injected over the chip for 240 s with a flow rate of 30 μL/min, followed by a 300 s dissociation time, and regeneration by injection of 10 mM Glycine pH 2.5 for 10 s.

For HLA-Cw4 binding experiments, Pierce Recombinant Protein A/G (Thermo Scientific) was coupled onto all flowpaths of a CM5 sensor chip via amine coupling to a total response of around 1500 RU. For each cycle, Fc-tagged KIR2DL1 and KIR2DS1 were immobilised onto this chip to a response of around 300 RU, and HLA-Cw4 flowed over for 240 s with a flow rate of 30 μL/min, followed by a 300 s dissociation time, and regeneration by injection of 10 mM Glycine pH 2.5 for 10 s.

For all experiments, a running buffer containing 20 mM Hepes pH 7.5, 300 mM NaCl, 0.005% v/v Tween-20 was used, and protein was desalted into this buffer using PD-10 columns (GE Healthcare) prior to injection over the chip. The response measured on a blank flow path, where no protein was immobilised, was subtracted from the experimental sensorgrams. Kinetic values were determined by fitting, using

BIAevaluation software (GE Healthcare), to a global 1:1 interaction model, allowing for the determination of the association rate constant (K_{on}), dissociation rate constant (K_{off}), and the affinity (K_d).

2.4 X-Ray Crystallography

To determine protein structures through X-ray crystallography, an incident beam of X-rays is passed through a protein, being diffracted by the electrons within the sample. A single molecule is a very weak diffractor of X-rays, and so a protein crystal is used, where the protein is organised in a regular and repeating lattice, which amplifies the signal to be collected as diffraction spots on a detector. The calculated position and intensity of each of these spots provides information about frequency and amplitude, respectively, and when this is combined with phase information, from either experimental phasing or molecular replacement, a three-dimensional map of the electron density within the crystal can be produced.

2.4.1 Crystallisation

To produce initial protein crystals, crystallisation trials were carried out using sitting-drop vapour diffusion in MRC 2-well crystallisation plates (Swissci, Hampton Research), with reservoirs of 50 μL per well, and drops consisting of 100 nL each of well solution and protein. Commercial sparse-matrix screens of well solutions (Molecular Dimensions) were used, and plates set up using a mosquito (TTP Labtech). The aim of these trials is to allow water to evaporate from the drops, increasing the concentration of protein and precipitant until supersaturation occurs, which can allow for crystal growth in the presence of nucleation sites. The broad commercial screens used contain a wide range of mixtures of precipitants, salts, buffers, and other additives which are known to produce conditions in which proteins are able to crystallise. Typically initial crystallisation screens were carried out using the

MIDAS, Morpheus, JCSG+, ProPlex, MemGold, and PACT screens (Molecular Dimensions).

To optimise the size or quality of crystals, optimisation plates were often set up. This usually involved setting up grid screens around an initial crystallisation condition, where precipitant and protein concentration gradients were used. In addition to this, a commercial 96-well additive screen, Silver Bullets (Hampton Research) was often used, where 50 nL of additive was added to each well of a plate as a way of modifying the crystallisation condition to attempt to improve crystal quality.

Crystal seeding was also used to attempt to resolve the issue of overnucleation. Here, initial crystals were harvested and broken up through vortexing with a Seed-Bead (Hampton Research), and the resulting crystal seeds were added to a new plate to provide nucleation sites for crystal growth.

2.4.2 Crystal Harvesting

To prevent damage caused to crystals through exposure to ionising X-ray radiation which can reduce the quality of diffraction through the production of free radicals which damage crystal contacts, data was collected at 100 K. Crystals are, therefore, cryocooled and stored in liquid nitrogen prior to data collection. To protect the crystals during the cooling process, a cryoprotectant is added to the drop prior to harvesting, excluding crystals grown in a condition which already contains a cryoprotectant. Those trialled included glycerol, 2-methyl-2,4-pentanediol (MPD), polyethylene glycol 400 (PEG400), and ethylene glycol, all added to the drop to a concentration of 25% v/v in the first instance.

2.4.3 Data Collection

Data were collected at synchtrotrons including Diamond Light Source, SOLEIL, and ESRF. Initial diffraction data was collected at 3 angles differing by 45° as an

initial measure of the diffraction quality of the crystal, and the resulting diffraction data autoprocessed by programmes such as Xia2, giving the spacegroup and resolution to which the crystal diffracted. Full datasets were collected for any crystal which diffracted to a resolution higher than around 4 Å, with data collected using 0.1° oscillations, typically across a total of 180°.

2.4.4 Data Processing

Processing of the data begins with indexing, to find the diffraction spots and determine the unit cell dimensions of the crystal lattice. For the RIFIN structure, this indexing was carried out using DIALS[217], while for the CIR structure, XDS was used[218]. The intensity of the reflections (I) and the background noise (σ/I) are then determined through integration of the raw pixel intensities. The quality of the data at this stage can be assessed through the signal-to-noise ratio, I/σ(I). As intensities drop at higher resolutions, the signal-to-noise ratio drops accordingly, and so this can be used as a measure of the usable high resolution limit of the dataset, where a cut-off of 2 is generally used.

The data is then scaled, to account for variations in diffraction power, intensity of the x-ray beam and the length of the x-ray path as the crystal is rotated. This scaling was carried out using Aimless[219] for the RIFIN dataset, and XSCALE[220] for the CIR dataset. Multiple observations of the same reflection are then merged into an average intensity.

The quality of the merged data can be assessed through the linear merging residual index, R_{merge}:

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)} \tag{2.1}$$

Here, N is the number of redundant observations for reflection *hkl*, and is the

average intensity of each reflection. This is, however, an imperfect measure of quality, as R_{merge} will result in higher R-factors, indicating poorer data quality, when the multiplicity, the number of measurements per unique reflection, is high. A modified version of this index is used instead to rectify this, R_{pim}, the precision-indicating R-factor. This incorporates an extra term, $\sqrt{\frac{1}{N-1}}$, to take multiplicity and the increasing precision that comes with more observations into account:

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)} \tag{2.2}$$

2.4.5 Phase Determination

Following scaling and merging, a complete hkl dataset is outputted with intensities, but the phase information required for calculation of electron density is lost during data collection. This calculation utilises the sum of the structure factors and phases of all diffraction points in reciprocal space, and relates this to the electron density as follows:

$$p(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| e^{-2\pi i[hx+ky+lz-\phi(hkl)]} \tag{2.3}$$

This phase information is typically recovered using one of two methods, molecular replacement or experimental phasing.

2.4.5.1 Molecular Replacement

Molecular replacement, as carried out in solving the structure of the RIFIN:LILRB1 complex, uses previously solved structures of related proteins as starting points for phase determination. A known search model, which usually has a sequence identity of at least 30%, is used, along with the known space group and estimated solvent content, calculated using the Matthews coefficient[221]. Phaser MR[222] then utilises

rotation and translation functions to attempt to fit the search model within the unit cell, determining the likelihood of each given placement compared to a model of random atoms.

A large number of possible solutions are produced, which are analysed through a log-likelihood of gain (LLG) score, along with Z-scores calculated through comparison of the LLG values with a random set of rotation and translation functions. The Z-score is defined as the number of standard deviations above the mean for the LLG values. Solutions with a Z-score of above 8 are considered likely to be correct, and these were then assessed in Coot[223] to confirm that the search model corresponds well to the calculated electron density. If true, this solution was then used for further model building and refinement.

2.4.5.2 Experimental Phasing

If a related search model is not available, as was the case for the CIR, where no structure of homologous proteins were available, phases can also be determined experimentally. During experimental phasing, heavy atoms, such as halides or heavy metals, introduced into the protein crystal can give rise to a different set of structure factor amplitudes relative to the reference amplitudes derived from the original crystal. This difference can be used to locate the heavy atom or anomalous scattering atoms, thereby solving the substructure and providing initial phases for structure determination. This method of exploiting differences between crystals is termed isomorphous replacement.

Heavy atoms can be introduced into the protein in a number of ways. These include soaking crystals in solutions containing, for example, heavy metal compounds, which interact with protein side chains. The crystals are then washed to remove unbound atoms, and then cryocooled for data collection. Alternatively, proteins can be labelled through incorporation of selenomethionine during expression, thereby introducing

selenium into the protein.

Alternatively, anomalous phasing experiments can be carried out, where heavy atoms, such as heavy metals, selenium introduced in selenomethionine, or the sulfur atoms found in native cysteine and methionine residues, can be exploited. Anomalous scattering by these atoms results in differences in magnitude between Bijvoet pairs, which are centrosymmetrically related pairs of structure factors. This allows phase determination from a single crystal, including native crystals through utilising internal sulfur atoms. As the anomalous scattering from sulfur is weak at the wavelengths typically used for data collection[224], a longer wavelength is used for these experiments, typically at least 2.7 Å. This method was used to solve the CIR structure, with SHELXD[225] used for phase determination.

2.4.6 Model Building and Refinement

Following experimental phasing, an empty electron density map is produced as, unlike when molecular replacement is used, there is no initial protein model from which to build. An initial model is therefore typically automatically produced using programmes such as Bucaneer, which traces a polyalanine chain into the density[226]. This polyalanine model, along with the initial search model output by molecular replacement, can then be rebuilt using Coot[223], with regions poorly fitting within the density removed, and empty regions of density added to. Disulphide bonds and bulky side chains, such as tryptophan side chains, can be used to position the protein sequence onto the model.

Rounds of rebuilding in Coot are combined with reciprical space refinement of the model in BUSTER[227], which utilises maximal likelihood methods to refine the model, whilst minimising deviation of the calculated data from the model, from the initial experimental data. The quality of data following refinement is measured using R-factors calculated for both the working dataset, R_{work}, and for a subset of the

experimental data which was reserved for validation and so not used in refinement, R_{free}. These R-factors are calculated as follows:

$$R = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} F_{hkl}^{obs}} \tag{2.4}$$

These R-factors should both improve and converge as the model improves, with typical final values of less than 0.3 for a 3 Å dataset indicating a good quality model.

2.5 Bioinformatics

All *Plasmodium* sequences for both bioinformatics and construct design were sourced from PlasmoDB[228], including 198 amino acid sequences from the *P. chabaudi* AS v3 genome assembly[88], 135 sequences from the *P. berghei* ANKA genome assembly[229], 1011 sequences from the *P. yoelii* 17X v3 genome assembly[170], 70 sequences from the *P. knowlesi* strain H genome assembly[158], 136 sequence s from the *P. malariae* UG01 genome assembly[159], 1495 sequences from the *P. ovale curtisi* GH01 genome assembly[159], 185 RIFIN sequences and 31 STEVOR sequences from the *P. falciparum* 3D7 genome assembly[230], 1086 sequences from the *P. vivax* P01[231], 487 sequences from the *P. reichenowi* G01 genome assembly[232] and 20 sequences from the *P. gallinaceum* 8A genome assembly. Inhibitory immune receptor sequences were sourced from UniProt[233].

For construct design, signal peptide prediction was carried out using the SignalP 4.0 server[234], and for transmembrane helix prediction the TMHMM Server V2.0 was used[235]. Protein disorder predictions were carried out using DISOPRED 3.1[236], PONDR, MFDp2[237], NORS2st[238], ESpritz[239], AUCpreD[240], SPOT[241], and Pr-DOS[242].

Sequence alignments were produced using MUSCLE v3.8.31 with default parameters[243,244], and visualised and analysed using Jalview[245]. Sequence alignment figures were produced using ESPript 3.0[167]. Entropy calculations were carried out using the Entropy-One server on the HIV Sequence Database. Sequence logos were produced using WebLogo 3[246]. Homology models were produced using structure-based threading on the Phyre2 server[247].

Evolutionary analyses were conducted in MEGA X[248,249]. The evolutionary history was inferred by using the Maximum Likelihood method and Whelan And Goldman + Freq. model[250]. The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join[251] and BioNJ[252] algorithms to a matrix of pairwise distances estimated using a JTT model[253], and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The trees were visualised and edited using FigTree v1.4.4[254].

Chapter 3

Production and Biophysical Characterisation of RIFINs

3.1 Summary

In order to carry out any structural or biophysical characterisation of RIFINs, the proteins first had to be recombinantly expressed and purified. To achieve this, expression was first trialled for a panel of RIFIN ectodomains, consisting of the region of protein between the predicted PEXEL cleavage site and transmembrane helix for each sequence. These correspond to the region of protein displayed on the surface of the infected red blood cell, after the signal sequence has been removed.

Expression of these ectodomains was trialled in a range of expression systems. Expression in the bacteria *Escherichia coli* only generated insoluble protein contained within inclusion bodies, while no correctly-folded protein was produced in either mammalian or insect cell systems.

Following these trials, attempts were instead made to produce smaller fragments of a RIFIN ectodomain, focusing on either the variable or constant region separately. The variable region of a LILRB1-binding RIFIN has previously been shown to bind to the receptor, while no binding was seen for the constant region[67].

Expression trials for the variable region of a LILRB1-binding RIFIN were again

unsuccessful in a mammalian cell system, and expression in *E. coli* was still confined to inclusion bodies. It was possible, however, to purify the aggregated protein from these inclusion bodies allowing its refolding to produce correctly folded protein which was able to bind to LILRB1, indicating it had retained its functionality. A second RIFIN variable domain, a fragment of a KIR2DL1- and KIR2DS1-binding RIFIN, could also be expressed and purified, in this case using mammalian cells. This was again able to bind to the relevant inhibitory immune receptors, indicating it was both correctly folded and functional.

In addition to this, the constant region of 4 RIFINs could be produced as soluble protein through secretion from a mammalian cell system, allowing for biophysical characterisation and crystallisation trials to be carried out for a panel of these constructs.

3.2 RIFIN Ectodomain Expression

3.2.1 Design of Ectodomain Constructs

Expression trials were carried out for a panel of RIFINs consisting of 6 shown to bind to LILRB1, and 3 shown to bind to LAIR1[67,255]. The 6 LILRB1-binding RIFINs were chosen following phylogenetic analysis of the 11 total known LILRB1-binding RIFINs, to ensure the panel chosen represented the diversity of the sequences **(Figure 3.1)**. There were only 3 known LAIR1-binding RIFINs, so all 3 were used for expression trials.

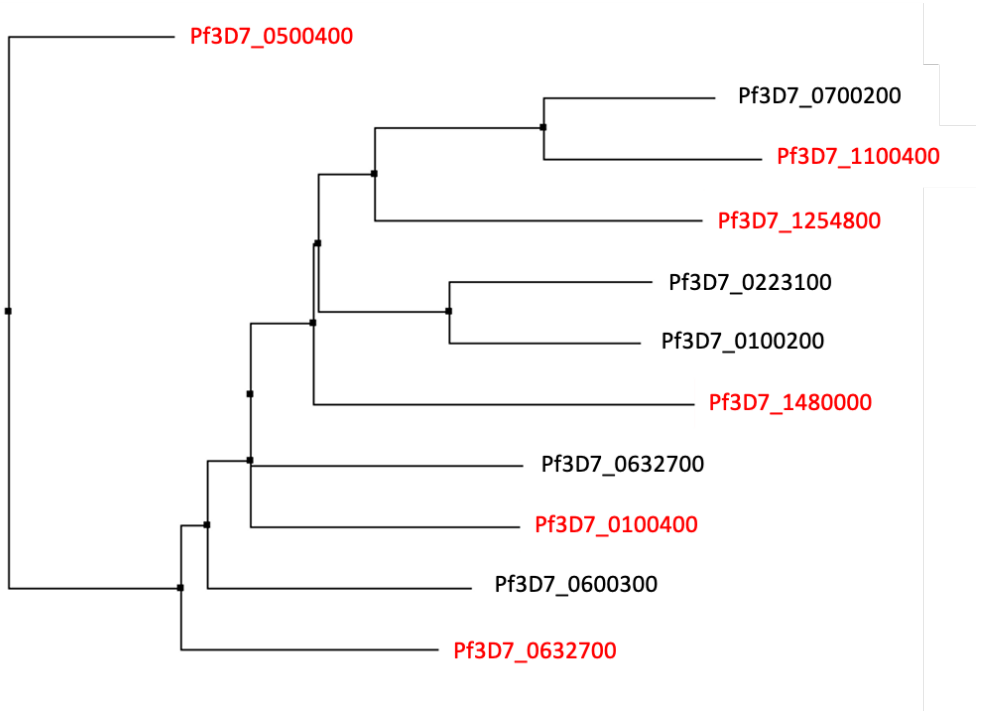


Figure 3.1: Phylogenetic Tree for All Known LILRB1-Binding RIFINs The chosen sequences are shown in red.

Constructs were designed to contain the region of each RIFIN which would be exposed on the infected red blood cell surface after processing and removal of the signal sequence. To achieve this, the consensus sequence for the PEXEL motif, RxLxE/Q/D[100], was identified within each RIFIN sequence **(Figure 3.2)**. This motif is cleaved by Plasmempsin V during the export of RIFINs to the host cell surface[256], with cleavage occurring at the C-terminus of the leucine residue, and so each construct was designed to begin at the residue following this.

In addition to this, transmembrane helix prediction was carried out using the TMHMM server[235] **(Figure 3.2)**. This allowed for the removal of the transmembrane and cytoplasmic regions of protein from the construct, so that the resultant constructs could be produced as soluble protein. The final ectodomain constructs, therefore, began directly following the PEXEL clevage site, and ended directly prior to the predicted transmembrane helix. A 6-His tag was added to each construct, to allow for purification using Ni-NTA affinity purification. These constructs ranged in molecular

weight from around 27 to 33 kDa. All protein constructs described in this thesis are listed in **Table A.1**.

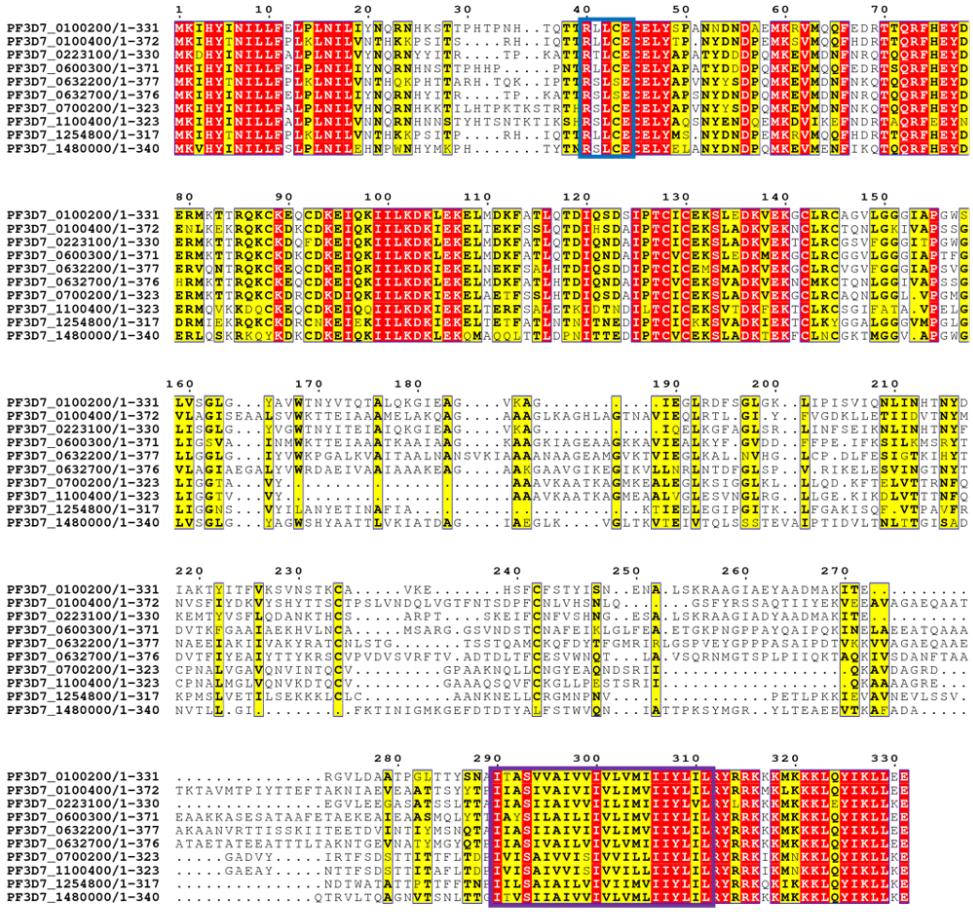


Figure 3.2: Sequence Alignment of LILRB1-Binding RIFINs The conserved PEXEL motif is highlighted in blue, and the transmembrane helices highlighted in purple. A red background for a column indicates a fully conserved residue, while a yellow background indicates conservation of residues with similar chemical properties. Figure prepared using ESPript 3.0[167].

3.2.2 Expression Trials of RIFIN Ectodomains in HEK Cells, *E. coli*, and S2 Cells

3.2.2.1 HEK293F Cell Trials

Initial expression trials were carried out using a cell line derived from human embryonic kidney cells, HEK293F cells, which has been successfully used for the expression of a number of *Plasmodium* proteins in the past[257], including members of the PfEMP1 family[258].

Expression in small scale trials was assessed through the purification of His-tagged protein from the HEK supernatant by Ni-NTA affinity, followed by SDS-PAGE analysis, and this demonstrated no appreciable expression for any of the constructs trialled **(Figure 3.3)**.

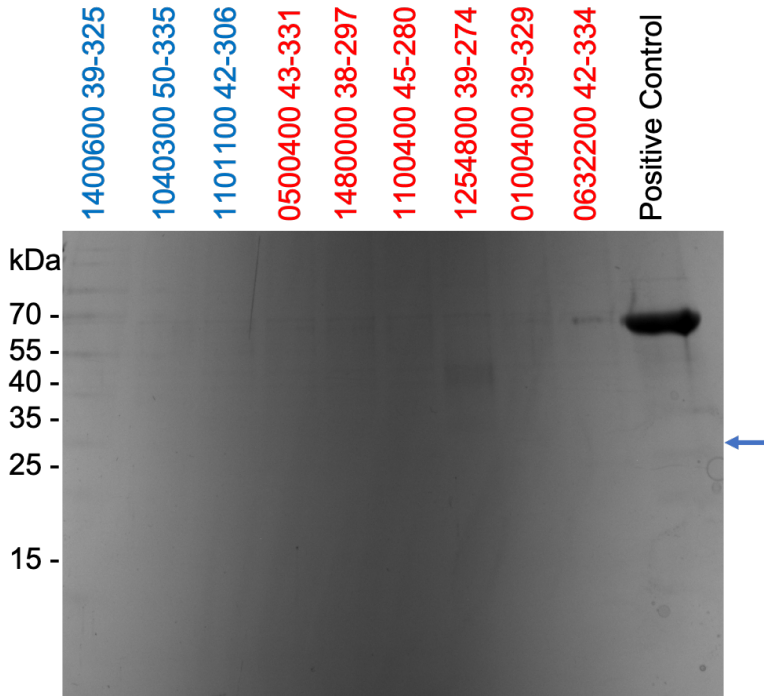


Figure 3.3: Expression Trial of RIFIN Ectodomain Constructs in HEK293F Cells SDS-PAGE gel showing no expression of 3 LAIR1-binding RIFIN (blue) and 6 LILRB1-binding (red) ectodomains by HEK293F cells. The arrow indicates the approximate expected position of bands, with constructs ranging from around 27 to 33 kDa in size. The positive control used was the CIR construct, PCHAS_0600600 1-268 with a CD4-tag.

3.2.2.2 Expi293F Cell Trials

In addition to the HEK293F cells, the Expi293F cell system was trialled for RIFIN expression. This utilises modified HEK293F cells, which are able to grow both more quickly, and to higher densities[259], commonly allowing for higher levels of recombinant protein expression.

Similar small scale expression trials were carried out in this system, which showed bands at around the expected molecular weights for 4 of the RIFINs, including 1 LAIR1-binding RIFIN, Pf3D7_1040300, the construct for which had an expected molecular weight of 32.5 kDa, and 3 LILRB1-binding RIFINs, Pf3D7_0500400, Pf3D7_1100400, and Pf3D7_1254800, which had expected molecular weights of 33.1,

27.4, and 27.9 kDa respectively **(Figure 3.4)**. However, upon scaling up expression for these three constructs, no protein could be successfully purified, suggesting that the protein seen in the trials was either unstable or incorrectly folded, and was aggregating or precipitating during purification.

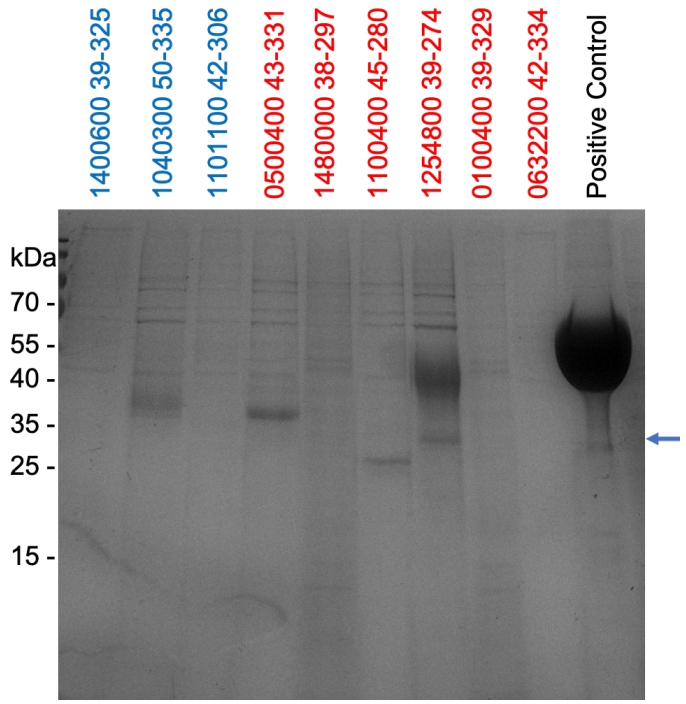


Figure 3.4: Expression Trial of RIFIN Ectodomain Constructs in Expi293F Cells SDS-PAGE gel assessing expression of 3 LAIR1-binding RIFIN (blue) and 6 LILRB1-binding (red) ectodomains by Expi293F cells. Expression at the expected molecular can be seen for 4 of the 9 RIFINs. The arrow indicates the approximate expected position of bands, with constructs ranging from around 27 to 33 kDa in size. The positive control used was the CIR construct, PCHAS_0600600 1-268 with a CD4-tag.

3.2.2.3 *E. coli* Trials

Expression trials were also carried out using the bacteria, *E. coli*. This system often allows for faster and larger scale expression of recombinant protein, but due to the relative simplicity of the host expression pathway, soluble expression of *Plasmodium* proteins can be challenging[260]. One common factor which can impede soluble protein expression is high molecular weight, and so four RIFINs with comparably small

ectodomains were chosen for trial in this system, Pf3D7_0100200, Pf3D7_0832800, Pf3D7_0832500, and Pf3D7_1040900, with expected molecular weights of 29.7, 26.7, 28.3, and 26.8 kDa respectively. These constructs again included a 6-His tag for Ni-NTA affinity purification, and also included a thrombin protease cleavage site to allow for the tag to be removed. Due to the presence of disulphide bonds in RIFINs, the SHuffle strain of *E. coli* was chosen, which has diminished cytoplasmic reduction pathways due to deletions of thioredoxin reductase (*trxB*) and glutathione reductase (*gor*), as well as an added disulphide bond isomerase, DsbC, which allows for the formation of disulphide bonds in the cytoplasm[261]. Small scale trials in this system, which included 3 different incubation temperatures of 16°, 25°, and 37°, showed no protein in the soluble fraction, although some expression could be seen in the insoluble fraction **(Figure 3.5)**.

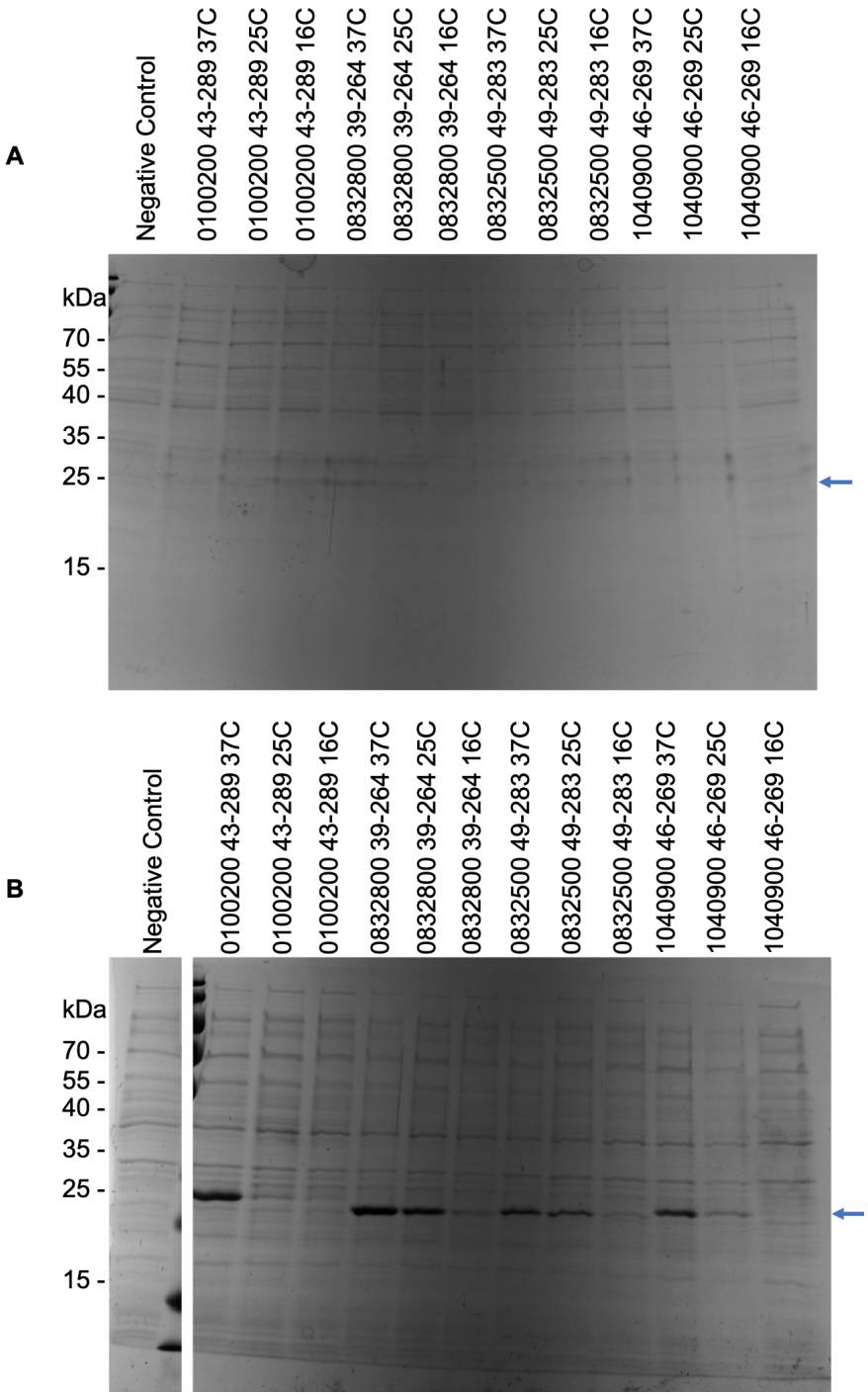


Figure 3.5: Expression Trial of RIFIN Ectodomain Constructs in *E. coli* SDS-PAGE gels assessing expression of 4 RIFIN ectodomains in the A) soluble and B) insoluble fractions from *E. coli*, at 3 different incubation temperatures. Expression can be seen in the insoluble fraction for all 4 RIFINs, with no soluble expression seen. The expected position of bands is indicated by arrows. The negative control used was the SHuffle strain of *E. coli*, with no IPTG induction prior to the overnight incubation.

3.2.2.4 S2 Cell Trials

Expression of one RIFIN ectodomain, that of Pf3D7_0100400, had previously been reported in Schneider 2 (S2) cells, an embryonic *Drosophila melanogaster* cell line[75]. Expression trials were, therefore, carried out for this along with 2 additional RIFINs, Pf3D7_1254800 and Pf3D7_1040300, which have been shown to bind LILRB1 and LAIR1 respectively[67,156]. Stable cell lines were established following transfection with each of these constructs. An N-terminal BiP signal sequence was included for each construct which allows binding of the BiP chaperone to guide secretion of the proteins from the cell[262].

Ni-NTA affinity purification from the supernatant demonstrated expression only of Pf3D7_0100400 **(Figure 3.6A)**. However, purification by size exclusion chromatography indicated that this protein was incorrectly folded. The protein eluted close to the void volume of the column which suggested a higher molecular weight than expected, indicative of incorrectly folded protein which has formed aggregates **(Figure 3.6B)**.

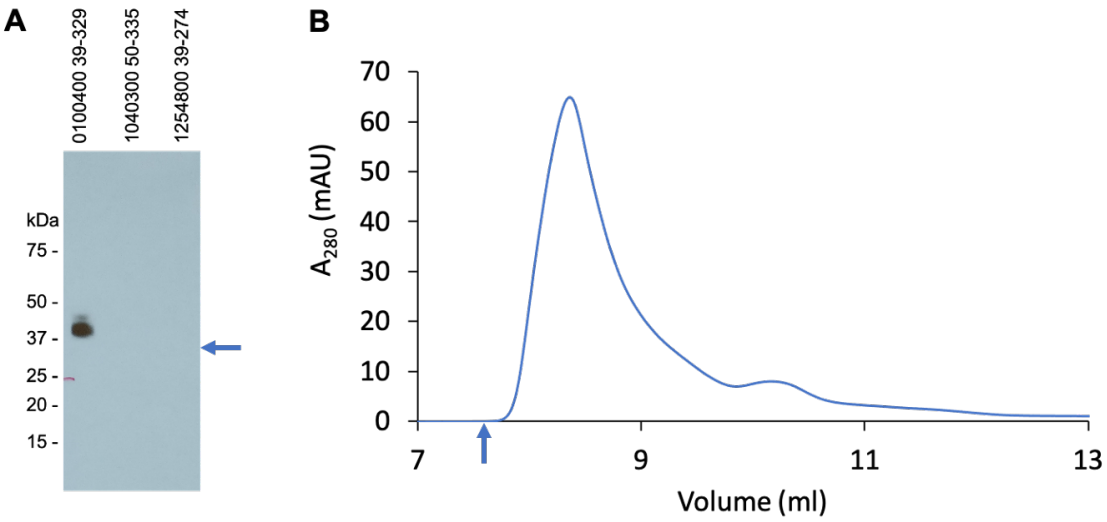


Figure 3.6: Expression Trial of RIFIN Ectodomain Constructs in S2 Cells A) SDS-PAGE gel assessing expression of 3 RIFIN ectodomains in S2 cell supernatant, with the expected position of bands indicated by an arrow, and B) SEC chromatogram for purification of Pf3D7_0100200 from S2 cell supernatant. The void volume for the Superdex 200 10/300 column used was approximately 7.5 ml, and is indicated by an arrow.

3.3 Characterisation of a LILRB1-Binding RIFIN Variable Region

Following the unsuccessful attempts to express and purify full-length RIFIN ectodomains, attempts were then made to instead produce smaller fragments. This had been previously achieved for a LILRB1-binding RIFIN, Pf3D7_1254800, where separate constructs for the constant and variable regions of the ectodomain were produced by refolding protein expressed insolubly in *E. coli* inclusion bodies[67]. Expression and purification of these fragments from *E. coli* was therefore trialled, and expression trials also carried out for both fragments in mammalian cell systems.

3.3.1 Expression and Purification

3.3.1.1 Expression Trials in Expi293F Cells

Small scale expression trials for the variable domain of Pf3D7_1254800, residues 165-274, were carried out in Expi293F cells. This showed very low levels of protein produced, although at a higher than expected molecular weight. The band seen on an SDS-PAGE gel indicated a molecular weight of around 30 kDa, more than double that of the predicted 13.4 kDa **(Figure 3.7A)**. This could be explained by the addition of extensive glycosylation during expression, and so an additional construct was produced which introduced the T261D mutation. This removed the only predicted glycosylation site within the sequence. This mutant retained the low level of expression in Expi293F cells, although the gel band on a non-reducing SDS-PAGE gel still indicated a significantly higher molecular weight than expected, at around 35 kDa **(Figure 3.7A)**, perhaps due to the formation of intermolecular disulphide bonds. In addition to this, when both the wild type and mutant constructs were purified by SEC, the peak again indicated a larger protein than expected, with no change after removing the glycosylation site **(Figure 3.7B)**. This suggested that the protein was not correctly folded, and that the increased size seen was not due to glycosylation, but instead because of the formation of multimeric aggregates.

3.3.1.2 Expression Trials in *E. coli*

Expression trials were then carried out in *E. coli*, again using the SHuffle strain to allow for disulphide bond formation. The same construct was used as for the mammalian cell trials, although a thrombin protease cleavage site was included between the RIFIN and the 6-His tag. These trials demonstrated no expression in the soluble fraction, with no band seen at the expected molecular weight of around 14.4 kDa, while some expression was seen in the insoluble fraction **(Figure 3.8)**.

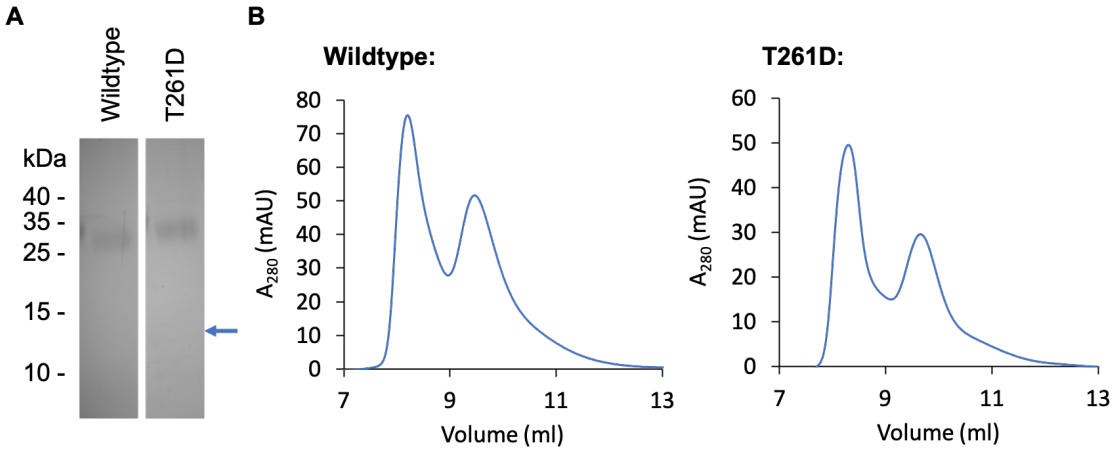


Figure 3.7: Expression Trial of a LILRB1-Binding RIFIN Variable Region in Expi293F Cells A) SDS-PAGE gels showing expression of Pf3D7_1254800 165-274 wild type and T261D by Expi293F cells. The arrow indicates the expected position of the bands at around 13.4 kDa. B) SEC chromatogram for purification of these constructs from Expi293F supernatant using an Superdex 75 10/300 column.

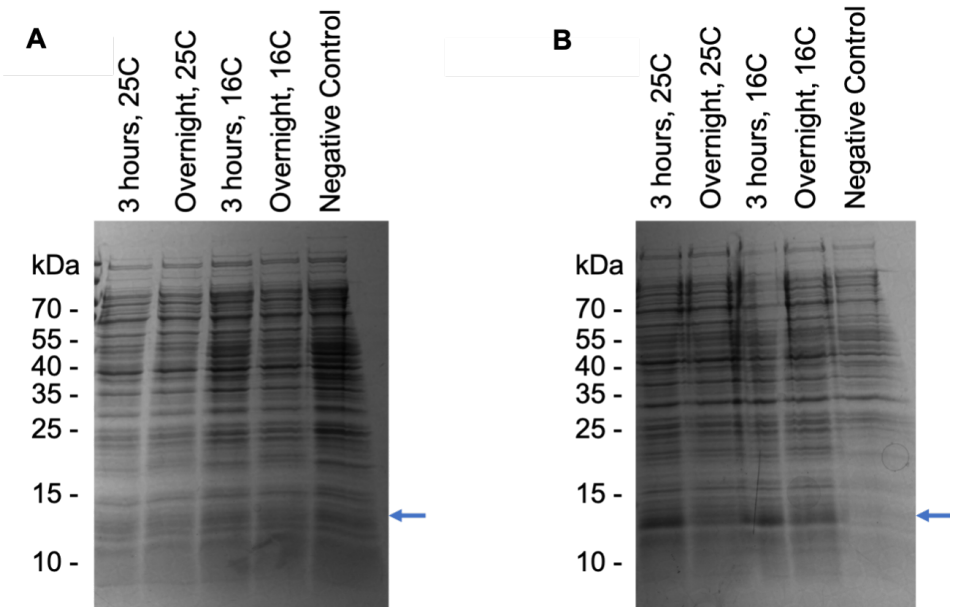


Figure 3.8: Expression Trial of a RIFIN Variable Region in *E. coli* SDS-PAGE gels assessing the expression in the A) soluble and B) insoluble fractions from *E. coli* at 2 incubation temperatures and times. The expected molecular weight of around 14.4 kDa is indicated by arrows. Expression can be seen in the insoluble fraction, but not in the soluble fraction. The negative control used was the SHuffle strain of *E. coli*, with no IPTG induction prior to the overnight incubation.

Attempts were then made to purify this variable region from inclusion bodies. These inclusion bodies are formed of aggregates of misfolded protein, and so any protein purified must be refolded in order to be functional. To achieve this, a range of refolding protocols were trialled. The first of these methods was on-column refolding, where after denaturation with 6 M guanidine, the protein is bound to Ni-NTA beads, and the guanidine concentration is slowly reduced to allow the bound protein to refold *in situ*. Flash dilution refolding was also trialled, where the denatured protein is diluted into a large volume of refolding buffer, to rapidly decrease the guanidine concentration to allow for refolding. The large decrease in protein concentration during flash dilution helps to reduce the risk of aggregation during refolding.

The most effective protocol used, as detailed in section 2.2.1.6, involved first purifying the unfolded protein by Ni-NTA affinity purification, followed by refolding by flash dilution into a buffer containing 1 M arginine, 100 mM Tris pH 8.5, 2 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 3 mM reduced glutathione, 0.3 mM oxidized glutathione. The addition of arginine to the flash dilution buffer helps to prevent protein aggregation[263], while the mixture of reduced and oxidised glutathione allows for reversible disulphide bond formation. The folded protein was then concentrated and buffer-exchanged by TFF, followed by further purification by SEC using a Superdex 75 10/300 column **(Figure 3.9)**.

3.3.2 Assessment of the Secondary Structure of the Variable Region

To confirm that the refolded variable region was correctly folded, circular dichroism spectroscopy was used to assess the secondary structure of the protein. The spectra from this was characteristic of an α-helical protein, with minima at both 208 and 222 nm, as expected from both secondary structure prediction[264], and from previous production of this protein fragment[67].

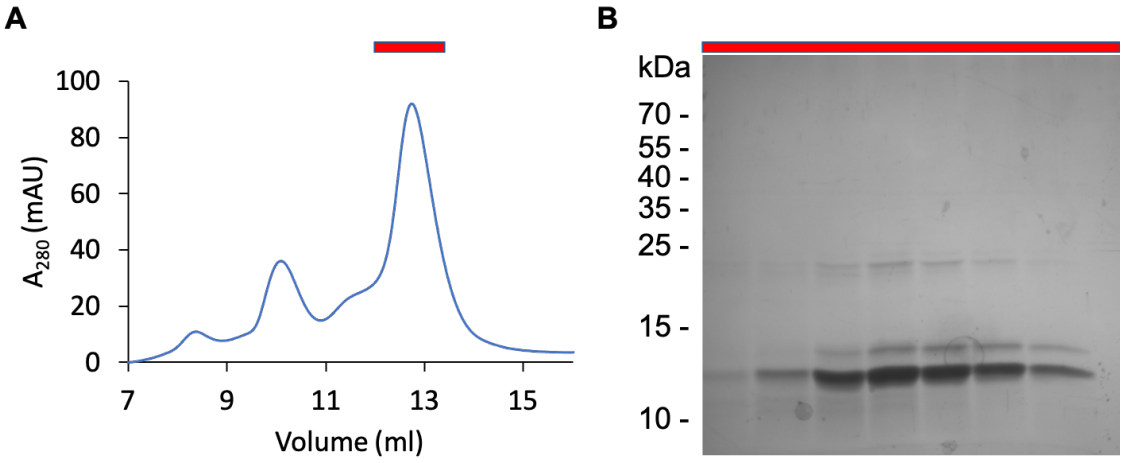


Figure 3.9: Refolding of a LILRB1-Binding RIFIN Variable Region from *E. coli* Inclusion Bodies A) SEC chromatogram showing purification of refolded Pf3D7_1254800 165-274 using a Superdex 75 10/300 column and B) SDS-PAGE gel showing fractions from the main peak, at around 13 ml, of this purification. The red bar above the SEC trace corresponds to the fractions shown in the SDS-PAGE gel.

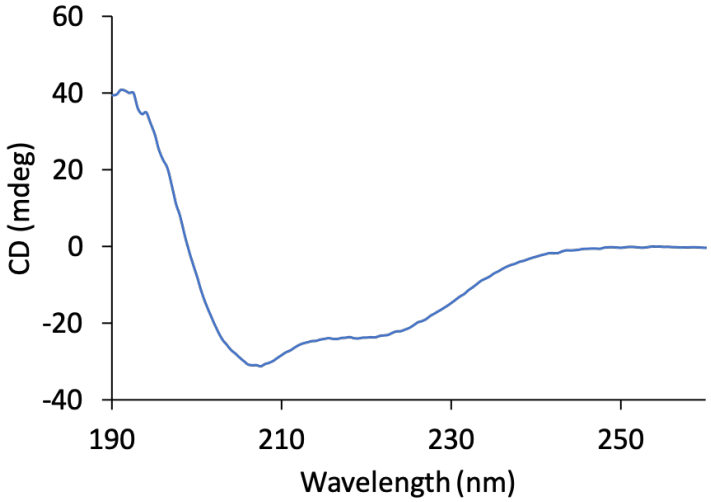


Figure 3.10: Circular Dichroism Spectrum for the Pf3D7_1254800 Variable Region

3.3.3 Expression and Purification of LILRB1

In order to test the ability of this variable region to bind LILRB1, 3 separate LILRB1 constructs were produced for expression in HEK293F cells. As described in section 1.5.1.1, the LILRB1 ectodomain consists of 4 Ig-like domains. The first of the constructs contained all 4 of these domains, with an expected molecular weight of around 45.1 kDa. Two shorter constructs were also designed, the first consisting of the 2 membrane distal domains, D1D2, and the second consisting of the 2 membrane proximal domains, D3D4, with expected molecular weights of 23.4 and 22.7 kDa respectively.

