UNIVERSITY OF OXFORD

Structural insights into innate immunity against African trypanosomes

Harriet Lane-Serff



A thesis presented for the degree of
Doctor of Philosophy

Department of Biochemistry
Keble College
Trinity Term 2017

"Poirot," I said. "I have been thinking."
"An admirable exercise my friend. Continue it."

Agatha Christie, *Peril at End House*

UNIVERSITY OF OXFORD

Abstract

Department of Biochemistry

Keble College

Doctor of Philosophy

Structural insights into innate immunity against African trypanosomes

by Harriet Lane-Serff

The haptoglobin-haemoglobin receptor (HpHbR) is expressed by the African trypanosome, *T. brucei*, whilst in the bloodstream of the mammalian host. This allows acquisition of haem, but also results in uptake of trypanolytic factor 1, a mediator of innate immunity against non-human African trypanosomes. Here, the structure of HpHbR in complex with its ligand, haptoglobin-haemoglobin (HpHb), is presented, revealing an elongated binding site along the membrane-distal half of the receptor. A ~50° kink allows the simultaneous binding of two receptors to one dimeric HpHb, increasing the efficiency of ligand uptake whilst also increasing binding site exposure within the densely packed cell surface. The possibility of targeting this receptor with antibody-drug conjugates is explored. The characterisation of the unexpected interaction between *T. congolense* HpHbR and its previously unknown ligand, haemoglobin, is also presented. This receptor is identified as an epimastigote-specific protein expressed whilst the trypanosome occupies the mouthparts of the tsetse fly vector. An evolutionary pathway of the receptor is proposed, describing how the receptor has changed to adapt to a role as a bloodstream form-specific protein in *T. brucei*. Apolipoprotein L1 (ApoL1) is the pore-forming component of the trypanolytic factors. An expression and purification protocol for ApoL1 is presented here, and the functionality of the protein established. Initial attempts to characterise the pores and structure of ApoL1 are described.

Acknowlegements

I am very thankful for the support I have had throughout this DPhil from colleagues, collaborators, friends and family.

Firstly, I must thank my supervisor, Matt Higgins. Over my five years in this lab I have always been appreciative of the time Matt has taken to offer guidance, training, technical advice and "grand plan" meetings whenever needed. For the wide-ranging opportunities I have been given throughout this project, from exploring new and unexpected experimental avenues, to exploring the beaches of Martha's vineyard, I will be forever grateful.

The work carried out by Mark Carrington's lab is an essential part of the story of this thesis. I am very grateful for the collaboration with this lab, and in particular Paula MacGregor; I am sure that I could not have had a more conscientious, efficient or patient collaborator. Both Paula and Mark's depth of knowledge has been crucial to the development of this project.

Rob Gilbert's insight and enthusiasm have shaped my views on pore-forming proteins, and I would like to thank him for all the expertise he has provided throughout this DPhil. Ed Lowe and David Staunton have been invaluable for troubleshooting crystallographic and biophysical techniques, respectively. For teaching me electrophysiology methods, I would like to thank Mercedes Romero-Ruiz and Hagan Bayley. Thank you to MedImmune, and specifically to Steve Rust and Andrea Gonzalez-Munoz, for their time whilst hosting me and for introducing me to phage display. Thank you also to Lea Barford and Nicky Hamp for performing the immunisation work. I am grateful for the funding I have received from the Wellcome Trust.

For technical advice, friendship and excellent cakes, I must thank all the members of the Higgins lab. In particular, thank you Clinton, Kate, Frank and Natalie for making life in the lab so enjoyable. Jon, for your calm and organised approach to everything, I am very grateful. Finally, thank you to Sebastian, who has been a big part of this project over the last few years, for invaluable discussions, camaraderie and unfaltering optimism.

To my "Wellfies", Verity, Paul and Helen, thank you for all the lunches, dinners, tissue culture room chats and tea breaks. I am also thankful to friends both in Oxford and further afield, from Cincinnati to London, and Manchester to Edinburgh, for all of the adventures we have shared over the last four years.

Chris, Sam and Imogen, thank you for reminding me of life outside of trypanosomes, and for providing a place to come and forget about them. Mum and Dad, for encouraging my passion for maths and science, and for all the phone calls you have answered during this DPhil, thank you.

Richard has been a huge part of this DPhil, helping me through the challenges and celebrating with me the triumphs. Throughout, he has been understanding, encouraging and supportive, and for this, I could not be more grateful.

Contents

Abstract ii

Acknowledgements iii

Table of contents iv

List of Figures ix

List of Tables xii

Abbreviations xiii

1 Introduction 1
- 1.1 Sleeping sickness . . . 1
  - 1.1.1 History and aetiology . . . 1
  - 1.1.2 Past epidemics and current status . . . 2
  - 1.1.3 Transmission . . . 4
  - 1.1.4 Symptoms . . . 4
- 1.2 Prevention and treatment of sleeping sickness . . . 5
  - 1.2.1 Control . . . 5
  - 1.2.2 Diagnosis . . . 6
  - 1.2.3 Drugs . . . 6
  - 1.2.4 Vaccine design . . . 8
- 1.3 African trypanosomes . . . 9
  - 1.3.1 The structure of the trypanosome cell . . . 9
  - 1.3.2 Life cycle . . . 10
  - 1.3.3 The endocytic pathway . . . 12
- 1.4 The cell surface of the bloodstream form trypanosome . . . 13
  - 1.4.1 Variant surface glycoproteins . . . 13
  - 1.4.2 VSG expression and antigenic switching . . . 15
  - 1.4.3 Surface receptors and transporters . . . 17

1.5 The trypanolytic factors of human serum . . . 18
1.5.1 Identification of the trypanolytic factors of human serum . . . 18
1.5.2 Mechanism of trypanosome lysis . . . 19
1.6 Apolipoprotein L1 . . . 20
1.6.1 Identification of apolipoprotein L1 . . . 20
1.6.2 Characterisation of ApoL1 pores . . . 22
1.6.3 Induction of cell death by ApoL1 . . . 22
1.6.4 Other roles of ApoL1 . . . 24
1.7 The haptoglobin-haemoglobin receptor . . . 25
1.7.1 Identification of the high affinity receptor responsible for TLF1 uptake . . . 25
1.7.2 The role of the haptoglobin-haemoglobin receptor . . . 27
1.7.3 The structure of the haptoglobin-haemoglobin receptor . . . 27
1.7.4 Haptoglobin and haptoglobin-related protein . . . 29
1.8 Resistance of *T. brucei* subspecies to NHS . . . 31
1.8.1 The serum resistance associated protein of *T. b. rhodesiense* . . . 31
1.8.2 The haptoglobin-haemoglobin receptor of *T. b. gambiense* . . . 32
1.8.3 Additional mechanisms of *T. b. gambiense* resistance . . . 33
1.9 Project aims . . . 34

2 Materials and Methods **36**
2.1 Cloning . . . 36
2.1.1 Bioinformatic analysis . . . 36
2.1.2 Genes and constructs . . . 37
2.1.3 Cloning vectors . . . 37
2.1.4 Polymerase chain reaction . . . 39
2.1.5 DNA agarose gel electrophoresis . . . 39
2.1.6 Restriction digestion and ligation . . . 40
2.1.7 Sequence- and ligation-independent cloning (SLIC) . . . 40
2.1.8 Transformation and sequencing . . . 41
2.2 Protein expression and purification . . . 41
2.2.1 SDS-PAGE and Western blotting . . . 41
2.2.2 Protein expression in *E. coli* . . . 42
2.2.3 Protein expression using baculovirus-infected *Sf9* cells . . . 46
2.2.4 Purification of haemoglobin . . . 47
2.2.5 Size exclusion chromatography . . . 48
2.2.6 Tobacco etch virus protease cleavage . . . 48
2.3 Antibody production, selection, expression and purification . . . 49
2.3.1 Hybridoma production, selection and expansion . . . 49
2.3.2 Phage display for single chain Fv panning selection . . . 51
2.3.3 Antibody selection . . . 52

2.4 Protein modification . . . 53
2.4.1 Protein methylation . . . 53
2.4.2 Protein biotinylation . . . 54
2.4.3 Fluorescent labelling of proteins . . . 54
2.4.4 Nanodisc production and protein insertion . . . 55
2.5 Biophysical methods . . . 56
2.5.1 Surface plasmon resonance . . . 56
2.5.2 Isothermal titration calorimetry . . . 57
2.5.3 SEC-MALLS . . . 58
2.5.4 Small-angle x-ray scattering . . . 61
2.5.5 Circular dichroism . . . 63
2.5.6 Cryo-electron microscopy . . . 63
2.5.7 Electrophysiology experiments . . . 64
2.6 X-ray crystallography . . . 68
2.6.1 Crystallisation and optimisation . . . 68
2.6.2 Data collection and reduction . . . 69
2.6.3 Phase determination . . . 72
2.6.4 Model building and refinement . . . 74

3 Structural basis of haptoglobin-haemoglobin binding by TbbHpHbR 77
3.1 Summary . . . 77
3.2 Expression and purification of TbbHpHbR and HpHb . . . 78
3.2.1 TbbHpHbR . . . 78
3.2.2 HpSP and HprSP . . . 80
3.2.3 Haemoglobin . . . 85
3.2.4 Haptoglobin-haemoglobin complex formation . . . 86
3.3 Functional validation of HpSPHb and TbbHpHbRt . . . 89
3.3.1 Assessing the affinities of truncated constructs using surface plasmon resonance . . . 89
3.4 Structural characterisation of TbbHpHbRt and TbbHpHbRt:HpSPHb . . . 90
3.4.1 TbbHpHbRt:HpSPHb complex formation . . . 90
3.4.2 Crystallisation . . . 91
3.4.3 Data collection and reduction . . . 94
3.4.4 Phasing and model refinement flow path . . . 95
3.4.5 HpSPHb phasing, model refinement and validation . . . 96
3.4.6 TbbHpHbRt phasing, model refinement and validation . . . 99
3.4.7 HpSPHb:TbbHpHbRt phasing, model refinement and validation . 102
3.4.8 The structure of TbbHpHbRt and TbbHpHbRt:HpSPHb . . . 104
3.4.9 Solution studies of the structure of HpHbR and its ligand using SAXS . . . 111
3.5 Testing the dimeric model of HpHb uptake by TbbHpHbR . . . 115

3.5.1 Solution studies of the stoichiometry of TbbHpHbR:HpHb using SEC-MALLS . . . 115
3.5.2 Trypanosome uptake experiments . . . 116
3.6 Discussion . . . 118

4 Structural and biophysical characterisation of anti-TbbHpHbR antibodies 121
4.1 Summary . . . 121
4.2 Generating a panel of anti-TbbHpHbR antibodies . . . 122
4.3 Biophysical characterisation of a panel of anti-TbbHpHbR antibodies . . . 124
4.3.1 Characterising the binding affinities of anti-TbbHpHbR mAbs . . . 124
4.3.2 Competition assays assessing the location of mAb binding . . . 125
4.4 Structural characterisation of TbbHpHbRt:Fab73 . . . 128
4.4.1 Fab:TbbHpHbRt complex formation . . . 128
4.4.2 Fab:TbbHpHbRt crystallisation . . . 129
4.4.3 Data collection and reduction . . . 130
4.4.4 Phasing, model refinement and validation . . . 131
4.4.5 The structure of Fab73:TbbHpHbRt . . . 133
4.5 Discussion . . . 135

5 Structural basis of haemoglobin binding by TcHpHbR 139
5.1 Summary . . . 139
5.2 Structural characterisation of TcHpHbRt:Hb . . . 140
5.2.1 TcHpHbR expression and purification . . . 140
5.2.2 Crystallisation . . . 142
5.2.3 Data collection and reduction . . . 143
5.2.4 Phasing, model refinement and validation . . . 143
5.2.5 The structure of TcHpHbRt:Hb . . . 146
5.3 Investigating the interaction between TcHpHbR and Hb . . . 149
5.3.1 Solution studies of the interactions of TcHpHbR by size exclusion chromatography . . . 149
5.3.2 Investigating the stoichiometry of the TcHpHbR:Hb interaction in solution . . . 152
5.3.3 Investigating the evolution of the ligand specificity of HpHbRs . . . 153
5.3.4 Investigating the role of HpHbR in *T. congolense* by assessing its expression profile and functionality . . . 156
5.4 Discussion . . . 160

6 Expression, purification and characterisation of Apolipoprotein L1 167
6.1 Summary . . . 167
6.2 Expression and purification of ApoL1 . . . 168
6.2.1 Expression of ApoL1 . . . 168

6.2.2 ApoL1 refolding . . . 171
6.2.3 Expression and purification of ApoL1 truncations . . . 174
6.2.4 Functionality of refolded ApoL1 . . . 177
6.2.5 The effect of detergent on ApoL1 homogeneity . . . 180
6.3 Generation of anti-ApoL1 antibodies . . . 183
6.3.1 Selection of anti-ApoL1 scFvs by phage display . . . 183
6.3.2 Production and selection of anti-ApoL1 mAbs using hybridomas . 186
6.4 Attempts to crystallise ApoL1 . . . 187
6.5 Establishing a method to visualise ApoL1 inserted into nanodiscs . . . 191
6.5.1 Production of nanodiscs . . . 191
6.5.2 Insertion of ApoL1 into nanodiscs . . . 193
6.5.3 Cryo-electron microscopy of nanodiscs . . . 194
6.6 Establishing an assay for the characterisation of ApoL1 pores . . . 197
6.6.1 Experimental setup . . . 197
6.6.2 Conductance of ApoL1 pores . . . 199
6.6.3 Kinetic analysis of ApoL1 pores . . . 200
6.7 Discussion . . . 205

7 Conclusions and perspectives 212
7.1 The haptoglobin-haemoglobin receptor of African trypanosomes . . . 212
7.1.1 Uptake of nutrients and TLF1 by TbbHpHbR . . . 212
7.1.2 A comparison of HpHbR from *T. brucei* and *T. congolense* . . . 213
7.1.3 Targeting HpHbR of *T. brucei* with antibody-drug conjugates . . . 214
7.1.4 The ligand binding specificity of TcHpHbR . . . 216
7.2 Apolipoprotein L1 . . . 217
7.2.1 Establishing a toolbox for the characterisation of ApoL1 . . . 217
7.2.2 Characteristics of the ApoL1 pore . . . 218
7.2.3 Steps towards ApoL1 structure determination . . . 218

A Primer sequences 247

B Publication arising from this thesis 248

List of Figures

1.1 The distribution of HAT in sub-Saharan Africa . . . 3
1.2 The structure of the trypanosome cell . . . 10
1.3 The lifecycles of *T. brucei* and *T. congolense* . . . 11
1.4 The structure of the *T. brucei* VSG . . . 14
1.5 The cell surface of the bloodstream form trypanosome . . . 18
1.6 The uptake and mechanism of action of TLF1 and TLF2 on *T. b. brucei* . . . 24
1.7 The structure of *T. congolense* haptoglobin-haemoglobin receptor . . . 28
1.8 The structure of porcine haptoglobin-haemoglobin . . . 30

2.1 Vector maps of pET-15b and pET-15bcon . . . 38
2.2 Vector map of pAcGP67Acon . . . 38
2.3 Electrophysiology experimental set up . . . 66
2.4 Pipeline for structure determination using molecular replacement . . . 75

3.1 Purification of truncated TbbHpHbR . . . 79
3.2 Expression of SeMet-labelled TbbHpHbR . . . 80
3.3 Designing a haptoglobin SP domain construct . . . 82
3.4 Expression trials for insect cell expression of HpSP and HprSP . . . 84
3.5 Deglycosylation trials of HpSP . . . 85
3.6 Purification of haemoglobin . . . 85
3.7 Purification of HpHb . . . 86
3.8 Purification of HpSPHb . . . 87
3.9 Discovery and removal of HprSP dimers by gel filtration . . . 88
3.10 SPR for the binding of ligands to immobilised TbbHpHbRt . . . 90
3.11 Formation of TbbHpHbRt:HpSPHb . . . 91
3.12 Crystals of HpSPHb . . . 92
3.13 Crystals of TbbHpHbRt . . . 93
3.14 Crystals of TbbHpHbRt:HpSPHb . . . 93
3.15 Ramachandran plot and statistics of the HpSPHb structure . . . 98
3.16 Fluorescence scans for heavy metal and SeMet TbbHpHbRt crystals . . . 100
3.17 Ramachandran plot and statistics of the TbbHpHbRt structure . . . 101
3.18 Ramachandran plot and statistics of the HpSPHb:TbbHpHbRt structure . . . 103
3.19 The structure of the *T. brucei* HpHbR . . . 104
3.20 Analysis of the kink of TbbHpHbR . . . 105
3.21 The structure of TbbHpHbRt:HpSPHb . . . 106
3.22 A missing loop in the structure of HpSPHb . . . 107

3.23 A model of the interaction between HpHbα with TbbHpHbR 108
3.24 Differences between haptoglobin and haptoglobin-related protein 109
3.25 A model for a complex of one HpHb dimer bound to two TbbHpHbRts . 110
3.26 SAXS analysis of TbbHpHbRt, HpSPHb and the complex they form . . . 112
3.27 SAXS analysis of HpHb, alone and in complex with TbHpHbRt 114
3.28 SEC-MALLS data to assess the stoichiometry of TbHpHbRt:HpHb . . . 116
3.29 Uptake of dimeric HpHb and monomeric HpSPHb 117
3.30 A model of the *T. brucei* cell surface 119

4.1 Uptake of mAbs into *T. b. brucei* via HpHbR 123
4.2 Killing of *T. b. brucei* by anti-TbbHpHbR ADCs 124
4.3 SPR curves for binding of Fabs to TbbHpHbRt 125
4.4 Competition SPR for the binding of mAbs to TbbHpHbRt 126
4.5 Competition SPR for the binding of mAbs and HpSPHb to TbbHpHbRt . 127
4.6 Purification of the complex formed between TbbHpHbRt and Fab77 . . . 128
4.7 Crystals of Fab73:TbbHpHbRt 129
4.8 Ramachandran plot and statistics of the Fab73:TbbHpHbRt structure . . . 133
4.9 The structure of TbbHpHbRt in complex with the Fab73 134
4.10 The interactions of TbbHpHbRt with Fab73 and HpSPHb 135
4.11 Interaction hotspots of TbbHpHbRt 137

5.1 Purification of TcHpHbR 141
5.2 Purification of TvHpHbR 141
5.3 Crystals of TcHpHbR:Hb 142
5.4 Ramachandran plot and statistics of TcHpHbR:Hb 146
5.5 The structure of two TcHpHbRs binding to one Hb tetramer 147
5.6 The conservation of TcHpHbR-binding residues across mammalian Hbs . 148
5.7 Analytical gel filtration of Hb with TcHpHbR 150
5.8 Analytical gel filtration of HpSPHb with TcHpHbR 151
5.9 Analytical gel filtration of HpHb with TcHpHbR 152
5.10 ITC data for TcHpHbR titrated into human Hb 153
5.11 SPR of HpHb and Hb binding to HpHbRs 155
5.12 Phylogenetic tree of trypanosome species and subspecies 155
5.13 SPR of bovine Hb binding to TcHpHbR 156
5.14 Western blot analysis of trypanosome TcHpHbR expression 157
5.15 Immunofluorescence analysis of TcHpHbR expression 158
5.16 The uptake of Hb into *T. congolense* epimastigotes 159
5.17 A model of TcHpHbR bound to HpSPHb 161
5.18 The tilt angles required for two TcHpHbRs to bind one HpHb or Hb . . . 162
5.19 Differences in the structures of TbHpHbR and TcHpHbR 163
5.20 The role of HpHbR during the lifecycles of *T. congolense* and *T. brucei* . 164

5.21 Comparison of the ligand binding of TbHpHbR and TcHpHbR 166

6.1 Purification of full length ApoL1 expressed by C43(DE3) cells 169
6.2 Purification of full length ApoL1 expressed by BL21 cells 170
6.3 Western blot assessment of the identity of ApoL1 171
6.4 Gel filtration of ApoL1 without detergent 171
6.5 Refolding trials of ApoL1 172
6.6 Purification of full length ApoL1 173
6.7 Prediction of ApoL1 secondary structure and disorder 174
6.8 Purification of N-terminally truncated ApoL1 by gel filtration 175
6.9 Purification of the PFD and CTD of ApoL1 by gel filtration 176
6.10 Summary of ApoL1 constructs 177
6.11 Circular dichroism of full length ApoL1, PFD and CTD 177
6.12 SPR of methylated SRA binding to ApoL1 178
6.13 MST of SRA binding to ApoL1 179
6.14 Killing of *T. b. brucei* by recombinant ApoL1 179
6.15 Detergent screening of ApoL1 181
6.16 SEC-MALLS of DDM-solubilised ApoL1 and truncations 182
6.17 An example of a phage ELISA result 184
6.18 SPR curves of the binding of a scFv to immobilised ApoL1 constructs . . 185
6.19 Purification of anti-ApoL1 scFv 186
6.20 SPR of mAb binding to ApoL1 constructs 187
6.21 ApoL1 crystallisation attempts 188
6.22 mAb purification and cleavage 189
6.23 The formation of ApoL1:Fab complexes 190
6.24 Microcrystals of ApoL1:Fab 190
6.25 Optimisation of nanodisc assembly 192
6.26 Optimisation of nanodisc lipid composition 193
6.27 Gel filtration of nanodiscs 194
6.28 Cryo-EM micrographs of empty nanodiscs 195
6.29 Single particle nanodisc images 196
6.30 Single channel traces of ApoL1 in a DPhPC bilayer 198
6.31 Single channel traces of ApoL1 in a soybean polar lipid extract bilayer . . 199
6.32 The conductance of ApoL1 pores 200
6.33 Kinetic analysis of the gating of ApoL1 pores 202
6.34 Further kinetic analysis of the gating of ApoL1 pores 202
6.35 Histograms of ApoL1 pore conductance 204
6.36 the structure of barrel stave pores and toroidal pores 209

List of Tables

3.1 Data processing statistics for HpSPHb, TbbHpHbRt and the complex . . . 95
3.2 Data refinement statistics for HpSPHb, TbbHpHbRt and the complex . . . 97
3.3 SAXS parameters for HpSPHb, TbbHpHbRt and their complex 113
3.4 SAXS parameters for HpHb alone and in complex with TbbHpHbRt . . . 115

4.1 Data processing statistics for Fab73:TbbHpHbRt 131
4.2 Data refinement statistics for Fab73:TbbHpHbRt 132

5.1 Data processing statistics for TcHpHbR:Hb 143
5.2 Data refinement statistics for TcHpHbR:Hb 145

6.1 Sequencing of scFv displayed by phage after 4 rounds of selection 183
6.2 The lipid composition of bloodstream form *T. b. brucei* 208

A.1 Table of primers used for cloning 247

Abbreviations

ABC	Dimethylamine-borane complex
ADC	Antibody-drug conjugate
ApoA1	Apolipoprotein A1
ApoL1	Apolipoprotein L1
BAP	Biotin acceptor peptide
BES	Bloodstream expression sites
β-OG	Octyl-β-glucoside
CCP	Complement control protein
CD	Circular dichroism
CHAPS	3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate
CMC	Critical micelle concentration
Cryo-EM	Cryo-electron microscopy
CTD	C-terminal domain
DDM	n-Dodecyl-β-D-maltopyranoside
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPG	1,2-dioleoyl-sn-glycero-3-phospho-1-rac-glycerol
DPhPC	1,2-diphytanoyl-sn-glycero-3-phosphocholine
ELISA	Enzyme-linked immunosorbent assay
ESAG	Expression site-associated genes
ESB	Expression site body
GARP	Glutamic acid/alanine-rich protein
GPI-anchor	Glycosylphosphatidylinositol-anchor
HAT	Human African Trypanosomiasis
Hb	Haemoglobin
HEK	Human embryonic kidney
HDL	High density lipoprotein
HGPRT	Hypoxanthine-guanine phosphoribosyletransferase
Hp, Hpr	Haptoglobin, haptoglobin-related protein
HpSP	Haptoglobin serine protease domain
HpHbR, HpHbRt	Haptoglobin-haemoglobin receptor, truncated
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISG	Invariant surface glycoprotein
ITC	Isothermal titration calorimetry
LDAO	LauryldimethylamineN-oxide

LDL	Low density lipoprotein
LLG	Likelihood gain score
mAb	Monoclonal antibody
MAD	Membrane-addressing domain
MALLS	Multi-angle laser light scattering
MST	Microscale thermophoresis
NHS	Normal human serum
NTD	N-terminal domain
ODC	Ornithine decarboxylase
PARP	Procyclic acidic repetitive protein
PCR	Polymerase chain reaction
PFD	Pore-forming domain
SAXS	Small angle X-ray scattering
scFv	Single chain variable fragment
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLIC	Sequence- and ligation-independent cloning
SPR	Surface plasmon resonance
SRA	Serum resistance associated protein
Tbb, Tbg, Tbr	*Trypanosoma brucei brucei, gambiense, rhodesiense*
TBE	Tris-borate EDTA
Tc	*Trypanosoma congolense*
TEV	Tobacco etch virus
TFZ	Translation function z-score
Tv	*Trypanosoma vivax*
TgsGP	*T. b. gambiense*-specific glycoprotein
TLF1, TLF2	Trypanolytic factor 1, 2
VSG	Variant surface glycoprotein
WHO	World Health Organization

Chapter 1

Introduction

1.1 Sleeping sickness

1.1.1 History and aetiology

The animal disease caused by African trypanosomes, nagana, has been impacting livestock production in Africa for millennia. Reports of a cattle disease resembling nagana date back as far as the second millennium BC [1]. However, the first description of Human African Trypanosomiasis (HAT) does not appear until the 14th century, in a text describing the suffering of the emperor of Mali: "Jata had been smitten by the sleeping illness...The sickness harms the patient and continues until he perishes" [2].

It was not until 1895 that the causative agent of the disease in animals was identified as *Trypanosoma brucei* [3]. When the parasites were first observed in human blood in 1901, they were initially thought to be worms, before later being named a new species, *Trypanosoma gambiense* [4]. In the decade that followed, a second cause of HAT, *T. rhodesiense*, and the two main causative agents of cattle nagana, *T. congolense* and *T.*

vivax, were identified, and the tsetse fly was discovered to be the transmission vector of the parsites [5]. Later, the classifications were altered, placing *T. gambiense*, *T. rhodesiense* and *T. brucei* as subspecies of *Trypanosoma brucei*.

1.1.2 Past epidemics and current status

The beginning of the 20th century saw major epidemics in a number of countries including Uganda, Congo Free State (now the Democratic Republic of Congo) and Cameroon, thought to be caused by the forced movement of populations that occurred at this time [6]. By the mid 1960s the disease was under control, with fewer than 5000 cases reported across the entire continent [7]. Unfortunately, this led to relaxed surveillance, allowing a resurgence of the disease during the 1970-90s. The number of cases has been steadily decreasing for the past two decades, with under 10,000 reported cases in 2009 for the first time in 50 years, and only 2804 reported cases in 2015, the bulk of which (86%) occurred in the Democratic Republic of Congo [7]. However, under-detection must be taken into account, particularly in areas of civil unrest and other disease epidemics, such as Guniea and South Sudan respectively. The number of actual cases is therefore estimated to be somewhere below 20,000, with 65 million people still at risk from the disease [7]. Human infections are most common in rural areas with suitable habitats for the vector, and where the human population has high exposure to tsetse flies through activites such as farming, fishing, hunting and collecting water [6]. *T. b. gambiense* is currently found in 24 countries in West and central Africa, and accounts for 97% of all reported cases. *T. b. rhodesiense* is found in 13 countries in eastern and southern Africa, and accounts for the remaining 3% of cases (Figure 1.1) [7].

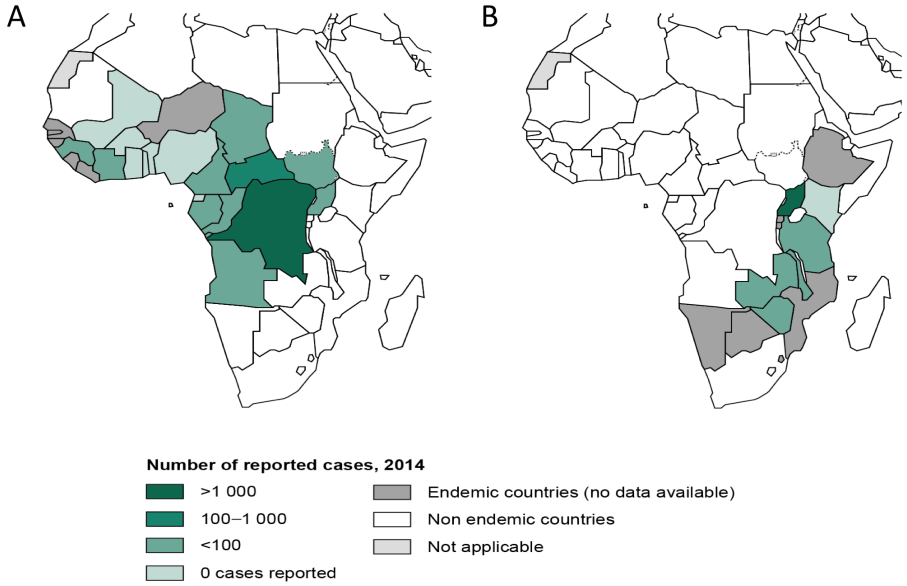


FIGURE 1.1: The distribution of cases of Human African Trypanosomiasis caused by (A) *T. b. gambiense* and (B) *T. b. rhodesiense* in 2014. This figure was adapted from [7].

Unlike HAT, which occurs at focal points, animal trypanosomiasis is widespread over areas where the tsetse fly is found [6]. It is the major constraint on livestock production in sub-Saharan Africa, causing loss of animal productivity and mortality if left untreated. The disease is estimated to be responsible for a loss of US$4.5 billion a year in East Africa alone [8]. Although there are, of course, many similarities between the human-infective species (*T. brucei*) and the main livestock species (*T. congolense* and *T. vivax*), they do differ in many ways, including differences in the surface proteins they express and the progression through lifecycle stages. Research into these species has not progressed as much as for their human-infective counterparts, and caution should be taken when extrapolating results from studies on *T. brucei* to other species [8].

1.1.3 Transmission

African trypanosomes are normally transmitted by the bite of the blood-sucking tsetse fly. *T. b. gambiense* can also be transmitted congenitally, whilst other routes such as blood transfusion and organ donation are poorly documented and rare, but possible [6]. *T. b. gambiense* is transmitted by *Glossina fuscipes* and *G. palpalis*, whilst *T. b. rhodesiense* is transmitted by *G. f. fuscipes*, *G. morsitans* and *G. pallipides* [6]. The flies are infected when they ingest trypanosomes from the blood or skin of mammals. A tsetse fly feeds on average every three days, and is therefore able to infect many people during its 2-3 month lifespan [6].

1.1.4 Symptoms

The symptoms displayed during HAT vary depending on the progression of the infection, the causative subspecies and the host response. During the first, haemolymphatic stage, the parasites are injected into the skin and undergo local multiplication [6]. After a few days the parasites spread via the blood and lymph system. In 88% of *T. b. rhodesiense* cases and 56% of *T. b. gambiense* cases a chancre forms at the site of the initial lesion, characterised by local erythema, edema, heat and tenderness [9]. Symptoms of this first stage include fever, headache, pruritus and lymphadenopathy. In *T. b. gambiense* infections, intermittent fever lasting 1-7 days occurs, separated by intervals of days-months.

To begin the second, meningoencephalitic stage, the parasites penetrate the blood-brain barrier, infecting the central nervous system and causing local inflammation and neurological damage [6]. Symptoms of the second stage include sleep cycle disruption, from which the disease gets its name, paralysis, and progressive mental deterioration. These symptoms lead to death of that patient in almost all cases unless effective treatment is given, although

healthy carriers and natural clearance of *T. b. gambiense* have been described [10].

T. b. rhodesiense causes an acute form of the disease, with progression to the second stage occurring within weeks and death within 6 months [11]. The main reservoirs for *T. b. rhodesiense* are domestic cattle and wild animals [7], and the parasite appears to be less well-adapted for long-term infection of humans. In contrast, *T. b. gambiense* results in a chronic infection, with a mean duration of 3 years [12].

1.2 Prevention and treatment of sleeping sickness

1.2.1 Control

The main control strategy for sleeping sickness is the active screening of at-risk populations, followed by treatment delivery [13]. Surveillance is generally carried out by national control programmes with support from the World Health Organisaiton (WHO). Collecting data on where infections have been detected is crucial to support the planning of control activities. The disease is often under-diagnosed due to limited access to remote areas, insufficient resources and the inaccuracy of detection [6].

As domestic livestock and wild animals are the main reservoir for *T. b. rhodesiense*, control strategies focusing solely on the human population are not as effective as for *T. b. gambiense*. Vector control, mainly through the use of low-cost insecticide-treated screens [14] and insecticide-treated cattle [15, 16], has been attempted. Although attempts to introduce sterile insects showed some success in Zanzibar, the feasibility of using this method continent-wide is debatable, as it would require major infrastructure and resources that are not currently available [17]. WHO has targeted the elimination of HAT as a public health problem by 2020. Although this may be achievable for *T. b. gambiense*, the large

animal reservoir of *T. b. rhodesiense* makes the elimination of HAT a far greater challenge [6].

1.2.2 Diagnosis

As the symptoms of HAT are not specific to this disease, reliable diagnostic tests are required. A serodiagnostic test for *T. b. gambiense*, the card agglutination test for trypanosomiasis, was developed 40 years ago and is based on the detection of specific antibodies [18]. The test is very quick and can be carried out using blood collected from a finger prick, plasma or serum, making it a convenient diagnostic tool for mobile screening teams. However, the test is not 100% specific and many results are false positives [19]. Currently there is no serodiagnostic test for *T. b. rhodesiense* that can be used in the field and infection can only be confirmed by microscope examination of stained blood or a chancre aspirate. For stage determination, which is essential for the correct prescription of treatment, a sample of cerebrospinal fluid collected by lumbar puncture is the only option for both subspecies [20].

Diagnosis of nagana is usually carried out by detection of symptoms in endemic regions [21]. The overall diversity of *T. congolense* and *T. vivax* means that a single diagnostic test to detect nagana is unlikely. It has been suggested that trying to find biomarkers that are associated with trypanosome infection could be a way round this problem, but no significant progress has been made with this to date [8].

1.2.3 Drugs

Five drugs are routinely used to treat HAT, and all are donated by the manufacturers and distributed by WHO and other organisations [22]. Choice of treatment depends on the stage

of infection and the subspecies that is causing the disease. Drugs used to treat second stage infections are generally more toxic, due to the requirement of these compounds to cross the blood-brain barrier.

Pentamidine isethionate has been used to treat first stage *T. b. gambiense* infection since the 1940s. It is a positively charged aromatic diamindine compound which accumulates in trypanosomes cells, where it is thought to disrupt mitochondrial processes, although the exact mechanism remains unclear. It has a good efficacy against *T. b. gambiense* of 95-98% and is generally well-tolerated, with only pain and transient swelling as side effects [23]. It is normally given intramuscularly once daily for 7 days.

The oldest drug used to treat HAT, suramin, is used against first stage *T. b. rhodesiense*. It was first used to treat HAT in 1922, and is a derivative of trypan blue. Although it is also effective against *T. b. gambiense*, it must be administered intravenously over the course of a month, so pentamidine is normally favoured. The side effects are generally mild and reversible, and can include fatigue, anaemia and thrombocytopenia [6]. It is thought to inhibit glycosomal enzymes involved in glycolysis, the only source of energy for the bloodstream form. ISG75, a surface protein with a currently unknown function, is responsible for uptake of suramin by the parasite [24].

Melarsoprol is currently the only treatment available for second stage *T. b. rhodesiense*, and has been in use since 1939. It is an organic arsenical drug and has a high toxicity, with 5-10% of patients experiencing post-treatment reactive encephalopathy [25]. Symptoms of this include seizures, high fever, headaches, nausea and vomiting, and half of patients that experience this die within 48 hours [26]. Increasing drug resistance has also meant that the treatment fails in 10-30% of cases in endemic regions [27].

Nifurtimox-eflornithine combination therapy has been the first-line treatment for second stage *T. b. gambiense* infection since 2009 [28, 29]. It has higher cure rates (95-98%)

and lower fatality rates (less than 1%) than either of the drugs alone. Moreover, the side effects are less severe, it is simpler to administer and is believed to avoid causing drug resistance. Nifurtimox is delivered orally, whilst eflornithine must be administered intravenously. The mechanism of action of nifurtimox is unclear, although it may involve the generation of free radicals [30]. Eflornithine binds the active site of ornithine decarboxylase (ODC), an enzyme involved in polyamine synthesis, necessary for cell division and proliferation [27]. The turnover of ODC is much more rapid in *T. b. rhodesiense*, 4.3 hours versus 18-19 hours, explaining its ineffectiveness against this subspecies [27]. The turnover rate in human cells is 10-30 mins, resulting in a low toxicity [31].

In general, the current drugs used to treat HAT require complex and prolonged administration, can cause severe side effects, and are in some cases are becoming less effective. Although the number of HAT cases is steadily decreasing, the discovery new treatments is therefore still of great importance. Target-based approaches focus on processes that are highly used in trypanosomes, such as endocytosis [32], the enzymes of the glycosome [33], and the glycosylphosphatidylinositol-anchor (GPI-anchor) synthetic pathway [34]. Cellular assays analysing the response of trypanosomes to compounds have identified two compounds which are currently in clinical trial: fexinidazole [35] and oxaborole SCYX-7158 [36]. Hopefully new candidates will be able to overcome some of the disadvantages of currently available HAT treatments.

1.2.4 Vaccine design

There are several problems concerning vaccine design for HAT prevention. Firstly, the parasite is coated in a thick layer of variant surface glycoprotein (VSG) that can be switched to avoid detection by the adaptive immune system (section 1.4.1). Secondly, the high rate

of endocytosis results in the rapid clearance of surface-bound IgG. Lastly, T and B cell responses are thought to be suppressed during trypanosome infection in a process predominantly mediated by macrophages [37, 38]. However, some attempts are still being made to pursue this option.

The two main options for vaccine targets are conserved VSG or invariant surface glycoprotein (ISG) sequences [39]. The former are known to be inaccessible to antibodies, but could still be targeted by Th1 cells. A DNA vaccine containing a single *T. brucei* ISG was shown to be partially protective in mice, dependent on the induction of a Th1 response to the ISG [40]. Although little success has been had so far, attempts to design a vaccine to protect against HAT are thus still being pursued.

1.3 African trypanosomes

1.3.1 The structure of the trypanosome cell

As the trypanosome progresses through the different lifecycle stages, the cell undergoes many changes to adapt to each environment it experiences. However, a number of general features of the trypanosome cell can be identified (Figure 1.2). One key feature of the trypanosome cell is the flagellum, comprised of the conventional 9 + 2 axoneme plus an associated paraflagellar rod structure [41]. The plasma membrane is divided into 3 main domains: the cell body, the flagellum, and the flagellar pocket. The flagellar pocket is an invagination at the base of the flagellum, and most invariant surface proteins are thought to be localised here [42, 43], offering some protection from the extraceullar environment due to the narrowing of the flagellar pocket at the apical end [44]. However, it is up for debate just how protected this environment really is, as large complexes (e.g. TLF1, 500 kDa

[45]) are able to enter this region and be endocytosed.

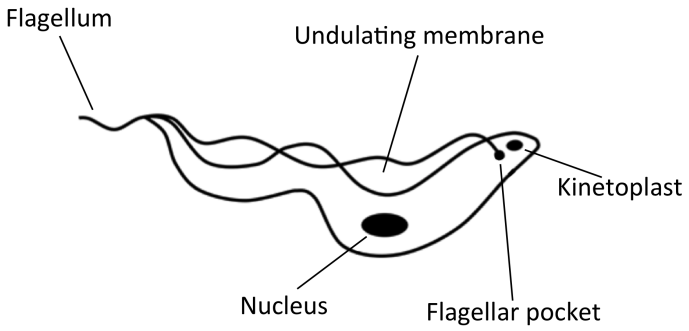


FIGURE 1.2: The structure of the trypanosome cell. This diagram represents the bloodstream form of the parasite due to the positioning of the kinetoplast relative to the nucleus and flagellar pocket, and the slender shape of the cell body. This figure was adapted from [46].

A distinct feature of the kinetoplastid family, of which the trypanosomes are a member, is a dilated region of the mitochondria called the kinetoplast, where kDNA is stored in a network of interlocked rings [47]. Kinetoplast repositioning relative to the nucleus and flagellar pocket occurs during differentiation between lifecycle stages [48].

The trypanosome cell contains peroxisome-related organelles called glycosomes, which contain many of the enzymes required for glycolysis, allowing a greater level of control of this metabolic process. This is the only source of energy for the bloodstream form trypanosome, and glycosomes are essential for parasite survival [49].

1.3.2 Life cycle

African trypanosomes belong to the *Trypanosomatidae* family of parasitic organisms, most of which are commensal rather than pathogenic [5]. Unlike other members of the family, African trypanosomes remain extracellular throughout their life cycle, resulting in a requirement for extra protection against the immune system of the host when in the bloodstream [6]. This comes in the form of a thick coat of a single protein, the VSG, discussed

in section 1.4.1.

The parasites have four main developmental stages: procyclics, epimastigotes, stumpy metacyclic trypomastigotes and slender bloodstream form trypomastogotes (Figure 1.3) [50]. The infected tsetse fly injects stumpy metacyclic trypomastigotes into the mammalian host during a blood meal. Once the parasites reach the bloodstream, they transform into proliferative, slender, bloodstream form trypomasigotes. The trypomastigotes divide via binary fission and can be ingested by the tsetse fly during a blood meal.

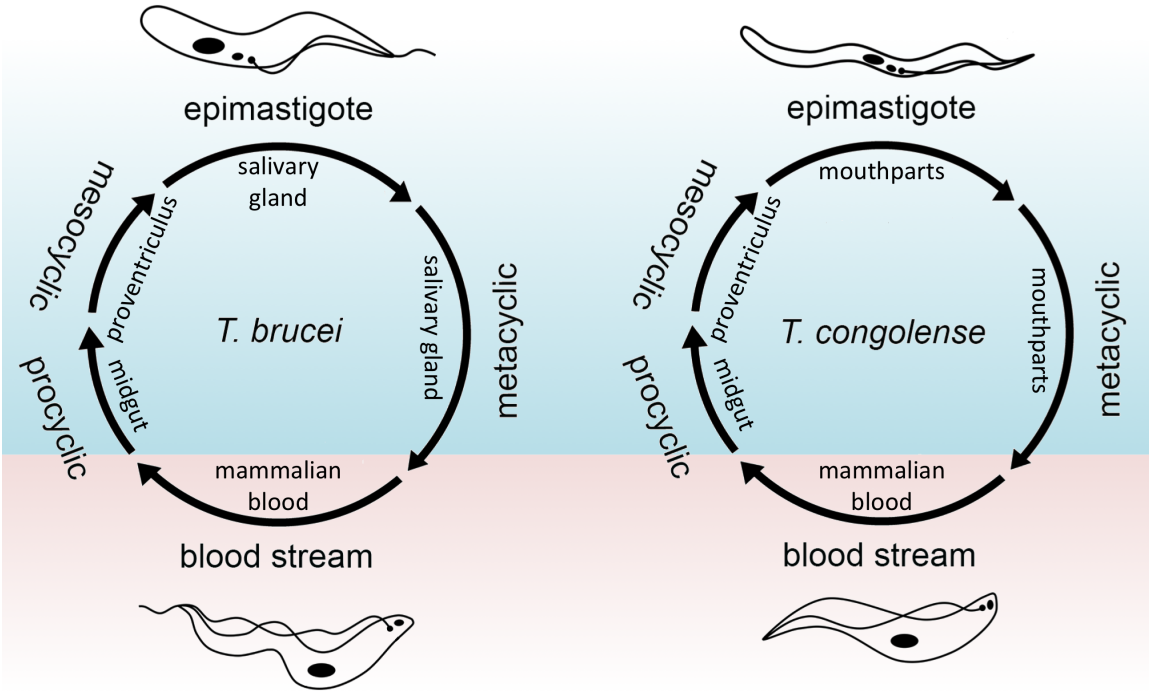


FIGURE 1.3: The lifecycles of *T. brucei* and *T. congolense*. The lifecycles are similar, with the exception of the latter stages of development in the tsetse fly, which occur in different locations. This figure was adapted from [46].

In the midgut of the fly, the parasites transform into proliferative, procyclic trypomastigotes, which no longer express VSG. Instead, the main protein of the cell surface is procyclic acidic repetitive protein (PARP) in *T. brucei* [51] and glutamic acid/alanine-rich

protein (GARP) in *T. congolense* [52], although the function of these is not clear. Here, the life cycles of *T. brucei* and *T. congolense* diverge [53, 54]. *T. brucei* parasites reach the salivary gland of the fly, where they differentiate into epimastigotes following asymmetric division. Epimastigotes divide whilst attached to the salivary gland wall, and some undergo meiosis and sexual exchange [55, 56]. In contrast, *T. congolense* procyclics migrate to the proboscis, part of the mouthparts of the fly, where differentiation into epimastigotes takes place [57, 58].

In both cases, these epimastigotes differentiate into non-proliferative, stumpy, metacyclic trypomastigotes expressing VSG as their major surface protein, and are able to infect a new human host. The trypanosomes reside in the fly for 20-30 days before they have undergone all the changes required to infect humans [59].

1.3.3 The endocytic pathway

Trypanosomes are able to take up macromolecules from the extracellular environment through the endocytic pathway. The parasites have specific receptors on their cell surface to mediate uptake of nutrients such as iron [60] and haem [43]. Endocytosis is thought to be restricted to the flagellar pocket at the anterior of the cell, where cargo is completely internalised in two minutes [61]. Following endocytosis, the cargo is delivered through the early and late endosomes before reaching the lysosomes in under five minutes [62]. Whilst cargo is degraded at this point, most GPI-anchored proteins are recycled via exocytic vesicles, which fuse with the flagellar pocket, thus allowing surface proteins such as VSGs to return to the cell membrane [63, 64].

1.4 The cell surface of the bloodstream form trypanosome

1.4.1 Variant surface glycoproteins

The cell surface of the bloodstream form of the African trypanosome is predominantly covered with a thick layer of a single protein, VSG, which is attached to the plasma membrane via a GPI-anchor. The trypanosome genome contains many different copies of the gene encoding this protein. Switching between the expression of different VSG genes enables the parasite to escape from adaptive immune responses of the host by replacing the antigens against which the response was mounted with different ones that have never been seen by the immune system of the host before. This process is known as antigenic switching. The intermittent fevers associated with the first stage of HAT infection are due to periodic peaks and troughs in parasitemia caused by antigenic switching [6].

VSG molecules make up 10% of the total protein content of the bloodstream form of *T. brucei*, with approximately 5 x 10^6 homodimers per cell, nearing the maximum possible packing density [65]. This dense surface layer is 12-15 nm thick and does not allow assembly of the membrane attack complex, acting as a physical barrier against the innate immune system of the host [66]. The VSGs of *T. brucei* are comprised of two domains. The N-terminal domain (NTD) is responsible for antigenic variation, whilst the C-terminal domain contains the GPI-anchor attachment site [67]. In contrast, *T. congolense* VSGs lack the C-terminal domain, although the GPI-anchor attachment site is still present. The structure of the *T. brucei* VSG NTD has been solved by x-ray crystallography, revealing a homodimer with a height of 100 Å (Figure 1.4) [68]. Each chain is comprised primarily of two α-helices, 70 Å in length, alongside a third, partially helical strand. Several short α-helices form the base at the C-terminus, whilst variable loops cover the membrane-distal end of the VSG homodimer. The short CTD, the structure of which has been determined

by nuclear magnetic resonance [69], is connected to the C-terminus of the NTD via an unstructured linker.

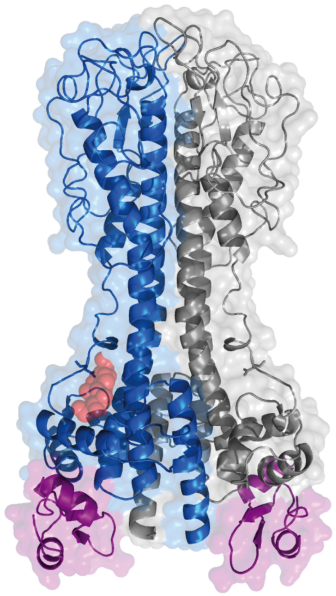


FIGURE 1.4: A model of the structure of *T. brucei* VSG MITat1.2. The two NTDs are shown in blue and grey, whilst the C-terminal domains are purple. An N-linked glycosylation is shown in red. This figure was taken from [69] and is based on the model from [70].

Individual VSGs are not exposed to the extracellular milieu for long, as the entire surface coat undergoes one round of internalisaiton and redistribution to the cell surface in 12.5 minutes [64]. When a VSG molecule is recognised by an antibody, this VSG:antibody complex rapidly relocates to the flagellar pocket, where it is internalised [71]. The antibody is then transported to the lysosome and degraded, allowing the unbound VSG to be recycled to the surface [63, 72]. VSG:antibody complexes are removed from the trypanosome surface at a much faster rate than unbound VSGs, taking only two minutes to fully remove a single layer of surface-bound IgG [73]. This is thought to be due to hydrodynamic force, causing bulkier protrusions from the trypanosome membrane to flow towards the anterior of the cell due to the swimming movement of the trypanosome, the result of the flagellum

[73].

The adaptive immune response to trypanosome infection is predominantly mediated by B cells. IgM antibodies appear earliest in infection and persist the longest, indicating that they are mainly responsible for the clearance of each variant [74]. In agreement with this, VSG-specific IgG does not appear until the first peak of parasitemia is cleared. Due to the dense packing of the protein, antibodies cannot access the conserved regions of the protein nearer the membrane. T cell responses are also only directed at the variable N-terminal regions of VSGs [75].

1.4.2 VSG expression and antigenic switching

VSG expression sites are polycistronic with an upstream promoter driving transcription through a number of expression site-associated genes (ESAGs) before reaching the telomere-proximal VSG gene. Transcription of VSG genes can only occur at one of approximately 15 of these specialised subtelomeric bloodstream expression sites (BESs) [76]. Inhibition studies first indicated that RNA polymerase I was responsible for this transcription, a theory which has subsequently been supported by biochemical studies [77–79].

Originally, the trypanosome genome was thought to contain over 1000 VSG genes [80]. However, once these regions were sequenced it became clear that the majority of these genes are incomplete and require assembly into functional mosaics before transcription can take place [81, 82]. This now puts the estimate of the total number of possible VSG genes encoded by the trypanosome genome far above the previous estimate, nearing infinite [83]. Telomeric VSGs appear to be expressed earlier in infection, whilst mosaic VSGs are only used later, although the mechanism for this preference is not well understood [84]. An

average of 28 different VSGs are estimated to be expressed by a trypanosome population at any single time during infection [84].

Multiple BESs have been identified, posing the question of how the trypanosome ensures only one is used at any one time. The answer is thought to be to do with nuclear location of the actively transcribed expression site. An extranucleolar RNA pol I transcription factory, the expression site body (ESB), is thought to be the location of active transcription of VSG genes [85]. Thus, restricting the association of only one telomeric expression site with the ESB is thought to ensure expression of only one VSG variant at any one time.

Approximately 1 in every 10^3-10^6 cells are estimated to switch VSG expression per population doubling [86]. Altering which BES is positioned at the ESB is thought to be influenced by nucleosomes and histone modification, as the active BES is relatively depleted of nucleosomes, and derepression of BES promoters can be triggered by inhibition of histone deacetylases [87–89].

Antigenic switching can also occur by altering which VSG gene is present in the active BES. This second mechanism can occur via three different processes. Firstly, VSG gene conversion can occur, where a copy of a silent, functional VSG gene replaces the BES-resident VSG gene [90]. A minor mechanism of VSG switching is reciprocal VSG recombination, which involves a crossover between a silenced telomeric VSG and an active one [91]. Lastly, segmental gene conversion can take place, the mechanism by which multiple pseudogenes are pieced together to generate a functional mosaic VSG gene [92, 93].

1.4.3 Surface receptors and transporters

A number of trypanosome surface receptors and transporters have been identified with roles in the uptake of nutrients from the mammalian bloodstream (Figure 1.5). The haptoglobin-haemoglobin receptor is the most well-characterised surface receptor of African trypanosomes [43, 94] and shall be discussed in section 1.7. A receptor for the essential nutrient, transferrin, has been identified [95–97] with ~3000 copies per cell [97]. The transferrin receptor is a heterodimeric protein formed of ESAG6 and ESAG7. The two share a high degree of sequence homology, with the most striking difference being a GPI-anchor attachment site at the C-terminus of ESAG6 [96]. A putative model of the structure of the transferrin receptor has been generated by homology modeling based on the VSG structure [60]. This revealed an elongated structure similar to that of the VSG NTDs. The genome of the trypanosome encodes up to 20 different transferrin receptors, but only one is expressed at any one time. It is thought that this may allow a receptor with specificity for the transferrin of the current mammalian host to be expressed [98].

Trypanosomes require the uptake of lipids for cell growth and division [99]. Endocytosis of both low density lipoprotein (LDL) and high density lipoprotein (HDL) via the flagellar pocket has been observed [99, 100]. A putative LDL receptor has been purified, but the gene has not yet been cloned [99]. Additionally, evidence for a lipoprotein scavenger receptor that can bind both LDL and HDL has been found using competition assays, but the protein responsible for this has not been identified.

A glucose transporter allows the uptake of glucose by the parasite [101]. The genome contains two different isoforms of the glucose transporter [102]. THT1 has a low affinity for glucose and is expressed by the bloodstream form, where glucose is abundant. THT2 has a high affinity for glucose and is expressed by the procyclic trypanosome in the

midgut of the tsetse fly, where glucose is scarce [103]. Both transporters are predicted to be comprised of 12 putative transmembrane α-helices, a common feature of the glucose transporter superfamily [104]. Transport of glucose across the plasma membrane of the bloodstream form is thought to be via facilitative diffusion [105].

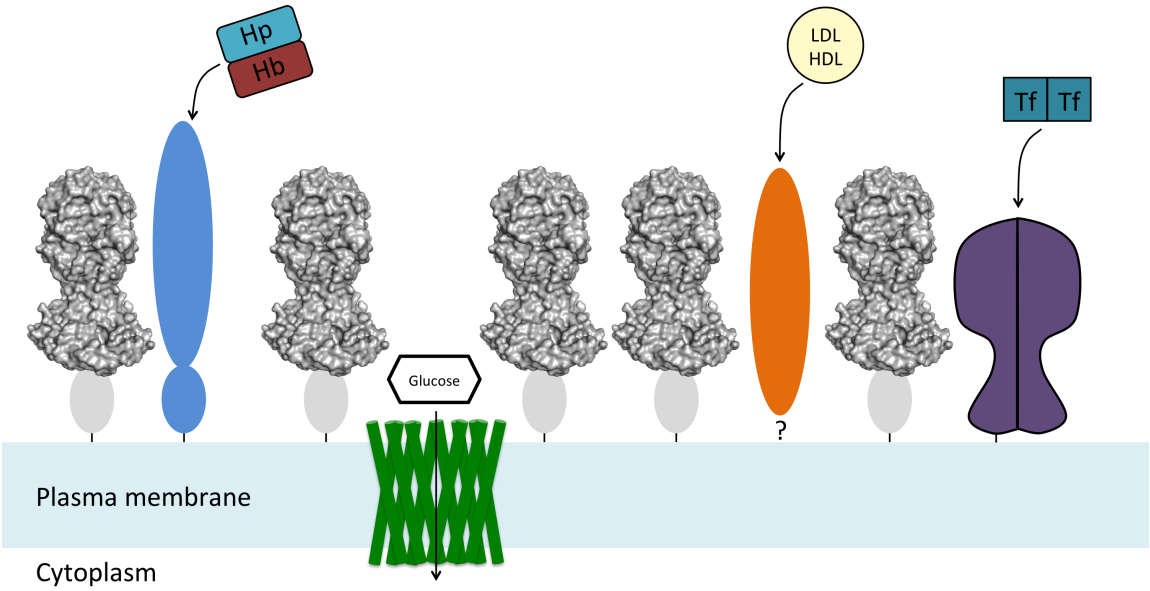


FIGURE 1.5: The cell surface of the *T. brucei* bloodstream form trypanosome showing the known surface receptors and transporters for uptake of nutrients. HpHbR (blue) and VSG (grey) are shown with C-terminal domains, as found in the *T. brucei* proteins. The black line connecting some receptors to the surface represents the GPI-anchor. The transferrin receptor and unknown LDL/HDL receptor are shown in purple and orange respectively. THT1 is shown in green. The surface representation of the VSG uses the x-ray structure of MITat1.2 (pdb:1VSG) [68].

1.5 The trypanolytic factors of human serum

1.5.1 Identification of the trypanolytic factors of human serum

Normal human serum (NHS) causes cell death of non-human African trypanosomes, such as *T. b. brucei* and *T. congolense*, in a process known as trypanolysis. Initial effects include the rapid flux of cations from the extracellular environment across the cell membrane, fol-

lowed by lysosome and cell swelling, fraying of the surface coat and the loss of intracellular organelles, before cell death is completed by lysis [106]. The plasma membrane is able to exclude molecules of 394 Da throughout this process, so must remain intact until lysis occurs [107].

Fractionation of human serum identified high-density lipoprotein (HDL) as the trypanocidal component [108, 109]. In support of this, serum from patients suffering from Tangier disease, which causes severe HDL deficiency, is not able to lyse trypanosomes [108]. This HDL appears to be specific to humans and some other primates, as the serum of other species cannot lyse trypanosomes [108].

Further separation of HDLs by gel filtration identified two different lipoprotein particles with trypanocidal activity: trypanolytic factors 1 (TLF1) and 2 (TLF2) [110]. Both contain haptoglobin-related protein (Hpr), Apolipoprotein A1 (ApoA1) and Apolipoprotein L1 (ApoL1). However, whilst TLF1 is lipid-rich and ~500 kDa, TLF2 is less than 1% lipid and ~1 MDa. Additionally, TLF2 also contains polyclonal IgM, which is not found in TLF1 [111]. The Hpr of TLF1 is bound to haemoglobin (Hb) [112], but it is currently unclear if the same is true for TLF2-bound Hpr [113].

1.5.2 Mechanism of trypanosome lysis

After addition of lytic HDL to *T. b. brucei*, a lag period of 20 minutes is seen before lysis begins, thought to be due to the transport of lytic HDL to its site of action [114]. The pathway along which this HDL is trafficked before lysis begins has been elucidated by incubating trypanosomes at different temperatures [45]. Lytic HDL is initially localised to membrane pits within the flagellar pocket, the site of endocytosis. Next, the particles move to small intracellular vesicles near the flagellar pocket, before fusion to form larger

vesicles, completing the journey through the endosomal system.

Prevention of endosomal acidification using either chloroquine or NH_4Cl inhibits the lytic action of both TLF1 and 2 [110]. This indicates that although the uptake pathways of TLF1 and 2 are different (section 1.7), the mechanism of trypanolysis is shared. Additionally, the vacuolar V-ATPase has been identified in an RNAi screen to find trypanosome proteins involved in cell lysis [115]. RNAi of V-ATPase components reduced the acidity of the endosomal/lysosomal compartments. Together, this data demonstrates that normal acidification of the endosomal pathway is crucial for trypanolysis.

Replacement of extracellular Na^+ (23 Da) with larger ions such as TMA^+ (74 Da) stops both trypanosome swelling and cell lysis caused by HDL, suggesting that the influx of small cations into the trypanosomes seen during the early stages of lysis is essential for cell death [107]. Planar lipid bilayer experiments have demonstrated that the addition of lytic HDL to liposomes results in the formation of cation-selective pores [107]. Replacement of of extracellular Cl^- (35 Da) with the larger gluconate ions (196 Da) also reduced swelling and lysis, indicating that anion influx is also important. Passive entry of Cl^- via Cl^- channels is thought to follow influx of Na^+, resulting in the influx of water responsible for cell swelling [107].

1.6 Apolipoprotein L1

1.6.1 Identification of apolipoprotein L1

ApoL1 was identified as the pore-forming component of trypanolytic HDL due to its interaction with serum resistance associated protein (SRA), the protein responsible for *T. b. rhodesiense* resistance to human serum (section 1.8.1) [116]. Since its identification,

many attempts have been made to elucidate the function of ApoL1 and how it causes lysis of trypanosomes.

Addition of recombinant ApoL1 to *T. b. brucei* below physiological concentrations of 8 μg/ml results in cell death, whilst *T. b. gambiense* and *T. b. rhodesiense* are unaffected [117]. The kinetics of this are delayed compared to that of NHS, probably due to the dynamics of uptake. Depletion of ApoL1 from NHS using SRA or anti-ApoL1 antibodies results in the total loss of lytic activity [116], confirming the importance of ApoL1 in trypanolysis.

A significant amount of work has been carried out to dissect the roles of different domains of ApoL1 [117]. Residues E77-W235 of ApoL1 are predicted to share a structural similarity with the pore-forming domain (PFD) of colicins. Expression of full length ApoL1 in bacteria results in cell death, whilst deletions in the predicted pore-forming domain abolish this toxic activity. Although deletion of residues A238-P304 renders ApoL1 inactive, its function is restored when this region is replaced with pelB, a membrane-targeting signal sequence. This indicates that this region, termed the membrane-addressing domain (MAD), may be responsible for targeting ApoL1 to trypanosome membranes. Deletion of P304-L398 has no affect on bactericidal activity. However, deletion of the PFD, the MAD, or this C-terminal domain (CTD) all result in a protein that is no longer able to lyse trypanosomes [118], demonstrating that all three are important for the function of ApoL1. This work has demonstrated that ApoL1 is the pore-forming component of TLF1 and 2.

1.6.2 Characterisation of ApoL1 pores

Attempts to characterise ApoL1 pores have been made using planar lipid bilayer experiments, using a pH gradient to mimic the lysosome [119]. Addition of ApoL1 to this system results in a conductance of 50-100 pS, which increases up to 3000-fold upon neutralisation of the *cis*-compartment and is reversed upon reacidification. No change in conductance is observed when ApoL1 is added at neutral pH, suggesting that acidic pH is required for initial interaction with the membrane. Additionally, pH neutralisation increases the probability of channel opening, which is also effected by voltage: channel opening was more likely at an applied potential of -80 mV compared to +80 mV, although the importance of this is unclear.

The channels observed in these experiments exhibited selectivity for K^+ over Cl^- ions, supporting the earlier work carried out using lytic HDL. Initial studies using recombinant ApoL1 had suggested that the channels were selective for anions [117]. However, these experiments were carried out with an ApoL1 C-terminal truncation that is not able to lyse trypanosomes, perhaps explaining this difference [119].

1.6.3 Induction of cell death by ApoL1

How these cation-selective pores result in trypanosome lysis is still unclear. Initially it was assumed that ApoL1 inserts into the lysosome membrane, resulting in lysosome swelling, which eventually causes rupture of the cell membrane [117]. However, the pH change required to maximise the conductance of ApoL1 pores indicates that, although insertion is likely to occur in the acidic lysosome, ApoL1 may then be trafficked to a different membrane to trigger cell lysis (Figure 1.6)[119]. One theory proposes that some ApoL1 may be trafficked to the plasma membrane due to high rates of endocytic recycling [119].

Here, it could cause cytoplasmic swelling due to the influx of water and ions across the plasma membrane. This concurs with the influx of cations across the plasma membrane observed during the early stages of trypanosome lysis.

Recently, depolarisation of the mitochondrial membrane was observed during trypanosome lysis [120]. ApoL1 has been detected in the mitochondria of *T. b. brucei*, but not *T. b. gambiense*, albeit at lower levels than is seen in the lysosome. Both recombinant ApoL1 and NHS induce the release of EndoG, a protein known to be released from the mitochondria of unicellular eukaryotes following mitochondrial membrane permeabilisation. RNAi of EndoG conferred resistance to trypanolysis by either NHS or recombinant ApoL1, but it did not stop lysosomal swelling or mitochondrial membrane depolarisation. Another theory has therefore been proposed, in which ApoL1 triggers trypanolysis by inducing mitochondrial membrane permeabilisation [120].

