Structural studies of Immune Mapped Protein 1 (IMP1) homologues in apicomplexan parasites

Stefi Victoria Benjamin

Department of Life Sciences, Imperial College London

A thesis submitted for the degree of Doctor of Philosophy (PhD), and the Diploma of Imperial College.

Abstract

Apicomplexan parasites are responsible for causing a number of diseases in humans and animals. *Toxoplasma gondii* is known to cause toxoplasmosis in humans, *Eimeria tenella* is known to cause coccidiosis in chickens and *Plasmodium falciparum* is known to cause malaria in humans. These parasites place a huge burden on global health and on the global economy. Current methods of treatment are losing effectiveness, raising the need for new ways to combat these diseases. Identifying vaccine candidates against apicomplexan parasites has been the focus of recent research. Immune Mapped Protein 1 (IMP1) is a novel antigenic protein which was first identified in *Eimeria maxima*. Recent research has shown that immunisation with IMP1 raises immunity against the parasite in a species specific manner. However, the structure and function of IMP1 is not known.

Here we report the structure of the IMP homologue from *Plasmodium falciparum* (PfIMP2) solved using solution state Nuclear Magnetic Resonance (NMR). The structure of PfIMP2 has revealed that IMP homologues have a conserved C-terminal IMP domain. Recent research has also revealed that the IMP domain is the immune-dominant region in the IMP homologues. We also report that there are two classes of IMP homologues within apicomplexans. This redundancy of the protein within the parasites suggests that IMP homologues play an important role within the parasite. *In vitro* experiments performed to gain some insight on the functional role played by IMP proteins within the parasites are also reported. Although the function of the IMP homologues is yet to be deduced, the immunogenic nature of these homologues makes them interesting candidates to study.

Declaration

I hereby declare that the work presented in this thesis is my own, except where specifically indicated in the text. The information gained from the work of others has been acknowledged in the references.

Stefi V Benjamin

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

I would like to give all the glory and honour to the Lord Jesus Christ for His abundant blessings in my life.

I would like to thank my Supervisor, Prof. Steve Matthews for giving me the opportunity to work on this project and for his tremendous support, patience and guidance throughout my PhD.

I would also like to thank the members of the Matthews lab for helping me at various stages of my PhD. I would like to especially thank Dr. Maud Henry for guiding me at the beginning of my PhD and also for proof reading my reports. I would like to thank Dr. Jan Marchant and Dr. Yingqi Xu for their help in running the NMR experiments and in data analysis.

I would like to thank Master Students, Louise Kerry and Felix Williams, for their contributions in sample preparation and data analysis, respectively.

I would also like to thank Dr. Jake Baum for giving me the opportunity to work in his lab to perform *in vivo* experiments. Special thanks to Dr. Chwen Ling Tay and Dr. Rama Desai for their help and guidance in performing these experiments.

I would like to thank my parents and my brother who supported me in every possible way, throughout my PhD. I would not be where I am today without your love, support, sacrifices and prayers.

Table of Contents

Abstract 2

Declaration 3

Copyright Declaration 4

Acknowledgements 5

Table of Contents 6

List of Figures 13

List of Tables 18

Abbreviations 19

Chapter 1 22

Apicomplexan Biology 22

1.1 Apicomplexans 23

1.1.1 Clinical Significance of Apicomplexans 24

1.2 Toxoplasma gondii 25

1.2.1 Life Cycle 25

1.2.2 Pathogenesis 26

1.3 The Glideosome 27

1.4 Invasion Mechanism 29

1.4.1 Initial attachment 30

1.4.2 Apical attachment 30

1.4.3 Commitment step 30

1.4.4 Moving Junction 31

1.4.5 Closure 31

1.4.6 Separation 31

1.5 SAG proteins 32

1.6 Microneme Proteins 33

1.6.1 MIC2 and M2AP 33

1.6.2 MIC1-MIC4-MIC6 36
1.6.3 Other microneme proteins 38
1.6.4 AMA1 39
1.7 Rhoptry and Rhoptry Neck Proteins 39
1.8 AMA1-RON2 complex 40
1.9 *Eimeria tenella* - Life Cycle 42
1.10 Plasmodium falciparum 43
1.10.1 *P. falciparum* – Life cycle 43
1.10.2 *P. falciparum* – Erythrocytic Schizogony 43
1.11 Vaccine development 44
1.11.1 Protein Vaccines 45
1.11.2 DNA Vaccines 46
1.11.3 Live, vectored Vaccines 47
1.12 Immune Mapped Protein 1 48
1.13 Project Aims and Objectives 49
Chapter 2 50
Nuclear Magnetic Resonance 50
2.1 Structural Biology Techniques 51
2.1.1 X-ray crystallography 51
2.1.2 Cryo-Electron Microscopy 55
2.2 NMR for Protein Structure Determination 60
2.2.1 Nuclear spin 60
2.2.2 The Classical model 60
2.2.3 The Quantum model 61
2.2.4 Bulk Magnetisation 62
2.2.5 Chemical Shift 63
2.2.6 Coupling 66

2.2.7 Relaxation mechanisms 68
2.2.8 Nuclear Overhauser Effect (NOE) 69
2.3 NMR Experiments 72
2.3.1 1D NMR 72
2.3.2 2D NMR 73
2.3.3 Three dimensional NMR 75
2.4 ARIA – Ambiguous Restraints for Iterative Assignment 84
Chapter 3 87
Materials & Methods 87
3.1 *T. gondii* IMP1 homologue 88
3.1.1 PCR amplification 88
3.1.2 Screening for positive colonies 89
3.1.3 Plasmid purification 90
3.1.4 Transformation of *E. coli* 90
3.1.5 Expression test 90
3.1.6 Expression 91
3.1.7 Protein Purification 91
3.1.8 SDS-PAGE 92
3.1.9 Limited Proteolysis 92
3.1.10 Membrane transfer for N-terminal amino acid sequencing 92
3.1.11 NMR experiments 92
3.1.12 Crystallisation trials 94
3.2 *Eimeria tenella* IMP1 homologue 94
3.2.1 PCR Amplification 94
3.2.2 Screening for positive colonies 94
3.2.3 NMR experiments 95
3.3 *P. falciparum* IMP2 homologue 95

3.3.1 PCR Amplification 95
3.3.2 Plasmid purification 96
3.3.3 Protein Purification 96
3.3.4 Factor Xa Cleavage 97
3.3.5 DSF Buffer Screen 97
3.3.6 Western blot 99
3.3.7 TEV cleavage 99
3.3.8 Site-directed mutagenesis 99
3.4 NMR 99
3.4.1 NMR Sample preparation 99
3.4.2 NMR experiments 100
3.5 ARIA 100
3.6 In vitro Experiments 101
3.6.1 Culturing *P. falciparum* parasites 101
3.6.2 Sorbitol synchronisation 101
3.6.3 Immunofluorescence assay 101
3.6.4 Pull down Assay 102
3.6.5 Silver staining 102
Chapter 4 103
Characterisation of coccidian IMP-like homologues 103
4.1 Characterisation of IMP1-like homologue in *E. tenella* 104
4.1.1 PCR Amplification 104
4.1.2 Colony screening 104
4.2.3 Expression test 105
4.1.4 Protein expression and purification 106
4.1.5 Limited proteolysis 107
4.1.6 ^{1}H 1D NMR 108

4.1.7 ^{15}N TROSY-HSQC 109
4.2 Characterisation of IMP1-like homologue in *T. gondii* 111
4.2.1 Bioinformatics analysis 111
4.2.2 Optimisation of TgIMP1 construct 114
4.2.3 Expression tests 114
4.2.4 ^{1}H 1D NMR 115
4.2.5 Limited proteolysis 117
4.2.6 Optimised TgIMP1 construct 118
4.2.7 Protein Expression and Purification 119
4.2.8 ^{1}H 1D NMR 120
4.2.9 ^{15}N TROSY-HSQC 121
4.2.10 First attempt at backbone assignment 122
4.3 Further NMR studies of TgIMP1 123
4.3.1 Deuteration of TgIMP1 clone 7 123
4.3.2 ^{15}N TROSY-HSQC 124
4.3.3 Backbone assignment of TgIMP1 clone 7 125
4.3.4 Crystallisation trials of TgIMP1 clone 7 127
4.4 Summary 128
Chapter 5 130
Characterisation of *Plasmodium falciparum* homologue of IMP (Class II) 130
5.1 Cloning, Protein expression and purification 131
5.1.1 PCR Amplification 131
5.1.2 Colony screening 131
5.1.3 Expression test 132
5.2 pET32Xa/LIC construct 134
5.2.1 Protein expression and purification 134
5.2.2 ^{1}H 1D NMR 135

5.3 pET30Xa/LIC construct 138
5.3.1 Protein expression and purification 138
5.3.2 ^{1}H 1D NMR 138
5.4 pNIC-ZB construct 140
5.4.1 Protein expression and purification 140
5.4.2 Western Blot analysis 143
5.4.3 Differential Scanning Fluorimetry (DSF) 143
5.4.4 ^{1}H 1D NMR 144
5.4.5 ^{15}N TROSY-HSQC of PfIMP2 with the His tag 145
5.4.6 TEV cleavage 146
5.4.7 ^{15}N TROSY HSQC of PfIMP2 without the His tag 147
5.4.8 PfIMP2 C71S mutant 148
5.5 Summary 150
Chapter 6 152
NMR assignments and Structure Determination of PfIMP2 152
6.1 NMR assignment 153
6.1.1 Backbone assignment 153
6.1.2 Side chain assignments 155
6.2 Structure calculation 159
6.3 Ramachandran plot 160
6.4 Solution state structure of PfIMP2 163
6.5 Structural features of PfIMP2 164
6.6 Modelled structures of homologues – TgIMP1 and EtIMP1 169
6.7 Summary 173
Chapter 7 175
Final Remarks 175
7.1 Immune Mapped Protein 1 (IMP1) – a novel immunogenic protein 176

7.2 Domain architecture of IMP1 176
7.3 Two classes of IMP proteins 178
7.4 PfIMP2 has a novel fold 178
7.5 Comparison of Class I and Class II IMP structures 180
7.6 Localisation and Pull down studies 181
7.7 Challenges and further work 182
Bibliography 185
Appendix I 197
Appendix II 217
A2.1 Culturing of parasites 217
A2.1.1 Transfection of parasites 217
A2.1.2 Sorbitol synchronisation 219
A2.1.3 Immunofluorescence Assay (IFA) 219
A2.1.4 Confocal Microscopy 220
A2.1.5 Pull down Assay 222
A2.2 Localisation and pull down studies 224
A2.2.1 PfIMP2 expression profile 224
A2.2.2 Localisation studies of PfIMP2 225
A2.2.3 Identification of PfIMP2 binding partners 226
A2.3 Summary 228
Appendix III 230

List of Figures

Figure 1.1: Structure of a *T. gondii* tachyzoite

Figure 1.2: Life cycle of *T. gondii*

Figure 1.3: Cartoon representation of the glideosome in *T. gondii*

Figure 1.4: Invasion process of *T. gondii*

Figure 1.5: Electrostatic surface representation of the SAG1 homodimer

Figure 1.6: Hydrophobic surface representation of the β domain of M2AP

Figure 1.7: (A) cartoon representation of the MIC1 C terminal domain/MIC6 EGF domain complex (B) cartoon representation of the MIC1 MAR domain bound to 3'SiaLacNac1-3 and (C) cartoon representation of the fifth Apple domain of MIC4

Figure 1.8: Cartoon representation of MIC5

Figure 1.9: (A) Cartoon and surface representations of AMA1 ectodomain and (B) cartoon representation of unbound AMA1 ectodomain and AMA1 ectodomain bound to RON2 fragment

Figure 1.10: Life cycle of *E. tenella*

Figure 1.11: Life cycle of *P. falciparum*

Figure 2.1.1: Schematic representation of X ray crystallography

Figure 2.1.2: Cartoon representation of the crystal structure of PvRBP2a

Figure 2.1.3: Schematic representation of Cryo-Electron Microscopy

Figure 2.1.4: Cartoon representation of the Cryo-EM structure of *Plasmodium falciparum* 20s proteasome

Figure 2.2.1: (A) magnetic field produced by a spinning nucleus and (B) precession of the spinning nucleus in the presence of an external magnetic field

Figure 2.2.2: Energy level diagram in the presence of an external magnetic field

Figure 2.2.3: The process of developing FID from a NMR sample

Figure 2.2.4: Schematic representation of averaged and unaveraged CSA

Figure 2.2.5: Decoupled and J-coupled NMR spectra

Figure 2.2.6: Schematic representation of NOE transitions

Figure 2.3.1: 1D NMR experiment – preparation and detection

Figure 2.3.2: Schematic representation of TROSY technique

Figure 2.3.3: Schematic showing a dipeptide

Figure 2.3.4a: Schematic representation of the CBCANH experiment

Figure 2.3.4b: Schematic representation of the CBCA(CO)NH experiment

Figure 2.3.4c: Schematic representation of the HNCO experiment

Figure 2.3.4d: Schematic representation of the HN(CA)CO experiment

Figure 2.3.4e: Schematic representation of the HBHA(CBCACO)NH experiment

Figure 2.3.4f: Schematic representation of the CC(CO)NH experiment

Figure 2.3.4g: Schematic representation of the HCCH-TOCSY experiment

Figure 2.4: Schematic representation of the various stages followed by ARIA

Figure 4.1.1: Annotated amino acid sequence of full length EtIMP1

Figure 4.1.2: Agarose gel results for colony PCR of EtIMP1 clone 2 (pET30 Xa/LIC)

Figure 4.1.3: SDS-PAGE results for expression test of EtIMP1 clone 2

Figure 4.1.4: SDS-PAGE results for Nickel affinity chromatography of EtIMP1 clone 2

Figure 4.1.5: (A) SDS-PAGE results for limited proteolysis of EtIMP1 clone 2 using chymotrypsin and (B) Annotated amino acid sequence of full length EtIMP1

Figure 4.1.6: ^{1}H 1D spectrum of EtIMP1 clone 2 (pET30 Xa/LIC)

Figure 4.1.7: 2D ^{15}N TROSY-HSQC spectrum of EtIMP1 clone 2 (^{13}C/^{15}N labelled)

Figure 4.2.1a: BLAST results and secondary structure prediction of the *Toxoplasma gondii* ME49 IMP1 homologue

Figure 4.2.1b: Sequence identity scores for the sequence alignments of the homologues shown in Figure 4.2.1a

Figure 4.2.2: SDS-PAGE results for full length TgIMP1 expressed in Rosetta2 cells

Figure 4.2.3: ^{1}H 1D spectrum of the full length TgIMP1 protein

Figure 4.2.4: ^{1}H 1D spectrum of TgIMP1 clone 6

Figure 4.2.5: (a) SDS-PAGE results for limited proteolysis of full length TgIMP1 with Chymotrypsin and (b) Annotated amino acid sequence of full length TgIMP1

Figure 4.2.6: Agarose gel results for colony PCR of TgIMP1 clone 7

Figure 4.2.7: (a) SDS-PAGE results for Nickel affinity chromatography of TgIMP1 clone 7 and (b) SDS-PAGE results for size exclusion chromatography of TgIMP1 clone 7

Figure 4.2.8: ^{1}H 1D spectrum of TgIMP1 clone 7 ($^{13}C/^{15}N$ labelled)

Figure 4.2.9: ^{15}N TROSY-HSQC spectrum of TgIMP1 clone 7 ($^{13}C/^{15}N$ labelled)

Figure 4.3.1a: SDS-PAGE results for TgIMP1 clone 7 expression in Rosetta2 cells in D_2O minimal media

Figure 4.3.1b: ^{1}H 1D spectrum of TgIMP1 clone 7 ($^{13}C/^{15}N/^{2}H$ labelled)

Figure 4.3.2: Overlay of the 2D ^{15}N HSQC spectra of the $^{13}C/^{15}N$ and $^{13}C/^{15}N/^{2}H$ labelled TgIMP1 clone 7 protein

Figure 4.3.3a: 2D ^{15}N TROSY-HSQC spectrum of double ($^{13}C/^{15}N$) labelled TgIMP1 clone 7

Figure 4.3.3b: Strips from the HN(CO)CACB and HNCACB experiments used for backbone assignments of triple labelled ($^{13}C/^{15}N/^{2}H$) TgIMP1 clone 7

Figure 4.3.4: Crystallisation trial results obtained for TgIMP1 clone 7

Figure 5.1.1: Agarose gel results for colony PCR of full length PfIMP2 cloned into pNIC-ZB vector

Figure 5.1.2 (A): SDS-PAGE results for full length PfIMP2 (pET32 Xa/LIC) expressed in ArcticExpress and BL21 cells

Figure 5.1.2 (B): SDS-PAGE results for full length PfIMP2 (pET32 Xa/LIC) expressed in Rosetta 2 cells

Figure 5.1.3: SDS-PAGE results for full length PfIMP2 (pNIC-ZB) in BL21 and Rosetta2 cells

Figure 5.2.1a: SDS-PAGE results for Nickel affinity chromatography of full length PfIMP2 (pET32 Xa/LIC) expressed in Rosetta2 cells

Figure 5.2.1b: SDS-PAGE results for Nickel affinity chromatography of full length PfIMP2 (pET32 Xa/LIC) expressed in BL21 cells

Figure 5.2.2a: ^{1}H 1D spectrum of PfIMP2 (pET32 Xa/LIC construct) expressed in BL21 cells

Figure 5.2.2b: ^{1}H 1D spectrum of PfIMP2 (pET32 Xa/LIC construct) expressed in Rosetta2 cells

Figure 5.2.2c: ^{1}H 1D spectra of PfIMP2 (pET32 Xa/LIC construct) expressed in Rosetta2 cells and thioredoxin

Figure 5.3.1: SDS-PAGE results for Nickel affinity chromatography of full length PfIMP2 (pET30 Xa/LIC) expressed in Rosetta2 cells

Figure 5.3.2: ^{1}H 1D spectrum of PfIMP2 (pET30 Xa/LIC construct) expressed in Rosetta 2 cells

Figure 5.3.3: 2D ^{15}N TROSY-HSQC spectrum of single (^{15}N) labelled PfIMP2 (pE30 Xa/LIC)

Figure 5.4.1a: SDS-PAGE results for Nickel affinity chromatography of full length PfIMP2 (pNIC-ZB) expressed in Rosetta2 cells

Figure 5.4.1b: SDS-PAGE results for size exclusion chromatography of unlabelled and $^{13}C/^{15}N$ labelled full length PfIMP2 (pNIC-ZB) expressed in Rosetta2 cells

Figure 5.4.1c: SDS-PAGE results for various concentrations of PfIMP2 sample

Figure 5.4.2: Western blot result of full length PfIMP2 (pNIC-ZB) with His tag

Figure 5.4.3: Buffer screen result using DSF for PfIMP2 (pNIC-ZB)

Figure 5.4.4: ^{1}H 1D spectrum of PfIMP2 (pNIC-ZB)

Figure 5.4.5: 2D ^{15}N TROSY-HSQC spectrum of $^{13}C/^{15}N$ labelled PfIMP2 (pNIC-ZB) protein with the N-terminal His tag

Figure 5.4.6: SDS-PAGE results for TEV cleavage and size exclusion chromatography of $^{13}C/^{15}N$ labelled full length PfIMP2 (pNIC-ZB) expressed in Rosetta2 cells

Figure 5.4.7: 2D ^{15}N TROSY-HSQC spectrum of double labelled PfIMP2 (pNIC-ZB) protein with and without the N-terminal His tag

Figure 5.4.8: SDS-PAGE results for Nickel affinity chromatography of full length PfIMP2 (pNIC-ZB) wild type and C71S mutant expressed in Rosetta2 cells

Figure 5.5: Labelled sequence alignment of PfIMP2 against the top six sequences from the BLAST search

Figure 6.1.1a: Assigned $^{1}H/^{15}N$ TROSY-HSQC spectrum of recombinant PfIMP2 without His tag

Figure 6.1.1b: Strips obtained from the CBCA(CO)NH and CBCANH experiments for double labelled ($^{13}C/^{15}N$) full length PfIMP2 for backbone assignment

Figure 6.1.2a: ^{13}C-NOESY and ^{15}N-NOESY strips for residue I35 of double labelled ($^{13}C/^{15}N$) full length PfIMP2

Figure 6.1.2b: Cartoon representation of PfIMP2 showing the short range and long range NOEs between the I35 (pink), W23 (cyan) and F73 (blue) residues in the PfIMP2 structure

Figure 6.1.2c: Cartoon representation of PfIMP2 showing the W23 (cyan), I35 (pink) and F73 (blue) residues

Figure 6.2: Stereo representation of the ensemble of PfIMP2 structures

Figure 6.3a: Ramachandran plot for the PfIMP2 structure

Figure 6.3b: Cartoon representation of the PfIMP2 structure highlighting H89

Figure 6.4a: Annotated cartoon representation of the PfIMP2 structure

Figure 6.4b: Cartoon representation of MDB1

Figure 6.5a: Cartoon representation of the structure of PfIMP2 showing the two predicted phosphorylation sites – S92 and S93

Figure 6.5b: Cartoon representation of the structure of PfIMP2 showing the only cysteine residue, C71

Figure 6.5c: Hydrophobic surface representation of PfIMP2

Figure 6.5d: Electrostatic surface representation of PfIMP2

Figure 6.5e: Surface representation of the level of conservation of PfIMP2

Figure 6.6a: Cartoon representation of the modelled structure of the IMP domain of TgIMP1 highlighting the insertions

Figure 6.6b: Cartoon representation of the modelled structure of the IMP domain of EtIMP1 highlighting the insertions

Figure 6.6c: Hydrophobic surface representation of the IMP domains of TgIMP1 and EtIMP1

Figure A2.1.1: Cartoon representation of the structure of the dihydrofolate reductase domain of *P. falciparum* dihydrofolate reductase-thymidylate synthase protein bound to WR99210

Figure A2.1.2: Schematic representation of direct and indirect IFA techniques

Figure A2.1.3: Schematic representation of the principle behind confocal microscopy

Figure A2.1.4: Schematic representation of the steps involved in a pull down assay

Figure A2.2.1: Expression profile of PfIMP2 in *P. falciparum* 3D7 following erythrocytic invasion

Figure A2.2.2: IFA results for the localisation of PfIMP2 in *P. falciparum*

Figure A2.2.3: SDS-PAGE and Western blot results of the pull down samples from the WT and PfIMP2 strains

List of Tables

Table 3.1: PCR conditions used to amplify the DNA of the IMP1 homologue from the T. gondii transcriptome

Table 3.2: Table showing the PCR reaction conditions set for the thermocycler.