3.3.3.1 Expression in HEK293F Cells

Expression trials for the 3 LILRB1 constructs in HEK293F cells showed strong expression of both the full-length and D3D4 constructs **(Figure 3.11)**. Expression could also be seen for the D1D2 construct, but this aggregated immediately upon purification, and so could not be produced successfully using this system. The bands for the full-length and D3D4 constructs were both at a slightly higher molecular weight than expected, likely due to glycosylation.

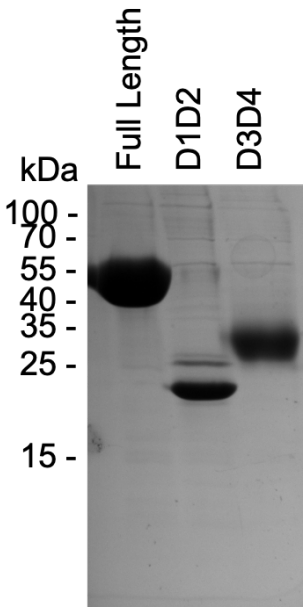


Figure 3.11: Expression Trial of LILRB1 Ectodomain Constructs in HEK293F Cells SDS-PAGE gel showing expression of 3 LILRB1 ectodomain constructs in HEK293F cells. The expected molecular weight of the 3 constructs was 45.1 kDa, 23.4 kDa, and 22.7 kDa for the full-length, D1D2, and D3D4 constructs respectively.

The full-length and D3D4 constructs could then be expressed and purified on a large scale, with further purification by SEC showing peaks at expected volumes, indicating monomeric, well-folded protein **(Figure 3.12)**.

3.3.3.2 Refolding of LILRB1 D1D2 from Insoluble Expression in *E. coli*

As the D1D2 construct could not be produced using a mammalian cell system, attempts were instead made to express this in *E. coli*. This required purification of the protein from inclusion bodies, and refolding by flash dilution using the same method utilised for the RIFIN variable region. This resulted in a clean SEC peak at the expected volume, again indicating monomeric well-folded protein **(Figure 3.13)**.

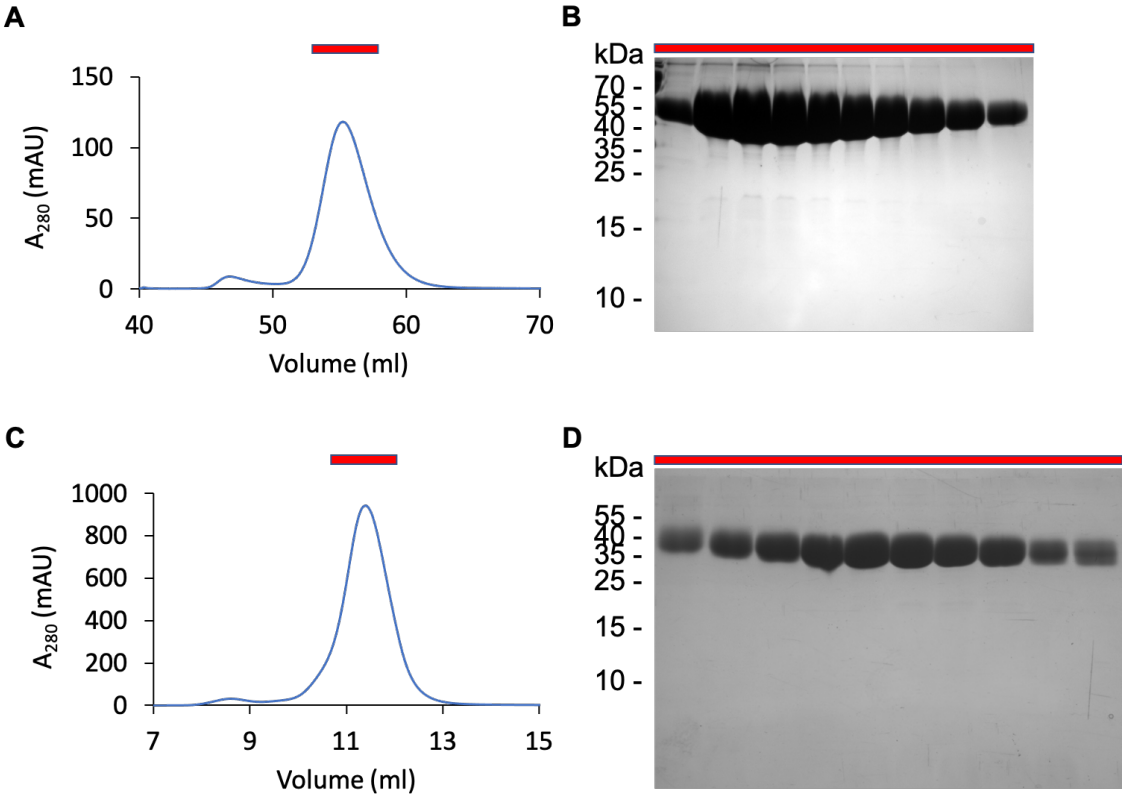


Figure 3.12: Purification of LILRB1 Full-Length and D3D4 Constructs from HEK293F Cells A) SEC chromatogram and B) SDS-PAGE gel for LILRB1 full-length purification from HEK293F cells using a Superdex 75 16/60 column. C) SEC chromatogram and D) SDS-PAGE gel for LILRB1 D3D4 purification from HEK293F cells using a Superdex 75 10/300 column. The red bar above the SEC traces correspond to the fractions shown in the SDS-PAGE gels.

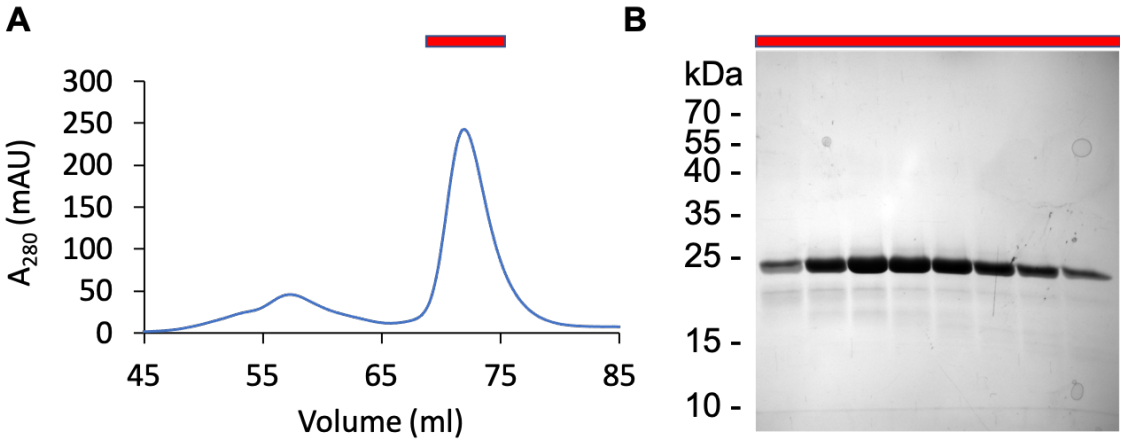


Figure 3.13: Refolding of LILRB1 D1D2 from *E. coli* Inclusion Bodies A) SEC chromatogram showing purification of refolded LILRB1 D1D2 using a Superdex 75 16/60 column, and B) SDS-PAGE gel showing fractions from the main peak, at around 72 ml, of this purification. The red bar above the SEC trace corresponds to the fractions shown in the SDS-PAGE gel.

3.3.4 Binding of the Variable Region to LILRB1

To assess the functionality of the Pf3D7_1254800 variable region, the ability of this construct to bind to LILRB1 was tested. This was achieved through two methods. The first, analytical gel filtration, involved mixing the RIFIN with the full-length LILRB1 ectodomain, and then running this over a Superdex 75 10/300 column. The two components were then analysed in the same way separately, and the resultant SEC traces were compared. The trace for LILRB1 showed a single peak at around 10.7 ml, while that for the RIFIN variable region showed a peak at around 13.8 ml. When mixed, the peak for the RIFIN at 13.8 ml disappeared, and the peak for LILRB1 shifted to around 10.3 ml **(Figure 3.14A)**. This shift indicated an increase in apparent size of LILRB1, and as the RIFIN peak disappeared, this suggested the RIFIN was successfully binding to the receptor. Analysis by SDS-PAGE, where fractions corresponding to both the RIFIN and LILRB1 peaks for all 3 traces were run, confirmed that this complex formation was occuring **(Figure 3.14B)**.

In addition to this, binding was quantitatively analysed by surface plasmon resonance

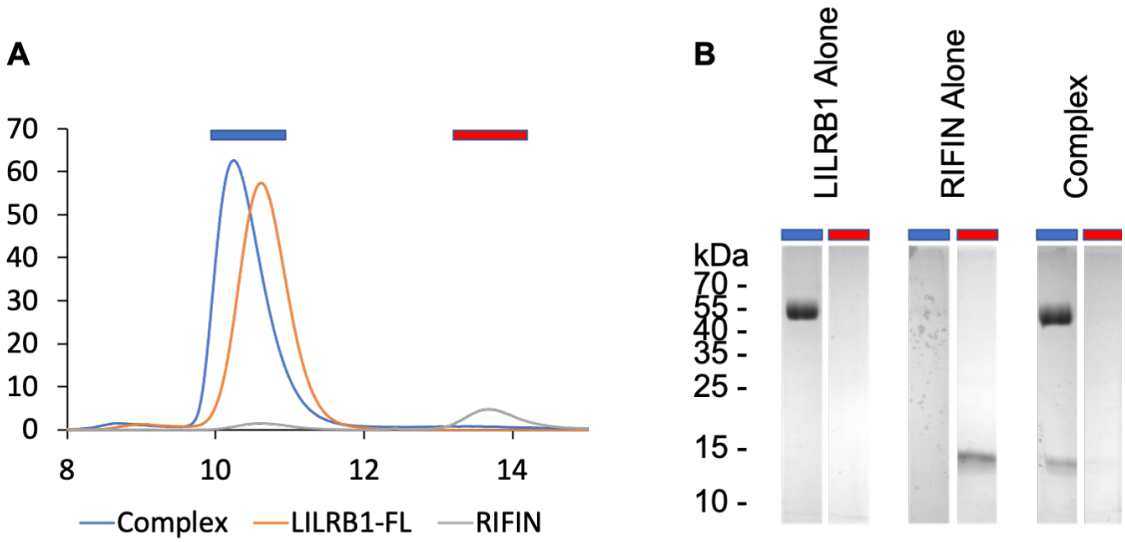


Figure 3.14: Analytical Gel Filtration Showing a RIFIN Variable Region Binding to LILRB1 A) SEC chromatogram and B) SDS-PAGE gel showing complex formation between the Pf3D7_1254800 variable region and LILRB1. The blue and red bars above the SEC trace correspond to the fractions shown in the SDS-PAGE gels. A Superdex 75 10/300 column was used for SEC.

(SPR). The 3 LILRB1 constructs were bound to separate flow-paths of the chip by amine coupling, with around 300 AU of the full-length LILRB1 ectodomain bound, and around 150 AU of each of D1D2 and D3D4 bound. This allowed for the assessment of binding affinity to LILRB1, as well as determining which region of LILRB1 the RIFIN binds. The RIFIN bound to both the full-length and D1D2 constructs with an affinity of around 1 μM **(Table 3.1)**, with no significant binding to the D3D4 construct **(Figure 3.15)**, showing that the RIFIN is binding to LILRB1 via its 2 membrane distal domains, in the same way as other known LILRB1 binding partners such as MHC class I molecules[191,194]. In addition to this, the 1 μM K_d value is comparable to that of the binding of MHC class I complexes to the receptor, which range from around 2 μM to around 7 μM[191].

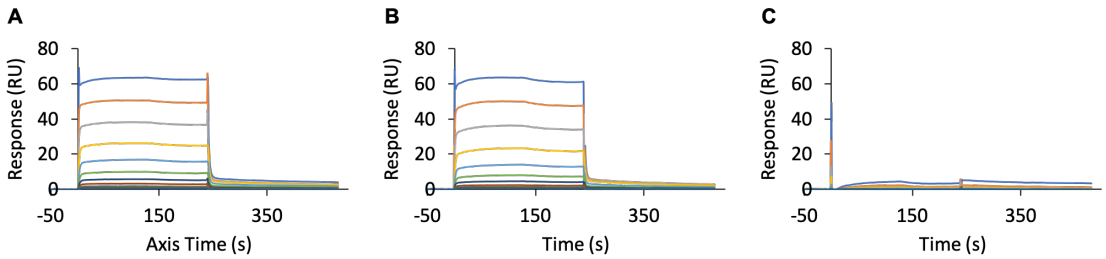


Figure 3.15: Surface Plasmon Resonance Showing a RIFIN Variable Region Binding to 3 LILRB1 Constructs SPR showing binding of the Pf3D7_1254800 variable region to A) full-length, B) D1D2, and C) D3D4 constructs of LILRB1. A 2-fold dilution series was used for each interaction, from 4 μM to 3.9 nM.

	$\mathbf{k}_{on}$ $(\mathbf{M}^{-1}s^{-1})$	$\mathbf{k}_{off}$ $(\mathbf{s}^{-1})$	$\mathbf{K}_d$ $(\mu\mathbf{M})$	χ^2 $(\mathbf{RU}^2)$
Full-Length	2.63×10^5	0.297	1.13	3.88
D1D2	4.37×10^5	0.579	1.33	4.04

Table 3.1: Kinetic Values for Binding of Pf3D7_1254800 Variable Region to LILRB1 Full-Length and D1D2 These values were calculated using fitting to a one-to-one kinetic binding model, with the quality of fitting described by the χ^2 value.

Following this analysis, crystallisation trials were carried out for the complex formed between the full-length LILRB1 ectodomain and the RIFIN variable region. This led to the determination of the structure of the complex, and is described in Chapter 4.

3.4 Characterisation of a RIFIN Constant Region

Having successfully produced the variable region of a RIFIN, the next aim was to produce the other portion of the ectodomain, the constant region. While the function of this portion of the ectodomain is currently unknown, any protein exported to the surface of an infected erythrocyte represents a significant risk to the parasite of

detection by the host immune system. As this region is significantly more conserved than the other half of the ectodomain, the immune evasion strategy of antigenic variation is less applicable, suggesting that this region may still play in important role in the function of the RIFINs. In order to characterise the constant region, the corresponding region of the LILRB1-binding RIFIN previously characterised, Pf3D7_1254800, was first produced. To improve the chances of crystallisation, purification was then additionally attempted for a panel of 4 other RIFIN constant regions for further characterisation and crystallisation trials.

3.4.1 Expression and Purification

3.4.1.1 Expression of the Constant Region of a LILRB1-Binding RIFIN

Initial expression trials for a RIFIN constant region were carried out using Pf3D7_1254800, with 2 separate constructs trialled. The first, 39-164, consisted of the entirety of the sequence between the cleaved PEXEL motif, and the variable region previously produced. A second, shorter construct, 39-139, removed a hydrophobic stretch from the C-terminus of this sequence which may impact expression. Each construct included a C-terminal 6-His tag for use in Ni-NTA affinity purification.

Small scale expression trials in HEK293F cells showed no expression of the longer construct, which had an expected molecular weight of 15.9 kDa, but there was a low level of expression of the shorter construct, with an expected molecular weight of 13.6 kDa **(Figure 3.16)**. This band was at a slightly higher molecular weight than expected, appearing at around 18 kDa, which is likely due glycosylation of the single predicted site within the construct. Expression of this construct was then scaled up to test purification.

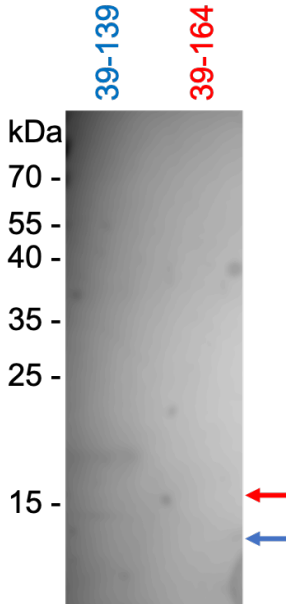


Figure 3.16: Expression Trial of the Constant Region of a LILRB1-Binding RIFIN in HEK293F Cells SDS-PAGE gel assessing expression of 2 constructs for the Pf3D7_1254800 constant region in HEK293F cells. There is no expression of the longer construct, with a low level of expression for the shorter construct. Arrows indicate the expected position of bands, with the blue arrow corresponding to the shorter construct, at around 13.6 kDa, and the red to the longer construct, at around 15.9 kDa.

Larger scale expression and purification of the 39-139 construct demonstrated by both SEC and SDS-PAGE analysis that the protein appeared to be monomeric and of the expected size **(Figure 3.17)**.

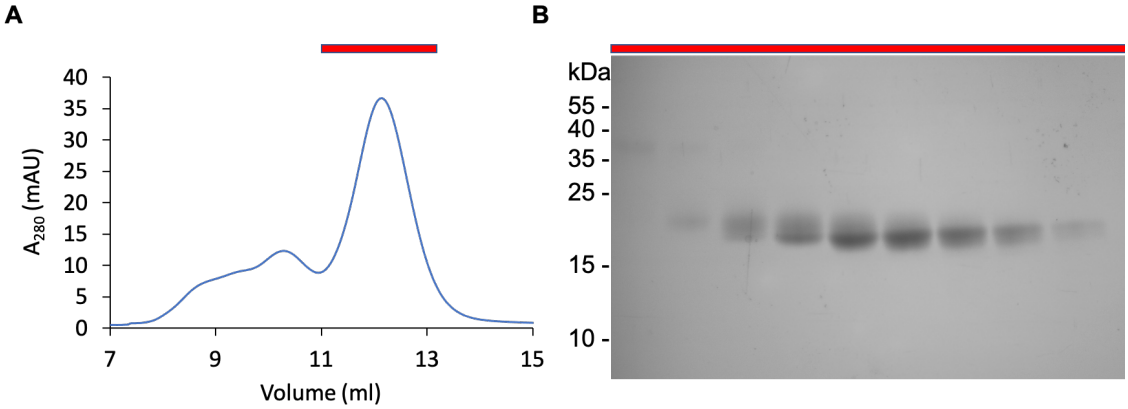


Figure 3.17: Purification of the Constant Region of a LILRB1-Binding RIFIN in HEK293F Cells A) SEC chromatogram showing purification of Pf3D7_1254800 39-139 using a Superdex 75 10/300 column and B) SDS-PAGE gel showing fractions from the main peak, at around 12 ml, of this purification. The red bar above the SEC trace corresponds to the fractions shown in the SDS-PAGE gel.

3.4.1.2 Assessment of the Secondary Structure of the Constant Region

To determine whether this constant region was correctly folded, CD spectroscopy was used to assess the secondary structure. This demonstrated that, like the variable region of the same RIFIN, the constant region is an alpha-helical protein, with minima at 208 and 222 nm **(Figure 3.18)**.

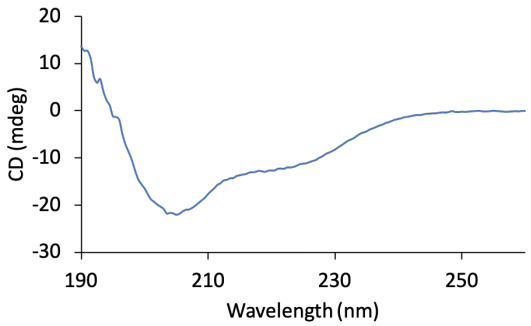


Figure 3.18: Circular Dichroism Spectrum for the Pf3D7_1254800 Constant Region

3.4.1.3 Binding of the Constant Region to LILRB1

To determine whether the constant region is involved in the interaction between this RIFIN and LILRB1, SPR was used to check for any interaction between the two, through directly coupling the full-length LILRB1 ectodomain to the chip as described previously, and flowing the RIFIN constant region over this. The RIFIN variable region was used as a control. This demonstrated that there was no significant response for any interaction of the constant region with LILRB1.

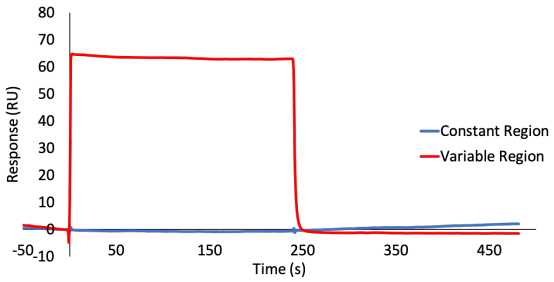


Figure 3.19: Surface Plasmon Resonance for the Interaction Between a RIFIN Constant Region and LILRB1 SPR curves for the interaction of the Pf3D7_1254800 constant, blue, and variable, red, regions with the full-length LILRB1 ectodomain.

3.4.2 Expression of 4 RIFIN Constant Regions

To increase the chances of producing crystals for a RIFIN constant region, a panel of 8 other constructs was also produced, consisting of the equivalent short and longer domain boundaries used for the Pf3D7_1254800 constant region expression trials from 4 other RIFINs.

The RIFINs used were Pf3D7_0401300, with expected molecular weights of 13.6 and 17.3 kDa for the shorter and longer boundaries respectively, Pf3D7_0500400, at 13.6 and 17.1 kDa, Pf3D7_1040800, at 10.8 and 14.1 kDa, and Pf3D7_1373400 at 13.4 and 17.0 kDa.

Initial expression trials in HEK293F cells showed strong expression for Pf3D7_1040800 40-115, with very weak expression for Pf3D7_0500400 44-145 and Pf3D7_1373400 39-

140. There was no expression for the equivalent shorter construct for Pf3D7_0401300, and no expression for any of the 4 longer constructs, which included the C-terminal hydrophobic stretch **(Figure 3.20A)**. Similar results were seen in Expi293F cells, with the same 3 constructs showing any expression, and the Pf3D7_1040300 construct being expressed at the highest level **(Figure 3.20B)**. In both cell systems, the expressed constructs all appear as two bands on the SDS-PAGE gels, the first at the expected molecular weight, and the second at a slightly higher molecular weight, suggesting the presence of two separate species. There are no predicted glycosylation sites within the sequences, so the identity of these species is unclear.

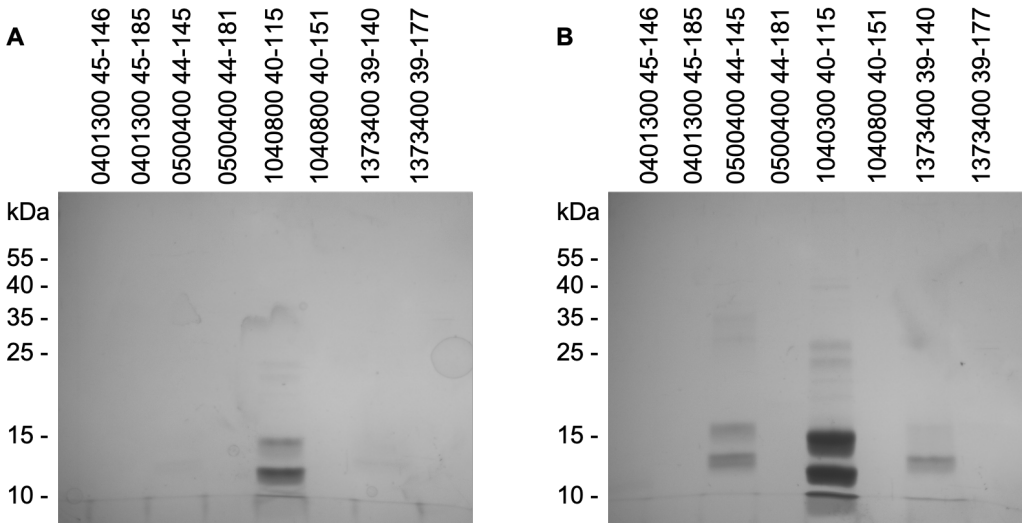


Figure 3.20: Expression Trials for RIFIN Constant Region Constructs in HEK293F and Expi293F Cells SDS-PAGE gels assessing expression of 8 RIFIN constant region constructs in A) HEK293F cells and B) Expi293F cells.

Large scale expression of the successfully expressed constructs, and subsequent purification by SEC, again revealed the presence of these two separate species, which was particularly clear for the Pf3D7_1040800 construct, where it could be seen in the SEC chromatogram, as well as the SDS-PAGE gel **(Figure 3.21)**. This severely limited the utility of these constructs for crystallography, where a homogeneous sample is required, and so crystallisation trials could only be carried out for the

Pf3D7_1040800 construct, where the two separate bands could be partially separated by SEC, allowing for fractions containing only one of the species to be isolated.

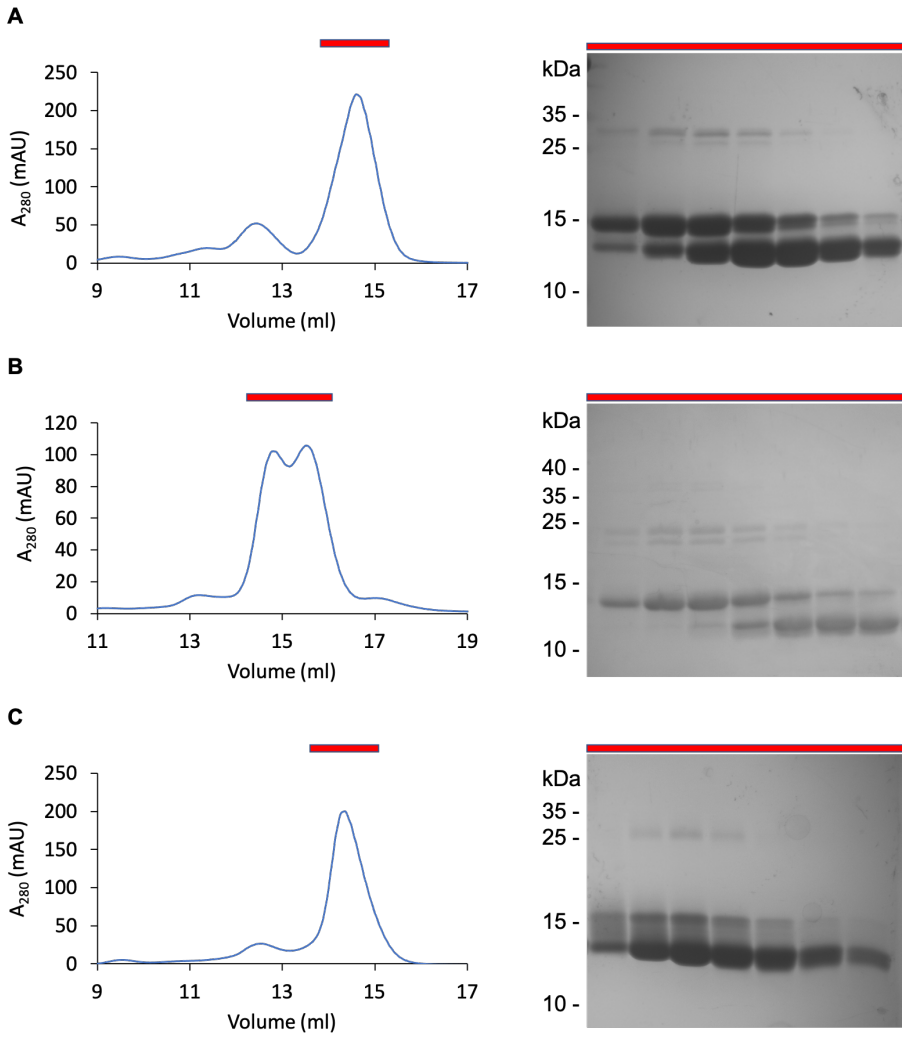


Figure 3.21: Purification of 3 RIFIN Constant Region Constructs from HEK293F Cells SEC chromatograms and SDS-PAGE gels for purification of A) Pf3D7_0500400 44-145, B) Pf3D7_1040800 40-115, and C) Pf3D7_1373400 39-140 using an S75 10/300 column. The red bars above the SEC traces correspond to the fractions shown in the SDS-PAGE gels.

3.4.3 Crystallisation Trials

Extensive crystallisation trials have been carried out in attempts to structurally characterise the RIFIN constant region, primarily for the Pf3D7_1254800 39-139 construct. These trials attempted crystallisation using vapour diffusion from sitting

drops in 96-well plates containing broad screens of conditions. The purified protein was concentrated to between 10 and 20 mg/mL, and several broad crystallisation screens used, at both 4°C and 18°C incubation temperatures. These were the commercial screens MIDAS, Morpheus, JCSG+, ProPlex, PACT, and MemGold from Molecular Dimensions. In addition to this, crystallisation trials were carried out using both glycosylated and deglycosylated protein. In total, 27 96-well plates were set up, representing a range of over 2500 conditions for this construct.

Trials were also carried out for Pf3D7_1040800 40-115. 10 96-well plates were set up for the Pf3D7_1040800 construct, representing both native and methylated protein at around 10 mg/mL, where methylation is aiming to reduce the flexibility of surface exposed lysine residues to improve crystal contact formation. Despite the large number of conditions trialled across both constructs, no crystals have yet been produced. This could be due to regions of flexibility or disorder within the protein which interfere with crystal contact formation. Limited proteolysis could be trialled in the future to attempt to overcome this. This relies on protease cleavage sites in flexible regions being exposed, while those within folded, stable parts of the protein fold are protected. The flexible regions can therefore be cleaved, while the core of the fold remains intact, potentially removing regions of the protein which are disrupting crystal contact formation, and thereby aiding crystallisation.

3.5 Characterisation of a KIR-binding RIFIN Variable Region

In addition to those RIFINs which bind to the receptors previously mentioned, RIFINs have now been identified which bind to a separate family of immune receptors, the killer-cell immunoglobulin-like receptors (KIRs) (H. Arase, personal communication).

The KIRs are a family of receptors found on NK cells which, as described in section 1.5, are key to mediation of the cytolytic activity of these cells. Binding of RIFINs to LILRB1 has already been shown to inhibit the NK cell response[67], and activation of inhibitory KIRs through RIFIN binding could provide another mechanism by which this can be achieved.

Two KIR-binding RIFINs have been identified. The first has been shown to bind to both an inhibitory receptor, KIR2DL1, and an activatory receptor, KIR2DS1, and will be referred to as the KIR2DL1/DS1-binding RIFIN. The second has been shown to interact with two inhibitory receptors, KIR2DL2 and KIR2DL3. This RIFIN will be referred to as the KIR2DL2/DL3-binding RIFIN.

Little is known about the interaction between these RIFINs and the KIRs, and their expression and purification would allow for biophysical and structural characterisation to help reveal details of the mechanism of their interaction.

3.5.1 Expression and Purification

To characterise the interaction between KIR receptors and the RIFINs, expression trials for both RIFINs were first carried out in the mammalian cell systems, HEK293F and Expi293F cells. These trials each included 3 separate constructs for each RIFIN, 2 representing the full-length ectodomain, the first containing the original signal sequence, 1-306 for the KIR2DL1/DS1-binding RIFIN and 1-294 for the KIR2DL2/DL3-binding RIFIN, and another after the signal sequence has been removed through PEXEL motif cleavage, 47-306 for the KIR2DL1/DS1-binding RIFIN and 41-294 for the KIR2DL2/DL3-binding RIFIN. In addition to this, a variable region construct, 144-306 and 138-294 for the KIR2DL1/DS1-binding and KIR2DL2/DL3-binding RIFINs respectively, was included. This was due to the fact that the LILRB1-binding site of Pf3D7_1254800 was contained within the variable region, while the constant region showed no receptor-binding capability.

The KIR2DL1/DS1-binding RIFIN variable region expressed in both systems, with a band seen at the expected molecular weight of 18.2 kDa, although no expression of either longer construct was seen **(Figure 3.22)**. For the KIR2DL2/DL3-binding RIFIN, no expression was seen for any of the 3 constructs in either cell system. As only the variable region of the KIR2DL1/DS1-binding RIFIN could be produced, this protein was focussed on for subsequent experiments.

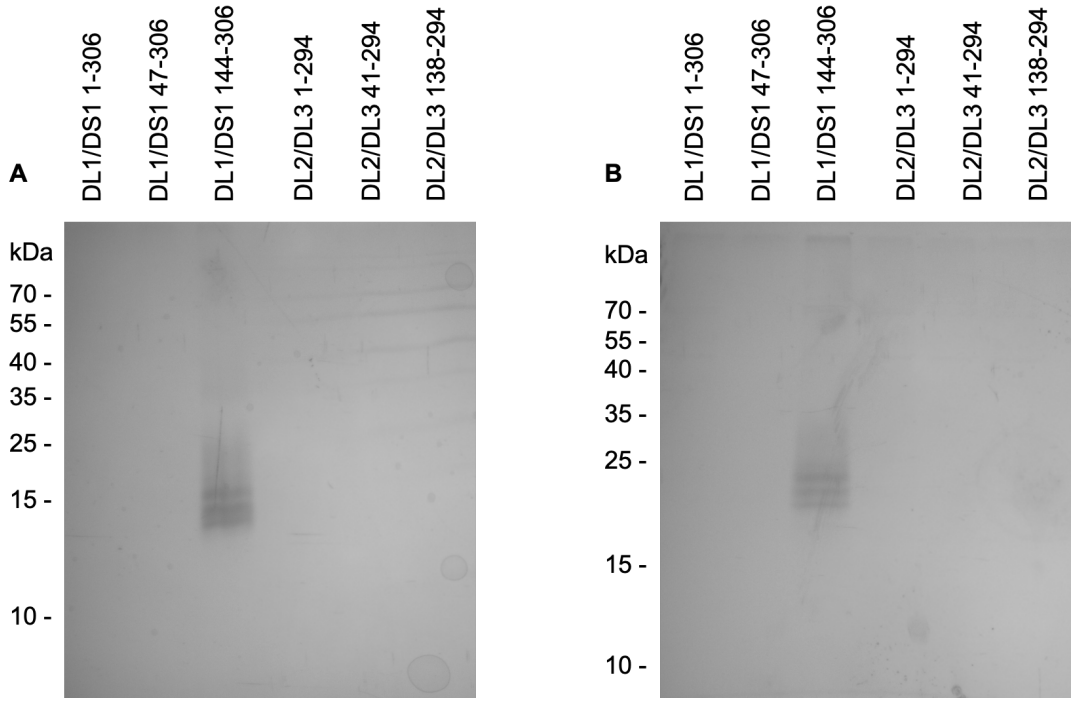


Figure 3.22: Expression Trial of KIR-Binding RIFIN Constructs in HEK293F and Expi293F Cells SDS-PAGE gels assessing expression of 3 constructs for the ectodomain of 2 KIR-binding RIFINs in A) HEK293F and B) Expi293F cells. Expression was seen for the variable region construct of the KIR2DL1/DS1-binding RIFIN in both cell systems.

Large scale expression of this variable region construct in HEK293F cells and purification by SEC suggested that this protein is likely monomeric and so correctly folded **(Figure 3.23A)**. The SDS-PAGE gel did, however, show that a range of species were present following purification, which are likely due to varying degrees of glycosylation **(Figure 3.23B)**.

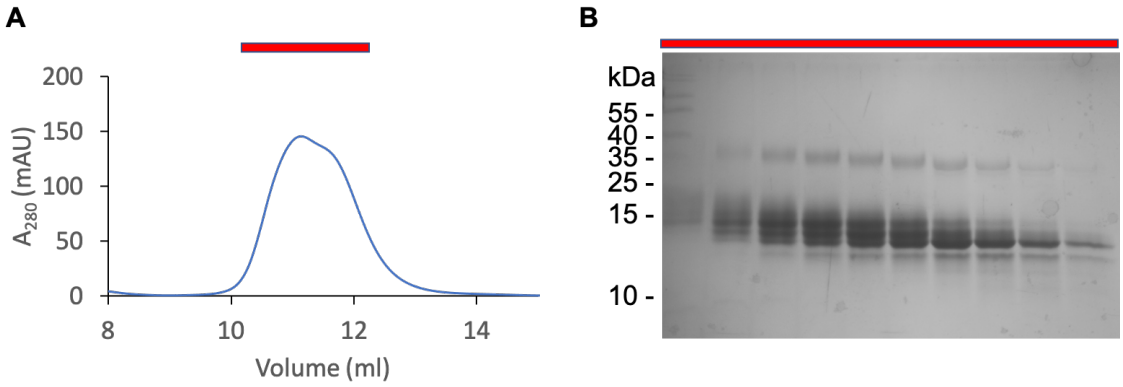


Figure 3.23: Purification of a KIR2DL1/DS1-Binding RIFIN Variable Region from HEK293F Cells A) SEC chromatogram and B) SDS-PAGE gel showing purification of the variable region of a KIR2DL1/DS1-binding RIFIN from HEK293F cells using a Superdex 75 10/300 column. The red bar above the SEC trace corresponds to the fractions shown in the SDS-PAGE gel.

3.5.2 Expression and Purification of KIR2DL1 and KIR2DS1

To test the binding of this variable region to KIR2DL1 and KIR2DS1, the full ectodomains of the two receptors, containing 2 Ig-like domains, was expressed in HEK293F cells. Both were strongly expressed in this cell system, and so were purified by Ni-NTA and subsequent size exclusion chromatography **(Figure 3.24)**.

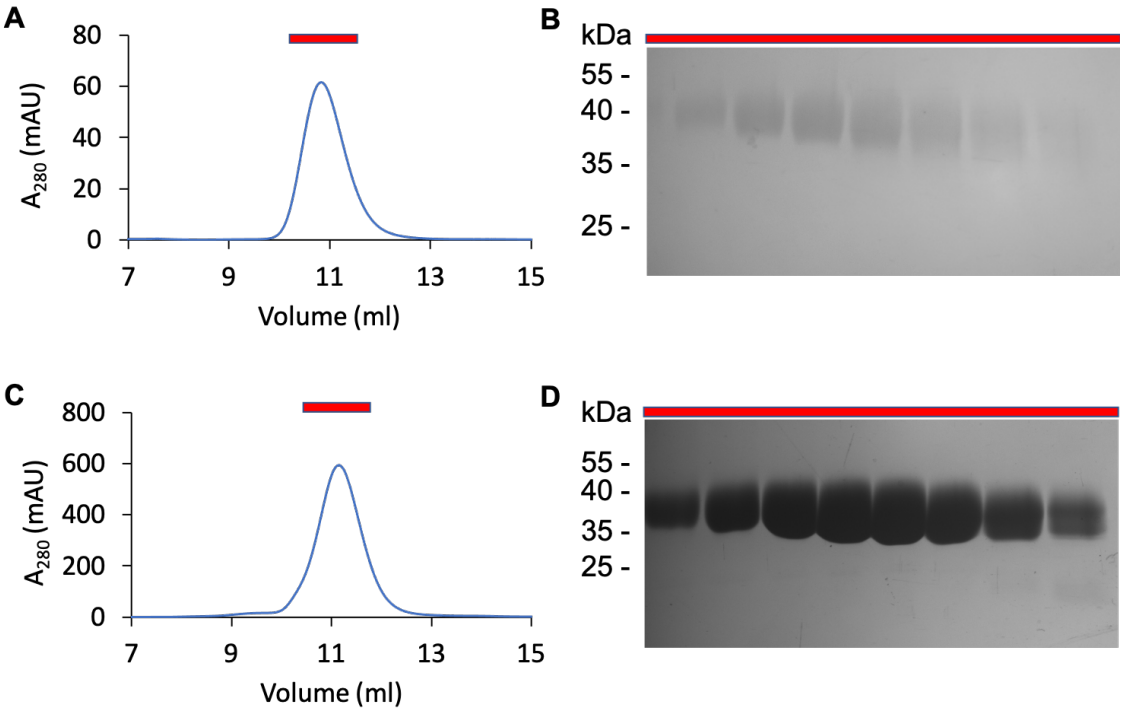


Figure 3.24: Purification of the KIR2DL1 and KIR2DS1 Ectodomains from HEK293F Cells SEC chromatograms and SDS-PAGE gels showing purification of the ectodomains of KIR2DL1 (A and B) and KIR2DS1 (C and D) from HEK293F cells using a Superdex 75 10/300 column. The red bars above the SEC traces correspond to the fractions shown in the SDS-PAGE gels.

3.5.3 Binding of the Variable Region to KIR2DL1 and KIR2DS1

Binding of the RIFIN variable region to the 2 KIR receptors was assessed using SPR. This was achieved through directly coupling the RIFIN variable region to a CM5 chip by amine coupling, and flowing a dilution series of both the KIR2DL1 and KIR2DS1 ectodomains over this. The RIFIN showed binding to both receptors **(Figure 3.25A, B)**, although with a roughly 5-fold stronger binding affinity for the inhibitory receptor as compared to the activatory receptor, with a K_d of 446 nM compared to 2.07 μM respectively **(Table 3.2)**.

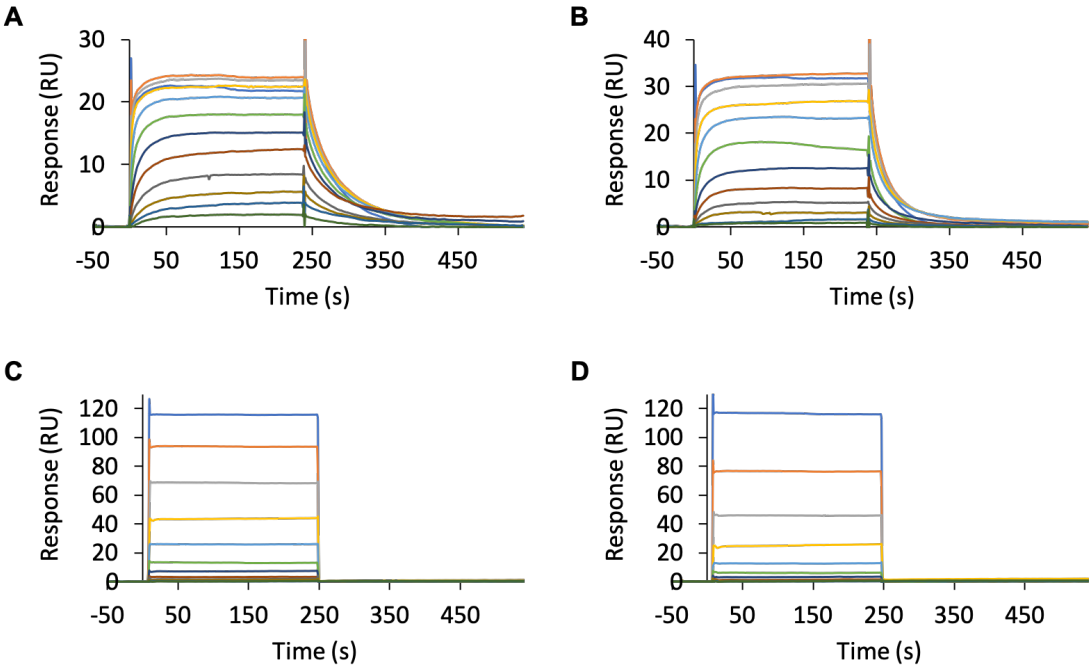


Figure 3.25: Binding of a RIFIN Variable Region and HLA-Cw4 to the KIR2DL1 and KIR2DS1 Ectodomains SPR curves showing the RIFIN variable region binding to A) the KIR2DL1 and B) KIR2DS1 ectodomain, and HLA-Cw4 binding to the C) KIR2DL1 and D) KIR2DS1 ectodomains. A two-fold dilution series was used for each interaction, from 50 μM to 24 nM for the RIFIN-binding experiments, and from 20 μM to 9.8 nM for the HLA-Cw4-binding experiments.

	$\mathbf{k}_{on}$ $(\mathbf{M}^{-1}s^{-1})$	$\mathbf{k}_{off}$ $(\mathbf{s}^{-1})$	$\mathbf{K}_d$ $(\mu\mathbf{M})$
RIFIN vs KIR2DL1	4.83×10^4	0.0213	0.446
RIFIN vs KIR2DS1	2.00×10^4	0.0397	2.07
HLA-Cw4 vs KIR2DL1	2.65×10^5	1.64	6.20
HLA-Cw4 vs KIR2DS1	2.50×10^6	58.4	23.2

Table 3.2: Kinetic Values for Binding of the KIR2DL1/DS1-Binding RIFIN Variable Region and HLA-Cw4 to KIR2DL1 and KIR2DS1 These values were calculated using fitting to a one-to-one kinetic binding model. Average values from three independent dilution series are shown for each interaction. Individual kinetic values from each dilution series, including χ^2 values describing quality of fitting, are given in **Table B.1**.

To compare these binding affinities to those of the natural ligand for these KIR receptors, MHC class I molecules[210], SPR was also used to determine the kinetics of

binding of HLA-Cw4 to the two receptors. Both the HLA-Cw4, and the Fc-tagged receptors used for this experiment were provided by Hisashi Arase, University of Osaka.

I first attempted to assess HLA-Cw4 using the same method used for the RIFIN, by directly coupling the HLA-Cw4 to a chip, and then flowing the receptor over this. However, this yielded negative results, where the MHC showed no interaction with either receptor. This is possibly due to denaturation of the complex at the low pH required for amine coupling to a CM5 chip. To rectify this, Protein A/G was instead coupled to a CM5 chip, allowing for the immobilisation of Fc-tagged KIR2DL1 and KIR2DS1. HLA-Cw4 could then be flowed over, and the interaction measured. This showed HLA-Cw4 bound to both receptors **(Figure 3.24 C, D)**, with a stronger affinity for KIR2DL1 than KIR2DS1, as seen for the RIFIN, with K_d values of 6.20 μM and 23.2 μM respectively. Binding to both receptors was more than 10-fold weaker than seen for the RIFIN variable region.