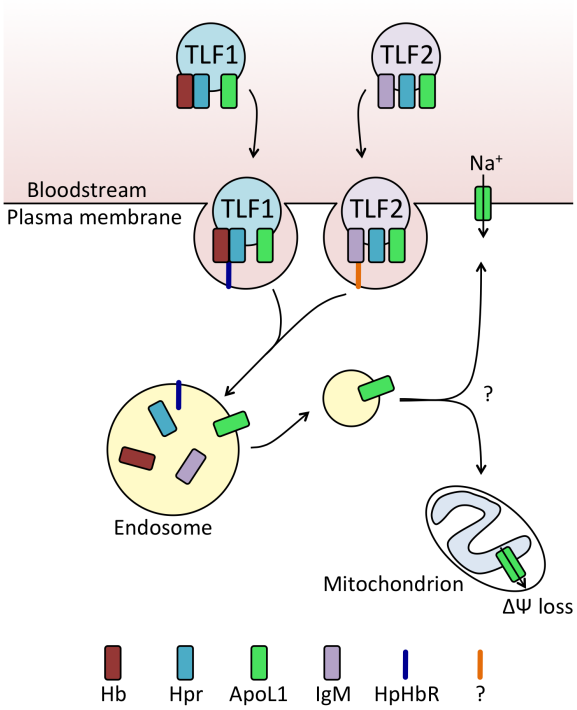


FIGURE 1.6: The uptake and mechanism of action of TLF1 and TLF2 in *T. b. brucei*. TLF1 is taken up via receptor-mediated endocytosis after binding HpHbR. The receptor for TLF2 is unknown, although it has been suggested that the IgM of TLF2 may bind VSG. ApoL1 is thought to insert into the membrane of the endosome or lysosome at acidic pH. Following this, membrane-inserted ApoL1 may traffic to the plasma membrane or mitochondria, where the conductance of the pores increases upon pH neutralisation. After dissociation of the ligand, HpHbR may be recycled to the cell surface, as occurs for VSGs. This figure was based on figures from [113, 119, 120].

1.6.4 Other roles of ApoL1

ApoL1 has been primarily studied due to its role in innate immunity towards many African trypanosomes. However, it is thought that the protein is likely to have other roles within the body. Two forms of kidney disease are associated with sequence variants of ApoL1, G1 and G2 [121]. These variants are common in African chromosomes, but absent from European chromosomes and the proteins they encode are able to lyse *T. b. rhodesiense* in vitro and do not bind SRA. How these variants cause kidney disease is not yet known, but

it is assumed that it is connected with a second role of ApoL1.

ApoL1 expression is induced by IFN-γ and TNF-α, proinflammatory cytokines [122]. Additionally, IFN-γ polarised macrophages that restrict HIV-1 replication also express ApoL1. Cotransfection of 293T cells with ApoL1 and HIV-1 results in the targeting of HIV-1 Gag for degradation. This indicates that ApoL1 may also play a role in the defence against other infectious agents.

Others believe that, due to similarities with the colicin pore-forming domain and eukaryotic Bcl-2 proteins, ApoL1 may be involved in apoptosis [123]. The Bcl-2 family includes anti-, pro-apoptotic and BH3-only protein, which act as sensors and initiators. Pro-apoptotic proteins cause depolarisation and permeabilisation of the mitochondrial membrane, causing release of cytochrome c, which in turn activates caspases. Both anti- and pro-apoptotic Bcl-2 family members have shown to have cation and anion selectivity *in vitro*, indicating that pore forming proteins are involved in apoptosis [124–126]. However, whether ApoL1 is involved in this process remains to be seen.

1.7 The haptoglobin-haemoglobin receptor

1.7.1 Identification of the high affinity receptor responsible for TLF1 uptake

Competition studies first suggested that there are two different receptors for TLF1 on the cell surface of trypanosomes [127]. The first has a high affinity for TLF1 with a low nM dissociation constant and ~350 copies per cell. The second is a low affinity receptor with a low μM dissociation constant and ~60,000 copies per cell. Unlabelled HDL competes for the low affinity receptor, indicating that it is likely to be an HDL receptor that binds to a

lipoprotein such as ApoA1. Uptake of TLF1 by the high affinity receptor is inhibited by haptoglobin-related protein (Hpr) in a dose-dependent manner, suggesting that this component of TLF1 may be involved in uptake, whilst no impact on TLF2 activity is seen. In fact, activity of TLF1 is totally suppressed when Hp is added at the concentration found in the blood, 0.2 mg/ml [110]. This suggests that the activity of TLF1 may only be important in cases where haptoglobin is depleted due to conditions such as haemolysis.

As both Hp and Hpr block uptake of TLF1 by the trypanosome, it was suggested that the high affinity receptor might be a haptoglobin-haemoglobin receptor (HpHbR). Indeed, such a protein was identified from trypanosome cell extracts by affinity chromatography using immobilised HpHb [43]. The gene for this GPI-anchored surface protein was subsequently identified in all kinetoplast species that live solely extracellularly in the bloodstream when in the mammalian host [43]. Disruption of HpHbR expression using knockouts or RNAi both inhibit uptake of HpHb, confirming the function of the receptor. There are an estimated 200-400 copies of HpHbR on the cell surface of *T. b. brucei*, agreeing well with the copy number of the high affinity receptor for TLF1 previously predicted. The protein appears to be localised to the flagellar pocket, the main site of endocytosis.

The absence of HpHbR renders *T. b. brucei* resistant to both NHS and TLF1, but not to TLF2 [43]. Overexpression of HpHbR in these cells caused increased sensitivity to both TLF1 and NHS. TLF1 activity was also blocked by recombinant HpHbR, HpHb and HprHb. Together, this confirms that HpHbR is the main entry route for TLF1, but not TLF2. It has been suggested that the IgM of TLF2 may bind to VSG, resulting in uptake via endocytosis [115].

1.7.2 The role of the haptoglobin-haemoglobin receptor

The affinity of the receptors from *T. congolense* and *T. brucei* for HpHb were found to be 1 μM and 8 μM respectively [94]. This was determined using surface plasmon resonance (SPR) with HpHb coupled to the chip, avoiding the possibility of the resultant affinity being affected by dimeric interactions. Indeed, when the orientation is reversed and the *T. b. brucei* receptor is coupled to the chip, a much higher apparent affinity of 4.5 nM is determined. This is very similar to the values originally predicted for the high affinity receptor of TLF1, suggesting that dimeric binding may also be important in the uptake of TLF1. Neither Hp nor Hb alone were able to interact with TbbHpHbR or be taken up by *T. b. brucei*, indicating that the binding site for the receptor is likely to be made up of both components [43].

The absence of TbbHpHbR slows the growth rate of trypanosomes in mice, whilst *in vitro* growth rates were unaffected. African trypanosomes are deficient in haem biosynthesis, despite expressing a number of haem proteins such as cytochromes P450 and b5 [43]. Haem is undetectable in HpHbR knockout trypanosomes, whilst it is incorporated into the proteins of wild type trypanosomes. The reduced growth rate of knockout cells in mice was still observed in SCID mice (those lacking functional B and T cells), but was alleviated when inhibitors of macrophage oxidative burst enzymes were added. The haem proteins of trypanosomes may therefore play a role in protection against oxidative stress induced by macrophages.

1.7.3 The structure of the haptoglobin-haemoglobin receptor

The structure of HpHbR from the animal trypanosome, *T. congolense*, has been determined by x-ray crystallography (Figure 1.7) [94]. TcHpHbR has an elongated structure, 112 Å

in length, formed primarily of a three-helix bundle. At the membrane-distal N-terminal end the receptor widens to form a compact head structure formed of three small α-helices and a connecting loop. The structure of the receptor is maintained by a core of conserved hydrophobic residues and a single disulphide bond. This architecture shows similarities with the N-terminal domain of the *T. brucei* VSG, which is comprised of two long helices and a third partially helical strand [68]. Additionally, GARP, the main surface protein of the *T. congolense* procyclic trypanosome found in the insect host, shares a similar basic structure [128].

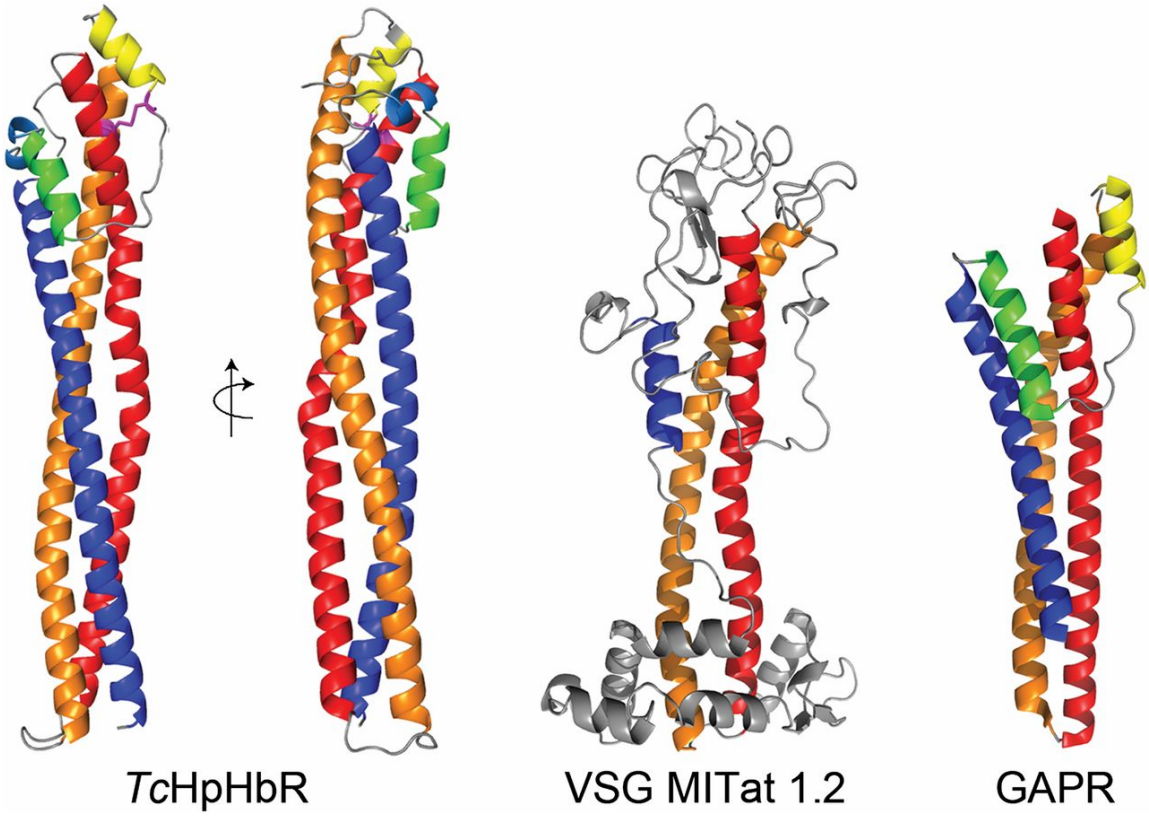


FIGURE 1.7: The structure of *T. congolense* haptoglobin-haemoglobin receptor (pdb:4E40). The receptor is comprised of an extended three-helix bundle and a compact head structure at the membrane-distal end. A conserved three-helical architecture can be seen between TcHpHbR, the N-terminal domain of VSG MITat1.2 (pdb:1VSG) and GARP (pdb:2Y44). This figure was taken from [94].

One main difference between *T. brucei* and *T. congolense* HpHbR can be gleaned from the sequence alone, which shows that TbbHpHbR has an additional 90 residues, predicted to form a C-terminal domain. This is another shared feature with the VSGs of these species: only the *T. brucei* VSG has a C-terminal domain of 70 additional residues. The lengths of these proteins determined by small angle x-ray scattering (SAXS) reflect these differences [94]. TcHpHbR has a D_{MAX} of 120 Å, versus 180 Å for TbbHpHbR. These dimensions suggest that TbbHpHbR will protrude above the VSG surface, which has a D_{MAX} of 130 Å.

The binding site for HpHb on TcHpHbR was identified by mutational studies of a conserved patch of residues 25 Å from the top of the receptor [94]. Mutation of one of the equivalent residues in TbbHpHbR also abolished binding of HpHb, indicating that the binding site is conserved. A position near the membrane-distal end of the receptor is not a surprising location for a ligand binding site, as this region is more likely to be more accessible in the context of the densely packed VSG coat.

1.7.4 Haptoglobin and haptoglobin-related protein

One component of the ligand for HpHbR, haptoglobin, is a protein synthesised in the liver and found in the plasma of mammals and a number of other species [129]. When haemoglobin is released into the plasma during haemolysis, it is bound by haptoglobin via an almost irreversible interaction, with a dissociation constant of 10^{-15} M [130]. The haem ring of haemoglobin can cause the generation of free radicals, which may result in oxidative damage [131]. The role of haptoglobin is therefore to protect against the possible damaging results of haemolysis. HpHb is scavenged by CD163 on macrophages, which mediates the endocytosis of HpHb but cannot bind Hp or Hb alone [132]. SAXS of this

receptor with HpHb indicates that two receptors can simultaneously bind one HpHb dimer, increasing the avidity of the interaction [133].

There are two main alleles of haptoglobin, which determine the multimerisation of the protein. HP1 forms dimers, whilst HP2, which is the result of a partial gene duplication of the complement control protein (CCP) domain, forms higher order polymers [134]. Although haptoglobin is synthesised as a single chain, the precursor is cleaved to release a single arginine residue, resulting in an $\alpha\beta$ heterodimer [129]. The structure of porcine HP1-1 in complex with haemoglobin has been solved, revealing a barbell shaped protein (Figure 1.8) [133]. The CCP domains of each molecule dimerise through a beta-strand swap, whilst the serine protease (SP) domains at either end interact with both subunits of the Hb $\alpha\beta$-dimers on the surface where these dimers would interact to form a haemoglobin tetramer.

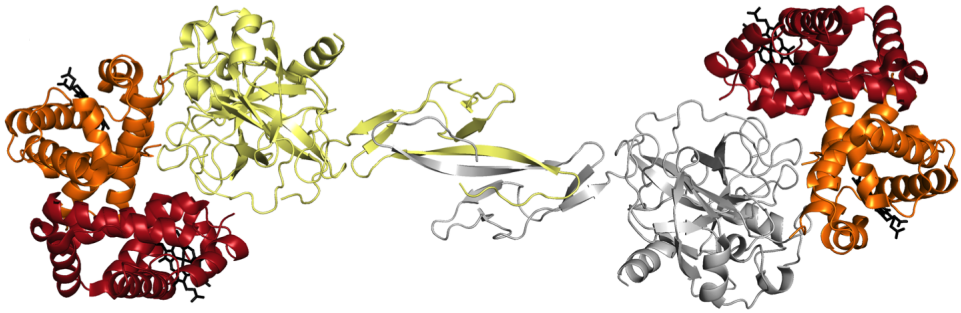


FIGURE 1.8: The structure of porcine haptoglobin-haemoglobin (pdb:4F4O) [133]. Hp molecules are in yellow and grey, haemoglobin α in orange and β in red. The CCP domains of haptoglobin can be seen in the centre of the structure forming a β-strand swap, whilst the SP domains at each end form all interactions with the haemoglobin dimer.

Haptoglobin-related protein is found in the blood at levels 50-fold lower than Hp [135]. The two share 90% sequence identity, and the binding site for Hb is predicted to be conserved, supported by the finding that the affinity is very similar. Hpr is thought to exist predominantly as an $\alpha\beta$-dimer, although some evidence for polymerisation has been seen.

The protein does not bind CD163, so cannot be cleared from the blood in the same way as HpHb [135].

1.8 Resistance of *T. brucei* subspecies to NHS

The two human-infective trypanosome species, *T. b. rhodesiense* and *gambiense*, are able to evade lysis by the trypanolytic factors. This can be accomplished by reducing uptake, or inhibiting the activity or trafficking of ApoL1 once inside the cell. The two subspecies have different mechanisms to accomplish this.

1.8.1 The serum resistance associated protein of *T. b. rhodesiense*

Comparisons of the mRNAs found in strains of *T. b. rhodesiense* that were resistant and sensitive to NHS led to the identification of the serum resistance associated protein (SRA) [136]. The abundance of the mRNA for this protein correlates with lytic HDL resistance [137]. The protein product of this gene is only detected in resistant strains of *T. b. rhodesiense*, and transfection of SRA into *T. b. brucei* is sufficient to confer resistance to NHS [138].

SRA is a VSG-like protein with an internal deletion of 126 amino acids [139]. This deletion occurs in the loops of the protein, and is therefore not thought to affect the overall structure of the two long helices common to all VSGs. Affinity chromatography using recombinant SRA and NHS identified ApoL1 as a binding partner and the two have since been shown to localise together in the lysosome [116].

An orthologue of ApoL1 was found in baboon serum, which is able to lyse *T. b. rhodesiense*, but not *T. b. gambiense* [118]. Baboon ApoL1 has two critical point mutations when compared to human ApoL1, with two residues replaced with lysines [140]. These

mutations are sufficient to prevent the binding of human ApoL1 to SRA *in vitro* and confer resistance of transfected mice to both *T. b. rhodesiense* and *T. b. brucei*. Together, this indicates that a direct interaction between ApoL1 and SRA inhibits ApoL1 function.

Inclusion of SRA in planar lipid bilayer experiments studying ApoL1 pore formation (section 1.6.2) has no effect on the small conductance seen upon ApoL1 addition [119]. However, with SRA present, no amplification of conductance is seen upon neutralisation of the *cis*-compartment. If the G2 variant of ApoL1 is used, which does not bind SRA, the normal increase in conductance is observed. This suggests that SRA does not inhibit the initial interaction of ApoL1 with the membrane, but does not allow ApoL1 pores to become fully functional upon neutralisation.

1.8.2 The haptoglobin-haemoglobin receptor of *T. b. gambiense*

A polymorphism in *T. b. gambiense* HpHbR was identified, L210S, that is unique to human-infective type 1 *T. b. gambiense* [94, 141]. This mutation reduces the K_D of the receptor for HpHb to 30 μM, whilst no binding of TLF1 is detected at near-physiological concentrations (0.5 μM) [94]. Based on the structure of TcHpHbR, L210 lies in the hydrophobic core of the receptor, just behind the binding site for HpHb. It is likely that replacing this leucine with a serine will disrupt the hydrophobic packing of the core, altering the binding site conformation. In addition to the effect of this polymorphism, the abundance of TbgHpHbR mRNA is 5-fold less than that of *T. b. brucei* or *T. b. rhodesiense* [142]. This suggests that a reduction in the copy number of TbgHpHbR may also contribute to reduced TLF1 uptake.

1.8.3 Additional mechanisms of *T. b. gambiense* resistance

Human-infective type 1 *T. b. gambiense* is resistant to killing by recombinant ApoL1, which is taken up by fluid phase uptake [143]. This subspecies must therefore have additional mechanisms of resistance to lysis that do not rely upon reduced uptake of TLF1. An attempt to find the protein responsible for this was carried out by searching for a truncated VSG unique to infective *T. b. gambiense*, similar to SRA of *T. b. rhodesiense* [144]. *T. b. gambiense*-specific glycoprotein (TgsGP) was identified, a protein predicted to have high structural similarity to the NTD of a *T. brucei* VSG and with a putative GPI-anchor site. The gene for this protein was found in all human-infective type 1 *T. b. gambiense* strains, but in no other strains or subspecies. TgsGP is a bloodstream form-specific protein, and is predominantly located in the flagellar pocket and endocytic compartments [144, 145].

A TgsGP knockout cell line was unable to survive in the presence of NHS or recombinant ApoL1, but could survive in non-lytic serum and TLF1 [145, 146]. NHS-mediated lysis of these cells was linked to swelling of endosomal vesicles, and could be inhibited by chloroquine [145]. *T. b. gambiense* expressing TbbHpHbR are able to grow in NHS or recombinant ApoL1, whilst growth is static in the presence of TLF1 or in hypohaptoglobinaemic serum, which favours TLF1 uptake [145, 146]. Therefore TgsGP-mediated resistance to ApoL1 can be bypassed by efficient TLF1 uptake, explaining why TgsGP alone is not sufficient to confer resistance of *T. b. brucei* to NHS, TLF1 or recombinant ApoL1 [145]. A combination of TgsGP and reduced TLF1 uptake appear to be required for protection from NHS. It has been suggested that the downregulation of TbgHpHbR may have been required for survival of *T. b. gambiense* in areas where malaria is prevalent, a disease that causes hypohaptoglobinaemia due to excessive haemolysis [145].

The mechanism through which TgsGP itself functions is still unclear. Although TgsGP

and ApoL1 have been found trafficking together from the flagellar pocket through the endocytic pathway, no interaction between the two has been seen *in vitro* [145]. In liposomes, TgsGP has been shown to increase membrane curvature. One theory is that this may inhibit ApoL1 insertion into membranes, although more work is required to test this [145].

1.9 Project aims

The haptoglobin-haemoglobin receptor of African trypanosomes is a protein of great interest. Firstly, it is a cell surface protein responsible for the uptake of nutrients by the trypanosome. How cell surface receptors function in the VSG layer whilst evading the immune system of the host is not well understood. Perhaps more important, however, is the role of HpHbR in the uptake of TLF1, one of the lipoprotein particles responsible for the innate immunity of humans to non-human African trypanosomes. The first aim of this project, therefore, was to determine the structure of this receptor in complex with its ligand, haptoglobin-haemoglobin. This would elucidate how a trypanosome receptor is able to efficiently take up nutrients in the context of the VSG layer, whilst additionally providing insight into the uptake of the innate immunity lipoprotein particle, TLF1.

Another aim of the project was to better characterise the haptoglobin-haemoglobin receptor of *T. congolense*. The body of knowledge on this species is unsurprisingly much smaller than on the human-infective species. However, the impact of *T. congolense* and animal nagana on livestock production is considerable and the importance of this disease should not be overlooked. Additionally, it is critical that assumptions are not made when extrapolating results from *T. brucei* to other species. I therefore aimed to determine the structure of TcHpHbR in complex with its ligand, as well as expanding the knowledge of the role of this receptor in this species.

Apolipoprotein L1 is the pore-forming component of TLF1 and 2, and causes lysis of non-human trypanosomes. The mechanism of action of ApoL1 is still unclear, and some of the current data is conflicting. I therefore wished to characterise this protein structurally and biophysically, in order to determine how it forms pores in trypanosome membranes, and how these pores lead to cell death.

Chapter 2

Materials and Methods

2.1 Cloning

2.1.1 Bioinformatic analysis

Before designing constructs, protein sequences were analysed to determine which boundaries to use. Secondary structure was predicted using JPred [147], whilst disorder was predicted using RONN [148]. Other information was obtained from the literature [117, 149]. Published structures available from the protein databank were viewed using Pymol [150]. To predict surfaces of a protein structure that were likely to form interactions, JET2 was used [151]. This predicts areas which are likely to form interactions based on sequence conservation, the propensity of amino acids to be present at an interface and the local geometry of the surface.

2.1.2 Genes and constructs

The genes for the *T. congolese*, *T. brucei brucei* and *T. vivax* haptoglobin-haemoglobin receptors had previously been amplified from genomic DNA (Prof. Mark Carrington, Cambridge). Mutants of these, including the point mutation found in *T. brucei gambiense*, were made by site-directed mutagenesis (SDM). Haptoglobin and haptoglobin-related protein serine protease domain genes were produced synthetically after codon optimisation for expression in insect cells (Life Technologies). The full length ApoL1 gene was produced synthetically (Life Technologies) after codon optimisation for either *E. coli* or human embryonic kidney (HEK) expression. Other ApoL1 constructs were produced by polymerase chain reaction (PCR) using these synthetic genes as templates.

2.1.3 Cloning vectors

For protein expression in *E. coli* either the pEt-15b vector (Novagen, Germany) or a modified version with the thrombin cleavage site replaced with a tobacco etch virus (TEV) protease cleavage site and several additional restriction enzyme sites (Prof. Matt Higgins, pEt15bcon) was used (Figure 2.1).

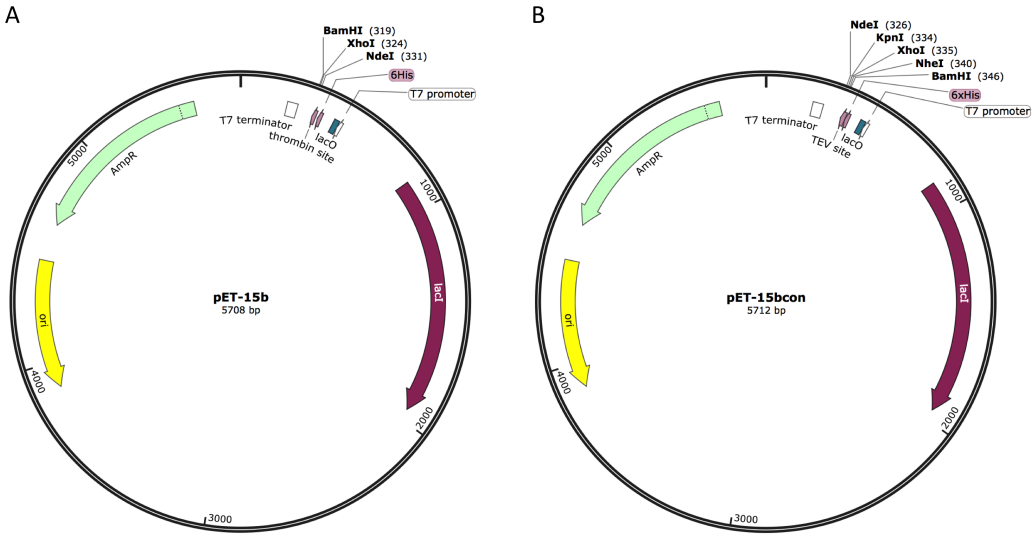


FIGURE 2.1: Vector maps of (A) pET-15b (Novagen) and (B) pET-15bcon, a modified version created by Prof. Matt Higgins. The vector maps were created using SnapGene (GSL Biotech).

For protein expression in *Sf9* cells, a modified pAcGP67A vector was used with the thrombin cleavage site replaced with a tobacco etch virus protease cleavage site (Prof. Matt Higgins) (Figure 2.2).

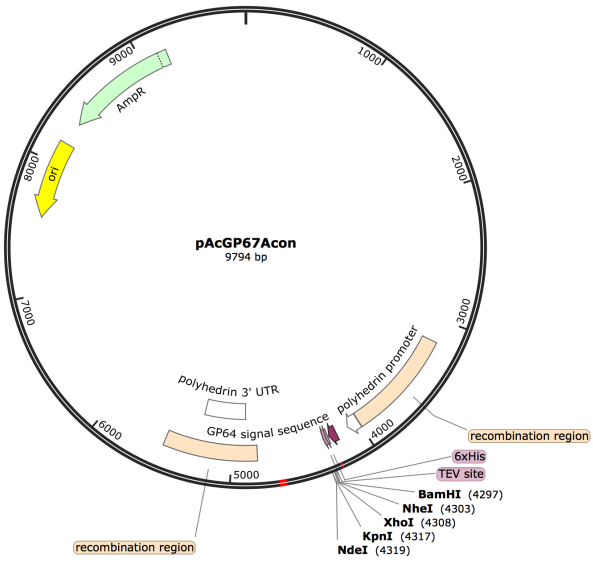


FIGURE 2.2: Vector map of pAcGP67Acon, a vector modified by Prof. Matt Higgins from pAcGP67A (BD Biosciences). The vector map was created using SnapGene (GSL Biotech).

2.1.4 Polymerase chain reaction

Where shorter constructs or the addition of tags was required, PCRs using specifically designed primers were carried out. Primers were commercially produced (Sigma-Aldrich). The protocol used KOD HotStart polymerase (EMB Millipore) and was carried out according to the manufacturer's instructions. The reaction mix contained 50 ng plasmid DNA, 0.4 μM of the forward and reverse primers, 1 mM $MgSO_4$, 0.2 mM dNTPs and 1 U of KOD HotStart DNA polymerase in KOD HotStart buffer (EMB Millipore). A typical reaction was comprised of an initial polymerase activation step at 95 °C for 3 min followed by a cycle of denaturation at 95 °C for 1 min, primer annealing at 45-55 °C for 1 min and extension at 70 °C for 1 min, repeated 30 times. A final extension stage of 3 min at 70 °C was included. For SDM reactions, the extension step was increased to 6 min. Following SDM PCRs the reaction product was treated with Dpn1 (NEB) for 5 hr at 37 °C before transformation. A list of primers used is shown in Appendix A.

2.1.5 DNA agarose gel electrophoresis

0.8% agarose gels were made by dissolving agarose in Tris-borate EDTA (TBE) buffer containing 89 mM Tris-borate, 2 mM EDTA pH 8. SYBR Safe (Thermo Fisher Scientific) was added 1:10,000 to the gel before setting. Samples were loaded using gel loading dye (NEB). Gels were run in TBE buffer at 100 mA for 45 min, before visualisation using a G-Box Chemi-XRQ imaging system (Syngene). 100 bp or 1 kbp DNA HyperLadders (Bioline) were used to assess the size of DNA bands. Where DNA extraction was required, the bands were cut from the gel using a clean razor blade whilst illuminated using a Dark Reader (Clare Chemical Research). DNA was then extracted using the QIAquick gel extraction kit (Qiagen).

2.1.6 Restriction digestion and ligation

Restriction enzyme double digests were carried out according to the manufactures protocols (NEB). Reactions were incubated at 37 °C for 4 hr (PCR products) or 16 hr (vectors), before purification by gel extraction (Qiagen) from a gel slice for vectors, or PCR purification (Qiagen) for PCR products. Before purification, cut vectors were treated with FastAP (Thermo Fisher Scientific) for 30 min at 37 °C. Ligations were completed using T4 DNA ligase (NEB) according to the manufacturer's instructions. Insert and vector were mixed in a 3:1 molar ratio with a total of 90 ng DNA, and incubated with 400 U T4 DNA ligase in T4 DNA ligase buffer (NEB) at 16 °C for 16 hr.

2.1.7 Sequence- and ligation-independent cloning (SLIC)

For N-terminally truncated ApoL1 and the pore-forming domain of ApoL1, several attempts at cloning by ligation failed. Instead, SLIC was used to insert the gene into the correct vector [152]. Primers were comprised of 25-30 nucleotides of the vector at one side of the gene insertion site, and 25-30 nucleotides of one end of the gene. A 2-step PCR was carried out due to the long length of the primers, so both primer annealing and DNA elongation were carried out at 72 °C. The PCR product and vector were both treated with T4 polymerase (NEB) for 30 min at room temperature. When no dNTPs are available, the exonuclease function of the polymerase is active and the ends of the DNA are trimmed to create ssDNA 5' overhangs. dCTP is added to terminate the reaction. The vector and insert were mixed in a 1:1 ratio and annealed in the presence of RecA, which increases the likelihood that annealing is successful, for 30 min at 37 °C. The product was transformed into library efficient DH5α *E. coli* cells (NEB) (Section 2.1.8).

2.1.8 Transformation and sequencing

Transformations for DNA production were generally completed using chemically competent DH5α cells (Life Technologies). SDM reactions were transformed into XL10-Gold ultracompetent cells (Stratagene), and SLIC cloning products were transformed into library efficiency DH5α cells (Life Technologies). *E. coli* strains used for expression are discussed in Section 2.2.2. 1 μl of the relevant plasmid, ligation mixture or SDM reaction was added to the cells, before incubation on ice for 30 min. Heat shock was then performed at 42 °C, generally for 45 s (time was varied based on manufacturer's instructions for individual strains), before incubation on ice for a further 2 min. 100 μl of Lysogeny Broth (LB) was added to the cells, which were then shaken at 37 °C for 1 hr. Cells were plated out on agar containing antibiotics (50 μg/ml ampicillin for all cell types, with an additional 35 μg/ml chloramphenicol for Origami B(DE3)pLysS cells) and incubated at 37 °C for 16 hr. For production of DNA, single colonies were picked and added to 3 ml of LB with antibiotics, and incubated shaking for 16 hr at 37 °C. After DNA extraction using a Miniprep kit (Qiagen), success of cloning was confirmed by Sanger sequencing (Source Bioscience).

2.2 Protein expression and purification

2.2.1 SDS-PAGE and Western blotting

Biphasic sodium-dodecyl sulphate polyacrylamide gels comprised of a 4% polyacrylamide stacking gel and a resolving gel of 12% or 15% polyacrylamide were run using a Mini-PROTEAN gel system (Bio-Rad). Samples were mixed with 6x loading dye containing 100 mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blue, 20% (v/v) glycerol,

200 mM β-mercaptoethanol and boiled at 95 °C for 4 min before loading alongside PageRuler prestained protein marker (Thermo Fisher Scientific). A running buffer of 30 mM Tris pH 7.5, 192 mM glycine, 0.1% SDS was used. Gels were run at 200 mV for 40 min and stained using InstantBlue (Expedeon).

For Western blots, gels were run in the same way but staining was not carried out. Instead, transfer to an Immobile PVDF transfer membrane (Merck Millipore) was carried out using a semi-dry Western blotting system (Bio-Rad) for 1 hr at 50 mA. The membrane was blocked for 1 hr at room temperature using protein-free blocking buffer (Pierce). The primary antibody was then diluted 1:5000 into this blocking buffer and another 1 hr incubation step was completed at room temperature. If a secondary antibody was required, incubation with this antibody was carried out in the same way as the primary, with a washing step in between the two. Three 10 min washing steps were then completed using 20 mM Tris pH 8, 500 mM NaCl, 0.2% Tween-20, 0.05% Triton x-100. Where the final antibody was conjugated to HRP, signal was detected by incubation with ECL Western blotting substrate (Pierce) followed by exposure of the film to Amersham Hyperfilm ECL (Pierce) for 1-30 s and development of the film using an automatic film processor (Eastman Kodak). For detection of biotinylation, phosphatase-conjugated anti-biotin antibody was used and signal was detected directly using Sigmafast BCIP/NBT tablets (Sigma-Aldrich).

2.2.2 Protein expression in *E. coli*

Expression strains

Expression of most *E. coli*-expressed proteins was carried out in BL21(DE3) cells (NEB). These cells express T7 polymerase upon induction with isopropyl β-D-1-thiogalactopyranoside (IPTG), allowing the expression of any genes

under the control of the T7 promoter. Origami B(DE3)pLysS cells (EMD Millipore) were used to express haptoglobin-haemoglobin receptors to aid formation of disulphide bonds. These cells contain mutations in *trx* and *gor* genes, which results in enhanced disulphide bond formation in the cytoplasm.

Small-scale expression trials

To assess protein expression at different temperatures and incubation times, colonies were picked from transformation plates and added to 100 ml antibiotic-containing LB. These cultures were incubated until an OD_{600} of 0.6-0.8 was reached, at which point they were induced with 1 mM IPTG. Incubation was continued for 3 or 16 hr at 18 °C, 25 °C, 30 °C and 37 °C. 1 ml samples from each temperature and time point were centrifuged at 12,000 x *g* for 2 min and the supernatants discarded. 1 ml of PBS was added to each pellet and sonicated for 5 x 1 s pulses at 50 mA using a Sonicator 3000 (Misonix). A sample of this was taken, representing the total cell fraction. The sonicated cells were then centrifuged at 12,000 x *g* for 2 min and a sample from the supernatant was taken, representing the soluble cell fraction. Expression levels were analysed by Western blot.

Large scale expression in *E. coli*

Large scale expression, based on the results of the small scale expression trials, was completed using 2-12 litres of culture. Colonies were picked from transformation plates, added to antibiotic-containing LB and incubated shaking overnight at 37 °C. 20 ml of overnight culture was added to each 1 litre of antibiotic-containing 2xYT media (Melford), and incubated shaking at 160 rpm at 37 °C. Cultures were induced with 1 mM isopropyl IPTG once an OD_{600} of 0.6-0.8 was reached, and the incubation temperature was reduced to the optimum temperature as determined by the small-scale trials. An incubation overnight at

18 °C was used for expression of TbbHpHbR and TvHpHbR, 3 hr at 30 °C for TcHpHbR and TEV protease, and overnight at 37 °C for ApoL1 constructs. Cells were harvested by centrifugation for 15 min at 6000 x *g*. Pellets were stored at -80 °C where required.

Soluble purification in *E. coli*

Pellets were resuspended in 20 mM Tris pH 8, 300 mM NaCl, 15 mM imidazole, 0.5% Triton x-100 and cells were disrupted by sonication using a Sonicator 3000 (Misonix). For TEV protease purification an additional 10% glycerol was included in the resuspension buffer. Disrupted cells were centrifuged for 20 min at 50000 x *g*. The protein was then purified by nickel affinity chromatography by applying the supernatant to Ni-NTA^{2+} affinity chromatography beads (Qiagen). The beads were washed with the resuspension buffer, followed by a wash buffer containing 20 mM Tris pH 8, 500 mM NaCl, 15 mM imidazole. The protein was eluted using 10 mM Tris pH 8, 300 mM NaCl, 200 mM imidazole. Where removal of the 6-His tag was required, the elution fractions were buffer exchanged using PD-10 desalting columns (GE Healthcare) into PBS.

Selenomethionine incorporation

For incorporation of selenomethionine, expression was carried out in minimal media supplemented with this amino acid based on the method in [153]. Overnight cultures were used to inoculate 2-12 litres of 2xYT media as with standard expression. Once an OD_{600} of 0.4 was reached, the cells were centrifuged 15 min at 6000 x *g*. The pellets were resuspended in minimal media containing 42 mM Na_2HPO_4, 22 mM KH_2PO_4, 17 mM NH_4Cl, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.04% glucose and 0.005% thiamine vitamin. Once an OD_{600} of 0.6-0.8 was reached, amino acids were added (L-Lysine 100mg/l, L-Phenylalanine 100mg/l, L-Threonine 100mg/l, L-Isoleucine 50mg/l, L-Leucine 50mg/l,

L-Valine 50mg/l, L-SelenoMethionine 50mg/l). After 15 min, the cultures were induced with 1 mM IPTG. Expression was carried out at 37 °C overnight due to slow growth in minimal media.

Insoluble purification in *E. coli*

Pellets were resuspended in 20 mM Tris pH 8, 300 mM NaCl, 15 mM imidazole, 0.5% Triton x-100 and cells were disrupted by sonication using a Sonicator 3000 (Misonix). Disrupted cells were centrifuged for 20 min at 50000 x g. The supernatant containing soluble proteins was discarded, and the pellet containing insoluble protein and cell debris was resuspended in 6 M guanidine-HCl, 10 mM Tris pH 8, 300 mM NaCl, 15 mM imidazole. These were incubated rolling at 4 °C for 1 hr, dounce homogenised and centrifuged for 30 min at 50000 x g.

The protein was then purified by nickel affinity chromatography by applying the supernatant to Ni-NTA^{2+} affinity chromatography beads (Qiagen). The beads were washed with the guanidine-containing resuspension buffer. The protein was eluted using 6 M guanidine-HCl, 10 mM Tris pH 8, 300 mM NaCl, 200 mM imidazole, before being concentrated to 1 ml per 6 litres of *E. coli* using an Amicon centrifuge filter with a molecular weight cut-off below half the molecular weight of the protein (EMD Millipore).

Rapid dilution refolding

For screening, 10 μl samples of guanidine-solubilised inclusion bodies were added to 200 μl of the Quickfold protein refolding buffers (AthenaES) which had been produced by Dr. Katherine Wright. Before addition of the protein, n-dodecyl-β-D-maltopyranoside (DDM) was added to each refolding buffer to a final concentration of 0.02%. To test success of refolding, 50 μl of these samples was run on Superdex 200 Increase 3.2/300 in

20 mM HEPES pH 7.5, 150 mM NaCl, 0.02% DDM.

The refolding buffer selected for flash dilution refolding of ApoL1 constructs, based on the results of the refolding trials, was 50 mM MES, 250 mM NaCl, 0.02% DDM, pH 6.0. After concentrating the protein solubilised in guanidine buffer, it was added dropwise to the flash dilution buffer to a final ratio of 1:100. The sample was constantly stirred whilst being recycled over a HisTrap FF 5 ml column (GE Healthcare) several times at 4 °C. The column was then washed with 20 mM citrate, 150 mM NaCl, 20 mM imidazole, 0.02% DDM pH 4.8. The protein was eluted using 20 mM citrate, 150 mM NaCl, 500 mM imidazole, 50 mM EDTA, 0.02% DDM pH 4.8. This was concentrated to 5 ml using an Amicon centrifuge filter (EMD Millipore) prior to gel filtration on a Superdex 200 16/600 in 25 mM citrate, 150 mM NaCl, 0.02% DDM, pH 4.8.

2.2.3 Protein expression using baculovirus-infected *Sf9* cells

Production of recombinant baculoviruses

Transfection of *Spodoptera frugiperda*-derived *Sf9* insect cells (Invitrogen) was carried out using the BaculoGold baculovirus expression system (BD Biosciences) following the manufacturer's instructions. Briefly, cells were seeded onto tissue culture plates and incubated with Transfection Buffer A. BaculoGold Baculovirus DNA was mixed with the construct cloned into the pAcGP67Acon vector (Figure 2.2) and Transfection Buffer B was added. This was transferred to the tissue culture plates. After incubation for 4 hr at 27 °C, the solution was removed and replaced with TNM-FH insect media (Sigma-Aldrich) followed by incubation for 5 days at 27 °C. Agarose plates were made containing *Sf9* cells and the transfection to perform plaque assays. Plaques were selected from these assays and added to insect cells to produce P1 viruses. Once the viruses were harvested they were stored at

4 °C for up to 2 months.

Large scale protein expression and purification using *Sf9* cells

Sf9 cells were maintained at 0.5 x 10^6 cells/ml in Sf-900 II medium (Thermo Fisher Scientific) by splitting three times a week. They were incubated at 27 °C shaking at 150 rpm. For large scale expression, P1 viruses were passaged twice more to amplify the virus, resulting in a P3 stock. 50 ml of this stock was added to 450 ml of *Sf9* cells one day after they had been passaged. After 3 days the cells were centrifuged at 5000 x *g* for 15 min.

To begin purification, the supernatant was filtered, before buffer exchange using a tangential flow filtration device (Pall) into 20 mM Tris pH 8, 300 mM NaCl. The elution from this was applied to Ni-NTA^{2+} affinity chromatography beads (Qiagen) and washed with 20 mM Tris pH 8, 500 mM NaCl, 15 mM imidazole, before elution using 20 mM Tris pH 8, 300 mM NaCl, 200 mM imidazole.

Deglycosylation

Proteins expressed in insect cells and required for crystallisation were subjected to deglycosylation to reduce the surface entropy. 20,000 U of both Endo H_f (NEB) and EndoF3 (Prof. Matt Higgins) were incubated with the protein to be deglycosylated in 1 mM $CaCl_2$, 1 mM $MgCl_2$, 100 mM HEPES pH 7.5 at 37 °C for 3 hr. Success of deglycosylation was confirmed by observing a band shift by SDS-PAGE.

2.2.4 Purification of haemoglobin

Hb was isolated from human blood by centrifugation at 1500 x *g* for 5 min. The cell pellet was resuspended in 20 mM HEPES pH 7.5, 150 mM NaCl and lysed using a Sonicator

3000 (Misonix). This was then centrifuged at 50000 x *g* for 20 min, and the supernatant was concentrated and gel filtered using a Superdex 200 16/600 run in 20 mM HEPES pH 7.5, 150 mM NaCl. SDS-PAGE showed that the protein content was almost exclusively Hb. Bovine haemoglobin was purchased as a lyophilised powder (Sigma-Aldrich).

2.2.5 Size exclusion chromatography

Size exclusion chromatography was generally used as the final step in the purification of all proteins and protein complexes. The sample was concentrated to 2-5 ml for application to a Superdex 16/600 column or 0.1-1 ml for a Superdex 10/300 column (GE Healthcare) using Amicon centrifugal filter devices with a molecular weight cut-off below half the molecular weight of the protein (EMD Millipore). A Superdex 75 or 200 was used depending on the molecular weight of the protein or complex being purified. Where a Superdex 200 3.2/300 column (GE Healthcare) was used for analytical gel filtration, a sample volume of 50 μl was used. In general, a running buffer of 20 mM HEPES pH 7.5, 150 mM NaCl was used for most proteins, whilst 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8 was most often used for ApoL1 constructs. Gel filtration was carried out using an ÄKTA purifier or pure (GE Healthcare). Elution fractions of 2 ml or 300 μl were collected for the Superdex 16/600 and 10/300 respectively. When required, purified proteins were stored at -80 °C after flash freezing in liquid nitrogen.

2.2.6 Tobacco etch virus protease cleavage

Proteins expressed using the modified pEt-15b vector (Figure 2.1B) or modified pAcGP67A vector (Figure 2.2) had an N-terminal 6-His tag followed by the recognition site for TEV protease. Where 6-His tag removal was required, the protein was mixed

with TEV protease in a 10:1 ratio, with 3 mM glutathione$_{ox}$, 0.3 mM glutathione$_{red}$ in PBS at 4 °C for 18 hr. The sample was applied to a Ni-NTA^{2+} affinity chromatography column and the flow-through collected. Any uncleaved protein, as well as the TEV protease, bound to the column and could be eluted in PBS, 200 mM imidazole.

2.3 Antibody production, selection, expression and purification

2.3.1 Hybridoma production, selection and expansion

Mouse inoculation and hybridoma fusion

Generation of antibodies using hybdrioma technology was attempted. Endotoxins were removed from purified proteins using high capacity endotoxin removal spin columns (Pierce) according to the manufacturer's instructions. Following this, Balb/c mice were immunised with the protein according to the schedule below:

Day 0: 100 micrograms of protein immunised in Complete Freund's adjuvant

Day 21: 50 micrograms of protein immunised in Incomplete Freund's adjuvant

Day 42: 50 micrograms of protein immunised in Incomplete Freund's adjuvant

Day 50: Small test bleed to check antibody levels by ELISA.

The protocol above describes the adjuvants used for immunisation using full length ApoL1 and was carried out by Nicky Hamp. For PFD and CTD, addivax was used instead. Lastly, for the complex of N-terminally truncated ApoL1:SRA, polyI:C adjuvant was used. Immunisation using these three immunogens was carried out by Lea Barford. The timecourse of injections also varied slightly from that shown above. Following the sacrifice of the

mouse, the spleen was harvested.

Splenocytes were fused with SP2/0 immortalised myeloma cells to produce hybridomas using the ClonaCell-HY hybridoma kit (Stemcell Technologies) according to the manufacturer's protocol. The spleen was disaggregated into a single cell suspension, followed by mixing with the myeloma cells. Addition of PEG resulted in fusion of the cell membranes, after which the cells were incubated at 37 °C for 16 hr. The cells were then plated out on a semi-solid medium and incubated at 37 °C for 14 days. Aminopterin in this medium inhibits dihydrofolate reductase, an enzyme required for the *de novo* DNA synthesis pathway. Therefore if cells are to divide they must use a salvage pathway which requires hypoxanthine-guanine phosphoribosyletransferase (HGPRT), an enzyme involved in the synthesis of IMP, a precursor of ATP and GTP. The myeloma cells used are negative for HGPRT, so are unable to divide, and unfused splenocytes have a short lifespan. Therefore only fused cells will survive.

Individual colonies were used to inoculate 200 μl of ClonaCell-HY Growth Medium E in a 96-well plate, and these were incubated at 37 °C for 3 days. The supernatants from these wells were used to perform enzyme-linked immunosorbent assay (ELISAs).

Hybridoma growth

Hybridoma cells were maintained at a density between 1-5 x 10^4 cells/ml in Clona-Cell-HY Medium A by splitting twice a week. Once a volume of 4 ml was reached, the medium was gradually exchanged to serum-free CD Hybridoma media (Gibco) containing 4 mM L-glutamine (Gibco) and 50 U/ml Pen-Strep (Gibco) over 3-4 passages. Hybridomas were grown to a volume of 2 litres and incubated for 6 days before harvesting by centrifugation for 15 min at 6000 x g. The supernatant was filtered and frozen at -20 °C.

Antibody purification and Fab production

The hybridoma supernatant was buffer exchanged using a tangential flow filtration device (Pall) into 20 mM Tris pH 8, 50 mM NaCl. Monoclonal antibodies (mAbs) were then purified using protein A/G agarose (Thermo Fisher Scientific). The elution from the previous step was applied to the beads, which were then washed using 20 mM Tris pH 8, 300 mM NaCl. mAbs were eluted in 1 ml fractions using 0.1 M glycine pH 2.5, immediately neutralised using 60 μl 1 M Tris pH 8.

For production of Fabs, the mAbs were buffer exchanged into 20 mM sodium phosphate, 10 mM EDTA pH 7 using PD-10 desalting columns (GE Healthcare) and concentrated to 20 mg/ml. Immobilized papain agarose (Thermo Fisher Scientific) was resuspended in 20 mM sodium phosphate, 10 mM EDTA, and mixed with the protein. 25 mM cysteine-HCl pH 8 was added and this was incubated at 37 °C for 16 hr. The flow-through from the beads contained the Fab and Fc fragments, as well as any uncleaved mAb. A HiTrap Protein A HP column (GE Healthcare) was used to remove any Fc or uncleaved mAb present. The flow-through from this contained the Fab fragments, as it is only the Fc region that binds to protein A.

2.3.2 Phage display for single chain Fv panning selection

Libraries containing phage expressing single chain variable fragments of antibodies (scFvs) were selected for the binding of constructs using phage display during a visit to MedImmune, supervised by Steve Rust and Andrea Gonzalez-Munoz. The BMV library was derived from bone marrow, peripheral blood lymphocytes and tonsil tissues [154]. The CL library was derived from spleenocyte cDNA [155]. The Ref library is a synthetic library in which only the CDR3s of the VH and VL have been randomised (MedImmune,

unpublished). The protein of interest had been previously randomly biotinylated on primary amines (PFD) or specifically biotinylated at the C-terminus (CTD and N-terminally truncated ApoL1). Streptavidin plates were coated with this protein at 10 μg/ml overnight at 4 °C before washing to remove unbound antigen. To remove any anti-streptavidin scFv-displaying phage, the phage were incubated with streptavidin-coated magnetic beads at room temperature for 1 hr. Both the plate and phage were blocked with 3% (w/v) skimmed milk powder in PBS, 0.02% DDM at room temperature for 1 hr. The phage were then incubated with the plate at room temperature for 1 hr before washing to remove bound phage. Bound phage were eluted by trypsin digestion at 37 °C for 30 min. The eluted phage were used to infect *E. coli* TG1 cells, before plating out onto a bioassay plate containing 2xTYAG agar and overnight incubation at 30 °C. To rescue the phage, the colonies grown overnight were used to inoculate 2xTYAG medium. Helper phage M13KO7trp were then added and after 1 hr the bacteria were transferred to 2xTYAK medium. This was grown overnight at 25 °C. The supernatant from this overnight culture contained phage ready for the next round of selection. Four rounds of selection were carried out for each construct.

2.3.3 Antibody selection

ELISAs were carried out for both hybridoma supernatant and scFv-displaying phage to test for reaction against the protein of interest. For selection of hybridomas, the protein of interest was immobilised on a maxisorp plate at 2 μg/ml, whilst for phage ELISAs, biotinylated protein at 10 μg/ml was coupled to a streptavidin-coated plate. Plates were incubated at 4 °C overnight before washing with PBS, 0.02% DDM, 0.05% Tween-20. For phage ELISAs, both the plate and phage to be tested were blocked with 3% (w/v) skimmed milk powder in PBS at room temperature for 1 hr, whilst for hybridoma ELISAs,

the plate only was blocked using protein-free blocking buffer (Pierce). The hybridoma supernatant or blocked phage was then transferred to the coated plate and incubated at room temperature for 1 hr. The plate was washed with PBS, 0.1% Tween-20, 0.02% DDM and the antibody was added, diluted 1:5000 in the blocking solution used. For phage ELISAs, this was an anti-M13-HRP antibody which recognises a component of the phage coat. For hybridoma ELISAs, this was an anti-mouse-peroxidase antibody (Sigma-Aldrich), which recognises the Fc region of the antibody. This was incubated for 1 hr at room temperature. After a final washing stage with PBS, 0.1% Tween-20, 0.02% DDM, TMB substrate (Thermo Fisher Scientific) was added to the wells and left for 10-20 min for the signal to develop. To stop the reaction, 0.5 M H_2SO_4 was added to each well. The signal was read using a SpectraMax M3 plate reader (Molecular Devices, UK) at 450 nm. Positive wells were judged to be those which displayed signal above 10% of the background level of the control wells.

2.4 Protein modification

2.4.1 Protein methylation

In order to carry out protein methylation, the protein was first diluted, if required, to below 0.5 mg/ml. 0.02 M dimethylamine-borane complex (ABC) and 0.04 M formaldehyde was then added. The sample was incubated for 2 hr at 4 °C, following which an additional 0.02 M ABC and 0.04 M formaldehyde was added. Finally, after another 2 hr incubation, 0.1 M ABC was added and the sample was incubated for 16 hr at 4 °C. The protein was then gel filtered, removing any remaining methylation reagents.

2.4.2 Protein biotinylation

For specific biotinylation, a biotin acceptor peptide (BAP) tag sequence was added by PCR to the C-termini of the constructs being used. Expression and purification was then carried out as before. 1 mg of protein at 30 μM was mixed with 20μg of BirA, 5 mM ATP and 300 μM biotin. This was incubated at room temperature overnight.

For non-specific biotinylation of primary amines, the protein was incubated with EZ-link NHS-biotin (Thermo Fisher Scientific) according to the manufacturer's instructions. The protein, at 2 mg/ml, was incubated with a 20-fold molar excess of biotin reagent at 4 °C for 2 hr.

All biotinylation reactions were performed in PBS, with the addition of 0.02% DDM when required. Desalting was carried out using a PD-10 desalting column (GE Healthcare) to remove excess biotin. Biotinylation was confirmed by Western blot using an anti-biotin antibody as the primary antibody.

2.4.3 Fluorescent labelling of proteins

Proteins were labelled in PBS, with 0.02% DDM where required. Labelling was carried out using the Alexa Fluor protein labelling kits (Thermo Fisher), extending the labelling time to 3 hr at room temperature followed by overnight incubation at 4 °C to increase labelling. Degree of labelling was calculated using the following equation:

$$\text{moles dye per mole protein} = \frac{A_\lambda}{\mathcal{E}_\lambda \text{ x protein concentration } (M)}$$

where A_λ is the absorbance at the absorption wavelength of the dye and $\mathcal{E}_\lambda$ is the extinction coefficient of the dye at this wavelength.

2.4.4 Nanodisc production and protein insertion

Nanodiscs were produced using a modified version of the method provided by Dr. Tao Ni from the Gilbert lab, STRUBI based on [156]. For the final method used for nanodisc production, 2.8 mg of lipids were mixed in chloroform in a ratio of 3DOPC:2DOPG (Avanti Polar Lipids) in a glass tube. The chloroform was evaporated using a stream of nitrogen gas, following which the lipids were resuspended in 25 mM sodium chelate, 20 mM HEPES pH 7.5, 100 mM NaCl. To fully resuspend the lipids, the tube was placed in an ultrasonic water bath (Ultrawave) and sonicated for 10 min. The sample was then heated to 42 °C, vortexed, and flash frozen in liquid nitrogen. These steps were repeated four times. 1 mg of the membrane scaffold protein, MSP1E3D1 (Sigma-Aldrich), was added to the sample. This was dialysed overnight against 25 mM citrate, 150 mM NaCl pH 4.8 buffer using a 7 kDa cut-off Slide-A-Lyzer (Thermo-Fisher Scientific). Next day, the sample was gel filtered using a Superdex 200 10/300 column using the same buffer as dialysis. Empty nanodiscs were stored at 4 °C for up to 1 week.

To insert ApoL1, the protein was diluted 1:5 or greater into purified nanodiscs to ensure the final concentration of DDM was well below the critical micelle concentration. ApoL1 was added to the nanodiscs at a molar ratio of 3:1. After overnight incubation at 4 °C, the sample was gel filtered using a Superdex 200 10/300 column in 25 mM citrate, 150 mM NaCl pH 4.8. ApoL1-containing nanodiscs were not stored and were used immediately.

2.5 Biophysical methods

2.5.1 Surface plasmon resonance

Surface plasmon resonance (SPR) is a technique commonly used to find the affinity and kinetics of the interaction between a ligand and an immobilised protein. A surface plasmon is an oscillation of electrons that can propagate on a thin film of conductor such as gold. The plasmon can interact with an evanescent wave in a way that is sensitive to the refractive index of the medium near the conductor surface. A prism is used to couple light from a beam to the surface plasmon and reflect some to an optical detector. Changes in the refractive index near the metal film alter the amount of light at the detector. Binding of prey molecules to the sensor chip increases the refractive index of the medium, thus altering the resonance angle for plasmon induction, which can be measured by the detector. It is important to also find the change in response when the ligand is flowed over the surface without any immobilised protein: this background value is subtracted from the response. The values for k_{on} and k_{off} can be found from the experiment, and therefore the K_D can be calculated:

$$K_D = \frac{k_{off}}{k_{on}}$$

SPR was performed using a Biacore T200 instrument and T200 control software (GE Healthcare). All experiments were completed using either a CAPture chip or protein A/G chip to immobilise the bait protein. The protein A/G chip was produced by immobilising ~800 RU of protein A/G (Thermo Fisher Scientific) to a CM5 chip (GE Healthcare) via primary amines before blocking all other sites with ethanolamine. Alternatively, protein was immobilised on the CAPture chip (GE Healthcare) according to the manufacturer's instructions via ssDNA conjugated to streptavidin. The bait in this case had been biotiny-

lated, either at a specific location or non-specifically at primary amines (see 2.5.1). In both cases, ~400-800 RU of the bait protein was immobilised on the chip.

In experiments where the K_D of an interaction was to be found, the ligand was flowed over the chip in a series of 1 in 2 dilutions, where the highest concentration was above the value of the K_D. For experiments which only aimed to confirm or disprove binding, a single curve at a concentration above the expected K_D was used. Generally, the running buffer used was 20 mM HEPES pH 7.5, 150 mM NaCl and 0.005% Tween-20. Where the experiment involved an ApoL1 construct binding to SRA, the running buffer used was 25 mM citrate, 150 mM NaCl, 0.005% Tween-20 pH 4.8. Ligands were injected for 120-240 s, followed by a 300 s dissociation. The chip was then regenerated. For the CAPture chip this was done using the regeneration reagents provided (6M guanidine HCl, 250 mM NaOH). For the protein A/G chip, 10 mM glycine pH 3 was used to dissociate bound antibody. Data analysis was completed using the T200 evaluation software (GE Healthcare).

2.5.2 Isothermal titration calorimetry

When performing isothermal titration calorimetry (ITC), one molecule is titrated into a solution containing the other and the enthalpy change is measured. Many injections of the ligand are carried out, producing a curve of the molar ratio of ligand:protein versus the enthalpy change of the injection. From the gradient of the curve the K_D of the interaction can be calculated, whilst the stoichiometry is equal to the molar ratio when the enthalpy change of injection is at half its maximum value.

Before beginning the experiment, the two interacting proteins were dialysed for 16 hr against 20 mM HEPES pH 7.5, 150 mM NaCl using Slida-A-Lyzer Dialysis Cassettes with

a molecular weight cut-off below half the molecular weight of the protein (Thermo Fisher Scientific). ITC experiments were completed using an iTC200 system (Malvern). The cell contained 300 μl of the less highly concentrated protein, whilst the syringe contained 50 μl of the more highly concentrated protein. The titrant was injected in 20 injections of 2 μl. The enthalpy change of each injection was plotted against the molar ratio of titrant:protein.

2.5.3 SEC-MALLS

Size exclusion chromatography multi-angle laser light scattering (SEC-MALLS) was used to determine the molecular weight of protein and protein:detergent samples. Purified samples were loaded onto a Superose 6 10/300 column (Section 3.5.1) or a Superdex 200 10/300 column (Section 6.2.5) (GE Healthcare) at a concentration of 0.05-0.15 mg/ml. A running buffer of 20 mM HEPES pH 7.5, 150 mM NaCl or, for ApoL1 samples which required detergent, 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8 was used. The elution was analysed as it emerged from the column by detecting laser light scattering at 662 nm wavelength at 8 scattering angles between 20.6° and 149.1° using a Heleos 8 instrument (Wyatt Technology, Germany). Where the amount of detergent bound was to be calculated, the refractive index was detected using an Optilab T-rEX (Wyatt Technology, Germany). ASTRA 6.1 (Wyatt Technology) was used to calculate molecular weights using the Zimm equation:

$$\frac{K^*c}{R(\theta,c)} = \frac{1}{M_w P(\theta)} + 2A_2c$$

where *R(θ,c)* is the excess Rayleigh ratio of the solution as a function of scattering angle, θ, and concentration of the solute, c, M_w is the molecular weight of the solute, A_2 is the second virial coefficient in the viral expansion of the osmotic pressure and *P(θ)* describes

the angular dependence of the scattered light. K^* is defined as follows:

$$K^* = 4\pi^2(dn/dc)^2 n_0^2 N_A^{-1} \lambda_0^{-4}$$

where *(dn/dc)* is the refractive index increment of the matter, N_A is Avogadro's constant and λ_0 is the vacuum wavelength of the laser. For simplification, the Zimm equation can be written as follows:

$$M_w = \frac{\Delta LS}{K\left(\frac{dn}{dc}\right)^2 \left(\frac{\Delta UV_{280}}{A_{280}}\right)}$$

at a 1 cm pathlength where A_{280} is the extinction coefficient of the protein in ml mg^{-1} cm^{-1}. When no detergent is present, only two detectors are required to determine molecular weight: the light scattering detector and a concentration sensitive detector, in this case a UV detector. *R(θ,c)* is directly proportional to the intensity of the scattered light in excess of the light scattered by the pure solvent, and *P(θ)* can be calculated from *R(θ,c)*. The value of *dn/dc* for proteins is 0.187 mg/ml, so the equation can be solved.

However, when studying a protein bound to detergent, the value of *dn/dc* depends on the proportion of detergent:protein found in the complex. Therefore a third detector, a refractometer, is required to solve the equation. In such cases, the equation can be written as follows:

$$M_{w,complex} = \frac{\Delta LS}{K(dn/dc)^2_{complex} c_{complex}}$$

where

$$M_{w,complex} = (1+\delta) M_w(protein)$$

and

$$c_{complex} = (1+\delta) c_{protein}$$

where δ is the amount of bound detergent compared to protein (g/g). Assuming that the detergent does not absorb at 280 nm, the concentration of the protein can be calculated as before, using the UV detector. The value of $dn/dc_{complex}$ can only be found if δ is known.

$$({}^{dn}/_{dc})_{complex} = \left(\frac{1}{1+\delta}\right)({}^{dn}/_{dc})_{protein} + \left(\frac{\delta}{1+\delta}\right)\delta({}^{dn}/_{dc})_{detergent}$$

By combining this equation with those describing the molecular weight and concentration of the complex, the following equation can be made:

$$M_{W,protein} = \frac{\Delta LS}{K(dn/dc)^2_{apparent}c_{protein}}$$

where $(dn/dc)_{apparent}$ describes how the refractive index of the solution changes as a function of protein concentration:

$$\left(\frac{dn}{dc}\right)_{apparent} = \left(\frac{dn}{dc}\right)_{protein} + \delta\left(\frac{dn}{dc}\right)_{detergent}$$

Assuming that the detergent does not absorb at 280 nm, this can be measured directly:

$$\left(\frac{dn}{dc}\right)_{apparent} = \frac{\Delta RI}{c_{protein}}$$

The molecular weight of the protein and, as dn/dc_{protein} and $dn/dc_{\text{detergent}}$ are known to be 0.187 mg/ml and 0.1435 respectively (for DDM), so the value of δ can now be calculated. This allows determination of the molecular weight of bound detergent [157].

2.5.4 Small-angle x-ray scattering

Small-angle x-ray scattering (SAXS) is a technique used to generate low resolution (50-10 Å) envelopes of a protein in solution. When an x-ray beam is applied to the sample, some of the x-rays are elastically scattered by atoms in the molecule. The resulting pattern of diffraction is recorded by a detector. The scattering pattern can be described by the relationship between the intensity of scattering, *I*, and the angular momentum transfer, *q*, which is defined as the following:

$$q = 4\pi sin\Theta/\lambda$$

where λ is the wavelength of the incident radiation and Θ is half the scattering angle between the incident and scattered radiation. This is used to produce 1D scattering curves for each dataset collected.