Table 3.3: Different conditions used for the buffer screen performed for PfIMP2 (pNIC-ZB)

Table 3.4: Tolerance levels set for the 13C- and15N- NOESY datasets recorded on the Bruker Avancell 800 (800 MHz)

Table 4.1: List of TgIMP1 constructs

Table 6.1: NMR restraints used for structure calculation and the structural statistics for the final structure of PfIMP2

Appendix Table 1: Primer sequences used for various constructs designed during the course of the project

Appendix Table 2: Resistance gene and tags present in the various vectors used in the project

Appendix Table 3: Recipes of all the buffers referred to in the thesis

Abbreviations

AA	-	Amino Acid
ACT	-	Artemisinin-based combination therapy
ADR	-	Ambiguous Distance Restraints
AIDS	-	Acquired immunodeficiency syndrome
AMA1	-	Apical Membrane Antigen 1
ARIA	-	Ambiguous Restraints for Iterative Assignment
BLAST	-	Basic Local Alignment Search Tool
Cryo-EM	-	Cryo Electron Microscopy
CSA	-	Chemical shift anisotropy
DAPI	-	4', 6-diamidino-2-phenylindole
DC	-	Dipolar Coupling
DNA	-	Deoxyribonucleic acid
DSF	-	Differential Scanning Fluorimetry
DTT	-	Dithiothreitol
EGF	-	Epidermal Growth Factor
ELC	-	Essential Light Chain
FID	-	Free Induction Decay
GAP	-	Glideosome associated proteins
GAPDH	-	Glyceraldehyde 3-phosphate dehydrogenase
GFP	-	Green Fluorescent Protein
GUI	-	Graphic User Interface
HA	-	Human Influenza Hemagglutinin
HMQC	-	Heteronuclear Multiple Quantum Coherence Spectroscopy
HSQC	-	Heteronuclear Single Quantum Coherence Spectroscopy
IFA	-	Immunofluorescence Assay
IFN	-	Interferon
IL	-	Interleukin

IMC	-	Inner Membrane Complex
IMP	-	Immune Mapped Protein
IMP1	-	Class I Immune Mapped Protein homologue
IMP2	-	Class II Immune Mapped Protein homologue
IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
LAMTOR1	-	late endosomal/lysosomal adaptor, MARK and mTOR activator 1
LIC	-	Ligation independent cloning
M2AP	-	MIC2 associated protein
MAD	-	Multiple Anomalous Dispersion
MAPS	-	Monolithic Active Pixel Sensors
MAR	-	Microneme Adhesive Repeat
MIC	-	Micronemal
MIR	-	Multiple Isomorphous Replacement
MJ	-	Moving Junction
MLC	-	Myosin Light Chain
MR	-	Molecular Replacement
MWCO	-	Molecular weight cut off
NADPH	-	Nicotinamide Adenine Dinucleotide Phosphate
NMR	-	Nuclear Magnetic Resonance
NOE	-	Nuclear Overhauser Effect
NOESY	-	Nuclear Overhauser Effect Spectroscopy
OD	-	Optical Density
PAGE	-	Polyacrylamide Gel Electrophoresis
PCR	-	Polymerase Chain Reaction
PV	-	Parasitophorous Vacuole
PVM	-	Parasitophorous vacuolar membrane
PvRBP	-	*Plasmodium vivax* Reticulocyte Binding Protein
RF	-	Radiofrequency

RNA	-	Ribonucleic acid
ROM	-	Rhomboid-like proteases
RON	-	Rhoptry Neck protein
ROP	-	Rhoptry protein
SAD	-	Single Anomalous Dispersion
SAG	-	GPI-anchored Surface antigens
SDS	-	Sodium dodecyl Sulphate
SIR	-	Single Isomorphous Replacement
SIRAS	-	Single Isomorphous Replacement with Anomalous Scattering method
SPG	-	Sulphated proteoglycans
SRS	-	SAG1-related sequence
SUB1	-	Subtilisin-like protein 1
TEV	-	Tobacco Etch Virus
TMS	-	Tetramethylsilane
TOCSY	-	Total Correlation Spectroscopy
TROSY	-	Transverse Relaxation Optimised Spectroscopy
TSR	-	Thrombospondin like repeats
UV	-	Ultraviolet
WHO	-	World Health Organisation
WT	-	Wild type

Chapter 1

Apicomplexan Biology

1.1 Apicomplexans

Apicomplexa is a phylum of intracellular parasites that cause a number of human and animal diseases. Apicomplexan parasites include *Plasmodium* species which cause malaria, *Toxoplasma gondii* which causes toxoplasmosis in immunocompromised individuals, *Eimeria* species which cause diseases in most vertebrates in a species-specific manner, *Theileria* species which cause diseases in cattle and *Cryptosporidium* species which cause diseases in animals and humans (Morrissette & Sibley, 2002). These parasites are polar cells and are characterised by the presence of a cytoplasmic complex at the apical end of the cell. This apical complex is known to be crucial for parasite replication and for invasion and consists of a microtubule organising centre called conoids and secretory organelles involved in invasion as shown in figure 1.1 (Morrissette & Sibley, 2002). As *T. gondii* is considered the model apicomplexan, most of the apicomplexan proteins and mechanisms have been widely studied in this parasite (Roos et al, 1999).

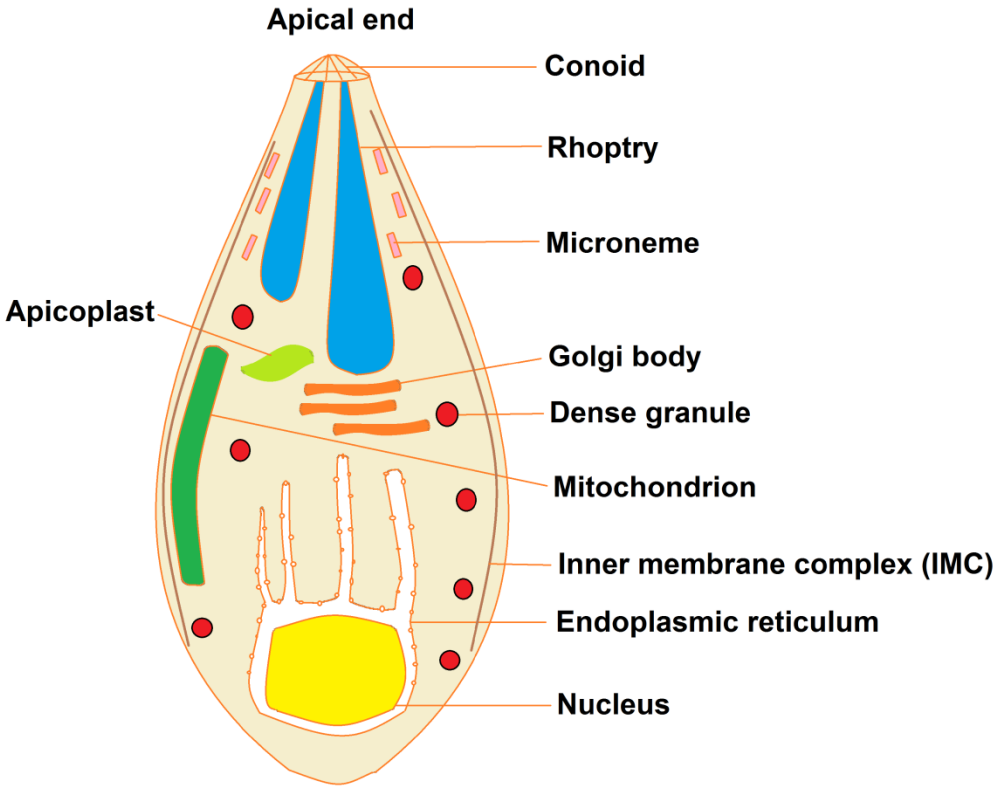


Figure 1.1: Figure showing the structure of a T. gondii tachyzoite as a model apicomplexan. The important organelles found across most apicomplexans are labelled in the figure.

1.1.1 Clinical Significance of Apicomplexans

The three apicomplexan parasites studied in this project are *Toxoplasma gondii, Eimeria tenella* and *Plasmodium falciparum*. *T. gondii* is responsible for causing toxoplasmosis in immunocompromised individuals, like AIDS patients, and congenital infections (Dzierszinski et al, 1999; Luft & Remington, 1992). The time from infection to the development of symptoms may take from 1-2 weeks. The disease can affect the brain, heart, lungs, eyes or liver (Liesenfeld, 2007). Toxoplasmosis can have a wide range of symptoms which are more severe in immunocompromised individuals than healthy individuals. It may also be responsible for causing toxoplasma encephalitis in AIDS patients (Luft & Remington, 1992). Toxoplasmosis is currently treated with chemotherapy but these drugs have toxic effects and reinfection occurs rapidly (Buxton & Innes, 1995).

Coccidiosis in chickens is caused by seven different species belonging to the *Eimeria* genus. *Eimeria tenella* is an intracellular parasite that causes cecal coccidiosis in young chickens (Yin et al, 2014). The symptoms of cecal coccidiosis are ruffled feathers, inappetance, weight loss, closed eyes, diarrhoea and blood in the faeces of infected chickens. *E. tenella* infections have a high morbidity and mortality rate, thus having a detrimental effect on the poultry industry. Parasite oocysts are shed in the faeces of infected chickens which undergo sporulation and are then transmitted to healthy chickens orally during feeding. One of the current methods of treatment is using coccidiostats (anti-protozoan agents). Resistance against currently used coccidiostats are on the rise raising the need for alternative drugs (Blake et al, 2011).

P. falciparum is the deadliest of the five human infecting malaria species - *P. falciparum, P. vivax, P. malariae, P. ovale* and *P. knowlesi*. *P. falciparum* is responsible for the majority of malaria related deaths. The parasite is spread from person to person by the female *Anopheles* mosquito, which is the vector for this parasite. Initial symptoms following

infection include fever, chills, nausea and headache (Ménard, 2013). If not treated immediately, *P. falciparum* infections can lead to severe illness and death. *P. falciparum* species have developed resistance against drugs that were used against malaria in the past which include chloroquine and sulfadoxione-pyrimethamine. The current method of treatment against uncomplicated *P. falciparum* infection is using artemisinin–based combination therapy (ACT). Resistance against ACT is also on the rise, raising the need for new drugs to treat malaria (Crutcher & Hoffman, 1996). Recent statistics from WHO have shown that there were 438000 malaria-related deaths worldwide in 2015 (WHO, 2015). This shows that *P. falciparum* causes a huge global health burden.

All three apicomplexan parasites have a huge impact on the global economy and cause a huge health burden. The current methods of treatment of these three diseases are losing effectiveness worldwide raising the need for alternative methods of treatment. A number of recent studies are focussed on the development of vaccines against these parasites. The recent developments in this area are discussed in detail in Section 1.11.

1.2 Toxoplasma gondii

1.2.1 Life Cycle

The definitive host of *T. gondii* is the cat which is where the sexual stages of the parasite take place. The parasites are released from the cat in the form of oocysts. The oocysts are ingested by intermediate hosts including humans. Oocysts can also be ingested by humans through consumption of uncooked or raw infected meat. Once the oocysts enter humans, they mature into tachyzoites which replicate rapidly. Within humans, the parasites can become encysted and stored in the muscles and the brain in another form called bradyzoites (Louis & Kami, 2007). An overview of the life cycle of *T. gondii* is shown in figure 1.2.

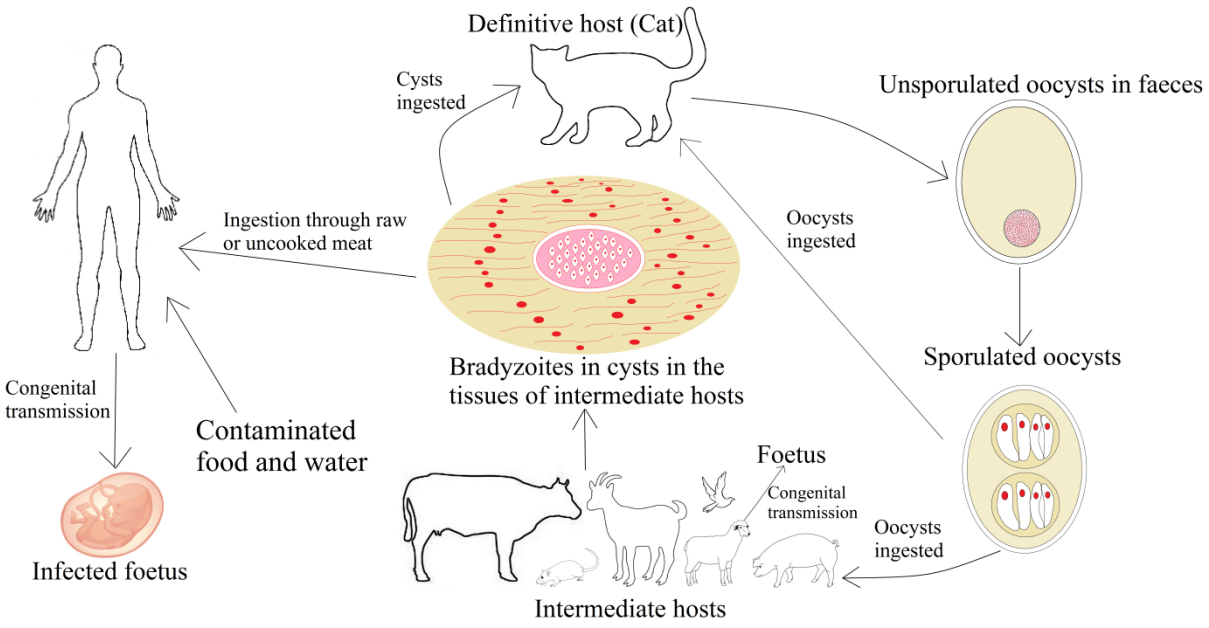


Figure 1.2: Figure showing the various stages in the life cycle of T. gondii and the various hosts involved. The stages of the parasite are labelled in the figure.

1.2.2 Pathogenesis

T. gondii is an obligate intracellular parasite and hence, it survives within the host cell evading the host immune system. *Toxoplasma* exists within humans in two life forms – tachyzoites and bradyzoites (Louis & Kami, 2007). Tachyzoites are the rapidly dividing, intracellular life form. They are crescent-shaped cells and are coated with three unit membranes forming a pellicle (Ferguson & Hutchison, 1987). The anterior end of the tachyzoite contains cytoskeletal elements, a possible microtubule complex called conoids and secretory organelles called rhoptries and micronemes. A third type of secretory organelle called dense granules are distributed everywhere in the cytoplasm (Woodison & Smith, 1990). These secretory organelles contain proteins required for the survival and invasion of the pathogen (Louis & Kami, 2007; Meissner et al, 2002).

Tachyzoites are responsible for causing acute infection and when this is controlled by the host immune response through the induction of IFN-γ (Suzuki et al, 1988), they become encysted

and remain within the muscles and the brain in a dormant form called the bradyzoite. The cysts can contain varying number of bradyzoites which are retained within the host for years and they are responsible for reinfection within immunocompromised hosts. This persistence of the pathogen within the host in a dormant form is one of the major reasons for difficulty in vaccine development against toxoplasmosis (Jackson & Hutchison, 1989). The invasive form of the pathogen is the tachyzoite and understanding its invasion process will demonstrate ways of controlling the infection.

1.3 The Glideosome

Toxoplasma adapts a substrate dependent locomotion called gliding motility and this is used for tissue migration and cell invasion. This locomotion is facilitated by the glideosome which is a multiprotein complex situated in the parasite pellicle, between the inner membrane complex (IMC) and the plasma membrane as shown in figure 1.3 (Williams et al, 2015). The glideosome consists of three components – the motor-receptor bridging proteins, motor complex proteins and the IMC proteins (Boucher & Bosch, 2015). Two motor-receptor bridging proteins have been identified so far – aldolase and glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Aldolase was reported to bind actin on one face and the C terminal end of microneme proteins via the active site (Buscaglia et al, 2003). However, recent studies have shown that mutants lacking aldolase exhibit normal gliding motility and invasion in a glucose-free medium. These results suggest that aldolase is necessary for energy metabolism and not for interacting with microneme proteins (Shen & Sibley, 2014). The interactions of GAPDH are yet to be characterised. The actin-myosin motor produces the force that is necessary for the gliding of the parasite. F-Actin polymerises to form unstable filaments, which are 100 nm in length, that act as the track for the glideosome motor. Myo A (Type XIV Myosin) pulls back on the short actin polymers causing the parasite to move forward. MyoA contains a neck and head region. The head region has a nucleotide binding pocket and the

neck region interacts with two essential light chains (ELC) – ELC1 and ELC2. ELC1 and ELC2 have been found to be involved in regulating and stabilising MyoA. ELC1 has also been reported to interact with Ca^{2+} and Myosin Light Chain 1 (MLC1) (Williams et al, 2015). Glideosome associated proteins (GAP) are involved in bridging the IMC to the actin myosin motor system. GAP45 acts as a tether between the IMC and the plasma membrane. The actin-myosin motor system is anchored to the IMC by transmembrane proteins – GAP40, GAP50 and GAPM family of proteins (Bullen et al, 2009). Gliding motility of the parasite relies on the interaction of the parasite with the host cell surface. Upon secretion of microneme proteins and their interaction with the host cell, the actin-myosin motor system pushes the parasite forward, whilst pushing the protein complex rearward.

