University of Oxford

Doctorate of Philosophy

Structure of the essential malaria invasion protein RH5 in complex with its erythrocyte receptor and inhibitory antibodies

Author:
Katherine Wright

Supervisor:
Matthew Higgins



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

in the

Department of Biochemistry

Christ Church College

December 2014

'But time is short, and science is infinite..."

Thomas Hardy (*Two on a Tower*)

UNIVERSITY OF OXFORD

Abstract

Department of Biochemistry

Doctor of Philosophy

Structure of the essential malaria invasion protein RH5 in complex with its erythrocyte receptor and inhibitory antibodies

by Katherine Wright

Invasion of host erythrocytes is an essential stage in the life cycle of *Plasmodium* parasites and in development of the pathology of malaria. The stages of erythrocyte invasion, including initial contact, apical reorientation, junction formation, and active invagination, are directed by the coordinated release of specialised apical organelles and their parasite protein contents. Among these proteins, and central to invasion by all species, are two parasite protein families, the reticulocyte-binding protein homologue (RH) and the erythrocyte-binding like (EBL) proteins, that mediate host-parasite interactions. RH5 from *Plasmodium falciparum* (PfRH5) is the only member of either family demonstrated to be necessary for erythrocyte invasion in all tested strains, through its interaction with the erythrocyte surface protein basigin. Indeed, antibodies targeting either PfRH5 or basigin can block parasite invasion with high efficiency *in vitro*, making PfRH5 an excellent candidate for a vaccine to protect against the most deadly form of malaria. Here I present crystal structures of PfRH5 in complex with basigin and with two distinct inhibitory antibodies. This is the first structure of any RH protein, revealing a novel fold in which two three-helical bundles come together to form a kite-like architecture. The two immunoglobulin domains of basigin and the inhibitory antibodies bind to one tip of the kite. These findings provide the first structural insights into erythrocyte binding by the *Plasmodium* RH protein family and identify novel inhibitory epitopes to guide the design of a new generation of vaccines against the blood-stage parasite.

Acknowledgements

I am deeply appreciative of the network of colleagues, collaborators, friends and family who have helped and supported me over the past four years.

First and foremost, I would like thank my supervisor, Matt Higgins, for his mentorship, guidance, seemingly boundless patience, and wizardly crystal-harvesting skills. I am so grateful for the training and advice he has given me over the years. All of the expression trials, cryo-cooling marathons, and synchrotron trips were worth it for the moments when we watched the structure of PfRH5 emerge in all of its glory! Thank you for all of your help along the way, Matt.

This project would not have been possible without Simon Draper's lab. Kathryn Hjerrild's assistance with purifying PfRH5 has been invaluable to the project, and her friendship has been invaluable to me. Simon has been an honorary second supervisor throughout. Rebecca Brown and Jing Jin performed GIA and ELISA experiments, and Julie Furze and Sandy Douglas provided hybridomas. Equally critical to the project was S2 cell help from Anette Strøbæk, Stine Clemmensen, and Wian de Jongh at ExpreS^{2}ion Biotechnologies.

Other people I would like to thank include Edward Lowe and David Staunton (Biochemistry Department), for their help with crystallography and analytical ultracentrifugation, respectively; Tom Waksman, for his assistance with production and analysis of the basigin mutants; Gavin Wright (Sanger Institute), for providing me with several plasmids; Louise Bird (Oxford Protein Production Facility) for her help with high-throughput protein expression trials; Maria Yamout (The Scripps Research Institute) for helpful discussions on construct design; Adam Round (ESRF) for his assistance with small angle X-ray scattering data collection; and beamline scientists Ralf Flaig and Jose Brandao-Neto (Diamond Light Source). I would also like to acknowledge Carol Robinson, for her help with non-dissociative mass spectrometry; Stuart Ferguson, for providing me with opportunities to teach and remain a part of Teddy Hall; and Simon Newstead, for his support as my Christ Church advisor. I am grateful for generous funding from the Wellcome Trust, Clarendon Fund, and Christ Church College.

I must of course thank my Higgins labmates—past and present—for support and camaraderie in and out of the lab, for excellent cakes, and for team spirit in our football matches against the Schnell lab. In particular, Frank and Livia have been the most helpful and kind postdocs (and friends) a graduate student could wish for! Thanks to Clinton and Harriet, my dear friends and my 'lobsters'; Jon, lab dad and cell whisperer; and Hong Kin.

My friends have helped to fill the past four years with fun experiences and happy memories: in particular, Tim, Callum, Serena, Jo, Ed, Kat, Kate, Emma, Narin, Marie and of course Jacob. I would also like to thank the Williamses and my sister-cousin Louise, for giving me a home away from home at Frankford Farmhouse.

Harry has lived through this D.Phil. with me—day by day, through the bad times and the good, from insect cells to unit cells—and has remained unbelievably patient, supportive, and encouraging throughout. I am tremendously grateful.

Thanks to my brother Nick, for reminding me what fun is when I forget. Most of all, thanks to my marvellous parents, for their unflagging support, and for instilling in me their love of science and of many other things. This thesis is dedicated to you two, with all of my love and gratitude.

Contents

Abstract ii

Acknowledgements iii

Contents iv

List of Figures ix

List of Tables xi

Abbreviations xii

1 Introduction 1

1.1 Malaria . . . 1

1.1.1 History and aetiology . . . 1

1.1.2 Global disease burden . . . 2

1.1.3 *Plasmodium* species . . . 2

1.1.4 Malaria genomes . . . 4

1.1.5 The *Plasmodium* lifecycle . . . 4

1.1.5.1 Exoerythrocytic schizogony . . . 4

1.1.5.2 Erythrocytic schizogony . . . 6

1.1.5.3 Gamogony . . . 9

1.1.5.4 Sporogony . . . 9

1.1.6 Disease manifestations and pathology . . . 10

1.1.6.1 Symptoms and pathology . . . 10

1.1.6.2 Host immune response and acquired immunity . . . 11

1.1.6.3 Genetic resistance . . . 12

1.1.7 Treatment and prevention . . . 12

1.1.7.1 Vector control . . . 13

1.1.7.2 Chemotherapy . . . 13

1.1.7.3 Vaccine candidates . . . 14

1.2 Erythrocyte invasion by *P. falciparum* . . . 16

1.2.1 Merozoite substructure . . . 17

1.2.2 Stages of erythrocyte invasion . . . 18

1.2.3 Malarial proteins involved in invasion . . . 22

1.2.3.1 Merozoite surface proteins . . . 22

1.2.3.2 Moving junction proteins and the actin-myosin motor . . . 23

1.2.3.3 Invasion ligands . . . 25
1.2.3.4 The EBL protein family . . . 28
1.2.4 Parasite immune evasion strategies during invasion . . . 29
1.3 The RH protein family . . . 31
1.3.1 Discovery . . . 31
1.3.2 General architecture . . . 32
1.3.3 RH proteins of *P. yoelii* . . . 33
1.3.4 RH proteins of *P. vivax* . . . 34
1.3.5 RH proteins of *P. falciparum* . . . 34
1.3.5.1 RH1 . . . 34
1.3.5.2 RH2a and RH2b . . . 35
1.3.5.3 RH4 . . . 36
1.4 *P. falciparum* RH5 . . . 37
1.4.1 Interaction with basigin . . . 38
1.4.2 Interaction with PfRIPR . . . 40
1.4.3 Immunogenicity and suitability as a vaccine candidate . . . 41
1.5 Structures of *Plasmodium* surface proteins . . . 43
1.6 Project aims . . . 47

2 Methods **48**
2.1 Bioinformatic analysis of PfRH5 . . . 48
2.2 Molecular biology and cloning . . . 48
2.2.1 Genes and constructs . . . 48
2.2.2 Cloning vectors . . . 49
2.2.3 Polymerase chain reaction (PCR) . . . 50
2.2.4 DNA agarose gel electrophoresis . . . 50
2.2.5 Restriction digestion, calf alkaline phosphatase treatment, and ligation . . . 50
2.2.6 Transformation, screening, and sequencing . . . 51
2.2.7 In-Fusion™ cloning . . . 51
2.2.8 RNA isolation and reverse transcription . . . 51
2.3 Protein expression and purification . . . 54
2.3.1 *E. coli* . . . 54
2.3.1.1 Expression strains . . . 54
2.3.1.2 Small scale expression trials . . . 54
2.3.1.3 SDS-PAGE and Western blotting . . . 55
2.3.1.4 Expression pipeline at the OPPF . . . 56
2.3.1.5 Large scale expression in *E. coli* . . . 56
2.3.1.6 Selenomethionine incorporation . . . 57
2.3.1.7 Soluble purification of 6-His-tagged proteins . . . 57
2.3.1.8 Tobacco etch protease (TEV) purification . . . 58
2.3.1.9 TEV cleavage . . . 58
2.3.1.10 Size exclusion chromatography . . . 59
2.3.1.11 Soluble purification of MBP-fusion protein . . . 59
2.3.1.12 Insoluble purification and refolding . . . 59
2.3.2 Human embryonic kidney (HEK) cells . . . 62
2.3.2.1 HEK cell maintenance . . . 62

2.3.2.2 Transient transfection of HEK cells . . . 62
2.3.3 Hybridomas . . . 62
2.3.3.1 Hybridoma growth . . . 62
2.3.3.2 Antibody purification . . . 63
2.3.3.3 Production of antibody fragments . . . 63
2.3.4 Baculovirus-infected insect cells . . . 63
2.3.4.1 Sf9 cell maintenance . . . 63
2.3.4.2 Baculovirus production . . . 64
2.3.4.3 Protein production and purification . . . 64
2.3.5 Schneider 2 cells . . . 64
2.3.5.1 Cell maintenance . . . 64
2.3.5.2 Stable transfection of S2 cells . . . 65
2.3.5.3 Protein expression in S2 cells . . . 65
2.3.5.4 Purification from S2 cells . . . 65
2.3.5.5 ConA purification . . . 66
2.3.5.6 Differential scanning fluorimetry . . . 66
2.3.5.7 Limited proteolysis . . . 67
2.3.5.8 Endoproteinase GluC cleavage . . . 67
2.3.5.9 Protein methylation . . . 67
2.4 Biophysical methods . . . 67
2.4.1 Surface plasmon resonance . . . 67
2.4.2 Analytical ultracentrifugation . . . 69
2.4.3 Native mass spectrometry . . . 70
2.4.4 Analytical gel filtration . . . 72
2.4.5 Small-angle X-ray scattering . . . 72
2.5 X-ray crystallography . . . 74
2.5.1 Crystallisation and crystal optimisation . . . 74
2.5.2 Data collection and initial processing . . . 76
2.5.3 Phase determination by molecular replacement . . . 78
2.5.4 Experimental phasing . . . 81
2.5.5 Model building and refinement . . . 82

3 Expression, Purification, and Functional Validation of PfRH5 85
3.1 Summary . . . 85
3.2 Bioinformatic survey of PfRH5 . . . 86
3.3 PfRH5 expression in *E. coli*, HEK cells, and baculovirus-infected insect cells . . . 89
3.3.1 Full-length and shortened full-length constructs . . . 90
3.3.2 Single-domain constructs . . . 93
3.3.3 Deletion constructs . . . 96
3.3.4 Co-expression and fusion of PfRH5 and basigin . . . 97
3.4 Schneider 2 insect cell expression and purification of PfRH5 . . . 99
3.4.1 Expression . . . 99
3.4.2 Purification . . . 100
3.4.3 Differential scanning fluorimetry . . . 100
3.4.4 Limited proteolysis and N-terminal sequencing . . . 102
3.4.5 Expression and purification of shortened constructs . . . 103

3.5 Validation of PfRH5 function and immunogenicity . . . 105
3.5.1 PfRH5 binding to basigin . . . 105
3.5.1.1 Expression and purification of basigin . . . 105
3.5.1.2 Analytical gel filtration . . . 106
3.5.1.3 Surface plasmon resonance . . . 107
3.5.2 Immunogenicity . . . 108
3.5.3 PfRH5 binding to monoclonal antibodies . . . 109
3.5.3.1 ELISA . . . 110
3.5.3.2 Production of monoclonal antibody fragments . . . 110
3.6 Discussion . . . 112

4 The Structure of PfRH5 and the RH Protein Fold 115
4.1 Summary . . . 115
4.2 PfRH5 preparation for crystallography . . . 115
4.2.1 Endoproteinase GluC cleavage . . . 116
4.2.2 Protein complex formation . . . 117
4.2.3 Lysine methylation . . . 117
4.3 PfRH5–9AD4 crystallisation, data collection, and data processing . . . 119
4.3.1 Crystallisation . . . 119
4.3.2 Data collection . . . 121
4.3.3 Data processing . . . 122
4.4 Solving and validating the PfRH5–9AD4 structure . . . 122
4.5 The structure of PfRH5 . . . 127
4.5.1 Fold architecture . . . 127
4.5.2 SAXS analysis . . . 129
4.6 The PfRH5 fold in other RH proteins . . . 133
4.6.1 Alignments and threading . . . 133
4.6.2 Expression of other RH proteins in S2 cells . . . 134
4.7 Discussion . . . 140

5 Structural Basis for Basigin Binding by PfRH5 143
5.1 Summary . . . 143
5.2 PfRH5–basigin crystallisation, data collection, and data processing . . . 143
5.2.1 Crystallisation . . . 143
5.2.2 Data collection . . . 145
5.2.3 Data processing . . . 146
5.2.4 Phasing attempts . . . 146
5.3 Solving and validating the PfRH5–basigin structure . . . 148
5.4 Analysis of the PfRH5–basigin structure . . . 151
5.4.1 SAXS analysis of flexibility in basigin binding . . . 151
5.5 Dissection of the basigin binding site . . . 154
5.5.1 A detailed tour of the binding site . . . 154
5.5.2 Naturally occurring polymorphisms and host tropism . . . 154
5.5.3 Mutagenesis of basigin . . . 156
5.6 A possible supramolecular complex . . . 161
5.7 Stoichiometry of the PfRH5–basigin complex . . . 162
5.7.1 Native mass spectrometry . . . 162

5.7.2 Analytical ultracentrifugation . . . 163
5.7.3 SAXS . . . 164
5.8 Discussion . . . 165

6 Structural Basis for PfRH5 Inhibition by Monoclonal Antibodies 168
6.1 Summary . . . 168
6.2 PfRH5–QA1 crystallisation, data collection, and data processing . . . 169
6.2.1 Crystallisation . . . 169
6.2.2 Data collection . . . 169
6.2.3 Data processing . . . 170
6.3 Solving and validating the PfRH5–QA1 structure . . . 171
6.4 The structure of PfRH5 bound to inhibitory antibodies . . . 173
6.4.1 PfRH5–QA1 . . . 174
6.4.2 PfRH5–9AD4 . . . 174
6.4.3 PfRH5–QA5 . . . 176
6.4.4 A comparison of the binding of three inhibitory antibodies . . . 177
6.5 Analysis of PfRH5 interactions with non-inhibitory antibodies . . . 180
6.6 Discussion . . . 183

7 Conclusions and Future Directions 185
7.1 The PfRH5 fold and an updated comparison of parasite surface proteins . 186
7.2 The PfRH5–basigin interaction and implications for erythrocyte invasion . 187
7.3 Implications for vaccine development . . . 190
7.4 Summary of the findings of this thesis . . . 191

A Primer Sequences 223

B Surface Plasmon Resonance Fit Information 229

C Publications Arising From The Thesis 232

List of Figures

1.1 Map of the population at risk of malaria . . . 2
1.2 Malaria lifecycle . . . 5
1.3 Merozoite ultrastructure . . . 18
1.4 Rhoptry development and ultrastructure . . . 19
1.5 Schematic depiction of erythrocyte invasion. . . . 20
1.6 EBL and RH ligands involved in invasion, and their erythrocyte receptors. 26
1.7 Schematic of EBL protein architecture. . . . 28
1.8 Schematic of PfRH protein architecture. . . . 32
1.9 Phylogenetic relationships among PvRBP and PfRH proteins. . . . 33
1.10 Structure of basigin. . . . 39
1.11 Architecture of PfRIPR. . . . 41
1.12 Model of the arrangement of PfRH5 epitopes recognised by inhibitory monoclonal antibodies. . . . 43
1.13 *Plasmodium* surface protein domain structures. . . . 45
1.14 EBA-175 dimerisation with sialyllactose in crystal structure. . . . 46

2.1 Maps of two commonly used expression plasmids. . . . 53
2.2 Model building and refinement pipeline. . . . 83

3.1 Predicted disorder and secondary structure of PfRH5. . . . 87
3.2 Sequence alignment of the PvRBP-2-like clade. . . . 88
3.3 Expression of full-length PfRH5 in three expression systems. . . . 91
3.4 An attempt to purify the PfRH5 N-terminal domain. . . . 94
3.5 Refolding trial of PfRH5 C-terminal domain. . . . 98
3.6 Co-expression trial of PfRH5 C-terminal domain with basigin. . . . 99
3.7 Purification of full-length PfRH5 from S2 insect cells. . . . 101
3.8 Differential scanning fluorimetry analysis of PfRH5 melting temperature. 102
3.9 Limited proteolysis of PfRH5. . . . 104
3.10 Purification of PfRH5ΔN and PfRH5ΔNL. . . . 105
3.11 Purification of basigin from *E. coli*. . . . 106
3.12 Analytical gel filtration of PfRH5 bound to basigin. . . . 107
3.13 Surface plasmon resonance to test basigin-binding affinities of PfRH5 variants. . . . 108
3.14 Growth inhibitory assays. . . . 109
3.15 Interaction of PfRH5 with a panel of mouse monoclonal antibodies. . . . 111
3.16 Purification of QA1 antibody fragment. . . . 112

4.1 Endoproteinase GluC cleavage of PfRH5ΔN and PfRH5ΔNL. . . . 117

4.2 Lysine methylation of PfRH5ΔNL. 118
4.3 Purifications of PfRH5ΔNL complexes. 119
4.4 Crystals of PfRH5ΔNL–9AD4. 120
4.5 PfRH5ΔNL–9AD4 diffraction image. 121
4.6 PfRH5ΔNL–9AD4 electron density maps. 125
4.7 Ramachandran Plot and Statistics of PfRH5ΔNL–9AD4 structure. 126
4.8 The PfRH5ΔNL structure. 127
4.9 Electrostatic surface potential of PfRH5ΔNL. 128
4.10 Disorder predictions and secondary structure of PfRH5. 130
4.11 Structural alignment of PfRH5 from different crystal structures. 131
4.12 PfRH5ΔNL in full-length PfRH5 SAXS envelope. 132
4.13 A conserved PfRH5-like fold in other *Plasmodium* RH proteins. 136
4.14 Threading models of other RH proteins. 137
4.15 Residues similar among RH proteins stabilise the fold. 138
4.16 PfRH2b and PfRH4 expression in S2 cells. 139

5.1 Crystals of PfRH5ΔNL–basigin. 144
5.2 PfRH5ΔNL–basigin diffraction image. 145
5.3 PfRH5ΔNL–basigin electron density map. 149
5.4 Ramachandran Plot and Statistics of PfRH5ΔNL–basigin structure. . . . 150
5.5 The PfRH5–basigin structure. 151
5.6 Two basigin conformations found in the asymmetric unit. 152
5.7 SAXS analysis of the PfRH5–basigin complex. 153
5.8 Close-up of the PfRH5–basigin binding site. 155
5.9 Location of PfRH5 polymorphisms, and residues of PfRH5 and basigin implicated in host tropism. 158
5.10 The owl monkey. 159
5.11 Basigin mutant proteins retain PfRH5 binding ability. 160
5.12 Arrangement of two PfRH5–basigin complexes in the asymmetric unit. . . 162
5.13 Native mass spectrometry analysis of PfRH5–basigin. 163
5.14 Sedimentation velocity analysis of the PfRH5–basigin complex. 164
5.15 Sedimentation equilibrium AUC analysis of PfRH5ΔN–basigin. 165

6.1 Crystals of PfRH5ΔNL–QA1. 170
6.2 PfRH5ΔNL–QA1 diffraction image. 171
6.3 PfRH5ΔNL–QA1 electron density map. 172
6.4 Ramachandran Plot and Statistics of PfRH5ΔNL–QA1 structure. 173
6.5 Crystal structures of PfRH5ΔNL bound to two inhibitory antibody fragments. 175
6.6 SAXS analysis of PfRH5–inhibitory antibody fragment complexes. 178
6.7 Comparison of PfRH5 binding sites for basigin and inhibitory antibodies. 179
6.8 SAXS analysis of PfRH5 complexed with non-inhibitory antibody fragments. 182

7.1 The DBL fold and the RH fold. 187

B.1 Surface plasmon resonance fits for the basigin binding of PfRH5 variants. 230
B.2 Surface plasmon resonance fits for PfRH5 binding of basigin variants. . . 231

List of Tables

2.1 Summary of Expression Plasmids 49
2.2 Protein refolding buffers. 61

3.1 PfRH5 full-length and shortened full-length expression results. 93
3.2 PfRH5 N- and C-terminal domain expression results. 95
3.3 PfRH5 deletion construct expression results. 97

4.1 PfRH5ΔNL–9AD4 data collection strategy. 121
4.2 PfRH5ΔNL–9AD4 data processing statistics. 122
4.3 PfRH5ΔNL–9AD4 refinement statistics. 124
4.4 Summary of RH protein sequence alignments and threading. 133

5.1 PfRH5ΔNL–basigin data collection strategy. 145
5.2 PfRH5ΔNL–basigin data processing statistics. 146
5.3 PfRH5ΔNL–basigin refinement statistics. 149
5.4 Table describing interactions between PfRH5 and basigin. 155

6.1 PfRH5ΔNL–QA1 data collection strategy. 169
6.2 PfRH5ΔNL–QA1 data processing statistics. 171
6.3 PfRH5ΔNL–QA1 refinement statistics. 172
6.4 Table describing interactions between PfRH5 and QA1. 174
6.5 Table describing interactions between PfRH5 and 9AD4. 176

A.1 Basigin primers 223
A.2 PfRH1, PfRH2b, and PfRH4 primers 223
A.3 Primers for PfRH5 expression in *E. coli* 224
A.4 Primers for PfRH5 expression in HEK and Sf9 cells 228
A.5 Primers for RT-PCR of hybridomas 228

Abbreviations

6-cys	**S**ix-**c**ysteine
AMA1	**A**pical **M**embrane **A**ntigen **1**
AUC	**A**nalytical **U**ltra**c**entrifugation
AVEXIS	**A**vidity-**B**ased **E**xtracellular **I**nteraction **S**creen
BSG	**B**asi**g**in
CIDR	**C**ysteine-**R**ich **I**nterdomain **R**egion
CR1	**C**omplement **R**eceptor **1**
DBL	**D**uffy-**B**inding-**L**ike
DBP	**D**uffy-**B**inding **P**rotein
EBA	**E**rythrocyte-**B**inding **A**ntigen
EBL	**E**rythrocyte-**B**inding **L**ike Proteins
EGF	**E**pidermal **G**rowth **F**actor-like
EMMPRIN	**E**xtracellular **M**atrix **M**etalloproteinase **In**ducer
Ig	**I**mmuno**g**lobulin
MSP	**M**erozoite **S**urface **P**rotein
Pf	***P**lasmodium **f**alciparum*
Pr	***P**lasmodium **r**eichenowi*
Pv	***P**lasmodium **v**ivax*
Py	***P**lasmodium **y**oelii*
RH	**R**eticulocyte-binding Protein **H**omologue
RIPR	**R**H5 **I**nteracting **Pr**otein
SAXS	**S**mall-**A**ngle **X**-ray **S**cattering
SNP	**S**ingle **N**ucleotide **P**olymorphism
SPR	**S**urface **P**lasmon **R**esonance

Chapter 1

Introduction

1.1 Malaria

1.1.1 History and aetiology

Malaria is an ancient scourge of humans, with descriptions of malaria-like symptoms found in Chinese manuscripts dating from 2700 BC and Mesopotamian clay tablets from 2000 BC. The ancient Greeks, including Homer and Hippocrates, eloquently chronicled the disease. For centuries, it was commonly accepted that malaria was the result of powerful miasmas emanating from swamps. The true aetiology of malaria was not discovered until the late 19th century, after the germ theory of Louis Pasteur and Robert Koch had gained traction in the scientific community, when Charles Louise Alphonse Laveran, a French army officer stationed in Algeria, observed crescent-shaped bodies in the blood of malaria patients, which disappeared when patients were given quinine [1, 2]. By 1890 researchers had identified the parasites which would later become known as *Plasmodium falciparum*, *Plasmodium vivax*, and *Plasmodium malariae*, three of the species which infect humans. Several years later, a British officer named Ronald Ross observed pigmented bodies on the stomach wall of a mosquito that had fed upon a malaria patient. These bodies seemed to rupture into 'rods' which then invaded the insect's salivary glands. Ross went on to demonstrate using avian parasite *Plasmodium relictum* that malaria parasites are transmitted by mosquitoes, which were then classified as female *Anopheles* mosquitoes by Bignami and Grassi [3–5].

1.1.2 Global disease burden

An estimated 3.4 billion people in 97 countries are at risk of malaria (Figure 1.1) [6]. 2012 saw approximately 207 million cases and approximately 627,000 malaria-related deaths, primarily in sub-Saharan Africa. Children are particularly at risk, with an estimated 428,000 children killed by malaria in 2012. Nonetheless, between 2000 and 2012, a redoubling of intervention efforts, including the distribution of long-lasting insecticidal nets, indoor spraying, and increased access to rapid diagnostic tests and artemisinin-based combination therapies, led to a decrease in the global incidence of malaria by 25% and the global mortality due to malaria by 42% [6].

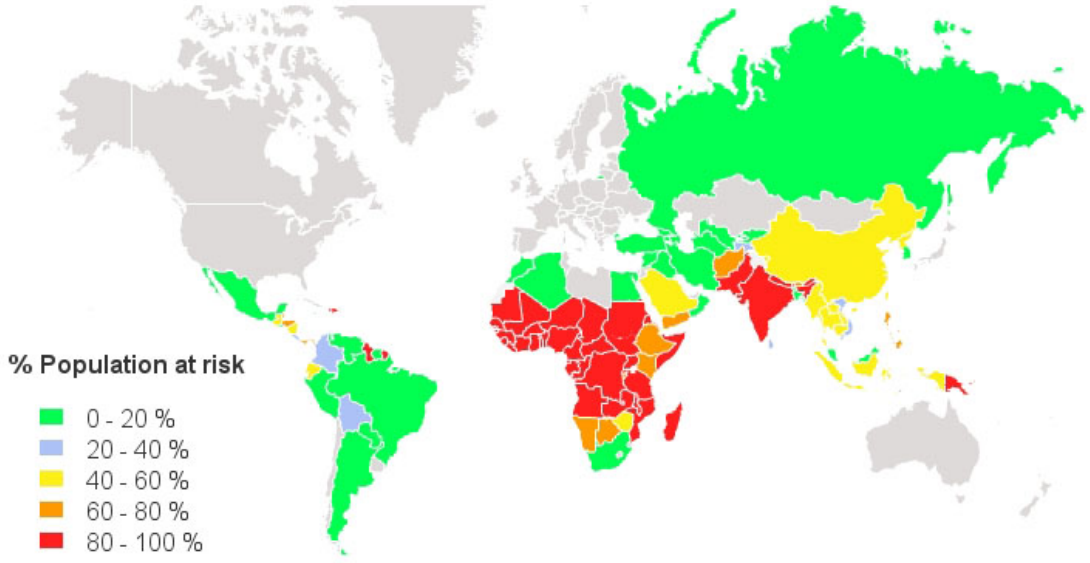


FIGURE 1.1: The percentage of the population at risk of malaria is shown by country. Image from the Global Malaria Mapper tool at `http://www.worldmalariareport.org/`.

1.1.3 *Plasmodium* species

Plasmodium parasites are members of the phylum Apicomplexa. Protozoans of this phylum are characterised by a series of organelles specialised for host cell invasion, known collectively as the apical complex (described in Section 1.2.1). Other Apicomplexan pathogens of medical and veterinary importance are the *Toxoplasma*, *Cryptosporidium*, *Eimeria*, and *Babesia* parasites.

Over 50 species comprise the *Plasmodium* family, of which five, *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*, are known to infect humans. Malaria parasites

that infect non-human primates, including *P. fragile*, *P. coatneyi*, and *P. knowlesi*, and those that infect rodents, such as *P. berghei* and *P. yoelii*, have been crucial to the development of laboratory animal models of human malaria [7–9].

P. falciparum is the most prevalent and the most deadly of the human-infective species, accounting for approximately 80% of cases worldwide [6, 10]. It is particularly fatal for children under five. In severe cases of *P. falciparum* infection, the hallmark symptoms of malaria, fever and anaemia, are accompanied by neurological complications which can lead to coma and death.

In contrast to *P. falciparum*, which predominates in Africa, *P. vivax* is more common in Asia and the Americas. It only infects reticulocytes, whereas *P. falciparum* can infect blood cells of all ages. It is unable to infect reticulocytes which lack the Duffy blood group antigen, leading to an absence from West Africa, where Duffy negativity is common [11]. Interestingly, unlike *P. falciparum*, which does not have a dormant lifecycle stage, so-called *P. vivax* hypnozoites can lie dormant in hepatocytes for months or years before relapse into blood-stage disease. Fewer fatalities are reported due to *P. vivax* compared to *P. falciparum*, but mounting evidence of severe manifestations of *P. vivax* disease, including respiratory complications and coma, is prompting rapid re-evaluation of the idea that *P. vivax* malaria is 'benign' [12–15].

P. malariae and *P. ovale* parasites cause relatively mild infections, only rarely progressing to clinical disease [16, 17]. These parasites are often found in mixed infections with *P. falciparum* and *P. vivax*. *P. ovale* resembles *P. vivax* in that it only infects reticulocytes, and can form latent hypnozoites. *P. malariae*, by contrast, prefers older erythrocytes, and appears to persist in the blood at a low level for long periods, leading to occasional recrudescence of disease [10].

P. knowlesi primarily infects Old World monkeys in Southeast Asia, but it has recently emerged that *P. knowlesi* is zoonotic, with the monkeys serving as a reservoir for human infection [18]. Many cases of *P. knowlesi* malaria in Malaysia were originally misclassified as *P. malariae* infections, as by light microscopy the early blood-stage parasites look similar. However, unlike *P. malariae*, which multiplies in the blood only every three days (quartan), *P. knowlesi* has a daily (quotidian) asexual replication cycle. As a result, if untreated, *P. knowlesi* can rapidly reach critically high and life-threatening densities in the blood.

As the *P. falciparum* species is the most common, the most virulent, and the most relevant to this thesis, it will be the main focus of the remainder of Section 1.1, unless otherwise specified.

1.1.4 Malaria genomes

The genomes of *P. falciparum*, *P. vivax*, *P. yoelii yoelii*, *P. knowlesi*, and, very recently, chimpanzee-infective *P. reichenowi* have been sequenced [19–23]. They encompass 23-27 million bases and approximately 5,500 genes on 14 chromosomes [24]. Low-complexity sequences abound, and the overall A+T content is unusually high. In the extreme case of *P. falciparum*, a remarkable 79.6% of the genome is A or T. Over the sequenced genomes, approximately 77% of predicted proteins are conserved. Interestingly, however, many of these proteins lack homologues in other organisms, and their cellular function is unknown [21].

1.1.5 The *Plasmodium* lifecycle

Plasmodium parasites cycle between two hosts, the human and the female *Anopheles* mosquito vector, and between sexual and asexual reproductive stages. Their lifecycle can be divided into four overall stages: the asexual stage in the human liver (exo-erythrocytic schizogony), the asexual stage in the human red blood cell (erythrocytic schizogony), the sexual stage in the mosquito (gamogony), and the asexual stage in the mosquito (sporogony) [10]. A generalised *Plasmodium* lifecycle is depicted schematically in Figure 1.2.

1.1.5.1 Exoerythrocytic schizogony

In the process of feeding on a human, the mosquito carrier inserts its proboscis and releases saliva containing sporozoites, primarily into the the skin [25, 26]. Sporozoites are the elongated (10-15 μM by 1 μM), motile, liver-infective forms of the parasite. They move by an unusual form of locomotion known as gliding motility, which is independent of cilia, flagella, or pseudopodia. In the current model of gliding motility, an actin-myosin motor is coupled to an extracellular substrate by the TRAP family of proteins: the motor pulls TRAP backward, but because TRAP is bound to a surface, the parasite

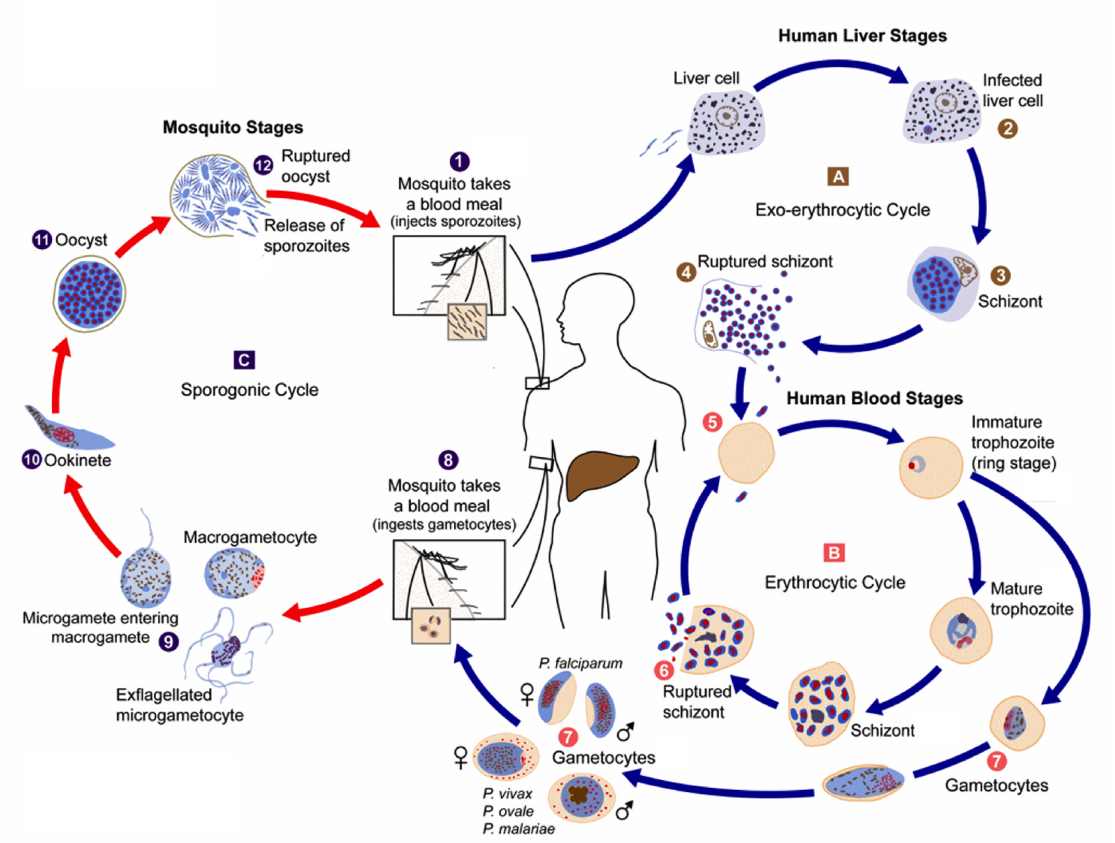


FIGURE 1.2: The lifecycle of *Plasmodium* parasites is divided into the human stages (blue arrows) and the mosquito stages (red arrows). In taking a blood meal, a female *Anopheles* mosquito inoculates the human with sporozoites (1). During the exo-erythrocytic cycle (A), the sporozoites invade (2) and mature into schizonts (3) within liver cells. The schizonts subsequently rupture to release merozoites into the blood stream. A merozoite invades a red blood cell (5), develops through the ring, trophozoite, and schizont stages, and ruptures to release a profusion of new merozoites to invade more red blood cells (the erythrocytic cycle, B). A small percentage of merozoites will develop into micro- and macrogametocytes, or male and female gametocytes, respectively (7), which are subsequently taken up by a feeding mosquito (8). In the mosquito gut, the gametocytes mature to gametes (which in the case of the microgametocyte involves exflagellation) and fuse to form a zygote (9). The zygote matures into a motile ookinete (10), which crosses the midgut epithelium and develops into an oocyst (11). The oocyte ruptures to release sporozoites (12), which migrate back to the salivary gland of the mosquito, where they will be injected into a human along with the saliva (1). Image from the Centers for Disease Control and Prevention (CDC) at `http://www.cdc.gov/malaria/about/biology/`.

moves forward [27, 28] (see Section 1.2.3.2). Most of the sporozoites remain at the site of the bite, but some enter the blood stream or the lymphatic system [29]. Interestingly, some of the rodent *P. berghei* and *P. yoelii* parasites remaining in the skin can develop into merozoites in skin cells, but whether they are capable of instigating a blood-stage infection is unresolved [30]. The merozoites in the lymphatic system end up in the proximal draining lymph node, and while they can invade lymphatic endothelial cells,

they cannot make merozoites therein [31].

The sporozoites that enter the bloodstream are carried to the liver sinusoid. Liver localisation is believed to be mediated by interactions between circumsporozoite protein (CSP), which forms a coat over the surface of the sporozoite, and heparan sulphate proteoglycans in the extracellular matrix of hepatocytes [32]. Here they wound the resident macrophages (Kupffer cells), cross the sinusoidal barrier, and traverse several hepatocytes before invading a hepatocyte productively to form an internalised parasitophorous vacuole in which to develop [33]. Hepatocyte cell traversal is a brute-force approach which involves the breakage of the membrane and movement through the cytoplasm [34]. The wounded cells release hepatocyte growth factor, which seems to stimulate productive invasion of hepatocytes and downstream intracellular development of the parasite [35, 36]. Productive invasion to form a vacuole might involve the cleavage of CSP to reveal a masked N-terminal domain which binds to the host cell [37].

When the sporozoite is safely entrenched in the parasitophorous vacuole, it develops into a trophozoite. After 1-2 weeks, an asexual fission process produces thousands of tiny merozoites (2.5 μM by 1.5 μM). During egress, hundreds of merozoites are packaged in vesicles known as merosomes, surrounded by the hepatocyte membrane, and released into the liver sinusoid. Merosomes survive passage through the heart and lodge in the pulmonary capillaries until they disintegrate, releasing merozoites into the blood circulation [38, 39]. In *P. vivax* and *P. ovale*, a parasite form called the hypnozoite can remain dormant within hepatocytes for years before relapse and release of merozoites into the bloodstream.

1.1.5.2 Erythrocytic schizogony

The merozoites rapidly invade red blood cells to inaugurate the erythrocytic schizogony phase. Invasion involves initial contact with a suitable host cell, reorientation so the apex of the parasite contacts the red blood cell, formation of a junction and invagination into a parasitophorous vacuole. The process relies upon recognition events mediated by host-parasite surface protein interactions. Erythrocyte invasion is central to this thesis, and will be discussed in detail in Section 1.2.

Once inside the erythrocyte, the parasite enters a 48 hour development cycle, which culminates in the release of a new generation of merozoites to infect further erythrocytes. First, the parasite enters the early trophozoite stage, also known as the 'signet ring' or ring stage due to its appearance in light microscopy, where a dark nucleus is encircled by cytoplasm. In reality, the parasite is cup-shaped at the stage. Small meals of host cell cytoplasm are obtained through a small ring at the surface of the parasite known as the cytostome. Haemoglobin is digested within a large digestive vacuole, and toxic free haem converted to inert haemozoin crystals [40]. The ring stage parasite morphs into the mid-trophozoite, which is more rounded, with irregular bulges at the surface [41]. This stage sees a great expansion in the rough endoplasmic reticulum to facilitate protein export (see below). Thereafter, the trophozoite becomes a schizont, the form which undergoes repetitive nuclear divisions to form approximately 16-20 nuclei (a process known as schizogony). The cytoplasm begins to form a series of merozoite buds, where the apical organelles of the merozoite, and later the nucleus, mitochondrion, and plastid, assemble (Section 1.2.1). Cytoskeletal elements form a cleavage furrow around each merozoite to demarcate it.

Egress of the merozoites involves the breakage of the parasitophorous vacuole and red blood cell membranes, in a mechanism not fully understood but generally described as 'explosive' [42–44]. Osmotic swelling is coupled with the destabilisation of the cytoskeleton underlying the erythrocyte membrane. The proteases of the serine repeat antigen (SERA) family have been implicated in cytoskeleton destruction [45]. It has been suggested that the increase in osmotic pressure leads to the opening of a pore in the erythrocyte membrane which ejects 1-2 merozoites; thereafter, the membrane begins to curl outward and evert, precipitating the release of the rest of the merozoites [42].

Throughout the intracellular blood stages, *Plasmodium* parasites modify the red blood cell into a suitable refuge in which to grow and divide, while remaining hidden from the immune system. This modification is achieved through the export of parasite proteins, many of which are targeted to the red blood cell by a *Plasmodium* export element (PEXEL) motif (also known as a host-targeting signal), RxLxE/Q/D [46, 47]. Export from the parasite vacuole may be via the *Plasmodium* Translocon for EXport (PTEX), a multi-protein complex which localises to the parasite vacuolar membrane and may form a pore by multimerisation of the EXP2 subunit [48, 49]. Recently, it was shown that when PTEX function is compromised (by knocking down components HSP101 or PTEX150),

export of both PEXEL-containing and PEXEL-negative proteins is markedly reduced [50, 51]. Once in the parasite cytoplasm, the proteins are thought to be ferried to their final destinations by other exported proteins acting as chaperones; vesicle-like structures and so-called Maurer's clefts, flat organelles tethered to the red blood cell membrane, might also be involved in trafficking [52, 53].

Exported proteins facilitate nutrient and waste exchange across the red blood cell membrane and strengthen and remodel the red blood cell cytoskeleton [54, 55]. Additionally, and unique to *P. falciparum*, members of a single protein family, the PfEMP1s, are targeted to protrusions on the red blood cell surface called knobs, where they mediate adhesive interactions with a variety of receptors on the vascular endothelium and, in pregnant women, the placenta [56, 57]. This cytoadherence of infected red blood cells prevents rapid clearance in the spleen [58]. The parasite-infected cell is sequestered in many organs and tissues, but it is sequestration in the placenta and the brain that is most commonly associated with severe pathology, causing placental and cerebral malaria, respectively [59, 60]. PfEMP1 are multi-domain proteins, with particular Duffy-binding-like (DBL) and cysteine-rich interdomain region (CIDR) domains mediating adhesion to different receptors (for instance, CD36, ICAM-1, EPCR, and CSA) expressed variably in different tissues. They are encoded by the highly variable *var* gene family, and only one variant is expressed at once; switching between *var* genes is an effective method of clonal antigenic variation to evade the host antibody response [61–63]. Furthermore, such switching may change the cytoadherence properties of the infected red cells, the location of their sequestration, and the pathological consequences of a bout of malaria [64]. PfEMP1 variant VAR2CSA binds to chondroitin sulphate A (CSA) on the placenta to cause placental malaria [65]. PfEMP1 variants containing certain tandem domains (called domain cassettes 8 and 13) have been linked to severe childhood disease, through their interaction with endothelial protein C receptor (EPCR) [66].

Occasionally, instead of progressing to the ring, trophozoite, and schizont stages, a penetrating merozoite instead follows a different developmental route to become a male microgametocyte or female macrogametocyte, the precursors of gametes [67]. The determinants of gametocyte formation are unclear, but members of the Apicomplexan AP2 family of transcription factors have been shown to be upregulated early in sexual differentiation [68]. Commitment to the asexual stage or to gametocyte development (male or female) occurs by the previous trophozoite stage [69, 70]. Like the asexual

stages, *P. falciparum* gametocytes are also sequestered by endothelial adhesion during maturation, primarily to the bone marrow and spleen [71, 72]. The mature gametocyte and its host red blood cell are released back into the peripheral circulation to be taken up by a feeding mosquito.

1.1.5.3 Gamogony

The sexual phase of *Plasmodium* development begins in the mosquito midgut, where the ingested male and female gametocytes develop into gametes and emerge from the human red blood cells in response to environmental and mosquito-specific triggers, including an increase in pH, a drop in temperature, and exposure to xanthurenic acid (gametocyte-activating factor) [73, 74]. Microgametocytes undergo a process called exflagellation, which begins with a series of mitoses to generate eight nuclei. Each nucleus, together with an axoneme and a small amount of cytoplasm, make up the motile microgamete. Meanwhile, the female gametocytes mature into macrogametes. A microgamete and macrogamete fuse to form a zygote, which then matures into a motile ookinete.

1.1.5.4 Sporogony

The ookinete invades the mosquito midgut epithelium, piercing the cell membrane and traversing the cytoplasm [75]. The mechanism of this cell traversal is probably conserved between sporozoite and ookinete, as it relies upon the same protein (CelTOS) [76]. The interaction(s) between ookinete and the epithelial cell which allow the invasion are not entirely clear, but ookinete proteins thought be involved include the membrane attack ookinete protein (MAOP), the circumsporozoite thrombospondin-related protein (CTRP), and TRAP [77, 78]. After the cell passage, wedged into the space between the basal lamina and the epithelium, the ookinete develops into the round oocyst. Development might be triggered by interactions with the basal lamina, namely interactions between parasite proteins Pfs25 and Pfs28 and mosquito laminin [79, 80]. A number of nuclear divisions occurs, leading to a multinucleate parasite. Infoldings of the oocyst plasma membrane form compartments called sporoblasts, from which haploid sporozoites bud. The sporozoites egress from the mature oocyst, in a process apparently mediated by proteases such as egression cysteine protease [81]. The sporozoites make their way to the salivary glands, but whether this is a chemotactic migration, or passive

movement in the haemocoel, is unresolved [82, 83]. They traverse the salivary gland epithelium to reach the salivary cavity and then the salivary ducts, where they are ready to be injected into a human.

1.1.6 Disease manifestations and pathology

1.1.6.1 Symptoms and pathology

Malaria manifests as a combination of host inflammatory reactions and anaemia [10]. The initial symptoms are nausea, fatigue, a slightly elevated temperature, muscle aches, vomiting and diarrhoea. The rupture of erythrocytes releases cellular debris, metabolic waste, and haemozoin crystals into the blood, triggering the host inflammatory response. This results in fever paroxysms which are usually synchronised with periodic ruptures of the erythrocytes after the initial blood stage invasion. (The initial invasion is not synchronous because the upstream liver stage parasites are often of different populations at different stages of development; how synchrony arises in future blood stage cycles is unclear.) Anaemia results from the large-scale destruction of red blood cells, combined with the lack of availability of haemoglobin as a result of its digestion by the parasite.

1-2% of infections progress to severe and life-threatening disease. Symptoms of severe malaria include respiratory distress, severe anaemia, coma, and multi-organ failure [64]. Severe disease is usually, though not exclusively, the result of *P. falciparum* infection, and stems from a high parasite load combined with microvascular sequestration (Section 1.1.5.2). Sequestration is mediated by the adhesion of an infected red blood cell to the vascular endothelium (cytoadhesion), to uninfected red blood cells (rosetting), and to other infected red blood cells (platelet-mediated clumping) [84]. The consequences of sequestration are vascular obstruction and poor tissue perfusion; lactic acidosis (probably resulting from anaerobic respiration in anoxic tissues) and other metabolic repercussions; and the release of pro-inflammatory cytokines such as tumour necrosis factor.

While most adults living in endemic regions have acquired some protective immunity against malaria, pregnancy causes severe symptoms in women and significant risk to their unborn babies. The binding of PfEMP1 variant VAR2CSA to CSA sequesters the parasite in the maternal vascular region of the placenta [65]. The pathological consequences include anaemia in the mother, preterm delivery, foetal growth restriction

(due to increased inflammation and decreased blood flow), low birth weight, and foetal and infant mortality [85].

Cerebral malaria, which most commonly affects African children, is defined by impaired consciousness or coma, and is fatal in 15-20% of cases. The survivors sustain permanent neurological damage. The mechanisms which lead to brain injury are incompletely understood, but sequestration of parasitized red blood cells in the brain is of central importance. PfEMP1 binding to ICAM-1 expressed on the endothelial lining has been implicated [60, 86].

1.1.6.2 Host immune response and acquired immunity

The innate immune response to erythrocyte stage infection involves the binding of pathogen-associated molecular patterns (PAMPs), namely glycosylphosphatidylinositol (GPI) and haemozoin, to host toll-like receptors (TLRs) to stimulate production of pro-inflammatory cytokines in dendritic cells and macrophages [87]. TLR activation also leads to upregulation of cell adhesion molecules such as ICAM-1 in vascular endothelial cells, which may, in turn, allow further binding of infected erythrocytes via PfEMP1. The adaptive immune response involves CD8+ T cells (pre-erythrocytic stage) and CD4+ T cells (erythrocytic stage) [10]. During severe infections, the balance between the T_H1 and T_H2 response becomes shifted in favour of T_H2.

In areas of endemic malaria transmission, naturally acquired immunity to the blood stages reduces parasitemia and prevents clinical disease [88]. Infants are protected by passive transfer of protective antibodies in their mothers' milk. (Upon weaning, however, these children are no longer protected, and at serious risk of severe or mortal malaria.) Acquired immunity takes the form of premunition, reliant upon continued, low-level infection to confer resistance; complete cure would render a previously immune individual susceptible [10]. Women become resistant to placental malaria over successive pregnancies, as they acquire antibodies against VAR2CSA which prevent binding of the infected red blood cells to the placenta [89].

Natural immunity to malaria is difficult to acquire and incredibly short-lived, in large part because of mechanisms the malaria parasite has evolved to evade immune detection. Antigenic diversity is a major part of immune evasion, and is achieved by three

main mechanisms. Firstly, the variable expression of the *var* genes presents different PfEMP1 molecules on the red blood cell surface (Section 1.1.5.2). Secondly, sequence polymorphisms in many parasite surface molecules (such as those involved in erythrocyte invasion; see Section 1.2) can limit the epitopes capable of eliciting an immune response protective against multiple strains. Finally, during erythrocyte invasion, different merozoite strains vary the combination of erythrocyte-binding proteins on the parasite surface (Section 1.2). Not only does this circumvent the immune response, but it also enables the parasite to bind to a breadth of erythrocyte receptors, variably expressed in human populations, to ensure invasion can proceed in spite of erythrocyte polymorphisms.

1.1.6.3 Genetic resistance

Several genetic conditions in the host can impair the ability of the parasite to thrive, including Duffy negativity, sickle-cell anaemia, and glucose-6-phosphate dehydrogenase deficiency. As aforementioned, Duffy negativity in West Africa leads to a dearth of *P. vivax* infection, which requires the Duffy antigen for erythrocyte invasion (Section 1.1.3). Heterozygote carriers of sickle cell anaemia have a survival advantage in malaria-endemic regions, for reasons which have long been unclear. Haem oxygenase-1, which is upregulated by the mutant haemoglobin, may play a role [90].

1.1.7 Treatment and prevention

In 1955, the World Health Organisation launched a global campaign to eradicate malaria. The main weapons in the eradication arsenal were a combination of anti-malarial drugs, most notably chloroquine, and vector control measures, including liberal use of insecticides such as DDT. In spite of early successes, the campaign was abandoned a mere fourteen years later due to major resurgences in many areas, and the goal scaled back to 'malaria control.' Resurgences stemmed in large part from widespread parasite resistance to chloroquine and mosquito resistance to insecticides [91]. In recent years, however, the concept of eradication has re-emerged. In 2007, for example, a consulting process to outline the research and development advances required for malaria eradication (malERA) was established [92]. Central, and perhaps vital, to any eradication strategy would be an effective vaccine against malaria, which is still conspicuously lacking.

The stages of the parasite lifecycle which are of most interest for drug and vaccine development are sporozoite invasion of the liver, gametocyte transmission to the mosquito, and the asexual blood stage. The shuttling of sporozoites and gametocytes between vectors involves, at most, several hundred parasites, as compared to the billions of parasites present during the blood stage. Further, targeting the gametocytes could block transmission and break the infection cycle. Effective interventions against the blood stage are critically important because it is during this stage that the symptoms and pathology of the disease arise (Section 1.1.6).

1.1.7.1 Vector control

The epidemiology of malaria infection in humans is entirely determined by the location, abundance, seasonal fluctuations, and feeding habits of the anopheline vector. Vector control has been essential to eradication of malaria in areas such as the USA and Europe, and continues to be important for disease control in affected areas. It primarily consists of long-lasting insecticidal nets and indoor residual spraying. The main insecticides used, pyrethroid insecticides and DDT, target the same voltage-gated sodium channel, and vector resistance is on the rise [93–95].

The insect-infective bacterium *Wolbachia* has raised hopes of a new form of vector control. *Wolbachia* do not naturally infect anopheline mosquitoes, but recent results suggest that a certain strain can in fact propagate in the malaria vector [96]. Infection with this bacterium impedes *Plasmodium* development within the mosquito, perhaps by precipitating an immune response which targets the parasite [96–98].

1.1.7.2 Chemotherapy

Many anti-malarial drugs target the process of haemoglobin digestion in the digestive vacuole during the red blood cell stage. The process releases free, oxidised haem (containing Fe(III)), a toxic compound capable of membrane damage [99]. The parasite therefore converts the haem into inert haemozoin crystals. The drug chloroquine, the front-line drug of the 1950s before resistance emerged, interferes with the crystallisation process [100]. Chloroquine resistance stems from mutations in the chloroquine resistance

transporter, PfCRT, an integral membrane protein in the digestive vacuole membrane which pumps chloroquine out of the vacuole and away from its site of action [101].

The artemisinins, the modern front-line drugs, also act on the haemoglobin digestion pathway [102]. They were developed from an ancient Chinese herb used to treat fevers, qinghaosu. They contain an endoperoxide which is cleaved by free iron (produced during haemoglobin digestion), releasing cytotoxic drug metabolites to damage the digestive vacuole membrane and interfere with haemozoin formation, along with other destructive downstream effects [102, 103]. Artemisinins are fast-acting against the blood stages, and seem to have some potency against the gametocyte stages, which could help to block transmission. However, they have a short half-life, and are thus usually prescribed in combinations with other drugs with longer half-lives (such as lumefantrine or amodiaquine), with the latter effective against recrudescences occurring after the artemisinin compound has been cleared. Recent, alarming reports of prolonged parasite clearance times in Western Cambodia warn of the rise of parasite resistance to artemisinins [104, 105]. A number of genomic loci have been associated with prolonged artemisinin clearance times, including a large expanse of *P. falciparum* chromosome 13 [106]. Recently, mutations in the kelch propeller domain have been correlated with artemisinin resistance, though the normal function of this protein, and the effect of these mutations, remains unknown [107].

Other drugs, such as pyrimethamine and sulfadoxine, are antifolates which inhibit enzymes in the essential folate synthesis pathway [108]. Pyrimethamine blocks the *P. falciparum* dihydrofolate reductase/tetrahydrofolate synthetase enzyme [109]. Parasite resistance to these drugs is also on the rise [110].

In light of the burgeoning resistance to many anti-malarials, development of novel drugs is an urgent priority. Drug discovery efforts centre on high-throughput screening of large compound libraries to identify potential new drugs, and, to a lesser extent, on modifying existing drugs to circumvent resistance [111].

1.1.7.3 Vaccine candidates

The remarkable morphological changes of *Plasmodium* parasites during their complex lifecycle, together with parasite antigenic variation and evasion of the host immune

response, preclude the development of sterile immunity in infected individuals (the phenomenon of premunition; Section 1.1.6.2). Thus, a malaria vaccine affording long-term, sterile protection is a formidable challenge, and one which has yet to be surmounted. In pursuit of a vaccine, three different types of candidate have been most heavily investigated: transmission-blocking vaccines, pre-erythrocytic vaccines, and blood-stage vaccines [112].

Transmission-blocking vaccines would target antigens of the gamete, zygote, and ookinete forms of the parasite, with the aim of preventing the transition to the oocyst and downstream sporozoites [113]. Transmission blocking immunity is predominantly mediated by vaccine-induced human antibodies against surface antigens, which act in the mosquito midgut stage to prevent fertilisation, or destroy the gametes or ookinetes [114]. Importantly, these vaccines would not protect the immunised individual, but rather, by abolishing transmission, would prevent the spread of malaria in a community. Ookinete surface proteins Pfs25 and Pfs28 from *P. falciparum*, and their *P. vivax* counterparts, are the leading candidates [115, 116]. Pfs25 loaded onto nanoparticles is in Phase I clinical trials [117]. Other candidates are Pfs48/45, Pfs230, and Hap2 [118].

Leading pre-erythrocytic vaccine candidates include whole, attenuated sporozoites and antigenic portions of circumsporozoite protein (CSP). The RTS,S vaccine, consisting of CSP fused to the hepatitis B surface antigen, has completed Phase III clinical trials, but confers only 30-50% protection, and that short-lived [119]. Another potential subunit vaccine candidate is CelTOS, which is expressed both in sporozoites and ookinetes (Section 1.1.5) [120]. The immunogenic potential of whole attenuated sporozoites has long been recognised, but the idea has received new impetus recently, with promising initial results in a clinical trial [121, 122]. Nonetheless, many potential stumbling blocks remain (for example, high dosage requirements, manufacturing bottlenecks, and the need for cryopreservation during vaccine delivery).

Current blood-stage vaccine targets are primarily cell-surface molecules implicated in erythrocyte invasion (Section 1.2), with the rationale that preventing invasion will prevent malarial disease. Apical membrane antigen 1 (AMA1), erythrocyte-binding antigen-175 (EBA-175), merozoite surface proteins (MSP-1, MSP-2, and MSP-3), serine repeat antigen 5 (SERA5), and glutamate-rich protein (GLURP) are all surface proteins currently in vaccine development [112]. To date, however, no blood-stage malaria vaccine

candidate has exhibited clinical efficiency in human trials [123]. These surface antigens are extremely polymorphic, and it is thought to be this genetic diversity which confounds efforts to produce a broadly effective vaccine [124, 125]. In marked contrast, merozoite protein PfRH5 has only very limited sequence diversity, and is capable of inducing invasion-blocking antibodies that are effective across the common variants of PfRH5 [126–129]. Since its discovery it has, therefore, received considerable attention as a vaccine candidate. This will be discussed in detail in Section 1.4.3.

Recently, the idea of whole attenuated blood-stage parasites has emerged. Immunisation with *P. yoelii* parasites genetically incapacitated by deletion of purine nucleoside phosphorylase or nucleoside transporter 1 protected against infection with infective parasites [130, 131]. Antibodies against another novel vaccine candidate, called schizont egress antigen-1 (PfSEA-1), have been shown to interfere with schizont rupture in infected red blood cells [132]. Anti-PfSEA-1 antibodies have been correlated with protection from severe disease and may reduce parasite load.

1.2 Erythrocyte invasion by *P. falciparum*

P. falciparum instigates disease in humans by invading erythrocytes. Invasion is rapid and efficient to evade the host immune response: within about two minutes of egress from one erythrocyte [44, 133], *P. falciparum* merozoites recognize and infiltrate another. The invasion process is a particularly attractive target for anti-malarial drug design and vaccine development for a number of reasons. Firstly, all of the pathological effects of malaria arise at the erythrocyte stage. The merozoites are transiently located outside the cell, rendering them more accessible to drugs and immune attack. Furthermore, the invasion machinery is obligatory and specific to the parasite. Finally, antibodies raised against merozoite antigens are capable of impeding erythrocyte invasion, and have been shown to be involved in protective immune responses [134, 135].

The fundamental steps of erythrocyte invasion were elucidated in the 1970s from live video microscopy [43] and electron microscopy [136, 137] of invading *P. knowlesi* and other, not human infective, parasites. Recently, major technological advances have filled in many gaps in our knowledge of this process, particularly in *P. falciparum*, which was historically refractory to culturing and genetic manipulation. These new

tools include methods to isolate viable, invasion-competent *P. falciparum* merozoites [138, 139]; genetic methods to produce parasites with specific deletions, truncations, or other alterations in putative invasion proteins to gauge their importance in the process [140]; and advanced microscopy techniques to better visualise invasion [141]. In spite of these advances, however, many gaps in our understanding have yet to be filled. The steps of invasion are highly coordinated, but exactly how they are triggered and synchronised is contentious. Though we have identified many of the key molecular players, we do not know exactly when and how they act. In the following sections, the cellular events of invasion will first be described, before implicated proteins, and any clues as to their mode and stage of action, are introduced.

1.2.1 Merozoite substructure

The merozoite is one of the smallest eukaryote cells known, yet it manages to package within it all of the machinery required for egress from a dying red blood cell and invasion of a new red blood cell. The overall shape has been likened to a lemon, with an apical prominence [41]. The apex contains a variety of secretory organelles crucial to the invasion process: namely, the dense granules, the micronemes, and the rhoptries (Figure 1.3). The merozoite also contains a mitochondrion, a plastid known as the apicoplast, ribosomes, and a nucleus, which are anchored in place by a minimal cytoskeleton. Around the merozoite is a bristly coat comprised of glycosylphosphatidylinositol (GPI) anchored membrane proteins, most prominently merozoite surface protein 1 (MSP-1) [142, 143]. Within the plasma membrane are an additional two membranes, which together form a flattened membranous sac called the inner membrane complex. Altogether, the three encompassing membranes are known as the pellicle. At the apical prominence are three protein rings called polar rings, which are connected to the pellicular cisterna, and to two subpellicular microtubules which extend along one side of the merozoite.

The secretory organelles are vesicles which fuse with the membrane to release their parasite protein contents during invasion. The largest of these are the rhoptries, two pear-shaped bodies which narrow into to the apical prominence (Figure 1.4). They consist of two distinct regions, the wide, electron-dense, granular posterior bulb and the narrow, electron-lucent anterior neck [144]. Intriguingly, rhoptry proteins can localise either to the rhoptry neck or the bulb domain, where they remain as distinct protein

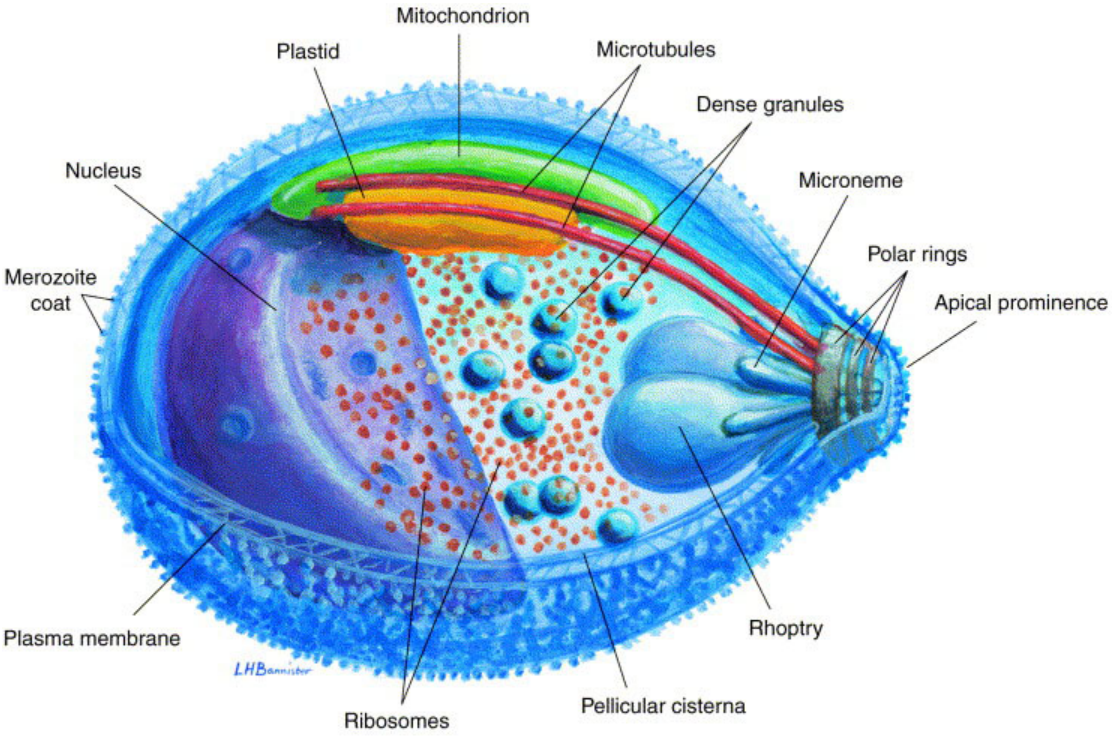


FIGURE 1.3: Schematic depiction of merozoite ultrastructure. Taken from [ref. 41].

populations that do not diffuse between the subcompartments [145]. How this compartmentalisation is established and maintained is unknown, but may be due to the different protein and lipid compositions in each compartment forming distinct aggregates which remain in their respective locations [144, 146]. Rhoptries are formed from the fusion of distinct, Golgi-derived vesicles, which then separate into the two subcompartments [144].

The other secretory vesicles are the micronemes, thin, elongated, and nestled against the neck of the rhoptries, and the dense granules, spherical and suspended in the cytoplasm (Figure 1.3). Distinct from (but originally confused with) dense granules are the exonemes, rounded vesicles implicated in schizont egress rather than merozoite invasion [147].

1.2.2 Stages of erythrocyte invasion

The invasion process—consisting of initial contact, apical reorientation, junction formation, and active penetration—is summarised in Figure 1.5. After egress (Section 1.1.5.2), non-motile merozoites are exposed to low potassium levels in the blood, which leads to internal calcium release and temporally coordinated secretion of the apical organelles:

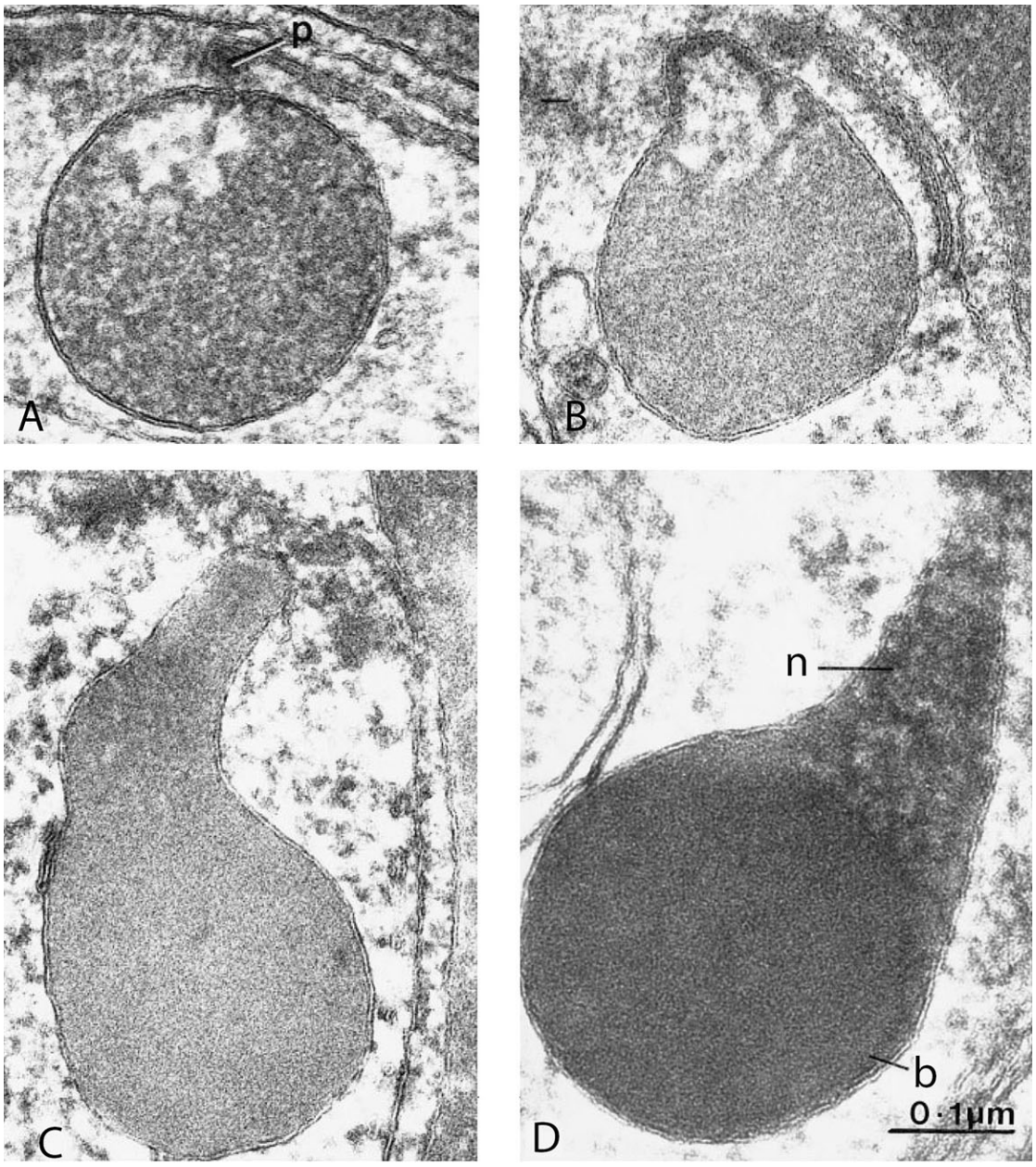


FIGURE 1.4: Electron micrographs of a section through the apex of a *P. falciparum* merozoite, revealing the stages of rhoptry development, starting from a sphere comprised of vesicles (A). Note the translucent region that will become the rhoptry neck, and the nascent polar rings (p). The rhoptry neck begins to develop in (B) and (C). In panel (D), an almost mature rhoptry has formed, with an electron-dense bulb (b) and a less dense neck (n). Taken from [ref. 144].

first the micronemes and then, when internal calcium levels are restored, subcompartments of the rhoptries (the dense granules are released after the invasion process is complete) [139, 141, 148]. Fusion of the micronemes and rhoptries releases proteins involved in red blood cell adhesion and invasion to the cell surface. The merozoites diffuse randomly until they meet an uninfected red blood cell, a process which takes up to one minute *in vitro* [43, 133]. Initial contact with the red blood cell is reversible and of low

affinity, but it is specific, as merozoites do not attach to the erythrocytes of non-host species [149]. Such specific interactions are probably mediated by GPI-anchored proteins coating the merozoite surface and their peripherally associated binding partners (Section 1.2.3.1). This first contact causes significant deformation of the surface of the red blood cell membrane, which partially enfolds the merozoite in a manner oddly reminiscent of the early stages of phagocytosis, of which process the red blood cell is not capable [43, 150]. These dynamic morphological changes may be the result of calcium influx and signalling in the red blood cell [150].