SAXS data was collected at PetraIII P12 at DESY, Hamburg, and at BM29 at the ESRF, Grenoble. This was completed for TbbHpHbRt and the Tb complex (TbbHpHbRt:HpSPHb) at 2 mg/ml, 1.5 mg/ml and 1 mg/ml and for HpSPHb at 2 mg/ml and 1 mg/ml in 20 mM HEPES pH 7.5, 150 mM NaCl. The scattering data at each concentration for each sample was averaged, and then the scattering of the buffer subtracted using PRIMUS [158, 159]. SAXS is a contrast method, where the signal is derived from the difference between the average electron density of the protein in buffer and the buffer alone. Care must therefore be taken when preparing the samples to ensure the composition of the sample buffer is very closely matched to the buffer sample used for subtractions. This is usually done by using the same batch of buffer for both the final gel filtration step of protein sample purification and the buffer sample measurements. Composite curves containing data from each concentration for one of the species were produced, generating a

zero concentration curve.

For the complex formed between two TbbHpHbRt molecules and dimeric HpHb, the process of data collection was slightly more complicated. Using the method of serial dilutions as above resulted in an averaged envelope of one or two receptors bound to HpHb. Therefore a different experimental set-up was required. HpHb was mixed with a six molar excess of TbbHpHbRt and concentrated to 100 μl. SAXS was then performed using an in-line SEC-SAXS set up. The sample at 1.5 mg/ml was injected onto a Superdex 200 10/300 column, and an x-ray scattering frame of the elution was collected every 2 s. Frames corresponding to the earliest eluting half of the peak were averaged, as they will have a greater fraction of two receptors bound to one HpHb dimer. Ten frames near the start of the run were averaged to produce a curve of the buffer scattering.

The Guiner approximation describes the relationship between intensity of scattering, *I*, and the radius of gyration, R_g and is true only at low values of *q* [160]:

$$lnI(q) = lnI(0) - \frac{q^2 R_g^2}{-3}$$

AutoRg [159] was used to generate Guinier plots of q^2 vs ln[*I(q)*], allowing prediction of R_g for each species. Gnom [161] transformed the 1D scattering curves into a particle distribution function, *P(r)*, by indirect Fourier transformation. The result of this is a histogram of distances, *r*, of all possible electron pairs in the protein. Thus the maximum distance on this graph is D_{MAX}, the predicted maximum length of the particle. A prediction of the Porod volume of the predicted particle was also determined from this plot [162]. An approximate molecular weight was estimated by dividing this value by 2. This plot, as well as the composite curves, was used by DAMMIF [163] to generate *ab initio* models of protein envelope. These models were built into an averaged model using the DAMAVER [164]

programme suite. DAMMIN [165] then produced a final model by minimising a function which describes differences between experimental data and scattering of the model. The envelope model was produced using Situs, and feature-based docking was completed using Sculptor [166].

2.5.5 Circular dichroism

Circular dichroism (CD) describes the difference in absorption of left- and right-handed circularly polarised light by chiral molecules and can be used to assess the secondary structure of a protein. The difference between the absorption of left- and right-handed circularly polarised light is expressed as the molar elipiticity at a particular wavelength. This refers to the change in the shape of the wave from circular to elliptical as the light is absorbed by the chiral molecule. Protein secondary structure gives rise to far UV CD due to the chiral environment of the peptide backbone. As many typical buffer components, such as Cl^- ions, absorb at these wavelengths, the buffer must be selected carefully to avoid this.

Ten CD spectra were collected using a Jasco J-720 Spectropolarimeter between 200-260 nm at 25 °C, and the results averaged. The protein was at 0.1-0.2 mg/ml in 100 mM phosphate buffer pH 7.5, 150 mM NaF, 0.02% DDM, held in a quartz cuvette (Starna Scientific Ltd). The high tension remained under 700 V in all cases.

2.5.6 Cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) is a technique used to image biological samples without the need for staining or crystallisation. The sample is prepared by pipetting a small volume onto a carbon film with holes in it, the grid, which is then blotted to reduce

the thickness of the liquid, and plunge frozen. The result is a thin layer of vitreous ice that contains the sample, ideally in many different orientations. Electrons are emitted by a source housed under a high vacuum and accelerated down the microscope column. The electrons are scattered when they interact with the sample and are subsequently focused by the electromagnetic lenses of the microscope. Cryogenic temperatures are used to reduce the radiation damage caused by the interaction of the electrons with the sample, which can result in the breaking of chemical bonds and the generation of free radicals. To carry out single particle analysis, individual particles on the images obtained are identified and sorted into groups representing different 2D projections of the particle. These are then combined to generate a 3D reconstruction of the structure.

Samples for cryo-electron microscopy were prepared using a Mark IV vitrobot (FEI). Various blotting protocols were trialled, with the best results obtained using a 3 s blotting time, 0.5 s wait and a blot force of -6. Both c-flat (Protochips) and Quantifoil holey carbon grids were tested, with the best images obtained using quantifoil 1.2/1.3 grids (Quantifoil). Images were collected using a Talos 200 kV cryo-TEM with Ceta camera (FEI) and 70° tilt cryo-holder (Gatan). Nanodisc concentrations of 0.05-1.5 mg/ml were tested, with the best particle density obtained at a concentration of 0.8 mg/ml. Micrographs were viewed and particles picked using EMAN2 [167].

2.5.7 Electrophysiology experiments

Planar lipid bilayer experiments can be used to study the effect of different conditions such as pH and lipid on the function of a protein pore. In this project, the data were collected using an Axopatch 200B patch-clamp amplifier which measures the ionic current as a continuous analogue signal. This was converted into binary numbers using a 16-bit

digitiser (Molecular Devices). Finally, the signal was displayed on a computer using the Clampex software (Molecular Devices). A sampling frequency of 10 Hz and a low pass filter with a frequency of 1 kHz was used. The sampling frequency determines how many times a second the signal is recorded, whilst the low pass frequency reduces the signal-to-noise but also the temporal resolution.

All experiments were carried out within a Faraday cage which was grounded and placed on a pneumatic anti-vibration table. This aims to stop any electrical or vibrational noise from the external environment from being detected by the apparatus and recorded. Ag/AgCl electrodes were embedded in an agar bridge, recommended when asymmetric conditions are being used. Ag/AgCl electrodes were prepared by soaking a 3 cm long, 1.5 mm diameter Ag wire in 5% hypochlorite solution overnight before use to coat the silver in AgCl. The electrode was then inserted into a 200 ul pipette tip filled with 3 M KCl, 3% w/v agarose with a rubber stopper at the end. The very tip of the pipette tip was then trimmed to ensure that the opening was at least 1 mm in diameter. After several uses, white patches appeared on the sides of the electrodes corresponding to degradation. At this point, the electrodes were replaced with fresh Ag/AgCl electrodes. The chambers used in these experiments were custom made out of Delrin, with the two 1 ml compartments separated by a 20 μm-thick PTFE film (Goodfellow UK) (Figure 2.3). At the centre of this film is an aperture ~100 μm in diameter.

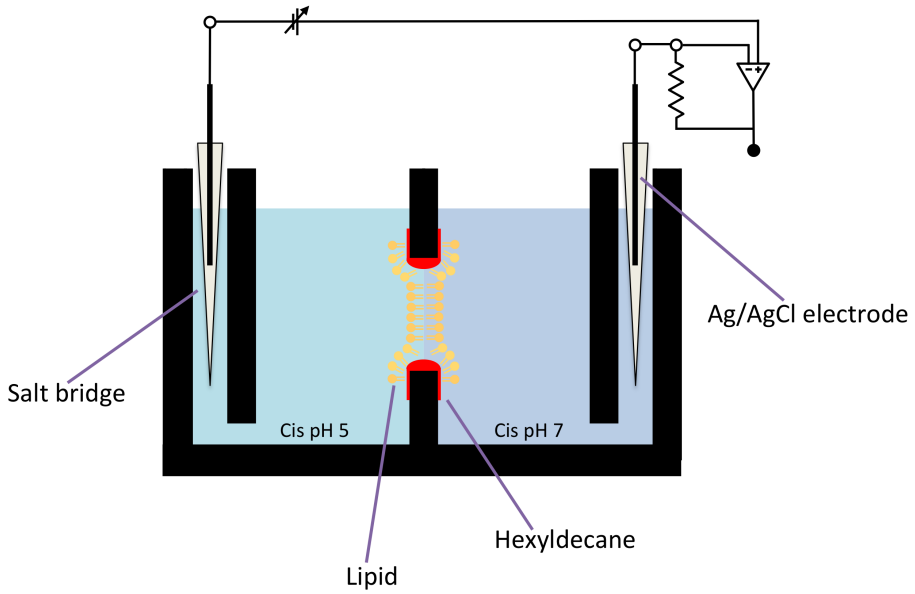


FIGURE 2.3: Electrophysiology experimental set up. The *cis*- and *trans*-compartments are separated by an aperture across which a planar lipid bilayer is formed. The membrane potential is controlled and the current is recorded by a voltage-clamp amplifier. This figure was adapted from [168].

To prepare bilayers, the aperture of the film was pre-painted with a single drop of 10% hexadecane in pentane. This formed an annulus around the rim of the aperture, allowing the bilayer to form an interface with the Teflon. After this dried, the *cis*-compartment was filled with 1 ml of 0.5 M KCl, 5 mM $CaCl_2$, 0.5 mM EDTA, 5 mM K-succinate, pH 5.3 whilst the *trans*-compartment was filled with 1 ml of 0.5 M KCl, 5 mM $CaCl_2$, 0.5 mM EDTA, 5 mM K-HEPES, pH 7.2, based on the conditions used in [119]. The lipid used to form the bilayer (Avanti Polar Lipids), at 10 mg/ml in pentane, was then added to the top of each compartment using a glass capillary. After the pentane had evaporated, the solution in one compartment was pipetted up to lower the level of the solution to below the aperture, after which the solution was slowly pipetted back into the compartment. After doing this several times in both compartments, a bilayer formed in the aperture of the film. The capacitance of the bilayer was then checked.

Once a bilayer was formed, its stability was tested by applying a potential of 200 mV. Stable bilayers showed no fluctuations in conductance at this voltage. To insert ApoL1, 0.1 μl of freshly refolded and gel-filtered ApoL1 was added to the solution in the *cis*-compartment. The concentration of ApoL1 was between 0.1-0.25 mg/ml. Next, a magnetic stirrer bar was used to mix the solution, following which a potential between -200 and -100 mV was applied to allow insertion of an ApoL1 molecule into the bilayer. Between each experiment the chamber was washed with water, followed by ethanol. Overnight soaks in ethanol were also completed.

To produce histograms of channel conductance, the amplitude of individual insertion events were recorded by finding the difference between the conductance before and after the event. These were then compiled into histograms and fit using a Gaussian curve to calculate the average conductance level.

$$f(x) = \sum_{i=1}^{n} A_i \frac{e^{-(x-\mu_i)^2/(2\sigma_i^2)}}{\sigma_i\sqrt{2\pi}} + C$$

The fit solves for the amplitude *A*, the Gaussian mean μ, the Gaussian standard deviation σ and the constant y-offset *C* for each component *i*.

Single channel searches were completed using the Clampfit software. This allowed the user to determine the conductance for level 0 and 1, after which the software identified all opening and closing events within a set timeframe. Events of less than 5 ms in duration were ignored, as these could represent noise alone. The software then fit a Gaussian curve to histograms for each level, allowing the calculation of the average dwell time at each level. Lastly, histograms were also produced to describe the occupancy of different conductance levels over a set timeframe, allowing confirmation of results obtained from the single channel searches.

2.6 X-ray crystallography

2.6.1 Crystallisation and optimisation

Protein crystals are required for x-ray crystallography, as a single molecule is a very weak diffractor of x-rays. The most common method of protein crystallisation is vapour diffusion. In this method, the purified protein is mixed with a buffer containing a precipitant, such that the concentration of the precipitant is just below that required to cause precipitation. Over time the water from the buffer evaporates, increasing the concentration of both the precipitant and the protein. Once the correct concentration of both these components is reached, nucleation occurs. The concentration of protein continues to decrease, reducing the chance of further nucleation and instead encouraging crystal growth.

Before setting up crystallisation trials, a sample of high purity and ideally high concentration (10 mg/ml) was obtained. Protein crystallisation often requires the testing of many different conditions before an initial hit is found. This was done by testing different 96-well commercial screens which employ the vapour diffusion method using a sitting drop setup. Some, such as JCSG+ (Molecular Dimensions), are sparse matrices designed by analysing which buffers have been most successful in the crystallisation of other proteins. Others, such as PACT (Molecular Dimensions), are grid screens which systematically test the effect of pH, anions and cations on crystallisation whilst keeping other conditions the same. These were set up rapidly using a mosquito crystallisation robot (TTP Labtech) and stored in a crystallisation tray incubator for automated imaging (Rigaku).

In some cases, optimisation of crystals was required. Although this can be done using grid screens sampling conditions around that which was most successful previously, in most cases the reservoir condition was not altered from the well solution which gave the initial crystals. Instead, seeds were generated by vortexing a crystal sample in 50 μl of

the well solution it was grown in, with a seed bead (Hampton Research).The Silver Bullet additive screen (Hampton Research) was also used. When attempting to obtain a heavy atom derivative dataset, native crystals were soaked in solutions of heavy atoms using the heavy atom screen kit (Hampton Research). Soaking was carried out overnight, followed by washing with the original crystallisation buffer to remove any unbound metals.

If the well condition did not already contain it, the crystal was transferred to a buffer containing a cryoprotectant, such as 30% glycerol. It was then cryo-cooled by plunging it into liquid nitrogen on a cryo-loop ensuring that only vitreous ice formed, avoiding damage to the crystal.

2.6.2 Data collection and reduction

Synchrotrons were used to collect all of the crystallographic data presented in this project using x-ray wavelengths of 0.9-1.1 Å. The crystals on the cryo-loops were mounted onto a goniometer in the path of these x-rays, allowing the crystal to be rotated precisely. Data collection was performed at 100 K to decrease the speed at which radiation damage occured and reduce disorder within the protein crystal. Pilatus detectors converted the diffracted x-rays into an electrical signal, which was recorded. Initially, three images were collected for each crystal at angles of 0°, 45° and 90°. These diffraction patterns were first inspected by eye to assess the quality of the spots and to what resolution spots could be seen. If spots were observed below 4 Å, a data collection strategy suggested by EDNA [169] or mosflm [170] was followed and data was collected using a rotation/oscillation method. The first frame is collected by oscillating the crystal through a small angle, recording all the resulting reflections. The crystal is then rotated to a new starting point which overlaps with the previous oscillation range, and the next frame is collected. This is re-

peated until all unique reflections have been recorded.

Complete datasets were indexed by iMOSFLM [171], XDS [172] or xia2 with DIALS [173, 174] to assign the unit cell dimensions. The spacing of reflections on the collected images are inversely related to the real-lattice spacings, allowing the unit cell dimensions to be calculated. iMOSFLM and XDS do this using spots found on 2-3 images, whilst DIALS finds the strongest spots from all images to use. A fast fourier transformation (FFT) method is employed by most programmes. The spot co-ordinates are converted into reciprocal-lattice scattering vectors, which are aligned along various directions relative to the x-ray beam. At directions which relate to a real space axis of the crystal, the vectors will lie in clusters at regular intervals along the direction. A FFT of these vectors will give a series of large peaks. The first non-origin peak will occur at a spacing that indicates the length of the real-space axis. Once the unit cell parameters have been determined, they are then refined using a greater subset of reflections. Next, the programmes carry out integration to determine the intensities of diffraction spots (I) and the background noise (σI). iMOSFLM uses 2D integration by profile fitting, whilst both XDS and DIALS use 3D integration, building up profiles across several adjacent images.

The space group was predicted by the integration programme and confirmed using Pointless [175]. This programme scores all possible space groups consistent with the crystal dimension restraints and then checks sets of reflections which may be systemically absent to suggest the best possible space group. The dataset was then scaled and, where 2D integration was carried out, merged. Scaling is required as the diffraction power, intensity of the x-ray beam and the length of the x-ray path may vary as the crystal is rotated during data collection. This was done using Scala [175] for datasets processed by iMOSFLM, and otherwise by Aimless [176]. The averaged intensities for each reflection are then converted into averaged structure factor amplitudes by Ctruncate [177].

The quality of the data can be assessed using a number of parameters. The traditional measure of data quality, R_{merge}, measures the spread of N independent measurements of the intensity of the reflection, $I_j(hkl)$, around their average, $\bar{I}(hkl)$:

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

However, a problem with this is that the value increases with multiplicity (discussed below), even though the data is likely to improve. Therefore the precision-indicating R-factor is often used, which gives a value independent of the multiplicity:

$$R_{PIM} = \frac{\Sigma_{hkl}\left(\frac{1}{N-1}\right)^{\frac{1}{2}}\Sigma_{i=1}^{N}|I_{(hkl)i} - \bar{I}_{(hkl)}|}{\Sigma_{hkl}\Sigma_{i=1}^{N}I_{(hkl)i}}$$

Another parameter often used is I/σ(I), a measure of the signal-to-noise ratio of the data. Historically, data was cut off at the resolution at which this value dropped below 2 [178]. However, a value of 1 or even lower is often used now, with $CC_{1/2}$ being used to determine at which resolution the data should be cut off at instead [179]. To calculate this, the dataset is randomly split in half:

$$CC_{1/2} = \Sigma(x - \langle x\rangle)(y - \langle y\rangle)/[\Sigma(x - \langle x\rangle)^2\Sigma(y - \langle y\rangle)^2]^{1/2}$$

where x and y are the intensities of unique reflections merged across the observations randomly assigned to subsets A and B, and $\langle$x$\rangle$ and $\langle$y$\rangle$ are their averages [180]. Examples have demonstrated that usually there is little information below a $CC_{1/2}$ of 0.2-0.4 [176].

Finally, a completeness of 99% or above and a multiplicity of over 5 are desirable [181]. The completeness is below 100% if there are any individual, symmetrically unique reflections that were not measured. The multiplicity describes the average number of mea-

surements recorded for each of these reflections. Having more than one measurement for each reflection reduces the error and therefore increases the accuracy of the final estimation of the intensity.

2.6.3 Phase determination

A structure factor, F_{hkl}, is a wave which describe the sum of a single reflection, *hkl*. The structure factor equation demonstrates that the intensity of every reflection depends on the electron density, $\rho(x,y,z)$, at every point in the cell:

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

This equation includes both the amplitudes, $|F_{hkl}|$, and phases, ϕ_{hkl}, of the structure factors. V is the volume of the unit cell of the crystal. However, the intensities recorded during data collection only describe the amplitude of the structure factors, whilst the phases are lost. This is known as the phase problem, and determining the phases is a major roadblock towards solving the structure of a protein.

In this thesis, molecular replacement was the method used to overcome this problem. This involves extracting the phases of a structurally similar protein and using these as starting phases. For this project the models used were either homologues from a different species, or the identical protein which was now being structurally characterised as part of a different complex.

The phases of the structure factors depend on the location of atoms within the unit cell. It is therefore critical that the phasing model is superimposed in the unit cell in the orientation closest to that of the protein of interest before the phases are calculated. To reduce the computer power required, this search is completed in two separate stages.

First, the orientation of the model relative to the unit cell is determined. In the past this was done by minimising the difference in Patterson functions, which can be calculated from structure factor amplitudes alone, between the data and the search model. The best orientation found using this search was then used to perform the translation search, which minimises the difference in structure factor amplitudes between the data and the search model, the R-value.

The programme used here for molecular replacement was Phaser [182], where the Patterson methods described above have been replaced with a maximum likelihood method. This method uses the principle that the best model is the one most consistent with the data. Consistency with the data is measured by the probability that the data would be measured given the model that is being tested. This method has advantages over the Patterson method, as it takes into account model errors and incompleteness, increasing the chance of finding the solution. However, it still uses a rotation search followed by a translation search as before. A packing search then determines whether there are any clashes within the crystal, before the solution is refined. The R-value of the output model describes how well the model fits the data, and is calculated as follows:

$$R = \frac{\Sigma||\mathbf{F}_{obs}| - |\mathbf{F}_{calc}||}{\Sigma|\mathbf{F}_{obs}|}$$

If the solution is correct, the R-value should be less than 45-50%. The likelihood that the solution is correct can also be assessed using the least likelihood gain score (LLG), which describes how much better the solution is than a random solution, and the translation function z-score (TFZ), which is how many standard deviations the solution LLG is above the mean value for a random set of translation functions.

2.6.4 Model building and refinement

Once an initial solution was found by molecular replacement, a cycle of model building in real space and restrained reciprocal space refinement was begun (Figure 2.4). REFMAC [183] was used for refinement of higher resolution structures, whilst Buster [184] and phenix.refine [185] were used for lower resolution structures. Refinement aims to find the best fit of the atomic model into the observed data whilst ensuring that the stereochemistry of the model is as good as possible. Maps are calculated using the phase estimates from the current atomic model. When the model is updated to fit the map, this changes the phases calculated from this model, resulting in changes in the new map. The refinement programme uses many restraints affecting parameters such as bond lengths, angles, planes, and some torsion angles. Other restraints can be applied by the user if desired. Non-crystallographic symmetry restraints are often used for moderate-low resolution data. Where there are several similar subunits in the unit cell, these can be identified and restrained to restrict how much they deviate from each other. Another restraint which can be used involves translation/libration/screw groups, which are used to allow rigid body motions of chains, domains or subunits relative to each other. If a higher resolution structure of one of the components in a complex has been determined previously, external restraints can be generated from this to restrict how much the current model deviates from this high resolution model.

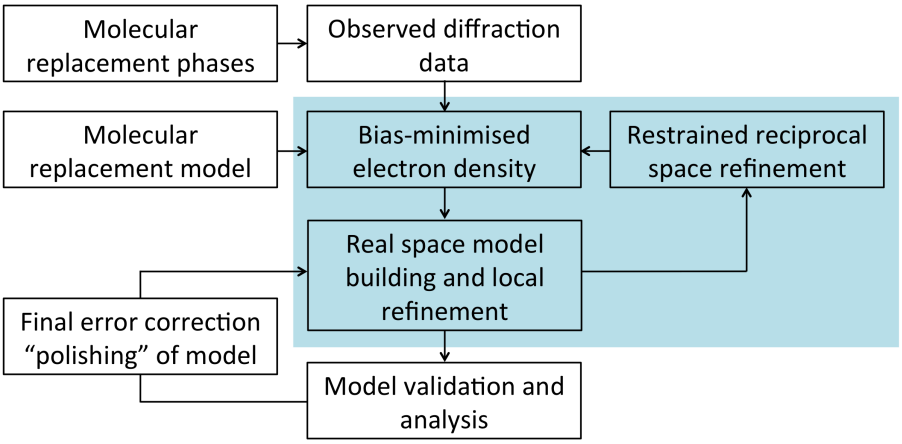


FIGURE 2.4: Pipeline for structure determination using molecular replacement. The cycle of refinement is highlighted in blue. This figure was adapted from [186].

Two electron density maps are used during model building. The first, $2F_{obs}$-F_{calc}, is the standard electron density map, and represents the contents of the crystal. The second is the difference density map, F_{obs}-F_{calc}. This is used to identify areas where the atomic model contains atoms which do not have any electron density, or where there is more electron density that requires atoms to be built into it. Model building was generally carried out in Coot ([187]). Buccaneer [188] can be used for automatic model building by comparing regions of the electron density map to maps of reference structures. This was used to place α-helices in the initial model building of TbbHpHbRt. Refinement ends when no significant changes can be made to the model based on the newest electron density map and the geometric parameters of the structure are within acceptable values.

Several tools are used to assess the quality of the model. To assess the fit of the model to the data the R-value is used, also called R_{work}:

$$R = \frac{\Sigma||\mathbf{F}_{obs}| - |\mathbf{F}_{calc}||}{\Sigma|\mathbf{F}_{obs}|}$$

The smaller the value of this parameter, the better the fit of the data to the model.

However, as the refinement programme is explicitly attempting to reduce R, this can lead to bias. To provide a less biased parameter, R_{free} is calculated in the same way, but only using 5% of the data which has not been used at any stage of structure determination. If the model is free from bias, the two values should be similar, with a slightly higher R_{free} value. If however R_{free} deviates significantly from R it indicates an over interpretation of the data. The geometry of the structure must also be assessed to confirm that bond angles, lengths and other values are within allowed ranges. A Ramachandran plot, generated by programmes such as PROCHECK [189], is used to visualise whether backbone backbone dihedral angles are within allowed regions for the amino acid type.

Chapter 3

Structural basis of haptoglobin-haemoglobin binding by TbbHpHbR

3.1 Summary

The first aim of my project was to determine how the *T. brucei brucei* haptoglobin-haemoglobin receptor (TbbHpHbR) is able to recognise its ligand, haptoglobin-haemoglobin (HpHb). Furthermore, I wanted to assess whether this binding mechanism was likely to be conserved for the uptake of the innate immune factor, trypanolytic factor 1 (TLF1), which contains haptoglobin-related protein (Hpr). I began by using information gained from sequence comparisons with proteins of known structure, namely porcine haptoglobin and *T. congolense* HpHbR (TcHpHbR), to design constructs containing only the domains predicted to be required for binding, increasing the likelihood of crystallisation. The truncated TbbHpHbR and Hp expressed solubly in *E. coli* and baculovirus-infected

Sf9 cells respectively, and the affinity for each other when Hp serine protease domain (HpSP) was mixed with Hb matched that of the full length proteins. I obtained the crystal structure of TbbHpHbR, which revealed a similar architecture to TcHpHbR, but with a distinct kink near the C-terminus. The crystal structure of the receptor with haptoglobin-haemoglobin demonstrated how the receptor recognises its ligand, and also shines some light on a possible role for the kink, allowing two receptors to simultaneously bind one dimeric haptoglobin-haemoglobin molecule, increasing efficiency of uptake.

3.2 Expression and purification of TbbHpHbR and HpHb

3.2.1 TbbHpHbR

To solve the structure of TbbHpHbR with and without its ligand, HpHb, using crystallography, the design of suitable constructs was required. Decreasing the disorder of a protein makes it more likely to crystallise, as disorder can hinder the packing into crystals. Sequence comparisons had revealed that, as with the variant surface glycoproteins (VSGs), TbbHpHbR has an additional C-terminal domain when compared to TcHpHbR [94]. As this domain would be positioned close to the cell membrane and as TcHpHbR can also bind HpHb, it was unlikely that this domain was involved in ligand binding. A truncated construct of TbbHpHbR (TbbHpHbRt), which lacked the C-terminal domain, was designed. The receptor had already been cloned into a modified pEt-15b vector with an N-terminal 6-His tag followed by a Tobacco Etch Virus (TEV) protease cleavage site (pEt-15b con). In order to produce the truncated version, the codon corresponding to residue 294 was mutated to a stop codon using site-directed mutagenesis. The full length *T. brucei* HpHbR

had been previously expressed in *E. coli*, so this was the first choice of expression system in which to test the truncated construct. The Origami B strain was used, which carries mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, greatly enhancing the formation of disulphide bonds in the cytoplasm. Indeed, TbbHpHbRt did express solubly in *E. coli* and was purified by Ni-NTA^{2+} affinity chromatography, followed by TEV protease cleavage to remove the flexible 6-His tag. A reverse Ni-NTA^{2+} column was used to remove TEV protease and any uncleaved TbbHpHbRt. The flow through from this was purified further by size exclusion chromatography (Figure 3.1). Although contaminants can still be seen, these are minor components.

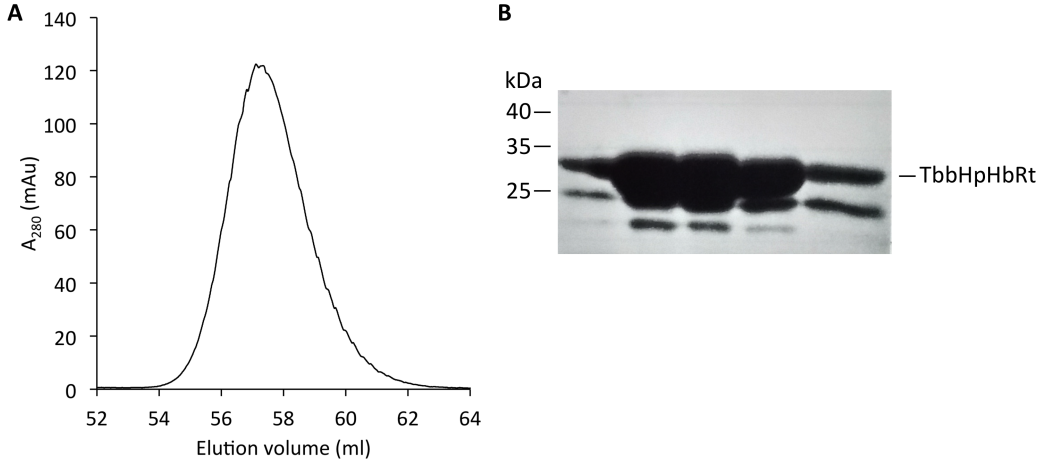


FIGURE 3.1: Purification of truncated TbbHpHbR. (A) TbbHpHbRt was purified by size exclusion chromatography using a Superdex 200 16/600 column. (B) Samples taken from the gel filtration peak elution fractions were analysed using SDS-PAGE. This revealed a large band at the correct molecular weight for TbbHpHbRt, and two minor bands which are likely to be contaminants or degradation products.

Selenomethionine-labelled TbbHpHbRt was required to attempt experimental phasing of a native x-ray crystallography dataset. This was done using a metabolic inhibition protocol, whereby methionine biosynthesis is blocked by growing *E. coli* in a high concentration of isoleucine, lysine and threonine. Selenomethionine is also added to the media, leading

to the incorporation of this in the place of methionine [153]. Expression was successful, although yields were approximately a tenth of the native level. The method was then adapted, only transferring the cells from LB into minimal media when the correct optical density for expression was reached. This did improve expression, but native levels were still not obtained (Figure 3.2).

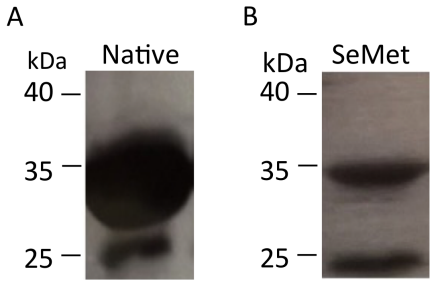
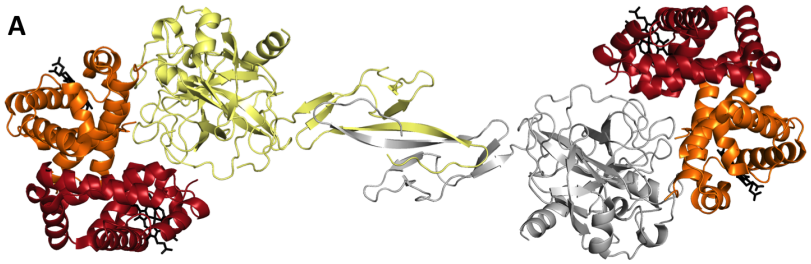


FIGURE 3.2: Comparison by SDS-PAGE of expression of (A) native TbbHpHbRt and (B) selenomethionine-labelled TbbHpHbRt. These samples were taken after Ni-NTA^{2+} affinity chromatography and before TEV protease cleavage. The band for the receptor, seen just under 35 kDa, is greatly decreased in the selenomethionie-labelled sample.

3.2.2 HpSP and HprSP

Haptoglobin is comprised of two domains, as first suggested by sequence analysis. The structure of porcine HpHb revealed a dumbbell shape, where the CCP domain forms a β-strand swap with a second haptoglobin molecule at the centre of the dumbbell to form a dimer [133] (Figure 3.3A). At each end the SP domains form all the contacts with the haemoglobin $\alpha\beta$ dimers. It was already known that neither haptoglobin nor haemoglobin could interaction with TbbHpHbR alone, only in a complex, suggesting that both formed interactions with the receptor [43]. I therefore set about designing a human HpSP construct to use for crystallisation, as this would ensure that when forming the complex with TbbHpHbRt the ratio of both components would be 1:1, and not a mixture of 1 or 2 receptors binding to each HpHb dimer. A sequence alignment of human haptoglobin 1-1

with porcine haptoglobin shows that the proteins are 80% sequence identical, and therefore are likely to share the same architecture (Figure 3.3B). The human HpSP construct was comprised of residues 89-347 of full length haptoglobin.



B

```
Porcine    1 MRALGAVVALLLCGQLFAAETGNEATDATDDSCPKPPEIPKGYVEHMVRY     50
             |.|||||:||||.||||||.::||:.||..||.|||||||..||||||.|||
Human      1 MSALGAVIALLLWGQLFAVDSGNDVTDIADDGCPKPPEIAHGYVEHSVRY     50

Porcine   51 HCQTYYKLRTAGDGVYTLDSNKQWTNKVTGEKLPECEAVCGKPKNPVDQV    100
             .|:.|||||||.|||||||:..|||.||..|:|||||||||||||.:.:|
Human     51 QCKNYYKLRTEGDGVYTLNDKKQWINKAVGDKLPECEAVCGKPKNPANPV    100

Porcine  101 QRIMGGSLDAKGSFPWQAKMISHHNLTSGATLINEQWLLTTAKNLRLGHK    150
             |||:||.|||||||||||||:||||||:||||||||||||||||.|.|.
Human    101 QRILGGHLDAKGSFPWQAKMVSHHNLTTGATLINEQWLLTTAKNLFLNHS    150

Porcine  151 NDTKAKDIAPTLRLYVGKKQEVEIEKVIFHPDNSTVDIGLIKLKQKVPVN    200
             .:..||||||||.|||||||.||||||:.||:.|.||||||||||||.||
Human    151 ENATAKDIAPTLTLYVGKKQLVEIEKVVLHPNYSQVDIGLIKLKQKVSVN    200

Porcine  201 ERVMPICLPSKDYVNVGLVGYVSGWGRNANLNFTEHLKYVMLPVADQEKC    250
             |||||||||||||..||.||||||||||||..||:||||||||||||::|
Human    201 ERVMPICLPSKDYAEVGRVGYVSGWGRNANFKFTDHLKYVMLPVADQDQC    250

Porcine  251 VQYYEGSTVPEKKTPKSPVGVQPILNEHTFCAGLSKYQEDTCYGDAGSAF    300
             :::||||||||||||||||||||||||||||||:||||||||||||||||
Human    251 IRHYEGSTVPEKKTPKSPVGVQPILNEHTFCAGMSKYQEDTCYGDAGSAF    300

Porcine  301 AVHDKDDDTWYAAGILSFDKSCRTAEYGVYVRVTSILDWIQTTIADN      347
             ||||.::||||||.||||||||||..|||||||:||||.||:|.|||:|
Human    301 AVHDLEEDTWYATGILSFDKSCAVAEYGVYVKVTSIQDWVQKTIAEN      347
```

FIGURE 3.3: Designing the human HpSP construct. (A) The structure of porcine HpHb (pdb:4F4O) [133]. Hp molecules are in yellow and grey, haemoglobin α in orange and β in red. The HpCPP domains can be seen in the centre of the structure forming a β-strand swap, whilst the SP domains at each end form all interactions with the haemoglobin dimer. (B) Sequence alignment of porcine (UniProtKB Q8SPS7) and human (UniProtKB A0A0C4DGL8) haptoglobin 1-1 proteins. Sequences were aligned using EMBOSS Needle pairwise sequence alignment. The sequences are 80% identical.

The construct was ordered as a synthetic gene codon-optimised for expression in insect cells. This was cloned into a pAcGP67A vector for expression in an *Sf9* baculovirus

system, modified to include an N-terminal 6-His tag and TEV protease cleavage site. This expression system was chosen as a similar system had been used previously to express full length haptoglobin [129]. *Sf9* cells were transfected following the BaculoGold protocol (BD Biosciences), and four HpSP-expressing P1 virus stocks were created from plaque-purified viruses. Expression trials were carried out in both *Sf9* and *Hi5* cells. Samples from both the total and soluble fractions were run on SDS-PAGE and Western blotted (Figure 3.4A). Plaque A had the strongest band, so this was amplified and used to transfect larger volumes of cells. As both *Sf9* and *Hi5* cells showed similar expression levels, *Sf9* cells were selected for simplicity as these are also the cells used for viral amplification. Filtered insect cell supernatant was buffer exchanged by tangential flow filtration, before purification of HpSP using an Ni-NTA^{2+} affinity column. As this appeared to be relatively pure by SPS-PAGE, no further purification was completed before formation of the HpSPHb complex.

As TLF1 contains haptoglobin-related protein instead of haptoglobin, it was also important to be able to produce this protein. A synthetic gene that corresponded to the SP domain of Hpr was therefore ordered. Cloning, transfection and expression trials were carried out using the same protocol as for HpSP, with 8 P1 viruses selected from plaque assays for expression trials (Figure 3.4B). Plaque 5 showed the highest expression level, so this P1 was frozen at -80 °C for large-scale expression at a later date. HprSP was then expressed and purified using the same protocol as HpSP.

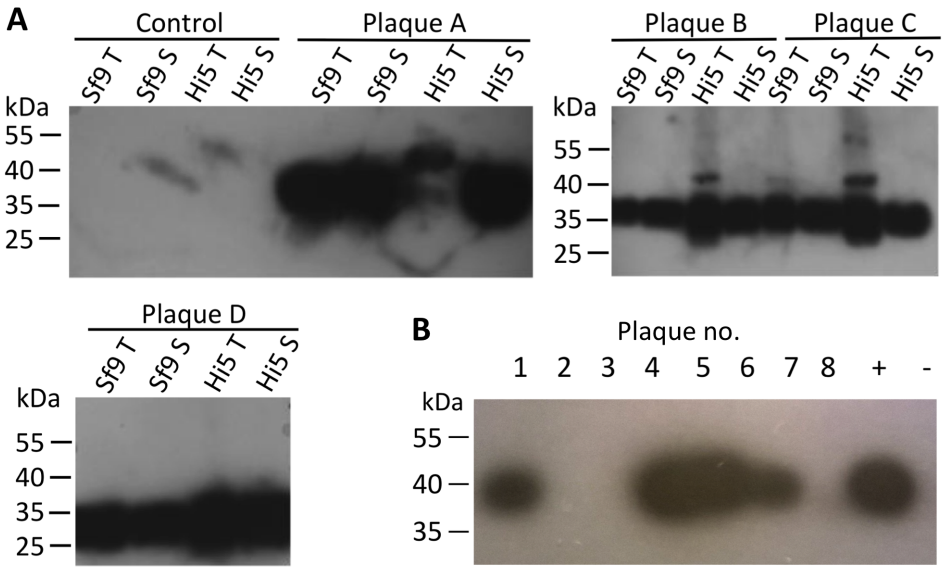


FIGURE 3.4: Expression trials for (A) HpSP and (B) HprSP in baculovirus-infected *Sf9* and *Hi5* cells. For HpSP, 4 different plaques were tested in both cell types, as well as a control containing no plaque. Total (T) and soluble (S) fractions were run on SDS-PAGE before Western blotting. Based on the results of HpSP, only soluble expression of HprSP by *Sf9* cells was assessed. This time a positive control of HpSP was used (+), as well as a negative control containing no plaque as before (-). Detection was performed using an anti-6-His antibody.

Due to the heterogeneity and flexibility of N-linked glycans, trimming them is often crucial before crystallisation can be achieved. Deglycosylation trials of HpSP using Endo H_f and Endo F3, which cleave within the chitobiose core of N-linked high mannose and complex oligosaccharides respectively, were carried out. Deglycosylation was successful, as seen by the shift of the gel band on SDS-PAGE (Figure 3.5). As incubating overnight did not increase the shift, 3 hour incubations were used for all further deglycosylations of HpSP, prior to mixing with Hb and size exclusion chromatography.

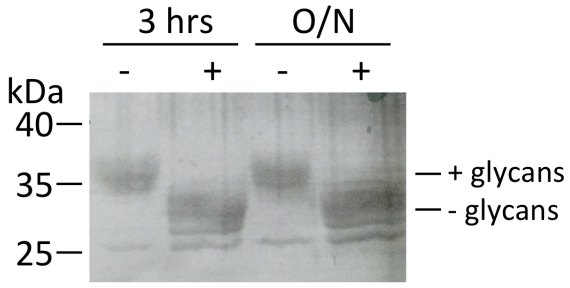


FIGURE 3.5: Deglycosylation trials of HpSP. HpSP was incubated with Endo H_f and Endo F3 at 37 °C (+). A control was completed without enzymes (-). A clear shift can be seen after 3 hours which does not increase after overnight incubation, suggesting that deglycosylation does not progress after 3 hours.

3.2.3 Haemoglobin

Haemoglobin is a protein that is easier to obtain from its natural source, rather than expressing it recombinantly, due it its high abundance and the complications of recombinant production due to the haem ring. To acquire the haemoglobin required for this project, red blood cells were pelleted from human blood by centrifugation, followed by sonication to lyse the cells. Purification was carried out by size exclusion chromatography. Haemoglobin was shown to be over 95% pure by SDS-PAGE (Figure 3.6).

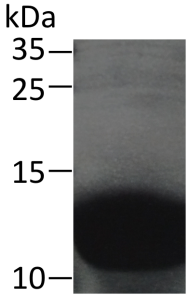


FIGURE 3.6: Purification of haemoglobin. SDS-PAGE of highly concentrated haemoglobin after purification by size exclusion chromatography using a Superdex 200 16/600 shows no visible contaminants.

3.2.4 Haptoglobin-haemoglobin complex formation

Hp has been shown to have a very high affinity for Hb [190], making the formation of the complex almost irreversible. To produce HpHb, excess Hb was mixed with type 1-1 haptoglobin and applied to a Superdex 200 16/600 gel filtration column (Figure 3.7). Haptoglobin 1-1 was used as it can only form dimers, whilst haptoglobin 2-1 and 2-2 form higher order multimers, increasing sample complexity. The HpHb complex (160 kDa) eluted in a separate peak from the excess haemoglobin (60 kDa), allowing separation of these two components.

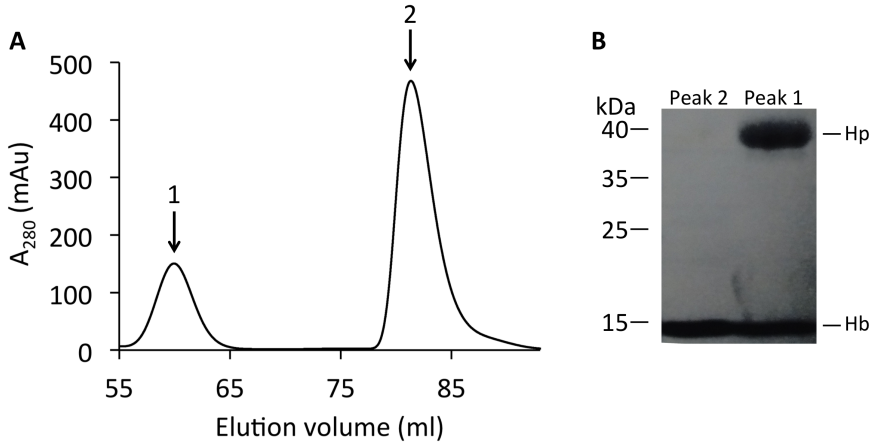


FIGURE 3.7: Purification of HpHb. (A) Hp was mixed with an excess of Hb before being applied to a Superdex 200 16/600 size exclusion column. HpHb eluted earlier, whilst the second peak contained excess Hb, as shown by SDS-PAGE (B). Hp 1-1 is a dimer where each haptoglobin molecule is comprised of two chains which dissociate under reducing conditions. One chain is ~40 kDa, seen on the gel, whilst the second chain is ~10 kDa and so would require a higher percentage gel to be observed.

HpSPHb was made by mixing HpSP with Hb. HpSPHb has a molecular weight of 60 kDa, very similar to the mass of a haemoglobin tetramer. Therefore separating these two components by size exclusion chromatography was not possible. Instead, when mixing HpSP with Hb an excess of HpSP was used, as free HpSP could be separated from HpSPHb by size exclusion chromatography. After mixing, the sample was applied to a Ni-NTA^{2+}

column and washed thoroughly with high salt buffer to remove any free haemoglobin. The elution from this column was applied to a Superdex 75 16/600 column (Figure 3.8). A peak corresponding to excess HpSP was observed, confirming that the larger molecular weight peak was unlikely to contain any free haemoglobin.

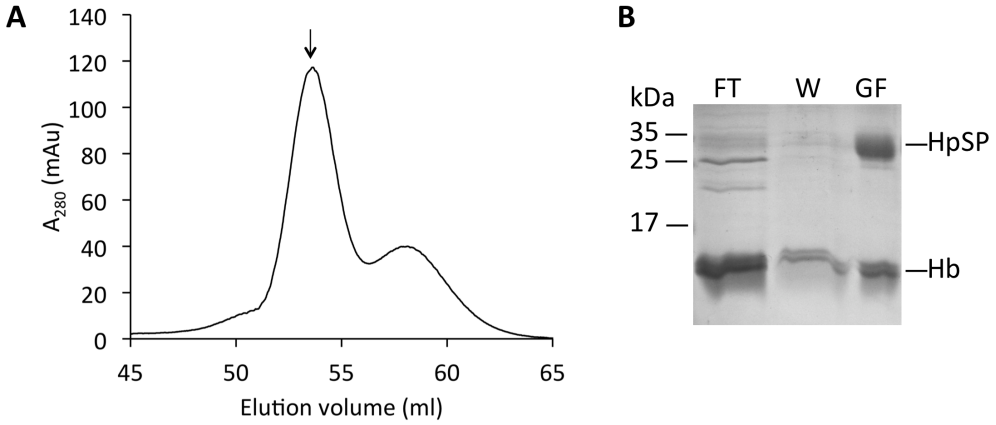


FIGURE 3.8: Purification of HpSPHb. (A) The gel filtration trace of HpSPHb after purification of the complex by Ni-NTA^{2+} affinity using the N-terminal His tag of HpSP shows two peaks. The first peak is HpSPHb, whilst the second peak contains excess HpSP (gel not shown). (B) Samples from the HpSPHb Ni-NTA^{2+} column (flow through and wash), and the gel filtration elution peak indicated by the arrow were run on SDS-PAGE. This demonstrates that the flow through and wash of the nickel column largely contain Hb, whilst the first gel filtration peak contains both components of the complex.

Initial attempts to make HprSPHb used the same protocol as for HpSPHb. However, when running the elution from the Ni-NTA^{2+} column on a Superdex 75 16/600 size exclusion column an additional peak was seen (Figure 3.9 A, B). Unbound HprSP had separated into two peaks, one of which (peak 2) overlapped with HprSPHb (peak 1) and one of which eluted later (peak 3). This suggests that HprSP is able to form multimers, accounting for the overlap of peaks 1 and 2. If the structure is, as expected, very similar to that of Hp, the multimerisation observed here is likely to be non-physiological. For all further purifications HprSP was first gel filtered using a Superdex 75 16/600 column (Figure 3.9C). This

separated the monomeric HprSP, which eluted at 60 ml, and the oligomeric HprSP, which eluted at 51 ml. Monomeric HprSP was then mixed with Hb following the same protocol as for HpSPHb. Oligomeric HprSP did not bind Hb, so was discarded.

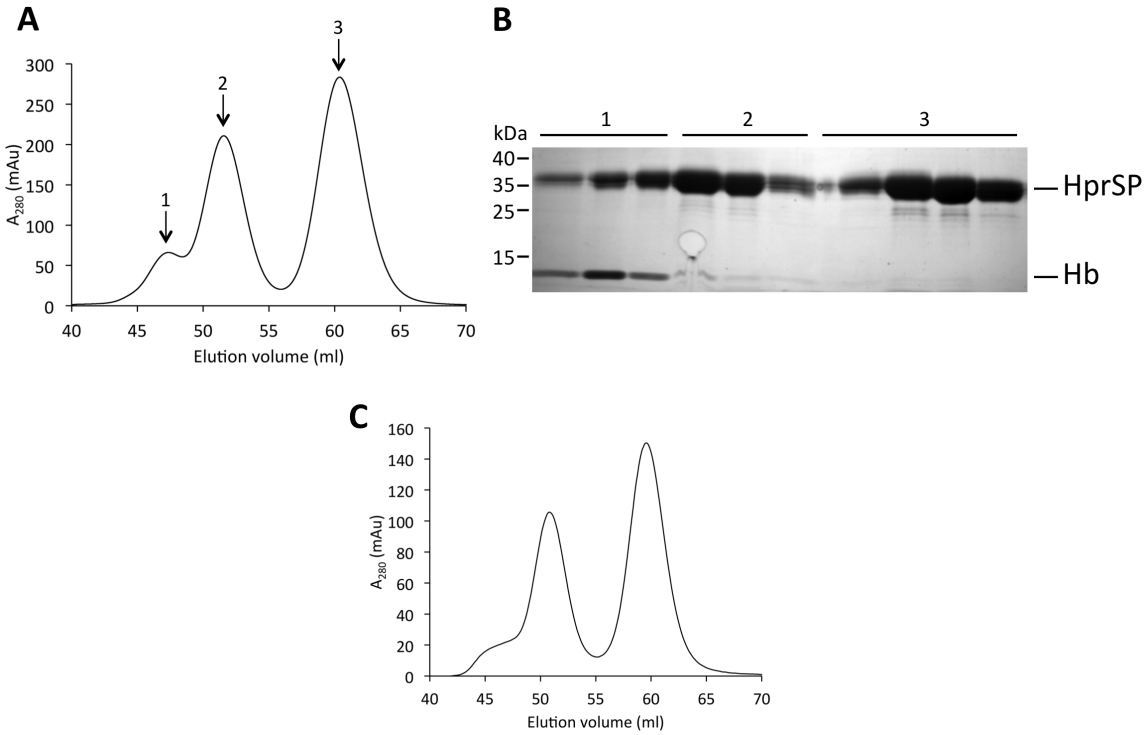


FIGURE 3.9: Discovery and removal of HprSP dimers by gel filtration. An attempt to produce HprSPHb using the same method used for HpSPHb gave three peaks on gel filtration using a Superdex 75 16/600 column (A). SDS-PAGE shows the that peak 1 contains HprSPHb, whilst peaks 2 and 3 both contain HprSP alone (B). In all subsequent purifications, HprSP was first subjected to size exclusion chromatography using a Superdex 75 16/600 column (C). This was able to separate monomeric and oligomeric HprSP.

3.3 Functional validation of HpSPHb and TbbHpHbRt

3.3.1 Assessing the affinities of truncated constructs using surface plasmon resonance

The affinity of full length TbbHpHbR for full length, dimeric HpHb, had already been found using SPR to be 1 μM [94]. To ensure that the truncated constructs used for crystallography contained the full binding site for the interaction, it was important to check that the affinity was unaltered when using these constructs. Additionally, haptoglobin normally undergoes a cleavage reaction, releasing R161, during maturation of the protein in the endoplasmic reticulum [129]. It was therefore also important to check that the lack of cleavage in this recombinant version did not alter the affinity.

To mimic the orientation that TbbHpHbR is displayed on the trypanosome cell surface, coupling the receptor to the surface of the SPR chip via its C-terminus was ideal. To accomplish this, a biotinylation acceptor peptide (BAP) tag was cloned into the construct at the C-terminus of the protein by PCR. After purification, this BAP-tagged protein was incubated with biotin and BirA enzyme, a biotin-protein ligase, overnight, resulting in the linking of a biotin moiety to the lysine of the BAP tag. CAPture chips can be used in SPR to couple biotinylated proteins to the chip. The CAPture chip is pre-coated with single-stranded DNA. A reagent containing complementary single-stranded DNA coupled to streptavidin is flown over the chip. This allows capture of the biotinylated protein, and also allows regeneration of the chip by disrupting the DNA duplex. This gave a K_D of 0.7 μM, comparable to the 1 μM affinity of full length TbbHpHbR for full length, dimeric HpHb (Figure 3.10A). This confirms that the HpSP domain and the TbbHpHbRt construct, along with Hb, contain the full binding sites for the interaction between the two.

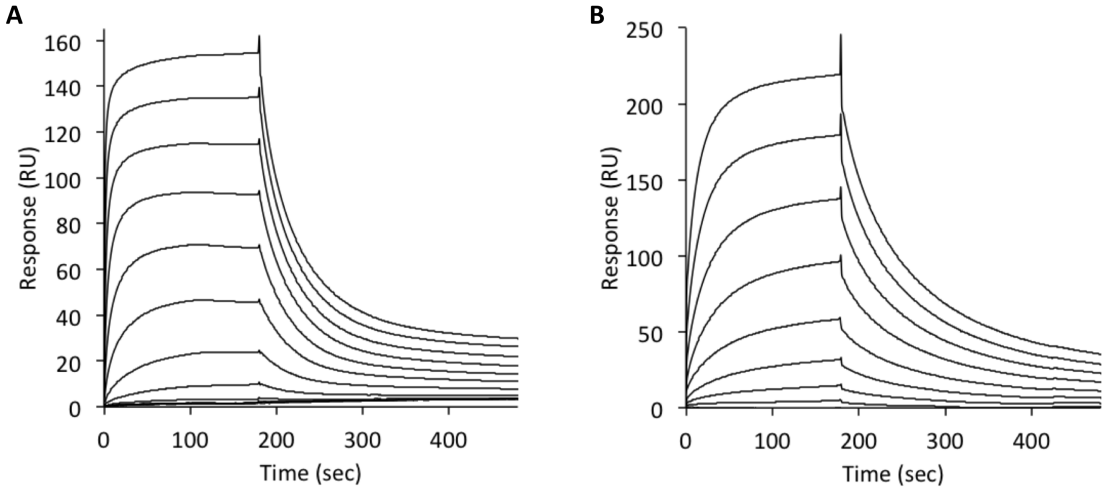


FIGURE 3.10: Surface plasmon resonance for the binding of ligands to immobilised TbbHpHbRt. Two-fold dilutions of (A) HpSPHb and (B) HprSPHb from a starting concentration of 16 μM were used to test the affinity for the truncated TbbHpHbR.

TLF1 does not contain haptoglobin, but instead haptoglobin-related protein, again in complex with haemoglobin. As HpHbR is known to be the uptake route for TLF1, it was therefore important to confirm that the receptor could bind HprHb. SPR was performed by flowing 1 in 2 dilutions of HprSPHb over a chip coated with biotinylated TbbHpHbRt (Figure 3.10B). This gave a K_D of 1.7 μM, comparable to the 0.7 μM affinity for HpSPHb, suggesting that the binding site for TbbHpHbR is conserved between HpHb and HprHb.

3.4 Structural characterisation of TbbHpHbRt and TbbHpHbRt:HpSPHb

3.4.1 TbbHpHbRt:HpSPHb complex formation

Excess HpSPHb was mixed with TbbHpHbRt before being applied to a Superdex 200 16/600 gel filtration column. Two peaks were observed (Figure 3.11). The higher molecu-

lar weight peak contained the complex of TbbHpHbRt:HpSPHb, whilst the second molecular weight peak contained excess HpSPHb.

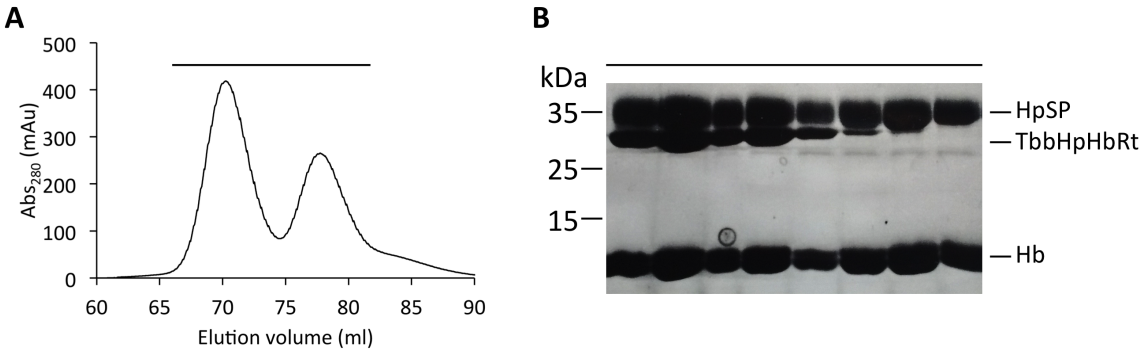


FIGURE 3.11: Formation of TbbHpHbRt:HpSPHb. (A) A gel filtration trace of TbbHpHbRt mixed with excess HpSPHb. (B) Samples from the elution volumes indicated by the horizontal line in A were run on SDS-PAGE. This showed that the first peak contained the complex, whilst the second peak contained excess HpSPHb.

3.4.2 Crystallisation

First, pre-crystallisation tests were carried out to determine what concentration of protein should be used for HpSPHb and TbbHpHbRt alone and in complex with each other. The results of this suggested that a concentration of 15 mg/ml would be suitable for all three. Initial crystallisation trials were carried out in 96-well format using vapour diffusion of sitting drops. The commercial screens used were JCSG+, ProPlex, MIDAS, Morpheus and PACT. These were set up at 4 °C for HpSPHb and TbbHpHbRt and at both 4 and 18 °C for the complex.

Small red cubic crystals appeared after 8 hours in well E2 of the JCSG+ screen containing HpSPHb (Figure 3.12). A 24-well optimisation plate was set up as a grid screen around this condition varying the concentrations of the buffer components, with the addition of seeds from the original well. After 3 days, larger cubic crystals appeared in this seeded plate. These were harvested and cryo-cooled in a cryoprotectant of the well

condition with an additional 30% glycerol.

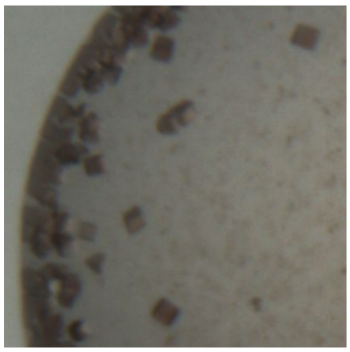


FIGURE 3.12: Images of HpSPHb crystals from well E2 of JCSG+ taken after 10 days. The condition contains 0.2 M NaCl, 0.1 M Na cacodylate pH 6.5, 2 M ammonium sulphate. The crystals in the optimisation plate were larger (20 μm), but similar in shape.

Small needle crystals appeared after 7 days in well G10 of the JCSG+ screen containing TbbHpHbRt. These grew larger and more regular in size over the following few days, and were harvested on day 15 (Figure 3.13A). More crystals were produced using the JCSG+ G10 condition with the silver bullet additives screen. Heavy metal soaking was carried out by soaking these native crystals in solutions of heavy metals overnight before cryo-cooling. SeMet crystals were grown in the same condition as native crystals, with the additive screen and seeds from native crystals (Figure 3.13B). These crystals were rougher around the edges but a number of harvestable crystals were still obtained.

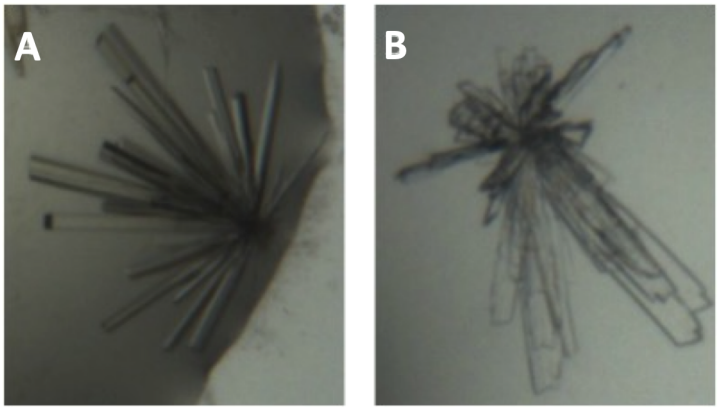


FIGURE 3.13: Crystallisation of TbbHpHbRt. (A) Native TbbHpHbRt and (B) selenomethionine-labelled TbbHpHbRt crystals were grown in the same condition. Both wells contain 0.15 KBr, 30% PEG 2000 MME. Although native and SeMet crystals were similar in length (200-300 μm), the SeMet crystals are less regular in shape, with less well-defined edges.

Crystals of TbbHpHbRt:HpSPHb appeared in Morpheus well B4 after 3 days of incubation at 18 °C. These crystals, which had a "broom-like" morphology, were used to seed a 96-well optimisation plate, using the same well condition combined with the silver bullet additive screen. This produced crystals of similar morphology in many wells, as well as some crystals with a "block-like" shape (Figure 3.14). SDS-PAGE confirmed that all three components were present in the crystals.

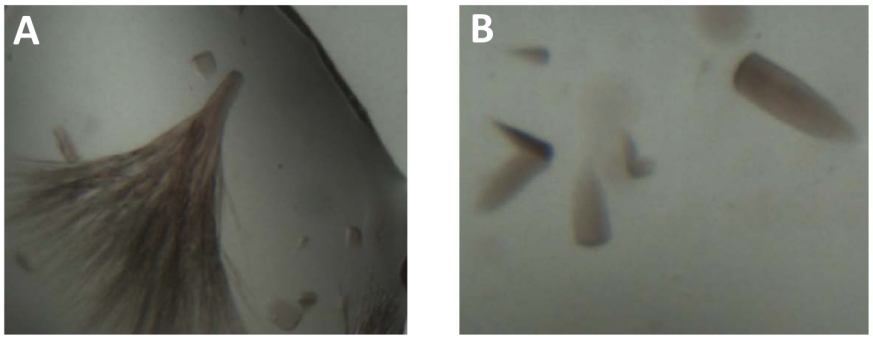


FIGURE 3.14: TbbHpHbRt:HpSPHb crystals. Crystals had either (A) a "broom-like" morphology or (B) a "block-like" morphology. All crystals were grown in 0.3 M NaFl, 0.3 M NaBr, 0.3 M NaI, 55 mM MES, 45 mM imidazole, 12.5% MPD, 12.5% PEG 1000, 12.5% PEG 3350, pH 6.5 combined with the silver bullet additive screen. After separation into individual crystals the longest dimension was 50-70 μm.

3.4.3 Data collection and reduction

HpSPHb crystals from the seeded 24-well plate diffracted well and a dataset to 2.05 Å resolution was collected. This dataset was initially indexed and integrated using iMOSFLM. The space group was identified by Pointless as $p3_121$ and the unit cell dimensions as a = b = 96.60 Å, c = 132.77 Å, $\alpha = \beta = 90.00°$, $\gamma = 120.00°$. Scala was then used to merge together multiple reflections from different diffraction patterns that are due to the same *hkl* coordinate, and to scale the intensities accordingly. The resolution was cut off at 2.05 Å to ensure that I/σ(I) remained above 2. The data reduction statistics are shown in Table 3.1. The dataset is 99.8% complete in the highest resolution shell, and the R_{PIM} is 37.4% for the highest shell and 8.1% overall. An overall value of less than 10% is preferable.

A TbbHpHbRt crystal from JCSG+ well G10 diffracted to 1.9 Å resolution and a dataset was collected. The data were initially indexed and integrated by iMOSFLM. Pointless was used to identify the space group as $p2_1$ and the unit cell dimensions as a = 27.90 Å b = 47.79 Å, c = 203.38 Å, $\alpha = \gamma = 90.00°$, $\beta = 92.79°$. Scala was then used to merge the data. These dimensions are large enough to accommodate a receptor, assuming similar dimensions to TcHpHbR. The resolution was cut off at 1.85 Å to ensure that I/σ(I) remained above 2 (Table 3.1). The dataset is 96.5% complete in the highest resolution shell, and the R_{PIM} is 42.9% for the highest shell and 4.5% overall.