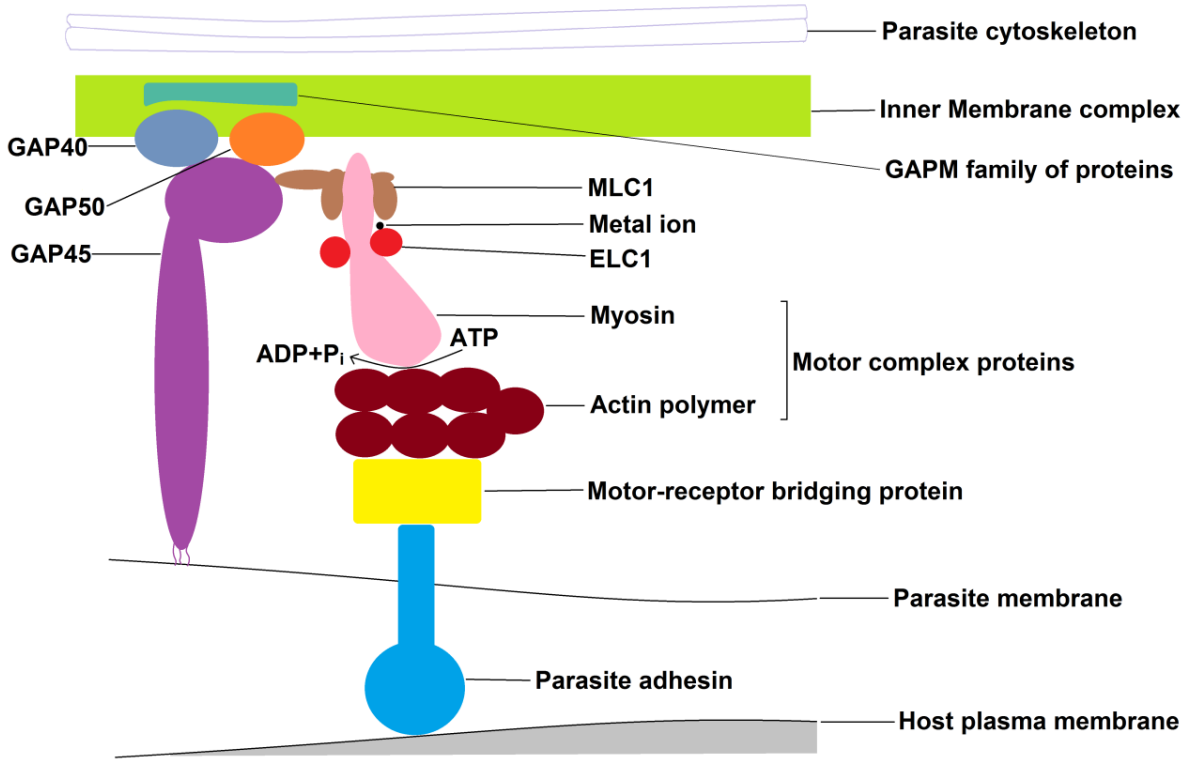


Figure 1.3: Figure showing a cartoon representation of the glideosome in T. gondii. The proteins forming this multi protein complex are labelled in the figure.

1.4 Invasion Mechanism

The invasion mechanism of *T. gondii* in humans has been widely studied giving insight into the roles of various proteins and the mechanisms that they are involved in. Invasion itself occurs in seven steps once the pathogen has located a host cell after gliding along the substrate as shown in figure 1.4 (Carruthers & Boothroyd, 2007). The parasite exists within the host cell within a structure called the parasitophorous vacuole (PV) and this structure protects the parasite within the host cell from degradation by lysosome fusion (Louis & Kami, 2007).

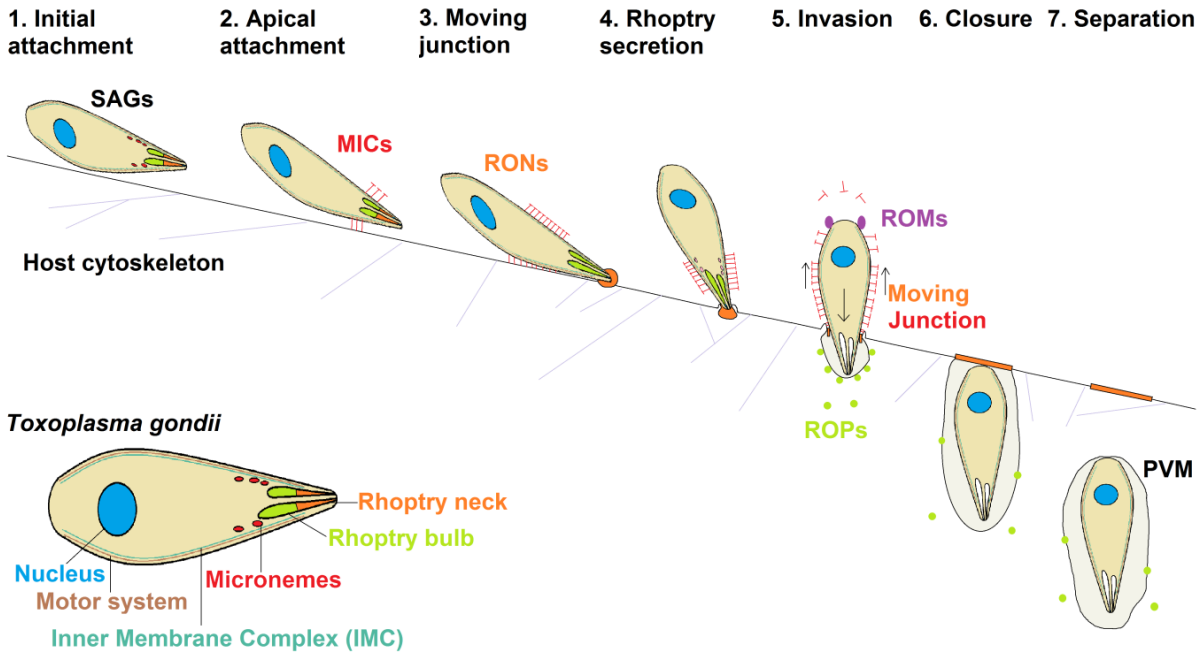


Figure 1.4: Figure showing the various stages of the invasion process of T. gondii. The organelles secreting proteins at different stages are labelled in the figure. The colour coding of the various organelles are also shown in the schematic of T. gondii in the bottom left of the figure.

1.4.1 Initial attachment

The surface of tachyzoites contains a number of proteins that belong to a family of GPI-anchored surface antigens (SAGs). These SAG proteins are almost evenly distributed along the surface of the pathogen and they form low-affinity, lateral interactions with the host surface. It is thought that these proteins may allow the pathogen to analyse the host cell surface for an optimal invasion site. This interaction is reversible and the pathogen can disengage from the host cell surface when conditions are not ideal for invasion (Carruthers & Boothroyd, 2007).

1.4.2 Apical attachment

The invasion process is highly polarised and is initiated at the apical end of the parasite. This step requires the proteins present within the microneme, called micronemal (MIC) proteins. Studies have shown that these MIC proteins are secreted to the parasite's apical surface in a calcium dependent manner (Lovett et al, 2002). Many MICs seem to have a conserved adhesin domain which aids in protein-protein or protein-carbohydrate interaction between the parasite and host cells (Soldati et al, 2001). Studies have also shown that a number of MIC proteins have adhesive domains which might allow binding to multiple receptors, and hence help form a robust interaction (Carruthers & Boothroyd, 2007).

1.4.3 Commitment step

It has been observed that the conoids extrude during invasion in a calcium dependent manner, possibly at the same time as MIC secretion. This extrusion brings the parasite and host plasma membranes into close proximity with each other. During this time, another micronemal protein called apical membrane antigen (AMA1) is involved in forming a third attachment which seems to be the commitment step for invasion (Carruthers & Boothroyd, 2007). AMA1 has been found to form complexes with four other proteins which are secreted from the neck region of rhoptries (RON proteins) (Alexander et al, 2005; Lebrun et al, 2005).

1.4.4 Moving Junction

The AMA1 and RON proteins form a complex as or just before they are secreted to the parasite's apical surface. There they cooperatively form the moving junction (MJ). The MJ is a ring-like band of intimate interaction between the parasite and the host cell and is similar in appearance to the mammalian tight junctions. One of the functions of this structure is to act as a 'sieve' that selectively removes many host and parasite proteins from the two plasma membranes as it moves down the parasite (Charron & Sibley, 2004; Mordue et al, 1999). Rhomboid-like proteases are released by the parasite for proteolysis of these protein complexes (Boucher & Bosch, 2015). ROM4 is one of the most crucial proteins for adhesin removal from the parasite surface and is necessary for efficient host invasion. However, recent mutagenesis studies have shown that MIC2 is removed from the parasite surface in the absence of ROM4 suggesting the presence of ROM-independent mechanism for adhesin removal (Shen et al, 2014). This 'stripping off' of certain proteins may be a possible explanation as to why the parasitophorous vacuolar membrane (PVM) does not fuse with host lysosomes or any other host vesicles, thus evading the host immune system (Carruthers & Boothroyd, 2007).

1.4.5 Closure

Once the MJ ring has completed its migration down the parasitophorous vacuole, closure is achieved by unknown means leaving a surface residue exposed (RON4) on the outside of the host cell (Carruthers & Boothroyd, 2007).

1.4.6 Separation

The PVM is then pinched off and separated from the host plasma membrane, releasing the PV into the host cell. The closure and separation steps together require 2 minutes and the machinery required for these steps remains unknown (Carruthers & Boothroyd, 2007). Once the parasite is within the host, encapsulated within the PVM, the secretion of dense granules

is upregulated to release proteins into the lumen of the PV (Carruthers & Sibley, 1997; Leriche & Dubremetz, 1990). These proteins modify the PV such that it can acquire nutrients from the host cell (Carruthers, 2002).

1.5 SAG proteins

SAG1 is a surface antigen protein that is part of a superfamily of SRS (SAG1-related sequence) proteins. The SRS family consists of 20 homologous proteins which have 25-30% sequence identity. All SRS proteins seem to be (glycosylphosphatidylinositol) GPI-anchored to the parasite surface (Jung et al, 2004). During acute infection, SAG1 and SAG2A are the most immune-dominant proteins within the SRS superfamily. SAG1 is critical for producing a hyper-immune response at the initial stages of infection. The crystal structure of SAG1 has been solved to 1.7 Å resolution and is shown in figure 1.5 (He et al, 2002).

It has also been found that SAG2A and SAG2B are the smallest SRS proteins and have one domain whereas a majority of the other SRS proteins are larger in size and have two SAG2A-like domains. The structure of SAG1 consists of two domains – DI and DII – which are very similar in their secondary structure composition. SAG1 has been found to form a homodimer which is initiated by the DI domains. Studies have also shown that SAG3 can bind sulphated proteoglycans (SPGs) but the exact identity of the SPG is unknown (He et al, 2002). The homodimer interface has a positively charged groove which could possibly bind SPGs as shown in figure 1.5.

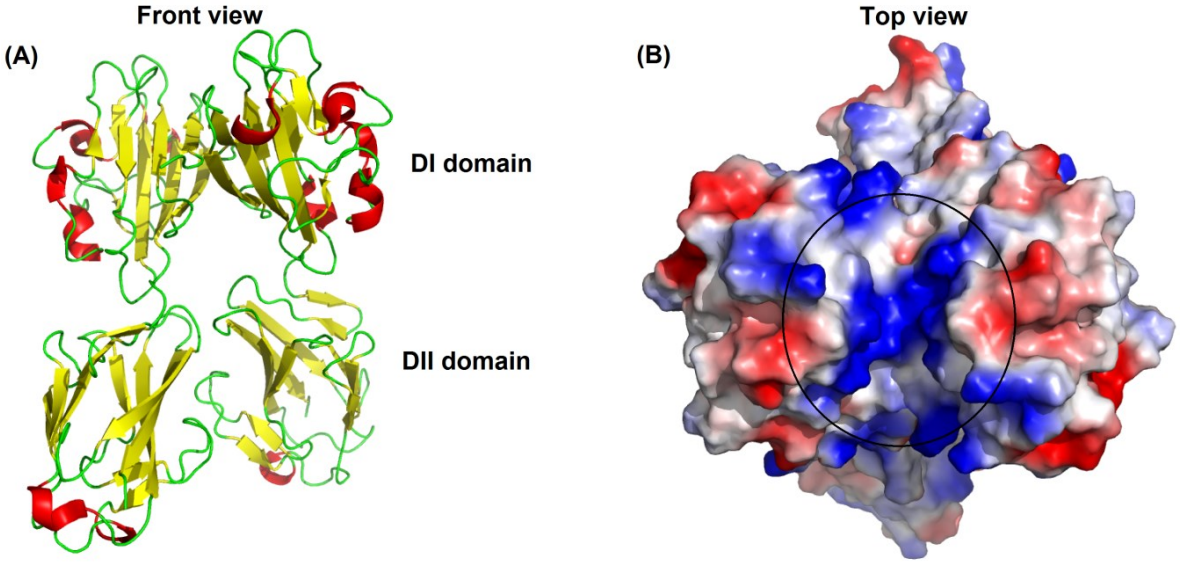


Figure 1.5: Figure showing (A) a cartoon representation of the SAG1 homodimer (alpha helices coloured red, beta sheets coloured yellow and loop regions coloured green) and (B) the electrostatic surface representation showing positively charged groove in the homodimer interface in blue (highlighted by the black circle)

1.6 Microneme Proteins

Micronemes are rod-shaped, secretory organelles which are the first to release proteins during invasion. Knockout studies have shown that MIC proteins are required for successful invasion. It has been found that microneme proteins form complexes such that one or more proteins in the complex have an adhesive function whereas the other protein contains a transmembrane domain that anchors the protein complex to the parasite actin-myosin motor system, thus forming a bridge between the parasite and host cell (Soldati et al, 2001). Microneme secretion is induced by a transient increase in intracellular Ca^{2+} concentration in the parasite (Lovett et al, 2002).

1.6.1 MIC2 and M2AP

The MIC2/M2AP complex has been shown to be necessary for invasion (Huynh et al, 2003). Knockout studies have shown that MIC2 is necessary for invasion and mutants lacking MIC2 were deficient in host cell attachment, invasion, gliding motility and the ability to support

lethal infection. MIC2 has also been found to be necessary for proper trafficking of M2AP (Huynh & Carruthers, 2006). The NMR solution state structure of the β domain of M2AP has been solved and is shown in figure 1.6 (Huynh et al, 2015). MIC2 is the transmembrane member of the MIC2/M2AP complex, with an extracellular region composed of an integrin-like I-domain followed by five thrombospondin like repeats (TSRs) and a degenerate sixth TSR near the transmembrane domain. The C-terminal cytosolic end of MIC2 is thought to associate with the actin-myosin motor of the parasite, but the molecular details of this interaction have not yet been characterised.

M2AP is a soluble protein with an N-terminal propeptide, followed by a β domain, and a predicted coil domain at the C-terminus (Huynh et al, 2015). M2AP does not appear to be involved in adhesion but has been found to be necessary for proper trafficking and secretion of MIC2 (Harper et al, 2006). The degenerate sixth TSR domain has been found to interact with M2AP by mutation and deletion studies. The M2AP structure shows that the β domain of M2AP is very similar in structure to MIC1. The structure of MIC1 β domain is similar to the galectin fold found in carbohydrate binding proteins (Saouros et al, 2005). However, MIC1 does not contain a carbohydrate binding domain but a hydrophobic patch. Similarly, the β domain of M2AP has a protein-protein interaction surface instead of a canonical carbohydrate-binding surface. This hydrophobic patch present in M2AP has been found to be crucial for interaction with MIC2 and is shown in figure 1.6 (Huynh et al, 2015).

MIC2 has been found to interact with two host proteins – LAMTOR 1 (late endosomal/lysosomal adaptor, MARK and mTOR activator 1) and RNaseH2B (ribonuclease H2 subunit B). These proteins have been found to be involved in many cellular processes, such as lysosome maturation, signalling transduction, and RNA catabolism (Wang et al, 2014). Although the molecular basis of this mechanism is unknown, it has been suggested that MIC2 is involved in altering host signal transduction, in addition to binding host cells.

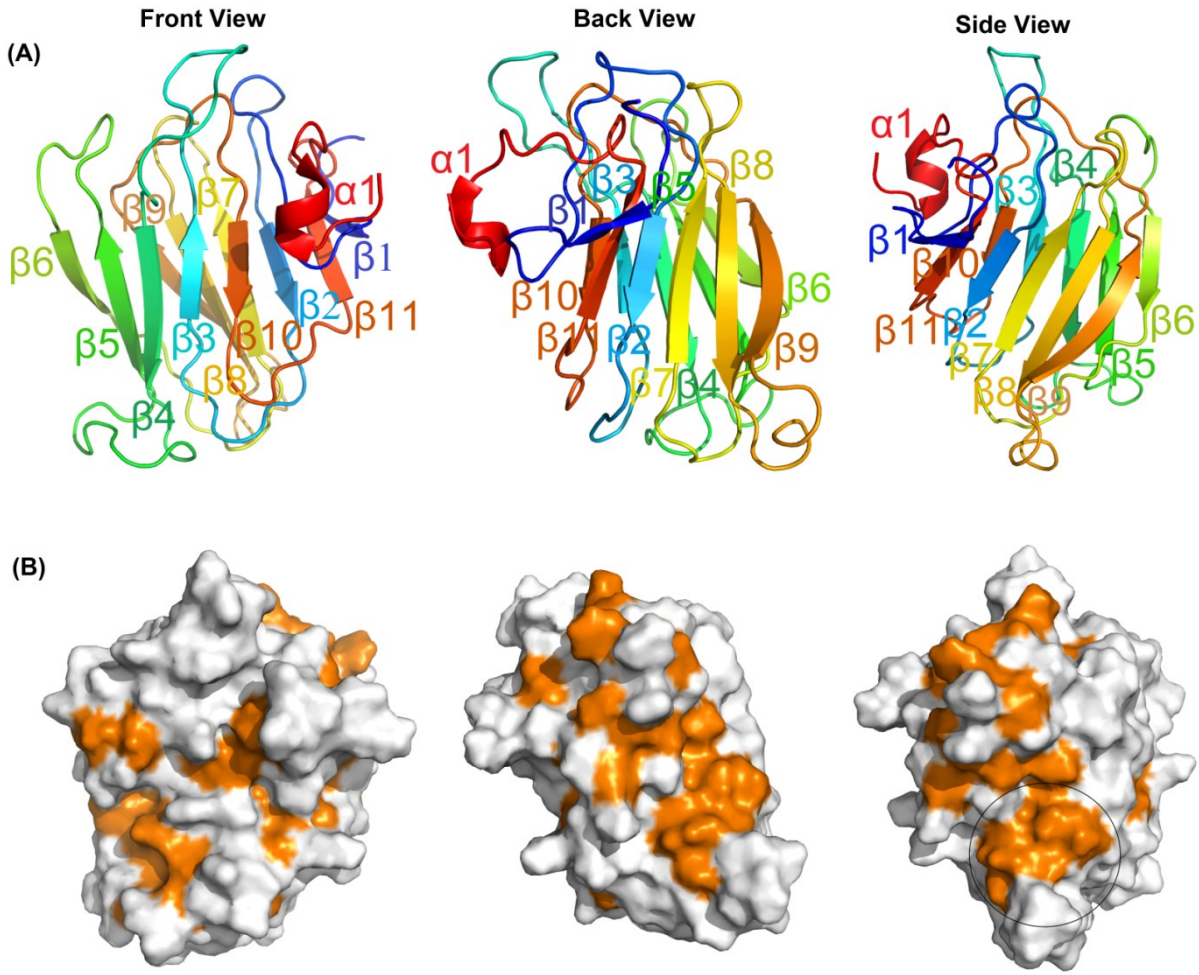


Figure 1.6: Figure showing (A) cartoon representation of the β domain of M2AP in rainbow colours from N to C terminus of the protein in three orientations and (B) the hydrophobic surface representation showing hydrophobic patches (orange) in the respective orientations. It has been found that the hydrophobic patch highlighted by the black circle is necessary to bind MIC2.