3.5.4 Crystallisation Trials

In order to structurally characterise the interaction between the RIFIN and the KIR2DL1 ectodomain, crystallisation trials have been carried out. Broad screens were set up at 10 mg/mL, both before and after deglycosylation, with a 4°C incubation temperature. The screens used were the MIDAS, Morpheus, JCSG+, and ProPlex screens from Molecular Dimensions. No crystals have been produced yet, which could be due to the relatively limited range of trials carried out. For future trials, a wider range of concentrations can be tested, along with additional broad screens. The trials can also be repeated at a higher incubation temperature of 18°C.

In addition to this, both the RIFIN and receptor used for these trials contained a C-terminal 6-His tag, which is likely to be flexible, and could be interfering with crystal contact formation. Engineering a protease cleavage site between the

protein and the tag would allow this to be excised, potentially removing a barrier to crystallisation.

Limited proteolysis, as described in section 3.4.3, could also be trialled to cleave any flexible regions of protein, either in loops or at the termini of each protein. This could again remove regions which are interfering with crystal contact formation, leaving the stable core of the proteins, and potentially aiding crystallisation.

3.6 Discussion

Expression of RIFINs remains a challenge. A number of attempts to produce the full ectodomain of these proteins using a range of expression systems proved unsuccessful, and so in order to study these proteins, shorter constructs, representing the 2 halves of the ectodomain separately, were resorted to. This proved more successful, with RIFIN constant regions produced using a mammalian cell system, and variable regions produced using either refolding from bacterial inclusion bodies, or again using a mammalian cell system. This allowed for the biophysical characterisation of both regions of the RIFIN ectodomain, as well as analysis of the interaction of RIFIN variable domains with 2 inhibitory immune receptors, LILRB1 and KIR2DL1, and an activatory receptor, KIR2DS1. Unfortunately, production of RIFINs remains a challenge, and this is still a significant barrier to studying this protein family.

A LILRB1-binding RIFIN variable region was shown to interact with LILRB1 with a similar affinity, around 1 μM, to the native ligand, MHC class I complexes, which bind with an affinity ranging from around 2 μM to around 7 μM. The constant region of the same RIFIN appears not to contribute to this interaction, showing no detectable binding to the receptor by SPR. In addition to this, the variable region was shown to interact solely with the 2 membrane distal domains of the receptor, with no interaction with domains 3 and 4, again similarly to the binding of MHC

class I complexes, which bind to the same portion of the ectodomain[191–193]. This interaction is explored further, with both structural and functional characterisation, in the following chapter.

A KIR2DL1- and KIR2DS1-binding RIFIN variable region was also produced, and its ability to bind to the two receptors was confirmed by SPR. The RIFIN showed more than 10-fold stronger binding to the receptors than the physiological binding of MHC complexes. In addition to this, the binding affinity for the inhibitory receptor was around 5-fold stronger than for the activatory receptor. This could mean that, in the presence of both receptors, the RIFIN would bind preferentially to the inhibitory receptor, and so would still be able to induce an inhibitory signal. In addition to this, KIR receptor signalling is mediated through the use of ITAM and ITIM motifs, as described in more detail in section 1.5. The inhibitory signal induced through ITIMs directly blocks activation signals induced by ITAM signalling[181], and so is typically dominant. This suggests that RIFIN binding would allow for the down-regulation of the NK cell response to an infected erythrocyte, despite the interaction between the RIFIN and the activatory receptor, KIR2DS1. The ability of this RIFIN to induce signalling through the KIR receptors still needs to be confirmed, however, as does the effect of RIFIN binding to KIRs on NK cell activity.

Crystallisation of this RIFIN in complex with the receptor will allow for this interaction to be investigated more closely, as well as revealing the extent of similarity of binding to the receptor as compared to both the physiological ligand, and of a LILRB1-binding RIFIN with its receptor. Further analysis of the effect of this RIFIN should also be carried out, such as through signalling assays, like the NFAT-GFP reporter assay used to confirm signalling induced by LILRB1-binding RIFINs[67], and NK cell activity assays to determine the effect of these RIFINs on NK cell activation.

Study of the RIFIN constant region has shown that, like the variable region, an

α-helical secondary structure is adopted, however beyond that very little is known about either the structure or function of this portion of the RIFIN ectodomain. It does not appear that the constant region is involved in ligand binding, although as ligands are not known for all RIFINs, this may not be universally true. Additional functional characterisation is therefore needed in the future, while crystallisation is also required to reveal the full structure of the RIFIN ectodomain.

Further attempts to produce the full-length RIFIN ectodomain are also needed, as this will allow for more biologically relevant study of the family, as opposed to simply focusing on one portion of the protein. This could also help to reveal the function of the constant region of the ectodomain, and to identify any contribution to either the stability of the variable region, or potentially to the interaction of RIFINs with their binding partners.

Chapter 4

Structural Basis for RIFIN Binding to LILRB1

4.1 Summary

To determine the structural basis for the modulation of LILRB1-signalling through binding by a RIFIN, the complex formed by the full LILRB1 ectodomain and the variable region of a RIFIN was crystallised. The structure of this complex was solved using molecular replacement, allowing both the first structural characterisation of any portion of a RIFIN, as well as identification of the interactions formed between the variable region and LILRB1.

This crystal structure identified a key loop within the RIFIN which forms the bulk of the interactions with LILRB1, and which is anchored in place by a conserved disulphide bond. Mutagenesis of this loop confirmed its importance for both binding and signalling through LILRB1, and similar mutations in other LILRB1-binding RIFINs suggested a conserved mode of binding for all LILRB1-binding RIFINs.

In addition to this, comparison of this structure with that of LILRB1 in complex with its physiological binding partner, an MHC class I complex, revealed a similar network of interactions formed within the two complexes, suggesting a form of molecular mimicry being utilised by the RIFIN. Further comparison using a supported lipid bilayer system to image NK cell contact formation again revealed a similarity in the

effect of RIFIN and MHC binding to LILRB1.

4.2 Crystallisation, Data Collection, and Structure Solution of the RIFIN Variable Region:LILRB1 Complex

4.2.1 Crystallisation

For crystallisation trials, the Pf3D7_1254800 RIFIN variable region, residues 165-274, and full-length LILRB1 ectodomain were mixed at a 1:1 ratio, incubated at room temperature for 20 min, and the complex then purified by SEC **(Figure 4.1)**. Fractions from the peak containing the complex were then pooled, and concentrated to between 6 and 20 mg/ml for crystallisation trials.

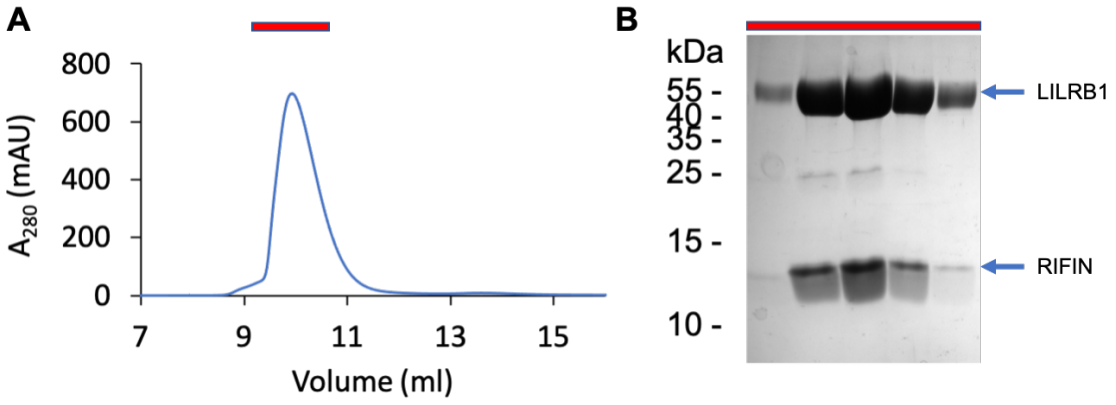


Figure 4.1: Purification of the RIFIN Variable Region:LILRB1 Complex A) SEC chromatogram showing purification of the Pf3D7_1254800 variable region:LILRB1 full-length ectodomain complex using a Superdex 75 10/300 column, and B) SDS-PAGE gel showing fractions corresponding to the red bar above the SEC trace which were pooled for crystallisation trials.

Crystallisation trials for the RIFIN variable region:LILRB1 complex were carried out across a wide range of conditions. These included a range of protein concentrations from 6 to 20 mg/mL, using both 4°C and 18°C incubation temperatures, and several

broad screens of conditions.

Crystals were observed after around 6 days in two separate conditions, both using a protein concentration of 10 mg/mL and in plates incubated at 4°C. One of these conditions was part of the MIDAS screen (Molecular Dimensions), and contained 0.1 M Tris pH 8.0, 20% w/v SOKALAN CP-42, and 5% v/v methanol. The second was part of the ProPlex screen (Molecular Dimensions), and contained 0.1 M sodium Hepes pH 7.5, 0.2 M ammonium acetate, 25% v/v isopropanol. Both conditions gave very small, bi-pyramidal crystals, around 20 μm across, which proved difficult to harvest for data collection.

In attempts to produce larger crystals which would be more suitable for harvesting, a number of optimisation plates were set up. These included plates containing gradients of protein concentration, from 6 mg/mL to 13 mg/mL, and precipitant conentration, either of SOKALAN CP-42 between 14% and 25% w/v, or of isopropanol, between 14% and 25% v/v.

These optimisation plates demonstrated that the best protein concentration appeared to be 6 mg/mL for both conditions, which allowed for slightly larger crystals to be produced, up to around 35 μm across, which could be more easily harvested **(Figure 4.2)**. These crystals were first cryoprotected through transfer to a drop of the original condition which had been supplemented with cryoprotectant at a concentration of 25% v/v. The cryoprotectants trialled were glycerol, 2-methyl-2,4-pentanediol (MPD), polyethylene glycol 400 (PEG400), and ethylene glycol. The crystals were then harvested using cyroloops and immediately cryocooled in liquid nitrogen.

4.2.2 Data Collection and Processing

Data were collected at the Proxima-1 beamline at SOLEIL at an incident X-ray wavelength of 0.97867 Å. Crystals grown using the condition containing 0.1 M sodium

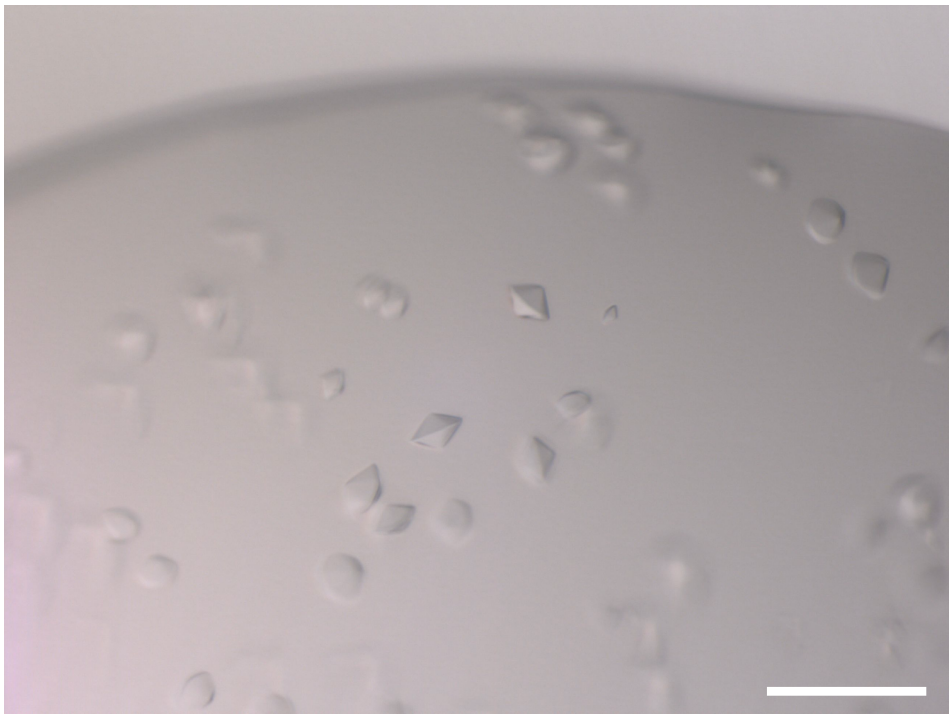


Figure 4.2: Crystals of the RIFIN Variable Region:LILRB1 Complex These crystals grew after around 6 days in 0.1 M sodium Hepes pH 7.5, 0.2 M ammonium acetate, 25% v/v isopropanol, using a protein concentration of 6 mg/mL, and incubation temperature of 4°C. The scale bar indicates a length of 100 μm.

Space group	$P6_1$ 2 2
Cell dimensions	
a, b, c (Å)	134.51, 134.51, 277.47
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.97857
Resolution (Å)	277.47-3.00 (3.16-3.00)
Total observations	602390 (89498)
Total unique reflections	30613 (4346)
R_{pim}	0.065 (0.647)
$CC_{1/2}$	0.99 (0.76)
I/σ *(I)*	9.4 (1.6)
Completeness (%)	100.0 (100.0)
Multiplicity	19.7 (20.6)
Wilson B-factor (Å^2)	90.83

Table 4.1: Data Collection Statistics for the RIFIN Variable Region:LILRB1 Structure Values in parentheses are for the highest resolution shell.

Hepes pH 7.5, 0.2 M ammonium acetate, and 25% v/v isopropanol, and cryoprotected using 25% v/v MPD, diffracted to the highest resolution, around 3 Å.

The data were indexed and integrated using DIALS[217], which identified a space group of $P6_1$ 2 2, with unit cell dimensions of a = 134.51 Å, b = 134.51 Å, and c = 277.47 Å, and angles of $\alpha = 90°$, $\beta = 90°$, and $\gamma = 120°$. Data were then scaled and merged using AIMLESS[219], which gave a complete dataset at a resolution of 3 Å, with a signal-to-noise ratio (I/σ (I)) of 9.4, and 1.6 in the highest resolution shell. The R_{pim}, the precision-indicating merging value which is as an indicator of data quality, is 0.065, below the generally accepted upper cutoff of 0.1, and $CC_{1/2}$ for the highest resolution shell was 0.76, well above the generally used 0.5 lower cutoff **(Table 4.1)**.

To solve the structure, molecular replacement was carried out using Phaser MR[222], with previously solved LILRB1 structures used for search models. Structures for the 4 individual domains of the LILRB1 ectodomain were extracted from the PDB entries 6EWA[194] and 4LL9[265]. Using the individual domains as separate search models helps to account for any differences in angles between domains between the search models and the RIFIN:LILRB1 structure.

One copy of each domain was found, which were positioned relative to one another as expected based on previous structures[194,265]. Extra density could also been seen for the helices of the RIFIN variable region. These helices were built, as polyalanine chains, using Coot[223], followed by refinement using BUSTER[227] to improve the initial phases and electron density map **(Figure 4.3)**. Subsequent cycles of building and refinement allowed the remaining loops and helices to be placed. The correct amino acids were substituted into the structure using the single disulphide bond to position the sequence. All residues of the RIFIN variable region could be placed within the model, other than 12 residues at the C-terminus, likely due to flexibility of this region. Residues 2-397 of the LILRB1 ectodomain could also be placed within the model. Glycans could also be seen at 3 sites on LILRB1, with these sites all found within domains 3 and 4, away from the RIFIN binding site.

Refinement of this final model gave R_{work} and R_{free} values of 0.234 and 0.249 respectively **(Table 4.2)**. The low RMSD values for both bond lengths, 0.011 Å, and bond angles, 1.25°, were indicative of a good quality model, as was the Ramachandran plot, indicating 93.8% of residues in favourable regions, and 6.2% in allowed regions, with no residues in disallowed regions **(Figure 4.4)**.

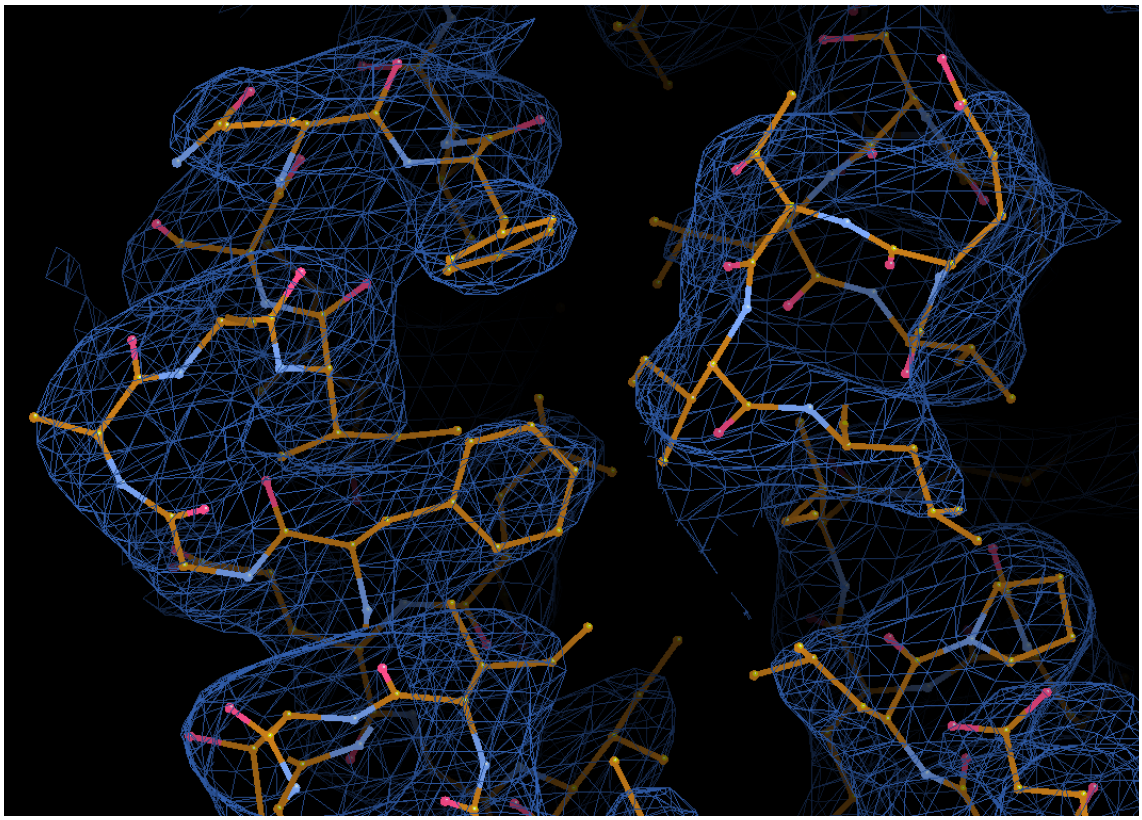


Figure 4.3: Final RIFIN Variable Region:LILRB1 Electron Density Map (2mF$_o$ - DF$_c$, contour 1σ), with Structure Superimposed The interface between a loop and helix of the RIFIN variable region is shown, atoms are coloured by element, with carbon in orange, nitrogen in blue, and oxygen in pink. This figure was prepared using Coot[223].

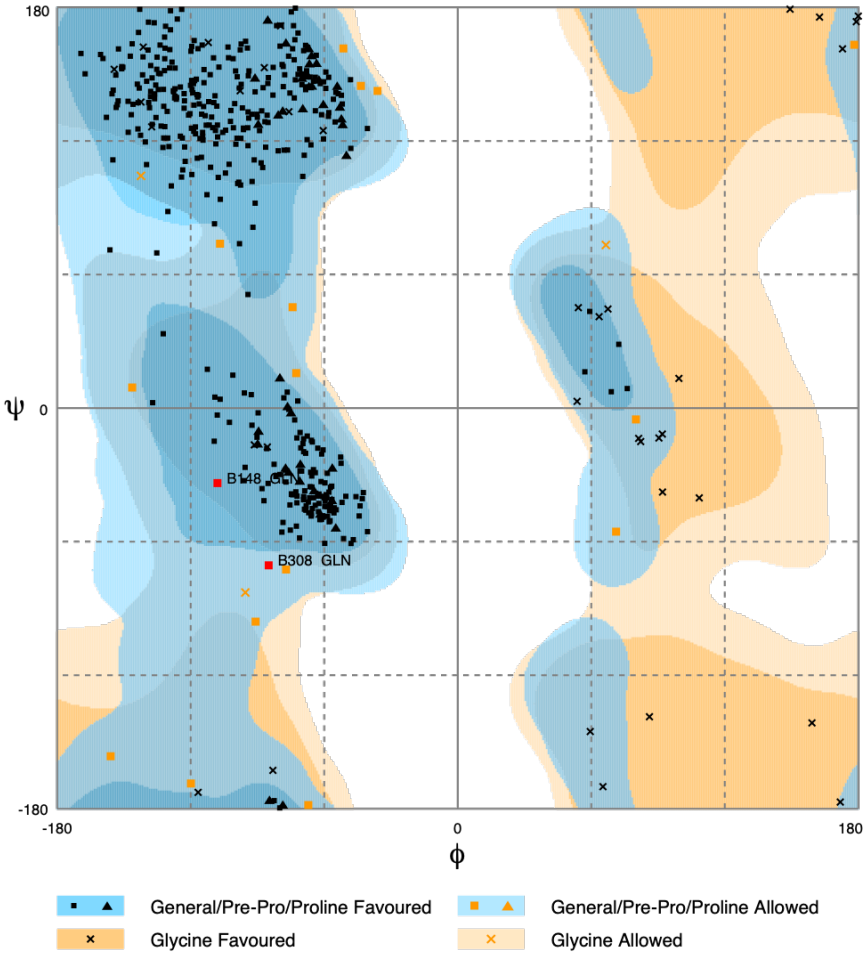


Figure 4.4: Ramachandran Plot for the RIFIN Variable Region:LILRB1 Structure

Reflections	30517
R_{work}/R_{free}	0.2340/0.2487
B-factor (Å^2)	
Protein	119.6
Glycans	176.0
R.m.s deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.25
Ramachandran plot	
Favoured (%)	93.8
Allowed (%)	6.2
Disallowed (%)	0.0

Table 4.2: Refinement Statistics for the RIFIN Variable Region:LILRB1 Structure

4.3 The Structure of the RIFIN Variable Region

The variable region of the RIFIN adopts a primarily α-helical structure, with 3 main helices packed together into a compact bundle **(Figure 4.5)**. The first helix is formed from residues 166 to 182, the second from residues 206 to 222, and the final helix from residues 242 to 259. These helices are connected by 2 long loops, both of which contain short α-helical sections. Two of the cysteines within this region of the RIFIN, C221 and C232, form a disulphide bond which anchors the complex loop between second and third main helices. A third cysteine remains unpaired, and is solvent exposed, which could potentially allow for the formation of a disulphide bond with the constant region within the full ectodomain. The constant region of Pf3D7_1254800 also contains an odd number of cysteines, 7, leaving an unpaired cysteine possibly allowing for this. The final 12 residues at the C-terminus of the

ectodomain are missing from the structure, suggesting flexibility. This could signify that the RIFIN ectodomain is connected to the transmembrane helix through a flexible linker, rather than being held rigidly in place relative to the membrane.

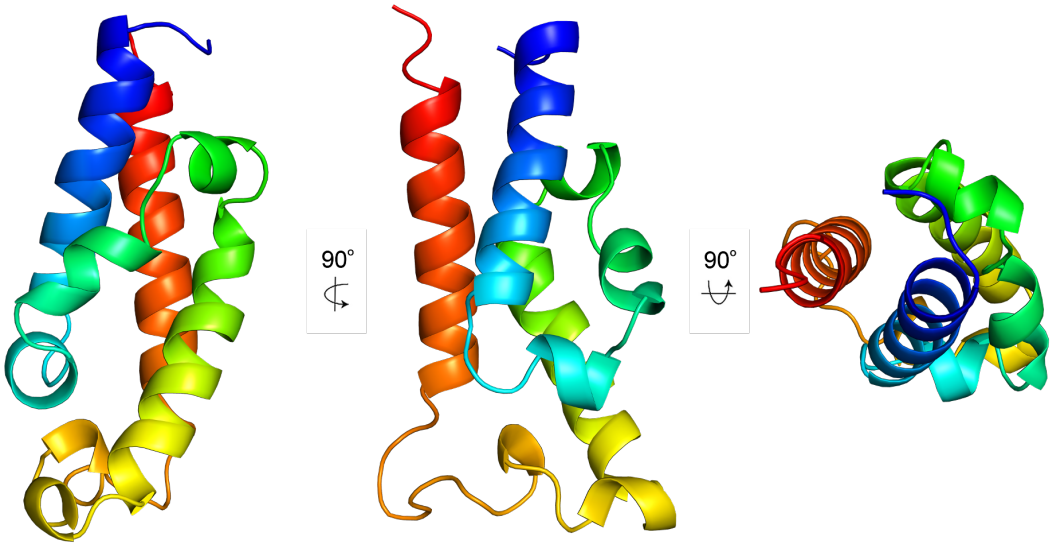


Figure 4.5: The Structure of a RIFIN Variable Region 3 views of the structure of the Pf3D7_1254800 variable region, coloured using a rainbow representation with the N-terminus in blue and the C-terminus in red.

4.3.1 Conservation of the Variable Region

To determine whether the fold of the variable region seen in the structure is conserved among all RIFINs, sequence alignments of the variable domains of all 185 RIFINs from the 3D7 strain of *P. falciparum* were produced using MUSCLE[244]. This demonstrated that very few residues within this region of the RIFINs are conserved, as shown by the sequence Logo for this alignment **(Figure 4.6A)**. The only highly conserved residues are the two cysteine residues which form the disulphide bond, C221 and C232. In addition to this, residue 203, a tyrosine in the structure, is well conserved as either a tyrosine or phenylalanine, both bulky hydrophobic residues which would act to stabilise the core of the fold. The Shannon sequence entropy for the RIFIN variable region was also calculated, where a lower entropy value corresponds to

a higher degree of conservation. Plotting these values onto the structure, where the most highly conserved residues, those with an entropy value of less than 0.5, are shown in red, further demonstrates the lack of conservation of the RIFIN fold **(Figure 4.7A)**. Again the only highly conserved residues are the 2 cysteines involved in disulphide bond formation and the core tyrosine. A few other residues on the N-terminal helix are also conserved, although to a lesser degree.

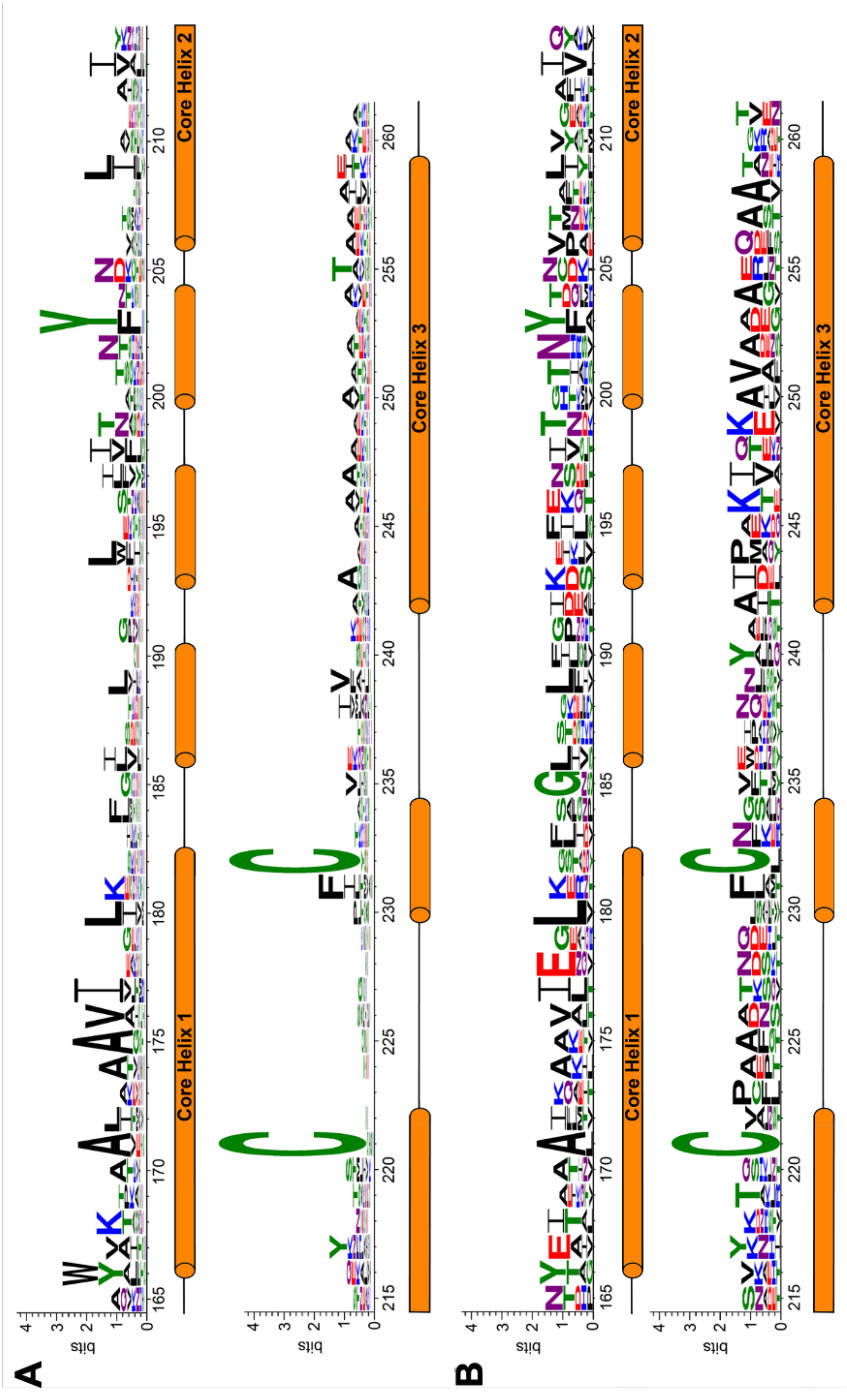


Figure 4.6: Sequence Logos for RIFIN Variable Regions Sequence Logos showing conservation for the portion of the variable region aligning with the structure for A) 185 RIFINs from the 3D7 strain of *P. falciparum*, and B) 11 LILRB1-binding RIFINs. The secondary structure for the variable region of the RIFIN Pf3D7_1254800 is shown underneath the Logo, with helices as cylinders and elements without a secondary structure shown as a line.

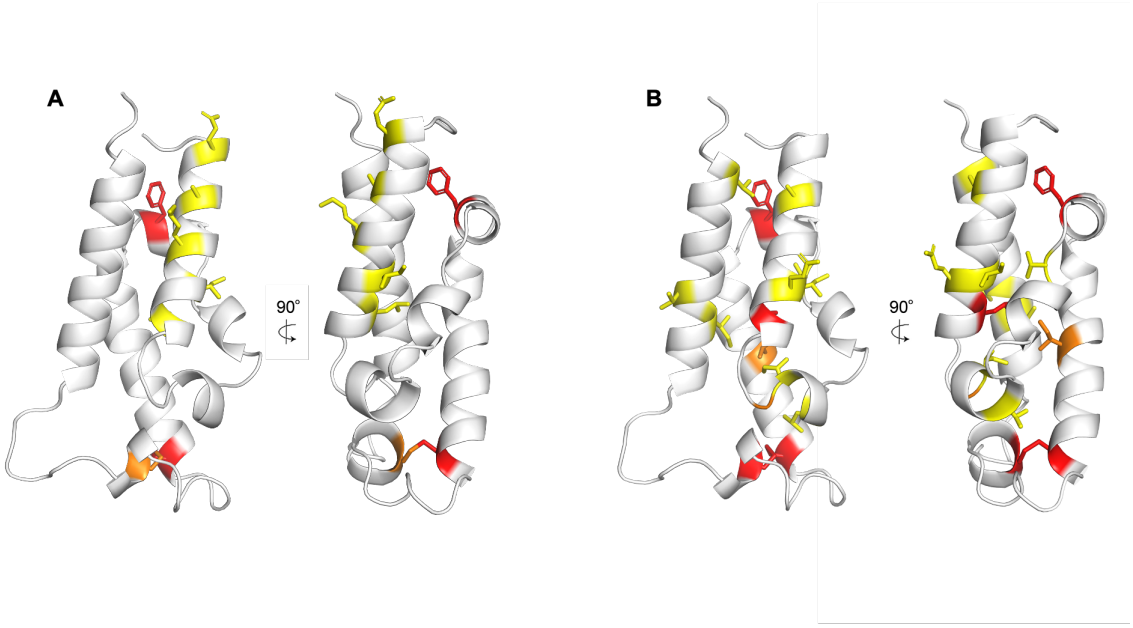


Figure 4.7: Conservation of the RIFIN Variable Region The structure of the RIFIN variable region with the most conserved residues coloured by sequence entropy for A) 185 RIFINs from the 3D7 strain of *P. falciparum*, and B) 11 LILRB1-binding RIFINs. Residues coloured in red have a sequence entropy of <0.5, those in orange of 0.5-0.75, and those in yellow of 0.75-1.0.

When looking just at the 11 RIFINs that are known to bind to LILRB1, the same lack of conservation is seen **(Figure 4.6B)**. The disulphide-forming cysteines remain conserved, as does Y203. There are also 2 more relatively well conserved core hydrophobic residues, L180 and I213 **(Figure 4.7B)**. In general, however, there are still very few residues which show a high degree of conservation across the variable region, making it difficult to assess the conservation of the fold of this region from sequence alignments alone.

Structure-based threading using Phyre2[247] was also used to assess conservation of the RIFIN variable region fold across the LILRB1-binding RIFINs **(Table 4.3)**. 5 members of this group can be aligned well to the entire length of the Pf3D7_1254800 variable domain structure, and modelled with high confidence scores, where a value of greater than 90% indicates with strong confidence that the sequence adopts the given fold[266]. This strongly suggests, therefore, that they will adopt the fold seen in

the structure.

Other LILRB1-binding RIFINs, however, gave contrasting results, with scores as low as around 54%. This could in part be explained by the particularly low sequence identities seen between RIFIN variable regions, as low as around 14% for some LILRB1-binding RIFINs, while sequence identities of over 30% are usually desirable for threading with Phyre2. Therefore while it does appear that many of the LILRB1-binding RIFINs do adopt the fold seen in the structure, the degree of conservation across all of the sequences is still unclear.

Sequence	**Sequence Identity (%)**	**Confidence (%)**	**Aligned Residues**
Pf3D7_0100200	20	96.49	97
Pf3D7_0100400	14	78.83	98
Pf3D7_0223100	22	96.46	99
Pf3D7_0500400	17	87.41	98
Pf3D7_0600300	16	76.57	97
Pf3D7_0632200	15	85.57	98
Pf3D7_0632700	19	54.57	93
Pf3D7_0700200	24	97.76	98
Pf3D7_1100400	25	97.71	99
Pf3D7_1480000	15	95.09	99
Average	18.7	87.65	97.6

Table 4.3: Structure-Based Threading of LILRB1-Binding RIFINs Sequence identities of the variable regions of LILRB1-binding RIFINs with the Pf3D7_1254800 variable region, and confidence scores for Phyre2 threading of these sequences using the Pf3D7_1254800 variable region structure. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266].

4.4 Analysis of the Interaction Between the RIFIN and LILRB1

All 4 immunoglobulin-like domains of the LILRB1 ectodomain can be seen in their entirety within the structure **(Figure 4.8)**. These are arranged in the same linear 'zig-zag' arrangement seen in other structures of the full-length ectodomain[191]. The C-terminus of the region of LILRB1 seen in the structure would be connected to the transmembrane helix via a 40 residue linker.

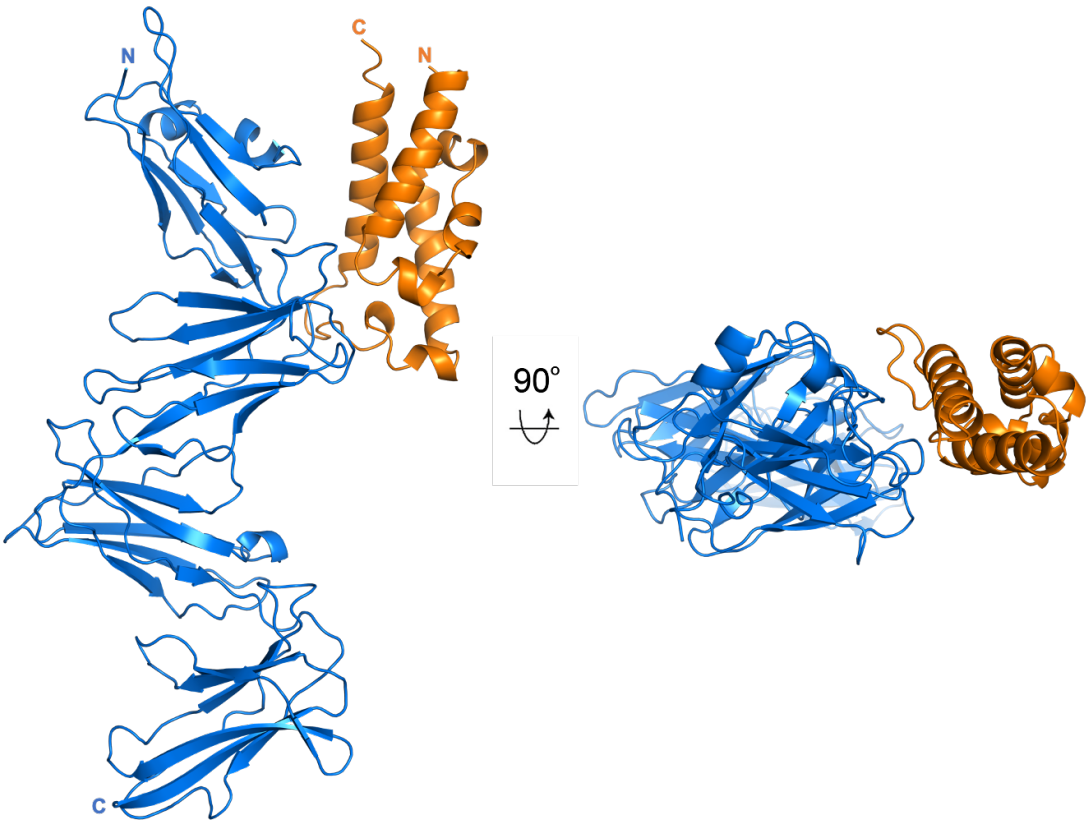


Figure 4.8: The Structure of the RIFIN Variable Region:LILRB1 Complex 2 views of the structure of the complex between the Pf3D7_1254800 variable domain, orange, and the ectodomain of LILRB1, blue.

The RIFIN variable region binding site is limited entirely to the 2 membrane-distal

domains, as expected from the SPR results showing binding to the D1D2 construct, but not the D3D4 construct **(Figure 3.15)**. Binding is mediated by the loops of LILRB1 which form the interface between these two domains, which forms interactions with 3 separate regions of the RIFIN. Residues from the long loop formed by residues 223 to 241 between the 2^{nd} and 3^{rd} main helices of the RIFIN, anchored in place by the single disulphide bond, contribute to around 50% of the 750 Å^2 binding surface. The remaining interactions are formed by a short section of loop C-terminal of the 1^{st} helix, and 3 side chains on the 3^{rd} helix - K245, K246, and E253. The majority of the interactions formed are hydrogen bonds, primarily utilising side chains of RIFIN residues, as well as the carbonyl group of G234 **(Table 4.4)**. There is also a hydrophobic component to the interaction, with the L230 side chain interacting with a hydrophobic pocket of LILRB1 formed by the side chains of 3 residues, Y99, V126, and F128, while an interaction is also formed between M235 of the RIFIN and Y99 of LILRB1.

RIFIN		**LILRB1**		
Residue	**Group**	**Residue**	**Group**	**Interaction**
N228	Side Chain NH_2	Q125	Backbone O	Hydrogen Bond
L230	Side Chain	Y99	Side Chain	Hydrophobic
		V126	Side Chain	Hydrophobic
		F128	Side Chain	Hydrophobic
R233	Side Chain N	Q125	Side Chain O	Hydrogen Bond
G234	Backbone O	I100	Backbone NH	Hydrogen Bond
M235	Side Chain	Y99	Side Chain	Hydrophobic
N236	Side Chain NH_2	G97	Backbone O	Hydrogen Bond
		A98	Backbone O	Hydrogen Bond
K245	Side Chain NH_2	E68	Side Chain OH	Hydrogen Bond
K246	Side Chain NH_2	E184	Side Chain OH	Hydrogen Bond
E253	Side Chain OH	K42	Side Chain NH_2	Hydrogen Bond

Table 4.4: Table of Interactions between the RIFIN Variable Region and LILRB1

Previous structures of LILRB1 alone (PDB code: 1G0X[267]), and in complex with a natural ligand, MHC class I molecules (PDB code: 6EWA[194]), have shown a small change in the angle, approximately 15°, between domains upon ligand binding **(Figure 4.9A)**. An angle change can also be seen upon binding of the RIFIN to LILRB1, although more subtle, at around 5°**(Figure 4.9B)**. The RIFIN is, therefore,

not inducing major conformational change within the LILRB1 ectodomain.

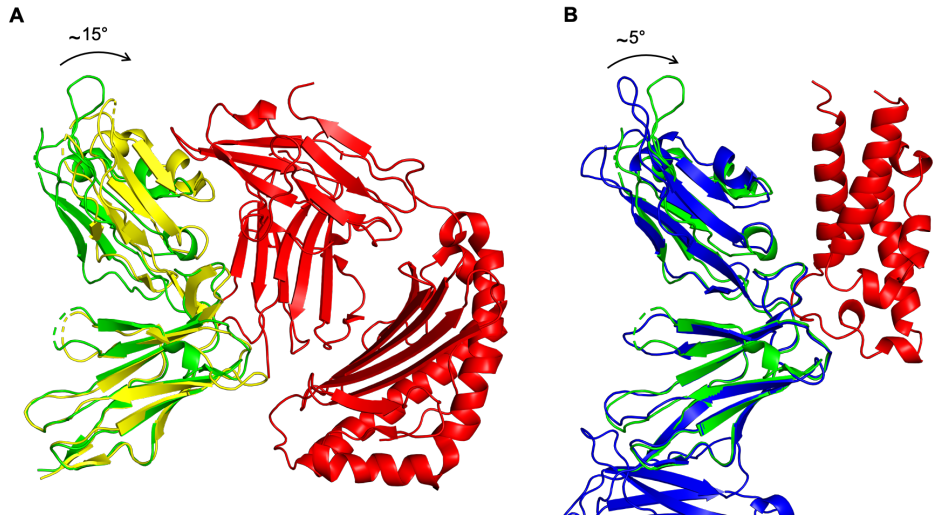


Figure 4.9: Change in Angle Between LILRB1 Domains Upon Ligand Binding The change in angle between domains 1 and 2 of the LILRB1 ectodomain upon A) MHC class I molecule binding, and B) RIFIN variable region binding. The unbound LILRB1 is shown in green, and ligands in red. LILRB1 D1D2 after MHC class I binding is in yellow, and LILRB1 after RIFIN binding in blue.

4.4.1 Confirmation of the Binding Interface

Based on the RIFIN variable region structure, a shorter construct was designed which contained residues 165-261. This removed the C-terminal residues which were missing from the structure, and so which were likely to be disordered. In addition to this, a mutation, C223S was introduced, to remove the unpaired cysteine. Both changes were made in an attempt to produce a more stable construct with a better yield from purification. This construct could be produced successfully using the same refolding protocol used for the RIFIN used for crystallisation trials, with a band seen on the SDS-PAGE gel at the expected molecular weight of 12.9 kDa **(Figure 4.10A, B)**. Analysis using CD confirmed that the construct adopted the expected α-helical secondary structure **(Figure 4.10C)**, and SPR also confirmed that this shorter construct could still bind to LILRB1 **(Figure 4.10D)**, with the same affinity as the slightly longer 165-274 construct previously described of around 1 μM **(Table**

4.5). This shorter construct, including the C223S mutation, was used in place of the wild-type RIFIN variable region for all subsequent experiments.

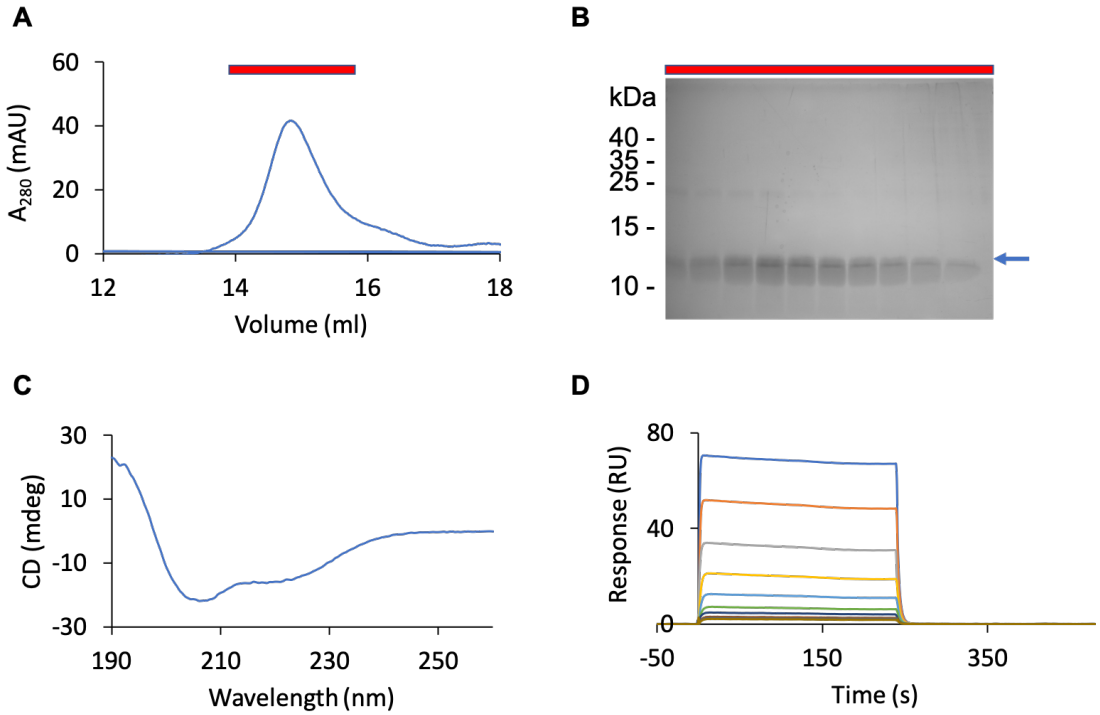


Figure 4.10: Purification and Biophysical Characterisation of the RIFIN C223S Mutant A) SEC chromatogram showing purification of the shorter Pf3D7_1254800 165-261 C223S construct, and B) SDS-PAGE gel showing fractions corresponding to the red bar above the SEC trace. The arrow indicates the expected position of bands on the SDS-PAGE gel for the expected molecular weight of 12.9 kDa. C) CD spectrum for this construct, and D) SPR curves showing binding to the LILRB1 ectodomain. A 2-fold dilution series was used for SPR, from 4 μM to 7.8 nM.