The best diffracting crystal of the complex was from the seeded additive optimisation plate containing the well condition of Morpheus well B4, and the additives 0.2% w/v 2,2 -thiodiglycolic acid, 0.2% w/v apidic acid, 0.2% w/v benzoic acid, 0.2% w/v oxalic acid anhydrous and 0.2% w/v terephthalic acid. This crystal diffracted to 3.1 Å resolution and a dataset was collected. The dataset was indexed and integrated by XDS, which selected a resolution cutoff of 3.1 Å, a space group of c2, and unit cell dimen-

sions of a = 223.4 Å, b = 56.59 Å, c = 65.29 Å, $\alpha = \gamma = 90.00°$, $\beta = 92.99°$. The I/σ(I) is 1.6, whilst the R_{PIM} values overall and for the highest resolution shell are 6.3% and 72.6% respectively (Table 3.1).

TABLE 3.1: Data processing statistics.

	HpSPHb	TbbHpHbR	TbbHpHbR:HpSPHb
Beamline	Diamond I04-1	Diamond I03	Diamond I03
Space group	$p3_121$	$p2_1$	c2
Cell dimensions			
a, b, c (Å)	96.6, 96.6, 132.77	27.90, 47.79, 203.38	223.4, 56.59, 65.29
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 92.79, 90.00	90.00, 92.99, 90.00
Resolution (Å)	2.05	1.85	3.1
Wavelength (Å)	0.916	0.9763	0.9750
R_{PIM} (%)	8.1 (37.4)	4.5 (42.9)	6.3 (72.6)
I/σ(I)	8.7 (2.3)	10.2 (2.0)	9.8 (1.6)
Completeness (%)	99.8 (100)	97.4 (96.5)	96.9 (97.1)
Multiplicity	9.6 (10.2)	3.1 (3.1)	3.2 (3.3)

3.4.4 Phasing and model refinement flow path

The order in which the three structures were phased and refined was not straightforward, so shall be outlined here. First, the HpSPHb dataset was phased by molecular replacement using the porcine HpHb structure as a search model [133]. The model was then refined. Phasing attempts of the TbbHpHbRt dataset by both molecular replacement using TcHpHbR as a search model [94] or experimental phasing using selenomethionine incorporation or heavy atom soaking were all unsuccessful. This was therefore temporarily abandoned whilst the structure of the complex was pursued. The complex dataset was successfully phased by molecular replacement using the human HpSPHb structure as a search model. After several rounds of refinement the electron density for TbbHpHbRt was strong enough that a three helical bundle could be built into this.

Returning to the higher resolution dataset for TbbHpHbRt alone, it was now possible to phase this by molecular replacement using the low resolution model obtained from the complex structure. After building and refinement of this model was complete, this high resolution structure, along with that of HpSPHb, was used to provide restraints during refinement of the complex structure.

3.4.5 HpSPHb phasing, model refinement and validation

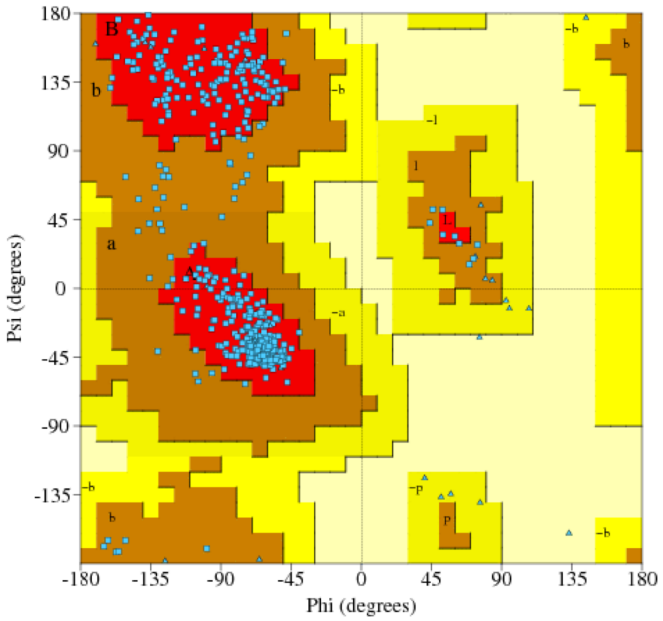
Molecular replacement for HpSPHb was straightforward, completed by Phaser using the porcine structure (pdb:4F4O) [133] as a search model. To generate a search model, the haem rings, oxygens and glycosylations were deleted, and the haptoglobin molecule was truncated so that only the SP domain remained. The sequence identity between the search model and human HpSPHb is 80%, so it was not surprising that molecular replacement was successful. The solution contained one copy of the complex in the asymmetric unit. The haem rings were visible in the electron density despite not being present in the search model, giving confidence that the solution was correct. A reduction in R_{work} from 55% to under 40% after rigid body refinement is a good indication that the model is correct [191]. In this case the value reduced from 45% to 30%, with an R_{free} of 34%. Again, this suggested that the model was correct.

The structure was then completed using iterative cycles of model building using Coot and refinement using REFMAC. Where required, the side chains were corrected from those found in the porcine model to those found in human HpHb. Water molecules and components of the crystallisation or cryo-cooling solution were added: a cacodylate ion, a sulphate ion, and two glycerol molecules. Side chain positions were altered to fit the density. Lastly, a glycosylation, Fucα1,6GlcNAc attached to N241, could be seen in the density

and was added to the model accordingly. This corresponds to the moiety that remains after treatment with Endo H_f and Endo F3. The structure contains all residues of haemoglobin α, and all but the first and last residues of haemoglobin β. The haptoglobin chain begins at the start of the sequence for HpSP, but is missing the N-terminal 6-His tag, which was not removed after purification. It is likely that this is flexible, and thus cannot be seen in the electron density. After refinement was completed, the R_{work} and R_{free} values had reduced to 18.0% and 22.4% respectively, whilst the root mean square deviations for bond lengths (0.02 Å) and angles (2°) were at or below accepted values (0.02 Å and 4°) (Table 3.2). The Ramachandran plot generated by PROCHECK revealed 89% of residues were within the allowed region, with the remaining 11% were in the additionally allowed regions, giving no cause for concern (Figure 3.15).

TABLE 3.2: Data refinement statistics.

	HpSPHb	TbbHpHbR	TbbHpHbR:HpSPHb
Resolution (Å)	2.05	1.85	3.1
No. reflections	43,170	44,685	17,302
R_{work} / R_{free} (%)	18.0 / 22.4	19.8 / 24.0	19.5 / 21.7
No. of protein residues in model	544	523	784
No. of copies in the asymmetric unit	1	2	1
R. m. s. deviations			
Bond lengths (Å)	0.020	0.017	0.012
Bond angles (°)	2.0	1.6	1.5



	No. of residues	Percentage
Allowed region	421	89.0
Additionally allowed region	52	11.0
Generously allowed region	0	0
Disallowed region	0	0
Residues not Gly or Pro	473	
End residues (excl. Gly or Pro)	6	
Gly residues	38	
Pro residues	27	
Total	544	

FIGURE 3.15: Ramachandran plot and statistics of the HpSPHb structure generated by PROCHECK [189]. Triangles represent proline and glycine residues, whilst all other amino acids are represented by squares. The red region indicates allowed angles, the brown region additionally allowed angles, the bright yellow region generously allowed angles and the pale yellow disallowed angles. Prolines and glycines with acceptable angles may appear in the pale yellow region, as they have different allowed angles to the other amino acids.

3.4.6 TbbHpHbRt phasing, model refinement and validation

To solve the structure of TbbHpHbRt, molecular replacement using TcHpHbR as a search model was first attempted. This search model was prepared by trimming all side chains to alanine using CHAINSAW, part of the CCP4 suite. Unfortunately this was not successful, perhaps unsurprising when the low sequence identity of 28% between the search model and TbbHpHbRt is considered.

In an attempt to obtain an anomalous dataset for TbbHpHbRt a number of heavy metal compounds were tested. In most cases diffraction was maintained at a high level but no absorption edge could be detected, suggesting the metal had not bound (Figure 3.16A). Unfortunately, the only condition which showed binding, OsO_4, also severely reduced the diffraction of the crystal, as very few spots, if any, were visible on the diffraction patterns. Other compounds showed only a very small edge at the correct position on the fluorescence scan and as thorough back-soaking was not completed after the overnight soak, this small amount of fluorescence is likely to be due to left over heavy metal solution, rather than any bound atoms. Therefore, although these crystals diffracted well, they could not be used for phase determination.

The SeMet crystals of TbbHpHbRt showed a strong absorption edge, suggesting that SeMet had been successfully incorporated during expression (Figure 3.16B). However, the quality of diffraction of these crystals varied, with the best reaching approximately 7 Å. Others did appear to have diffraction spots at higher resolution, but in these cases the spots were streaky, and the data could not be indexed. Only 10 of these crystals were harvestable, and it is possible that if the sample size had been bigger, one may have been found that diffracted suitably well.

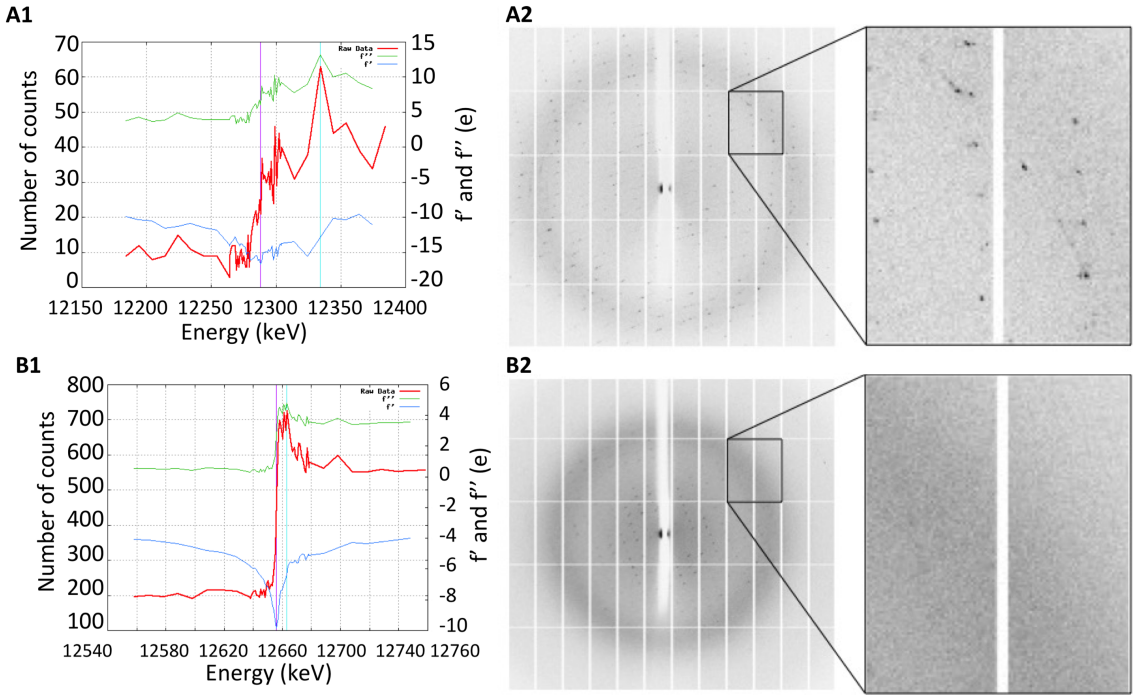
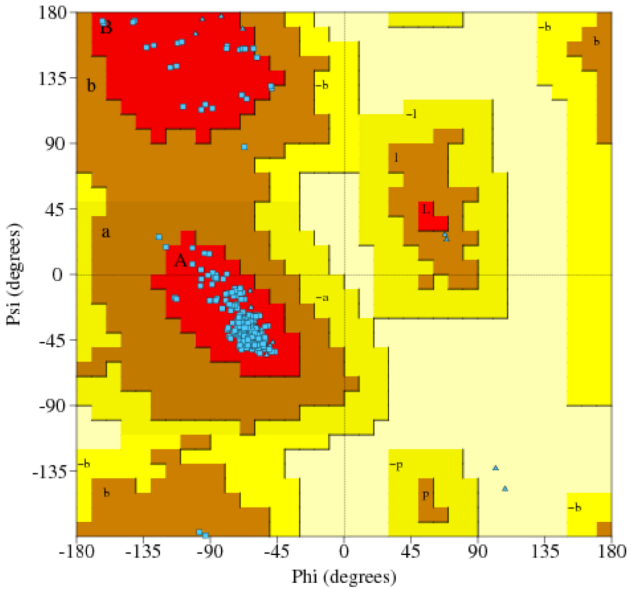


FIGURE 3.16: Attempts to obtain a phasing dataset for TbbHpHbRt. (1) Fluorescence scans and (2) diffraction patterns for (A) native TbbHpHbRt crystals soaked in heavy metals and (B) selenomethionine TbbHpHbRt crystals. The crystal used in (A) had been soaked in $Hg(CN)_2$ overnight. Heavy metal soaking was not successful, as no clear absorption edge is seen. Incorporation of selenomethionine was successful, but the resolution of the data is significantly worse.

Once the complex structure had been solved, a fresh attempt to phase this dataset by molecular replacement could be carried out using the polyalanine model of TbbHpHbRt obtained from the complex structure as a search model. This was successful, and the solution contained two copies of the receptor in the asymmetric unit. During cycles of refinement using REFMAC the location of bulky side chains and the disulphide bond present in the head structure of the receptor were revealed, allowing the correct sequence to be built into the structure. After several rounds of refinement the density of loops at each end of the helical bundle were also visible, allowing all residues of the receptor to be built, with the exception of two residues at the N-terminus of both molecules, two residues at the C-terminus of molecule A, and one residue from the C-terminus of molecule B. Water

molecules were also added to the model. The final R_{work} and R_{free} values were 19.8% and 24.0% respectively, whilst the root mean square deviation for bond lengths was 0.017 Å and for angles was 1.6° (Table 3.2). The Ramachandran plot showed only 1.2% of residues fell into the additionally allowed region, whilst the remaining 98.8% were in the allowed region (Figure 3.15).



	No. of residues	Percentage
Allowed region	485	98.8
Additionally allowed region	6	1.2
Generously allowed region	0	0
Disallowed region	0	0
Residues not Gly or Pro	491	
End residues (excl. Gly or Pro)	4	
Gly residues	28	
Pro residues	0	
Total	523	

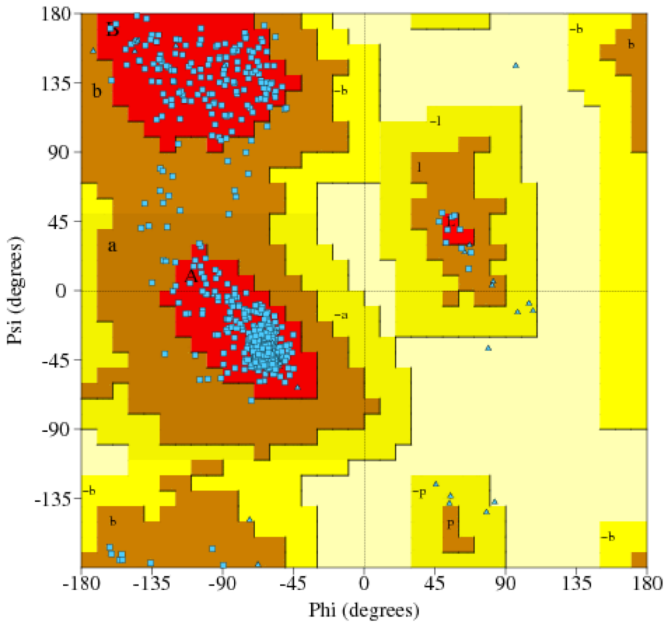
FIGURE 3.17: Ramachandran plot and statistics of the truncated TbbHpHbR structure generated by PROCHECK [189].

3.4.7 HpSPHb:TbbHpHbRt phasing, model refinement and validation

Molecular replacement for the complex formed between HpSPHb and TbbHpHbRt was attempted using the structure of human HpSPHb obtained above. A search model was produced by removing all water molecules and other buffer components, the glycosylation and the haem rings. This model comprised roughly two thirds of the molecular weight of the complex, so was likely to be sufficient. Indeed, a single solution was output, and looked promising. This solution contained a single copy of the receptor and HpSPHb in the asymmetric unit. The positions of receptor helices were identified using Buccaneer [188], allowing a polyalanine model of *T. brucei* HpHbR to be built by a cycle of model building and refinement in Coot and Buster as the electron density improved. Once this polyalanine model had been used to solve the structure of TbbHpHbRt alone using the high resolution dataset, this high resolution model was used to guide building of the complex structure, and was used as a restraint for refinement, along with the high resolution structure of HpSPHb.

No significant structural changes were seen between the individual components and their positions in the complex, with the exception of side chain movements. The haemoglobin α-chain is complete, whilst the β-chain is lacking only the first and final four residues. As with the structure of HpSPHb alone, the haptoglobin chain starts where the HpSP sequence starts, missing the purification tag, and contains all residues at the C-terminus. The receptor is lacking three residues at the N-terminus and two at the C-terminus. The final R_{work} and R_{free} values were 19.5% and 21.7% respectively (Table 3.2). These are perhaps lower than expected for a structure of this moderate resolution. However, this can be explained by the fact that two high resolution structures of the individual components were used as restraints during refinement. The root mean square deviations for

bond lengths (0.012 Å) and angles (1.5°) were very respectable, whilst the Ramachandran plot showed 93.0% of residues in the allowed region and 7.0% in the additionally allowed region (Figure 3.18).



	No. of residues	Percentage
Allowed region	652	93.0
Additionally allowed region	49	7.0
Generously allowed region	0	0
Disallowed region	0	0
Residues not Gly or Pro	701	
End residues (excl. Gly or Pro)	8	
Gly residues	50	
Pro residues	25	
Total	784	

FIGURE 3.18: Ramachandran plot and statistics of the HpSPHb:TbbHpHbRt structure generated by PROCHECK [189].

3.4.8 The structure of TbbHpHbRt and TbbHpHbRt:HpSPHb

The overall structure of TbbHpHbRt is similar to that of TcHpHbR, comprising an elongated three-helical bundle 118 Å long (Figure 3.19A). At the membrane-distal end, the receptor forms a compact head structure which includes two additional short helices. This head structure is very similar to the head structure of TcHpHbR, with a root mean square deviation between the two of 1.1 Å (Figure 3.19B).

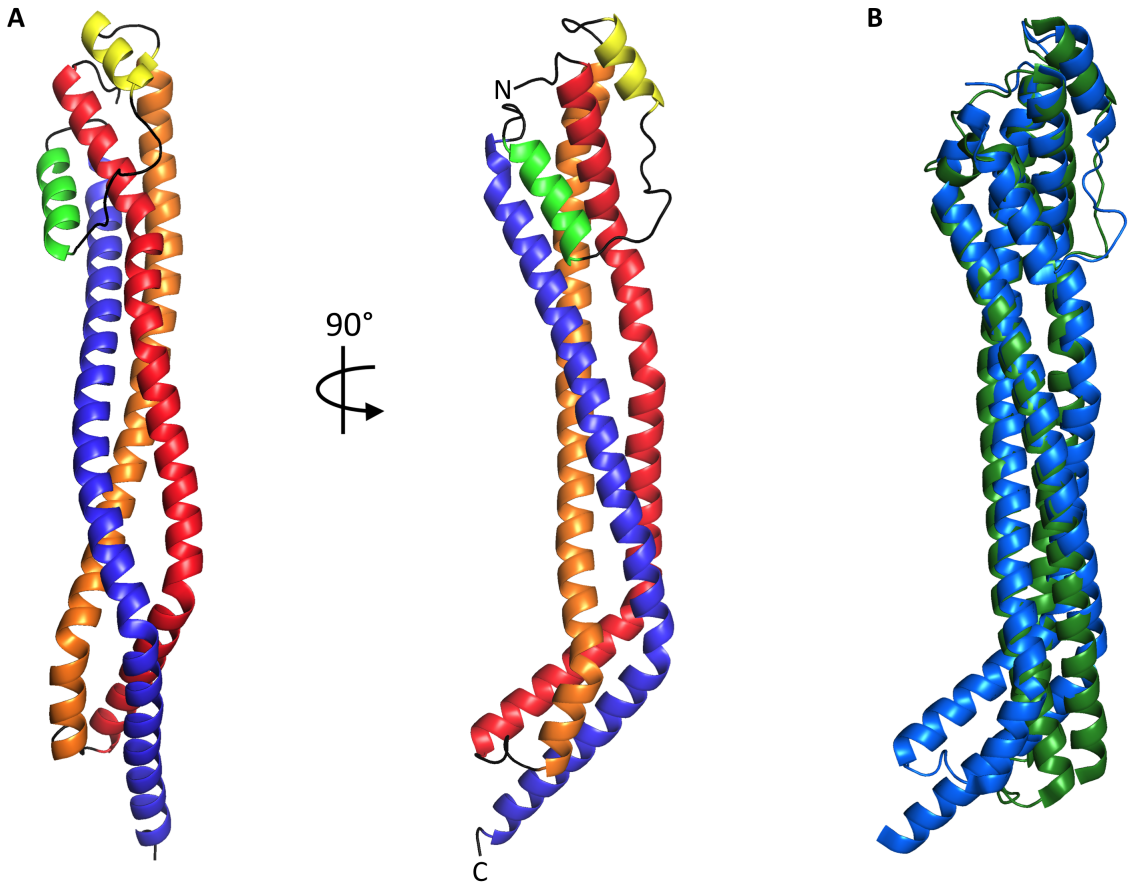


FIGURE 3.19: The structure of the *T. brucei* HpHbR. (A) The three main helices form an elongated bundle with a ~50° kink towards the membrane-proximal C-terminal end and an N-terminal head structure. (B) Alignment of the TbHpHbR and TcHpHbR structures. *T. brucei* is shown in blue and *T. congolense* in green. The root mean square deviation between the head structures of the two is 1.1 Å, demonstrating that the structures are very similar.

The most striking difference between the HpHbRs of the two species is the 50° kink near the membrane-proximal end of TbbHpHbRt. This feature is seen in both the TbbHpHbRt structure, and the structure of the complex, suggesting that it is a rigid feature. The packing of the three helices against each other is the result of a pattern of hydrophobic and hydrophilic residues on each helix, with the helices packing together in such a way as to form a hydrophobic core (Figure 3.20). In *T. brucei* this pattern is disrupted, which causes a distortion in the paths of the helices, resulting in the rigid kink.

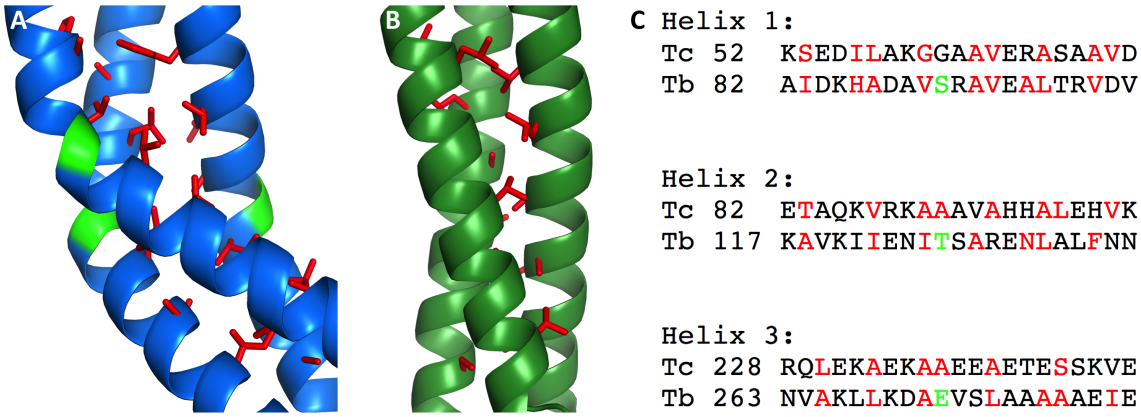


FIGURE 3.20: The *T. brucei* receptor has a kink due to a change in the pattern of hydrophobic and hydrophilic residues. Hydrophobic residues are coloured red, whilst the kink site is coloured green. This reveals that in all helices a hydrophobic residue appears earlier in the sequence following the kink, causing a tighter helix which results in the formation of a kink.

Previously, mutational studies suggested that the HpHb binding site was found near the membrane-distal end of TbbHpHbR [94]. This is as expected, as binding must occur whilst the receptor is in the context of a densely packed layer of VSG protein. However, the structure of the complex formed between TbbHpHbRt and HpSPHb reveals that the binding site extends down the length of the receptor (Figure 3.21). The residues identified by mutational studies form part of the haemoglobin β-subunit binding site, which has a surface area of 745 Å^2, whilst no interactions are seen with the haemoglobin α-subunit. Side chains from helix I of the receptor account for a large proportion of this

mainly hydrophilic interaction, with helix II and the loop between helices III and IV making additional contacts. The binding site on haemoglobin is a groove formed by helices C and F of the β-subunit. The propionate chains of the haem group also make direct contacts with the receptor, accounting for 140 Å^2 of the total contact are for Hb.

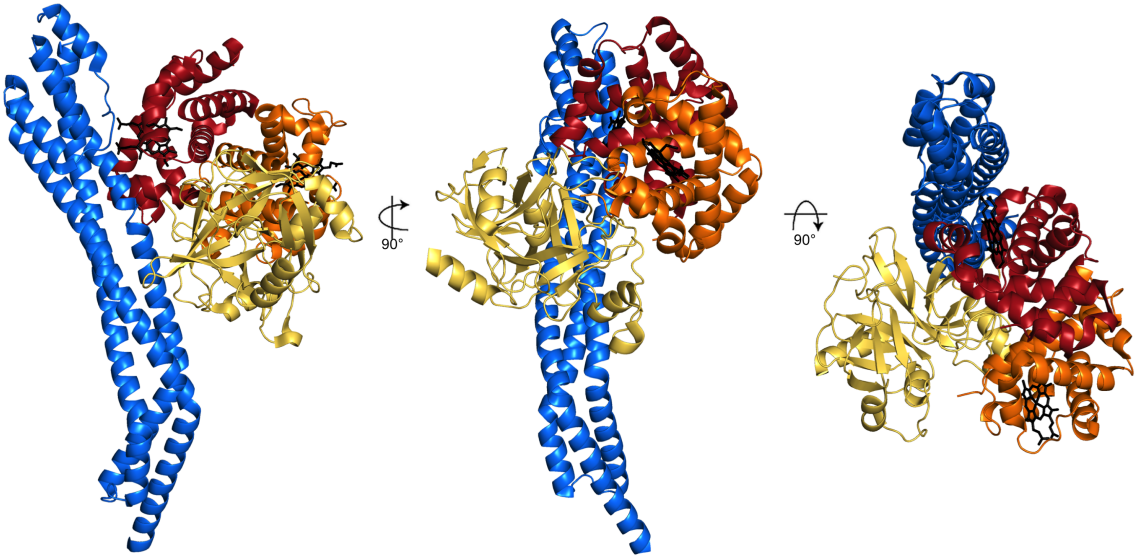


FIGURE 3.21: The structure of TbbHpHbRt:HpSPHb. TbbHpHbRt is in blue, HpSP in yellow, haemoglobin α-subunit in orange and β-subunit in red.

The membrane-proximal binding site interacts with the haptoglobin SP subunit, and has a total surface area of 505 Å^2. Again these contacts are mainly made by helix I of the receptor, although in this case the interactions are mainly hydrophobic in nature. During maturation of haptoglobin a natural cleavage event occurs which results in the release of arginine 161. In this structure, residues 154-170 are missing in this structure, suggesting that this loop is flexible or is susceptible to cleavage. The former is more likely as these residues can be seen in the structure of HpSPHb alone. This loop is distant from both the haemoglobin and TbbHpHbt binding site, so the cleavage event is unlikely to affect the affinity of HpSP for its binding partners (Figure 3.22). The contribution of both haemoglobin and haptoglobin to the binding site for TbbHpHbR explains why the receptor

can only bind the complex of the two, and not the individual components.

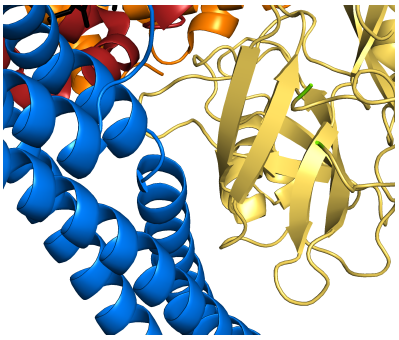


FIGURE 3.22: The site of natural cleavage of Hp is missing from the structure. The two residues in green, 153 and 171, indicate where the electron density becomes to weak to model. Cleavage normally occurs at residues 160 and 162, resulting in the release of residue 161.

As the haemoglobin subunits have very similar structures, it was important to check whether HpHb could bind to the receptor with the haemoglobin α-subunit making contacts with the upper binding site of the receptor. A model to investigate this was produced by aligning the haemoglobin α-subunit of HpSPHb with the β-subunit seen in the crystal structure (Figure 3.23). However, in this orientation the haptoglobin SP subunit is distant from the receptor and could not make the contacts required for this interaction to occur. Therefore it seems likely that the only binding mode for HpHb and TbbHpHbR is the one seen in the crystal structure.

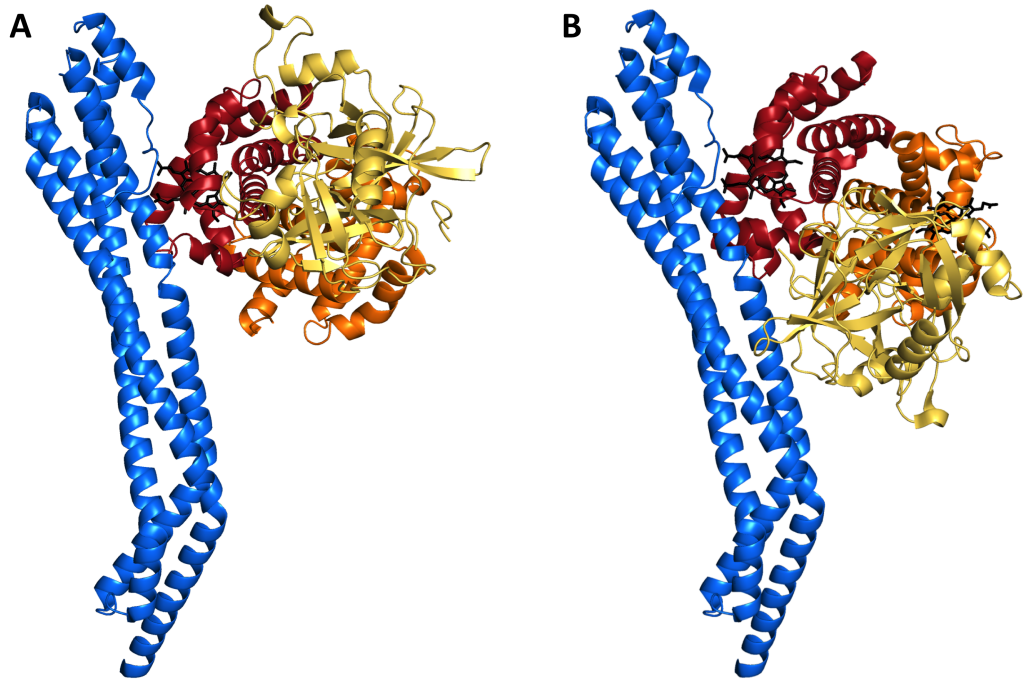


FIGURE 3.23: Haptoglobin cannot interact with TbbHpHbR when the haemoglobin subunits are flipped. (A) The model was produced by aligning the haemoglobin α subunit of HpSPHb with the β-subunit in the complex structure. Although the contacts made by haemoglobin are maintained, the haptoglobin molecule is now too distant from the receptor to make contacts. (B) shows the orientation seen in the crystal structure, with both the haemoglobin β subunit and haptoglobin forming contacts with the receptor.

TLF1 contains haptoglobin-related protein rather than haptoglobin, in complex with haemoglobin. The SP domain of Hpr contains 16 amino acid substitutions compared to Hp (Figure 3.24). The structure of the complex reveals that none of these substitutions are found near the receptor binding site. This, along with the very similar affinity determined by SPR, suggests that TLF1 binds to the receptor using the same binding site as HpSPHb.

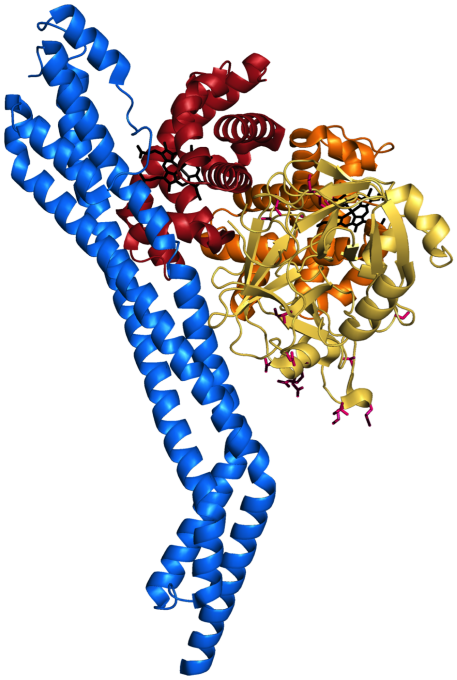


FIGURE 3.24: Differences between haptoglobin and haptoglobin-related protein. Residues that are different between the two proteins are coloured in pink in this structure of HpSPHb. None of these are near the binding interface with TbbHpHbR.

The structure of the TbbHpHbRt:HpSPHb complex also provides insight into the purpose of the kink seen in the TbbHpHbR. One consequence is to increase the effective diameter of the receptor, which may push apart the VSG molecules surrounding it and therefore provide greater access of the extracellular environment to the binding site. The inside of the kink is positioned on the opposite side of the receptor to the binding site. This orientation results in even greater exposure of the HpHb binding site, making it more accessible to the ligand.

Another role of the kink can be seen when two of the complex structures are docked onto the structure of dimeric porcine HpHb (Figure 3.25). In this model, the kink causes the C-termini of both receptors to point in the same direction. Although the structure of the C-terminal domain of this receptor is unknown, it is likely that this orientation allows

two receptors to bind to one dimeric HpHb. This could increase the efficiency of uptake of HpHb by an avidity mechanism, a theory which is explored in the latter part of this chapter.

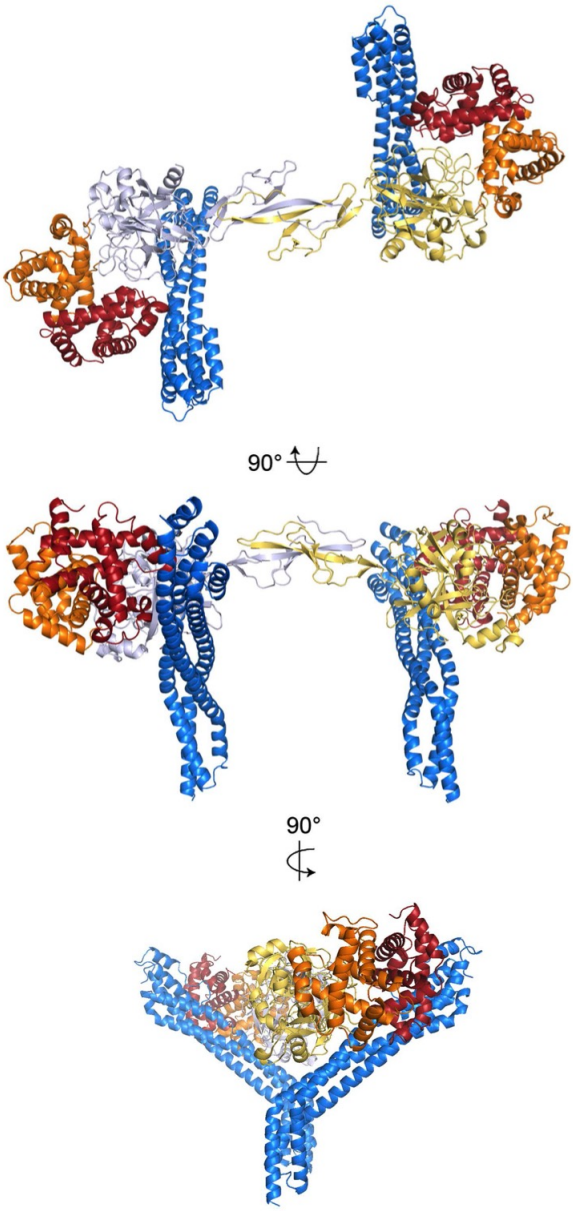


FIGURE 3.25: A model for a complex of one HpHb dimer bound to two receptors. This was generated by docking the structure of the TbHpHbR:HpSPHb complex onto that of porcine HpHb (pdb:4F4O) [133]. The C-termini of both receptors point in the same direction, suggesting that this orientation may be possible with both receptors anchored to the membrane through the C-terminal domains.

3.4.9 Solution studies of the structure of HpHbR and its ligand using SAXS

Small angle x-ray scattering (SAXS) can be used to validate the structures modelled using x-ray crystallography, confirming that these models are similar to the structures of the proteins in solution. For HpSPHb, TbbHpHbRt, the complex formed between them and HpHb alone, this was relatively straightforward. Data collection was carried out on 2-fold dilutions from high concentrations of 2 mg/ml. The scattering curves were averaged, and the buffer scattering subtracted. The scattering curves for each species were then scaled and combined to produce a single scattering curve (Figure 3.26A).

These curves were used to generate twenty *ab initio* models of each protein, which were averaged to produce molecular envelopes. The crystal structures of all three fit well into the envelopes. In particular, the kink of TbbHpHbRt can also be seen in the SAXS envelope, further supporting the assumption that this is a rigid feature of the structure.

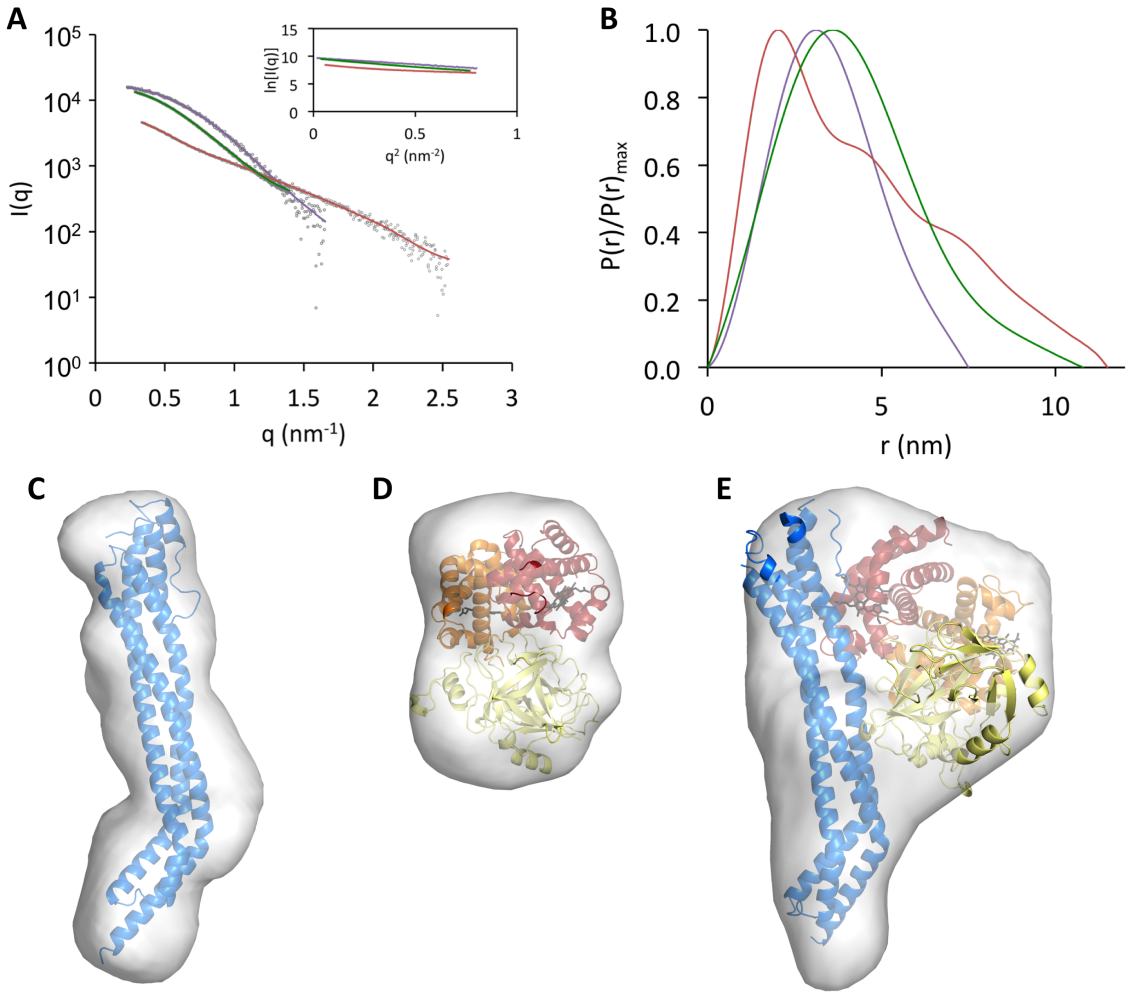


FIGURE 3.26: Small angle x-ray scattering analysis of TbbHpHbRt, HpSPHb and the complex they form. Curves corresponding to TbbHpHbRt, HpSPHb and the complex formed between them are shown in red, purple and green respectively. (A) The theoretical scattering calculated from *ab initio* reconstructions superimposed into experimental scattering data. Guinier plots are shown as an insert. (B) Distance distribution functions derived from small angle x-ray scattering. The *ab initio* molecular envelopes calculated from scattering data of (C) TbbHpHbRt, (D) HpSPHb and (E) the complex formed between them. The envelopes were visualised using Situs, and the crystal structures were docked into these using Sculptor.

The apparent molecular weights were calculated from the volumes predicted from Porod plots (Table 3.3). Although the calculated weights are low, they do correlate well with each other, as the values of the two components add to give the value of the complex.

TABLE 3.3: Summary of SAXS parameters for HpSPHb, TbbHpHbRt and their complex. The radius of gyration (R_g) was calculated by AutoRg. The maximum particle dimension (D_{MAX}) and Porod volume were calculated by GNOM. M_{Wapp} was calculated by assuming an average density of 0.5 Da/Å^3.

	Mw (kDa)	R_g (nm)	D_{MAX} (nm)	Volume (nm^3)	M_{Wapp} (kDa)
TbbHpHbRt	32.2	3.3	11	43	21.55
HpSPHb	59.7	2.6	7.5	75	37.5
TbbHpHbRt:HpSPHb	91.8	3.1	11.5	112	56.14

For the complex formed between two TbbHpHbRt molecules and dimeric HpHb, the process of data collection was slightly more complicated. The method of serial dilutions as above resulted in a C-shaped envelope. The vertical part of the "C" fit the crystal structure of dimeric porcine HpHb well. However, the arms of the "C" were too short to fit the receptor. This suggested that the complex was in equilibrium, and that the envelope obtained was an average of one, two or no receptors bound to one HpHb dimer. Therefore a different experimental set-up using inline SEC-SAXS was required, where SAXS data is collected from the sample as it emerges from a size exclusion column. Frames corresponding to the earliest eluting half of the peak containing TbbHpHbRt in complex with HpHb were averaged as they will have a greater fraction of two receptors bound to one HpHb dimer. The buffer scattering was subtracted, and the rest of the analysis was completed as described previously.

The envelope confirms that two TbbHpHbRts are able to bind one HpHb dimer in solution (Figure 3.27). The dimensions of the C-shaped envelope for the TbbHpHbRt:HpHb complex are 165 Å across the top, and 120 Å for each of the arms. This matches the lengths of both HpHb (175 Å) and TbbHpHbRt (118 Å) from crystal structures well. As before, the calculated molecular weights add up to the value for the complex (Table 3.4).

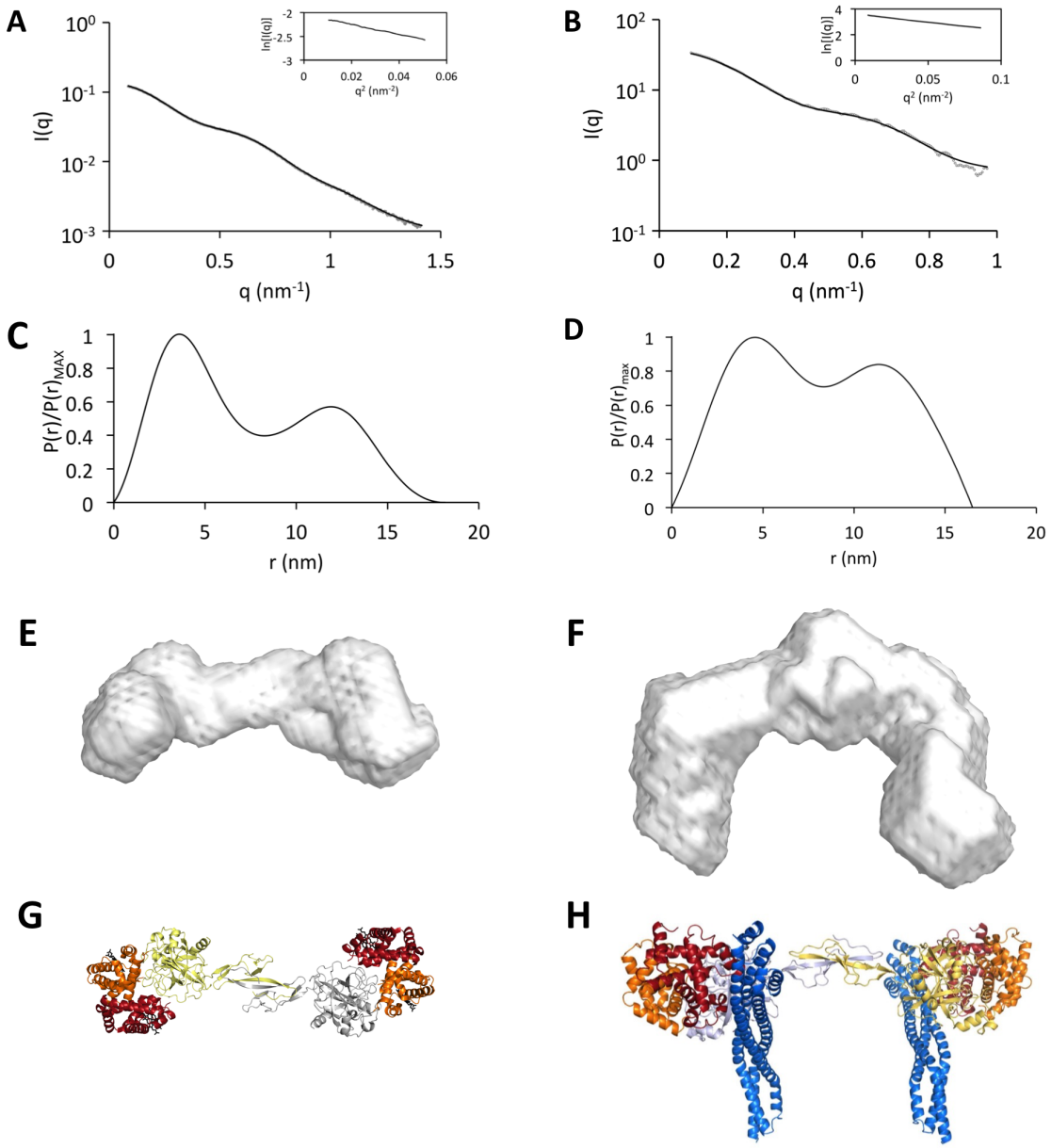


FIGURE 3.27: Small angle x-ray scattering analysis of HpHb, alone and in complex with TbHpHbR. (A) and (B) show the theoretical scattering calculated from *ab initio* reconstructions for HpHb and TbbHpHbRt:HpHb respectively, superimposed into experimental scattering data. Guinier plots are shown as an insert. (C) and (D) show distance distribution functions of HpHb and TbbHpHbRt:HpHb respectively, derived from small angle x-ray scattering. (E) and (F) show the *ab initio* molecular envelopes calculated from scattering data from HpHb and TbbHpHbRt:HpHb complex respectively, with (G) and (H) showing the crystal structure and model of porcine HpHb (pdb:4F4O) [133] and HpHb bound to two receptors respectively for comparison. The envelopes were visualised using Situs, and the crystal structures were docked into these using Sculptor.

TABLE 3.4: Summary of SAXS parameters for HpHb alone and in complex with TbbHpHbRt. The radius of gyration (R_g)was calculated by AutoRg. The maximum particle dimension (D_{MAX}) and Porod volume were calculated by GNOM. M_{Wapp} was calculated by assuming an average density of 0.5 Da/Å^3.

	Mw (kDa)	R_g (nm)	D_{MAX} (nm)	Volume (nm^3)	M_{Wapp} (kDa)
HpHb	153	5.9	18.1	290	145
TbbHpHbRt:HpHb	217	6.3	16.5	370	185

3.5 Testing the dimeric model of HpHb uptake by TbbHpHbR

3.5.1 Solution studies of the stoichiometry of TbbHpHbR:HpHb using SEC-MALLS

To further investigate the complex formed between TbbHpHbRt and HpHb in solution, SEC-MALLS was performed. This was completed for both components alone, as well as a mixture of the two with an excess of TbbHpHbRt. This revealed approximate masses of 30 kDa and 150 kDa for TbbHpHbRt and HpHb respectively, whilst the complex of the two had a mass of 210 kDa. This data suggests that two receptors are capable of binding to one HpHb dimer in solution, supporting the data obtained by SAXS.

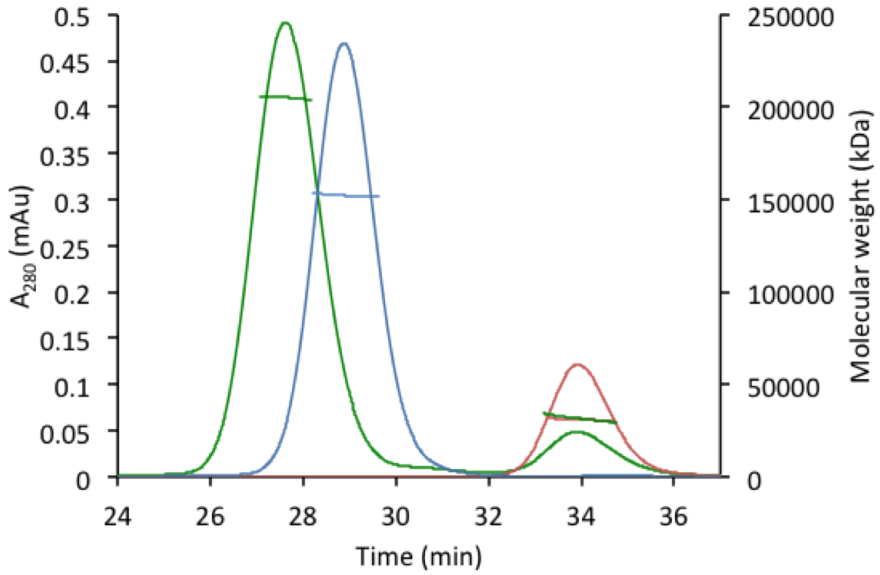


FIGURE 3.28: SEC-MALLS data to assess the stoichiometry of the TbHpHbRt:HpHb complex. Runs were completed using a Superdex 200 10/300 column connected to a multi-angle laser light scattering detector. HpHb (blue) has a molecular weight of 150 kDa, whilst the truncated TbbHpHbRt (red) has a molecular weight of 30 kDa. The two peaks seen for the trace corresponding to the complex of both components (green), made using a molar excess of TbbHpHbRt, have molecular weights of 210 kDa and 30 kDa. The latter corresponds to excess TbbHpHbRt. The former corresponds to two TbbHpHbRt molecules bound to one HpHb molecule.

3.5.2 Trypanosome uptake experiments

To test whether the dimeric binding seen using SEC-MALLS and SAXS can occur at the trypanosome cell surface, the efficiency of uptake of monomeric and dimeric HpHb was assessed. Fluorescent labelling of ligands was carried out using the Alexa-488 protein labelling kit. Hp and HpSP were labelled alone before mixing with Hb and purification of the complexes as normal. Uptake experiments were then performed by Dr Paula MacGregor in the Carrington lab, Cambridge. *T. brucei* were incubated with increasing concentrations of either fluorescently-labelled ligand. Bovine serum albumin (BSA) was used as a negative control to assess fluid phase uptake. The experiment was then repeated in a TbbHpHbR$^{-/-}$ null cell line in which both copies of the receptor were disrupted. Uptake

was monitored via flow cytometry. At these concentrations negligible uptake of BSA was observed, and uptake of HpSPHb and HpHb into the null cell line was comparable to that of BSA. This demonstrates that any uptake of HpSPHb and HpHb into wild type cells was due to TbbHpHbR. In wild type *T. brucei* cells, uptake of HpHb reached saturation below a concentration of 4 nM (Figure 3.29). At the concentration range tested, uptake of HpSPHb was very low at 4 nM and had not yet saturated by 62.5 nM, the highest concentration tested. This data demonstrates that uptake of dimeric HpHb is far more efficient than monomeric HpSPHb, suggesting that two receptors are able to bind one HpHb molecule at the trypanosome cell surface. The average serum concentration of HpHb is below the dissociation constant of the receptor, 1 μM. Therefore a mechanism to increase avidity, such as this one, would be very beneficial to the parasite, decreasing the apparent dissociation constant to 4 nM and allowing efficient uptake at physiological concentrations.

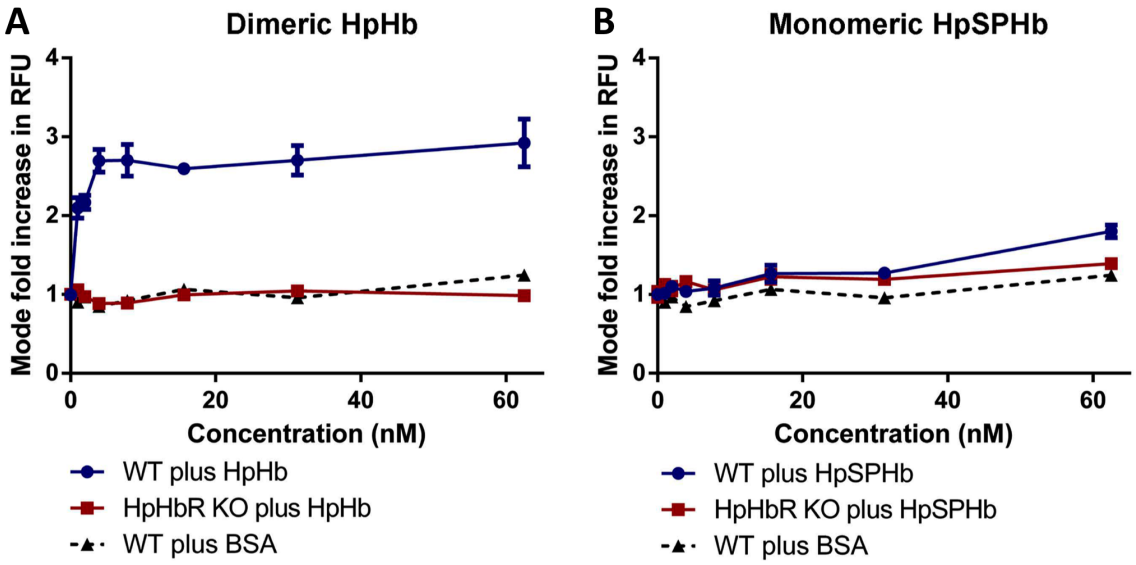


FIGURE 3.29: Uptake of (A) dimeric HpHb and (B) monomeric HpSPHb. Uptake of fluorescently-labelled ligands was monitored via flow cytometry. Each uptake experiment was carried out in triplicate. The error bars show the standard error of the mean, n=3. This figure was produced by Dr. Paula MacGregor.

3.6 Discussion

The work presented in this chapter reveals the first structure of a trypanosome receptor in complex with its ligand. At a similar time, this structure was also determined by another lab group [192], which revealed an almost identical interaction. The bloodstream form of *T. brucei* is coated in a layer of variant surface glycoprotein that is thought to reach near-maximum density. It had been presumed, therefore, that the ligand binding site of TbbHpHbR would be positioned at the membrane-distal tip of the receptor, allowing access to the ligand within the context of the VSG layer by protruding above it. Indeed, mutational studies had previously revealed the location of residues important for binding to be in this region [94]. However, the structure reveals that this is the upper binding site which contacts the haemoglobin β-subunit. There is a second, lower binding site which had not been identified previously, and which contacts the haptoglobin component of HpHb. The presence of two separate binding sites, one for each component, suggests why the receptor can only take up the complex of HpHb, and neither of the individual components. Thus, due to the second site for haptoglobin, the binding site extends further down the receptor than was previously predicted, approximately halfway along its length. This proposes the question of how HpHb can reach the binding site if it is found within the VSG layer.

A second surprising finding from the structure was a 50° kink in TbbHpHbR one third of the way from the membrane-proximal end. There are several possible roles for this. Firstly, the kink may act to physically push apart the surrounding VSG molecules, allowing easier access of the ligand to the binding site. A similar mechanism is thought to be used by the transferrin receptor, this time using glycans to increase the accessibility of the binding site [193]. This could explain how the receptor can access its ligand despite the binding site appearing to be buried in the VSG layer (Figure 3.30).

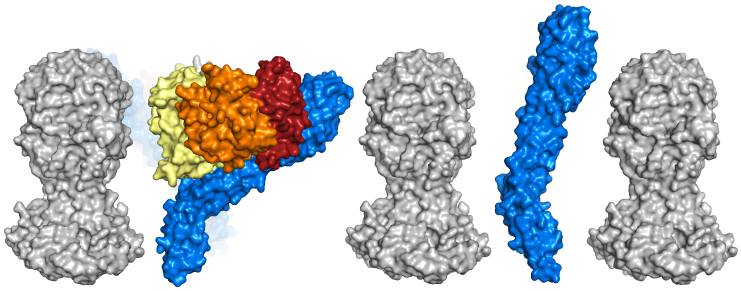


FIGURE 3.30: Model showing the possible orientations of VSGs and TbbHpHbR on the trypanosome cell surface. The relative heights suggest that TbbHpHbR may extend slightly above the VSG layer. However, the structure of the TbbHpHbR C-terminus is unknown.

Another possible role was discovered by aligning two copies of the complex structure with the structure of dimeric HpHb. This revealed that when two receptors bind to one ligand, the C-termini of the receptors are pointing in the same direction. Although the structure of the attachment of the receptors to the membrane is still unknown, as the structure of the C-terminal domain of TbbHpHbR has not been solved, it does suggest that dimeric binding may be possible at the trypanosome surface. This may allow increased efficiency of uptake. SEC-MALLS and SAXS data supported this theory, demonstrating that two receptors could bind to one dimeric HpHb in solution.

For this proposal to be correct, two receptors must come into close proximity on the cell surface, to allow them to both bind the same HpHb molecule. The copy number of TbbHpHbR is 200-400, and it is concentrated in the flagellar pocket. The surface area of the flagellar pocket is 4.3 μm^2, resulting in a density of one receptor per 0.022 μm^2 [65]. Although the diffusion constant for TbbHpHbR is not known, the value for VSG, which is also GPI-anchored, is 0.01 μm^2/sec, as measured by fluorescence recovery after photobleaching experiments [194]. This means that two receptors would come in contact with each other every 2 seconds, much more rapidly than the $t_{1/2}$ for monovalently bound HpHb, 70-100 secs, as measured by SPR.

The greatest support for the theory that dimeric binding of HpHb by TbbHpHbR occurs at the cell surface comes from the uptake experiments. These showed that uptake of dimeric HpHb saturated at 4 nM, whilst uptake of monomeric HpHb was minimal at this concentration, and still increasing at 62.5 nM. The value of 4 nM matches perfectly with the value obtained using SPR when TbbHpHbR is couple to the chip and dimeric HpHb is flown over the surface, a setup which would allow dimeric binding [94].

Thus, the kink of TbbHpHbR is likely to act both to expose the binding site to the extracellular environment, as well as to increase the efficiency of HpHb uptake. If these possible roles for the kink are correct, it seems surprising that the *T. congolense* HpHbR has not also evolved this adaptation. This will be investigated in Chapter 5.

Chapter 4

Structural and biophysical characterisation of anti-TbbHpHbR antibodies

4.1 Summary

In chapter 3 the binding site for HpHb on the *T. brucei* HpHbR was revealed by x-ray crystallography. This site must therefore be exposed to the extracellular milieu when the protein is at the surface of the trypanosome in the bloodstream of the mammalian host and is able to interact with large protein complexes, making it a possible target for antibodies. As proteins attached to the surface of trypanosomes, including HpHbR, are readily endocytosed, antibody-drug conjugates (ADCs) are a good candidate for therapeutics. In this chapter I began by biophysically characterising the interactions of a panel of 6 anti-TbbHpHbR monoclonal antibodies (mAbs) which had been shown to be taken up by *T. b. brucei*. The results showed that all 6 bound an overlapping site on TbbHpHbR. Addi-

tionally, all antibodies competed with HpHb for interaction with the receptor, suggesting that this region of the receptor is particularly prone to forming interactions. I determined the structure of the Fab fragment of one of these antibodies in complex with the receptor. This confirmed that the binding site did indeed overlap that of HpHb. Any further design of protein therapeutics which target TbHpHbR should therefore focus on this site.

4.2 Generating a panel of anti-TbbHpHbR antibodies

The haptoglobin-haemoglobin receptor of African trypanosomes is a good target for antibody-drug conjugates (ADCs) as the receptor is surface-exposed, has no homologue in humans [43] and undergoes high rates of endocytosis [64, 73]. One potential downside, however, is that the receptor is not essential for growth of the parasite [43]. Nevertheless, it is a good candidate for proof-of-principle experiments to demonstrate that targeting trypanosome surface proteins with ADCs leads to cell death. In order to attempt this, a panel of antibodies against the receptor was obtained in a series of experiments carried out by Dr. Paula MacGregor (Carrington lab) and Dr. Andrea L. Gonzalez-Munoz (MedImmune). Six antibodies were selected using phage display based on their specificity for binding the N-terminal domain of TbbHpHbR (TbbHpHbRt). To do this, TbbHpHbRt was chemically biotinylated and coupled to a streptavidin plate, before incubation with phage, washing, and elution of bound phage. Several rounds of panning were carried out and 6 clones were selected and assessed for their ability to be taken up by *T. brucei brucei* (Figure 4.1). All were taken up by wild type cells, whilst uptake into HpHbR -/- cells was not detectable, demonstrating that uptake is a result of the specific interaction with TbbHpHbR.

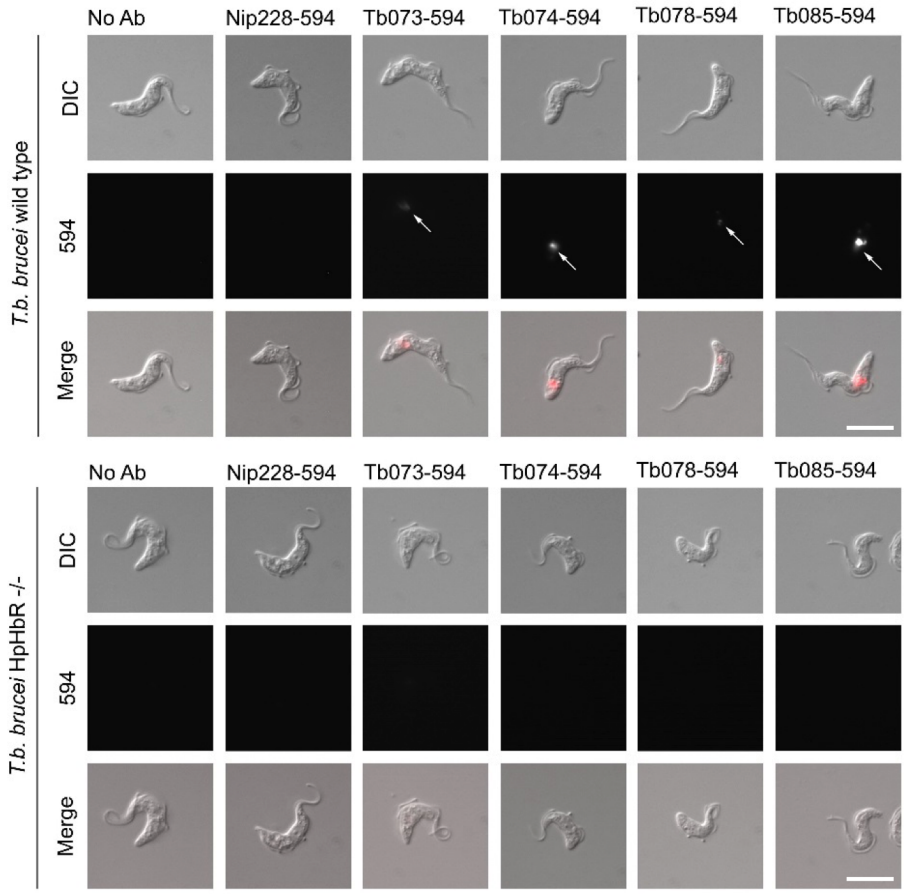


FIGURE 4.1: The uptake of monoclonal antibodies into *T. b. brucei* via endocytosis in complex with the haptoglobin-haemoglobin receptor. Trypanosomes were incubated with 10 nM Alexa-594-labelled antibodies before imaging by microscopy. No uptake of the control antibody (Nip228) was observed. Uptake of anti-TbbHpHbR antibodies into wild type cells was observed, whilst no uptake into HpHbR -/- cells was detected. Here, uptake of 4 of the 6 antibodies being assessed is shown. The other two were tested subsequently and gave the same result. Scale bars represents 10 μm. This figure was produced by Dr. Paula MacGregor.