1.6.2 MIC1-MIC4-MIC6

Another microneme protein complex which has been extensively studied is the MIC1-MIC4-MIC6 complex. MIC1 and MIC4 are soluble adhesins whereas MIC6 does not have any adhesive functions. MIC6 has a transmembrane domain and acts as an escorter and mutagenesis studies have shown that MIC6 is necessary for proper trafficking of MIC1 and MIC4 (Reiss et al, 2001). MIC6 contains a transmembrane domain, followed by an acidic domain and three EGF-like domains. MIC4 is a soluble propeptide consisting of six Apple/PAN domains. It is processed extracellularly when the sixth Apple/PAN domain is cleaved off.

MIC1 consists of two MAR (Microneme Adhesive Repeat) domains characterised by X-ray crystallography (Garnett et al, 2009), followed by a galectin-like domain at the C-terminus which has been characterised by NMR (Saouros et al, 2005). One of the MAR domains is type I and the other one is type II. The C-terminal domain is a rigid β barrel consisting of two β sheets. This galectin-like domain does not have a carbohydrate binding domain but a hydrophobic patch as shown in figure 1.7 (Saouros et al, 2005). The hydrophobic patch from two MIC1 proteins stabilise the second and third EGF-like domains of MIC6 by interacting with them and is shown in figure 1.7 (Sawmynaden et al, 2008). The two MAR domains of MIC1 bind sialylated oligosaccharides and can differentiate between different linkages (Friedrich et al, 2010). The structure of MIC1 MAR domain bound to 3'SiaLacNAc1-3 has been solved by X-ray crystallography and is shown in Figure 1.7 (Garnett et al, 2009).

MIC4 interacts with terminal galactose residues via the fifth Apple/PAN domain. The solution structure of the fifth Apple/PAN domain of MIC has been solved by NMR and is shown in figure 1.7 (Marchant et al, 2012). Following this, TgSUB1 is secreted and it cleaves MIC4 between the fourth and fifth Apple/PAN domain releasing the galectin domain (Lagal

et al, 2010). It has been suggested that this would aid the parasite in evading the host immune response activated by the galectin (Marchant et al, 2012; Marchant et al, 2008).

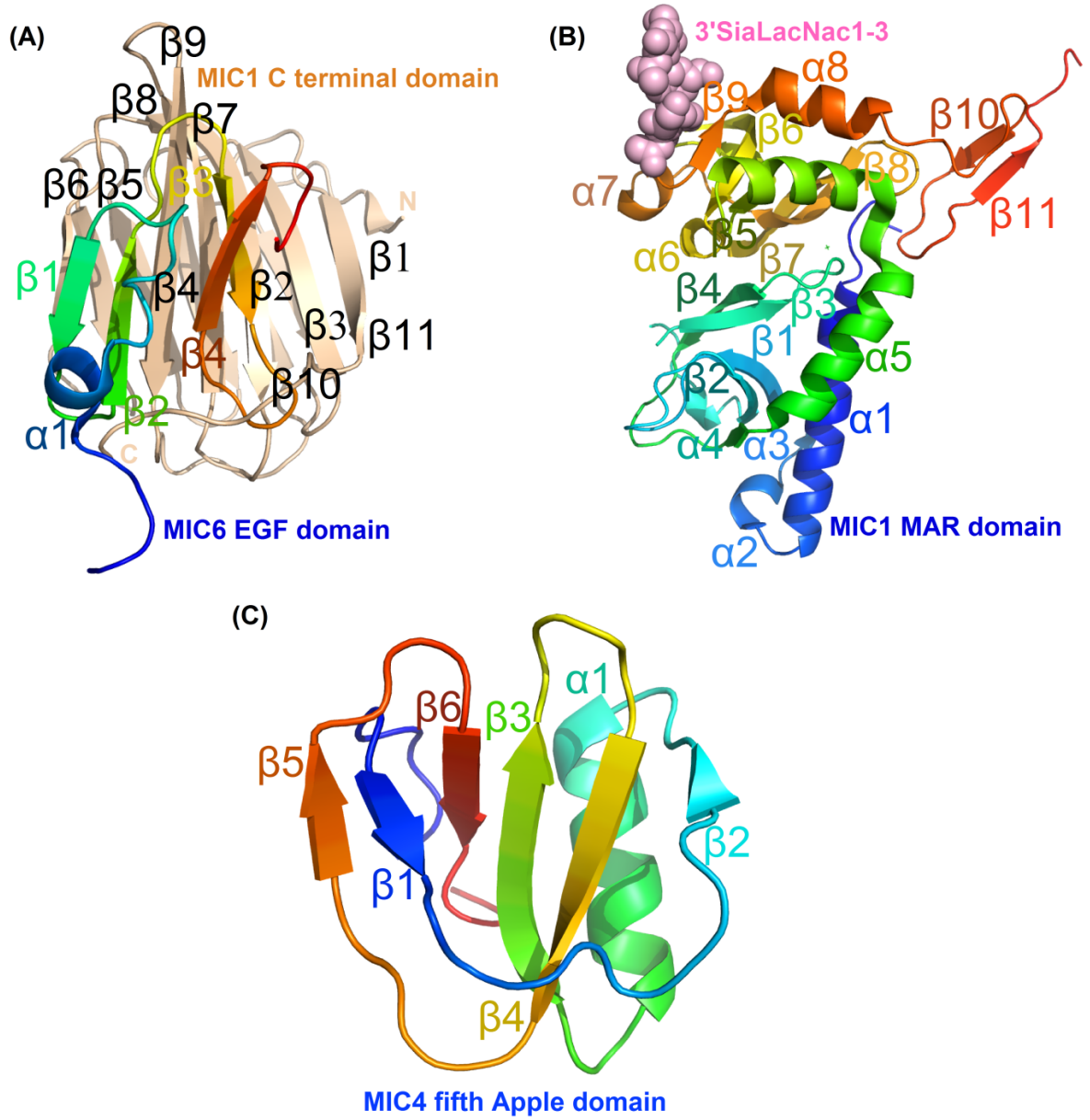


Figure 1.7: Figure showing (A) cartoon representation of the complex formed by MIC1 C terminal domain (wheat) and the MIC6 EGF domain (rainbow colours from N to C terminus) (B) cartoon representation of the MIC1 MAR domain (rainbow colours from N to C terminus) bound to 3'SiaLacNac1-3 (shown as pink spheres) and (C) cartoon representation of the fifth Apple domain of MIC4 (rainbow colours from N to C terminus)

1.6.3 Other microneme proteins

MIC5 is another microneme protein which has been found to have a role in inhibiting TgSUB1 (Saouros et al, 2012). The NMR solution state structure of MIC5 has been solved and it has been found that the structure of MIC5 is similar to that of subtilisin prodomains as shown in figure 1.8. Mutation studies have also shown that MIC5 contains a C-terminal flexible peptide which inserts into the active site of TgSUB1 inhibiting it (Saouros et al, 2012).

MIC3 is a soluble adhesive protein that has been found to form a complex with MIC8, an escorter protein. MIC3 is exocytosed from the micronemes and secreted all over the surface of the parasite. The structure of MIC3 has five partially overlapping EGF-like domains and a chitin binding-like domain, which can be involved in protein–protein or protein–carbohydrate interactions (Garcia-Réguet et al, 2000). MIC8 has a transmembrane domain and it interacts with MIC3 forming a complex in the transit during the secretory pathway. MIC8 is necessary for efficient invasion as it is required for the formation of the moving junction (Meissner et al, 2002).

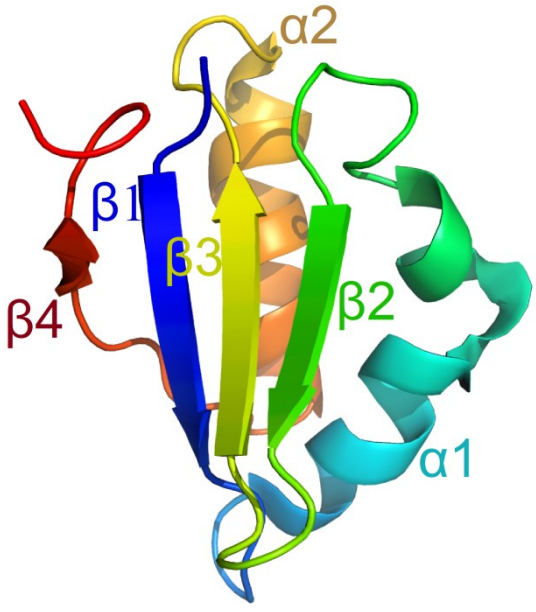


Figure 1.8: Figure showing cartoon representation of MIC5 in rainbow colours from N to C terminus of the protein

1.6.4 AMA1

AMA1 is a transmembrane protein which is secreted from micronemes and localised to the parasite membrane upon invasion. Structural studies of AMA1 homologues from *T. gondii*, *P. vivax* and *P. falciparum* have shown that the ectodomain of the protein has a hydrophobic trough (Tonkin et al, 2011). The crystal structure of the ectodomain of TgAMA1 was solved to 2 Å resolution and is shown in figure 1.9. The ectodomain consists of three domains – DI, DII and DIII. The DI domain consists of α helices, short β twists and coiled coils held together by three disulphide bonds. The DII domain consists of a loop region called the DII loop, followed by the DII core domain. The DII loop is stabilised by a disulphide bond and the DII core also contains a disulphide bond. The DII core domain is followed by a long linker that connects the DII domain to the DIII domain. The long linker facilitates the positioning of the DIII domain at the posterior end of the ectodomain. The DIII domain consists of three disulphide bonds that form a highly stable cysteine knot (Crawford et al, 2010).

1.7 Rhoptry and Rhoptry Neck Proteins

Rhoptries are bulb-shaped organelles which secrete proteins which are localised either to their duct shaped neck or the bulb (Louis & Kami, 2007). The proteins secreted from the neck are called rhoptry neck or RON proteins and those secreted from the bulb are called ROP proteins. The RON proteins are known to be mainly involved in the formation of the moving junction whereas the ROP proteins have a number of other functions including virulence and host cell signalling (Carruthers & Sibley, 1997). A complex of RON2, RON4, RON5 and RON8 proteins is necessary for successful invasion. Out of these RON proteins, RON8 is the only one which is conserved only in coccidian parasites. Mutagenesis studies have suggested that RON8 is necessary for stabilising the initial host-parasite interaction (Straub et al, 2011). RON2 is a membrane protein that inserts into the plasma membrane of the host cell with the

N-terminal end in the host cytoplasm and the C-terminal end forming a complex with AMA1 (Tyler & Boothroyd, 2011). RON4, RON5 and RON8 proteins are injected into the host cell bound to RON2 as a complex (Straub et al, 2011; Tyler & Boothroyd, 2011). The ectodomain of AMA1 binds RON2 and the C-terminal end of AMA1 interacts with the actin-myosin motor system within the parasite (Tyler & Boothroyd, 2011).

1.8 AMA1-RON2 complex

The crystal structure of AMA1 in complex with RON2 peptide has been solved to 1.7 Å resolution (Tyler & Boothroyd, 2011). This structure shows that the RON2 peptide interacts with AMA1 in a U shaped conformation. It has also been found that AMA1 undergoes a significant conformational change for the RON2 peptide to bind as shown in figure 1.9. In the unbound form, the DII loop in AMA1 is situated in a hydrophobic groove in the structure. In the bound form, the DII loop is displaced from this position exposing a basic binding interface for the acidic residues of RON2. The exposed binding surface complements RON2 structurally and electrostatically (Tyler & Boothroyd, 2011). Following the formation of the moving junction, the conoids in the parasite extrude, putting a large amount of stress on the AMA1-RON2 complex. The extrusion of the conoids is dependent on a rise in Ca^{2+} concentration (Mondragon & Frixione, 1996). It is suggested that the complementarity of the AMA1-RON2 binding interface, in terms of charge and conformation makes the interaction robust enough to withstand the stress of the extrusion (Tyler & Boothroyd, 2011).

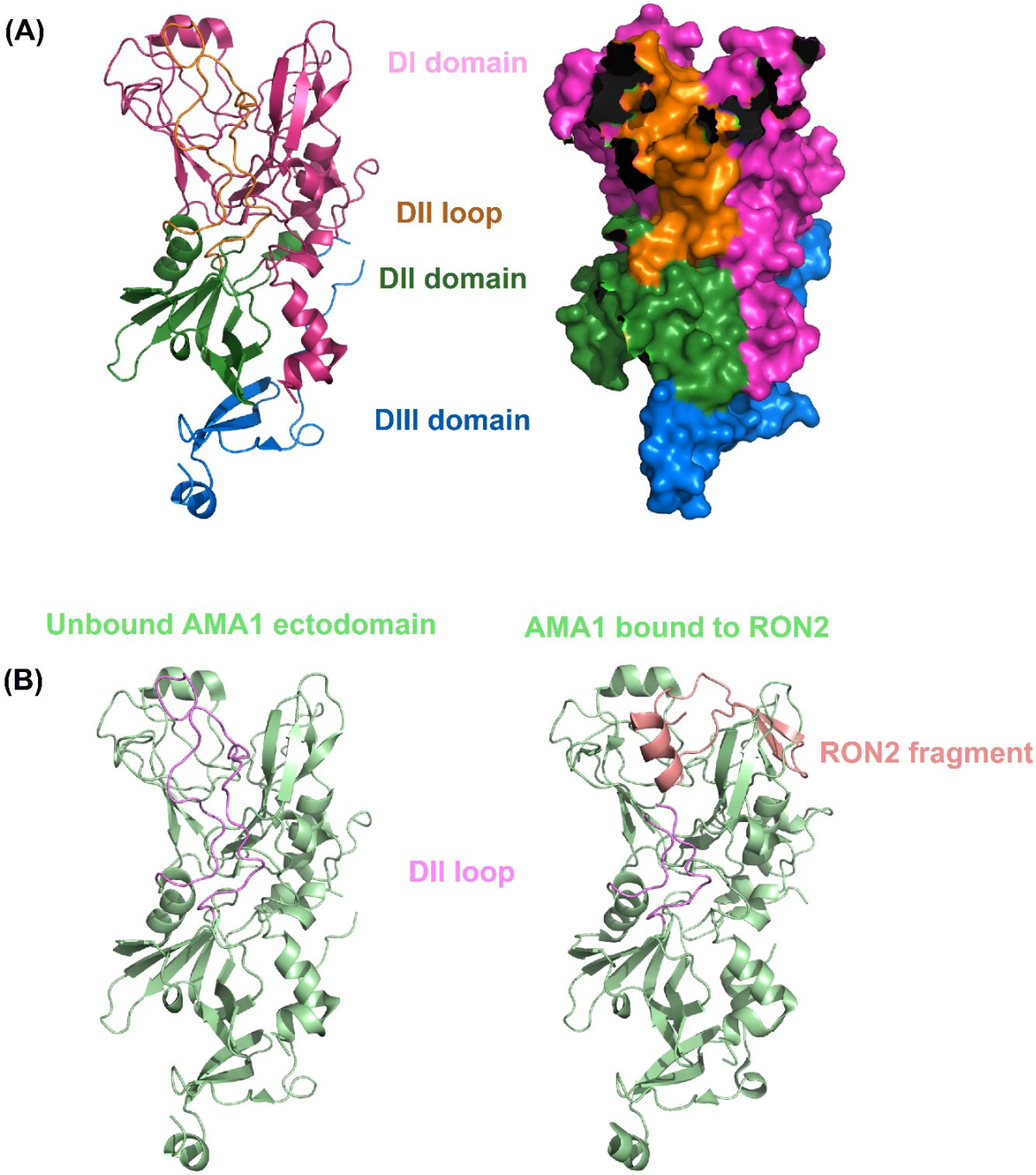


Figure 1.9: Figure showing (A) cartoon representation (left) and surface representation (right) of AMA1 ectodomain with DI domain in magenta, DII loop in gold, DII domain in green and DIII domain in blue and (B) cartoon representation of unbound AMA1 ectodomain shown in green and DII loop shown in purple and AMA1 ectodomain bound to RON2 fragment (pink)

1.9 *Eimeria tenella* - Life Cycle

The *Eimeria* species under study in this project is *Eimeria tenella.* Although *E. necatrix* is the most pathogenic parasite, *E. tenella* is also highly pathogenic in chickens as pathology is primarily associated with the second generation schizonts on account of their size and deep location within the lamina propia. Schizony in chickens has been found to occur in the crypts and lamina propria of the ceca and the Lieberkühn of the small intestine, respectively (Gerhold, 2014). Parasites are ingested by chickens as oocysts. Once the oocysts reach the gut of the chicken, sporozoites are released which invade the epithelial cells lining the intestine. Following this, the sporozoites undergo replication to form merozoites. Merozoites can invade fresh host cells and differentiate into male and female gametes. The gametes fuse to form a zygote which matures into an oocyst. Oocysts are then released in the faeces of the chicken (Fernando, 1990) as shown in Figure 1.10.

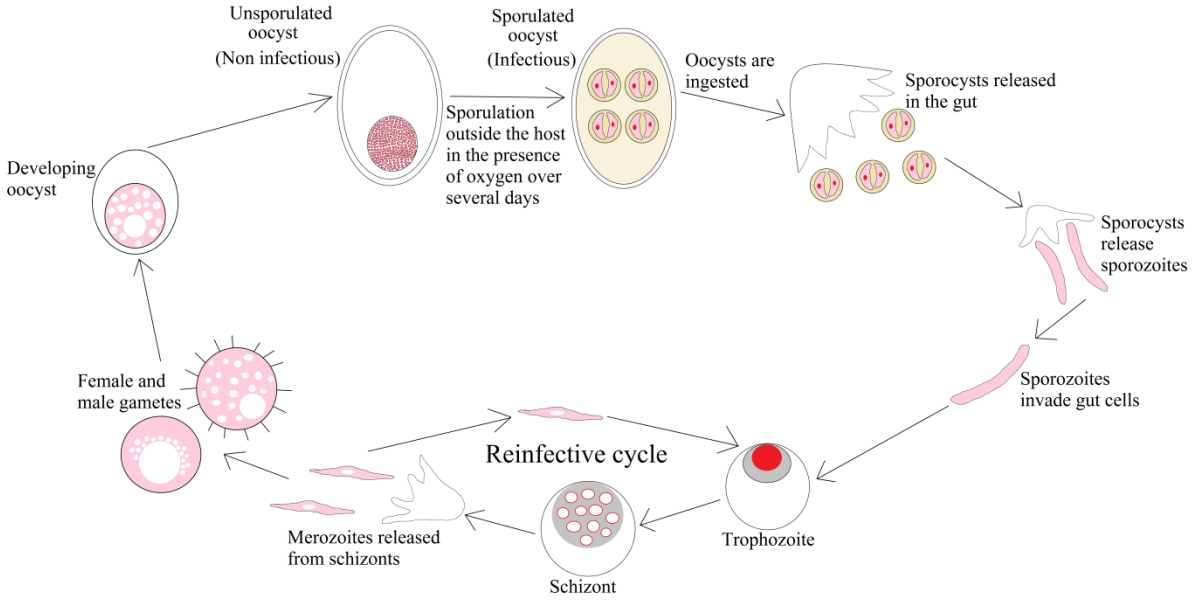


Figure 1.10: Figure showing the various stages in the life cycle of E. tenella within chickens. The stages of the parasite are labelled in the figure.