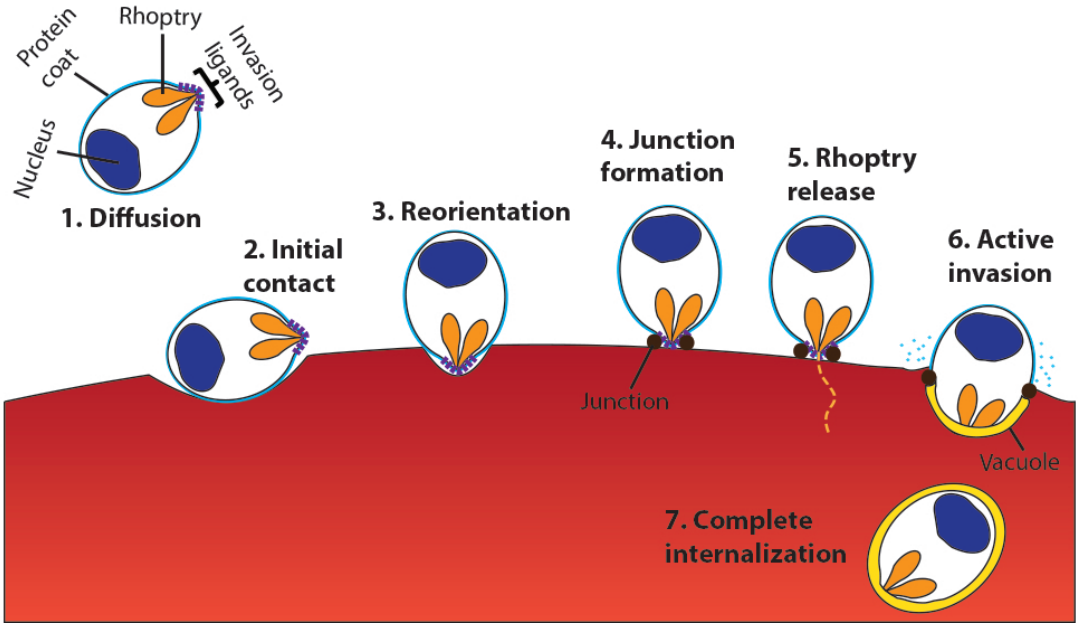


FIGURE 1.5: The stages of erythrocyte invasion are shown schematically. A newly egressed merozoite (1) contains apical organelles such as the rhoptries (orange), a basal nucleus (blue-purple), and a bristly protein coat (cyan). The release of apical organelles in a coordinated manner exposes invasion ligands (purple lines) at the apex. Initial contact (2) with an erythrocyte (red) is mediated by the protein coat. The merozoite then reorients so that its apex contacts the red blood cell (3). A tight junction (a close apposition of the two membranes) forms (4), and the rhoptries are released into the red blood cell (5). Active invasion proceeds via a parasite actinomyosin motor (6). The parasite ends up in a parasitophorous vacuole (yellow), completely encompassed within the red blood cell (7). Adapted from [ref. 151].

The next stage of invasion is the reorientation of the merozoite so that its apex contacts the red blood cell [43]. How the reorientation occurs is still unclear, though it appears to be a passive process that does not involve the parasite motor (see below) [149]. One hypothesis is that erythrocyte deformations facilitate merozoite 'tumbling' along the surface of the red blood cell [150, 151]. It may be that a succession of weak merozoite attachments occur at random until a final, tight contact with the apex forms. Or perhaps a gradient of coat proteins on the merozoite surface establishes a gradient of interaction

affinities leading up to the irreversible apical contact, which promotes movement toward the apex and prevents the merozoite rolling backward [150]. Another model focuses on the curvature of the membrane, positing that adhesive interactions between parasite and erythrocyte might induce local membrane curvature and provide energetic impetus to bring the most curved part of the membrane (the apex) into contact with the erythrocyte [152]. Regardless of the mechanism, once reorientation has occurred (usually in ∼6 seconds), the red blood cell returns to its normal biconcave shape, no longer exhibiting the deformability of the earliest stages pre-invasion, and the process of active invasion begins [43].

Indeed, at some point after apical reorientation, the parasite becomes 'committed' to invasion, at which point the erythrocyte and parasite can no longer dissociate, and invasion becomes an inevitability [141]. It is unclear, however, what molecular event or process might herald invasion commitment. Adhesive proteins released from the secretory organelles, including members of the reticulocyte-binding protein homologue (RH) and erythrocyte binding-like (EBL) protein families, are localised to the apex of the parasite, where they mediate specific contacts with erythrocyte surface proteins and may have a role in reorientation and/or subsequent steps of invasion (Section 1.2.3.3). The binding of RH/EBL proteins to their target erythrocyte receptors has been proposed to signal invasion commitment, precipitating downstream events without any further checkpoints. This theory is based on EBA-175 or PfRH4 knockout experiments in which downstream junction formation was not seen [141]. Another possibility is that the complex of AMA1 and RON2 constitutes the master switch of invasion (see next paragraph) [153]. Any signalling resulting from flipping the invasion 'on-switch' is likewise unknown, though might occur through the cytoplasmic tails of parasite adhesion proteins such as AMA1 and PfRH2a/b [148, 154].

The parasite apex is now poised above a slight depression, the invasion pit, on the red blood cell surface. A tight connection forms between the parasite and erythrocyte plasma membranes [155]. This so-called 'tight' (because it resembles a tight junction) or 'moving' junction forms a ring encircling the apex of the parasite. It is formed of parasite proteins which are inserted into the parasite plasma membrane (AMA1) and the erythrocyte plasma membrane (RON2 and associated RON proteins; Section 1.2.3.2). Thus, the parasite provides both components of the junctional complex, in both cell membranes. The moving junction is linked to the parasite's internal actin-myosin motor,

which powers the progressive movement of the moving junction backwards, away from the apex, and thus advances the parasite further into the host cell [156]. Meanwhile, the rhoptry bulb contents are released into the host cell, this lipid-rich stockpile probably incorporated into the developing parasitophorous vacuole membrane (PVM) [157]. The moving junction might also play a role in the formation of the PVM, and in determining its protein composition (which is primarily GPI-anchored or multi-pass transmembrane proteins, rather than single-pass, type I transmembrane proteins, at least in *Toxoplasma gondii*) [158, 159]. As the parasite enters the erythrocyte, its surface protein coat is shed by the action of proteases, namely subtilisin 2 (PfSUB2) released from the rhoptries [137, 160]. Finally, the parasite is completely enveloped in the parasitophorous vacuole inside the erythrocyte, and both the erythrocyte membrane and the PVM seal. The entire invagination process takes ∼17 s to complete [133].

1.2.3 Malarial proteins involved in invasion

Over fifty *Plasmodium* proteins have been implicated in the invasion process, through genomic, proteomic, or transcriptional analyses. Among them are the merozoite surface proteins, the moving junction proteins, and invasion ligands (including the RH and EBL protein families). The former are present on the cell prior to invasion, while the latter two classes are contained within the apical organelles, their release exquisitely timed during invasion.

1.2.3.1 Merozoite surface proteins

Proteins on the surface of the merozoite are plausible candidates to mediate the initial steps of invasion. It is logical that these coat proteins should act as initial ligands, rather than as more specific adhesins in the apical junction later in invasion, since they are distributed across the parasite surface; however, direct evidence that they function at this very initial stage is scarce, in part because the interactions are of such low affinity and difficult to reproduce *in vitro* [151]. Some are not evenly spread over the surface of the parasite, but rather enriched at the apex, and might reasonably be conjectured to serve a role in reorientation [143]. Consistent with a role in erythrocyte contact, coat proteins include domains which are commonly involved in protein–protein interactions, such as Duffy binding-like (DBL) domains (also present in the EBL family of proteins;

Section 1.2.3.4), epidermal growth factor-like (EGF) domains, and six-cysteine (6-Cys) domains [161]. Furthermore, some of these proteins are clearly functionally important or essential, as evidenced by a failure to knock out the corresponding gene, or the ability of antibodies raised against them to block invasion [142, 162].

Merozoite surface proteins can be divided into GPI-anchored proteins and peripherally associated molecules. There are nine GPI-anchored surface proteins, including the most abundant and most studied, merozoite surface protein 1 (MSP-1) [163]. MSP-1 has been implicated in attachment for invasion, perhaps through its interaction with Band 3 on the red blood cell [164], though the importance of this interaction has yet to be confirmed. A complicated series of cleavage events by proteases PfSUB1 and PfSUB2 generates a number of processed forms of MSP-1 during erythrocyte invasion, culminating in the formation of a 19-kDa, EGF-domain-containing C-terminal fragment which remains anchored to the membrane. These processed forms are important for formation of complexes with peripheral surface molecules (such as MSP-3, MSP-6, and MSP-7) [165]. Furthermore, MSP-1 has been demonstrated to be essential for parasite survival, and antibodies targeting processed fragments of MSP-1, particularly the 19-kDa fragment, block invasion *in vitro* [142, 166, 167]. MSP-2, another GPI-anchored protein, appears to be intrinsically disordered, and self-assembles into fibrils which could be a major component of the dense merozoite protein coat [168].

Peripherally associated proteins are grouped into three main families, based on their similarity to MSP-3, MSP-7, or the serine repeat antigen (SERA) protease-like family [169]. Which of these make direct contact with the red blood cell, and their cognate binding partners, are currently being investigated. Two members of the MSP-3 family, MSPDBL-1 and MSPDBL-2, contain an additional DBL domain relative to MSP-3 which binds to erythrocytes *in vitro*, and elicit invasion-blocking antibodies [170–172].

1.2.3.2 Moving junction proteins and the actin-myosin motor

Proteins involved in the moving junction and coordination of the actin-myosin motor have been dubbed ‘invasins,’ in contrast to the proteins which mediate adhesive contacts to the red blood cell (‘invasion ligands’ or ‘adhesins’; Section 1.2.3.3) [161]. The most studied among the putative invasins is apical membrane antigen-1 (AMA1), a microneme-derived, type I transmembrane protein localised to the moving junction. It

is expressed in all infective zoite forms of Apicomplexa, and is generally agreed to be essential [173]; however, a recent report asserts that AMA1 is dispensable for erythrocyte invasion by *P. berghei* merozoites, a finding which necessitates further study [174]. Furthermore, AMA1 is the target of invasion-blocking antibodies and a key vaccine candidate [173, 175]. Many possible functions for this protein have been proposed, including erythrocyte binding, apical reorientation, rhoptry secretion, and moving junction formation. It may signal downstream invasion events through the phosphorylation of its cytoplasmic tail by protein kinase A [148, 176]. Its potential role in moving junction formation—which has earned it the sobriquet 'invasin'—stems from its interaction with a set of proteins emanating from the rhoptry neck, known as the RON complex (RON2, 4, 5, and 8) [177]. The RON complex is inserted into the host erythrocyte by an unknown mechanism, with RON2 acting as a membrane-buried anchor and other members exposed to the host cell cytoplasm [155]. A direct contact between AMA1 and RON2, consisting of a loop of RON2 fitting into a hydrophobic groove in AMA1, brings together the two cell membranes [141, 178–180]. The importance of this interaction is underscored by the ability of anti-AMA1 antibodies, or peptide mimics of the RON2 loop, to block invasion [180, 181]. The AMA1-RON2 complex remains associated with the tight junction as it moves to the posterior [141].

The actin-myosin motor, which is responsible not only for invasion but also for the gliding motility of ookinetes and sporozoites (Section 1.1.5.1), is located between the parasite plasma membrane and the inner membrane complex, and consists of myosin A (a single-headed, Apicomplexan-specific class XIV myosin), an associated myosin light chain (called MTIP), and treadmilling, unusually short actin filaments [152]. How the moving junction and the internal actin-myosin motor are linked—in essence, how host cell recognition is coupled to motor function—is still a matter of debate. The current model for coupling and invasion is as follows [155]. Members of the microneme-derived *Plasmodium* thrombospondin-related anonymous protein (TRAP) family, including merozoite TRAP (MTRAP) and *Plasmodium* thrombospondin-related apical merozoite protein (PTRAMP), are released into the parasite plasma membrane, where they mediate adhesive interactions with host surface receptors. The acidic cytoplasmic tail of TRAP proteins may connect to aldolase, which in turn may be linked to filamentous actin [182]. How exactly (or indeed, if) the AMA1-RON complex interaction

is linked to the motor—whether AMA1 too binds to aldolase, or to some other cytoplasmic molecule—is unknown [155]. Myosin, which is tethered to the inner membrane complex through its interactions with gliding-associated proteins 45 and 50 (GAP45 and GAP50), then walks along the filamentous actin, propelling the bound adhesive proteins backward, and hence the parasite forward into the cell [183]. Imaging studies reveal an elegant ring of parasite actin following the moving junction to the posterior of the cell during invasion [141]. Another aspect of this process which has yet to be elucidated is a potential role of the host cell cytoskeleton—connected through the RON proteins—in movement. In support of some sort of role, host cell actin and actin-associated proteins such as Arp2/3 localise to the moving junction [184].

1.2.3.3 Invasion ligands

In contrast to the molecules with a clear or assumed role in the mechanics of invasion, another set of proteins mediates specific and crucial interactions with red blood cell receptors. These proteins are primarily derived from two protein families, the EBLs and the RHs [185]. They are variously described as 'invasion ligands,' 'adhesins,' or 'adhesion proteins' [161]. All we know for certain about the function of these proteins is that they are released from secretory organelles (the micronemes for EBLs, the rhoptries for RHs) to partner molecules on the red blood cell surface, and that these protein-protein interactions are central to the process of invasion [169, 185] (Figure 1.6). We do not know conclusively whether they exist solely as adhesives, with a great number of relatively low-affinity (0.1-1 μM) interactions creating a stable interaction with the red blood cell; or whether they trigger downstream events of invasion, perhaps through signalling pathways in parasite or erythrocyte (hence their multiple, somewhat non-descriptive monikers). However, recent studies have begun to point at a role in merozoite signalling. The rise in internal calcium after egress (Section 1.2.2) triggers microneme release, exposing EBL proteins (such as EBA-175) to their red blood cell receptors (glycophorin A in the case of EBA-175) [139]. These interactions, in turn, appear to restore intracellular calcium levels, and thus trigger the release of rhoptry proteins such as PfRH2b. This model suggests that the EBL proteins function first, followed by RH proteins—this fits with the differently timed release of micronemes versus rhoptries. Further, all *Plasmodium* species express at least one member of each family, strongly

suggesting that the the two families have distinct (yet important) functions. Super-resolution microscopy coupled with EBA-175 and PfRH4 knockouts has suggested a role for these proteins in apical attachment, tight junction formation, and rhoptry release into the host cell [141], and has led to the suggestion that RH and EBL proteins signal the 'commitment' to invasion (Section 1.2.2); what this means in molecular terms is only just beginning to be realised.

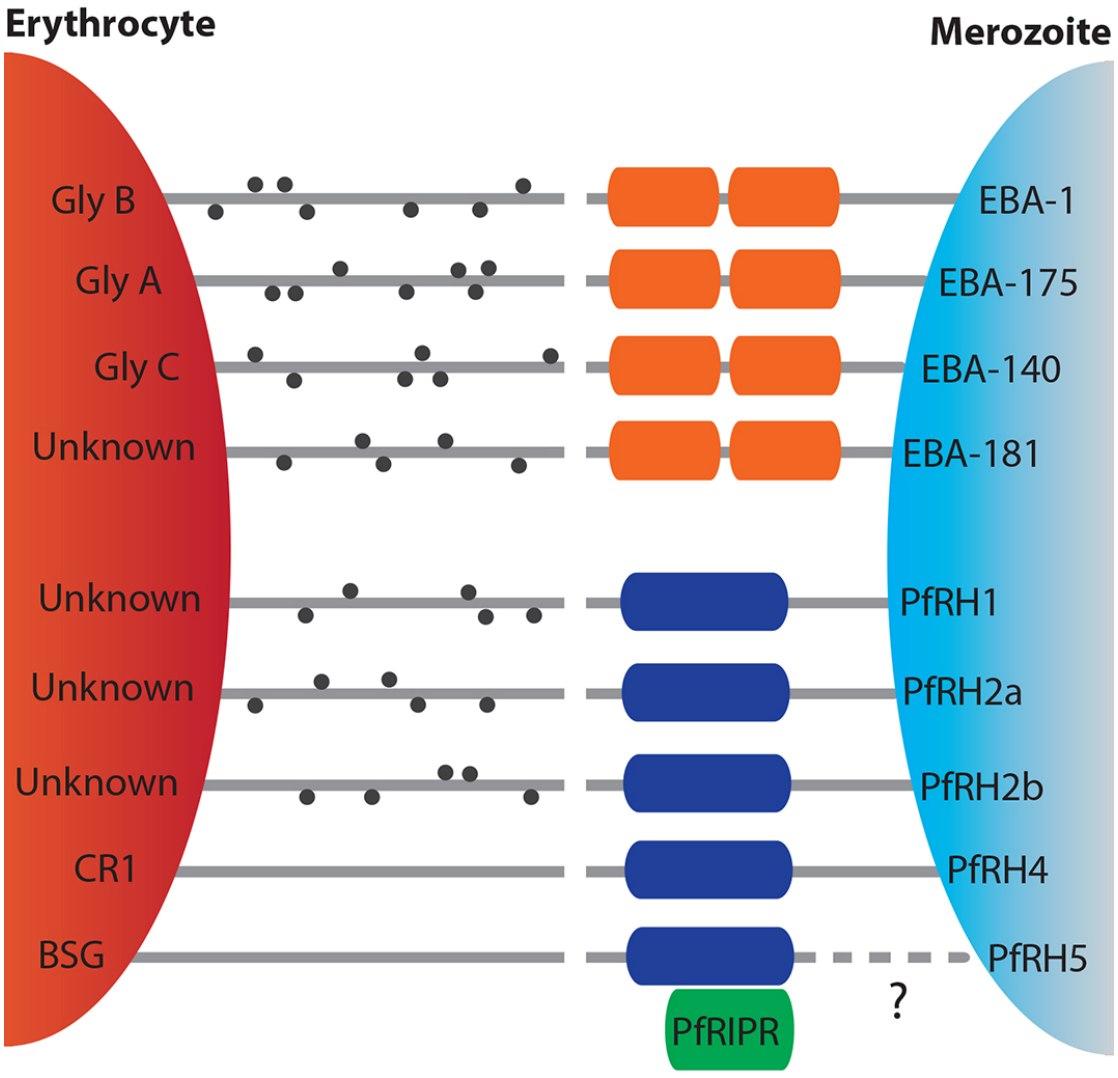


FIGURE 1.6: Merozoite invasion ligands of the EBL and RH protein families (right) bind to specific red blood cell receptors (left), some of which have yet to be identified. For those interactions which are dependent upon sialic acid moieties on the erythrocyte receptor, sialic acids are shown as dark grey circles. [Note that CR1 and basigin (BSG) are, however, glycosylated.] The DBL domains of the EBL proteins are shown as orange blocks, and the RH homology domain as a blue block. The PfRH5 binding partner, PfRIPR is shown in green; the dotted line indicates an unknown linkage of PfRH5 to the merozoite membrane. Adapted from [ref. 185].

Though the RH and EBL proteins are vitally important for invasion to occur, they

exhibit a marked functional redundancy. The genes encoding these proteins can all be disrupted in at least some parasite strains, with one notable exception, the *PfRH5* gene [134, 186, 187]. The genetic deletion of one of these ligands, or the treatment of erythrocytes with enzymes to remove or inactivate certain receptors, often results in a given parasite strain 'switching' to use alternative ligands to invade. This phenomenon was historically described as the merozoite utilising alternative 'invasion pathways' (i.e. invasion ligands or combinations of ligands), some of which, like EBA-175, were sialic acid-dependent (that is, neuraminidase treatment of erythrocyte surface receptors prevented invasion), and others of which, like PfRH4, were sialic acid-independent. In one of the most striking examples of functional redundancy, in some strains EBA-175 deletion results in increased transcription and protein expression of PfRH4 to substitute for the loss of EBA-175 [134, 188]. Similarly, when erythrocytes are treated with neuraminidase to strip the sialic acids from glycophorin A and ablate EBA-175 binding, parasites respond by upregulating PfRH4; but subsequently, upon prolonged exposure to untreated erythrocytes, the same parasites begin to silence PfRH4 expression again [188]. Thus, a given parasite strain can switch the ligands it uses for invasion. (This also suggests that, to some extent, the RH and EBL ligands can functionally complement each other, at least *in vitro*.) Different strains vary the combination of surface ligands they express, and hence the population of red blood cell receptors that will be recognised [161]. Differential expression and switching of invasion ligands are an effective means of immune evasion, and also a mechanism to tolerate erythrocyte surface receptor polymorphisms (many of which presumably arose through the evolutionary pressure applied by the parasite). The canonical example of erythrocyte polymorphism is Duffy antigen negativity, common in West Africa, which precludes *P. vivax* infection because of this species's reliance upon a single EBL family member (DBP) [11]. An evolutionary war is clearly occurring, in which the host is modifying its erythrocyte crenelations, and the parasite updating and expanding its arsenal of siege weapons.

There are other putative invasion ligands which do not belong to either the EBL or the RH family. For example, PfAARP is a predicted intrinsically disordered protein released from the rhoptries to the apical junction, where it is believed to form a specific interaction with an as yet unidentified erythrocyte receptor [189]. The following section will introduce the EBL family; the RHs will be discussed in further detail in Section 1.3.

1.2.3.4 The EBL protein family

All *Plasmodium* species contain at least one member of the EBL protein family. Notable examples include *P. vivax* DBP and the *P. falciparum* EBLs. Members of this family contain one (in the case of PvDBP) or two (in the case of the PfEBLs) N-terminal DBL domains, compact, cysteine-rich, and unique to *Plasmodium*, which mediate interactions with red blood cell receptors (Section 1.5). N-terminal to the DBL domain is a signal sequence for protein export to the cell surface, and C-terminal is a cysteine-rich region, followed immediately a transmembrane helix and a cytoplasmic domain (Figure 1.7).

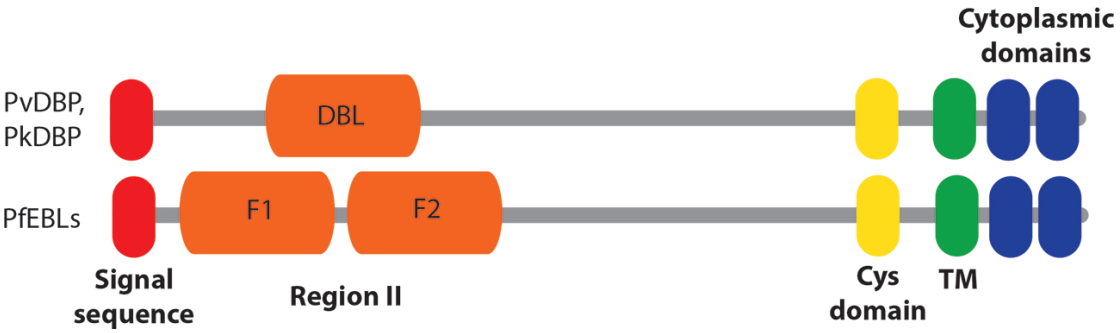


FIGURE 1.7: EBL proteins contain either one DBL domain (orange) in their erythrocyte receptor binding Region II (PvDBP and PkDBP) or two (domains I and II of PfEBLs). Additionally, they contain an N-terminal signal sequence (red), a cysteine-rich domain (yellow), a transmembrane helix (green), and cytoplasmic domains (blue). Adapted from [ref. 190].

The first identified member of the EBL protein family was Duffy-Binding Protein (DBP) from *P. vivax*, which interacts with the Duffy antigen receptor for chemokines (DARC) on the reticulocyte surface [11]. Merozoites of this parasite species fail to invade reticulocytes lacking DARC (found in Duffy negative individuals, primarily of African descent), seeming to stagnate after apical reorientation but before junction formation [149]. Surprisingly, however, *P. vivax* parasites capable of invading Duffy-negative erythrocytes were recently discovered in Madagascar, contraverting the notion that this invasion ligand–erythrocyte receptor combination is absolutely required, and suggesting that it can in some cases be bypassed [191].

The *P. falciparum* EBL protein family includes EBA-175, EBA-140 (BAEBL), EBA-181 (JESEBL), and EBL-1 [192]. These proteins interact with erythrocyte surface glycophorins; glycophorin A, B, and C make contact with EBA-175, EBA-1, and EBA-140, respectively [185]. In all three cases, interaction is dependent upon negatively-charged

sialic acid moieties decorating the surface of the glycophorins. (This is in marked contrast to the PfRH proteins, including PfRH4 and PfRH5, which do not require sialic acid modification of their red blood cell receptors to interact.) Individual EBL-glycophorin interactions are of moderate affinity (e.g. 260 nM for EBA-175–glycophorin A [193]). The glycophorins are abundant on the surface of red blood cells, with up to 1 million copies estimated per erythrocyte, suggesting that multiple EBL-glycophorin interactions and enhanced avidity may be important for invasion [185].

Invasion ligands EBA-175 and PvDBP have been suggested to undergo functional dimerisation upon receptor binding [190, 194], which, in turn, could instigate downstream signalling pathways. In the case of EBA-175, invasion-inhibitory monoclonal antibodies have been shown to interact both with the binding site for sialic acids, and with the dimerisation interface, implying that hindrance of dimerisation might contribute to invasion inhibition, along with blockade of ligand binding [195]. This hints at an important function of dimerisation for the EBL ligands. However, neither the crystal structure, nor any solution experiment, has proffered evidence that EBA-140 dimerises [196]. There is no evidence, as yet, that any of these proteins dimerise *in vivo*. Thus, further investigation will be required to determine the role and extent of dimerisation in this ligand family, as well as its downstream effects.

EBA-175 has been shown to be proteolytically processed during invasion. A rhomboid protease, PfROM4, cleaves EBA-175 within its transmembrane domain to shed the invasion ligand from the merozoite surface [197]. Why this occurs, and how it affects function and/or dimerisation, remains a mystery.

Before introducing the second of the two families of invasion ligands, the RH family, I will briefly discuss mechanisms by which the parasite evades host immune surveillance during erythrocyte invasion.

1.2.4 Parasite immune evasion strategies during invasion

One of the primary reasons for the lack of sterile immunity in exposed populations, and a major obstacle to the development of an effective vaccine, is the ability of the parasite to evade the host immune response. One of the primary mechanisms by which *P. falciparum* avoids detection is *var* gene switching and differential PfEMP1 surface expression once

entrenched within the erythrocyte (Section 1.1.5.2). During invasion, the merozoite is extracellular, albeit transiently, and thus vulnerable. The three main mechanisms by which immune evasion is achieved during invasion are sequence polymorphism, functional redundancy in the use of invasion ligands, and reduced antigenicity.

Many of the proteins involved in invasion—including the MSPs and AMA1—are extremely polymorphic, with multiple SNPs and allelic variants [198]. As a result, a single individual can be simultaneously infected with distinct populations of merozoites exhibiting different epitopes. In endemic regions, malaria-immune individuals contain high titres of antibody against MSPs, and yet they are not completely protected from infection, only from clinical disease, suggesting that the antibodies are strain-restricted because of MSP polymorphism [198, 199]. Similarly, attempts to develop an AMA1 vaccine have been frustrated by the fact that anti-AMA1 antibodies appear to be strain-specific, only protecting against strains homologous to the strain used to raise them [175]. Sequence polymorphism in AMA1 creates immunologically distinct molecules which nonetheless maintain their function in invasion.

The redundancy of RH and EBL family proteins (Section 1.2.3.3) is another mechanism of immune evasion. Merozoites can utilise alternative invasion ligands to invade erythrocytes, an observation predicated on the fact that different strains respond differentially to a change in the erythrocyte receptor repertoire (by enzyme treatment or by using erythrocytes with defined blood group variations) [198]. This adaptation not to rely on a single ligand, or a single RH/EBL combination, may be one of the parasite's answers to host antibodies and host polymorphism. The large repertoire of invasion ligands, as well as their differential importance and usage in different strains, make mounting an effective immune response a big challenge. Any vaccine against these proteins would need to be multivalent, incorporating many of these ligands, to have a chance of being strain-transcendent.

In contrast to the other RH and EBL proteins, PfRH5 does not appear to serve a redundant function, as it is essential in all strains so far tested [187]; nor does it seem to be under significant immune pressure, as it is minimally polymorphic [126, 129]. Antibodies raised against PfRH5 block invasion, but are rarely seen during natural infections [198]—however, a recent study has identified anti-PfRH5 antibodies acquired during natural infection which are linked to protection from severe disease [200]. Overall,

this has led to the hypothesis that RH5 is somehow minimally immunogenic during natural infection, but the mechanism by which this might be accomplished is unknown.

Limited duration of exposure of antigens may be a general mechanism to minimise either the induction or effector functions of the immune response. Indeed, the entire invasion process is complete, and the parasite safely hidden in an erythrocyte, in a matter of minutes. Moreover, the release of different populations of proteins from subcompartments of the rhoptries and subpopulations of the micronemes could allow exquisite temporal control of the invasion process, presumably only exposing potentially antigenic invasion ligands as they are required. Invasion commonly necessitates the association of two or more parasite proteins, which often emanate from different secretory organelles (e.g. micronemal AMA1 binds to rhoptry neck RON2 and, by extension, the remainder of the RON complex [155]; and rhoptry-derived PfRH5 binds to PfRIPR from the micronemes [202]). Purely hypothetically, such pairings could constitute an additional level of control over antigen exposure; for example, one parasite protein could shield a functional, erythrocyte-binding, or immunogenic surface of its partner until it is needed.

1.3 The RH protein family

1.3.1 Discovery

The binding of PvDBP to DARC, while crucial for invasion (in most cases; see Section 1.2.3.4), does not explain the tropism of *P. vivax* for reticulocytes over mature erythrocytes, since DARC is expressed on erythrocytes of all ages. A search was thus initiated for an upstream, 'host cell selection' step, before PvDBP–DARC engagement [192]. Two large, apically expressed proteins were identified that bound solely to reticulocytes, regardless of their Duffy phenotype [203]. They were named reticulocyte-binding proteins 1 and 2 (PvRBP-1 and PvRBP-2). Subsequently, homologues of these proteins were identified in *P. yoelii*, *P. falciparum*, and other *Plasmodium* species. In all species, these proteins originate from the rhoptries, and are released to the merozoite apex during erythrocyte invasion, where they mediate contacts with the red blood cell. Not all of their erythrocyte receptors have been identified to date; however, RH protein–receptor binding events have been probed indirectly through enzyme treatment of the erythrocyte surface. Neuraminidase has been used to remove sialic acid moieties, and trypsin and

chymotrypsin to digest protein, to see which components of the receptor are required for RH protein binding [198].

1.3.2 General architecture

The general features of these proteins, as deduced from the gene sequence, include an N-terminal signal sequence, a single transmembrane helix and a cytoplasmic domain (Figure 1.8). Between the signal sequence and the transmembrane helix lies an ectodomain of variable (but usually considerable) length. A ~500 amino acid region near the N-terminus constitutes the 'RH homology region,' a region of low sequence identity (16-20% across the RHs from *P. falciparum*, for example). PfRH5 is an exception among the RH proteins identified to date, in that it is much shorter than its homologues and is not membrane-linked (Section 1.4) [187]. An unrooted phylogenetic tree of *P. falciparum* and *P. vivax* RH proteins reveals two distinct clades, a PvRBP-1-like clade, and a PvRBP-2-like clade (Figure 1.9) [187].

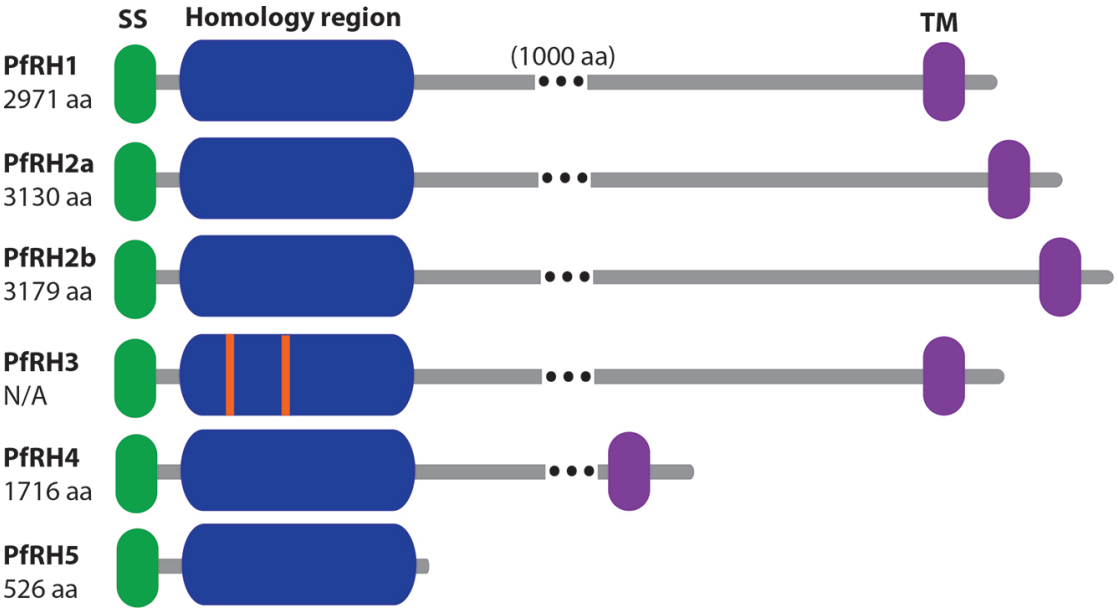


FIGURE 1.8: Architecture of the *P. falciparum* RH proteins (members of the family from other species are similar). The proteins are linked by a ~500 amino acid N-terminal region of low sequence identity (blue). They also contain an N-terminal signal sequence (green), and, in all but PfRH5, a C-terminal transmembrane helix and adjacent cytoplasmic domain. The ellipses indicate a 1000 amino acid spacer. PfRH3 is a pseudogene (orange lines denote frame-shifts). Adapted from [ref. 187].

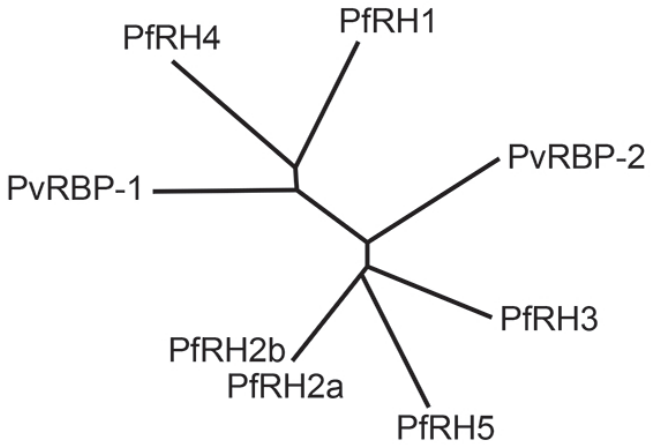


FIGURE 1.9: Unrooted phylogenetic tree, showing the relationship among *P. vivax* and *P. falciparum* RH proteins. They can be divided into a PvRBP-1-like clade (PvRBP-1, PfRH4 and PfRH1) and a PvRBP-2-like clade (PvRBP-2, PfRH2a/b, PfRH3, PfRH5). Adapted from [ref. 187].

1.3.3 RH proteins of *P. yoelii*

Before the discovery of PvRBP-1 and PvRBP-2, a novel 235-kDa rhoptry protein, Py235, had been identified in rodent-infective *P. yoelii*. Anti-Py235 monoclonal antibodies derived from mice resistant to the lethal YM parasite strain conferred protection in passive transfer experiments, by preventing YM invasion into mature erythrocytes and limiting the infection to reticulocytes [204]. Furthermore, immunisation of mice with Py235 protein or protein fragments induced a protective immune response that allowed the mice to survive a challenge with YM parasites [205]. Together, these experiments strongly indicated that Py235 plays a role in host cell selection and the virulence of the parasite [206]. When PvRBP-1 and PvRBP-2 were cloned, Py235 was quickly identified as a homologue of the *P. vivax* proteins, containing the characteristic RH homology domain [207, 208]. The RH protein family is thus evolutionarily ancient enough to exist both in human and rodent parasites.

The *P. yoelii* genome contains 14 copies of the *Py235* gene, indicating that there is a family of related Py235 proteins [192]. Interestingly, in a novel mechanism of clonal phenotypic variation, merozoites originating from a single schizont appear to have the ability to express distinct Py235 members, perhaps a survival strategy of the parasite to ensure successful host cell invasion [209]. Binding of a Py235 protein to mouse erythrocytes has been observed [208], and this binding appears to be sensitive to trypsin and chymotrypsin cleavage, but insensitive to sialic acid removal by neuraminidase,

suggesting binding requires a protein receptor but not any sialic acid decorations [210]. Of note, several Py235 proteins are expressed in sporozoites during hepatocyte invasion, and thus might be involved in this process; however, no RH protein from any other species has yet been implicated in sporozoite invasion [211].

1.3.4 RH proteins of *P. vivax*

Because these proteins appear to bind specifically to reticulocytes and not to mature erythrocytes [203], the RH proteins of *P. vivax* are believed to be responsible for the parasite's reticulocyte preference, through their interactions with as yet unknown, but presumably reticulocyte-restricted, receptors [206]. Completion of the *P. vivax* genome sequence unearthed additional members of the PvRBP family, and prompted a slight change in nomenclature [20, 212]: now the family contains seven members (PvRBP-1a, PvRBP-1b, PvRBP-2a, PvRBP-2b, PvRBP-2c, PvRBP-2d, and PvRBP-3). Of these, PvRBP-2d and PvRBP-3 are pseudogenes. More so than the other members of the family, PvRBP-2c (originally PvRBP-2) exhibits a high level of polymorphism, indicating that it is under significant selection pressure [212].

1.3.5 RH proteins of *P. falciparum*

The six currently known *P. falciparum* RH proteins, PfRH1, PfRH2a, PfRH2b, PfRH3, PfRH4, and PfRH5, were identified by bioinformatics (Figure 1.8) [169, 206]. PfRH3 is a pseudogene with frame-shifts at the 5' end of the gene; it is transcribed, and its mRNA processed, but is not translated [213]. PfRH1-4 will be described below; Section 1.4 is dedicated to PfRH5.

1.3.5.1 RH1

The *PfRH1* gene is found on chromosome 4, and the protein it encodes localises to the apex of the merozoite after release from the rhoptries [214]. The identity of its erythrocyte receptor is not known, but the binding event has been shown to be reliant upon sialic acids and resistant to either trypsin or chymotrypsin treatment [169]. A putative erythrocyte binding region was identified by expression of PfRH1 fragments on the surface of COS cells and assessment of red blood cell binding [215]. The region

observed to bind erythrocytes is near the N-terminus and encompasses residues 500-833 (out of over 2700); however, further verification will be required once the erythrocyte receptor has been discovered.

Targeted gene disruption of *PfRH1* changes the invasion ligand—receptor profile (from a 'sialic acid-dependent' to a 'sialic acid-independent' invasion pathway), and furthermore, PfRH1 is expressed at variable levels in strains utilising different invasion pathways [206, 216]. In contrast to the apparent importance of PfRH1 in *P. falciparum* invasion, however, the homologue of RH1 in *P. reichenowi*—a simian parasite closely related to *P. falciparum*—is a pseudogene [217]. This is a striking example of the incredible plasticity of this family, and the ability of parasites to vary their usage of individual family members to increase the chance of productive invasion [218].

Like EBA-175 and many other invasion ligands, PfRH1 undergoes proteolytic processing during invasion. The ∼360 kDa protein is cleaved into a soluble 240 kDa fragment and a 120 kDa membrane-bound fragment, both of which are localised to the merozoite apex [219]. Some molecules of PfRH1 are also cleaved in the transmembrane domain, perhaps the result of PfROM4 action, while others appear to remain membrane-bound and to be carried into the red blood cell [219].

1.3.5.2 RH2a and RH2b

Comparative analysis with *PvRBP-2* identified two related genes on *P. falciparum* chromosome 13, which encode proteins PfRH2a and PfRH2b [220]. These genes, perhaps the products of a gene duplication event, are identical for the first 8 kb but have divergent 3' ends, corresponding to the final 500 amino acids of the protein, including part of the ectodomain, the transmembrane region, and the cytoplasmic tail [221]. The resultant PfRH2a and PfRH2b proteins are over 350 kDa, and are partitioned in the rhoptry neck [186]. Like PfRH1, they are apically localised during invasion but their specific erythrocyte receptor(s)—mediating invasion through a chymotrypsin-sensitive, trypsin-insensitive, and sialic acid-independent mechanism—is unknown [221]. Different groups have proposed overlapping N-terminal regions of PfRH2a/b (residues 446-557 [222] or residues 495-860 [221]) to contain the erythrocyte binding region. Antibodies raised against these regions (recombinantly expressed in *E. coli*) reduce efficiency of invasion into enzyme-treated erythrocytes in growth inhibition assays [221, 222]. Moreover, the

N-terminus of the protein contains the highest concentration of polymorphisms, implying diversifying selection as a result of immune pressure [223].

Different isolates express *PfRH2a/b* to differing extents, hinting at a strategy of phenotypic variation to evade the immune system and cope with erythrocyte receptor polymorphism [186]. Antibodies against PfRH2a/b partially inhibit invasion *in vitro*, and also seem to play a role in natural infections: high levels of these antibodies were correlated with protection from symptomatic malaria and high-density parasitaemia [223]. Targeted disruption of the *PfRH2b* gene has a marked effect on the ligands used for invasion, as demonstrated by a change in enzyme sensitivity; however, disruption of *PfRH2a* in the same strain has no effect upon ligand usage (though in other strains it could not be knocked out) [186]. The different relative importance of PfRH2a versus PfRH2b in a particular strain is surprising given that most of their vast expanse of ectodomain is identical. This phenomenon may derive, at least in part, from their different cytoplasmic domains, a notion supported by experiments with chimeric parasites [154]. Involvement of the cytoplasmic domain hints tantalisingly at a possible signalling mechanism in the merozoite during invasion, though any corroborating evidence is thus far lacking.

PfRH2a and PfRH2b are also enzymatically processed, both in the ectodomain to release a soluble ∼85 kDa N-terminal fragment in the schizont stage, and in the transmembrane domain (potentially by PfROM4) during invasion [221, 222]. It has been proposed, based on studies with enzyme-treated erythrocytes, that several of these fragments (not just those containing the proposed N-terminal erythrocyte binding site, above) are capable of binding erythrocytes with different enzyme sensitivities, suggesting interactions with distinct erythrocyte receptors, though this requires more rigorous verification [221, 224].

1.3.5.3 RH4

PfRH4 is on chromosome 4, and encodes a ∼220 kDa, rhoptry-derived and apically localised protein [225]. Like its counterparts, PfRH4 is processed during invasion to a ∼160 kDa soluble fragment [226]. PfRH4 upregulation (and switch to a 'sialic acid-independent pathway' [227]) is observed following EBA-175 knockout [188], and antibodies against the PfRH4 binding region have been shown to block erythrocyte binding [228] and invasion [226] of neuraminidase-treated erythrocytes, testament to its role in certain invasion pathways. Antibodies against PfRH4 are acquired during natural infection,

and high levels of such antibodies are strongly linked to protection from symptomatic malaria [229].

PfRH4 was the first RH protein to have its erythrocyte receptor, complement receptor 1 (CR1), identified [230]. The CR1-binding region of PfRH4 was narrowed down to an N-terminal fragment consisting of residues 28-766, or even more minimally, 104-465 [226, 230]. The two bind with a dissociation constant of 2.9 μM. CR1 is a complement regulator, which, among other functions, ferries C3b/C4b-opsonised particles to the liver and the spleen to be phagocytosed [231]. The most common isoform of CR1 is divided into 30 complement control protein modules (CCPs), also known as sushi domains or short consensus repeats, ∼60-70 residues long. The binding site for PfRH4 was mapped first to CCP1-3, and then to CCP1 alone, found at the N-terminus of CR1, distal from the membrane [232]. PfRH4 binding to CCP1-3 does not interfere with the weak binding of C3b or C4b in this region, and a ternary complex of PfRH4/CR1/C4b has been observed [233]. In contrast, PfRH4 does seem to compromise CR1's decay accelerating activity of C3 and C5 convertases. However, the role of complement components in malaria infection remains unclear, as the addition of C3b/C4b has no effect on invasion *in vitro* [233], and invasion proceeds with equal efficiency in normal and heat-inactivated (and hence complement-inactivated) serum [140]. (Worth mentioning, a distinct region of CR1 (CCP10 and CCP17) is also a PfEMP1 receptor involved in rosetting, the attachment of infected erythrocytes and non-infected erythrocytes [234].)

1.4 *P. falciparum* RH5

Bioinformatic scrutiny of the *P. falciparum* genome and a genetic cross between two *P. falciparum* strains revealed a gene near *PfRH4* on chromosome 4, which codes for another, atypical member of the RH protein family [187, 235, 236]. Christened PfRH5, this ∼63 kDa protein is a much abridged version of its homologues, comprised of the N-terminal signal sequence and the RH homology region, but lacking the remainder of the ectodomain, the transmembrane helix, and the cytoplasmic domain (Figure 1.8). No RH5 orthologues have been found in other human-infective or rodent-infective malarias; only African ape–infective species in subgenus *Laverania* (e.g. *P. reichenowi*) appear to contain PfRH5 [187]. PfRH5 emanates from the rhoptry body and localises to the merozoite apical membrane, where it remains peripherally attached by interactions with

unknown proteins [187, 202, 237]. It also seems to undergo proteolysis from the N-terminus during invasion, to form a ∼45 kDa fragment [187, 202]. Both full-length PfRH5 and the 45 kDa processed fragment are capable of binding red blood cells; indeed, one study suggests that the 45 kDa fragment binds at higher levels in erythrocyte binding assays [187]. An erythrocyte-binding 28 kDa fragment has also been reported [235].

PfRH5 is expressed in all strains tested [187]. Significantly, unlike all the other *PfRHs* and *PfEBLs*, the *PfRH5* gene has not been successfully disrupted in any strain so far, implying that its protein product might be universally essential [187]. Presumably, therefore, PfRH5 must play a distinct role in invasion compared to the other RH and EBL ligands, though its function (or that of the other invasion ligands, for that matter) is unknown. Antibodies against PfRH5 appear to inhibit tight attachment of merozoites to erythrocytes, suggesting a role in this process [237].

1.4.1 Interaction with basigin

Erythrocyte basigin was identified as the red blood cell receptor for PfRH5 using an avidity-based extracellular interaction screen (AVEXIS) [238]. Also known as extracellular matrix metalloproteinase inducer (EMMPRIN), CD147 or BSG, basigin is a type I integral membrane protein and an immunoglobulin (Ig) superfamily member with a wide tissue distribution. Its interactions with other proteins contribute to such diverse biological processes as spermatogenesis, fertilisation, neural network development, tumour metastasis and angiogenesis, and human immunodeficiency virus infection [239]. Basigin has been linked to integrins, monocarboxylate transporters, and cyclophilins, and many other binding partners [239, 240]. Interestingly, basigin was recently identified as the endothelial adhesion receptor for *Neisseria meningitidis* type IV pili [241].

Basigin consists of either two (short isoform) or three (long isoform) Ig domains. The crystal structure of the short isoform reveals two canonical Ig domains, joined by a 5-residue flexible linker (Figure 1.10A). The flexible linker permits these two Ig domains to change their orientation relative to each other (Figure 1.10B). Both domains of the short isoform (but not the third Ig domain of the long isoform) are required for PfRH5 binding, and the short isoform is thought to predominate on the red blood cell surface [238]. Binding of PfRH5 is not dependent on glycan decorations on basigin, as their removal by deglycosylation or mutation of predicted glycosylation sites has no effect on

the interaction [238]. Five single nucleotide polymorphisms (SNPs) in the short isoform of human basigin have been described; of these, two have an effect on PfRH5 binding. One (L90P) causes misfolding of the N-terminal Ig domain, and another (E92K, on the loop connecting strands F and G in the N-terminal Ig domain) leads to a two-fold lower PfRH5:basigin affinity [238]. (E92K is the distinguishing mutation of the Ok^{a-} blood group, seen in eight Japanese families—this polymorphism is presumably not a major mechanism of malaria resistance!) The addition of soluble basigin or anti-basigin antibodies, or lentiviral knockdown of erythrocyte basigin, potently blocks invasion *in vitro*, underscoring the importance of the PfRH5:basigin interaction [238].

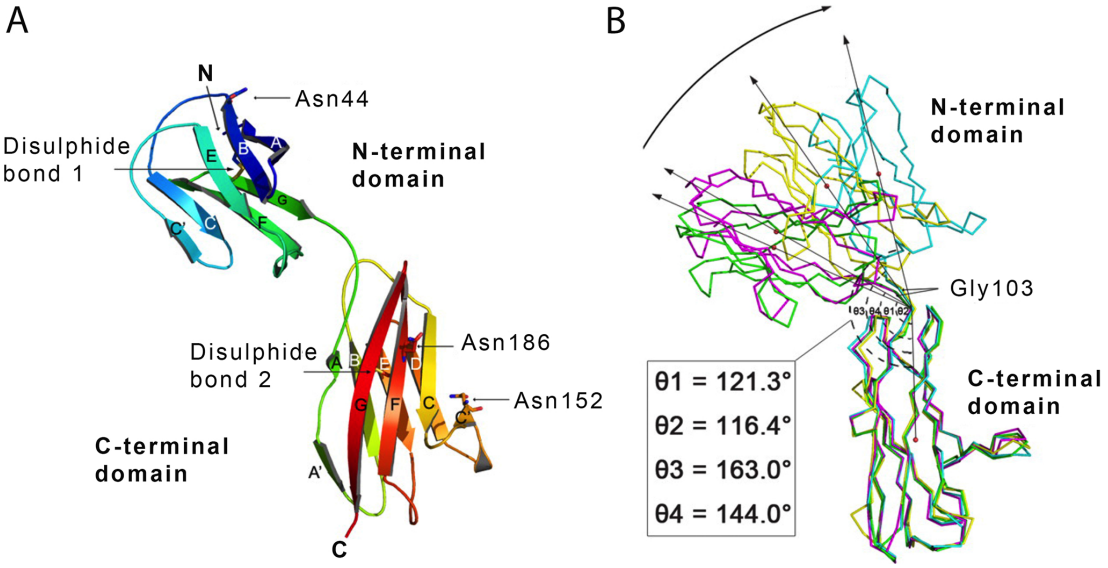


FIGURE 1.10: (A) The basigin crystal structure, coloured in rainbow from N-terminus (blue) to C-terminus (red) consists of two Ig domains connected by a 5-residue linker. Two disulphide bonds and three potential glycosylation sites (Asn44, Asn186, and Asn152) are shown as sticks. (B) Comparison of the four chains in the crystal structure, prepared by superimposing the C-terminal domain. Flexibility in the linker is evident in the different positions of the N-terminal domains relative to the C-terminal domains, and the different 'elbow' angles (θ1-θ4). Gly103 is one of the residues in the linker. Taken from [ref. 239].

Analysis of sequence variants of human and primate basigin, and PfRH5 from various parasite strains, can identify determinants of host tropism—the ability of parasites to invade erythrocytes of different species. *P. falciparum* is occasionally able to infect chimpanzees, and never able to infect gorillas; strikingly, the ability of PfRH5 to bind chimpanzee basigin (weak binding) or gorilla (no binding) exactly recapitulates this infection pattern [242]. Comparison of the sequence of human basigin with that of chimpanzee and gorilla basigin identified only very few polymorphic residues, some of which (including F27, Q100, +H103, and K191 of the human sequence) were correlated

with ability to bind PfRH5 [242]. These may be located in or near the PfRH5 binding site.

PfRH5 is minimally polymorphic, with only twelve non-synonymous SNPs found in 227 field isolates, and only five at frequencies of 10% or greater [126, 128, 129] (Section 1.4.3). Sequenced laboratory strains exhibit eight distinct SNPs that are associated with increased ability to invade New World monkey (genus *Aotus*) erythrocytes, sometimes in combination [235, 243]. These residues (namely, 48, 197, 204, 347, 358, 362, 407, and 429) may be important for basigin binding, though of course further validation will be required.

1.4.2 Interaction with PfRIPR

Recently, purification of PfRH5 from parasite culture revealed the presence of a parasite binding partner, accordingly named *P. falciparum* RH5 interacting protein (PfRIPR) [202]. PfRIPR is ∼123 kDa, with 10 predicted EGF domains and 87 cysteines (Figure 1.11). It is localised to the micronemes in the schizont stage, but contains an N-terminal signal sequence to promote its release. Since PfRH5 abides in the rhoptries, these proteins do not meet until they are released during invasion [202]. Like many proteins involved in invasion, PfRIPR appears to be proteolytically processed, between the second and third EGF domain to form two ∼65 kDa fragments which remain associated with each other [202]. The PfRIPR complex is capable of binding both full-length PfRH5 and its 45 kDa cleaved component. As neither PfRH5 nor PfRIPR contain a transmembrane domain, yet together are peripherally associated with the membrane [202], there must be an additional component, or additional components, of the complex that mediate merozoite membrane anchoring. The PfRIPR–PfRH5 complex might dissociate upon encountering erythrocyte basigin, however, as PfRIPR does not appear to attach to red blood cells [202]. Antibodies targeting PfRIPR potently block merozoite invasion, and the *PfRIPR* gene has proved refractory to disruption. While the importance of PfRIPR in invasion has thus been demonstrated, its specific role is less clear. Does it function to shield the basigin binding site or immunologically vulnerable surfaces of PfRH5 from immune attack? Is it a simply a chaperone stabilising the fold of PfRH5? Does it have a role in signalling in the merozoite? Further experiments will be required to begin to answer these questions.



FIGURE 1.11: PfRIPR is 123 kDa (1086 amino acids) and contains 10 EGF domains (purple). The N-terminal signal sequence is shown in blue. Proteolysis is thought to occur between the second and third EGF domain. Adapted from [ref. 202].

1.4.3 Immunogenicity and suitability as a vaccine candidate

Since its discovery, PfRH5 has received a lot of attention as a potential blood-stage malaria vaccine target. As mentioned above, it is unique among the RHs and EBLs in that it cannot be knocked out, and is expressed in every strain so far tested [187]. Moreover, unusually for malaria invasion proteins, it is remarkably conserved, with only five common non-synonymous SNPs [126, 128, 129]. Antibodies against either PfRH5 [126–128, 237, 244] or basigin [238] potently block invasion, emphasising the importance of PfRH5, and its interaction with basigin in particular. Crucially, antibodies raised against one PfRH5 variant neutralise parasites of all tested heterologous strains expressing PfRH5 with the full gamut of SNPs [126, 127]. In one study, which compared ten blood-stage vaccine candidates, including front-runners PfAMA1 and PfMSP-1, the antibodies induced by PfRH5 were by far the most effective at neutralising parasites; furthermore, anti-PfAMA1 antibodies failed to confer the strain-transcendent protection seen for anti-PfRH5 antibodies [127]. Antibodies against PfRH5 can also act synergistically with antibodies against other merozoite antigens, most notably PfRH4 or PfRH2a/b [128], raising the possibility of a multivalent vaccine incorporating PfRH5 or part of PfRH5.

Invasion-inhibitory anti-PfRH5 antibodies were generated by immunisation of rabbits with either viral vectors expressing full-length PfRH5 [127], or by full-length PfRH5 recombinant protein [126, 244]. Initial attempts to immunise with two different fragments of PfRH5 were not successful: a bacterially expressed fragment comprised of residues 31-174 raised no antibodies capable of inhibiting invasion [236], and a viral vector expressing a fragment of residues 191-359 led to only very modestly inhibitory antibodies [127]. However, it is unclear whether these constructs were correctly folded. High-resolution structural information could significantly aid efforts to generate a smaller fragment of

PfRH5 capable of inducing a comparable, or better, inhibitory immune response relative to the full-length protein.

The finding that vaccine-induced anti-PfRH5 antibodies are capable of blocking invasion launched a search for similar antibodies naturally acquired during human infection. An early study failed to find any antibodies recognising PfRH5 in malaria-exposed Kenyans; coupled with its low level of polymorphism, this lead to the suggestion that PfRH5 is not inherently immunogenic *in vivo*, perhaps due to limited exposure during invasion (Section 1.2.4) [127]. However, more recently, anti-PfRH5 antibodies have been identified in malaria-exposed Malian children and adults [200]. The level of anti-PfRH5 antibodies strongly correlates with clinical outcome, and these antibodies can inhibit parasite growth *in vitro*. How widespread such antibodies are, and how big a role they play in malaria infections worldwide, is unknown.

A number of monoclonal antibodies have now been raised against PfRH5, many of which (so-called QA1, QA5, 6BF10, 9AD4, and 2AC7) are inhibitory of invasion in *in vitro* growth inhibition assays [237]. Of the inhibitory monoclonal antibodies, several (QA1, QA5, and 6BF10) can directly interfere with the PfRH5–basigin interaction in a protein–protein interaction assay. The others (9AD4 and 2AC7) do not appear to directly impede basigin binding *in vitro*, but they might sterically interfere with the interaction during invasion, when both PfRH5 and basigin are immobilised on their respective cell surfaces. Further antibodies were identified, called RB3 and 4BA7, which proved not to be invasion-inhibitory. By assessing the competition between these various monoclonal antibodies for binding to PfRH5, a model for the arrangement of the epitopes on the surface of PfRH5 was generated [237] (Figure 1.12). Further, the epitopes of several monoclonal antibodies were tentatively mapped using overlapping PfRH5 peptides [237]. Altogether, these results promote PfRH5 as an extremely promising vaccine candidate.

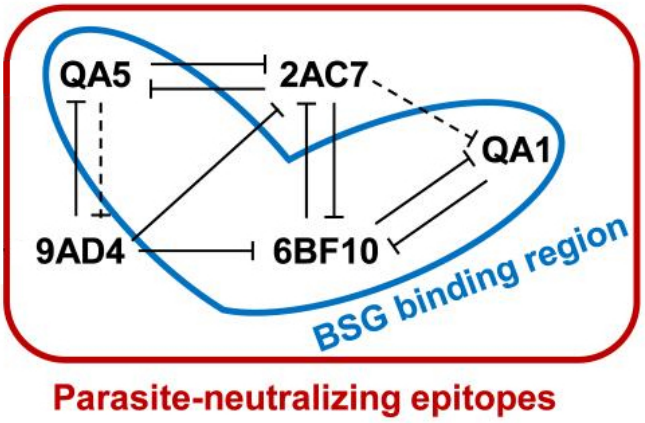


FIGURE 1.12: Model of the arrangement of PfRH5 epitopes recognised by invasion-inhibitory monoclonal antibodies. Solid lines indicate complete competition between the two antibodies for PfRH5 binding, dotted lines indicate partial competition, and the absence of a line indicates no competition. All of the epitopes are associated with parasite neutralisation (red outline), and some are in the putative basigin binding site (blue outline). Not shown are the non-inhibitory antibodies 4BA7 and RB3. Taken from [ref. 237].

1.5 Structures of *Plasmodium* surface proteins

Malaria surface proteins are comprised of a variety of modules to mediate adhesion to host cells. Some of these—such as thrombospondin type I repeats (TSP-I), PAN domains, and EGF domains—are also found in proteins with adhesive functions in higher organisms. Others—most notably the DBL domain and the 6-cys domain—are unique to subsets of the Apicomplexans, or *Plasmodium* in particular. Some proteins appear to lack structure, or, in the case of the RH proteins, to adopt a structure which has not yet been determined.

The DBL domain is a prominent, cysteine-rich host-binding module unique to *Plasmodium*, present both in the EBL family of erythrocyte invasion ligands, and in PfEMP1 surface protein variants responsible for cytoadhesion. EBL family members contain either one (PvDBP and PkDBP) or two (PfEBLs) DBL domains, while PfEMP1s contain a string of DBLs interspersed with cysteine-rich interdomain regions (CIDRs) [190]. The ten structures of DBL domains solved to date have revealed a conserved core fold, in spite of low sequence identity, which contains three subdomains (Figure 1.13A) [245]. Subdomain 2 forms a four-helical bundle around which subdomain 1 drapes itself, in a manner not conserved across the domains. Subdomain 3 is packed against subdomain 2, and consists of two long helices and a shorter, partial helix, stapled together

by three disulphide bonds. Three additional disulphides are contained in subdomains 1-2. Those few residues which are conserved (only ∼4%) or similar (∼15%) are directed inwards, stabilising the core of the fold [245]. The exposed faces of the helices exhibit significant diversity, and additional loops and helices decorate various DBL domains. Thus, the conserved core acts as a scaffold for presentation of highly variable surfaces, allowing these domains to bind a variety of cell surface receptors, and to circumvent immune detection. Interestingly, in DBP, amino acids found to be highly polymorphic in field isolates (i.e. amino acids presumably under immune pressure) map onto the opposite side of the structure from the DARC binding site [246]. This is in keeping with 'just-in-time' release of DBP from the micronemes, immediate engagement of the DARC receptor, and burying of the DARC binding site away from host antibodies. For at least some of the EBL family of invasion ligands, including EBA-175, dimerisation seems to occur upon receptor binding, which in turn may have functional consequences yet to be determined [190, 194] (Figure 1.14).

Other *Plasmodium* surface protein modules include the EGF domain and the 6-cys domain. The proteolytically processed C-terminal fragment of MSP-1 consists of two EGF domains which pack together via a hydrophobic interface, and is linked to the membrane via a GPI anchor (Figure 1.13C) [248, 249]. PfRIPR is predicted to contain ten EGF domains [202] The 6-cys (or s48/45) domain is found in surface proteins of many *Plasmodium* lifecycle stages, including sporozoites, gametes, and merozoites, and is related to *T. gondii* surface molecules in the surface antigen related sequence (SRS) superfamily [250, 251]. The 6-cys structure is exemplified in *Plasmodium* blood-stage protein Pf12, and consists of a compact β-sandwich, formed of two sheets of mixed parallel and anti-parallel strands, joined by two of the three disulphides [251].

The erythrocyte binding domains of AMA1 (domains I and II) are mixed α-β modules called PAN (or apple) domains [247]. A PAN domain contains a five-stranded β sheet which wraps around a twelve-residue helix. Strikingly, AMA1 contains long loop extensions emanating from the PAN domain core, which could act as immune decoys to divert the antibody response (Figure 1.13B). The most highly polymorphic residues of AMA1 ring the conserved hydrophobic cleft, the binding pocket for RON2 [179, 247].

Other domains found in microneme proteins were identified based on vertebrate homologues. The short TSR-I repeat is found in multiple copies in, for example, TRAP and

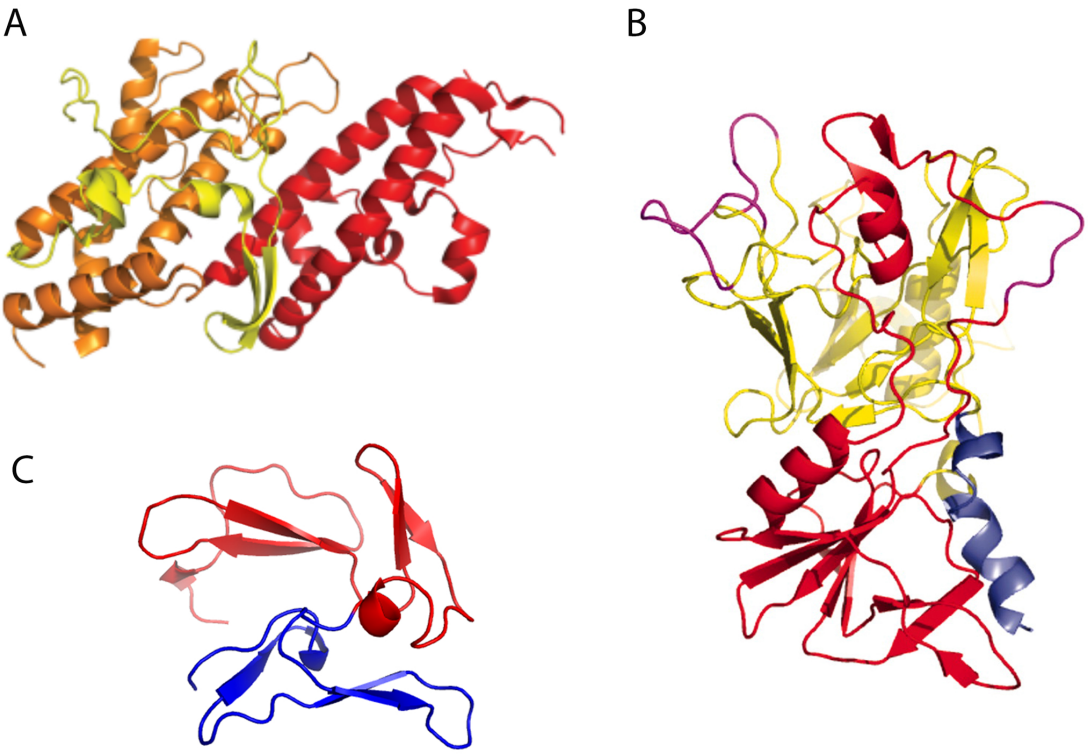


FIGURE 1.13: (A) The structure of a DBL domain, here the DBL3X domain from var2csa (PDB code: 3BQI), revealing three subdomains, subdomain 1 (yellow), subdomain 2 (orange), and subdomain 3 (red). Taken from [ref. 245]. (B) The structure of AMA1 (PDB code: 1Z40), consisting of two PAN domains, domain I (yellow) and domain II (red). Disordered loops are coloured in violet, and an N-terminal extension helix in blue. Taken from [ref. 247]. (C) The structure of the C-terminal fragment of MSP-1 (PDB code: 1B9W), with two EGF domains coloured in red (N-terminal) and blue (C-terminal). Image created using Pymol.

CSP [252]. Integrin insertion (I)-domains (also known as von Willebrand factor (vWF) type A domains) are predicted in several *Plasmodium* proteins (including TRAP), and consist of a 200-amino acid, mixed α/β module which harbours a divalent metal-binding site [253].

Other parasite surface proteins are defined by their lack of structure. Apicomplexans in general and mammalian-infective *Plasmodium* species in particular are predicted to have unusually high numbers of intrinsically disordered proteins (IDPs) and long repetitive sequences [254]. Such disordered proteins could make interactions with host proteins to facilitate invasion, while preventing the development of efficient host immunity. One such protein is PfAARP, a merozoite ligand which has been implicated in erythrocyte invasion, seeming to bind to a specific, but currently unidentified, cell surface receptor [189]. Antibodies raised against PfAARP can hinder invasion, indicating that at least partially efficacious antibodies can be raised against such proteins [189].

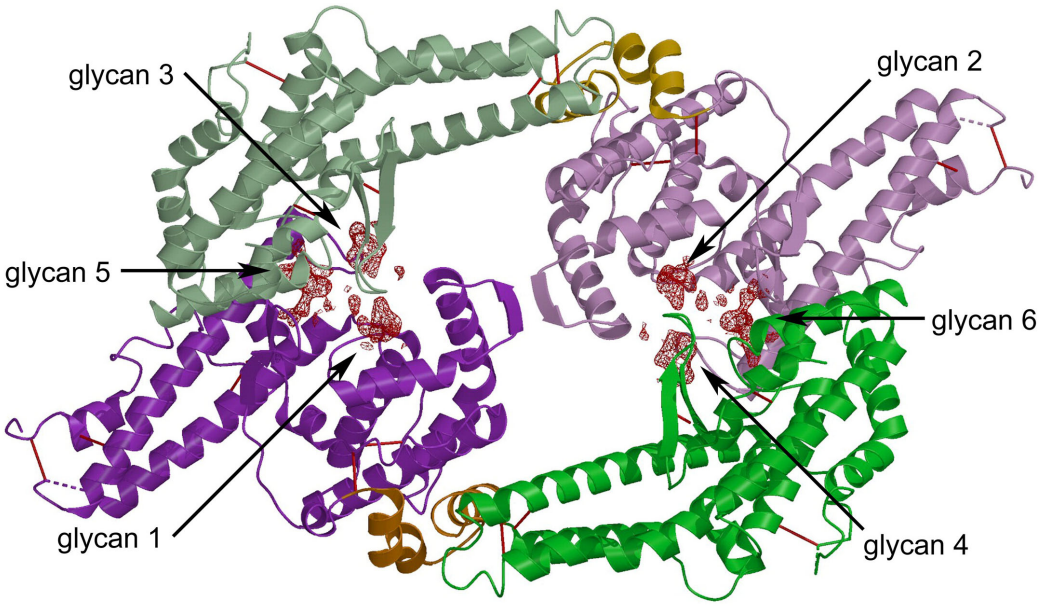


FIGURE 1.14: The structure of the EBA-175 RII domain was solved in the presence of α-2,3-sialyllactose, analogous to the sialic acid moieties on its red blood cell binding partner, glycophorin A, that are requisite for binding (PDB code: 1ZRO). EBA-175 RII consists of two tandem DBL domains, F1 (green or light green) and F2 (purple or light purple), joined by a short helical linker (gold or light gold). Two EBA-175 molecules join to form a dimer, with the six glycan binding sites at the dimer interface (labeled glycans 1-6). Thus, dimerisation presumably occurs only upon glycophorin A binding, and may have a functional role in subsequent invasion events. Taken from [ref. 190].

As more structures of parasite surface proteins are determined, broad themes are beginning to emerge. A relatively few domains are exploited repeatedly, with their few similar residues maintaining the (often α-helical, but sometimes β-sheet-containing) fold, and their extreme surface variation yielding them diversity of function and ligand-binding capabilities [245]. These domains can latch on to their host cell partners tightly ($\sim$nM range, as in the case of PfEMP1 domains mediating cytoadhesive interactions), or weakly ($\sim\mu$M range, as in the case of RH proteins and AMA1 making red blood cell contacts but potentially also fulfilling other roles during invasion). Their antigenic diversity, meanwhile, foils the immune system and prevents immunoglobulin binding. Extended loops and disordered segments could act as shields or decoys to protect important residues from immune recognition. Dimerisation may be of functional significance, at least for some members of the EBL family of invasion ligands.

1.6 Project aims

Our understanding of *Plasmodium* surface protein structure is limited to the few domains—like the DBL domain, the PAN domain, and the 6-cys domain—that have been visualised at high resolution. There is a complete dearth of structural information on the RH protein family, which, together with the DBL-containing EBL family, constitute the two crucial families mediating erythrocyte invasion across the *Plasmodium* species. Standing out among the EBL and RH invasion ligands is *P. falciparum* RH5, the only demonstrably essential protein. Antibodies that bind to PfRH5, and potentially prevent it from binding its erythrocyte receptor, basigin, can potently inhibit erythrocyte invasion. These findings have generated intense excitement about PfRH5 as a next-generation blood-stage malaria vaccine target and emphasised the need for structural information to guide rational immunogen design.

The aim of my D.Phil. project was, firstly, to solve the high-resolution crystal structure of PfRH5, to determine its fold, and to assess whether it is likely to be conserved across the other *Plasmodium* RH proteins. The second goal was structural characterisation of the PfRH5 interaction with basigin, to determine the molecular basis of erythrocyte binding by PfRH5. Finally, I set out to perform structural and biophysical analysis of PfRH5 in complex with the monoclonal antibodies that block erythrocyte invasion and parasite growth, to identify and analyse the PfRH5 epitopes that correlate with protection, towards the design of a new generation of PfRH5 immunogens and a possible vaccine.

Chapter 2

Methods

2.1 Bioinformatic analysis of PfRH5

PSI-BLAST [255] and FUGUE [256] searches were used to predict homologues of, or recognisable domains in, PfRH5. Disorder predictions were carried out using IUPred [257, 258], Disprot VL3H [259], and POODLE-L [260]. Potential helical regions were predicted using Prof [261], Porter [262], PSIPRED [263], and Jpred [264]. The predictions were compared to look for consensus. The SignalP 4.0 server [265] was used to generate signal peptide predictions. Sequence alignments were generated using the CLUSTAL Omega alignment package [266] or the T-COFFEE server [267]. Alignment figures were generated with Espript [268]. Structure-based threading was performed with the Phyre2 server [269].