$\mathbf{k}_{on}$ $(\mathbf{M}^{-1}s^{-1})$	$\mathbf{k}_{off}$ $(\mathbf{s}^{-1})$	$\mathbf{K}_d$ $(\mu\mathbf{M})$	χ^2 $(\mathbf{RU}^2)$
2.88×10^5	0.299	1.04	3.76

Table 4.5: Kinetic Values for Binding of the Pf3D7_1254800 C223S Mutant to LILRB1 These values were calculated using fitting to a one-to-one kinetic binding model, with the quality of fitting described by the χ^2 value.

To confirm that the binding interface seen in the crystal structure is accurate for binding both in solution, and in a cellular context, mutations were designed which

were predicted, using the RIFIN:LILRB1 structure, to block binding. These primarily focused on introducing bulky residues to sterically hinder the interaction at 2 sites, P184 and G234 **(Figure 4.11)**. Expression and refolding trials were carried out for 4 separate mutations, 3 of G234, and 1 of P184. Monomeric protein could be produced for the 3 G234 mutants, G234R, G234W, and G234Y. Refolding of the P184W mutant produced a band on the non-reducing SDS-PAGE gel at around 25 kDa, much higher than the anticipated size of around 14 kDa, perhaps indicating the formation of intermolecular disulphide bonds **(Figure 4.12)**.

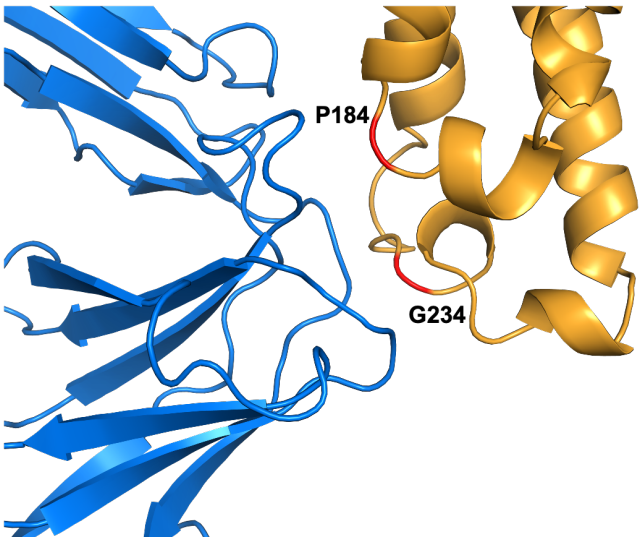


Figure 4.11: Location of the P184 and G234 Residues of the RIFIN Within the Binding Interface The two residues of interest are shown in red.

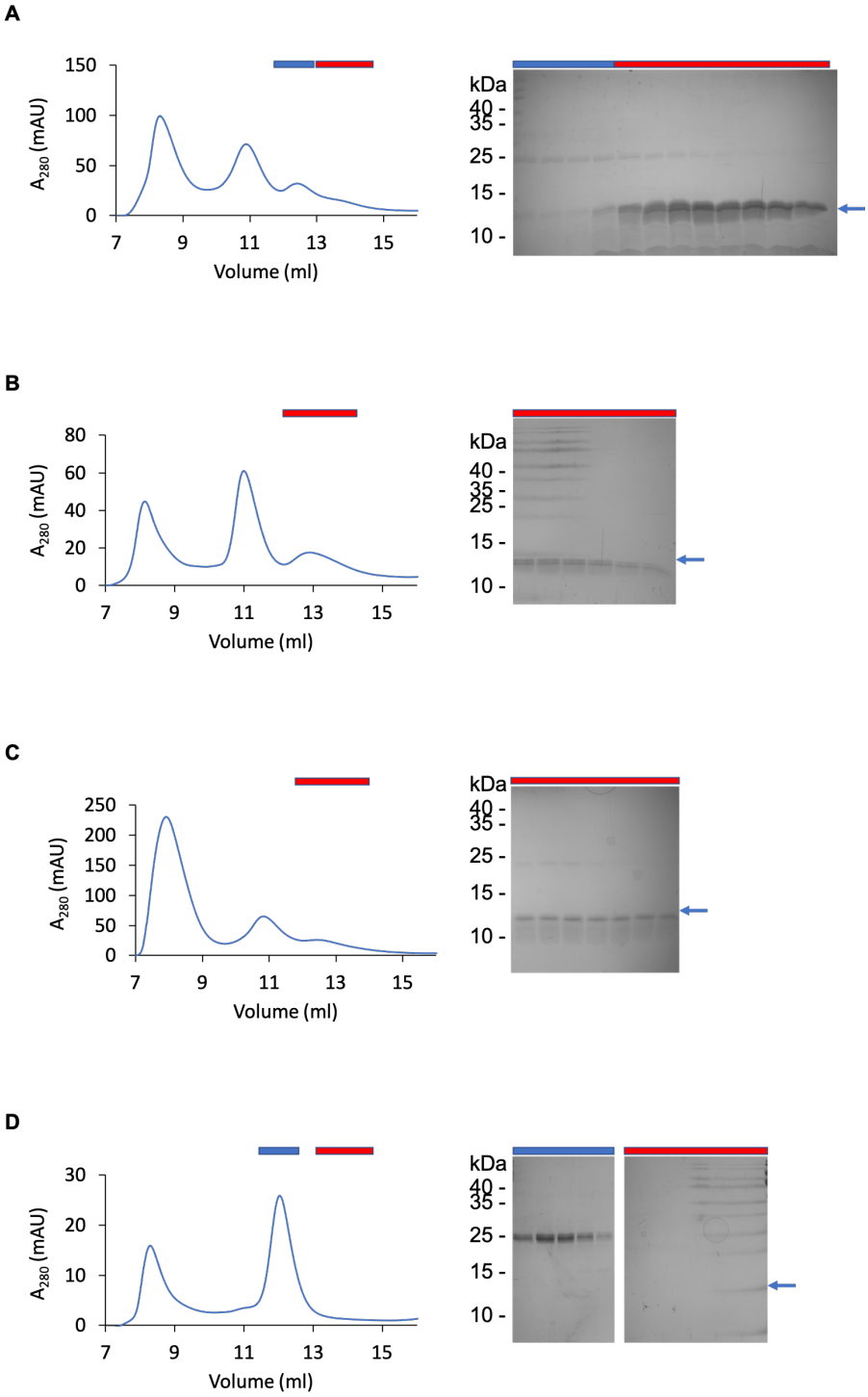


Figure 4.12: Purification of Binding Site Mutants for the RIFIN Variable Region SEC chromatograms and SDS-PAGE gels for the purification of the A) G234R, B) G234W, C) G234Y, and D) P184W mutants of the Pf3D7_1254800 variable region using a Superdex 75 10/300 column. The bars above the SEC traces indicate the fractions shown on the SDS-PAGE gels, and arrows indicate the expected position of bands for the expected molecular weight of around 12.9 kDa. Higher molecular weight bands seen on the left-hand lanes of the SDS-PAGE gel of B), and the right hand lanes of the SDS-PAGE gel of D) are due to bleed-over of markers.

SPR was then carried out using these mutants to test whether the ability to bind LILRB1 had been blocked as expected **(Figure 4.13)**. This demonstrated significantly reduced binding for all 3 mutants, with close to undetectable binding for G234R and G234W. Due to the significantly better yield of the G234R mutant from refolding, this construct was chosen for use during further analysis.

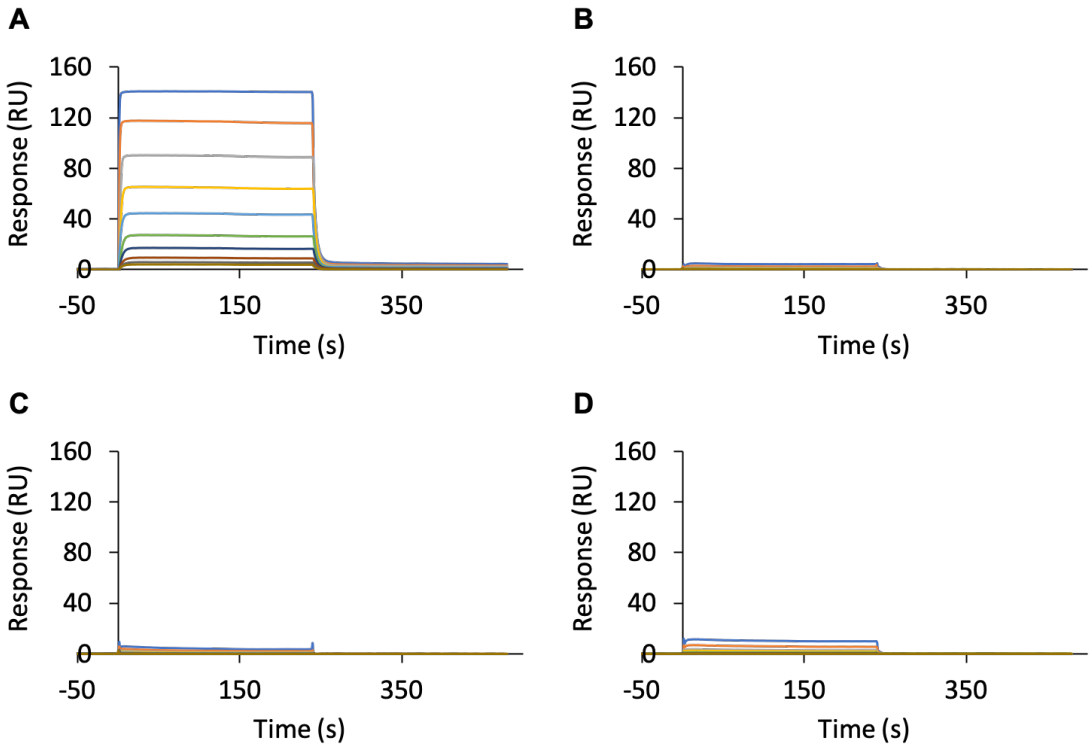


Figure 4.13: Surface Plasmon Resonance Showing Binding of RIFIN Variable Region Mutants to LILRB1 SPR curves showing the Pf3D7_1254800 variable region binding to the LILRB1 ectodomain. The curves shown are for A) C223S, B) C223S+G234R, C) C223S+G234W, and D) C223S+G234Y. A 2-fold dilution series was used for each interaction, from 4 μM to 3.9 nM.

To confirm that this mutation did not significantly alter the fold of the protein, CD was carried out **(Figure 4.14)**. This confirmed that the mutation had not affected the secondary structure of the protein, as expected due to the position of the G234 residue within a loop, where the side chain is facing away from the core of the protein, and due to the residue lying within a region of the Ramachandran plot which is also favourable for arginine residues. CD was not carried out for the G234W

and G234Y due to a shortage of protein, and they were not used for subsequent experiments.

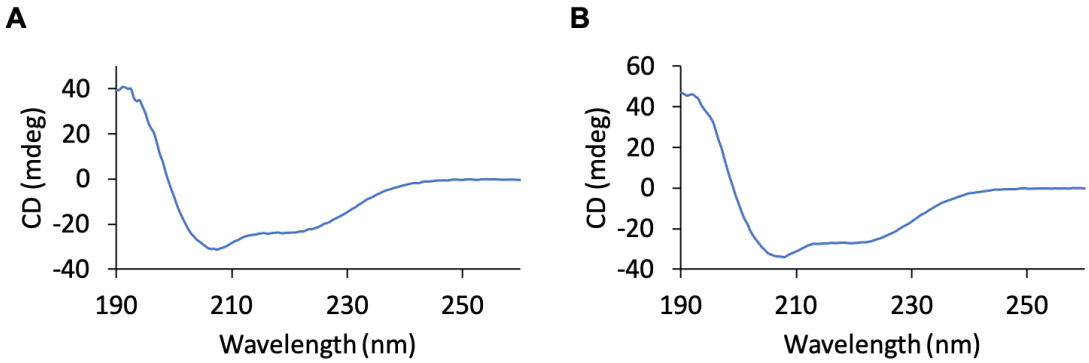


Figure 4.14: Circular Dichroism Spectrum for the G234R Mutant of the RIFIN Variable Region CD spectra comparing the secondary structure of the A) C223S and B) C223S+G234R mutants of the Pf3D7_1254800 variable region.

The effect of this mutation on signalling through LILRB1 was also tested using a cellular assay, the nuclear factor of activated T cells (NFAT)-GFP reporter system[268]. This was carried out by a collaborator, Akihito Sakoguchi in the lab of Hisashi Arase, at the Department of Immunochemistry, University of Osaka. In this assay, the extracellular domain of LILRB1 is fused to the transmembrane and intracellular domains of paired immunoglobulin-like receptor β (PILRβ). Upon activation of LILRB1, the intracellular region of PILRβ associates with the adaptor protein DAP12[269]. This leads to activation of NFAT, resulting in expression of GFP. Signalling through LILRB1, therefore, results in expression of GFP, and so produces a fluorescent signal. This assay had been previously used to demonstrate a signalling response induced by binding of this RIFIN variable region to LILRB1[67].

This assay demonstrated that, while the C223S construct was able to induce strong GFP expression, indicating induction of LILRB1-mediated signalling, the G234R mutation was sufficient to completely block this signalling **(Figure 4.15)**. This again confirms that the G234R mutation is able to block binding of the RIFIN to

LILRB1, indicating that the binding interface seen in the crystal structure is the key interface involved in the induction of signalling through LILRB1 binding.

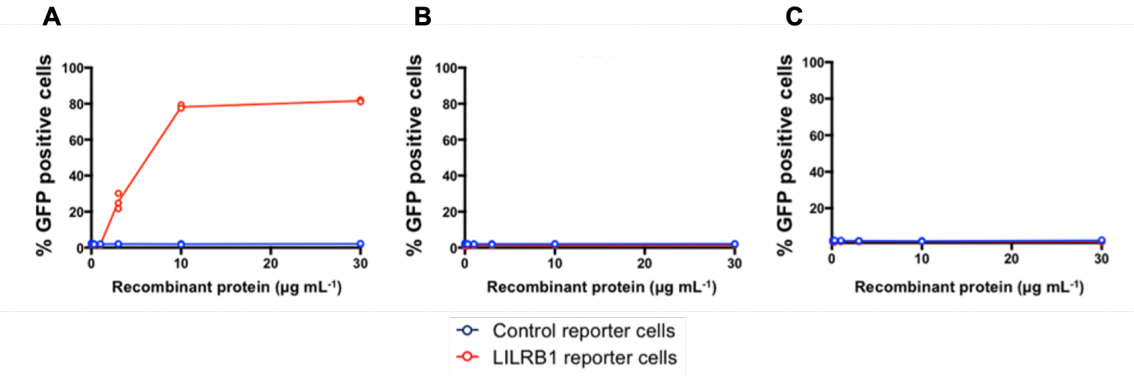


Figure 4.15: NFAT-GFP Reporter Assay Showing the Effect of the G234R Mutation on Signalling Through LILRB1 The effect on signalling of binding of the Pf3D7_1254800 variable region, both A) C223S and B) C223S+G234R, and C) the Pf3D7_1254200 variable region, a RIFIN which does not bind LILRB1[67]. The mean of 3 independent measurements are shown for a range of concentrations.

4.5 The LILRB1-Binding Mode of Other RIFINs

The majority of residues involved in LILRB1-binding by the RIFIN are found within the loop anchored in place by the disulphide bond within the variable region. This disulphide bond was highly conserved both throughout all RIFINs, and within the known LILRB1-binding RIFINs **(Figure 4.16)**. This high level of conservation suggests that the loop could be key for the functionality of these proteins, potentially by mediating binding both for LILRB1-binding RIFINs, and for those binding to other receptors. To test this, mutations were introduced into other LILRB1-binding RIFINs to mimic the G234R mutation used previously. If the binding mode of these RIFINs is the same as that seen in the structure, then introducing a bulky side chain after the second cysteine of the disulphide bond would be predicted to sterically hinder binding in the same way as does G234R.

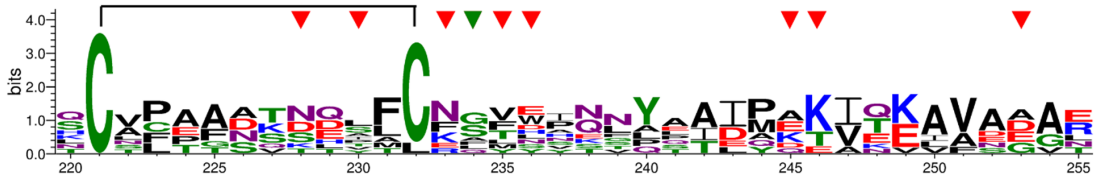


Figure 4.16: A Sequence Logo for Residues from LILRB1-Binding RIFINs corresponding to Residues 220-255 of Pf3D7_1254800 Residues forming interactions with LILRB1 are indicated with red triangles, the G234 residue is indicated with a green triangle, and the disulphide bond is shown as a bracket.

4.5.1 Expression and Purification of the Wild-type and Mutant Variable Regions

Three RIFINs were chosen to test this, with a range of amino acids in the equivalent position in the sequence: Pf3D7_0100200 with the mutation S244R, Pf3D7_0100400 with the mutation L264R, and Pf3D7_1100400 with the mutation G240R. Attempts were then made to produce both the wild type and mutant variable regions for these RIFINs, using the same refolding protocol used for the Pf3D7_1254800 variable region. For both Pf3D7_0100200 and Pf3D7_1100400 this was unsuccessful, but for Pf3D7_0100400, both the wild type and L264R mutant could be produced in a monomeric form **(Figure 4.17)**.

To confirm that the Pf3D7_0100400 variable region was folded correctly, for both the wild type and L264R constructs, CD was carried out. This showed that this protein adopts a primarily α-helical structure, as expected, and that the L264R mutation did not affect this **(Figure 4.18)**.

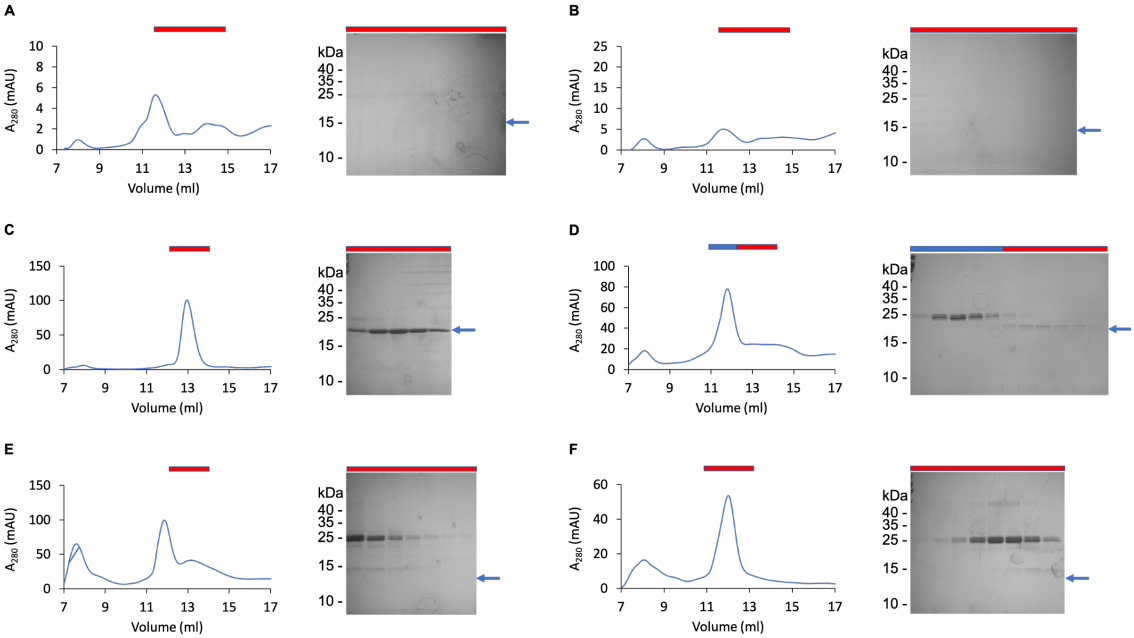


Figure 4.17: Purification of Wild Type and Binding Site Mutants for 3 LILRB1-Binding RIFIN Variable Regions SEC chromatograms and SDS-PAGE gels for purification of A) Pf3D7_0100200 WT and B)S244R, C) Pf3D7_0100400 WT and D) L264R, and E) Pf3D7_1100400 WT and F) G240R using a Superdex 75 10/300 column. The bars above the SEC traces indicate the fractions shown on the SDS-PAGE gels. Arrows indicate the expected position of bands on the SDS-PAGE gels, at around 15 kDa for Pf3D7_0100200, 20 kDa for Pf3D7_0100400, and 13 kDa for Pf3D7_1100400. For Pf3D7_0100400, both WT and L264R, the fractions indicated with the red bar were those pooled for future use.

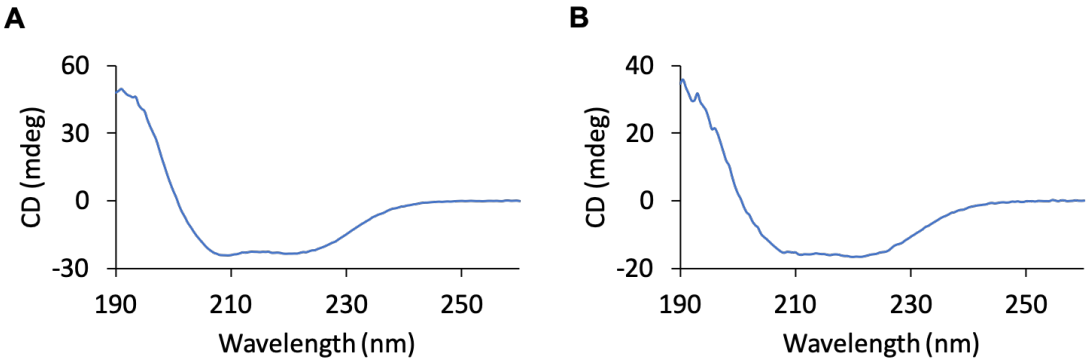


Figure 4.18: Circular Dichroism for the A) Wild Type and B) L264R Mutant of the Pf3D7_0100400 Variable Region

4.5.2 Effect of Predicted Binding-Site Mutations on the Interaction of RIFIN Variable Regions with LILRB1

The effect of the L264R mutation on the interaction with LILRB1 was then assessed by SPR. This showed that the wild type variable region was able to bind to the full-length LILRB1 ectodomain **(Figure 4.19)**. The affinity of this interaction was, however, more than 10-fold lower than binding of Pf3D7_1254800, with a K_d of around 14.5 μM for Pf3D7_0100400 binding to LILRB1 **(Table 4.6)**, compared with around 1 μM for Pf3D7_1254800 binding to the same receptor. The introduction of the L264R mutation was able to successfully disrupt binding **(Figure 4.19)**, which suggests that the Pf3D7_0100400 variable region is using the same loop as Pf3D7_1254800 to form interactions with LILRB1.

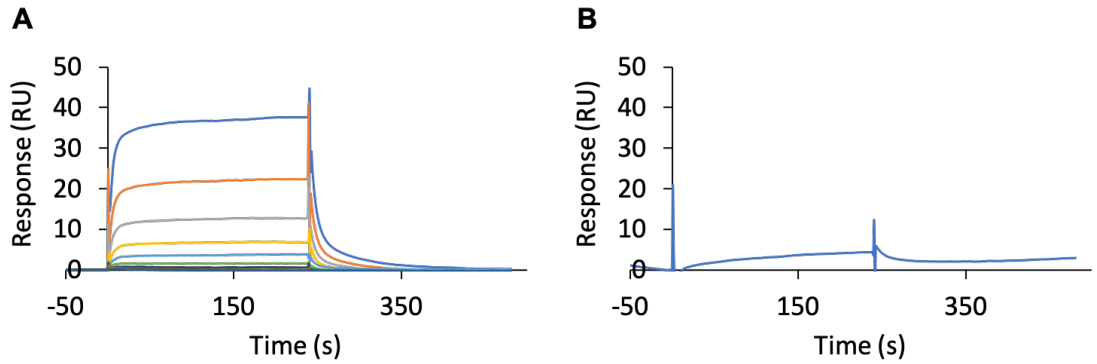


Figure 4.19: Surface Plasmon Resonance for the Binding of the A) Wild Type and B) L264R Mutant of the Pf3D7_0100400 Variable Region to the Full-Length LILRB1 Ectodomain

Surprisingly, SPR also showed that the wild type Pf3D7_0100400 variable region bound to the D3D4 LILRB1 construct, but not to the D1D2 construct **(Figure 4.20)**. This is in contrast to both the Pf3D7_1254800 variable region seen in the structure, and natural ligands of LILRB1, which bind exclusively to the 2 membrane distal domains of the receptor[191–193]. These results suggest, therefore, that while this RIFIN may bind to LILRB1 using the same region of the RIFIN, it is binding to a separate site on LILRB1. Whether binding through domains 3 and 4 is able to result in inhibitory signalling through LILRB1 is not known, and so whether or not this RIFIN is able to provoke the same response as Pf3D7_1254800 through binding to LILRB1 remains unclear.

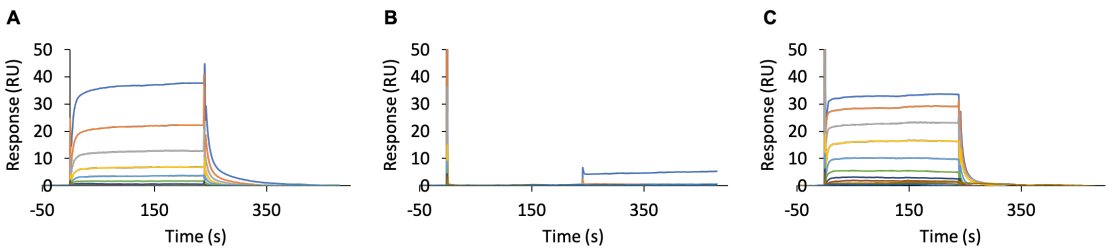


Figure 4.20: Surface Plasmon Resonance for the Binding of the Pf3D7_0100400 Variable Region to 3 LILRB1 Constructs SPR showing binding of the Pf3D7_0100400 variable region to A) full-length, B)D1D2, and C) D3D4 constructs of LILRB1.

	$\mathbf{k}_{on}$ ($\mathbf{M}^{-1}s^{-1}$)	$\mathbf{k}_{off}$ ($\mathbf{s}^{-1}$)	$\mathbf{K}_d$ (μM)	χ^2 ($\mathbf{RU}^2$)
Full-Length	4.10×10^3	$5.93\times10^{-}2$	14.5	0.459
D3D4	5.72×10^4	$8.19\times10^{-}2$	1.43	10.3

Table 4.6: Kinetic Values for Binding of the Pf3D7_0100400 Variable Region to LILRB1 Full-Length and D3D4 These values were calculated using fitting to a one-to-one kinetic binding model, with the quality of fitting described by the χ^2 value.

In addition to these studies using recombinant protein, the conservation of binding mode for LILRB1-binding RIFINs was analysed utilising transgenic parasites expressing either wild type RIFINs, or those incorporating predicted binding site mutations. These experiments were carried out by Akihito Sakoguchi in the lab of Hisashi Arase, University of Osaka. *P. falciparum* 3D7 parasites were transfected with a specific *rif* gene, and erythrocytes infected with these parasites were then tested for their ability to bind to Fc-tagged LILRB1 by flow cytometry.

The binding of three different RIFINs, both with and without predicted binding site mutations, to LILRB1 was tested in this way. As a positive control, Pf3D7_1254800 was used both with and without the G234R mutation described previously. Two other LILRB1-binding RIFINs were also tested, Pf3D7_0100200 with and without the mutation S244R, and Pf3D7_1100400 with and without the mutation G240R. For all 3 RIFINs, the full-length sequence was used.

This gave similar results to the experiments using recombinant protein. All 3 wild-type RIFINs were able to bind to LILRB1, as expected, with LILRB1 binding abolished for all 3 RIFINs when the predicted binding site mutations were introduced **(Figure 4.21)**. This again suggests that, despite the limited sequence identity across the family, the same binding mode is used across LILRB1-binding RIFINs, with binding mediated primarily by the loop stabilised by the conserved disulphide bond.

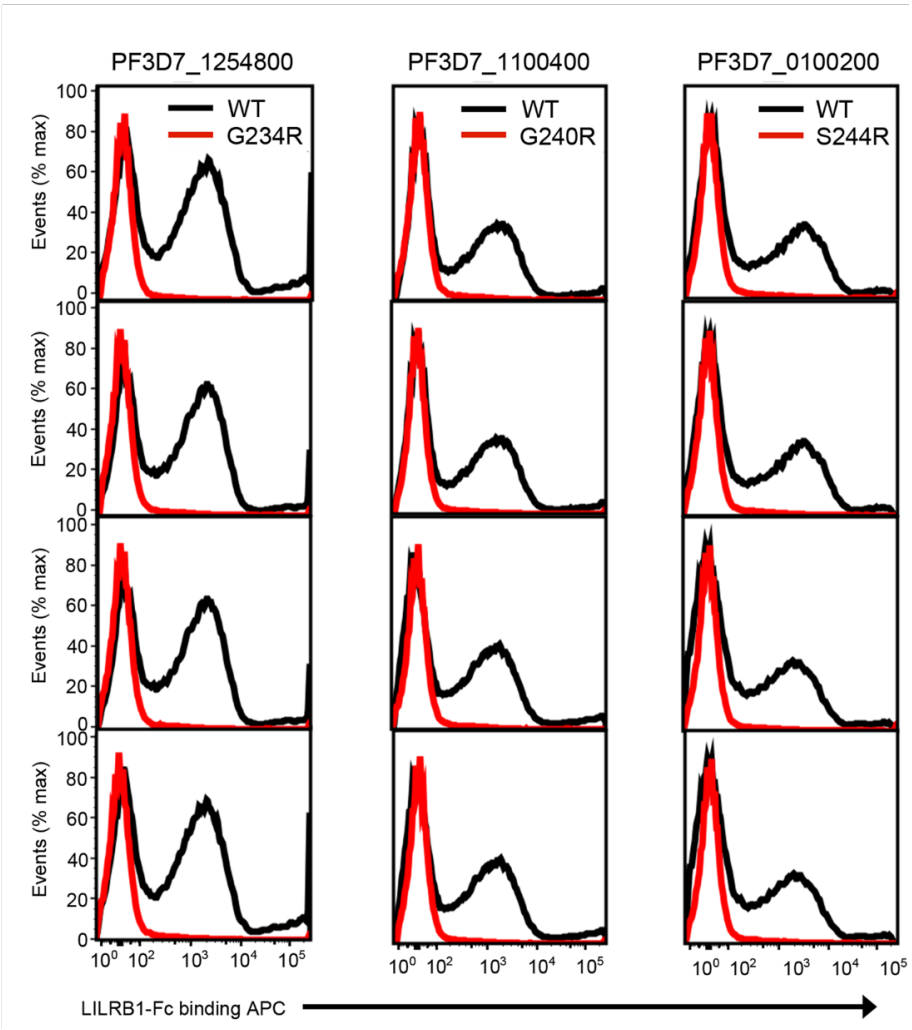


Figure 4.21: Flow Cytometry Analysis of the Binding of LILRB1-Fc to Erythrocytes Infected with Transgenic *P. falciparum* Expressing Specific RIFINs Four independent experiments are shown for each RIFIN, with the results for parasites expressing wild type RIFINs shown in black, and those for parasites expressing mutant RIFINs shown in red.

4.5.2.1 Crystallisation of the Pf3D7_0100400 Variable Region:LILRB1 Complex

To further explore the unexpected interaction between the Pf3D7_0100400 variable region with LILRB1, with binding to the D3D4 region of the ectodomain, as opposed to the usually seen interaction with the D1D2 region, attempts were made to crystallise the complex formed between the RIFIN and the full-length LILRB1

ectodomain.

For crystallisation trials, the Pf3D7_0100400 variable region and full-length LILRB1 ectodomain were mixed at a 1:1 ratio, incubated at room temperature for 20 min, and the complex then purified by SEC **(Figure 4.22)**. Fractions from the peak containing the complex were then pooled, and concentrated to 10 mg/ml for crystallisation trials.

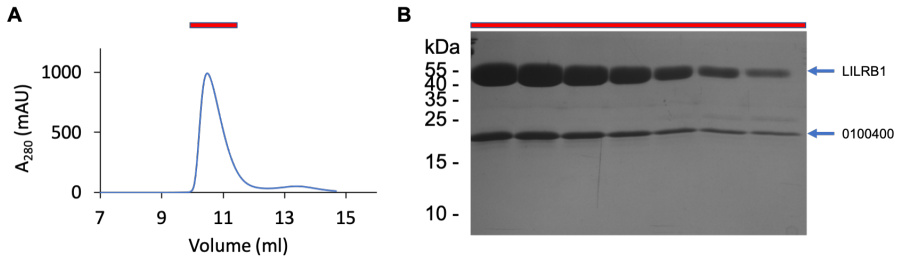


Figure 4.22: Purification of the Pf3D7_0100400 Variable Region:LILRB1 Complex A) SEC chromatogram showing purification of the Pf3D7_0100400 variable region:LILRB1 full-length ectodomain complex using a Superdex 75 10/300 column, and B) SDS-PAGE gel showing fractions corresponding to the red bar above the SEC trace which were pooled for crystallography.

A range of crystallisation screens were trialled at both 4°C and 18°C, and crystals appeared after around two weeks in a condition containing an amino acid mix (20 mM DL-Glutamic acid monohydrate, 20 mM DL-Alanine, 20 mM Glycine, 20 mM DL-Lysine monohydrochloride, and 20 mM DL-Serine), 20% v/v PEG 500 MME, 10 % w/v PEG 20000, 61.2 mM MES, and 38.8 mM Imidazole, at a pH of 6.5, and an incubation temperature of 18°C **(Figure 4.23)**.

As these crystals were quite small, less than 100 μm in length, and very thin, generally less than 10 μm, they were challenging to both harvest and collect data from, and so an optimisation plate was also set up to attempt to produce larger crystals. This plate included a protein concentration gradient, from 6.5 mg/ml to 10 mg/ml, and a gradient of PEG500 MME and PEG2000, from 12.7% v/v and 6.3 % w/v to 20% v/v and 10% w/v respectively. Crystals only formed at the higher precipitant

concentrations, with protein concentration having little effect on either the size or number of crystals, although all crystals on this plate were still very thin, again usually less than 10 μm, so further optimisation is required.

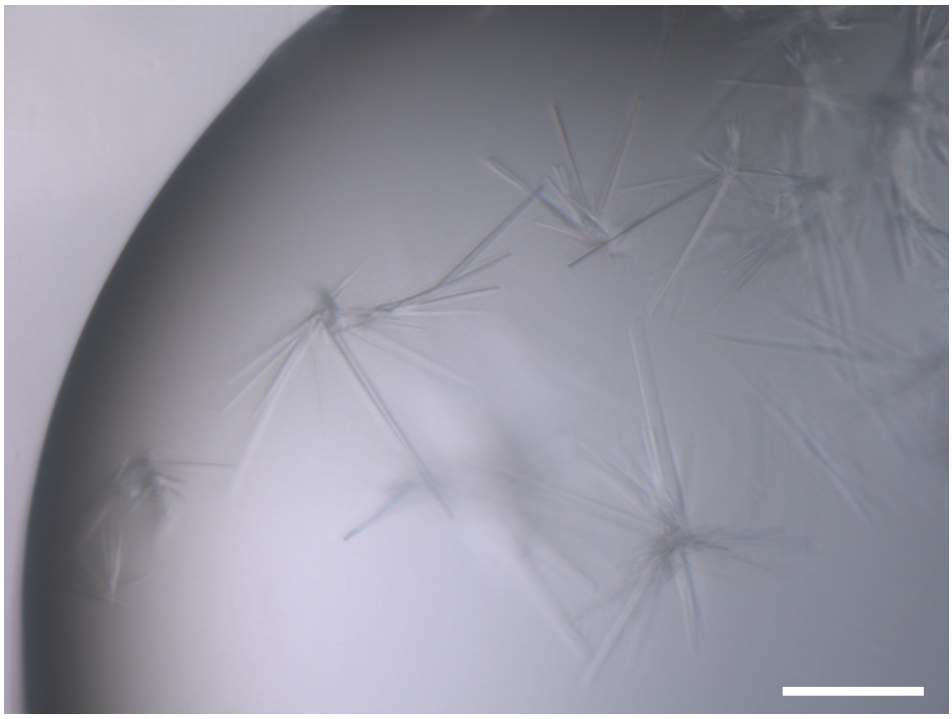


Figure 4.23: Crystals of the Pf3D7_0100400 Variable Region:LILRB1 Complex These crystals grew after two weeks in a condition containing an amino acid mix (20 mM DL-Glutamic acid monohydrate, 20 mM DL-Alanine, 20 mM Glycine, 20 mM DL-Lysine monohydrochloride, and 20 mM DL-Serine), 20% v/v PEG 500 MME, 10 % w/v PEG 20000, 61.2 mM MES, and 38.8 mM Imidazole, at a pH of 6.5, and an incubation temperature of 18°C, with a protein concentration of 10 mg/ml. The scale bar indicates a length of 100 μm.

4.6 Comparison with the interaction between MHC Class I and LILRB1

The previously mentioned cellular GFP-reporter assay suggests that RIFIN binding to LILRB1 is capable of inducing signalling through the receptor. This is despite the complete lack of both sequence and structural similarity between the RIFIN variable region and the physiological ligand of LILRB1, MHC class I molecules. Both

the α chain and β_2-microglobulin (β_2M) chain of MHC complexes are primarily composed of β-strands, while the RIFIN variable region is almost entirely α-helical. Despite this, there are a number of strong similarities between the interactions the two ligands form with LILRB1.

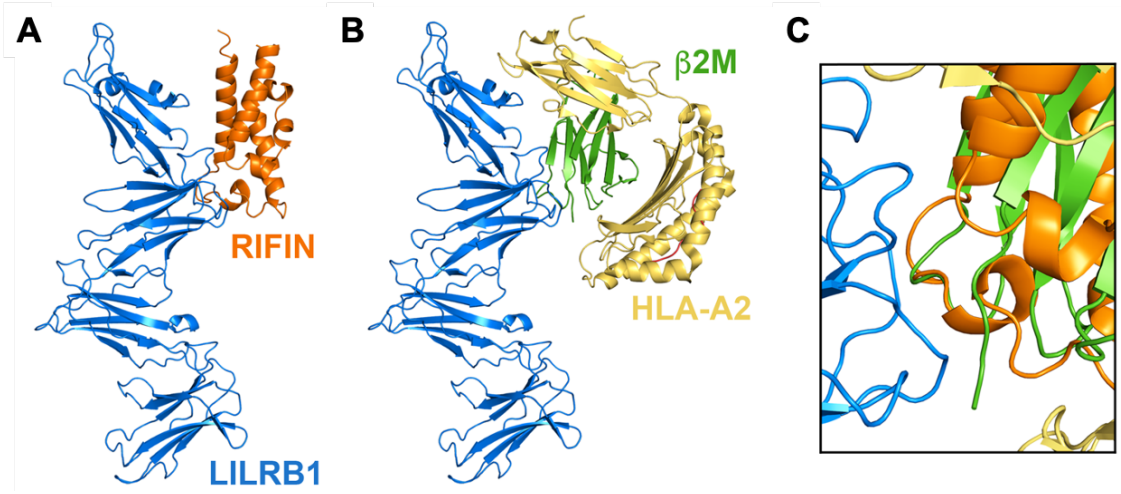


Figure 4.24: A Structural Comparison of Binding to LILRB1 by a RIFIN Variable Region and an MHC Class I Complex A) The RIFIN variable region:LILRB1 complex structure, B) The HLA-A2:LILRB1 complex structure (PDB code: 6EWA[194]), and C) a close up of the alignment of the 2 complexes.

To compare the interaction of the RIFIN with that of MHC complexes, the structure of the HLA-A2:LILRB1 complex, PDB entry 6EWA[194], was aligned to that of the RIFIN:LILRB1 complex **(Figure 4.24)**. This demonstrates the shared binding site for the 2 ligands on the receptor, with both binding to the hinge region between the 1^{st} and 2^{nd} domains of LILRB1, and with both forming binding sites of similar areas, around 750 Å^2 for the RIFIN, and 833 Å^2 for HLA-A2.

In addition to this, a large number of interactions are shared between the two complexes. All interactions formed with LILRB1 by HLA-A2 are hydrogen bonds, and the vast majority of these are through residues of the β_2M chain **(Table 4.7)**. For the majority of these interactions, equivalent hydrogen bonds are formed by the RIFIN variable region, with 8 of the 11 functional groups of LILRB1 which form interactions with the β_2M chain of HLA-A2 also forming interactions with the

RIFIN.

Additionally, the structure of the RIFIN loop that forms a number of the interactions with the receptor bears strong similarity to that of the F-G strand loop of the β_2M chain **(Figure 4.24C)**. This section of loop is involved in the formation of multiple backbone-mediated hydrogen bonds in both structures, with G234 and N236 of the RIFIN, forming near identical interactions with I100 and A98 of LILRB1 respectively as residues T86 and S88 of β_2M.

These factors combined suggest that the RIFINs, despite a total lack of sequence similarity to MHC class I molecules, are mimicking the mode of binding of the physiological ligand, potentially explaining the ability of the RIFIN to modulate signalling through binding to LILRB1.

HLA-A2		**LILRB1**		
β_2M Chain:				
Residue	**Group**	**Residue**	**Group**	**Interaction**
Q2	Side Chain O	Q125	Side Chain NH_2	Hydrogen Bond
	Backbone NH	Q125	Backbone O	Hydrogen Bond
	Backbone O	A127	Backbone NH	Hydrogen Bond
T86	Backbone O	I100	Backbone NH	Hydrogen Bond
S88	Side Chain OH	G97	Backbone O	Hydrogen Bond
	Backbone NH	A98	Backbone O	Hydrogen Bond
Q99	Side Chain NH_2	Q18	Side Chain O	Hydrogen Bond
K91	Side Chain NH_2	E184	Side Chain O	Hydrogen Bond
I92	Backbone O	W67	Side Chain NH	Hydrogen Bond
K94	Side Chain NH_2	E68	Side Chain O	Hydrogen Bond
M99	Backbone O	K42	Side Chain NH_2	Hydrogen Bond
α3 Domain:				
Residue	**Group**	**Residue**	**Group**	**Interaction**
A193	Backbone O	T43	Side Chain OH	Hydrogen Bond
D196	Backbone NH	Y79	Side Chain OH	Hydrogen Bond
A193	Side Chain OH	K41	Backbone O	Hydrogen Bond

Table 4.7: Table of Interactions between HLA-A2 and LILRB1
These interactions are seen in PDB entry 6EWA, and are separated into those formed by the β_2-microglobulin chain and those formed by the α3 domain of the MHC class I complex. Residues of LILRB1 and interactions highlighted in red are those also seen in the RIFIN variable region:LILRB1 structure.

4.6.1 Analysis of the NK Cell Immunological Synapse using TIRF Microscopy

To further explore the similarity between the interactions of the RIFIN and HLA-A2 with LILRB1, and to assess whether a shared mechanism is used to stimulate LILRB1-mediated signalling, total internal reflection fluorescence (TIRF) microscopy was used to image the interaction between natural killer (NK) cells and a supported lipid bilayer (SLB) upon which the ligands had been immobilised.

This allowed for the analysis of the recruitment of both the ligands and the receptor to a mimic of the immunological synapse formed by NK cell contact formation with a target, as well as of the activation, or inhibition, of NK cell activity upon contact formation.

This microscopy was carried out by Alex Morch and James Felce of Mike Dustin's group at the Kennedy Institute of Rheumatology, University of Oxford, using RIFIN which I produced, and HLA-A2 provided by Edward Jenkins of Simon Davis's group at the MRC Weatherall Institute of Molecular Medicine, University of Oxford.

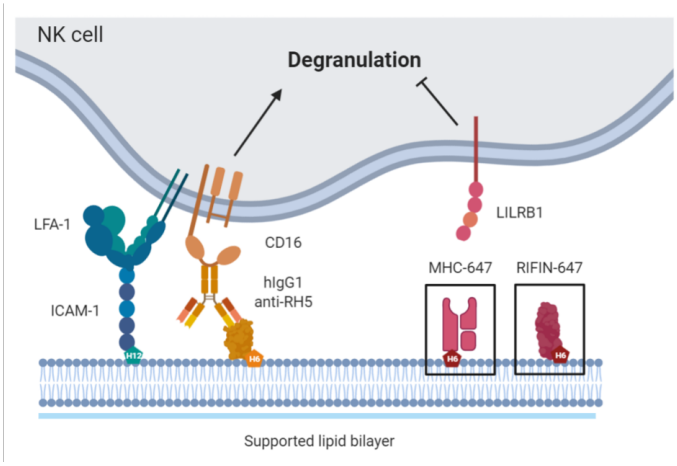


Figure 4.25: Supported Lipid Bilayer Setup for TIRF Microscopy Figure created using BioRender.