1.10 Plasmodium falciparum

1.10.1 *P. falciparum* – Life cycle

P. falciparum has a complex life cycle which involves asexual and sexual development both in the female *Anopheles* mosquito and within humans as shown in figure 1.11. The parasite is transmitted to humans by the vector, as sporozoites, during a blood meal. The sporozoites then enter the liver, invade and multiply within the hepatocyte cells in the liver. This stage of multiplication is called the pre-erythrocytic development. The parasites are released from the hepatocytes, as merozoites, into the bloodstream. The merozoites then invade the erythrocytes and multiply within the erythrocytes asexually. A fraction of these merozoites form gametocytes possibly in the bone marrow as the immature gametocytes are present in the bone marrow (Aguilar et al, 2014). The mature gametocytes are released into the blood stream and can be found in the peripheral blood stream. The parasites can be transmitted to the female *Anopheles* mosquito only as mature gametocytes, which occurs during the blood meal. Once the gametocytes enter the mosquito, the gametocyte either matures into a female macrogamete or into eight male microgametes. The gametes fuse within the mosquito midgut to form a zygote, which develops into a motile ookinete. The ookinete then penetrates the midgut wall and forms oocysts. The oocysts grow in size and burst releasing sporozoites. These sporozoites enter the salivary gland of the mosquito ready to be transmitted to humans during a blood meal, thus completing the cycle (Bousema et al, 2014).

1.10.2 *P. falciparum* – Erythrocytic Schizogony

P. falciparum undergoes two stages of development within humans – pre-erythrocytic and erythrocytic development. During the erythrocytic development stage, the merozoite goes through four stages – ring, trophozoite, immature and mature schizont as shown in figure 1.11. The ring or immature trophozoite is the first stage which then develops into a mature trophozoite. The trophozoite cannot digest the haem but digests the globin and the haem is

stored as a haemozoite crystal. The trophozoite then matures into a schizont. The morphology of these different stages are different and can be seen in Figure 1.11 (Petri et al, 2006).

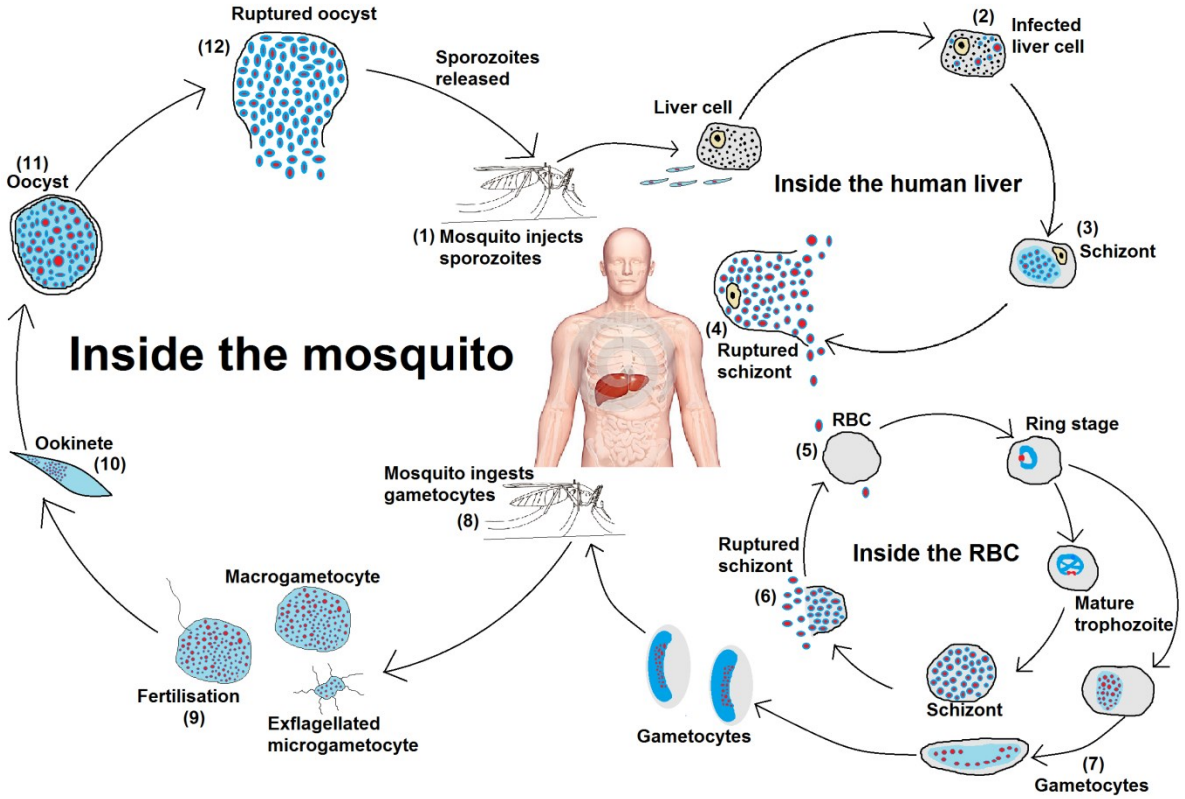


Figure 1.11: Figure showing the various stages in the life cycle of P. falciparum within the mosquito and humans. The stages of the parasite are labelled in the figure.

1.11 Vaccine development

Efforts are currently being made to develop vaccines against toxoplasmosis, coccidiosis and malaria. Most of the vaccine studies which are taking place are species specific without any cross reactivity. There are three types of approaches to develop vaccines – using live or attenuated parasites, DNA vaccines and protein vaccines. Although numerous studies have been performed in all three approaches, only the studies involving protein vaccines will be discussed in detail in this thesis.

1.11.1 Protein Vaccines

Toxoplasmosis in immunocompromised individuals is usually treated with anti-parasitic drugs like pyrimethamine, spiramycin and sulfadiazine. Since these drugs can cause toxic hypersensitivity reactions and are teratogenic, alternative methods to treat toxoplasmosis are becoming a rising necessity (Lim & Othman, 2014). It has been established that the IFN-γ secreting CD8+ cells are the major subset of cytotoxic T cells necessary for providing long term protective immunity against toxoplasmosis (Brown & McLeod, 1990; Gazzinelli et al, 1991). This suggests that any antigen that raises a Th1-type immune response is a good candidate to battle pathogenesis and provide protection against chronic infection.

The most important protein antigen which was initially used for vaccination studies was the SAG proteins. Immunisation of mice with SAG1, with QuilA (Khan et al, 1991) and cholera toxin (Debard et al, 1996) as adjuvants, showed reduced brain cysts after infection. QuilA is a saponin based adjuvant which induces strong cytotoxic CD8+ lymphocyte responses (Kensil, 1996). However, the use of a non-toxic adjuvant, labile enterotoxin, showed decreased protection and much lower reduction in brain cysts (Bonenfant et al, 2001). However, in recent years it has been found that using multimeric protein antigens provide better protection against infection. Recent studies have shown that the oral administration of TgMIC1, TgMIC4 and TgMIC6 as a combination vaccine is effective as it increases host survival rate and reduces brain cyst burden in mice. The same study also reports that mice immunised with these proteins elicit a strong Th1-type immune response and an equilibrated production of IL-10 (Pinzan et al, 2015). Another recent approach to developing an effective vaccine has been the use of short peptides of different parasite proteins in combination. A recent study has shown that the use of the T/B cell epitopes from TgAMA1 and TgRON2 as a combination vaccine in mice leads to increased host survival rate and reduced brain cyst burden. The study also reports that mice immunised with a combination of TgAMA1 and TgRON2 epitopes

showed a stronger immune response than the mice immunised with one of the two protein epitopes. The combination vaccine has been found to increase both Th1 and Th2 type immune responses (Zhang et al, 2015). Although many antigens that induce a Th1 immune response to the parasite have been identified, none of them have been able to provide sterile immunity against infection.

AMA1 has also been found to induce antibodies and protection against malarial infection in a strain-specific manner in rodent and monkey models. Thus, AMA1 is considered a potential blood stage malarial vaccine. Phase 2 trials of AMA1-Combination 1 (C1) malaria vaccine were performed in children in Mali. This vaccine is a combination of the ectodomains of AMA1 homologues from two divergent strains of *P. falciparum* - FVO and 3D7 – in equal amounts, adjuvanted with Alhydrogel. The results from this study showed that this vaccine is only moderately immunogenic and a more immunogenic formulation of the vaccine may be more likely to be protective (Sagara et al, 2009). Another recent study tested the efficacy of FMP2.1/AS01 vaccine in controlling parasite multiplication rate in mosquito-bite controlled human malaria infection (CHMI) models. FMP2.1 is a protein vaccine based on the sequence of AMA1 from the *P. falciparum* 3D7 strain. This vaccine also demonstrated no efficacy after blood stage CHMI (Payne et al, 2016).

1.11.2 DNA Vaccines

DNA vaccines against toxoplasmosis would be beneficial because they are much cheaper in terms of cost (Lim & Othman, 2014). DNA vaccines generally induce a Th1-type immune response in addition to an increase in IFN-γ and IL-2 production. The major disadvantage of using DNA vaccines is that the immune response induced by these vaccines is generally too weak in humans. Similar to proteins vaccines, DNA vaccines encoding a combination of multiple antigens have also been found to be more immunogenic than those encoding a single antigen. Studies have shown that the use of a multimeric DNA vaccine encoding the T cell

epitopes of a combination of micronemal proteins – MIC2, MIC3, MIC4, M2AP and AMA1 – showed 84% reduction in brain cyst burden and longer survival time in mice than DNA vaccine encoding a single antigen (Beghetto et al, 2005). Recent studies have also shown that DNA vaccine encoding 12 T and B cell epitopes from SAG1, GRA2, GRA7 and ROP16 sequences provide partial protection against acute infection in mice. This vaccine has been found to increase the Th1 type immune response. However, this vaccine did not provide complete protection (Cao et al, 2015). Another study has found that the use of an adjuvant with the DNA vaccine increases the efficacy of the vaccine. A DNA vaccine encoding MIC8 showed increased survival time in mice when administered with mIL-21 and mIL-15 cytokines than when administered on its own (Li et al, 2014; Liu et al, 2010). Although many more multi-epitope DNA vaccines have been found, sterile immunity against toxoplasmosis has not been achieved with DNA vaccines so far.

1.11.3 Live, vectored Vaccines

Another vaccination strategy used against toxoplasmosis is the use of live, vectored vectors as vehicles to carry and express antigenic epitopes from *T. gondii*. A number of attenuated vectors being used in various studies tend to be bacteria and viruses. This method of vaccination tends to be more effective due to the intrinsic adjuvant properties of the vectors. Studies have shown that delivering the recombinant GRA4 antigen in attenuated vaccinia vaccine provided 100% survival from infection and no brain cysts 45 days post infection in mice. On the other hand, vaccination with DNA vaccine encoding GRA4 antigen only provided 70% survival from infection in mice (Zhang et al, 2007).

Currently, the only successful vaccine against toxoplasmosis is Toxovax (Schering-Plough Animal Health Ltd, New Zealand). Toxovax is a live, attenuated vaccine which is used to vaccinate sheep against toxoplasmosis infections. *T. gondii* infection in sheep leads to congenital transmission resulting in abortions and stillbirths. Toxovax consists of a particular

strain of *T. gondii* which can only undergo limited multiplication in sheep and has lost the ability to form bradyzoites (Buxton & Innes, 1995; Buxton et al, 1991). This vaccine is administered to uninfected sheep prior to mating and induces protective immunity for 18 months following a single shot. The vaccine needs to be repeated after 2 years (HiszczyŃSka-Sawicka et al, 2014). Although this method of vaccination is effective, there are many safety concerns regarding the use of attenuated viruses and parasites as vaccines against toxoplasmosis. The danger of the attenuated parasite reverting back to the active form is an ever present risk.

1.12 Immune Mapped Protein 1

Immune Mapped Protein 1 (IMP1) is an immunogenic protein that was first identified in *Eimeria maxima.* It has been shown to confer immunity against *E. maxima* infections through unknown mechanisms (Blake et al, 2011). An orthologue of this IMP1 protein has also been recently identified in the *T. gondii* RH strain (TgIMP1), in addition to other apicomplexan parasites. Recent studies have shown that immunisation of BALB/c mice with the pcDNA-IMP1 from the RH strain of *T. gondii* has increased the survival time of mice challenged with the lethal *T. gondii* RH strain tachyzoites compared to the control mice that were immunised with the empty plasmid. pcDNA is a vector which is designed for high level, constitutive protein expression in mammalian cell lines (ThermoFisher, UK) The vaccine causes an increase in IFN-γ, IL-2, and IL-10 levels, which are indicative of humoral and Th1 type response. The same study has also shown that TgIMP1 is a 43 kDa protein that seems to localise at the parasite plasma membrane and it induces both cellular and humoral immune responses in mice (Cui et al, 2012). Predicted orthologues of this IMP1 protein have also been identified in the genome of *T. gondii* ME49 strain, which is one of the strains under study in this project.

Studies have shown that a fragment of IMP1 from *Eimeria tenella* Houghton strain fused with a molecular adjuvant produces a stronger cellular immune response against *E. tenella* infections in chickens in a species-specific manner. This further confirms that IMP1 is an antigenic protein which produces an immunogenic response in the host by some unknown mechanism (Yin et al, 2013). A recent study by the same group has shown that chickens immunised with a C-terminal fragment (AA151-AA387) of EtIMP1 (EtIMP1-C) showed reduced faecal oocyst shedding and diminished intestinal pathology in comparison to non-immunised chickens. The study has also confirmed that the efficacy of EtIMP1-C was the same as that of full length EtIMP1 suggesting that the C-terminal region is the immune-dominant region in the protein (Yin et al, 2014).

1.13 Project Aims and Objectives

The immunogenic properties of IMP1 made this protein an interesting candidate to study it further. The aims of this project are as follows:

- Clone, express and purify the IMP1 homologues from *T. gondii, E. tenella* and *P. falciparum* for structural studies.
- Study the structure of IMP1 using NMR and X-ray Crystallography. Solving the structure of IMP1 will allow us to compare and analyse the IMP1 structure with existing protein structures. The structure of IMP1 might also guide us towards its possible function.
- Few studies have reported that IMP1 proteins are localised to the inner leaflet of the plasma membrane. The immunogenic nature of the protein, however, suggests that the protein has to be exposed on the parasite surface. Since the IMP1 protein lacks a transmembrane domain, it was hypothesised that the protein might localise to the plasma membrane by interacting with a transmembrane protein. Perform localisation studies and identify potential binding partners of IMP1 in the host organism.

Chapter 2

Nuclear Magnetic Resonance

2.1 Structural Biology Techniques

Protein structure determination is an important step in understanding protein biology. The function of a protein is determined by the three dimensional structure of the protein. The three major techniques which are used to determine protein structure to atomic resolution are X-ray crystallography, Nuclear Magnetic Resonance (NMR) and Cryo-Electron Microscopy (Cryo-EM).

2.1.1 X-ray crystallography

X-ray crystallography is a technique that is used to solve the structure of macromolecular proteins to atomic resolution. In order to study a protein using this technique, high quality protein crystals are required. Protein crystallisation is generally achieved by bringing the protein solution to a super-saturation state. In the super-saturation state, the protein is "forced" out of solution to form crystals. Protein purity and homogeneity are important factors that determine the quality of the protein crystal (Drenth, 2010). Super-saturation can also be achieved by varying several conditions such as protein concentration, pH, temperature, salt concentration, buffers etc. When the protein is in the right condition, it forms crystals which can be used for structural studies (Abola et al, 2000). Advances in automation of high throughput protein crystallisation in the past decade allow us to screen a vast range of crystallisation conditions rapidly. Very low amounts of pure protein are required to screen each condition.

Once a protein crystal is obtained, the crystal is shot with an X-ray beam. The X-rays in the beam are diffracted elastically by the electron cloud surrounding the nuclei in the protein (Drenth, 2010). The intensity of the diffracted X-rays relies on the interference of the rays diffracted by the atoms in the protein. These diffracted rays are then recorded by a detector (as shown in figure 2.1.1) and are called the X-ray diffraction pattern. A backward transformation of the diffracted waves will provide us with information about the electron

density of the protein. The diffraction pattern provides information about the amplitude of the diffracted rays but not the phase. The inability to determine the phase of the diffracted waves from the diffraction pattern is termed the phase problem (Drenth, 2010).

The phase problem is usually solved by using Isomorphous Replacement (MIR and SIR), Molecular Replacement (MR) or Anomalous Dispersion (MAD and SAD). MR is used to calculate the phase when the structure of a close homologue of the protein is known. Isomorphous Replacement involves recording the diffraction pattern of crystals of the protein with heavy atoms in known positions. Anomalous Dispersion uses the native crystal but involves recording diffraction patterns with the same crystal at different wavelengths of the X-ray. In MIR and anomalous displacement, the heavy atoms and variations in the wavelength cause changes in the diffraction pattern, respectively. Comparison of the two diffraction patterns will allow determination of the phase of the diffracted waves (Drenth, 2010).

Although X-ray crystallography is an effective method for structure determination of proteins to atomic resolution, the technique requires a vast amount of pure protein to screen numerous crystallisation conditions prior to successful crystallisation. Crystallisation to produce protein crystals which diffract to high resolution is a limiting step and requires homogenous protein samples of high purity. The diffraction quality of a crystal also determines the resolution of the final structure. A good quality crystal can produce a high resolution structure. However, this technique does have its advantages. There is no size limit on the protein sample that can be studied using this technique.

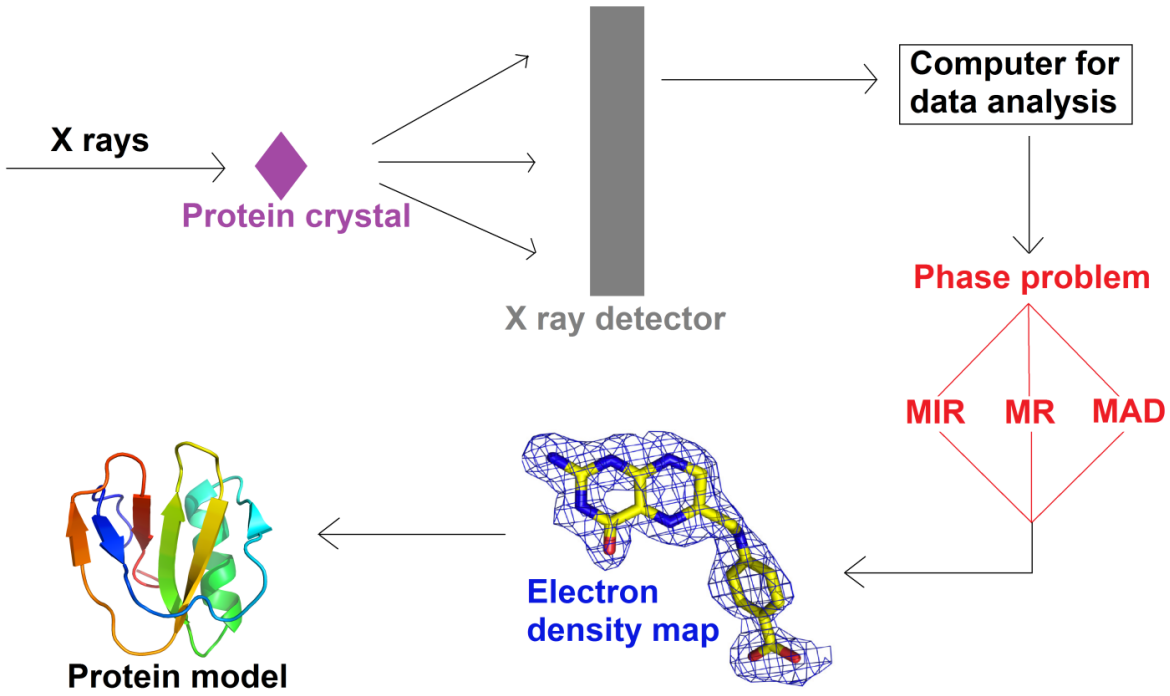


Figure 2.1.1: Figure showing an overview of the various stages in protein structure calculation using X-ray crystallography

The structure of a protein from a *Plasmodium* parasite was recently solved using X-ray crystallography. The symptoms that occur during malarial infections are caused by blood stage infection by merozoites. *Plasmodium* species merozoites invade human blood cells in a multistep process. The initial step in this invasion process is host selectivity by the parasite. PvRBP (Reticulocyte binding protein) is a family of adhesin proteins that are involved in host cell selectivity. The full length PvRBP2a protein is 286 kDa in size. The protein construct used for crystallographic studies included amino acids 169 to 1135 with a theoretical molecular weight of 98.4 kDa (Gruszczyk et al, 2016).