Next, the toxicity towards trypanosomes of two of these antibodies in ADC format was assessed (Figure 4.2) by Dr. Paula MacGregor. The toxin alone is able to cross cell membranes, so is highly toxic. Addition of the linker and a non-specific antibody decrease the ability of the toxin to kill trypanosomes due to reduced uptake. However, addition of the anti-TbbHpHbR antibodies results in a dramatic decrease in the IC_{50} of the toxin,

demonstrating that uptake has increased. The HpHbR -/- cells show no increase in cell death when the toxin is coupled to an anti-TbbHpHbR antibody, confirming that uptake of the ADCs into wild type cells is dependent on the receptor.

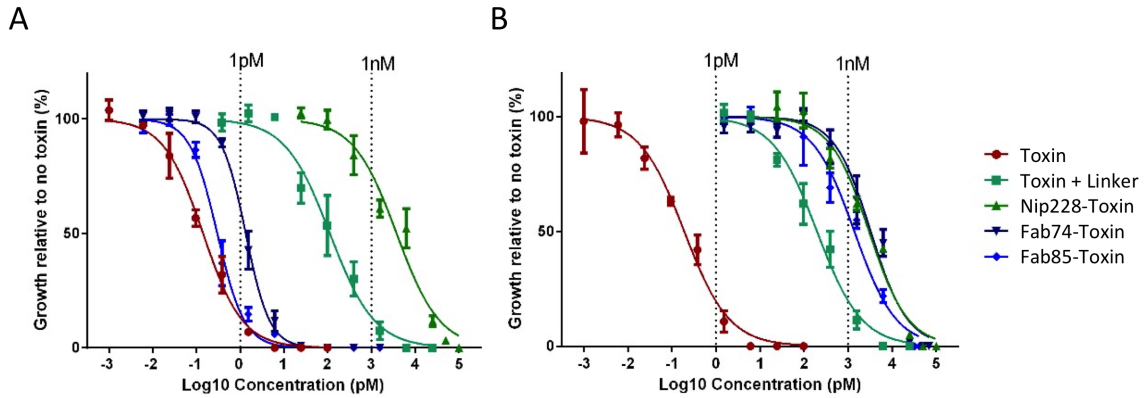


FIGURE 4.2: Killing of *T. b. brucei* by a toxin linked to monoclonal antibodies against TbbHpHbRt. The assay was carried out using both (A) wild type and (B) HpHbR -/- trypanosomes. Adding the linker increases the IC_{50} of the drug, and adding a non-specific antibody as well as this increases it further. Adding anti-TbbHpHbRs reduced the IC_{50} in wild type cells. This figure was produced by Dr. Paula MacGregor.

4.3 Biophysical characterisation of a panel of anti-TbbHpHbR antibodies

4.3.1 Characterising the binding affinities of anti-TbbHpHbR mAbs

The data from the experiments above were very promising, so further characterisation of these antibodies was carried out. To determine the affinity of each mAb for TbbHpHbRt, SPR was performed. TbbHpHbRt was specifically biotinylated at the C-terminus and coupled to a streptavidin-coated CAPture chip (see 3.3.1). Fab fragments of the mAbs provided by MedImmune were then flowed over the immobilised TbbHpHbRt. The K_Ds of the Fabs for TbbHpHbRt were all high affinity, ranging from 0.2-89.2 nM (Figure 4.3).

Despite this, the shapes of the curves varied significantly, with some having very fast on- and off-rates (e.g. 85), whilst the kinetics of others were slower (e.g. 73). Nonetheless, a panel of high affinity antibodies targeting TbbHpHbR had clearly been established.

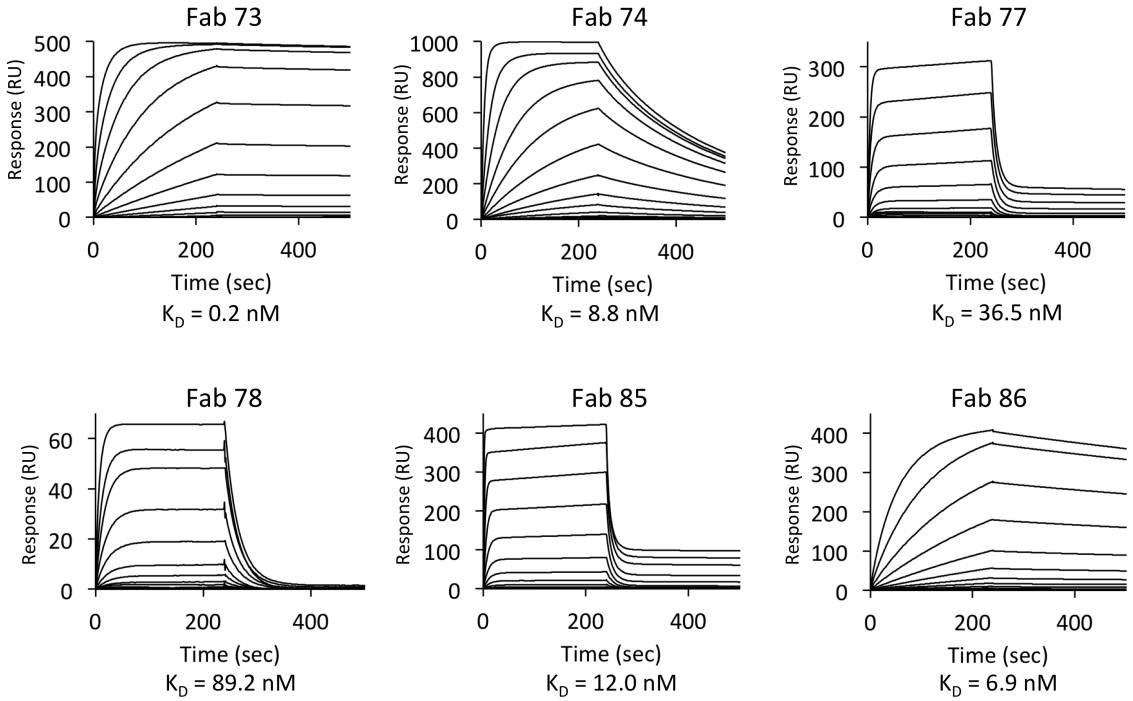


FIGURE 4.3: Surface plasmon resonance curves for binding of Fabs to TbbHpHbRt. One in two dilutions of the Fabs were flowed over a chip coated with receptor from 316 nM. K_Ds were calculated by kinetically fitting the curves to a model.

4.3.2 Competition assays assessing the location of mAb binding

To investigate whether all antibodies recognise different areas of TbbHpHbR, competition assays were performed using SPR. As mAb73 had the slowest off rate and the highest affinity, binding of the other Fabs was tested relative to this one. For this experiment, a protein A/G chip was used. To produce this, protein A/G was coupled to a CM5 chip via primary amines. This allowed each mAb to be immobilised on the chip in turn, and the ligand flowed over. To regenerate the chip, glycine pH 3 was flowed over it, causing dissociation

of bound mAb. After immobilising the mAb on the chip, a mixture of TbbHpHbRt with excess Fab73 was then flowed over (Figure 4.4). If the antibodies bound to the same location on TbbHpHbR as Fab73, no binding would be observed for the complex. TbbHpHbRt alone and Fab73 alone were used as positive and negative controls respectively.

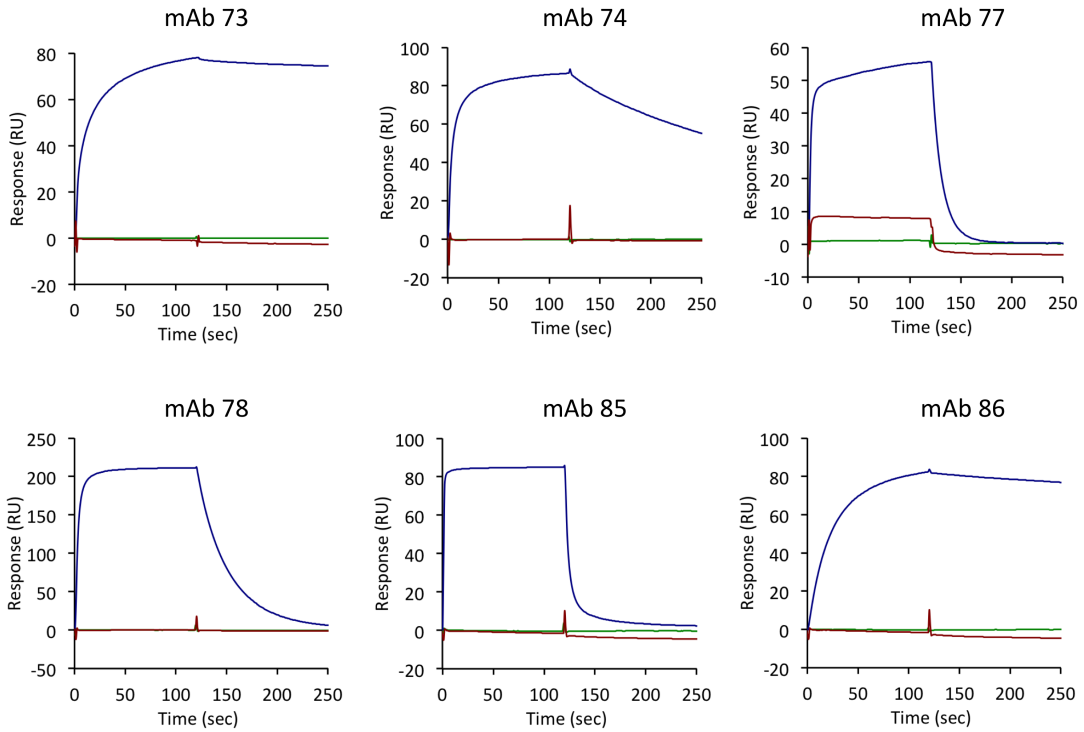


FIGURE 4.4: Surface plasmon resonance curves for the binding of TbbHpHbRt (blue), Fab73 (green), or the complex of the two (red) to immobilised mAb. Fab73 was at 2 μM and TbbHpHbRt at 1 μM.

In all cases, the controls performed as expected. When mixed with Fab 73, TbbHpHbRt was unable to bind to any of the other mAbs. This suggests that all the mAbs in this panel bind to a region on the TbbHpHbRt surface that overlaps the epitope for Fab73. If the antibodies had been selected for uptake into *T. b. brucei* trypanosomes, this would not be surprising, as it is likely than only a small region of the surface of the receptor is exposed at the cell surface due to the dense packing of proteins. However, as the antibodies were only

selected for binding to TbbHpHbRt, during which the whole surface of the receptor would be exposed, it is far more surprising. This indicates a strong propensity of this region for the formation of interactions when compared to other regions of the surface.

To investigate further which region of TbbHpHbRt interacted with the mAbs, a competition assay with HpSPHb was performed. A similar setup to the previous experiment was used, this time with a mixture of TbbHpHbRt and excess HpSPHb flowed over the immobilised mAbs (Figure 4.5). TbbHpHbRt alone and HpSPHb alone were used as positive and negative controls, and behaved as expected. No binding was observed for TbbHpHbRt:HpSPHb in any case, suggesting that all mAbs compete with HpSPHb for binding to TbbHpHbRt. Again, this indicates that this region of the receptor has a higher propensity to form high affinity interactions.

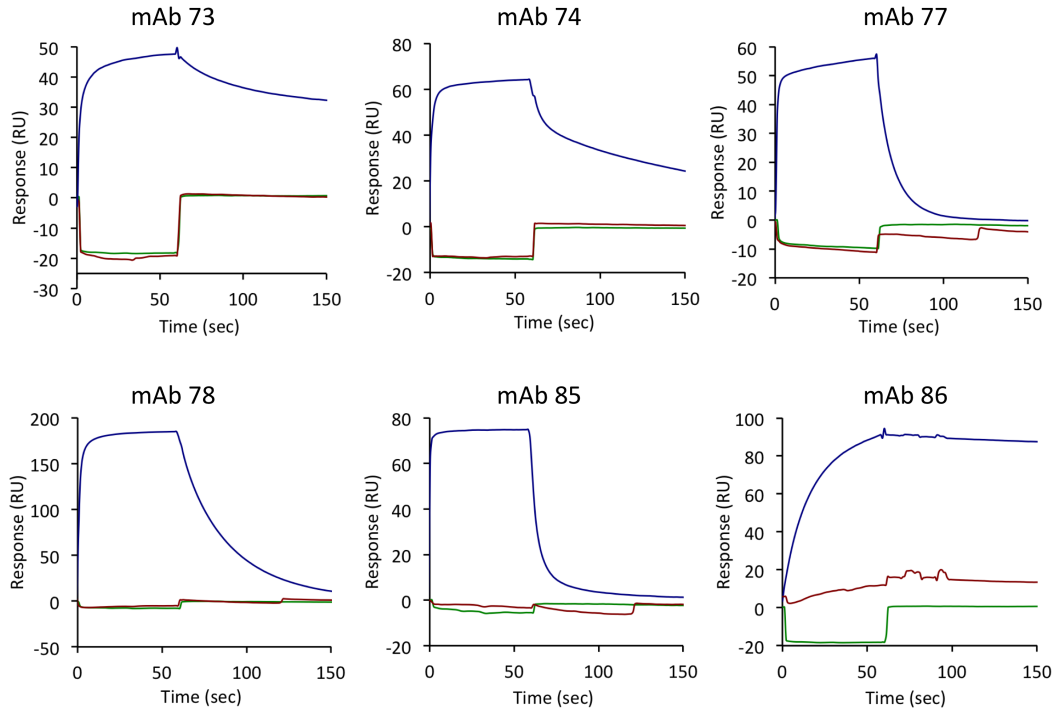


FIGURE 4.5: Surface plasmon resonance curves for the binding of TbbHpHbRt (blue), HpSPHb (green), or the complex of the two (red) to immobilised mAb. HpSPHb was at 10 μM and TbbHpHbRt at 1 μM.

4.4 Structural characterisation of TbbHpHbRt:Fab73

4.4.1 Fab:TbbHpHbRt complex formation

The structure of one of the Fabs in complex with TbbHpHbR would be ideal to confirm that the binding site overlaps with that of HpSPHb, as indicated by SPR. To achieve this, TbbHpHbRt and each of the 6 Fabs in turn were mixed with one component in excess, before being applied to a Superdex 200 16/600 gel filtration column (Figure 4.6). Ideally an excess of Fab was used, as this elutes later than TbbHpHbRt due to the elongated shape of the latter, and would therefore result in greater peak separation from the complex. However, in cases where only a small amount of Fab was available, excess receptor was used.

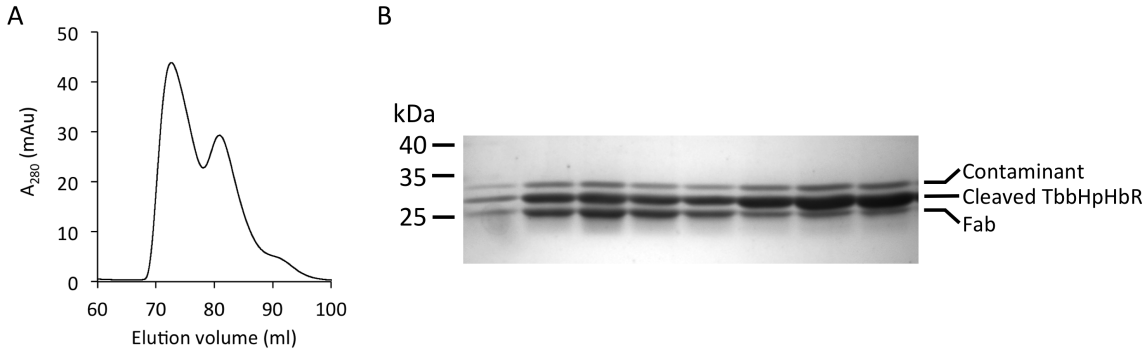


FIGURE 4.6: Purification of the complex formed between TbbHpHbRt and Fab77. (A) Fab77 was mixed with a molar excess of receptor and gel filtered using a Superdex 200 16/600 in 20 mM HEPES pH 7.5, 150 mM NaCl. (B) The identity of each gel filtration peak was confirmed by SDS-PAGE. As the contaminant is in both the complex and excess TbbHpHbRt peak, it is likely to be uncleaved TbbHpHbRt. Other Fab:TbbHpHbRt complexes were purified in the same way.

SDS-PAGE revealed which elution fractions contained the complex. This was completed for all six Fab:TbbHpHbRt complexes.

4.4.2 Fab:TbbHpHbRt crystallisation

Initial crystallisation trials were carried out in 96-well format using vapour diffusion of sitting drops. Each Fab:TbbHpHbRt complex was concentrated to 10 mg/ml and used to set up several broad screens. James Harrison, an undergraduate student, assisted with crystallisation trials for several Fabs, under my guidance. The commercial screens JCSG+, ProPlex, Morpheus and PACT were used, where enough material was obtained, at both 4 and 18 °C. The most successful complex was Fab73:TbbHpHbRt, which is perhaps not surprising as Fab73 has the highest affinity for TbbHpHbRt of all the Fabs in the panel. Crystals appeared in several conditions in column 9 of Morpheus at 18 °C, which share the common components of 0.1 M Tris BICINE pH 8.5, 20% PEG 500 MME and 10% PEG 20K. After 2 days crystals appeared and they were harvested on day 16 (Figure 4.7A). Crystals from well B9 which, in addition to the common components of this column, contained 30 mM NaFl, 30 mM NaBr and 30 mM NaI, diffracted the best, reaching 4 Å resolution.

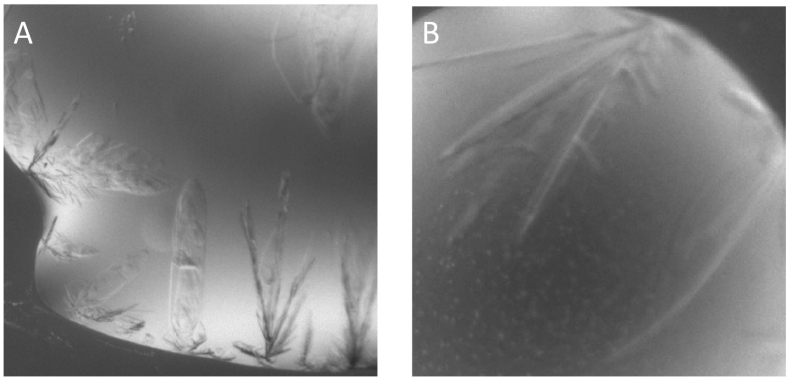


FIGURE 4.7: Crystals of Fab73:TbbHpHbRt. (A) Crystals grew in well B9 of the Morpheus plate at 18 °C in 2 days and by day 16 had grown into overlapping plates and needles 350-450 μm in length. The well condition contained 0.1 M Tris BICINE pH 8.5, 20% PEG 500 MME, 10% PEG 20K, 30 mM NaFl, 30 mM NaBr and 30 mM NaI. (B) Crystals appeared within 14 hours in F11 of the seeded additive screen, and had grown into 400-500 μm long needles by day 6. In addition to the well condition of B9, this optimisation plate drop contained condition F11 of the silver bullet screen: 0.25% w/v sodium 1-pentanesulfonate monohydrate, 0.25% w/v 4-aminobutyric acid, 0.25% w/v cytosine, 0.25% salicylamide, 0.02 M Hepes sodium pH 6.8.

A plate using this well condition, seeds from this well and the silver bullet additive screen was set up at 18 °C. Crystals appeared within 14 hours and were harvested after 6 days. These crystals were less overlapping than those from the broad screen (Figure 4.7B). No additional cryoprotectant was required for cryo-cooling of these crystals as the well condition contained 20% PEG 500 MME.

4.4.3 Data collection and reduction

A Fab73:TbbHpHbRt crystal from optimisation plate well F11 diffracted well and a dataset to 2.91 Å resolution was collected. The additive screen for this drop contained 0.25% w/v sodium 1-pentanesulfonate monohydrate, 0.25% w/v 4-aminobutyric acid, 0.25% w/v cytosine, 0.25% salicylamide, 0.02 M Hepes sodium pH 6.8. The dataset was indexed and integrated by Xia2 with DIALS. The space group was identified as $p2_1$ and the unit cell dimensions as a = 71.73 Å, b = 160.67 Å, c = 93.33 Å, $\alpha = \gamma = 90.00°$, $\beta = 91.66°$. The data reduction statistics are shown in table 4.1. The dataset is 100% complete in the highest resolution shell, and the R_{PIM} is 65.7% for the highest shell and 12.8% overall. The I/σ(I) and $CC_{1/2}$ are 1.2 and 60% for the highest resolution shell respectively.

TABLE 4.1: Data processing statistics.

	TbbHpHbRt:Fab73
Beamline	Diamond I24
Space group	$p12_11$
Cell dimensions	
a, b, c (Å)	a = 71.73, b = 160.67, c = 93.33
α, β, γ (°)	α = γ = 90.00 β = 91.66
Resolution (Å)	2.91
Wavelength (Å)	1.0389
R_{PIM} (%)	12.8 (65.7)
I/σ(I)	5.1 (1.2)
$CC_{1/2}$	97.5 (60.0)
Completeness (%)	99.9 (100.0)
Multiplicity	3.4 (3.4)

4.4.4 Phasing, model refinement and validation

Molecular replacement for Fab73:TbbHpHbRt was carried out by Phaser using the structure of TbbHpHbRt alone (pdb:4X0J) [195] and a Fab (pdb:4U0R) [196]. To prepare the TbbHpHbRt search model all water molecules were removed. To prepare the Fab search model all water molecules were removed and the sidechains were trimmed to a single methyl group to produce a polyalanine model using CHAINSAW. Additionally, the loops between each region of secondary structure were removed. Molecular replacement was successful, with two copies of both the receptor and Fab in the asymmetric unit. Initially Buster was used for refinement, and during these refinement cycles the location of disulphide bonds and bulky side chains of the Fab were revealed. This allowed the sequence of Fab73 to be built into the model, mutating the alanines of the model to their correct identity. Later in refinement phenix.refine was also used, as this greatly improved the geometry of the model. In both cases, the high resolution structure of TbbHpHbRt alone (pdb:4X0J)

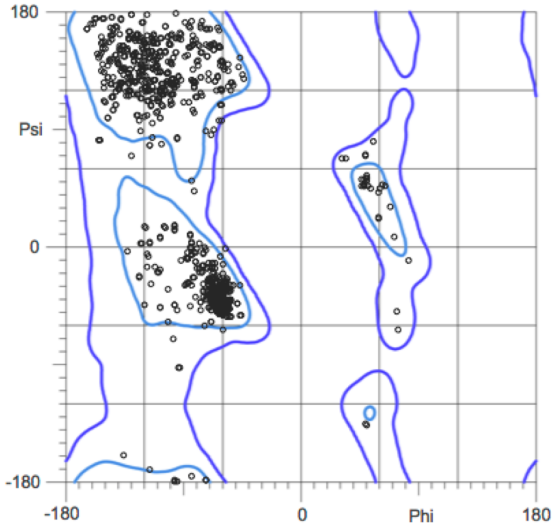
[195] and a high resolution Fab structure with high sequence identity (pdb:4LLW) [197] were used to provide restraints.

For the heavy chain of the Fab, all but the first residue of chain E was visible and of the light chain only the first residue of chain F and the final residue of both chains C and F were missing. The receptors are complete, except for one residue at the N-terminus of chain A, three residues at the N-terminus of chain D, and three residues at the C-terminus of chain A. In addition to the protein chains, the structure contains seven PEG molecules and 50 water molecules. The final R_{work} and R_{free} values were 19.1% and 23.3% respectively, whilst the root mean square deviations for bond lengths and angles were 0.009 Å and 1.14° respectively (Table 4.2).

TABLE 4.2: Data refinement statistics.

	TbbHpHbRt:Fab73
Resolution (Å)	2.9
No. reflections	46,263
R_{work} / R_{free} (%)	19.1 / 23.3
No. of protein residues in model	1397
R. m. s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.14

The Ramachandran plot shows that 96.8% of residues are in the favoured region, whilst 3.2% are in the allowed region and none are in the disallowed region (Figure 4.8).



	No. of residues	Percentage
Favoured region	1344	96.8
Allowed region	41	3.2
Disallowed region	0	0

FIGURE 4.8: Ramachandran plot and statistics of the Fab73:TbbHpHbRt structure generated by Molprobity [198].

4.4.5 The structure of Fab73:TbbHpHbRt

The crystal structure of Fab73:TbbHpHbR reveals the binding site for a monoclonal antibody against TbbHpHbR, beginning approximately 35 from the membrane-distal tip and ends just above the distinct kink of the receptor (Figure 4.9). The structure of TbbHpHbRt is largely unchanged from the structure of the receptor alone, with a root mean square deviation between the two of 2.2 Å. The binding site on the receptor is predominantly formed by helix I, with additional contributions from helix III. This is similar to the binding site for haptoglobin-haemoglobin, which again is predominantly formed by helix I, this time with additional contacts made by helix II. As expected, the variable loops of both the heavy and light chains of the antibody are involved in the interaction, contributing 640 Å^2 and 415

Å^2 to the total surface area of the binding site respectively. CDR1 and 3 of both the heavy and light chain predominantly form hydrophilic interactions, whilst CDR2 of the heavy chain forms mainly hydrophobic interactions. CDR2 of the light chain is too distant from TbbHpHbRt to be involved in the interaction, accounting for the reduced surface area for the interaction of this chain.

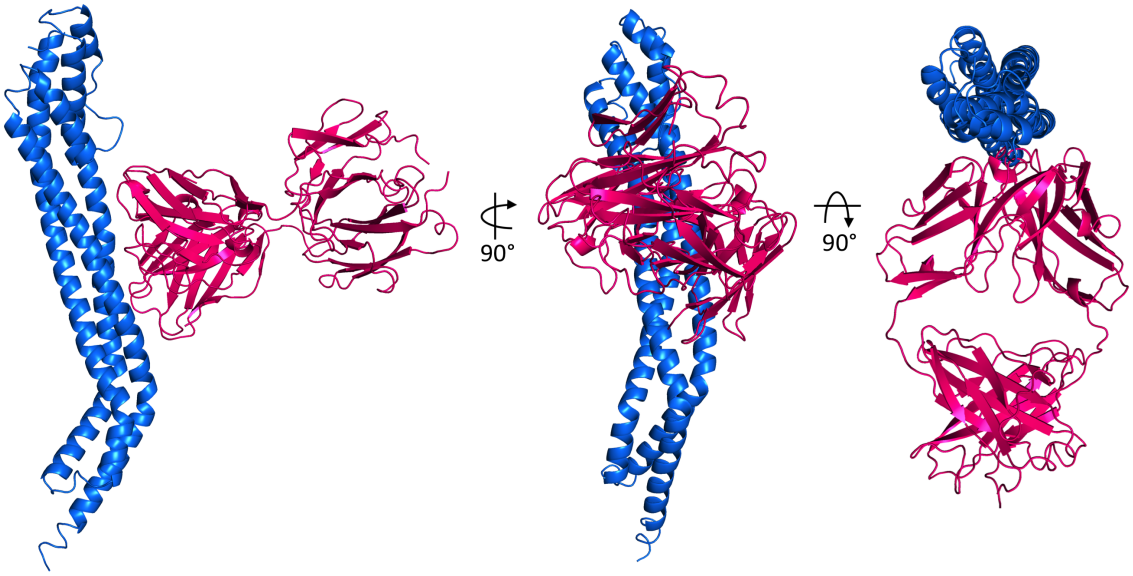


FIGURE 4.9: The structure of TbbHpHbRt in complex with the Fab fragment of an anti-TbbHpHbRt antibody, mAb73. The receptor is coloured blue and Fab73 pink.

This structure confirms that the binding site for Fab73 overlaps with the binding site for HpHb on TbbHpHbR (Figure 4.10). Haptoglobin interacts with a very similar region of TbbHpHbR on helix I to the heavy chain of Fab73. The binding sites for both the light chain of Fab73 and haemoglobin are further towards the N-terminus of the receptor, although on different sides. From the SPR results in section 4.3, it can be assumed that all mAbs in the panel bind to a region similar to this.

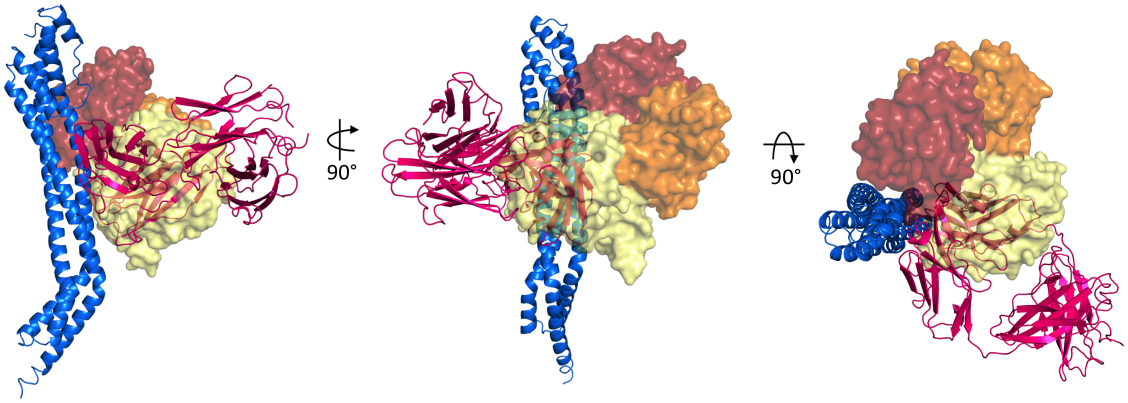


FIGURE 4.10: A comparison of the interactions of TbbHpHbRt (blue) with Fab73 (pink) and HpSPHb (surface representation, haemoglobin-α orange, haemoglobin-β, haptoglobin yellow). The model was produced by aligning the receptors in the complex structures of TbbHpHbRt:Fab73 and TbbHpHbRt:HpSPHb.

4.5 Discussion

The haptoglobin-haemoglobin receptor of African trypanosomes is a surface protein known to bind and take up its ligand from the extracellular environment by endocytosis [43]. Targeting this receptor with ADCs is therefore appealing, as the surface of the receptor must be accessible to proteins in the blood, and anything that binds to the receptor is likely to be taken up rapidly by the trypanosome due to the high rate of endocytosis [64, 73]. Furthermore, a trypanosome surface protein such as this has no homologue in humans, negating the possibility that the ADCs will be taken up by host cells, and thus reducing the toxicity of the treatment. As ADCs targeting cancers are being developed, the safety and efficacy of the linkers and drugs used have already been tested [199–201]. Therefore only the specificity of the antibody requires alterations to allow these ADCs to be adapted to become anti-trypanosomal [202].

With this goal in mind, a panel of six anti-TbbHpHbR monoclonal antibodies was produced and uptake of all six antibodies into trypanosomes was demonstrated at nM con-

centrations. Binding studies demonstrated that all the antibodies have a low nM affinity for TbbHpHbRt. In contrast, low pM ADC concentrations resulted in trypanosome cell death, suggesting that only a very small number of toxin molecules are required to kill the cell. The results also show that all the mAbs of the panel all recognise overlapping regions on TbbHpHbR, one that overlaps with the binding site for HpHb. The crystal structure of Fab73:TbbHpHbRt confirmed this, with the binding site location clearly overlapping with that of haptoglobin.

The six antibodies were selected for binding to TbbHpHbRt *in vitro* and only subsequently was uptake and successful killing of trypanosomes confirmed. It is therefore surprising that all six bind to such a similar location, and that, conveniently, it is a site that is exposed on when the receptor is at the cell surface of the trypanosome. Perhaps it is the location and identity of the amino acids in this patch that make it particularly able to form interactions, characteristics which would have developed in order to bind HpHb. When the structure of a protein is known, interaction hotspots can be predicted by Jet2 [151]. This identified two regions of the TbbHpHbR that are likely to form interactions, both of which are in positions that are likely to be exposed to the extracellular environment in the context of the cell surface (Figure 4.11). SC3, which is predicted based on the propensity of amino acids to be present at an interface and circular variance, which measures the local geometry of the surface, overlaps considerably with the binding sites for both Fab73 and HpSPHb. To a lesser extent SC1, which was predicted based predominantly on evolutionary conservation of amino acids, also overlaps with the binding site for HpHb. Assessing the characteristics of amino acids appears, in this case, to be a good indicator for the ability of a site to form interactions.

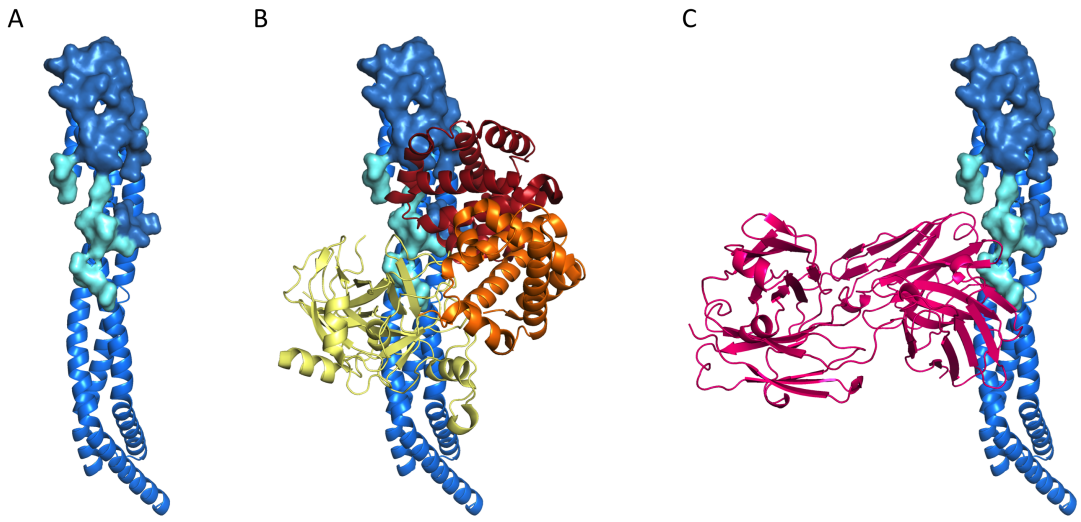


FIGURE 4.11: Interaction hotspots of TbbHpHbRt correlate with known interaction sites. (A) Two different interaction hotspots were predicted using Jet2 [151], shown as surfaces in sky-blue (SC1) and aquamarine (SC2), overlaid on the structure of TbbHpHbRt (marine). The binding sites for (A) HpSPHb (Hp yellow, Hb orange/red) and (B) Fab73 (hot pink) overlap with SC3, and to some extent SC1.

A consequence of this is that any ADC that targets TbbHpHbR will be competing for binding of both HpHb and TLF1. It will therefore be important to ensure that administration of this treatment does not render the patient more susceptible to infection by those trypanosome species that are sensitive to TLF1. However, as the treatment would only be administered during infection, and as TLF2 uptake will remain unaffected, this should not be a major problem.

T. b. brucei was used for these experiments due to the extensive knowledge about the structure and function of the receptor and as a null cell line for the receptor was already in place [43, 192, 195]. However, to provide a suitable treatment against trypanosomiasis, the drug must function against the human-infective subspecies *T. b. rhodesiense* and *T. b. gambiense*. When incubated with the former, these ADCs are able to kill the trypanosomes, albeit with a higher IC_{50} (low pM). This was expected, as the HpHbRs of *T. b. brucei*

and *T. b. rhodesisense* are identical. However, when incubated with *T. b. gambiense*, the ADCs fail to lower the IC_{50} to below the value of the toxin coupled to the linker alone. This is not surprising, due to the known differences in the binding affinities and expression levels of HpHbR in *T. b. gambiense* [94, 142, 203]. Targeting this trypanosome may therefore require antibodies against a more abundantly expressed surface protein such as the transferrin receptor [60]. The work here has furthered characterisation of the interactions involved in the first example of targeting trypanosomes using ADCs, demonstrating that receptor-mediated nutrient uptake can be exploited for the delivery of drugs at picomolar concentrations into trypanosome cells.

Chapter 5

Structural basis of haemoglobin binding by TcHpHbR

5.1 Summary

As described in chapter 3, I discovered that the *T. b. brucei* haptoglobin-haemoglobin receptor has a kink a third of the way along the receptor from the C-terminus, a structural feature not seen in the *T. congolense* receptor. The work in this chapter aimed to investigate why this difference had evolved. To achieve this, I attempted to crystallise the *T. congolense* receptor in complex with HpSPHb. Surprisingly, the crystals obtained instead contained a complex between the receptor and a haemoglobin tetramer alone. ITC experiments confirmed that this interaction between haemoglobin and TcHpHbR can take place in solution and SPR demonstrated that HpHbR from a more distantly related trypanosome, *T. vivax*, can also bind Hb alone, suggesting that binding of haemoglobin by TbbHpHbR has been lost during the parasite's evolution. To establish whether the differences in both structure and ligand specificity between HpHbRs reflects differences in function, the ex-

pression profile and functionality of TcHpHbR was investigated by our collaborators in the Carrington lab, Cambridge. This led to the discovery that TcHpHbR is expressed at far greater levels when the parasite occupies the mouth parts of the tsetse fly, rather than in the bloodstream of the mammalian host, and it functions as a haemoglobin receptor *in vivo*. Together, this data allowed an evolutionary history of the receptor to be proposed, evolving from a haemoglobin receptor expressed by the epimastigote to a haptoglobin-haemoglobin receptor expressed by bloodstream form *T. brucei*.

5.2 Structural characterisation of TcHpHbRt:Hb

5.2.1 TcHpHbR expression and purification

The cloning, expression and purification protocols for this protein had been established previously [94]. The construct lacked the GPI-anchor site, and had a C-terminal 6-His tag separated from the rest of the protein sequence by a TEV protease cleavage site. The protein was expressed solubly in *E. coli* and purified by Ni-NTA^{2+} affinity chromatography, followed by TEV protease cleavage to remove the flexible 6-His tag. A reverse Ni-NTA^{2+} column was used to remove TEV protease and any uncleaved TcHpHbRt. The flow through from this was purified further by size exclusion chromatography (Figure 5.1A). SDS-PAGE confirmed that the major component of the peak was TcHpHbR (Figure 5.1B).

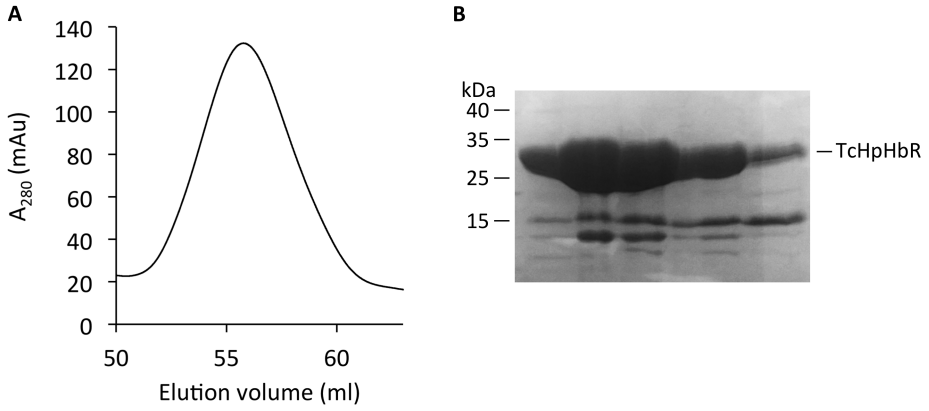


FIGURE 5.1: Purification of TcHpHbR. (A) TcHpHbR was purified by size exclusion chromatography using a Superdex 200 16/600 column in 20 mM Hepes pH 7.5, 150 mM NaCl. (B) Samples taken from the gel filtration peak fractions were analysed using SDS-PAGE.

HpHbR from another trypanosome species that infects livestock, *T. vivax*, was also expressed and purified. Although this was not used for crystallisation, it was used for the binding studies in section 5.3.3. Again the construct lacked the GPI-anchor site, and had a C-terminal 6-His tag separated from the rest of the protein sequence by a TEV protease cleavage site. The protocol used for expression and purification was identical to that used for TcHpHbR (Figure 5.2).

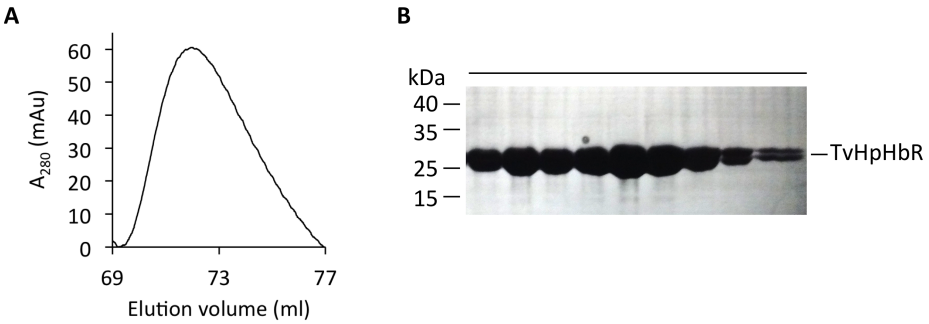


FIGURE 5.2: Purification of TvHpHbR. (A) TvHpHbR was purified by size exclusion chromatography using a Superdex 200 16/600 column in 20 mM Hepes pH 7.5, 150 mM NaCl. (B) Samples taken from the gel filtration peak fractions were analysed using SDS-PAGE.

5.2.2 Crystallisation

Purified TcHpHbR and HpSPHb were mixed in equimolar ratios and the concentration required for crystallisation was determined using a pre-crystallisation test. This suggested a concentration of 12.5 mg/ml. Broad screens were set up using JCSG+, ProPlex, MIDAS, Morpheus and PACT, all at 4 °C. After 20 hours, red crystals were seen in condition E2 of the ProPlex plate containing 0.1 M Na citrate pH 5, 8% PEG 8000 (Figure 5.3A). Crystals from this well diffracted to 8 Å, so this condition was chosen for optimisation. A 96-well optimisation plate was set up using a seed stock made from crystals in well E2 of the ProPlex plate, using this same well condition and the silver bullets additive screen. Larger, overlapping, plate-shaped crystals appeared rapidly in these drops (Figure 5.3B). Crystals were separated using an acupuncture needle, before cryo-cooling in the well condition with an additional 30% glycerol.

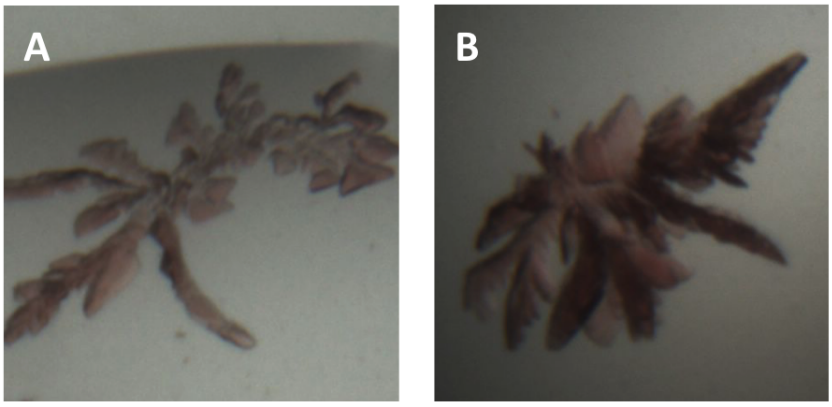


FIGURE 5.3: Crystallisation of TcHpHbR:Hb. (A) Crystals appeared in well E2 of the ProPlex broad screen after 24 hours. This condition contained 0.1 M Na citrate pH 5, 8% PEG 8000. (B) Crystals appeared in the optimisation plate containing the well condition of ProPlex E2, with the addition of a seed stock made from the original crystals and the silver bullet screen.

5.2.3 Data collection and reduction

The best diffracting crystal was from the seeded additive optimisation plate containing the well condition of ProPlex well E2, and the additives 0.005 M gadolinium(III) chloride hexahydrate, 0.005 M samarium(III) chloride hexahydrate, 0.05 M benzamidine hydrochloride, 0.25% w/v salicin. This crystal diffracted to 3.2 Å resolution and a dataset was collected. This dataset was initially indexed and integrated by iMOSFLM. Pointless was used to identify the space group as $p22_12_1$ and the unit cell dimensions as a = 72.75 Å b = 127.30 Å, c = 172.42 Å, $\alpha = \beta = \gamma = 90.00°$. Scala was used to merge the data. The resolution was cut off at 3.2 Å to ensure that I/σ(I) remained above 2 (Table 3.1). The dataset is 99.1% complete in the highest resolution shell, and the R_{PIM} is 41.0% for the highest shell and 7.3% overall (Table 5.1).

TABLE 5.1: Data processing statistics.

	TcHpHbR:Hb
Beamline	Diamond I03
Space group	$p22_12_1$
Cell dimensions	
a, b, c (Å)	72.75, 127.30, 172.42
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	3.2
Wavelength (Å)	0.976
R_{PIM} (%)	7.3 (41.0)
I/σ(I)	7.3 (2.4)
Completeness (%)	99.0 (99.1)
Multiplicity	2.7 (3.0)

5.2.4 Phasing, model refinement and validation

Molecular replacement was first attempted using both TcHpHbR (pdb:4E40) [94] and human HpSPHb (pdb:4X0L) [195] structures as search models. To prepare these models,

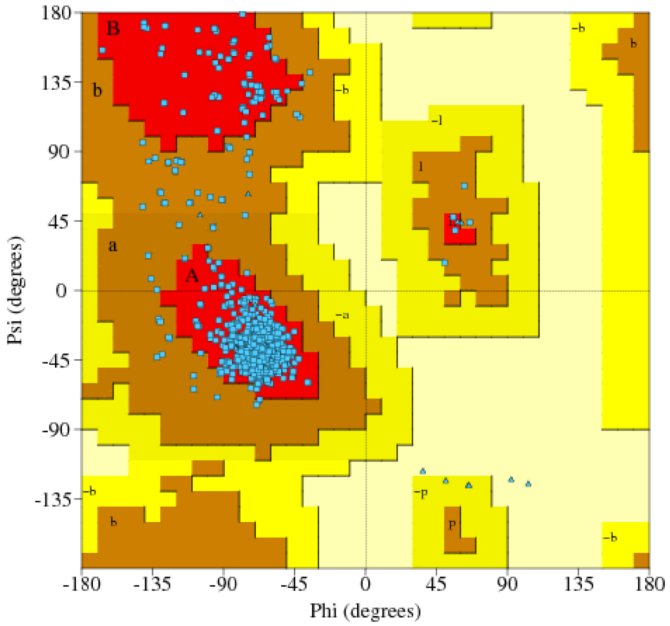
water molecules were removed from both, whilst glycosylations, haem rings and oxygen molecules were removed from HpSPHb. Phaser was able to find a solution and, after inspection of the electron density map, the solution appeared to be correct for the position of the receptor and the two haemoglobin subunits. However, the density for haptoglobin fit the model very poorly, suggesting that this solution was in fact incorrect. A second attempt to phase the dataset was made, this time using TcHpHbR and an oxyhaemoglobin tetramer (pdb:1HHO) [204], as models. The haem rings and oxygen molecules were removed from the haemoglobin search model. Again a solution was found and this time inspection of the electron density showed all chains of the model fit well. This correct solution contained two receptors and one haemoglobin tetramer in the asymmetric unit.

As all the amino acids present in the structure were also in the search model, model building was fairly simple. The loop between helices I and II was removed and rebuilt to match the electron density, and side chain positions were adjusted. Haem rings and oxygen molecules were added to the structure. Refinement and rebuilding was completed using Buster and Coot respectively. The structures of TcHpHbR (pdb:4E40) [94] and Hb (pdb:1HHO) [204], were provided target restraints during refinement. Both receptors are missing the first ten residues, whilst receptor chain A is missing the final residue, and chain B is missing the final eight, probably due to the flexibility of the termini. Of the haemoglobin chains, chains E (Hbα) and F (Hbβ) are missing the terminal residues, but all other residues are present.

After refinement was completed, the R_{work} and R_{free} values had reduced to 20.4% and 23.5% respectively, whilst the root mean square deviations for bond lengths (0.01 Å) and angles (1.21°) were below accepted values (0.02 Å and 4°) (Table 5.2). The Ramachandran plot generated by PROCHECK calculated that 6.9% of residues were in the additionally allowed region, whilst the remaining 93.1% were in the allowed region, with none in the generously allowed or disallowed regions (Figure 5.4).

TABLE 5.2: Data refinement statistics.

	TcHpHbR:Hb
Resolution (Å)	3.2
No. reflections	25,191
R_{work} / R_{free} (%)	20.4 / 23.5
No. of protein residues in model	1064
R. m. s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.21



	No. of residues	Percentage
Allowed region	896	93.1
Additionally allowed region	66	6.9
Generously allowed region	0	0
Disallowed region	0	0
Residues not Gly or Pro	962	
End residues (excl. Gly or Pro)	10	
Gly residues	60	
Pro residues	32	
Total	1064	

FIGURE 5.4: Ramachandran plot and statistics of TcHpHbR:Hb generated by PROCHECK [189].

5.2.5 The structure of TcHpHbRt:Hb

The structure of the complex between TcHpHbR and Hb revealed that two receptors make identical contacts to the two binding sites on the Hb tetramer (Figure 5.5). The conformations of both proteins are unaltered from that of the unbound state, with root-mean-square

deviations of 0.6 Å and 0.7 Å for the receptors and oxyhaemoglobin, respectively. The haemoglobin molecule is in the oxygenated state.

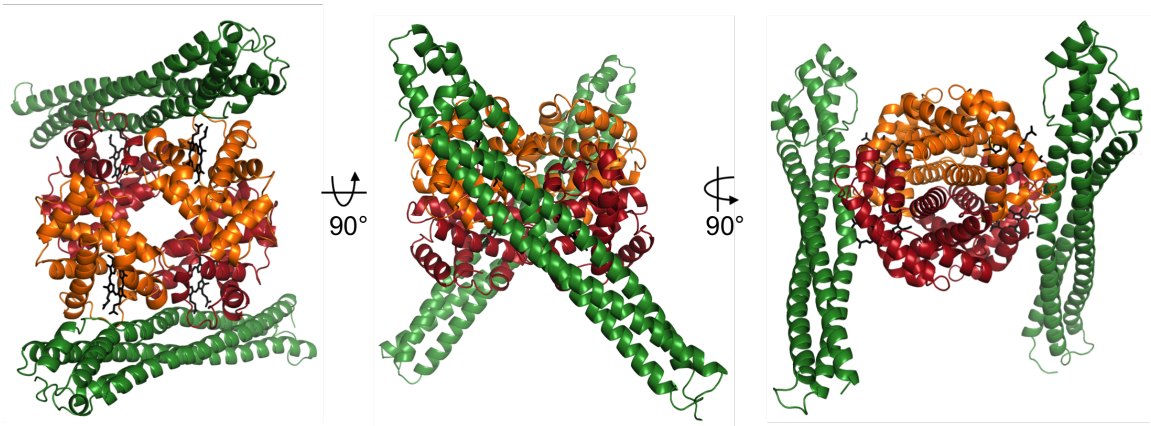


FIGURE 5.5: The structure of two TcHpHbRs binding to one Hb tetramer. TcHpHbR is green, the Hb α-subunits are orange and Hb β-subunits are red. Both Hb subunits make contacts with TcHpHbR.

As with the binding of HpSPHb, the interaction surface can be divided into two subsites. The upper binding site contacts the α-subunit of one haemoglobin dimer, whilst the lower binding site contacts the β-subunit of the other haemoglobin dimer. Therefore, either the receptor selectively binds haemoglobin tetramers, or causes two dimers to assemble when they bind the receptor. This structure contains human haemoglobin, but *T. congolense* instead infects other species, such as cow and goat. Sequence alignments revealed that all haemoglobin residues that are involved in the interaction are conserved between humans, cow, sheep, horse and camel, and therefore the mode of binding is highly likely to be conserved (Figure 5.6).

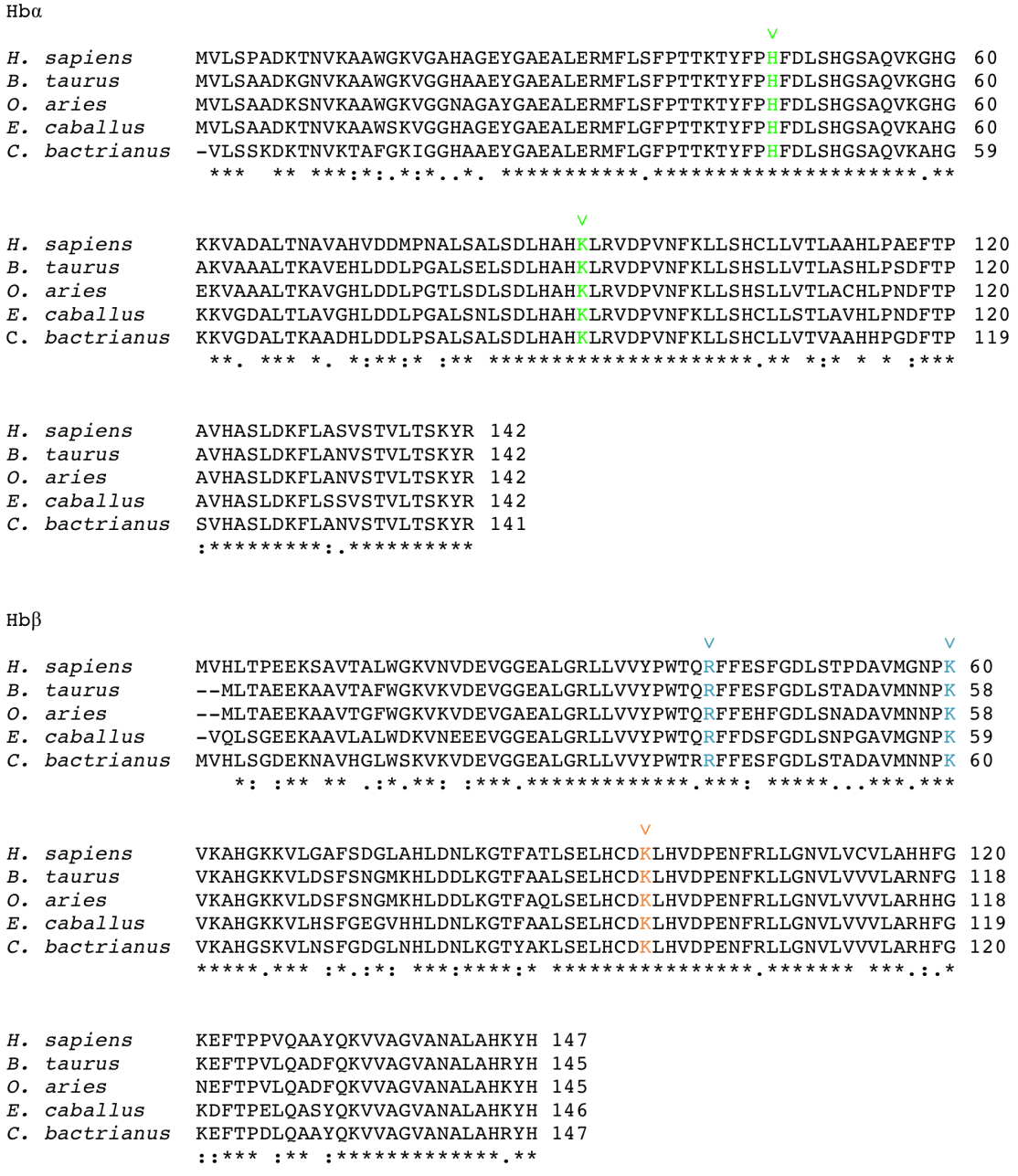


FIGURE 5.6: TcHpHbR-binding residues are conserved across mammalian haemoglobins. This sequence alignment was completed using ClustalW2. Protein sequences for different species were obtained from UniProt. The side chains of residues in green interact with TcHpHbR in the TcHpHbR:Hb structure. The side chains of residues in blue interact with TbbHpHbR in the TbbHpHbRt:HpSPHb structure. The side chains of residues in orange interact with both TcHpHbR and TbbHpHbR.

The membrane-distal binding site, contacting the haemoglobin α-subunit, has a surface area of 735 Å^2. The contacts are mainly hydrogen bonds and electrostatic interactions, including interactions between the receptor and the propionate chains of the haem group. This upper binding site is similar in surface area, chemical nature and location on both the receptor and haemoglobin to the membrane-distal binding site of TbbHpHbR for HpHb, despite the haemoglobin subunit being different in each case (α in this case, and β in the TbbHpHbRt:HpHb structure).

The membrane-proximal binding site, contacting the haemoglobin β-subunit, has a surface area of 650 Å^2. The contacts are mainly hydrophobic, although again interactions with propionate chains of the haem group can be seen. This binding site is further from the membrane surface than the binding site for haptoglobin on TbbHpHbR. The haptoglobin binding site is also smaller (505 Å^2), which is likely to account for the greater affinity of TcHpHbR for Hb.

5.3 Investigating the interaction between TcHpHbR and Hb

5.3.1 Solution studies of the interactions of TcHpHbR by size exclusion chromatography

The surprising discovery that the crystal structure of the complex contained TcHpHbR with an Hb tetramer, rather than HpSPHb, required investigation. Was this an artefact of crystallisation, or did this interaction occur in solution? This was initially tested by mixing Hb and TcHpHbR and applying this sample to a gel filtration column (Figure 5.7). Two peaks are seen, which SDS-PAGE confirms contain the complex formed between TcH-

pHbR and Hb in the higher molecular weight peak, whilst the second peak contains excess TcHpHbR. This confirms that this interaction does occur in solution, and that the affinity is high enough for the complex to remain intact during size exclusion chromatography.

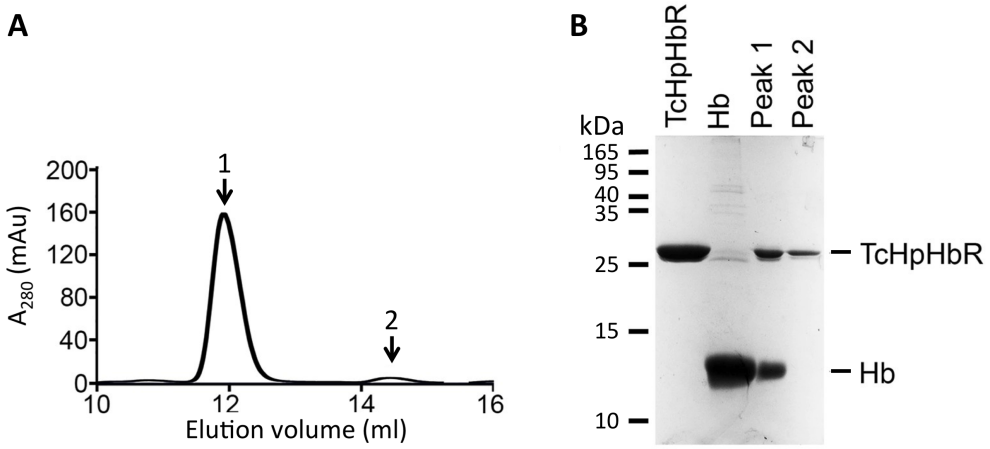


FIGURE 5.7: Analytical gel filtration demonstrating the consequence of mixing Hb with TcHpHbR. (A) The sample was applied to a Superdex 200 10/300 column. (B) Samples from peak fractions were run on SDS-PAGE.

It was still not clear why this interaction was seen in the crystal structure, as TcHpHbR had been mixed with HpSPHb before setting up the crystallisation screens. Either sufficient amounts of contaminating Hb were present in the HpSPHb sample, or the presence of TcHpHbR was causing disassembly of HpSPHb into free HpSP and Hb tetramers. To investigate these possibilities, HpSPHb was mixed with TcHpHbR and applied to a gel filtration column (Figure 5.8A). As expected, two peaks are seen. However, SDS-PAGE shows that the first peak contains the complex of TcHpHbR with Hb, whilst the second peak contains excess TcHpHbR and free HpSP, which cannot be separated due to their similar molecular weights (Figure 5.8B). This explains why the crystal structure contained the receptor bound to haemoglobin, rather than HpSPHb.

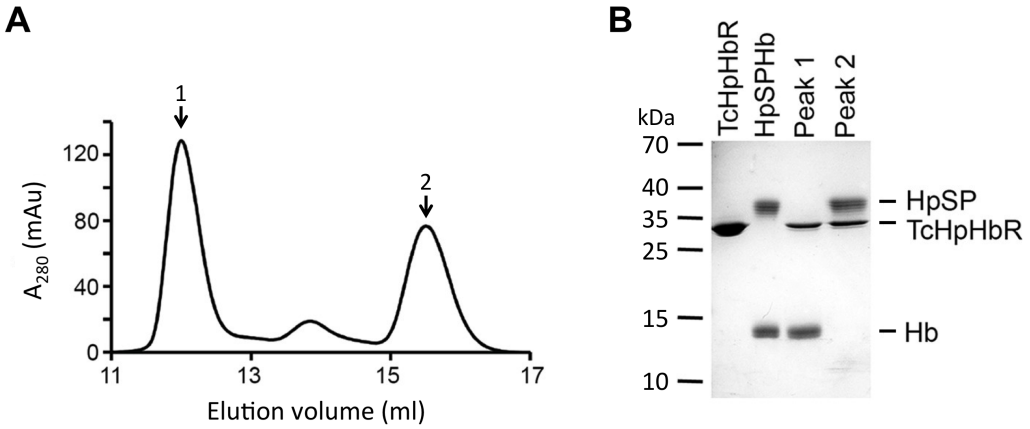


FIGURE 5.8: Analytical gel filtration demonstrating the consequence of mixing HpSPHb and TcHpHbR. (A) The sample was applied to a Superdex 200 10/300 column. (B) Samples from peak fractions were run on SDS-PAGE.

Native haptoglobin is cleaved to form α- and β-subunits, whilst the recombinant HpSP used in this study does not undergo this cleavage. This gel filtration was therefore repeated, this time using full length dimeric HpHb (Figure 5.9). Peak 1 contained only HpHb, whilst all TcHpHbR was in the second peak. This indicates two things. Firstly, the disassembly of haptoglobin-haemoglobin complexes seen for the recombinant protein, HpSPHb, would not happen *in vivo*, as it does not occur with native HpHb. Secondly, the affinity of TcHpHbR for HpHb is not high enough to keep the complex together during gel filtration and must therefore be lower than the affinity of the receptor for Hb.