2.2 Molecular biology and cloning

2.2.1 Genes and constructs

The basigin gene was amplified from human cDNA. Over the course of the project, two different synthetic genes encoding full-length PfRH5 were used. One, used for expression in *E. coli*, baculovirus-infected insect cells, and HEK cells, was optimised for mammalian cell expression, and had four putative N-linked glycosylation sites mutated to glutamine (S. Draper). The other, synthesised later, was optimised for, and used for,

Vector	Expression system	Resistance	N-tags	C-tags	5' RS	3' RS	Source
pET15b$_{con}$	*E. coli*	Amp	6-His–TEV	None	BamHI	NdeI	[270]
pLou3	*E. coli*	Amp	6-His–MBP–TEV	None	BamHI	HindIII	S. Newstead
pRSFDuet-1	*E. coli*	Kan	6-His	None	BamHI	HindIII	EMD Millipore
pOPINF	*E. coli*, Sf9, HEK	Amp	6-His	None	N/A	N/A	OPPF
pOPINM	*E. coli*, Sf9, HEK	Amp	6-His–MBP	None	N/A	N/A	OPPF
pOPINS3C	*E. coli*, Sf9, HEK	Amp	6-His–SUMO	None	N/A	N/A	OPPF
pOPINTRX	*E. coli*, Sf9, HEK	Amp	6-His–TRX	None	N/A	N/A	OPPF
pOPINFB	*E. coli*, Sf9, HEK	Amp	6-His	Bap	N/A	N/A	OPPF
pExpreS^2.1$_{con}$	S2	Zeo	BiP leader	None	EcoRI	NotI	ExpreS^{2}ion Bio
pAcGP67A$_{con}$	Sf9	Amp	gp67ss–6-His–TEV	none	BamHI	NdeI	M. Higgins
pHLsec	HEK (transient)	Amp	Secretion signal	6-His	AgeI	KpnI	[271]
pEFbos$_{con}$	HEK (transient)	Amp	6-His–TEV	None	BamHI	NheI	M. Higgins

TABLE 2.1: Summary of expression plasmids. Subscript 'con' indicates that the multiple cloning site has been altered. Abbreviations are as follows: N-tags, N-terminal tags; C-tags, C-terminal tags; RS, restriction site; MBP, maltose binding protein; TRX, thioredoxin; 6-His, hexa-histidine tag; Bap, biotinylation peptide; gp67ss, gp67 secretion signal sequence.

Drosophila Schneider 2 (S2) cell expression; in this gene, the four N-linked glycosylation consensus sequences (N–X–S/T) were inactivated by threonine-to-alanine mutations (Life Technologies). The *Drosophila*-optimised gene was flanked by an EcoRI site and Kozak sequence at the 5' end, and the coding sequence for a hexa-His tag and a NotI site at the 3' end. Further *Drosophila*-optimised synthetic genes, encoding shorter fragments of PfRH5 (PfRH5ΔN and PfRH5ΔNL; see Chapter 3), were also synthesised (Life Technologies).

For expression of other RH proteins in S2 cells, parts of the PfRH2b, PvRBP-1 and PvRBP-2 genes were optimised for *Drosophila* expression and synthesised (Life Technologies). PfRH1 and PfRH4 synthetic genes optimised for mammalian expression were provided by G. Wright and S. Draper, respectively.

2.2.2 Cloning vectors

The main vector for *E. coli* expression was a modified pET15b (pET15b$_{con}$), which encodes an N-terminal hexa-histidine tag followed by a tobacco etch virus (TEV) protease cleavage site [270] (Figure 2.1A). For S2 expression, the pExpreS^2.1 vector (ExpreS^{2}ion Biotechnologies, Horsholm, Denmark), or a modified version thereof, was used (Figure 2.1B). These and other vectors used are summarised in Table 2.1.

2.2.3 Polymerase chain reaction (PCR)

PCR primers used to generate the various basigin and PfRH5 constructs for *E. coli*, baculovirus-infected insect cell, and HEK cell expression are summarised in Appendix A. PCR reactions were performed using KOD HotStart polymerase (EMD Millipore), according to manufacturer's protocols. A typical reaction contained 50 ng plasmid DNA, 0.3 μM of the forward and reverse primers, 0.2 mM dNTPs, 1 mM $MgSO_4$, and 1 U of KOD HotStart DNA polymerase in KOD HotStart buffer (EMD Millipore). The PCR reactions were carried out in an MJ Mini Personal Thermal Cycler (Bio-Rad), commencing with 2 minutes at 95°C, continuing with 30 thermal cycles (20 s at 95°C, 10 s at 55-65°C, 2 minutes at 70°C), and finishing with 5 minutes at 70°C. For site-directed mutagenesis, the elongation time during each cycle was extended to 7 minutes. Successful amplification of DNA fragments was assessed by agarose gel electrophoresis.

2.2.4 DNA agarose gel electrophoresis

An 0.8% agarose gel was prepared in TBE buffer (89 mM Tris pH 8.0, 89 mM boric acid, 2 mM EDTA). After melting and cooling to ∼60°C, 5 μg/ml ethidium bromide was added, and the gel allowed to set. The gel was run in TBE buffer at 65 V for 45 minutes. Bands were visualised with a G:Box Chemi-XRQ imaging system (Syngene).

2.2.5 Restriction digestion, calf alkaline phosphatase treatment, and ligation

Restriction digestion reactions included 3-5 μg DNA and 20 U of the relevant restriction enzymes (NEB) in CutSmart buffer (NEB), and were incubated at 37°C for 4 hours (for plasmids) or overnight (for PCR products). Digested DNA was purified (Qiagen Gel Extraction kit), either directly or from an agarose gel slice. Vectors were subsequently treated with 10 U calf alkaline phosphatase for 1 hour at 37°C, before plasmid purification. To ligate the relevant insert and vector, 200 U of T4 DNA ligase and 1xT4 DNA ligase buffer (NEB) were added to a mixture containing ∼50 ng of each component (but weighted according to their relative molecular weights). The ligations were incubated at 16°C overnight, or for 10 minutes at room temperature.

2.2.6 Transformation, screening, and sequencing

Chemically competent DH5α cells (Life Technologies) or XL10-Gold ultracompetent cells with supplemented β-mercaptoethanol (Stratagene) were transformed with 2-5 μl of the ligation reaction. In brief, after addition of the DNA, the cells were incubated on ice for 30 minutes before heat shock for 45 seconds at 45°C. The cells were returned to the ice for 2 minutes, then 1 ml of Lysogeny Broth (LB) media was added. The cells were incubated with shaking for 1 hour at 37°C, plated on agar plates containing the requisite antibiotic (either 50 μg/ml carbenicillin, 35 μg/ml chloramphenicol, or 25 μg/ml kanamycin), and incubated in a static incubator at 37°C overnight. Single colonies were selected and used to inoculate 3 ml of antibiotic-containing media; the cultures were grown overnight, shaking, at 37°C, before DNA extraction (Qiagen Miniprep kit). Clones were screened by colony PCR (in which 0.5 μl of cells were added directly to a PCR reaction) or restriction digestion, before final verification by Sanger sequencing (Source Biosciences, Nottingham, UK).

2.2.7 In-Fusion^{TM} cloning pipeline at the Oxford Protein Production Facility

Cloning was carried out in 96-well, high-throughput format as described previously [272] at the Oxford Protein Production Facility (OPPF), Harwell, UK. Five vectors were used (Table 2.1). PCR reactions, purification of PCR products with AMPure XP magnetic beads (Beckman-Coulter, High Wycombe, Buckinghamshire, UK), In-Fusion^{TM} cloning (Clontech-Takara Bio, Saint-Germain-en-Laye, France), and transformation into OmniMaxII competent cells (Invitrogen) were as described previously [272]. Plasmids were prepared using a Qiagen BioRobot 8000 and Qiagen TurboPrep kits (Qiagen) and screened by PCR.

2.2.8 RNA isolation and reverse transcription

To isolate total RNA, 1-10 x 10^6 hybridoma cells were harvested (1,000xg) and resuspended in 1 ml ice-cold phosphate buffered saline (PBS). The cells were spun again and the PBS aspirated, to be replaced with 1 ml Trizol (Life Technologies). The reaction was incubated for 5 minutes at room temperature. 250 μl chloroform was added and

the mixture shaken vigorously before another 5 minute incubation. Next, the sample was centrifuged at 10,000 rpm for 5 minutes. The resultant aqueous phase was retained, 550 μl of isopropanol was added, and the mixture was incubated for 5 minutes at room temperature before another spin (14,000 rpm) for 30 minutes. After the spin, the liquid was poured off and 1 ml 75% ethanol in DEPC-treated water added. The sample was centrifuged at 9500 rpm for 5 minutes, the ethanol removed, and the pellet allowed to dry for 5-10 minutes, after which time 25 μl water was added to redissolve the pellet. The A_{260}/A_{280} ratio was ascertained to be >1.8.

The RNA thus purified was used as a template in a SuperScript III reverse transcriptase reaction (Life Technologies), following manufacturer's protocols. In brief, 500 μg RNA was mixed with 2 pmol of reverse primers (specific for mouse κ light or IgG2a heavy chains) and 0.5 mM dNTPs, then incubated for 5 minutes at 65°C, followed by 1 minute on ice. At this point, 1x First-Strand Buffer, 50 μM DTT, 40 U RNaseOUT (Life Technologies), and 200 U Superscript III reverse transcriptase were added to the reaction, which was then incubated at 55°C for 60 minutes. The reverse transcriptase reaction was stopped by heating to 70°C for 15 minutes. 2 μl of the resultant cDNA was used as a template for PCR (Section 2.2.3). The PCR product was separated by agarose gel electrophoresis, and the bands corresponding to the immunoglobulin cDNA excised, purified, and sent for Sanger sequencing (Source Bioscience).

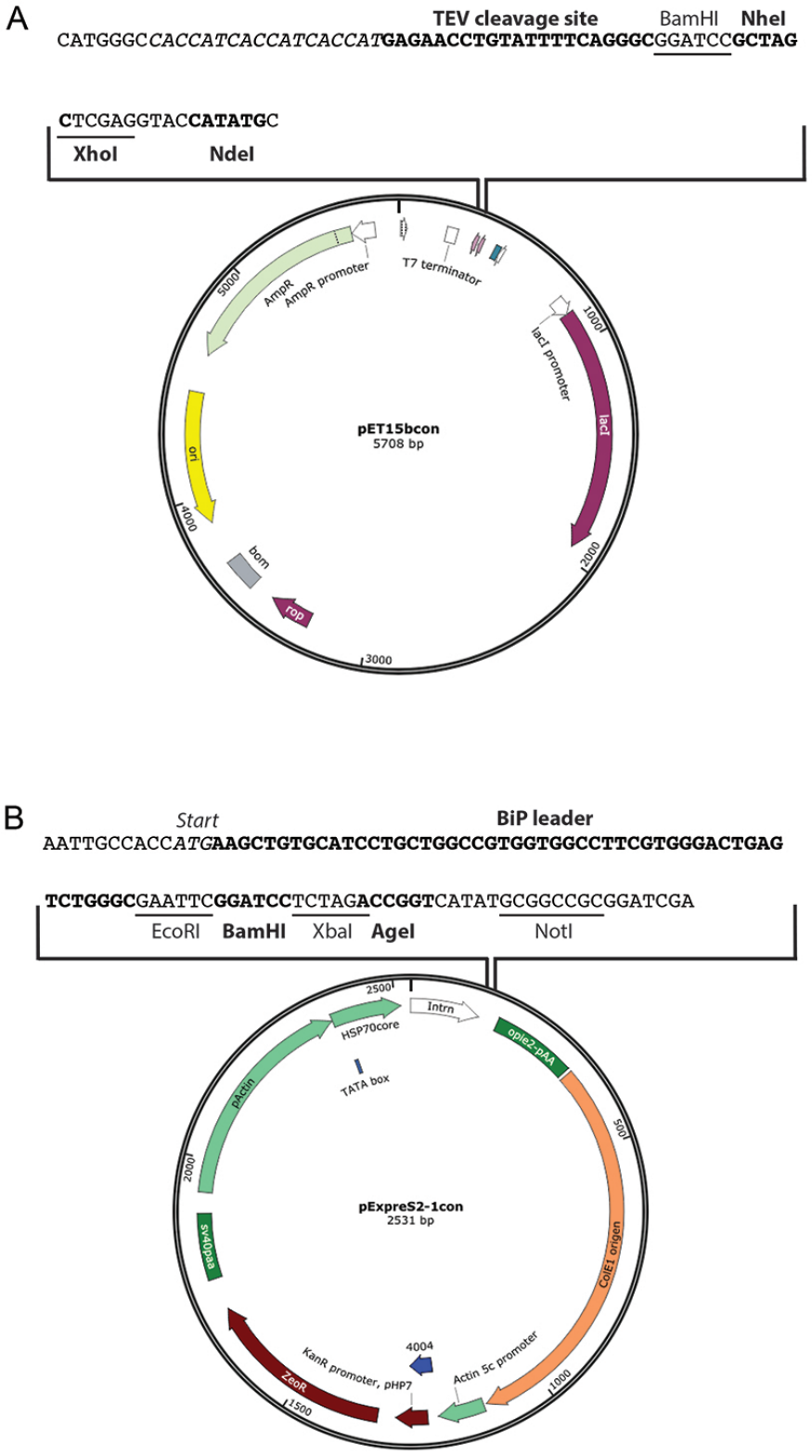


FIGURE 2.1: Maps of expression plasmids pET15b$_{con}$ (A) and pExpreS^2.1$_{con}$ (B), their multiple cloning sites modified by Dr. Matthew Higgins. (A) pET15b$_{con}$ contains an N-terminal 6-His tag followed by a TEV site. (B) pExpreS^2.1$_{con}$ contains a BiP signal sequence upstream of the multiple cloning site.

2.3 Protein expression and purification

2.3.1 *E. coli*

2.3.1.1 Expression strains

Expression strains used included methionine-auxotrophic B834 and Rosetta 2(DE3)pLysS (Novagen). Rosetta strains supply tRNAs for codons rarely used in *E. coli* but common in eukaryotes. Other cell lines were optimised for expression of disulphide bond-containing proteins: Novagen Origami B(DE3) cells, which contain mutations in the thioredoxin reductase gene *trxB* and the glutathione reductase gene *gor*; and B834 pMJ59 cells, which express Erv1p sulphydryl oxidase and the human protein disulphide isomerase PDI [273]. Rosetta-gami 2(DE3) cells (Novagen) combine the properties of Origami and Rosetta strains. Finally, ArcticExpress cells (Agilent), encoding protein folding chaperonins Cpn60 and Cpn10 from antarctic bacterium *Oleispira antarctica*, were used in an attempt to increase solubility of expressed proteins. The strains were transformed as described in Section 2.2.6.

2.3.1.2 Small scale expression trials

To assess the level of protein expression, colonies from fresh transformation plates were used to inoculate antibiotic-containing LB cultures. The cultures were grown at 37°C to OD_{600} = 0.6-1.0 and induced with 0.5-1 mM IPTG. Cultures were then incubated for 3 hours or overnight, at either 13°C (for ArcticExpress cells), 25°C, 30°C, or 37°C. Cells were harvested by spinning at 8,000 x *g* for 3 minutes.

The expression level was tested by one of two different methods, using either small-scale nickel-nitrotriacetic acid (Ni^{2+}-NTA) columns, or by Western blot. For the former, the pellet from 5 ml cells was resuspended in 1 ml BugBuster (Novagen), incubated with nutation for 20 minutes at room temperature. Insoluble cell debris was removed by centrifugation at 17,000 x *g* for 20 minutes at 4°C. To test soluble expression, the supernatant was applied directly to His SpinTrap columns (GE Healthcare), the column washed (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% (v/v) Tween-20), and specifically bound protein eluted with the same buffer containing 250 mM imidazole. To test insoluble expression, the pellet from the spin was resuspended (50

mM Tris pH 8.5, 300 mM NaCl, 8 M urea, 30 mM imidazole, 0.5% Triton-X-100 and 1 mM TCEP), and the solution incubated for 1 hour nutating at room temperature before centrifugation for 1 hour at 17,000 x *g*. The supernatant from this spin was applied to a His SpinTrap column, which was subsequently washed in resuspension buffer lacking Triton and TCEP. Insoluble protein was eluted with the same buffer containing 300 mM imidazole. Eluted soluble and insoluble proteins were then analysed by SDS-PAGE.

To test expression by Western blot, cell pellets from 1 ml culture were resuspended in 1 ml PBS and sonicated using a Misonix Sonicator 3000 for five 5-sec pulses. A sample of the sonicated cells was reserved (whole cell fraction). The sonicated cells were centrifuged for 5 minutes at 13,000 rpm. Samples of the supernatant (soluble fraction) and of the whole cell fraction were analyzed by SDS-PAGE.

2.3.1.3 Polyacrylamide gel electrophoresis and Western blotting

Protein samples were mixed with 2x loading dye (100 mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanol), boiled at 95°C for 4 minutes, and loaded into hand-cast 12% polyacrylamide gels in a Bio-Rad Mini-PROTEAN gel system. The running buffer consisted of 30 mM Tris pH 7.5, 192 mM glycine, and 0.1% SDS. Gels were run at 200 V for 45 minutes, and then stained with InstantBlue (Expedeon).

For Western blotting to detect 6-His-tagged fusion proteins, protein was transferred onto an Immobilon polyvinylidene fluoride (PVDF) membrane (Millipore) using a Bio-Rad TransBlot Semi-Dry system. The transfer buffer was 48 mM Tris pH 9.2, 39 mM glycine, and 20% methanol; the PVDF membrane was first rinsed in 100% methanol before submersion in blotting buffer. The membrane and the gel were placed between Whatman blotting paper, and a current of 50 mA applied for 1 hour. The membrane was then blocked in penta-His-HRP conjugate blocking buffer (Qiagen) for one hour at room temperature or overnight at 4°C. After blocking, the blot was washed three times for 10 minutes each (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.05% Tween 20, and 0.2% Triton-X-100). Thereafter, the blot was incubated for 1 hour at room temperature with mouse penta-His-HRP conjugate IgG antibody (Qiagen; 1:5000 dilution) and washed three more times in wash buffer for 10 minutes each. Finally, the blot was developed using the ECL Prime Western Blotting Detection Kit (Amersham), with bands visualised on Amersham

Hyperfilm ECL. Occasionally, instead of using the penta-His-HRP conjugate system, a combination of mouse penta-His IgG antibody (Qiagen; 1:1000 dilution) and anti-mouse IgG-alkaline phosphatase conjugate antibody (Sigma; 1:2000 dilution) was used. In this case, the blot was developed using SIGMAFAST BCIP/NBT tablets (Sigma-Aldrich).

2.3.1.4 Expression pipeline at the OPPF

Expression constructs verified by PCR were transformed into B834 (DE3), Rosetta 2(DE3)pLysS (Novagen), and B834 pMJ59 *E. coli*. Expression conditions, both for auto-induction with Overnight ExpressTM Instant TB medium (Novagen) and for induction with IPTG, were as described [272]. Modifications to the published protocol included the supplementation of the medium with 35 μg/ml chloramphenicol for B834 pMJ59 cells, to maintain pMJ59 plasmid. B834 pMJ59 cells were pre-induced with 0.2% L-arabinose 30 minutes prior to induction with IPTG as described.

Soluble expression levels were gauged using established protocols [272]. In brief, 1 ml of culture was harvested by centrifugation at 6,000 x *g* for 10 minutes at 4°C and pelleted cells frozen for 30 minutes at -80°C. Thereafter, pellets were screened for soluble protein expression using the BioRobot 8000 with Qiagen Ni-NTA magnetic beads. Proteins eluted from the magnetic beads were analyzed using SDS-PAGE (Invitrogen NuPAGE Novex 10% Bis-Tris Midi gels). Gels were stained with InstantBlue (Expedeon).

To analyse insoluble protein expression, insoluble pellets were defrosted and solubilized in buffer containing 50 mM Tris pH 8.5, 300 mM NaCl, 8 M urea, 30 mM imidazole, 0.5% Triton X 100 and 1 mM TCEP. The plate was agitated at 700 rpm for 1 hour at room temperature using the THeonyx root (Aviso Gmbh, Gera, Germany), then centrifuged for 1 hour at 6,500 x *g* at 4°C. Insoluble protein was purified using the Qiagen Ni-NTA Superflow 96 BioRobot kit. The protein was eluted in 50 mM Tris pH 8.5, 8 M urea, and 300 mM imidazole, and analyzed by SDS-PAGE.

2.3.1.5 Large scale expression in *E. coli*

Once a suitable expression regimen was identified, the growth was scaled up (2-12 L) to produce the quantities required for biophysical analysis and crystallisation. Freshly transformed cells were used to inoculate 50-100 ml of LB medium containing the relevant

antibiotics. The next day, 10-20 ml of the overnight culture was diluted into each litre of 2x Yeast-Tryptone medium (2xYT, Melford). The cultures were incubated at 37°C, shaking at 200 rpm, until they reached an OD_{600} of 0.6-1.0, when they were induced with 1 mM IPTG. The temperature was reduced to that at which the protein best expressed in the small-scale trial; likewise, the cells were harvested after 3 hours or overnight depending on the small-scale results. After harvesting by centrifugation at 5,500 x *g* for 20 minutes in a Beckman Coulter Avanti J-20 centrifuge (JLA-8.1000 rotor), the cell pellets were stored at -80°C.

2.3.1.6 Selenomethionine incorporation

For selenomethionine labelling of basigin, freshly transformed Origami B cells (containing basigin in the $pET15b_{con}$ vector) were used to inoculate 160 ml LB cultures containing 50 μg/ml carbenicillin, and left to grow overnight at 37°C. The next day, 20 ml of the culture was diluted into each of 8 litres of 2xYT with carbenicillin. The large cultures were grown to OD_{600} = 0.6-1.0, then centrifuged for 15 minutes at 5,000 rpm. The cell pellet was resuspended in minimal medium. 1 litre of minimal medium contained autoclaved M9 (5 g KH_2PO_4, 1 g NH_4Cl, 0.5 g NaCl), 2 mM autoclaved $MgSO_4$, 10 μM autoclaved $CaCl_2$, 2 g glucose, 5 ml MEM vitamins (Gibco), and 50 μg/ml carbenicillin. The resuspended cells were incubated for 30 minutes at 37°C, before adding amino acids including selenomethionine (100 mg L-lysine, 100 mg L-phenylalanine, 100 mg L-threonine, 50 mg L-leucine, 50 mg L-valine, and 50 mg selenomethionine per litre). After a further 15 minute incubation to allow inhibition of methionine synthesis to commence, cells were induced with 1 mM IPTG and incubated at 25°C overnight. They were harvested and stored as usual (Section 2.3.1.5).

2.3.1.7 Soluble purification of 6-His-tagged proteins

Cell pellets were thawed and resuspended in 20 mM Tris pH 8.0, 300 mM NaCl, 15 mM imidazole, and 0.5% Triton-X-100, supplemented with 250 U Benzonase nuclease (Sigma) and a protease inhibitor cocktail (Thermo Scientific). Cells were disrupted by sonication using a Misonix Sonicator 3000, and the resulting cell debris removed by pelleting at 20,000 rpm for 30 minutes at 4°C in the Ti 30.50 rotor in a Beckman

Coulter centrifuge. Overexpressed, 6-His-tagged proteins were then purified by nickel-nitrilotriacetic acid (Ni^{2+}-NTA; Qiagen) affinity chromatography. The supernatant from the spin was applied to a Ni^{2+}-NTA column. The beads were washed with resuspension buffer (lacking Benzonase and protease inhibitors), and then in a higher salt buffer (20 mM Tris pH 8.0, 500 mM NaCl, 15 mM imidazole). Specifically-bound protein was eluted (20 mM Tris pH 8.0, 150 mM NaCl, 200 mM imidazole), and then buffer-exchanged using a PD-10 desalting column (GE Healthcare) into phosphate buffered saline (PBS), before overnight cleavage with TEV protease.

2.3.1.8 Tobacco etch protease (TEV) purification

6-His-tagged TEV protease was expressed (Section 2.3.1.5) and purified by Ni^{2+}-NTA chromatography followed by buffer exchange into PBS (Section 2.3.1.7) essentially as described. The only difference to the above protocols was the addition of 10% glycerol to the resuspension buffer. Notably, no reducing agent was added during the purification, since the surface proteins being cleaved with TEV protease contain disulphide bonds. The protein was aliquoted, frozen in liquid nitrogen, and stored at -80°C.

2.3.1.9 TEV cleavage

The pET15b$_{con}$ vector encodes a TEV cleavage site between the N-terminal 6-His tag and the protein of interest. The 6-His tag was therefore removed from the protein of interest by overnight TEV protease cleavage. 1 mg TEV protease was added for every 10 mg of the Ni^{2+}-NTA-purified protein of interest (Section 2.3.1.7); the reaction was supplemented with a redox buffer (3 mM reduced and 0.3 mM oxidised glutathione), to maintain a balance between the reducing conditions required for TEV protease function and the oxidising conditions required to maintain disulphide bonds. The mixture was incubated at 4°C overnight, and then applied to a second Ni^{2+}-NTA column. The material that did not bind to the column, containing the protein of interest without its His-tag, was retained; the elution, containing His-tagged TEV protease and any uncleaved protein, was discarded. The flow-through was concentrated using an Amicon Ultra centrifugal filter device, with a molecular mass cut-off at least three times smaller than the molecular mass of the protein, in preparation for size exclusion chromatography.

2.3.1.10 Size exclusion chromatography

Size exclusion chromatography was used as a polishing, final step of protein purification. The protein was loaded on a Superdex 200 16/60 prep-grade column (GE Healthcare), pre-equilibrated in 20 mM HEPES pH 7.5, 150 mM NaCl. 2 ml fractions were collected and analysed by SDS-PAGE; fractions judged to contain pure protein were pooled and concentrated with an Amicon Ultra centrifugal filter device. Finally, pure, concentrated protein was aliquoted, frozen in liquid nitrogen, and stored at -80°C.

2.3.1.11 Soluble purification of MBP-fusion protein

The N-terminal half of PfRH5 was purified from *E. coli* as a 6-His-maltose-binding-protein (MBP) fusion, encoded by the pOPINM vector. In between MBP and the protein of interest is a 3C protease cleavage site. After Ni^{2+}-NTA chromatography (Section 2.3.1.7), the protein was further purified by applying to a column packed with amylose resin (NEB), to which MBP binds specifically. After the column was washed (20 mM Tris pH 8.0 and 150 mM NaCl), the protein was eluted with wash buffer containing 10 mM maltose. 6-His-tagged 3C protease (provided by S. Newstead) and glutathione mixture (Section 2.3.1.9) were added to the amylose column elution and the mixture incubated overnight at 4°C. A second Ni^{2+}-NTA column was run the next day, and the material that did not bind the column was concentrated and further purified by gel filtration (Section 2.3.1.10). Fractions were analysed by SDS-PAGE.

2.3.1.12 Insoluble purification from inclusion bodies and refolding

For attempts to purify and refold insolubly expressed PfRH5, inclusion bodies pelleted along with the cell debris (Section 2.3.1.7) were washed once in resuspension buffer and re-pelleted. The inclusion bodies were refolded either on-column, or by flash dilution. For on-column refolding, the inclusion bodies were resuspended in guanidine-containing buffer (6 M guanidine, 20 mM Tris pH 8.0, 300 mM NaCl, 15 mM imidazole, 5 mM β-mercaptoethanol) using a Dounce homogeniser. The sample was stirred for 60 minutes at room temperature, and then centrifuged to remove insoluble material (20,000 rpm for 1 hour). The supernatant was applied to a Ni^{2+}-NTA column. The column was washed sequentially with buffers containing a glutathione mixture (Section 2.3.1.9) and stepwise

decreasing amounts of guanidine, until a final wash into buffer with no guanidine. The putatively refolded protein was eluted in 20 mM Tris pH 8.0, 150 mM NaCl, and 200 mM imidazole. The rest of the purification was identical to that for a soluble protein (Section 2.3.1.7).

For flash dilution, the inclusion bodies were resuspended in urea-containing buffer (50 mM Tris pH 8.5, 300 mM NaCl, 8 M urea, 30 mM imidazole, 5 mM β-mercaptoethanol) by Dounce homogenisation, stirred for 60 minutes at room temperature, and centrifuged to remove insoluble material (20,000 rpm for 1 hour). The supernatant was applied to a Ni^{2+}-NTA column, washed with resuspension buffer, and eluted in 50 mM Tris pH 8.5, 8 M urea, and 300 mM imidazole. Once purified, the elution was diluted 20-fold into $\sim$20 different refolding buffers (Table 2.2), 15 of which were formulated according to the QuickFold protein refolding kit (Athena Enzyme Systems). These buffers included various additives previously shown to aid correct refolding, including arginine, sucrose, polyethylene glycol 3,550, and various salts. The diluted samples were incubated overnight at 4°C; the next day, they were concentrated as necessary, and analysed by SDS-PAGE.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
50 mM MES pH 6	✓	✓	✓	✓	✓	✓	✓												✓	
50 mM M Tris pH 8.5								✓	✓	✓	✓	✓	✓	✓	✓			✓		
50 mM CAPS pH 10.5																✓	✓			✓
9.6 mM NaCl	✓	✓	✓					✓	✓	✓	✓									
240 mM NaCl				✓	✓	✓	✓					✓	✓	✓	✓					
500 mM NaCl																	✓	✓	✓	
0.4 mM KCl	✓	✓	✓					✓	✓	✓	✓					✓				
10 mM KCl				✓	✓	✓	✓					✓	✓	✓	✓					
2 mM MgCl2	✓	✓		✓			✓	✓		✓				✓	✓					
2 mM CaCl2	✓	✓		✓			✓	✓		✓				✓	✓					
0.4 M sucrose			✓		✓	✓		✓		✓				✓	✓					
0.5 M arginine		✓		✓		✓			✓	✓			✓	✓			✓	✓	✓	✓
1 mM EDTA			✓		✓	✓			✓		✓	✓	✓							
0.05% PEG 3,350		✓	✓			✓	✓	✓	✓			✓		✓						
0.5% Triton-X-100	✓		✓	✓		✓		✓			✓		✓	✓						
0.75 M guanidine	✓		✓		✓		✓		✓	✓			✓	✓						
1 mM DTT	✓		✓		✓		✓		✓		✓		✓		✓					
1 mM GSH/ 0.1 mM GSSH		✓		✓		✓		✓		✓		✓		✓		✓	✓	✓	✓	✓

TABLE 2.2: Protein refolding buffers (1-20) were formulated according to the above table. A check mark indicates the presence of the given component in the buffer. The recipes for buffers 1-15 were taken from the QuickFold protein refolding kit (Athena Enzyme Systems).

2.3.2 Human embryonic kidney (HEK) cells

2.3.2.1 HEK cell maintenance

Adherent HEK293 cells were maintained in 6-well plates (a 10 cm^2 surface area) in Dulbecco's Modified Eagle Medium (DMEM; PAA) supplemented with 10% foetal calf serum (FCS; Gibco), 100 U penicillin and 0.1 mg/ml streptomycin (Sigma). The cells were split at 80% confluence. First, the cells were detached by replacement of their medium with Dulbecco's PBS-trypsin. Once the cells detached (∼5 min), they were centrifuged at 1000 rpm for 5 minutes. The medium was aspirated from the cell pellet, and the cells resuspended in fresh medium to the desired density.

2.3.2.2 Transient transfection of HEK cells

Plasmids pEFbos$_{con}$ and pHLsec [271] were used for transient HEK transfections. 3 μl Fugene 6 transfection reagent (Promega) was diluted into 62.5 μl serum-free DMEM drop-wise, and the mixture incubated at room temperature for 5 minutes before the addition of 2 μg DNA and a further 20 minute incubation to allow DNA:Fugene complexation. During the incubation, the medium on the cells was replaced with fresh serum-free medium. The DNA-Fugene mixture was then added drop-wise to cells (80% confluent) in 6-well plates in serum-free DMEM. The next day, the medium was changed to fresh serum-free DMEM to remove the transfection reagent. After 4-5 days, the medium (containing the secreted protein of interest) was collected and analysed for protein expression by Western blot. Similar procedures were employed for transient transfections using transfection reagents lipofectamine (Invitrogen) and polyethyleneimine (Sigma).

2.3.3 Hybridomas

2.3.3.1 Hybridoma growth

Anti-PfRH5 monoclonal antibodies were described previously [237]. Hybridomas expressing 9AD4, QA1, QA5, RB3 and 4BA7 were grown in DMEM (Sigma) supplemented with 4 mM L-glutamine (Sigma), 0.01 M HEPES (Life Technologies), 100 U penicillin and 0.1 mg/ml streptomycin (Sigma), and 20% foetal calf serum (Gibco). The cells

were maintained at a density between 1-10 x 10^5 cells per ml. They were amplified, transferred into serum-free CD Hybridoma medium (Life Technologies) with glutamine, penicillin, and streptomycin, and allowed to grow unmolested for 7-10 days, secreting antibodies all the while. The cells were then harvested by centrifugation at 5,500 rpm for 20 minutes, and the supernatant retained for further processing.

2.3.3.2 Antibody purification

The cell culture supernatant was exchanged into 20 mM phosphate pH 7.0 with a tangential flow filtration device (Pall). The sample was then loaded onto a HiTrap Protein G HP column (GE Healthcare), eluted in 0.1 M glycine-HCl pH 3.0, and immediately neutralised with 0.1 M Tris pH 8.0. The sample was exchanged into 100 mM phosphate pH 6.4, 300 mM NaCl, 2 mM EDTA, 5 mM L-cysteine pH 6.4, and 1.5 mM β-mercaptoethanol using PD-10 columns (GE Healthcare).

2.3.3.3 Production of antibody fragments

Antibody fragments (Fabs) were generated by addition of papain agarose (Sigma), and overnight incubation at 37°C. The papain agarose was removed by centrifugation, and the sample loaded onto a HiTrap Protein A HP column (GE Healthcare). Fab fragments do not bind the column as they lack the protein A-binding fragment crystallisable (Fc) region. The flow-through from the protein A column was concentrated using an Amicon centrifugal filter device (molecular mass cut-off 10,000 kDa) and gel filtered on a Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES pH 7.5 and 150 mM NaCl. Fractions containing the Fab fragment were pooled, concentrated again, and flash-frozen for storage at -80°C.

2.3.4 Baculovirus-infected insect cells

2.3.4.1 Sf9 cell maintenance

Spodoptera frugiperda-derived Sf9 insect cells (Invitrogen) were split three times a week to 0.5 x 10^6 cells/ml. They were maintained in Sf-900 II medium.

2.3.4.2 Baculovirus production

The PfRH5 gene was subcloned into the pAcGP67A$_{con}$ vector, situated after an N-terminal secretion signal, a 6-His tag, and a TEV cleavage site. Gene expression is controlled by the *Autographa californica* nuclear polyhedrosis virus (AcNPV) polyhedrin promoter, a very late promoter that mediates high-level expression. To create stable baculoviruses, Sf9 cells were co-transfected with the PfRH5–pACGP67A$_{con}$ construct and linearized Baculogold baculovirus DNA (Pharmingen), in the manner described by the manufacturer. The Baculogold DNA contains a lethal deletion which must be rescued by homologous recombination with complementary pACGP67A DNA, thus introducing the gene of interest into the baculovirus DNA. Recombinant baculoviruses were subsequently isolated by plaque purification and passaged three times in Sf9 cells to obtain high titre stocks (2-4 x 10^8 plaque forming units per ml).

2.3.4.3 Expression in baculovirus-infected cells and purification

Sf9 insect cells were grown in 500 ml suspension cultures in 2 L roller bottles at 27°C with shaking at 150 rpm. Once the cells reached a density of 1 x 10^6 cells/ml, they were infected with virus at a multiplicity of infection of 5-10. After 72 hours, cells were harvested by centrifugation at 5,500 rpm for 10 minutes, and the supernatant, containing secreted PfRH5, retained for purification. For purification, the supernatant was buffer exchanged via tangential flow filtration (Section 2.3.5.4) and applied to a Ni^{2+}-NTA column. Thereafter, the purification procedure was as described in Section 2.3.1.7.

2.3.5 Schneider 2 cells

2.3.5.1 Cell maintenance

Schneider 2 (S2) cells are derived from late-stage *Drosophila melanogaster* embryos [274]. Untransfected S2 cells were maintained in EX-CELL 420 insect cell medium with L-glutamine (Sigma) supplemented with 10% FCS (Gibco), 100 U penicillin and 0.1 mg/ml streptomycin (Sigma), and split every 3-4 days to 5 million cells/ml. Stable cell lines, by contrast, were maintained in medium lacking FCS and split to 8 million cells/ml.

2.3.5.2 Stable transfection of S2 cells

In work performed by ExpreS^{2}ion Biotechnologies, Hørsholm, Denmark, various constructs of PfRH5 were subcloned into the pExpreS^2.1 vector and then transfected into S2 cells. I performed the same subcloning and transfection procedures for constructs expressing other RH proteins. The transfection protocol was as follows: 50 μl ExpreS^2 Insect-TRx5 (ExpreS2ion) transfection reagent was added to 5 ml of cells at 2 million cells/ml in serum-free EX-CELL 420 medium, followed by 12.5 μg DNA. This was gently mixed and transferred to a canted neck flask. After 4 hours, 12.5% FCS was added, and after 24 hours, 1.5 mg/ml zeocin was added to begin the antibiotic selection. A polyclonal cell line was selected over three weeks, with 2 ml medium replaced with fresh zeocin-containing medium twice a week.

2.3.5.3 Protein expression in S2 cells

The stable S2 cell lines constitutively express and secrete the protein of interest. The cell lines were expanded in EX-CELL 420 medium without FCS or zeocin. The volume could be scaled according to the amount of protein required: from 20 ml for an expression trial, to upwards of 10 litres to produce protein for crystallography. Four days after the final split, the cells were harvested, and the supernatant retained for Western blotting (for small-scale expression trials) or protein purification. To produce selenomethionine-labelled protein, cells were transferred into methionine-deficient D Series medium (Expression Systems), with no methionine supplement; when they were expanded to the desired volume, their medium was supplemented with 200 mg/ml selenomethionine, and the supernatant was harvested 4 days later.

2.3.5.4 Purification from S2 cells

Before purification, the cell culture supernatant was concentrated ∼10-20-fold and exchanged into 20 mM phosphate, pH 7.4, 250 mM NaCl, 20 mM imidazole, and 0.005% Tween-20, using a tangential flow filtration device (Pall). At this stage, the supernatant was either flash-frozen in liquid nitrogen and stored at -80°C, or used directly. The buffer-exchanged supernatant was loaded onto a Ni^{2+}-NTA chromatography column (HisTrap FF; GE Healthcare). The column was washed with 20 mM phosphate, pH 7.4,

250 mM NaCl, 40 mM imidazole, and 0.005% Tween-20, and specifically-bound protein eluted in the same buffer with 400 mM imidazole.

2.3.5.5 ConA purification

During PfRH5 purification, the eluate from Ni^{2+}-NTA chromatography contained a variety of primarily high-molecular-weight contaminants. However, because its glycan sites were mutated, PfRH5 was a lone unglycosylated protein among the largely glycan-decorated S2 proteins, and could be purified further by ConA lectin-binding chromatography. The Ni^{2+}-NTA elution was applied to a HiTrap ConA 4B column (GE Healthcare), and the flow-through, containing non-binding PfRH5, retained. (The contaminant S2 proteins were released from the column by the addition of methyl-α-D-glucopyranoside.) Thereafter, PfRH5 destined for biophysical analysis was subjected to gel filtration on a Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES pH 7.5 and 150 mM NaCl, concentrated with an Amicon centrifugal filter device (molecular weight cut-off, 10,000 Da), flash-frozen and stored at -80°C.

2.3.5.6 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was employed to optimise the buffer conditions for PfRH5 purification [275]. PfRH5 was diluted into a large number of buffers, systematically varied in salt concentration or pH, for a final concentration of 20 μg/ml in 50 μl in a 96-well plate. SYPRO Orange fluorescent dye (Life Technologies; excitation 473 nm, emission 610 nm) was diluted 5000-fold into the protein samples and appropriate buffer blanks. The plate was placed in a real-time PCR machine, and the temperature ramped up degree by degree (26-80°C), while the sample was illuminated with 473 nm light, and the fluorescence at 610 nm monitored. The SYPRO Orange dye binds to hydrophobic regions of the protein, and hence when the protein unfolds the dye molecules can access its hydrophobic core, and the fluorescence increases. Thus, this method monitors the unfolding transition of the protein in the different buffers.

2.3.5.7 Limited proteolysis

The effect of proteases trypsin, chymotrypsin, carboxypeptidase B, carboxypeptidase Y, and endoproteinase GluC (Sigma) on pure PfRH5 protein was monitored over time. 75 μg PfRH5 was mixed with 1.5 μg of each protease in 20 mM HEPES pH 7.5 and 150 mM NaCl, and the reaction mixtures incubated at 37°C. Samples were taken after 0 min, 15 min, 60 min, 4 hours, and overnight, and analysed by SDS-PAGE and InstantBlue staining.

2.3.5.8 Endoproteinase GluC cleavage

To prepare a PfRH5 construct (PfRH5ΔNL; see Chapter 3) for crystallisation, the sample was cleaved overnight with endoproteinase GluC (Sigma) at 20°C in Ni^{2+}-NTA elution buffer (20 mM phosphate, pH 7.4, 250 mM NaCl, 400 mM imidazole, 0.005% Tween-20). The next day, the protein was centrifuged at 17,000 x g to pellet any precipitate.

2.3.5.9 Protein methylation

To chemically methylate PfRH5 protein complexes, 0.02 M dimethylamine-borane complex (ABC) and 0.04 M formaldehyde were added and the reaction incubated at 4°C for 1 hour. An additional 0.02 M ABC and 0.04 M formaldehyde were added; after another 1 hour, 0.01 M ABC was added and the reaction was incubated at 4°C overnight. Finally, the complexes were gel filtered and concentrated as detailed above (Section 2.3.5.5).

2.4 Biophysical methods

2.4.1 Surface plasmon resonance

The phenomenon of surface plasmon resonance (SPR) is exploited to analyse macromolecular interactions [276]. When plane-polarised light strikes the interface between two transparent media of different refractive index above a critical angle of incidence, the light emanating from the medium with the higher reflective index is reflected (total

internal reflection). However, some of the incident light penetrates into the medium with the lower refractive index, as a so-called evanescent wave. When the interface between the two media is coated with a thin layer of a conducting material, such as a metal, the parallel-polarised component of the evanescent wave enters the metal layer. At a certain angle of incidence, the delocalised surface electrons, or plasmons, of the metal are excited, and the evanescent wave greatly magnified. As a result, at this resonance angle the observed intensity of reflected light is greatly reduced. In a typical SPR experiment, the two media on the sensor chip are a glass prism and a flow cell containing the biochemical solution, separated by a thin gold layer. The resonance angle depends upon the refractive index of the flow cell near the sensor surface, which in turn is directly influenced by the mass of protein bound to the surface. To detect a binding event, a ligand is immobilised on the chip, and its binding partner, the analyte, is flowed over the surface. As the ligand binds the analyte, the accumulation of protein on the chip surface leads to an increased refractive index, and a changed SPR angle. The change in SPR angle during complex association and dissociation is measured over time, and the result plotted as resonance units (RUs) versus time.

Surface plasmon resonance experiments were carried out using a Biacore T200 instrument (GE Healthcare). To confirm basigin binding by PfRH5 constructs, experiments were performed at 20°C in 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20, 2 mg/ml dextran, and 1 mg/ml salmon sperm DNA (ssDNA; Sigma Aldrich). Basigin was immobilized on a CM5 chip (GE Healthcare) by amine coupling (GE Healthcare kit) to a total of 850-1000 RU. A concentration series of each PfRH5 variant (8, 4, 2, 1, 0.5, 0.25, 0.125, and 0.0625 μM) was injected over the basigin-coated chip for 120 s at 30 μl/min, followed by a 300 s dissociation time. The chip surface was then regenerated with 30 s of 2 M NaCl. Specific binding of each variant was obtained by subtracting the response from a blank surface from that of the basigin-coated surface. The kinetic sensorgrams were fitted to a global 1:1 interaction model, allowing determination of the dissociation constant, K_d, using BIAevaluation software 1.0 (GE Healthcare).

To test the binding of basigin mutants to PfRH5, the experiment was similar to the above, but PfRH5 was immobilised to the sensor chip instead of basigin. The buffer did not contain EDTA, dextran, or ssDNA, but the samples did contain 1 mg/ml ssDNA.

2.4.2 Analytical ultracentrifugation

Analytical ultracentrifugation is a powerful tool to characterise the self-assocation and hetero-association of macromolecules. When a sedimentation force is applied to macromolecules in solution, a boundary is formed between the macromolecules and the solvent, which migrates down the sample cell at a rate that can be measured in sedimentation velocity experiments. If the rotor speed is lowered and the sedimentation force decreased, sedimentation of the molecules down the cell is countered by back-diffusion–a phenomenon exploited in sedimentation equilibrium experiments.

Sedimentation velocity experiments give information about molecular size, shape, and interactions in solution. The spinning rotor exerts a sedimentation force on the macromolecule, equal to $m(1-\bar{v}\rho)\omega^2 r$, where m is the macromolecule mass, $\bar{v}$ is the partial-specific volume of the macromolecule, ρ is the solvent density, ω is the rotor angular velocity, and r is the distance from the centre of rotation. The $m(1-\bar{v}\rho)$ term is the effective molecular mass (taking into account solvent displacement). At terminal velocity, the sedimentation force will be balanced by the frictional force, and the mean velocity is constant. Since the frictional coefficient, f, is related to the diffusion coefficient D ($f = RT/N_A D$, where N_A is Avogadro's number, T is temperature, and R is the gas constant), this can be described by the equation:

$$v = \frac{DM(1-\bar{v}\rho)\omega^2 r}{RT}$$

where M is the molecular weight. The sedimentation coefficient (measured in Svedbergs, where 1 S = 10^{-13} s) is defined as $s = v/\omega^2 r$. Hence, we arrive at the Svedberg equation:

$$\frac{s}{D} = \frac{M(1-\bar{v}\rho)}{RT}$$

By combining the Svedberg equation with Fick's diffusion laws, it is possible to derive the Lamm equation, which describes the time dependence of the concentration, c, in a sedimentation velocity run. The program SEDFIT [277, 278] solves Lamm equations to fit experimental data and determine s and D. Sedimentation velocity can also be used to analyse macromolecular complexes and mixed species, by plotting the differential sedimentation coefficient distribution ($c(s)$) against a range of s values [278]. The

frictional ratio, f/f_0, can be calculated experimentally, and gives an indication of the molecular shape in solution: globular proteins have $f/f_0 \approx 1.2 - 1.3$, while elongated proteins have $f/f_0 > 1.5$.

In sedimentation equilibrium experiments, sedimentation and diffusion reach equilibrium, and D is eliminated from the Lamm equation. For a single, monodisperse macromolecule, a simple equation is the result:

$$M(1 - \bar{v}\rho)\omega^2 = \frac{2RTd(\ln c)}{dr^2}$$

where M is the macromolecule's molarity and c is the concentration of the solute at radial distance, r, from the centre of rotation. A plot of $\ln c$ against r^2 produces a straight line with a slope proportional to M. Sedimentation equilibrium can therefore provide an accurate estimate of the molecular weight of a macromolecule or a macromolecular complex uninfluenced by shape.

Samples for analytical ultracentrifugation were prepared at concentrations of 0.5-1.0 mg/ml in 20 mM HEPES (pH 7.5), and 150 mM NaCl. Experiments were carried out by David Staunton, using a Beckman An-60 Ti rotor in a Beckman Optima XL-1 analytical ultracentrifuge at 20°C. For both sedimentation velocity and sedimentation equilibrium runs, sedimentation behaviour was monitored by both UV absorbance (280 nm) and Rayleigh interference optics. Sedimentation equilibrium runs proceeded for 20 hours at different speeds.

Data were processed as follows. First, the buffer density and viscosity were calculated with SEDNTERP [279]. The sample partial specific volumes were estimated from the amino acid sequences, also using SEDNTERP. For sedimentation velocity data, multiple scans (recorded using interference optics) were fit to a continuous size distribution using SEDFIT [277, 278]. Sedimentation equilibrium data were fit to an ideal monodisperse model using SEDPHAT [280].

2.4.3 Native mass spectrometry

Ions of macromolecular complexes can be formed through electrospray ionization (ESI) and introduced and maintained intact in a mass spectrometer [281]. Crucially, because

it is a soft ionization technique, noncovalent molecular interactions can be preserved, allowing analysis of large, noncovalent complexes. In ESI, high (usually positive) electrostatic charging is applied to the tip of a capillary, leading to the emission of droplets of protein and solvent. As the drops travel down pressure and potential differences, they gradually desolvate until multiply-charged protein ions remain [282]. Nanoflow ESI is a miniaturized version of conventional ESI, which lowers the flow rate and thereby creates droplets of smaller size [283].

Hybrid mass spectrometers, which contain both a quadrupole (Q) mass filter and a time-of-flight (TOF) analyser, are used to analyse intact macromolecular complexes. The quadrupole consists of four parallel rods, and in normal mass spectrometry mode, functions as a simple ion guide, funnelling the ions to the TOF analyser. The TOF analyser is oriented orthogonally to the quadrupole. Ions with different m/z values will have different velocities and will thus reach the detector at different times. To correct for the large initial spread of kinetic energy of the ions, an ion mirror, or reflectron, is introduced. For ions with identical m/z values but different velocities, the faster ion will penetrate more deeply into the decelerating region of the reflectron, thus increasing its trajectory relative to the slower ion. As a result of these corrections, both ions should reach the microchannel plate (MCP) detector at identical times.

Samples for native mass spectrometry analysis were buffer-exchanged into 200 mM ammonium acetate pH 7.0 using BioSpin 6 columns (Bio-Rad). The samples were introduced into the instrument using nano-ESI gold-coated glass capillaries. These capillaries were prepared in-house, using a micropipette puller (Flaming/Brown P-97, Sutter Instruments) to pull glass tubes (outer diameter 1 mm, inner diameter 0.78 mm; Harvard Apparatus Ltd.) The pulled capillaries were coated with a thin layer of gold using an SEM sputter-coater (Polaron). The tapered end of the capillary was then trimmed under a stereomicroscope, to a diameter of 2-5 μm at the tip. 2 μl of sample (at a concentration between 1-10 μM) was introduced into the capillary.

Mass spectrometry data were collected on a modified Q-ToF II mass spectrometer (Waters, Manchester, UK), in positive mode, under conditions optimized for the efficient desolvation and transmission of high-molecular weight protein complexes. Typical instrument settings were: capillary 1.5-1.65 kV, sample cone 100-190 V, extraction cone 1-10 V, backing pressure 2.5-3.5 bar. When necessary, a backing pressure was applied to

the capillary to initiate or maintain a stable spray. The instrument was calibrated with 100 mg/ml cesium iodide. Data were acquired and processed using MassLynx software (Waters, Manchester, UK).

2.4.4 Analytical gel filtration

Gel filtration with a Superdex 200 10/300 GL column (GE Healthcare) was used to verify the formation of a PfRH5:basigin complex. Individual protein components, and the complex between them, were analysed in subsequent runs, and the traces overlaid to diagnose the change in elution profile upon complexation.

2.4.5 Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) can provide information about macromolecular shape, size, folding, aggregation, and assembly state in solution [284, 285]. The comparably low-resolution (∼50-10 Å) information gained from SAXS analysis can be complementary to high-resolution crystallographic studies (Section 2.5). As in X-ray crystallography, a highly collimated X-ray beam is used. In a SAXS experiment, the beam is applied to a biological sample in solution. Most of the X-rays pass directly through the solution, but some are elastically scattered by individual atoms in a molecule; the resultant coherent scattering can constructively or destructively interfere. The pattern of scattered radiation at very small angles (∼0.1-10°) is recorded on a detector. It is rotational averaging of the molecules free in solution that limits the resolution, in contrast to the ordered crystalline packing of proteins in X-ray crystallography.

SAXS is a contrast method, in which the scattering signal is derived from a combination of the electron density of the macromolecules and of the bulk solvent. The scattering pattern can be described by a relationship between the intensity (I) and the momentum transfer, q (units Å^{-1}): $q = 4\pi \sin\theta/\lambda$, where λ is the wavelength of the incident radiation, and θ is half the scattering angle between incident and scattered radiation. Since θ is small and λ fixed, $I(q)$ versus q is effectively the intensity as a function of scattering angle. The scattering profile can be subjected to an indirect Fourier transform to give the interatomic distance distribution function $P(r)$, which defines the frequency of interatomic vector lengths (r) in a protein. $P(r)$ calculation relies upon several assumptions:

namely, that $P(r) = 0$ when $r = 0$ and when $r = D_{max}$, where D_{max} is the maximum linear dimension. The $P(r)$ function gives information about the shape and volume of a protein.

Two useful parameters related to shape and size of the scattering particle are the zero-angle scattered intensity, $I(0)$, and the radius of gyration, R_g. The former is the intensity of radiation which is scattered through the zero-angle, and is determined by extrapolation. The R_g parameter represents the root-mean-squared distance of all scatterers from the centre of the particle, weighted by their scattering density. It is thus shape-dependent, and gives information about the mass distribution of a particle. $I(0)$ and R_g can be determined either by the Guinier approximation, an equation which results in a linear fit of $\ln I(q)$ versus q^2 for small q values only [286], or from the $P(r)$ function, a method which uses the full range of q values.

Small-angle X-ray scattering (SAXS) data were collected at beamline BM29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Scattering was detected using a Pilatus image reader at 20°C. Full-length PfRH5 complexes with basigin and with antibody fragments were prepared by mixing the individual proteins in equimolar amounts, and gel filtering on a Superdex 200 16/60 column in 20 mM HEPES pH 7.5, 150 mM NaCl. PfRH5–9AD4 samples were prepared by mixing PfRH5 and 9AD4 in equimolar amounts for 10 minutes. 60 μl samples were prepared in serial dilutions between 0.2-2.0 mg/ml.

Ten consecutive frames of 10 s each were recorded for each protein sample, with buffer samples measured in between. The frames were inspected for any evidence of radiation damage, and any affected images were excluded from further processing. Data were normalised to the intensity of the incident beam, images were averaged, and buffer scattering subtracted using PRIMUS [287, 288]. Guinier plots were inspected to assess the presence of aggregation [286]. Composite curves were generated by scaling and merging appropriate data sets. The radius of gyration (R_g) was determined from the Guinier plot using AutoRg [288]. The distance distribution function ($P(r)$) was obtained by indirect Fourier transform, generating an estimate of D_{max} and the Porod volume of the hydrated particle [288, 289]. An estimate of the molecular weight was obtained by dividing the Porod volume by 1.7. Twenty ab initio shape reconstructions were generated with DAMMIF [290] and averaged with DAMAVER [291]. The normalized spatial

discrepancy (NSD) parameter was used to diagnose the similarity of the reconstructions [292]. SITUS and SCULPTOR [293] were used to dock atomic structures into the SAXS envelopes.

2.5 X-ray crystallography

X-ray crystallography, like SAXS, involves an incident beam of X-rays being elastically scattered, or diffracted, in many directions, by the clouds of electrons enveloping the atoms in the sample. Unlike in SAXS, however, the sample is an ordered array of molecules arranged in a crystal lattice. During the experiment, the crystal is systematically rotated on a goniometer, leading to a set of diffraction patterns on the detector made up of diffraction spots or reflections. The three-dimensional projection of diffraction patterns make up a so-called reciprocal space. The positions and intensities of the reflections (providing information about frequency and amplitude, respectively) are combined with phase information to produce a three-dimensional map of the electron density within the crystal. In the following sections, practical aspects of the crystallography pipeline—from crystallisation to model building and refinement—are briefly considered.

2.5.1 Crystallisation and crystal optimisation

Protein crystals are grown under conditions in which molecules are driven out of solution in a reversible process. A highly concentrated protein solution (mg/ml) is mixed with a solution containing a precipitant, leading to a supersaturated solution. If suitable nucleation sites are present, crystals may grow. Different formats of crystallisation are possible, of which the most common is the vapour diffusion method, accomplished either in hanging-drop or in sitting-drop plates. A droplet containing protein is mixed with the crystallisation solution and equilibrated with a larger reservoir containing the crystallisation solution. As water evaporates from the drop, the concentration of both protein and precipitant increases until supersaturation occurs.

Protein crystallisation is a purely empirical process, which in practice entails screening a large number of crystallisation ‘cocktails’, containing various mixtures of different precipitants, buffers, additives, and detergents. Other variables can also be systematically

varied, such as protein concentration and temperature of crystallisation. The most common types of crystallisation screens are sparse matrix screens, which include conditions that have had prior crystallisation success and do not have a rational experimental design; and grid screens, in which two experimental parameters are varied in a systematic two-dimensional matrix. Commercial kits of sparse matrix screens in 96-well plate format, such as Morpheus, JCSG+, and Proplex (Molecular Dimensions), are often used to set up initial crystallisation trials. Once initial crystals have formed, optimisation of the initial crystallisation conditions is usually required to yield diffraction-quality crystals. Grid screens around the original crystallisation condition can be used to find the optimal growth conditions. Initial crystals can be harvested, added to a so-called Seed-Bead (Hampton Research), and vortexed to generate crystal seeds which, when added to fresh crystallisation drops, can act as nucleation sites for another round of crystallisation. Additive screens, like Silver Bullets (Hampton Research), can also be used for optimisation.

Protein crystals are extremely sensitive to ionising X-ray radiation, and during the course of an experiment, free radicals are formed that break the ordered crystal contacts and destroy diffraction potential. Data are now routinely collected at cryogenic temperatures to slow the destructive progress of these radicals, and lengthen the lifetime of the crystal. Hence, before X-ray exposure, protein crystals are 'fished' out of the crystallisation drop with so-called cryoloops and flash-cooled in liquid nitrogen. In order to prevent ice formation on cryocooling (which would destroy the diffraction potential of the crystal though the disruption of internal order, and the change in density) the crystal is first cryoprotected—that is, transferred briefly into a solution of mother liquor containing a sufficiently high concentration of a cryoprotectant, such as low molecular weight polyethylene glycol, ethylene glycol, or glycerol. These cryoprotectants act as anti-freeze, causing the water in the mother liquor to vitrify instead of freezing. In practice, a variety of cryoprotectants may have to be assayed, as different cryoprotectants can differentially affect the crystal order. In an ideal case, the mother liquor would already contain a suitable cryoprotectant, which could be supplemented as necessary to reach a suitable concentration for cryoprotection.

2.5.2 Data collection and initial processing

Data are collected at synchrotrons such as Diamond Light Source, typically with X-rays of wavelengths ∼0.9-1.0 Å. Initially, two images are taken, at 0° and 90°, of a mounted crystal in a cryoloop. If the resolution is sufficient, a data collection strategy is then devised, based upon the crystal orientation, instrument parameters, and the results of automated initial indexing (see below). Full datasets are collected only for single crystals that exhibit high-resolution diffraction and low mosaicity (see below), by oscillating the crystal through a small angle, recording the reflections in one frame, then oscillating and recording another frame, until all unique reflections have been recorded. For molecular replacement phasing, single-wavelength datasets are sufficient. Redundancy—multiple independent measurements of equivalent reflections—is vital to accuracy.

Processing of the data begins with indexing, cell refinement, and integration, accomplished by program iMOSFLM [294, 295]. Indexing allows the assignment of unit cell dimensions (a, b, and c, and the α, β, and γ angles between them), and gives information about the minimal symmetry of the crystal. 2-3 images at different rotation angles (and ∼50-100 strong reflections) are usually sufficient for unambiguous indexing. The angles of reflection, and thus the spacings of reflections observed on the detector, are determined by the reciprocal lattice spacings, which in turn are inversely related to the real-lattice spacings. Thus, the spacing between reflections can give precise information about the unit cell dimensions. In practice, a Fast Fourier transform algorithm is employed to find a consistent set of reciprocal lattice vectors; together with information about the origin of the image (inferred from the position of the incident primary beam on the detector), this determines the orientation matrix in reciprocal space, which is subsequently tested against further frames [296].

The unit cell parameters are refined, using one or several wedges (3-5°) of data. iMOSFLM also performs post-refinement, which 'rescues' partially measured reflections. Partial reflections arise because a reciprocal lattice point acts more like a finite sphere than an infinitesimal point, so a given oscillation during data collection will only partially record some reflections. Post-refinement assigns a partiality fraction to each partial reflection, which can be used as a correction factor to estimate its full intensity. Further, iMOSFLM generates an estimate of the mosaic spread of the crystal. (Mosaicity arises because crystals are actually built from domains which are imperfectly aligned.)

Next, the raw pixel intensities of the diffraction spots are integrated to determine the intensities of the reflections (I) and the background values (σI).

Next, the space group is determined using program Pointless [297]. The space group describes the symmetry of the contents of the unit cell. For example, crystals of space group $P2_12_12_1$ consist of a primitive lattice containing three perpendicular two-fold screw axes. (Along each screw axis, the asymmetric unit is rotated 180° and translated one-half of the length of the axis.) Symmetry elements result in systematically absent (zero-intensity) reflections, which can be inventoried to determine the space group. Pointless first scores possible lattice centring and point groups (into which space groups are classified based on their combination of rotational symmetry operators, ignoring the translational element of screw operators) according to putative symmetry-equivalent reflections. In the next step, it analyses systematically absent reflections to accurately determine the space group.

Data reduction is the next step. The measured intensities are spread over a number of frames; variations in diffraction power, the intensity of the X-ray beam, and the length of the X-ray path in different crystal orientations can result in inconsistencies in the absolute intensities observed. Hence, reflections of the same index are compared between frames, and the intensities rescaled accordingly. This process, known as scaling, is accomplished by program scala [297, 298]. Scala also merges multiple observations of the same reflection into a single, average intensity. The program Truncate [297] calculates the average structure factor amplitudes from the averaged intensities. A structure factor F is the sum of all the atomic scattering contributions to a given diffracted X-ray that produces a particular reflection, and is proportional to the square root of the intensity.

Data collection and merging statistics are important indicators of data quality. These statistics are reported not only for the entire dataset, but also for resolution shells (divisions of the dataset containing an equal number of reflections), since intensities are attenuated as resolution increases. The signal-to-noise ratio, $I/\sigma I$, drops accordingly at higher resolutions; a cut-off of $\sim$2 is reasonable. Redundancy, based upon the number of independent observations of a given reflection, is reported. The quality of the merged data—either for the dataset or for an individual resolution shell—is given by R_{merge}, the linear merging residual index:

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I_{(hkl)i} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{(hkl)i}}$$

where N is the number of redundant observations for reflection hkl, and $\bar{I}$ is the average intensity of each reflection. It is always inflated in the higher resolution shells because of the greater relative error in measuring weak reflections. A modified version of R_{merge}, which takes into account the multiplicity and the increasing precision as more and more observations are merged, is the precision-indicating merging R-value, or R_{pim}:

$$R_{pim} = \frac{\sum_{hkl} (\frac{1}{N-1})^{\frac{1}{2}} \sum_{i=1}^{N} |I_{(hkl)i} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{(hkl)i}}$$

which simply involves the incorporation of an extra term, $(\frac{1}{N-1})^{\frac{1}{2}}$, into the formula for R_{merge}.

Several statistics are reported relative to the 'batch number' (that is, the image number or time) to give an indication of the extent of beam damage over the course of the data collection. For example, in the absence of damage, R_{merge} should remain roughly constant across batch numbers, but will increase over time with radiation damage. The decision of where to truncate the dataset to compromise between beam damage and dataset completeness is somewhat empirical, but programs such as CHEF in xia2 can be helpful [299].

2.5.3 Phase determination by molecular replacement

The complex structure factors can be computed by summation of all of the scattered partial waves from all atoms of the unit cell. However, during data collection, the phase information for these structure factors is lost, and we are only able to calculate the scalar structure factor amplitude, by measuring the intensities of waves scattered from planes (hkl) in the crystal that give rise to reflection hkl on the detector. Once the phase is obtained, a Fourier transform relates the structure factors to the electron density as follows:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i((hx+ky+lz)-\phi_{hkl})}$$

where $\rho(x, y, z)$ is the electron density at position xyz in the unit cell, V is the volume of the unit cell, $|F_{hkl}|$ is the structure factor amplitude (proportional to $\sqrt{I_{hkl}}$), and ϕ_{hkl} is the missing phase of the reflection at coordinates hkl.

There are several strategies to recover the lost phase information, and thus convert structure factors to electron density. Molecular replacement is one such method, which borrows starting phases from a related (generally sequence identity ∼30% or greater) known structure. The known search model is rotated and translated in the unit cell or the asymmetric unit to arrive at a solution, representing the best fit between the calculated diffraction data from the molecular-replaced model and the data from the unknown structure. For structures with very high similarity to the model, this has become reasonably facile with programs such as Phaser [300]; structures with poor structural similarity, flexibly hinged subunits, or multiple copies of the same or different subunits, can be more challenging. The search model must be prepared to be maximally similar to the structure of interest, by removing solvent atoms, substituting non-similar amino acids, pruning unrelated regions, and separating flexibly linked domains as necessary.

There are two ways to implement a molecular replacement search: either a six-dimensional brute-force approach, which calculates the structure factors for the probe in different orientations and locations in the unit cell and computes a correlation score for each position; or the more widely used rotation-translation approach, which divides this search into a more manageable three-dimensional rotation search followed by a three-dimensional translation search, and is less computationally taxing.

In order to orient the search model, traditional molecular replacement relies upon calculation of the Patterson map, a map reconstruction that, crucially, does not require phases but only amplitudes, and contains interatomic distance information. The Patterson function $P(u, v, w)$ is as follows:

$$P(u, v, w) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

where (u, v, w) are coordinates at a particular point on the Patterson map. The Patterson contains interatomic distance vectors between all pairs of molecules, and self-vectors between equivalent atoms related by non-crystallographic symmetry (resulting from the presence of more than one identical molecule in the asymmetric unit). As a result, the

Patterson map is highly complex and cannot be used for structure determination (in the absence of heavy atoms), but it is characteristic for a given molecule and can be used as a 'fingerprint'.

For the rotation search, the intramolecular vectors (short distances of ∼10-4 Å) are analysed, since longer, intermolecular vectors will vary for each unit cell packing scenario. The search is performed in a suitably-sized 'box' to include only these intermolecular vectors. To determine the orientation of the search model, its Patterson map is rotated systematically, and the match to the Patterson map of the unknown molecule diagnosed by the linear correlation coefficient (CC_E) or the R-factor (see below). It is also possible to rotate the Patterson map of the unknown onto itself to determine the orientation of multiple copies related by non-crystallographic symmetry (a Patterson self-rotation search). These Patterson rotation searches are implemented by direct or fast rotation functions. A simple rotation function R can be expressed as:

$$R(\phi, \psi, \chi) = \int_{u,v,w} P^{target}(u, v, w) P^{model}[(u, v, w)(\phi, \psi, \chi)] dudvdw$$

At rotation angles (ϕ, ψ, χ), the rotation function R is equal to the integral of the product of two Patterson functions, that of the target and that of the model, the latter rotated by (ϕ, ψ, χ). Multiple solutions (derived from the highest CC_E or R-factor values) are then funnelled into the Patterson translation searches.

In the translation search, intermolecular vectors between the oriented molecules in the unit cell of the Patterson map become relevant, and intramolecular vectors become unnecessary. A fast translation function has thus been developed which excludes self-vectors. The location is then determined by matching intermolecular Patterson vectors using this fast translation function. The algorithm is sensitive to errors in rotation, hence the use of multiple rotation solutions. Packing functions implement an overlap penalty. The molecular replacement solutions then undergo rigid-body refinement to improve the orientation and translation of the molecule(s), and are re-ranked accordingly. Rigid-body refinement entails the minimisation of the sum of residuals squared between the calculated and observed structure factors. The final R-factor (R_{free}; see Section 2.5.5) should be less than 45-50% if the solution is correct. It is given by:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

In Phaser [300], Patterson methods have been replaced by maximum likelihood molecular replacement functions that significantly improve the target function. Maximum likelihood functions take errors and model incompleteness into account during the rotation and translation searches, which increases the chance of finding a solution. Since Phaser is more likely to succeed with poorer models, the final R-factor of a correct solution can be 55% or even higher.

The phases can be improved drastically by density modification procedures such as solvent flattening, with programs such as PARROT [301]. In brief, this involves tracing an envelope around the protein, creating a 'solvent mask', outside of which the density is set to a constant value. The masked map is back-transformed to structure factors, made up of the observed amplitudes but considerably improved phases, which are then transformed into a better map, and so on until convergence is reached.

2.5.4 Experimental phasing

Phase information can also be recovered experimentally, with no need for prior structural information or a related search model. In marker atom substructure methods, heavy atoms or anomalous scatterers act as marker atoms which cause the electronic structure to differ from an isomorphous reference structure. This gives rise to a different set of atomic structure factor amplitudes relative to the reference amplitudes; the resultant difference data can be used to locate the marker atom, and hence solve the substructure. This permits the phasing equations of all of the protein reflections to be solved.

Marker atoms can be heavy metal atoms that are co-crystallised with a protein, or soaked into pre-existing crystals to generate derivative crystals. Metals can react with specific side chains: for example, mercury interacts with cysteine sulphydryls. Metal soaks are largely trial–and–error, to find a metal that binds with high occupancy to several sites without destroying the diffraction. At the end of the soak, which can last hours to days, the crystals are back-soaked into mother liquor, before cryoprotection and cryo-cooling.

Proteins can also be derivatised for anomalous phasing (see below) with heavy halides (iodine or bromine) or heavy alkali metal ions (caesium or rubidium), by very briefly soaking the crystal in a high concentration solution before cryo-cooling. Proteins are often labeled with selenomethionine for anomalous phasing; labelling is achieved by expressing the protein in the presence of excess selenomethionine.

Substructure solution by exploiting the intensity differences between a native and an isomorphous derivatised crystal is isomorphous replacement (single, called SIR, or multiple, MIR). Heavy atoms (or native sulphurs or selenomethionines in the protein) can also be exploited in anomalous phasing experiments, more popular now than isomorphous replacement, in part because only a single crystal is required, bypassing the need for isomorphism. Anomalous scattering leads to an anomalous difference in magnitude between a given structure factor and its centrosymmetrically related mate (a so-called Bijvoet pair), the basis of single-wavelength anomalous diffraction (SAD). Furthermore, near absorption edges, a given reflection will have a different intensity at different wavelengths, leading to dispersive differences, a property exploited in multi-wavelength anomalous diffraction (MAD). Isomorphous replacement and anomalous diffraction can also be combined (SIRAS, MIRAS). Programs such as autoSHARP are complete structure solution packages, which solve the marker atom substructure and calculate initial protein phases [302].

2.5.5 Model building and refinement

Experimental phasing results in an initial, empty electron density map with poor phases, while molecular replacement results in an initial map with phases entirely derived from the repositioned search model. The molecular replacement phases are probably incomplete due to model trimming or missing regions, and in many places incorrect. The figure of merit statistic indicates the probability that a phase angle is correct. We account for this model bias by using maximum likelihood-based functions. The σ_A variance term accounts for both errors and incompleteness of the model, and allows calculation of so-called maximum likelihood Fourier map coefficients. Electron density maps calculated from these Fourier coefficients are minimally biased.