NK cells were chosen for this analysis due to their high level of LILRB1 expression, as well as their relevance to the blood stage of a malaria infection, where NK cell

activation is believed to occur through antibody dependent cell-mediated cytotoxicity (ADCC)[155,270]. To mimic the contacts formed by NK cells during ADCC, the SLB was coated with multiple components **(Figure 4.25)**. NK cells are triggered by the binding of CD16 on their surface to the Fc region of IgG upon binding to a target[271]. To mimic this interaction, PfRH5 was coupled directly to the SLB, and an anti-RH5 human IgG, R5.016[272], then added. This allowed the Fc region of the antibody to be displayed directly for CD16 binding. While RH5 itself is not physiologically relevant here, it had no effect on NK cell activation **(Figure 4.26)**, while providing a target for human IgG to allow display of the Fc region. ICAM-1 was also coupled to the SLB, which forms an adhesive interaction with LFA-1 on the NK cell surface[273]. Finally, equal densities of fluorescently labelled ligand were coupled, at 100 molecules/μm^2, with the non-binding G234R RIFIN mutant also tested. As a control, SLBs were coated with ICAM-1, PfRH5, and human IgG, but no additional ligand. NK cells isolated from 3 independent donors were incubated with the SLBs, which were then fixed, and stained for either LILRB1, to analyse the accumulation of the receptor at the immunological synapse, or perforin, to assess the degree of activation of the cells upon contact formation[274].

4.6.1.1 Effect of RIFIN Binding on RIFIN and LILRB1 Accumulation at the Immunological Synapse

A common mechanism of immune receptor regulation is through regulation of recruitment of the receptor to the immunological synapse[275,276]. It is not known, however, if this mechanism is key to LILRB1 signalling, and so LILRB1 was fluorescently labelled with anti-LILRB1 antibodies to allow for its localisation to be assessed after the binding of either the RIFIN variable region or MHC class I **(Figure 4.27)**. As these ligands were also fluorescently labelled, their recruitment was also visualised.

Imaging the contacts formed at the SLB by NK cells showed a significant accumulation

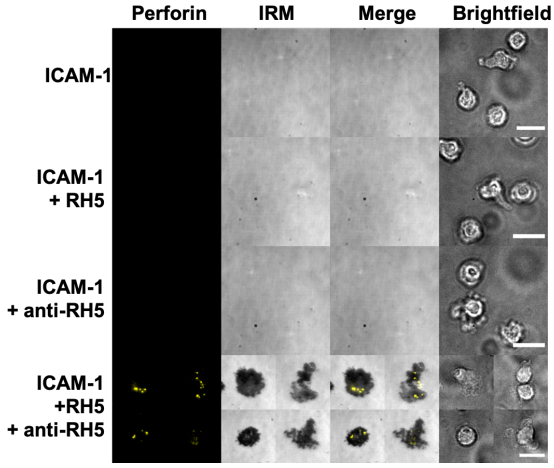


Figure 4.26: Effect of RH5 and anti-RH5 Antibody on NK Cell Activation NK cell contact formation and perforin release (yellow) in the presence of ICAM-1, RH5, and the anti-RH5 antibody, R5.016, demonstrating that both RH5 and R5.016 are necessary and sufficient for contact formation and cell activation. The scale bars indicate a length of 10 μm.

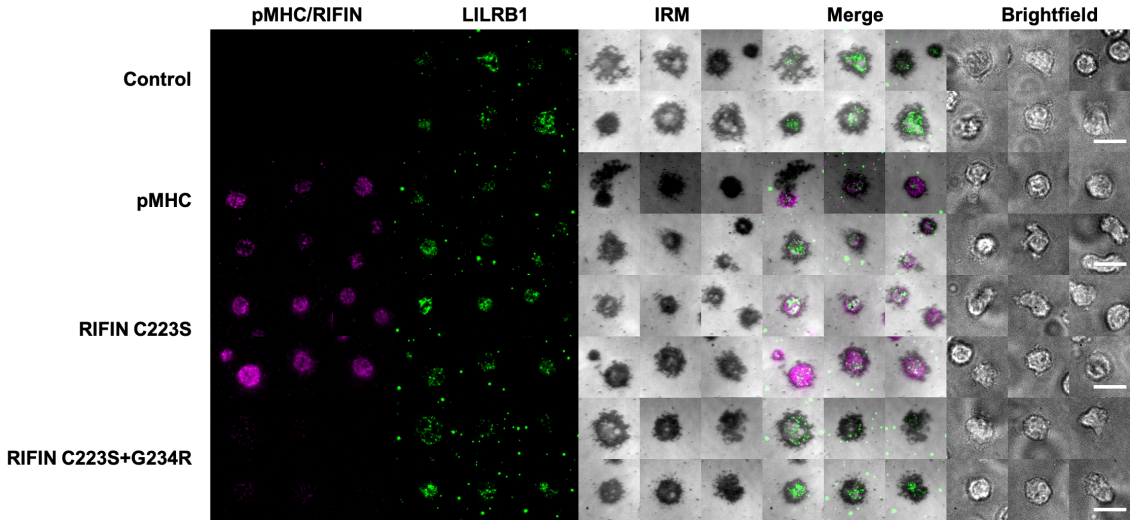


Figure 4.27: Images of RIFIN, MHC, and LILRB1 Localisation at the Immunological Synapse Formed Between NK Cells and a Supported Lipid Bilayer MHC, wildtype RIFIN, and the G234R mutant are shown in magenta, and LILRB1 in green. The scale bars indicate a length of 10 μm.

of both MHC and the RIFIN C223S construct at the contact sites, with almost no G234R enrichment seen **(Figure 4.28)**, as expected due to the inability of the G234R mutant to bind to the receptor. Positive, although not complete, co-localisation with LILRB1 was seen for both the MHC complex and RIFIN C223S, significantly above that for the G234R mutant. These results together suggest that the MHC and RIFIN both localise to the synapse, and this can be prevented by blocking binding to LILRB1. By contrast, no significant change in LILRB1 accumulation at the immunological synapse was observed upon either MHC or RIFIN binding when compared to the control. There was a small change in LILRB1 accumulation in the presence of the G234R mutant as compared to RIFIN C223S, significant to $P<0.05$, this difference is slight, and there is no significant change from the control in the presence of either RIFIN C223S or G234R. These results suggest that recruitment to the immunological synapse is not a key mechanism for regulation of LILRB1-mediated signalling.

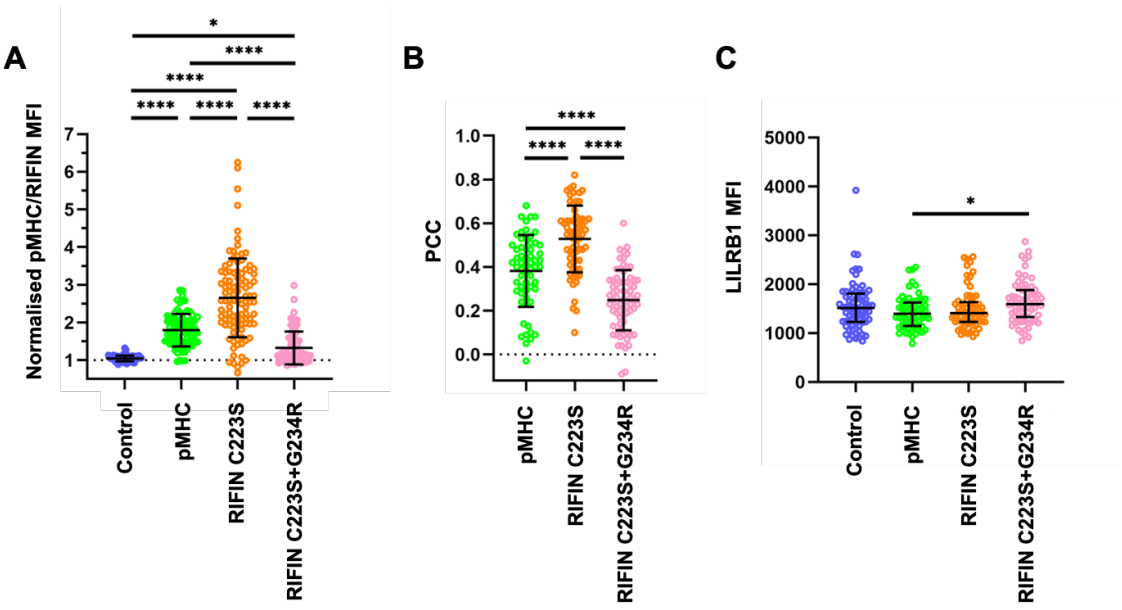


Figure 4.28: Analysis of Quantity and Co-Localisation of the RIFIN Variable Region, MHC, and LILRB1 at the NK Cell Immunological Synapse A minimum of 15 cells from each of 3 donors per condition were analysed. A) The quantity of RIFIN C223S, RIFIN C223S+G234R, and MHC at the contact area. B) The degree of co-localisation of RIFIN C223S, RIFIN C223S+G234R, and MHC with LILRB1. C) The quantity of LILRB1 at the contact area in the presence of RIFIN C223S, RIFIN C223S+G234R, and MHC. Tukey's test was carried out for A and B, with the mean ± 1 SD shown, and Dunn's test was carried out for figure C, with the median and interquartile range shown.

4.6.1.2 Effect of RIFIN Binding on Perforin Release

Activation of the NK cells was monitored through the labelling and observation of perforin accumulation at the contact sites **(Figure 4.29)**.

Perforin is a pore-forming protein which is recruited to the immunological synapse upon activation of NK cells[274], and forms pores in target cells, allowing cytotoxic proteases, such as granzyme B, to be released into the target cell cytoplasm[277]. Accumulation of perforin at the contact site formed between NK cells and the SLB therefore acts as a marker of NK cell activation.

There was a significant reduction in perforin accumulation in the presence of both

MHC and RIFIN C223S when compared to both the control, and the G234R mutant **(Figure 4.30)**. This indicates a clear inhibition of NK cell activity by both MHC class I and the RIFIN, with this effect removed through the disruption of LILRB1 binding.

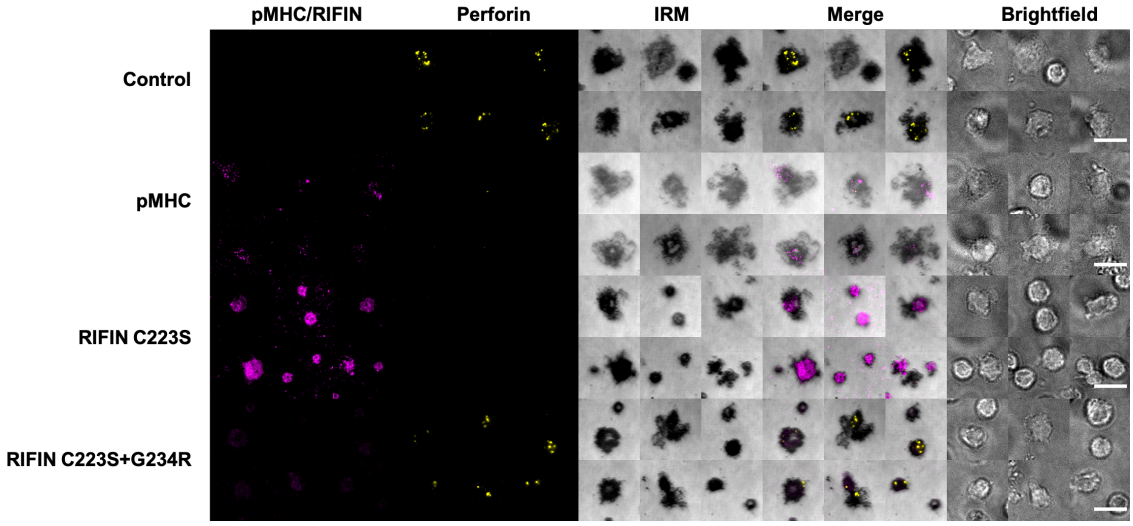


Figure 4.29: Images of Perforin Accumulation at the Immunological Synapse Formed Between NK Cells and a Supported Lipid Bilayer MHC, RIFIN C223S, and RIFIN C223S+G234R are shown in magenta, and perforin in yellow. The scale bars indicate a length of 10 μm.

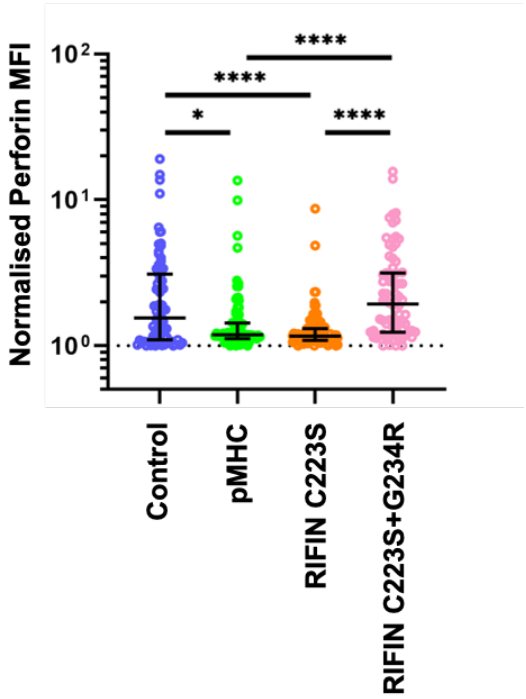


Figure 4.30: Analysis of the Quantity of Perforin at the NK Cell Immunological Synapse in the Presence of the RIFIN Variable Region and MHC A minimum of 15 cells from each of 3 donors per condition were analysed. Dunn's test was carried out, with the median and interquartile range shown.

These results combined support the claim that RIFINs are mimicking the mechanism of binding to LILRB1 used by MHC class I molecules, with both similar interactions at the binding interface, and similar behaviour and effect at the immunological synapse.

4.7 Discussion

Crystallisation of the RIFIN variable region in complex with the full-length ectodomain of LILRB1 has allowed for the determination of the first structure of part of a RIFIN, as well as its interaction with the inhibitory immune receptor. This revealed that the RIFIN, which adopts an almost entirely α-helical fold, interacts with LILRB1 primarily through a long loop anchored in place by the single disulphide bond within

the variable domain.

These interactions appear to be direct molecular mimicry of the physiological binding of the MHC class I complex to LILRB1, with binding loops of both ligands adopting remarkably similar conformations, and the network of hydrogen bonds formed by the RIFIN directly mimicking many of those formed by the β_2M chain of the MHC complex. This similarity in binding could help to explain the elevation in serum β_2M levels in malaria patients[278]. Excess β_2M may be able to block binding to LILRB1 by RIFINs, thereby reducing the effect of RIFIN binding on the immune response.

The similarity with the interaction of the MHC complex with LILRB1 was further seen in a cellular context, with the RIFIN appearing to stimulate LILRB1-mediated signalling in NK cells in a similar way to MHC class I. Here, both the RIFIN and MHC were recruited to the immunological synapse, with disruption of LILRB1 binding preventing this from occurring, and this binding allows for significant inhibition of perforin accumulation at the synapse, indicating strong inhibition of the NK cell response. While the RIFIN is not unique in its ability to mimic binding of MHC molecules to LILRB1, with viruses such as human cytomegalovirus also producing LILRB1-binding proteins[195] for immune modulation[196], the RIFIN differs in achieving this despite its complete lack of either sequence or structural homology with the native ligand.

A number of questions surrounding the RIFINs remain unanswered however. It appears likely that the overall fold of the RIFIN variable region is conserved, as high confidence homology models could be produced for 5 of the 10 sequences tested. High confidence models could not be produced, however, for the remaining sequences, likely due to the lack of sequence identity across the family, and this does make it challenging to conclusively state that the RIFIN variable region fold is conserved. Structural studies of other members of the family would help to confirm the degree of structural homology.

It also appears that the binding mode of RIFINs to LILRB1 does remain conserved, with point mutations in the predicted binding loop sufficient to disrupt binding to the receptor in four sequence diverse RIFINs, assessed using both recombinant protein and transgenic parasites. This comes despite extensive sequence diversity across residues shown in the structure to form interactions with LILRB1. This could suggest a similarity with ligand-binding by the PfEMP1 family, where the binding sites are also often highly sequence diverse, but retain chemical and structural properties to enable binding to a common ligand[134–136].

This sequence diversity has made it difficult to identify key binding motifs for LILRB1 binding, and so prediction of LILRB1-binding remains a challenge. In addition to this, the binding of the Pf3D7_0100400 variable domain to the D3D4 region of LILRB1 raises the question of whether the mechanism of LILRB1 activation by RIFINs is conserved, as no previous ligand for the receptor has been shown to bind to this region of LILRB1. Further structural studies of other RIFINs, both LILRB1-binding and those which bind other receptors, are necessary, therefore, to identify whether this fold is conserved, and whether a common mechanism of receptor activation is used by the RIFIN family.

Crystals have been produced for the Pf3D7_0100400 variable domain in complex with the LILRB1 ectodomain, the structure of which will provide some answers to these questions. They are, however, so far too small for data collection, and so significant optimisation is needed to improve the quality of these. This includes further optimisation of precipitant and protein concentrations, and the pH of the crystallisation solution, as well as other methods including seeding new plates with the existing crystals to provide nucleation sites for crystal growth. This would allow lower protein concentrations to be used, so that excessive nucleation can be avoided, which is currently allowing the large number of small crystals to form[279].

Additionally, it remains to be seen whether the binding of Pf3D7_0100400 to the

D3D4 region of LILRB1 is able to induce signalling through LILRB1 in the same way as the binding of MHC class I molecules or the Pf3D7_1254800 variable region to the D1D2 region. Testing signalling through the NFAT-GFP reporter system would allow this to be determined, as would studying perforin release at the immunological synapse formed by NK cells using the supported lipid bilayer system.

Chapter 5

The Structure of a CIR

5.1 Summary

The next major aim of this D.Phil was to structurally characterise a member of the PIR family, to allow for analysis of structural conservation across the PIRs, as well as to assess whether the PIR and RIFIN structures are homologous. To achieve this a CIR, a member of the PIR family from *P. chabaudi*, was expressed and crystallised, allowing for structure determination, and subsequent bioinformatic analysis of the family.

To determine the structure of a member of the CIR family, expression trials were first carried out, allowing for production of recombinant protein using a mammalian cell system. Crystallisation trials were then carried out for multiple representatives of the family, allowing a 2.15 Å structure to be solved using sulphur-SAD phasing.

This structure, the first for a member of the PIR family, revealed that the CIR adopts a novel, primarily α-helical fold. Sequence analysis and homology modelling suggests that this is likely to be conserved across the CIR family, while the surface of the fold shows a high degree of diversity, as expected for a family of variant surface antigens.

Similar analysis suggested conservation of the fold across the PIR superfamily, from both rodent- and human-infective *Plasmodium* species, while the RIFIN and

STEVOR families from *P. falciparum*, along with putative PIR proteins from the ape-infective *P. reichenowi* and the avian-infective *P. gallinaceum* are confirmed to not be members of the PIR superfamily.

5.2 Expression and Purification of CIRs

Expression trials were carried out using HEK293F cells, using constructs representing the full ectodomain provided by Jean Langhorne, Francis Crick Institute. Constructs for 7 His-tagged CIRs were provided, along with an additional 8 constructs where a C-terminal CD4-tag was added to boost expression. All constructs had predicted glycosylation sites removed through mutation of the serine or threonine residue of the NxS/T motif to alanine.

Small scale expression trials showed strong expression for 5 constructs both with and without the CD4-tag **(Figure 5.1)**, and so all future expression was using those constructs without the tag, to remove the need for an additional cleavage step to remove the tag prior to any crystallisation trials.

For further analysis the 3 shortest constructs that expressed to high levels, believed to be the most likely to crystallise due to potentially shorter, and so less flexible, loops, were chosen. The chosen constructs were residues 1-264 of PCHAS_1200500, with an expected molecular weight of 31.1 kDa, residues 1-235 of PCHAS_0007200, with an expected molecular weight of 28.9 kDa, and residues 1-296 of PCHAS_0400300, with an expected molecular weight of 32.0 kDa.

These 3 constructs were then expressed and purified on a large scale, with purification by Ni-NTA affinity chromatography and SEC **(Figure 5.2)**. Purification of each of the 3 constructs by SEC produced peaks at expected volumes for their size, indicating monomeric, well-folded protein.

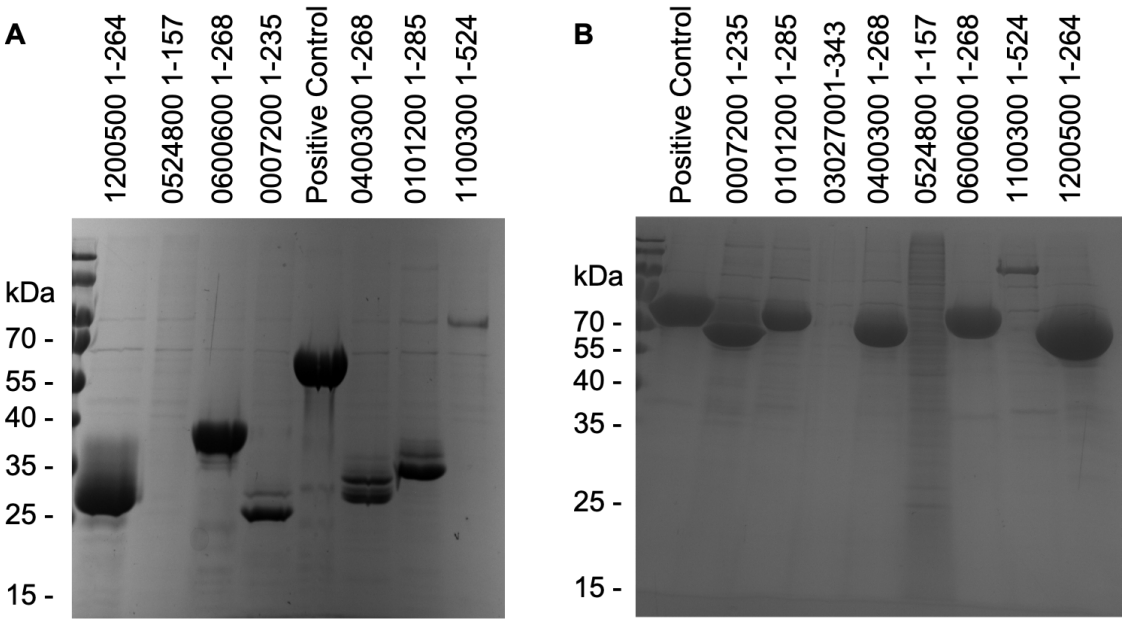


Figure 5.1: Expression Trial of CIR Ectodomain Constructs in HEK293F Cells A) without and B) with a C-terminal CD4-tag. Strong bands could be seen for 5 of the CIRs, both without and with the CD4-tag, at the expected molecular weights of between 20 kDa and 52 kDa for those without the tag, and between 44 kDa and 84 kDa for those with the tag. The positive control used was the CIR construct, PCHAS_0600600 1-268 with a CD4-tag.

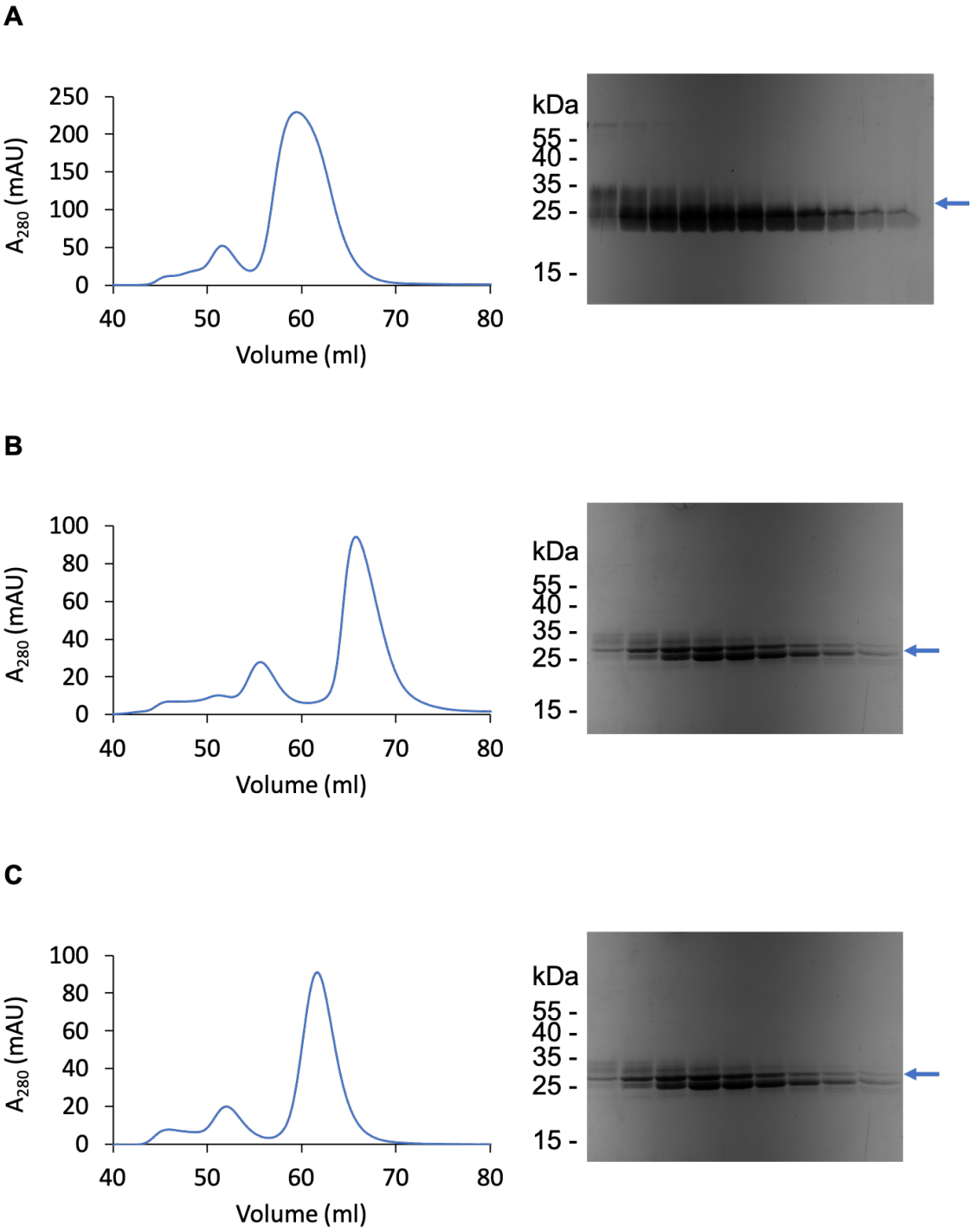


Figure 5.2: Purification of 3 CIRs from HEK293F Cells SEC chromatogram and SDS-PAGE gels for purification from HEK293F cells for A) PCHAS_1200500, B) PCHAS_0007200, and C) PCHAS_0400300 using a Superdex 75 10/300 column. The red bars above the SEC traces correspond to the fractions shown in the SDS-PAGE gels. The expected position of bands on the SDS-PAGE gels at 31.1 kDa, 28.9 kDa, and 32.0 kDa respectively for PCHAS_1200500, PCHAS_0007200, and PCHAS_0400300 are indicated by arrows.

5.3 Assessment of the Fold of CIRs

To determine the secondary structure adopted by the CIRs, CD spectroscopy was carried out. This showed that all 3 proteins adopted strongly α-helical structures, with the characteristic minima at both 208 and 222 nm **(Figure 5.3)**.

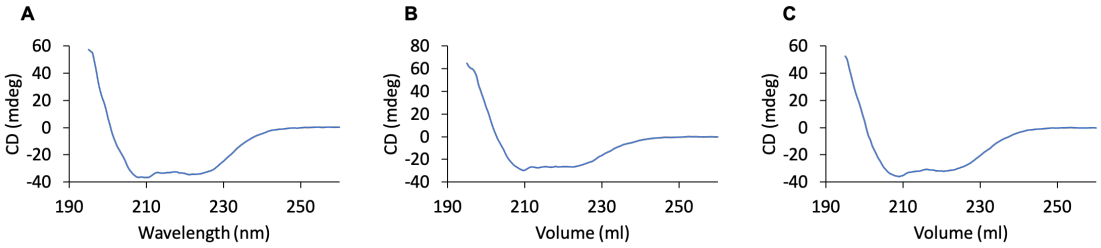


Figure 5.3: Circular Dichorism Spectra for 3 CIR Ectodomains
A) PCHAS_1200500, B) PCHAS_0007200, and C) PCHAS_0400300

5.4 CIR Crystallisation, Data Collection, and Data Processing

5.4.1 Crystallisation of the CIR, PCHAS_1200500

For structural characterisation of the CIRs, extensive crystallisation trials were carried out. These included broad screens at concentrations ranging from 6 to 40 mg/mL, at both 4°C and 18°C incubation temperatures, using both methylated and native protein, for all 3 of the chosen CIRs.

Crystals formed for PCHAS_1200500 after around 40 days in a condition from the MIDAS screen (Molecular Dimensions) **(Figure 5.4)**. The well solution for this condition consisted of 0.1 M HEPES pH 6.5, 45% w/v poly(acrylic acid sodium salt). A protein concentration of 34 mg/mL was used, after methylation of surface lysine residues to reduce flexibility and limit disruption of crystal contact formation.

To optimise these crystals, to both produce a large number for data collection and

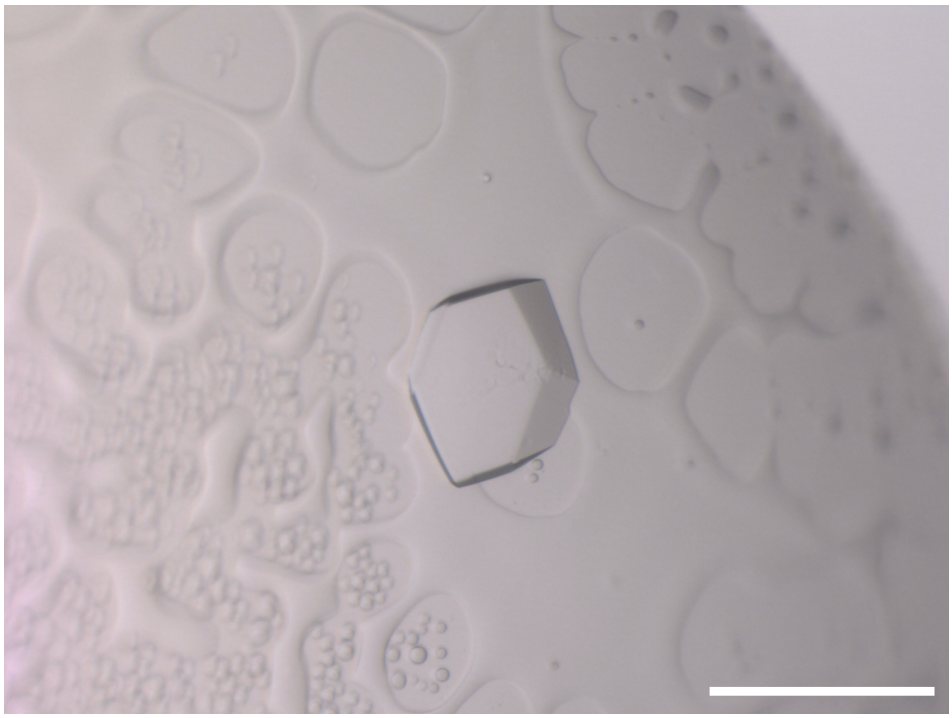


Figure 5.4: A Crystal of the CIR, PCHAS_1200500 These crystals grew after around 40 days in 0.1 M HEPES pH 6.5, 45% w/v poly(acrylic acid sodium salt), with methylated protein at 34 mg/mL and an incubation temperature of 4°C. The scale bar indicates a length of 100 μm.

increase the size and quality of the crystals, a number of optimisation strategies were trialled. These included gradients of protein concentration, pH, and precipitant concentration, as well as the addition of an additive screen, the Silver Bullets screen (Hampton Research). This is a screen of small molecules and excipients which have been previously shown to promote lattice formation and crystallisation. In addition to this, seeding was tested, where the original crystals are harvested and vortexed in the presence of a seed bead, to break up the crystals. These crystal fragments can then be added to the well solution, where they provide nucleation points to promote crystal formation. None of these strategies, however, succeeded in improving the quality of crystals produced in this case.

For data collection, a number of cryoprotection strategies were trialled, including soaking the crystals in drops of well solution supplemented with ethylene glycol and

glycerol to a concentration of 25% v/v. The poly(acrylic acid sodium salt) in the crystallisation condition had not been previously reported as cryoprotective. Crystals were, however, also harvested without additional cryoprotection to test this, and these diffracted to the highest resolution, so for final data collection no additional cryoprotectant was used prior to cyrocooling.

5.4.2 Data Collection and Processing

Data were collected at the ID23 beamline at ESRF, with the data indexed and scaled using GrenADES fastproc[280], giving a dataset at a resolution of 2.15 Å. A spacegroup of $P2_1\ 2_1\ 2_1$ was identified, with unit cell dimensions of a = 60.20 Å, b = 62.58 Å, and c =199.27 Å, and angles of α= 90°, β= 90°, and γ= 90°. The dataset had a signal-to-noise (I/σ (I)) of 4.8, and 1.4 for the highest resolution data shell, an R_{pim} of 0.04, and a $CC_{1/2}$ of 0.55 for the highest resolution shell **(Table 5.1)**. These values are all indicative of a high quality dataset.

5.4.3 Solving the CIR Structure

As there are no predicted structural homologues of the CIR, molecular replacement could not be used to solve the structure. Phases instead had to be determined experimentally. A number of strategies were tried for experimental phasing, which attempted to exploit anomalous differences between Bijvoet pairs, or through differences in the magnitude of structure factors between slightly different crystals, such as between crystals before and after the incorporation of heavy atoms.

The first of these strategies tried here was through modifying the crystals by introducing heavy atoms, such as metals or halides, the second through modifying the protein itself through the incorporation of selenium in the form of selenomethionine, and finally, through utilising the native sulphur atoms within the 5 predicted disulphide bonds and 6 methionine residues found within each copy of the CIR.

Space group	$P2_1\ 2_1\ 2_1$
Cell dimensions	
a, b, c (Å)	60.20, 63.58, 199.27
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.97166
Resolution (Å)	49.84-2.15 (2.22-2.15)
Total observations	79795 (6894)
Total unique reflections	42523 (3621)
R_{pim}	0.040 (0.632)
$CC_{1/2}$	1.00 (0.55)
I/σ *(I)*	4.8 (1.3)
Completeness (%)	99.8 (99.9)
Multiplicity	1.9 (1.9)
Wilson B-factor (Å^2)	54.4

Table 5.1: Data Collection Statistics for the CIR, PCHAS_1200500 Values in parentheses are for the highest resolution shell.

5.4.3.1 Heavy Atom Derivatives

Initial attempts at experimental phasing involved the incorporation of heavy atoms into crystals. A wide range of heavy atom-containing compounds were trialled, including those containing metals such as mercury, thallium, lead, platinum, and gold, or halides such as bromide. Typically crystals were harvested and transferred into drops containing between 0.05 and 1 mM of the compound, and were soaked for either 1 hr or overnight. The crystals were then transferred back into a drop of well solution to wash away any excess, unbound additives, and then cryocooled in liquid nitrogen for data collection.

Data were collected for a large number of crystals following these soaks, at the ID30B beamline at ESRF, and the I03 and I04 beamlines at Diamond Light Source. This typically gave one of two outcomes. Either the crystals no longer diffracted, so useful data could not be collected, or fluorescence scans indicated that no heavy atom had bound within the crystals.

For a few crystals, which had been soaked in either a platinum-containing compound such as K_2PtBr_4 or K_2PtBr_6, or a mercury-containing compound such as K_2HgI_4, a fluorescence edge scan revealed the presence of the metal within the crystal, and the crystal still diffracted to a reasonable resolution, but downstream processing showed no anomalous scattering, suggesting that the excess heavy atom solution had not been washed away completely. These attempts were also, therefore, unsuccessful.

In addition to this, a new crystal plate was set up with bromide incorporated into the well solution, in an attempt to grow crystals in the presence of the halide to increase the likelihood of incorporation. No crystals, however, grew under these conditions.

5.4.3.2 Selenomethionine Incorporation

To incorporate selenomethionine into the CIR, a modified HEK293F transfection protocol was used[281]. This involved switching to a media which was available deficient in methionine, and so RPMI-1640 was used in place of Freestyle F17 media. To confirm that the CIR would still express in this new media, a small scale expression trial was carried out. This showed that, when compared to Freestyle F17, significantly less protein was produced, but it was still enough to warrant testing selenomethionine incorporation **(Figure 5.5 A)**.

An additional step was also included in the protocol, where the cells were centrifuged and resuspended in this media without methionine, followed by a 12 hour incubation to deplete the intracellular methionine stores. The cells were then centrifuged again and resuspended in the same media supplemented with selenomethionine.

Two slightly different protocols were trialled. For the first, the 12 hour methionine depletion step was carried out immediately prior to transfection, and for the second, this occurred immediately following transfection. Both trials succeeded in producing some protein, but significantly more was produced when the second method was used **(Figure 5.5 A,B)**. Purification by SEC confirmed that the resulting protein was still folded as expected **(Figure 5.5 C,D)**, and the incorporation of selenomethionine was confirmed by native mass spectrometry, which showed an increase in molecular weight of 282 Da, from 30935 Da to 31217 Da, very close to the expected 286 Da increase from the incorporation of 6 selenomethionine residues.

Despite the success of selenomethionine incorporation, crystallisation of this derivative was never trialled, as the structure was solved using sulphur-SAD phasing before this could be carried out.

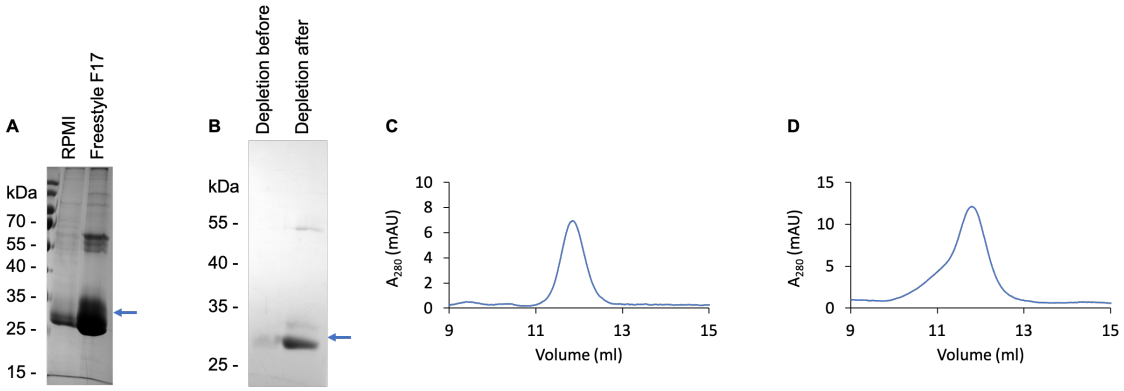


Figure 5.5: Selenomethionine Incorporation into the CIR, PCHAS_1200500 A) Comparison of PCHAS_1200500 expression from HEK293F cells grown in RPMI-1640 with those grown in Freestyle F17. B) Comparison of PCHAS_1200500 expression from HEK293 cells depleted of methionine prior to transfected with those depleted after transfection. Arrows indicate the expected position of bands in these SDS-PAGE gels at around 31 kDa. C) and D) SEC chromatograms for purification of PCHAS_1200500 expressed in HEK293F cells depleted of methionine C) prior to transfection and D) after transfection using a Superdex 75 10/300 column.

5.4.3.3 Sulphur-SAD

Sulphur-SAD phasing utilises the anomalous scattering of sulphur atoms found natively within the structure, as part of cysteine and methionine residues. This anomalous scattering increases with wavelength[282], and so a long-wavelength beamline was needed for data collection. Data were collected at the I23 beamline at Diamond Light Source using an incident X-ray wavelength of 2.76 Å, much longer than the around 0.9 to 1 Å typically used for data collection. Data were indexed and scaled using XDS[218], giving a resolution of 3 Å. The CC_{anom} value of 0.62, and anomalous signal value of 1.54 both indicated strong enough anomalous signal for phase determination **(Table 5.2)**. Phases were then obtained using SHELXD[225], allowing for an initial model to be built using Buccaneer[226], with 3 copies of the CIR in the asymmetric unit of the crystal.

Buccaneer placed a number of helices into the resulting electron density map, and after some rebuilding and extension of these helices in Coot[223], the model was refined

Space group	$P2_1\ 2_1\ 2_1$
Cell dimensions	
a, b, c (Å)	63.81, 65.61, 199.27
α, β, γ (°)	90, 90, 90
Wavelength (Å)	2.75520
Resolution (Å)	62.41-3.00 (3.20-3.00)
Total observations	387923 (57599)
Total unique reflections	31506 (5127)
R_{pim}	0.062 (1.143)
$CC_{1/2}$	1.00 (0.74)
I/σ *(I)*	22.51 (1.82)
Completeness (%)	96.0 (88.9)
Multiplicity	12.3 (11.2)
CC_{anom}	0.62 (0.02)
Anomalous Signal *(|F(+)-F(-)|)/σ)*	1.54 (0.70)

Table 5.2: Sulphur-SAD Data Collection Statistics for the CIR, PCHAS_1200500 Values in parentheses are for the highest resolution shell.

using BUSTER[227], giving improved phases and an improved electron density map **(Figure 5.6)**. This allowed for the remaining helices and some loops of one copy of the CIR molecule to be built. This model could then be used as a search model for molecular replacement using Phaser MR[222], and the earlier mentioned native dataset at a resolution of 2.15 Å. The higher resolution dataset could then be used to complete building of the structure, with correct amino acids assigned using the disulphide bonds as a guide.

After a number of rounds of rebuilding and refinement in Coot and BUSTER, all 3 copies of the CIR could be built. For the first copy, residues 4-241 could be placed. The remaining 22 residues at the C-terminal end were largely predicted to be disordered, explaining their absence from the structure. For the second copy,

residues 3-242 could be placed, although a small section of loop containing residues 44-47 are missing, and for the final copy, residues 6-239 were placed, with 24-28 and 75-77 missing, again sections of loops.

After refinement, the final model gave an R_{work} value of 0.2185, and an R_{free} of 0.2410, and RMSD values of 0.009 Å and 1.01° for bond lengths and angles respectively **(Table 5.3)**. These values all indicate a good quality model. In addition to this, 95.3% of residues were in a favoured region of the Ramachandran plot, with the remaining 4.7% in an allowed region, with no residues in disallowed regions **(Figure 5.7)**.

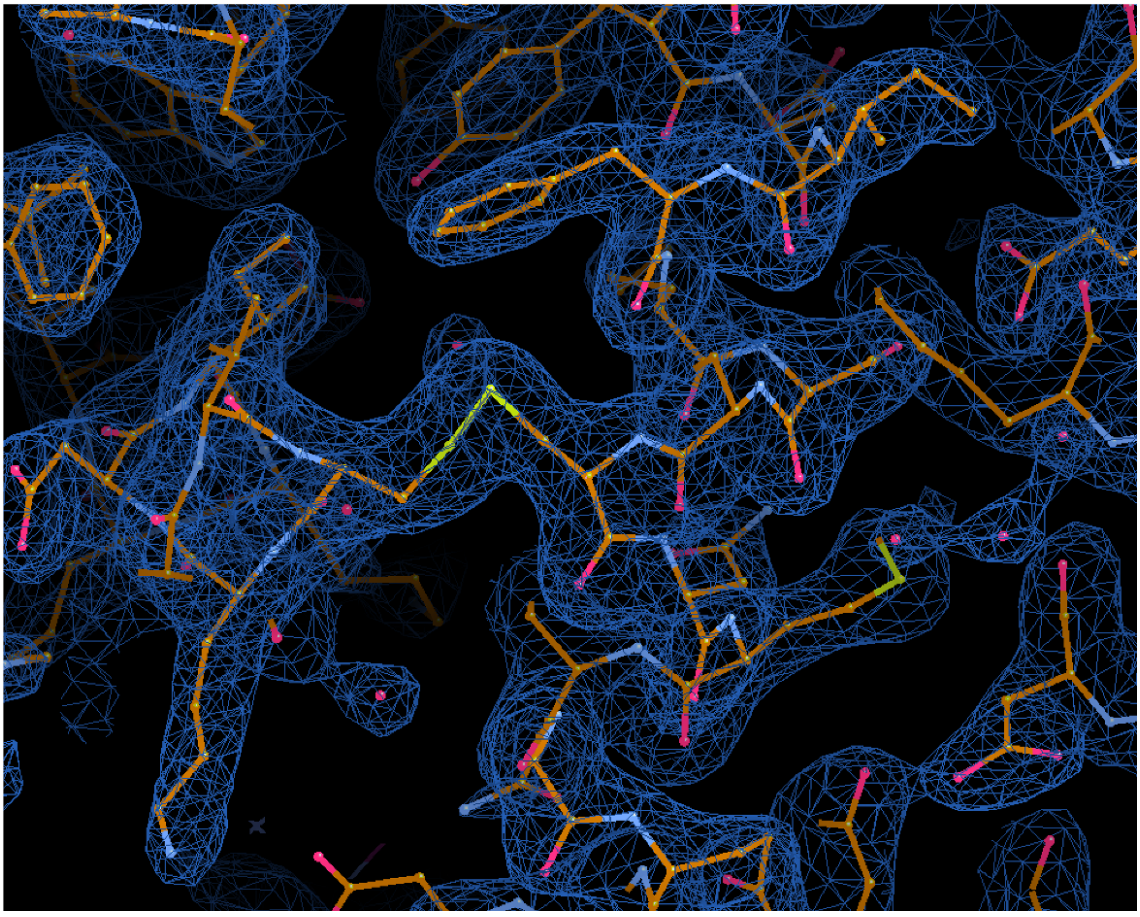


Figure 5.6: Final CIR Electron Density Map (2mF$_o$ - DF$_c$, contour 1σ), with Structure Superimposed The interface between two helices of the CIR is shown, atoms are coloured by element, with carbon in orange, nitrogen in blue, and oxygen in pink. This figure was prepared using Coot[223].