The structure of *Plasmodium vivax* RBP2a (PvRBP2a) protein was solved using the Single Isomorphous Replacement with Anomalous Scattering method (SIRAS) using iodine-derivatised crystal and refinement was performed against the native diffraction data set. The resolution of the final structure was 2.12 Å. The high resolution structure of the PvRBP2a protein solved by X-ray crystallography is shown in Figure 2.1.2. The protein contains two

sub-domains which are related to each other by pseudo two-fold rotational symmetry. The N-terminal domain consists of the 3_{10} helix, the beta hairpin, α1 to α3b helices and the C-terminal domain consists of α4 to α7 helices. Although the crystal structure contains two molecules in the asymmetric unit (Figure 2.1.2) with extensive contact over an area of 1349 $Å^2$, the same study goes on to show that the protein exists as a monomer in solution. This confirms that the dimeric existence of the protein is a crystal artefact which is not biologically relevant (Gruszczyk et al, 2016).

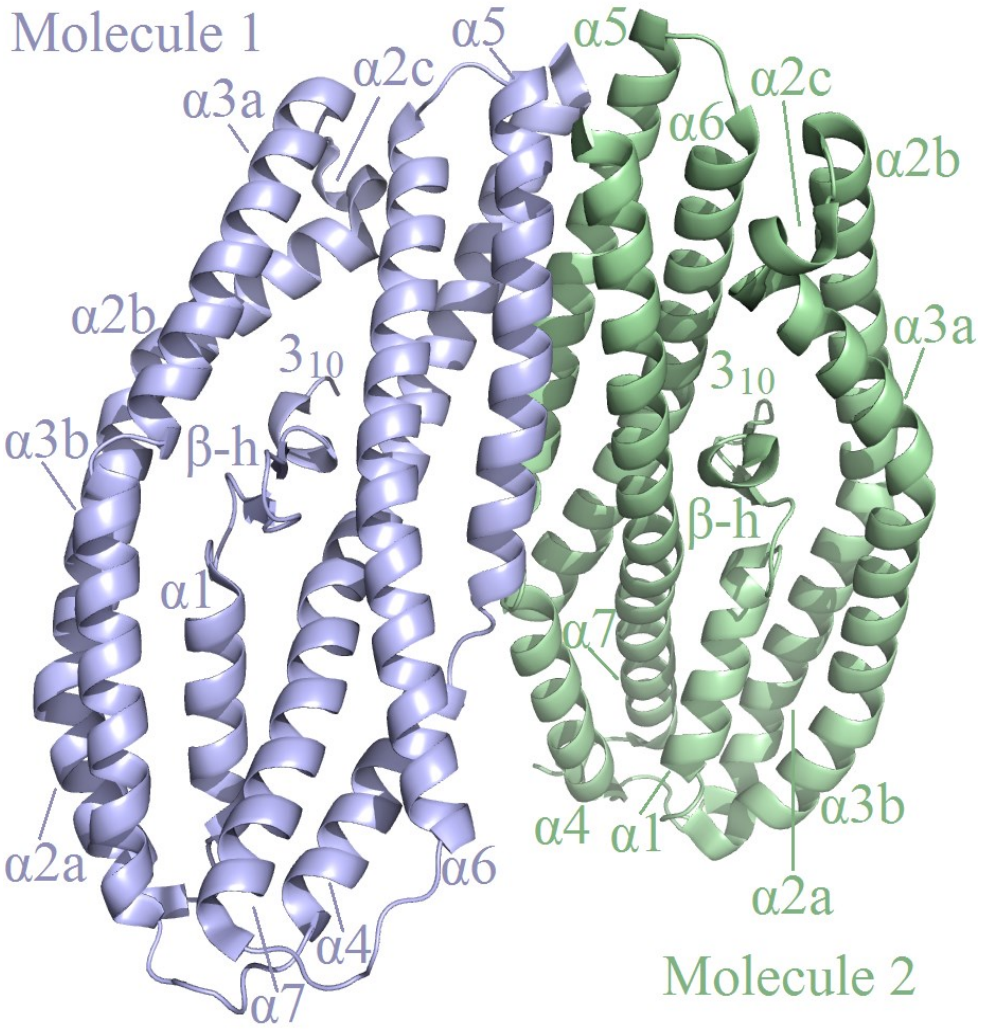


Figure 2.1.2: Figure showing a cartoon representation of the crystal structure of PvRBP2a protein with the secondary structure elements labelled. The beta hairpin in the structure has been labelled β-h. The two molecules found in the asymmetric unit of a crystal are shown. One molecule is coloured purple and the other coloured green. However, this dimer is not biologically relevant (Gruszczyk et al, 2016).

2.1.2 Cryo-Electron Microscopy

Cryo Electron Microscopy (Cryo-EM) is another technique used for protein structure determination. Although this technique has limitations due to difficulty in sample preparation and limitations in resolution, there is no limit on the size of the sample that can be studied using this technique.

Sample preparation is a crucial step in Cryo-EM as sample quality affects the quality of the data. Once a homogenous sample is obtained, the sample is prepared on an EM grid. These are copper grids which have a thin layer of carbon support film deposited on them. The protein sample rests on this carbon film. In Cryo-EM studies, a holey carbon grid is used which is a grid in which the carbon is deposited in such a way that there are holes in the carbon layer. The holes in the carbon layer allow the solvent to form pockets wherein the protein remains hydrated and can, therefore, be frozen in its native state (Thompson et al, 2016).

The sample is frozen rapidly so that the water that is present in and around the protein sample is frozen in a vitreous state. Ice formation can occur if the sample is not frozen fast enough or if the sample warms up, causing the water to devitrify. This can have detrimental effects on sample integrity and image quality. Cryo-immobilisation of the sample is usually achieved by blotting to produce a thin layer of liquid containing the sample and plunging it into a cryogen to freeze it. This is performed using commercially available freezing apparatus. It is essential to ensure that the sample is completely embedded in liquid in order to prevent radiation damage to the exposed regions in the sample (Thompson et al, 2016).

The protein sample is then bombarded with an electron beam and the electrons are scattered elastically. The last decade has seen numerous advances in the detectors used to record electron scattering. Photographic films were originally used as detectors in cryo-EM. However, they have recently been replaced by Monolithic Active Pixel Sensors (MAPS).

These sensors are also referred to as ‘direct electron detectors’ and the electrons pass through a thin semi-conductor membrane where they deposit energy. This is detected by electronics fabricated into the membrane. Behind the membrane, there is a support matrix which back-scatters the electrons into the membrane. This back-scattering leads to noise and can be eliminated by thinning the support matrix in a process known as back-thinning (Bai et al, 2015).

Commercial detectors which have a total thickness of 50 µm and allow all the electrons to completely pass through them are now available. Since low doses of electrons are enough to saturate these detectors, the dosage needs to be spread out over multiple low doses. Movie-mode functionality is also provided by the latest detectors. Development of improved detectors and improvements in image characterisation and alignment algorithms have made cryo-EM a viable structural biology tool (Bai et al, 2015). A schematic representation of the various stages of data analysis in cryo-EM is shown in figure 2.1.3 (Pintilie, 2009).

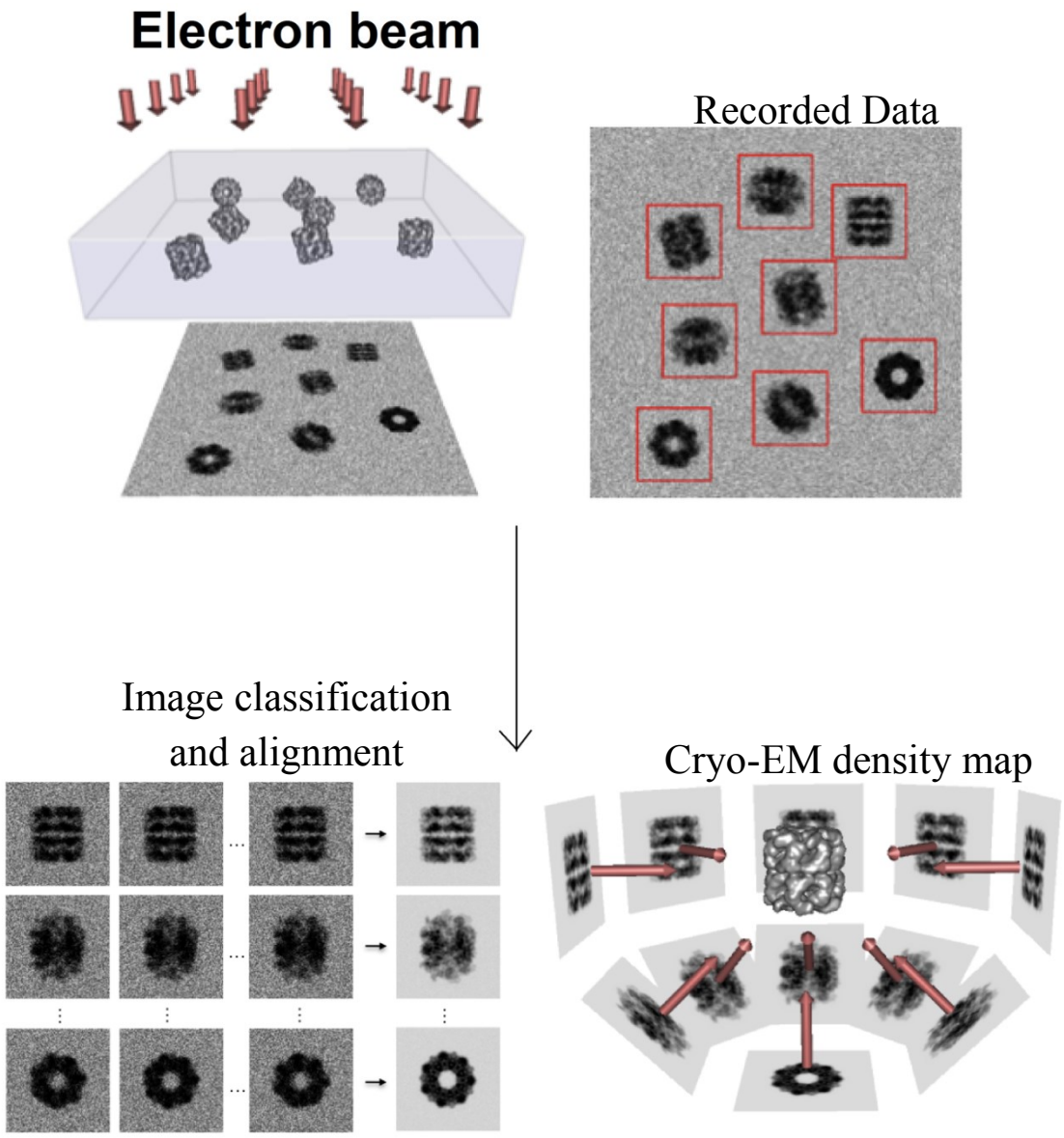


Figure 2.1.3: Figure showing an overview of various stages in protein structure calculation using Cryo - Electron Microscopy. Source: (Pintilie, 2009)

Cryo-EM is becoming a prominent tool in deducing the structure of large protein complexes. The structure of the 20S proteasome in *P. falciparum* bound to the Mor-WLW vinyl sulphone inhibitor has been solved using Cryo-EM and single particle analysis as shown in figure 2.1.4 (Li et al, 2016). The structure of the inhibitor bound human 20S proteasome was also solved previously using a similar strategy (da Fonseca & Morris, 2015). The *Plasmodium* proteasome is a protease complex, which is crucial for protein turnover within the parasite. Due to the rapid multiplication of the parasite within the host cell, rapid protein

turnover is required for parasite survival. This makes the proteasome an important anti-malarial drug target (Voges et al, 1999).

The study has also shown that the inhibitor under study can selectively inhibit the parasite proteasome with a minimum impact on the human proteasome. The basis behind selective inhibition was understood by analysing and comparing the high resolution structures of the parasite and human proteasomes. The structures revealed that steric restraints in the active site of the parasite and human 20S proteasome are the cause for selective inhibition (Li et al, 2016).

The molecular weight of the *P. falciparum* 20S proteasome – inhibitor complex is 760 kDa. The protein concentration that was used to prepare the sample was 0.1 mg/ml. It is evident that the structure of such a large complex cannot be solved by NMR due to the size restraints in this technique. Crystallography, on the other hand, requires large quantities of protein samples for screening and optimising crystallisation conditions. This high resolution structure is a fine example that shows that cryo-EM is a valuable tool that can be used for high resolution structure determination of large complexes which are not suited for structural studies by NMR and crystallography (Li et al, 2016). This study is an excellent example that demonstrates that cryo-EM is increasingly becoming an important structural biology technique that can be used to obtain high resolution structures.

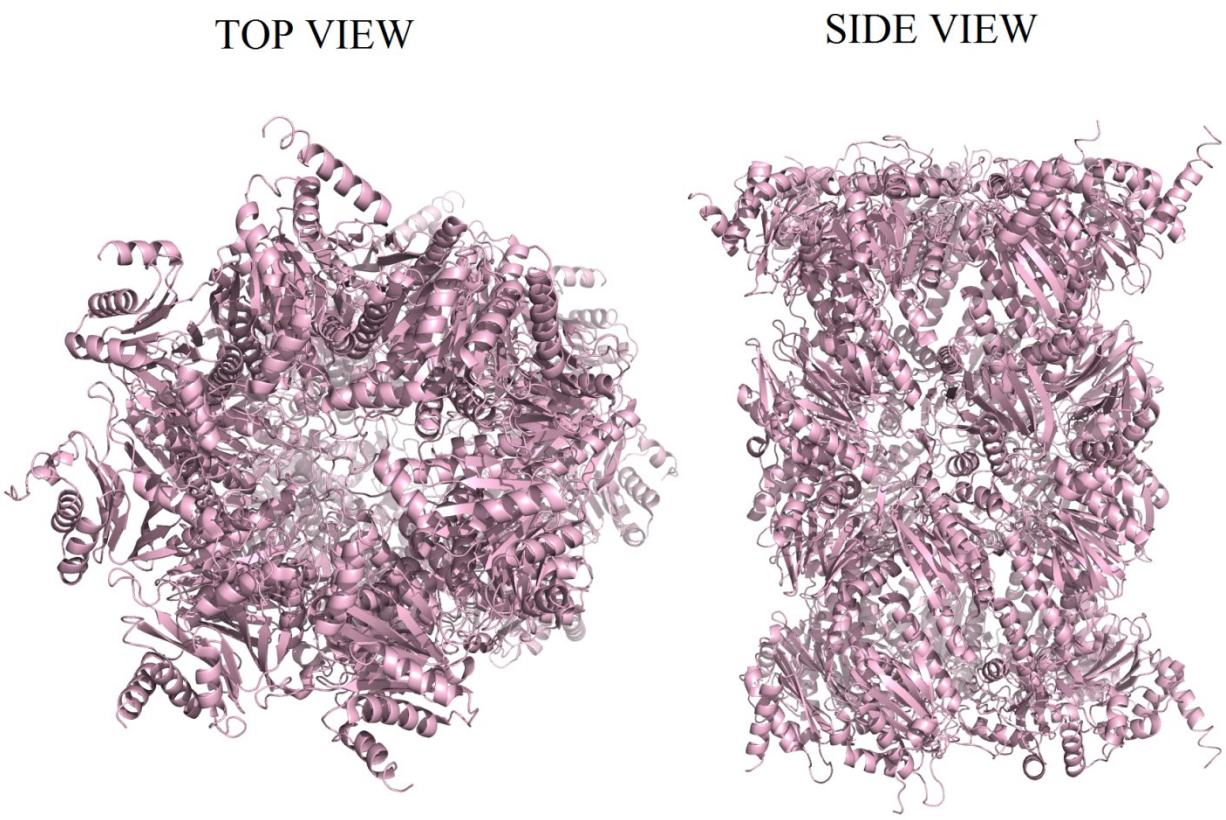


Figure 2.1.4: Figure showing two views of the cartoon representation of the cryo-EM structure of Plasmodium falciparum 20s proteasome (Li et al, 2016).

2.2 NMR for Protein Structure Determination

NMR is a powerful technique which allows the study of protein structures in solution. This technique provides information about the dynamics of the protein and allows us to study protein-protein interactions in addition to protein structure in atomic detail. The Matthews' lab mainly uses X-ray crystallography and NMR for structural studies. Since the protein structure was solved using NMR, this technique is described in detail. Most of the technical details about NMR have been taken from Keeler, 2010 and Gunther, 2013 unless otherwise specified.

2.2.1 Nuclear spin

The identity of an atomic nucleus is defined by two aspects – the atomic number and the mass number. These define the number of protons and neutrons present within a particular nucleus. The protons and neutrons within the nucleus have an intrinsic property called spin which gives rise to an overall nuclear spin. When the number of protons and the number of neutrons is even, their spins are paired and the nucleus has no overall spin. In atoms where the number of neutrons plus the number of protons is odd, the nuclei have a half integer spin. In atoms where the number of neutrons plus the number of protons is even, the nuclei have integer spin. For purposes of protein structure determination, we work with ½ spin nuclei - ^{1}H, ^{13}C and ^{15}N (Keeler, 2010).

2.2.2 The Classical model

A spin ½ nucleus is a spinning charge which creates a small magnetic field. When a spin ½ nucleus experiences an external magnetic field (B_0), the small magnetic field of the nucleus interacts with B_0. The torque from B_0 causes the spinning nucleus to precess around the axis of B_0 as shown in Figure 2.1.1. The frequency of this precession is called the Larmor frequency (ν_0). The strength of the magnetic field generated by a particular nucleus is termed magnetogyric ratio, γ and it is different for every nucleus. The larmor frequency of a nucleus

is directly proportional to B_0 and γ and can be calculated using the following equation (Gunther, 2013):

$$\nu_0 = \frac{\gamma B_0}{2\pi} \quad \text{(Equation 1)}$$

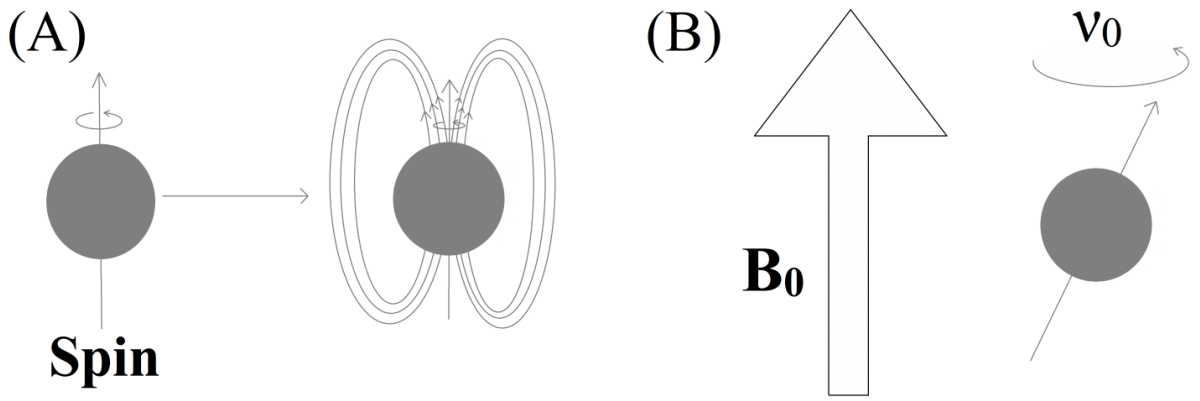


Figure 2.2.1: Figure depicting (A) magnetic field produced by a spinning nucleus and (B) precession of the spinning nucleus in the presence of an external magnetic field, B_0. Adapted from (Keeler, 2010).