Building an atomic structure model into the electron density and real-space fitting is alternated with reciprocal space refinement of the model (Figure 2.2). Model building

is manually accomplished in programs such as Coot [303], usually by first building a Cα backbone into regions of obvious secondary structure. The program Buccaneer [304] automatically traces polyalanine protein chain fragments into the electron density. Large side-chains (e.g. tryptophan) and disulphide bridges can help to assign the correct sequence. Building into visible loops and connecting pieces of secondary structure comes next. After a continuous protein chain has been threaded through the electron density, residues can be fit into the electron density with local real space refinement. Likewise, stretches of backbone can be regularised according to real space geometry. Incorrect torsion angles are made manifest and fixed insofar as possible.

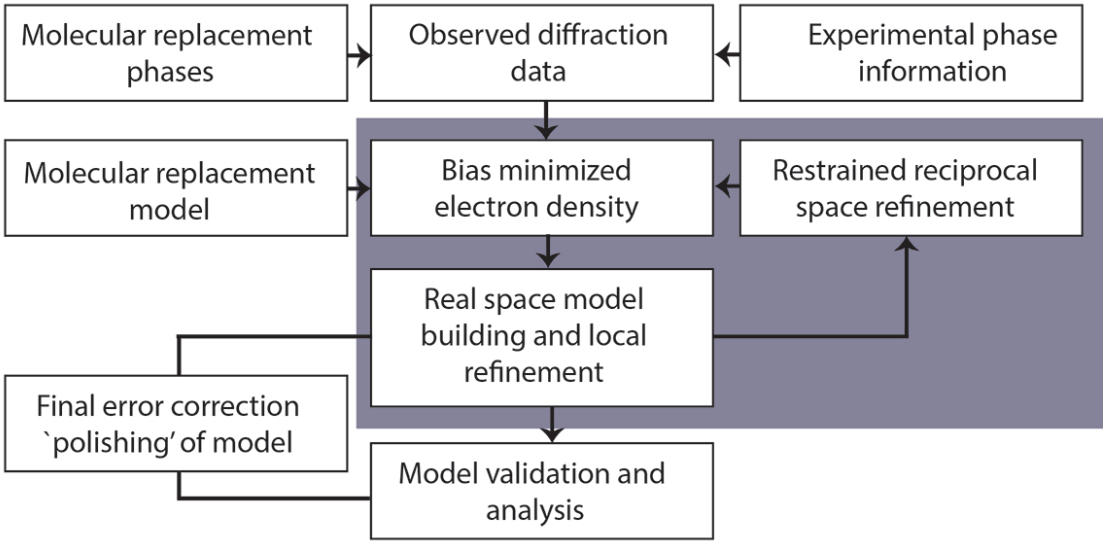


FIGURE 2.2: Flow diagram depicting the processes of model building and refinement. Model building occurs in real space by fitting residues into the electron density; refinement, by contrast, occurs in reciprocal space, and involves minimisation of the differences between observed and calculated structure factors. Model building and refinement are performed in an interative cycle (grey box), until the model contains as few errors are possible. Stereochemical restraints are also inputted into reciprocal space refinement. Taken and adapted from [305].

Once the initial model is built, a restrained refinement in reciprocal space is carried out. The refinement optimises the parameters of individual atoms (namely, the coordinates xyz and the B-factors), as well as global parameters (including the overall B-factor and bulk solvent corrections). The B-factor is a measure of the atomic mobility of an atom, or the vectorial displacement from its average position. In practice, this positional displacement arises through thermal motion, the existence of several partially occupied conformations, or lattice disorder. These parameters are altered to optimise the fit between experimental observations and their computed equivalents. Experimental observations are not only the experimental structure factor amplitudes (the data), but also

a variety of restraints valid for all models. These latter include stereochemical restraints (allowed bond angles, bond lengths, and chirality etc.), non-crystallographic symmetry restraints (which take into account the geometry and redundancy between identical molecules in the unit cell), and B-factor restraints. For refinement to be possible, there must be more data points n than there are parameters p to optimise (since refinement is effectively the solving of p simultaneous equations); thus, the more restraints we can add, and the higher the data-to-parameter ratio, the better.

In statistical terms, the refinement procedure aims to minimise the global linear residual between the structure factor amplitudes determined experimentally and those computed from the model structure (the R-factor; see Section 2.5.3). The R-factors are calculated not only for the working dataset (R_{work}), but also for a subset ($\sim$5%) of the experimental data that was set aside for cross-validation (R_{free}). R_{free} is related to the mean phase error, and is thus a good metric to describe phase (and model) accuracy. Maximum likelihood statistics are almost universally employed for reciprocal refinement procedures, since they account for both errors and incompleteness in the model. A typical round of reciprocal space refinement involves cycles of parameter adjustment until convergence is reached. The program Buster uses Bayesian statistical methods for reciprocal space refinement [306].

After reciprocal space refinement and recalculation of the electron density map with maximum likelihood map coefficients, the resultant electron density map will be better. Inspection of the electron density and the difference map allow further real space refinement and model building, followed by another round of reciprocal space refinement, and the cycle of model building and refinement continues.

When suitable R-factors have been obtained, further real space refinement is not possible, and there are no outliers in the disallowed regions on a Ramachandran plot (determined by programs such as PROCHECK [307]), refinement is 'abandoned,' [308] rather than ever truly completed. In addition to the final R_{work} and R_{free}, other refinement statistics reported include the number of unique reflections, which is estimated based on the volume of the unit cell and the symmetry, or calculated from the structure factors. The root mean square deviation (r.m.s.d.) of bond lengths and angles is also reported.

Chapter 3

Expression, Purification, and Functional Validation of PfRH5

3.1 Summary

The first step towards any biophysical or structural analysis of PfRH5 was its recombinant overexpression and purification, followed by confirmation that the resultant pure protein is functional. I used information gleaned from a bioinformatics survey to design a number of constructs lacking putative disordered regions or subdivided into potential globular domains. Initially, I transformed this collection of constructs into fast-growing *E. coli*, only to find the proteins expressed insolubly, if at all. Expression in mammalian cells and baculovirus-infected insect cells was no more successful. However, the Schneider 2 cell system proved capable of expressing soluble, secreted PfRH5. Using these cells, I produced both the full-length protein and two shortened constructs, engineered to lack disordered regions in order to increase the likelihood of crystallisation. These constructs could be purified to homogeneity by a multi-step purification, proceeding through two affinity chromatography steps and gel filtration. All three constructs were demonstrated to bind the erythrocyte receptor basigin, and to a panel of mouse monoclonal antibodies recognising PfRH5, indicating that they are all correctly folded and functional. Furthermore, they are capable of inducing a polyclonal immunoglobulin response which potently ablates parasite growth *in vitro*. This emphasises that the disordered regions

are not essential for function or immunogenicity, and makes the constructs lacking them suitable candidates for crystallisation trials.

3.2 Bioinformatic survey of PfRH5

To guide the design of PfRH5 protein constructs, a comprehensive bioinformatics analysis was carried out. PSI-BLAST searches did not identify any homologues of PfRH5 outside the *Plasmodium* RH protein family. Further, FUGUE, an algorithm for identifying distant structural homologues based on the protein sequence, did not identify a known fold within the RH protein sequence. Thus, in the absence of information from non-existent homologues, we sought to annotate the sequence of PfRH5 using a combination of disorder predictions, secondary structure predictions, and alignments with other PfRH proteins.

Long disordered regions, rich in lysines, glutamates, aspartates, and serines, are common in malaria proteins [254, 309]. Indeed, PfRH5 is predicted to contain two such disordered amino acid stretches, one at the N-terminus (residues ∼44-80) and one in a central region of the protein (residues ∼255-295) (Figure 3.1). Another region near the N-terminus (residues ∼120-133) might also be disordered. Secondary structure prediction suggested that PfRH5 is predominantly helical, with several very short β-strands (Figure 3.1).

Phylogenetic analysis of RH protein sequences indicates a closer evolutionary relationship between PfRH5, PfRH2b, pseudogene PfRH3 and PvRBP-2 than other *P. falciparum* and *P. vivax* RH proteins [187] (Figure 1.9). An alignment of these sequences, together with RH5 from *P. reichenowi* (one of the only other species known to contain RH5), is shown in Figure 3.2. It reveals that the disordered central region of PfRH5 is missing in its homologues, suggesting it may be an insertion into a loop or linker between two domains. Curiously, exactly three cysteine residues are fully conserved in this clade; an additional, almost conserved cysteine—only missing in PfRH2b—is also found (PfRH5 C317). This suggests the formation of at least one disulphide bond, and probably two in most homologues. PfRH5 contains five or six cysteines, depending on the strain (residues 329 and 204, the latter of which is polymorphic and sometimes encoded by tyrosine). The remaining cysteine(s) are not conserved across the homologues.

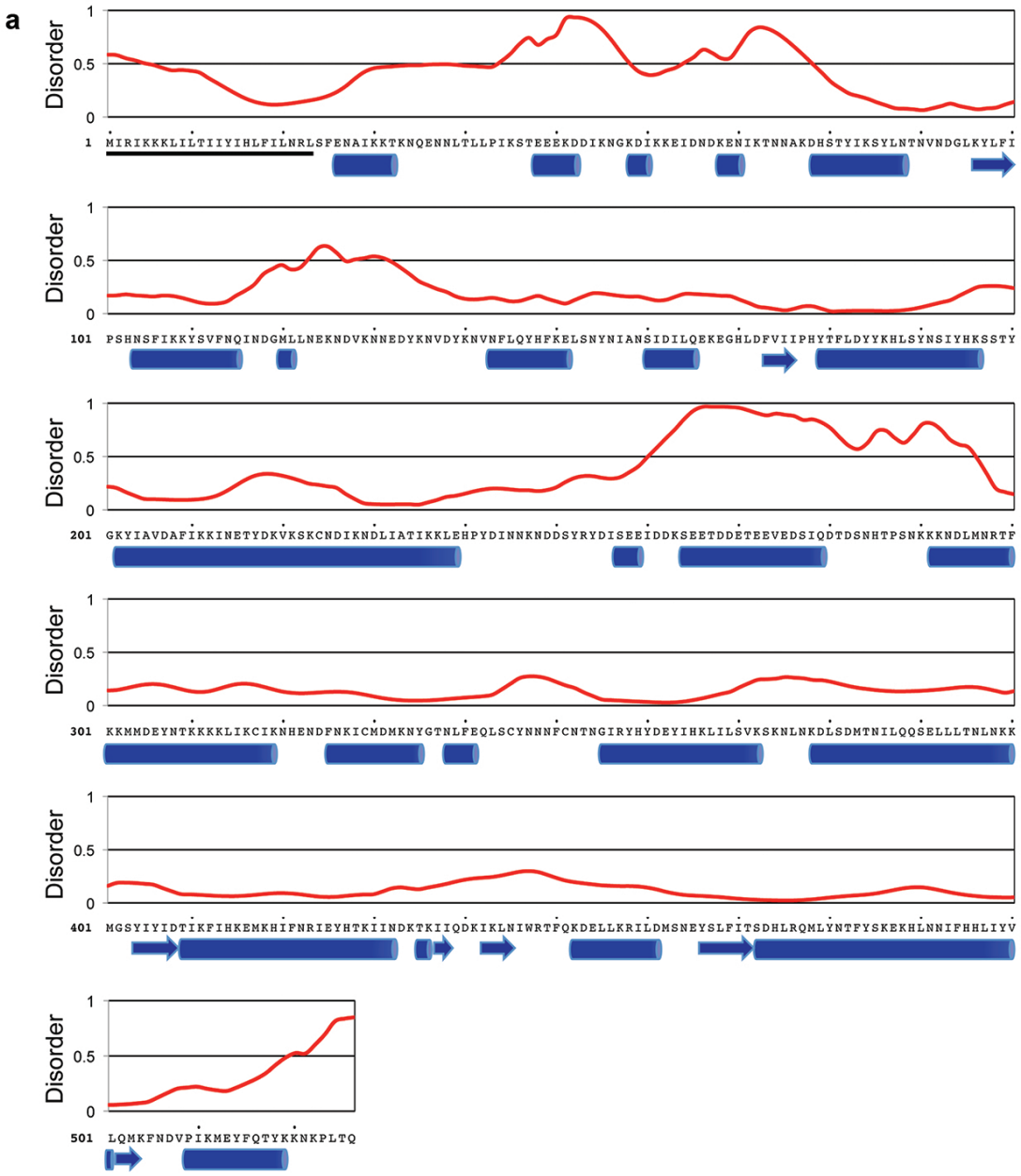


FIGURE 3.1: The predicted long-range disorder [260] is shown above, and the predicted secondary structure [263] below, the sequence of PfRH5. Two regions (residues ∼44-80 and ∼255-295; the exact boundaries vary by prediction program) are predicted to be disordered (values >0.5). Residues 120-133 might also be disordered. Blue rods indicate predicted α-helical secondary structure, and blue arrows indicate β-strands. The signal sequence [265] is underlined in black.

Thus, PfRH5 probably forms two disulphide bonds, and contains 1-2 unpaired cysteines.

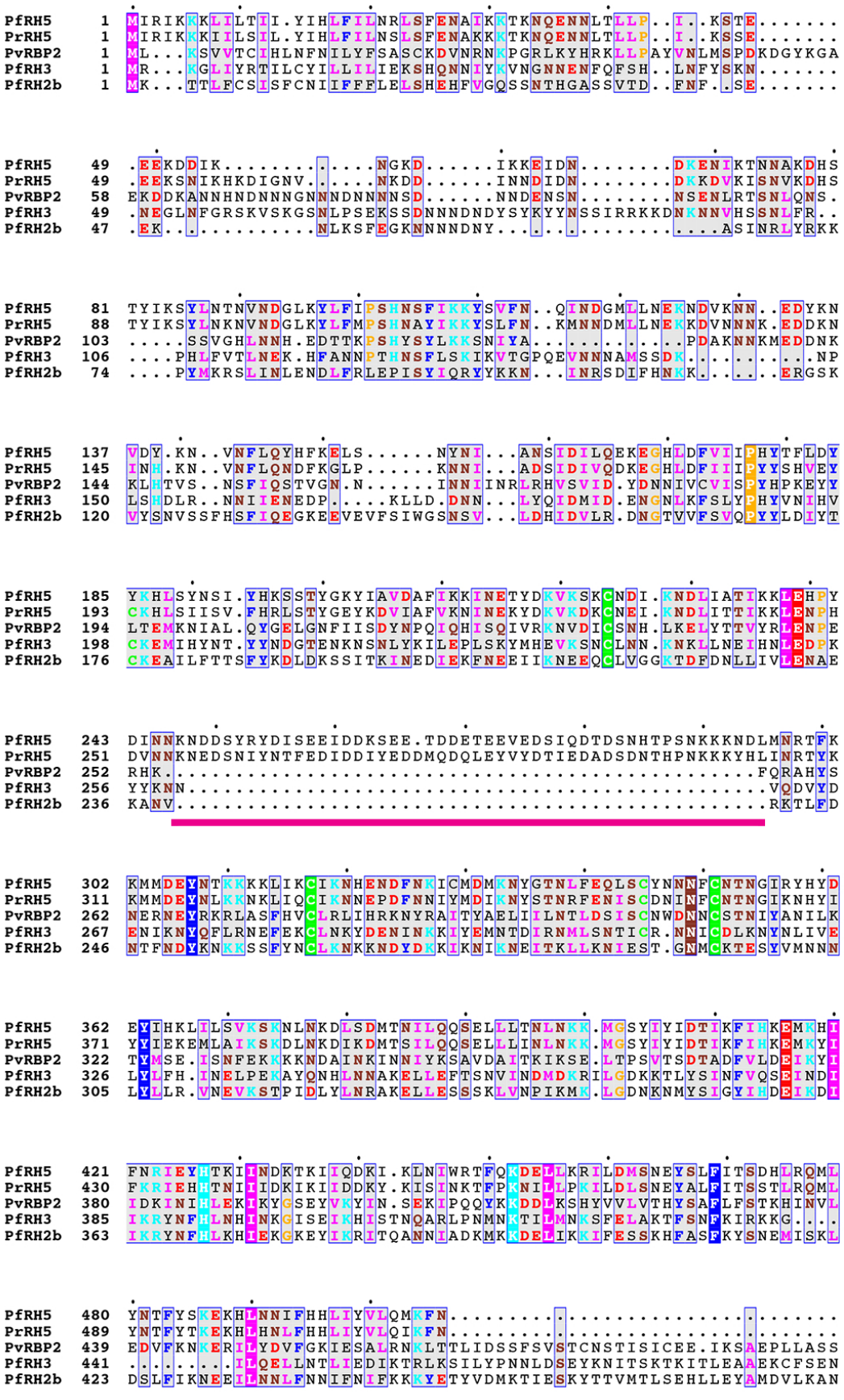


FIGURE 3.2: Sequences of the PvRBP-2-like clade were aligned using T-COFFEE [267] and the figure prepared using Espript [268]. The pink underline indicates the inserted loop in PfRH5. Residues are coloured by their physiochemical properties: HKR, cyan; DE, red; STNQ, maroon; AVLIM, pink; FYW, blue; PG, orange; C, green. Similar residues (global similarity score >0.5) are boxed in blue, and identical residues are shown in white, highlighted with the appropriate colour.

3.3 Expression and purification of PfRH5 from *E. coli*, HEK cells, and baculovirus-infected insect cells

The expression of functional, soluble *Plasmodium* proteins remains a major hindrance in the study of the biology of malaria and a bottleneck in the design of novel antimalarial drugs or vaccines [310]. This is one reason for the relative dearth of structures of plasmodial surface proteins. I adopted a multi-pronged approach to PfRH5 expression: I would assay different expression systems (bacterial, insect, and mammalian cells), different expression tags, different constructs designed based on bioinformatics (Section 3.2), and different N- and C-terminal boundaries for these constructs.

Escherichia coli generally remains the preferred host for recombinant protein expression, due to the speed and ease of genetic manipulation. Soluble expression of plasmodial proteins is often impeded in *E. coli* by high molecular weight, protein disorder, basic isoelectric point (pI), codon usage, lack of necessary post-translational modifications, and lack of homology to *E. coli* proteins [310]. Nonetheless, there have been notable successes in expression of malarial proteins in *E. coli*, including extracellular proteins: for example, the structure of AMA-1 was elucidated using protein refolded from inclusion bodies [247].

Expression in eukaryotic cells can circumvent some of the potential problems of bacterial expression: pathways of folding, assembly, and post-translational modification should more closely resemble these processes in *Plasmodium*. In baculovirus-infected insect cell expression systems, the gene of interest is subcloned into a transfer vector, which is co-transfected with viral DNA encoding the *Autographa californica* nuclear polyhedrosis virus into *Spodoptera frugiperda* Sf9 cells. Homologous recombination leads to stable integration of the gene of interest into the viral genome, and gene expression—driven by a strong viral promoter—in the infected Sf9 cells. *Plasmodium* proteins successfully expressed in the baculovirus system include MSP-1 [311], AMA-1 [312], and the VAR2CSA PfEMP1 variant [313].

Mammalian cells have been used extensively to target proteins to the cell surface, to aid our understanding of the molecular basis of host cell binding. For example, red blood cell binding of EBA-175, AMA-1, and PfAARP, or domains of them, have all been analysed by expression on the surface of COS-7 cells [310]. Human embryonic kidney 293 (HEK)

cells have been used to secrete recombinant proteins such as PfRH5 (see below) [238]. However, mammalian cells often give substantially lower yields than other expression systems (e.g. yeast), and can be harder to scale up, which can limit their usefulness in preparative expression [310].

A previous study produced full-length PfRH5 (residues 26-526, with only the N-terminal signal sequence removed) in HEK cells in quantities sufficient for biophysical analysis [238]. The construct was tagged at the N-terminus with immunoglobulin domains 3-4 of mouse Cd4. It was shown to be functional in basigin-binding experiments. Other attempts to synthesise PfRH5 recombinantly have involved refolding either the full-length protein or various fragments from *E. coli* inclusion bodies [187, 236, 244]. However, these may not be correctly folded, as no evidence was proffered to support correct refolding or functionality (e.g. circular dichroism, or basigin-binding experiments for the study [244] published after the discovery of the basigin interaction).

3.3.1 Full-length and shortened full-length constructs

In an attempt to recapitulate or improve upon the published HEK cell expression, I first attempted to express full-length PfRH5, residues 26-526, lacking only the N-terminal signal sequence—the same construct boundaries as described previously [238]—in three different expression systems. Instead of the Cd4 tag, I used a 6-His tag for these initial experiments. The reasons for the replacement were two-fold. Firstly, the bulky and flexibly attached Cd4 tag would likely prevent crystallisation, and would need to be removed. However, there were concerns about the removability of the Cd4 tag (G. Wright, personal communication), and about a possible loss of solubility of PfRH5 upon its removal. Secondly, the 6-His tag is a short, convenient tag that can be used in all expression systems. The gene used for these constructs, and for all subsequent constructs in this section, was a synthetic gene codon optimised for mammalian expression. The *PfRH5* gene was cloned into *E. coli* expression vector pET15b$_{con}$, encoding an N-terminal His tag; into Sf9 vector pAcGP67A$_{con}$, encoding an N-terminal His tag; and into mammalian expression vectors pEF$_{bos}$ and pHLsec, encoding an N-terminal and a C-terminal His tag, respectively.

Expression was diagnosed by Western blot; representative expression results for these constructs are shown in Figure 3.3. In *E. coli*, the protein was overwhelmingly insoluble,

and appeared moreover to be degraded into smaller fragments during expression. Some expression was observed in Sf9 cells; none at all was observed in HEK cells. Attempts to purify PfRH5 from the soluble or insoluble fraction of *E. coli*, or from Sf9 or HEK supernatants, were unsuccessful: no protein was visible by Coomassie stain after Ni^{2+}-NTA chromatography (data not shown).

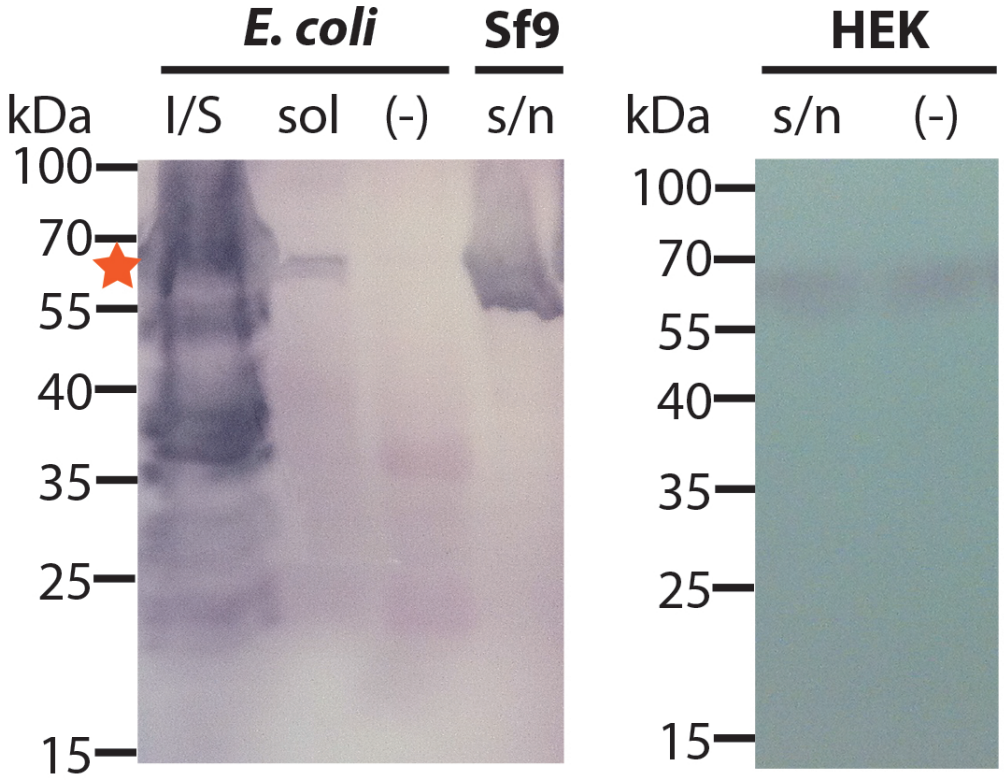


FIGURE 3.3: Anti-6-His Western blots to assess expression of full-length PfRH5 (26-526) in three expression systems. The orange star indicates the expected position of migration of PfRH5 (∼63 kDa). In *E. coli*, most protein expressed in Origami B (DE3) cells from the pET15b$_{con}$ vector is insoluble (I/S) rather than soluble (sol), and appears to be degraded into smaller fragments. Data from other bacterial strains are similar. The soluble protein from *E. coli* and the protein from Sf9 cell supernatant (s/n) (expressed from plasmid pAcGP67A$_{con}$) could not be purified. No protein was observed in HEK cell supernatant; the faint band also visible in the negative control [(-)] is a cross-reaction of the anti-His antibody with BSA from foetal calf serum in the growth media. This blot shows protein expressed from the mammalian pEFbos$_{con}$ vector, but results for the pHLsec vector were comparable.

Next, the construct was pruned significantly at the N-terminus, to remove the predicted disordered region(s) at the beginning of the protein (Section 3.2). A first round of shortened constructs removed the first disordered region (residues ∼44-80), by beginning at residue 68 or 76. I also designed constructs with slightly shorter C-termini (ending at

residue 511 or 522), to remove increasingly disordered residues at the extreme C-terminus (Figure 3.1).

I decided *E. coli* was the best expression system in which to test these constructs, because of its speed and ease of use. Many of these constructs were cloned and expressed in a high-throughput pipeline at the Oxford Protein Production Facility (OPPF). In the OPPF pipeline, constructs with each combination of N- and C-terminal boundaries were subcloned into several related vectors with different expression tags (the pOPIN series). All vectors contained an N-terminal 6-His tag, either by itself (pOPINF), or with maltose binding protein (pOPINM), SUMO (pOPINS3C), or thioredoxin (pOPINTRX). The final vector contained a supplemental C-terminal biotinylation peptide (pOPINFB). Soluble and insoluble expression of these constructs were tested in a variety of *E. coli* cell lines. Depending on the construct, these could include B834, methionine auxotrophs useful for selenomethionine labelling; B834 pMJ59 cells, containing a plasmid encoding a disulfhydryl oxidase and a disulphide isomerase to aid disulphide bond production; Origami B (DE3) cells, containing mutations to aid disulphide bond production; Rosetta cells, supplying tRNAs for commonly used codons in eukaryotes that are underused in *E. coli*; Rosetta-gami 2 (DE3) cells, combining the properties of Rosettas and Origamis; and ArcticExpress cells, which encode folding chaperoninins and can be grown at low temperatures to aid protein folding. The level of expression was assessed by high-throughput, small-scale Ni^{2+}-NTA chromatography, either directly from the clarified cell lysate (to test soluble expression), or from inclusion bodies solubilised in buffer containing 8 M urea (to test insoluble expression).

Expression results were disappointing (Table 3.1, construct numbers 12-26): no soluble or insoluble expression of any of the shortened full-length constructs was observed in any of the cell types tested. Construct 68-526 was also transiently transfected into HEK cells; insoluble expression within the cells was observed, but no soluble protein was secreted.

A further two constructs were designed, which additionally removed the next putative disordered region (residues ∼120-133), by beginning at residue 140 or 150. These constructs are estimated to be ∼40 kDa, similar to the 45-kDa fragment isolated from parasite supernatant [187, 202]. These constructs, subcloned into pET15b$_{con}$ with an N-terminal 6-His tag, did not express either (Table 3.1, constructs 57-58). Thus, production of a shortened full-length construct of PfRH5 was not successful in an *E. coli*

No.	Vector	Organism	aa_N	aa_C	Sol.	Insol.	Cells
1	$pET15b_{con}$	*E. coli*	26	526	no	yes	OB
12	pOPINF	*E. coli*	68	511	no	no	
13	pOPINFB	*E. coli*	68	511	no	no	
14	pOPINM	*E. coli*	68	511	no	no	
15	pOPINF	*E. coli*	68	511	no	no	
16	pOPINFB	*E. coli*	68	511	no	no	
17	pOPINM	*E. coli*	68	511	no	no	
18	pOPINF	*E. coli*	68	526	no	no	
19	pOPINFB	*E. coli*	68	526	no	no	
20	pOPINM	*E. coli*	68	526	no	no	
21	pOPINF	*E. coli*	76	511	no	no	
22	pOPINM	*E. coli*	76	511	no	no	
23	pOPINF	*E. coli*	76	522	no	no	
24	pOPINM	*E. coli*	76	522	no	no	
25	pOPINF	*E. coli*	76	526	no	no	
26	pOPINM	*E. coli*	76	526	no	no	
56	$pET15b_{con}$	*E. coli*	76	522	no	no	
57	$pET15b_{con}$	*E. coli*	140	522	no	no	
58	$pET15b_{con}$	*E. coli*	150	522	no	no	
81	pHLsec	HEK	26	526	no	no	
83	pHLsec	HEK	68	526	no	yes	
85	$pEFbos_{con}$	HEK	26	526	no	no	
86	$pAcGP67A_{con}$	Sf9	26	526	no	no	

TABLE 3.1: PfRH5 full-length and shortened full-length expression results. The construct number corresponds to the primer tables in Appendix A. Cell types in which expression was observed are indicated. Other abbreviations: no., construct number; aa_N and aa_C, the N- and C-terminal boundaries of the construct, respectively; sol., soluble protein expression; insol., insoluble protein expression; OB, Origami B (DE3) cells.

expression screen.

3.3.2 Single-domain constructs

Meanwhile, I also attempted to cut PfRH5 into two distinct domains, by severing the extended disordered linker or loop (residues ∼255-295). N-terminal constructs began after the first putative disordered region in the N-terminus (residues 68 or 76), and ended at residue 250. C-terminal constructs began at residue 255, 277, 288, 292, 322, 347, or 366, and terminated at residue 511, 522, or 526. Cloning and expression trials were initially carried out at the OPPF (Section 3.3.1).

The putative N-terminal domain (residues 76-250) expressed solubly as a fusion with maltose binding protein (MBP) in several *E. coli* strains (Table 3.2, construct 10).

Attempts to purify it using a combination of Ni^{2+}-NTA chromatography and amylose chromatography, followed by cleavage of the 6-His-MBP tag with 3C protease, were an unmitigated failure, with no protein remaining after protease cleavage. This suggests that MBP might have been stabilising the protein in a soluble form; upon its removal, the protein precipitated. The fact that this construct did not express with any other tag lends credence to this idea.

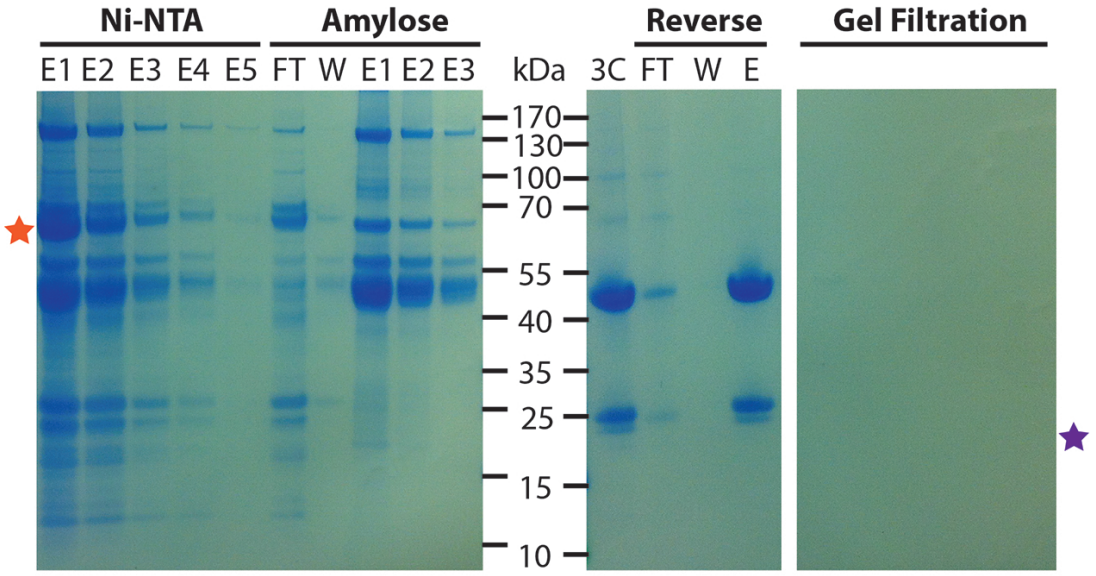


FIGURE 3.4: A representative attempt to purify the PfRH5 N-terminal domain with an N-terminal MBP tag. The protein was first subjected to Ni^{2+}-NTA chromatography (Ni-NTA), followed by amylose chromatography. The fusion protein should be ~63 kDa (orange star): some contaminants also running at this molecular weight (perhaps chaperones) flow through the amylose column, leaving a small amount of the fusion protein, which elutes from the amylose column (E1-E3). The protein was cleaved overnight with 3C protease (the recognition site for which is inserted between 6-His-MBP and the N-terminal domain of PfRH5), leaving 6-His-MBP (43 kDa). It should also leave ~21 kDa PfRH5 N-terminal domain (purple star), but the protein visible at this molecular weight is in fact 6-His-tagged 3C protease, which elutes from a tandem Ni-NTA-amylose column (reverse) along with MBP. Very little protein (presumably some or all of which is 3C protease instead of PfRH5) flows through the Ni-NTA-amylose reverse column, and no PfRH5 N-terminal domain is visible in gel filtration fractions. Abbreviations as follows: FT, flow-through; W, wash; E, elution; 3C, 3C protease-cleaved sample; reverse, reverse Ni-NTA and amylose chromatography.

The putative C-terminal domain (residues 292-522) consistently expressed in inclusion bodies, with the range of tags and in a number of cell lines (Table 3.2, constructs 38-42). The construct with the 6-His tag (construct 38) was selected as the most tractable for protein refolding studies (without unwieldy MBP or other large tags), and was cloned into pET15b$_{con}$ and expressed in B834 pMJ59 cells. To purify the 6-His tagged C-terminal domain from the insoluble fraction, I first solubilised the inclusion bodies in

No.	**Vector**	**Domain**	**aa$_N$**	**aa$_C$**	**Sol.**	**Insol.**	**Cells**
2	pOPINF	NTD	68	250	no	no	
3	pOPINFB	NTD	68	250	no	no	
4	pOPINTRX	NTD	68	250	no	yes	R
5	pOPINM	NTD	68	250	no	no	
6	pOPINS3C	NTD	68	250	no	yes	
7	pOPINF	NTD	76	250	no	yes	
8	pOPINFB	NTD	76	250	no	yes	
9	pOPINTRX	NTD	76	250	no	yes	
10	pOPINM	NTD	76	250	yes	yes	MJ,R,RG
11	pOPINS3C	NTD	76	250	no	yes	R
27	pOPINF	CTD	255	526	no	no	
28	pOPINM	CTD	255	526	no	no	
29	pOPINF	CTD	277	526	no	no	
30	pOPINM	CTD	277	526	no	no	
31	pOPINF	CTD	288	526	no	no	
32	pOPINM	CTD	288	526	no	no	
33	pOPINF	CTD	292	511	no	no	
34	pOPINFB	CTD	292	511	no	no	
35	pOPINTRX	CTD	292	511	no	no	
36	pOPINM	CTD	292	511	no	no	
37	pOPINS3C	CTD	292	511	no	no	
38	pOPINF	CTD	292	522	no	yes	R
39	pOPINFB	CTD	292	522	no	yes	R
40	pOPINTRX	CTD	292	522	no	yes	R,RG
41	pOPINM	CTD	292	522	no	yes	R,RG,AE
42	pOPINS3C	CTD	292	522	no	yes	R,RG
43	pOPINF	CTD	292	526	no	no	
44	pOPINFB	CTD	292	526	no	no	
45	pOPINTRX	CTD	292	526	no	no	
46	pOPINM	CTD	292	526	no	no	
47	pOPINS3C	CTD	292	526	no	no	
48	pOPINF	CTD	296	526	no	no	
49	pOPINM	CTD	296	526	no	no	
50	pOPINF	CTD	322	526	no	no	
51	pOPINM	CTD	322	526	no	no	
52	pOPINF	CTD	347	526	no	no	
53	pOPINM	CTD	347	526	no	no	
54	pOPINF	CTD	366	526	no	no	
55	pOPINM	CTD	366	526	no	no	
59	pET15b$_{con}$	NTD	76	250	no	yes	MJ
60	pET15b$_{con}$	CTD	292	522	no	yes	MJ
82	pHLsec	NTD	68	250	no	yes	HEK
84	pHLsec	CTD	292	526	no	yes	HEK

TABLE 3.2: PfRH5 N- and C-terminal domain expression results. The construct number corresponds to the primer tables in Appendix A. Cell types in which expression was observed are indicated. Other abbreviations as in Table 3.1, with additional *E. coli* cell types R, Rosetta; RG, Rosetta-gami; MJ, B834 pMJ cells; AE, Arctic Express cells.

buffer containing either 6 M guanidine or 8 M urea. I then attempted to refold on a Ni^{2+}-NTA column, by reducing the amount of 6 M guanidine in the buffer in a stepwise manner, until there was none remaining; only trace amounts of protein remained after on-column refolding, which were not visible by gel filtration. I also tried refolding by flash dilution: protein first purified by Ni^{2+}-NTA chromatography in 8 M urea was rapidly diluted into a larger (20x) volume. Small-scale trials were carried out to test flash dilution into 20 different buffers, and the results analysed by SDS-PAGE and Coomassie staining (Figure 3.5). Attempts to scale-up the refolding process to larger quantities of PfRH5 were thwarted by an inability to concentrate the protein in Amicon concentrators (which is a common diagnostic of aggregated, misfolded protein), and a subsequent lack of any peak apart from the void volume on a Superdex 200 16/60 gel filtration column. It was not possible to perform circular dichroism analysis of the nominally refolded protein, since buffer exchange–by either concentration or gel filtration—was not possible. These abortive attempts to functionally refold the putative C-terminal domain were thus abandoned.

The N-terminal (76-250) and C-terminal (292-522) constructs were also cloned into pHLsec for HEK cell expression trials. The proteins were retained within cells insolubly rather than secreted (Table 3.2, constructs 82 and 84), and were not pursued further.

3.3.3 Deletion constructs

It is quite possible the the central disordered region is in fact an intradomain loop, rather than an interdomain linker, and that an insertion into this loop in PfRH5 expanded it relative to the other RH proteins (Figure 3.2). The complete lack of expression of full-length and shortened full-length PfRH5 constructs in *E. coli*, when both the N-terminal domain and the C-terminal domain (which exclude the central disordered region between them) expressed, at least insolubly, suggests that the central disordered region might be hindering expression; indeed, low-complexity disordered regions commonly do so in *E. coli* [310]. Thus, I next designed constructs with an internal truncation of the disordered loop, based on alignment with the other RH proteins: residues 243-291, 246-294, 240-288, 247-295, or 243-291 were omitted, from full-length or shortened full-length constructs (residues 26-522, 76-522, 140-522, or 150-522).

The constructs were cloned into pET15b$_{con}$ and expressed in B834 pMJ59 cells. All but one of the deletion constructs expressed in inclusion bodies (Table 3.3). Removal of the loop clearly made full-length PfRH5 expression, albeit insoluble expression, possible in *E. coli*. Expression of constructs 64 (residues 76-522, lacking 248-296) and 69 (residues 140-522, lacking 248-296) was scaled up, and protein refolding attempted. The inclusion bodies were solubilised in 8 M urea buffer, purified by Ni^{2+}-NTA chromatography, and flash diluted into refolding buffer 13 without 0.5% Tween (Table 2.2). The protein could not be concentrated, however: all of the protein stuck to the Amicon centrifugal filter device membrane, implying unsuccessful refolding.

No.	**Vector**	**Deletion**	**aa$_N$**	**aa$_C$**	**Sol.**	**Insol.**	**Cells**
61	pET15b$_{con}$	FLΔ243-291	76	522	no	yes	MJ
62	pET15b$_{con}$	FLΔ246-294	76	522	no	yes	MJ
64	pET15b$_{con}$	FLΔ248-296	76	522	no	yes	MJ
65	pET15b$_{con}$	FLΔ247-295	76	522	no	yes	MJ
66	pET15b$_{con}$	FLΔ243-291	140	522	no	yes	MJ
67	pET15b$_{con}$	FLΔ246-294	140	522	no	yes	MJ
68	pET15b$_{con}$	FLΔ240-288	140	522	no	yes	MJ
69	pET15b$_{con}$	FLΔ248-296	140	522	no	yes	MJ
70	pET15b$_{con}$	FLΔ247-295	140	522	no	yes	MJ
71	pET15b$_{con}$	FLΔ243-291	150	522	no	yes	MJ
72	pET15b$_{con}$	FLΔ246-294	150	522	no	yes	MJ
73	pET15b$_{con}$	FLΔ240-288	150	522	no	yes	MJ
74	pET15b$_{con}$	FLΔ248-296	150	522	no	yes	MJ
75	pET15b$_{con}$	FLΔ247-295	150	522	no	yes	MJ
76	pET15b$_{con}$	FLΔ243-291	26	522	no	yes	MJ
77	pET15b$_{con}$	FLΔ246-294	26	522	no	no	
79	pET15b$_{con}$	FLΔ248-296	26	522	no	no	MJ
80	pET15b$_{con}$	FLΔ247-295	26	522	no	yes	MJ

TABLE 3.3: PfRH5 deletion construct expression results. The deleted residues are indicated by Δ. The construct number corresponds to the primer tables in Appendix A. Other abbreviations as in Table 3.1.

3.3.4 Co-expression and fusion of PfRH5 and basigin

Several lines of reasoning converged on the idea of co-expressing PfRH5 and basigin. Co-expression is commonly used for conformational stabilisation [314], and basigin expresses prolifically in *E. coli*. Furthermore, the previously published construct [238] contained a Cd4 tag, consisting of two Ig domains (domains 3-4): could the tag be mimicking basigin, which is also comprised of two Ig domains, or even PfRIPR, the parasite binding partner of PfRH5, comprised of EGF domains? (The tag did not prevent basigin

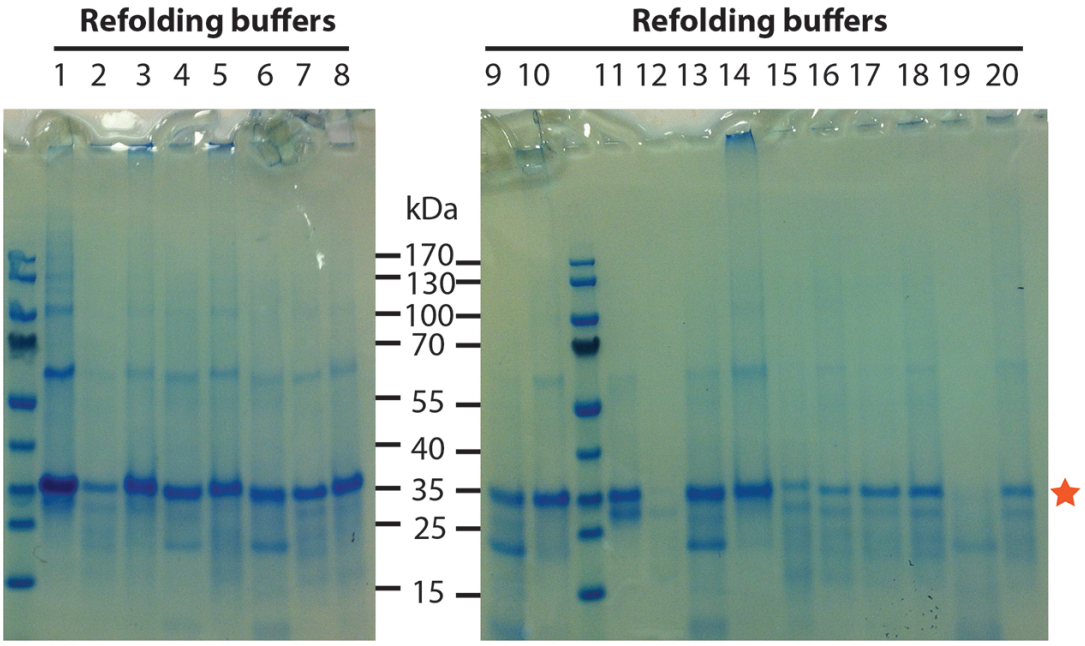


FIGURE 3.5: Refolding trial of the PfRH5 C-terminal domain into 20 different refolding buffers, formulated as in Table 2.2. The loss of the correct protein band at ~28 kDa (orange star), or the presence of aggregated protein stuck in the wells of the Coomassie-stained SDS-PAGE gel, indicated unsuccessful refolding attempts.

binding in surface plasmon resonance assays [238], though of course some interference or competitive effect cannot be ruled out based solely on the kinetic parameters.)

Co-expression is usually accomplished using bicistronic vectors, in which the two proteins are expressed coordinately from an operon, or by dual vector expression, in which the proteins are expressed from two different plasmids, with different antibiotic resistance markers. I adopted the latter approach, cloning basigin into the kanamycin-resistant pRSFDuet-1 vector, and co-transforming with the C-terminal domain of PfRH5 (292-522) in ampicillin-resistant pET15b$_{con}$. The result was soluble expression of basigin, or insoluble expression of PfRH5, but not both; no soluble PfRH5 was produced (Figure 3.6). Further optimisation of this expression, and expression of basigin with full-length PfRH5 and deletion constructs, was rendered unnecessary by the advent of S2 cell expression of PfRH5 (Section 3.4).

Beyond co-expression, I also considered the idea of fusing PfRH5 and basigin together. Fusion would obviate the need for finicky optimisation of dual vector expression; furthermore, easily soluble basigin might act almost as a solubility tag for PfRH5, akin to MBP; but better than MBP, it could specifically bind to PfRH5 to further stabilise it. I set out to fuse PfRH5 (deletion constructs 64, 69, 74, 79) and basigin, in both orientations (N-

and C-terminal). My strategy for the fusion cloning was to use primers with overhangs containing an AscI site (GGCGCGCCG), part of a glycine-alanine-proline sequence, to form the linker between PfRH5 and basigin. However, this cloning proved challenging, and remained incomplete when expression of PfRH5 in S2 cells began.

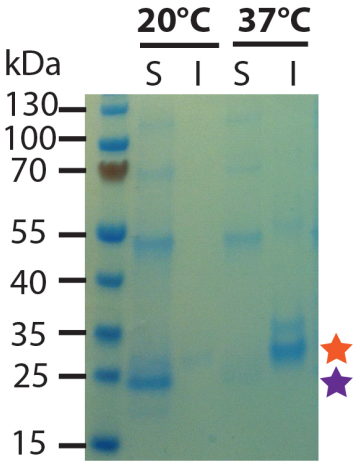


FIGURE 3.6: Co-expression trial of the PfRH5 C-terminal domain with basigin. Dual vector expression was employed, in which two different plasmids encoding each binding partner were co-transformed into *E. coli* for expression trials. Expression was diagnosed by small-scale Ni^2-NTA chromatography, followed by SDS-PAGE and Coomassie staining. The soluble (S) and insoluble (I) expression in Rosetta cells at two different temperatures are shown. The PfRH5 C-terminal domain (by orange star, ~28 kDa) expresses insolubly at 37°C, and basigin (by purple star; ~20 kDa) expresses solubly at 20°C.

3.4 Schneider 2 insect cell expression and purification of PfRH5

3.4.1 Expression

A major turning point in the project was the discovery of PfRH5 expression in Schneider 2 (S2) cells, a *Drosophila* embryonic cell line. The company ExpreS^{2}ions (Hørsholm, Denmark) cloned full-length PfRH5 (residues 26-526) into a proprietary expression vector, pExpreS^2.1. The *PfRH5* gene was resynthesised, codon optimised for *Drosophila* expression, with its four putative glycosylation motifs eliminated by threonine-to-alanine mutations. In the construct encoded by pExpreS^2.1, N-terminal to the protein sequence is a BiP signal sequence, which allows the BiP chaperone to bind and mediate protein secretion, and C-terminal to the protein sequence is a 6-His tag. The vector also contains

a zeocin-resistance marker, which permits selection both in *E. coli* (for cloning) and in S2 cells.

PfRH5-pExpreS^2.1 was transfected into S2 cells at ExpresS^{2}ions. Successfully transfected cells were selected by addition of zeocin to the growth media. After three weeks, a stable, polyclonal cell line was generated. Protein expression is under the control of a constitutive promoter (pActin), and hence the cells needed only to be expanded, allowed to express protein for 4 days, and harvested. PfRH5 could then be purified from the supernatant. I received the PfRH5-pExpreS^2.1 cell line from ExpresS^{2}ions and began to produce large quantities of supernatant for subsequent purification.

3.4.2 Purification

The first step of the purification was Ni^{2+}-NTA chromatography. The elution contained an assortment of S2 contaminant proteins (Figure 3.7A), which could not be entirely removed by gel filtration. Thus, another step was needed to intercalate between Ni^{2+}-NTA chromatography and gel filtration. Several candidate purification procedures were trialled. Ion exchange chromatography and heparin affinity chromatography were unsuccessful purification steps, accomplishing no separation of PfRH5 from contaminants. Since PfRH5 is unglycosylated (due to mutation of its glycosylation sites), lectin affinity resins, including wheat germ agglutinin and concanavalin A, were assayed next. Concanavalin A-conjugated sepharose proved an effective means of isolating PfRH5, which flowed through the column without binding the lectin, leaving the glycan-festooned S2 contaminants behind (Figure 3.7B). A subsequent polishing step of gel filtration yielded highly pure PfRH5 (Figure 3.7C). During expression and purification, some of the PfRH5 was degraded to a 45-kDa fragment—analogous to the fragment isolated from parasite culture [187]—presumably through the action of trace proteases.

3.4.3 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) with SYPRO Orange dye was used to optimise buffer conditions for PfRH5. SYPRO Orange binds to hydrophobic regions of proteins. The protein sample is progressively heated, and as it denatures, its hydrophobic core is exposed to the dye, and the fluorescence increases markedly. The melting temperature

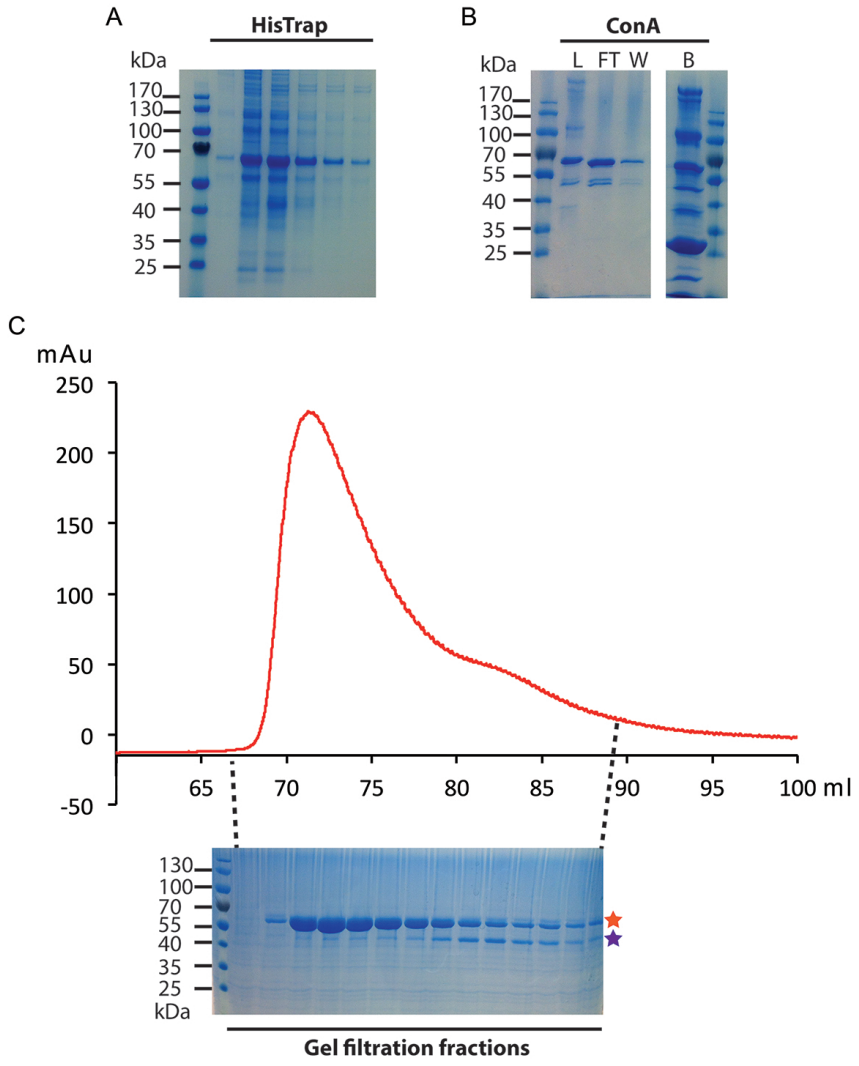


FIGURE 3.7: Purification of full-length PfRH5 from S2 insect cells. The contaminants remaining after elution from a HisTrap Ni^{2+}-NTA column (A) were removed by ConA lectin-binding affinity chromatography (B): unglycosylated PfRH5 and 45-kDa fragment flowed through the column (FT), while glycosylated S2 proteins bound to the ConA beads (boiled beads, lane labeled B). Other abbreviations: L, load; W, wash. Gel filtration followed (C): the trace from a Superdex 200 16/60 column is shown (red). The main peak is full-length PfRH5, but it contains a shoulder of 45-kDa fragment, as witnessed in an SDS-PAGE gel of successive fractions (below; orange star is full-length PfRH5 and purple star is 45-kDa fragment).

can be estimated as the temperature at which half of the protein is unfolded (i.e. half of the maximal fluorescence). DSF analysis suggested that PfRH5 was marginally more stable at pH 7.0 and 7.5 than at pHs 6.0-6.5 and 8.0-8.5 (Figure 3.8A), and destabilised at pH 5.2. As a result, purification and storage of PfRH5 were carried out at pH 7.0-7.5. By contrast, the concentration of salt (between 50-300 mM) had no notable effect upon the stability of the protein (Figure 3.8B). In general, salt concentrations were maintained at 250 mM during purification, and dropped to 150 mM NaCl during gel filtration for

subsequent analysis.

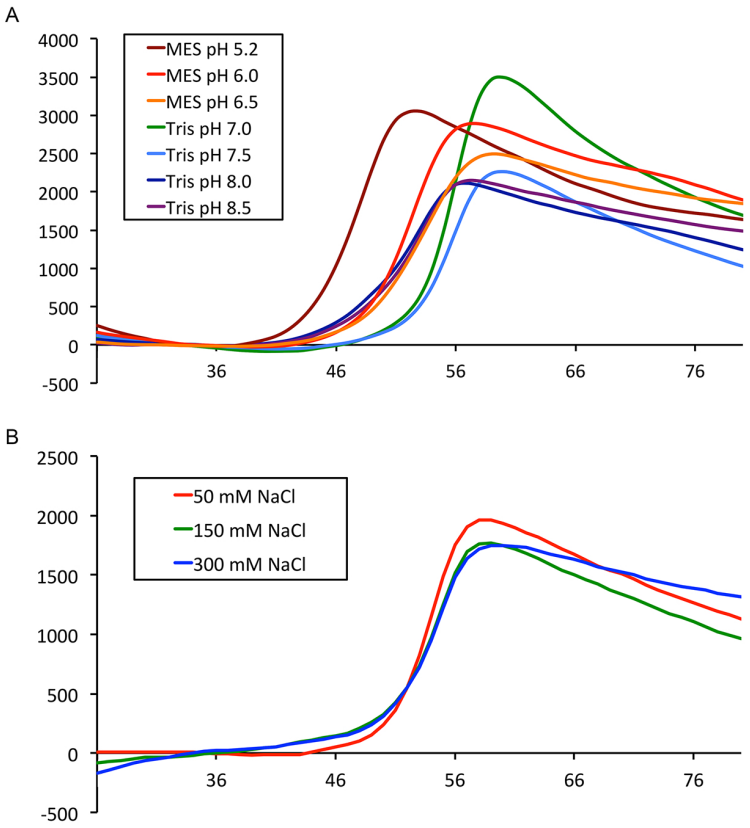


FIGURE 3.8: Microscale differential scanning fluorimetry analysis of PfRH5 melting temperature using SYPRO Orange dye. (A) The stability of PfRH5 at different pHs. All reactions contained 20 mM of the buffer indicated in the figure legend, and 50 mM NaCl. PfRH5 exhibits the highest melting temperature (estimated as the temperature at half-maximal fluorescence) at pH 7.0 and pH 7.5 (results for Tris shown; HEPES similar). It is considerably destabilised at pH 5.2 (MES shown, but acetate similar). (B) The amount of NaCl (50, 150, or 300 mM) does not appear to affect PfRH5 stability. Reactions shown also contained 20 mM Tris pH 8.0, but results in other buffers were similar.

3.4.4 Limited proteolysis and N-terminal sequencing

I next set out to identify potential smaller, stable fragments of PfRH5, with the goal of removing disordered regions and the C-terminal 6-His tag, which might hinder crystallisation. I performed a time-course of proteolytic digestion with enzymes trypsin, chymotrypsin, carboxypeptidase Y, and endoproteinase GluC. Flexible, disordered regions and loops with suitable protease recognition sites should be cleaved more readily, leaving folded domains intact. Samples were taken at different time points and analysed by SDS-PAGE (Figure 3.9). Discrete domains or subdomains were identified following trypsin (∼45-kDa and ∼25-kDa fragments), chymotrypsin (∼40-kDa, ∼30-kDa, and

∼18-kDa fragments), and GluC treatment (∼55-kDa and ∼45-kDa fragments). Many of these fragments were rapidly generated and rapidly degraded further. Notably, the GluC-generated, ∼45-kDa fragment was stable after 60 minutes; GluC cleavage of PfRH5 will be revisited in Chapter 4.

Four of the resultant fragments, together with full-length, uncleaved PfRH5, and the 45-kDa fragment formed during purification (Figure 3.7), were subjected to N-terminal sequencing (Protein & Nucleic Acid Chemistry Facility, Biological Sciences, Cambridge). Full-length PfRH5 was confirmed to begin at residue 26, as expected. The data for the 45-kDa fragment formed during purification were impossible to interpret unambiguously, suggesting a mixture of N-termini, presumably resulting from a plethora of protease-sensitive sites in a flexible, accessible region. Likewise, Edman degradation results were an uninterpretable mixture for the ∼25-kDa trypsin-treated fragment and the ∼45-kDa GluC-treated fragment, again indicative of a mixture of N-termini.

However, the chymotrypsin fragments proved to be formed from a clean cleavage, and could be sequenced. The ∼40-kDa fragment began at residue K140 (and was cleaved after bulky Y139, according to chymotrypsin's predilection). This residue follows both N-terminal potentially disordered regions, and was precisely one of the start sites trialled in *E. coli*-expressed constructs (Section 3.3.1). The ∼30-kDa fragment derived from chymotrypsin cleavage began at residue 253 (following Y252), located in the putative disordered loop in the middle of the protein. Thus, the three regions identified by bioinformatics as disordered are indeed exposed and protease-sensitive.

3.4.5 Expression and purification of shortened constructs

Since disordered regions are a hindrance to ordered packing in crystals, the next step was to remove them from PfRH5, towards the long-term goal of crystallisation and structure determination by crystallography. Guided by the bioinformatic disorder predictions (Section 3.2) and the N-terminal sequencing of chymotrypsin-cleaved fragments (Section 3.4.4), I therefore designed two new constructs for expression in S2 cells. PfRH5ΔN (residues 140-526) lacked the N-terminal disordered regions, and began at residue 140, like the ∼40-kDa chymotrypsin-cleaved fragment. PfRH5ΔNL (140-526, lacking 248-296) was generated by removal of the internal disordered loop (248-296) in addition to

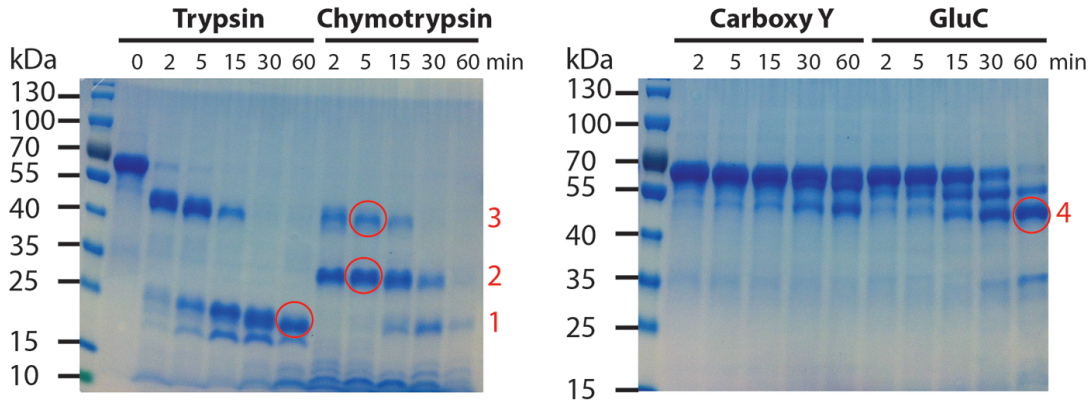


FIGURE 3.9: Limited proteolysis of PfRH5. Full-length PfRH5 (∼60 kDa) was mixed with different proteases, and incubated at 37°C. Samples taken at 2, 5, 15, 30, and 60 minutes were analysed by SDS-PAGE. Four stable subdomains (1-4, indicated by red circles and numbers) were sent for N-terminal sequencing. The larger tryptic fragment was not sent for sequencing. The N-termini of fragments 1 and 4 could not be determined unambiguously, but the N-termini of fragment 2 and fragment 3 derived from chymotrypsin cleavage could. The N-terminal disordered regions had been truncated in fragment 2, and the disordered loop cleaved in fragment 3. Abbreviations: Carboxy Y, carboxypeptidase Y; GluC, endoproteinase GluC.

the N-terminal truncation. The boundaries of the loop truncation were selected based on previous *E. coli*-expressed deletion constructs (Section 3.3.3).

Synthetic genes were prepared, and cloning and transfection carried out by ExpreS^{2}ion Biotechnologies, as described for full-length PfRH5 (Section 3.4.1). I then amplified the cell lines, harvested the supernatant, and purified the proteins as described above. The final gel filtration step of the purification is shown in Figure 3.10. Interestingly, PfRH5ΔN and PfRH5ΔNL both run anomalously the Superdex column, eluting at 95 ml for PfRH5ΔN and 105 ml for PfRH5ΔNL, where one might expect a ∼10 kDa protein to elute, not a ∼40-kDa protein like the PfRH5 fragments. This strongly suggests that these proteins interact transiently with the dextran-conjugated agarose resin. While full-length PfRH5 eluted at ∼72 ml (Figure 3.7), approximately where a ∼60 kDa protein should, purely qualitative observation suggested that it too might interact with and even aggregate on the Superdex resin. A generous proportion of full-length protein was lost at the gel filtration step of purification, and was not contained in the void volume; furthermore, the column required thorough cleaning afterward to release proteinaceous aggregate. Together, these observations imply an interaction between PfRH5 and the resin; however, the addition of a binding partner—be it basigin or an anti-PfRH5 antibody—restored normal behaviour during gel filtration (see Chapter 4).

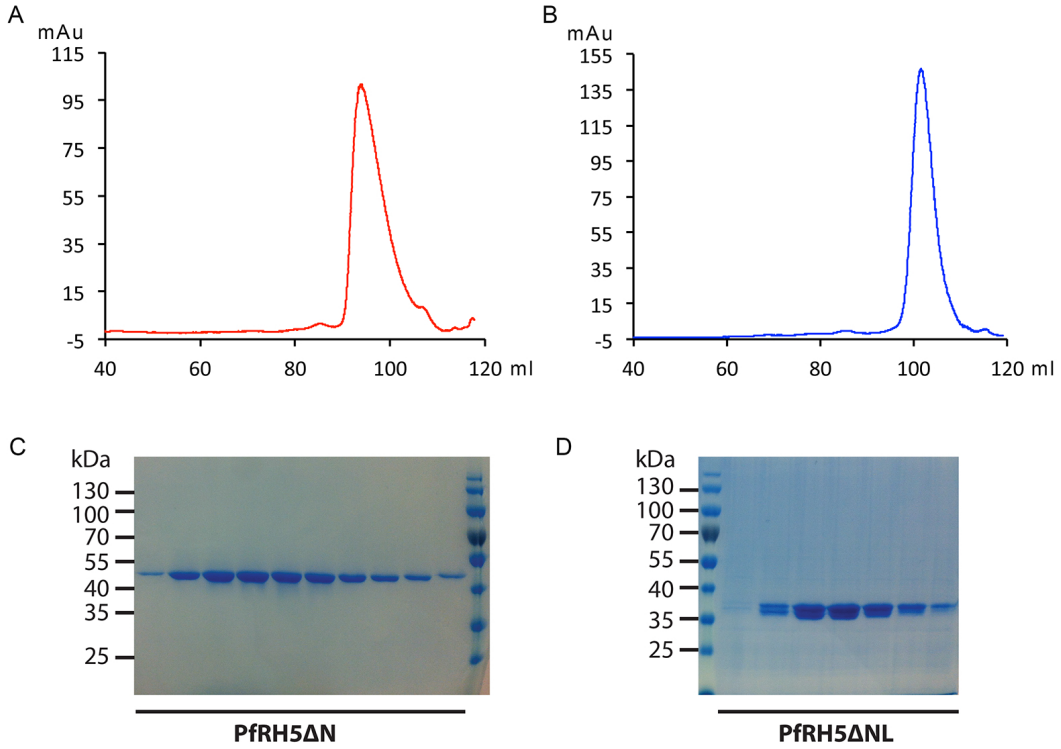


FIGURE 3.10: Gel filtration of PfRH5ΔN and PfRH5ΔNL. The Superdex 200 16/60 gel filtration trace (A, B) and the resultant fractions analysed on an SDS-PAGE gel (C, D), are shown for PfRH5ΔN (A, C) and PfRH5ΔNL (B, D). Note in both cases that the protein elutes much later than anticipated based on the molecular weight.

3.5 Validation of PfRH5 function and immunogenicity

Three pure versions of PfRH5—full-length, PfRH5ΔN and PfRH5ΔNL—were thus generated. I next tested whether they were capable of binding to basigin, capable of raising an antibody response that could block parasite invasion, and capable of binding anti-PfRH5 monoclonal antibodies.

3.5.1 PfRH5 binding to basigin

3.5.1.1 Expression and purification of basigin

In order to test the basigin binding of PfRH5, I expressed basigin with an N-terminal 6-His tag (in vector pET15b$_{con}$) in *E. coli*, using the boundaries of the construct previously crystallised [239]. Soluble expression in Origami B (DE3) cells and subsequent purification—involving Ni^{2+}-NTA chromatography, 6-His tag removal, and gel

filtration—were straightforward. The purification procedure is summarised in Figure 3.11.

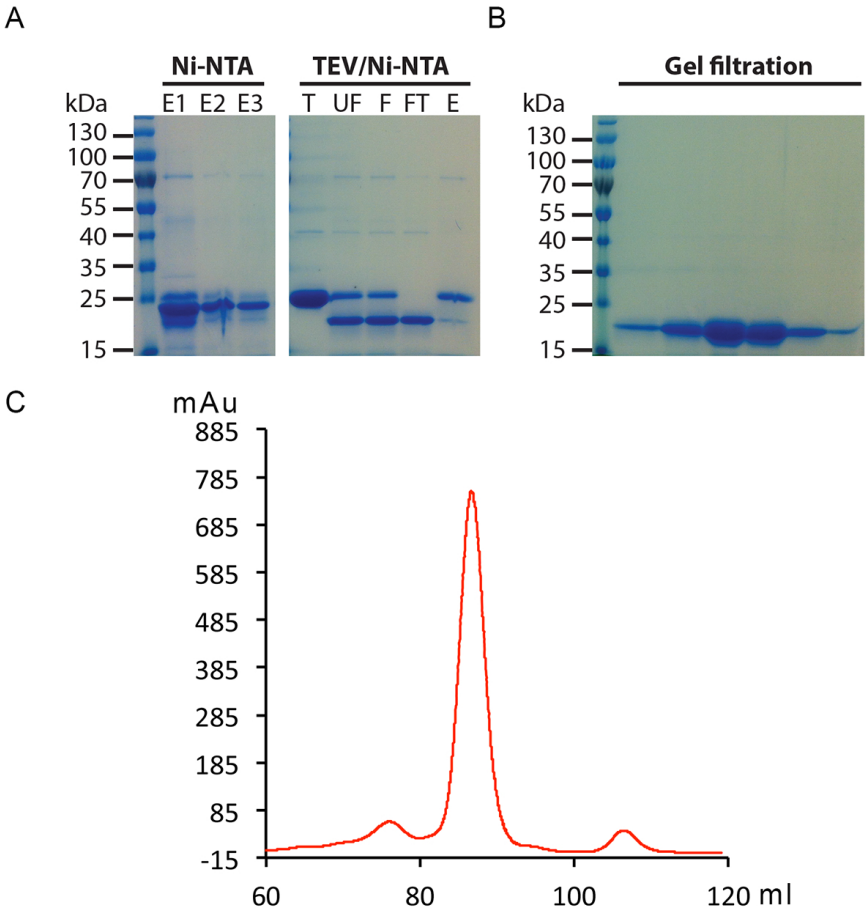


FIGURE 3.11: Purification of basigin from *E. coli*. (A) Basigin was eluted from a Ni^{2+}-NTA column (E1-E3) before the N-terminal 6-His tag was cleaved with TEV protease (T) overnight at 4°C. The next day, the sample (UF, unfiltered) was filtered through a 0.2 μm filter (F), then loaded onto another Ni^{2+}-NTA column. Basigin eluted in the flow-through (FT) and 6-His-tagged TEV protease in the elution (E). (B) and (C) show the results of subsequent gel filtration: the trace (C) and the fractions analysed by SDS-PAGE (B). Basigin is ~20 kDa.

3.5.1.2 Analytical gel filtration

The first confirmation that PfRH5 produced in S2 cells could bind basigin was attained through gel filtration. PfRH5 and basigin were mixed together and loaded onto a Superdex 200 16/60 column. The resultant trace was compared to that of PfRH5 or basigin alone. In the mixed sample, there was a clear peak shift corresponding to a PfRH5-basigin complex (Figure 3.12). Thus, PfRH5 and basigin form a stable complex which remains intact during gel filtration. A comparable peak representing a basigin complex was also observed with each of the shortened variants (data not shown).

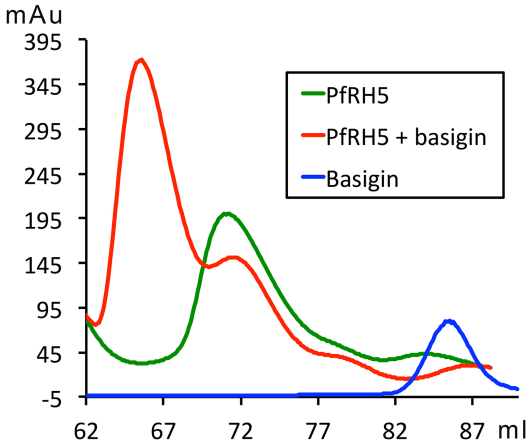


FIGURE 3.12: Analytical gel filtration of PfRH5 bound to basigin. Samples were run on an Superdex 200 16/60 prep grade column. There is a clear peak shift upon addition of basigin to PfRH5.