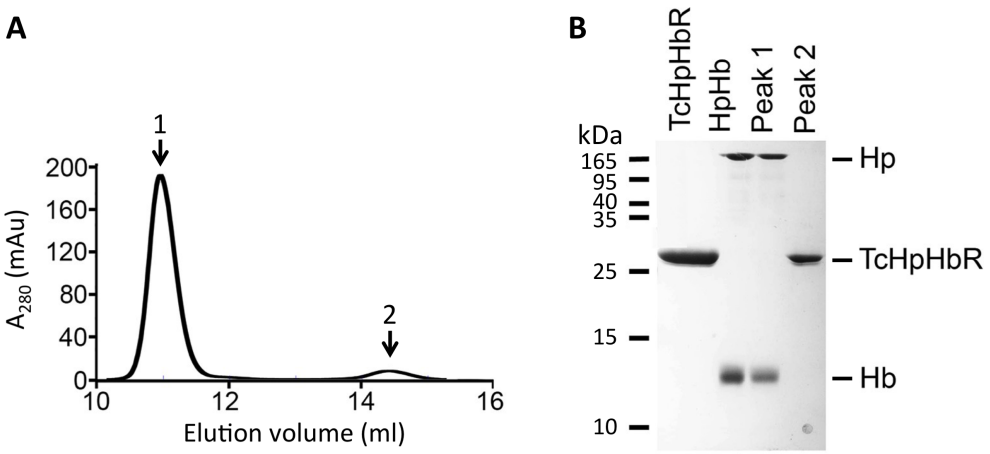


FIGURE 5.9: Analytical gel filtration demonstrating the consequence of mixing HpHb and TcHpHbR. (A) The sample was applied to a Superdex 200 10/300 column. (B) Samples from peak fractions were run on SDS-PAGE.

In summary, TcHpHbR is able to promote the disassembly of HpSPHb into its two components *in vitro* by favourably binding Hb over HpSPHb . This will remove free Hb from the solution, causing more HpSPHb to disassemble in order to maintain the equilibrium between HpSPHb and its free components. As the same is not seen for HpHb, recombinant HpSP must have a weaker affinity for Hb that native Hp does, presumably due to the cleavage of Hp that normally occurs during its processing. The receptor would therefore be unable to cause the dissociation of haemoglobin from haptoglobin *in vivo*.

5.3.2 Investigating the stoichiometry of the TcHpHbR:Hb interaction in solution

The crystal structure of TcHpHbR:Hb showed two receptors binding to one haemoglobin tetramer. To assess whether this stoichiometry occurs in solution, isothermal titration calorimetry was performed. TcHpHbR was injected into a cell containing Hb (Figure 5.10). The stoichiometry showed 2 receptors binding to one Hb tetramer. This confirms

that the complex seen in the crystal structure can also occur in solution, with one receptor binding to each of the two identical binding sites of the tetramer. The K_D was estimated to be 3 nM, approximately 1000-fold higher than the affinity of the Tc receptor for HpHb, which was previously measured by ITC to be 3 μM [94]. This much higher affinity indicates that the haptoglobin-haemoglobin receptor may have evolved from an ancestral receptor that primarily bound haemoglobin.

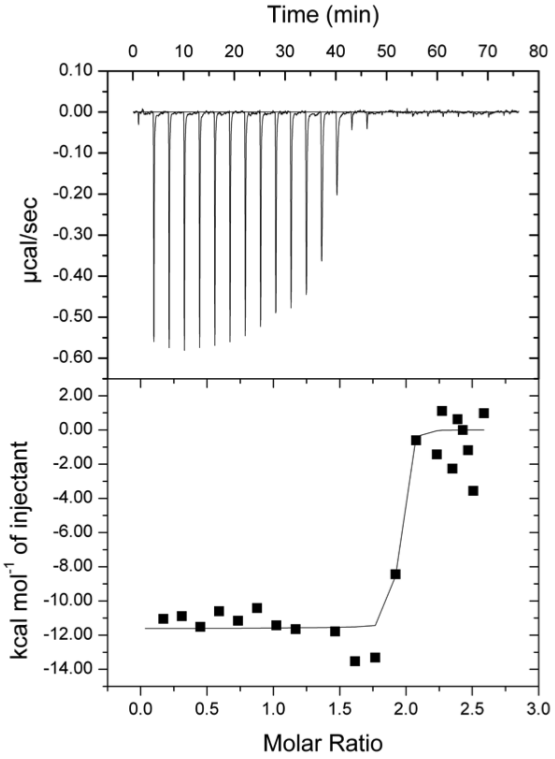


FIGURE 5.10: Isothermal titration calorimetry data for TcHpHbR titrated into human Hb. The cell contained 18 μM Hb whilst the syringe contained 250 μM TcHpHbR. The molar ratio indicates that two TcHpHbR can bind to a single Hb tetramer.

5.3.3 Investigating the evolution of the ligand specificity of HpHbRs

TbbHpHbR had previously been shown to bind only the complex formed between Hp and Hb, and neither of the individual components [43]. Since it had now been shown that

TcHpHbR was able to bind Hb alone, an experiment was designed to determine whether the ancestral HpHbR was likely to have been an Hb or HpHb receptor. SPR was carried out by immobilising the receptors of *T. brucei brucei*, *T. congolense* and *T. vivax* on a chip through a C-terminal biotin moiety (Figure 5.11). This allowed them to be presented on the chip in the same orientation that they are presented on the surface of the cell. HpHb and Hb were then flowed over the chip to investigate binding. The receptor of *T. vivax* was chosen as this trypanosome species diverged from *T. brucei* and *T. congolense* before they diverged from each other (Figure 5.12) [205]. As expected, TbbHpHbR interacted with HpHb, but not with Hb alone. However, both TcHpHbR and TvHpHbR interacted with both HpHb and Hb alone. As the experiment was performed in this orientation, the binding curves recorded correspond to a mixture of monovalent and bivalent binding for both ligands, as there are two receptor binding sites on both of them. Unfortunately the opposite orientation was not possible, as the tetramer of Hb fell apart when coupled to the chip. Therefore affinities cannot be found from these results. However, the shapes of the curves do indicate that the stability of the complex with Hb is much greater than with HpHb, as supported by the affinity found by ITC for TcHpHbR.

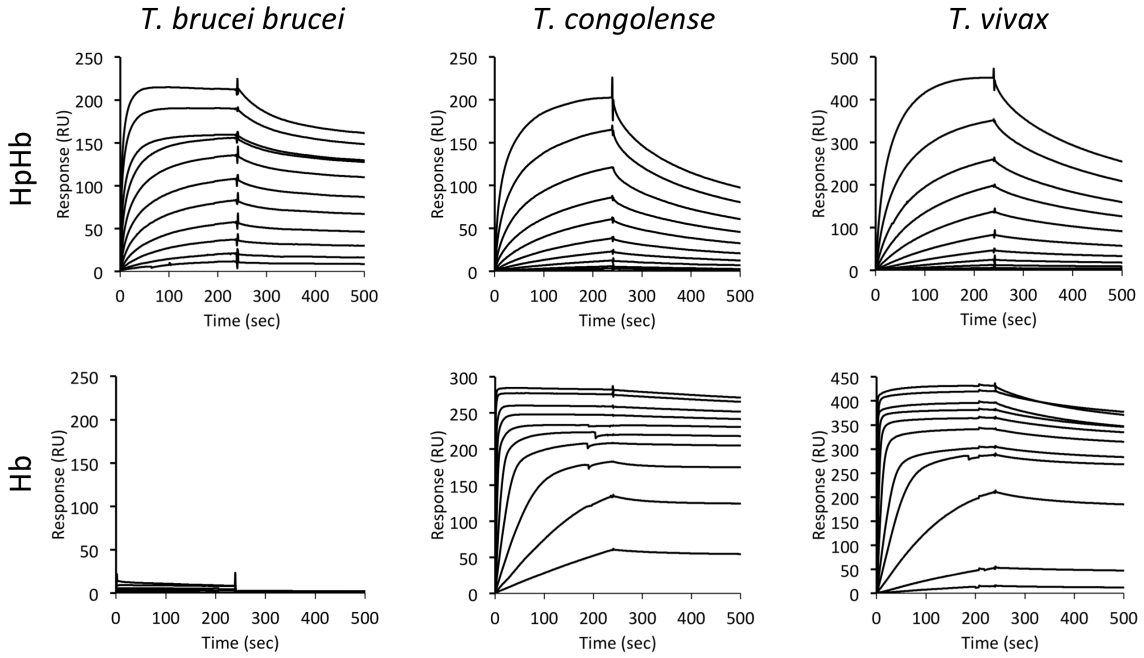


FIGURE 5.11: Surface plasmon resonance curves for binding of HpHb and Hb to HpHbRs. One in two dilutions of the ligands were flowed over a chip coated with receptor, starting from 1 μM.

As *T. vivax* diverged from *T. brucei brucei* and *T. congolense* before they diverge from each other (Figure 5.12), the results suggest that TbbHpHbR has lost the ability to bind Hb alone since the species diverged from *T. congolense*, as it is more unlikely that both *T. congolense* and *T. vivax* gained the ability to bind Hb separately.

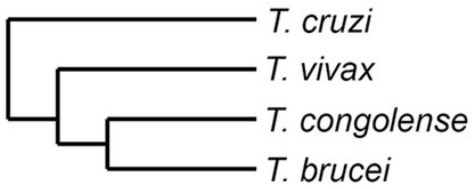


FIGURE 5.12: A phylogenetic tree indicating the evolutionary history of the trypanosome species and subspecies under study [205].

T. congolense does not infect humans, but is instead a livestock pathogen. Sequence alignments had demonstrated that the residues identified in the crystal structure as interacting with TcHpHbR were unaltered between a variety of species haemoglobins. However,

it was important to confirm this experimentally. SPR was therefore performed, this time flowing bovine Hb over the chip coated with TcHpHbR (Figure 5.13). As before, the response curves revealed a strong interaction with fast on- and slow off-rates. This confirms that TcHpHbR also has a high affinity for bovine Hb.

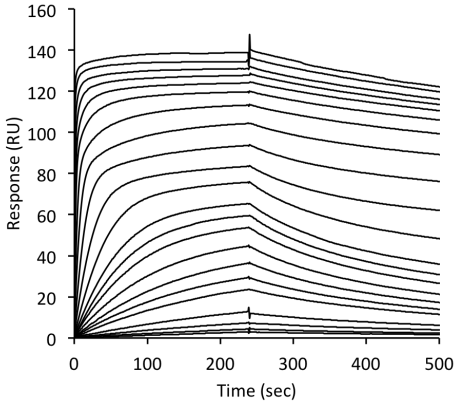


FIGURE 5.13: Surface plasmon resonance curves for binding of bovine Hb to TcHpHbR. One in two dilutions of bovine Hb were flowed over a chip coated with TcHpHbR, starting from 1 μM.

5.3.4 Investigating the role of HpHbR in *T. congolense* by assessing its expression profile and functionality

In the human bloodstream, levels of free haemoglobin are very low due to the presence of Hp, which binds Hb with high affinity [190]. It was therefore unclear what the purpose of a haemoglobin receptor could be for *T. congolense* if, like *T. brucei*, the receptor was expressed during the bloodstream lifecycle stage. Proteomic analysis of *T. congolense* had previously shown that TcHpHbR was not detectable in the bloodstream form, but was an abundant protein in epimastigotes [206]. Similarly, transcriptome analysis of *T. vivax* indicated that TvHpHbR mRNA was most abundant in epimastigotes [207]. The epimastigotes of *T. congolense* develop in the mouth parts of the tsetse fly, unlike those of *T. brucei* which are found in the salivary glands. In the mouthparts of the fly the parasites

will be exposed to a bloodmeal, which could be a source of free haemoglobin. We therefore decided to investigate during which lifecycle stage the receptor was present and functional in *T. congolense* in a series of experiments detailed in this section, all of which were carried out by Dr Paula MacGregor in the Carrington lab, Cambridge. First, Western blot analysis of culture-derived epimastigotes indicated not only that TcHpHbR is only expressed at detectable levels in the epimastigote form (Figure 5.14A), but that an average of ~5-9 x 10^5 copies of TcHpHbR are expressed per cell (Figure 5.14B), about 1000-fold more than in *T. brucei* [43, 127].

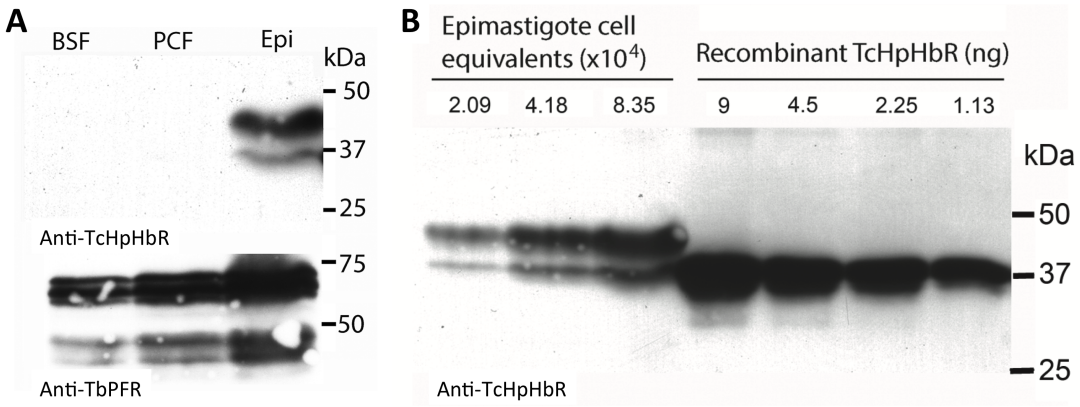


FIGURE 5.14: Western blot analysis of TcHpHbR expression. (A) Expression can only be detected in epimastigote forms, and in neither bloodstream nor procyclic. (B) Expression levels were quantified by Western blot, estimating that an average of ~5-9 x 10^5 copies of TcHpHbR are expressed per epimastigote. This figure was produced by Dr. Paula MacGregor.

Expression was then detected using using anti-TcHpHbR polyclonal sera and a fluorescently-tagged anti-rabbit IgG secondary antibody. Differentiation to epimastigote form can be identified due to a repositioning of the kinetoplast from a position posterior to the nucleus in procyclic forms, to juxtanuclear or anterior to the nucleus in epimastigotes [58, 208]. This demonstrated that in all cells where repositioning had occurred, the receptor was expressed across the whole cell surface, including the flagellum. This can be seen in both non-permeabilised culture-derived epimastigotes (Figure 5.15D arrowhead) as

well as permeabilised epimastigotes isolated from the tsetse fly mouthparts (Figure 5.15C arrowhead). Parasites from the mouthparts that had not yet undergone differentiation to the epimastigote stage (Figure 5.15D arrow), and parasites isolated from the proventriculus and midgut of the tsetse fly (Figure 5.15A and B) did not express TcHpHbR at detectable levels. The observation that TcHpHbR expression is detected over the entire cell surface of the epimastigotes differs from the expression pattern of TbHpHbR, which is thought to be predominantly restricted to the flagellar pocket [43].

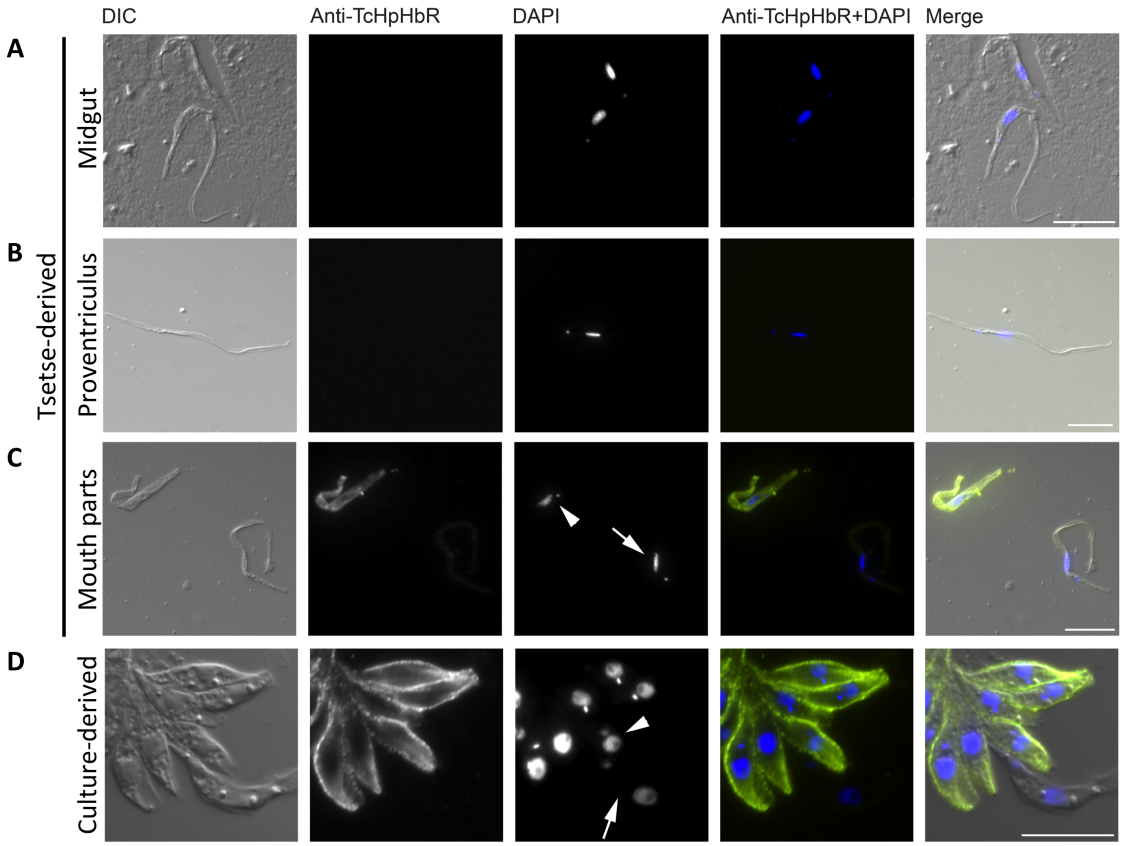


FIGURE 5.15: Immunofluorescence analysis of TcHpHbR expression in *T. congolense*. TcHpHbR was detected using anti-TcHpHbR polyclonal sera and a fluorescently-tagged anti-rabbit IgG secondary antibody. A, B and C show parasites isolated from the tsetse fly and fixed using paraformaldehyde (permeabilised cells), whilst D shows culture-derived parasites fixed using methanol (non-permeabilised cells). Arrows indicate trypomastigotes, whilst arrowheads highlight epimasigotes. All scale bars represent 10 μm. This figure was produced by Dr. Paula MacGregor.

To determine whether TcHpHbR enables the internalisation of haemoglobin, culture-derived epimastigotes were incubated with 10 nM Alexa-488-labelled Hb or BSA. Internalisation of Hb, but not BSA, was observed specifically in epimastigote forms but not in trypomastigote forms (Figure 5.16). This demonstrates that TcHpHbR is able to internalise Hb at low nanomolar concentrations, as would be predicted based on the affinity obtained from ITC of 3 nM.

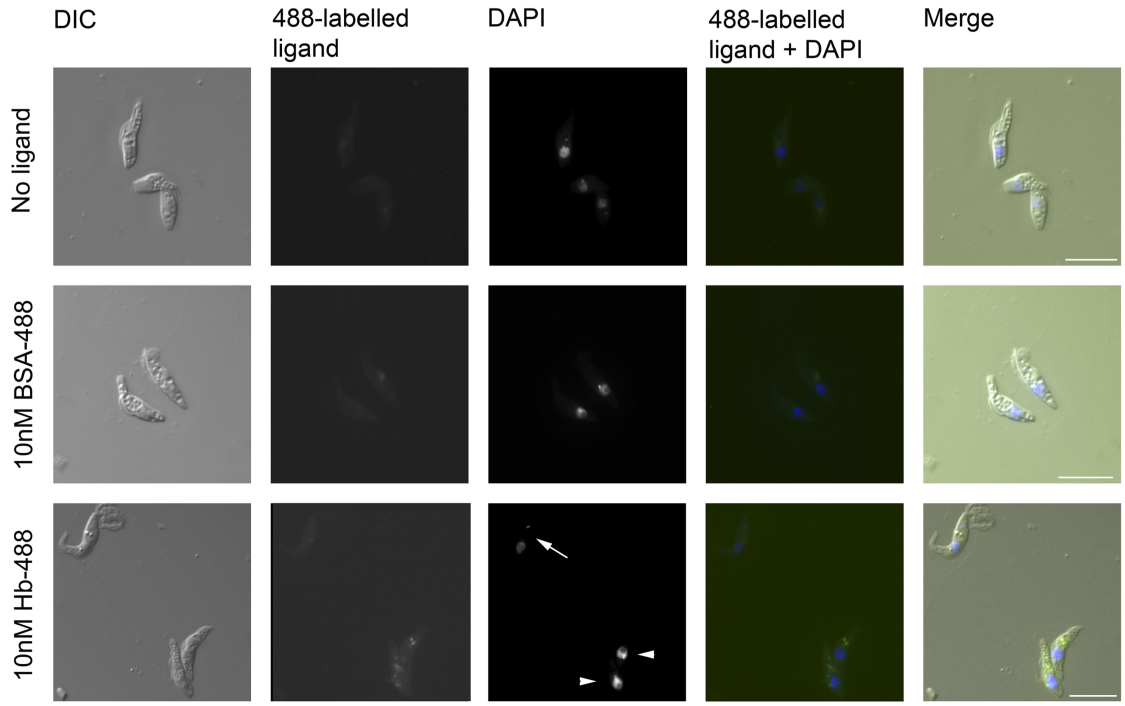


FIGURE 5.16: The uptake of Alexa-488-labelled Hb (Hb-488) and Alexa-488-labelled BSA (BSA-488) into *T. congolense* culture-derived epimastigotes. Uptake of Hb-488 was detected at 10 nM in epimastigotes (arrowhead) but not trypomastigotes (arrow) (centre panel). No fluid phase uptake of BSA-488 at 10 nM was observed into any cells (bottom panel). All scale bars represent 10 μm. This figure was produced by Dr. Paula MacGregor.

In summary, this data indicates that TcHpHbR is indeed a haemoglobin receptor which is able to internalise its ligand. It is highly expressed in the epimastigote form of *T. congolense* at ~1000-fold higher levels than TbHpHbR in bloodstream *T. brucei* and is found across the whole surface of the parasite.

5.4 Discussion

The results in this chapter have identified significant differences between HpHbR of *T. congolense* and *T. brucei*. TbHpHbR is expressed by the bloodstream form of *T. brucei*, enabling the parasite to take up haem in the form of HpHb. The receptor is also responsible for the uptake of TLF1, resulting in cell death of the *T. brucei brucei* subspecies. It had been expected that HpHbR of other species would have a similar expression pattern and function. However, the results described here demonstrate that the receptor has an approximately 1000-fold greater affinity for haemoglobin than for haptoglobin-haemoglobin and *T. congolense* can take up Hb at low nanomolar concentrations. Additionally, the receptor is predominantly expressed by the epimastiogte form when in the mouth parts of the tsetse fly, rather than by the bloodstream form in the mammalian host. These results have subsequently been confirmed by others [209]. Similar results are seen for TvHpHbR, which can also bind haemoglobin and is preferentially expressed in epimastigotes, at least at the mRNA level [207].

Although the structure revealed in this chapter shows the complex between TcHpHbR and Hb, it can also be used to gain insight into the binding of this receptor to HpHb. It was already known that docking the haemoglobin α-subunit of HpHb onto the same subunit in the complex structure would result in the haptoglobin molecule being too far from the receptor to make contacts, as this had been shown for the TbbHpHbR complex structure (Figure 3.23). Therefore a model was generated by aligning the HpHb haemoglobin β-subunit onto the haemoglobin α-subunit that contacts the upper binding site in the TcHpHbR:Hb structure (Figure 5.17). In this orientation, the residues of the haemoglobin α-subunit that make contacts with the membrane-distal binding site are replaced with chemically similar residues of the β-subunit. The Hp is positioned adjacent to the receptor, and a hydrophobic

patch on the surface of TcHpHbR is seen in the same location as the patch on TbbHpHbR, allowing interaction with the hydrophobic patch of Hp. This suggests that TcHpHbR interacts with HpHb using similar binding sites to those seen in the TbbHpHbR. The finding that when the α-subunit of Hb is replaced with the β-subunit in this structure all the same interactions can still occur suggests that Hb is likely to be able to bind TcHpHbR in either orientation, with either an α-subunit or β-subunit interacting with each binding site. Despite this, the orientation seen in the crystal structure is likely to represent a preferred orientation, as otherwise a mixture of both would have been observed.

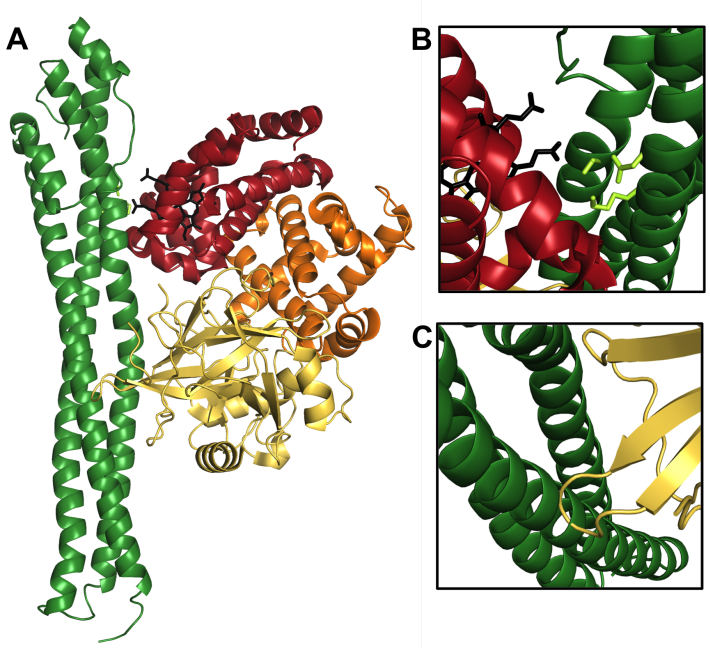


FIGURE 5.17: A model of TcHpHbR bound to HpSPHb, based on the TcHpHbR:Hb structure. The β-subunit of HpSPHb has been aligned with the α-subunit of Hb in the TcHpHbR:Hb structure that interacts with the receptor.

The TbbHpHbR:HpHb structure revealed that when two receptors are bound to one dimeric HpHb molecule, the C-termini of the receptors point in the same direction due to the kink in the receptor. This is likely to allow dimeric binding to occur in the context of the cell surface. As the kink is not present in TcHpHbR, a model was created using the

porcine HpHb structure [133] bound to two TcHpHbRs to assess whether this interaction would still be possible when both receptors are anchored to the cell membrane (Figure 5.18). The model demonstrates that a large tilt angle would be required for this dimeric interaction to take place, so it may not be possible *in vivo*. The K_D of TcHpHbR for HpHb (8 μM) is already lower than that of TbbHpHbR (1 μM), indicating that dimeric binding would surely be even more necessary. This suggests that binding and uptake of HpHb in the bloodstream of the mammalian host may not be a crucial role of the receptor in *T. congolense*.

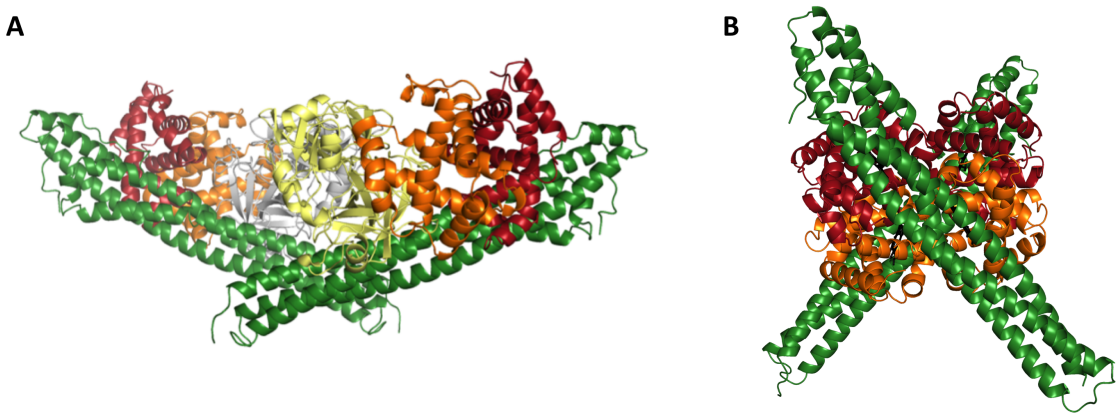


FIGURE 5.18: A comparison of the tilt angles required for two TcHpHbRs to bind one HpHb or Hb. (A) A model of two TcHpHbRs bound to one HpHb dimer. This was produced by aligning two copies of the model produced in the previous figure with the porcine structure of dimeric HpHb (pdb:4F4O) [133]. (B) Two TcHpHbRs bound to a single Hb tetramer, as seen in the crystal structure. The tilt angle required for dimeric binding is much greater for HpHb than for Hb.

By comparing the TcHpHbR:Hb structure to that of TbbHpHbRt:HpSPHb, insight can be gained into the changes that have occurred to cause loss of Hb binding by TbbHpHbR (Figure 5.19). TbbHpHbR was aligned with TcHpHbR in the structure of the complex with Hb. This revealed that the path of helix I in TbbHpHbR is altered so that it no longer forms the membrane-proximal binding site for the Hb β-subunit as the interacting residues are too far apart. This will reduce the surface area of the interaction by nearly half, so it is

unsurprising that this interaction is not observed.

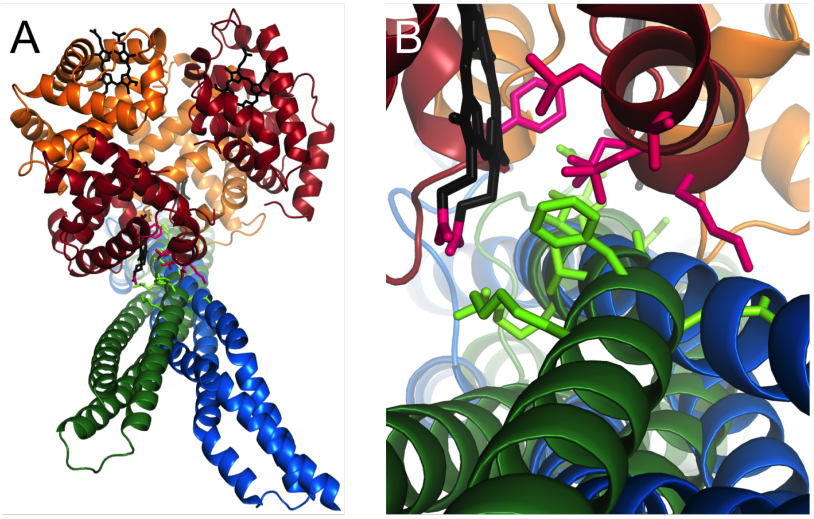


FIGURE 5.19: The path of helix I significantly diverges between TbHpHbR and TcHpHbR. (A) This model was produced by aligning the TcHpHbR:Hb and TbHpHbR:HpSPHb complexes, using the haemoglobin subunit that interacts with the membrane distal binding site to carry out the alignment. Residues shown in bright green or pink are those that interact in the TcHpHbR:Hb structure. (B) shows a close up with the C-termini of the receptors off the bottom of the image.

The change in the path of helix I could be a result of two different things. Firstly, as discussed in chapter 3, TbbHpHbR has a 50° kink approximately a third of the way from the C-terminus. To form the kink may have required a change in the path of helix I. Secondly, the affinity of TbbHpHbR for HpHb had previously been shown by SPR to be 1 μM, higher than the affinity of TcHpHbR (8 μM). Therefore it may be that the adaptions in TbbHpHbR that resulted in this higher affinity also caused this alteration in the path of helix I, causing loss of Hb binding as a side effect.

Together this data suggests that TcHpHbR represents an ancestral form of the receptor, the purpose of which is to bind haemoglobin during a bloodmeal in the tsetse fly mouthparts, providing the parasite with a source of haem (Figure 5.20). Expression levels are ~1000-fold higher than for TbHpHbR and the protein can be detected over the entire surface of the cell, not just the flagellar pocket. The bloodmeal is very transient, and it may

be possible that this expression pattern is required in order to capture enough haemoglobin during the brief period of exposure. This is in contrast with *T. brucei* where HpHbRs will be constantly exposed to their ligand in the bloodstream. It is unclear how the bloodstream form of *T. congolense* is able to obtain haem. It may be that there is a very low level of TcHpHbR expressed during this stage, albeit below detectable level, but perhaps enough to provide the required haem. Alternatively, it may be possible that enough HpHb is taken up by fluid phase uptake to fulfil the needs of the parasite.

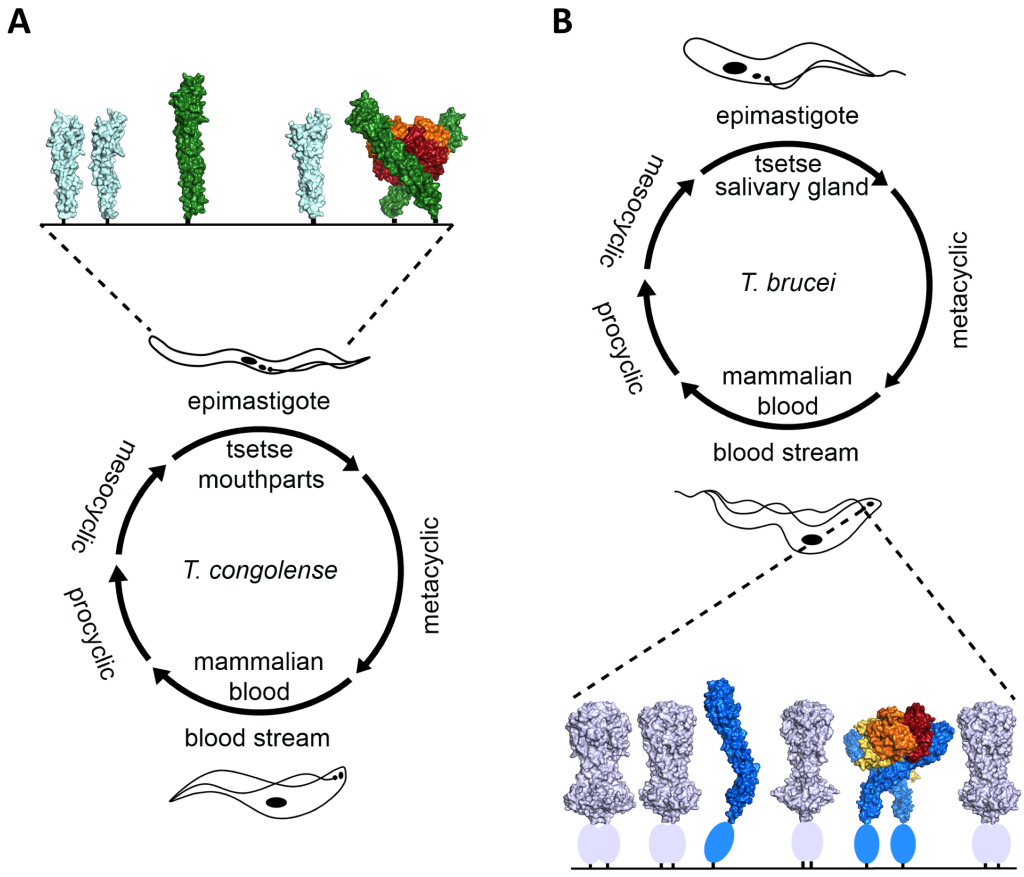


FIGURE 5.20: A comparison of the role of HpHbR during the lifecycle of (A) *T. congolense* and (B) *T. brucei*. The *T. congolense* epimastigote expresses an Hb-binding form of the receptor across the cell surface, whilst in the bloodstream form of *T. brucei* the receptor is predominantly found in the flagellar pocket and can only bind HpHb. This figure is taken from [46].

These results allow an evolutionary pathway of HpHbR to be proposed. In the ancestral trypanosome, the receptor is expressed by the epimastigote across the whole surface of the cell. The receptor has a high affinity for haemoglobin, which it is able to bind and internalise during a bloodmeal, due to the location of the epimastigotes in the mouth parts of the tsetse fly. The first step is the change in this location to the salivary glands, as occurs in the lifecycle of *T. brucei*. This may perhaps be to avoid toxic components of the bloodmeal, or to avoid competition with other trypanosome species. In this new location, HpHbR becomes redundant as there is no access to the bloodmeal, and thus haemoglobin. Expression of the receptor is therefore altered to occur instead in the bloodstream form of the trypanosome. However, as most of the haemoglobin in the blood is in complex with haptoglobin, the receptor undergoes evolutionary changes which increase its affinity for this new ligand. This is evidenced by the difference in affinity of TcHpHbR (8 μM) and TbHpHbR (1 μM) for HpHb [94].

On the surface of the *T. congolense* epimastigote, two receptors can bind to one haemoglobin tetramer with only a moderate tilt around the GPI-anchor. However, the surface of the bloodstream form trypanosome is densely packed with VSG molecules, so the tilt that would be required for two TcHpHbRs to bind one dimeric HpHb molecule would be unfeasible (Figure 5.18). TbHpHbR has overcome this by developing a ~50° kink towards the membrane-proximal C-terminus of the receptor. Thus when two receptors bind to one dimeric HpHb molecule, the C-termini of both receptors are pointing in the same direction. This allows dimeric binding of HpHb without the need for a large tilt angle at the GPI-anchor (Figure 5.21). TbHpHbR also has a C-terminal domain that probably increases the distance of the binding site from the cell surface, enabling better access of the ligand.

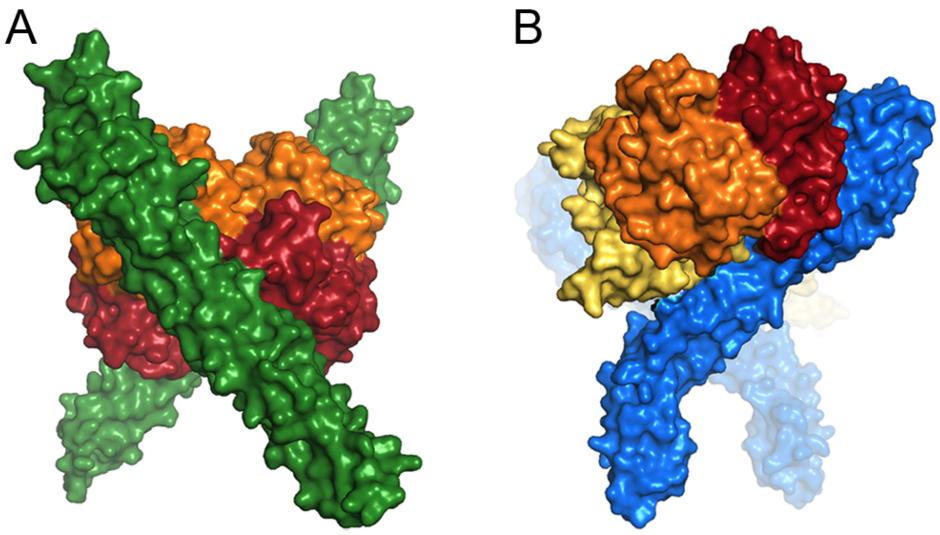


FIGURE 5.21: Comparison of the ligand binding of (A) TbHpHbR and (B) TcHpHbR. TcHpHbR (green) must tilt at the GPI-anchor site to allow dimeric binding of haemoglobin (α orange, β red). TbbHpHbR (blue) has overcome this by introducing a kink, allowing dimeric binding of HpHb (Hp yellow) without tilting.

The kink may also have another role in the context of the cell surface, pushing apart the densely packed VSG molecule on the surface to allow better access to the HpHb binding site. However, a side effect of these alterations to the structure of the receptor is to lose the ability to bind haemoglobin, probably due to the alteration in the path of helix I (Figure 5.19). Finally, the expression levels of HpHbR were greatly reduced to ~200-400 per cell, and expression is restricted to the flagellar pocket. Both of these changes are likely to be an attempt to evade the immune system of the human host by minimising detection of the receptor.

Therefore, HpHbR has evolved from an epimastigote-expressed haemoglobin receptor into a haptoglobin-haemoglobin receptor expressed during the bloodstream stage of *T. brucei*. Due to this change in lifecycle stage, the receptor has undergone a number of alterations to function efficiently in this different surface environment.

Chapter 6

Expression, purification and characterisation of Apolipoprotein L1

6.1 Summary

Apolipoprotein L1 (ApoL1) is the protein responsible for the innate immunity of humans and some other primates to most species of African trypanosome. It is known to form pores in lipid bilayers, but the structure of these pores and how they result in cell death remains unclear. I therefore wished to characterise this protein both structurally and biophysically, to determine how ApoL1 forms pores and in which membrane this takes place. Initially, I needed to establish a method for the expression and purification of ApoL1, which I succeeded in doing using *E. coli* cells to insolubly express the protein, before refolding by flash dilution. I then used a number of methods to verify that the refolded protein was functional. To aid both structural and functional characterisation, I aimed to produce a panel of antibodies against ApoL1. I was able to obtain one such antibody using phage display, and a second was obtained from hybridomas. Crystallisation of ApoL1 has been

unsuccessful to date, so instead I began investigating the use of nanodiscs in conjunction with cryo-electron microscopy as a tool to visualise this protein and the pores it forms. Lastly, I established an electrophysiology assay which can be used to assess the ability of ApoL1 to form pores in a variety of different lipid bilayers and pHs, as well as investigating the roles of different domains of ApoL1 in pore function. Initial results suggest that lipid composition of the bilayer in which ApoL1 inserts alters the kinetics of pore opening.

6.2 Expression and purification of ApoL1

6.2.1 Expression of ApoL1

Although expression and purification of ApoL1 had been demonstrated previously, the protocols used were complex [117, 119]. It was therefore important to establish a new protocol for large scale expression of ApoL1. To begin with, expression of ApoL1 in *E. coli* was attempted using a non-codon optimised gene corresponding to residues 28-398, as residues 1-27 have been identified as a signal sequence [149]. A pEt-15b plasmid was used, which contains an N-terminal 6-His tag followed by a thrombin protease cleavage site. Induction of ApoL1 expression in BL21(DE3) cells at 25 °C led to cell death, with only a minute pellet obtained. This indicated that expression had probably been successful, but that the protein was toxic to these cells. Next, expression in both baculovirus-infected *Sf9* cells, S2 insect cells and HEK 293F cells was attempted. In all cases, no expression was observed by Western blot. Lastly, C43(DE3) *E. coli* cells were used, as these contain mutations that were phenotypically selected for conferring tolerance to the expression of toxic proteins. A very small amount of protein was seen in this case after purification by soluble Ni-NTA^{2+} affinity chromatography, but this was lost during the gel filtration

purification step (Figure 6.1). This was probably at least partially due to the lack of detergent in the running buffer, which was later found to be crucial for the solubility of ApoL1.

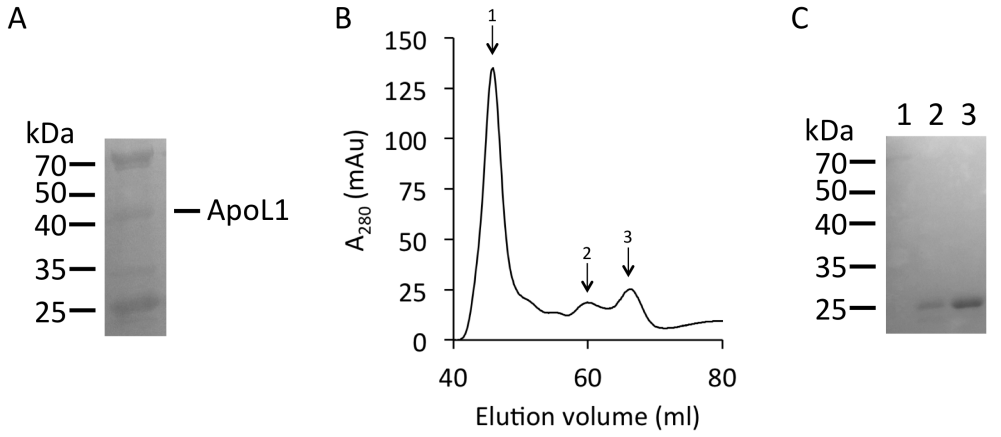


FIGURE 6.1: Purification of full length ApoL1 expressed by C43(DE3) cells. (A) A very small amount of protein at the correct molecular weight can be seen in the Ni-NTA^{2+} affinity chromatography elution by SDS-PAGE, although the identity of the band was not confirmed using another method. (B) Gel filtration gave a very large peak in the void. The sample was applied to a Superdex 75 16/600 in 20 mM Hepes pH 7.5, 150 mM NaCl. (C) SDS-PAGE of fractions collected from the three gel filtration peaks showed that none contained the correct protein.

As BL21(DE3) cells appeared to be capable of expressing ApoL1, albeit in such a way that was toxic to the cells, this expression system was focused on. This was similar to the expression system of Thomson et al. [119], which also used insoluble expression in *E. coli*, although the refolding method was different. To encourage insoluble expression, an induction temperature of 37 °C overnight was used, and this time the ApoL1 gene was codon-optimised for *E. coli*. These alterations aimed to increase expression levels and promote insoluble expression, making it less likely that ApoL1 would kill the cells due to the ability of the correctly folded protein to form pores in membranes. Using this protocol, cell death no longer occurred, and the protein appeared to be expressed both solubly and insolubly (Figure 6.2). However, soluble expression was much more impure, and mass

spectrometry had revealed a contaminant in previous expression attempts, EF-Tu, with a molecular weight of 43.3 kDa. This is very similar to the molecular weight of full length ApoL1, 41.1 kDa. Therefore ApoL1 was only purified from insoluble material, as this was more pure, thus avoiding contamination issues. Whilst this expression protocol avoided the initial problem of cell death, optimisation of the refolding of ApoL1 was now required. ApoL1, solubilised in 6 M guanidine buffer, was purified by Ni-NTA^{2+} affinity chromatography. The elution from this could then be used in refolding trials.

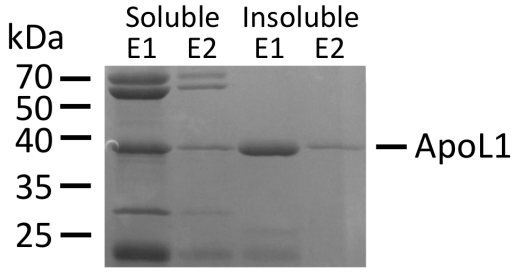


FIGURE 6.2: Purification of full length ApoL1 expressed by BL21 cells and a codon-optimised gene. SDS-PAGE of elution fractions following purification by Ni-NTA^{2+} affinity chromatography suggests that both soluble and insoluble expression was successful. However, soluble ApoL1 is much more impure.

A Western blot of the Ni-NTA^{2+} affinity chromatography elution from insoluble material was carried out. The primary antibody used was either anti-6His (Figure 6.3 A) or anti-ApoL1 C-terminal peptide (Figure 6.3 B). The anti-ApoL1 C-terminal peptide was a polyclonal sera produced in rabbit (Covalab). Both confirmed the identity of the band as ApoL1, and this was subsequently verified by mass spectrometry.

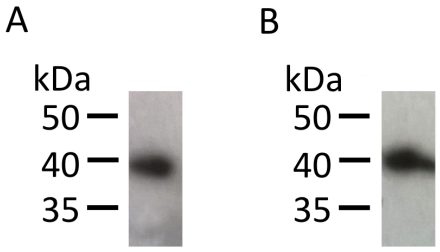


FIGURE 6.3: Western blots assessing the identity of ApoL1. The primary antibodies used were (A) anti-6His HRP-conjugate monoclonal and (B) anti-ApoL1 C-terminal peptide polyclonal sera. For the latter, a secondary anti-rabbit peroxidase-conjugate monoclonal was used for detection.

However, when this protein was gel filtrated, it eluted in the void volume (Figure 6.4). When the purification was next completed, 0.02% n-dodecyl-β-D-maltopyranoside (DDM) was added, as carried out by Thomson et al. [119]. This successfully overcame the issue of aggregation (Figure 6.6).

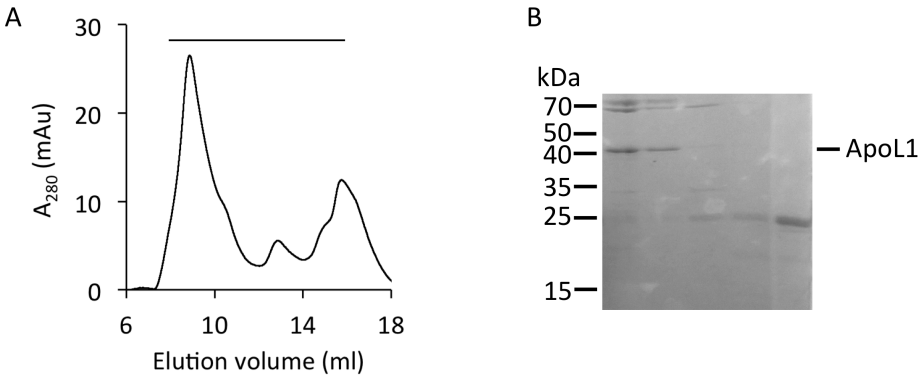


FIGURE 6.4: Gel filtration of ApoL1 in 20 mM Hepes pH 7.5, 150 mM NaCl using a Superdex 200 16/600. (A) The gel filtration trace shows a large proportion of the protein eluting in the void volume. The horizontal bar indicates which fractions were run on SDS-PAGE. (B) SDS-PAGE confirms that ApoL1 elutes in the void volume.

6.2.2 ApoL1 refolding

Initially, refolding was successfully carried out by flash dilution into PBS, 0.02% DDM. Overnight dialysis into this buffer was also attempted, but resulted in complete precipi-

tation of the protein. A number of different flash dilution refolding buffers were tested to see whether an increased proportion of ApoL1 could be refolded successfully. The buffer screen used was the QuickFold protein refolding buffer screen. Buffers containing triton x-100 were not selected, due to the difficulty of removing this detergent by gel filtration. Instead, 0.02% DDM was added to each refolding buffer. To test refolding, 10 μl of guanidine-solubilised ApoL1 was rapidly added to 190 μl of the refolding buffer and mixed. After incubation at room temperature for 1 hour, the sample was gel filtered. As such a small volume was being injected onto the gel filtration column, the peak height of ApoL1 was likely to vary not only due to differences in refolding, but also due to inaccuracies in the volume injected. Therefore, the ratio of the ApoL1 peak relative to the peak of a smaller molecular weight contaminant was assessed in order to rank the suitability of each refolding buffer (Figure 6.5A).

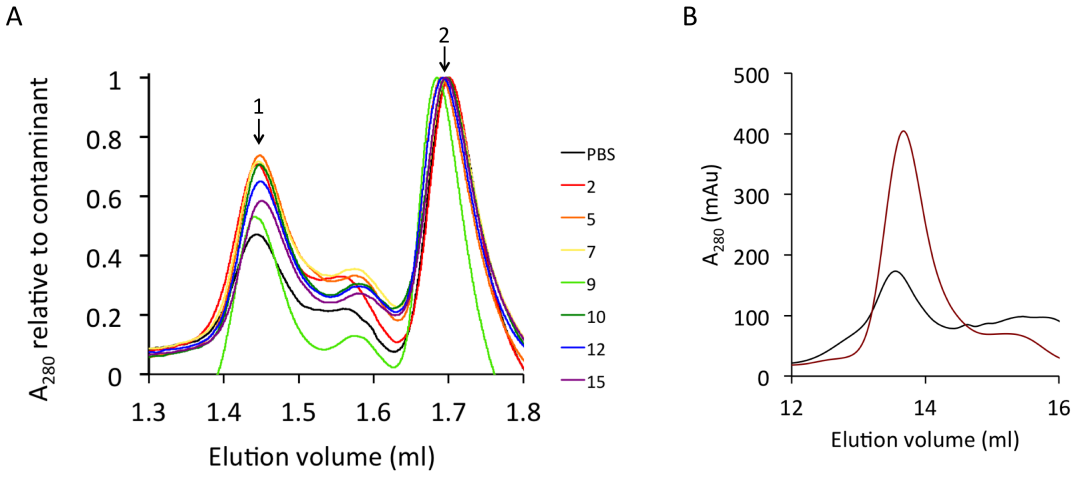


FIGURE 6.5: Gel filtration curves for ApoL1 refolded by flash dilution into different refolding buffers. (A) The curves are shown relative to the maximum UV signal for each curve, which relates to elution of the contaminant at 1.7 ml (peak 2). ApoL1 elutes at 1.45 ml (peak 1). This allowed the buffers which increased the relative proportion of ApoL1 to contaminant to be identified. These gel filtration runs used a Superdex 200 3.2/300 (B) Using 50 mM MES pH 6, 250 mM NaCl, 0.02% DDM (red) instead of PBS, 0.02% DDM (black) greatly increased the yield of a single purification of N-terminally truncated ApoL1 (see section 6.2.3). These gel filtration runs used a Superdex 200 10/300.

Buffers 5, 7, 2 and 10 appeared to be the best at refolding ApoL1 relative to the contaminant. The most common features of these buffers were a pH of 6 and approximately 250 mM salt. The refolding buffer designed from this screen contained 50 mM MES pH6, 250 mM NaCl, 0.02% DDM. As ApoL1 does not contain any cysteines, a reducing agent was not necessary. The improvements in yield can be clearly seen for a single purification of N-terminally truncated ApoL1, increasing from 100 μg/l of *E. coli* to 300 μg/l of *E. coli* (Figure 6.5B).

Expression, purification and refolding of full length ApoL1 was successful using the protocol established above (Figure 6.6). A pH 4.8 buffer was generally used for gel filtration, as this is approximately the pH ApoL1 would experience in the lysosome. However, the protein was also successfully gel filtered at neutral pH, and this did not appear to affect yield.

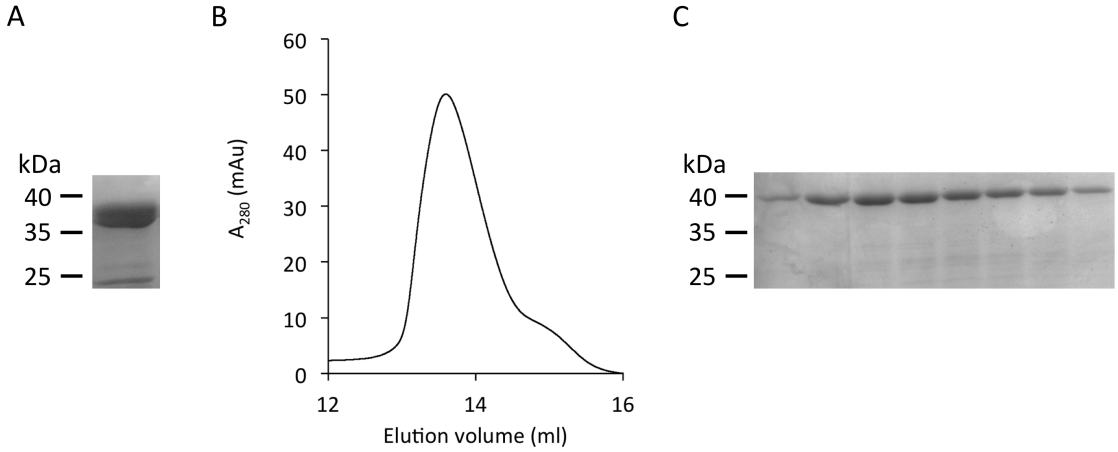


FIGURE 6.6: Purification of full length ApoL1. (A) ApoL1 (41 kDa) was purified by Ni-NTA^{2+} affinity chromatography followed by flash dilution refolding. (B) The protein was further purified by gel filtration in 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8 using a Superdex 200 10/300. (C) Purity of the gel filtration elution fractions was was confirmed by SDS-PAGE.

6.2.3 Expression and purification of ApoL1 truncations

Truncations of ApoL1 were required both to improve the chances of crystallisation by reducing flexibility of the protein, and to allow the different domains of the protein to be assessed in functional assays to determine their roles. To design different constructs of ApoL1, the sequence was used to predict regions of disorder and secondary structure (Figure 6.7).

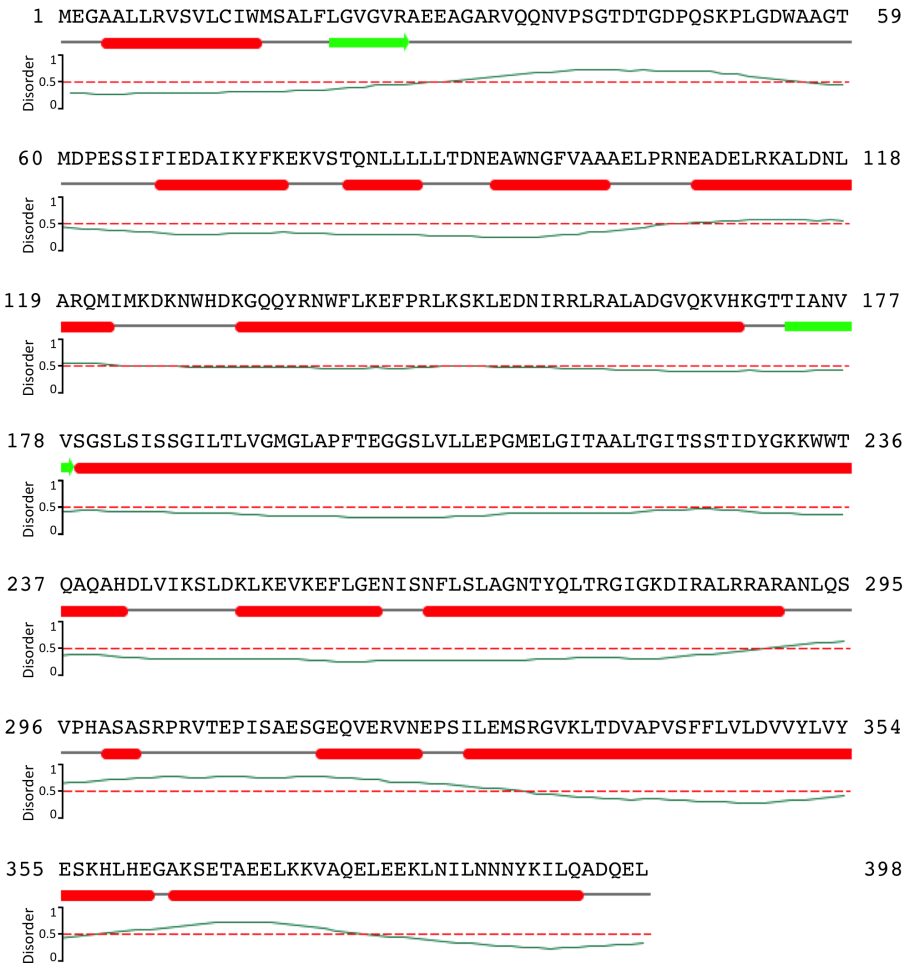


FIGURE 6.7: Prediction of ApoL1 secondary structure and disorder. Secondary structure was predicted by JPred [147]. Red cylinders represent α-helices, whilst green arrows represent β-sheets. Disorder was predicted by RONN [148]. Values above 0.5 are predicted to be disordered.

Residues 1-27 have been identified as a signal sequence [149]. Following this, the region between 28-57 is predicted to be disordered. In fact, ApoL1 of two different masses has been found when this protein is extracted from human blood; the predominant form contains residues 28-398, whilst the other contains residues 54-398, lacking this disordered N-terminus [149]. It is unclear whether these two forms have a different activity, and it is also possible that the cleavage occurred during isolation of the protein, rather than *in vivo*. Therefore, an obvious choice to potentially increase protein stability was an N-terminally truncated construct. Based on the disorder prediction, this construct contained residues 60-398. Expression and purification was carried out following the same protocol as for full length ApoL1 (Figure 6.8).

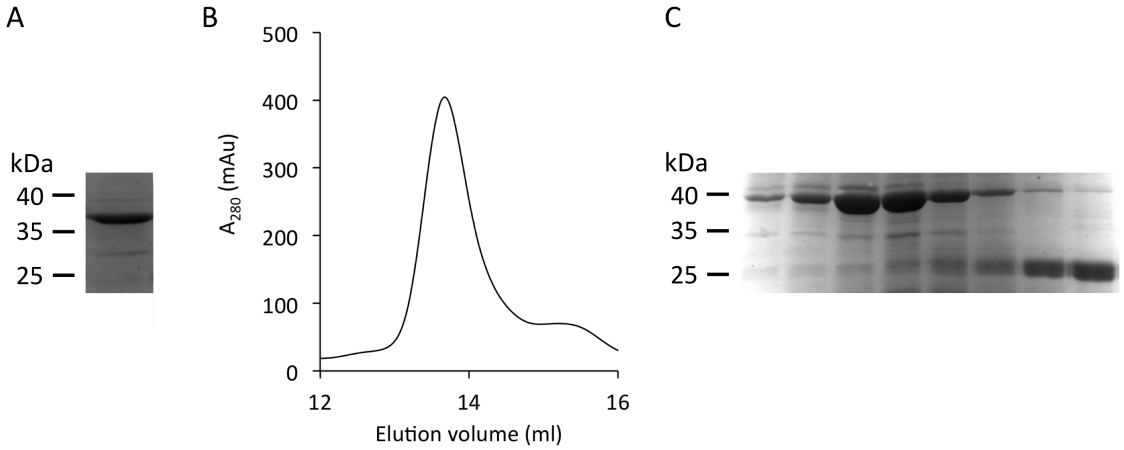


FIGURE 6.8: Purification of N-terminally truncated ApoL1 (38 kDa) by gel filtration. (A) Truncated ApoL1 was purified by Ni-NTA^{2+} affinity chromatography followed by flash dilution refolding and (B) gel filtration in 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8 using a Superdex 200 10/300. (C) Purity of the gel filtration elution fractions was assessed by SDS-PAGE.

Next, constructs corresponding to predicted domains of ApoL1 were designed. Region E77-W235 had been previously predicted to share structural similarity with the pore-forming domain (PFD) of colicin A, residues V389-H592. It is therefore predicted to have a crucial role in pore formation by ApoL1 and results have indicated that it is bac-

tericidal [117]. A construct was designed containing residues 60-242. Finally, to assess the structure and function of the part of the protein not thought to resemble colicin A, a DNA construct comprising of 243-398 was produced by Dr Sebastian Zoll, named the C-terminal domain (CTD). The CTD is thought to be predominantly responsible for interaction with SRA [116], and may also be important for interaction with the membrane in which the protein inserts [117]. For both of these constructs, expression and purification was successful following the same protocol as used for full length ApoL1 (Figure 6.9).

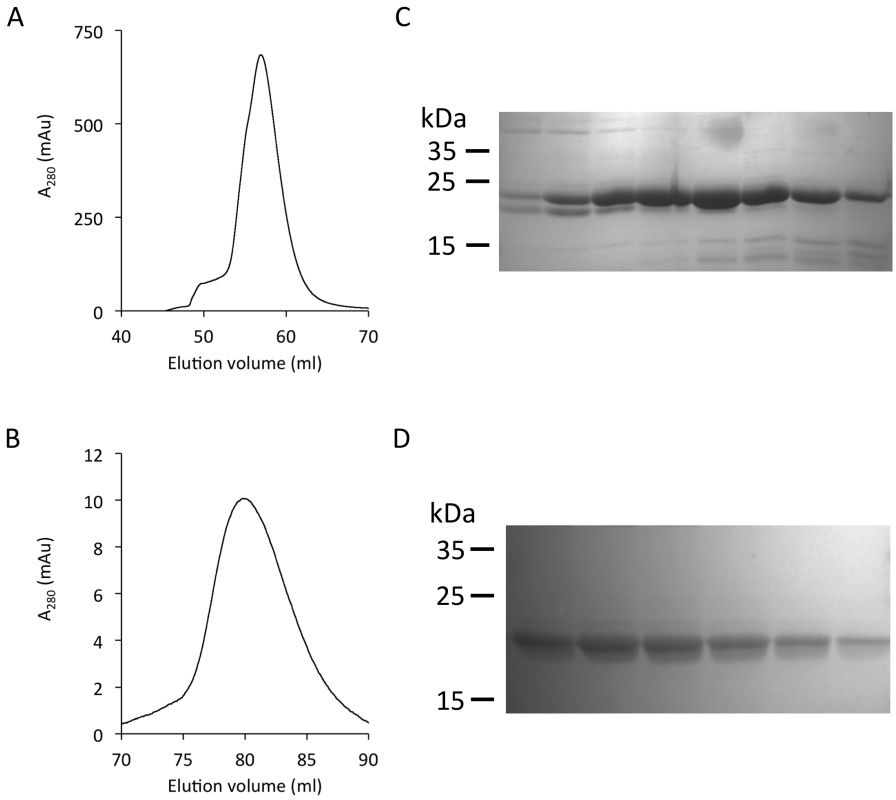


FIGURE 6.9: Purification of the pore-forming domain (20 kDa) and C-terminal domain (18 kDa) of ApoL1 by gel filtration. After Ni-NTA^{2+} affinity chromatography and flash dilution refolding, (A) PFD and (B) CTD were purified by gel filtration using a Superdex 75 16/600 and a Superdex 75 16/600, respectively, in 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8. The purity of the gel filtration elution fractions of (C) PFD and (D) CTD was confirmed by SDS-PAGE.