Reflections	42587
R_{work}/R_{free}	0.212/0.236
Average B-factor (Å^2)	64.0
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.93
Ramachandran plot	
Favoured (%)	96.2
Allowed (%)	3.8
Disallowed (%)	0.0

Table 5.3: Refinement Statistics for the CIR, PCHAS_1200500

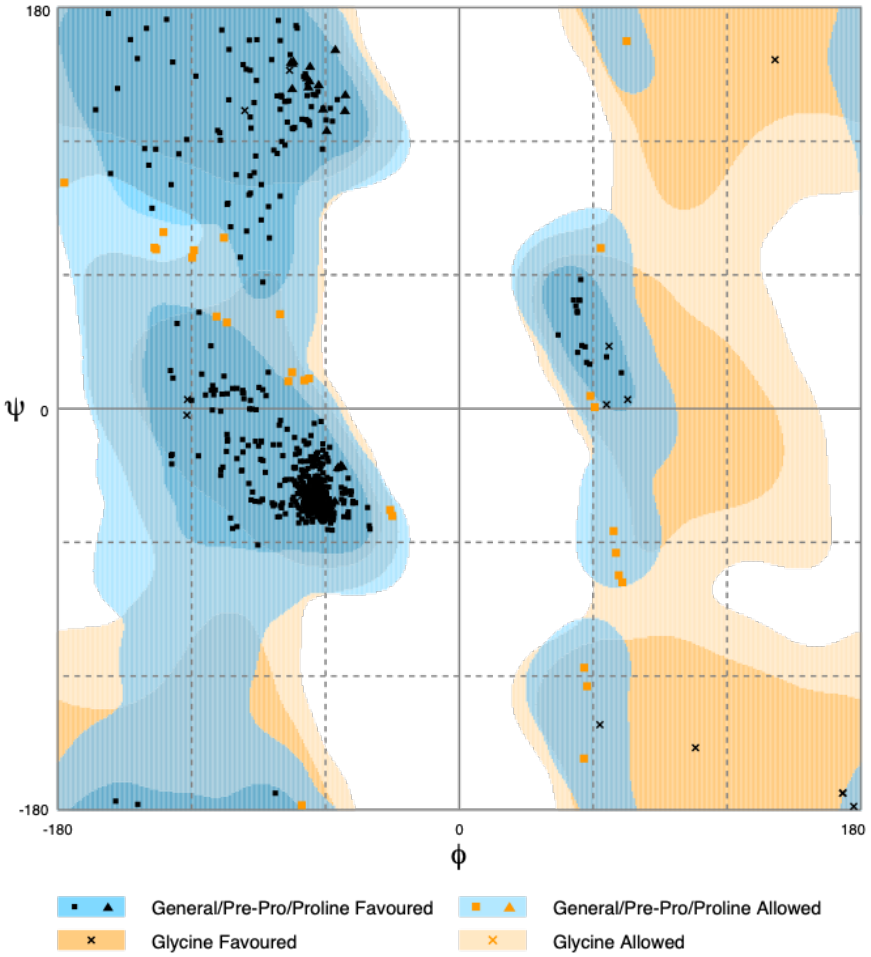


Figure 5.7: Ramachandran Plot for the Structure of the CIR, PCHAS_1200500

5.5 The Structure of a CIR

The ectodomain of the CIR adopts a highly compact α-helical fold, consisting of 8 main helices, arranged into two distinct clusters **(Figure 5.8)**. The N-terminal half of the ectodomain consists of 5 helices, packed tightly together around a central, core helix. These helices are connected by extensive loops, which are largely devoid

of any secondary structure. The C-terminal half of the ectodomain forms a bundle of 3 parallel helices, connected by much shorter loops. These two halves of the protein pack tightly together, with 1 of the 5 disulphide bonds of the fold stabilising the junction between the two halves. The C-terminus of the fold consists of a lengthy stretch of residues lacking in secondary structure, which packs against the interface between helices 7 and 8. Two of the remaining disulphide bonds can be found within the long loop connecting the first and second helices, with another stabilising the shorter loop between the second and third helices. The final disulphide bond connects helices 7 and 8. A further 22 residues are missing from the C-terminus of the protein, suggesting these residues may act as a flexible linker between this domain and the transmembrane helix, allowing the CIR ectodomain to remain flexible relative to the membrane.

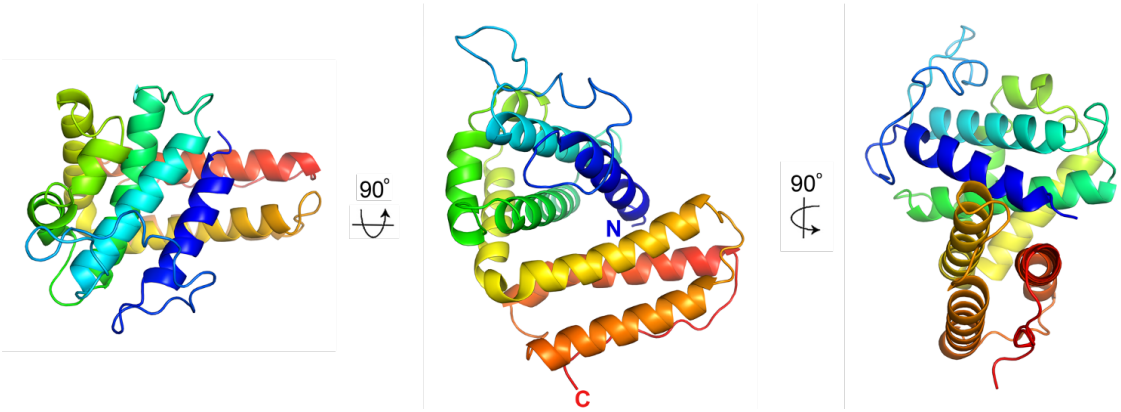


Figure 5.8: The Structure of a CIR Three views of the structure of the ectodomain of the CIR, PCHAS_1200500, coloured using a rainbow representation with the N-terminus in blue and the C-terminus in red.

The most flexible region of the fold, which contains the stretch of residues with the highest crystallographic B factors, **(Figure 5.9)**, is the loop between helices 1 and 2. The second and third copies of the protein within the asymmetric unit are both missing portions of this loop, with chain B missing residues 45-48, and chain C missing residues 25-35, which supports the idea that this is a particularly flexible region of protein. The 3 copies are otherwise very similar, with RMSD values of 0.807

Å and 0.943 Å for alignment of chains B and C respectively to chain A. This loop is situated on the opposite side of the fold to the C-terminus, where the ectodomain is connected to the transmembrane helix, and so it is likely that, if exported to the infected erythrocyte membrane, this surface would be most exposed to the host immune response. A high degree of disorder or flexibility would limit the ability of antibodies to bind, and so could limit the efficacy of an immune response against the CIR.

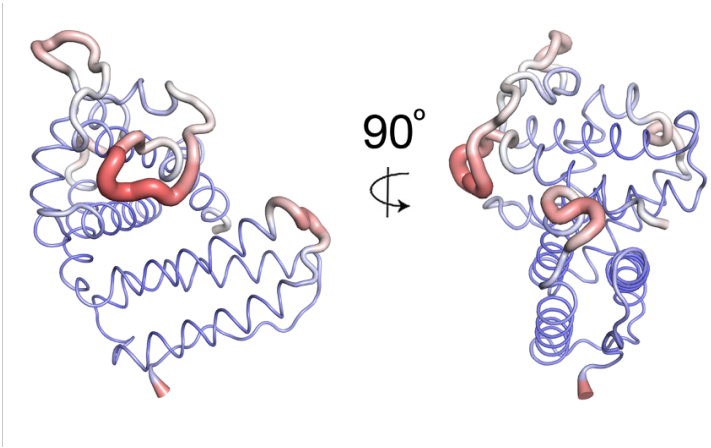


Figure 5.9: Flexibility within the CIR Structure Two views of the structure of the ectodomain of the CIR, PCHAS_1200500, represented as a wire with the thickness scaled according to the crystallographic B factor of this region of the model. Wide, red regions of wire indicate more disordered regions of the protein domain.

To determine whether this fold has been seen previously in other protein structures, the DALI server was used to search the Protein Data Bank for homologues of the CIR structure[283]. The highest scoring results all contained 3-helical bundles aligning to the C-terminal portion of the CIR. No result had a Z-score above the cut-off for a protein of this length of 19.8, where the cut-off used is $\frac{n}{10} - 4$, n being the number of residues in the structure, 238.[284]. This indicates that the CIR adopts a novel fold.

5.6 Conservation Among CIRs

5.6.1 Sequence Conservation

To determine the degree of structural conservation across the CIR family, a sequence alignment was produced for all 198 CIRs from the AS strain of *P. chabaudi* using MUSCLE[244]. The sequence Logo for this alignment revealed that, while there is a large degree of diversity within the family, a number of residues are highly conserved **(Figure 5.10)**. The majority of conserved residues are bulky hydrophobic residues, along with four cysteine residues. These cysteine residues correspond to those forming two disulphide bonds, Cys12-Cys165, and Cys42-Cys52. Plotting the sequence entropy of this family onto the CIR structure demonstrates that the conserved residues, with low sequence entropy, nearly universally correspond to the core of the protein, while the surface of the ectodomain, by contrast, shows almost no conservation **(Figure 5.11)**. This pattern of conservation indicates a conserved core structure, with the hydrophobic residues and disulphide bonds stabilising the positioning and packing of the helices of the fold, while the surface of the fold is highly diverse.

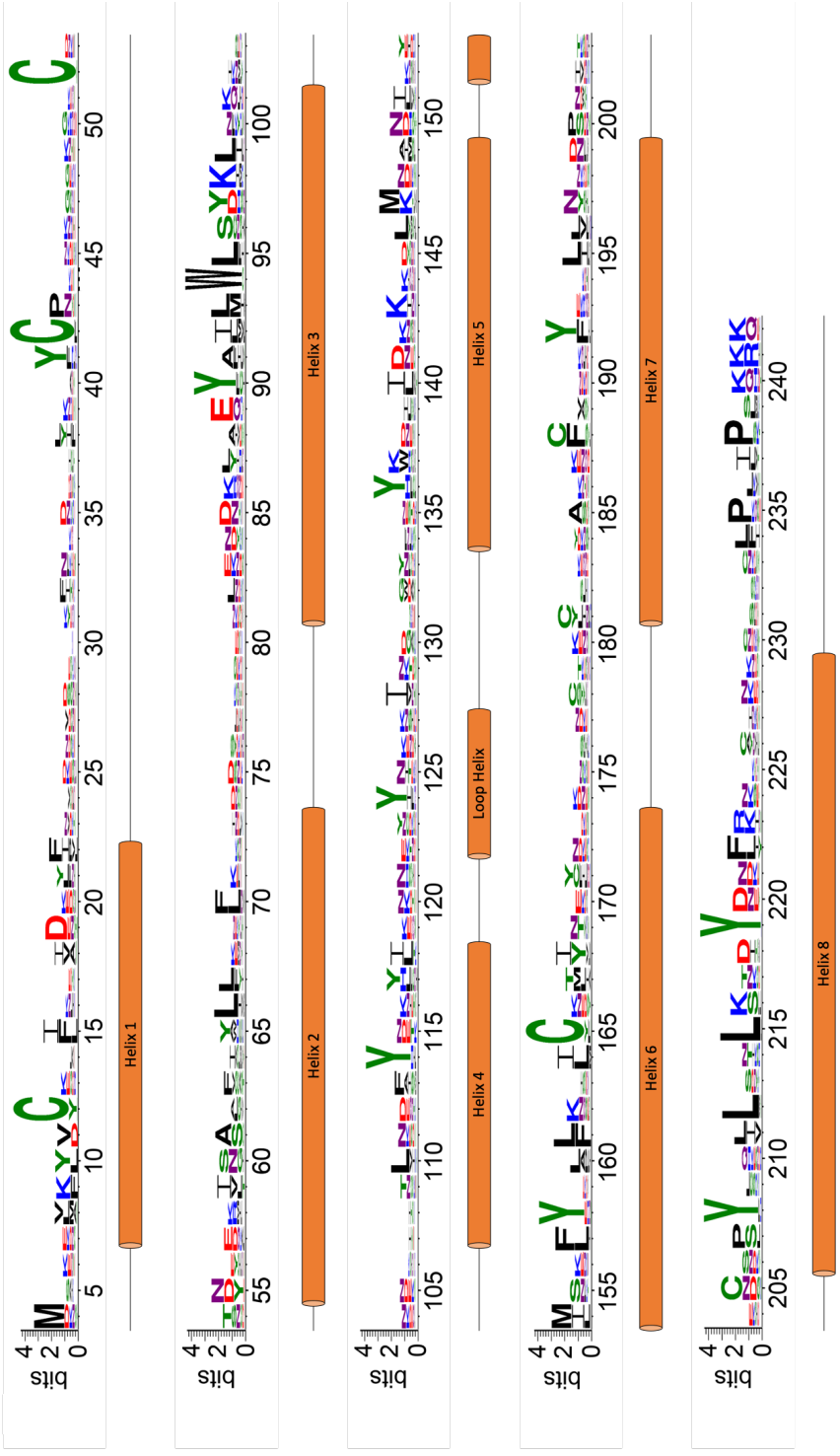


Figure 5.10: Sequence Logo for the Ectodomain of CIRs A sequence logo showing conservation for the portion of the ectodomain aligning with the structure of the CIR for all 198 CIRs from the AS strain of *P. chabaudi*. The secondary structure for the CIR, PCHAS_1200500, is shown underneath the Logo, with helices as cylinders and elements without a secondary structure shown as a line.

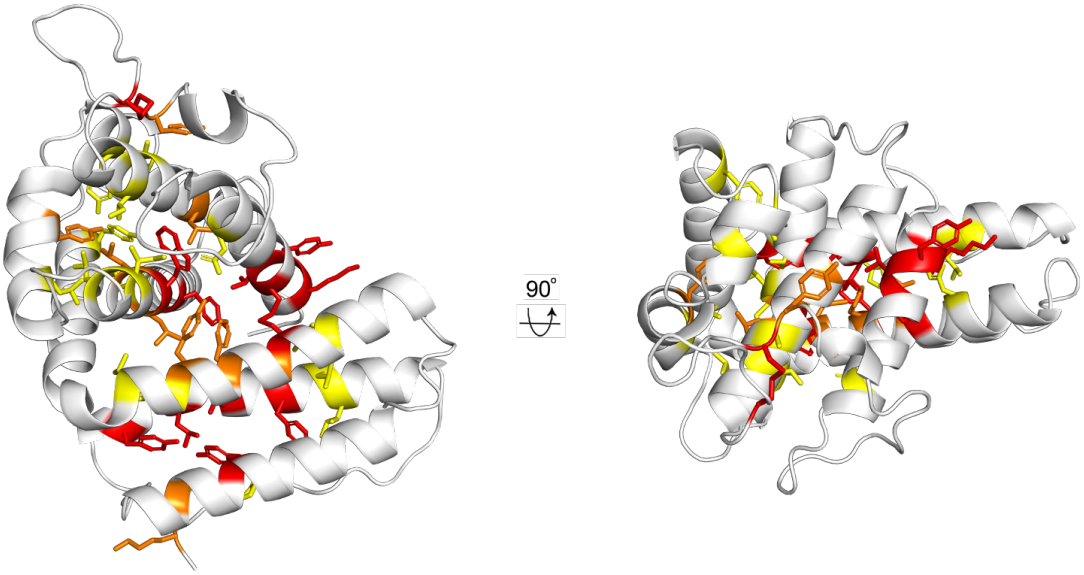


Figure 5.11: Conservation of the Ectodomain of CIRs The structure of the CIR with the most conserved residues coloured by sequence entropy for all 198 CIRs from the AS strain of *P. chabaudi*. Residues coloured in red have a sequence entropy of <0.5, those in orange of 0.5-0.75, and those in yellow of 0.75-1.0.

The diverse surface of the fold is also clear when looking at the length of the loops between helices across the CIR family **(Table 5.4)**. For the majority of these loops, there is a large degree of diversity in their lengths, in particular for those loops found on the surface of the protein distal from the membrane attached C-terminus. For example, the diverse and flexible loop between the first and second helices ranges in length by 15 residues, from 25 to 40, while the loop between helices 4 and 5 ranges from as short as 2 residues, to 18. This again indicates a high degree of plasticity on the surface of the CIR fold, as expected for a family under the immune pressure caused from exposure on the surface of a host red blood cell.

Loop	Range of Length (Residues)	Average Length (Residues)
1-2	25-40	31.5
2-3	0-6	3.6
3-4	3-16	6.0
4-5	2-18	9.7
5-6	1-2	2.0
6-7	6-8	7.0
7-8	4-8	5.7

Table 5.4: Variation in Loop Lengths Among CIRs

5.6.2 Structure-Based Threading

To confirm that the core structure of the CIR is conserved, structure-based threading was then carried out using Phyre2[247] for 10 members of the family. These were selected following phylogenetic analysis of all CIRs from the AS strain of *P. chabaudi*, with the 10 sequences chosen to represent the wide range of diversity across the family **(Figure 5.14B)**. Homology models could be produced for all sequences with 100% confidence **(Table 5.6)**, aligning across the full-length of the fold, providing further support for structural conservation within the family.

The position of this domain within the full length of the CIR sequences is also highly conserved, with the fold consistently found within 5 residues of the N-terminus of all CIR sequences. The length of protein between the domain and the predicted transmembrane helix, however, varies significantly across the family, from as short as just 18 residues, up to 1117 residues **(Table 5.5)**, and the length of this linker is the primary difference between the S and L families of CIRs. For example, for the S7 family, which contains the CIR seen in the structure, PCHAS_1200500, and is associated with acute infection in mice[88], the length of this linker ranges from 18 to 47 residues. By contrast, the shortest seen in the L1 family, associated with chronic

infection in mice[88], is 59 residues, and this stretch of protein can reach over 1000 residues for some members of this family.

Family	Range of Length (Residues)	Average Length (Residues)
L1	59-1117	533.5
L2	99-103	101.0
L4	45-430	98.6
S1	18-47	25.0
S3	18-18	18.0
S7	20-68	36.7
All CIRs	18-1117	242.3

Table 5.5: Variation in Lengths of the Linker Between the CIR Fold and the Transmembrane Helix Among CIR Families

Disorder predictions for members of each family indicate that the only well-ordered region of the ectodomain is the fold seen in the structure **(Figure 5.12)**, and the connection to the membrane is predicted to be through a disordered region of protein, indicating that the CIRs all share a common architecture **(Figure 5.13)**, with a conserved N terminal fold connected to the membrane by a variable-length flexible linker.

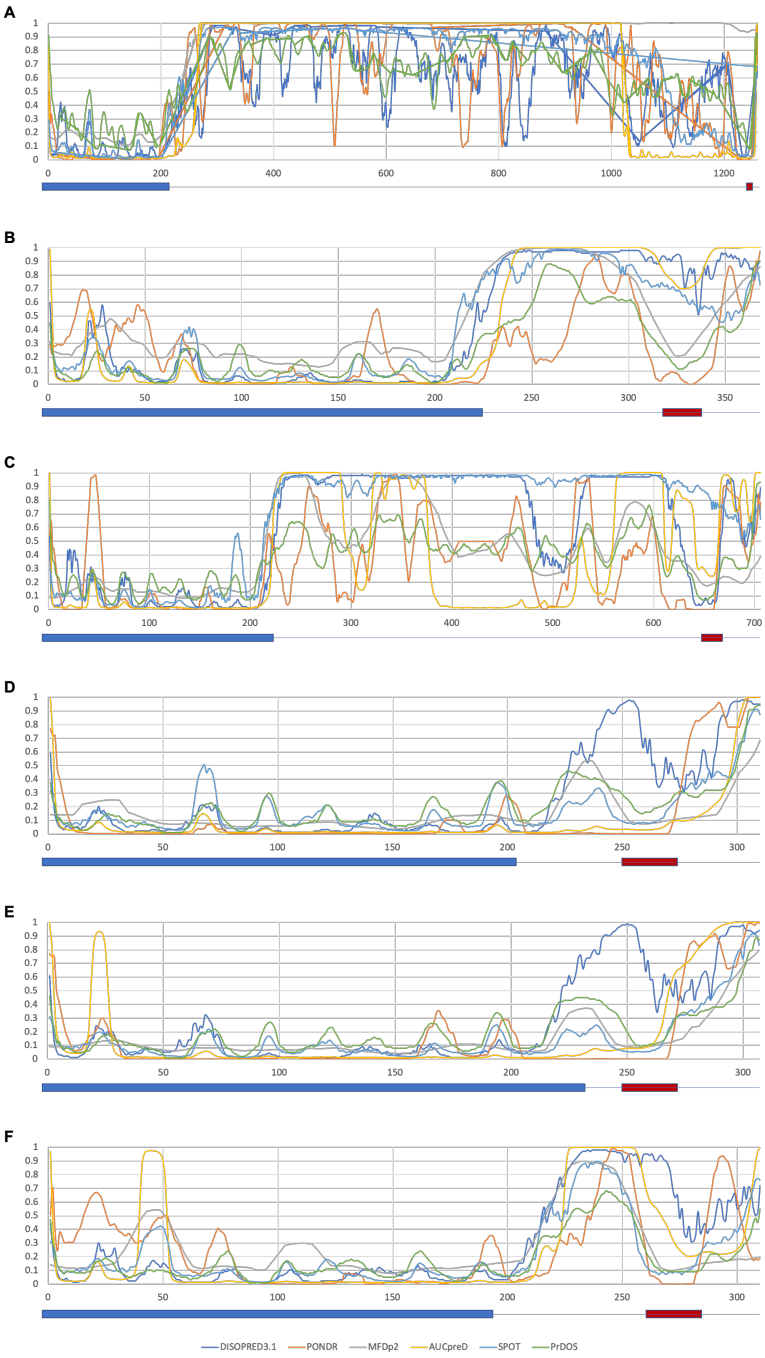


Figure 5.12: Disorder Predictions for Representatives of Each CIR Familiy Analysis using 6 separate predictors of disorder for the CIR sequence most closely related by sequence identity to the consensus sequence for each CIR family, A)L1 - PCHAS_0601000, B) L2 - PCHAS_0318500, C) L4 - PCHAS_0301800, D) S1 - PCHAS_0700300, E) S3 - PCHAS_1000300, and F) S7 - PCHAS_0500200. The blue bar represents the region of each sequence which aligns to the PCHAS_1200500 structure, and the red box represents the predicted transmembrane helix for each sequence.



Figure 5.13: Conserved Protein Architecture of the CIRs The protein architecture of 2 CIRs, each most closely related by sequence identity to the consensus sequence for the given family. A) The L1 CIR, PCHAS_0601000, and B) the S7 CIR, PCHAS_0500200. The blue box represents the conserved N-terminal fold seen in the PCHAS_1200500 structure, and the red box represents the predicted transmembrane helix for each sequence. The black line represents disordered regions of protein, including the variable length linker connecting the N-terminal domain to the membrane.

5.7 Conservation Among Other Putative PIR Families

Having determined that the fold seen in the structure is likely to be conserved across the CIR family, the next aim was to determine whether this could be considered a characteristic "PIR fold", found across putative PIR families from a range of *Plasmodium* species, and whether this could be used to better define membership of the family. While putative PIRs have been identified in a range of human- and rodent-infective *Plasmodium* species, whether PIR homologues are present in the *P. falciparum* genome is less clear. It has been suggested that the RIFINs and STEVORs could be part of the family, although a lack of protein sequence similarity with other putative PIRs has cast doubt on this theory. As structure is often more conserved than sequence, the CIR structure could be used to more clearly explore the relationship between potential members of the PIR family.

To accomplish this, similar phylogenetic analysis used for the CIRs was carried out for putative PIR families from 2 other species of rodent-infective *Plasmodium* from the *Vinkeia* clade **(Figure 5.14C, D)**, along with four species from the monkey- and human- infective *Plasmodium* clade **(Figure 5.14E-H)**, two species from the ape-infective *Laverania* **(Figure 5.14I-K)**, which included both the STEVOR and RIFIN families from *P. falciparum*, and finally the avian-infective *P. gallinaceum* from the *Haemamoeba* clade **(Figure 5.14L)**.

This covered the putative PIR proteins from a wide range of species from across the *Plasmodium* evolutionary tree **(Figure 5.14A)**, including all known human-infective species of the parasite, excluding *P. ovale wallikeri*, for which PIR sequences were not available on PlasmoDB. Each of these sequences was then subjected to structure-based threading using Phyre2[247].

5.7.1 Is the CIR Protein Fold Conserved Across Rodent-Infective *Plasmodium* Species?

The most closely related *Plasmodium* species to *P. chabaudi* are two other rodent-infective species, *P. berghei* and *P. yoelii*. The genomes of both species contain large families of *pir* genes, with 135 such sequences in the genome of the ANKA strain of *P. berghei* and 1011 in the genome of the 17X strain of *P. yoelii*. Homology models could be produced using Phyre2 to very high degrees of confidence for all chosen sequences from both species, with an average confidence score of 99.9% for *P. berghei*, and 100% for *P. yoelii* **(Table 5.6, Figure 5.15)**. This strongly suggests that the fold seen in the CIR structure is conserved across PIR proteins from both of these species, along with the CIRs.

Aligning the PIR sequences from both of these species against the CIR structure reveals a similar pattern of conservation as seen for the CIR family, with a number of hydrophobic residues corresponding to the core of the fold showing a high degree

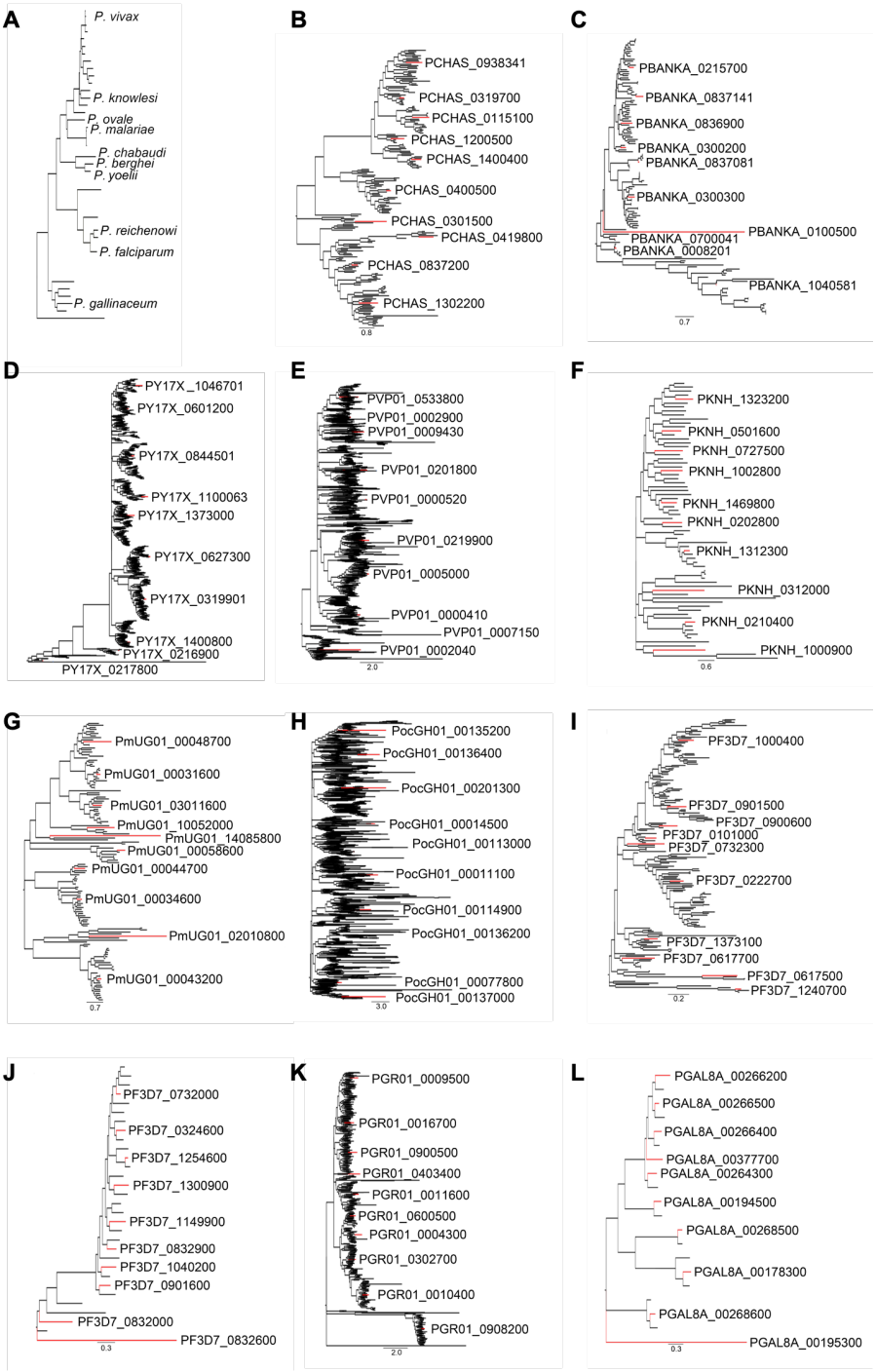


Figure 5.14: Sequence Analysis of Putative PIR Families A) An evolutionary tree for *Plasmodium* species based on Krief et al.[17], and evolutionary trees, based on sequence analysis of putative PIR families to allow for selection of 10 diverse sequences from each of b) *P. chabaudi*, C) *P. berghei*, D) *P. yoelii*, E) *P. vivax*, F) *P. knowlesi*, G) *P. malariae*, H) *P. ovale curtisi*, I) *P. falciparum* (RIFINs), J) *P. falciparum* (STEVORs), K) *P. reichenowi*, and L) *P gallinaceum*. The ten members of each family used for modelling studies are shown as red lines and are labelled.

Family	**Average Sequence Identity (%)**	**Average Confidence (%)**	**Average Number of Aligned Residues**
P. chabaudi	34.2	100	227.2
P. berghei	29.3	99.9	226
P. yoelii	29.2	100	225
P. knowlesi	12.8	98.7	207
P. malariae	13.2	97.7	209
P. ovale curtisi	12.9	98.3	206
P. vivax	15.1	97.9	192
P. falciparum (STEVOR)	14.7	83.0	178
P. falciparum (RIFIN$_C$)	20.8	1.06	89.9
P. reichenowi	16.9	66.4	180
P. gallinaceum	19.9	42.5	147

Table 5.6: Structure-Based Threading of PIRs Average sequence identities of 10 representatives from each potential PIR family with the PCHAS_1200500 ectodomain, and average confidence scores for Phyre2 threading of these sequences using the PCHAS_1200500 structure. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266]. The average sequence length threaded onto the structure is also shown. RIFIN$_C$ refers to the RIFIN constant region. The data for each individual sequence can be seen in **Table C.1**.

of conservation **(Figure 5.16)**. In addition to this, the same four cysteine residues corresponding to the Cys12-Cys165 and Cys42-Cys52 disulphide bonds which are conserved in the CIRs, are conserved in the PIR sequences from both *P. berghei* and *P. yoelii*. This strongly supports the data from the structure-based threading in Phyre2, again suggesting a conserved fold among these PIR families.

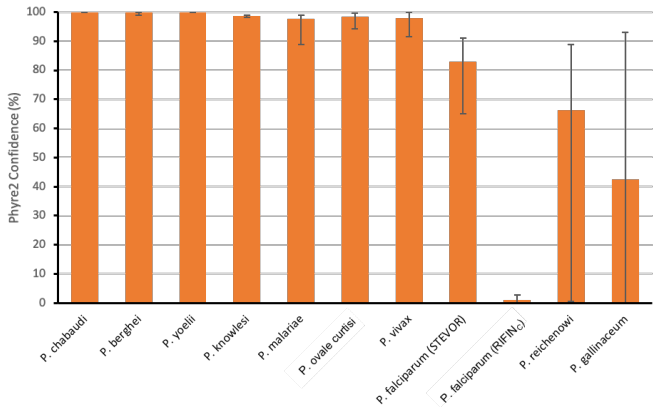


Figure 5.15: Phyre2 Threading Confidence Scores for Putative PIR Families Average confidence scores for threading of 10 representatives from putative PIR proteins of different *Plasmodium* species. Error bars show the range of scores obtained for each family, and the 90% high-confidence cutoff limit [266] is represented with a dashed line. RIFIN$_C$ refers to the RIFIN constant region. The data for each individual sequence can be seen in **Table C.1**.

5.7.2 Is the CIR Protein Fold Conserved Across Human-Infective *Plasmodium* Species?

The same analysis was then carried out for four human-infective species of *Plasmodium* from the *Plasmodium* subgenus. Phylogenetic analsis was carried out for 70 sequences from the genome of strain H of *P. knowlesi*, 136 sequences from the UG01 strain of *P. malariae*, 1495 sequences from the GH01 strain of *P. ovale curtisi*, and 1086 sequences from the P01 strain of *P. vivax*. Homology models could again be produced with a high degree of confidence based on the CIR structure for the 10 sequences chosen for each species, with an average confidence score of around 98.7% for sequences from *P. knowlesi*, 97.7% for *P. malariae*, 98.3% for *P. ovale curtisi*, and 97.9% for *P. vivax*, all above the 90% cut-off used for a high-confidence model [266].

This again suggests that the fold seen in the structure is conserved across all of these families, and therefore that the PIR sequences of all 4 of these human-infective *Plasmodium* species, along with the 3 rodent-infective species previously discussed, can be considered part of a larger family, with the fold seen in the structure acting

as a characteristic PIR fold.

This appears to be despite a relative lack of sequence conservation within PIRs for these species, when compared with those from the rodent-infective species. While there are some hydrophobic residues conserved across all of these species, including for example, that corresponding to W94 of the CIR structure, there is generally much more variability, particularly for the PIRs of *P. ovale curtisi* and *P. vivax* **(Figure 5.16)**. When combined with the homology modelling data, this suggests that the PIR superfamily is closely related by conservation of structure, despite a distinct lack of sequence conservation.

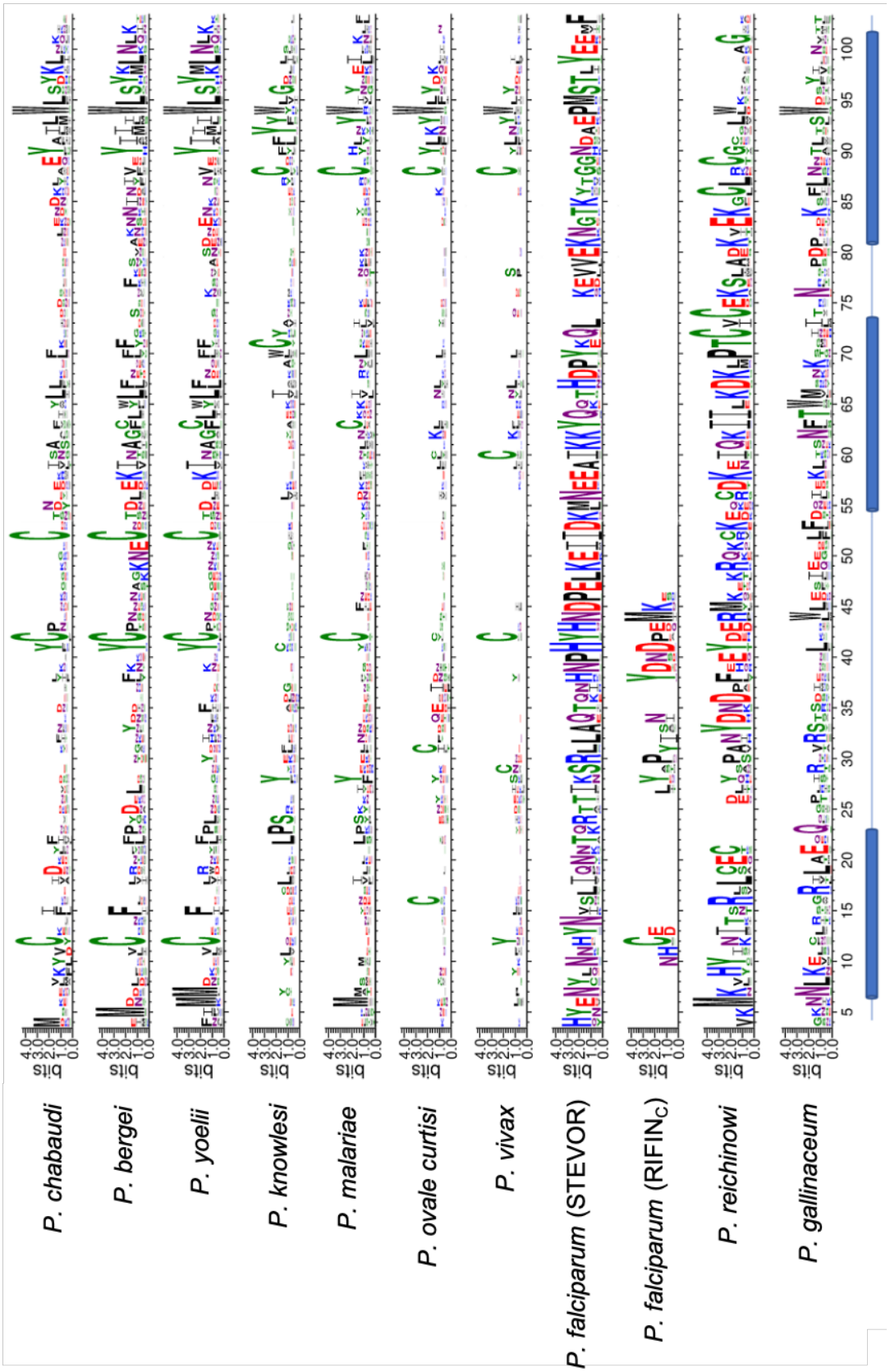


Figure 5.16: Sequence Logo for the Ectodomain of Putative PIRs. Page 1 of 3

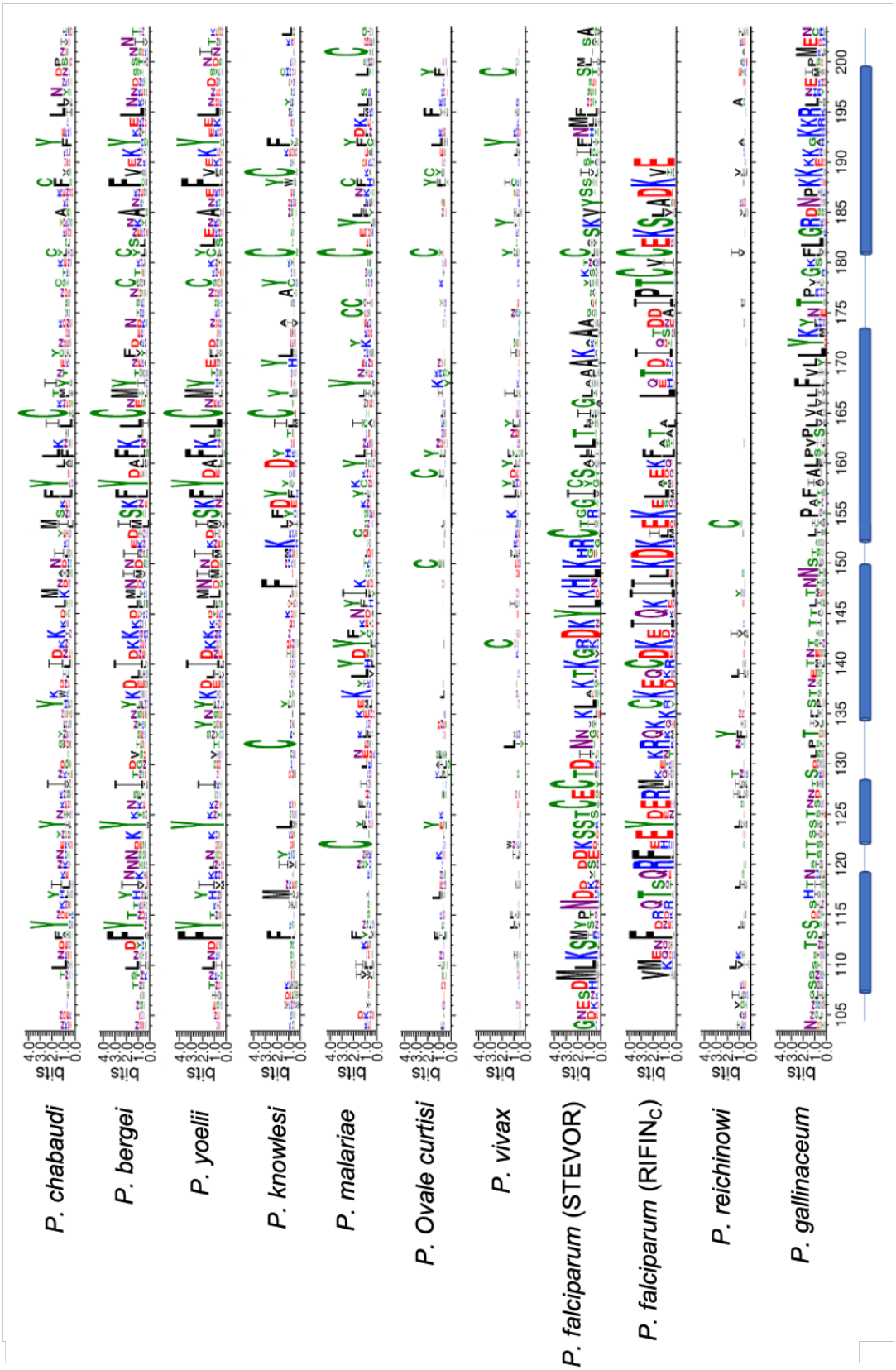


Figure 5.16: Sequence Logo for the Ectodomain of Putative PIRs. Page 2 of 3

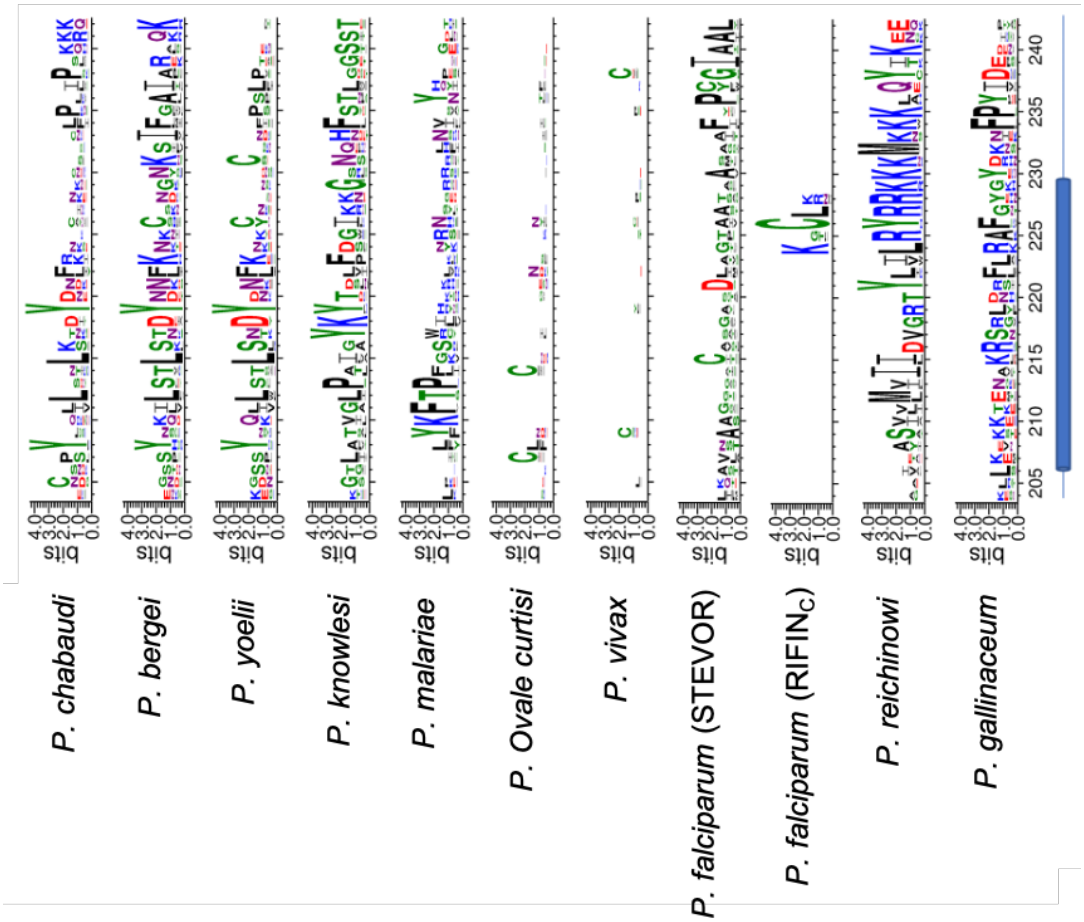


Figure 5.16: Sequence Logo for the Ectodomain of Putative PIRs. Page 3 of 3 A sequence logo showing conservation for the portion of the ectodomain aligning with the structure of the CIR for all putative PIRs from different *Plasmodium* species. The sequences used were from *P. chabaudi* (198 sequences from the AS v3 genome assembly), *P. berghei* (135 sequences from the ANKA genome assembly), *P. yoelii* (1011 sequences from the 17X v3 genome assembly), *P. knowlesi* (70 sequences from the strain H genome assembly), *P. malariae* (136 sequences from the UG01 genome assembly), *P. ovale curtisi* (1495 sequences sequences from the GH01 genome assembly), *P. vivax* (1086 sequences from the P01 genome assembly), *P. falciparum* (185 RIFIN constant region sequences and 31 STEVOR sequences from the 3D7 genome assembly), *P. reichinowi* 487 sequences from the G01 genome assembly), and *P. gallinaceum* (20 sequences from the 8A assembly). RIFIN$_C$ refers to the RIFIN constant region. The secondary structure for the CIR PCHAS_1200500 is shown underneath the Logo, with helices as cylinders and elements without a secondary structure shown as a line.