3.5.1.3 Surface plasmon resonance

Each variant thus appears to bind to basigin; to test whether they bind with similar affinities, I performed surface plasmon resonance (SPR). Basigin was bound to the chip surface by amine coupling and each variant of PfRH5 flowed over (Figure 3.13; Appendix B). Full-length PfRH5 exhibited a 0.35 μM affinity for basigin, in the same approximate range as the published affinity of PfRH5 for basigin of 1 μM [238]. Similar to full-length PfRH5, PfRH5ΔN bound basigin with 0.25 μM affinity, indicating that the N-terminal segment truncated in this construct is not contributing to basigin binding.

PfRH5ΔNL behaved slightly differently, with a slower off-rate than either full-length or PfRH5ΔN, and an incomplete dissociation from the chip surface (Figure 3.13C). This hints at a non-specific interaction of this variant with the chip surface. Nonetheless, its affinity for basigin was in the same range as that of the other two variants (0.1 μM), suggesting that the internal loop, like the N-terminal disordered region, makes no contribution to basigin binding. During preparation of PfRH5ΔNL for crystallisation, the protein was methylated and cleaved with endoproteinase GluC (Chapter 4). Analysed by SPR, the methylated, cleaved form of PfRH5ΔNL exhibits an association-dissociation curve (fast on-rate, fast off-rate) that more closely resembles the full-length and PfRH5ΔN curves than the unmethylated PfRH5ΔNL curve (Figure 3.13D). This suggests that lysines are not important for the PfRH5–basigin interaction. Perhaps methylation of the myriad lysines of PfRH5 prevents their binding non-specifically to the chip surface; though why this non-specific binding effect was seen only with PfRH5ΔNL

and not the other two variants is unclear. (In all three cases, salmon sperm DNA was added to the running buffer; without the negatively-charged DNA, there was incomplete dissociation from the chip surface, suggesting that some non-specific binding of all three variants to the negatively-charged carboxyl groups on the chip surface might have been occurring.) The methylated protein has an affinity for basigin of 1.4 μM, similar to the affinity of full-length PfRH5 for basigin.

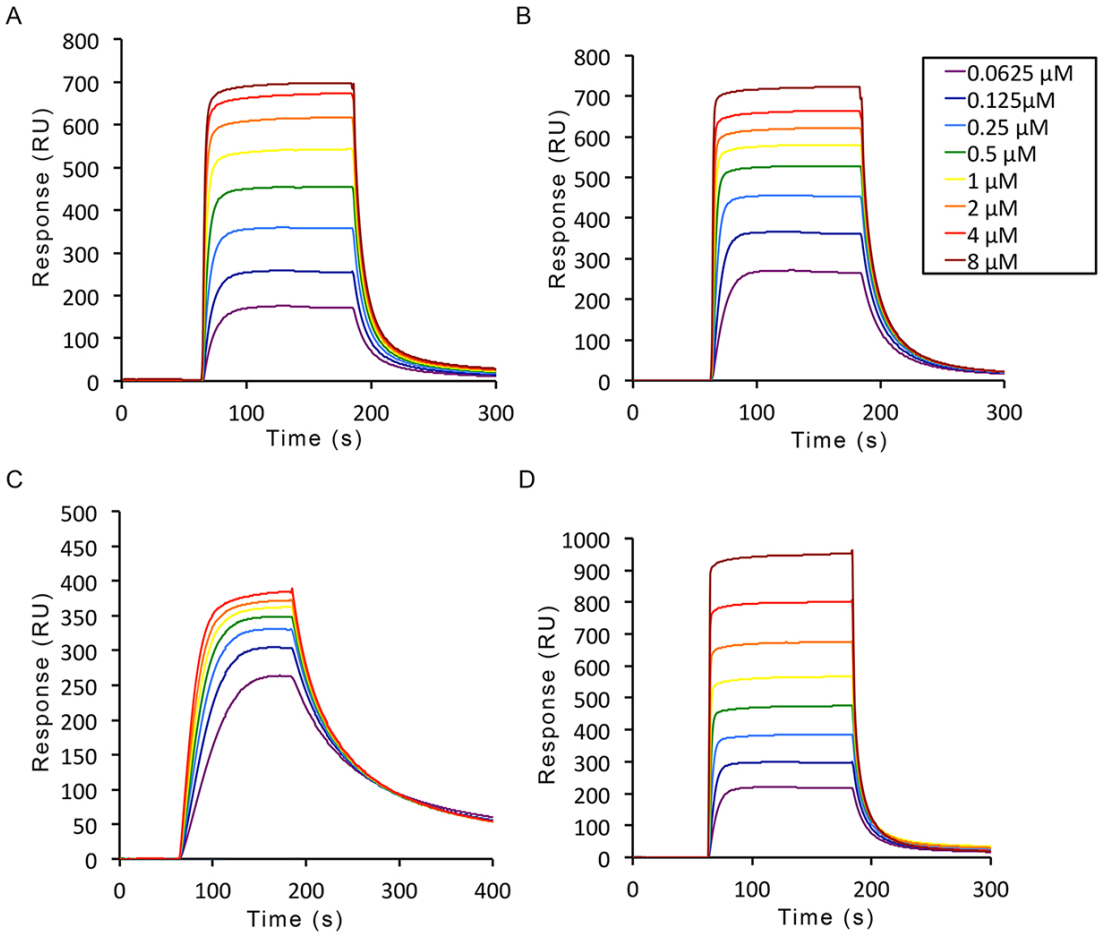


FIGURE 3.13: Surface plasmon resonance to test basigin-binding affinities of full-length PfRH5 (0.35 μM; A), PfRH5ΔN (0.25 μM; B), PfRH5ΔNL (0.1 μM; C), and methylated and GluC-cleaved PfRH5ΔNL (1.4 μM; D). Note that the slower off-rate of unmethylated PfRH5ΔNL relative to the other variants is returned to normal by methylation (D).

3.5.2 Immunogenicity

Next, I wanted to test whether PfRH5 from S2 cells was capable of eliciting a polyclonal IgG immune response which blocked parasite invasion *in vitro*. In experiments performed by Rebecca Brown at the Jenner Institute (Oxford, United Kingdom), rabbit polyclonal

IgG were raised against each of the three protein variants, which were then tested for their ability to neutralise parasites in a growth-inhibitory (GIA) assay. IgG raised against all three variants showed a potent inhibitory effect, both against homologous strain 7G8, and heterologous strain 3D7 (Figure 3.14). The observed growth inhibition was similar to that of IgG raised previously by immunisation of rabbits with viral vectors expressing full-length PfRH5 [127], or full-length PfRH5 recombinant protein [126, 244]. Thus, PfRH5 from S2 cells induces a growth-inhibitory immune response, even in the absence of the two disordered regions. Indeed, the concentration of antibody required to give 50% GIA (EC_{50}) is actually highest for IgG raised against full-length PfRH5: antibodies against both shortened variants, particularly PfRH5ΔNL, appear to be more potent. Why this is the case is unknown; it may be that the extra disordered regions are in fact masking some of the inhibitory epitopes in the full-length protein.

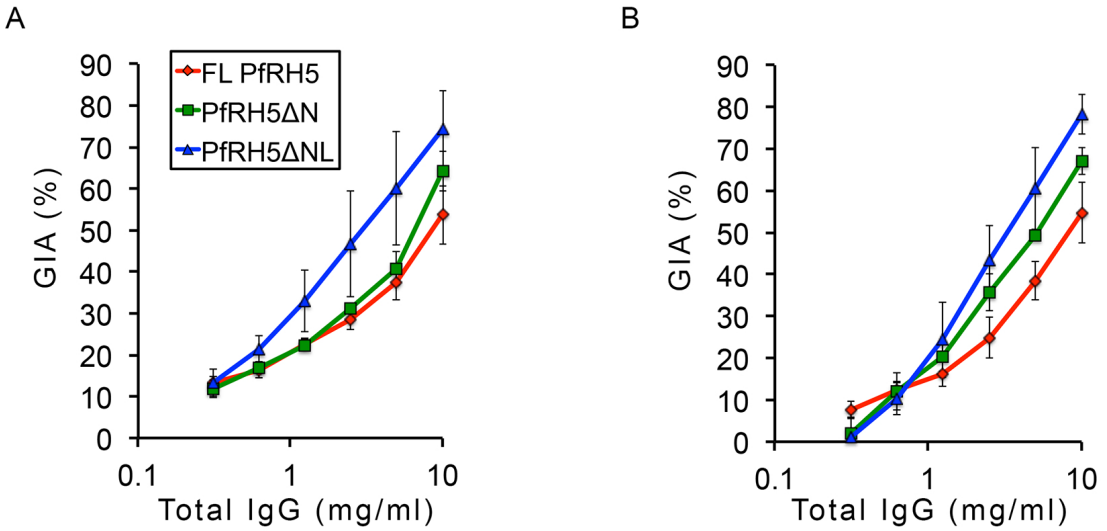


FIGURE 3.14: *In vitro* growth inhibition activity (GIA) of IgG from rabbits immunised with the three PfRH5 variants against 3D7 (A) and 7G8 (B) *P. falciparum* strains. The error bars are standard error of the mean (n = 4). The half-maximal effective concentration (EC_{50}) is given by the concentration of IgG which gives 50% GIA. FL, full-length. Experiments performed by R. Brown.

3.5.3 PfRH5 binding to monoclonal antibodies

Mouse monoclonal antibodies were previously generated against PfRH5, some of which potently blocked parasite growth [237]. I next tested the binding of S2-produced PfRH5, PfRH5ΔN and PfRH5ΔNL to these antibodies using ELISA.

3.5.3.1 ELISA

A previous study speculatively mapped the epitopes of anti-PfRH5 monoclonal antibodies by testing antibody binding to overlapping PfRH5 peptides [237]. Putative epitopes recognised by antibodies 9AD4 (residues 200-213) and QA5 (353-361) were located in the predicted ordered portion of the protein (not the disordered regions). By contrast, the epitope of RB3 was in the N-terminal disordered region (residues 50-78) and the epitope of 4BA7 in the internal disordered loop (residues 242-270). Other antibodies, 2AC7 and QA1, did not bind to any linear peptides.

To test whether these six monoclonal antibodies recognised full-length PfRH5 and the shortened constructs, ELISAs were performed by collaborators at the Jenner Institute (Figure 3.15). Antibodies 2AC7, 9AD4, QA1 and QA5 interacted with all three PfRH5 variants, indicating that their binding sites are located in the ordered regions, not the disordered regions removed from PfRH5ΔN and PfRH5ΔNL. Notably, these antibodies are all potently inhibitory of parasite growth *in vitro*. Thus, the shortened versions of PfRH5 still contain the epitopes targeted by inhibitory monoclonal antibodies.

Antibody 4BA7 binds to full-length and PfRH5ΔN, but not to PfRH5ΔNL, consistent with its epitope being located in the disordered loop truncated in PfRH5ΔNL. Antibody RB3, with its epitope in the disordered N-terminus, should bind only to full-length PfRH5, since both shorter variants lack this region; however, it does not appear to bind full-length PfRH5, perhaps due to degradation of PfRH5 to the 45-kDa fragment, weak binding, or a combination of the two. Neither of these two antibodies are growth-inhibitory. Interestingly, the absolute ELISA response appears to be higher in general for the shorter variants than for full-length PfRH5, a phenomenon similar to that seen in the GIA assay (Section 3.5.2). Thus, the epitopes correlated with growth inhibition are retained in the shorter variants, and the removal of the loops only removes epitopes recognised by non-inhibitory monoclonal antibodies.

3.5.3.2 Production of monoclonal antibody fragments

Anti-PfRH5 monoclonal antibodies for biophysical analysis and crystallisation trials were produced in hybridomas. For all subsequent analysis, they needed to be cleaved into

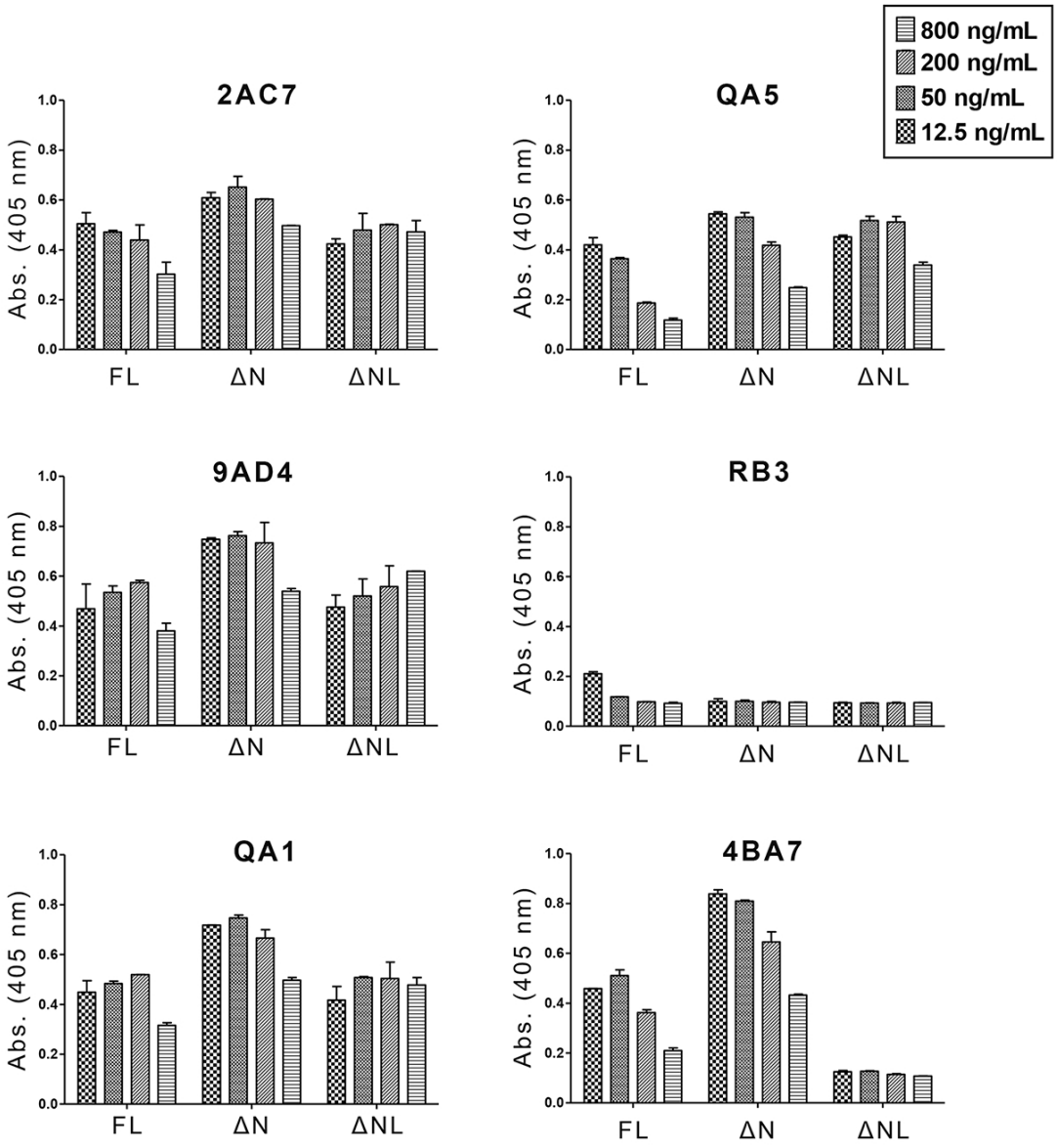


FIGURE 3.15: Interaction of PfRH5 with a panel of mouse monoclonal antibodies was investigated using ELISA. Six mAbs that bind PfRH5 were coated on an ELISA plate and probed using full-length PfRH5 (FL), PfRH5ΔN (ΔN), or PfRH5ΔNL (ΔNL) at concentrations of 12.5, 50, 200, or 800 ng/ml as indicated in the legend. Binding was detected using polyclonal anti-PfRH5 antibody [315]. The error bars are standard error of mean (n=3). These experiments were performed and this figure created by J. Jin.

antibody fragments, containing only a single epitope-binding site. This was accomplished by addition of papain agarose to protein G-purified monoclonal antibodies. The antibody fragments, separated from the fragment crystallisable region by protein A chromatography, were then gel filtered (Figure 3.16) to produce pure protein for downstream analysis (Chapters 4 and 6).

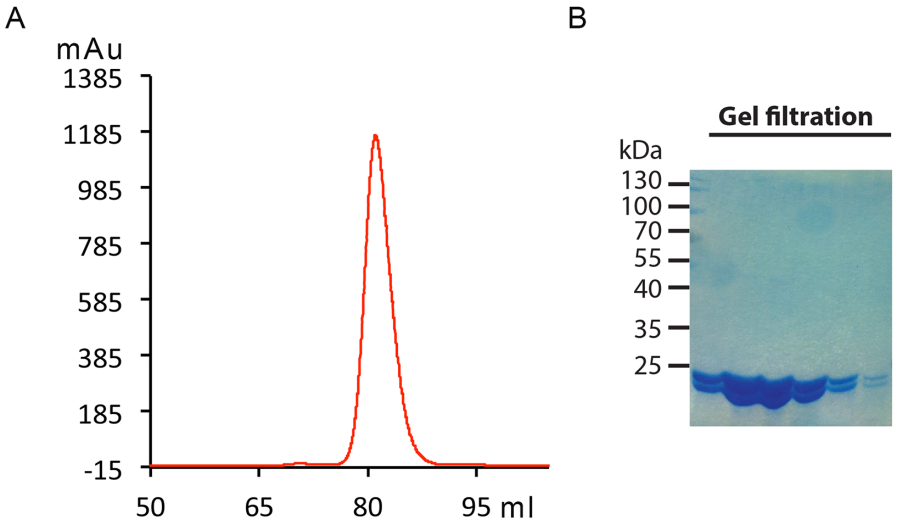


FIGURE 3.16: Purification of QA1 antibody fragment. Monoclonal antibody QA1 was produced from a hybridoma, purified by protein G chromatography, then incubated overnight with papain to produce antibody (Fab) fragments. The fragment crystallisable region (Fc) was removed by protein A chromatography. Finally, the Fab fragment was loaded onto a Superdex 200 16/60 prep grade gel filtration column. The trace (A) and the resultant fractions, visualised with SDS-PAGE (B) are shown. The two bands correspond to the heavy and light chain components (each ~25 kDa) of the Fab fragment (~50 kDa). Purification of the other antibody fragments proceeded similarly.

3.6 Discussion

Bioinformatics analysis of PfRH5 identified regions of predicted disorder, at the N-terminus and in the centre of the protein. Since such protein disorder can interfere both with protein expression and with crystallisation attempts, these regions were cropped to create various smaller constructs of PfRH5, initially in genetically tractable and rapid-growing *E. coli*. Removal of the N-terminal disordered region to form shortened full-length constructs, removal of the internal loop, or both, did not give rise to soluble protein expression; nor did truncating the protein into two putative domains separated by the disordered loop.

The truncation approach was validated, however, once a suitable expression system for PfRH5 was discovered. The S2 cell system allowed the production of full-length PfRH5 as well as two shorter versions, PfRH5ΔN (lacking the N-terminal disordered region) and PfRH5ΔNL (lacking N-terminus and disordered loop). Given that constructs with the same truncations as PfRH5ΔN and PfRH5ΔNL failed to express solubly in *E. coli*, the bacterial expression system may simply not be suitable to express and properly fold this protein: presumably, as the PfRH proteins have no known homologues outside of

the *Plasmodium* genus, no proteins like it are produced by the bacterium. The S2 cell is capable of correctly folding and secreting the protein, though it too does not contain any homologous proteins; this may be due, at least in part, to the fact that S2 cells are eukaryotic and hence more closely resemble *Plasmodium* in their mechanism of protein expression and secretion.

The three protein variants produced in S2 cells each bound to basigin with a similar affinity, as determined by SPR. Though these interactions are of relatively low affinity (low μM range), each variant forms a stable complex with basigin during gel filtration, hinting that co-purification and co-crystallisation trials would be possible, and even advisable, given the peculiar behaviour of these proteins on gel filtration columns (Chapter 4). The fact that the affinity was not altered by removal of either putative disordered region suggests that these regions are dispensable to basigin binding. In keeping with this finding, the ∼45-kDa fragment produced in parasite culture by cleavage of the N-terminal disordered region is still capable of binding to the erythrocyte surface and to PfRIPR [187, 202]. The central disordered region is an insertion in PfRH5 relative to other RH protein sequences, suggesting that it is not essential to the fold; the SPR analysis makes it clear that it is not essential for ligand binding either.

If the disordered regions are not essential to PfRH5 binding to basigin, what is their role? The *Plasmodium* genome encodes a glut of intrinsically disordered protein regions [310], both in secreted invasion proteins but also in cytoplasmic proteins. For the secreted proteins, it may be a mechanism of immune diversion by distraction, with long, extended regions serving as immunogenic decoys. The N-terminal region of PfRH5 appears to be removed during parasite invasion [187, 202]: this cleavage event might be simply a result of the protease sensitivity that leads this region to be cleaved during purification also; or perhaps it is somehow functional. The N-terminus could mask a binding site or interaction surface, not for basigin, since the affinities of full-length PfRH5 and PfRH5ΔN are identical and the binding curves similar, but for another binding partner—perhaps PfRIPR or another protein as yet unidentified. Productive protease digestion to reveal functional subdomains is common in *Plasmodium* invasion proteins, including MSP-1 and perhaps other RH proteins [219, 316]. Further studies will need to determine the contribution, if any, of the disordered regions to PfRH5 function.

Monoclonal antibodies previously raised against full-length PfRH5 using viral-vectored

vaccines [237] proved, in all but one case, to be reactive with full-length PfRH5 generated from S2 cells. Monoclonal antibody RB3, which binds to the N-terminal disordered region, did not react with full-length PfRH5, but this matched previous ELISA experiments where it showed only poor reactivity [237], perhaps partly due to protein degradation to the 45-kDa fragment. RB3 did not bind to either of the variants lacking the N-terminal disordered region, as expected. Moreover, monoclonal antibody 4BA7, with its epitope in the internal disordered loop, did not bind to PfRH5ΔNL. Antibodies QA1, QA5, 2AC7, and 9AD4 bound to each PfRH5 protein variant. Unlike RB3 and 4BA7, these four antibodies are inhibitory of invasion, and hence, the truncated variants contain all of the inhibitory epitopes recognised by these monoclonal antibodies. Furthermore, each of the variants was capable of inducing a polyclonal IgG response that blocked parasite growth. Together, these data signify that the truncated PfRH5ΔN and PfRH5ΔNL proteins are functional in binding basigin, immunogenic, and contain the epitopes recognised by inhibitory monoclonal antibodies; that they are, in short, suitable targets for crystallisation and biophysical studies, and suitable immunogens for vaccine trials.

Chapter 4

The Structure of PfRH5 and the RH Protein Fold

4.1 Summary

The overarching aim of this research was to solve the crystal structure of PfRH5, ideally bound to basigin and to invasion-blocking antibodies. In this chapter, I first describe how the shortened protein construct PfRH5ΔNL, introduced in Chapter 3, was modified for crystallisation by protease trimming, chemical methylation, and mixing with a binding partner, either basigin or an antibody fragment. I then report crystallisation of PfRH5ΔNL bound to antibody fragment 9AD4, and describe how these crystals were used to solve the PfRH5ΔNL structure. (The structure of the antibody fragment complex will be described in Chapter 6.) PfRH5 is revealed to be a novel fold, which is likely to be present in other RH proteins, giving a first insight into the structure of this family of invasion proteins.

4.2 PfRH5 preparation for crystallography

Removal of disordered regions and correct construct boundaries are of fundamental importance to protein crystallography, because protein disorder can interfere with ordered packing into crystals. Therefore, for crystallisation trials, I decided to focus primarily

on the smallest of the PfRH5 constructs, PfRH5ΔNL, lacking both the N-terminal disordered region and the internal loop (Chapter 3). A number of additions to the basic purification procedure were introduced to prepare the protein for crystallography.

4.2.1 Endoproteinase GluC cleavage

My construct boundaries for PfRH5ΔN and PfRH5ΔNL were speculative, based upon limited proteolysis results, sequence alignments, and disorder predictions. They were not necessarily lacking all disordered residues from N- and C-termini—indeed, they contained, at the very least, a 6-His tag at their C-terminus—and did not necessarily represent the optimum crystallisation constructs. I therefore attempted limited proteolysis of these constructs with endoproteinase GluC (Section 3.4.4), to see whether I could trim off any floppy ends or the 6-His tag (Figure 4.1). PfRH5ΔN is cleaved from ∼47 kDa to ∼35 kDa (Figure 4.1A), perhaps the result of cleavage in the internal loop, because this cleavage event is not observed in PfRH5ΔNL. Some trimming of PfRH5ΔNL (∼40 kDa) is, however, observed, forming a slightly smaller fragment (Figure 4.1B). Notably, both proteins lose their 6-His tag during this process, implying that cleavage is occurring (at least) at the C-terminus.

The trimmed fragment of PfRH5ΔNL was particularly promising, losing its 6-His tag and perhaps other disordered parts of N- and C-termini, but otherwise remaining intact. This fragment is stable after overnight cleavage, making large-scale endoproteinase GluC cleavage a possibility, without the need to rapidly quench the reaction. The small-scale proteolysis trial in Figure 4.1 was performed at 37°C, but I discovered upon trying to scale up that incubation at 22°C overnight was preferable, if the protein was to survive the experience without rampant precipitation. I optimised the ratio of GluC to PfRH5ΔNL, ascertaining that a 1:1000 ratio was sufficient for full cleavage overnight (data not shown). During subsequent purifications, before gel filtration of PfRH5ΔNL, I inserted an overnight endoproteinase GluC cleavage step.

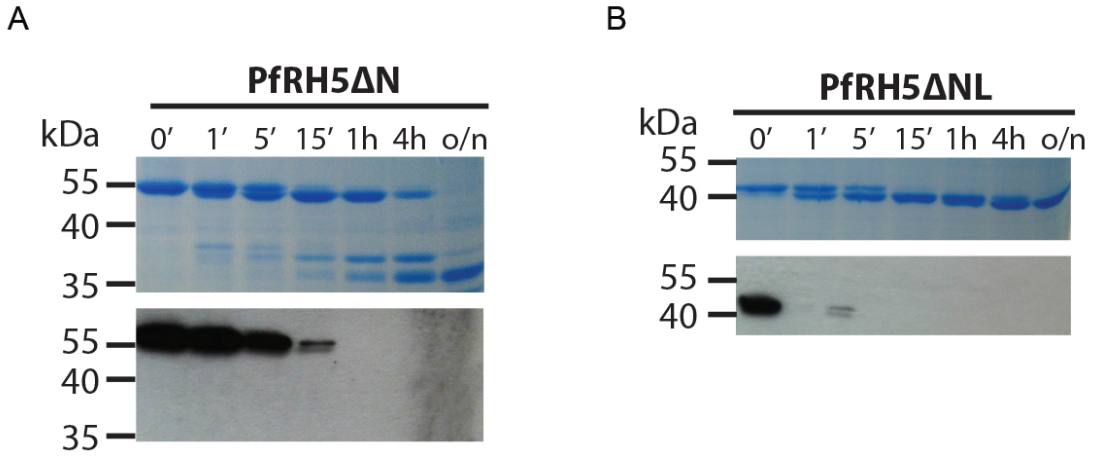


FIGURE 4.1: Endoproteinase GluC cleavage of PfRH5 variants PfRH5ΔN (A) and PfRH5ΔNL (B), over a time course at 37°C (with samples taken at 0, 1, 5, 15 minutes, 1 hour, 4 hours, and overnight). In each case, the Coomassie-stained gel is shown above an anti-6-His Western blot. Over time, PfRH5ΔN is gradually cleaved to a smaller fragment which might correspond to cleavage in the internal loop. PfRH5ΔNL is cleaved overnight to a slightly smaller fragment. Both proteins lose their C-terminal 6-His tag.

4.2.2 Protein complex formation

All three PfRH5 constructs proved much more tractable in the presence of a binding partner, either basigin or monoclonal antibodies. For example, in the absence of a binding partner, all three variants would precipitate at a concentration above ∼2 mg/ml, but once complexed they could be concentrated to above 10 mg/ml. Furthermore, binding partners—notably antibody fragments—can act as 'crystallisation chaperones,' which can suppress molecular motion and provide large surface areas for crystal packing interactions. Therefore, after GluC cleavage, either basigin or one of a number of monoclonal antibody fragments was added to PfRH5ΔNL.

4.2.3 Lysine methylation

Another strategy I exploited for crystallisation was lysine methylation. Flexible, solvent-exposed amino acids, and especially large and charged residues such as lysine, can interfere with the formation of an ordered crystal lattice. Methylation of exposed lysine residues by a simple chemical reaction has proven effective in reducing this so-called surface entropy, and led to crystallisation successes [317]. Given that PfRH5ΔNL contains 48 lysines, this seemed like a good strategy for PfRH5 crystallisation. When

PfRH5ΔNL, in complex with basigin or an antibody fragment, was incubated with borane dimethylamine complex and formaldehyde overnight at 4°C, a gel shift was observed (Figure 4.2). This procedure did not appear to adversely affect the protein, causing no visible precipitation.

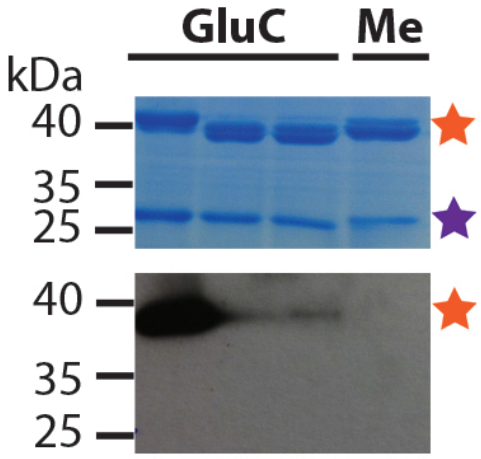


FIGURE 4.2: Lysine methylation of a PfRH5ΔNL–basigin complex. PfRH5ΔNL (orange star) and basigin (purple star) are first GluC cleaved, and then methylated by the addition of borane dimethylamine complex and formaldehyde overnight at 4°C, giving the band shift evident in the rightmost lane (labelled 'Me'). Results for PfRH5ΔNL–antibody fragment complexes were comparable.

In summary, to prepare protein for crystallisation, PfRH5ΔNL was purified as described in Chapter 3 by Ni^{2+}-NTA and concanavalin A chromatography, before its ends were trimmed overnight with endoproteinase GluC. The next day, a binding partner (basigin or antibody fragment) was added, and the lysines chemically methylated overnight. On the third day, the protein complex underwent gel filtration to produce a pure sample (Figure 4.3), ready for crystallisation trials. The crystallisation pipeline for the complex of PfRH5ΔNL with antibody fragment 9AD4, which led to the determination of the structure of PfRH5, is discussed in the following sections.

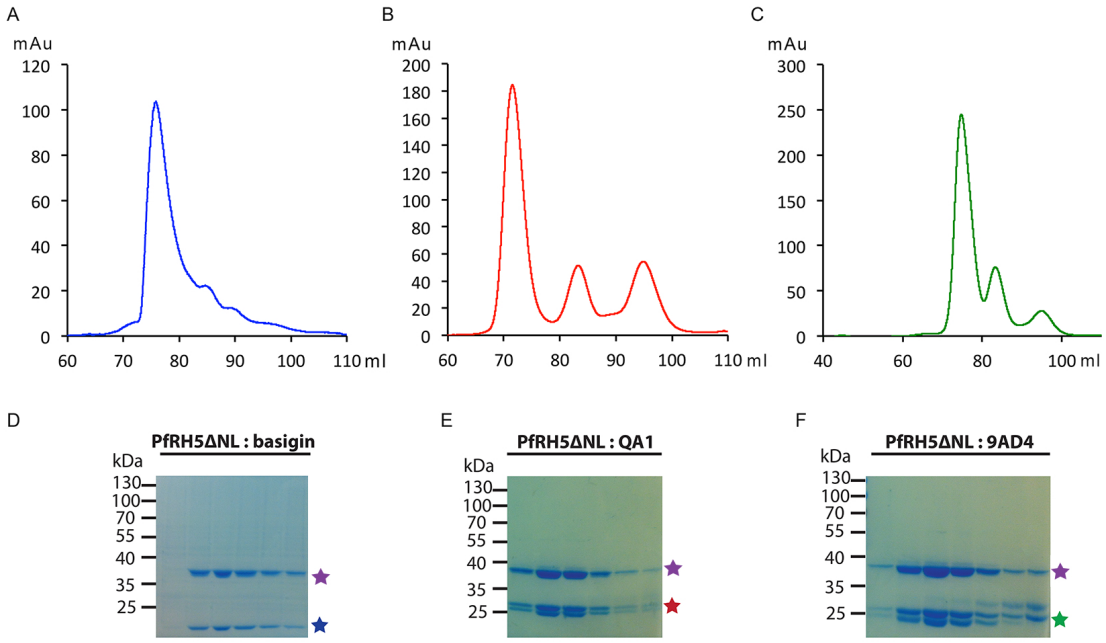


FIGURE 4.3: Purifications of PfRH5ΔNL together with basigin (A, D) and two inhibitory antibody fragments, QA1 (B, E) and 9AD4 (C, F). The gel filtration traces are shown in A-C, with the resultant fractions of the large peak (eluting at ∼75 ml) analysed by SDS-PAGE in D-F. In the SDS-PAGE gels, PfRH5ΔNL is indicated by a purple star, basigin by a blue star, QA1 by a red star, and 9AD4 by a green star. There is residual fragment crystallisable region remaining in the PfRH5ΔNL–9AD4 purification (F), located in between the two antibody fragment chains at ∼25 kDa and PfRH5ΔNL at ∼40 kDa; any contaminated fractions were excluded from the pool and subsequent crystallisation trials.

4.3 PfRH5–9AD4 crystallisation, data collection, and data processing

4.3.1 Crystallisation

Crystallisation was achieved using vapour diffusion in sitting drops. Initial trials of PfRH5ΔNL–9AD4 crystallisation were carried out in a 96-well plate format using commercially available screens, ProPlex, Morpheus, and JCSG+ (Molecular Dimensions). Crystallisation plates were incubated at either 4°C or 18°C. Long, thin crystals appeared after three days at 4°C, in one of the conditions of the JCSG+ crystallisation screen, containing 24% PEG 1500 and 20% (v/v) glycerol (Figure 4.4A). The crystals were in a messy pile, and clustered around the edges of the drop.

In an attempt to grow more crystals, and to optimise them to be larger, with more potential to diffract to high resolution, some of the initial crystals were harvested, mixed

with a Seed Bead, and vortexed to make a stock of seeds to nucleate further crystals in new crystallisation drops. Protein, well solution, and seeds were then mixed with an additive screen, consisting of various combinations of small molecules which can influence protein solubility and crystallisation properties. The optimised crystals adopted several morphologies, including pseudo-parallelograms and longer, rod-like crystals (Figure 4.4B).

The optimised crystals were harvested using a cryoloop, and rapidly plunged into liquid nitrogen. (The well solution was already cryo-protected, so there was no need to exchange the crystal into a different solution before cryocooling.) The crystals were stored in liquid nitrogen until their outing to the synchrotron.

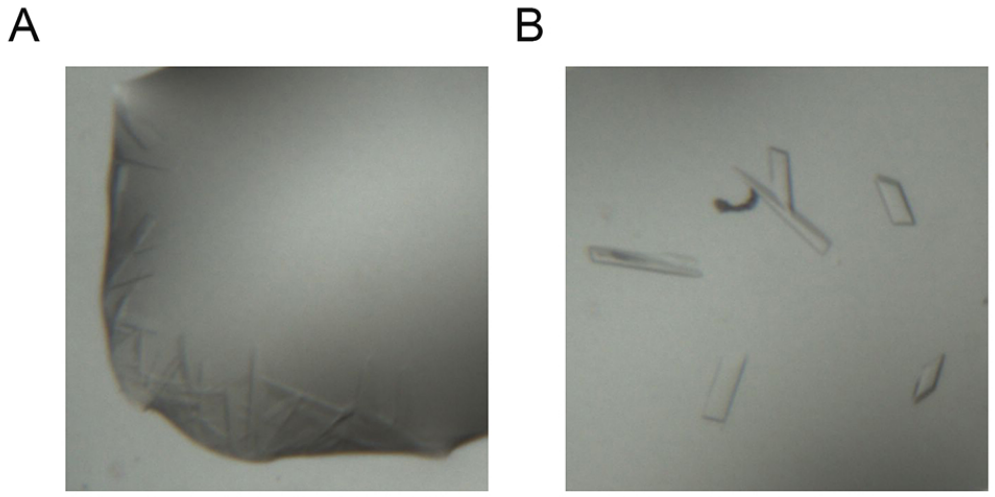


FIGURE 4.4: Crystals of the PfRH5ΔNL–9AD4 complex. (A) Initial crystals were grown in 24% PEG 1500 and 20% (v/v) glycerol in a JCSG+ crystallisation screen at 4°C. The crystallisation drops were set up by mixing 100 nl of 8 mg/ml protein complex with 100 nl of well solution. Crystals appeared after three days, and were ∼60 μm x 130 μm (visible dimensions). (B) Optimised crystals were grown with crystal seeds generated by harvesting the crystals in (A) and vortexing in the presence of a SeedBead (Hampton Research). The well solution was the same as in (A), with the addition of a Silver Bullet additive mixture (Hampton Research; in the well shown, the additive solution contained 0.2% betaine, 0.2% L-glutamic acid, 0.2% L-proline, 0.2% taurine, 0.2% trimethylamine N-oxide, and 0.02 M sodium HEPES pH 6.8). Crystal drops were set up with 100 nl of 8 mg/ml complex, 50 nl of well solution, 50 nl of seeds, and 50 nl of Silver Bullets additives. The crystals began to form after 3 days and were fully formed after 8 days. The parallelogram crystals (seen at right) had dimensions of ∼100 μm x 50 μm (visible dimensions).

4.3.2 Data collection

X-ray diffraction data were collected at the Diamond Light Source on beamline I04. A crystal was mounted and two images collected 90° apart, with a 0.1° oscillation and a 0.1 second exposure. Three parallelogram-like crystals (Figure 4.4B) diffracted to ∼2.3-2.5 Å, and were summarily indexed by program iMOSFLM to space group P222, with unit cell dimensions a = 39.789 Å, b = 86.701 Å, c = 324.834 Å (or similar). For one of these crystals, a full dataset was collected with the strategy suggested by automated data collection strategy program EDNA (Table 4.1). One of the recorded diffraction images is shown in Figure 4.5.

Beamline	I04
Wavelength	0.97949 Å
Aperture	50 μm
Omega Start	164°
Oscillation	0.1°
Exposure	0.1 s
Transmission	30%
No. of images	3600

TABLE 4.1: PfRH5ΔNL–9AD4 data collection strategy.

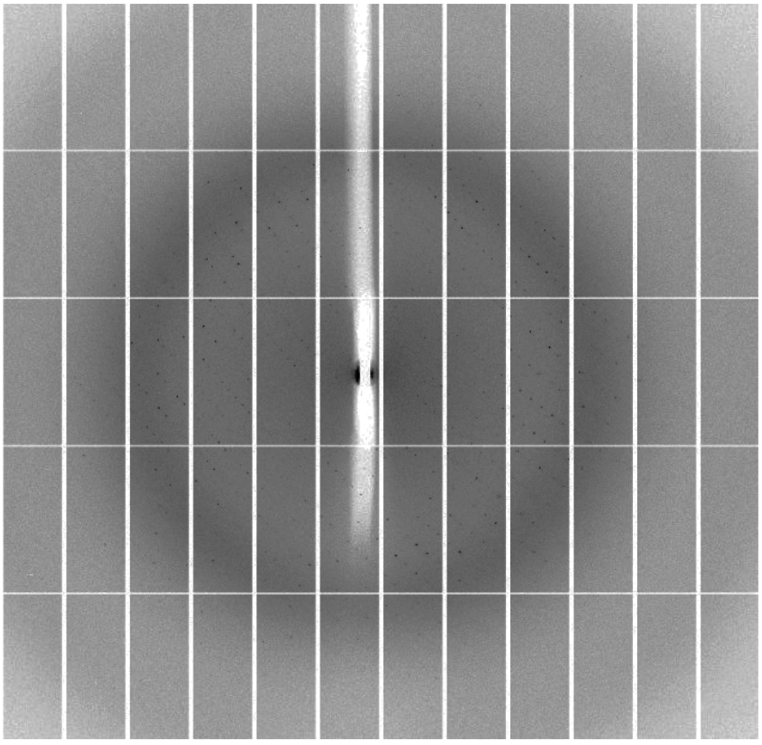


FIGURE 4.5: PfRH5ΔNL–9AD4 diffraction image from beamline I04 at Diamond Light Source.

4.3.3 Data processing

First, the data were indexed and integrated in iMOSFLM. The program identified the space group as P222, with refined unit cell lengths a = 39.79 Å, b = 86.66 Å, c = 324.04 Å, and angles $\alpha = \beta = \gamma = 90°$. The first 800 images were integrated to convert raw pixels into intensities, which were then fed into program Pointless to more accurately determine the space group as $P2_12_12_1$ using systematic absences. Multiple observations of a given reflection were scaled and merged into a single, average intensity using program scala. During this process, 5% of the reflections were reserved for R_{free} calculation. The program Truncate then calculated the average structure factors from the average intensities. The resolution was set to 2.3 Å, to ensure that the signal-to-noise ratio ($I/\sigma I$) equalled or exceeded a value of 2.0 in the highest resolution shell.

The data reduction statistics are summarised in Table 4.2. The signal-to-noise ratio is 9.8 for the full dataset, and 2.0 for the highest resolution shell. The precision-indicating merging value, R_{pim}—an indicator of data quality—is 0.047 for the full dataset (<0.1 is preferable), and 0.375 for the high resolution shell. The dataset is 96.2% complete. Matthews coefficient analysis suggested a solvent content of 62%, and the presence of a single PfRH5ΔNL–9AD4 complex per asymmetric unit.

Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	39.79, 86.66, 324.04
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	41.86–2.30 (2.42-2.30)
$\mathbf{R}_{merge}$	0.098 (0.769)
$\mathbf{R}_{rim}$	0.047 (0.859)
$\mathbf{R}_{pim}$	0.047 (0.375)
$\mathbf{CC}_{1/2}$	0.997 (0.590)
$\mathbf{I/\sigma I}$	9.8 (2.0)
Completeness (%)	96.2 (83.3)
Redundancy	5.0 (4.4)

Table 4.2: PfRH5ΔNL–9AD4 data processing statistics. Values in parentheses indicate the highest resolution shell.

4.4 Solving and validating the PfRH5–9AD4 structure

To solve the structure, molecular replacement with an antibody fragment was attempted. 9AD4 makes up over 50% of the electron density of the PfRH5ΔNL–9AD4 complex: if

the molecular replacement was successful, it was possible that the resultant phases would be good enough to begin to see electron density for parts of PfRH5.

First, a suitable search model had to be generated. A PSI-BLAST search of the structures of the Protein Data Bank suggested that the light chain most resembled PDB entry 3QQ9, and the heavy chain, entry 3HR5. The 3QQ9 light chain and 3HR5 heavy chain were trimmed to mutate or remove any dissimilar residues or insertions compared to 9AD4 (mainly in the hypervariable loops), and a composite antibody fragment created.

To locate the 9AD4 antibody fragment and produce initial phases, the search model was fed into program Phaser, first as a single molecule, and then with constant domains and variable domains separated into two inputs. In the case of the deconstructed antibody fragment, the molecular replacement gave a single solution, in which the two halves of the antibody fragment were put back together perfectly. Rigid body refinement (using program REFMAC5 [318]) gave an R-factor of 49.3%, and restrained refinement gave 49% and 45% for R_{work} and R_{free}, respectively. In general, R-values of less than 50% indicate a reasonable likelihood of a correct solution.

The initial map was solvent flattened using PARROT, and visualised in Coot. Electron density for several helices of PfRH5 was observed (Figure 4.6A). Helices were built into the initial PfRH5 density automatically using Buccaneer, and manually using Coot (Figure 4.6B). Reciprocal-space refinement with BUSTER was used to improve the initial phases and the electron density map. Model building in real space was next, and an iterative cycle of model building and refinement was inaugurated. PfRH5 helices grew and were joined together, and alanines were replaced with the correct amino acid sequence, divined from the presence of disulphide bonds and large side chains, including a single tryptophan. The hypervariable loops of 9AD4 were built alongside their epitope. Water molecules (210 total) were added to the structure, and side-chain rotamers corrected as required.

Following the final refinement, the R-factors were 22.1% (R_{work}) and 25.8% (R_{free}) (Table 4.3). The average B-factor for all atoms (53.0 Å^2), and the root mean square deviations for bond lengths (0.014 Å) and angles (1.7°), were generally indicative of good stereochemical quality. (A high quality structure will have rmsd values of <0.02 Å for bond lengths and <4° for bond angles.) The Ramachandran plot, generated with PROCHECK, contained 93.9% of residues within the most favourable regions, 5.8% in

the additional allowed regions, 0.2% in the generously allowed regions, and 0.2% in the disallowed regions (Figure 4.7). (The residue in the disallowed region is in the antibody structure, distant from the PfRH5 binding site.) The majority of the PfRH5 residues are located in the α-helical region of the plot, in keeping with the observed structure. The final electron density, with the refined structure, is shown in Figure 4.6C.

The final structure encompasses residues 160-504 of PfRH5, with residues 243-299 missing (from the truncated internal loop). PfRH5ΔNL began at residue 140 and ended at 526 (with a C-terminal His tag), but a combination of GluC cleavage and flexibility accounts for the trimming at each terminus. Similarly, a few more residues were absent at each end of the truncated internal loop than were originally deleted in this construct (248-296). No other residues are missing. The structure of the 9AD4 complex will be revisited in Chapter 6. For the remainder of this chapter, I will focus on the structure of the PfRH5 component of the complex, as a first insight into the fold of the RH protein family.

Resolution (Å)	41.86-2.30
No. reflections ($\mathbf{R}_{free}$)	49336 (2462)
$\mathbf{R}_{work}$/ $\mathbf{R}_{free}$ (%)	22.10 / 25.83
No. protein atoms	5795
No. water molecules	210
Average B-factors (Å^2)	53.0
R.m.s deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.7

TABLE 4.3: PfRH5ΔNL–9AD4 refinement statistics.

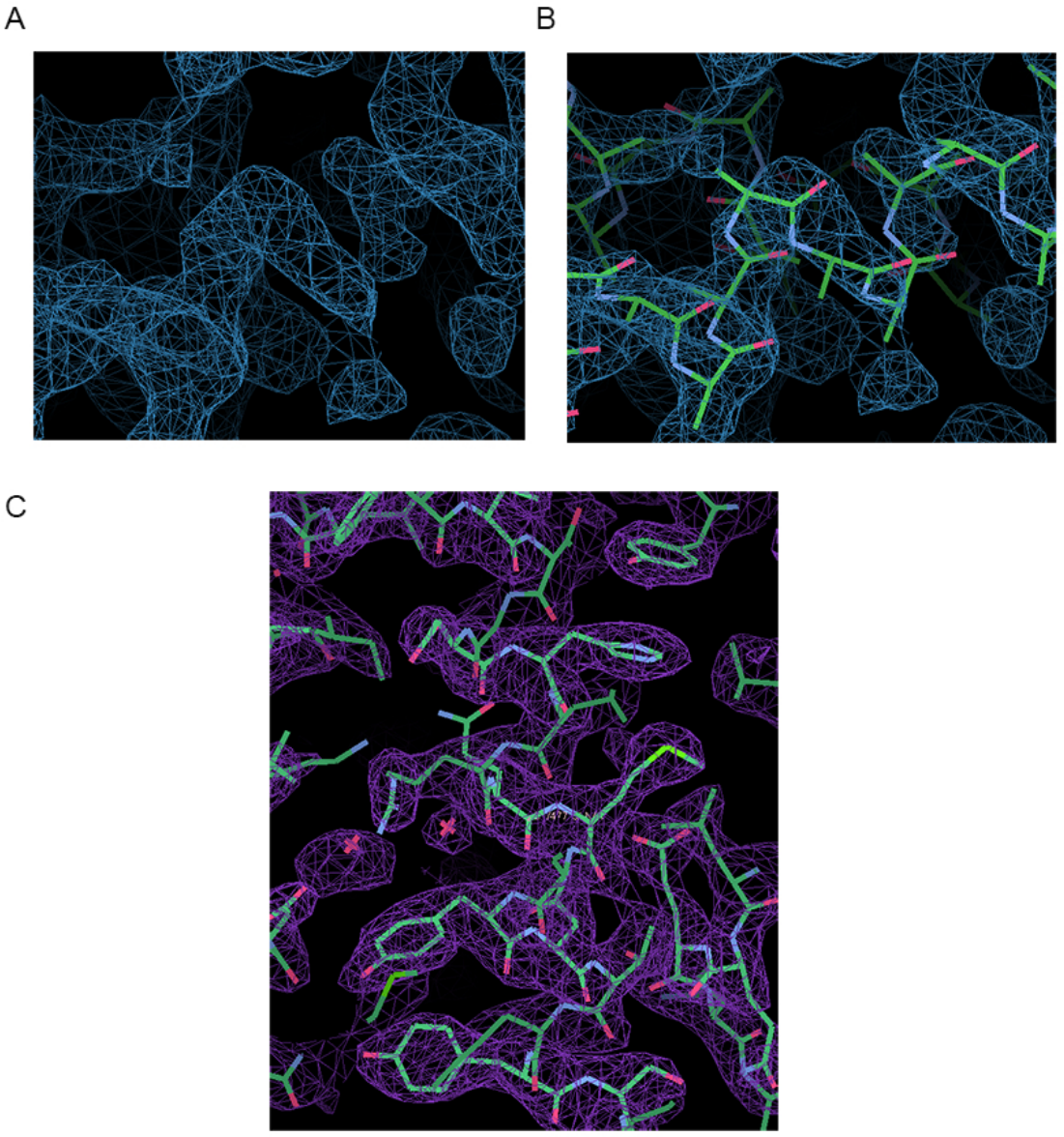
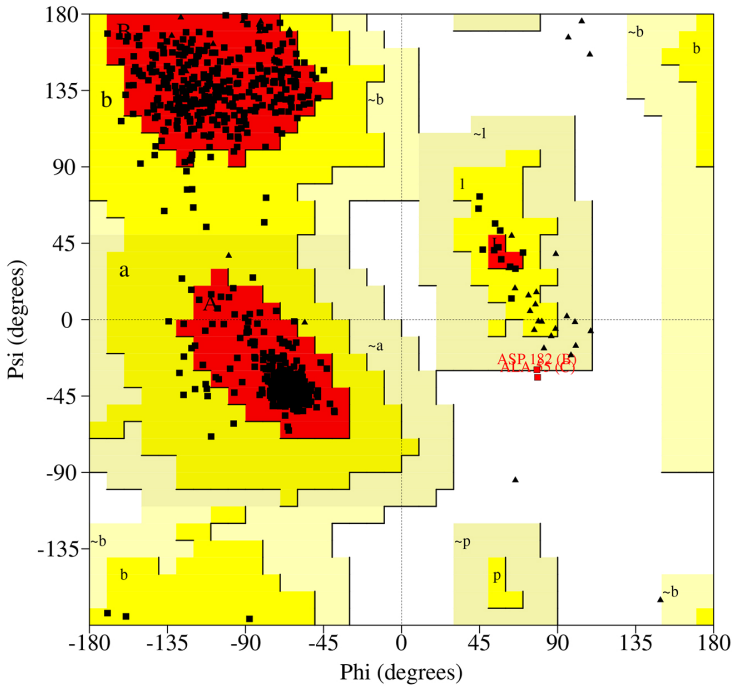


FIGURE 4.6: PfRH5ΔNL–9AD4 electron density maps (all $2F_o$ - F_c, contour 1σ). (A) Initial map derived from molecular replacement of antibody fragment and density modification with PARROT. (B) The initial map with a polyalanine helix built into the density for PfRH5ΔNL. (C) The final electron density superimposed with the final PfRH5ΔNL–9AD4 structure. A helix of PfRH5 is shown. Atoms are coloured green (carbon), blue (nitrogen), or pink (oxygen). Pink crosses represent water molecules. The figure was prepared using Coot [303].



	No. Residues	Percentage
Most favoured regions	613	93.9
Additional allowed regions	38	5.8
Generously allowed regions	1	0.2
Disallowed regions	1	0.2
Residues not Gly or Pro	653	
End-residues (excl. Gly or Pro)	8	
Gly residues	39	
Pro residues	26	
Total	726	

FIGURE 4.7: Ramachandran plot and statistics of PfRH5ΔNL–9AD4 structure, from PROCHECK [307].

4.5 The structure of PfRH5

4.5.1 Fold architecture

The unveiling of PfRH5 reveals a rigid, flat, 'kite-shaped' molecule (Figure 4.8). It manifests a pseudo-two-fold rotation symmetry, with each half predominantly built from a three-helical bundle. The outermost helices are markedly kinked or bent to give the characteristic kite shape. The N-terminal half begins with a short, two-stranded β-sheet that crosses the long axis of the kite at the centre of the molecule. This is followed by a single, short helix and then by two long, kinked helices. Between the second and third helices is the flexible loop (truncated in this construct), which contains 58 residues in full-length PfRH5. The C-terminal half is simpler, consisting of three long helices, one kinked, that span the entire length of the domain and finishing with a flexible C-terminus. One disulphide bond stabilises the loop that links the two halves of the structure, while another links the second and third helices.

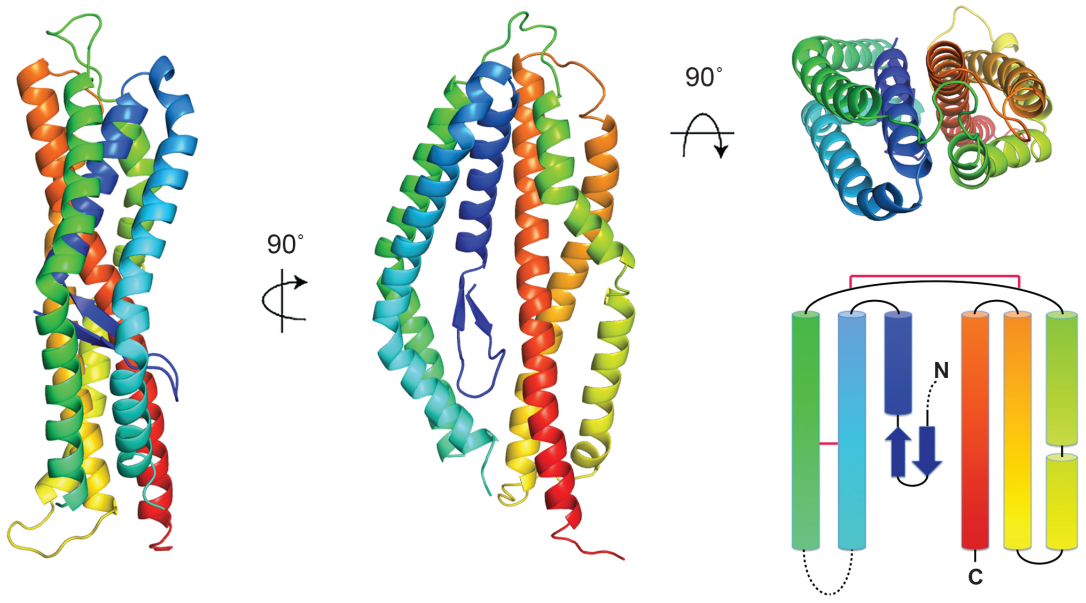


FIGURE 4.8: Three views of the PfRH5ΔNL structure, and a schematic topology diagram, coloured in rainbow from blue (N-terminus) to red (C-terminus). Disulphide bonds are indicated by red lines, and the excised loop by a dotted line on the topology diagram.

A representation of the electrostatic potential of the protein surface reveals a relatively even distribution of negative and positive charges, apart from a single, positively-charged

patch on one surface of the protein (Figure 4.9). Whether this charged patch has a functional role remains to be seen.

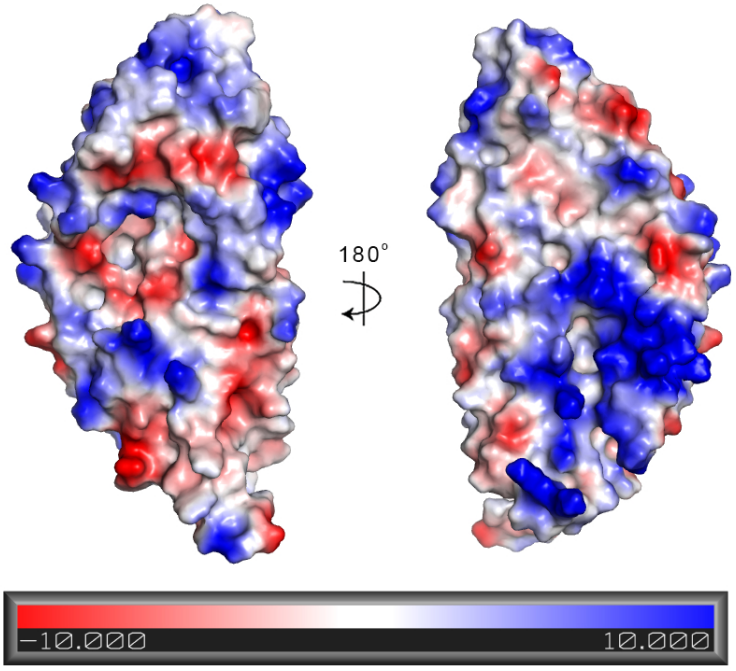


FIGURE 4.9: The electrostatic surface potential of PfRH5ΔNL, generated in Pymol. Scale ranges between -10 k_bT/e_c (red) and +10 k_bT/e_c (blue), where k_b is the Boltzmann constant, T is the temperature (Kelvin), and e_c is the charge of an electron.

The ∼94903 protein structures solved to date are classified into only 1393 families of unique folds (as of August 2014; based on SCOP analysis of the Protein Data Bank). To determine whether the PfRH5 fold has been previously observed, the Protein Data Bank was scoured for any structural homologues, using the DALI server [319]. The top-scoring hits (e.g. SipB from *Salmonella enterica*, PDB code 3TUL) were all three-helical bundles, aligning only ∼120 residues out of the 300 residues of PfRH5ΔNL, and had alignment Z-scores hovering around 7 (a strong match for a 300-residue structure would have a Z-score of $\frac{300}{10} - 4 = 26$). No structure showed the characteristic kite shape. Hence, I conclude that PfRH5 adopts a novel fold. This is consistent with the fact that PSI-BLAST searches (undertaken before the structure was determined) unearthed no homologues apart from *Plasmodium* surface proteins implicated in erythrocyte invasion (Section 3.2).

It is interesting to assess how well the PfRH5ΔNL structure conforms to my early bioinformatic predictions (Section 3.2). Overall, the secondary structure predictions are remarkably similar to the observed secondary structure, correctly predicting one of the

β-strands, many of the helices, and even some of the helix breaks (Figure 4.10). The disorder predictions are also pleasingly accurate, with small increases in the disorder propensity of the PfRH5 loops.

The structure of the PfRH5 core appears to be relatively rigid. In succeeding chapters, two more crystal structures of PfRH5 complexes will be introduced, with four more molecules of PfRH5 in the two new asymmetric units. The five copies of PfRH5 found in the three different crystals aligned with an r.m.s.d. of 0.9 Å over 95% of the residues (Figure 4.11). The only differences are in the organisation of the C-terminus (residue 496-end) and the loop linking helices 4 and 5 (residues 396-406), which adopt a different position in each of the five copies due to crystal packing (Chapter 5-6).

4.5.2 SAXS analysis

To investigate how the structure of PfRH5ΔNL compares with full-length PfRH5—that is, to determine the location of the truncated regions—I performed small-angle X-ray scattering analysis (SAXS). This technique uses X-ray scattering to extract information about molecular size and shape in solution. The scattering curve can be used for *ab initio* modelling of the protein shape at low resolution. The crystal structure of PfRH5ΔNL was docked into the SAXS envelope of PfRH5 generated in this manner (Figure 4.12). The envelope exhibits the same flat structure, but appears elongated relative to PfRH5ΔNL. The extra density in the SAXS envelope is likely due to the C-terminal residues removed from PfRH5ΔNL by endoproteinase GluC; the flexible loop; and perhaps also part of the ∼160-residue, extended N-terminus (even though the N-terminus eventually joins at the centre of the kite).

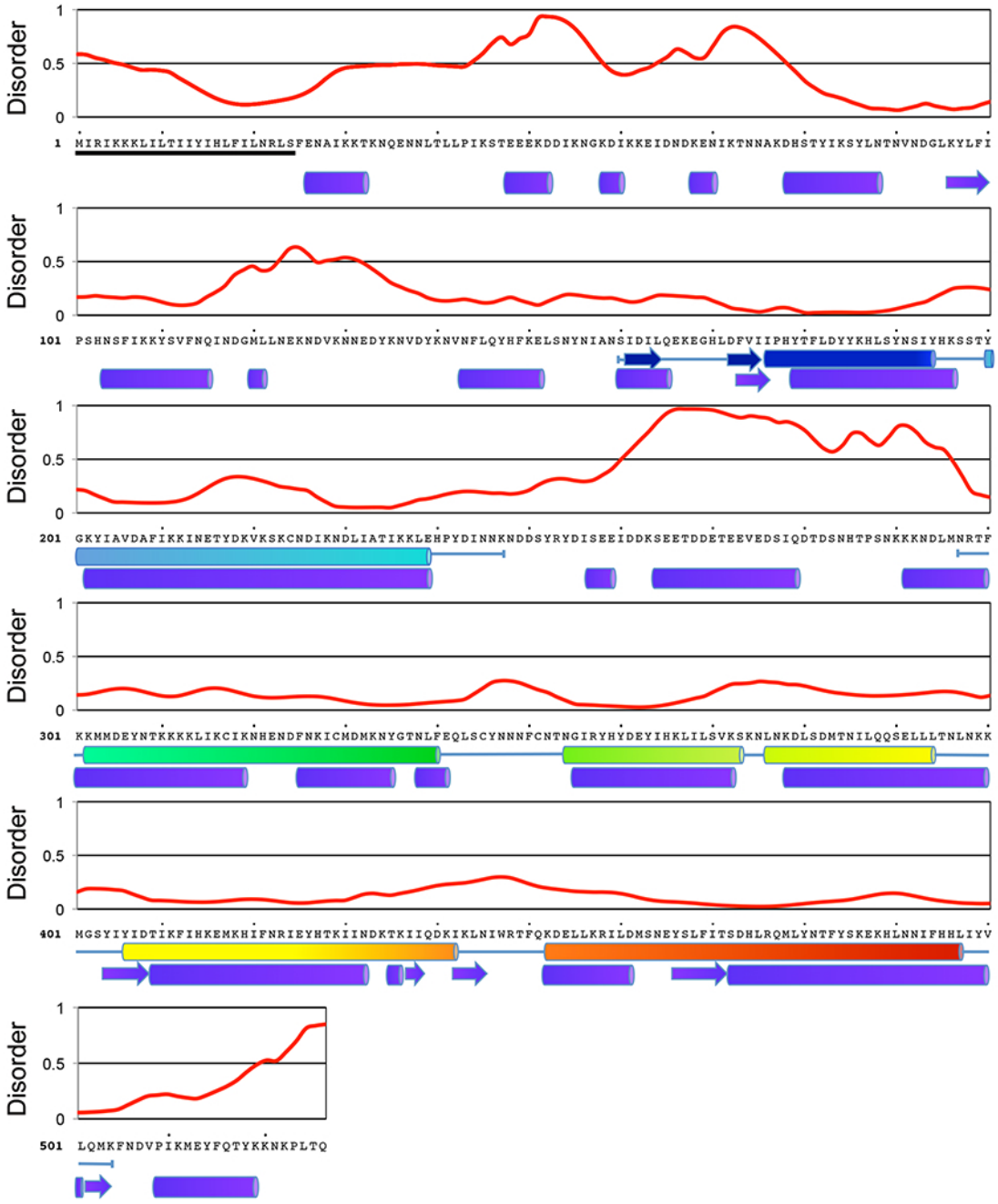


FIGURE 4.10: Disorder predictions and secondary structure of PfRH5. Predictions of long-range disorder (Section 3.2) are shown above the sequence of PfRH5, with values >0.5 indicative of disorder. The residues visible in the PfRH5ΔNL crystal structure are shown below the PfRH5 sequence as secondary structure elements (sheets as blue arrows, and helices as tubes in rainbow coloring) linked by blue lines. The missing loop (248-296) is shown as a break in the blue line. Underneath the observed secondary structure is the predicted secondary structure for comparison (purple; Section 3.2). The secretion signal sequence is indicated (black underline).

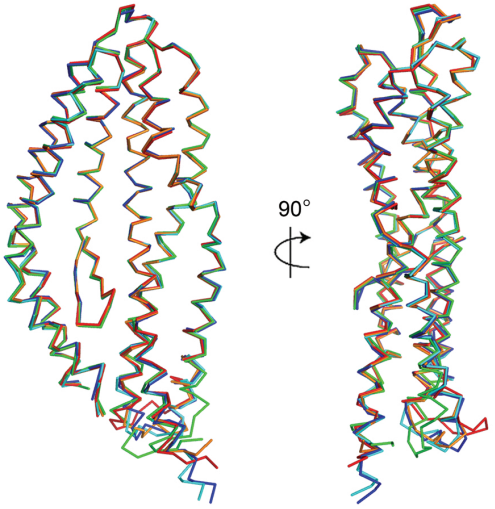
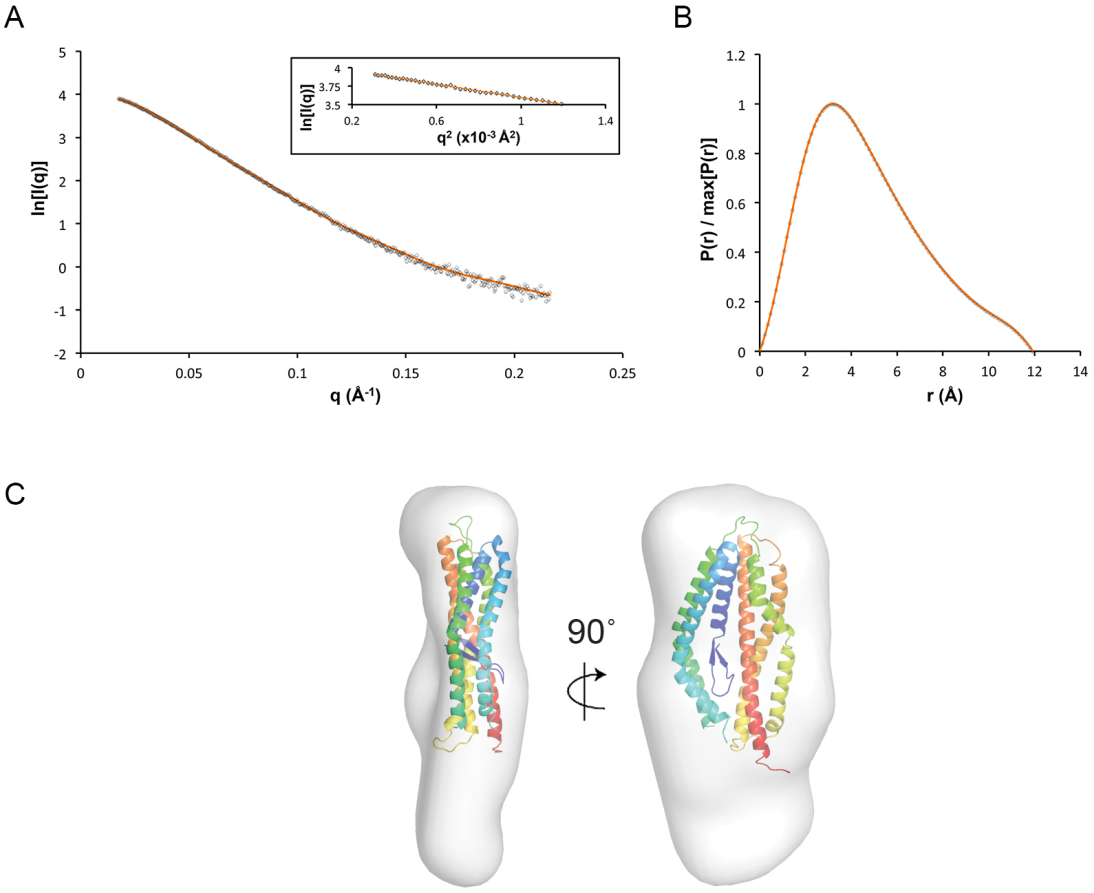


FIGURE 4.11: Structural alignment of PfRH5 from different crystal structures. PfRH5ΔNL from the 9AD4 complex (green) structure was aligned with four more PfRH5 molecules from two additional crystal structures. [Red and orange molecules are from the PfRH5–basigin structure (see Chapter 5); and blue and cyan molecules are from the PfRH5ΔNL–QA1 structure (see Chapter 6).] The alignment was performed in Coot [303], giving an r.m.s.d. of 1.7 Å. The C-terminus and the loop between helices 4 and 5 were the only regions showing significant differences. For the remaining 95% of PfRH5, the RMSD is 0.9 Å.



D

	Rg (nm)	D_{max} (nm)	Porod volume (nm^3)	MW_{app} (kDa)	NSD	χ
PfRH5	3.7	11.9	104.5	61.4	0.692±0.036	1.01

FIGURE 4.12: SAXS analysis of PfRH5. (A) The theoretical scattering calculated from the average of 20 *ab initio* reconstructions of full-length PfRH5 (orange line) plotted with the experimental scattering intensity curves (diamonds). The data are presented as the natural logarithm of the intensity. The Guinier plot is displayed in the inset. (B) The distance distribution function, $P(r)$, of PfRH5. (C) PfRH5ΔNL structure (rainbow) docked into a SAXS envelope of full-length PfRH5 (grey), revealing a bulge of extra SAXS density at the bottom of the structure, near the C-terminus and the excised loop. (D) Summary of SAXS parameters. The radius of gyration (R_g) was determined from the Guinier plot using AutoRg, and the maximum particle dimension (D_{max}) and the Porod volume were calculated using GNOM. An estimate of the molecular weight was obtained by dividing the Porod volume by 1.7. *Ab initio* modeling was used to generate 20 shape reconstructions from the data. The normalised spatial discrepancy parameter (NSD) diagnoses the similarity of these models. The models were averaged and the fit of the average model to the experimental data is indicated by the χ value. Situs and Sculptor were used to manually fit the PfRH5ΔNL crystal structure into the PfRH5 SAXS model.