A summary of the ApoL1 constructs designed and expressed is shown in Figure 6.10.

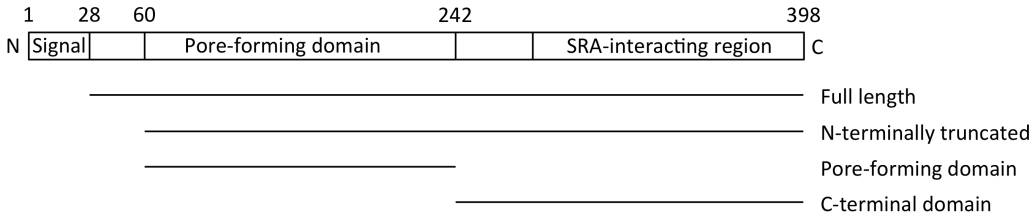


FIGURE 6.10: Summary of ApoL1 constructs. The diagram of ApoL1 domains is based on the figure from [113]. All constructs were expressed, purified and refolded following the same protocol.

6.2.4 Functionality of refolded ApoL1

As the ApoL1 being used was refolded from insoluble material, it was particularly important to confirm that it was functional. The first indication of this was gained from circular dichroism, which indicated that the full length protein, PFD and CTD were all predominantly α-helical (Figure 6.11). This is in agreement with the secondary structure predicted from the sequence (Figure 6.7).

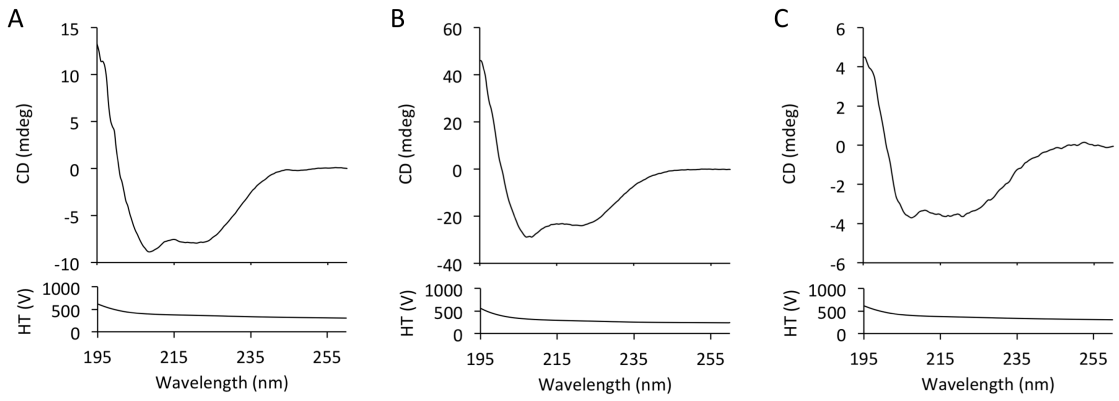


FIGURE 6.11: Circular dichroism of (A) full length ApoL1, (B) the pore-forming domain and (C) the C-terminal domain. All show predominantly α-helical secondary structure. Protein was at a concentration of 0.1-0.2 mg/ml in 100 mM phosphate buffer pH 7.5, 150 mM NaF, 0.02% DDM. The high tension is below the acceptable value of 700 V in all cases.

Next, the ability of full length recombinant ApoL1 to bind its interaction partner, SRA, was assessed. Firstly, SPR was performed by flowing SRA over immobilised ApoL1 (Figure 6.12). ApoL1 was randomly biotinylated at primary amines, and coupled to a CAPture chip. As the CAPture chip is coated with single-stranded DNA covalently linked to streptavidin, the surface is negatively charged. SRA is positively charged and therefore bound non-specifically to the chip. To overcome this issue, methylated SRA was used, reducing the overall charge of the protein. However, as this is not a physiological modification, the K_D obtained from this (0.1 nM) may not reflect the true value of the interaction *in vivo*. It does still clearly show that recombinant ApoL1 can interact with this methylated SRA.

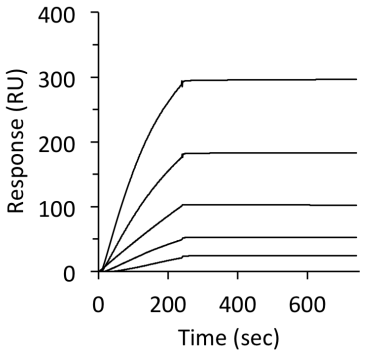


FIGURE 6.12: Surface plasmon resonance binding curves of the interaction between methylated SRA and immobilised ApoL1. 1 in 2 dilutions of SRA were used, starting with a concentration of 125 nM. The running buffer contained 25 mM citrate, 150 mM NaCl, 0.005% Tween-20, pH 4.8.

To confirm that the interaction occurred with unmethylated SRA and obtain an accurate K_D, microscale thermophoresis (MST) was performed by Dr. Sebastian Zoll (Figure 6.13). This was carried out using Alexa-488-labelled SRA and unlabelled N-terminally truncated ApoL1 in a buffer of 25 mM citrate pH, 150 mM NaCl, 0.02% DDM pH 4.8. A K_D of 328 nM was found.

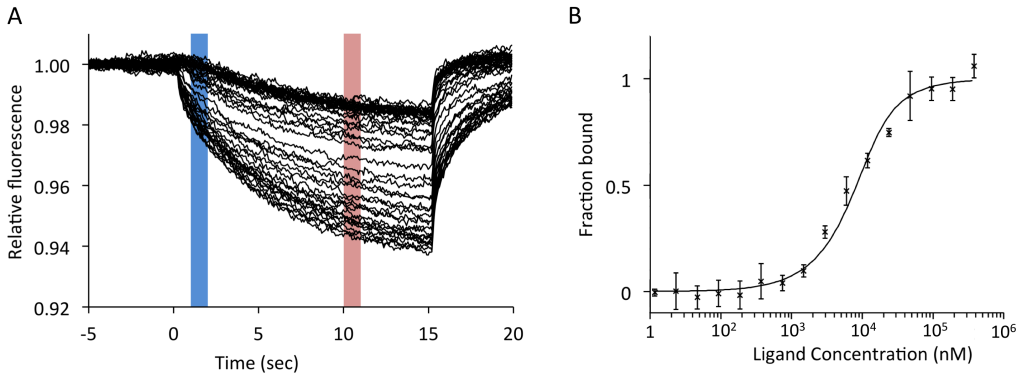


FIGURE 6.13: Microscale thermophoresis of fluorescently labelled SRA binding to ApoL1. (A) Fluorescence time traces. The fluorescence after thermo-diffusion (F_1, red) was normalised to the fluorescence after the T-jump (F_0, blue). (B) Dose response curves produced using the values obtained from (A). The experiment was carried out in triplicate. This figure was produced by Dr. Sebastian Zoll.

To assess the ability of recombinantly produced and refolded ApoL1 to kill trypanosomes, *T. b. brucei* were incubated with a range of ApoL1 concentrations in an experiment conducted by Prof. Mark Carrington (Figure 6.14A). This demonstrated that recombinant ApoL1 was able to kill *T. b. brucei* at low nM concentrations, similar to concentrations previously demonstrated to kill trypanosomes [117].

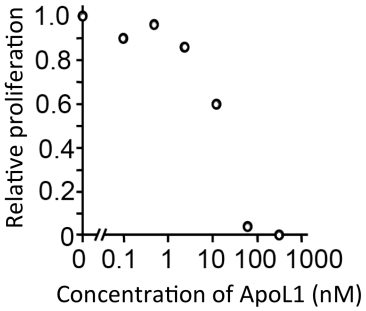


FIGURE 6.14: Recombinant ApoL1 kills *T. b. brucei*. Assay of *T. b. brucei* proliferation over 24 hr in culture in the presence of different ApoL1 concentrations (where 1 = 5-fold increase in cell density). This figure was produced by Prof. Carrington.

Together, this data indicates that recombinant ApoL1 is correctly folded and able to function in the same way as native ApoL1.

6.2.5 The effect of detergent on ApoL1 homogeneity

For purification of ApoL1 and its truncations, all buffers from flash dilution onwards included 0.02% DDM, as the presence of detergent was required to stop aggregation of the protein. However, it was important to screen a number of different detergents to assess their ability to keep ApoL1 soluble, as detergents with different characteristics may be useful for different experiments. For example, detergents with shorter chains were more likely to succeed in the crystallisation of ApoL1. Purified ApoL1 was gel filtered in a number of different detergents and detergent concentrations (Figure 6.15). In this experimental setup ApoL1 is likely to be bound by mixed detergent micelles containing DDM as well as the detergent in which it is being gel filtered. However, it is more high throughput than methods that would completely exchange the detergent, and still gives a good indication of how the protein will behave in different detergents.

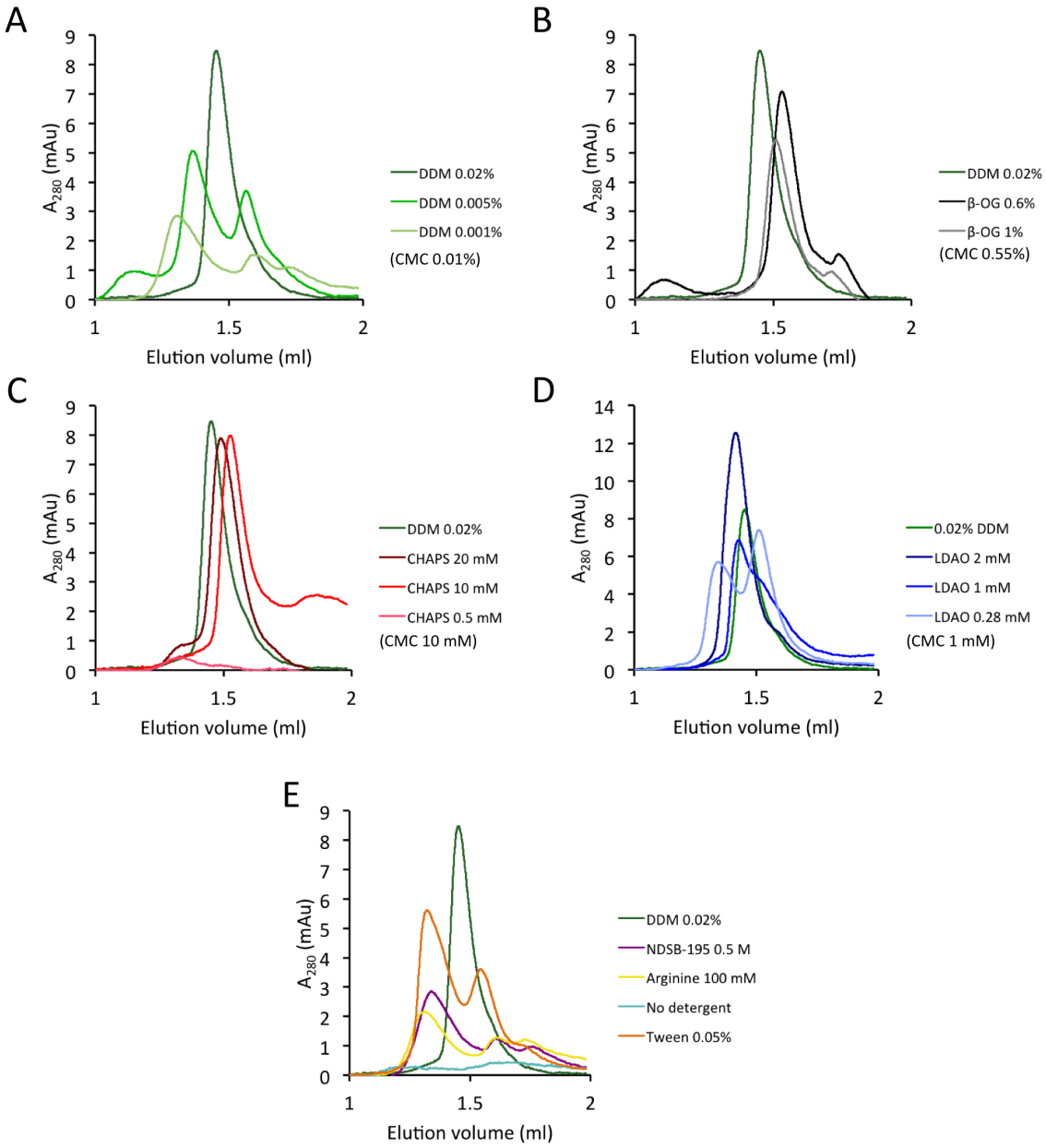
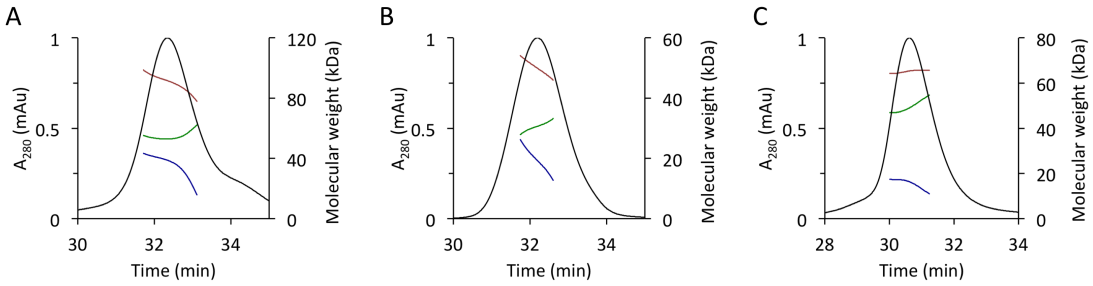


FIGURE 6.15: Gel filtration curves for ApoL1 in gel filtered in 20 mM Hepes, 150 mM NaCl and a variety of different detergents using a Superdex 200 3.2/200. In all cases the ApoL1 sample had already been gel filtered in 20 mM Hepes pH 7.5, 150 mM NaCl, 0.02% DDM.

In all cases, keeping the detergent concentration at twice the critical micelle concentration (CMC) was crucial. The detergents in which ApoL1 was

most stable were n-dodecyl β-D-maltoside (DDM), octyl-β-glucoside (β-OG), 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) and lauryldimethylamine N-oxide (LDAO). This gave a good panel of detergents to be used in crystallisation trials, and may be useful for other experiments.

SEC-MALLS was used to assess how much detergent was bound to each construct (Figure 6.16). Using this technique, the overall molecular mass can be separated into contributions made by protein and detergent using a differential refractive index detector. The value given by this detector allows the ratio of protein:detergent to be determined, as the refractive index of pure protein and pure detergent is known.



D

	Mw (kDa)	Mw_{app} (kDa)	Detergent (kDa)	Total (kDa)
ApoL1	41	40	53	93
PFD	20	21	30	51
CTD	18	17	48	65

FIGURE 6.16: SEC-MALLS traces for (A) full length ApoL1 (B) PFD and (C) the CTD. All runs were completed using a Superdex 200 10/300 with a running buffer of 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8. The data was analysed using the ASTRA software to find the total molecular weight (red), the contribution of detergent (green) and the contribution of protein (blue). (D) Comparison of the calculated molecular weight of the protein with the apparent weight found by SEC-MALLS.

The results indicate that a large amount of DDM is bound, in all cases more than the molecular weight of the protein. This is likely to make crystallisation difficult, reducing the exposed surface area which can make protein-protein contacts to stabilise a crystal. The

molecular weight of the protein component of PFD, and to some extent full length ApoL1, increases towards the left of the peak. This indicates that some aggregation may be present. In all cases, the protein component decreases and detergent component increases towards the right of the peak. This is because the detergent micelle, with a molecular weight of 50-70 kDa, elutes shortly after the protein peak.

6.3 Generation of anti-ApoL1 antibodies

6.3.1 Selection of anti-ApoL1 scFvs by phage display

Attempts to obtain anti-ApoL1 antibodies by phage display were completed during a visit to MedImmune. N-terminally truncated ApoL1 was biotinylated non-specifically on primary amines, and this was then immobilised on a streptavidin-coated plate. Three libraries of phage containing the genes for randomly generated single chain variable fragments of antibodies (scFvs) were then deselected for streptavidin-binding, and both the phage and ApoL1-coated plates were blocked to reduce non-specific interactions. The phage were incubated with plates before washing, followed by trypsin digestion to elute bound phage. After expansion of these phage by infection into *E. coli*, this cycle was repeated a total of four times. 88 phage from each library were then selected for sequencing to analyse diversity.

TABLE 6.1: Sequencing of scFv displayed by phage after 4 rounds of selection. The BMV library was derived from bone marrow, peripheral blood lymphocytes and tonsil tissues [154]. The CL library was derived from spleenocyte cDNA [155]. The Ref library is a synthetic library in which only the CD3s of the VH and VL have been randomised (MedImmune, unpublished).

Library	Total no.	Failed	Frame shifts	Stop codons	Truncations	Remaining no.	Repeats	Diversity
Ref	88	10	0	3	0	75	1	99%
CS	88	0	5	4	7 (V_H)	72	2	97%
BMV	88	27	0	11	0	50	8	84%

After each round of phage display panning the phage should be enriched for anti-ApoL1 phage, resulting in a decrease in the diversity of scFv sequences that they display. Ref and CS libraries do not show much decrease in diversity, suggesting that further rounds of phage display are required. Unfortunately, due to time constraints, this was not possible. Instead, 66 of the successfully sequenced phage were selected for ELISA (Figure 6.17). ELISAs were carried out at both pH 7 and 4.8, and were conducted against the full length protein, PFD and CTD. Out of the 66 tested, 57 reacted significantly to N-terminally truncated ApoL1 and not to strepatavidin. Signal was considered significant if it was greater than 110% of the background signal. Of these, some reacted to both the CTD and PFD. These were discarded, as this is likely to be the result of non-specific binding. Of the others, 5 were selected that bound the CTD at both pHs, 5 that bound only N-terminally truncated ApoL1 at both pHs, and one each that bound the N-terminally truncated protein only, or N-terminally truncated ApoL1 and CTD at pH 7 only. No phage bound PFD and N-terminally truncated ApoL1 only.

	Strep	NTT	PFD	CTD	Strep	NTT	PFD	CTD	Strep	NTT	PFD	CTD
A	0.041	1.4	0.049	0.192	0.042	1.533	0.053	0.322	0.545	1.471	0.623	1.044
B	0.051	1.797	0.063	0.295	0.051	1.622	0.049	0.116	0.051	1.421	0.063	0.163
C	1.577	1.715	1.736	1.619	0.044	1.484	0.056	0.076	0.047	1.536	0.066	0.222
D	0.072	1.753	0.091	0.192	0.049	1.708	0.06	0.159	0.055	1.492	0.056	0.282
E	0.047	1.441	0.054	0.108	0.044	1.461	0.046	0.086	0.044	1.715	0.05	0.168
F	0.087	1.427	0.054	0.088	0.052	1.551	0.049	0.075	0.05	1.597	0.071	0.655
G	1.772	1.717	1.632	1.617	0.077	1.282	0.048	0.705	0.064	1.833	0.062	0.297
H	0.062	1.191	0.058	0.143	0.056	0.053	0.05	0.053	0.047	0.058	0.063	0.043
	1	2	3	4	5	6	7	8	9	10	11	12

FIGURE 6.17: An example of the readings for a phage ELISA plate with incubation steps performed at pH 4.8. The phage were tested for binding to streptavidin alone (Strep), N-terminally truncated ApoL1 (NTT), PFD and CTD. In this example, the phage incubated in wells C1-4 is able to bind streptavidin alone, so is discarded. The phage incubated in wells C5-8 is able to bind the N-terminally truncated ApoL1, but neither PFD nor CTD. The phage incubated in wells G5-8 is able to bind both N-terminally truncated ApoL1 and CTD.

The scFvs displayed by the selected phage were purified by Ni-NTA^{2+} affinity chromatography by MedImmune. All were then tested by SPR for binding to N-terminally truncated ApoL1, PFD and CTD (Figure 6.18). In all cases, the ApoL1 protein was randomly biotinylated on primary amines and coupled to a CAPture chip. scFvs were flowed over the chip at concentrations of 1.5 μM in a pH 7.5 running buffer. Of the 12 scFvs tested, only one demonstrated an affinity for ApoL1 (scFv8). As predicted from the ELISA, this scFv recognised the CTD. The affinity from this single curve was calculated to be 1 μM, although a concentration series is required to confirm this.

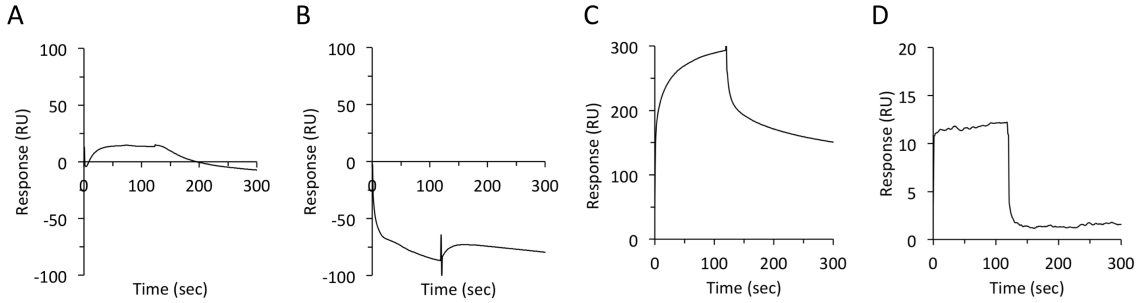


FIGURE 6.18: Surface plasmon resonance curves of the binding of a scFv to immobilised ApoL1 constructs. Binding to (A) N-terminally truncated ApoL1 and (C) the C-terminal domain occurs, whilst (B) the pore-forming domain does not bind. (D) scFv expressed and purified from T7 shuffle cells also appears to bind truncated ApoL1, although the shape of the binding curves is different. The scFv was at 1.5 μM for A, B and C, and at 20 μM in D.

To produce large amounts of this scFv, the gene was cloned into a pEt-15b vector for expression in T7 Shuffle cells. This was successful, and the yield was very high. After purification by Ni-NTA^{2+} affinity chromatography, the protein was gel filtered (Figure 6.19). To confirm that this protein was still able to recognise ApoL1, SPR was once again performed (Figure 6.18D). The binding curve indicates that both the on- and off-rate of binding for the scFv expressed in T7 shuffle cells have increased dramatically. The interaction should be confirmed using a different technique before this scFv is pursued further.

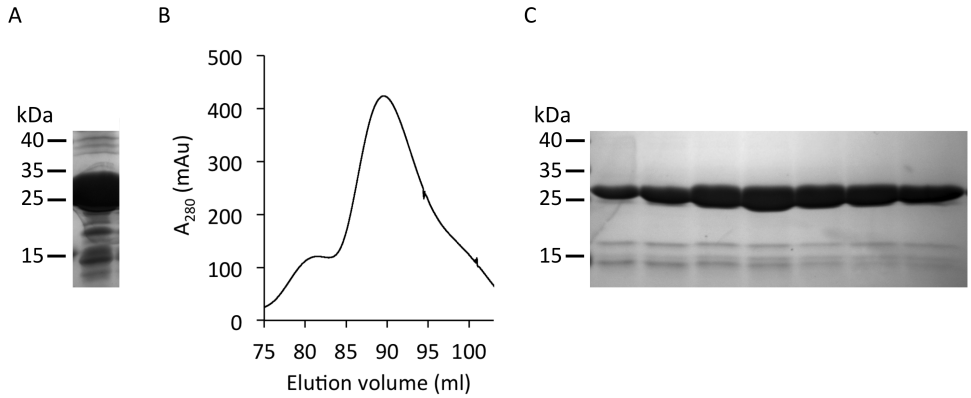


FIGURE 6.19: Purification of anti-ApoL1 scFv. The scFv was purified first (A) by Ni-NTA^{2+} affinity chromatography followed by (B) gel filtration using a Superdex 200 16/600. (C) Purity of gel filtration elution fractions was confirmed by SDS-PAGE.

6.3.2 Production and selection of anti-ApoL1 mAbs using hybridomas

Several attempts were made to obtain anti-ApoL1 antibodies from hybridomas. ApoL1, N-terminally truncated ApoL1, PFD, CTD and ApoL1:SRA were all used as immunogens, injected into mice after purification and endotoxin removal. Hybridoma generation for the last two constructs was carried out by Dr. Sebastian Zoll. Three injections of the protein sample mixed with adjuvant were given to each mouse over 42 days, before the spleens were harvested on day 50. Splenocytes were fused with immortalised myeloma cells to produce hybridomas using the ClonaCell-HY hybridoma kit (Stemcell Technologies). Successful fusion to form hybridomas was then selected for, and individual colonies were used to immunise each well of several 96 well plates. 500-1000 colonies were selected for each construct, except PFD which, surprisingly, did not generate any colonies.

Once each culture had been expanded, the supernatant was tested for reactivity to ApoL1 by ELISA. A single antibody against ApoL1:SRA was identified. This was tested

by SPR, and shown to bind both N-terminally truncated ApoL1 and the CTD (Figure 6.20).

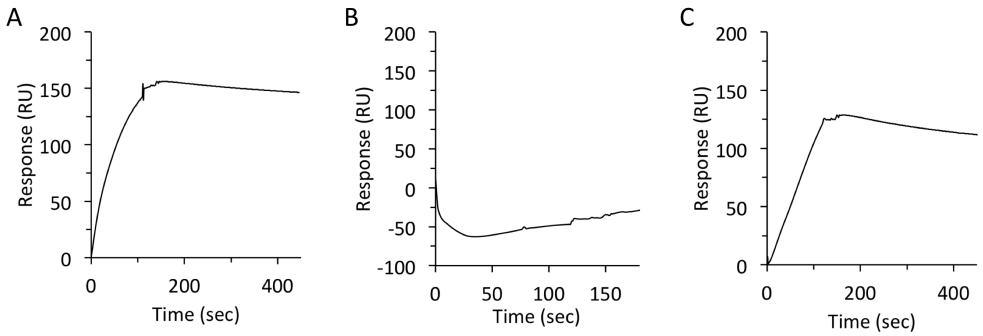


FIGURE 6.20: Surface plasmon resonance binding curves of a mAb binding to immobilised ApoL1 constructs. Binding to (A) N-terminally truncated ApoL1 and (C) the C-terminal domain occurs, whilst (B) the pore-forming domain does not bind. This was completed using concentrated supernatant rather than purified protein, so the concentration of antibody was unknown.

6.4 Attempts to crystallise ApoL1

Many attempts were made to crystallise ApoL1, using different detergents and truncations of ApoL1. Unless stated otherwise, crystallisation was attempted at 4 °C due to the instability of the protein. In all cases, trials were carried out in 96-well format using vapour diffusion of sitting drops. Crystallisation of full length ApoL1 was not attempted, due to the predicted disorder of the N-terminal region. Both full length and N-terminally truncated ApoL1 precipitate when the protein is concentrated to above ~4 mg/ml, a very low concentration at which to attempt crystallisation. One plate of N-terminally truncated ApoL1 at 4 mg/ml in 1% β-OG was set up using the MemGold 2 broad screen. In an attempt to overcome the issue of precipitation, N-terminally truncated ApoL1 was methylated. This time 1.6% β-OG was used, but a concentration of only 3.5 mg/ml was reached. Three broad screens were set up using MemGold 2, Morpheus and JCSG+.

PFD alone was more likely to succeed in crystallisation as it was predicted to be a

single domain, thus minimising flexible loops. First, 0.02% DDM was used and the broad screens MemGold 2, Morpheus, PACT and JCSG+ were set up at 10 mg/ml. Using PFD at 15 mg/ml in 0.05% LDAO, plates were set up using MemGold 2, JCSG+, Morpheus and PACT, this time at both 4 and 18 °C. Lastly, plates were set up at 9 mg/ml in 0.6% NG using MemGold 1, MemGold 2 and JCSG+. No crystals were seen in any of these conditions. Although the result of the crystallisation trial differed from plate to plate, in general 70% of the drops phase separated, 10% were clear, and 20% resulted in precipitation (Figure 6.21). A large amount of phase separation is not surprising, given the presence of detergent [210].

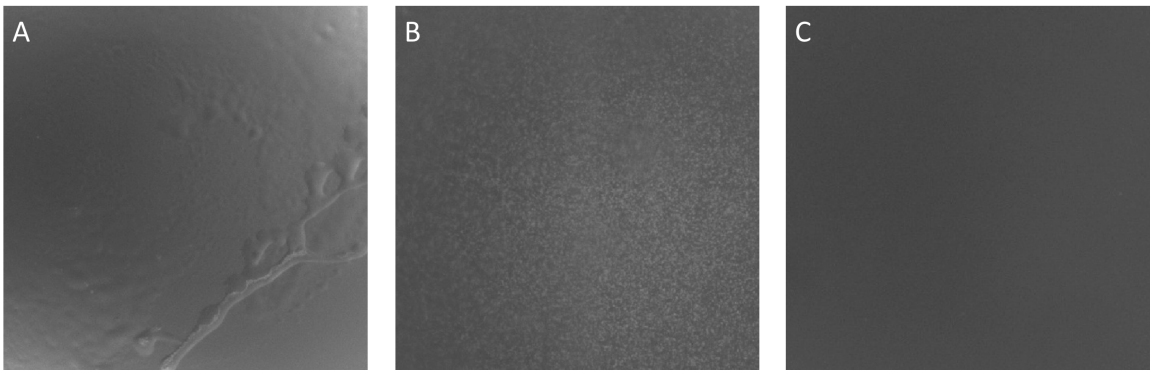


FIGURE 6.21: The results of ApoL1 crystallisation attempts. (A) Most drops resulted in phase separation, whilst others caused (B) precipitation of (C) no effect.

Adding a Fab fragment or scFv to the N-terminally truncated ApoL1 decreases the proportion of detergent in the complex, and it was hoped that it would increase the possibility of forming the protein:protein contacts required for crystallisation. As the antibody obtained from the hybridoma appeared to have a slower off-rate for the interaction with ApoL1 than the scFv and large scale expression was possible, this was focused on. To produce a quantity of protein suitable for crystallisation, the hybridoma was expanded to several litres. After purification of the antibody by tangential flow filtration and protein A/G affinity chromatography, the full length antibody was cleaved by overnight papain

digestion (Figure 6.22). The Fab fragments were then separated from the Fc regions and any remaining full length antibody using a protein A column.

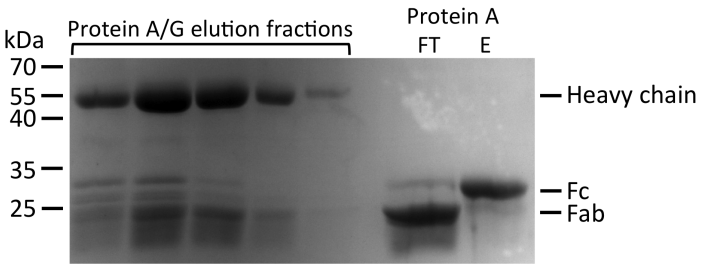


FIGURE 6.22: SDS-PAGE of mAb purification and cleavage. Five elution fractions containing mAb were eluted from the protein A/G beads. Following overnight papain cleavage, the Fc and Fab were separated by their ability to bind to a protein A column.

For later crystallisation experiments, a HEK 293F expression system for the Fab had been established by Dr. Sebastian Zoll. In these cases, the protein was purified by tangential flow filtration and Ni-NTA^{2+} affinity chromatography. Once the Fab was pure, it was mixed with N-terminally truncated ApoL1. The excess of one component was then separated by gel filtration to isolate the complex (Figure 6.23). Ideally excess Fab was used, as this resulted in better peak separation. However, this was not always possible due to limiting amounts of Fab. The complex was concentrated to 10 mg/ml. Plates were set up at both 4 °C and 18 °C using JCSG+ and MemGold 2, and at 4 °C only using PACT and Morpheus.

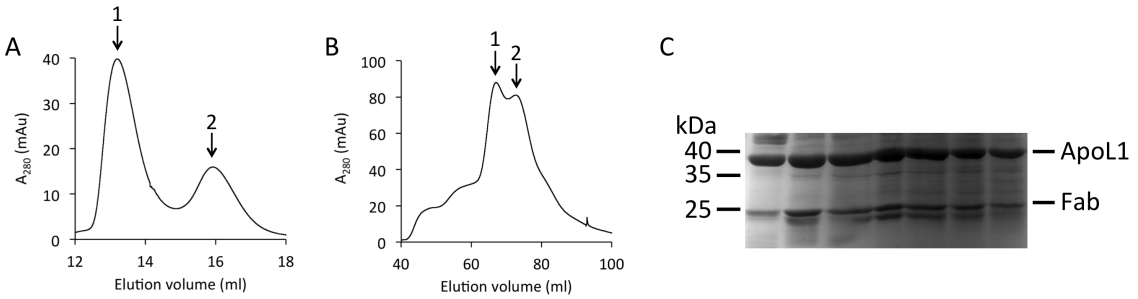


FIGURE 6.23: The formation of N-terminally truncated ApoL1:Fab complexes. Complexes were made using Fab produced using both a (A) hybridoma expression system and (B) a HEK expression system. In (A) the second peak contains excess Fab, whilst in (B) the second, overlapping peak contains N-terminally truncated ApoL1. In both cases, peak 1 contains the complex. (C) Complex formation was confirmed by SDS-PAGE of the gel filtration elution fractions of peak 1.

The most promising condition was well A9 of the broad screen PACT, containing 0.2 M LiCl, 0.1 M NaAc pH 5, 20 % PEG 6000. Under the microscope, very small spearhead-shaped crystals could be seen (Figure 6.24). Well A9 had similar crystals, with an identical well condition except the buffer, which was 0.1 M MES in this case. When trying to harvest these crystals it became apparent that they were only semi-solid, and could not be picked up using a cryoloop. Optimisation plates were set up using the well condition of A9 with the silver bullet additive screen and detergent screen. Unfortunately this has been unsuccessful to date.

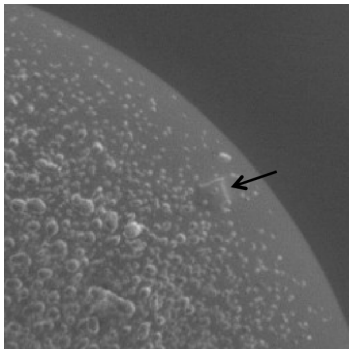


FIGURE 6.24: Microcrystals under 10 μm long grown from N-terminally truncated ApoL1:Fab grew in 0.2 M LiCl, 0.1 M NaAc pH 5, 20 % PEG 6000 amongst the phase separation. Prodding with a cryoloop demonstrated that these crystals were only semi-solid and could not be harvested.

6.5 Establishing a method to visualise ApoL1 inserted into nanodiscs

6.5.1 Production of nanodiscs

Crystallisation of ApoL1 had proved difficult, probably partly due to the large amount of detergent bound to each molecule of protein. Additionally, this technique would not allow visualisation of the structure of ApoL1 pores. Nanodiscs combined with cryo-electron microscopy were investigated as a method to image ApoL1 inserted into a lipid bilayer. Nanodiscs are formed from several copies of a membrane scaffold protein which circularise to surround a lipid bilayer. Once this structure has been formed, the protein of interest can be added at such a dilution factor that the concentration of the detergent is reduced to below the CMC, promoting the insertion of the protein into the bilayer of the nanodisc.

The protocol used for production of nanodiscs was adapted from a method provided by Dr. Tao Ni from the Gilbert lab (STRUBI), based on [156]. Briefly, lipids are mixed in chloroform before drying. They are resuspended in 25 mM Na cholate, 20 mM Hepes pH 7.5, 100 mM NaCl and the membrane scaffold protein, MSP1E3D1, is added. After incubation for several hours, the solution is dialysed to remove cholate. The nanodiscs are then gel filtered. In this protocol, the lipids used are 3 1,2-dioleoyl-sn-glycero-3-phosphocholine: 2 1,2-dioleoyl-sn-glycero-3-phospho-1'-rac-glycerol (3DOPC:2DOPG) and both the dialysis buffer and gel filtration buffer are 20 mM Hepes pH 7.5, 150 mM NaCl. As ApoL1 had been previously shown to insert into soybean polar lipid extract bilayers at pH 5 [119], nanodiscs were also produced using this lipid mixture (Figure 6.25).

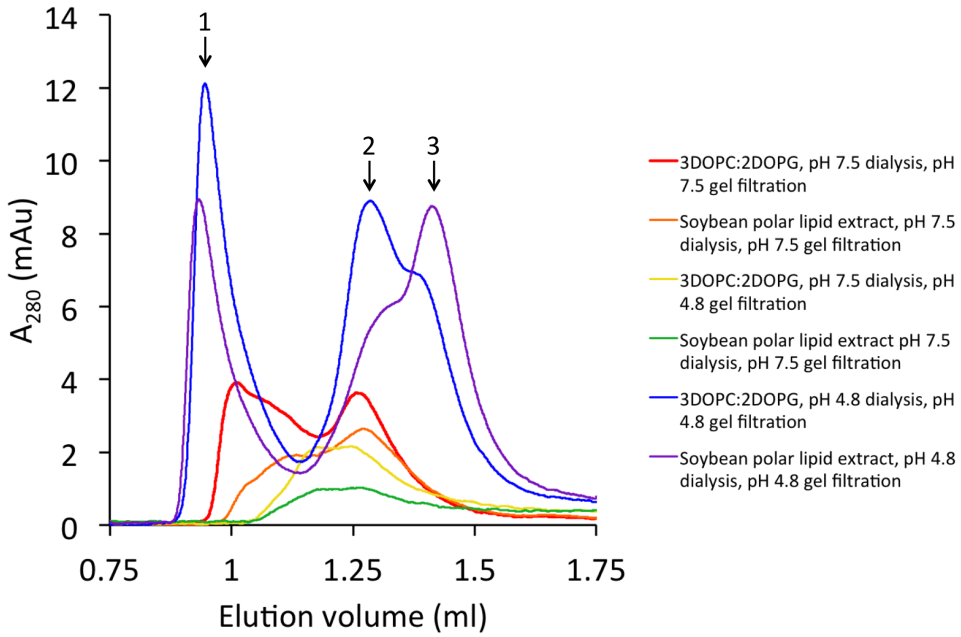


FIGURE 6.25: Gel filtration of nanodiscs prepared using different lipids, dialysis buffer pH and gel filtration running buffer pH. Size exclusion chromatography was completed using a Superdex 200 3.2/200. The buffers used for dialysis and gel filtration were 25 mM citrate, 150 mM NaCl pH 4.8 and 20 mM Hepes, 150 mM NaCl pH 7.5. The void volume is 0.9 ml (peak 1), whilst correctly formed nanodiscs elute at 1.3 ml (peak 2) as subsequently demonstrated by cryo-EM of samples from this elution volume (Figure 6.28). Peak 3 is likely to contain MSP that has not been incorporated into nanodiscs.

2DOPG:3DOPC gave a sharper peak than soybean polar lipid extract, no matter which pH was used. The peak where the nanodiscs are dialysed and gel filtered at low pH is sharper than where the nanodiscs were dialysed at neutral pH and gel filtered at low pH, although the size of the void peak has also increased. Therefore to insert ApoL1, which is thought to require pH 5 or below, both dialysis and gel filtration were carried out at pH 4.8.

Next, the ratio of MSP:lipid used was investigated. From the initial ratio of 1:2.8, the proportion of lipid was increased in an attempt to promote nanodisc formation (Figure 6.26).

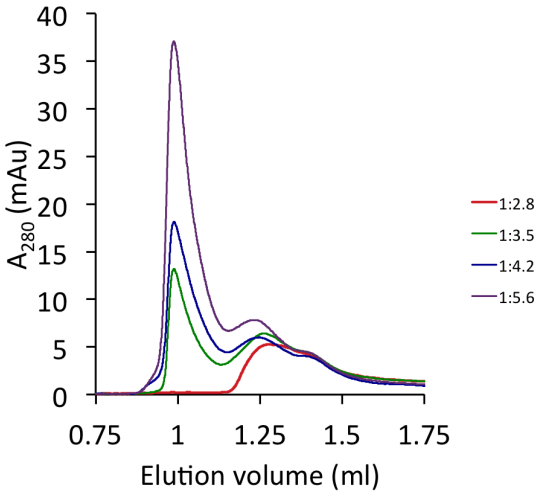


FIGURE 6.26: Gel filtration of nanodiscs prepared using different MSP:lipid ratios. The void volume peak increases in size with the amount of lipid used. Gel filtration was carried out using a Superdex 3.2/300 and a running buffer of 25 mM citrate, 150 mM NaCl, 0.02% DDM pH 4.8.

Instead, the size of the void peak in fact increased as the proportion of lipid was increased. This suggests that the void peak contains lipid micelles, which have been shown to absorb UV at 280 nm [211], rather than aggregated membrane scaffold protein. Therefore the original ratio of 1:2.8 was used in all future nanodisc purifications.

6.5.2 Insertion of ApoL1 into nanodiscs

After gel filtration, ApoL1 was incubated with the nanodiscs at 4 °C overnight. The dilution factor was 5-fold, reducing the DDM concentration to 0.004%, well below the CMC of 0.01%. The sample was then gel filtered and a shift in the peak was seen from 11.3 ml for empty nanodiscs, to 10.9 ml for those containing ApoL1 (Figure 6.27). SDS-PAGE confirmed the presence of both components. Each nanodisc should contain two molecules of MSP1E3D1, the membrane scaffold protein, each of which have a molecular weight of 32.6 kDa. As ApoL1 elutes as aggregates when gel filtered without detergent, the presence

of ApoL1 in this elution peak strongly suggests that the protein has successfully inserted into the nanodiscs.

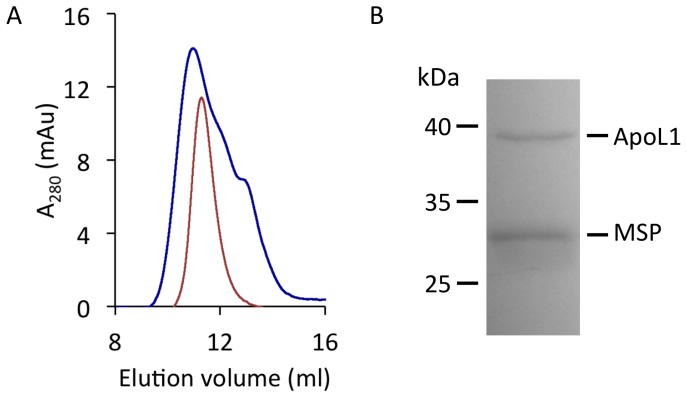


FIGURE 6.27: Gel filtration of nanodiscs. (A) A peak shift can be seen in the gel filtration of empty nanodiscs (red, 11.3 ml) and nanodiscs containing ApoL1 (blue, 10.9 ml). (C) The presence of ApoL1 was confirmed by SDS-PAGE.

6.5.3 Cryo-electron microscopy of nanodiscs

Once optimisation of nanodisc production was complete, preparation of samples and grids of a suitable quality for cryo-EM was focused on. This was carried out using empty nanodiscs. Several different nanodisc concentrations, grids and blotting protocols were tested. A concentration of 0.8 mg/ml gave the best results (Figure 6.28). Micrographs were imaged using an FEI Talos 200 kV cryo-TEM with FEI Ceta camera and STEM detector. Although ice and contaminating liposomes can be seen, the density of nanodiscs is suitable. The nanodiscs should be 9.7-12.9 nm in diameter according to the manufacturer of the membrane scaffold protein (Sigma).

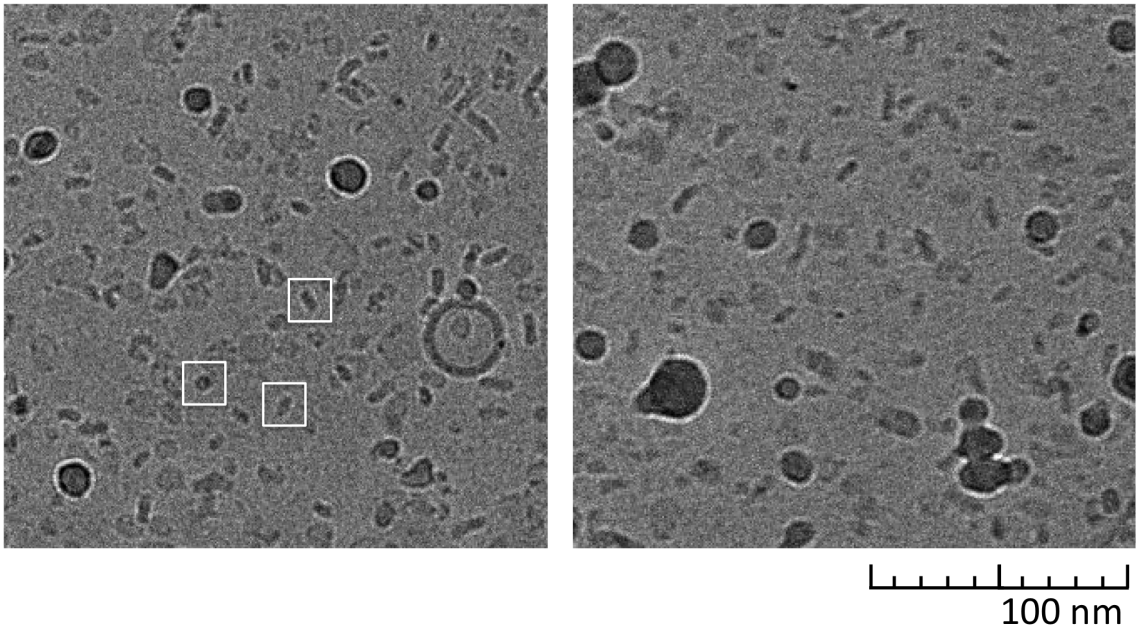


FIGURE 6.28: Two micrographs of empty nanodiscs. Both ice and contaminating liposomes can be seen in each. Three nanodiscs are indicated in the left image by the white boxes. The nanodiscs were at a concentration of 0.8 mg/ml. Data were collected using an FEI Talos 200 kV cryo-TEM with FEI Ceta camera and STEM detector.

Particle picking was then carried out using the EMAN2 software package [167], with a box size of 12.8 nm Approximately 2500 particles were picked from 40 micrographs (Figure 6.29). The particles vary in shape, from elongated ovals to circular shapes. This suggests that the particles are being imaged from different views, the ovals corresponding to side views and the circular shapes being topdown views.

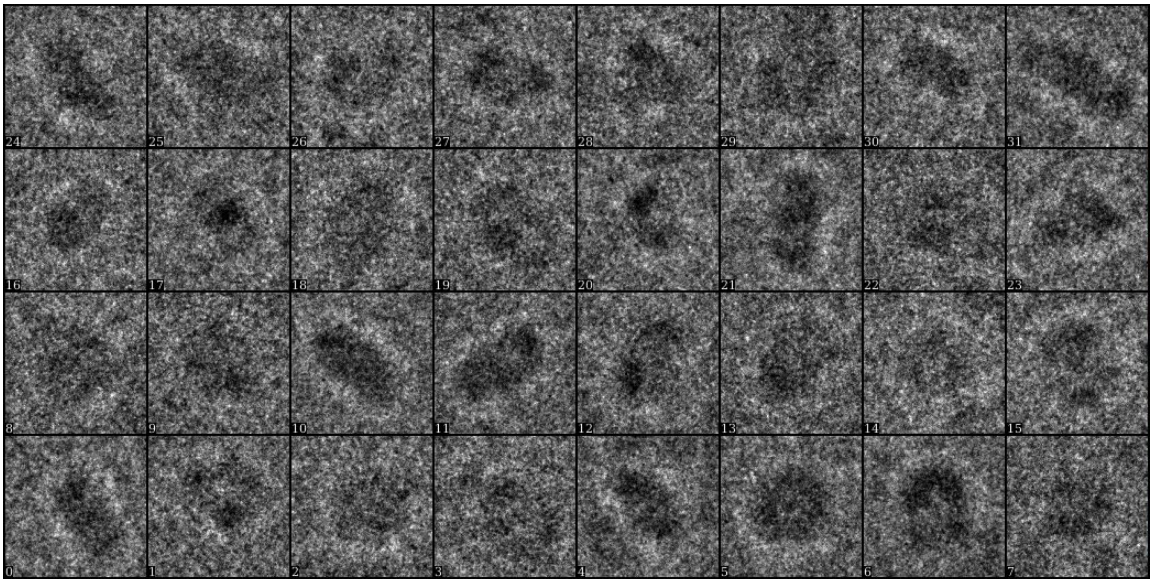


FIGURE 6.29: A selection of particles from micrographs imaging empty nanodiscs. Each box is 12.8 nm in wide. Nanodiscs are expected to be 9.7-12.9 nm in diameter.

Although the particles are the correct size, their diameter does vary significantly. This can be seen by comparing the size of the particles in the top left and top right of the figure. The homogeneity of nanodisc diameter has been improved greatly by the use of covalently circularised nanodiscs, which in turn increases the quality of the 3D reconstructions that can be produced [212]. Before continuing further with this project, circularised nanodiscs will therefore be produced to improve the homogeneity of the sample. However, the work here does demonstrate that ApoL1 can insert into nanodiscs, and that such nanodiscs can be imaged by cryo-electron microscopy.

6.6 Establishing an assay for the characterisation of ApoL1 pores

6.6.1 Experimental setup

To characterise the pores formed when Apolipoprotein L1 inserts into a lipid bilayer, electrophysiology experiments were carried out using artificial planar lipid bilayers in collaboration with Mercedes Romero-Ruiz from the Bayley lab. These experiments aimed to assess the affect of lipid composition, pH and ApoL1 deletions on the kinetics of the ApoL1 pore. A bilayer is formed across a small hole in a Teflon film that separates two compartments. In these experiments the *cis*-compartment contained 0.5 M KCl, 5 mM $CaCl_2$, 0.5 mM EDTA, 5 mM K-succinate, pH 5.3 whilst the *trans*-compartment contained 0.5 M KCl, 5 mM $CaCl_2$, 0.5 mM EDTA, 5 mM K-Hepes, pH 7.2, based on previous experiments by Thompson et al. [119]. In the setup used, a fixed voltage is applied and the transmembrane current is recorded. Using this technique many characteristics of the pores can be studied, such as the effect of different lipids, buffer pHs and whether truncations of the pore-forming protein are still able to form these functional pores. In this section, an assay that can be used to assess these characteristics is established and the preliminary analysis of the effect of lipid composition on pore characteristics is completed.

Up to 5 μl of ApoL1 was added to the *cis* chamber, which had a volume of 1 ml. Addition of the buffer in which ApoL1 was in, 20 mM citrate pH 4.8, 150 mM NaCl, 0.02% DDM caused no fluctuations in current at any applied potential. After addition of the protein, the solution was stirred with a magnetic bar and a potential applied. Often when an insertion event occurred, the channel either closed rapidly, or was followed by the insertion of several more ApoL1 molecules. Although data obtained from these ex-

periments could be used to assess the conductance of the pores, it was essential to obtain a single channel insertion in order to assess the pore kinetics. This was achieved on several occasions. In all cases, a sampling frequency of 10 kHz and a low pass filter with a frequency of 1 kHz was used.

Three different lipid compositions were used in the following experiments. The first, 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) was used as DPhPC bilayers are very well-characterised and have good stability (Figure 6.30).

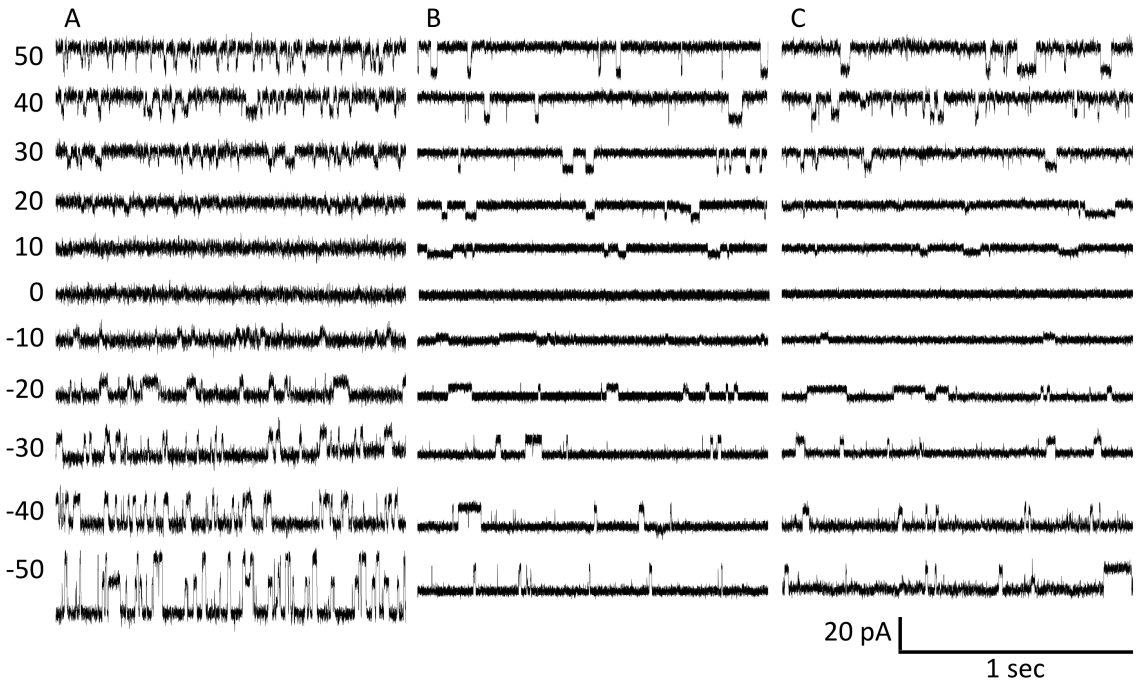


FIGURE 6.30: Representative single channel current traces from three independent experiments demonstrating the gating of a single ApoL1 pore in a DPhPC bilayer. All three look similar, although in experiment A a second ApoL1 molecule had inserted demonstrated by the third conductance level seen at -50 mV.

Soybean polar lipid extract was chosen as this had been used in previous electrophysiology studies of ApoL1 [119] (Figure 6.31).

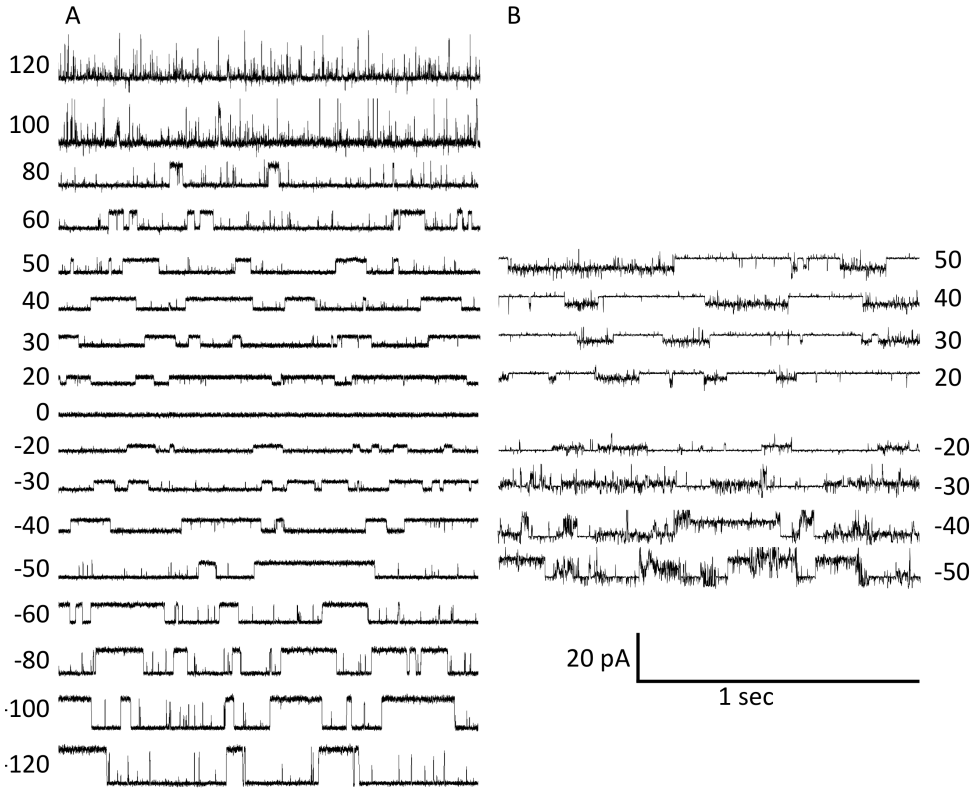


FIGURE 6.31: Representative single channel current traces from two independent experiments demonstrating the gating of a single ApoL1 channel in a soybean polar lipid extract bilayer. Both look similar, although the second experiment is much noisier and so cannot be used for further kinetic analysis.

Lastly, a mixture of 3DOPC:2DOPG was chosen as this was being used to form nanodiscs for this project. Therefore it was essential to show that the protein was active in bilayers formed from this mixture of lipids. Although insertions were seen in this lipid, not enough data was collected for kinetic analysis of single pores, so only the conductance of these pores was assessed.

6.6.2 Conductance of ApoL1 pores

The conductance of the pore describes the rate of the flow of ions through the pore. This can be calculated by dividing the observed current by the applied voltage. In order to assess

this, the difference in conductance before and after each event is calculated. A histogram is then produced displaying the frequency at which different conductance amplitudes occur (Figure 6.32). This was carried out for pores inserted into DPhPC, polar soybean lipids and 3DOPC:2DOPG. The major peak occurs at values of 0.15 nS, 0.13 nS and 0.12 nS respectively, indicating that this is the conductance of a single channel. Further peaks can be seen in each histogram, which relate to multiple insertions. The conductance of a single pore in all three lipid compositions is therefore very similar and also relates well to the value that was previously obtained for soybean polar lipids of 0.05-0.1 nS [119].

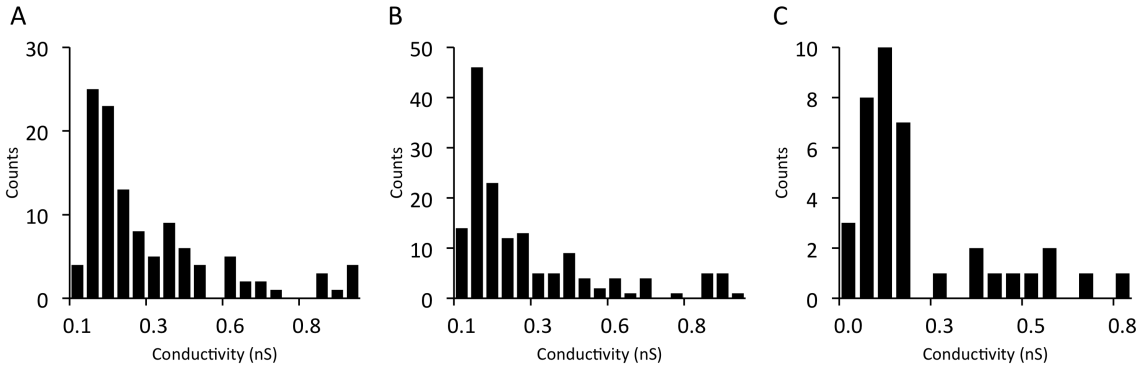


FIGURE 6.32: The conductance of ApoL1 pores in (A) DPhPC, (B) soybean polar lipid extract and (C) 3DOPC:2DOPG bilayers. The average conductance of a single pore in each lipid compositions was 0.15 $\pm$ 0.006 nS, 0.13 $\pm$ 0.004 nS and 0.12 $\pm$ 0.004 nS respectively. The histograms were produced by plotting conductance of insertion events against frequency.

6.6.3 Kinetic analysis of ApoL1 pores

Kinetic analysis can be completed to assess the frequency at which gating of the inserted pore occurs, and the average time the pore spends open (at level 1) and closed (at level 0) before another gating event occurs. Kinetic analysis of the gating of ApoL1 pores was completed for both DPhPC and soybean polar lipid extract. Firstly, traces showing the gating of these pores can be inspected. Three repeats of this experiment were carried out for DPhPC (Figure 6.30). All three appear to be very similar, with the pores open the vast

majority of the time at all applied potentials tested. In experiment A a second pore must be present, as a third conductance level can be seen at -50 mV. Therefore this experiment cannot be used for further kinetic analysis. Instead experiment B was chosen, as it had the most complete results.

Two repeats of this experiment were carried out for soybean polar lipid extract (Figure 6.31). The two repeats look very similar, with the pores open at negative applied potentials, with longer closures observed as the applied potential is increased above 0, until no openings can be seen in the trace at 120 mV. Experiment B is fairly noisy, perhaps due to other ApoL1 molecules attempting to insert or other background noise. Therefore experiment A was used for further kinetic analysis as the events could be more easily detected.

To investigate whether the observations made from the traces are correct, the average dwell times of the pore openings at each applied potential were calculated for a representative 20 second time period (Figure 6.33). In DPhPC, the dwell time for level 1 is much greater than that of level 0, and neither the dwell times nor the number of events are affected by changes in the applied potential. In contrast, both dwell times are low at negative applied potentials for the pore inserted into a soybean polar lipid extract bilayer, and as the applied potential is increased above 0 the number of events decreases and the dwell time for level 0 increases. This confirms the observations made from the traces.

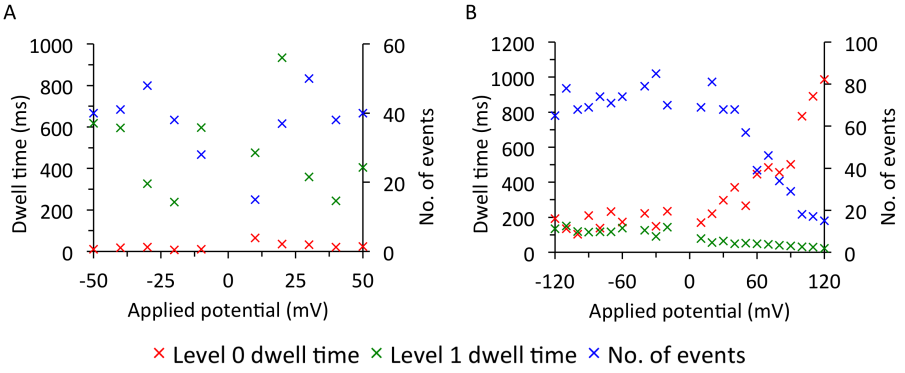


FIGURE 6.33: Kinetic analysis of the gating of ApoL1 pores inserted into (A) a DPhPC bilayer and (B) a soybean polar lipid extract bilayer. Analysis was conducted by identifying all gating events within a 20 second time period and the average dwell times were calculated for each conductance level. Level 0 refers to the conductance level of the closed pore, whilst level 1 refers to the conductance level of the open pore.

The data used in Figure 6.33A only went to 50 mV and -50 mV. Therefore experiment C from Figure 6.30 was also analysed. For this experiment, only 10 seconds of data at each applied potential voltage was available. However, the pattern is the same as that seen in Figure 6.33A, with a fairly constant number of events, approximately half the number as seen in Figure 6.33A due to the shorter timeframe, and a larger level 1 dwell time than level 0.

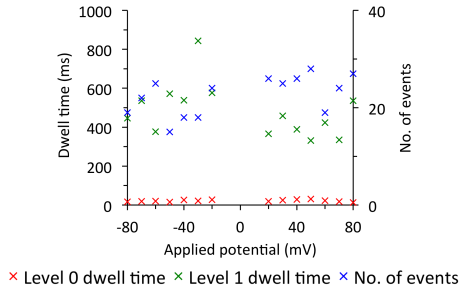


FIGURE 6.34: Kinetic analysis of the gating of ApoL1 pores inserted into a DPhPC bilayer. Analysis was conducted by identifying all gating events within a 10 second time period.