2.2.3 The Quantum model

According to quantum mechanics, a nucleus with an overall spin of *I*, can have 2*I*+1 possible quantum states. Nuclei with ½ spin, therefore, have two possible quantum states. In the absence of a magnetic field, both states have the same energy in thermal equilibrium. When a spin ½ nucleus experiences an external magnetic field (B_0), the two states have different energy levels. The +½ state is aligned with the axis of B_0 and is the lower energy state (α) and the -½ state is aligned against B_0 and is the higher energy state (β). In the absence of B_0, the nuclei are in random orientations as there is no preferable one. In the presence of B_0, when the sample is in thermal equilibrium, the population of nuclei in the α state is slightly more than the population in the β state as shown in Figure 2.1.2. The energy difference between the two states can be calculated using the following equation derived from equation 1:

$$\Delta E = h\nu_0 = h\frac{\gamma B_0}{2\pi} \quad \text{(Equation 2)}$$

where h is Planck's constant (Gunther, 2013).

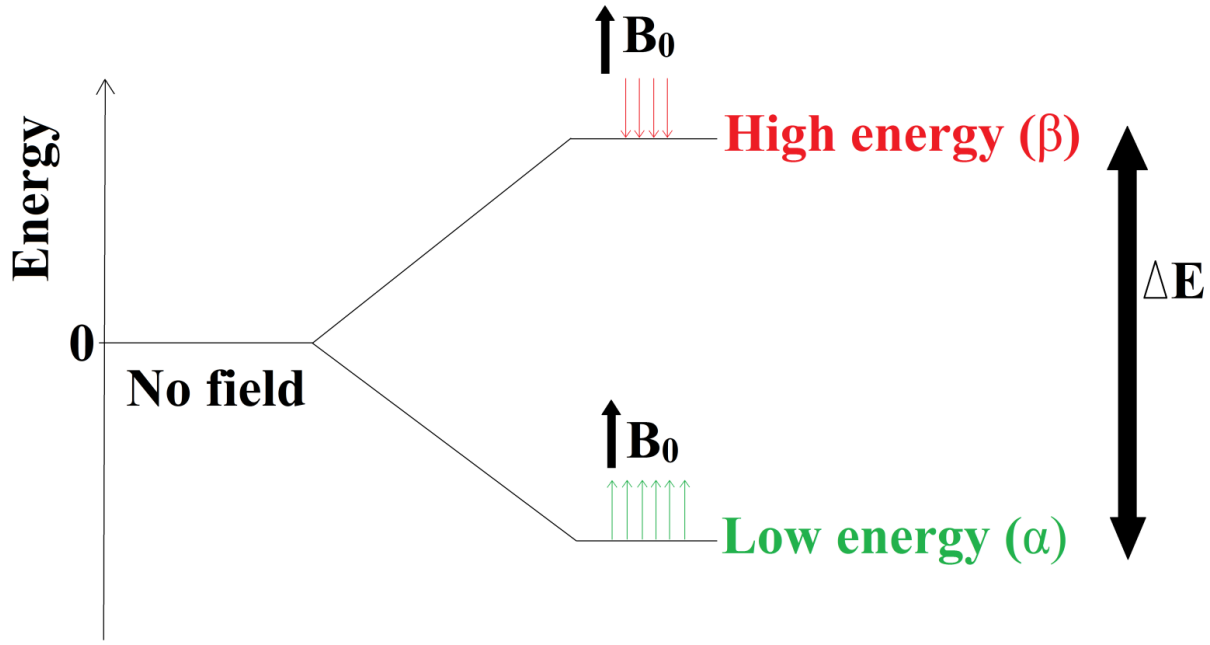


Figure 2.2.2: Figure showing a depiction of the population distribution of nuclei in an energy diagram in the presence of an external magnetic field, B_0. Adapted from (Gunther, 2013).

2.2.4 Bulk Magnetisation

In the absence of a magnetic field, the nuclei do not have a preferential orientation. When a sample is placed in a magnetic field, the nuclei preferentially adopt the lower energy state by aligning with B_0 as shown in Figure 2.1.2. This preferential alignment occurs slowly over time and the net magnetisation of the sample is aligned along the direction of the B_0, which is along the z axis. This is called the bulk magnetisation or net magnetisation of the sample.

Radio frequency pulse

The detector that records the signal in NMR experiments is a small coil that is placed along the x-y plane. In order to record a signal, the net magnetisation needs to be in the x-y plane. The detector coil applies a radiofrequency (RF) pulse that rotates the net magnetisation from the z axis by 90°, aligning it along the y axis. When the pulse is removed, the net magnetisation precesses in the x-y plane around the B_0 axis (z axis) at the larmor frequency.

The frequency of precession is the signal that is recorded by the detector. The precessing net magnetisation vector induces an oscillating voltage in the receiver coil which is detected on an oscilloscope as free induction decay (FID). This process has been shown in Figure 2.1.3. Fourier transform, which is a mathematical function, converts the time domain in the FID to a frequency domain in the NMR spectrum. This whole experiment can be repeated as many times as necessary to obtain a stronger signal (Gunther, 2013).

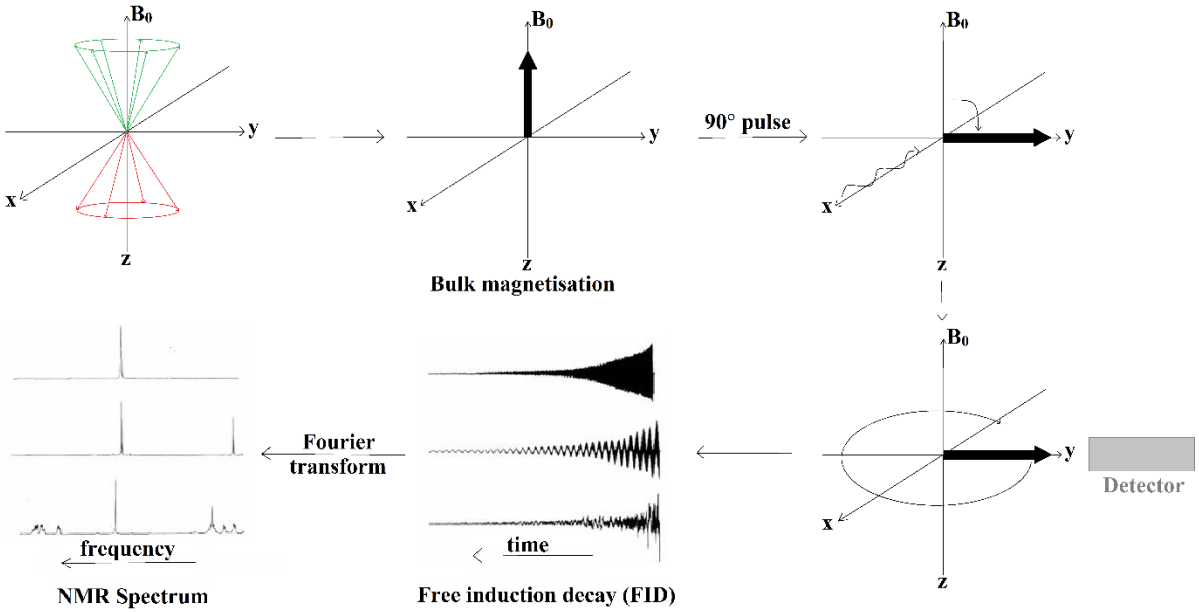


Figure 2.2.3: Figure depicting the process of developing free induction decay (FID) from an NMR sample. The bulk magnetisation of the sample is shown in the second part of the figure. The application of the 90° radiofrequency pulse rotates the bulk magnetisation to the y axis. Precession of the bulk magnetisation in the x-y plane is recorded by the detector as a FID. The FID is converted to an NMR spectrum through Fourier transform. Adapted from Gunther, 2013.

2.2.5 Chemical Shift

The signal that is detected in NMR experiments is the frequency of precession of the net magnetisation of the sample. Based on equation 1, this resonant frequency of a particular nucleus in the sample is dependent on B_0. This means that the resonant frequency of a nucleus will not be consistent when the data is recorded in magnets of different strengths. In order to make the measurement independent of B_0, the position of the signal from a nucleus is

expressed relative to the signal from a universally accepted standard reference compound. This value is called the chemical shift, δ, and the unit of chemical shift is ppm (parts per million). The standard that is used for ^{1}H and ^{13}C is Tetramethylsilane (TMS). The position of a particular peak in a spectrum is calculated using the following equation:

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^{6} \quad \text{(Equation 3)}$$

where ν is the frequency of the nucleus of interest and ν_{ref} is the frequency of the standard reference compound. Based on equation 3, the peak of the reference compound will always appear at 0. Since δ is a ratio between two frequencies, δ of a nucleus is no longer dependent on B_0. The 10^6 in the equation allows for the unit to be ppm. δ of a nucleus is affected by shielding and anisotropy effects (Gunther, 2013).

Shielding effects

As established by equation 1, the resonant frequency of a nucleus is dependent on the B_0 experienced by the nucleus. In an atom, the nucleus is surrounded by a cloud of electrons which circulate around the nucleus. Since the electrons are also charged particles, they also generate their own magnetic field in B_0 which alters the magnetic field experienced by the nucleus that they surround. This effect caused by the electrons is called the shielding effect. This effect alters the magnetic field experienced by the nucleus as follows:

$$B_{eff} = B_0(1 - \sigma) \quad \text{(Equation 4)}$$

where B_{eff} is the effective magnetic field experienced by the nucleus and σ is the shielding constant which has units of ppm. This alters equation 1 as follows:

$$\nu_0 = \frac{\gamma B_0(1-\sigma)}{2\pi} \quad \text{(Equation 5)}$$

Based on equation 3, a change in the resonant frequency of the nucleus will affect the δ of the nucleus. This variation in δ allows us to identify and differentiate nuclei present in different environments in the sample (Gunther, 2013).

Anisotropic effects

As established already, the electronic environment around a nucleus is anisotropic which means that the chemical shift of the nucleus varies depending on the orientation of the molecule with respect to B_0. In solution state NMR, anisotropic effects are averaged out by the fast tumbling rate of the molecules due to Brownian motion. This averaging leads to each nucleus having one sharp isotropic peak in the NMR spectrum. In large molecules, the tumbling rate is slower than in smaller molecules. In such cases, chemical shift anisotropy leads to each nucleus having more than one chemical shift which is observed as a broad peak on the NMR spectrum as shown in Figure 2.1.4. Chemical shift anisotropy is one of the reasons for the size limitation in NMR experiments (Gunther, 2013).

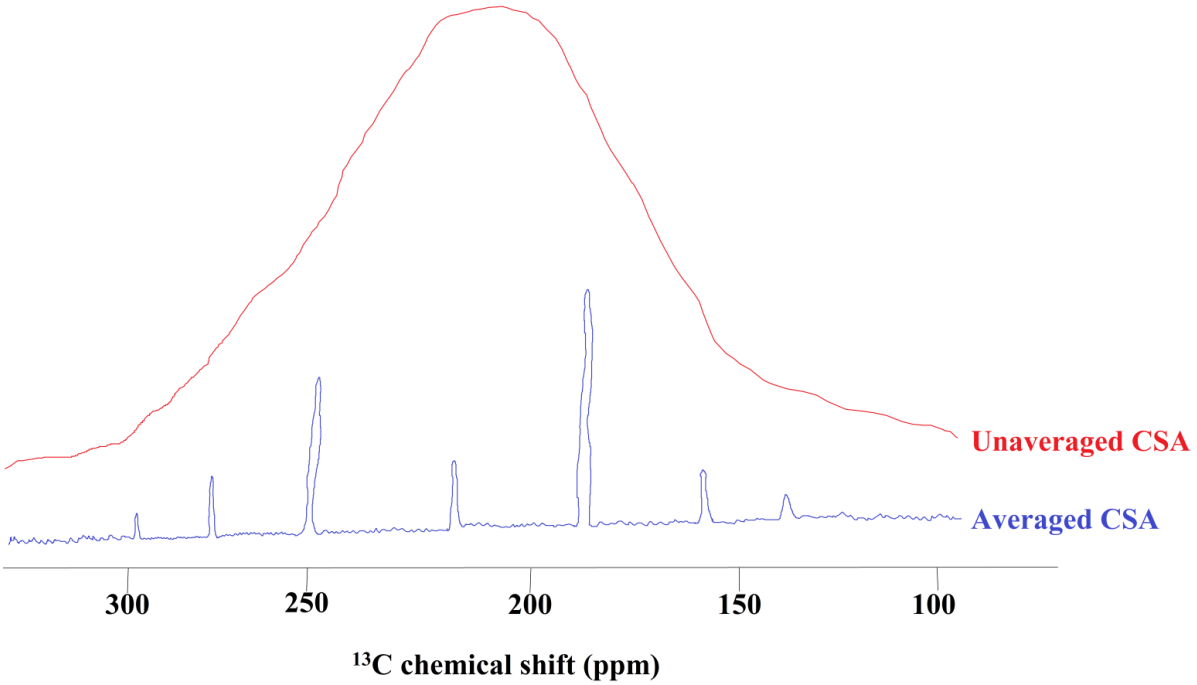


Figure 2.2.4: A schematic showing the NMR spectra produced by a sample when CSA is not averaged (red) and when the CSA is averaged (blue).

2.2.6 Coupling

Another aspect that a spinning nucleus is affected by is coupling. Coupling occurs when a spinning nucleus is in close proximity to another spinning nucleus. When two spins (spinning nuclei) are in close proximity such that the energy state of one nucleus affects the energy state of the neighbouring nucleus, the two nuclei are coupled. There are two mechanisms by which two nuclei can be coupled – Scalar coupling and dipolar coupling.

Scalar coupling

Scalar coupling between nuclei is mediated by chemical bonds. When two spins are connected by chemical bonds, we see multiplets in the NMR spectrum. If two spin-half nuclei are coupled, each of the individual peaks of the two nuclei becomes a doublet. Each peak is split into two peaks which are at equal distances from the chemical shift of the nuclei. The distance between the two peaks is called the coupling constant, *J*. Each peak is at a distance of J/2 from the chemical shift of the nucleus. For nucleus 1, one peak corresponds to when

nucleus 2 is in a spin "up" state and the other peak corresponds to when nucleus 2 is in a spin "down" state. Similarly, for nucleus 2, one peak corresponds to when nucleus 1 is in the spin "up" state and the other peak corresponds to when nucleus 1 is in the spin "down" state. This phenomenon has been shown in figure 2.1.5. When nucleus 2 is in the "up" state, the B_{eff} experienced by nucleus 1 is more than B_0 leading to a higher frequency signal leading to a higher chemical shift for nucleus 1 than expected. The reverse of this phenomenon happens when nucleus 2 is in the "down" state. The B_{eff} experienced by nucleus 1 is lower than B_0 leading to a lower frequency signal and a lower chemical shift than expected. The same phenomenon occurs to nucleus 2 depending on the spin state of nucleus 1 (Keeler, 2010).

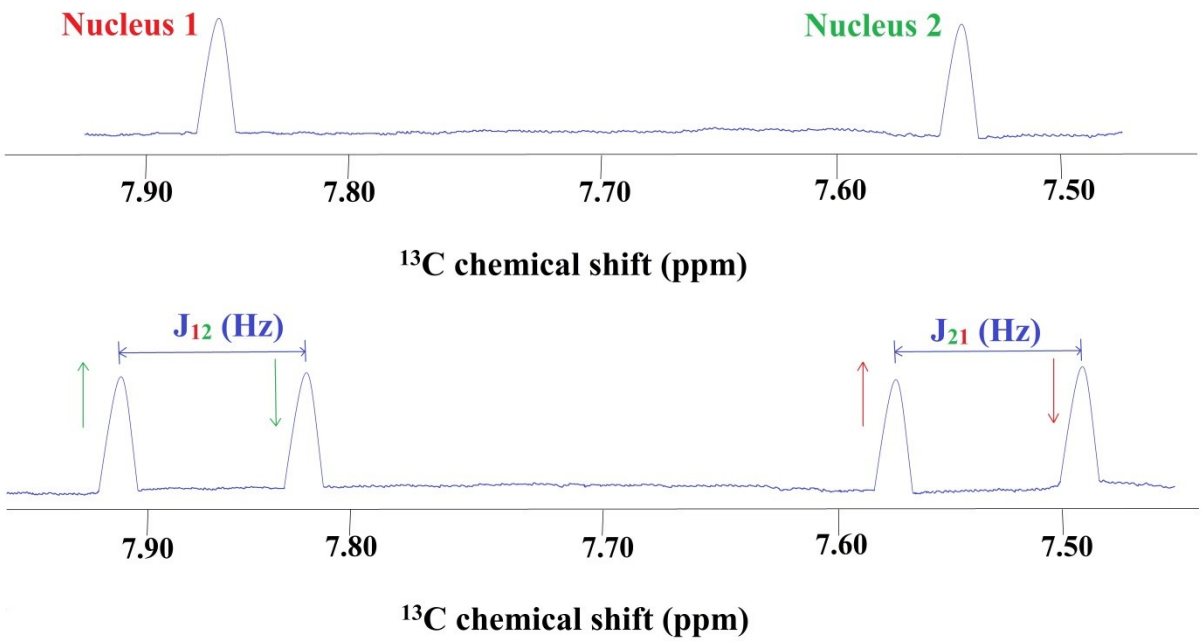


Figure 2.2.5: Schematic representation of the NMR spectrum of two nuclei which are decoupled (top) and which are J coupled (bottom). The green arrows represent the spin orientation of Nucleus 2 and the red arrows represent the spin orientation of Nucleus 1. The different orientations of the nuclei can lead to peak splitting. Adapted from Keeler, 2010.

Dipolar coupling

Dipolar coupling is an effect that happens through space. When two spins are in close proximity in space, the magnetic field of one spin affects the magnetic field experienced by

the other spin. Similar to scalar coupling, this depends on the orientation of one spin relative to the other. The strength of dipolar coupling between nuclei is affected by three factors:

- The distance between the two spins, r. The strength of dipolar coupling is inversely proportional to r^3
- The magnetogyric ratio, γ, of both the nuclei
- The direction of the axis that connects the two nuclei relative to the axis of B_0

Dipolar coupling is only effective up to a distance of 5Å. In solution state NMR, dipolar coupling is usually not observed as it is averaged out by fast tumbling rates due to Brownian motion (Keeler, 2010).

2.2.7 Relaxation mechanisms

Relaxation is the process by which the net magnetisation of the sample returns to equilibrium. At equilibrium, the net magnetisation is aligned along the z axis and is stationary. There are two mechanisms of relaxation that the sample undergoes – longitudinal relaxation and transverse relaxation.

Longitudinal relaxation

The return of the net magnetisation to its equilibrium along the z axis is called longitudinal relaxation. Within the NMR sample, there are numerous spin nuclei which generate their own magnetic field by spinning called the local field. This means that every spin in the sample experiences a magnetic field from numerous nuclei which are in close proximity to this spin. Therefore, the effective magnetic field experienced by a particular spin in the sample at any particular time is variable. When a spin experiences this varied magnetic field, it leads to changes in orientation of the spin. Overall, this could lead to random reorientations of spins at random times within the sample. When the sample is at equilibrium, the net magnetisation of the sample is aligned along the z axis. When a RF pulse is applied, the net magnetisation is

rotated to the y axis. The RF pulse increases the energy in the sample as the number of spins oriented along the axis of B_0, the low energy state, is decreased. In order to return to equilibrium, the system needs to lose energy. This loss of energy occurs by interaction between the spin energy of local fields and energy of thermal molecular motion. The two energy reservoirs come to thermal equilibrium with one another. This method of relaxation is called longitudinal relaxation or spin-lattice relaxation. The rate of longitudinal relaxation is called T_1 (Gunther, 2013).