4.6 The PfRH5 fold in other RH proteins

4.6.1 Alignments and threading

As members of the Plasmodium RH family show little sequence identity [187, 236], a combination of sequence alignments (Figure 4.13) and structure-based threading (Table 4.4) was used to determine whether other members may also contain the PfRH5 fold. For each of *P. falciparum* RH1, RH4, RH2b, RH2a and RH3 (a pseudogene); *P. vivax* RBP-1 and RBP-2; *P. reichenowi* RH5; and *P. yoelii* Py01365, PfRH5-like domains were identified with high confidence at the N-terminus of the protein. The sequence identity relative to PfRH5 ranges from 10-14% (PvRBP1) to 70-73% (PrRH5), but in all cases threading gave exceptionally high confidence scores for the adoption of the PfRH5 fold (98%-100%). Threaded models of other members of the PvRBP-2 clade are very similar to the PfRH5 structure, while those of the PvRBP-1 clade are more divergent, with additional loops and helix breaks (Figure 4.14).

	Clustal Omega	**Phyre2 threading**		
	Identity (%)	Identity (%)	Confidence (%)	Aligned residues
PfRH5	100	100	-	-
PrRH5	70	73	100	287
PfRH2a	22	27	100	283
PfRH2b	22	27	100	283
PfRH3	20	23	100	267
PfRH4	17	13	99	267
PfRH1	15	16	98	269
PvRBP1	14	10	100	280
PvRBP2	17	16	100	284
Py01365	18	16	100	283

TABLE 4.4: Summary of RH protein sequence alignments and threading. The percentage identity between each query RH protein sequence and PfRH5, derived from both the Clustal Omega sequence alignment and the Phyre2 threading, is shown in the first two columns. Also shown is the confidence in the threading prediction; a value of >90% indicates strong confidence that the query sequence adopts the given fold [269]. The final column indicates the number of residues of the query sequence threaded onto the PfRH5 fold.

Consistent with the low sequence identity, there are no residues totally conserved across this protein family. Not even the disulphide-forming cysteine residues are completely conserved, suggesting that the disulphides are not necessary to adopt or stabilise the fold. However, one of the disulphides, that connecting helices two and three, is conserved in PfRH2a/b and PvRBP2, other members of the PvRBP2-like clade, a reminder of their

closer evolutionary relationship with PfRH5 (Section 1.3). Notably, residues showing similarity (that is, with similar physicochemical properties) are located primarily in the interior of the domain (Figure 4.15). They are likely to act to stabilise the fold, while exposing a variable surface to allow diversification of function.

PfRH5 is the only member of the RH family comprised solely of this kite-like fold: the other members are significantly larger (Section 1.3). Does the N-terminal PfRH5 domain identified in these other PfRH proteins also function as the ligand-binding domain? For PfRH4, the only other RH protein whose erythrocyte receptor has been identified, the putative PfRH5 fold overlaps with a large region known to contain the complement receptor 1 (CR1) binding site [230]. Therefore, the PfRH5-like domain is an excellent candidate for the ligand-binding module in other RH family members.

4.6.2 Expression of other RH proteins in S2 cells

A lack of any recognisable domains in the larger RH proteins previously forestalled the rational design of protein constructs. Now, however, discovery of the PfRH5 fold, and the reasonable supposition that it may be the ligand-binding domain, makes design of smaller, more manageable, and perhaps functionally important sections of these proteins possible. Furthermore, the success of PfRH5 expression in S2 cells suggests that these cells are capable of folding this domain. I therefore decided to attempt expression of PfRH1, PfRH2a, PfRH4, PvRBP-1, and PvRBP-2 in S2 cells. *Drosophila*-optimised versions of the *P. vivax* genes and PfRH2b were synthesised; mammalian-optimised genes of PfRH1 and PfRH4 were provided by G. Wright and S. Draper, respectively. In a preliminary set of experiments initiated during the process of thesis writing, the genes were cloned into pExpreS^2.1$_{con}$ and the plasmid transfected into S2 cells.

Two separate attempts to transfect PfRH1, PvRBP-1 and PvRBP-2 were unsuccessful: all of the cells died over the course of three weeks. Since expression is constitutive, it may be that the protein product is somehow toxic to the cells; perhaps an inducible S2 expression system could be assayed next.

PfRH2b and PfRH4 transfections were, by contrast, successful, insofar as a subsection of cells did survive. An expression trial revealed that PfRH2b was expressed in a soluble form, while PfRH4 was not (Figure 4.16). Perhaps the domain boundaries

of PfRH4, a more distant relative of PfRH5 than PfRH2b, are not quite correct, and further constructs will need to be designed in future; or perhaps the gene will need to be re-optimised for *Drosophila* expression. The erythrocyte receptor of PfRH2b has not yet been identified, so once the S2-produced protein has been purified, the next experiment will be to identify its erythrocyte receptor by screening commercially available erythrocyte surface protein libraries.

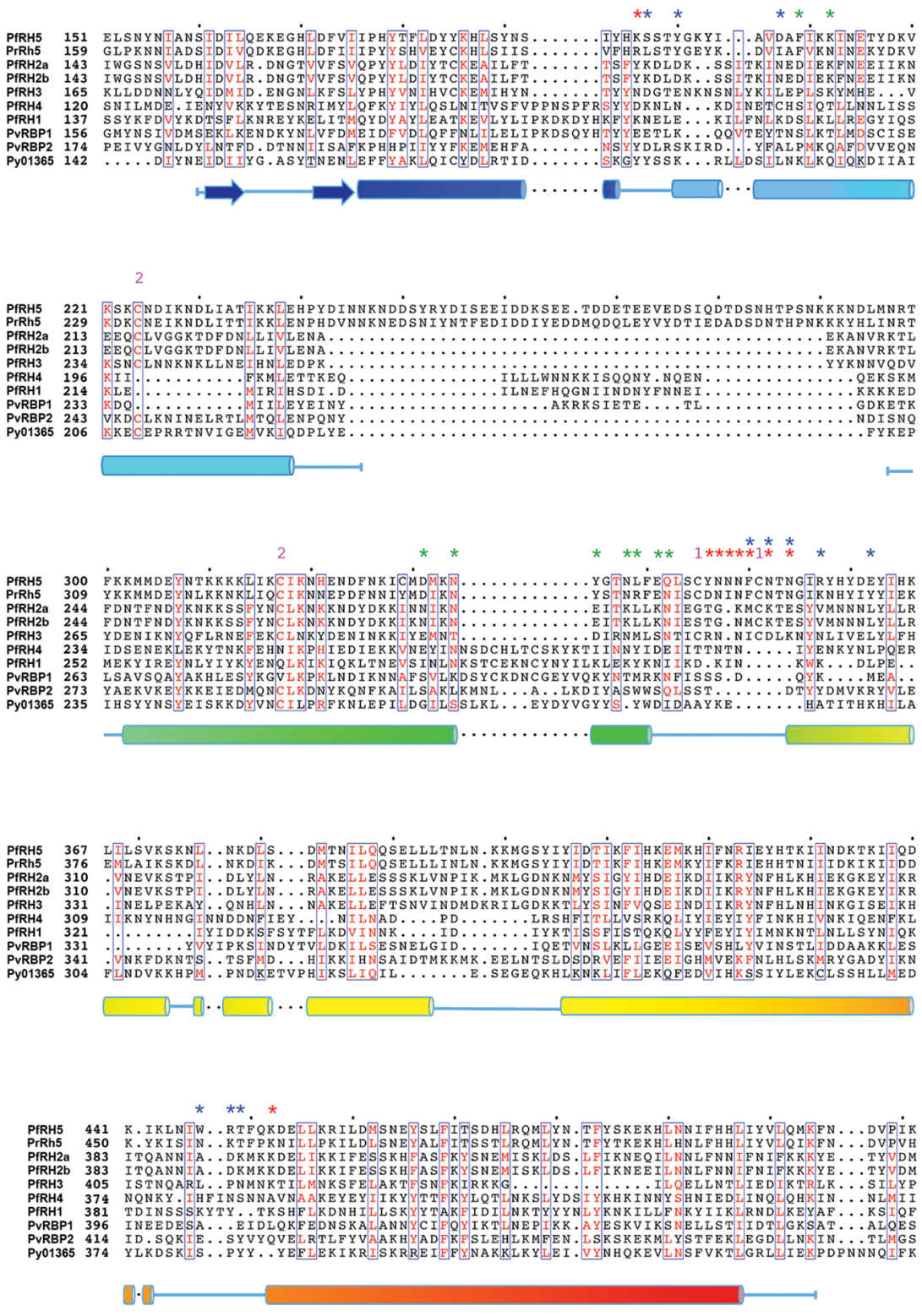


FIGURE 4.13: A conserved PfRH5-like fold in other *Plasmodium* RH proteins. *P. falciparum* RH1, RH4, RH2b, RH2a and RH3 (a pseudogene); *P. vivax* RBP-1 and RBP-2; *P. reichenowi* RH5; and *P. yoelii* Py01365 were aligned using Clustal Omega [266] and were threaded using the Phyre2 server [269], giving more than 98% confidence of fold conservation over >260 residues in each case. The secondary structure of PfRH5 is shown below the sequence in a rainbow colour scheme. Residues from PfRH5 that interact with basigin, QA1 and 9AD4 are indicated above the sequence by blue, red or green stars respectively (see Chapters 5-6). Cysteine residues that make disulphide bonds are indicated by pink numbers, with residues sharing the same number forming a disulphide bond.

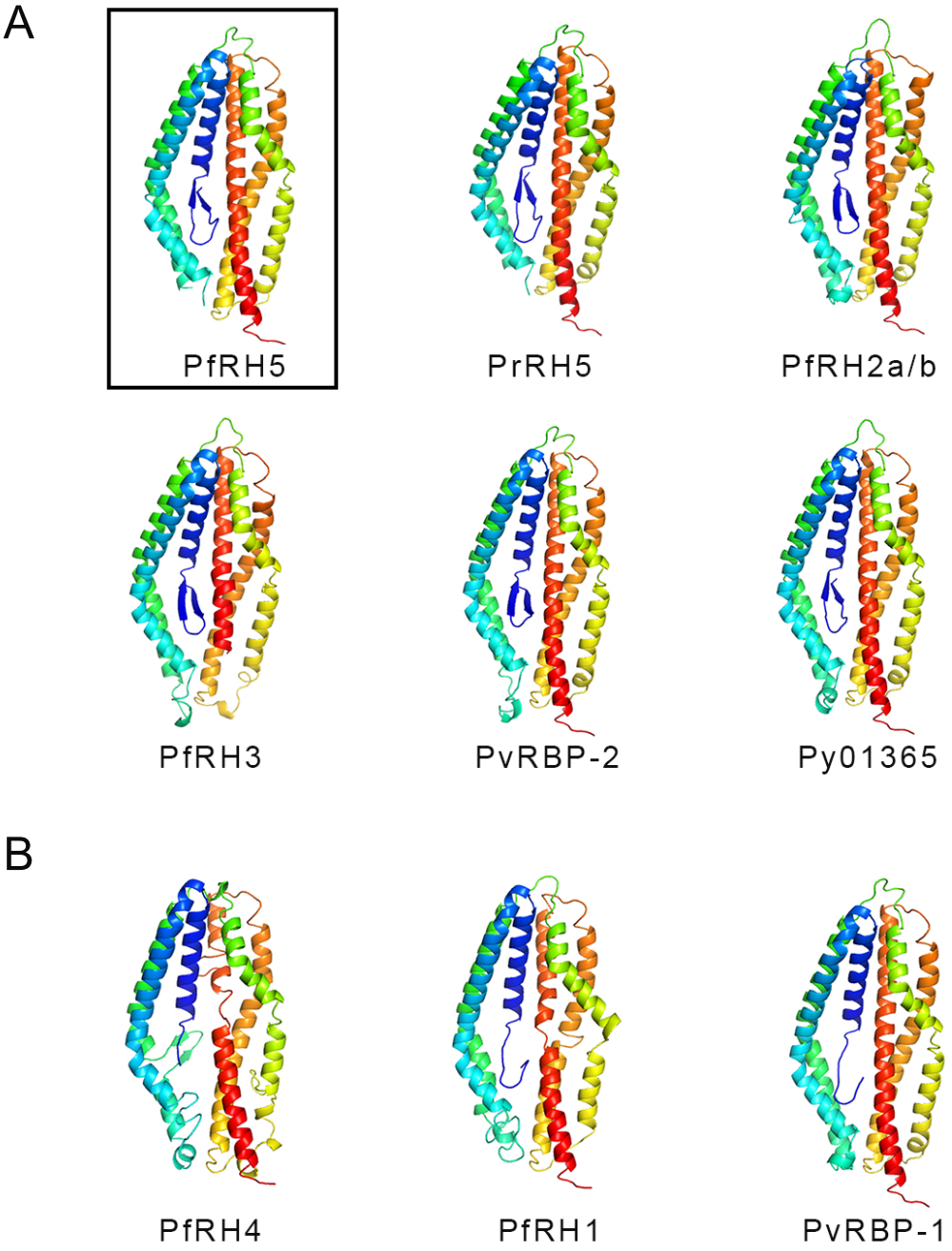


FIGURE 4.14: Models of other RH proteins generated by Phyre2 threading. The PfRH5 structure is boxed. Shown are (A) other members of the PvRBP-2 clade, including PfRH5, PfRH2a/b (which are identical across the PfRH5 homology region), PfRH3 (a pseudogene), PvRBP-2, and Py01365; and (B) members of the PvRBP-1 clade, including PfRH4, PfRH1, and PvRBP-1. Note that the overall fold is conserved, but there is considerable more variation relative to PfRH5 in the PvRBP-1 clade than the PvRBP-2 clade.

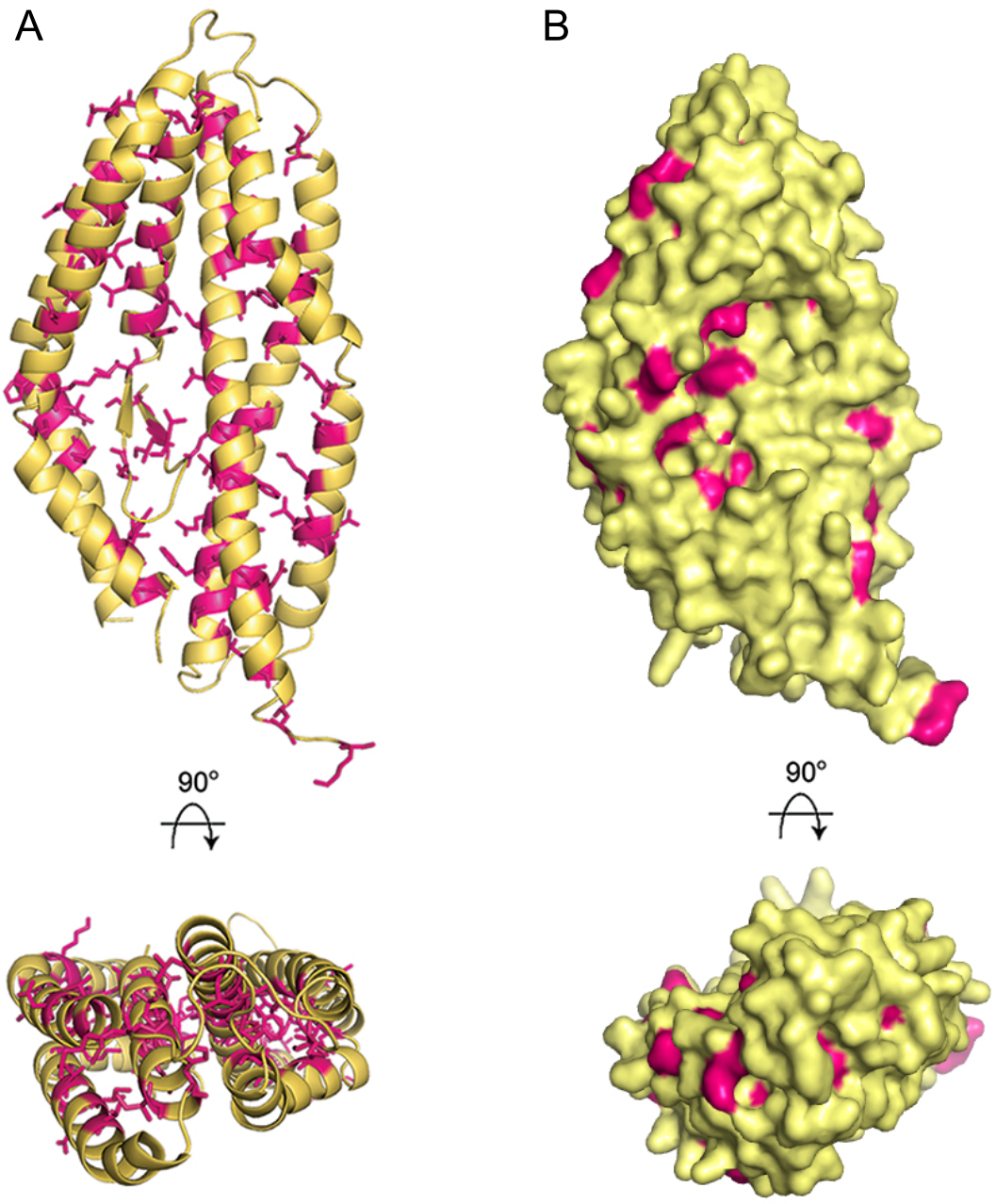


FIGURE 4.15: Residues similar across RH proteins are primarily located in the interior of the PfRH5 fold. (A) PfRH5 is shown in yellow, with residues similar among RH proteins (Figure 4.13) highlighted as pink sticks. (B) is a surface representation, with colours as in (A).

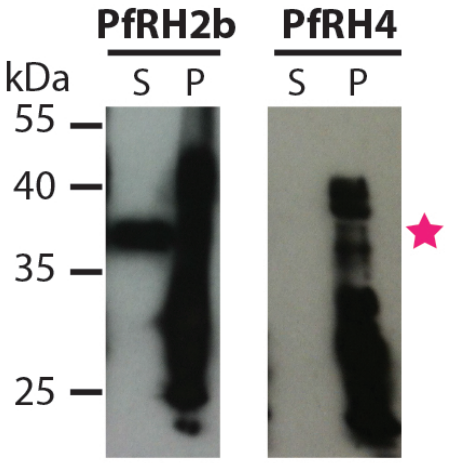


FIGURE 4.16: PfRH2b and PfRH4 expression in insect cells. Cells were grown for 4 days before harvesting by centrifugation. The supernatant (S; containing secreted protein) and cell pellet (P) were subjected to SDS-PAGE and Western blot analysis. Both proteins should be ∼35 kDa.

4.7 Discussion

Ultimately, crystallisation of PfRH5 was facilitated by a number of modifications to the protein, genetic, proteolytic and chemical. The gene was modified both to remove glycosylation sites, which enabled the effective Concanavalin A purification step (Section 3.4), and to remove predicted disordered regions at the N-terminus and in an internal loop. Constructs with these disordered regions still present did not produce crystals: for example, unsuccessful crystallisation plates were set up with full-length PfRH5 complexed with basigin, or with PfRH5ΔN together with basigin, and sometimes also antibody fragment 4BA7, which binds the disordered loop.

The proteolytic processing of PfRH5ΔNL with endoproteinase GluC, which removed the C-terminal 6-His tag, and perhaps more residues at one or both termini, was probably important for subsequent crystallisation. As discussed in Chapters 5-6, crystal packing, particularly in the PfRH5ΔNL–basigin crystals, led to the close apposition of the PfRH5ΔNL C-termini. Clearly, differences in GluC cleavage between protein preparations, or within a single preparation (due to incomplete cleavage or cleavage at different residues, either the glutamate or the aspartate near the C-terminus) could influence crystallisation and crystal packing. Indeed, I noticed that only one particular preparation of PfRH5ΔNL–basigin produced crystals with aspirations to high-resolution diffraction (Chapter 5).

Crystallisation was also facilitated by mixing PfRH5 with a binding partner, either basigin or an inhibitory antibody fragment. The function of complexation was threefold: firstly, to stabilise PfRH5, which on its own is prone to stickiness (e.g. sticking to gel filtration resin) and aggregation (e.g. aggregating upon concentration); secondly, to serve as a sort of 'crystallisation chaperone,' to provide additional crystal contacts; and thirdly, to allow molecular replacement with a known structure as a search model. Perhaps the most effective single modification was chemical methylation of the protein. Crystallisation trials with unmethylated and methylated PfRH5ΔNL–basigin side-by-side revealed completely different solubility properties: the unmethylated complex was much more prone to precipitation, and did not crystallise. Of course, these modifications acted in concert. For example, methylation of PfRH5ΔNL made it more stable on its own, allowing it to be concentrated to 8 mg/ml for crystallisation trials; but without a binding partner, it did not crystallise.

Crystals of PfRH5ΔNL–9AD4 diffracted reproducibly, and the structure was solved by molecular replacement of the antibody fragment to 2.3 Å resolution, with refinement statistics comparable to those of other high-quality crystal structures at the same resolution. PfRH5ΔNL adopts a novel, kite-like fold, with a pseudo-two-fold rotation symmetry linking together two bundles of three helices each. Comparison with a SAXS envelope of full-length PfRH5 reveals extra density at one end of the kite in the full-length protein, near the C-terminus and the loop truncated in PfRH5ΔNL. Even though the N-terminus will eventually connect to the β-sheet at the centre of the kite, it may be that the ∼160 N-terminal residues missing in the crystal structure also contribute to the density at the bottom of the molecule, though there is no solid evidence to support this yet. The role, if any, of the N-terminal extension, is unknown: perhaps part of its role could be to protect or mask a binding site at the foot of the molecule from immune recognition.

One of the main conclusions of this chapter—based on sequence alignments and threading—is that other RH proteins, from *P. falciparum* and other *Plasmodium* species, appear to include the same PfRH5 kite-like fold. No residues or disulphide bonds are totally conserved across the family, but similar, primarily hydrophobic residues cluster in the interior of the domain, where they presumably act to stabilise the fold. The surface of the domain, by contrast, is not conserved, and could permit diversity of binding properties and function. The domain is located at the N-terminus of these extremely long relatives of PfRH5, where it may act to bind specifically to erythrocyte receptors. To test this, I have begun to express the putative PfRH5-like domain from other RH proteins of *P. falciparum* and *P. vivax*. If expression is successful, I will be able to confirm whether PfRH4 binds to complement receptor 1 using this domain. For the other RH proteins such as PfRH2b, whose erythrocyte receptors are currently unknown, I will attempt to identify their receptors by screening erythrocyte surface protein libraries for binding to the PfRH5 domain.

How does the structure of the PfRH5 domain compare to other families of *Plasmodium* surface proteins? Like the DBL domains which make up the EBL family of invasion proteins, the RH protein fold is largely α-helical, providing a robust scaffold for the variable surface residues to mediate the different interactions involved in invasion. The surfaces of the DBL domains of PfEMP1 and the PAN domains of AMA1 are decorated with extended loops, perhaps as an immune diversion. The PfRH5ΔNL domain has

only short loops, but the removed loop and N-terminus might function in an analogous manner.

Chapter 5

Structural Basis for Basigin Binding by PfRH5

5.1 Summary

To determine where basigin binds to PfRH5, PfRH5ΔNL was crystallised in complex with basigin, and the structure solved to 3.1 Å by molecular replacement. After inspection of the binding site, basigin mutant proteins were generated to investigate the importance of individual interactions in the overall complex affinity. Finally, the stoichiometry of the complex was explored by crystallographic and solution methods.

5.2 PfRH5–basigin crystallisation, data collection, and data processing

5.2.1 Crystallisation

PfRH5–basigin complexes were prepared as described in Chapters 3-4. A number of broad screens for PfRH5ΔNL–basigin crystallisation were set up, including JCSG I-IV, Morpheus, INDEX, MIDAS, ProPlex, and PACT (Hampton Research). Crystals were obtained in the ProPlex screen at 4°C, in a condition containing 0.2 M lithium sulphate, 0.1 M MES pH 6.0, and 20% PEG 4000. The ProPlex screen is formulated based upon a protein complex crystallisation database, and is intended to facilitate the crystallisation

of protein complexes. These crystals were harvested, pulverised, and used to seed further crystallisation drops which also contained Silver Bullets additives (Hampton Research).

The optimised crystals were of several morphologies, among them a square block and an elongated rectangle with narrowed ends, rather like the hull of a boat (Figure 5.1). The prow and stern of the boat-like crystals were somewhat jagged, suggesting that they might not be well-ordered. Crystals of each form were harvested, a procedure which proved extremely problematic, as many of the crystals were stuck to the plastic of the crystallisation plate. Many techniques were developed to attempt to dislodge them, including gentle prodding, more vigorous levering, and gouging the plastic nearby to try to pop them off; perhaps least successful of all the methods was the application of dry ice below the crystallisation plate, which lead to the freezing of the well solution. Once freed by some method or other, for better or for worse, crystals were transferred into a drop of well solution supplemented with 25% ethylene glycol to cryoprotect them, and cryo-cooled in liquid nitrogen.

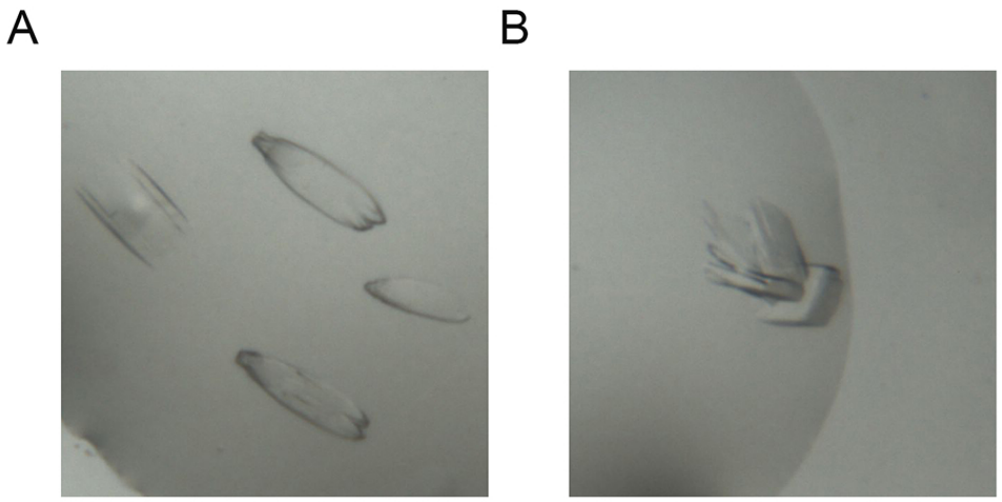


FIGURE 5.1: Crystals of the PfRH5ΔNL–basigin complex. Initial crystals were grown in 0.2 M lithium sulphate, 0.1 M MES pH 6.0, and 20% PEG 4000 in a ProPlex crystallisation screen at 4°C. The crystallisation drops were set up by mixing 100 nl of 8 mg/ml protein complex with 100 nl of well solution. Crystals appeared after approximately one week. Crystal drops were set up with 100 nl of 8 mg/ml complex, 50 nl of well solution, 50 nl of seeds (generated from the initial crystals), and 50 nl of Silver Bullets additives. (A) and (B) show two of this next generation of crystallisation drops. (A) Some crystals manifest a square morphology (top left) and others resemble the hull of a boat with jagged prow and stern. (B) A fan of crystals emanates from a square block crystal. The individual crystals of the fan were dislodged by prodding the block crystal.

5.2.2 Data collection

The crystals were taken to Diamond Light Source beamline I04-1. In general, they proved to be poor diffractors (7-20 Å), particularly the dingy dinghy crystals. Fortunately, however, one of the thin crystals radiating from the block in Figure 5.1B diffracted to ∼3.1 Å. iMOSFLM classified it in the P222 space group, with unit cell dimensions of a = 75.5 Å, b = 109.79 Å, and c = 152.20 Å. A full dataset of 900 images was collected according to EDNA's suggestion (Table 5.1); one of the images is shown in Figure 5.2.

Beamline	I04-1
Wavelength	0.92000 Å
Aperture	70 μm
Omega Start	49°
Oscillation	0.2°
Exposure	0.2 s
Transmission	47%
No. of images	900

Table 5.1: PfRH5ΔNL–basigin data collection strategy.

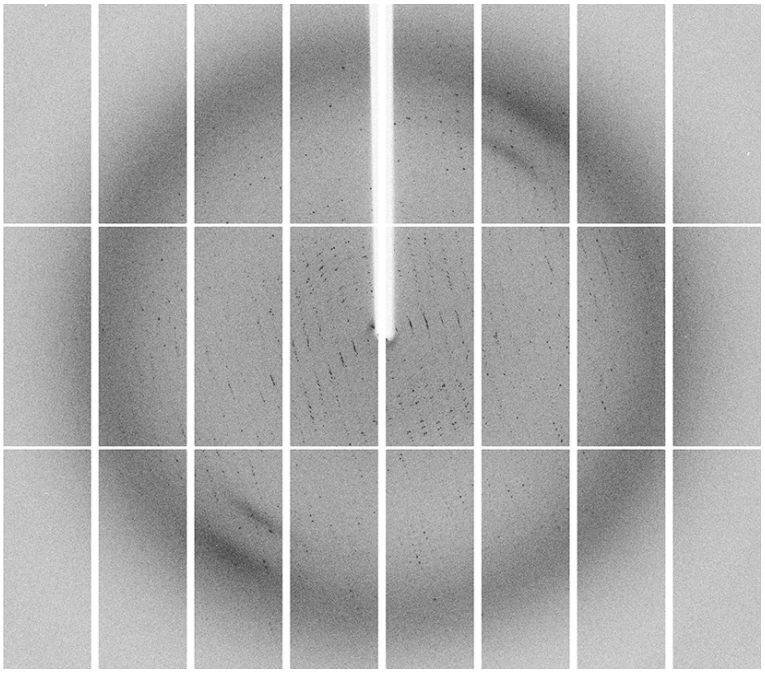


Figure 5.2: PfRH5ΔNL–basigin diffraction image from beamline I04-1 at Diamond Light Source.

5.2.3 Data processing

The data were processed using the same pipeline as for PfRH5ΔNL–9AD4 (Section 4.3.3). The refined unit cell parameters, given by iMOSFLM, were a = 75.28 Å, b = 109.36 Å, and c = 151.95 Å, with $\alpha = \beta = \gamma = 90°$. After integration of all 900 images, Pointless determined the space group as $P2_12_12_1$. During scaling and merging, the resolution was set to 3.1 Å, giving $I/\sigma I = 2.7$ in the highest resolution shell. R_{pim} was 0.08 for the full dataset, and 0.297 for the highest resolution shell (Table 5.2). The dataset was 99.8% complete, with a 6-fold redundancy. Calculation of the Matthews coefficient gave a solvent estimate of 54%, with two PfRH5ΔNL–basigin complexes in the asymmetric unit.

Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	75.28, 109.36, 151.95
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	67.46–3.10 (3.27-3.10)
$\mathbf{R}_{merge}$	0.184 (0.690)
$\mathbf{R}_{rim}$	0.201 (0.753)
$\mathbf{R}_{pim}$	0.080 (0.297)
$\mathbf{CC}_{1/2}$	0.988 (0.761)
$\mathbf{I/\sigma I}$	9.4 (2.7)
Completeness (%)	99.8 (99.8)
Redundancy	6.1 (6.2)

Table 5.2: PfRH5ΔNL–basigin data processing statistics.

5.2.4 Phasing attempts

The PfRH5ΔNL–basigin crystals actually predated the PfRH5ΔNL–9AD4 crystals; and so, before the structure of PfRH5 was solved using the diffraction of the latter, I attempted molecular replacement and experimental phasing methods with the diffraction of the former. For molecular replacement, the two immunoglobulin domains of basigin (PDB code 3B5H) were separated into two different search models, since the linker between them is flexible and they can adopt many orientations relative to each other. Each domain only represented $\frac{1}{6}$ of the total scattering mass, and furthermore there were two of each domain per asymmetric unit (according to the Matthews analysis), making this a challenging molecular replacement. Neither Phaser nor MolRep produced solutions

containing all four immunoglobulin domains; nor did any of the solutions manifest low enough R-factors to be considered correct.

Therefore, a experimental phasing campaign was launched. Experimental phasing attempts to exploit anomalous differences between Bijvoet pairs, or differences in the structure factor magnitudes from crystals with slightly different contents (e.g. with and without heavy metals bound). The campaign had two fronts: either existing crystals were modified by the addition of metals, halides, or xenon; or the proteins were modified to incorporate selenomethionine and new crystals generated. Additives assayed included various mercury and platinum compounds, samarium, europium, dysprosium, praseodymium, holmium, cadmium, iridium, gold, lead, xenon, iodide, and bromide. A crystal was transferred into a droplet of well solution containing 0.1–1 mM of the heavy metal, placed in a hanging drop set-up over well solution, and soaked for 24 hours. After the soak, the crystal was transferred into a drop of well solution, incubated for 5 minutes to remove any unbound metal, transferred into cryo-protected well solution, and cryo-cooled. For bromide and iodide soaks, the crystal was transferred into cryo-protected well solution containing 0.6 M sodium halide, and frozen directly after 10-30 seconds. Xenon was incorporated using a xenon source at Diamond Light Source.

I prepared selenomethionine derivatives of basigin in *E. coli* and of PfRH5ΔNL in S2 cells. I managed to grow crystals of selenomethionine-containing basigin complexed with native PfRH5ΔNL, but no crystals formed when both components contained selenomethionine. Finally, I also attempted to pre-derivatise basigin with heavy metal compounds: first, I incubated basigin with various mercury and platinum compounds and performed a SDS-PAGE gel-shift assay to determine which metals had strongly bound to the protein. Basigin modified with HgO, $Hg(CN)_2$, or dichloro(ethylenediamene)platinum (II) was retarded in its progress through the gel, indicating that metal binding had occurred. Basigin modified with these compounds was then mixed with PfRH5ΔNL and crystals grown. The resultant selenomethionine-, Hg- or Pt-derivatised crystals were harvested as usual.

The crystals were taken to various beamlines at Diamond Light Source and the European Synchrotron Radiation Facility (ESRF), and thus began a seemingly endless succession of abysmally diffracting crystals—over 210 in total. When, occasionally, a crystal diffracted

to within 5 Å (as occurred, for example, with one $K_2Pt(CN)_4$ and several bromide-derivatised crystals), a fluorescence scan, or downstream attempts to locate sites with SHELXD, revealed no bound metal or halide. Fortunately, the structure determination of PfRH5ΔNL–9AD4 (Section 4.4) allowed molecular replacement to be revisited, and the experimental phasing project was abandoned with alacrity.

5.3 Solving and validating the PfRH5–basigin structure

PfRH5ΔNL and the two individual immunoglobulin domains of basigin were used as search models in molecular replacement of the native PfRH5ΔNL–basigin dataset (Section 5.2.3). First, two copies of PfRH5ΔNL were placed by Phaser; rigid body refinement of the solution gave R-factors of 42% (R_{work}) and 45% (R_{free}), which suggested the solution was correct. Oddly, however, there was a demonstrable clash between the two PfRH5 molecules, at one end, near their C-termini. A loop (residues 395-406) and the C-terminus (498-end) were removed from the search model and the molecular replacement repeated. The same solution was obtained, with even lower R-factors (∼35%), suggesting that the C-terminus and this loop adopt different conformations in PfRH5ΔNL–basigin and PfRH5ΔNL–9AD4 structures.

Next, Phaser was used to search for two copies of each of the basigin immunoglobulin domains. The two N-terminal basigin domains were located, at the tips of the two PfRH5 kites (distal from the C-terminus). Only one of the basigin C-terminal domains was placed by Phaser; MolRep was therefore used to dock the other. The refinement and model building cycle was launched. Model building included rebuilding of the C-terminus and the 295-406 loop, which, indeed, were in a different conformation than in the PfRH5ΔNL–9AD4 structure.

The final electron density is shown in Figure 5.3, and the final refinement statistics in Table 5.3. The R_{work} and R_{free} values were 21.71% and 24.45%, respectively. The structure was stereochemically sound, as the r.m.s. deviations were 0.01 Å(bond lengths) and 1.5° (bond angles), and the average B-factor for all atoms was 50.0 $Å^2$. The Ramachandran plot showed 93% of residues were located in the most favoured regions and the remaining 7% in the additional allowed regions (Figure 5.4).

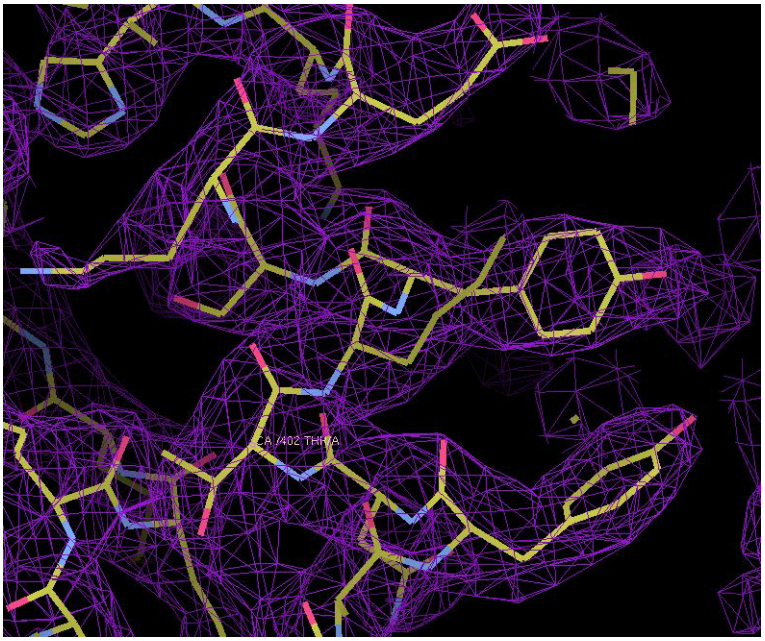
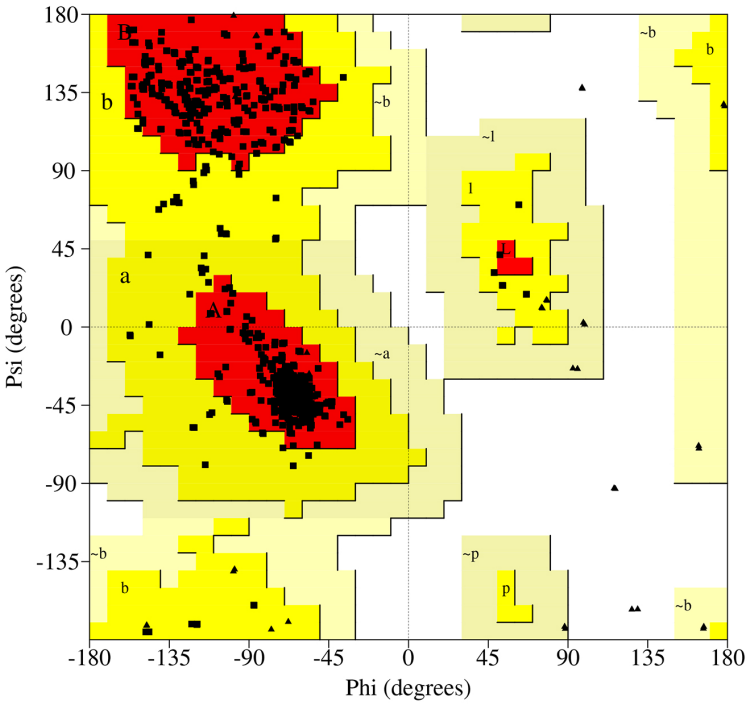


FIGURE 5.3: Final PfRH5ΔNL–basigin electron density map ($2F_o$ - F_c, contour 1σ), with structure superimposed. A helix of PfRH5 is shown. Atoms are coloured yellow (carbon), blue (nitrogen), or pink (oxygen). The figure was prepared using Coot [303].

Resolution (Å)	67.46-3.10
No. reflections ($\mathbf{R}_{free}$)	23357 (1186)
$\mathbf{R}_{work}$/ $\mathbf{R}_{free}$ (%)	21.71 / 24.45
No. atoms	7629
Average B-factors (Å^2)	50.0
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.5

TABLE 5.3: PfRH5ΔNL–basigin refinement statistics.



	No. Residues	Percentage
Most favoured regions	801	93.0
Additional allowed regions	60	7.0
Generously allowed regions	0	0
Disallowed regions	0	0
Residues not Gly or Pro	861	
End-residues (excl. Gly or Pro)	12	
Gly residues	40	
Pro residues	20	
Total	933	

FIGURE 5.4: Ramachandran plot and statistics of PfRH5ΔNL–basigin structure, from PROCHECK [307].

5.4 Analysis of the PfRH5–basigin structure

The binding site for basigin lies at the tip of PfRH5, distant from the flexible loop and C-terminus, with both of the basigin domains and the linker making direct contacts with PfRH5 (Figure 5.5). This is consistent with previous studies showing that both basigin domains are required for PfRH5 binding [238]. Most of the contact occurs between the N-terminal domain of basigin and the loops at the tip of PfRH5. Interestingly, because of inherent flexibility of the linker, the C-terminal domain of basigin actually adopts different orientations in the two copies of the complex in the asymmetric unit of the crystal (Figure 5.6). When the two complexes are aligned based on the PfRH5 structure, the C-termini of basigin are separated maximally by ∼18 Å.

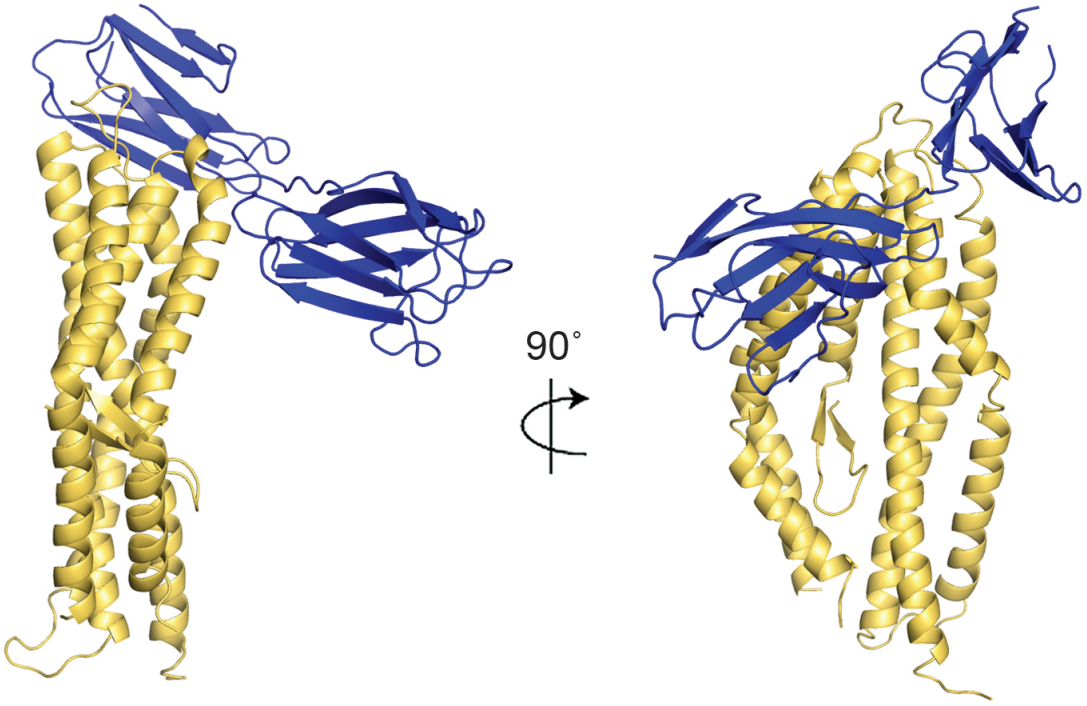


FIGURE 5.5: The structure of PfRH5ΔNL (yellow) bound to basigin (blue).

5.4.1 SAXS analysis of flexibility in basigin binding

SAXS analysis provided further evidence of the flexibility of the basigin C-terminal domain in binding to PfRH5 (Figure 5.7). Data were collected for full-length PfRH5 on its own and in complex with basigin. The resultant scattering was used to calculate the distance-distribution function ($P(r)$), and thence to calculate twenty *ab initio* models, which were averaged to give a molecular envelope. The PfRH5 SAXS model

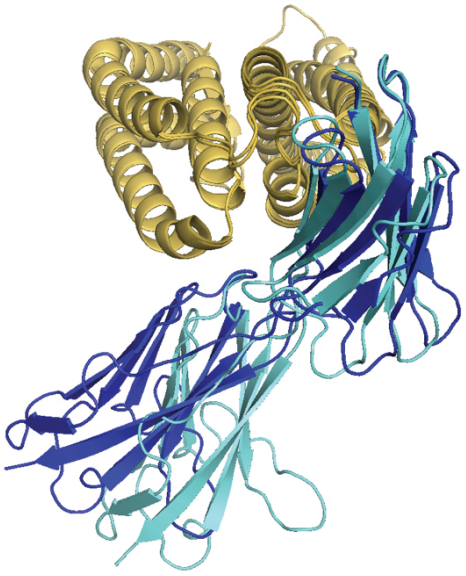
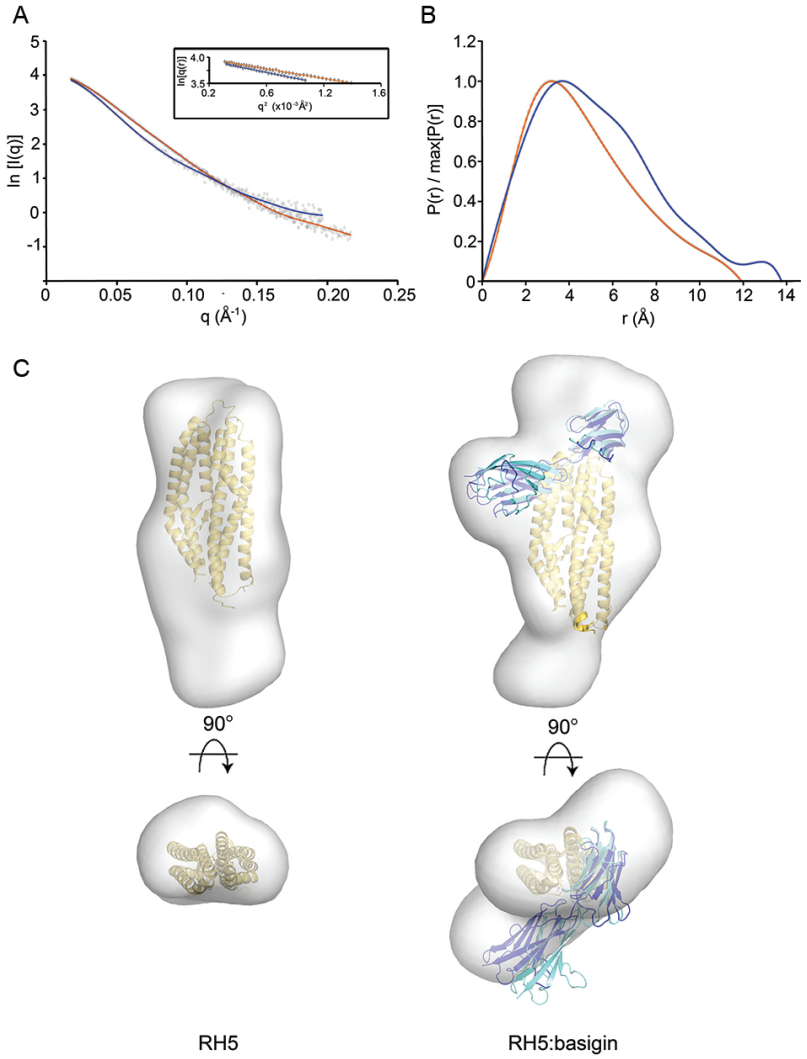


FIGURE 5.6: A top view of the PfRH5ΔNL–basigin complex showing the two conformations of basigin (blue and cyan) found in the asymmetric unit, aligned on PfRH5ΔNL.

nearly perfectly envelops the flat, kite-shaped PfRH5ΔNL crystal structure, with extra density probably corresponding to regions removed from the crystallisation construct (Section 4.5.2). By contrast, while PfRH5 and the N-terminal domain of basigin fit into the PfRH5–basigin envelope, the C-terminal domain (in both orientations observed in the crystal structure) only partially fits (Figure 5.7C). This is consistent with a more flexible interaction of this domain with the PfRH5 helices. The extent of this flexibility, and the existence of further PfRH5-binding modes besides the two observed in the crystal, is a matter of speculation.



D

	Rg (nm)	D_{max} (nm)	Porod volume (nm^3)	MW_{app} (kDa)	NSD	χ
PfRH5	3.7	11.9	104.5	61.4	0.692±0.036	1.01
PfRH5–basigin	4.1	13.8	138.3	81.3	0.715±0.034	0.74

FIGURE 5.7: SAXS analysis of the PfRH5–basigin complex. (A) The theoretical scattering calculated from the average of 20 *ab initio* reconstructions (continuous lines, with PfRH5 in orange and PfRH5–basigin in blue) plotted with the experimental scattering intensity curves (diamonds). The data are presented as the natural logarithm of the intensity. Guinier plots are displayed in the inset. (B) The distance distribution function, $P(r)$, of PfRH5 and PfRH5–basigin. (C) To the left, the crystal structure of PfRH5ΔNL (yellow) was docked into the average *ab initio* SAXS envelope of full-length PfRH5 (grey). Extra density corresponding to some or all of the truncated regions is visible at the bottom of the kite-like structure, near the C-terminus. To the right, the crystal structure of PfRH5ΔNL–basigin is docked into the average *ab initio* SAXS envelope of full-length PfRH5–basigin (grey). Basigin molecules from the two PfRH5ΔNL–basigin complexes in the asymmetric unit (blue and cyan) are superimposed based on the structure of PfRH5ΔNL. (D) Summary of SAXS parameters. The radius of gyration (R_g) was determined from the Guinier plot using AutoRg, and the maximum particle dimension (D_{max}) and the Porod volume were calculated using GNOM. An estimate of the molecular weight was obtained by dividing the Porod volume by 1.7. *Ab initio* modeling was used to generate 20 shape reconstructions from the data. The normalised spatial discrepancy parameter (NSD) diagnoses the similarity of these models. The models were averaged and the fit of the average model to the experimental data is indicated by the χ value. Situs and Sculptor were used to manually fit the crystal structures into the SAXS models.

5.5 Dissection of the basigin binding site

5.5.1 A detailed tour of the binding site

The two domains of basigin wrap around the tip of PfRH5 like a delicately draped stole (Figure 5.5), forming a contact area of ∼1350 Å^2. The majority of the contact occurs between strands A and G of the N-terminal domain of basigin and the loops at the tip of PfRH5 (Figure 5.8). These contacts mostly involve hydrogen bonding to the backbone of basigin by the side chains of PfRH5 and the backbone of PfRH5 residues 447-449. Residues F350 and W447 of PfRH5 stabilise the interaction by packing into hydrophobic pockets on basigin. The interactions are summarised in Table 5.4. Notably, very few basigin side chains are involved in the contact, reducing the chance of basigin escape mutants that prevent PfRH5 binding and impair parasite invasion.

Both the C-terminal domain of basigin and the side chain of H102 in the linker also make direct contacts with PfRH5. The three loops at the tip of the C-terminal domain (linking strands B and C, strands D and E and strands F and G) interact with the second and fourth helices of PfRH5. This interaction is stabilised by hydrogen bonds and by a hydrophobic patch contributed by residues VPP from the BC loop. However, in each orientation of the C-terminal domain seen in the crystal, only two of these loops contact PfRH5. Chain B interacts through the BC and DE loops (a contact area of ∼650 Å^2) while chain D interacts through the BC and FG loops (∼480 Å^2) (Figure 5.8). None of the three glycosylation sites of basigin—at residue 44 in the N-terminal domain, and residues 152 and 186 in the C-terminal domain—are located in or near the binding site for PfRH5.

5.5.2 Naturally occurring polymorphisms and host tropism

PfRH5 is highly conserved, with just twelve non-synonymous SNPs found in 227 isolates from across the world, and only 5 at frequencies of 10% or greater [126, 128, 129]. These SNPs are distributed across the surface of the domain, with just one, Y203, forming part of the basigin binding site (Figure 5.9). In spite of its location however, Y203 has no effect on PfRH5–basigin affinity, as demonstrated by a previous study comparing the basigin-binding affinity of PfRH5 from the 3D7 strain (containing C203) and the 7G8 or

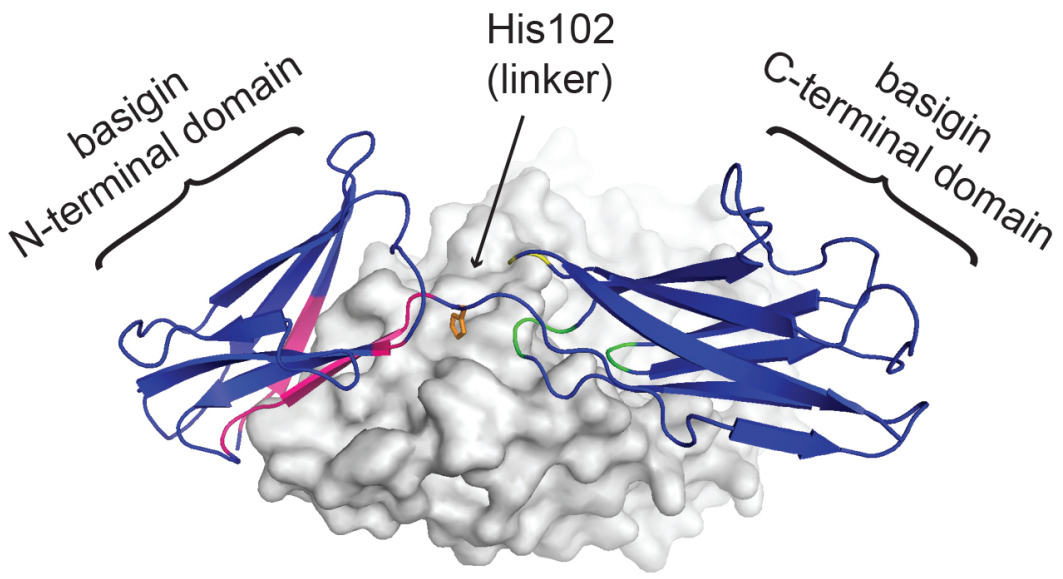


FIGURE 5.8: Close-up of the PfRH5–basigin binding site. Basigin residues in the N-terminal domain (pink), the linker (His102, orange stick), and the C-terminal domain (green) contact PfRH5 (grey surface). In the alternative basigin conformation in the asymmetric unit, the yellow loop contacts PfRH5.

PfRH5			Basigin			
Chain	**Residue**	**Group**	**Chain**	**Residue**	**Group**	**Type of interaction**
			N-terminal domain:			
A / C	S197	side chain	B / D	Q100	side chain NH2	hydrogen bond
A / C	S197	side chain	B / D	E84	side chain	hydrogen bond
A / C	F350	side chain	B / D		hydrophobic pocket	hydrophobic
A / C	N352	side chain CO	B / D	N98	backbone NH	hydrogen bond
A / C	N354	side chain NH2	B / D	N98	backbone CO	hydrogen bond
A / C	R357	side chain	B / D	V26	backbone CO	hydrogen bond
A / C	W447	side chain	B / D		hydrophobic pocket	hydrophobic
A / C	W447	backbone CO	B / D	T28	backbone NH	hydrogen bond
A / C	R448	side chain	B / D	T25	side chain	hydrogen bond
A / C	T449	side chain	B / D	V26	backbone NH	hydrogen bond
A / C	T449	backbone NH	B / D	V26	backbone CO	hydrogen bond
			Linker:			
A / C	Y200	side chain	B / D	H102	side chain	hydrogen bond
			C-terminal domain:			
A / C		hydrophobic pocket	B / D	V131	side chain	hydrophobic
A		hydrophobic patch	B	P133	side chain	hydrophobic
A	D207	side chain	B	Q164	side chain NH2	hydrogen bond
C	E362	side chain	D	K191	side chain	hydrogen bond
C	E362	side chain	D	S190	side chain	hydrogen bond

TABLE 5.4: Table describing interactions between PfRH5 and basigin.

GB4 strains (containing Y203) [126]. Thus, the few naturally occurring polymorphisms of PfRH5 are not clustered in the basigin binding site, and do not appear to modulate basigin binding.

Several studies have investigated how SNPs in PfRH5 and in basigin affect host tropism—that is, which species of non-human primates that *P. falciparum* can invade. In sequenced laboratory strains (as opposed to the field strains described above), eight different PfRH5 SNPs have been associated with increased ability of *P. falciparum* to invade New World monkey (*Aotus*) erythrocytes [235, 243] (Figure 5.10). A number of these residues (I204, N347, Y358 and E362) are close to, or within the basigin binding site, and changes here are likely to affect the interaction of PfRH5 with basigin from different species (Figure 5.9).

Similarly, basigin residues which, when mutated, affect PfRH5 affinity (F27 and Q100) [242] are also located at the interface between the binding partners (Figure 5.9C). (These residues were identified by comparing gorilla, chimpanzee, and human basigin sequences.) An insertion (+H103) next to H102 in the basigin linker reduces the affinity for PfRH5 (this insertion is found in gorilla basigin, which does not bind PfRH5). Therefore, as expected, many of the residues previously linked to host tropism in PfRH5 and basigin are located at the interface between them.

5.5.3 Mutagenesis of basigin

Both domains of basigin contribute to PfRH5 binding [238]. To determine their relative contributions, and to analyse the effect of disrupting individual interactions on the overall affinity, I designed five basigin mutations, two in each of the domains and one in the linker (Figure 5.11C). In the N-terminal domain, T28F and G95F were intended to sterically disrupt packing of PfRH5 loops against basigin. In particular, T28F is located where a PfRH5 loop forms a short strand which adds to a basigin β-sheet. H102A abolishes the specific interaction of the basigin linker with PfRH5. V131D should disrupt the VPP loop interaction common to both C-terminal domain binding modes. Finally, S190D should remove an interaction that occurs in only one of the C-terminal domain binding modes.

The mutagenesis was performed using the QuikChange method. Expression and purification of the mutant proteins, and downstream analysis by SPR and circular dichroism, were performed by a summer student, T. Waksman, under my guidance. The mutant proteins were all properly folded, according to circular dichroism (Figure 5.11B). PfRH5 binding was assessed using SPR, with PfRH5 bound to the chip surface, and the various

basigin mutant proteins flowed over. Interestingly, all of the variants bound to PfRH5 (Figure 5.11A). T28A was the only variant with a significantly altered affinity, binding ∼six-fold more weakly than wild-type, with both a slower on-rate and a faster off-rate (Figure 5.11D). This implies that the β-strand addition is probably marginally more influential on binding than the other interactions, but nevertheless patently unessential. (Perhaps worth mentioning, H102A appears to have a faster on-rate compared to wild-type–could the mutation render the linker more flexible, and allow more rapid interaction, or more favourable relative orientation, of neighbouring N- and C-terminal domains?)

One of the most striking characteristics of the PfRH5–basigin interface is the relative lack of involvement of basigin side chains in the interaction (Table 5.4); the ability of each of these basigin mutant proteins to bind PfRH5 lends credence the idea that basigin cannot easily escape PfRH5 by mutation, with erythrocyte invasion the inevitable consequence.

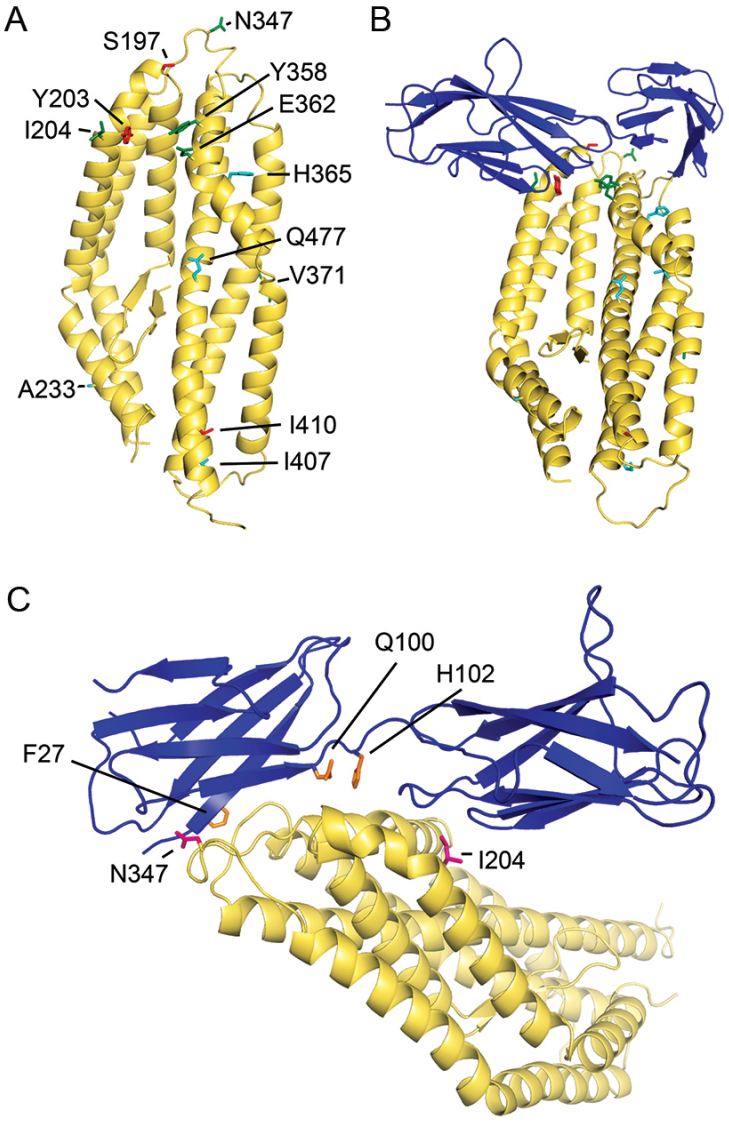
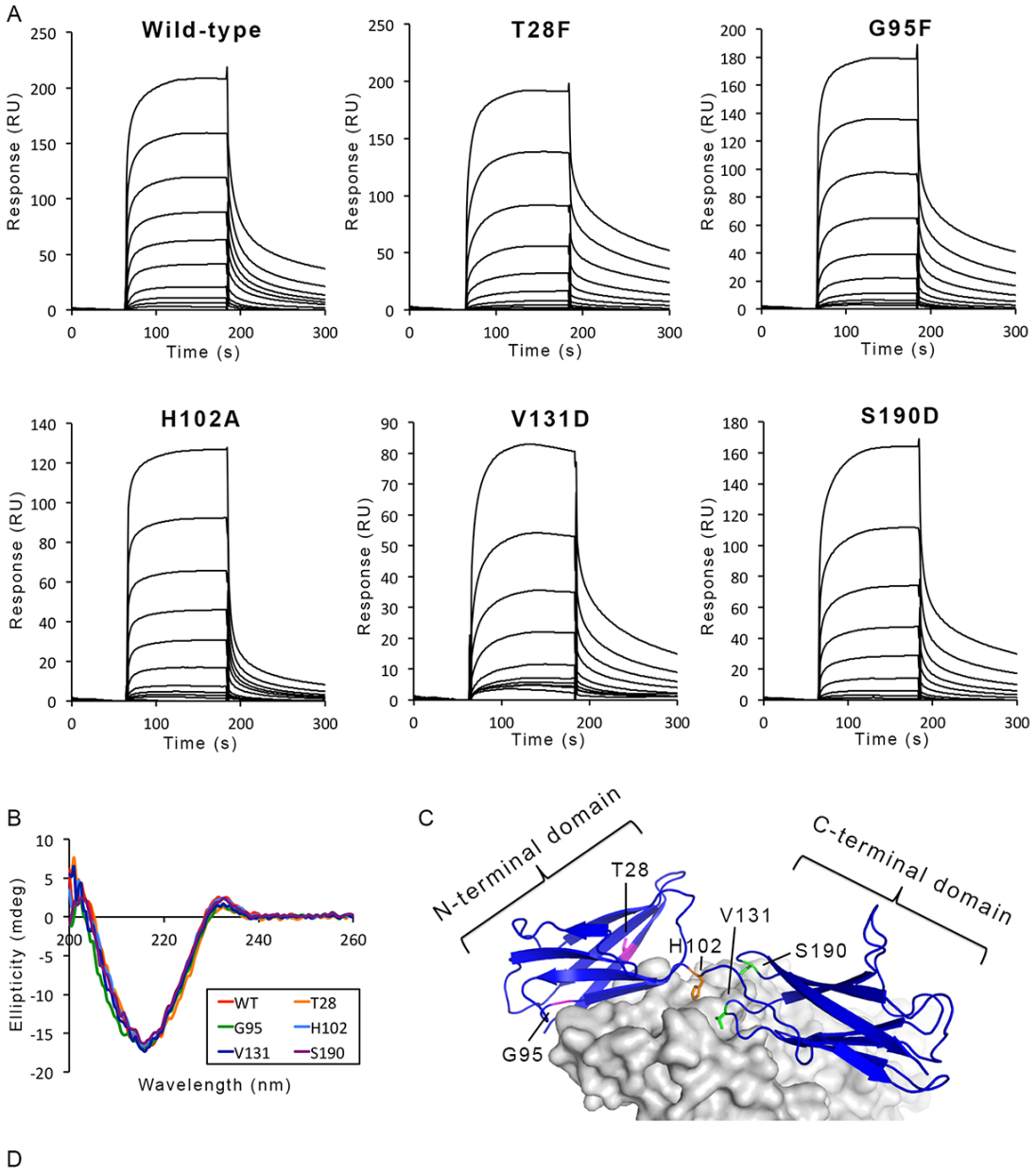


FIGURE 5.9: Location of PfRH5 polymorphisms, and residues of PfRH5 and basigin implicated in host tropism. (A–B) Indicated are the locations of PfRH5 SNPs that are common (10% frequency or greater; red sticks) or uncommon (blue sticks) among 227 field isolates [126, 128, 129], as well as additional SNPs observed in lab strains (green sticks) [235, 243]. (B) Basigin (blue) is shown in addition to PfRH5 (yellow). SNPs Y203, I204, N347, Y358, and E362 are localized in or near the PfRH5–basigin interface. Not visible in this orientation is lab strain polymorphism K429. (C) Highlighted are basigin residues F27, Q100, and H102 (orange sticks). Mutation of F27 or Q100, or an insertion adjacent to H102 (+H103), affects the affinity for PfRH5 [242]. Also shown are two SNPs of PfRH5, namely N347 and I204 (pink sticks), found in the PfRH5:basigin binding interface and linked to the strain's ability to invade Aotus monkey erythrocytes [235].



FIGURE 5.10: The *Aotus* owl monkey. Taken from `http://www.monkeyland.co.za`.



D

	k_{on} $(M^{-1}s^{-1})$	k_{off} (s^{-1})	K_d (μM)
Wild-type	2100	0.0043	2.0
T28F	856	0.0102	12.0
G95F	1040	0.0030	3.0
H102A	3520	0.0063	1.8
V131D	1120	0.0028	2.5
S190D	1800	0.0029	1.6

FIGURE 5.11: Basigin mutant proteins retain PfRH5 binding ability. (A) Surface plasmon resonance analysis of the binding of basigin mutant proteins to PfRH5 bound to the chip surface. The kinetic on- and off-rates and overall dissociation constant (K_d) derived from SPR are given in (D). (B) Circular dichroism confirms that all of the basigin mutant proteins are properly folded, all manifesting the characteristic minimum at 217 nm for β-sheets. (C) The mutated residues are shown on the structure of basigin (blue) in complex with PfRH5 (grey surface). Residues in the basigin N-terminal immunoglobulin domain, linker, and C-terminal domain are shown in pink, orange, and green, respectively. T28F disrupts a strand-strand packing interaction, leading to a faster off-rate, and is the only mutation with a notable effect on the affinity. H102A in the linker actually appears to increase the on-rate; but the overall affinity remains similar to wild-type.

5.6 A possible supramolecular complex

The two PfRH5–basigin complexes in the asymmetric unit of the crystal pack together through a series of basigin-mediated contacts, including a ∼911 Å^2 interface between the two basigin C-terminal domains, which brings their C-termini into close proximity (Figure 5.12). As basigin is linked to the erythrocyte membrane through a C-terminal helix, it is tempting to speculate that this complex, containing two PfRH5 molecules and two basigin molecules, assembles during erythrocyte invasion, perhaps mediating a signalling event in either parasite or erythrocyte to trigger an essential downstream process. I therefore set out to analyse the stoichiometry of PfRH5–basigin by a number of methods, to see whether a 2:2 complex is formed.

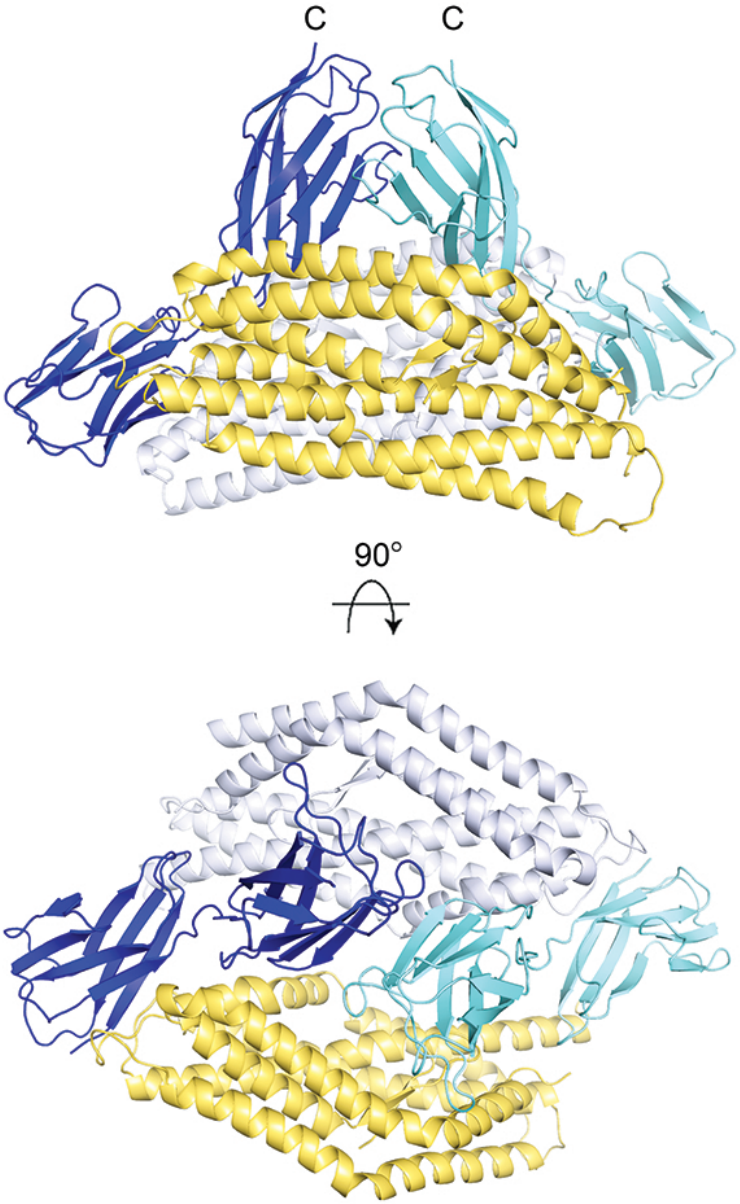


FIGURE 5.12: Arrangement of two PfRH5–basigin complexes in the asymmetric unit of the crystal. One complex, shown in yellow (PfRH5) and blue (basigin), interacts with the second, shown in silver (PfRH5) and cyan (basigin), primarily through packing between the two C-terminal domains of basigin. The two C-termini of basigin are in close proximity (top view).