To visualise the proportion of time each pore spends at the two conductance levels, histograms can be plotted. These take into account the conductance frequency for the

entire 20 second timeframe being analysed, rather than looking at gating events alone, although they should of course be related. For DPhPC (Figure 6.35A), at both negative and positive applied potentials the peak sizes remain unchanged as the magnitude of the applied voltage is increased, with the channels remaining open for a large proportion of the time.

The result for soybean polar lipid extract is very different (Figure 6.35B). At low potentials, either positive or negative, the pore is partially open. As the amplitude of the positive applied potential is increased, the conductance remains at level 0 an increased proportion of the time, whilst the peak for level 1 ultimately becomes negligible. As the applied voltage becomes more negative, the size of the peaks at both level 0 and level 1 remain approximately unchanged. These results therefore confirm the conclusions of the dwell time analysis.

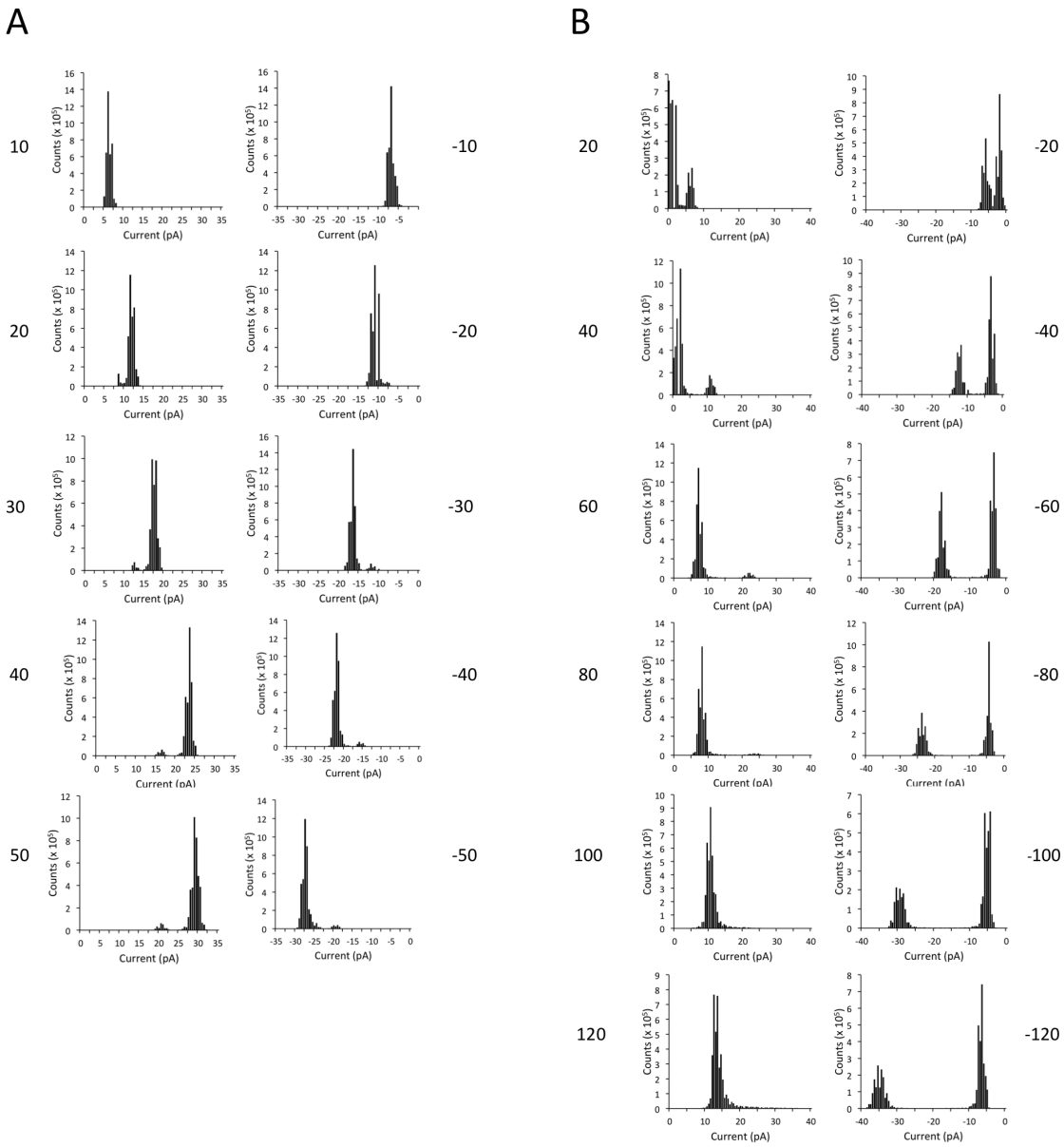


FIGURE 6.35: Histograms of the conductance frequency of an ApoL1 pore inserted into a (A) DPhPC bilayer and (B) polar soybean lipid bilayer. Level 0 is the peak nearest 0 pA, whilst the peak further from 0 pA represents level 1.

6.7 Discussion

Establishing an expression protocol for ApoL1 that provided enough pure protein for structural studies was a great challenge. The protein was known to be toxic when expressed in *E. coli* cells [117], and it was therefore unsurprising that expression of folded protein by insect or mammalian cells was either toxic or non-existent. Previously, recombinant expression of ApoL1 had been carried out using a cell-free system [117] which was unlikely to generate sufficient material for these experiments. An alternative method using insoluble expression was published in 2015 [119], which was likely to give more success. A new protocol using a codon-optimised gene was established, overcoming the need for BL21(DE3)-RIL cells. Overnight expression at 37 °C was used to encourage insoluble expression of the protein, minimising the toxicity of the functional protein. By optimising the refolding buffer used for flash dilution, the yield of refolded ApoL1 was increased dramatically.

To increase the chances of crystallisation and also to test the functions of different regions of ApoL1, truncations of the protein were made. The first of these was the N-terminal truncation, beginning the protein at residue 60. This truncation corresponds with that observed in ApoL1 extracted from human blood [149]. It is unclear whether this truncation is found *in vivo* or whether it occurred during extraction and purification of the protein. If the former, it will be important to compare the functionality of this construct versus the full length to establish the purpose of this truncation. If the latter, this truncation may be important in increasing the stability and reduce the flexibility of the protein so that it can be used for structural studies.

Additionally, constructs corresponding to the pore-forming domain, which is thought to have structural similarities to pore-forming colicins, and the rest of the protein, here

referred to as the C-terminal domain, were also produced. Again, these truncations could be useful both for assessing the roles of different regions of the protein, and for increasing the stability and decreasing the flexibility of the protein for use in structural studies.

As the proteins were produced insolubly and subsequently refolded, it was essential to confirm the they were correctly folded and functional. First indications that this was the case were gleaned from CD experiments, in which all constructs demonstrated α-helical secondary structure, corresponding with sequence-based predictions. Both SPR and MST experiments showed the ability of recombinant ApoL1 to interact with its binding partner, SRA. From the latter of these experiments an affinity of 328 nM was calculated. Perhaps the most compelling data to confirm the functionality of recombinant ApoL1 came from the killing assay. This demonstrated that recombinant ApoL1 was able to kill *T. brucei brucei* at low nM concentrations. Thus, recombinant ApoL1 is capable of its known function: causing trypanosome cell death.

To aid both functional and structural experiments, the production of a panel of antibodies against ApoL1 was desirable. However, after several attempts using both hybridoma and phage display methods, it became apparent that ApoL1 was difficult to target with antibodies. There could be several reasons for this. For example, as will be discussed in the next paragraph, ApoL1 must bind a large amount of detergent in order to stay soluble. This has two consequences. Firstly, it suggests that a large proportion of the surface area is covered by detergent, and thus hidden from possible binding partners. Secondly, it indicates that the protein may not remain correctly folded and soluble when this detergent is diluted, for example when injected into the bloodstream of a mouse. Although it is found in the human bloodstream, this is in a lipoprotein particle, where lipids are expected to play a role in hiding hydrophobic surfaces, allowing the protein to remain correctly folded. It is interesting to note that no colonies were observed on the plates corresponding to spleeno-

cytes of the mouse injected with PFD, and that no scFvs were identified which bound to the PFD. Some regions of the CTD must be exposed to allow interaction with SRA. However, surely some regions of the PFD must also be exposed to allow insertion into lipid bilayers, unless this is triggered by a rearrangement that begins when the CTD interacts with the bilayer. The results here are therefore inconclusive, but noteworthy.

Purification of ApoL1 was successfully completed adding 0.02% DDM to all buffers from the refolding step onwards. SEC-MALLS indicated that a large amount of detergent was bound to each ApoL1 molecule. This is in contrast to pore-forming colicins which do not require detergent to remain soluble, and are thought to undergo a conformational rearrangement in order to expose those regions which interact with the bilayer [213]. Protein-protein interactions form the majority of contacts in protein-detergent crystals. Therefore the greater the ratio of protein:detergent and the lower the molecular weight of the detergent micelles, the more likely that crystallisation will be successful [214]. Crystallisation of N-terminally truncated ApoL1 and PFD was attempted in a selection of detergents which included DDM but focused on those with shorter alkyl chains such as β-OG. Another attempt at increasing the protein:detergent ratio was carried out by adding a Fab fragment to the complex. This had a little more success, with semi-solid microcrystals forming in two conditions. However, optimisation is required, if possible, in order to obtain crystals which can be harvested.

To overcome the issues posed by attempting to study a membrane protein requiring detergent, other methods were investigated. Firstly, ApoL1 was inserted into nanodiscs, which consist of a ring of protein surrounding a lipid bilayer. Essentially, this allows ApoL1 to remain soluble in a buffer lacking detergent. Insertion of ApoL1 into these bilayers was successful, and initial imaging of empty nanodiscs was encouraging. However, the nanodiscs appeared heterogeneous in diameter. Covalently circularised nanodiscs will

be pursued in the future in an attempt to increase homogeneity of the sample [212]. Additionally, ApoL1 inserted into nanodiscs could also be useful for other experiments, such as SPR, to study interactions of the protein when embedded in a lipid bilayer.

Lastly, electrophysiology was used to study the characteristics of ApoL1 pores in lipid bilayers. Together, the data from these experiments show a clear difference in the characteristics of the ApoL1 pore when embedded in a bilayer formed of different lipids. In DPhPC the dwell time of pore opening remains fairly constant at all applied potentials tested, whereas in soybean polar lipid extract the higher the positive applied potential, the less time the pore is open for. This agrees with Thomson et al. [119], who noted that channel opening in soybean polar lipid extract was seen more frequently at an applied potential of -80 mV than +80 mV. The conductance of the pore is not affected by different lipid compositions. This is expected, as the diameter of the pore is likely to remain similar in all lipids. Although the lipid composition of different trypanosome membranes has not been characterised, the overall lipid composition of bloodstream *T. b. brucei* has been determined (Table 6.2) [215]. This demonstrates that the composition of polar soybean lipid extract is not dissimilar from the total lipid extract of the trypanosome. Therefore it is possible that this bilayer best reflects that which ApoL1 would experience in its native environment.

TABLE 6.2: Lipid composition of polar soybean lipid extract (Avanti Polar Lipids) and bloodstream form (BSF) *T. b. brucei* total cell extract [215].

Lipid	Soybean polar (%)	*T. b. brucei* BSF (%)
Phosphatidylcholine	45.7	45
Phosphatidylethanolamine	22.1	15
Sphingomyelin	0	10
Phosphatidylinositol	18.4	6
Phosphatidylserine	0	4
Other	13.8	20

Lipid composition of bilayers is known to have a big impact on toroidal pore formation, a pore structure where the lining of the pore is partially protein and partially lipid. In these circumstances the lipid bilayer must curve round between the two monolayers in order to form the interface between the lipid and the environment (Figure 6.36). Thus, lipids with a greater positive curvature promote pore formation, whilst lipids with a greater negative curvature inhibit pore formation [216, 217]. The pore-forming domain of ApoL1 is thought to be structurally similar to that of colicins and Bax [116], which are known to form toroidal pores [216, 217]. It is therefore thought that ApoL1 is likely to form pores with a similar structure, and it is thus unsurprising that lipid composition plays an important role in the functionality of the pore. However, more lipid compositions must be assessed before any strong conclusions can be drawn on the impact of lipid composition on ApoL1 pore functionality.

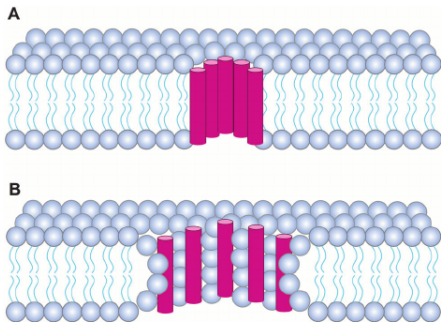


FIGURE 6.36: A cartoon demonstrating the structure of (A) a typical barrel stave pore and (B) a toroidal pore. Lipids play a direct role in the surface of the toroidal pore. This figure was taken from [218].

A second factor that must be considered is the potential of the membrane in which ApoL1 functions, and how this is likely to affect its action. There are currently three proposed membranes in which ApoL1 pores may function: the lysosomal membrane [117] which has a membrane potential of 19 mV [219], the plasma membrane [119] which has a membrane potential of -76 mV [220], and the mitochondrial membrane [120] which

has a membrane potential of -180 mV [221]. These membrane potential values are with respect to the outside of the lysosome, cell or mitochondrial matrix. To understand what membrane potential the ApoL1 pore will be experiencing, it is important to think about from which side the protein will have been inserted. For the lysosome, the protein will have inserted from inside the compartment, so it will experience a membrane potential of 19 mV. For the other two possible membranes, it is likely that the protein will not directly insert into them, but instead will the membrane it is already embedded in will fuse with the membrane. In the case of the mitochondria, the orientation of ApoL1 will remain the same as in the lysosome, so a membrane potential of -180 mV will be experienced. However, fusion with the plasma membrane will reverse the orientation, with the region of the pore that faced the inside of the lysosome now facing the extracellular environment. In this case the membrane potential will therefore be +76 mV. Assuming that the lipid composition of these membrane is similar to that of the soybean polar lipid extract, this would indicate that the channel is likely to be the most active in the mitochondria, and the least active at the cell membrane. It is therefore crucial to determine in which membrane this protein functions, and to confirm in which orientation the protein would be inserted into this membrane. Only at that stage could definite conclusions about the functionality of ApoL1 pores be drawn.

The establishment of this assay has been an important step forward, but further work is required to properly characterise the role that lipid composition has on activity of the pore. Additionally, these assays could be expanded to assess the role of pH, the activity of different truncations of ApoL1, and the effect of SRA inhibition on insertion and gating of the ApoL1 pore.

The work in this chapter has come a long way, establishing a successful protocol for the production and purification of ApoL1 and a number of truncations. Although crys-

tallisation has not been successful, a new method to visualise ApoL1 pores in nanodiscs using cryo-electron microscopy has been optimised. Lastly, functional characterisation of ApoL1 has begun, using an electrophysiology assay. In the future this could be used to investigate the role of pH, lipid and ApoL1 domains on pore functionality, giving insights into how ApoL1 forms the pores that results in death of the trypanosome cell.

Chapter 7

Conclusions and perspectives

7.1 The haptoglobin-haemoglobin receptor of African trypanosomes

HpHbR is a GPI-anchored, cell surface protein found across the African trypanosome species. The receptor is expressed by the bloodstream form of *T. brucei*, enabling the parasite to take up HpHb whilst in the mammalian bloodstream. However, the receptor is also the uptake route for TLF1, a human lipoprotein particle that plays an important role in the innate immunity of humans to non-human African trypanosomes such as *T. b. brucei*.

7.1.1 Uptake of nutrients and TLF1 by TbbHpHbR

This thesis presents the first structure of a trypanosome cell surface receptor in complex with its ligand. The structure of TbbHpHbRt:HpSPHb revealed an extended binding site along the membrane-distal half of the receptor. The upper binding site, which had been previously identified by mutational studies [94], contacts the β-subunit of haemoglobin, in-

cluding the haem ring itself, ensuring only haem-loaded complexes are taken up. The lower binding site contacts the SP domain of haptoglobin, predominantly through hydrophobic interactions, perhaps suggesting why this binding site is less conserved. This structure was confirmed by another lab group at around the same time [192].

The HprHb of TLF1, the human innate immunity lipoprotein particle, is responsible for uptake of the particle via HpHbR [43, 127]. The sequences of Hpr and Hp are very similar and the residues of Hp that contact TbbHpHbR are identical in Hpr, indicating that HprHb is likely to interact with HpHbR by forming the same interactions as HpHb. The complex structure therefore also reveals the binding mechanism for uptake of TLF1.

7.1.2 A comparison of HpHbR from *T. brucei* and *T. congolense*

The structure of TbbHpHbRt allowed a comparison of the receptors from human-infective *T. brucei* and the animal trypanosome *T. congolense*. This revealed a very similar architecture of an extended three-helix bundle with a compact head structure at the membrane-distal end. As had already been noted for TcHpHbR, this architecture is similar to that of VSG and GARP surface proteins [94]. It is likely that other trypanosome surface proteins share this same basic structure.

However, the structure revealed a significant difference between HpHbR of the two species too. A ~50° kink in TbbHpHbR one third of the way from the membrane-proximal end could be seen, which is not a feature of TcHpHbR. TbbHpHbR has to function in the context of the densely packed layer of VSG molecules on the surface of the bloodstream form trypanosome. One possible role of the kink may be to push apart the VSG molecules surrounding the receptor, resulting in better exposure of the binding site to the ligand. This is not a new concept: the same role has been proposed for the glycans of the transferrin

receptor [193].

A second function of the kink of TbbHpHbR is to allow dimeric binding of HpHb without requiring a significant tilt at the GPI-anchor. Uptake of dimeric HpHb by *T. brucei* was shown to saturate at much lower concentrations than uptake of monomeric HpHb. Thus, the kink of TbbHpHbR appears to have evolved to increase the efficiency of HpHb uptake in the context of the VSG layer of the bloodstream form trypanosome. The number of HprHb molecules in a TLF1 particle is unknown, and it is unclear whether Hpr forms dimers in the same way as Hp. However, original estimations for the dissociation constant of the high affinity receptor of trypanosomes for TLF1 were in the low nM region [127], suggesting that dimeric binding may also be important for uptake of TLF1.

7.1.3 Targeting HpHbR of *T. brucei* with antibody-drug conjugates

The available drugs to treat Human African Trypanosomiasis are not always effective, and often cause severe side effects. It is important that research into potential new drugs is persued. Antibody-drug conjugates targeting trypanosome surface receptors are appealing for several reasons. Firstly, the surface proteins of trypanosomes are known to undergo high rates of endocytosis, allowing rapid uptake of the ADCs [64, 73]. Secondly, the trypanosome surface proteins that have been identified so far do not have any homologues in humans, reducing the possible toxicity of this treatment. An additional advantage of these drugs is that they are already being developed for the treatment of cancers, ensuring that the efficacy and safety of the linkers and drugs have already been tested [199–201]. To adapt these ADCs to target trypanosomes, only the specificity of the conjugated antibody needs to be changed.

Using HpHbR to assess the possibility of ADCs as a treatment for HAT is practical as this is the most well-studied trypanosome cell surface receptor [43, 94, 192, 195]. Several antibodies have been generated against TbbHpHbR, all of which recognise an overlapping region. Interestingly, this region also contains the lower binding site for HpHb, as demonstrated by competition assays and structural determination of a Fab:receptor complex. This is surprising, as the antibodies were selected for binding any region on the surface of TbbHpHbR, not for uptake into trypanosomes. The advantage of this finding is that this region is known to be exposed in the context of the VSG layer, so uptake is likely to occur. Indeed, the results of both uptake and trypanosome killing assays indicate that this is the case.

This study was carried out as a proof-of-principle experiment to demonstrate that the killing of African trypanosomes with low concentrations of ADCs is possible. However, to persue this further, some adjustments must be made. Firstly, the antibodies in this study were selected for binding to TbbHpHbR from the non-human species, *T. b. brucei*. Although the ADCs still showed some efficacy against *T. b. rhodesisense*, they were much less effective against *T. b. gambiense*. *T. b. brucei* was used in this study due to the ease of handling and presence of a receptor null cell line. In subsequent studies, it will be important to focus on the efficacy of ADCs towards human-infective trypanosome subspecies. The HpHbR of *T. b. gambiense* has a lower affinity for HpHb and is also expressed at lower levels [94, 142]. It may therefore be difficult to select antibodies that can interact with this receptor with a great enough affinity to be effective as an ADC. Alternatively, more abundant surface proteins may have to be targeted, such as the transferrin receptor [60], in order to progress this work further. An additional advantage of the transferrin receptor is that transferrin is an essential growth factor for bloodstream form trypanosomes, so reduction of transferrin uptake may also effect trypanosome growth [95].

7.1.4 The ligand binding specificity of TcHpHbR

A second structure of a trypanosome receptor:ligand complex was determined during the work of this thesis. Surprisingly, this revealed the structure of TcHpHbR in complex with a haemoglobin tetramer. It had been assumed that, as shown for the *T. b. brucei* receptor [43], TcHpHbR would not bind Hb alone. However, subsequent studies demonstrated that the affinity of this receptor for free Hb is in fact higher than for HpHb. Immunofloures-cence assays confirmed that *T. congolense* can indeed take up Hb via HpHbR, but that this occurs during the epimastigote lifecycle stage, and the receptor cannot be detected in the bloodstream form of this parasite. HpHbR from a more distant trypanosome, *T. vivax*, was also able to bind Hb, indicating that the HpHbRs of these trypanosomes represent an ancestral haemoglobin receptor, which has subsequently evolved to become an HpHb receptor in *T. brucei*.

Together, the data allows an evolutionary scheme for HpHbR to be proposed. The receptor began as an epimastigote-specific protein whilst the trypanosomes were in the mouthparts of the tsetse fly. Here, the high affinity of the receptor for haemoglobin al-lowed binding and uptake of this ligand from a bloodmeal of the fly as it flowed over the trypanosome surface. At some stage, the location of epimastigote differentiation moved from the mouthparts to the salivary gland of the fly, perhaps to avoid toxic components of the bloodmeal or competition with other trypanosome species. As the source of haem had been lost, expression of the receptor by the bloodstream form began. To increase the affin-ity of HpHbR for the now more abundant ligand, HpHb, and to increase the avidity of the interaction by dimeric binding, the path of helix I altered. A consequence of this was the loss of haemoglobin binding. As the receptor was exposed to the adaptive immune system of the mammalian host, the expression level was reduced and localisation restricted to the

flagellar pocket. This study has led to a better understanding of how this protein, with an important role at the host-parasite interface, has been able to adapt to different functions in different trypanosome species.

7.2 Apolipoprotein L1

Apolipoprotein L1 is able to form pores in trypanosome membranes, resulting in cell death. The human-infective subspecies, *T. b. rhodesisense* and *T. b. gambiense*, have evolved mechanisms of evading ApoL1 activity. A greater understanding of how this protein works, and how these subspecies are able to overcome this, will go a long way towards understanding the complex interplay between the human trypanolytic factors and African trypanosomes.

7.2.1 Establishing a toolbox for the characterisation of ApoL1

A number of different expression protocols for ApoL1 have been used over the last two decades, varying from cell-free to mammalian expression systems [117, 121]. However, the most promising system to produce large amounts of protein used insoluble *E. coli* expression [119]. The protocol established in this thesis was adapted from this, and allows the expression, refolding and purification of miligram quantities of functional ApoL1. Expression and purification of different truncations of the protein has been carried out, allowing individual domains to be studied separately, as well as improving the chances of crystallisation.

An anti-ApoL1 antibody and a scFv have been obtained, which will be useful tools for structural studies and may also be beneficial for functional studies. The difficulty of anti-ApoL1 antibody selection may reflect the large amount of detergent required to keep

the protein in solution and the instability of the protein.

7.2.2 Characteristics of the ApoL1 pore

An electrophysiology assay has been established to study the characteristics of ApoL1 pores in lipid bilayers. Preliminary data suggests that the kinetics of pore opening is affected by the lipid composition of the bilayer. This could indicate an important role for lipid in the structure of the pore, as is seen for the toroidal pores formed by colicins [216]. The applied membrane potential also appears to effect the kinetics of pore opening. Before these results can be properly interpreted, it is crucial to establish in which membrane of the trypanosome ApoL1 functions, as this will determine the physiological voltage that the pore experiences. Additionally, more lipid compositions must be tested before any definite conclusions can be drawn. Once the optimum conditions have been determined, this assay can be used to assess the effect of ApoL1 and SRA truncations on the kinetics of the ApoL1 pore, leading to a better understanding of how these proteins function.

7.2.3 Steps towards ApoL1 structure determination

Crystallisation of ApoL1 has been attempted using a number of different truncations, detergents and crystallisation screens. Concentration of ApoL1 to values usually used for crystallisation was impossible due to precipitation. However, this problem was largely overcome when a complex was formed with an antibody Fab fragment. The second advantage of this is that it increases the proportion of protein surface that is exposed, rather than coated in detergent, increasing the chances of the formation of crystal contacts. Only initial attempts at the crystallisation of this complex have been carried out so far. Further attempts to crystallise ApoL1 should focus on this complex, or the complex formed with

SRA, which may have similar stabilising effects to the Fab fragment.

The use of nanodiscs to study membrane proteins has been developed over the last 15 years [156, 222]. Insertion of a membrane protein into nanodiscs allows single particle cryoEM to be carried out, treating each nanodisc as a particle [223]. Insertion of ApoL1 into nanodiscs was successful and the sample concentration required for cryoEM was optimisied. However, the diameter of the nanodiscs was not consistent, increasing the difficulty of single particle reconstruction. In the future, circularised nanodiscs should be used to increase the homogeneity of the sample and therefore the quality of the reconstruction obtainable [212].

Although determination of the ApoL1 structure and the pore it forms has been unattainable so far, progress has been made towards this goal and there are many more avenues to pursue.

Bibliography

1. Molyneux, D., Pentreath, V. & Doua, F. African trypanosomiasis in man. *Manson's tropical Diseases* **20,** 1171–1196 (1996).

2. Williams, B. African trypanosomiasis. *The Wellcome Trust illustrated history of tropical diseases.* 178–191 (1996).

3. Bruce, S. D. *Preliminary report on the tsetse fly disease or nagana, in Zululand* (Bennett & Davis, 1895).

4. Dutton, J. E. Preliminary note upon a trypanosome occurring in the blood of man (1902).

5. Steverding, D. The history of African trypanosomiasis. *Parasites & Vectors* **8,** 1–8 (2008).

6. Büscher, P., Cecchi, G., Jamonneau, V. & Priotto, G. Human African Trypanosomiasis. *The Lancet* **17,** 31510–6 (2017).

7. *Human African Trypanosomiasis* <http : / / www . who . int / trypanosomiasis_african/> (visited on 08/09/2017).

8. Auty, H., Torr, S. J., Michoel, T., Jayaraman, S. & Morrison, L. J. Cattle trypanosomosis : the diversity of trypanosomes and implications for disease epidemiology and control Trypanosome species of relevance to cattle. **34,** 587–598 (2015).

9. Urech, K., Neumayr, A. & Blum, J. Sleeping sickness in travelers-do they really sleep? *PLoS neglected tropical diseases* **5,** e1358 (2011).

10. Jamonneau, V. *et al.* Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. *PLoS neglected tropical diseases* **6,** e1691 (2012).

11. Odiit, M., Kansiime, F. & Enyaru, J. Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. *East African medical journal* **74,** 792–795 (1997).

12. Checchi, F., Filipe, J. A., Barrett, M. P. & Chandramohan, D. The natural progression of Gambiense sleeping sickness: what is the evidence? *PLoS neglected tropical diseases* **2,** e303 (2008).

13. Simarro, P. P. *et al.* Monitoring the progress towards the elimination of gambiense human African trypanosomiasis. *PLoS neglected tropical diseases* **9,** e0003785 (2015).

14. Courtin, F. *et al.* Reducing human-tsetse contact significantly enhances the efficacy of sleeping sickness active screening campaigns: a promising result in the context of elimination. *PLoS neglected tropical diseases* **9,** e0003727 (2015).

15. Wendo, C. Uganda revises cattle treatment to protect humans from sleeping sickness. *The Lancet* **359,** 239 (2002).

16. Magona, J. & Walubengo, J. Mass-treatment and insecticide-spraying of animal reservoirs for emergency control of Rhodesiense sleeping sickness in Uganda. *Journal of vector borne diseases* **48,** 105 (2011).

17. Kabayo, J. P. Aiming to eliminate tsetse from Africa. *Trends in parasitology* **18,** 473–475 (2002).

18. Magnus, E., Van Meirvenne, N., Vervoort, T., Le Ray, D., Wéry, M., *et al.* Use of freeze-dried trypanosomes in the indirect fluorescent antibody test for the serodiagnosis of sleeping sickness. *Ann Soc Belg Méd Trop* **58,** 103–109 (1978).

19. Jamonneau, V. *et al.* Accuracy of individual rapid tests for serodiagnosis of gambiense sleeping sickness in West Africa. *PLoS neglected tropical diseases* **9,** e0003480 (2015).

20. Mumba Ngoyi, D., Menten, J., Pyana, P. P., Büscher, P. & Lejon, V. Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. *Tropical Medicine & International Health* **18,** 778–782 (2013).

21. Eisler, M. C., Magona, J. W. & Revie, C. W. Diagnosis of cattle diseases endemic to sub-Saharan Africa: evaluating a low cost decision support tool in use by veterinary personnel. *PloS one* **7,** e40687 (2012).

22. Sudarshi, D. & Brown, M. Human African trypanosomiasis in non-endemic countries. *Clinical Medicine* **15,** 70–73 (2015).

23. Pohlig, G. *et al.* Efficacy and safety of pafuramidine versus pentamidine maleate for treatment of first stage sleeping sickness in a randomized, comparator-controlled, international phase 3 clinical trial. *PLoS neglected tropical diseases* **10,** e0004363 (2016).

24. Horn, D. High-throughput decoding of drug targets and drug resistance mechanisms in African trypanosomes. *Parasitology* **141,** 77–82 (2014).

25. Steverding, D. The development of drugs for treatment of sleeping sickness: a historical review. *Parasites & vectors* **3,** 15 (2010).

26. Balasegaram, M. *et al.* Effectiveness of melarsoprol and eflornithine as first-line regimens for gambiense sleeping sickness in nine Medecins Sans Frontieres programmes. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103,** 280–290 (2009).

27. Babokhov, P., Sanyaolu, A. O., Oyibo, W. A., Fagbenro-Beyioku, A. F. & Iriemenam, N. C. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. *Pathogens and global health* **107,** 242–252 (2013).

28. Priotto, G. *et al.* Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *The Lancet* **374,** 56–64 (2009).

29. Franco, J. R. *et al.* Monitoring the use of nifurtimox-eflornithine combination therapy (NECT) in the treatment of second stage gambiense human African trypanosomiasis. *Res Rep Trop Med* **3,** 93–101 (2012).

30. Hall, B. S., Bot, C. & Wilkinson, S. R. Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites. *Journal of Biological Chemistry* **286,** 13088–13095 (2011).

31. Barrett, M. P., Boykin, D. W., Brun, R. & Tidwell, R. R. Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. *British journal of pharmacology* **152,** 1155–1171 (2007).

32. Frearson, J. A. *et al.* N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature* **464,** 728 (2010).

33. Eisenthal, R. & Cornish-Bowden, A. Prospects for antiparasitic drugs the case of Trypanosoma brucei, the causative agent of African sleeping sickness. *Journal of Biological Chemistry* **273,** 5500–5505 (1998).

34. Smith, T. K., Crossman, A., Brimacombe, J. S. & Ferguson, M. A. Chemical validation of GPI biosynthesis as a drug target against African sleeping sickness. *The EMBO journal* **23,** 4701–4708 (2004).

35. Torreele, E. *et al.* Fexinidazole–a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness. *PLoS neglected tropical diseases* **4,** e923 (2010).

36. Jacobs, R. T. *et al.* SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. *PLoS neglected tropical diseases* **5,** e1151 (2011).

37. Sileghem, M., Darji, A., Hamers, R., Winkel, M. V. D. & Baetselier, P. D. Dual role of macrophages in the suppression of interleukin 2 production and interleukin 2 receptor expression in trypanosome-infected mice. *European journal of immunology* **19,** 829–835 (1989).

38. Corsini, A., Clayton, C., Askonas, B. A. & Ogilvie, B. M. Suppressor cells and loss of B-cell potential in mice infected with Trypanosoma brucei. *Clinical and experimental immunology* **29,** 122 (1977).

39. Black, S. J. & Mansfield, J. Prospects for vaccination against pathogenic African trypanosomes. *Parasite immunology* **38,** 735–743 (2016).

40. Lança, A. S. C. *et al.* Trypanosoma brucei: immunisation with plasmid DNA encoding invariant surface glycoprotein gene is able to induce partial protection in experimental African trypanosomiasis. *Experimental parasitology* **127,** 18–24 (2011).

41. Bastin, P., Sherwin, T. & Gull, K. Paraflagellar rod is vital for trypanosome motility. *Nature* **391,** 548 (1998).

42. Borst, P. & Fairlamb, A. Surface receptors and transporters of Trypanosoma brucei. *Annual Reviews in Microbiology* **52,** 745–778 (1998).

43. Vanhollebeke, B. *et al.* A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. *Science* **320,** 677–81 (2008).

44. Bonhivers, M., Nowacki, S., Landrein, N. & Robinson, D. R. Biogenesis of the trypanosome endo-exocytotic organelle is cytoskeleton mediated. *PLoS biology* **6,** e105 (2008).

45. Hager, K. M. *et al.* Endocytosis of a cytotoxic human high density lipoprotein results in disruption of acidic intracellular vesicles and subsequent killing of African trypanosomes. *The Journal of cell biology* **126,** 155–167 (1994).

46. Higgins, M. K., Lane-Serff, H., MacGregor, P. & Carrington, M. A Receptor's Tale: An Eon in the Life of a Trypanosome Receptor. *PLoS pathogens* **13,** e1006055 (2017).

47. Meyer, H. The Fine Structure of the Flagellum and Kinetoplast-Chondriome of Trypanosoma (Schizotrypanum) cruzi in Tissue Culture. *Journal of Eukaryotic Microbiology* **15,** 614–621 (1968).

48. Field, M. C. & Carrington, M. The trypanosome flagellar pocket. *Nature reviews. Microbiology* **7,** 775 (2009).

49. Michels, P. A., Bringaud, F., Herman, M. & Hannaert, V. Metabolic functions of glycosomes in trypanosomatids. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1763,** 1463–1477 (2006).

50. Matthews, K. R. 25 years of African trypanosome research: From description to molecular dissection and new drug discovery. *Molecular and biochemical parasitology* **200,** 30–40 (2015).

51. Acosta-Serrano, A. *et al.* The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proceedings of the National Academy of Sciences* **98,** 1513–1518 (2001).

52. Hehl, A., Pearson, T. W., Barry, J. D., Braun, R. & Roditi, I. Expression of GARP, a major surface glycoprotein of Trypanosoma congolense, on the surface of Trypanosoma brucei: characterization and use as a selectable marker. *Molecular and biochemical parasitology* **70,** 45–58 (1995).

53. Hoare, C. A. *et al. The trypanosomes of mammals. A Zoological Monograph.* (Blackwell Scientific Publications, 1972).

54. Rotureau, B. & Van Den Abbeele, J. Through the dark continent: African trypanosome development in the tsetse fly. *Frontiers in cellular and infection microbiology* **3** (2013).

55. Peacock, L. *et al.* Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. *Proceedings of the National Academy of Sciences* **108,** 3671–3676 (2011).

56. Peacock, L., Bailey, M., Carrington, M. & Gibson, W. Meiosis and haploid gametes in the pathogen Trypanosoma brucei. *Current Biology* **24,** 181–186 (2014).

57. Jefferies, D., Helfrich, M. & Molyneux, D. Cibarial infections of Trypanosoma vivax and T. congolense in Glossina. *Parasitology research* **73,** 289–292 (1987).

58. Peacock, L., Cook, S., Ferris, V., Bailey, M. & Gibson, W. The life cycle of Trypanosoma (Nannomonas) congolense in the tsetse fly. *Parasites & vectors* **5,** 109 (2012).

59. Aksoy, S., Gibson, W. C. & Lehane, M. J. Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. *Advances in parasitology* **53,** 1–83 (2003).

60. Steverding, D. The transferrin receptor of Trypanosoma brucei. *Parasitology international* **48,** 191–198 (2000).

61. Rodrigues J.C.F. Godinho J.L.P., d. S. W. *Biology of Human Pathogenic Trypanosomatids: Epidemiology, Lifecycle and Ultrastructure* (Springer, 2014).

62. Webster, P. Endocytosis by African trypanosomes. II. Occurrence in different life-cycle stages and intracellular sorting. *European journal of cell biology* **49,** 303–10 (1989).

63. Arun, P., JEFFRIES, T. R., Mark, C., *et al.* Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in Trypanosoma brucei. *Biochemical Journal* **374,** 443–451 (2003).

64. Engstler, M. *et al.* Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei. *J Cell Sci* **117,** 1105–1115 (2004).

65. Grünfelder, C. G. *et al.* Accumulation of a GPI-Anchored Protein at the Cell Surface Requires Sorting at Multiple Intracellular Levels. *Traffic* **3,** 547–559 (2002).

66. Ferrante, A. & Allison, A. Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite immunology* **5,** 491–498 (1983).

67. Hovel-Miner, G., Mugnier, M., Papavasiliou, F. N., Pinger, J. & Schulz, D. in *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations* 23–46 (Springer, 2015).

68. Freymann, D. *et al.* 2.9 Å resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. *Journal of molecular biology* **216,** 141–160 (1990).

69. Schwede, A., Macleod, O. J., MacGregor, P. & Carrington, M. How does the VSG coat of bloodstream form African trypanosomes interact with external proteins? *PLoS pathogens* **11,** e1005259 (2015).

70. Chattopadhyay, A. *et al.* Structure of the C-terminal domain from Trypanosoma brucei variant surface glycoprotein MITat1. 2. *Journal of Biological Chemistry* **280,** 7228–7235 (2005).

71. Barry, J. D. Capping of variable antigen on Trypanosoma brucei, and its immunological and biological significance. *Journal of cell science* **37,** 287–302 (1979).

72. O'Beirne, C., Lowry, C. M. & Voorheis, H. P. Both IgM and IgG anti-VSG antibodies initiate a cycle of aggregation–disaggregation of bloodstream forms of Trypanosoma brucei without damage to the parasite. *Molecular and biochemical parasitology* **91,** 165–193 (1998).

73. Engstler, M. *et al.* Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131,** 505–515 (2007).

74. Dempsey, W. L. & Mansfield, J. Lymphocyte function in experimental African trypanosomiasis. V. Role of antibody and the mononuclear phagocyte system in variant-specific immunity. *The Journal of Immunology* **130,** 405–411 (1983).

75. Schleifer, K., Filutowicz, H., Schopf, L. & Mansfield, J. Characterization of T helper cell responses to the trypanosome variant surface glycoprotein. *The Journal of Immunology* **150,** 2910–2919 (1993).

76. Hertz-Fowler, C. *et al.* Telomeric expression sites are highly conserved in Trypanosoma brucei. *PloS one* **3,** e3527 (2008).

77. Kooter, J. M. & Borst, P. α-Amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic acids research* **12,** 9457–9472 (1984).

78. Zomerdijk, J. C., Kieft, R., Shiels, P. G. & Borst, P. Alpha-amanitin-resistant transcription units in trypanosomes: a comparison of promoter sequences for a VSG gene expression site and for the ribosomal RNA genes. *Nucleic Acids Research* **19,** 5153–5158 (1991).

79. Günzl, A. *et al.* RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryotic Cell* **2,** 542–551 (2003).

80. Van der Ploeg, L. *et al.* An analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic acids research* **10,** 5905–5923 (1982).

81. Barry, J. *et al.* What the genome sequence is revealing about trypanosome antigenic variation. *Biochemical Society Transaction* **33,** 986–989 (2005).

82. Marcello, L. & Barry, J. D. Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome research* **17,** 1344–1352 (2007).

83. Taylor, J. E. & Rudenko, G. Switching trypanosome coats: what's in the wardrobe? *TRENDS in Genetics* **22,** 614–620 (2006).

84. Mugnier, M. R., Cross, G. A. & Papavasiliou, F. N. The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science* **347,** 1470–1473 (2015).

85. Navarro, M., Peñate, X. & Landeira, D. Nuclear architecture underlying gene expression in Trypanosoma brucei. *Trends in microbiology* **15,** 263–270 (2007).

86. Horn, D. Antigenic variation in African trypanosomes. *Molecular and biochemical parasitology* **195,** 123–129 (2014).

87. Wang, Q.-P., Kawahara, T. & Horn, D. Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. *Molecular microbiology* **77,** 1237–1245 (2010).

88. Alsford, S. & Horn, D. Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. *Nucleic acids research* **40,** 10150–10160 (2012).

89. Povelones, M. L., Gluenz, E., Dembek, M., Gull, K. & Rudenko, G. Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in Trypanosoma brucei. *PLoS pathogens* **8,** e1003010 (2012).

90. McCulloch, R., Rudenko, G. & Borst, P. Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Molecular and cellular biology* **17,** 833–843 (1997).

91. Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N. & Steinert, M. Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. *Nature* **316,** 562–564 (1985).

92. Thon, G., Baltz, T. & Eisen, H. Antigenic diversity by the recombination of pseudogenes. *Genes & development* **3,** 1247–1254 (1989).

93. Thon, G., Baltz, T., Giroud, C. & Eisen, H. Trypanosome variable surface glycoproteins: composite genes and order of expression. *Genes & development* **4,** 1374–1383 (1990).

94. Higgins, M. K. *et al.* Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 1905–10 (2013).

95. Schell, D., Borowy, N. K. & Overath, P. Transferrin is a growth factor for the bloodstream form ofTrypanosoma brucei. *Parasitology research* **77,** 558–560 (1991).

96. Steverding, D. *et al.* ESAG 6 and 7 products of Trypanosoma brucei form a transferrin binding protein complex. *European journal of cell biology* **64,** 78–87 (1994).

97. Steverding, D., Stierhof, Y.-D., Fuchs, H., Tauber, R. & Overath, P. Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei. *The Journal of Cell Biology* **131,** 1173–1182 (1995).

98. Bitter, W., Gerrits, H., Kieft, R. & Borst, P. The role of transferrin-receptor variation in the host range of Trypanosoma brucei. *Nature* **391,** 499 (1998).

99. Coppens, I., Baudhuin, P., Opperdoes, F. & Courtoy, P. J. Receptors for the host low density lipoproteins on the hemoflagellate Trypanosoma brucei: purification and

involvement in the growth of the parasite. *Proceedings of the National Academy of Sciences* **85,** 6753–6757 (1988).

100. Gillett, M. P. & Owen, J. S. Characteristics of the binding of human and bovine high-density lipoproteins by bloodstream forms of the African trypanosome, Trypanosoma brucei brucei. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1123,** 239–248 (1992).

101. Bringaud, F. & Baltz, T. A potential hexose transporter gene expressed predominantly in the bloodstream form of Trypanosoma brucei. *Molecular and biochemical parasitology* **52,** 111–121 (1992).

102. Bringaud, F. & Baltz, T. Differential regulation of two distinct families of glucose transporter genes in Trypanosoma brucei. *Molecular and cellular biology* **13,** 1146–1154 (1993).

103. Cazzulo, J.-J. Aerobic fermentation of glucose by trypanosomatids. *The FASEB Journal* **6,** 3153–3161 (1992).

104. Gould, G. W. & Holman, G. D. The glucose transporter family: structure, function and tissue-specific expression. *Biochemical Journal* **295,** 329 (1993).

105. SEYFANG, A. & DUSZENKO, M. Specificity of glucose transport in Trypanosoma brucei. *The FEBS Journal* **202,** 191–196 (1991).

106. Rifkin, M. Trypanosoma brucei: biochemical and morphological studies of cytotoxicity caused by normal human serum. *Experimental parasitology* **58,** 81–93 (1984).

107. Molina-Portela, M. d. P., Lugli, E. B., Recio-Pinto, E. & Raper, J. Trypanosome lytic factor, a subclass of high-density lipoprotein, forms cation-selective pores in membranes. *Molecular and biochemical parasitology* **144,** 218–26 (2005).

108. Rifkin, M. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proceedings of the National Academy of Sciences* **75,** 3450–3454 (1978).

109. Hajduk, S. *et al.* Lysis of Trypanosoma brucei by a toxic subspecies of human high density lipoprotein. *Journal of Biological Chemistry* **264,** 5210–5217 (1989).

110. Raper, J., Nussenzweig, V. & Tomlinson, S. The main lytic factor of Trypanosoma brucei brucei in normal human serum is not high density lipoprotein. *Journal of Experimental Medicine* **183,** 1023–1029 (1996).

111. Raper, J., Fung, R., Ghiso, J., Nussenzweig, V. & Tomlinson, S. Characterization of a novel trypanosome lytic factor from human serum. *Infection and immunity* **67,** 1910–6 (1999).

112. Widener, J., Nielsen, M. J., Shiflett, A., Moestrup, S. K. & Hajduk, S. Hemoglobin is a co-factor of human trypanosome lytic factor. *PLoS pathogens* **3,** e129 (2007).

113. Pays, E., Vanhollebeke, B., Uzureau, P., Lecordier, L. & Pérez-Morga, D. The molecular arms race between African trypanosomes and humans. *Nature reviews. Microbiology* **12,** 575–84 (2014).

114. Rifkin, M. Trypanosoma brucei: some properties of the cytotoxic reaction induced by normal human serum. *Experimental parasitology* **46,** 189–206 (1978).

115. Lecordier, L. *et al.* Identification of Trypanosoma brucei components involved in trypanolysis by normal human serum. *Molecular microbiology* **94,** 625–636 (2014).

116. Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D. P., *et al.* Apolipoprotein LI is the trypanosome lytic factor of human serum. *Nature* **422,** 83 (2003).

117. Perez-Morga, D. *et al.* Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* **309,** 469–72 (2005).

118. Lecordier, L. *et al.* C-terminal mutants of apolipoprotein LI efficiently kill both Trypanosoma brucei brucei and Trypanosoma brucei rhodesiense. *PLoS pathogens* **5,** e1000685 (2009).

119. Thomson, R. & Finkelstein, A. Human trypanolytic factor APOL1 forms pH-gated cation-selective channels in planar lipid bilayers: relevance to trypanosome lysis. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 2894–9 (2015).

120. Vanwalleghem, G. *et al.* Coupling of lysosomal and mitochondrial membrane permeabilization in trypanolysis by APOL1. *Nature communications* **6** (2015).

121. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329,** 841–845 (2010).

122. Taylor, H. E., Khatua, A. K. & Popik, W. The innate immune factor apolipoprotein L1 restricts HIV-1 infection. *Journal of virology* **88,** 592–603 (2014).

123. Vanhollebeke, B. & Pays, E. The function of apolipoproteins L. *Cellular and molecular life sciences* **63,** 1937–1944 (2006).

124. Schendel, S. L. *et al.* Ion channel activity of the BH3 only Bcl-2 family member, BID. *Journal of Biological Chemistry* **274,** 21932–21936 (1999).

125. Schendel, S. L. *et al.* Channel formation by antiapoptotic protein Bcl-2. *Proceedings of the National Academy of Sciences* **94,** 5113–5118 (1997).

126. Minn, A. J. *et al.* Bcl-xL forms an ion channel in synthetic lipid membranes. *Nature* **385,** 353 (1997).

127. Drain, J., Bishop, J. R. & Hajduk, S. L. Haptoglobin-related protein mediates trypanosome lytic factor binding to trypanosomes. *Journal of Biological Chemistry* **276,** 30254–30260 (2001).

128. Loveless, B. C. *et al.* Structural Characterization and Epitope Mapping of the Glutamic Acid/Alanine-rich Protein from Trypanosoma congolense: defining assembly on the parasite cell surface. *Journal of Biological Chemistry* **286,** 20658–20665 (2011).

129. Heinderyckx, M., Jacobs, P. & Bollen, A. Secretion of glycosylated human recombinant haptoglobin in Baculovirus-infected insect cells. *Molecular biology reports* **13,** 225–232 (1988).

130. Hwang, P. K. & Greer, J. Interaction between hemoglobin subunits in the hemoglobin. haptoglobin complex. *Journal of Biological Chemistry* **255,** 3038–3041 (1980).

131. Sadrzadeh, S., Graf, E., Panter, S. S., Hallaway, P. & Eaton, J. Hemoglobin. A biologic fenton reagent. *Journal of Biological Chemistry* **259,** 14354–14356 (1984).

132. Kristiansen, M., Graversen, J. H., Jacobsen, C., Sonne, O., *et al.* Identification of the haemoglobin scavenger receptor. *Nature* **409,** 198 (2001).

133. Andersen, C. B. F. *et al.* Structure of the haptoglobin-haemoglobin complex. *Nature* **489,** 456–9 (2012).

134. Maeda, N., Yang, F., Barnett, D. R., Bowman, B. H. & Smithies, O. Duplication within the haptoglobin Hp2 gene. *Nature* **309,** 131–135 (1984).

135. Muranjan, M., Nussenzweig, V. & Tomlinson, S. Characterization of the human serum trypanosome toxin, haptoglobin-related protein. *Journal of Biological Chemistry* **273,** 3884–3887 (1998).

136. De Greef, C., Imberechts, H., Matthyssens, G., Van Meirvenne, N. & Hamers, R. A gene expressed only in serum-resistant variants of Trypanosoma brucei rhodesiense. *Molecular and Biochemical Parasitology* **36,** 169–176 (1989).

137. Milner, J. D. & Hajduk, S. L. Expression and localization of serum resistance associated protein in Trypanosoma brucei rhodesiense. *Molecular and biochemical parasitology* **104,** 271–283 (1999).

138. Van Xong, H. *et al.* A VSG expression site–associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell* **95,** 839–846 (1998).

139. Campillo, N. & Carrington, M. The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from Trypanosoma brucei rhodesiense. *Molecular and biochemical parasitology* **127,** 79–84 (2003).

140. Thomson, R., Molina-Portela, P., Mott, H., Carrington, M. & Raper, J. Hydrodynamic gene delivery of baboon trypanosome lytic factor eliminates both animal and human-infective African trypanosomes. *Proceedings of the National Academy of Sciences* **106,** 19509–19514 (2009).

141. Symula, R. E. *et al.* Trypanosoma brucei gambiense group 1 is distinguished by a unique amino acid substitution in the HpHb receptor implicated in human serum resistance. *PLoS neglected tropical diseases* **6,** e1728 (2012).

142. Kieft, R. *et al.* Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. *Proceedings of the National Academy of Sciences* **107,** 16137–16141 (2010).

143. Capewell, P. *et al.* Differences between Trypanosoma brucei gambiense groups 1 and 2 in their resistance to killing by trypanolytic factor 1. *PLoS neglected tropical diseases* **5,** e1287 (2011).

144. Berberof, M., Pérez-Morga, D. & Pays, E. A receptor-like flagellar pocket glycoprotein specific to Trypanosoma brucei gambiense. *Molecular and Biochemical Parasitology* **113,** 127–138 (2001).

145. Uzureau, P. *et al.* Mechanism of Trypanosoma brucei gambiense resistance to human serum. *Nature* **501,** 430–4 (2013).

146. Capewell, P. *et al.* The TgsGP gene is essential for resistance to human serum in Trypanosoma brucei gambiense. *PLoS pathogens* **9,** e1003686 (2013).

147. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic acids research* **43,** W389–W394 (2015).

148. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics (Oxford, England)* **21,** 3369–76 (2005).

149. Duchateau, P. N. *et al.* Apolipoprotein L, a new human high density lipoprotein apolipoprotein expressed by the pancreas Identification, cloning, characterization, and plasma distribution of apolipoprotein L. *Journal of Biological Chemistry* **272,** 25576–25582 (1997).

150. Schrödinger, L. *The PyMOL molecular graphics system, version 1.3 r1* (2010).

151. Ripoche, H., Laine, E., Ceres, N. & Carbone, A. JET2 Viewer: a database of predicted multiple, possibly overlapping, protein–protein interaction sites for PDB structures. *Nucleic acids research* **45,** D236–D242 (2017).

152. Li, M. Z. & Elledge, S. J. SLIC: a method for sequence-and ligation-independent cloning. *Gene Synthesis: Methods and Protocols* **852,** 51–59 (2012).

153. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* **229,** 105–124 (1993).

154. Vaughan, T. J. *et al.* Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature biotechnology* **14,** 309–314 (1996).

155. Lloyd, C. *et al.* Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. *Protein Engineering, Design & Selection* **22,** 159–168 (2008).

156. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Letters* **2,** 853–856 (2002).

157. Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46,** 73–82 (2008).

158. 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).

159. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* **50** (2017).

160. Guiner, A., Fournet, G. & Walker, C. *Small angle scattering of X-rays* (J. Wiley & Sons, 1955).

161. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography* **25,** 495–503 (1992).

162. Porod, G. in *Small angle X-ray scattering* (Academic Press, 1982).

163. 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).

164. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography* **36,** 860–864 (2003).

165. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal* **76,** 2879–2886 (1999).

166. 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).

167. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157,** 38–46 (2007).

168. BA, K. *Planar lipid bilayer electrophysiology* <https://mcb.berkeley.edu/labs/krantz/equipment/blm.html>.

169. Incardona, M.-F. *et al.* EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation* **16,** 872–879 (2009).

170. Leslie, A. G. & Powell, H. R. Processing diffraction data with MOSFLM. *Evolving methods for macromolecular crystallography* **245,** 41–51 (2007).

171. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* **67,** 271–81 (2011).

172. Kabsch, W. XDS. en. *Acta crystallographica. Section D, Biological crystallography* **66,** 125–32 (2010).

173. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43,** 186–190 (2010).

174. Waterman, D. G. *et al.* Diffraction-geometry refinement in the DIALS framework. *Acta Crystallographica Section D: Structural Biology* **72,** 558–575 (2016).

175. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* **62,** 72–82 (2006).

176. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69,** 1204–1214 (2013).

177. French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **34,** 517–525 (1978).

178. 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).

179. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).

180. Assmann, G., Brehm, W. & Diederichs, K. Identification of rogue datasets in serial crystallography. *Journal of applied crystallography* **49,** 1021–1028 (2016).

181. Müller, P. Practical suggestions for better crystal structures. *Crystallography Reviews* **15,** 57–83 (2009).

182. McCoy, A. J. *et al.* Phaser crystallographic software. en. *Journal of applied crystallography* **40,** 658–674 (2007).

183. 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).

184. Bricogne, G. *et al. BUSTER version 2.10.0.* Global Phasing Ltd, Cambridge (2011).

185. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* **68,** 352–367 (2012).

186. Rupp, B. *Biomolecular crystallography: principles, practice, and application to structural biology* (Garland Science, 2009).

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

188. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. en. *Acta crystallographica. Section D, Biological crystallography* **62,** 1002–11 (2006).

189. 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).

190. Asleh, R., Guetta, J., Kalet-Litman, S., Miller-Lotan, R. & Levy, A. P. Haptoglobin genotype- and diabetes-dependent differences in iron-mediated oxidative stress in vitro and in vivo. *Circulation Research* **96,** 435–441 (2005).

191. Dodson, E. The befores and afters of molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **64,** 17–24 (2008).

192. Stødkilde, K., Torvund-Jensen, M., Moestrup, S. K. & Andersen, C. B. Structural basis for trypanosomal haem acquisition and susceptibility to the host innate immune system. *Nature communications* **5,** 5487 (2014).

193. Mehlert, A., Wormald, M. R. & Ferguson, M. A. Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of Trypanosoma brucei. *PLoS pathogens* **8,** e1002618 (2012).

194. Buelow, R., Overath, P. & Davoust, J. Rapid lateral diffusion of the variant surface glycoprotein in the coat of Trypanosoma brucei. *Biochemistry* **27,** 2384–2388 (1988).

195. Lane-Serff, H., MacGregor, P., Lowe, E. D., Carrington, M. & Higgins, M. K. Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. en. *eLife* **3,** e05553 (2014).

196. Wright, K. E. *et al.* Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature* **515,** 427 (2014).

197. Lewis, S. M. *et al.* Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface. *Nature biotechnology* **32,** 191–198 (2014).

198. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66,** 12–21 (2010).

199. Sutherland, M. S. K. *et al.* SGN-CD33A: a novel CD33-targeting antibody–drug conjugate using a pyrrolobenzodiazepine dimer is active in models of drug-resistant AML. *Blood* **122,** 1455–1463 (2013).

200. Flynn, M. J. *et al.* ADCT-301, a Pyrrolobenzodiazepine (PBD) Dimer–Containing Antibody–Drug Conjugate (ADC) Targeting CD25-Expressing Hematological Malignancies. *Molecular cancer therapeutics* **15,** 2709–2721 (2016).

201. Jeffrey, S. C. *et al.* A potent anti-CD70 antibody–drug conjugate combining a dimeric pyrrolobenzodiazepine drug with site-specific conjugation technology. *Bioconjugate chemistry* **24,** 1256–1263 (2013).

202. Andrews, K. T., Fisher, G. & Skinner-Adams, T. S. Drug repurposing and human parasitic protozoan diseases. *International Journal for Parasitology: Drugs and Drug Resistance* **4,** 95–111 (2014).

203. DeJesus, E., Kieft, R., Albright, B., Stephens, N. A. & Hajduk, S. L. A single amino acid substitution in the group 1 Trypanosoma brucei gambiense haptoglobin-hemoglobin receptor abolishes TLF-1 binding. *PLoS pathogens* **9** (ed Hill, K. L.) e1003317 (2013).

204. Shaanan, B. Structure of human oxyhaemoglobin at 2 1resolution. *Journal of molecular biology* **171,** 31–59 (1983).

205. Stevens, J., Noyes, H., Schofield, C. & Gibson, W. The molecular evolution of Trypanosomatidae. *Advances in parasitology* **48,** 1–53 (2001).

206. Eyford, B. A. *et al.* Differential protein expression throughout the life cycle of Trypanosoma congolense, a major parasite of cattle in Africa. *Molecular and biochemical parasitology* **177,** 116–125 (2011).

207. Jackson, A. P. *et al.* Global gene expression profiling through the complete life cycle of Trypanosoma vivax. *PLoS neglected tropical diseases* **9,** e0003975 (2015).

208. Vickerman, K. Developmental cycles and biology of pathogenic trypanosomes. *British medical bulletin* **41,** 105–114 (1985).

209. Yamasaki, S. *et al.* Characterization of an epimastigote-stage-specific hemoglobin receptor of Trypanosoma congolense. *Parasites & vectors* **9,** 299 (2016).

210. Newby, Z. E. *et al.* A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nature protocols* **4,** 619 (2009).

211. Reeve, J., Kligman, L. H. & Anderson, R. Are natural lipids UV-screening agents? *Applied Microbiology and Biotechnology* **33,** 161–166 (1990).

212. Nasr, M. L. *et al.* Covalently circularized nanodiscs for studying membrane proteins and viral entry. *nature methods* **14,** 49–52 (2017).

213. Ghosh, P., Mel, S. F. & Stroud, R. M. The domain structure of the ion channel-forming protein colicin Ia. *Nature Structural & Molecular Biology* **1,** 597–604 (1994).

214. Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41,** 388–397 (2007).

215. Richmond, G. S. *et al.* Lipidomic analysis of bloodstream and procyclic form Trypanosoma brucei. *Parasitology* **137,** 1357–1392 (2010).

216. Sobko, A. A., Kotova, E. A., Antonenko, Y. N., Zakharov, S. D. & Cramer, W. A. Effect of lipids with different spontaneous curvature on the channel activity of colicin E1: evidence in favor of a toroidal pore. *FEBS letters* **576,** 205–10 (2004).

217. Basañez, G. *et al.* Bax-type apoptotic proteins porate pure lipid bilayers through a mechanism sensitive to intrinsic monolayer curvature. *Journal of Biological Chemistry* **277,** 49360–49365 (2002).

218. Aguilella, V. M., Verdiá-Báguena, C. & Alcaraz, A. Lipid charge regulation of nonspecific biological ion channels. *Physical Chemistry Chemical Physics* **16,** 3881–3893 (2014).

219. Koivusalo, M., Steinberg, B. E., Mason, D. & Grinstein, S. In situ measurement of the electrical potential across the lysosomal membrane using FRET. *Traffic* **12,** 972–982 (2011).

220. Nolan, D. P. & Voorheis, H. P. Factors that determine the plasma-membrane potential in bloodstream forms of Trypanosoma brucei. *The FEBS Journal* **267,** 4615–4623 (2000).

221. Chen, L. B. Mitochondrial membrane potential in living cells. *Annual review of cell biology* **4,** 155–181 (1988).

222. Denisov, I. G. & Sligar, S. G. Cytochromes P450 in nanodiscs. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1814,** 223–229 (2011).

223. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504,** 107 (2013).

Appendix A

Primer sequences

TABLE A.1: Table of primers used for cloning

Construct	Direction	Restriction site	Tag	Method	Sequence
TbbHpHbRt-pEt15bcon	f	N/A	N/A	SDM	gagatgaagcgctaggggaacccgatc
TbbHpHbRt-pEt15bcon	r	N/A	N/A	SDM	gatcgggttcccctagcgcttcatctc
TbbHpHbRt-BAP-pEt15bcon	f	BamHI	N/A	Ligation	cgggatccggcgctgagggtttaaaaaccaaagac
TbbHpHbRt-BAP-pEt15bcon	r	NdeI	BAP	Ligation	cgggtaccctactcgtgccattcgattttctgagcttcgaaaa tatcgttcaggccgcgcttcatctcttcctgtacttttgtttc
TcHpHbRt-BAP-pEt15bcon	f	BamHI	N/A	Ligation	cgggatccatggaaaatctgtacttccaaggcgctg
TcHpHbRt-BAP-pEt15bcon	r	KpnI	BAP	Ligation	cgcatatgttactcgtgccattcgattttctgagcttcgaaaa tatcgttcaggcctgaggattctgtctcaaccttgctgc
TvHpHbRt-BAP-pEt15bcon	f	BamHI	N/A	Ligation	cgggatccggcaacatcaccgccgaggacg
TvHpHbRt-BAP-pEt15bcon	r	NdeI	BAP	Ligation	cgcatatgttactcgtgccattcgattttctgagcttcgaaaa tatcgttcaggccggagccgttaccgccag
N-terminally truncated ApoL1-pEt15b	f	NdeI	N/A	SLIC	cagcggcctggtgccgcgcggcagccatatggacccggaaagc agcatttttatcgaaga
N-terminally truncated ApoL1-pEt15b	r	BamHI	N/A	SLIC	ttcctttcgggctttgttagcagccggatccttatcacagttc ttgatcggcctgcagga
PFD-pEt15b	f	N/A	N/A	SLIC	caggcacaggcacatgattaagtgattaaaagtctgg
PFD-pEt15b	r	N/A	N/A	SLIC	ccagacttttaatcacttaatcatgtgcctgtgcctg
CTD and N-terminally truncated ApoL1-BAP-pEt15b	r	BamHI	BAP	Ligation	cgggatccctactcgtgccattcgattttctgagcttcgaaaa tatcgttcaggcctcacagttcttgatcggcctgcagg

Appendix B

Publication arising from this thesis

Lane-Serff, H., MacGregor, P., Lowe, E. D., Carrington, M., & Higgins, M. K. Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. *Elife*, **3**, e05553 (2014).

Lane-Serff, H., MacGregor, P., Peacock, L., Macleod, O. J., Kay, C., Gibson, W., Higgins, M. K. & Carrington, M. (2016). Evolutionary diversification of the trypanosome haptoglobin-haemoglobin receptor from an ancestral haemoglobin receptor. *Elife*, **5**, e13044.

Higgins, M. K., **Lane-Serff, H.**, MacGregor, P., & Carrington, M. A Receptors Tale: An Eon in the Life of a Trypanosome Receptor. *PLoS pathogens*, **13(1)**, e1006055 (2017).