5.7.3 Is the CIR Protein Fold Conserved Across the RIFINs and STEVORs?

A comparison to the RIFIN and STEVOR families from *P. falciparum* was then carried out. For the RIFINs, the structure of the CIR was first compared with that of the Pf3D7_1254800 variable region (**Figure 5.17**). Both structures are composed of a tightly packed series of α-helices, connected by an extensive series of loops. Additionally, the C-terminal portion of the CIR structure forms a three-helical bundle, similar to that seen in the RIFIN structure. However, the topology and arrangement of these helices is very different in the RIFIN, as are the loops between them. In this region of the CIR, the maximum length of loops seen across the family is 8 residues (**Table 5.4**), whereas in the RIFIN structure, the loop between the second and third helices is significantly longer, at 18 residues.

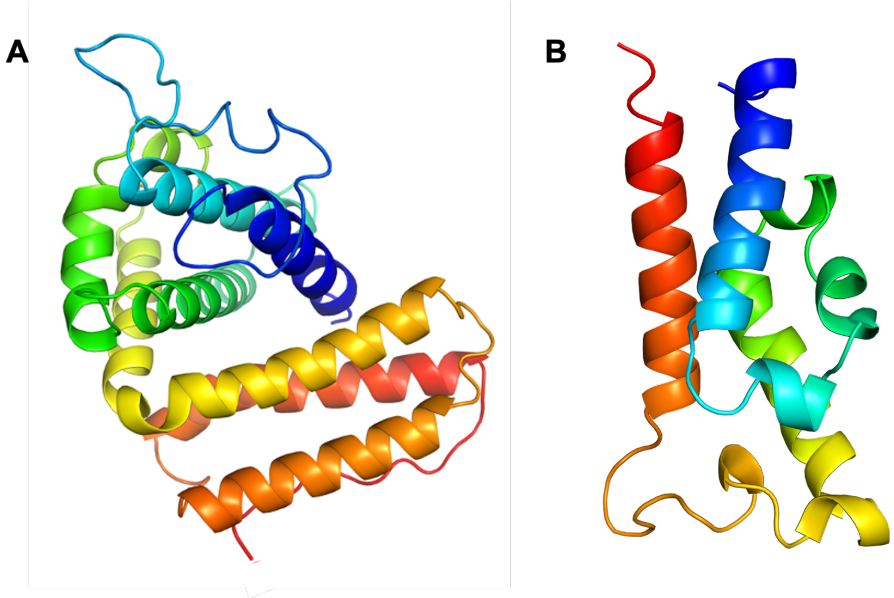


Figure 5.17: Comparison of the CIR and RIFIN Variable Region Structures The differing structures of A) the CIR, PCHAS_1200500, and B) the RIFIN, Pf3D7_1254800, variable region.

A comparison of the constant region of the RIFINs also demonstrates a probable lack of structural similarity. The constant region of the Pf3D7_1254800 RIFIN is

just 124 residues in length, significantly shorter than the domain seen in the CIR structure, and so this would be unable to adopt the CIR fold.

Furthermore, structure-based threading of 10 RIFIN constant region sequences onto the two separate halves of the CIR structure demonstrates it is unlikely that this region of the RIFIN is adopting the structure of either the N-terminal 5 helices, or C-terminal 3 helices of the CIR fold (**Table 5.7**), with average confidence scores of around 0.4% and 0.3% respectively.

Region of CIR Structure	**Average Sequence Identity (%)**	**Average Confidence (%)**	**Average Number of Aligned Residues**
N Terminal	20.3	0.41	82.2
C Terminal	16.8	0.28	71.6

Table 5.7: Structure-Based Threading of the RIFIN Constant Region Sequence identities of 10 representative RIFIN constant region sequences with the two regions of the PCHAS_1200500 ectodomain, and confidence scores for Phyre2 threading of these sequences using the same two regions of the PCHAS_1200500 structure. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266]. The length of sequence threaded onto the structure is also shown. the data for each indiviual sequence can be seen in **Table C.2**.

As neither the RIFIN constant or variable regions seem to show any structural similarity to the CIR, it is unlikely that the two families are related, and these findings therefore suggest that the RIFINs should not be considered be members of the PIR superfamily.

Structure-based threading of sequences from the STEVOR family using the CIR structure gives similar results, with an average confidence score for the 10 sequences selected of 83%, below the cut-off for high confidence modelling of 90% (**Table 5.6**). This again suggests that the STEVORs are not structurally related to the CIRs, and therefore should not be considered as PIR proteins, suggesting that the PIRs are entirely absent from the *P. falciparum* genome.

5.7.4 Is the CIR Protein Fold Conserved Across the *Laverania*?

Following this, the putative PIR sequences from another member of the *Laverania*, *P. reichenowi*, were analysed to determine whether members of the PIR superfamily were present in any species within this clade, which also contains *P. falciparum*. There were 487 putative PIR proteins within the G01 genome assembly of *P. reichenowi*, and structure-based threading of 10 of these sequences onto the CIR structure again revealed a lack of structural homology, with an average confidence score of just 66% (**Table 5.6**). Modelling of these same sequences using the RIFIN variable region could also not be achieved with high-confidence, with an average score of 71.5%, and this suggests that these proteins should be considered as neither PIRs, nor as RIFIN homologues (**Table C.3**).

5.7.5 Is the CIR Protein Fold Conserved Across Avian-Infective *Plasmodium* Species?

Finally, to attempt to determine when the PIR superfamily evolved, and whether this family emerged after the *Laverania* split from the other rodent- and simian-infective species, or was lost following this split, 10 putative PIR sequences from the avian-infective *P. gallinaceum* genome were modelled. This species diverged prior to the emergence of the *Laverania* (**Figure 5.14A**), and so the presence of PIRs here would suggest an older protein family which was lost from the *Lavernia* genomes. Structure-based threading of these sequences, however, demonstrated a lack of homology with the CIR, with an average confidence of 42.5%, suggesting again that these sequences should not be considered members of the PIR superfamily.

These results all indicate a lack of PIR homologues in both the *Laverania* and *P. gallinaceum*, which suggests that the PIR superfamily emerged following the

divergence of the *Plasmodium* and *Vinkeia* clades from the *Laverania* (**Figure 5.18**).

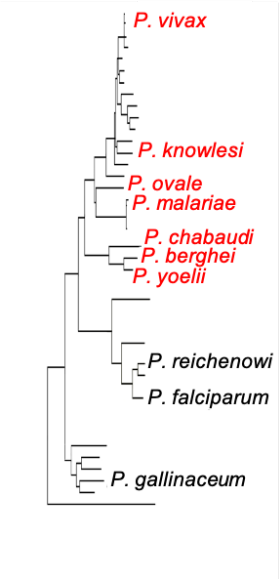


Figure 5.18: Evolutionary Tree of *Plasmodium* Species with Those Expressing PIR Proteins in Red Figure adapted from Krief et al.[17]

5.8 Discussion

The production of a panel of CIR ectodomains in mammalian cells allowed for extensive crystallisation trials to be carried out, eventually leading to the crystallisation of the PCHAS_1200500 ectodomain after slight modification of the protein by methylation.

While extensive attempts to solve the structure by derivatisation of the crystals with heavy atoms was unsuccessful, the use of a long wavelength beamline to utilise the anomalous scattering of the sulphur atoms found natively within the cysteine and methionine residues of the protein allowed for the first structure of a PIR protein to be solved.

This revealed that the CIR adopts a novel fold, composed of 8 tightly packed α-helices connected by a series of largely flexible loops. This fold can be described

as containing two distinct regions - an N-terminal portion composed of 5 α-helices and lengthy, highly flexible loops, and a C-terminal portion containing 3 helices connected by much shorter loops, connecting to the predicted transmembrane by a stretch of residues lacking in secondary structure running parallel to two of the helices, followed by another 22 residues which were both missing from the structure, and predicted to be disordered. This fold was stabilised by a series of 5 disulphide bonds, which act to both stabilise loops and the interface between helices.

Analysis of conservation of this fold within the CIR family revealed that while there is very limited sequence conservation across the CIRs, there were a number of well conserved hydrophobic residues found within the core of the fold, acting to stabilise the packing of helices, along with 2 well conserved disulphide bonds. The surface of the protein, by contrast, was highly variable, both in sequence and in the lengths of loops, as expected for proteins believed to be under immune pressure on the surface of the host erythrocytes. Structure-based threading added further support to the idea of a conserved CIR fold, with all tested sequences producing very high confidence homology models.

Conservation of the fold could be seen across putative PIRs from a number of other *Plasmodium* species. Sequence analysis and homology modelling of families from the rodent-infective *P. berghei* and *P. yoelii* share the same fold as seen in the CIR structure, as do those from the majority of human-infective species of *Plasmodium*, with putative PIRs from *P. knowlesi*, *P. malariae*, *P. ovale*, and *P. vivax* all shown with high degrees of confidence to adopt the same fold. This suggests that the CIR fold can be considered to be a conserved PIR fold, characteristic of members of this superfamily, and found across a wide variety of *Plasmodium* species.

Analysis of putative PIR sequences from *P. falciparum* and *P. reichenowi* demonstrated a lack of homology with this PIR fold, and so the absence of any members of the PIR superfamily from the *Laverania*. The PIRs can not, therefore, be used

as a model for the function of either the RIFIN or STEVOR families, although the rodent PIRs still retain utility as an animal model for the activity of the PIRs from other human-infective *Plasmodium* species, including the VIRs from *P. vivax*.

Similar analysis of sequences from the *P. gallinaceum* genome helped to pinpoint when in the evolutionary history of the *Plasmodium* parasite these PIRs evolved, which is likely to be following the divergence of the *Laverania* from the rodent- and simian-infective species of the parasite.

This CIR structure has, therefore, helped to shed significant light on the PIR family, by revealing a characteristic PIR fold which has both allowed for validation of membership of the family, and helped to uncover its evolutionary history.

Despite this, there is still a clear lack of knowledge surrounding the function of the PIRs. While PIRs from *P. chabaudi* have been implicated in influencing the course of an infection in mice[88,168], the mechanism by which this occurs is unknown. A PIR from *P. vivax* has been shown to bind ICAM-1 leading to adhesion of infected erythrocytes within the vasculature[90,171], allowing for avoidance of clearance in the spleen, providing a possible explanation. However, this has yet to be demonstrated for PIRs from other *Plasmodium* species, or as a universal mechanism of action of PIRs in *P. vivax*. Further functional analysis, therefore, is clearly required to identify the function of the family as a whole, the mechanism by which this is carried out, and any interactions the PIRs are forming with both other *Plasmodium* proteins, and host cell components.

Chapter 6

Conclusions and Future Directions

During an infection with the causative against of malaria, the *Plasmodium* parasite, the red blood cell is invaded and undergoes major modifications. Among these are the export of proteins to the erythrocyte membrane, at great risk to the parasite due to direct exposure of these antigens to the host immune response. These proteins include the RIFIN family of *P. falciparum*[74], and the PIR family of a variety of human- and rodent-infective *Plasmodium* species[165].

The RIFINs have been implicated in the direct modulation of the host immune response, through interactions with inhibitory immune receptors such as LILRB1 and LAIR1[67], and the PIR family has been shown to play a role in the determination of virulence and course of an infection[88,168]. The main aim of my thesis was to shed light on these two families through carrying out structural studies of members of the PIR family, to aid in analysis of conservation and function of members of the family, along with structural characterisation of the interaction between RIFINs and inhibitory immune receptors, to determine the molecular mechanism behind their ability to influence the host immune response.

6.1 Characterisation of the Interaction Between RIFINs and LILRB1

6.1.1 The RIFIN Variable Region Structure

Characterisation of the RIFINs relied entirely on the ability to recombinantly express and purify the proteins, and this proved a major challenge throughout the project. This ultimately lead to instead separately producing different regions of the RIFIN ectodomain. Biophysical characterisation of the separate constant and variable regions of a LILRB1-binding RIFIN by SPR showed that only the variable-region is able to bind to the receptor.

The variable-region of this RIFIN could then be crystallised in complex with the LILRB1 ectodomain, allowing for the determination of the first structure of any portion of a member of the RIFIN family, as well as revealing the molecular mechanism for receptor binding by the protein. The RIFIN variable region adopted a tightly-packed α-helical structure, with 3 helices connected by long loops, both of which contained more helical sections.

Analysis of conservation of this fold across the RIFINs was challenging due to the complete lack of sequence identity, although the single disulphide bond within the variable region was highly conserved, as was a single core hydrophobic residue. Structure-based threading of RIFIN variable regions onto the structure was also unclear, with high quality models produced for 5 of the 10 sequences tested. The ability to produce high quality homology models for half of the tested sequences does suggest that the RIFIN variable region fold is likely to be conserved across the family, with the poor quality of the remaining models possibly due to the lack of sequence identity making homology modelling challenging.

Structural studies of the semiconserved region of the ectodomain, as well as further

attempts to produce and characterise the full-length ectodomain would allow for a better assessment of structural conservation, as well as perhaps providing an indication of the function of the semiconserved region, as information surrounding this is still very lacking.

6.1.2 The RIFIN:LILRB1 Interaction

The structure of the RIFIN variable domain with LILRB1 also allowed for the determination of the molecular mechanism of binding of the RIFIN to the receptor. This revealed a binding interface formed primarily through hydrogen bonds between side chains of the RIFIN with both side chain and backbone groups of LILRB1, along with an additional hydrophobic component through the placement of a leucine side chain into a hydrophobic pocket on LILRB1. This interaction showed remarkable similarity to that of MHC molecules with the same receptor, with both HLA-A2 and the RIFIN forming interactions with largely the same residues of LILRB1. This is in spite of the lack of structural homology between the two proteins, with the RIFIN adopting a primarily α-helical structure, in contrast to the majority β-sheet structure adopted by the β_2M-chain of the MHC complex which is forming the bulk of the interaction with the receptor. The structure did reveal, however, that despite the difference in fold, the RIFIN was structurally mimicking one aspect of the MHC complex, with the RIFIN binding loop closely resembling a loop of the β_2M chain involved in the formation of multiple hydrogen bonds with LILRB1.

The use of both a reporter assay and microscopic imaging of a model immunological synapse revealed that the RIFIN was able to induce signalling through the receptor, again mimicking the activity of the native ligand, MHC class I. RIFIN binding was sufficient to significantly reduce perforin mobilisation by NK cells, demonstrating the potential of this interaction to interfere with the immune response against a red blood cell infected with *P. falciparum*.

While poor sequence identity again made it challenging to determine whether the RIFIN utilised a conserved mode of binding to the receptor, mutagenesis studies of key residues in the binding site allowed for some answers to be obtained. Introduction of a bulky side chain into the binding loop of the RIFIN used for structural studies was sufficient to block any detectable binding to LILRB1. This single mutation was also able to inhibit both signalling through LILRB1 in the NFAT-GFP reporter assay, and the reduction in perforin mobilisation in the supported lipid bilayer assay.

Similar mutagenesis of three other RIFINs, one produced recombinantly and two in transgenic parasites, was able to prevent binding of these RIFINs to LILRB1. This suggests a conserved mode of binding to the receptors, with the RIFINs utilising the same binding loop, anchored in place by a conserved disulphide bond, in order to interact with LILRB1. This suggests a similarity with the interactions formed by the PfEMP1 family, which often utilise a highly sequence diverse binding site, but maintain structural and chemical similarity to allow for ligand binding[134–136].

The RIFIN produced recombinantly for this mutagenesis analysis, however, unexpectedly showed binding to the 2 membrane proximal domains of LILRB1, in contrast to both the RIFIN in the structure, and all other known binding partners of the receptor which bind to those domains furthest from the membrane. This raises the question of what exactly the function of this RIFIN is. Is it still capable of inducing signalling through LILRB1 in this way? If signalling can be induced, by what mechanism is this achieved following RIFIN binding? Further studies of this RIFIN, in both functional assays, such as supported lipid bilayer microscopy, signalling assays, and structural studies would help to answer these question, as well as explain how the RIFIN is able to interact with LILRB1 in a way not previously seen by any other ligand.

The same is also true of the newly identified KIR-binding RIFINs - the effect of these proteins on the activity of NK cells is currently unknown, and conducting similar

experiments to look at the effect of the RIFINs on the immunological synapse would shed light on their functionality. Mutagenesis studies using the KIR2DL1/DS1-binding RIFIN variable region would also be useful in determining whether the conserved binding loop used by the RIFINs for LILRB1 interaction is utilised for other receptors, or whether the RIFINs instead use an alternate binding mode for interaction with the KIR receptors.

6.2 Characterisation of a PIR Protein

6.2.1 The CIR Structure

The second major aim of this D.Phil. was the structural characterisation of a member of the PIR family. This was achieved through expression and crystallisation of a CIR from *P. chabaudi*. The main hurdle of this project was during experimental phasing, where numerous attempts at heavy atom derivitisation of the CIR crystals were unsuccessful. This problem was eventually solved through the use of a long-wavelength beamline allowing for phasing using sulfur-SAD.

This structure revealed that, like the RIFIN, the CIR adopted a compact, highly α-helical structure, composed of two distinct helical clusters. This is a novel fold, consisting of an N-terminal region of 5 helices connected by long loops, largely lacking in secondary structure, which is packed against a C-terminal three-helical bundle connected by much shorter loops. The structure is stabilised by 5 disulphide bonds, 3 of which are involved in stabilising loops, with the final 2 lying at the interface between helices.

Sequence analysis revealed that the fold is likely to be conserved among CIRs, with a number of hydrophobic residues found within the core of the domain, along with 2 of the disulphide bonds, showing relatively high degrees of confirmation. Structure based threading of CIR sequences provided further support for this, with high

confidence homology models produced for all tested CIRs.

6.2.2 Conservation of the PIRs

Homology appears to be equally clear for PIR proteins from the majority of the families tested, with those from the other 2 rodent-infective species, *P. berghei* and *P. yoelii*, along with 4 human-infective species, *P. knowlesi*, *P. malariae*, *P. ovale curtisi*, and *P. vivax*, again threading successfully onto the CIR structure. This was not the case for the two VSA families tested from *P. falciparum*, with neither the RIFINs or the STEVORs appearing to show homology with the CIR, further challenging the assertion that they belong within the PIR family.

There are still a number of questions to be answered surrounding the PIR proteins, however. Functional information is scarce, and although they have been implicated in modulation of parasite virulence[168], and the outcome of infection with *P. chabaudi*[88], a mechanism for how this may occur is unclear. Structural studies of other CIRs, including both those associated with chronic infection, and those associated with acute infection, may help identify key differences which could provide some functional insight.

The binding of VIRs to ICAM-1 could provide a mechanistic explanation[90,171] for the effect of CIRs on parasite virulence, supported by reported binding of CIRs to uninfected red blood cells[164] suggesting the PIRs are contributing to infected cell cytoadhesion, perhaps in a similar way to the PfEMP1[126]. Structural studies of VIRs with this endothelial receptor would confirm this, as well as demonstrate whether or not a similar binding mode is used for this interaction as used by members of the PfEMP1 family. Additionally, the discovery of other PIR binding partners would open up further avenues to investigate their function, such as explaining how the CIRs are interacting with red blood cells, and whether this interaction is sufficient to explain their influence on parasite virulence.

6.3 Summary of the Findings of this Thesis

In this thesis, the first structure of a RIFIN was revealed, showing the adoption of a compact α-helical fold, along with a molecular mechanism of interaction of this protein with the inhibitory immune receptor LILRB1, with this mechanism conserved across all tested RIFINs. The structure also revealed that the mode of binding mimics that of the native ligand for the receptor, MHC class I molecules, allowing the RIFIN to induce signalling through the receptor, leading to inhibition of the host immune response. The structure of a CIR was also determined, the first of any member of the PIR family, showing the adoption of a novel α-helical fold. While no structural homologues of the CIR were identified in *P. falciparum*, the fold is conserved across the CIRs, as well as across PIRs from a wide range of both human- and rodent-infective *Plasmodium* species.

Appendix A

Constructs and Primers

Protein	Construct	Vector	Expected Molecular Weight (kDa)	Known Protein Binding Partner	Primers
RIFIN Ectodomain Constructs:					
Pf3D7_1040300	50-335	pHLsec	32.5	LAIR1 [156]	F: AAAAAACCGGTTGCGAGTGCGAGCTGTACTC
					R: AAAAAGGTACCACTCTGGTTGCTCATG
Pf3D7_1101100	42-306	pHLsec	30.8	LAIR1 [67]	F: AAAAAACCGGTTGCGAGTGCGAACTGTACG
					R: AAAAAGGTACCTGTTTGGTAGCCGCCGTAG
Pf3D7_1400600	39-325	pHLsec	32.6	LAIR1 [156]	F: AAAAAACCGGTTGCGAGTGCGAGCTGTACTCC
					R: AAAAAGGTACCAGGGGTGGTAGC
Pf3D7_0100400	39-329	pHLsec	33.7	LILRB1 [67]	F: AAAAAACCGGTTGCGAGTGCGAGCTGTACAC
					R: AAAAAGGTACCTGTATAGTAGGAGGTGGTGGCGG
Pf3D7_0500400	43-331	pHLsec	33.1	LILRB1 [67]	F: AAAAAACCGGTTATGAGTGCGGCCTGTACAGC
					R: AAAAAGGTACCTGTCTGCTTGCTCATGAAGATGGTG
Pf3D7_0632200	42-334	pHLsec	33.3	LILRB1 [67]	F: AAAAAACCGGTAGCGAGTGCGAACTGTACG
					R: AAAAAGGTACCTGTCTGGTTGCTCATGTAGATGGTG
Pf3D7_1100400	45-280	pHLsec	27.4	LILRB1 [67]	F: AAAAAACCGGTTGCGAGTGTGAACTGTACGCC
					R: AAAAAGGTACCGTCGGTCAGAAAGGCGGTGA
Pf3D7_1254800	39-274	pHLsec	27.9	LILRB1 [67]	F: AAAAAACCGGTTGCGAGTGCGAGCTGTACATG
					R: AAAAAGGTACCGTTGGTGAAGAAGGTGGTAGGGG
Pf3D7_1480000	38-297	pHLsec	30.2	LILRB1 [67]	F: AAAAAACCGGTTGCGAGTGCGAGCTGTATGAG
					R: AAAAAGGTACCTGTGGTCAGGTTGCTGGTCAC
Pf3D7_0100200	43-289	pET15b	29.7	LILRB1 [67]	F: AAAAACATATGTGCGAGTGCGAGCTGTACTC
					R: AAAAAGGATCCTCAGGCATTGCTGTATGTTGTCAGG
Pf3D7_0832500	49-283	pET15b	28.3	-	F: AAAAACATATGTGCGAGTGCGACCCCTAC
					R: AAAAAGGATCCTCAGGCGATATTGAAGCTGGTG
Pf3D7_0832800	39-264	pET15b	26.7	-	F: AAAAACATATGAGCGAGTGCGACACCTACAC
					R: AAAAAGGATCCTCATGTGCTGTAGGCGGTGC
Pf3D7_1040900	46-269	pET15b	26.8	-	F: AAAAACATATGAGCGAGTGCGACATCAGAAG
					R: AAAAAGGATCCTCAGCCGGTGAAGTAGGCTC
Pf3D7_0100400	39-329	pExpreS2-1	33.5	LILRB1 [67]	F: AAAAAGAATTCTGCGAGTGCGAGCTGTACAC

Protein	Construct	Vector	Expected Molecular Weight (kDa)	Known Protein Binding Partner	Primers
					R: AAAAAGCGGCCGCTTATG-TATAGTAGGAGGTGGTGGCG
Pf3D7_1040300	50-335	pExpreS2-1	33.0	LAIR1[156]	F: AAAAAGAATTCTGCGAGT-GCGAGCTGTACTC
					R: AAAAAGCGGCCGCT-TACTGGTTGCTCATGAA-GATGGTGTTG
Pf3D7_1254800	39-274	pExpreS2-1	28.6	LILRB1[67]	F: AAAAAGAATTCTGCGAGT-GCGAGCTGTACATG
					R: AAAAAGCGGCCGCT-TAGTTGGTGAAGAAGGTG-GTAGGGG
RIFIN Constant Region Constructs:					
Pf3D7_0401300	45-146	pHLsec	13.6	-	F: AAAAAACCGGTTGC-GAGTGCGAGCTGTACG
					R: AAAAAGGTAC-CTCTCAGGCACACTTTTTC-CACCT
Pf3D7_0401300	45-185	pHLsec	17.3	-	F: AAAAAACCGGTTGC-GAGTGCGAGCTGTACG
					R: AAAAAGGTACCCATAGCG-GCTTTTGTAGCGGC
Pf3D7_0500400	44-145	pHLsec	13.6	LILRB1[67]	F: AAAAAACCGGTTATGAGT-GTGGCCTGTACAGCC
					R: AAAAAGGTAC-CTCTCAGGCAGCC-CTTTTCCA
Pf3D7_0500400	44-181	pHLsec	17.1	LILRB1[67]	F: AAAAAACCGGTTATGAGT-GTGGCCTGTACAGCC
					R: AAAAAGGTACCAATAGCG-GCCTCAATGGCG
Pf3D7_1040800	40-115	pHLsec	10.8	-	F: AAAAAACCGGTAGCGAGT-GCAACGTGTACAT
					R: AAAAAGGTACC-CTTCAGGCAGCCCTTTTC-CAC
Pf3D7_1040800	40-151	pHLsec	14.1	-	F: AAAAAACCGGTAGCGAGT-GCAACGTGTACAT
					R: AAAAAGGTACCAATG-GCCAGATCAATAGCGGC
Pf3D7_1254800	39-139	pHLsec	13.6	LILRB1[67]	F: AAAAAACCGGTTGC-GAGTGCG
					R: AAAAAGGTACC-CTTCAGGCAGGTCTTTTCAATCT-TATCGG
Pf3D7_1254800	39-164	pHLsec	15.9	LILRB1[67]	F: AAAAAACCGGTTGC-GAGTGCG
					R: AAAAAGGTACCG-GCCAGGATGTACACGCTG
Pf3D7_1373400	39-140	pHLsec	13.4	-	F: AAAAAACCGGTTGC-GAGTGTGAACTGTACGC
					R: AAAAAGGTAC-CTTTCAGGCAGCCCTTTTC-CAC
Pf3D7_1373400	39-177	pHLsec	17.0	-	F: AAAAAACCGGTTGC-GAGTGTGAACTGTACGC
					R: AAAAAGGTAC-CTTCAGCCAGTTCTT-TAGCAGCG
RIFIN Variable Region Constructs:					
Pf3D7_1254800	165-274	pHLsec	13.5	LILRB1[67]	F: AAAAAACCGGTAACTAC-GAGACAATCAACGCCTT-TATCG
					R: AAAAAGGTACCGTTGGT-GAAGAAGGTGGTAGGGG
Pf3D7_1254800	165-274 T261D	pHLsec	13.5	LILRB1[67]	F: AGCAGCGTGAACGAT-GACTGGGCCACAGCTACA
					R: TGTAGCTGTGGCCCAGT-CATCGTTCACGCTGCT

Protein	Construct	Vector	Expected Molecular Weight (kDa)	Known Protein Binding Partner	Primers
Pf3D7_1254800	165-274	pET15b	14.4	LILRB1[67]	F: AAAAACATATGAACTAC-GAGACAATCAACGCCTT-TATCG
					R: AAAAAGGATCCGTTGGT-GAAGAAGGTGGTAGG
Pf3D7_1254800	165-261 C223S	pET15b	12.9	LILRB1[67]	F: AAAAAGCTGTGCCT-GAGCGCCGCCAACAAGAAC
					R: GTTCTTGTTGGCG-GCGCTCAGGCACAGCTTTTT
Pf3D7_1254800	165-261 C223S+P184W	pET15b	13.0	LILRB1[67]	F: GAACTGGAAG-GCATCTGGGGCATCAC-CAAGCTG
					R: CAGCTTGGTGATGCC-CCAGATGCCTTCCAGTTC
Pf3D7_1254800	165-261 C223S+G234R	pET15b	13.0	LILRB1[67]	F: GAGCTGCTGTGCAGAA-GAATGAACCCCAACGTG
					R: CACGTTGGGGTTCATTCTTCT-GCACAGCAGCTC
Pf3D7_1254800	165-261 C223S+G234W	pET15b	13.1	LILRB1[67]	F: GAGCTGCTGTGCAGATG-GATGAACCCCAACGTG
					R: CACGTTGGGGTTCATC-CATCTGCACAGCAGCTC
Pf3D7_1254800	165-261 C223S+G234Y	pET15b	13.0	LILRB1[67]	F: GAGCTGCTGTGCAGATA-CATGAACCCCAACGTG
					R: CACGTTGGGGTTCATG-TATCTGCACAGCAGCTC
Pf3D7_0100200	170-288	pET15b	15.1	LILRB1[67]	F: AAAAACATATGAAC-TACGTGACCCAGACAGCTC
					R: AAAAAGGATCCTTAGCTG-TATGTTGTCAGGCCAGG
Pf3D7_0100200	170-288 S244R	pET15b	15.2	LILRB1[67]	F: CACAGCTTCTGCTTCCG-TACCTACATCTCCAAC
					R: GTTGGAGATGTAGGTACG-GAAGCAGAAGCTGTG
Pf3D7_0100400	165-329	pET15b	20.2	LILRB1[67]	F: AAAAACATATGGTGTG-GAAAACCACAGAAATCGCC
					R: AAAAAGGATCCTTATG-TATAGTAGGAGGTGGTGGCG
Pf3D7_0100400	165-329 L264R	pET15b	20.2	LILRB1[67]	F: GATCCTTTCTGCAACCGT-GTGCACAGCAACCTG
					R: CAGGTTGCTGTGCA-CACGGTTGCAGAAAGGATC
Pf3D7_1100400	169-275	pET15b	13.3	LILRB1[67]	F: AAAAACATATGGCCGCT-GTGAAAGCCGC
					R: AAAAAGGATCCTTAGGT-GATGGTTGTAGAGTCGCT-GAAG
Pf3D7_1100400	169-275 G240R	pET15b	13.4	LILRB1[67]	F: CAGGTGTTCTG-TAAACGTCTGCTGCCTGA-GAGC
					R: GCTCTCAGGCAGCA-GACGTTTACAGAACACCTG
KIR2DL1/DS1-Binding RIFIN	1-306	pHLsec	34.8	KIR2DL1, KIR2DS1	F: AAAAAACCGGTAT-GAAGCTGCAGTACTTCAA-GATCCTG
					R: AAAAAGGTACCAGTCT-GACAAGCGGCATAGGC
KIR2DL1/DS1-Binding RIFIN	47-306	pHLsec	29.3	KIR2DL1, KIR2DS1	F: AAAAAACCGGTAGCGAGT-GCGATCTGCAC
					R: AAAAAGGTACCAGTCT-GACAAGCGGCATAGGC
KIR2DL1/DS1-Binding RIFIN	144-306	pHLsec	18.2	KIR2DL1, KIR2DS1	F: AAAAAACCGGTAGCTG-GAAGTCTGTGGCCC
					R: AAAAAGGTACCAGTCT-GACAAGCGGCATAGGC
KIR2DL2/DL3-Binding RIFIN	1-294	pHLsec	33.4	KIR2DL2, KIR2DL3	F: AAAAAACCGGTAT-GAAGCTGCACTACAGCAA-GATTCT
					R: AAAAAGGTACCGGTCT-GAAAGCTGGCGTATTTGG
KIR2DL2/DL3-Binding RIFIN	41-294	pHLsec	28.7	KIR2DL2, KIR2DL3	F: AAAAAACCGGTAACGAGT-GCGACCTGTACATG

Protein	Construct	Vector	Expected Molecular Weight (kDa)	Known Protein Binding Partner	Primers
					R: AAAAAGGTACCGGTCTGAAAGCTGGCGTATTTGG
KIR2DL2/DL3-Binding RIFIN	138-294	pHLsec	17.6	KIR2DL2, KIR2DL3	F: AAAAAACCGGTGGATGGAAGATCGCCGCTACAAC R: AAAAAGGTACCGGTCTGAAAGCTGGCGTATTTGG
Inhibitory Immune Receptor Constructs:					
LILRB1	24-417 (FL)	pHLsec	44.8	MHC Class I[183], UL18[195], BST2[184], angiopoietin-like proteins[185], S100A9[186], myelin inhibitors[187], β-amyloids[188], DENV[198], RIFIN[67]	F: AAAAAACCGGTGGACATCTGCCCAAAC R: AAAAAGGTACCCAGCTCGAGAGGATCGGATGG
LILRB1	24-221 (D1D2)	pHLsec	23.5	MHC Class I[183], UL18[195], BST2[184], angiopoietin-like proteins[185], S100A9[186], myelin inhibitors[187], β-amyloids[188], DENV[198], RIFIN[67]	F: AAAAAACCGGTGGACATCTGCCCAAAC R: AAAAAGGTACCGCCCAGCACAAGCAGTTCC
LILRB1	24-221 (D1D2)	pET15b	24.4	MHC Class I[183], UL18[195], BST2[184], angiopoietin-like proteins[185], S100A9[186], myelin inhibitors[187], β-amyloids[188], DENV[198], RIFIN[67]	F: AAAAACATATGGGACATCTGCCCAAACCTAC R: AAAAAGGATCCTTAGCCCAGCACAAGCAGTTCC
LILRB1	222-417 (D3D4)	pHLsec	22.7	MHC Class I[183], UL18[195], BST2[184], angiopoietin-like proteins[185], S100A9[186], myelin inhibitors[187], β-amyloids[188], DENV[198], RIFIN[67]	F: AAAAAACCGGTGTGTCCAAGAAACCCAGCCTGTC R: AAAAAGGTACCCAGCTCGAGAGGATCGGATGG
KIR2DL1	27-221	pHLsec	22.9	MHC Class I[208], RIFIN	F: AAAAAACCGGTAGAAAACCTAGCCTGCTGGCC R: AAAAAGGTACCGGTCACGGACACCAGCAG
KIR2DS1	27-221	pHLsec	23.0	MHC Class I[208], RIFIN	F: AAAAAACCGGTAGAAAACCTAGCCTGCTGGCC R: AAAAAGGTACCTGTCACGGACACCAGCAG
CIR Constructs (Provided by Jean Langhorne, Francis Crick Institute):					
PCHAS_0007200	1-235	pTT3	31.6	-	-
PCHAS_0007200	1-235	pTT3 (with CD4-tag)	52.7	-	-
PCHAS_0101200	1-285	pTT3	34.7	-	-
PCHAS_0101200	1-285	pTT3 (with CD4-tag)	58.5	-	-
PCHAS_0400300	1-268	pTT3	32.0	-	-
PCHAS_0400300	1-268	pTT3 (with CD4-tag)	55.7	-	-

Protein	Construct	Vector	Expected Molecular Weight (kDa)	Known Protein Binding Partner	Primers
PCHAS_0524800	1-157	pTT3	20.2	-	-
PCHAS_0524800	1-157	pTT3 (with CD4-tag)	43.9	-	-
PCHAS_0600600	1-268	pTT3	31.6	-	-
PCHAS_0600600	1-268	pTT3 (with CD4-tag)	55.3	-	-
PCHAS_1100300	1-524	pTT3	52.3	-	-
PCHAS_1100300	1-524	pTT3 (with CD4-tag)	84.0	-	-
PCHAS_1200500	1-264	pTT3	31.1	-	-
PCHAS_1200500	1-264	pTT3 (with CD4-tag)	54.9	-	-

Table A.1: Summary of Protein Constructs and Primers. A list of all protein constructs used in this thesis, including sequences of primers using for their cloning. The F and R refer to forward and reverse primers respectively.

Appendix B

Kinetic Values for Binding of the KIR2DL1/DS1-Binding RIFIN Variable Region and HLA-Cw4 to KIR2DL1 and KIR2DS1

	$\mathbf{k}_{on}$ ($\mathbf{M}^{-1}s^{-1}$)	$\mathbf{k}_{off}$ ($\mathbf{s}^{-1}$)	$\mathbf{K}_d$ (μM)	χ^2 ($\mathbf{RU}^2$)
RIFIN vs KIR2DL1				
1^{st} Repeat	5.46×10^4	0.0219	0.400	0.577
2^{nd} Repeat	4.79×10^4	0.0213	0.444	0.465
3^{rd} Repeat	4.22×10^4	0.0209	0.494	0.589
RIFIN vs KIR2DS1				
1^{st} Repeat	1.48×10^4	0.0401	2.71	12.1
2^{nd} Repeat	2.32×10^4	0.0398	1.72	0.767
3^{rd} Repeat	2.20×10^4	0.0392	1.79	0.675
HLA-Cw4 vs KIR2DL1				
1^{st} Repeat	2.13×10^5	1.29	6.08	0.992
2^{nd} Repeat	1.91×10^5	1.22	6.36	7.35
3^{rd} Repeat	3.91×10^5	2.41	6.15	0.377
HLA-Cw4 vs KIR2DS1				
1^{st} Repeat	2.19×10^6	45.8	20.9	0.623
2^{nd} Repeat	5.12×10^5	12.5	24.4	3.17
3^{rd} Repeat	4.79×10^6	117	24.4	0.251

Table B.1: Kinetic Values for Binding of the KIR2DL1/DS1-Binding RIFIN Variable Region and HLA-Cw4 to KIR2DL1 and KIR2DS1 These values were calculated using fitting to a one-to-one kinetic binding model, with the quality of fitting described by the χ^2 value.

Appendix C

CIR Modelling Data

Family	Sequence	Sequence Identity (%)	Confidence (%)	Number of Aligned Residues
P. chabaudi	PCHAS_0115100	32	100	229
	PCHAS_0301500	29	100	224
	PCHAS_0319700	37	100	231
	PCHAS_0400500	24	100	217
	PCHAS_0419800	23	100	217
	PCHAS_0837200	18	100	223
	PCHAS_0938341	33	100	235
	PCHAS_1200500	97	100	239
	PCHAS_1302200	20	100	225
	PCHAS_1400400	29	100	232
P. berghei	PBANKA_0008201	31	100	235
	PBANKA_0100500	10	98.94	221
	PBANKA_0215700	32	100	226
	PBANKA_0300200	33	100	229
	PBANKA_0300300	30	100	222
	PBANKA_0700041	33	100	234
	PBANKA_0836900	31	100	227
	PBANKA_0837081	39	100	224
	PBANKA_0837141	32	100	227
	PBANKA_1040581	22	100	211
P. yoelii	PY17X_0216900	24	100	224
	PY17X_0217800	20	100	211
	PY17X_0319901	33	100	226
	PY17X_0601200	30	100	230
	PY17X_0627300	30	100	223
	PY17X_0844501	32	100	224
	PY17X_1046701	34	100	223
	PY17X_1100063	27	100	225
	PY17X_1373000	29	100	233
	PY17X_1400800	33	100	228
P. knowlesi	PKNH_0202800	11	98.9	215
	PKNH_0210400	13	98.76	209
	PKNH_0312000	16	98.16	200
	PKNH_0501600	15	98.33	218
	PKNH_0727500	11	98.93	206
	PKNH_1000900	12	98.49	192
	PKNH_1002800	11	98.79	215
	PKNH_1312300	14	98.74	208
	PKNH_1323200	11	98.86	219
	PKNH_1469800	14	98.93	189
P. malariae	PmUG01_00031600	16	98.6	213
	PmUG01_00034600	12	98.79	212
	PmUG01_00043200	12	99.03	201

Family	Sequence	Sequence Identity (%)	Confidence (%)	Number of Aligned Residues
	PmUG01_00044700	12	98.64	216
	PmUG01_00048700	12	98.66	218
	PmUG01_00058600	16	88.94	188
	PmUG01_02010800	14	98.43	211
	PmUG01_03011600	12	98.95	212
	PmUG01_10052000	12	98.95	219
	PmUG01_14085800	14	98.02	202
P. ovale	PocGH01_00011100	8	98.68	196
	PocGH01_00014500	12	98.75	212
	PocGH01_00077800	16	98.93	209
	PocGH01_00113000	13	98.72	223
	PocGH01_00114900	11	97.22	183
	PocGH01_00135200	14	99.04	212
	PocGH01_00136200	10	98.94	207
	PocGH01_00136400	11	99.56	223
	PocGH01_00137000	16	99.4	225
	PocGH01_00201300	18	94.15	173
P. vivax	PVP01_0000410	14	99.79	223
	PVP01_0000520	23	91.57	139
	PVP01_0002040	11	98.83	210
	PVP01_0002900	16	98.7	222
	PVP01_0005000	18	96.2	85
	PVP01_0007150	13	98.29	210
	PVP01_0009430	18	98.84	210
	PVP01_0201800	12	99.16	206
	PVP01_0219900	12	98.38	200
	PVP01_0533800	14	98.93	217
P. falciparum (RIFIN)	PF3D7_0101000	19	78.79	182
	PF3D7_0222700	10	67.3	217
	PF3D7_0617500	11	63.04	166
	PF3D7_0617700	15	53.94	197
	PF3D7_0732300	16	73.13	182
	PF3D7_0900600	20	74.67	192
	PF3D7_0901500	25	65.21	200
	PF3D7_1000400	18	66.79	194
	PF3D7_1240700	16	78.64	201
	PF3D7_1373100	18	79.34	187
P. falciparum (RIFIN-C)	PF3D7_0101000	22	1.35	86
	PF3D7_0222700	15	0.11	73
	PF3D7_0617500	18	2.64	93
	PF3D7_0617700	26	0.97	86
	PF3D7_0732300	23	0.88	97
	PF3D7_0900600	19	0.67	97
	PF3D7_0901500	18	0.64	98
	PF3D7_1000400	20	1.51	87
	PF3D7_1240700	29	0.83	84
	PF3D7_1373100	18	1.01	98
P. falciparum (RIFIN-V)	PF3D7_0101000	17	22.73	132
	PF3D7_0222700	10	23.26	136
	PF3D7_0617500	23	35.57	137
	PF3D7_0617700	11	8.1	120
	PF3D7_0732300	15	9.65	137
	PF3D7_0900600	14	22.32	136

Family	Sequence	Sequence Identity (%)	Confidence (%)	Number of Aligned Residues
	PF3D7_0901500	14	25.5	150
	PF3D7_1000400	14	34.28	149
	PF3D7_1240700	11	3	137
	PF3D7_1373100	14	13.83	139
P. falciparum (STEVOR)	PF3D7_0324600	16	86.35	165
	PF3D7_0732000	13	89.78	165
	PF3D7_0832000	16	86.67	151
	PF3D7_0832600	17	79.8	180
	PF3D7_0832900	16	85.8	173
	PF3D7_0901600	14	85.22	187
	PF3D7_1040200	14	64.98	193
	PF3D7_1149900	14	91.04	181
	PF3D7_1254600	13	77.74	192
	PF3D7_1300900	14	82.47	188
P. reichenowi	PRG01_0004300	17	0.43	96
	PRG01_0009500	19	86.73	188
	PRG01_0010400	14	53.72	191
	PRG01_0011600	17	77.21	194
	PRG01_0016700	20	67.01	192
	PRG01_0302700	15	76.59	200
	PRG01_0403400	19	63.49	180
	PRG01_0600500	19	77.01	196
	PRG01_0900500	14	72.57	198
	PRG01_0908200	15	88.83	163
P. gallinaceum	PGAL8A_00178300	19	61.74	184
	PGAL8A_00194500	24	82.81	135
	PGAL8A_00195300	9	0	77
	PGAL8A_00264300	20	6	105
	PGAL8A_00266200	23	34.55	136
	PGAL8A_00266400	20	19.3	171
	PGAL8A_00266500	23	33.72	145
	PGAL8A_00268500	21	92.97	199
	PGAL8A_00268600	22	91.52	216
	PGAL8A_00377700	18	2.51	101

Table C.1: Structure-Based Threading of Putative PIRs Sequence identities of 10 representatives from each potential PIR family with the PCHAS_1200500 ectodomain, and confidence scores for Phyre2 threading of these sequences using the PCHAS_1200500 structure. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266]. The length of sequence threaded onto the structure is also shown. RIFIN-C and RIFIN-V refer to the constant and variable regions respectively of the selected sequences.

Sequence	Sequence Identity (%)	Confidence (%)	Number of Aligned Residues
Modelled vs N Terminal Lobe of CIR			
Pf3D7_0101000	22	0.45	86
Pf3D7_0222700	26	0.40	61
Pf3D7_0617500	20	0.36	87
Pf3D7_0617700	22	0.29	86
Pf3D7_0732300	16	0.36	86
Pf3D7_0900600	19	0.41	90
Pf3D7_0901500	20	0.41	90
Pf3D7_1000400	21	0.53	90
Pf3D7_1240700	16	0.43	56
Pf3D7_1373100	21	0.43	90
Modelled vs C Terminal Lobe of CIR			
Pf3D7_0101000	21	1.35	73
Pf3D7_0222700	12	0.11	59
Pf3D7_0617500	13	2.64	72
Pf3D7_0617700	22	0.97	73
Pf3D7_0732300	16	0.88	73
Pf3D7_0900600	18	0.67	73
Pf3D7_0901500	18	0.64	73
Pf3D7_1000400	19	1.51	73
Pf3D7_1240700	15	0.83	74
Pf3D7_1373100	14	1.01	73

Table C.2: Structure-Based Threading of RIFINs Sequence identities of 10 representatives of the RIFIN constant region with the two lobes of the PCHAS_1200500 ectodomain, and confidence scores for Phyre2 threading of these sequences using the same regions of the PCHAS_1200500 structure. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266]. The length of sequence threaded onto the structure is also shown.

Sequence	Sequence Identity (%)	Confidence (%)	Number of Aligned Residues
PRG01_0004300	24	1.08	76
PRG01_0009500	14	78.74	98
PRG01_0010400	17	76.94	99
PRG01_0011600	15	86.35	98
PRG01_0016700	15	89.89	98
PRG01_0302700	18	92.64	99
PRG01_0403400	19	96.49	98
PRG01_0600500	14	85.98	98
PRG01_0900500	13	90.92	98
PRG01_0908200	19	15.98	88

Table C.3: Structure-Based Threading of Putative PIRs from *P. reichenowi* onto the RIFIN Variable Region Structure Sequence identities of 10 putative PIR sequences from *P. reichenowi* with the variable region of the RIFIN Pf3D7_1254800, and confidence scores for Phyre2 threading of these sequences using the structure of the same RIFIN variable region. A value of >90% indicates with strong confidence that the sequence adopts the given fold[266]. The length of sequence threaded onto the structure is also shown.