5.7 Stoichiometry of the PfRH5–basigin complex

5.7.1 Native mass spectrometry

Accurate mass measurements of intact protein complexes can be obtained by native mass spectrometry using electrospray, a gentle ionisation technique. The PfRH5–basigin complex exhibited a molecular mass of ∼80 kDa by this technique, consistent with a 1:1

complex of ∼60-kDa PfRH5 and ∼20-kDa basigin (Figure 5.13C). No 2:2 complex was observed. Of unclear significance, basigin formed dimers, either alone or in the presence of PfRH5 (Figure 5.13A,C). This could be an artefact of the gas phase, though basigin dimers, formed through interactions between N-terminal domains, have been reported previously [239, 320].

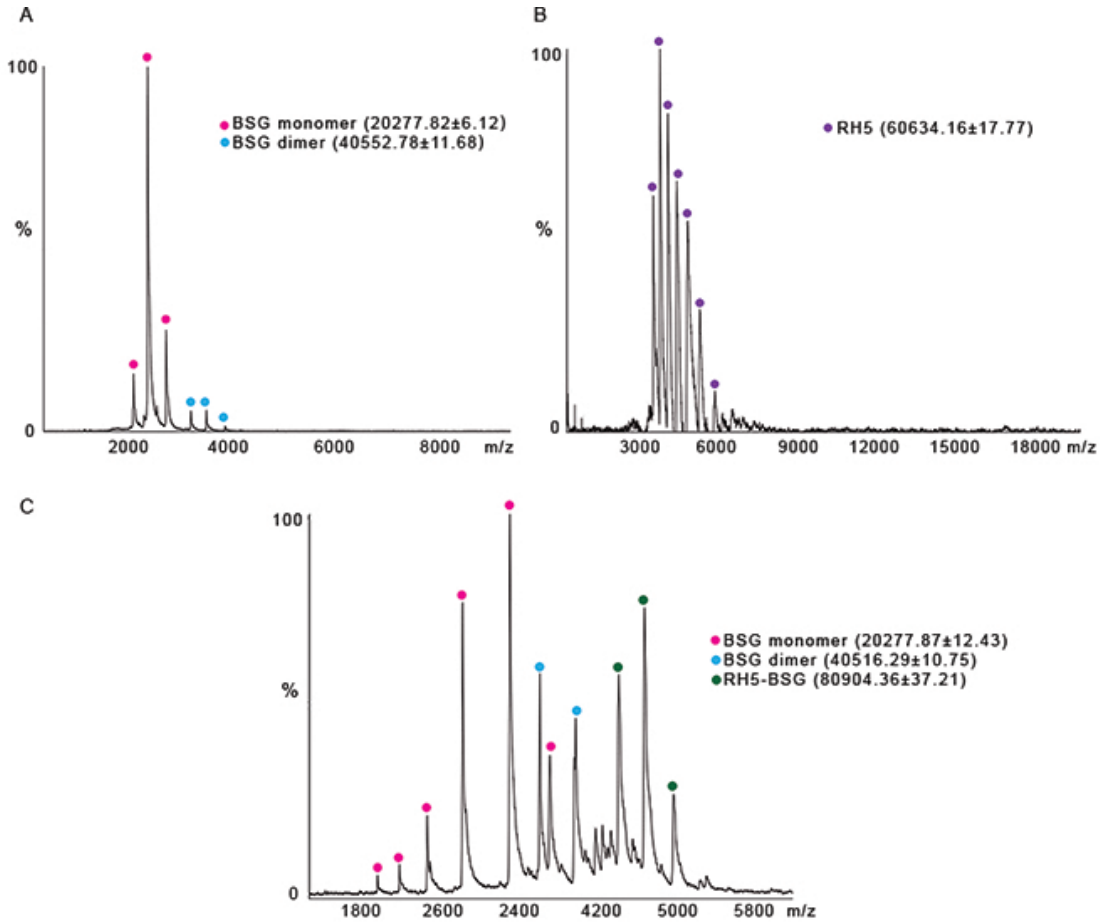
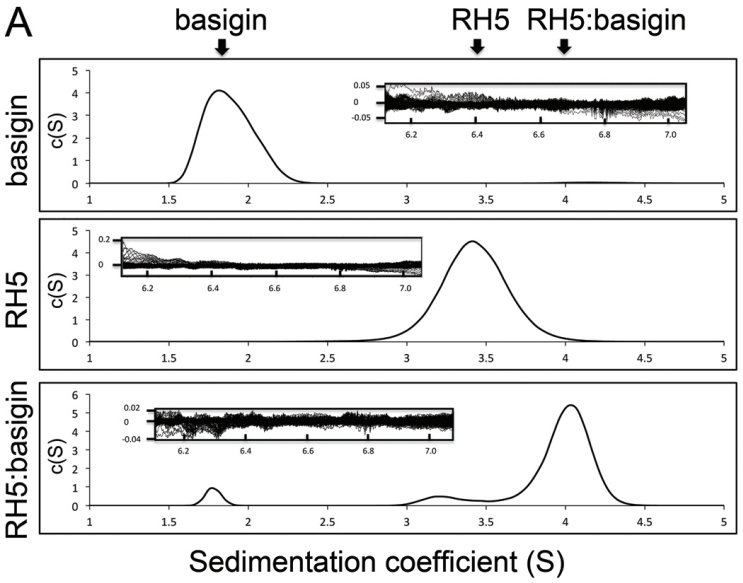


FIGURE 5.13: Native mass spectrometry analysis of (A) basigin (BSG), (B) PfRH5, and (C) a PfRH5–basigin complex. Note the presence of a basigin dimer, either in the presence or absence of PfRH5. No 2:2 complex of PfRH5–basigin is observed, however. All proteins were analysed in 200 mM ammonium acetate at a concentration between 1-10 μM.

5.7.2 Analytical ultracentrifugation

To test the PfRH5–basigin stoichiometry in solution, I performed analytical ultracentrifugation, both sedimentation velocity—to give an indication about molecular size and shape—and sedimentation equilibrium—to give an accurate estimate of the molecular

weight of the complex. Sedimentation velocity results are shown in Figure 5.14. The estimated molecular weight for PfRH5–basigin was again ∼80 kDa, indicative of a 1:1 complex. Sedimentation equilibrium was performed with shortened construct PfRH5ΔN, because degradation of full-length PfRH5 was observed during the prolonged course of the run. The results, like the velocity results, pointed to a 1:1 complex (Figure 5.15).



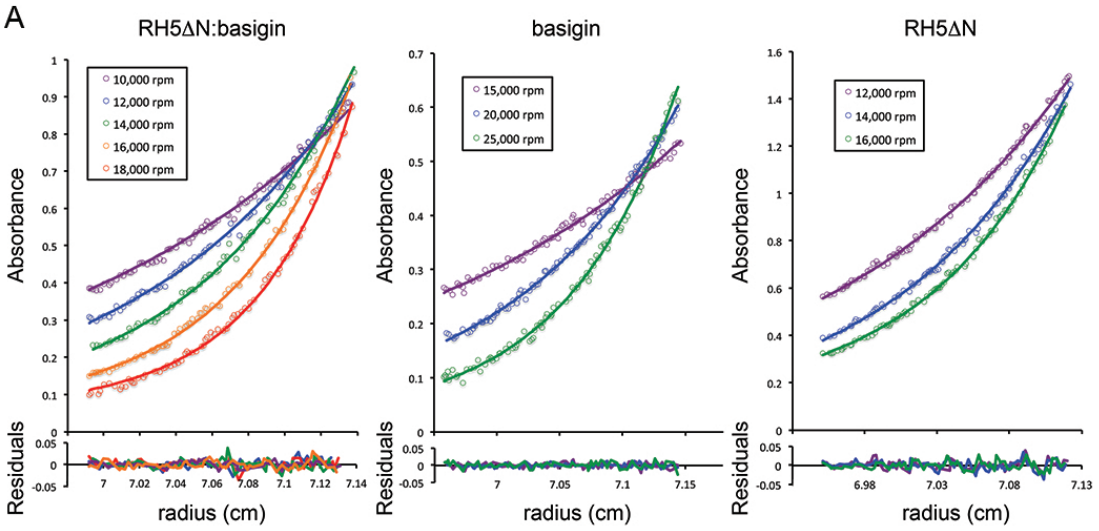
B

	Frictional coefficient ($\frac{f}{f0}$)	Sedimentation coefficient (S)	MW_{app} (kDa)
PfRH5	1.36	3.7	51.7
basigin	1.37	2.0	21.5
PfRH5:basigin	1.58	4.3	80.2

FIGURE 5.14: Sedimentation velocity analysis of the PfRH5–basigin complex. (A) The continuous sedimentation coefficient distributions that best fit the data are shown for basigin (top), full-length PfRH5 (middle), and a gel filtered PfRH5:basigin complex (bottom). The inset shows the fitting residuals. (B) A summary of the sedimentation coefficients, frictional coefficients, and apparent molecular weights of PfRH5, basigin, and the complex. Worth mentioning, the frictional ratios of each of PfRH5 and basigin (∼1.36-1.37) indicate reasonably globular proteins, only very marginally elongated, which is consistent with their crystal structures. The PfRH5–basigin complex is more markedly elongated (frictional ratio 1.58), as expected given that basigin stacks on top of PfRH5.

5.7.3 SAXS

The SAXS data presented in Section 5.4.1 could also be used to obtain a rough estimate of the molecular weight, by dividing the Porod volume by 1.7. For PfRH5–basigin, this suggested, once again, a 1:1 complex (Figure 5.7D). Furthermore, the resultant molecular envelope can only accommodate a single PfRH5–basigin complex. Thus, it



B

	$\mathbf{MW}_{app}$ **(kDa)**	**Reduced** χ^2
basigin	21.0	1.52
PfRH5ΔN	51.4	4.68
PfRH5ΔN–basigin	70.2	5.11

FIGURE 5.15: Sedimentation equilibrium AUC analysis of PfRH5ΔN–basigin. PfRH5ΔN (residues 140-526), basigin, and a gel filtered PfRH5ΔN–basigin complex were analysed. The runs lasted 20 hours at different speeds, as indicated in the inset legends. UV absorbance was monitored at 280 nm. The residuals are shown below fitted data. The calculated molecular weights are consistent with the formation of a 1:1 complex between PfRH5ΔN and basigin. The sedimentation equilibrium data were fit to an ideal monodisperse model using SEDPHAT. The reduced χ^2 value for the fit of the model is shown in the rightmost column. The calculated molecular weights are consistent with the formation of a 1:1 complex between PfRH5ΔN and basigin.

appears that PfRH5 and basigin do not form a 2:2 complex at the low concentrations (less than 24 μM) tested, in either solution or in the gas phase.

5.8 Discussion

PfRH5ΔNL–basigin crystals were, in general, unattractive by eye and disappointing by X-ray. One native crystal fortunately diffracted to 3.1 Å, but the odd conditions of its growth—germinated as an off-shoot from a block crystal—made it impossible to replicate. The low proportion of good diffractors, when combinatorially coupled to the vast range of possible phasing atoms, made the task of derivatisation for experimental phasing statistically absurd. The poor diffraction of these crystals probably stems from disorder and mosaicity. Indeed, variable crystal packing could have resulted from the

near-clash of the C-terminal ends of neighbouring PfRH5 molecules, and might have been exacerbated by different C-termini resulting from GluC cleavage. Moreover, some of the crystals were clearly multiple crystals fused laterally, and could not be teased apart.

Fortunately, the solution of the PfRH5ΔNL structure (Chapter 4) allowed the PfRH5ΔNL–basigin structure to be solved by molecular replacement. Basigin wraps around the tip of PfRH5, making contact along the length of its immunoglobulin domains and the linker joining them. Particularly in the C-terminal domain, only very few contacts (either hydrogen bonds, or hydrophobic interactions) link the two molecules, consistent with a low (μM) affinity. The C-terminal domain, furthermore, appears to interact flexibly with PfRH5. Basigin participates in the interaction almost entirely through its backbone, and mutation of five basigin residues at the interface—to disrupt individual interactions, primarily sterically—did not abolish binding. Taken together, these data suggest that basigin cannot mutate to escape PfRH5 binding, and is inevitably suborned for downstream invasion.

One of the main reasons that PfRH5 is such an excellent vaccine candidate is its lack of polymorphism, especially relative to other *Plasmodium* surface proteins [126, 128, 129]. The few polymorphisms found in the wild are evenly distributed across the structure, not enriched in the basigin binding site; critically, those in the binding site do not affect basigin binding. The structure confirms that a PfRH5 vaccine should be capable of inducing an immune response that is SNP- and strain-independent.

The functional consequence of PfRH5–basigin binding is still unresolved. PfRH5ΔNL and basigin are provocatively arranged in the crystal asymmetric unit as a 2:2 complex, bringing the membrane-proximal C-terminal domains of basigin together. Could this be a functional oligomerisation, dimerising the cytoplasmic domains of basigin to incite a signalling cascade? Or, conversely, does this bring together parasite binding partners of PfRH5 to trigger a retrograde signal into the merozoite? The organisation of the 2:2 complex would leave one face of PfRH5 available for binding to PfRIPR and other, as yet unidentified, partners. Could such higher-order organisation be part of a larger clustering of invasion ligands with their red blood cell receptors? However, this wave of speculation breaks on the reality that solution methods provide no evidence of a 2:2 complex, at least at concentrations up to 24 μM. It is of interest that dimerisation is

believed to be important for the function of EBA-175, but EBA-175 exists primarily as a monomer in solution, according to analytical ultracentrifugation analysis [190]. Whether such a supramolecular PfRH5–basigin complex assembles at the high local concentration of the junction that forms during invasion remains a tantalising question for the future.

Chapter 6

Structural Basis for PfRH5 Inhibition by Monoclonal Antibodies

6.1 Summary

Monoclonal antibodies against PfRH5 have been classified into those that block invasion and parasite growth, and those that do not [237]. In this chapter, I first describe how the structure of PfRH5ΔNL bound to inhibitory antibody QA1 was solved using molecular replacement. Together with the structure of PfRH5ΔNL–9AD4, the solution of which was detailed in Chapter 4, and with SAXS data of PfRH5 bound to different inhibitory antibodies, I next investigate how distinct inhibitory antibodies bind to PfRH5. Then, as a counterpoint, I use SAXS to show how non-inhibitory antibodies interact with PfRH5. Altogether, I use this chapter to build a picture of PfRH5 as an immunogen, identifying the epitopes and surfaces that could be of great utility in designing a vaccine.

6.2 PfRH5–QA1 crystallisation, data collection, and data processing

6.2.1 Crystallisation

Broad screens to effect crystallisation of PfRH5ΔNL–QA1 proceeded much as they did for PfRH5ΔNL–9AD4 (Section 4.3.1). Initial clusters of crystals (reminiscent of feather-dusters) were formed after three days at 18°C, in a solution from the Morpheus crystallisation screen, containing a 0.12 M mix of 1,6-hexanediol, 1-butanol, 1,2-propanediol, 2-propanol, 1,4-butanediol, 1,3-propanediol; 0.1 M MES-imidazole (pH 6.5); 20% (w/v) PEG 550 MME; and 10% (w/v) PEG 20,000 (Figure 6.1A).

Optimisation with crystal seeds and the Silver Bullets additive screen (Section 4.3.1) lead to single, rod-like crystals (Figure 6.1B). The optimised crystals were cryoprotected in well solution with a total of 32% (w/v) PEG 550 MME and 16% (w/v) PEG 20,000 before cryocooling in liquid nitrogen.

6.2.2 Data collection

Sixteen PfRH5ΔNL–QA1 crystals were exposed on beamline I04 at Diamond Light Source. One of them diffracted to 3.1 Å. iMOSFLM suggested the P222 space group, and unit cell dimensions a = 64.972 Å, b = 137.299 Å, c = 227.539 Å. A full dataset was collected based on the EDNA suggestion (Table 6.1). One of the 1800 images is shown in Figure 6.2.

Beamline	I04
Wavelength	0.97949 Å
Aperture	50 μm
Omega Start	87°
Oscillation	0.1°
Exposure	0.1 s
Transmission	50%
No. of images	1800

Table 6.1: PfRH5ΔNL–QA1 data collection strategy.

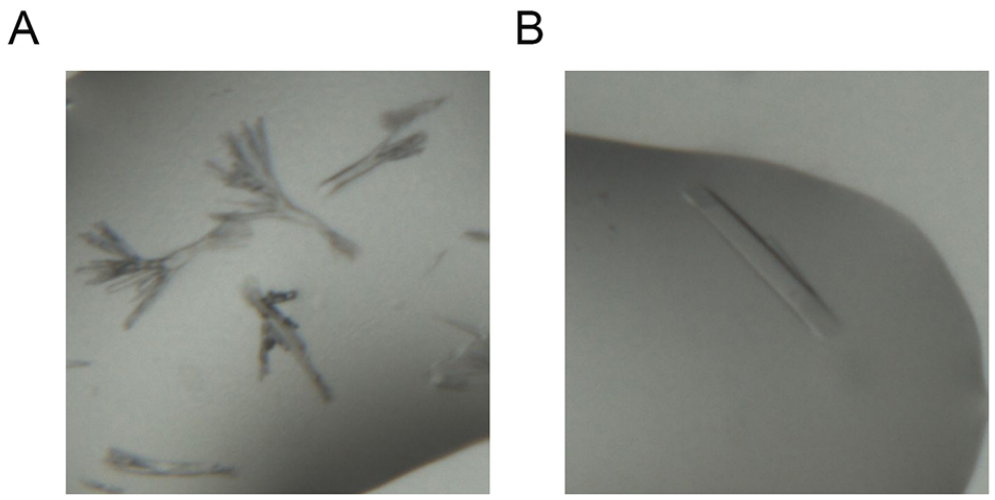


FIGURE 6.1: Crystals of PfRH5ΔNL–QA1 complex. (A) Initial crystals were grown in a 0.12 M mix of 1,6-hexanediol, 1-butanol, 1,2-propanediol, 2-propanol, 1,4-butanediol, 1,3-propanediol; 0.1 M MES-imidazole (pH 6.5); 20% (w/v) PEG 550 MME; and 10% (w/v) PEG 20,000 in a Morpheus screen at 18°C. The crystallisation drops were set up by mixing 100 nl of 8 mg/ml protein complex with 100 nl of well solution. Crystals appeared after three days. (B) Optimised crystals were grown with crystal seeds generated by harvesting the crystals in (A) and vortexing in the presence of a SeedBead (Hampton Research). The well solution was the same as in (A), with the addition of a Silver Bullet additive mixture (Hampton Research; in the well shown the additive solution contained 0.2% D-sorbitol, 0.2% glycerol, 0.2% glycine, 0.2% myo-inositol, 0.2% sarcosine, and 0.02 M sodium HEPES pH 6.8). Crystal drops were set up with 100 nl of 8 mg/ml complex, 50 nl of well solution, 50 nl of seeds, and 50 nl of Silver Bullets additives. The crystal (visible dimensions ∼375 μm x 50 μm) began to form after 3 days and was fully formed after 8 days.

6.2.3 Data processing

Data processing proceeded as described in Section 4.3.3; the statistics are summarised in Table 6.2. The refined unit cell parameters were a = 65.12 Å, b = 137.30 Å, c = 228.58 Å, with all angles 90°; images 1-1400 were integrated. Pointless identified the space group as $P22_12_1$ based on systematic absences. The resolution was 3.10 Å, with an $I/\sigma I$ ratio of 2.3 in the highest resolution shell. R_{pim} was 0.049 overall, and 0.3805 in the highest shell. The dataset was 95.5% complete, with a redundancy of 4.5. Finally, Matthews analysis indicated the presence of two complexes in the asymmetric unit, with 57.3% solvent.

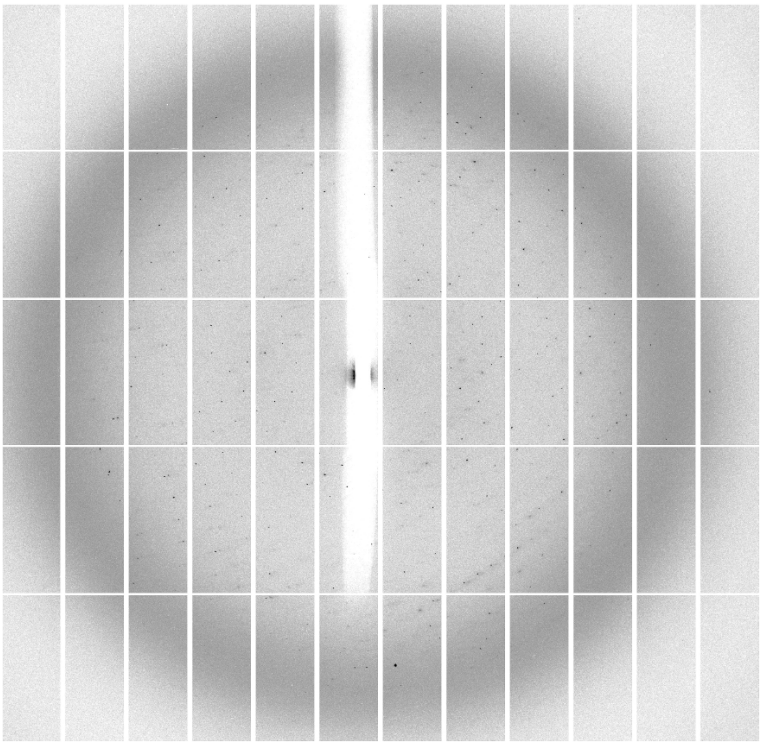


FIGURE 6.2: PfRH5ΔNL–QA1 diffraction image from beamline I04 at Diamond Light Source.

Space group	$P22_12_1$
Cell dimensions	
a, *b*, *c* (Å)	65.12, 137.30, 228.58
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	47.24–3.10 (3.27-3.10)
$\mathbf{R}_{merge}$	0.096 (0.773)
$\mathbf{R}_{rim}$	0.108 (0.866)
$\mathbf{R}_{pim}$	0.049 (0.3805)
$\mathbf{CC}_{1/2}$	0.996 (0.768)
$\mathbf{I}/\sigma\mathbf{I}$	9.2 (2.3)
Completeness (%)	95.5 (96.3)
Redundancy	4.5 (4.6)

TABLE 6.2: PfRH5ΔNL–QA1 data processing statistics.

6.3 Solving and validating the PfRH5–QA1 structure

The structure was solved using molecular replacement, with PfRH5ΔNL and an antibody fragment as search models. The antibody fragment model was prepared as a composite of the light chain of PDB entry 1I7Z and the heavy chain of 2ZN9, based on their similarity to the light and heavy chain of QA1, respectively, with hypervariable loops removed and any dissimilar residues mutated. Phaser was used to locate, first, two copies of PfRH5ΔNL, with a loop and the C-terminus removed (Section 5.3), and second, the

separated constant and variable regions of the antibody fragment. The antibodies were reassembled, and rigid body refinement gave R-factors below 50%.

Refinement and model building proceeded using Coot and BUSTER as described in previous chapters, and led to the final model shown in Figure 6.3, and the refinement statistics in Table 6.3. R_{work} and R_{free} were 23.6% and 28.1%, respectively. The B-factors for all atoms averaged 101.0 Å^2; the bond lengths (0.013 Å) and bond angles (1.9°) were indicative of a good model. PROCHECK confirmed that 86% of residues were in the most favourable regions, 13.3% in the additional allowed regions, and 0.7% in the generously allowed regions (Figure 6.4).

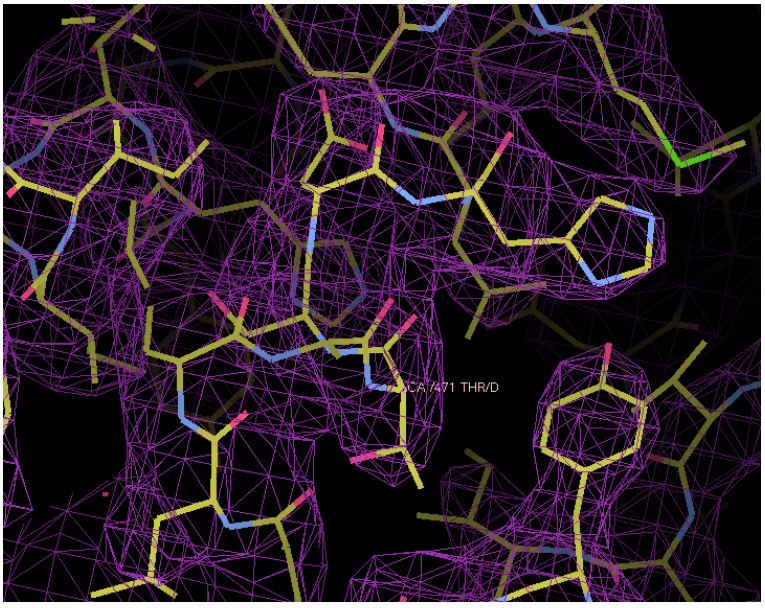
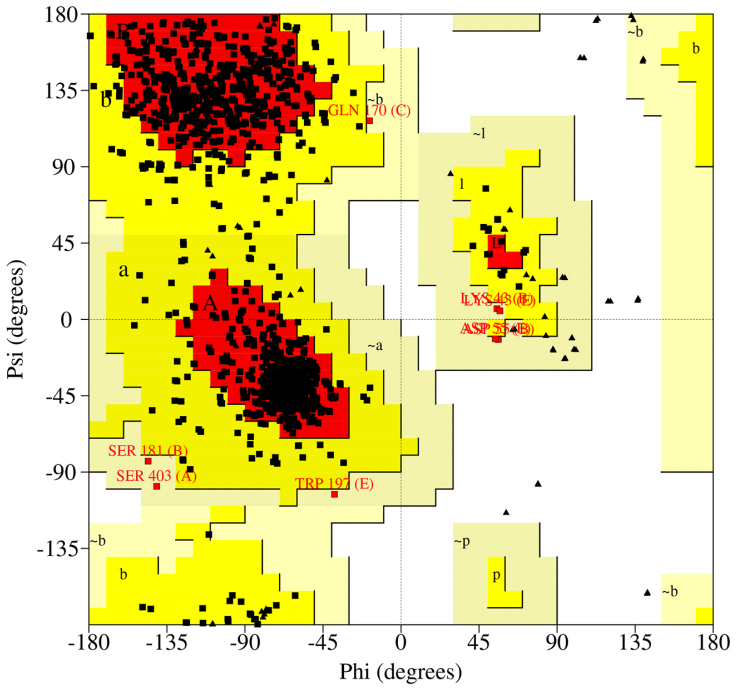


FIGURE 6.3: Final PfRH5ΔNL–QA1 electron density map ($2F_o$ - F_c, contour 1σ), superimposed with the final PfRH5ΔNL–QA1 structure. A helix of PfRH5 is shown. Atoms are coloured yellow (carbon), blue (nitrogen), or pink (oxygen). The figure was prepared using Coot [303].

Resolution (Å)	47.24-3.10
No. reflections (R_{free})	36259 (1793)
R_{work}/ R_{free} (%)	23.6 / 28.1
No. atoms	11478
Average B-factors (Å^2)	101.0
R.m.s deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.9

TABLE 6.3: PfRH5ΔNL–QA1 refinement statistics.



	No. Residues	Percentage
Most favoured regions	1112	86.0
Additional allowed regions	172	13.3
Generously allowed regions	8	0.7
Disallowed regions	0	0
Residues not Gly or Pro	1292	
End-residues (excl. Gly or Pro)	17	
Gly residues	72	
Pro residues	52	
Total	1433	

FIGURE 6.4: Ramachandran plot and statistics of PfRH5ΔNL–QA1 structure, from PROCHECK [307].

6.4 The structure of PfRH5 bound to inhibitory antibodies

In the manner described in the preceding section and in Chapter 4, the crystal structures of PfRH5ΔNL bound to two different invasion-inhibitory antibody fragments, 9AD4 and QA1, were solved. The crystal structures were analysed together with SAXS data of full-length PfRH5 bound to three different inhibitory antibody fragments, to begin to map the regions of PfRH5 that can be targeted in vaccine development.

6.4.1 PfRH5–QA1

Monoclonal antibody QA1 has been shown to directly block basigin binding by an *in vitro* protein–protein interaction assay, known as the AVidity-based EXtracellular Interaction Screen (AVEXIS), in addition to preventing parasite growth *in vitro* [237]. The crystal structure shows that QA1 binds at the tip of PfRH5, projecting along the PfRH5 long axis (Figure 6.5A, left). Its hypervariable loops make contact to PfRH5 through a loop (residues 346-352) and the tips of helices 2 and 6 (residue 198 and 452) (Figure 6.5B, left). Notably, this binding site partially overlaps that of the basigin N-terminal domain, explaining the previous AVEXIS result [237]. The interactions between QA1 and PfRH5ΔNL are summarised in Table 6.4.

The binding of QA1 to full-length PfRH5 was studied using SAXS (Figure 6.6). The resultant molecular envelope is consistent with the crystal structure, with the characteristic extra density corresponding to the truncated regions of PfRH5ΔNL. The radius of gyration (R_g) and the maximum particle dimension (D_{max}) are markedly increased compared to PfRH5 alone, consistent with the elongated architecture of the complex.

PfRH5			**QA1**			
Chain	**Residue**	**Group**	**Chain**	**Residue**	**Group**	**Type of interaction**
			Heavy Chain:			
A / D	K196	side chain	B / E	N101	side chain CO	hydrogen bond
A / D	K196	side chain	B / E	D31	backbone CO	hydrogen bond
A / D	S197	backbone NH	B / E	D53	side chain	hydrogen bond
A / D	Y346	side chain	B / E	Y33	side chain	pi stacking
A / D	N348	side chain CO	B / E	G105	backbone NH	hydrogen bond
A / D	N352	side chain	B / E	Y33	side chain	hydrogen bond
A / D	N352	side chain	B / E	Y59	side chain	hydrogen bond
A / D	N354	backbone NH	B / E	Y57	side chain	hydrogen bond
A / D	K452	side chain	B / E	D104	side chain	hydrogen bond
			Light Chain:			
A / D	N347	side chain NH_2	C / F	S95	backbone CO	hydrogen bond
A / D	N348	backbone NH	C / F	Y36	side chain	hydrogen bond
A / D	N349	backbone NH	C / F	Y34	side chain	hydrogen bond
A / D	N349	side chain CO	C / F	W96	side chain	hydrogen bond
A / D	F350	side chain	C / F		hydrophobic surface	hydrophobic
A / D	Q451	backbone CO	C / F	Y34	side chain	hydrogen bond

Table 6.4: Table describing interactions between PfRH5 and QA1.

6.4.2 PfRH5–9AD4

Antibody 9AD4 does not impede basigin binding by AVEXIS [237]. Nevertheless, 9AD4 is one of the most potent antibodies currently available for inhibiting parasite growth:

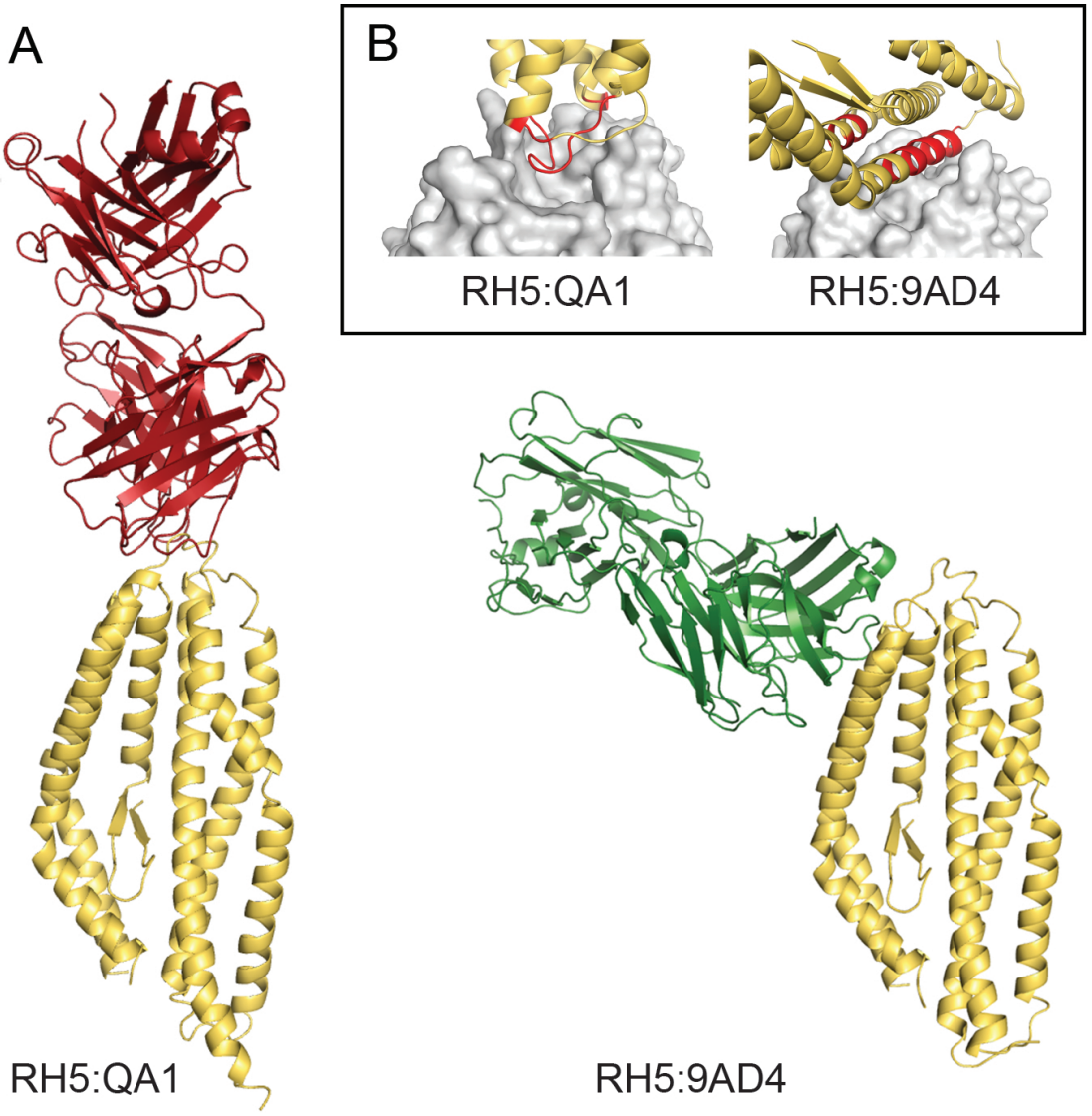


FIGURE 6.5: Crystal structures of PfRH5ΔNL (yellow) bound to inhibitory antibody fragments QA1 (red; left) and 9AD4 (green; right). Close-up views of the PfRH5 epitopes (red on yellow PfRH5) are shown in the inset box, with the Fab fragment in each case shown as a grey surface.

for example, by comparison to QA1, which achieved only 38% growth inhibitory activity (GIA) at a concentration of 500 μg/ml, 9AD4 achieved 50% GIA at only 62 μg/ml. The crystal structure of PfRH5ΔNL–9AD4 reveals that this antibody fragment interacts with PfRH5 helices 2 (residues 209-213) and 3 (residues 331-341), binding close to the tip of PfRH5 on its lateral edge (Figure 6.5A-B, right). The specific interactions involved are listed in Table 6.5. Superposition of the PfRH5–basigin structure onto that of PfRH5–9AD4 reveals that, although the C-terminal domain of basigin also interacts with helix 2, the 9AD4 binding site does not overlap with that of basigin (Figure 6.7A), in keeping

with the AVEXIS data. However, it may be that the close proximity of the 9AD4 binding site to that of basigin leads to a steric block by the whole IgG molecule, that prevents binding when PfRH5 is immobilised on the parasite surface and basigin is tethered on the erythrocyte membrane.

A previous study attempted to map the epitopes of anti-PfRH5 monoclonal antibodies by testing their binding to overlapping, linear peptides [237]. Some antibodies, such as QA1, did not bind to any of these peptides, strongly suggesting conformational and/or discontinuous epitopes. 9AD4 appeared to bind to a peptide consisting of residues 353-361. The crystal structure reveals these residues to be located at the top of helix 4, on a completely different face of the kite from the 9AD4 binding site. Thus, the peptide previously identified was not, in fact, part of the binding site, underscoring the danger of linear epitope mapping experiments and the importance of three-dimensional structural information to confidently identify epitopes.

SAXS analysis of 9AD4 together with full-length PfRH5 generates an envelope into which the crystal structure snugly docks (Figure 6.6). The R_g and D_{max} values indicate a complex elongated relative to free PfRH5, but not as extended as the PfRH5–QA1 structure, much as expected based on their relative shapes.

PfRH5			**9AD4**			
Chain	**Residue**	**Group**	**Chain**	**Residue**	**Group**	**Type of interaction**
			Heavy Chain:			
A	A205	backbone CO	B	Y103	backbone CO	hydrogen bond (via H_2O)
A	F209	side chain	B		hydrophobic surface	hydrophobic
A	Y335	side chain	B	Y103	backbone CO	hydrogen bond
A	N338	side chain CO	B	W107	side chain	hydrogen bond
A	L339	side chain	B		hydrophobic surface	hydrophobic
A	E341	side chain	B	S52	side chain	hydrogen bond
A	E341	side chain	B	M54	backbone CO	hydrogen bond
A	E341	side chain	B	A55	backbone CO	hydrogen bond
A	E341	side chain	B	Y56	backbone CO	hydrogen bond
A	Q342	side chain CO	B	N53	side chain NH_2	hydrogen bond
A		hydrophobic pocket	B	F101	side chain	hydrophobic
			Light Chain:			
A	K212	side chain	C	Y32	backbone CO	hydrogen bond
A	D331	side chain	C	Y32	side chain	hydrogen bond
A	N334	side chain NH_2	C	Y31	side chain	hydrogen bond (via H_2O)

TABLE 6.5: Table describing interactions between PfRH5 and 9AD4.

6.4.3 PfRH5–QA5

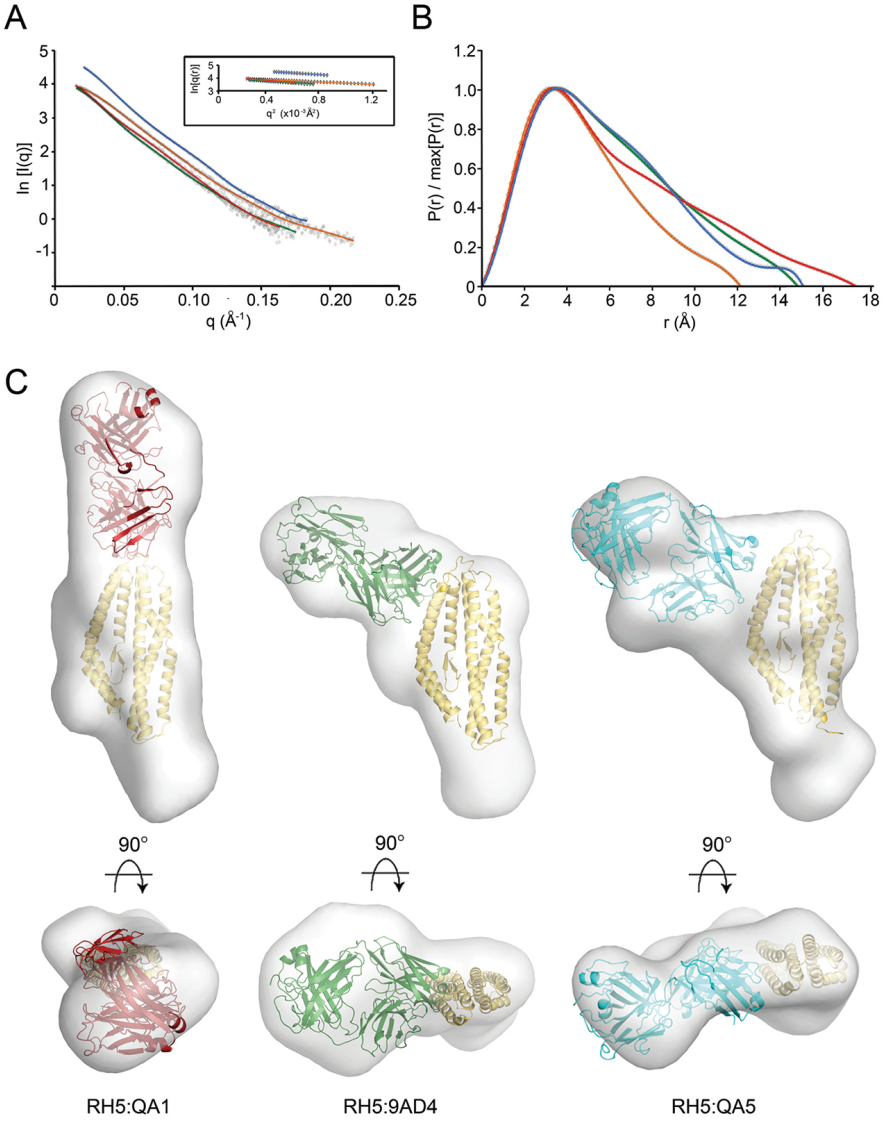
A third antibody, QA5, is known to block both basigin binding by AVEXIS and parasite growth by GIA [237]. A previous study suggested QA5 to interact with a linear

PfRH5 peptide comprised of residues 201-213 from helix 2 [237]. Attempts to crystallise the PfRH5ΔNL–QA5 complex were unsuccessful. I therefore docked the structure of PfRH5ΔNL and an antibody fragment into a SAXS-derived molecular envelope of the PfRH5–QA5 complex (Figure 6.6). On this occasion, the SAXS model was compatible with the linear epitope previously identified, which is located in the interface between PfRH5 and the antibody fragment (Figure 6.7B). Values of R_g and D_{max} for PfRH5–QA5 resemble those of the PfRH5–9AD4 complex, consistent with these two antibodies binding along the side of PfRH5 in a similar manner, and the complexes adopting a similar overall shape (Figure 6.6D).

The SAXS-based model proposes that QA5 binds in between 9AD4 and the C-terminal domain of basigin, and overlaps with the binding sites of both (Figure 6.7A-B). This is consistent with its ability to block basigin binding by AVEXIS; furthermore, the epitopes of QA5 and 9AD4 have been shown to overlap by a sandwich SPR-based assay [237].

6.4.4 A comparison of the binding of three inhibitory antibodies

Assembling together the data from the preceding sections, it is clear that inhibitory antibodies can bind to distinct sites on the PfRH5 molecule. They can block the binding of the N-terminal domain (e.g. QA1) or the C-terminal domain (e.g. QA5) of basigin, or can bind nearby without directly overlapping the basigin binding site (e.g. 9AD4) (Figure 6.7C). These regions of PfRH5 should all be included in any immunogen used to raise inhibitory antibody responses to block basigin binding and parasite invasion.



D

	R_g (nm)	D_{max} (nm)	Porod volume (nm^3)	MW_{app} (kDa)	NSD	χ
PfRH5	3.7	11.9	104.5	61.4	0.692±0.036	1.01
PfRH5–9AD4	4.6	14.6	154.5	90.9	0.759±0.55	0.95
PfRH5–QA1	4.9	17.2	152.4	89.6	0.789±0.032	0.97
PfRH5–QA5	4.4	14.8	157.9	92.9	0.643±0.023	0.89

FIGURE 6.6: SAXS analysis of the PfRH5–inhibitory antibody fragment complexes. (A) The theoretical scattering calculated from the average of 20 *ab initio* reconstructions (continuous lines, with PfRH5 in orange, PfRH5:9AD4 in green, PfRH5:QA1 in red, and PfRH5:QA5 in blue) plotted with the experimental scattering intensity curves (diamonds). The data are presented as the natural logarithm of the intensity. Guinier plots are displayed in the inset. (B) The distance distribution function, $P(r)$. (C) The crystal structures of PfRH5ΔNL–QA1(left) and PfRH5ΔNL–9AD4 (middle) were docked into the corresponding full-length PfRH5–antibody fragment envelopes (grey). PfRH5ΔNL is shown in yellow, QA1 in red, and 9AD4 in green. PfRH5ΔNL and an antibody fragment (shown in cyan) were manually docked into the PfRH5–QA5 SAXS envelope to generate a model of the PfRH5–QA5 structure (right). (D) Summary of SAXS parameters. The radius of gyration (R_g) was determined from the Guinier plot using AutoRg, and the maximum particle dimension (D_{max}) and the Porod volume were calculated using GNOM. An estimate of the molecular weight was obtained by dividing the Porod volume by 1.7. *Ab initio* modeling was used to generate 20 shape reconstructions from the data. The normalised spatial discrepancy parameter (NSD) diagnoses the similarity of these models. The models were averaged and the fit of the average model to the experimental data is indicated by the χ value. Situs and Sculptor were used to manually fit the crystal structures into the SAXS models.

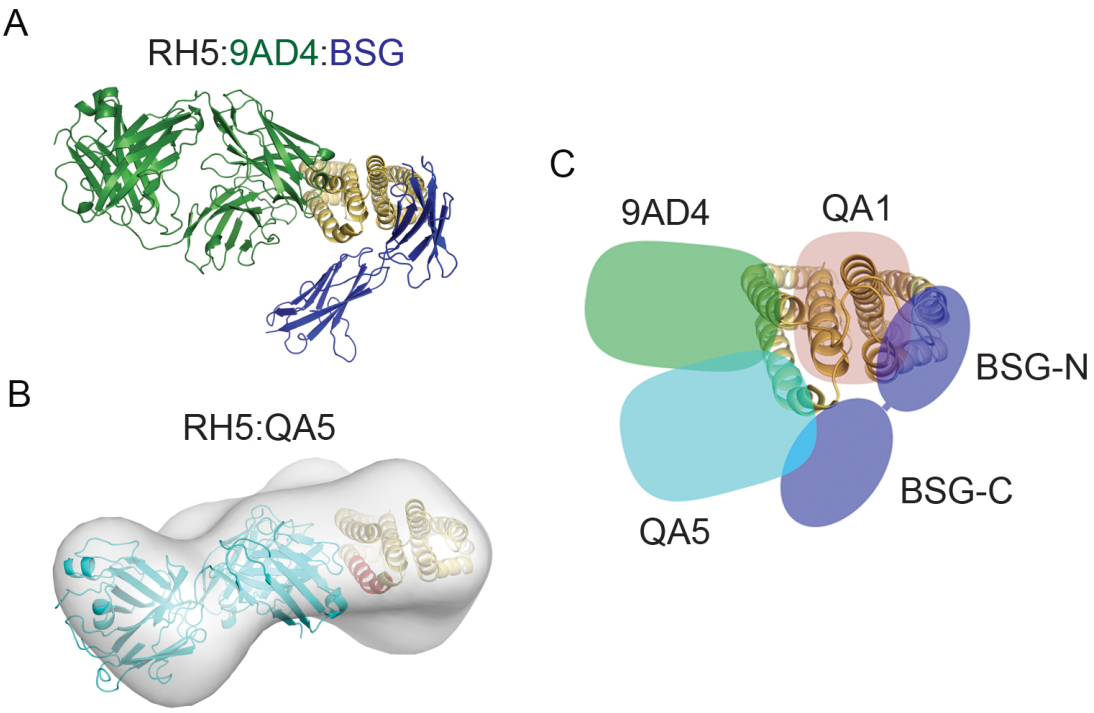


FIGURE 6.7: Comparison of PfRH5 binding sites for basigin and inhibitory antibodies. (A) Top view of PfRH5ΔNL–9AD4 crystal structure with superimposed basigin (blue), which was aligned based on the structure of PfRH5. (B) Top view of a model of the PfRH5–QA5 interaction, docked into a molecular envelope derived from SAXS. The putative QA5 epitope is highlighted in red [237]. (C) A schematic diagram showing the binding sites for the N- and C-terminal domains of basigin (BSG-N and BSG-C, respectively; blue), QA1 (red), and 9AD4 (green) and QA5 (cyan), on the structure of PfRH5ΔNL (yellow), showing QA1 and QA5 to overlap with the basigin binding site on PfRH5.

6.5 Analysis of PfRH5 interactions with non-inhibitory antibodies

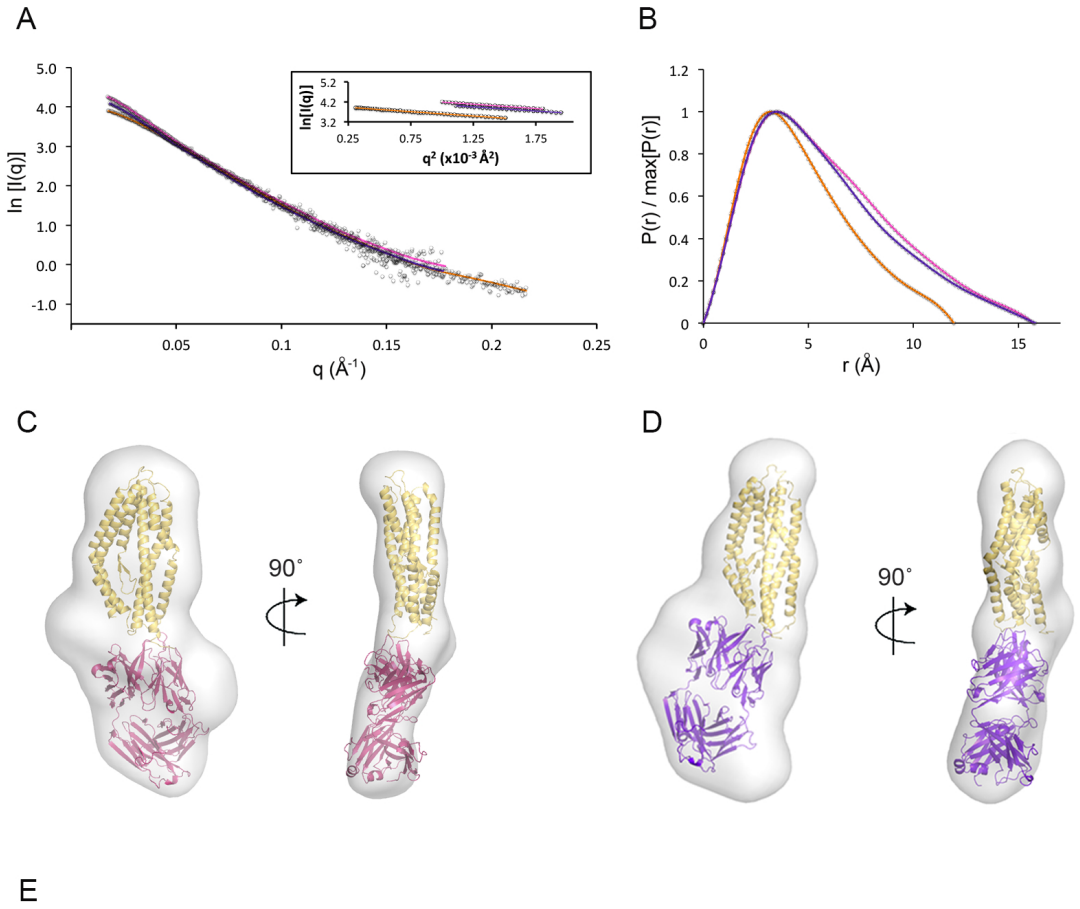
Other antibodies that target PfRH5 are not inhibitory of erythrocyte invasion. The epitope of RB3 was mapped to residues 50-78 [237], residues predicted to be disordered within the region truncated from PfRH5ΔN and PfRH5ΔNL (Figure 3.1). Antibody 4BA7 appears to bind residues 242-270 [237], in the disordered loop removed from PfRH5ΔNL. In these cases, given that these regions are both predicted to be disordered, it may be that linear peptide mapping is more reliable. Furthermore, 4BA7 did not bind to PfRH5ΔNL by ELISA (Section 3.5.3.1), consistent with the removal of its epitope from this construct. Oddly, RB3 did not bind to any of the PfRH5 constructs by ELISA, which might be a reflection of a preparation of protein with more degradation to the 45-kDa fragment, lacking the RB3 epitope (Chapter 3); it did, previously, bind weakly to full-length PfRH5 [237].

Thus, with reasonable confidence in location of their epitopes, I used SAXS analysis to explore the complexes of PfRH5 with 4BA7 and RB3 (Figure 6.8). In both cases, an increase in the Porod volume, and a consequent increase in the apparent molecular weight, relative to PfRH5 alone indicated binding of the antibody (Figure 6.8E). The binding site of RB3 therefore had not been wholly cleaved in the protein preparation used for SAXS, as perhaps occurred in the ELISA experiments. The crystal structure of PfRH5ΔNL, and that of a representative antibody fragment, were docked into SAXS-derived molecular envelopes.

The envelope for PfRH5–4BA7 is thin and elongated, indicating that the antibody is adding to one end of the PfRH5 structure. The structures can be docked such that 4BA7 binds to the bottom of PfRH5, near the C-terminus and the disordered loop containing its epitope (Figure 6.8C). In this conformation, PfRH5 fits the density very well; the extra density normally seen at its C-terminus is now occupied by the antibody. Some extra density to either side of the antibody suggests some flexibility in the interaction of 4BA7 with PfRH5, consistent with the loop to which it binds being flexible.

The PfRH5–RB3 molecular envelope is broadly similar to the PfRH5–4BA7 envelope. It is likewise elongated, suggesting antibody binding at one end of PfRH5 (Figure 6.8D). This lends credence to the notion that the extended N-terminus missing from the

PfRH5ΔNL crystal structure is actually located at the foot of PfRH5, near the C-terminus, even though it must eventually connect to the β-sheet structure at the centre of the kite. PfRH5ΔNL and an antibody fragment were correspondingly docked in this conformation. Like in the PfRH5–4BA7 SAXS model, there is extra density surrounding the antibody fragment, which might be due to flexibility of the PfRH5 N-terminus allowing the antibody to flop about like a hooked fish out of water. The SAXS analysis of the non-inhibitory antibodies is thus consistent with interactions at the far end of the PfRH5 molecule, distant from the basigin binding site.



E

	R_g (nm)	D_{max} (nm)	Porod volume (nm^3)	MW_{app} (kDa)	NSD	χ
Rh5	3.7	11.9	104.5	61.4	0.692±0.036	1.01
Rh5:4BA7	4.5	15.7	155.5	91.5	0.755±0.042	0.81
Rh5:RB3	4.4	15.8	146.3	86.1	0.696±0.045	0.99

FIGURE 6.8: SAXS analysis of PfRH5 complexed with non-inhibitory antibody fragments. (A) The theoretical scattering calculated from the average of 20 *ab initio* reconstructions (continuous lines, with PfRH5 in orange, PfRH5:4BA7 in pink, and PfRH5:RB3 in purple) plotted with the experimental scattering intensity curves (diamonds). The data are presented as the natural logarithm of the intensity. Guinier plots are displayed in the inset. (B) The distance distribution function, $P(r)$. (C) The crystal structure of PfRH5ΔNL (yellow) and an antibody fragment (pink) were docked into the PfRH5–4BA7 SAXS envelope to generate a model of the PfRH5–4BA7 structure. (D) PfRH5ΔNL with an antibody fragment (purple) docked into the PfRH5–RB3 SAXS envelope .(E) Summary of SAXS parameters. The radius of gyration (R_g) was determined from the Guinier plot using AutoRg, and the maximum particle dimension (D_{max}) and the Porod volume were calculated using GNOM. An estimate of the molecular weight was obtained by dividing the Porod volume by 1.7. *Ab initio* modeling was used to generate 20 shape reconstructions from the data. The normalised spatial discrepancy parameter (NSD) diagnoses the similarity of these models. The models were averaged and the fit of the average model to the experimental data is indicated by the χ value.

6.6 Discussion

In order to design a vaccine based upon PfRH5, it is important to identify the epitopes associated with inhibition of invasion and parasite growth. Several distinct monoclonal antibodies against PfRH5 were shown to inhibit invasion [237]. For some of these inhibitory antibodies, including 9AD4 and QA5, speculative epitopes were identified based upon binding to overlapping, linear PfRH5 peptides; others, including QA1, did not bind to any linear peptide. To more accurately pinpoint the epitopes recognised by these antibodies, I first ascertained that the crystallisation construct PfRH5ΔNL included these epitopes (Chapter 3), and then solved the crystal structure of PfRH5ΔNL bound to antibody fragments 9AD4 (Chapter 4) and QA1 (this chapter).

The epitope of antibody QA1 was discontinuous, consisting of a loop and the ends of two helices at the tip of the PfRH5 kite, explaining why no linear epitope could mimic it. 9AD4 bound to two helices along the edge of PfRH5; however, neither helix contained the previously identified epitope, which was located in a more distant helix. It is not clear why 9AD4 interacted with this peptide *in vitro*, but not the peptides corresponding to its binding site. This emphasises that antibody binding sites on a folded protein are generally three-dimensional surfaces, and cannot be adequately re-created by peptides alone.

Nonetheless, a SAXS-derived model of the PfRH5–QA5 structure is consistent with the peptide epitope previously identified. A co-crystal structure of PfRH5ΔNL bound to QA5 would provide further confirmatory evidence, but initial efforts to crystallise this complex were not successful. In the absence of a structure, hydrogen-deuterium exchange mass spectrometry could be used to confirm the QA5 epitope. In this technique, individual proteins and the complex between them are allowed to exchange with deuterated solvent for periods of time ranging from several minutes to several hours, cleaved into peptides and run on a mass spectrometer. Peptides that are 'protected' from solvent exchange in the complex (relative to the individual components) are probably located in the binding site.

9AD4, QA1, and QA5 thus bind to three different regions of PfRH5. The QA1 binding site overlaps with that of the N-terminal domain of basigin, and the QA5 binding site overlaps with that of the C-terminal domain of basigin, corroborating previous ELISA

data that suggest these antibodies directly impede basigin binding [237]. While the 9AD4 binding site does not directly overlap that of the basigin C-terminal domain, it is nearby, implying that the antibody probably sterically interferes with PfRH5 binding to basigin on the surface of the red blood cell. Overall, a large surface on the upper half of the PfRH5 kite has been identified that can elicit inhibitory antibodies.

The classification of PfRH5 epitopes of other monoclonal antibodies might further extend this surface. According to a sandwich competition SPR assay, the 2AC7 epitope overlaps with the QA5 and possibly QA1 epitopes (Figure 1.12), suggesting a location intermediate between the binding sites of the two basigin domains, but—according to ELISA data—not overlapping either of them [237]. The 6BF10 antibody binds an epitope overlapping with the QA1 and 9AD4 binding sites, and furthermore blocks basigin binding by ELISA, and thus may bind near the N-terminal binding site [237]. The 2AC7 and 6BF10 epitopes also overlap each other. These epitopes could be more explicitly mapped with crystallography or hydrogen-deuterium exchange mass spectrometry, to expand our understanding of inhibitory epitopes of PfRH5. (Initial attempts to crystallise 2AC7 with PfRH5ΔNL were, however, unsuccessful.)

Monoclonal antibodies RB3 and 4BA7 do not block basigin binding or inhibit invasion. These non-inhibitory antibodies bind to epitopes in the disordered region at the N-terminus, and in the disordered loop, respectively [237]. SAXS analysis suggests both of these disordered regions are located near the C-terminus, at the bottom of the PfRH5 kite. This is consistent with the extra density seen in the SAXS envelope of full-length PfRH5 relative to the crystal structure of PfRH5ΔNL (Section 4.5.2). Thus, these antibodies fail to inhibit invasion because they bind at the opposite end of the molecule from the basigin binding site. Furthermore, the binding site for antibody RB3 is at least partially cleaved off during invasion, when the 45 kDa fragment is formed [187, 202].

Chapter 7

Conclusions and Future Directions

Malaria caused by *Plasmodium falciparum* remains a major global killer, leading to an estimated 1 million deaths each year, primarily of African children under the age of five [6]. An effective vaccine, still lacking, is thus an urgent medical imperative. During its complicated lifecycle in both mosquitoes and humans, the parasite undergoes remarkable transformations to invade different cell types, including the human red blood cell. Invasion of red blood cells is a particularly attractive stage of the life cycle for vaccine intervention, since the parasite is transiently exposed to the host immune system—before it successfully penetrates and hides within the erythrocyte—and since this stage involves a profusion of parasite-specific proteins that are potential vaccine targets [169].

Central among these parasite proteins are members of the EBL and RH protein families, surface proteins that interact with cognate red blood cell receptors [185]. While they are important for red blood cell invasion, however, their function seems to be largely redundant. Unique among them, PfRH5 is the only member of either family that appears, through its interaction with erythrocyte surface protein basigin, to be essential for invasion [187, 238]. Antibodies against PfRH5 potently block parasite invasion and subsequent growth [127, 128, 237, 244]. PfRH5 is therefore an extremely promising vaccine candidate, and early-stage trials of its efficacy are underway.

During my D.Phil., I optimised recombinant expression of PfRH5, and solved the crystal structure of PfRH5 together with its erythrocyte receptor, basigin, and with two different

antibodies that prevent parasite invasion. I also identified the binding sites of further antibodies using SAXS. These studies have given a first insight into the RH protein fold, provided a molecular glimpse of the PfRH5–basigin interaction, and characterised PfRH5 inhibitory epitopes to aid future vaccine development.

7.1 The PfRH5 fold and an updated comparison of parasite surface proteins

PfRH5 forms a flat, kite-shaped, α-helical bundle not previously observed. This fold appears to be conserved in other RH proteins from a number of *Plasmodium* species, in spite of low sequence identity. Similar residues point inward to stabilise the fold, while surface-exposed residues are variable, allowing diversification of binding potential and function. Furthermore, threading models of the structures of other RH proteins suggest some diversity in surface features, loops, and helix disruptions (Chapter 4). In the future, it would be instructive to solve the structure of the PfRH5-like domain from other RH proteins, to verify these models and to begin to diagnose the extent of conservation and diversity in this protein fold. It would be interesting to determine whether this domain serves as the ligand-binding domain in all of the RH proteins, and if so, whether the same face of the domain is involved in ligand contacts in each case. RH proteins are generally much longer than PfRH5: further biophysical and structural studies could probe other regions of these proteins, to see if they have a distinct shape and function.

There are some superficial similarities between the RH domain fold and the DBL domains found in the other family of invasion proteins, the EBL family. Both domains are comprised of a helical scaffold, the architecture of which is stabilised by a few similar (and, for the DBLs, conserved) residues. To some degree both domains can be described by two separate helical bundles—which, in the case of PfRH5, might have arisen from a gene duplication event (Figure 7.1). The surface of these domains can be highly variable, liberally strewn with polymorphisms or variable loops. Together, these two protein families highlight the utility of a robust α-helical fold, which remains intact during diversification of surface-exposed residues.

The DBL domain is also found in other proteins outside of the EBL family, such as the PfEMP1s; it remains to be seen whether the RH fold is more widely exploited in the *Plasmodium* genus, or even outside of it. Bioinformatic searches are confounded somewhat by the low sequence identity (10-20%) among the RH proteins; but perhaps the elucidation of the structures of several family members will facilitate future searches.

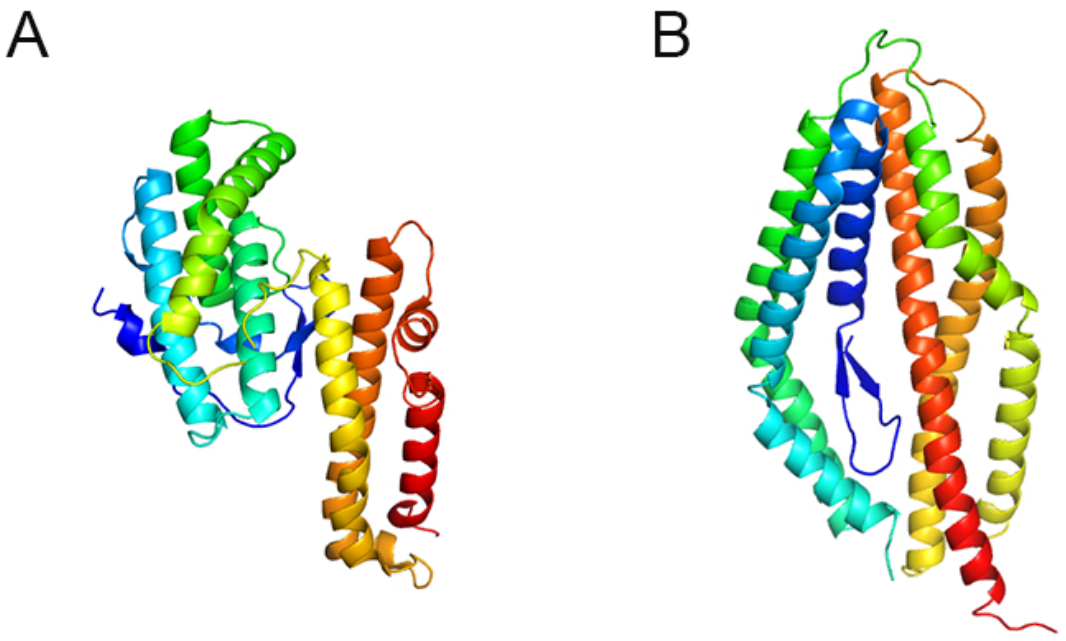


FIGURE 7.1: The DBL fold (the N-terminal DBL domain of EBA-175 [190]) (A) and the RH fold (B), coloured in rainbow from blue (N-terminus) to red (C-terminus).

7.2 The PfRH5–basigin interaction and implications for erythrocyte invasion

The co-crystal structure reveals that PfRH5 contacts its erythrocyte receptor, basigin, across its two immunoglobulin domains. The interaction is immutable, as far as basigin is concerned: PfRH5 contacts basigin almost exclusively through its protein backbone, and mutation of nearby side chains does not greatly affect binding (Chapter 5). As a consequence, it is highly unlikely that basigin 'escape mutants,' which have lost their ability to bind PfRH5, could emerge to protect the host from invasion and infection. Thus, the PfRH5–basigin interaction is not only essential for invasion; it is also seemingly inevitable.

The importance of the PfRH5–basigin interaction is well established, and now, the crystal structure has revealed its molecular basis. However, the function of PfRH5–basigin during invasion remains a mystery. How is it temporally and spatially regulated? What are the downstream effects of this binding event? What is the lifetime of the complex, and what happens after invasion is complete?

To begin with, there is some debate about where PfRH5 is localised within the rhoptry subcompartments, when it gets released, and when it binds basigin. What happens to the complex at the later stages of invasion, after invagination? Super-resolution microscopy using some of the monoclonal antibodies exploited in this thesis could go some way to answering these questions.

The PfRH5–basigin interaction has a dissociation constant in the 0.1-1 μM range [238], with rapid on- and off-rates. Similar binding characteristics are commonly observed for interactions involved in signalling processes, such as binding of a T-cell receptor to a peptide–MHC complex [321]. By contrast, adhesion events, such as those mediated by PfEMP1, typically have dissociation constants in the low nM range, and very slow off-rates. This suggests that the PfRH5–basigin interaction does not serve a purely adhesive purpose, and makes a signalling function at least plausible.

The crystal structure reveals an intriguing dimerisation interface between two C-terminal immunoglobulin domains of basigin, to generate a 2:2 complex of PfRH5–basigin, and to bring together the cytoplasmic tails of basigin. It remains to be determined whether this dimerisation indeed occurs during invasion, and if so, what its effect might be. Dimerisation could mediate a signalling event in the erythrocyte—signalling, for example, to the cytoskeleton, to mould the erythrocyte in readiness for invasion. Crystallographic and solution dimers of several of the EBL proteins have also been observed, indicating dimerisation might be a more general mechanism of action of invasion proteins [190, 194]. However, none of the putative dimers have been detected *in vivo* to date. Future experiments will need to address the existence and functional significance of PfRH5–basigin dimerisation.

PfRH5 is one member of a larger family of *P. falciparum* proteins involved in red blood cell binding, but is the only one of them demonstrably essential to invasion. This implies that its function must at least subtly differ from that of the other RH proteins. PfRH5 lacks the transmembrane helix and cytoplasmic domain of the others: it is interesting

to speculate that the other RH proteins function in retrograde signalling into the merozoite, while PfRH5 signals forward into the erythrocyte, perhaps through dimerisation of basigin. Of course, further experiments will be needed to reveal the truth.

If the function of PfRH5 is unique among the RH proteins, then why has it not been identified in other *Plasmodium* species, outside of the subgenus *Laverania*? Perhaps there are PfRH5 homologues in other species, which remain obscure due to low sequence identity; or perhaps other species have another protein which binds to basigin. It is also possible that other species do not rely upon PfRH5–basigin as the *Laverania* parasites do, and its function is not as unique in the family of RH proteins as its essentiality in *P. falciparum* implies.

How the EBL family functions relative to the RH family will also need to be investigated. There is evidence to suggest that members of these families have at least partially overlapping functions, and act at the same stage of invasion: the *PfRH4* gene is upregulated when *EBA-175* is genetically disrupted, and the invasion pathway switches from sialic acid-dependent (involving EBA-175) to sialic acid-independent (involving PfRH4) [188]. Members of both families bind their receptors with similar affinities ($0.1 - 1\mu M$ range). Though glycophorins are more abundant on the cell surface than complement receptor 1 and basigin, the number of interactions formed may not be a limiting factor in invasion. It would be interesting to compare the rate of invasion of parasites using EBA-175 and those using PfRH4, to assess the true extent of the redundancy between these two protein families.

Thus, overall, there is little evidence so far to determine or distinguish the functions of PfRH5, other RH proteins, and the EBL proteins, but this work fuels speculation and future experimentation on the subject. Several advances made in this thesis research will facilitate such experiments. An insect cell expression system has allowed expression of PfRH5 in large quantities, and this system can now be extended to express other PfRH proteins, such as PfRH2b (Chapter 4). Moreover, based on the PfRH5ΔNL–basigin structure, it is now possible to design mutants of PfRH5 that do not bind basigin. Transgenic parasites expressing non-binding PfRH5 could be used to determine the effect of disrupting the PfRH5–basigin interaction on erythrocyte invasion. In a similar manner, transgenic erythroid cells containing basigin mutated to disrupt PfRH5 binding,

or to disrupt the putative dimer interface, could be used to probe the function of PfRH5 binding and basigin dimerisation.

7.3 Implications for vaccine development

PfRH5 induces a potent invasion-inhibitory antibody response. To harness and optimise this response for vaccine development, it would be beneficial to design a shorter construct of PfRH5 which still contains inhibitory epitopes, in the hope of producing an immunogen that can be mass-produced at low cost. PfRH5ΔNL presented in this work is a first-generation truncated construct, designed to lack disordered regions. By analysing the structure of PfRH5 in complex with three different antibodies that inhibit invasion, it is possible to identify the surfaces of PfRH5 which are capable of inducing an inhibitory antibody response, and thus to design even more minimal constructs. The surface identified overlaps with the basigin binding site, at the top of the PfRH5 kite (Chapter 6). Both of the three-helical bundles of the PfRH5 kite contribute to this surface: perhaps, therefore, it would be possible to design optimised and stabilised versions of these three-helical bundles individually, to assess their immunogenicity and potency. Ideally, a smaller construct of PfRH5 would be cheaper and easier to produce, and the *E. coli* expression system could be revisited to express these three-helical bundles. It may also be possible to graft inhibitory epitopes onto a scaffold protein for *E. coli* expression [322]. In any case, the structures of PfRH5 bound to different inhibitory antibodies makes rational immunogen design a real possibility.

7.4 Summary of the findings of this thesis

PfRH5 is a parasite protein crucial for red blood cell invasion, and the onset of malaria disease symptoms. In this thesis, the structure of PfRH5 was elucidated, revealing a novel, kite-like fold that is predicted to be conserved across a family of parasite invasion proteins. The structure of PfRH5 together with its erythrocyte receptor, basigin, exposed a remarkable molecular binding site in which few basigin side chains participate, making mutation of basigin to escape invasion a virtual impossibility. Antibodies that block parasite invasion bind at or close to this site. Immunogens containing these regions of PfRH5 will be important components of a vaccine to prevent *P. falciparum* erythrocyte invasion, thereby crippling the parasite responsible for the deadliest form of human malaria.

Bibliography

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

[2] Laveran, A. *Un nouveau parasite trouvé dans le sang des malades atteints de fièvre palustre: origine parasitaire des accidents de l'impaludisme* (1881).

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

[4] Ross, R. *et al.* Memoirs: with a full account of the great malaria problem and its solution. *Memoirs: with a Full Account of the Great Malaria Problem and its Solution.* (1923).

[5] Grassi, G. B. *Studio di uno zoologo sulla malaria...*, vol. 3 (R. Accademia dei lincei, 1900).

[6] World Health Organisation. World malaria report 2013. *Geneva: WHO* (2013).

[7] Cox, F., Wernsdorfer, W., McGregor, I. *et al.* Major animal models in malaria research: rodent. *Malaria: Principles and Practice of Malariology. Volume 2.* 1503–1543 (1988).

[8] Craig, A. G. *et al.* The role of animal models for research on severe malaria. *PLoS Pathogens* **8**, e1002401 (2012).

[9] Langhorne, J. *et al.* The relevance of non-human primate and rodent malaria models for humans. *Malaria Journal* **10**, 23 (2011).

[10] Bogitsh, B. J., Carter, C. E. & Oeltmann, T. N. *Human Parasitology* (Academic Press, 2012).

[11] Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks: the Duffy-blood-group genotype, FyFy. *New England Journal of Medicine* **295**, 302–304 (1976).

[12] Price, R. N. *et al.* Vivax malaria: neglected and not benign. *The American Journal of Tropical Medicine and Hygiene* **77**, 79–87 (2007).

[13] Tjitra, E. *et al.* Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Medicine* **5**, e128 (2008).

[14] Genton, B. *et al.* *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. *PLoS Medicine* **5**, e127 (2008).

[15] Tan, S. O. *et al.* Thrombocytopaenia in pregnant women with malaria on the Thai-Burmese border. *Malaria Journal* **7**, 209 (2008).

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

[17] Trape, J.-F. *et al.* The Dielmo project: a longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of Senegal. *American Journal of Tropical Medicine and Hygiene* **51**, 123–137 (1994).

[18] Cox-Singh, J. *et al.* *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases* **46**, 165–171 (2008).

[19] Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002).

[20] Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763 (2008).

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

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

[23] Otto, T. D. *et al.* Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nature communications* **5** (2014).

[24] Winzeler, E. A. Malaria research in the post-genomic era. *Nature* **455**, 751–756 (2008).

[25] Sidjanski, S. & Vanderberg, J. P. Delayed migration of *Plasmodium* sporozoites from the mosquito bite site to the blood. *The American Journal of Tropical Medicine and Hygiene* **57**, 426–429 (1997).

[26] Vanderberg, J. P. & Frevert, U. Intravital microscopy demonstrating antibody-mediated immobilisation of *Plasmodium berghei* sporozoites injected into skin by mosquitoes. *International Journal for Parasitology* **34**, 991–996 (2004).

[27] Kappe, S. H., Buscaglia, C. A. & Nussenzweig, V. *Plasmodium* sporozoite molecular cell biology. *Annu. Rev. Cell Dev. Biol.* **20**, 29–59 (2004).

[28] Sultan, A. A. *et al.* TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. *Cell* **90**, 511–522 (1997).

[29] Amino, R. *et al.* Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nature Medicine* **12**, 220–224 (2006).

[30] Gueirard, P. *et al.* Development of the malaria parasite in the skin of the mammalian host. *Proceedings of the National Academy of Sciences* **107**, 18640–18645 (2010).

[31] Amino, R. *et al.* Imaging malaria sporozoites in the dermis of the mammalian host. *Nature Protocols* **2**, 1705–1712 (2007).

[32] Pradel, G., Garapaty, S. & Frevert, U. Proteoglycans mediate malaria sporozoite targeting to the liver. *Molecular Microbiology* **45**, 637–651 (2002).

[33] Ménard, R. *et al.* Looking under the skin: the first steps in malarial infection and immunity. *Nature Reviews Microbiology* **11**, 701–712 (2013).

[34] Mota, M. M. *et al.* Migration of *Plasmodium* sporozoites through cells before infection. *Science* **291**, 141–144 (2001).

[35] Carrolo, M. *et al.* Hepatocyte growth factor and its receptor are required for malaria infection. *Nature Medicine* **9**, 1363–1369 (2003).

[36] Mota, M. M., Hafalla, J. C. & Rodriguez, A. Migration through host cells activates *Plasmodium* sporozoites for infection. *Nature Medicine* **8**, 1318–1322 (2002).

[37] Coppi, A. *et al.* The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *The Journal of Experimental Medicine* **208**, 341–356 (2011).

[38] Sturm, A. *et al.* Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science* **313**, 1287–1290 (2006).

[39] Baer, K., Klotz, C., Kappe, S. H., Schnieder, T. & Frevert, U. Release of hepatic *Plasmodium yoelii* merozoites into the pulmonary microvasculature. *PLoS Pathogens* **3**, e171 (2007).

[40] Francis, S., Sullivan Jr, D. & Goldberg, D. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual Review of Microbiology* **51**, 97–123 (1996).

[41] 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).

[42] Abkarian, M., Massiera, G., Berry, L., Roques, M. & Braun-Breton, C. A novel mechanism for egress of malarial parasites from red blood cells. *Blood* **117**, 4118–4124 (2011).

[43] Dvorak, J. A., Miller, L. H., Whitehouse, W. C. & Shiroishi, T. Invasion of erythrocytes by malaria merozoites. *Science* **187**, 748–750 (1975).

[44] Glushakova, S., Yin, D., Li, T. & Zimmerberg, J. Membrane transformation during malaria parasite release from human red blood cells. *Current Biology* **15**, 1645–1650 (2005).

[45] Blackman, M. J. Malarial proteases and host cell egress: an 'emerging'cascade. *Cellular Microbiology* **10**, 1925–1934 (2008).

[46] Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).

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

[48] 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).

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

[50] Beck, J. R., Muralidharan, V., Oksman, A. & Goldberg, D. E. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. *Nature* **511**, 592–595 (2014).

[51] Elsworth, B. *et al.* PTEX is an essential nexus for protein export in malaria parasites. *Nature* **511**, 587–591 (2014).

[52] Elsworth, B., Crabb, B. S. & Gilson, P. R. Protein export in malaria parasites: an update. *Cellular Microbiology* **16**, 355–363 (2014).

[53] Maier, A. G., Cooke, B. M., Cowman, A. F. & Tilley, L. Malaria parasite proteins that remodel the host erythrocyte. *Nature Reviews Microbiology* **7**, 341–354 (2009).

[54] Nguitragool, W. *et al.* Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665–677 (2011).

[55] Silva, M. D. *et al.* A role for the *Plasmodium falciparum* RESA protein in resistance against heat shock demonstrated using gene disruption. *Molecular Microbiology* **56**, 990–1003 (2005).

[56] Baruch, D. I. Adhesive receptors on malaria-parasitized red cells. *Best Practice & Research Clinical Haematology* **12**, 747–761 (1999).

[57] Newbold, C. *et al.* Cytoadherence, pathogenesis and the infected red cell surface in *Plasmodium falciparum*. *International Journal for Parasitology* **29**, 927–937 (1999).

[58] Langreth, S. G. & Peterson, E. Pathogenicity, stability, and immunogenicity of a knobless clone of *Plasmodium falciparum* in colombian owl monkeys. *Infection and Immunity* **47**, 760–766 (1985).

[59] Beeson, J. G. *et al.* *Plasmodium falciparum* isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties. *Journal of Infectious Diseases* **180**, 464–472 (1999).

[60] Newbold, C. *et al.* Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. *The American Journal of Tropical Medicine and Hygiene* **57**, 389–398 (1997).

[61] Baruch, D. I. *et al.* Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).

[62] Smith, J. D. *et al.* Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).

[63] 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).

[64] Miller, L. H., Baruch, D. I., Marsh, K. & Doumbo, O. K. The pathogenic basis of malaria. *Nature* **415**, 673–679 (2002).

[65] Salanti, A. *et al.* Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *The Journal of Experimental Medicine* **200**, 1197–1203 (2004).

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

[67] Talman, A. M., Domarle, O., McKenzie, F. E., Ariey, F. & Robert, V. Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malaria Journal* **3**, 24 (2004).

[68] Painter, H. J., Campbell, T. L. & Llinás, M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Molecular and Biochemical Parasitology* **176**, 1–7 (2011).

[69] Bruce, M., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100**, 191–200 (1990).

[70] Silvestrini, F., Alano, P. & Williams, J. Commitment to the production of male and female gametocytes in the human malaria parasite *Plasmodium falciparum*. *Parasitology* **121**, 465–471 (2000).

[71] Smalley, M., Abdalla, S. & Brown, J. The distribution of *Plasmodium falciparum* in the peripheral blood and bone marrow of Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **75**, 103–105 (1981).

[72] 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).

[73] Billker, O. *et al.* Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* **392**, 289–292 (1998).

[74] Vlachou, D., Schlegelmilch, T., Runn, E., Mendes, A. & Kafatos, F. C. The developmental migration of *Plasmodium* in mosquitoes. *Current Opinion in Genetics & Development* **16**, 384–391 (2006).

[75] Vlachou, D. *et al.* Real-time, *in vivo* analysis of malaria ookinete locomotion and mosquito midgut invasion. *Cellular Microbiology* **6**, 671–685 (2004).

[76] Kariu, T., Ishino, T., Yano, K., Chinzei, Y. & Yuda, M. CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. *Molecular Microbiology* **59**, 1369–1379 (2006).

[77] Dessens, J. T. *et al.* CTRP is essential for mosquito infection by malaria ookinetes. *The EMBO Journal* **18**, 6221–6227 (1999).

[78] Kadota, K., Ishino, T., Matsuyama, T., Chinzei, Y. & Yuda, M. Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16310–16315 (2004).

[79] Tomas, A. M. *et al.* P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. *The EMBO Journal* **20**, 3975–3983 (2001).

[80] Vlachou, D. *et al.* *Anopheles gambiae* laminin interacts with the P25 surface protein of *Plasmodium berghei* ookinetes. *Molecular and Biochemical Parasitology* **112**, 229–237 (2001).

[81] Aly, A. S. & Matuschewski, K. A malarial cysteine protease is necessary for *Plasmodium* sporozoite egress from oocysts. *The Journal of Experimental Medicine* **202**, 225–230 (2005).

[82] Akaki, M. & Dvorak, J. A. A chemotactic response facilitates mosquito salivary gland infection by malaria sporozoites. *Journal of Experimental Biology* **208**, 3211–3218 (2005).

[83] Rodriguez, M. H. & Hernández-Hernández, F. d. l. C. Insect–malaria parasites interactions: the salivary gland. *Insect Biochemistry and Molecular Biology* **34**, 615–624 (2004).

[84] Rowe, J. A., Claessens, A., Corrigan, R. A. & Arman, M. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert Reviews in Molecular Medicine* **11**, e16 (2009).

[85] Rogerson, S. J., Hviid, L., Duffy, P. E., Leke, R. F. & Taylor, D. W. Malaria in pregnancy: pathogenesis and immunity. *The Lancet Infectious Diseases* **7**, 105–117 (2007).

[86] 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).

[87] Schofield, L. Intravascular infiltrates and organ-specific inflammation in malaria pathogenesis. *Immunology and Cell Biology* **85**, 130–137 (2007).

[88] Cohen, S., McGregor, I. & Carrington, S. Gamma-globulin and acquired immunity to human malaria. *Nature* **192**, 733–737 (1961).

[89] Fried, M., Nosten, F., Brockman, A., Brabin, B. J. & Duffy, P. E. Maternal antibodies block malaria. *Nature* **395**, 851–852 (1998).

[90] Ferreira, A. *et al.* Sickle hemoglobin confers tolerance to *Plasmodium* infection. *Cell* **145**, 398–409 (2011).

[91] Nájera, J. A., González-Silva, M. & Alonso, P. L. Some lessons for the future from the global malaria eradication programme (1955–1969). *PLoS Medicine* **8**, e1000412 (2011).

[92] malERA Consultative Group on Modeling *et al.* A research agenda for malaria eradication: modeling. *PLoS Medicine* **8**, e1000403 (2011).

[93] Hargreaves, K. *et al.* *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Medical and Veterinary Entomology* **14**, 181–189 (2000).

[94] Tripet, F. *et al.* Longitudinal survey of knockdown resistance to pyrethroid (kdr) in Mali, West Africa, and evidence of its emergence in the Bamako form of *Anopheles gambiae* ss. *The American Journal of Tropical Medicine and Hygiene* **76**, 81–87 (2007).

[95] Martinez-Torres, D. *et al.* Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* ss. *Insect Molecular Biology* **7**, 179–184 (1998).

[96] Bian, G. *et al.* *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. *Science* **340**, 748–751 (2013).

[97] Hughes, G. L., Koga, R., Xue, P., Fukatsu, T. & Rasgon, J. L. *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS Pathogens* **7**, e1002043 (2011).

[98] Kambris, Z. *et al.* *Wolbachia* stimulates immune gene expression and inhibits *Plasmodium* development in *Anopheles gambiae*. *PLoS Pathogens* **6**, e1001143 (2010).

[99] Fitch, C. *et al.* Lysis of *Plasmodium falciparum* by ferriprotoporphyrin IX and a chloroquine-ferriprotoporphyrin IX complex. *Antimicrobial Agents and Chemotherapy* **21**, 819–822 (1982).

[100] Martin, R. *et al.* Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325**, 1680–1682 (2009).

[101] Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular Cell* **6**, 861–871 (2000).

[102] Klonis, N. *et al.* Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. *Proceedings of the National Academy of Sciences* **108**, 11405–11410 (2011).

[103] Hartwig, C. L. *et al.* Accumulation of artemisinin trioxane derivatives within neutral lipids of *Plasmodium falciparum* malaria parasites is endoperoxide-dependent. *Biochemical Pharmacology* **77**, 322–336 (2009).

[104] Noedl, H. *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *New England Journal of Medicine* **359**, 2619–2620 (2008).

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

[106] Cheeseman, I. H. *et al.* A major genome region underlying artemisinin resistance in malaria. *Science* **336**, 79–82 (2012).

[107] Ariey, F. *et al.* A molecular marker of artemisinin-resistant plasmodium falciparum malaria. *Nature* **505**, 50–55 (2014).

[108] Yuthavong, Y., Kamchonwongpaisan, S., Leartsakulpanich, U. & Chitnumsub, P. Folate metabolism as a source of molecular targets for antimalarials. *Future Medicine* (2006).

[109] Yuvaniyama, J. *et al.* Insights into antifolate resistance from malarial DHFR-TS structures. *Nature Structural & Molecular Biology* **10**, 357–365 (2003).

[110] Triglia, T., Menting, J. G., Wilson, C. & Cowman, A. F. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences* **94**, 13944–13949 (1997).

[111] Flannery, E. L., Chatterjee, A. K. & Winzeler, E. A. Antimalarial drug discovery–approaches and progress towards new medicines. *Nature Reviews Microbiology* **11**, 849–862 (2013).

[112] Arama, C. & Troye-Blomberg, M. The path of malaria vaccine development: challenges and perspectives. *Journal of Internal Medicine* **275**, 456–466 (2014).

[113] Carter, R., Mendis, K. N., Miller, L. H., Molineaux, L. & Saul, A. Malaria transmission-blocking vaccines—how can their development be supported? *Nature Medicine* (2000).

[114] Carter, R. Transmission blocking malaria vaccines. *Vaccine* **19**, 2309–2314 (2001).

[115] 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).

[116] Hisaeda, H. *et al.* Antibodies to malaria vaccine candidates Pvs25 and Pvs28 completely block the ability of *Plasmodium vivax* to infect mosquitoes. *Infection and Immunity* **68**, 6618–6623 (2000).

[117] Shimp Jr, R. L. *et al.* Development of a Pfs25-EPA malaria transmission blocking vaccine as a chemically conjugated nanoparticle. *Vaccine* **31**, 2954–2962 (2013).

[118] Sinden, R. E., Carter, R., Drakeley, C. & Leroy, D. The biology of sexual development of *Plasmodium*: the design and implementation of transmission-blocking strategies. *Malaria Journal* **11**, 70 (2012).

[119] The RTS,S Clinical Trials Partnership. A phase 3 trial of RTS, S/AS01 malaria vaccine in African infants. *The New England Journal of Medicine* **367**, 2284–95 (2012).

[120] Bergmann-Leitner, E. S., Legler, P. M., Savranskaya, T., Ockenhouse, C. F. & Angov, E. Cellular and humoral immune effector mechanisms required for sterile protection against sporozoite challenge induced with the novel malaria vaccine candidate CelTOS. *Vaccine* **29**, 5940–5949 (2011).

[121] Nussenzweig, R., Vanderberg, J., Most, H. & Orton, C. Protective immunity produced by the injection of x-irradiated sporozoites of *Plasmodium berghei*. *Nature* (1967).

[122] Seder, R. A. *et al.* Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science* **341**, 1359–1365 (2013).

[123] Bruder, J. T., Ango, E., Limbach, K. J. & Richie, T. L. Molecular vaccines for malaria. *Human Vaccines* **6**, 54–77 (2010).

[124] Takala, S. L. *et al.* Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. *Science Translational Medicine* **1**, 2ra5–2ra5 (2009).

[125] Takala, S. L. & Plowe, C. V. Genetic diversity and malaria vaccine design, testing, and efficacy: Preventing and overcoming 'vaccine resistant malaria'. *Parasite Immunology* **31**, 560 (2009).

[126] Bustamante, L. Y. *et al.* A full-length recombinant *Plasmodium falciparum* PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. *Vaccine* **31**, 373–379 (2013).

[127] 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).

[128] Williams, A. R. *et al.* Enhancing blockade of *Plasmodium falciparum* erythrocyte invasion: assessing combinations of antibodies against PfRH5 and other merozoite antigens. *PLoS Pathogens* **8**, e1002991 (2012).

[129] Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).

[130] Aly, A. S., Downie, M. J., Mamoun, C. B. & Kappe, S. H. Subpatent infection with nucleoside transporter 1-deficient *Plasmodium* blood stage parasites confers sterile protection against lethal malaria in mice. *Cellular Microbiology* **12**, 930–938 (2010).

[131] Ting, L.-M., Gissot, M., Coppi, A., Sinnis, P. & Kim, K. Attenuated *Plasmodium yoelii* lacking purine nucleoside phosphorylase confer protective immunity. *Nature Medicine* **14**, 954–958 (2008).

[132] Raj, D. K. *et al.* Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection. *Science* **344**, 871–877 (2014).

[133] Gilson, P. R. & Crabb, B. S. Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. *International Journal for Parasitology* **39**, 91–96 (2009).

[134] 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).

[135] Persson, K. E. *et al.* Variation in use of erythrocyte invasion pathways by *Plasmodium falciparum* mediates evasion of human inhibitory antibodies. *The Journal of Clinical Investigation* **118**, 342–351 (2008).

[136] Bannister, L., Butcher, G., Dennis, E. & Mitchell, G. Structure and invasive behaviour of *Plasmodium knowlesi* merozoites *in vitro*. *Parasitology* **71**, 483–491 (1975).

[137] Ladda, R., Aikawa, M. & Sprinz, H. Penetration of erythrocytes by merozoites of mammalian and avian malarial parasites. *The Journal of Parasitology* 633–644 (1969).

[138] Boyle, M. J. *et al.* Isolation of viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and advance vaccine and drug development. *Proceedings of the National Academy of Sciences* **107**, 14378–14383 (2010).

[139] Singh, S., Alam, M. M., Pal-Bhowmick, I., Brzostowski, J. A. & Chitnis, C. E. Distinct external signals trigger sequential release of apical organelles during erythrocyte invasion by malaria parasites. *PLoS Pathogens* **6**, e1000746 (2010).

[140] Boyle, M. J., Wilson, D. W. & Beeson, J. G. New approaches to studying *Plasmodium falciparum* merozoite invasion and insights into invasion biology. *International Journal for Parasitology* **43**, 1–10 (2013).

[141] Riglar, D. T. *et al.* Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host & Microbe* **9**, 9–20 (2011).

[142] O'Donnell, R. A., Saul, A., Cowman, A. F. & Crabb, B. S. Functional conservation of the malaria vaccine antigen MSP-119 across distantly related *Plasmodium* species. *Nature Medicine* **6**, 91–95 (2000).

[143] Sanders, P. R. *et al.* Distinct protein classes including novel merozoite surface antigens in raft-like membranes of *Plasmodium falciparum*. *Journal of Biological Chemistry* **280**, 40169–40176 (2005).

[144] Bannister, L., Hopkins, J., Fowler, R., Krishna, S. & Mitchell, G. Ultrastructure of rhoptry development in *Plasmodium falciparum* erythrocytic schizonts. *Parasitology* **121**, 273–287 (2000).

[145] Preiser, P., Kaviratne, M., Khan, S., Bannister, L. & Jarra, W. The apical organelles of malaria merozoites: host cell selection, invasion, host immunity and immune evasion. *Microbes and Infection* **2**, 1461–1477 (2000).

[146] Ngô, H. M., Yang, M. & Joiner, K. A. Are rhoptries in Apicomplexan parasites secretory granules or secretory lysosomal granules? *Molecular Microbiology* **52**, 1531–1541 (2004).

[147] Yeoh, S. *et al.* Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell* **131**, 1072–1083 (2007).

[148] Treeck, M. *et al.* Functional analysis of the leading malaria vaccine candidate AMA-1 reveals an essential role for the cytoplasmic domain in the invasion process. *PLoS Pathogens* **5**, e1000322 (2009).

[149] Miller, L. H., Aikawa, M., Johnson, J. G. & Shiroishi, T. Interaction between cytochalasin b-treated malarial parasites and erythrocytes. Attachment and junction formation. *The Journal of Experimental Medicine* **149**, 172–184 (1979).

[150] Lew, V. L. & Tiffert, T. Is invasion efficiency in malaria controlled by pre-invasion events? *Trends in Parasitology* **23**, 481–484 (2007).

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

[152] Farrow, R. E. *et al.* The mechanism of erythrocyte invasion by the malarial parasite, *Plasmodium falciparum*. In *Seminars in Cell Developmental Biology*, vol. 22, 953–960 (Elsevier, 2011).

[153] Srinivasan, P. *et al.* Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proceedings of the National Academy of Sciences* **108**, 13275–13280 (2011).

[154] Dvorin, J. D., Bei, A. K., Coleman, B. I. & Duraisingh, M. T. Functional diversification between two related *Plasmodium falciparum* merozoite invasion ligands is determined by changes in the cytoplasmic domain. *Molecular Microbiology* **75**, 990–1006 (2010).

[155] Besteiro, S., Dubremetz, J.-F. & Lebrun, M. The moving junction of Apicomplexan parasites: a key structure for invasion. *Cellular Microbiology* **13**, 797–805 (2011).

[156] Keeley, A. & Soldati, D. The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. *Trends in Cell Biology* **14**, 528–532 (2004).

[157] Bannister, L., Mitchell, G., Butcher, G. & Dennis, E. Lamellar membranes associated with rhoptries in erythrocytic merozoites of *Plasmodium knowlesi*: a clue to the mechanism of invasion. *Parasitology* **92**, 291–303 (1986).

[158] Mordue, D. G., Desai, N., Dustin, M. & Sibley, L. D. Invasion by *Toxoplasma gondii* establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring. *The Journal of Experimental Medicine* **190**, 1783–1792 (1999).

[159] Charron, A. J. & Sibley, L. D. Molecular partitioning during host cell penetration by *Toxoplasma gondii*. *Traffic* **5**, 855–867 (2004).

[160] Harris, P. K. *et al.* Molecular identification of a malaria merozoite surface sheddase. *PLoS Pathogens* **1**, e29 (2005).

[161] Cowman, A. F., Berry, D. & Baum, J. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *The Journal of Cell Biology* **198**, 961–971 (2012).

[162] Blackman, M. J., Scott-Finnigan, T. J., Shai, S. & Holder, A. A. Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. *The Journal of Experimental Medicine* **180**, 389–393 (1994).

[163] Holder, A. The precursor to major merozoite surface antigens: structure and role in immunity. *Progress in Allergy* **41**, 72 (1988).

[164] 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).

[165] Kauth, C. W., Epp, C., Bujard, H. & Lutz, R. The merozoite surface protein 1 complex of human malaria parasite *Plasmodium falciparum*: interactions and arrangements of subunits. *Journal of Biological Chemistry* (2003).

[166] Blackman, M. J., Heidrich, H.-G., Donachie, S., McBride, J. & Holder, A. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. *The Journal of Experimental Medicine* **172**, 379–382 (1990).

[167] Egan, A. F., Burghaus, P., Bruilhe, P., Holder, A. A. & Riley, E. M. Human antibodies to the 19 kDa C-terminal fragment of*Plasmodium falciparum* merozoite surface protein 1 inhibit parasite growth *in vitro*. *Parasite Immunology* **21**, 133–139 (1999).

[168] Low, A. *et al.* Merozoite surface protein 2 of *Plasmodium falciparum*: Expression, structure, dynamics, and fibril formation of the conserved N-terminal domain. *Biopolymers* **87**, 12–22 (2007).

[169] Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–766 (2006).

[170] Wickramarachchi, T. *et al.* A novel *Plasmodium falciparum* erythrocyte binding protein associated with the merozoite surface, PfDBLMSP. *International Journal for Parasitology* **39**, 763–773 (2009).

[171] Hodder, A. N. *et al.* Insights into Duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from *Plasmodium falciparum*. *Journal of Biological Chemistry* **287**, 32922–32939 (2012).

[172] Sakamoto, H. *et al.* Antibodies against *Plasmodium falciparum* antigen PfMSPDBL1 inhibit merozoite invasion into human erythrocytes. *Vaccine* **30**, 1972–1980 (2012).

[173] Triglia, T. *et al.* Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. *Molecular Microbiology* **38**, 706–718 (2000).

[174] Bargieri, D. Y. *et al.* Apical membrane antigen 1 mediates Apicomplexan parasite attachment but is dispensable for host cell invasion. *Nature Communications* **4** (2013).

[175] Thera, M. A. *et al.* A field trial to assess a blood-stage malaria vaccine. *New England Journal of Medicine* **365**, 1004–1013 (2011).

[176] Leykauf, K. *et al.* Protein kinase A dependent phosphorylation of apical membrane antigen 1 plays an important role in erythrocyte invasion by the malaria parasite. *PLoS Pathogens* **6**, e1000941 (2010).

[177] Alexander, D. L., Mital, J., Ward, G. E., Bradley, P. & Boothroyd, J. C. Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles. *PLoS Pathogens* **1**, e17 (2005).

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

[179] Tonkin, M. L. *et al.* Host cell invasion by Apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. *Science* **333**, 463–467 (2011).

[180] Vulliez-Le Normand, B. *et al.* Structural and functional insights into the malaria parasite moving junction complex. *PLoS Pathogens* **8**, e1002755 (2012).

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

[182] Starnes, G. L., Coincon, M., Sygusch, J. & Sibley, L. D. Aldolase is essential for energy production and bridging adhesin-actin cytoskeletal interactions during parasite invasion of host cells. *Cell Host & Microbe* **5**, 353–364 (2009).

[183] Baum, J., Papenfuss, A. T., Baum, B., Speed, T. P. & Cowman, A. F. Regulation of Apicomplexan actin-based motility. *Nature Reviews Microbiology* **4**, 621–628 (2006).

[184] Gonzalez, V. *et al.* Host cell entry by Apicomplexa parasites requires actin polymerization in the host cell. *Cell Host & Microbe* **5**, 259–272 (2009).

[185] Tham, W.-H., Healer, J. & Cowman, A. F. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends in Parasitology* **28**, 23–30 (2012).

[186] Duraisingh, M. T. *et al.* Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *The EMBO Journal* **22**, 1047–1057 (2003).

[187] Baum, J. *et al.* Reticulocyte-binding protein homologue 5–an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *International Journal for Parasitology* **39**, 371–380 (2009).

[188] Stubbs, J. *et al.* Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science* **309**, 1384–1387 (2005).

[189] Wickramarachchi, T., Devi, Y. S., Mohmmed, A. & Chauhan, V. S. Identification and characterization of a novel *Plasmodium falciparum* merozoite apical protein involved in erythrocyte binding and invasion. *PLoS ONE* **3**, e1732 (2008).

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

[191] Ménard, D. *et al.* *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proceedings of the National Academy of Sciences* **107**, 5967–5971 (2010).

[192] Gaur, D., Mayer, D. & Miller, L. H. Parasite ligand–host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *International Journal for Parasitology* **34**, 1413–1429 (2004).

[193] Wanaguru, M., Crosnier, C., Johnson, S., Rayner, J. C. & Wright, G. J. Biochemical analysis of the *Plasmodium falciparum* erythrocyte-binding antigen-175 (EBA175)-Glycophorin-A interaction: Implications for vaccine design. *Journal of Biological Chemistry* **288**, 32106–32117 (2013).

[194] Batchelor, J. D., Zahm, J. A. & Tolia, N. H. Dimerization of *Plasmodium vivax* DBP is induced upon receptor binding and drives recognition of DARC. *Nature Structural & Molecular Biology* **18**, 908–914 (2011).

[195] Chen, E., Paing, M. M., Salinas, N., Sim, B. K. L. & Tolia, N. H. Structural and functional basis for inhibition of erythrocyte invasion by antibodies that target *Plasmodium falciparum* EBA-175. *PLoS Pathogens* **9**, e1003390 (2013).

[196] Lin, D. H., Malpede, B. M., Batchelor, J. D. & Tolia, N. H. Crystal and solution structures of *Plasmodium falciparum* erythrocyte-binding antigen 140 reveal determinants of receptor specificity during erythrocyte invasion. *Journal of Biological Chemistry* **287**, 36830–36836 (2012).

[197] O'Donnell, R. A. *et al.* Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *The Journal of Cell Biology* **174**, 1023–1033 (2006).

[198] Wright, G. J. & Rayner, J. C. *Plasmodium falciparum* erythrocyte invasion: Combining function with immune evasion. *PLoS Pathogens* **10**, e1003943 (2014).

[199] Fowkes, F. J., Richards, J. S., Simpson, J. A. & Beeson, J. G. The relationship between anti-merozoite antibodies and incidence of*Plasmodium falciparum* malaria: A systematic review and meta-analysis. *PLoS Medicine* **7**, e1000218 (2010).

[200] Tran, T. M. *et al.* Naturally acquired antibodies specific for *Plasmodium falciparum* reticulocyte-binding protein homologue 5 inhibit parasite growth and predict protection from malaria. *Journal of Infectious Diseases* **209**, 789–798 (2014).

[201] Hafalla, J. C., Silvie, O. & Matuschewski, K. Cell biology and immunology of malaria. *Immunological reviews* **240**, 297–316 (2011).

[202] Chen, L. *et al.* An EGF-like protein forms a complex with PfRh5 and is required for invasion of human erythrocytes by*Plasmodium falciparum*. *PLoS Pathogens* **7**, e1002199 (2011).

[203] Galinski, M. R., Medina, C. C., Ingravallo, P. & Barnwell, J. W. A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* **69**, 1213–1226 (1992).

[204] Freeman, R. R., Trejdosiewicz, A. J. & Cross, G. A. Protective monoclonal antibodies recognising stage-specific merozoite antigens of a rodent malaria parasite. *Nature* (1980).

[205] Holder, A. A. & Freeman, R. R. Immunization against blood-stage rodent malaria using purified parasite antigens. *Nature* (1981).

[206] Gunalan, K., Gao, X., Yap, S. S. L., Huang, X. & Preiser, P. R. The role of the reticulocyte-binding-like protein homologues of *Plasmodium* in erythrocyte sensing and invasion. *Cellular Microbiology* **15**, 35–44 (2013).

[207] Keen, J. K., Sinha, K. A., Brown, K. N. & Holder, A. A. A gene coding for a high-molecular mass rhoptry protein of *Plasmodium yoelii*. *Molecular and Biochemical Parasitology* **65**, 171–177 (1994).

[208] Ogun, S. A. & Holder, A. A. A high molecular mass *Plasmodium yoelii* rhoptry protein binds to erythrocytes. *Molecular and Biochemical Parasitology* **76**, 321–324 (1996).

[209] Preiser, P., Jarra, W., Capiod, T. & Snounou, G. A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* **398**, 618–622 (1999).

[210] Ogun, S. A., Scott-Finnigan, T. J., Narum, D. L. & Holder, A. A. *Plasmodium yoelii:* effects of red blood cell modification and antibodies on the binding characteristics of the 235-kDa rhoptry protein. *Experimental Parasitology* **95**, 187–195 (2000).

[211] Preiser, P. *et al.* Stage-specific transcription of distinct repertoires of a multigene family during *Plasmodium* life cycle. *Science* **295**, 342–345 (2002).

[212] Kosaisavee, V. *et al.* Genetic diversity in new members of the reticulocyte binding protein family in Thai *Plasmodium vivax* isolates. *PloS ONE* **7**, e32105 (2012).

[213] Taylor, H. M. *et al. Plasmodium falciparum* homologue of the genes for *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins, which is transcribed but not translated. *Infection and Immunity* **69**, 3635–3645 (2001).

[214] Rayner, J. C., Vargas-Serrato, E., Huber, C. S., Galinski, M. R. & Barnwell, J. W. A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *The Journal of Experimental Medicine* **194**, 1571–1582 (2001).

[215] Gao, X. *et al.* Antibodies targeting the PfRH1 binding domain inhibit invasion of*Plasmodium falciparum* merozoites. *PLoS Pathogens* **4**, e1000104 (2008).

[216] Triglia, T., Duraisingh, M. T., Good, R. T. & Cowman, A. F. Reticulocyte-binding protein homologue 1 is required for sialic acid-dependent invasion into human erythrocytes by *Plasmodium falciparum*. *Molecular Microbiology* **55**, 162–174 (2005).

[217] Rayner, J. C., Huber, C. S., Galinski, M. R. & Barnwell, J. W. Rapid evolution of an erythrocyte invasion gene family: the *Plasmodium reichenowi* Reticulocyte Binding Like (RBL) genes. *Molecular and Biochemical Parasitology* **133**, 287–296 (2004).

[218] Rayner, J. C. The merozoite has landed: reticulocyte-binding-like ligands and the specificity of erythrocyte recognition. *Trends in Parasitology* **25**, 104–106 (2009).

[219] Triglia, T., Tham, W.-H., Hodder, A. & Cowman, A. F. Reticulocyte binding protein homologues are key adhesins during erythrocyte invasion by *Plasmodium falciparum*. *Cellular Microbiology* **11**, 1671–1687 (2009).

[220] Rayner, J. C., Galinski, M. R., Ingravallo, P. & Barnwell, J. W. Two *Plasmodium falciparum* genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proceedings of the National Academy of Sciences* **97**, 9648–9653 (2000).

[221] Sahar, T. *et al.* Plasmodium falciparum reticulocyte binding-like homologue protein 2 (PfRH2) is a key adhesive molecule involved in erythrocyte invasion. *PLoS ONE* **6**, e17102 (2011).

[222] Triglia, T. *et al.* *Plasmodium falciparum* merozoite invasion is inhibited by antibodies that target the PfRh2a and b binding domains. *PLoS Pathogens* **7**, e1002075 (2011).

[223] Reiling, L. *et al.* Evidence that the erythrocyte invasion ligand PfRh2 is a target of protective immunity against *Plasmodium falciparum* malaria. *The Journal of Immunology* **185**, 6157–6167 (2010).

[224] Gunalan, K., Gao, X., Liew, K. J. & Preiser, P. R. Differences in erythrocyte receptor specificity of different parts of the *Plasmodium falciparum* reticulocyte binding protein homologue 2a. *Infection and Immunity* **79**, 3421–3430 (2011).

[225] Kaneko, O., Mu, J., Tsuboi, T., Su, X. & Torii, M. Gene structure and expression of a *Plasmodium falciparum* 220-kDa protein homologous to the *Plasmodium vivax* reticulocyte binding proteins. *Molecular and Biochemical Parasitology* **121**, 275–278 (2002).

[226] Tham, W.-H. *et al.* Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *Plasmodium falciparum* into human erythrocytes. *Infection and Immunity* **77**, 2427–2435 (2009).

[227] Gaur, D. *et al.* Upregulation of expression of the reticulocyte homology gene 4 in the *Plasmodium falciparum* clone Dd2 is associated with a switch in the erythrocyte invasion pathway. *Molecular and Biochemical Parasitology* **145**, 205–215 (2006).

[228] Gaur, D. *et al.* Recombinant *Plasmodium falciparum* reticulocyte homology protein 4 binds to erythrocytes and blocks invasion. *Proceedings of the National Academy of Sciences* **104**, 17789–17794 (2007).

[229] Reiling, L. *et al.* The *Plasmodium falciparum* erythrocyte invasion ligand PfRh4 as a target of functional and protective human antibodies against malaria. *PloS ONE* **7**, e45253 (2012).

[230] Tham, W.-H. *et al.* Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proceedings of the National Academy of Sciences* **107**, 17327–17332 (2010).

[231] Krych-Goldberg, M. & Atkinson, J. P. Structure–function relationships of complement receptor type 1. *Immunological Reviews* **180**, 112–122 (2001).

[232] Park, H. J. *et al.* Using mutagenesis and structural biology to map the binding site for the *Plasmodium falciparum* merozoite protein PfRh4 on the human immune adherence receptor. *Journal of Biological Chemistry* **289**, 450–463 (2014).

[233] Tham, W.-H. *et al.* *Plasmodium falciparum* uses a key functional site in complement receptor type-1 for invasion of human erythrocytes. *Blood* **118**, 1923–1933 (2011).

[234] 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).

[235] Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host & Microbe* **4**, 40–51 (2008).

[236] Rodriguez, M., Lustigman, S., Montero, E., Oksov, Y. & Lobo, C. A. PfRH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE* **3**, e3300 (2008).

[237] Douglas, A. D. *et al.* Neutralization of *Plasmodium falciparum* merozoites by antibodies against PfRH5. *J Immunol* **192**, 245–58 (2014).

[238] Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* **480**, 534–537 (2011).

[239] Yu, X.-L. *et al.* Crystal structure of HAb18G/CD147 implications for immunoglobulin superfamily homophilic adhesion. *Journal of Biological Chemistry* **283**, 18056–18065 (2008).

[240] Yurchenko, V., Constant, S. & Bukrinsky, M. Dealing with the family: CD147 interactions with cyclophilins. *Immunology* **117**, 301–309 (2006).

[241] Bernard, S. C. *et al.* Pathogenic neisseria meningitidis utilizes CD147 for vascular colonization. *Nature Medicine* (2014).

[242] Wanaguru, M., Liu, W., Hahn, B. H., Rayner, J. C. & Wright, G. J. Rh5–basigin interaction plays a major role in the host tropism of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences* **110**, 20735–20740 (2013).

[243] Hayton, K. *et al.* Various PfRH5 polymorphisms can support *Plasmodium falciparum* invasion into the erythrocytes of owl monkeys and rats. *Molecular and Biochemical Parasitology* **187**, 103–110 (2013).

[244] Reddy, K. S. *et al.* Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infection and Immunity* **82**, 152–64 (2014).

[245] Higgins, M. K. & Carrington, M. Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein Science* **23**, 354–365 (2014).

[246] Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E. & Sharma, A. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* **439**, 741–744 (2005).

[247] Bai, T. *et al.* Structure of AMA1 from *Plasmodium falciparum* reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12736–12741 (2005).

[248] Chitarra, V., Holm, I., Bentley, G. A., Pêtres, S. & Longacre, S. The crystal structure of C-terminal merozoite surface protein 1 at 1.8 Å resolution, a highly protective malaria vaccine candidate. *Molecular Cell* **3**, 457–464 (1999).

[249] Morgan, W. D. *et al.* Solution structure of an EGF module pair from the *Plasmodium falciparum* merozoite surface protein 1. *Journal of Molecular Biology* **289**, 113–122 (1999).

[250] Gerloff, D. L., Creasey, A., Maslau, S. & Carter, R. Structural models for the protein family characterized by gamete surface protein Pfs230 of*Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13598–13603 (2005).

[251] Arredondo, S. A. *et al.* Structure of the *Plasmodium* 6-cysteine s48/45 domain. *Proceedings of the National Academy of Sciences* **109**, 6692–6697 (2012).

[252] Naitza, S., Spano, F., Robson, K. & Crisanti, A. The thrombospondin-related protein family of Apicomplexan parasites: the gears of the cell invasion machinery. *Parasitology Today* **14**, 479–484 (1998).

[253] Tomley, F. M. & Soldati, D. S. Mix and match modules: structure and function of microneme proteins in Apicomplexan parasites. *Trends in Parasitology* **17**, 81–88 (2001).

[254] Feng, Z.-P. *et al.* Abundance of intrinsically unstructured proteins in *P. falciparum* and other Apicomplexan parasite proteomes. *Molecular and Biochemical Parasitology* **150**, 256–267 (2006).

[255] Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

[256] Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology* **310**, 243–257 (2001).

[257] Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).

[258] Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology* **347**, 827–839 (2005).

[259] Peng, K. *et al.* Optimizing long intrinsic disorder predictors with protein evolutionary information. *Journal of Bioinformatics and Computational Biology* **3**, 35–60 (2005).

[260] Hirose, S., Shimizu, K., Kanai, S., Kuroda, Y. & Noguchi, T. POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions. *Bioinformatics* **23**, 2046–2053 (2007).

[261] Rost, B. Review: protein secondary structure prediction continues to rise. *Journal of Structural Biology* **134**, 204–218 (2001).

[262] Pollastri, G. & Mclysaght, A. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* **21**, 1719–1720 (2005).

[263] Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* **292**, 195–202 (1999).

[264] Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **36**, W197–W201 (2008).

[265] Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1–6 (1997).

[266] Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7** (2011).

[267] Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**, 205–217 (2000).

[268] Gouet, P., Courcelle, E., Stuart, D. I. *et al.* Espript: analysis of multiple sequence alignments in postscript. *Bioinformatics* **15**, 305–308 (1999).

[269] Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).

[270] Higgins, M. K. Overproduction, purification and crystallization of a chondroitin sulfate A-binding DBL domain from a *Plasmodium falciparum* var2csa-encoded PfEMP1 protein. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **64**, 221–223 (2008).

[271] 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).

[272] Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* **35**, e45 (2007).

[273] Veggiani, G. & de Marco, A. Improved quantitative and qualitative production of single-domain intrabodies mediated by the co-expression of Erv1p sulfhydryl oxidase. *Protein Expression and Purification* **79**, 111–114 (2011).

[274] Schneider, I. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *Journal of Embryology and Experimental Morphology* **27**, 353–365 (1972).

[275] Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**, 2212–2221 (2007).

[276] Serdyuk, I. N., Zaccai, N. R. & Zaccai, J. *Methods in Molecular Biophysics: Structure, Dynamics, Function* (Cambridge University Press, 2007).

[277] Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* **78**, 1606–1619 (2000).

[278] Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. & Schubert, D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophysical Journal* **82**, 1096–1111 (2002).

[279] Laue, T., Shah, B., Ridgeway, T. & Pelletier, S. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Royal Society of Chemistry, 1992).

[280] Vistica, J. *et al.* Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Analytical Biochemistry* **326**, 234–256 (2004).

[281] Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chemical Reviews* **107**, 3544–3567 (2007).

[282] Dole, M. *et al.* Molecular beams of macroions. *The Journal of Chemical Physics* **49**, 2240–2249 (1968).

[283] Wilm, M., Neubauer, G. & Mann, M. Parent ion scans of unseparated peptide mixtures. *Analytical Chemistry* **68**, 527–533 (1996).

[284] Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191–285 (2007).

[285] Jacques, D. A. & Trewhella, J. Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Science* **19**, 642–657 (2010).

[286] Guinier, A., Fournet, G., Walker, C. B. & Yudowitch, K. L. *Small-angle scattering of X-rays* (Wiley New York, 1955).

[287] Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. Primus: a windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277–1282 (2003).

[288] Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied Crystallography* (2007).

[289] Porod, G. General theory. In *Small angle X-ray scattering* (Academic Press, 1982).

[290] Franke, D. & Svergun, D. I. Dammif, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**, 342–346 (2009).

[291] Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **39**, 277–286 (2006).

[292] Kozin, M. B. & Svergun, D. I. Automated matching of high-and low-resolution structural models. *Journal of Applied Crystallography* **34**, 33–41 (2001).

[293] Birmanns, S., Rusu, M. & Wriggers, W. Using sculptor and situs for simultaneous assembly of atomic components into low-resolution shapes. *Journal of Structural Biology* **173**, 428–435 (2011).

[294] Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* **67**, 271–281 (2011).

[295] Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallographica Section D: Biological Crystallography* **62**, 48–57 (2005).

[296] Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* **55**, 1690–1695 (1999).

[297] Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235–242 (2011).

[298] Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* **62**, 72–82 (2005).

[299] Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography* **67**, 282–292 (2011).

[300] McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674 (2007).

[301] Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Progress in Biophysics and Molecular Biology* **72**, 245–270 (1999).

[302] Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica Section D: Biological Crystallography* **59**, 2023–2030 (2003).

[303] Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486–501 (2010).

[304] Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography* **62**, 1002–1011 (2006).

[305] Rupp, B. *Biomolecular crystallography* (Garland Science New York, 2010).

[306] Bricogne, G. *et al.* Buster version 2.11. 2. *Cambridge, United Kingdom: Global Phasing Ltd* (2011).

[307] Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283–291 (1993).

[308] Sheldrick, G. M. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography* **64**, 112–122 (2007).

[309] Dunker, A. K. *et al.* Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **19**, 26–59 (2001).

[310] Birkholtz, L.-M. *et al.* Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria Journal* **7**, 197 (2008).

[311] Pizarro, J. *et al.* Crystal structure of a Fab complex formed with PfMSP1-19, the c-terminal fragment of merozoite surface protein 1 from *Plasmodium falciparum*: A malaria vaccine candidate. *Journal of molecular biology* **328**, 1091–1103 (2003).

[312] Narum, D. L., Welling, G. W. & Thomas, A. W. Ion-exchange–immunoaffinity purification of a recombinant baculovirus *Plasmodium falciparum* apical membrane antigen, PF83/AMA-1. *Journal of Chromatography A* **657**, 357–363 (1993).

[313] Barfod, L. *et al.* Human pregnancy-associated malaria-specific B cells target polymorphic, conformational epitopes in VAR2CSA. *Molecular Microbiology* **63**, 335–347 (2007).

[314] Hunt, I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification* **40**, 1–22 (2005).

[315] Wright, K. E. *et al.* Structure of malaria invasion protein rh5 with erythrocyte basigin and blocking antibodies. *Nature* (2014).

[316] McBride, J. S. & Heidrich, H.-G. Fragments of the polymorphic Mr 185 000 glycoprotein from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. *Molecular and Biochemical Parasitology* **23**, 71–84 (1987).

[317] Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617–1622 (2006).

[318] Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2184–2195 (2004).

[319] Holm, L. & Rosenström, P. Dali server: conservation mapping in 3d. *Nucleic Acids Research* **38**, W545–W549 (2010).

[320] Luo, J. *et al.* Structure of the EMMPRIN N-terminal domain 1: Dimerization via β-strand swapping. *Proteins: Structure, Function, and Bioinformatics* **77**, 1009–1014 (2009).

[321] Stone, J. D., Chervin, A. S. & Kranz, D. M. T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. *Immunology* **126**, 165–176 (2009).

[322] Azoitei, M. L. *et al.* Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *Journal of Molecular Biology* **415**, 175–192 (2012).

Appendix A

Primer Sequences

Construct	Direction	Rest. site	Sequence
BSG–pET$_{con}$	f	BamHI	ccggatccgctgccggaaccgtgttc
BSG–pET$_{con}$	r	NdeI	cccatatgctagtgtgaccgcactctcagg
BSG–pRSF-Duet1	f	BamHI	ccggatccgctgccggaaccgtgttc
BSG–pRSF-Duet1	r	HindIII	ccaagcttctagtgtgaccgcactctcagg
BSG(T28)–pET$_{con}$	f	N/A	gccggaaccgtgttcttcaccgtggaggacctg
BSG(T28)–pET$_{con}$	r	N/A	caggtcctccacggtgaagaacacggttccggc
BSG(G95)–pET$_{con}$	f	N/A	ctgcccgagcctatgttcaccgccaatatccag
BSG(G95)–pET$_{con}$	r	N/A	ctggatattggcggtgaacataggctcgggcag
BSG(H102)–pET$_{con}$	f	N/A	gccaatatccagctggccggccctccaagagtg
BSG(H102)–pET$_{con}$	r	N/A	cactcttggagggccggccagctggatattggc
BSG(V131)–pET$_{con}$	f	N/A	tgcaagagcgagagcgatccccctgtgaccgac
BSG(V131)–pET$_{con}$	r	N/A	gtcggtcacaggggatcgctctcgctcttgca
BSG(S190)–pET$_{con}$	f	N/A	tgcaacggcaccagcgacaagggctccgaccag
BSG(S190)–pET$_{con}$	r	N/A	ctggtcggagcccttgtcgctggtgccgttgca

Table A.1: Summary of basigin primers, including mutagenesis (T28, G95, H102, V131, S190) primers.

Construct	Direction	Rest. site	Sequence
PfRH1–pExpreS^2.1$_{con}$	f	EcoRI	aagaattcagctacaagttcgatgtgtacaaggac
PfRH1–pExpreS^2.1$_{con}$	r	NotI	ttgcggccgcctagtggtgatggtggtggtgcttgaaggcatactctttcaggtc
PfRH2b–pExpreS^2.1$_{con}$	f	EcoRI	aagaattctccaactccgtgctggaccac
PfRH2b–pExpreS^2.1$_{con}$	r	NotI	ttgcggccgcctagtggtgatggtggtggtgagtctcgtacttcttcttgaagatgtt
PfRH4–pExpreS^2.1$_{con}$	f	EcoRI	aagaattcagcaacatcctgatggacgagatc
PfRH4–pExpreS^2.1$_{con}$	r	NotI	ttgcggccgcctagtggtgatggtggtggtggttgttgattttgtgctgcagctg

Table A.2: Summary of PfRH1, PfRH2b, and PfRH4 primers for S2 cell expression.

TABLE A.3: Primers for PfRH5 expression in *E.coli*. Abbreviations: FL, full-length; NTD, N-terminal domain; CTD, C-terminal domain; aa_N and aa_C, N- and C-terminal residues, respectively; dir., direction.

No.	Vector	Domain	aa_N	aa_C	Dir.	Sequence
1	pET15b$_{con}$	FL	26	526	f	ccggatccgagaacgccatcaagaaaaccaagaaccaggaaaac
					r	cccatatgctactgggtcaggggcttgttcttctt
2	pOPINF	NTD	68	250	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
3	pOPINFB	NTD	68	250	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	agatgtcgttcaggccGTCGTCGTTCTTGTTGTTGATGTCG
4	pOPINTRX	NTD	68	250	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
5	pOPINM	NTD	68	250	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
6	pOPINS3C	NTD	68	250	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
7	pOPINF	NTD	76	250	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
8	pOPINFB	NTD	76	250	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	agatgtcgttcaggccGTCGTCGTTCTTGTTGTTGATGTCG
9	pOPINTRX	NTD	76	250	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
10	pOPINM	NTD	76	250	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
11	pOPINS3C	NTD	76	250	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaGTCGTCGTTCTTGTTGTTGATGTCG
12	pOPINF	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
13	pOPINFB	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	agatgtcgttcaggccCTTGATTGGCACGTCGTTGAACTTC
14	pOPINM	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
15	pOPINF	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
16	pOPINFB	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	agatgtcgttcaggccCTTGTTCTTCTTGTAGGTCTGGAAG
17	pOPINM	FL	68	511	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
18	pOPINF	FL	68	526	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGCC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
19	pOPINFB	FL	68	526	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGCC
					r	agatgtcgttcaggccCTGGGTCAGGGGCTTGTTCTTCTTG
20	pOPINM	FL	68	526	f	aagttctgtttcagggcccgAAAGAGAACATCAAGACCAACAACGCC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
21	pOPINF	FL	76	511	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
22	pOPINM	FL	76	511	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
23	pOPINF	FL	76	522	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC

Continued on next page...

No.	Vector	Domain	aa_N	aa_C	Dir.	Sequence
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
24	pOPINM	FL	76	522	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
25	pOPINF	FL	76	526	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
26	pOPINM	FL	76	526	f	aagttctgtttcagggcccgGCCAAGGACCACAGCACCTACATC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
27	pOPINF	CTD	255	526	f	aagttctgtttcagggcccgGACATCTCCGAGGAAATCGACGAC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
28	pOPINM	CTD	255	526	f	aagttctgtttcagggcccgGACATCTCCGAGGAAATCGACGAC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
29	pOPINF	CTD	277	526	f	aagttctgtttcagggcccgAGCATCCAGGACACCGACAGCAAC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
30	pOPINM	CTD	277	526	f	aagttctgtttcagggcccgAGCATCCAGGACACCGACAGCAAC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
31	pOPINF	CTD	288	526	f	aagttctgtttcagggcccgAGCAACAAGAAAAAGAACGACCTGATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
32	pOPINM	CTD	288	526	f	aagttctgtttcagggcccgAGCAACAAGAAAAAGAACGACCTGATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
33	pOPINF	CTD	292	511	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
34	pOPINFB	CTD	292	511	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	agatgtcgttcaggccCTTGATTGGCACGTCGTTGAACTTC
35	pOPINTRX	CTD	292	511	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
36	pOPINM	CTD	292	511	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
37	pOPINS3C	CTD	292	511	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGATTGGCACGTCGTTGAACTTC
38	pOPINF	CTD	292	522	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
39	pOPINFB	CTD	292	522	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	agatgtcgttcaggccCTTGTTCTTCTTGTAGGTCTGGAAG
40	pOPINTRX	CTD	292	522	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
41	pOPINM	CTD	292	522	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
42	pOPINS3C	CTD	292	522	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTTGTTCTTCTTGTAGGTCTGGAAG
43	pOPINF	CTD	292	526	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
44	pOPINFB	CTD	292	526	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	agatgtcgttcaggccCTGGGTCAGGGGCTTGTTCTTCTTG
45	pOPINTRX	CTD	292	526	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
46	pOPINM	CTD	292	526	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
47	pOPINS3C	CTD	292	526	f	aagttctgtttcagggcccgAAGAACGACCTGATGAACCGGACC
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG

Continued on next page...

No.	Vector	Domain	aa_N	aa_C	Dir.	Sequence
48	pOPINF	CTD	296	526	f	aagttctgtttcagggcccgATGAACCGGACCTTCAAGAAAATGATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
49	pOPINM	CTD	296	526	f	aagttctgtttcagggcccgATGAACCGGACCTTCAAGAAAATGATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
50	pOPINF	CTD	322	526	f	aagttctgtttcagggcccgGAGAACGACTTCAACAAAATCTGCATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
51	pOPINM	CTD	322	526	f	aagttctgtttcagggcccgGAGAACGACTTCAACAAAATCTGCATG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
52	pOPINF	CTD	347	526	f	aagttctgtttcagggcccgAACAACAACTTCTGCAACACCAACGG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
53	pOPINM	CTD	347	526	f	aagttctgtttcagggcccgAACAACAACTTCTGCAACACCAACGG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
54	pOPINF	CTD	366	526	f	aagttctgtttcagggcccgAAGCTGATCCTGAGCGTGAAGTCTAAG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
55	pOPINM	CTD	366	526	f	aagttctgtttcagggcccgAAGCTGATCCTGAGCGTGAAGTCTAAG
					r	atggtctagaaagctttaCTGGGTCAGGGGCTTGTTCTTCTTG
56	pET15b$_{con}$	FL	76	522	f	ccggatccgccaaggaccacagcacctacatc
					r	gggctagcctacttgttcttcttgtaggtctggaag
57	pET15b$_{con}$	FL	140	522	f	ccggatccaagaatgtgaacttcctgcagtaccac
					r	gggctagcctacttgttcttcttgtaggtctggaag
58	pET15b$_{con}$	FL	150	522	f	ccggatccaaagagctgagcaactacaatatcgcc
					r	gggctagcctacttgttcttcttgtaggtctggaag
59	pET15b$_{con}$	NTD	76	250	f	ccggatccgccaaggaccacagcacctacatc
					r	gggctagcctagtcgtcgttcttgttgttgatgtcg
60	pET15b$_{con}$	CTD	292	522	f	ccggatccaagaacgacctgatgaaccggacc
					r	gggctagcctacttgttcttcttgtaggtctggaag
61	pET15b$_{con}$	FLΔ243-291	76	522	f	aagaacgacctgatgaaccggacc
					r	gtaggggtgttccagcttcttgatggt
62	pET15b$_{con}$	FLΔ246-294	76	522	f	ctgatgaaccggaccttcaagaaaatg
					r	gttgatgtcgtaggggtgttccag
63	pET15b$_{con}$	FLΔ240-288	76	522	f	aacaagaaaaagaacgacctgatgaaccgg
					r	ttccagcttcttgatggtggcgatcag
64	pET15b$_{con}$	FLΔ248-296	76	522	f	aaccggaccttcaagaaaatgatggac
					r	cttgttgttgatgtcgtaggggtgttc
65	pET15b$_{con}$	FLΔ247-295	76	522	f	atgaaccggaccttcaagaaaatgatggac
					r	gttgttgatgtcgtaggggtgttccag
66	pET15b$_{con}$	FLΔ243-291	140	522	f	aagaacgacctgatgaaccggacc
					r	gtaggggtgttccagcttcttgatggt
67	pET15b$_{con}$	FLΔ246-294	140	522	f	ctgatgaaccggaccttcaagaaaatg
					r	gttgatgtcgtaggggtgttccag
68	pET15b$_{con}$	FLΔ240-288	140	522	f	aacaagaaaaagaacgacctgatgaaccgg
					r	ttccagcttcttgatggtggcgatcag
69	pET15b$_{con}$	FLΔ248-296	140	522	f	aaccggaccttcaagaaaatgatggac
					r	cttgttgttgatgtcgtaggggtgttc
70	pET15b$_{con}$	FLΔ247-295	140	522	f	atgaaccggaccttcaagaaaatgatggac
					r	gttgttgatgtcgtaggggtgttccag
71	pET15b$_{con}$	FLΔ243-291	150	522	f	aagaacgacctgatgaaccggacc
					r	gtaggggtgttccagcttcttgatggt
72	pET15b$_{con}$	FLΔ246-294	150	522	f	ctgatgaaccggaccttcaagaaaatg

Continued on next page...

No.	Vector	Domain	aa_N	aa_C	Dir.	Sequence
					r	gttgatgtcgtaggggtgttccag
73	pET15b$_{con}$	FLΔ240-288	150	522	f	aacaagaaaaagaacgacctgatgaaccgg
					r	ttccagcttcttgatggtggcgatcag
74	pET15b$_{con}$	FLΔ248-296	150	522	f	aaccggaccttcaagaaaatgatggac
					r	cttgttgttgatgtcgtaggggtgttc
75	pET15b$_{con}$	FLΔ247-295	150	522	f	atgaaccggaccttcaagaaaatgatggac
					r	gttgttgatgtcgtaggggtgttccag
76	pET15b$_{con}$	FLΔ243-291	26	522	f	aagaacgacctgatgaaccggacc
					r	gtaggggtgttccagcttcttgatggt
77	pET15b$_{con}$	FLΔ246-294	26	522	f	ctgatgaaccggaccttcaagaaaatg
					r	gttgatgtcgtaggggtgttccag
78	pET15b$_{con}$	FLΔ240-288	26	522	f	aacaagaaaaagaacgacctgatgaaccgg
					r	ttccagcttcttgatggtggcgatcag
79	pET15b$_{con}$	FLΔ248-296	26	522	f	aaccggaccttcaagaaaatgatggac
					r	cttgttgttgatgtcgtaggggtgttc
80	pET15b$_{con}$	FLΔ247-295	26	522	f	atgaaccggaccttcaagaaaatgatggac
					r	gttgttgatgtcgtaggggtgttccag

[End of table.]

No.	Vector	Domain	aa_N	aa_C	Dir.	Primer
81	pHLsec	FL	26	526	f	aaaccggtgagaacgccatcaagaaaaccaagaaccaggaaaac
					r	ttggtaccgccctgaaaatacaggttctcctgggtcaggggcttgttcttcttgtaggt
82	pHLsec	NTD	68	250	f	aaaccggtaaagagaacatcaagaccaacaacgccaag
					r	ttggtaccgccctgaaaatacaggttctcgtcgtcgttcttgttgttgatgtcgtaggg
83	pHLsec	FL	68	526	f	aaaccggtaaagagaacatcaagaccaacaacgccaag
					r	ttggtaccgccctgaaaatacaggttctcctgggtcaggggcttgttcttcttgta
84	pHLsec	CTD	292	526	f	aaaccggtaagaacgacctgatgaaccggaccttc
					r	ttggtaccgccctgaaaatacaggttctcctgggtcaggggcttgttcttcttgta
85	$pEFbos_{con}$	FL	26	526	f	ccggatccgagaacgccatcaagaaaaccaagaaccaggaaaac
					r	ccgctagcctactgggtcaggggcttgttcttctt
86	$pAcGP67A_{con}$	FL	26	526	f	ccggatccgagaacgccatcaagaaaaccaagaaccaggaaaac
					r	ccaagcttctactgggtcaggggcttgttcttctt

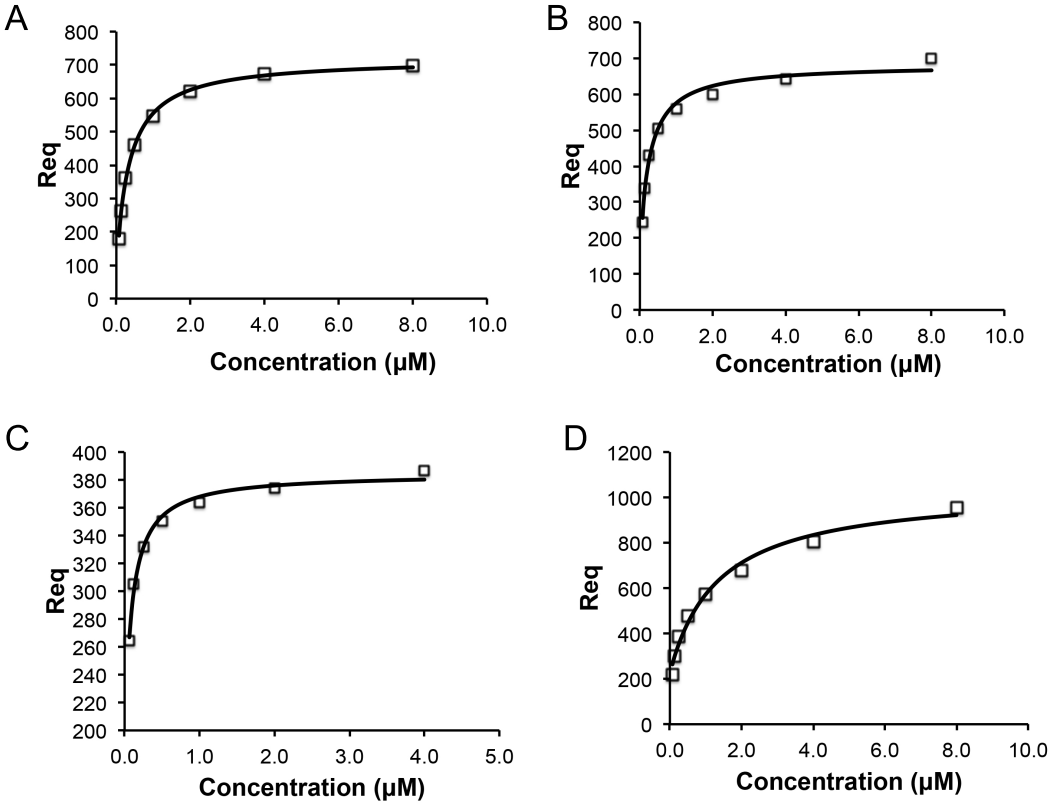
TABLE A.4: Summary of primers for PfRH5 expression in HEK and Sf9 cells. Abbreviations are as in Table A.3.

Hybridoma	Chain	Forward primer	Reverse primer
9AD4	κ	caatctccagcttctttggc	acactcattcctgttgaagctc
9AD4	IgG2a	gtcgagtctgggggaggc	aagagttaggtgctgggcatttg
QA1	κ	gacacagtctcctgcttcc	acactcattcctgttgaagctc
QA1	IgG2a	gtggaatctgggggaggc	aagagttaggtgctgggcatttg

TABLE A.5: Summary of primers for RT-PCR of hybridomas 9AD4 and QA1.

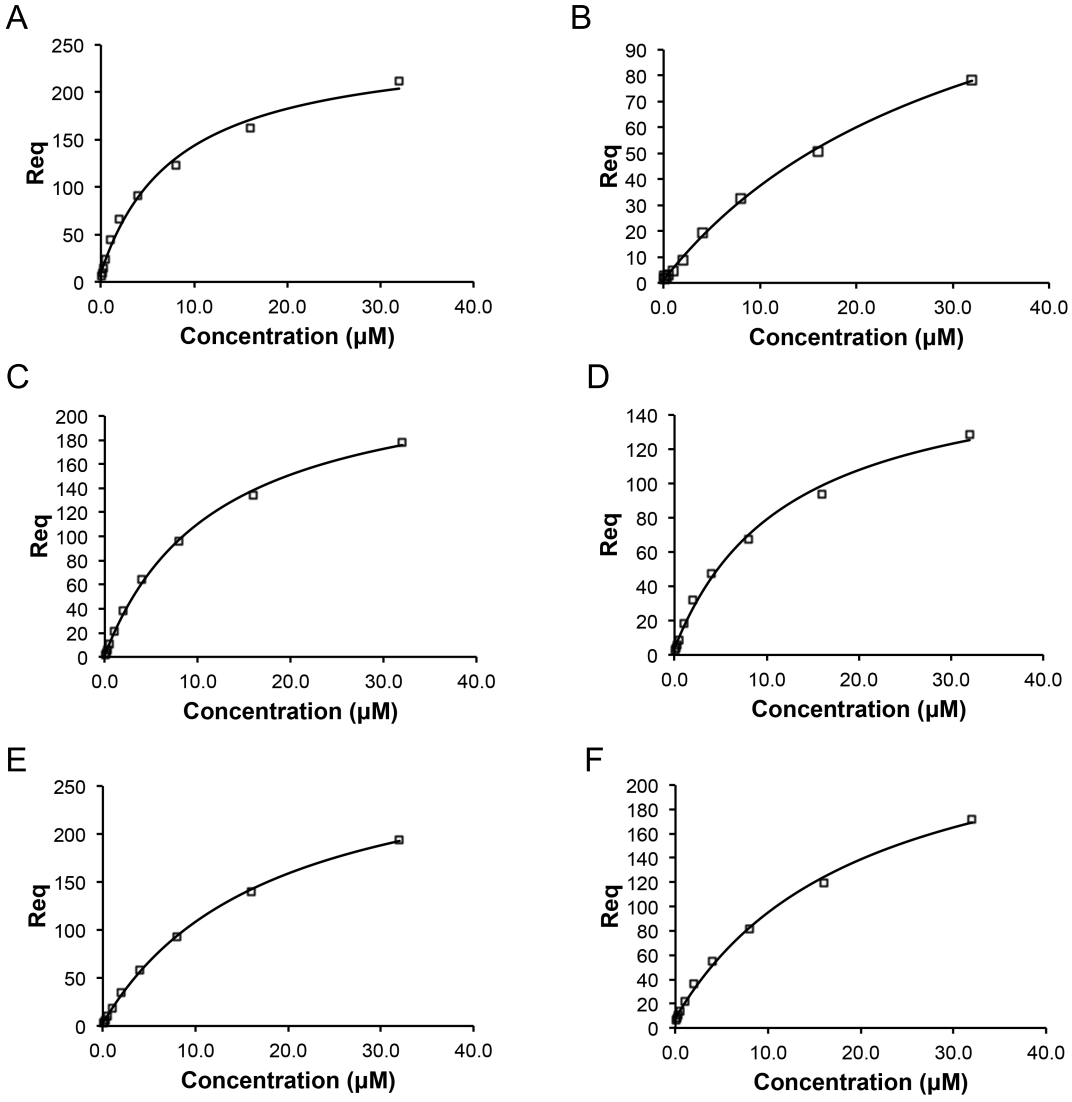
Appendix B

Surface Plasmon Resonance Fit Information



	K_d (μM)	$\mathbf{R}_{max}$ **(RU)**	$\mathbf{R}_{surface}$ **(RU)**
Full-length PfRH5	0.35	625.5	850
RH5ΔN	0.25	532.9	850
RH5ΔNL	0.10	196.9	850
RH5ΔNL methylated	1.4	811.8	850

FIGURE B.1: Steady state affinity fits for the basigin binding of full-length PfRH5 (A), PfRH5ΔN (B), PfRH5ΔNL (C), and methylated and GluC-cleaved PfRH5ΔNL (D). Basigin was bound to the chip surface and the PfRH5 variants flowed over. The table at the bottom indicates the dissociation constant (K_d), the maximum response (R_{max}), and the units of basigin bound to the sensor surface ($R_{surface}$).



	$\mathbf{R}_{max}$ **(RU)**	$\mathbf{R}_{surface}$ **(RU)**
Wild-type	244.1	1300
T28F	155.8	1300
G95F	239.7	1300
H102A	168.5	1300
V131D	297.3	1300
S190D	262.5	1300

FIGURE B.2: Steady state affinity fits for the PfRH5 binding of wild-type basigin (A), and basigin variants T28F (B), G95F (C), H102A (D), V131D (E), and S190D (F). Full-length PfRH5 was bound to the chip surface, and the basigin variants flowed over. A kinetic fit was also performed, with similar results (Figure 5.11D). The table at the bottom indicates the maximum response (R_{max}), and the units of PfRH5 bound to the sensor surface ($R_{surface}$).

Appendix C

Publications Arising From The Thesis

Wright, K. E., Hjerrild, K. A., Bartlett, J., Douglas, A. D., Jin, J., Brown, R. E., Illingworth, J. J., Ashfield, R., Clemmensen, S. B., de Jongh, W. A., Draper, S. J., Higgins, M. K. Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature* (2014). doi:10.1038/nature13715