Appendix D
Publications Arising from this Thesis

Harrison, T. E., Mørch, A. M., Felce, J. H., Sakoguchi, A., Reid, A. J., Arase, H., Dustin, M. L. & Higgins M. K. Structural basis for RIFIN-mediated activation of LILRB1 in malaria. *Nature* **587**, 309-312 (2020). https://doi.org/10.1038/s41586-020-2530-3

Harrison T. E., Reid A. J., Cunningham D., Langhorne J. & Higgins M. K. Structure of the *Plasmodium*-interspersed repeat proteins of the malaria parasite. *Proceedings of the National Academy of Sciences*, (2020). https://doi.org/10.1073/pnas.2016775117

Bibliography

[1] Cox, F. E. History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors* **3**, 5 (2010).

[2] Hempelmann, E. & Krafts, K. Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malaria Journal* **12**, 232 (2013).

[3] Ross, R. The role of the mosquito in the evolution of the malarial parasite. *The Lancet* **152**, 488–490 (1898).

[4] Carter, R. & Mendis, K. N. Evolutionary and Historical Aspects of the Burden of Malaria. *Clinical Microbiology Reviews* **15**, 564–594 (2002).

[5] WHO. World Malaria Report 2019. Geneva. Tech. Rep. (2019).

[6] Cibulskis, R. E. *et al.* Malaria: Global progress 2000 – 2015 and future challenges. *Infectious Diseases of Poverty* **5**, 61 (2016).

[7] Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

[8] Lyke, K. E. Steady progress toward a malaria vaccine. *Current Opinion in Infectious Diseases* **30**, 463–470 (2017).

[9] Arisue, N. & Hashimoto, T. Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitology International* **64**, 254–259 (2015).

[10] Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. Plasmodium Biology. *Cell* **115**, 771–785 (2003).

[11] Milner, D. A. Malaria Pathogenesis. *Cold Spring Harbor perspectives in medicine* **8** (2018).

[12] Barber, B. E., Grigg, M. J., William, T., Yeo, T. W. & Anstey, N. M. The Treatment of Plasmodium knowlesi Malaria. *Trends in Parasitology* **33**, 242–253 (2017).

[13] Barber, B. E. *et al.* A Prospective Comparative Study of Knowlesi, Falciparum, and Vivax Malaria in Sabah, Malaysia: High Proportion With Severe Disease From Plasmodium Knowlesi and Plasmodium Vivax But No Mortality With Early Referral and Artesunate Therapy. *Clinical Infectious Diseases* **56**, 383–397 (2013).

[14] Lover, A. A., Baird, J. K., Gosling, R. & Price, R. N. Malaria Elimination: Time to Target All Species. *The American Journal of Tropical Medicine and Hygiene* **99**, 17–23 (2018).

[15] Mueller, I., Zimmerman, P. A. & Reeder, J. C. Plasmodium malariae and Plasmodium ovale – the 'bashful' malaria parasites. *Trends in Parasitology* **23**, 278–283 (2007).

[16] Sutherland, C. J. *et al.* Two Nonrecombining Sympatric Forms of the Human Malaria Parasite Plasmodium ovale Occur Globally. *The Journal of Infectious Diseases* **201**, 1544–1550 (2010).

[17] Krief, S. *et al.* On the diversity of malaria parasites in African apes and the origin of Plasmodium falciparum from bonobos. *PLoS Pathogens* (2010).

[18] Perkins, S. L., Sarkar, I. N. & Carter, R. The phylogeny of rodent malaria parasites: Simultaneous analysis across three genomes. *Infection, Genetics and Evolution* **7**, 74–83 (2007).

[19] Valkiūnas, G. & Iezhova, T. A. Keys to the avian malaria parasites. *Malaria Journal* **17**, 212 (2018).

[20] Beier, J. C. Malaria parasite development in mosquitoes. *Annual Review of Entomology* **43**, 519–543 (1998).

[21] Jin, Y., Kebaier, C. & Vanderberg, J. Direct microscopic quantification of dynamics of Plasmodium berghei sporozoite transmission from mosquitoes to mice. *Infection and immunity* **75**, 5532–9 (2007).

[22] Amino, R. *et al.* Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nature medicine* **12**, 220–4 (2006).

[23] Amino, R. *et al.* Imaging malaria sporozoites in the dermis of the mammalian host. *Nature protocols* **2**, 1705–12 (2007).

[24] Vaughan, A. M., Aly, A. S. & Kappe, S. H. Malaria Parasite Pre-Erythrocytic Stage Infection: Gliding and Hiding. *Cell Host & Microbe* **4**, 209–218 (2008).

[25] Pradel, G., Garapaty, S. & Frevert, U. Proteoglycans mediate malaria sporozoite targeting to the liver. *Molecular microbiology* **45**, 637–51 (2002).

[26] Pradel, G., Garapaty, S. & Frevert, U. Kupffer and stellate cell proteoglycans mediate malaria sporozoite targeting to the liver. *Comparative hepatology* **3 Suppl 1**, S47 (2004).

[27] Carrolo, M. *et al.* Hepatocyte growth factor and its receptor are required for malaria infection. *Nature medicine* **9**, 1363–9 (2003).

[28] Baer, K., Klotz, C., Kappe, S. H. I., Schnieder, T. & Frevert, U. Release of Hepatic Plasmodium yoelii Merozoites into the Pulmonary Microvasculature. *PLoS Pathogens* **3**, e171 (2007).

[29] Silvie, O., Mota, M. M., Matuschewski, K. & Prudêncio, M. Interactions of the malaria parasite and its mammalian host. *Current Opinion in Microbiology* **11**, 352–359 (2008).

[30] Collins, W. E. & Jeffery, G. M. Plasmodium malariae: Parasite and Disease. *Clinical Microbiology Reviews* **20**, 579–592 (2007).

[31] Cowman, A. F. & Crabb, B. S. Invasion of Red Blood Cells by Malaria Parasites. *Cell* **124**, 755–766 (2006).

[32] Goel, V. K. *et al.* Band 3 is a host receptor binding merozoite surface protein 1 during the Plasmodium falciparum invasion of erythrocytes. *Proceedings of the National Academy of Sciences* **100**, 5164–5169 (2003).

[33] Batchelor, J. D. *et al.* Red Blood Cell Invasion by Plasmodium vivax: Structural Basis for DBP Engagement of DARC. *PLoS Pathogens* **10**, e1003869 (2014).

[34] Lopaticki, S. *et al.* Reticulocyte and Erythrocyte Binding-Like Proteins Function Cooperatively in Invasion of Human Erythrocytes by Malaria Parasites. *Infection and Immunity* **79**, 1107–1117 (2011).

[35] 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. *Journal of Biological Chemistry* **285**, 14815–14822 (2010).

[36] Aikawa, M., Miller, L., 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).

[37] Bannister, L., Hopkins, J., Fowler, R., Krishna, S. & Mitchell, G. A Brief Illustrated Guide to the Ultrastructure of Plasmodium falciparum Asexual Blood Stages. *Parasitology Today* **16**, 427–433 (2000).

[38] Liu, Z., Miao, J. & Cui, L. Gametocytogenesis in malaria parasite: commitment, development and regulation. *Future Microbiology* **6**, 1351–1369 (2011).

[39] Thomson, J. & Robertson, A. The structure and development of plasmodium falciparum gametocytes in the internal organs and peripheral circulation. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **29**, 31–40 (1935).

[40] Siciliano, G. *et al.* Critical Steps of Plasmodium falciparum Ookinete Maturation. *Frontiers in Microbiology* **11** (2020).

[41] Vaughan, J. A. Population dynamics of Plasmodium sporogony. *Trends in Parasitology* **23**, 63–70 (2007).

[42] White, N. J. *et al.* Malaria. *The Lancet* **383**, 723–735 (2014).

[43] Dondorp, A. M. *et al.* The Relationship between Age and the Manifestations of and Mortality Associated with Severe Malaria. *Clinical Infectious Diseases* **47**, 151–157 (2008).

[44] Cooke, B., Coppel, R. & Wahlgren, M. Falciparum Malaria: Sticking up, Standing out and Out-standing. *Parasitology Today* **16**, 416–420 (2000).

[45] 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* **98**, 1805–1810 (2001).

[46] Doumbo, O. K. *et al.* High Levels of Plasmodium falciparum Rosetting in All Clinical Forms of Severe Malaria in African Children. *The American Journal of Tropical Medicine and Hygiene* **81**, 987–993 (2009).

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

[48] Idro, R., Marsh, K., John, C. C. & Newton, C. R. J. Cerebral Malaria: Mechanisms of Brain Injury and Strategies for Improved Neurocognitive Outcome. *Pediatric Research* **68**, 267–274 (2010).

[49] White, M. T. *et al.* Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. *Parasites & Vectors* **4**, 153 (2011).

[50] Mitchell, S. N. *et al.* Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT Resistance in Anopheles gambiae. *PLoS ONE* **9**, e92662 (2014).

[51] Walker, T. & Moreira, L. A. Can Wolbachia be used to control malaria? *Memórias do Instituto Oswaldo Cruz* **106**, 212–217 (2011).

[52] Herren, J. K. *et al.* A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. *Nature Communications* **11**, 2187 (2020).

[53] Sullivan, D. J., Gluzman, I. Y., Russell, D. G. & Goldberg, D. E. On the molecular mechanism of chloroquine's antimalarial action. *Proceedings of the National Academy of Sciences* **93**, 11865–11870 (1996).

[54] Martin, R. E. *et al.* Chloroquine Transport via the Malaria Parasite's Chloroquine Resistance Transporter. *Science* **325**, 1680–1682 (2009).

[55] Cui, L. & Su, X.-z. Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-infective Therapy* **7**, 999–1013 (2009).

[56] Dondorp, A. M. *et al.* Artemisinin Resistance in Plasmodium falciparum Malaria. *New England Journal of Medicine* **361**, 455–467 (2009).

[57] Laurens, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Human Vaccines & Immunotherapeutics* **16**, 480–489 (2020).

[58] Schuerman, L. RTS,S malaria vaccine could provide major public health benefits. *The Lancet* **394**, 735–736 (2019).

[59] Srinivasan, P. *et al.* A malaria vaccine protects Aotus monkeys against virulent Plasmodium falciparum infection. *npj Vaccines* **2**, 14 (2017).

[60] Douglas, A. D. *et al.* The blood-stage malaria antigen PfRH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody. *Nature Communications* **2**, 601 (2011).

[61] Douglas, A. D. *et al.* A PfRH5-Based Vaccine Is Efficacious against Heterologous Strain Blood-Stage Plasmodium falciparum Infection in Aotus Monkeys. *Cell Host & Microbe* **17**, 130–139 (2015).

[62] Chaturvedi, N., Bharti, P., Tiwari, A. & Singh, N. Strategies & recent development of transmission-blocking vaccines against Plasmodium falciparum. *Indian Journal of Medical Research* **143**, 696 (2016).

[63] Arakawa, T. *et al.* Nasal Immunization with a Malaria Transmission-Blocking Vaccine Candidate, Pfs25, Induces Complete Protective Immunity in Mice against Field Isolates of Plasmodium falciparum. *Infection and Immunity* **73**, 7375–7380 (2005).

[64] Lennartz, F. *et al.* Structural basis for recognition of the malaria vaccine candidate Pfs48/45 by a transmission blocking antibody. *Nature Communications* **9**, 3822 (2018).

[65] McLeod, B. *et al.* Potent antibody lineage against malaria transmission elicited by human vaccination with Pfs25. *Nature Communications* **10**, 4328 (2019).

[66] del Portillo, H. A. *et al.* The role of the spleen in malaria. *Cellular Microbiology* **14**, 343–355 (2012).

[67] Saito, F. *et al.* Immune evasion of Plasmodium falciparum by RIFIN via inhibitory receptors. *Nature* **552**, 101–105 (2017).

[68] Brown, K. N. & Brown, I. N. Immunity to Malaria: Antigenic Variation in Chronic Infections of Plasmodium knowlesi. *Nature* **208**, 1286–1288 (1965).

[69] Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. *The Journal of Experimental Medicine* **159**, 1567–1575 (1984).

[70] 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).

[71] Magowan, C., Wollish, W., Anderson, L. & Leech, J. Cytoadherence by Plasmodium falciparum-infected erythrocytes is correlated with the expression of a family of variable proteins on infected erythrocytes. *The Journal of Experimental Medicine* **168**, 1307–1320 (1988).

[72] 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**, 292–295 (1997).

[73] Gardner, M. J. Chromosome 2 Sequence of the Human Malaria Parasite Plasmodium falciparum. *Science* **282**, 1126–1132 (1998).

[74] Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with

Plasmodium falciparum. *Proceedings of the National Academy of Sciences* **96**, 9333–9338 (1999).

[75] Goel, S. *et al.* RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. *Nature Medicine* **21**, 314–321 (2015).

[76] Cheng, Q. *et al.* Measurement of Plasmodium falciparum Growth Rates in Vivo: A Test of Malaria Vaccines. *The American Journal of Tropical Medicine and Hygiene* **57**, 495–500 (1997).

[77] Niang, M., Yan Yam, X. & Preiser, P. R. The Plasmodium falciparum STEVOR Multigene Family Mediates Antigenic Variation of the Infected Erythrocyte. *PLoS Pathogens* **5**, e1000307 (2009).

[78] Niang, M. *et al.* STEVOR Is a Plasmodium falciparum Erythrocyte Binding Protein that Mediates Merozoite Invasion and Rosetting. *Cell Host & Microbe* **16**, 81–93 (2014).

[79] Winter, G. *et al.* SURFIN is a polymorphic antigen expressed on Plasmodium falciparum merozoites and infected erythrocytes. *Journal of Experimental Medicine* **201**, 1853–1863 (2005).

[80] Bachmann, A. *et al.* A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in Plasmodium falciparum-infected erythrocytes. *Malaria Journal* **14**, 274 (2015).

[81] Al-Khedery, B., Barnwell, J. W. & Galinski, M. R. Antigenic Variation in Malaria. *Molecular Cell* **3**, 131–141 (1999).

[82] Frech, C. & Chen, N. Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics* **14**, 427 (2013).

[83] Korir, C. C. & Galinski, M. R. Proteomic studies of Plasmodium knowlesi SICA variant antigens demonstrate their relationship with P. falciparum EMP1. *Infection, Genetics and Evolution* **6**, 75–79 (2006).

[84] del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. *Nature* **410**, 839–842 (2001).

[85] Janssen, C. S., Barrett, M. P., Turner, C. M. R. & Phillips, R. S. A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 431–436 (2002).

[86] Janssen, C. S. Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Research* **32**, 5712–5720 (2004).

[87] Ansari, H. R. *et al.* Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmod-

ium malariae and with other Plasmodium species. *International Journal for Parasitology* **46**, 685–696 (2016).

[88] Brugat, T. *et al.* Antibody-independent mechanisms regulate the establishment of chronic Plasmodium infection. *Nature Microbiology* **2**, 16276 (2017).

[89] Lin, J.-w. *et al.* Genomic and transcriptomic comparisons of closely related malaria parasites differing in virulence and sequestration pattern. *Wellcome Open Research* **3**, 142 (2018).

[90] Bernabeu, M. *et al.* Functional analysis of Plasmodium vivax VIR proteins reveals different subcellular localizations and cytoadherence to the ICAM-1 endothelial receptor. *Cellular Microbiology* (2012).

[91] Kyes, S., Horrocks, P. & Newbold, C. Antigenic Variation at the Infected Red Cell Surface in Malaria. *Annual Review of Microbiology* **55**, 673–707 (2001).

[92] Matthews, K. M., Pitman, E. L. & Koning-Ward, T. F. Illuminating how malaria parasites export proteins into host erythrocytes. *Cellular Microbiology* **21**, e13009 (2019).

[93] Kyes, S., Pinches, R. & Newbold, C. A simple RNA analysis method shows var and rif multigene family expression patterns in Plasmodium falciparum. *Molecular and Biochemical Parasitology* **105**, 311–315 (2000).

[94] Kaviratne, M., Khan, S. M., Jarra, W. & Preiser, P. R. Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells. *Eukaryotic Cell* (2002).

[95] Dzikowski, R., Frank, M. & Deitsch, K. Mutually Exclusive Expression of Virulence Genes by Malaria Parasites Is Regulated Independently of Antigen Production. *PLoS Pathogens* **2**, e22 (2006).

[96] Cabral, F. J. & Wunderlich, G. Transcriptional memory and switching in the Plasmodium falciparum rif gene family. *Molecular and Biochemical Parasitology* **168**, 186–190 (2009).

[97] Blythe, J. E. *et al.* Plasmodium falciparum STEVOR Proteins Are Highly Expressed in Patient Isolates and Located in the Surface Membranes of Infected Red Blood Cells and the Apical Tips of Merozoites. *Infection and Immunity* **76**, 3329–3336 (2008).

[98] Bhattacharjee, S., Stahelin, R. V., Speicher, K. D., Speicher, D. W. & Haldar, K. Endoplasmic Reticulum PI(3)P Lipid Binding Targets Malaria Proteins to the Host Cell. *Cell* **148**, 201–212 (2012).

[99] Russo, I. *et al.* Plasmepsin V licenses Plasmodium proteins for export into the host erythrocyte. *Nature* **463**, 632–636 (2010).

[100] Chang, H. H. *et al.* N-terminal processing of proteins exported by malaria parasites. *Molecular and Biochemical Parasitology* **160**, 107–115 (2008).

[101] Kaur, J. & Hora, R. '2TM proteins': an antigenically diverse superfamily with variable functions and export pathways. *PeerJ* **6**, e4757 (2018).

[102] Grüring, C. *et al.* Uncovering Common Principles in Protein Export of Malaria Parasites. *Cell Host & Microbe* **12**, 717–729 (2012).

[103] de Koning-Ward, T. F. *et al.* A newly discovered protein export machine in malaria parasites. *Nature* **459**, 945–949 (2009).

[104] Gehde, N. *et al.* Protein unfolding is an essential requirement for transport across the parasitophorous vacuolar membrane of Plasmodium falciparum. *Molecular Microbiology* **71**, 613–628 (2009).

[105] Bullen, H. E. *et al.* Biosynthesis, Localization, and Macromolecular Arrangement of the Plasmodium falciparum Translocon of Exported Proteins (PTEX). *Journal of Biological Chemistry* **287**, 7871–7884 (2012).

[106] Elsworth, B. *et al.* Proteomic analysis reveals novel proteins associated with the Plasmodium protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component, PTEX150. *Cellular Microbiology* **18**, 1551–1569 (2016).

[107] Peng, M., Cascio, D. & Egea, P. F. Crystal structure and solution characterization of the thioredoxin-2 from Plasmodium falciparum, a constituent of an essential parasitic protein export complex. *Biochemical and Biophysical Research Communications* **456**, 403–409 (2015).

[108] Chisholm, S. A. *et al.* The malaria PTEX component PTEX88 interacts most closely with HSP101 at the host–parasite interface. *The FEBS Journal* **285**, 2037–2055 (2018).

[109] Batinovic, S. *et al.* An exported protein-interacting complex involved in the trafficking of virulence determinants in Plasmodium-infected erythrocytes. *Nature Communications* **8**, 16044 (2017).

[110] Knuepfer, E., Rug, M., Klonis, N., Tilley, L. & Cowman, A. F. Trafficking of the major virulence factor to the surface of transfected P falciparum–infected erythrocytes. *Blood* **105**, 4078–4087 (2005).

[111] Bhattacharjee, S., van Ooij, C., Balu, B., Adams, J. H. & Haldar, K. Maurer's clefts of Plasmodium falciparum are secretory organelles that concentrate virulence protein reporters for delivery to the host erythrocyte. *Blood* **111**, 2418–2426 (2008).

[112] Tilley, L., Sougrat, R., Lithgow, T. & Hanssen, E. The Twists and Turns of Maurer's Cleft Trafficking in P. falciparum-Infected Erythrocytes. *Traffic* **9**, 187–197 (2007).

[113] Vincensini, L. *et al.* Proteomic Analysis Identifies Novel Proteins of the Maurer's Clefts, a Secretory Compartment Delivering Plasmodium falciparum Proteins to the Surface of Its Host Cell. *Molecular & Cellular Proteomics* **4**, 582–593 (2005).

[114] Wickert, H. Evidence for trafficking of PfEMP1 to the surface of -infected erythrocytes via a complex membrane network. *European Journal of Cell Biology* **82**, 271–284 (2003).

[115] Otto, T. D. *et al.* Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Research* **3**, 52 (2018).

[116] Kamaliddin, C. *et al.* From genomic to LC-MS/MS evidence: Analysis of PfEMP1 in Benin malaria cases. *PLOS ONE* **14**, e0218012 (2019).

[117] Templeton, T. J. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. *Molecular and Biochemical Parasitology* **166**, 109–116 (2009).

[118] Mayer, C., Slater, L., Erat, M. C., Konrat, R. & Vakonakis, I. Structural Analysis of the Plasmodium falciparum Erythrocyte Membrane Protein 1 (Pf EMP1) Intracellular Domain Reveals a Conserved Interaction Epitope. *Journal of Biological Chemistry* **287**, 7182–7189 (2012).

[119] Smith, J. D. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. *Molecular and Biochemical Parasitology* **195**, 82–87 (2014).

[120] Joergensen, L. *et al.* Surface Co-Expression of Two Different PfEMP1 Antigens on Single Plasmodium falciparum-Infected Erythrocytes Facilitates Binding to ICAM1 and PECAM1. *PLoS Pathogens* **6**, e1001083 (2010).

[121] van Schravendijk, M. *et al.* Characterization and localization of Plasmodium falciparum surface antigens on infected erythrocytes from west African patients. *Blood* **78**, 226–236 (1991).

[122] Leech, J. H., Barnwell, J. W., Aikawa, M., Miller, L. H. & Howard, R. J. Plasmodium falciparum malaria: association of knobs on the surface of infected erythrocytes with a histidine-rich protein and the erythrocyte skeleton. *Journal of Cell Biology* **98**, 1256–1264 (1984).

[123] Crabb, B. S. *et al.* Targeted Gene Disruption Shows That Knobs Enable Malaria-Infected Red Cells to Cytoadhere under Physiological Shear Stress. *Cell* **89**, 287–296 (1997).

[124] Oberli, A. *et al.* A Plasmodium falciparum PHIST protein binds the virulence factor PfEMP1 and comigrates to knobs on the host cell surface. *The FASEB Journal* **28**, 4420–4433 (2014).

[125] Parish, L. A., Mai, D. W., Jones, M. L., Kitson, E. L. & Rayner, J. C. A member of the Plasmodium falciparum PHIST family binds to the erythrocyte cytoskeleton component band 4.1. *Malaria Journal* **12**, 160 (2013).

[126] 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* **97**, 1766–1771 (2000).

[127] Figenschau, S. L. *et al.* ICAM1 expression is induced by proinflammatory cytokines and associated with TLS formation in aggressive breast cancer subtypes. *Scientific Reports* **8**, 11720 (2018).

[128] Mo, M. *et al.* The C-Terminal Segment of the Cysteine-Rich Interdomain of Plasmodium falciparum Erythrocyte Membrane Protein 1 Determines CD36 Binding and Elicits Antibodies That Inhibit Adhesion of Parasite-Infected Erythrocytes. *Infection and Immunity* **76**, 1837–1847 (2008).

[129] Andersen, M., Lenhard, B., Whatling, C., Eriksson, P. & Odeberg, J. Alternative promoter usage of the membrane glycoprotein CD36. *BMC molecular biology* **7**, 8 (2006).

[130] Turner, L. *et al.* Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* **498**, 502–505 (2013).

[131] Laszik, Z., Mitro, A., Taylor, F. B., Ferrell, G. & Esmon, C. T. Human Protein C Receptor Is Present Primarily on Endothelium of Large Blood Vessels. *Circulation* **96**, 3633–3640 (1997).

[132] 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).

[133] Ayres Pereira, M. *et al.* Placental Sequestration of Plasmodium falciparum Malaria Parasites Is Mediated by the Interaction Between VAR2CSA and Chondroitin Sulfate A on Syndecan-1. *PLOS Pathogens* **12**, e1005831 (2016).

[134] Lennartz, F., Smith, C., Craig, A. G. & Higgins, M. K. Structural insights into diverse modes of ICAM-1 binding by Plasmodium falciparum -infected erythrocytes. *Proceedings of the National Academy of Sciences* **116**, 20124–20134 (2019).

[135] Lau, C. K. *et al.* Structural Conservation Despite Huge Sequence Diversity Allows EPCR Binding by the PfEMP1 Family Implicated in Severe Childhood Malaria. *Cell Host & Microbe* **17**, 118–129 (2015).

[136] Hsieh, F.-L. *et al.* The structural basis for CD36 binding by the malaria parasite. *Nature Communications* **7**, 12837 (2016).

[137] Ochola, L. B. *et al.* Specific Receptor Usage in Plasmodium falciparum Cytoadherence Is Associated with Disease Outcome. *PLoS ONE* **6**, e14741 (2011).

[138] Wichers, J. S. *et al.* Dissecting the Gene Expression, Localization, Membrane Topology, and Function of the Plasmodium falciparum STEVOR Protein Family. *mBio* **10** (2019).

[139] Cheng, Q. *et al.* stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Molecular and Biochemical Parasitology* **97**, 161–176 (1998).

[140] Sanyal, S. *et al.* Plasmodium falciparum STEVOR proteins impact erythrocyte mechanical properties. *Blood* **119**, e1–e8 (2012).

[141] Nash, G. B., O'Brien, E., Gordon-Smith, E. C. & Dormandy, J. A. Abnormalities in the mechanical properties of red blood cells caused by Plasmodium falciparum. *Blood* **74**, 855–61 (1989).

[142] Khattab, A. & Meri, S. Exposure of the Plasmodium falciparum clonally variant STEVOR proteins on the merozoite surface. *Malaria Journal* **10**, 58 (2011).

[143] Wahlgren, M., Goel, S. & Akhouri, R. R. Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. *Nature Reviews Microbiology* **15**, 479–491 (2017).

[144] Fernandez, V., Hommel, M., Chen, Q., Hagblom, P. & Wahlgren, M. Small, Clonally Variant Antigens Expressed on the Surface of the Plasmodium falciparum–Infected Erythrocyte Are Encoded by the rif Gene Family and Are the Target of Human Immune Responses. *Journal of Experimental Medicine* **190**, 1393–1404 (1999).

[145] Lavazec, C., Sanyal, S. & Templeton, T. J. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Nucleic Acids Research* **34**, 6696–6707 (2006).

[146] Joannin, N., Abhiman, S., Sonnhammer, E. L. & Wahlgren, M. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics* **9**, 19 (2008).

[147] Andersson, A. *et al.* Membrane integration and topology of RIFIN and STEVOR proteins of the Plasmodium falciparum parasite. *The FEBS Journal* **287**, 2744–2762 (2020).

[148] Boddey, J. A. *et al.* Role of Plasmepsin V in Export of Diverse Protein Families from the Plasmodium falciparum Exportome. *Traffic* **14**, 532–550 (2013).

[149] Petter, M. *et al.* Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Molecular and Biochemical Parasitology* **156**, 51–61 (2007).

[150] Bachmann, A. *et al.* Temporal Expression and Localization Patterns of Variant Surface Antigens in Clinical Plasmodium falciparum Isolates during Erythrocyte Schizogony. *PLoS ONE* **7**, e49540 (2012).

[151] Mwakalinga, S. B. *et al.* Expression of a type B RIFIN in Plasmodium falciparum merozoites and gametes. *Malaria Journal* **11**, 429 (2012).

[152] Zhou, A. E. *et al.* Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. *mSphere* **4** (2019).

[153] Abdel-Latif, M. S., Dietz, K., Issifou, S., Kremsner, P. G. & Klinkert, M.-Q. Antibodies to Plasmodium falciparum Rifin Proteins Are Associated with Rapid Parasite Clearance and Asymptomatic Infections. *Infection and Immunity* **71**, 6229–6233 (2003).

[154] Abdel-Latif, M. S., Köhler, C., Luty, A. J. F., Cabrera, G. & Kremsner, P. G. Antibodies to rifin: a component of naturally acquired responses to Plasmodium falciparum variant surface antigens on infected erythrocytes. *The American Journal of Tropical Medicine and Hygiene* **71**, 179–186 (2004).

[155] Arora, G. *et al.* NK cells inhibit plasmodium falciparum growth in red blood cells via antibody-dependent cellular cytotoxicity. *eLife* **7** (2018).

[156] Tan, J. *et al.* A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* **529**, 105–109 (2016).

[157] Hsieh, F.-L. & Higgins, M. K. The structure of a LAIR1-containing human antibody reveals a novel mechanism of antigen recognition. *eLife* **6** (2017).

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

[159] Rutledge, G. G. *et al.* Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).

[160] Fonager, J. *et al.* Transcription and alternative splicing in the yir multigene family of the malaria parasite Plasmodium y. yoelii: Identification of motifs suggesting epigenetic and post-transcriptional control of RNA expression. *Molecular and Biochemical Parasitology* **156**, 1–11 (2007).

[161] Cunningham, D. A. *et al.* Host immunity modulates transcriptional changes in a multigene family (yir) of rodent malaria. *Molecular Microbiology* **58**, 636–647 (2005).

[162] Pasini, E. M. *et al.* Proteomic and Genetic Analyses Demonstrate that Plasmodium berghei Blood Stages Export a Large and Diverse Repertoire of Proteins. *Molecular & Cellular Proteomics* **12**, 426–448 (2013).

[163] Di Girolamo, F. *et al.* Plasmodium lipid rafts contain proteins implicated in vesicular trafficking and signalling as well as members of the PIR superfamily, potentially implicated in host immune system interactions. *PROTEOMICS* **8**, 2500–2513 (2008).

[164] Yam, X. Y. *et al.* Characterization of the Plasmodium Interspersed Repeats (PIR) proteins of Plasmodium chabaudi indicates functional diversity. *Scientific Reports* **6**, 23449 (2016).

[165] Cunningham, D., Lawton, J., Jarra, W., Preiser, P. & Langhorne, J. The pir

multigene family of Plasmodium: Antigenic variation and beyond. *Molecular and Biochemical Parasitology* **170**, 65–73 (2010).

[166] Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature* **455**, 757–763 (2008).

[167] Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).

[168] Spence, P. J. *et al.* Vector transmission regulates immune control of Plasmodium virulence. *Nature* (2013).

[169] Yoeli, M., Hargreaves, B., Carter, R. & Walliker, D. Sudden increase in virulence in a strain of Plasmodium berghei yoelii. *Annals of Tropical Medicine & Parasitology* **69**, 173–178 (1975).

[170] Otto, T. D. *et al.* A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biology* **12**, 86 (2014).

[171] Carvalho, B. O. *et al.* On the cytoadhesion of plasmodium vivax-infected erythrocytes. *Journal of Infectious Diseases* (2010).

[172] Healy, J. I. & Goodnow, C. C. POSITIVE VERSUS NEGATIVE SIGNALING BY LYMPHOCYTE ANTIGEN RECEPTORS. *Annual Review of Immunology* **16**, 645–670 (1998).

[173] Ravetch, J. V. Immune Inhibitory Receptors. *Science* **290**, 84–89 (2000).

[174] Brown, D., Trowsdale, J. & Allen, R. The LILR family: modulators of innate and adaptive immune pathways in health and disease. *Tissue Antigens* **64**, 215–225 (2004).

[175] Zhao, J. *et al.* The MHC class I-LILRB1 signalling axis as a promising target in cancer therapy. *Scandinavian Journal of Immunology* **90** (2019).

[176] Yawata, M., Yawata, N., Abi-Rached, L. & Parham, P. Variation within the human killer cell immunoglobulin-like receptor (KIR) gene family. *Critical reviews in immunology* **22**, 463–82 (2002).

[177] Love, P. E. & Hayes, S. M. ITAM-mediated Signaling by the T-Cell Antigen Receptor. *Cold Spring Harbor Perspectives in Biology* **2**, a002485–a002485 (2010).

[178] Smith-Garvin, J. E., Koretzky, G. A. & Jordan, M. S. T Cell Activation. *Annual Review of Immunology* **27**, 591–619 (2009).

[179] Long, E. O. Negative signaling by inhibitory receptors: the NK cell paradigm. *Immunological Reviews* **224**, 70–84 (2008).

[180] Kang, X. *et al.* Inhibitory leukocyte immunoglobulin-like receptors: Immune checkpoint proteins and tumor sustaining factors. *Cell Cycle* **15**, 25–40 (2016).

[181] Kumar, S. Natural killer cell cytotoxicity and its regulation by inhibitory receptors. *Immunology* **154**, 383–393 (2018).

[182] Hirayasu, K. & Arase, H. Functional and genetic diversity of leukocyte immunoglobulin-like receptor and implication for disease associations. *Journal of Human Genetics* **60**, 703–708 (2015).

[183] Hudson, L. E. & Allen, R. L. Leukocyte Ig-Like Receptors – A Model for MHC Class I Disease Associations. *Frontiers in Immunology* **7** (2016).

[184] Cao, W. *et al.* Regulation of TLR7/9 responses in plasmacytoid dendritic cells by BST2 and ILT7 receptor interaction. *The Journal of Experimental Medicine* **206**, 1603–1614 (2009).

[185] Zheng, J. *et al.* Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development. *Nature* **485**, 656–660 (2012).

[186] Arnold, V. *et al.* S100A9 protein is a novel ligand for the CD85j receptor and its interaction is implicated in the control of HIV-1 replication by NK cells. *Retrovirology* **10**, 122 (2013).

[187] Atwal, J. K. *et al.* PirB is a Functional Receptor for Myelin Inhibitors of Axonal Regeneration. *Science* **322**, 967–970 (2008).

[188] Kim, T. *et al.* Human LilrB2 Is a -Amyloid Receptor and Its Murine Homolog PirB Regulates Synaptic Plasticity in an Alzheimer's Model. *Science* **341**, 1399–1404 (2013).

[189] Zhang, J. *et al.* Leukocyte immunoglobulin-like receptors in human diseases: an overview of their distribution, function, and potential application for immunotherapies. *Journal of Leukocyte Biology* **102**, 351–360 (2017).

[190] Jones, D. C. *et al.* HLA Class I Allelic Sequence and Conformation Regulate Leukocyte Ig-Like Receptor Binding. *The Journal of Immunology* **186**, 2990–2997 (2011).

[191] Wang, Q. *et al.* Structures of the four Ig-like domain LILRB2 and the four-domain LILRB1 and HLA-G1 complex. *Cellular and Molecular Immunology* (2019).

[192] Willcox, B. E., Thomas, L. M. & Bjorkman, P. J. Crystal structure of HLA-A2 bound to LIR-1, a host and viral major histocompatibility complex receptor. *Nature Immunology* **4**, 913–919 (2003).

[193] Dulberger, C. L. *et al.* Human Leukocyte Antigen F Presents Peptides and Regulates Immunity through Interactions with NK Cell Receptors. *Immunity* **46**, 1018–1029.e7 (2017).

[194] Mohammed, F., Stones, D. H. & Willcox, B. E. Application of the immunoregulatory receptor LILRB1 as a crystallisation chaperone for human class I MHC complexes. *Journal of Immunological Methods* **464**, 47–56 (2019).

[195] Cosman, D. *et al.* A Novel Immunoglobulin Superfamily Receptor for Cellular and Viral MHC Class I Molecules. *Immunity* **7**, 273–282 (1997).

[196] Prod'homme, V. *et al.* The Human Cytomegalovirus MHC Class I Homolog UL18 Inhibits LIR-1 + but Activates LIR-1 NK Cells. *The Journal of Immunology* **178**, 4473–4481 (2007).

[197] Yang, Z. & Bjorkman, P. J. Structure of UL18, a peptide-binding viral MHC mimic, bound to a host inhibitory receptor. *Proceedings of the National Academy of Sciences* **105**, 10095–10100 (2008).

[198] Chan, K. R. *et al.* Leukocyte immunoglobulin-like receptor B1 is critical for antibody-dependent dengue. *Proceedings of the National Academy of Sciences* **111**, 2722–2727 (2014).

[199] Van der Vuurst de Vries, A.-R., Clevers, H., Logtenberg, T. & Meyaard, L. Leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) is differentially expressed during human B cell differentiation and inhibits B cell receptor-mediated signaling. *European Journal of Immunology* **29**, 3160–3167 (1999).

[200] Meyaard, L. *et al.* LAIR-1, a Novel Inhibitory Receptor Expressed on Human Mononuclear Leukocytes. *Immunity* **7**, 283–290 (1997).

[201] Brondijk, T. H. C. *et al.* Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood* **115**, 1364–1373 (2010).

[202] Lebbink, R. J. *et al.* Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *Journal of Experimental Medicine* **203**, 1419–1425 (2006).

[203] Rowley, A. T. *et al.* Effects of Surface-Bound Collagen-Mimetic Peptides on Macrophage Uptake and Immunomodulation. *Frontiers in Bioengineering and Biotechnology* **8** (2020).

[204] Pende, D. *et al.* Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation. *Frontiers in Immunology* **10** (2019).

[205] Parham, P. Immunogenetics of killer cell immunoglobulin-like receptors. *Molecular Immunology* **42**, 459–462 (2005).

[206] Hsu, K. C. *et al.* Killer Ig-Like Receptor Haplotype Analysis by Gene Content: Evidence for Genomic Diversity with a Minimum of Six Basic Framework Haplotypes, Each with Multiple Subsets. *The Journal of Immunology* **169**, 5118–5129 (2002).

[207] Lanier, L. L., Corliss, B. C., Wu, J., Leong, C. & Phillips, J. H. Immunoreceptor DAP12 bearing a tyrosine-based activation motif is involved in activating NK cells. *Nature* **391**, 703–707 (1998).

[208] Moretta, A. *et al.* Receptors for HLA class-I molecules in human natural killer cells. *Annual Review of Immunology* **14**, 619–648 (1996).

[209] Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G. & Sun, P. D. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* **405**, 537–543 (2000).

[210] Fan, Q. R., Long, E. O. & Wiley, D. C. Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1–HLA-Cw4 complex. *Nature Immunology* **2**, 452–460 (2001).

[211] Vivian, J. P. *et al.* Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature* **479**, 401–405 (2011).

[212] William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 60–89 (1990).

[213] Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D: Biological Crystallography* **62**, 1243–1250 (2006).

[214] Walter, T. S. *et al.* Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **14**, 1617–1622 (2006).

[215] Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876–2890 (2007).

[216] Nguyen, H., Park, J., Kang, S. & Kim, M. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. *Sensors* **15**, 10481–10510 (2015).

[217] Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package. *Acta Crystallographica Section D: Structural Biology* **74**, 85–97 (2018).

[218] Kabsch, W. *et al.* XDS. *Acta Crystallographica Section D Biological Crystallography* **66**, 125–132 (2010).

[219] Winn, M. D. *et al.* Overview of the CCP4 suite and current developments (2011).

[220] Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography* (2010).

[221] Matthews, B. Solvent content of protein crystals. *Journal of Molecular Biology* **33**, 491–497 (1968).

[222] McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674 (2007).

[223] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486–501 (2010).

[224] Basu, S. *et al.* Long-wavelength native-SAD phasing: opportunities and challenges. *IUCrJ* **6**, 373–386 (2019).

[225] Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallographica Section D: Biological Crystallography* **58**, 1772–1779 (2002).

[226] Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography* **62**, 1002–1011 (2006).

[227] Bricogne, G. *et al.* BUSTER version 1.10.0. *Cambridge, United Kingdom: Global Phasing Ltd.* (2011).

[228] Aurrecoechea, C. *et al.* PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Research* **37** (2009).

[229] Fougère, A. *et al.* Variant Exported Blood-Stage Proteins Encoded by Plasmodium Multigene Families Are Expressed in Liver Stages Where They Are Exported into the Parasitophorous Vacuole. *PLoS Pathogens* **12**, e1005917 (2016).

[230] Böhme, U., Otto, T. D., Sanders, M., Newbold, C. I. & Berriman, M. Progression of the canonical reference malaria parasite genome from 2002–2019. *Wellcome Open Research* **4**, 58 (2019).

[231] Auburn, S. *et al.* A new Plasmodium vivax reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. *Wellcome Open Research* **1**, 4 (2016).

[232] Otto, T. D. *et al.* Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nature Communications* **5**, 4754 (2014).

[233] The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506–D515 (2019).

[234] Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786 (2011).

[235] 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**, 567–580 (2001).

[236] Jones, D. T. & Cozzetto, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).

[237] Mizianty, M. J., Peng, Z. & Kurgan, L. MFDp2. *Intrinsically Disordered Proteins* **1**, e24428 (2013).

[238] Yachdav, G. *et al.* PredictProtein - An open resource for online prediction of protein structural and functional features. *Nucleic Acids Research* **42** (2014).

[239] Walsh, I., Martin, A. J., Di domenico, T. & Tosatto, S. C. Espritz: Accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).

[240] Wang, S., Ma, J. & Xu, J. AUCpreD: Proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. In *Bioinformatics*, vol. 32, i672–i679 (2016).

[241] Hanson, J., Paliwal, K. K., Litfin, T. & Zhou, Y. SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning. *Genomics, Proteomics and Bioinformatics* (2020).

[242] Ishida, T. & Kinoshita, K. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research* **35** (2007).

[243] Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).

[244] Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019).

[245] Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

[246] Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Research* **14**, 1188–1190 (2004).

[247] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).

[248] Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549 (2018).

[249] Stecher, G., Tamura, K. & Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular biology and evolution* **37**, 1237–1239 (2020).

[250] Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**, 691–699 (2001).

[251] Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* (1987).

[252] Gascuel, O. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14**, 685–695 (1997).

[253] Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282 (1992).

[254] Rambaut, A. FigTree v. 1.4.4 (2018).

[255] Pieper, K. *et al.* Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. *Nature* **548**, 597–601 (2017).

[256] Boddey, J. A. *et al.* An aspartyl protease directs malaria effector proteins to the host cell. *Nature* **463**, 627–631 (2010).

[257] Müller-Sienerth, N. *et al.* A panel of recombinant proteins from human-infective Plasmodium species for serological surveillance. *Malaria Journal* **19**, 31 (2020).

[258] Srivastava, A., Durocher, Y. & Gamain, B. Expressing Full-Length Functional PfEMP1 Proteins in the HEK293 Expression System. 307–319 (2012).

[259] Zhao, X. *et al.* Transient expression of Fc-fused human glycoprotein 130 in Expi293F suspension cells. *Protein Expression and Purification* **124**, 41–47 (2016).

[260] Birkholtz, L. M. *et al.* Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria Journal* **7**, 197 (2008).

[261] Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* **11**, 753 (2012).

[262] Ohmuro-Matsuyama, Y. & Yamaji, H. Modifications of a signal sequence for antibody secretion from insect cells. *Cytotechnology* **70**, 891–898 (2018).

[263] Yamaguchi, H. & Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4**, 235–251 (2014).

[264] Bultrini, E. *et al.* Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction. *BMC Genomics* **10**, 445 (2009).

[265] Nam, G. *et al.* Crystal structures of the two membrane-proximal Ig-like domains (D3D4) of LILRB1/B2: alternative models for their involvement in peptide-HLA binding. *Protein & Cell* **4**, 761–770 (2013).

[266] Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the web: A case study using the phyre server. *Nature Protocols* **4**, 363–373 (2009).

[267] Chapman, T. L., Heikema, A. P., West, A. P. & Bjorkman, P. J. Crystal structure and ligand binding properties of the D1D2 region of the inhibitory receptor LIR-1(ILT2). *Immunity* **13**, 727–736 (2000).

[268] Ohtsuka, M. *et al.* NFAM1, an immunoreceptor tyrosine-based activation motif-bearing molecule that regulates B cell development and signaling. *Proceedings of the National Academy of Sciences* **101**, 8126–8131 (2004).

[269] Kogure, A., Shiratori, I., Wang, J., Lanier, L. L. & Arase, H. PANP is a novel O-glycosylated PILRα ligand expressed in neural tissues. *Biochemical and Biophysical Research Communications* **405**, 428–433 (2011).

[270] Burrack, K. S., Hart, G. T. & Hamilton, S. E. Contributions of natural killer cells to the immune response against Plasmodium. *Malaria Journal* **18**, 321 (2019).

[271] Vivier, E., Tomasello, E., Baratin, M., Walzer, T. & Ugolini, S. Functions of natural killer cells. *Nature Immunology* **9**, 503–510 (2008).

[272] Alanine, D. G. *et al.* Human Antibodies that Slow Erythrocyte Invasion Potentiate Malaria-Neutralizing Antibodies. *Cell* **178**, 216–228.e21 (2019).

[273] Barber, D. F., Faure, M. & Long, E. O. LFA-1 Contributes an Early Signal for NK Cell Cytotoxicity. *The Journal of Immunology* **173**, 3653–3659 (2004).

[274] Voskoboinik, I., Whisstock, J. C. & Trapani, J. A. Perforin and granzymes: Function, dysfunction and human pathology. *Nature Reviews Immunology* **15**, 388–400 (2015).

[275] Grakoui, A. *et al.* The immunological synapse: A molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).

[276] Dustin, M. L. & Baldari, C. T. The immune synapse: Past, present, and future. In *Methods in Molecular Biology*, vol. 1584, 1–5 (2017).

[277] Spicer, B. A., Conroy, P. J., Law, R. H., Voskoboinik, I. & Whisstock, J. C. Perforin—A key (shaped) weapon in the immunological arsenal. *Seminars in Cell & Developmental Biology* **72**, 117–123 (2017).

[278] Chiewsilp, D., Chiewsilp, P., Boonpucknavig, S. & Rugsakul, T. Elevation of beta2-microglobulin in malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **82**, 688–689 (1988).

[279] Bergfors, T. Seeds to crystals. *Journal of Structural Biology* **142**, 66–76 (2003).

[280] Monaco, S. *et al.* Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *Journal of Applied Crystallography* **46**, 804–810 (2013).

[281] Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P. & Nikolov, D. B. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Science* **15**, 2008–2013 (2006).

[282] Rose, J. P., Wang, B.-C. & Weiss, M. S. Native SAD is maturing. *IUCrJ* **2**, 431–440 (2015).

[283] Holm, L. DALI and the persistence of protein shape. *Protein Science* **29**, 128–140 (2020).

[284] Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780–2781 (2008).