Transverse relaxation

The decay of transverse magnetisation (along the y axis) to zero to reach equilibrium is called transverse relaxation. When the net magnetisation is rotated from the z axis to the y axis, net magnetisation is referred to as transverse relaxation. This magnetisation needs to return to equilibrium before the net magnetisation can return to equilibrium along the z axis. The transverse magnetisation is coherent following the 90° pulse and this coherence is lost to reach equilibrium. The loss of coherence occurs due to inhomogeneity in the magnetic field and interactions between spins due to variations in the local field. This loss of coherence does not occur by spin-lattice relaxation. The transverse magnetisation returns to its equilibrium value of zero by exponential decay. The rate of transverse relaxation is called T_2 (Gunther, 2013).

2.2.8 Nuclear Overhauser Effect (NOE)

Nuclear Overhauser effect or NOE is the enhancement or depletion of signal of a spin in a dipolar coupled spin system through magnetisation transfer via cross relaxation. The intensity of NOE is inversely proportional to the sixth power of the distance between the spins and is effective when the distance between the two spins is less than 5 Å.

NOE transitions

When a spin (A) is in close proximity to another spin (B), saturation of B by low power irradiation will perturb the population distribution such that the populations of B_α and B_β are equal. In order to return to equilibrium, magnetisation is transferred from B to A via cross relaxation. This magnetisation transfer leads to enhancement or depletion of signal from A. This effect is called the Nuclear Overhauser Effect (NOE).

There are two pathways through which cross relaxation can occur – double quantum transition and zero quantum transition (Figure 2.1.6). If both A and B are in the β state, cross relaxation causes both spins to be flipped to the α state, this transition is called the double quantum transition (W_2). This transition is possible when the neighbouring spins are tumbling at the sum of the frequency of the two spins ($\nu_A+\nu_B$) which is approximately twice the larmor frequency. This is a significant pathway in small molecules (MW < 1000 Da) as they tumble rapidly. The second scenario is when A is in the α state, B is in the β state and cross relaxation causes A to be flipped to the β state and B to be flipped to the α state. This transition is called the zero quantum transition (W_0). This occurs when the tumbling rate is close to the difference in frequency of the two spins ($\nu_A-\nu_B$) and is usually observed with large molecules as they tumble slowly.

Double quantum relaxation leads to an increase in the population difference of A following relaxation, which causes an enhancement in the signal from A. On the other hand, zero quantum relaxation leads to a decrease in the population difference of A following relaxation, which causes a decrease in the signal intensity from A. Relaxation can also occur by spin-lattice relaxation mechanisms via single quantum transitions (W_1). W_1 transitions decrease the intensity of NOEs and need to be minimised when recording NOE data (Keeler, 2010).

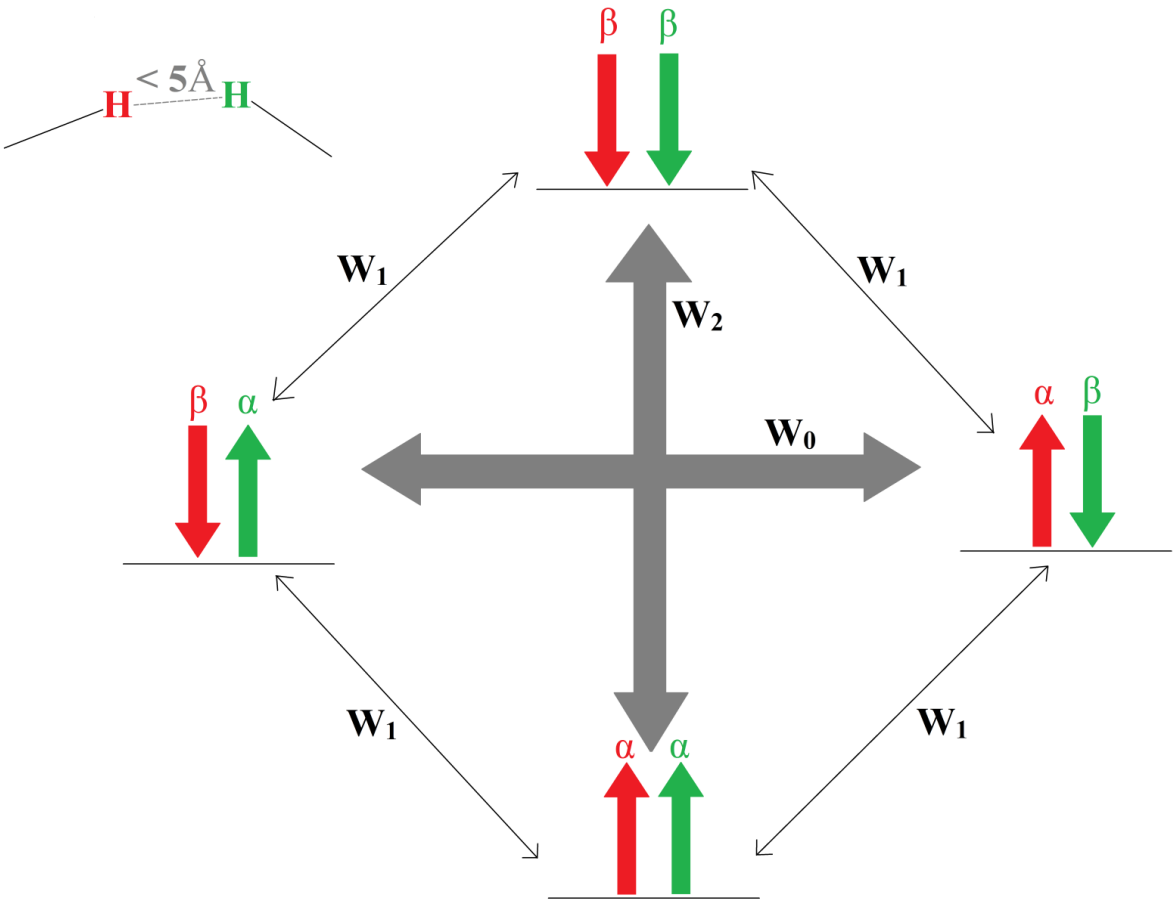


Figure 2.2.6: Figure depicting the various transitions of two nuclei which are less than 5Å apart in space. Nucleus 1 is depicted in red and nucleus 2 is depicted in green. The spin states are labelled in the figure. W_0 represents zero transition, W_1 represents single transition and W_2 represents double transition. Adapted from Keeler, 2010.

2.3 NMR Experiments

NMR experiments can provide vital information about the folded state and structure of a protein in solution. Different experiments provide different levels of information about the protein.

2.3.1 1D NMR

A 1D NMR experiment consists of two phases – preparation and detection as shown in Figure 2.2.1. In a 1D experiment, preparation involves applying a 90° pulse to the sample along the x-axis. This rotates the net magnetisation of the spins from the z-axis to the y-axis. This magnetisation rotates around the B_0 axis (z-axis) and this precession is recorded as a signal and forms the detection phase of the experiment. The magnetisation eventually loses coherence due to T_2 relaxation, leading to decay of the signal. This is why the recorded signal is referred to as the free induction decay (FID). The experiment is repeated numerous times and the signal is averaged to increase the signal to noise ratio. The FID signal is then subject to Fourier transform to obtain the 1D NMR spectrum of the particular nucleus (Gunther, 2013).

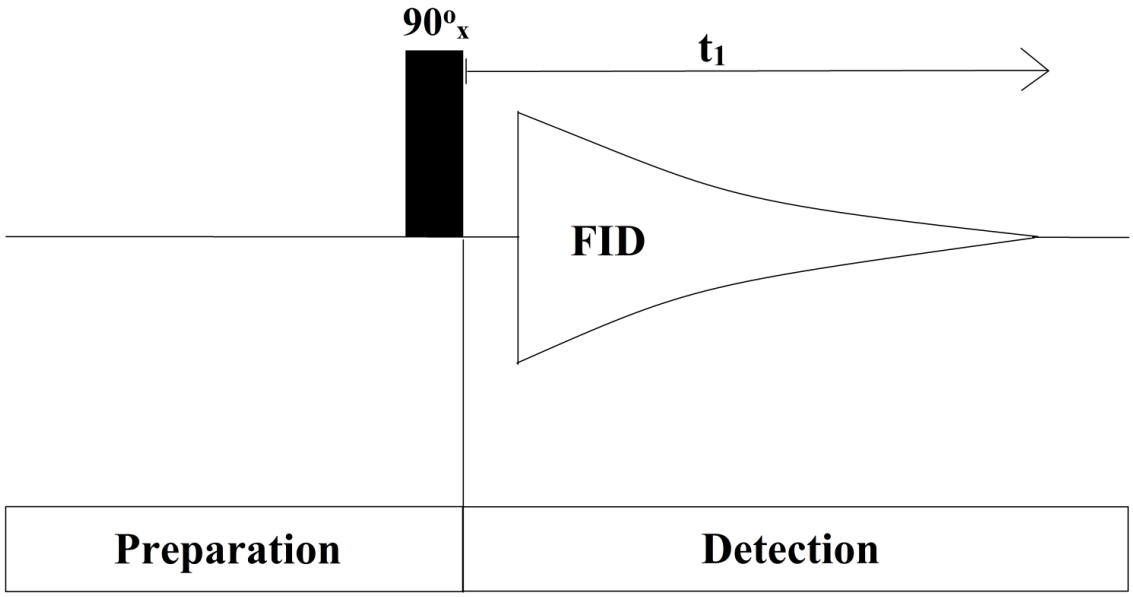


Figure 2.3.1: Figure depicting the two phases of recording a 1D NMR experiment – preparation and detection. The waiting time between pulses is termed evolution time, t_1. Adapted from Gunther, 2013.

2.3.2 2D NMR

A 2D NMR spectrum has two frequency axes – the horizontal axis is the direct frequency recorded as a FID and the vertical axis is the indirect frequency. There are four steps in recording a 2D NMR spectrum – preparation, evolution, mixing and detection. In order to record a 2D NMR spectrum with nuclei A and B, a 90° pulse is applied to A to rotate the net magnetisation of A to the x-y plane. Following this, the chemical shift of A is indirectly measured in the evolution step. During the mixing step, the magnetisation is transferred from A to B either via scalar coupling or NOE using different pulse sequences. The chemical shift of nucleus B is then measured in the final detection step. This experiment is repeated numerous times and the waiting time before recording the magnetisation of A (evolution time, t_1) is incremented by a fixed amount each time. This delay in evolution time is subjected to Fourier transform separately which gives the frequency of nucleus A (Gunther, 2013).

TROSY-HSQC

Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) is a 2D NMR experiment that involves transfer of magnetisation from one nucleus to another via scalar coupling. When studying proteins by NMR, HSQC is usually recorded between ^{1}H and ^{13}C or ^{1}H and ^{15}N. In both cases, ^{1}H frequency is the direct frequency and ^{13}C or ^{15}N is the indirect frequency. In ^{1}H-^{15}N HSQC, magnetisation is transferred from ^{1}H to ^{15}N, evolved on ^{15}N and transferred back to ^{1}H, which is then detected. Thus, it correlates the ^{1}H to the ^{15}N atom, to which the ^{1}H atom is bound, on the protein backbone. Every peak on the HSQC spectrum corresponds to the amide ^{1}H of an amino acid residue in the protein, with the exception of proline. The HSQC can, therefore, be used as a preliminary experiment to check if the protein sample can be studied further by NMR.

Size is a major limitation in NMR mainly due to the issue with signal overlap. As the chemical shift range for a specific nucleus is fixed, signals recorded from a large protein will

be heavily overlapped. This is further complicated by the fact that large proteins tumble slowly in comparison to small proteins. Large molecules have short T_2 relaxation times which lead to broadening of signals and loss of resolution. In addition to this, at high magnetic fields, chemical shift anisotropy (CSA) and dipolar coupling (DC) do not average out for large molecules. Decoupling the amide ^{1}H and ^{15}N atoms produces only one peak. In the absence of decoupling, peaks of large proteins appear as multiplets due to J coupling. These multiplets have different shapes because of cross correlation between CSA and DC. Transverse Relaxation Optimised Spectroscopy (TROSY) is a technique that selects the component in the signal wherein CSA and DC cancel each other out as shown in Figure 2.2.2, giving a sharp peak for every amide ^{1}H in the protein sequence in the HSQC (Keeler, 2010).

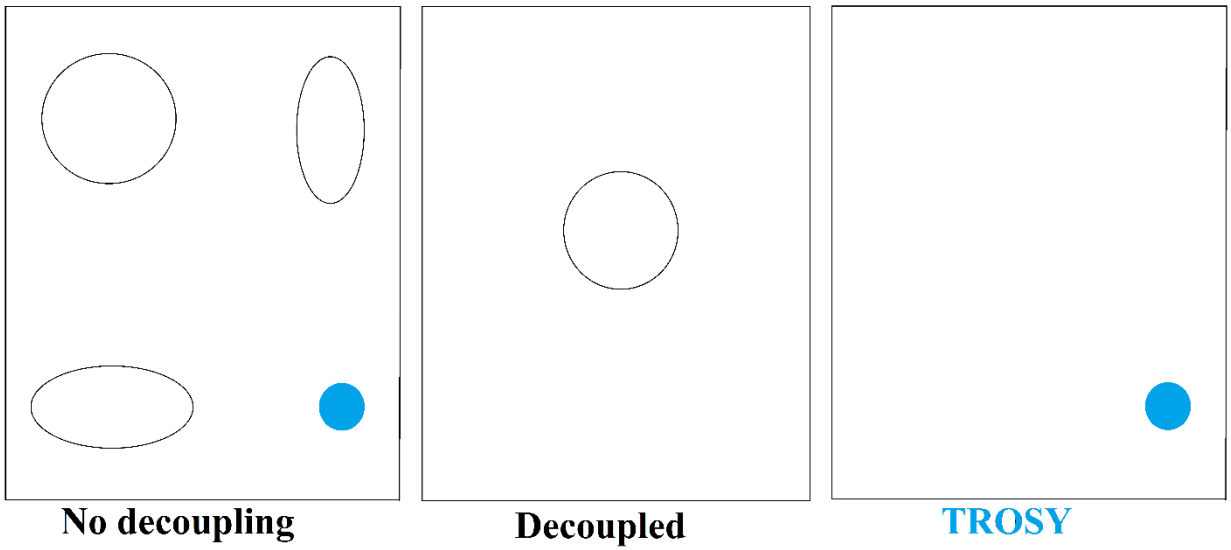


Figure 2.3.2: Figure depicting the signal recorded for a large protein when there is no decoupling (left panel), when the sample is decoupled (middle panel) and when the TROSY technique is used (right panel). Adapted from (Tomas, 2010).

NOESY

NOESY experiments provide valuable information which is required for structure determination of proteins using NMR. ^{15}N-NOESY and ^{13}C-NOESY experiments are used for

calculating distance restraints for structure calculation. Every strip in the ^{15}N-NOESY and ^{13}C-NOESY spectra contains signal from all the ^{1}H which are less than 5 Å away from the ^{1}H bound to the amide ^{15}N or backbone ^{13}C atom respectively.

2.3.3 Three dimensional NMR

A three dimensional NMR experiment can be described as a 2D NMR experiment with an additional evolution and mixing time (t_2) step between the first mixing time and the final detection step. The two indirect frequency dimensions are created by gradually incrementing the two evolution times, t_1 and t_2. There are two types of 3D NMR experiments which are crucial for structure determination – triple resonance and NOESY-HSQC experiments. Triple resonance experiments are vital in assigning the chemical shifts to the appropriate backbone and side chain atoms in a sequential manner. All the triple resonance experiments used for the purposes of this project have been described in this section. NOESY-HSQC experiments are used to obtain distance restraints for structure calculation. The NOESY-HSQC spectra used in this project have also been described in this section.

Triple Resonance Experiments

Triple resonance experiments are experiments which correlate three different nuclei – ^{1}H, ^{13}C and ^{15}N. The signals recorded in triple resonance experiments are intense as magnetisation is transferred via scalar coupling. Chemical shift assignment of atoms is generally done in two stages – backbone assignment and side chain assignment. Backbone assignment involves assignment of atoms which are present in the backbone of the protein sequence. These include the amide ^{1}H and ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and the carbonyl ^{13}C atoms as shown in figure 2.2.3. Following the assignment of these atoms, the side chain atoms are assigned. The nomenclature used for triple resonance experiments describes the pathway of magnetisation transfer and will be described further in this section. The atoms mentioned in brackets only relay the magnetisation and the signal from these atoms are not recorded.

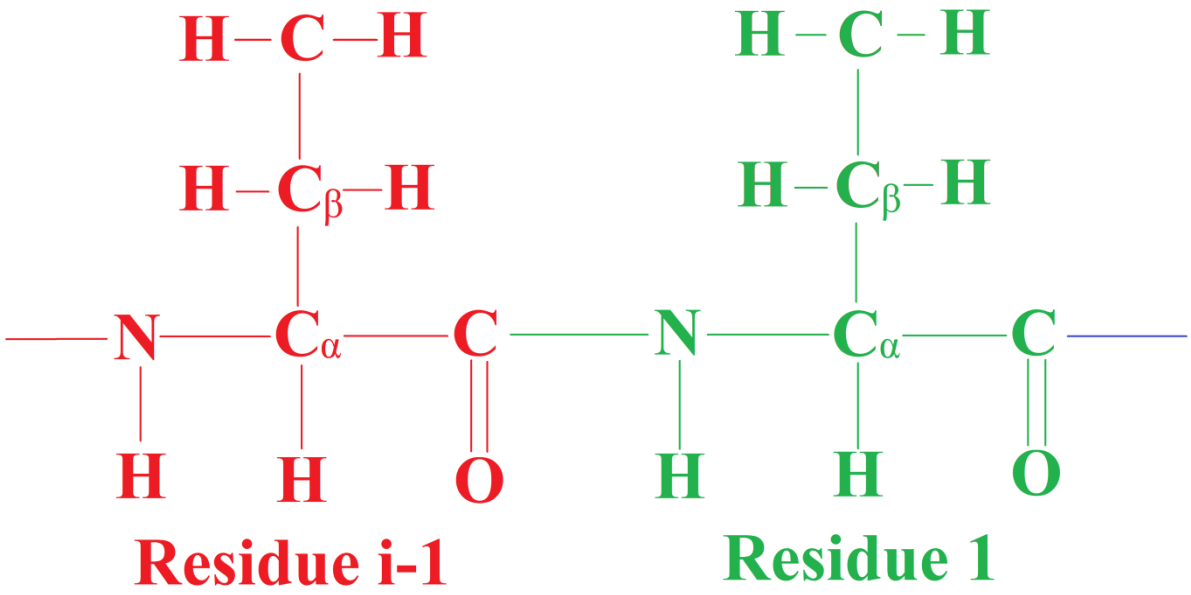


Figure 2.3.3: Schematic showing two consecutive amino acids in a protein sequence. The atoms studied during NMR experiments are labelled in the figure. The atoms of Residue i-1 are shown in red and the atoms of Residue 1 are shown in green.

CBCANH and CBCA(CO)NH

In the CBCANH experiment, the magnetisation is transferred from H_α and H_β to the C_α and C_β respectively. The magnetisation is then transferred from C_β to C_α to the amide ^{15}N and ^{1}H for detection. The chemical shift evolution is performed together on the two ^{13}C atoms so that they appear on one dimension. Chemical shift evolution is also performed on ^{15}N and ^{1}H, which form the two other dimensions of the spectrum. As shown in figure 2.2.4a, magnetisation transfer occurs from the $^{13}C_\alpha$ and $^{13}C_\beta$ atoms of the current and the previous residues in the sequence. Thus, every strip in this spectrum will contain four peaks – two peaks corresponding to the C_α and C_β atoms of the previous residue and two peaks corresponding to the C_α and C_β atoms of the current residue. This spectrum is used to assign the C_α and C_β atoms in conjunction with the CBCA(CO)NH spectrum (Grzesiek & Bax, 1993).

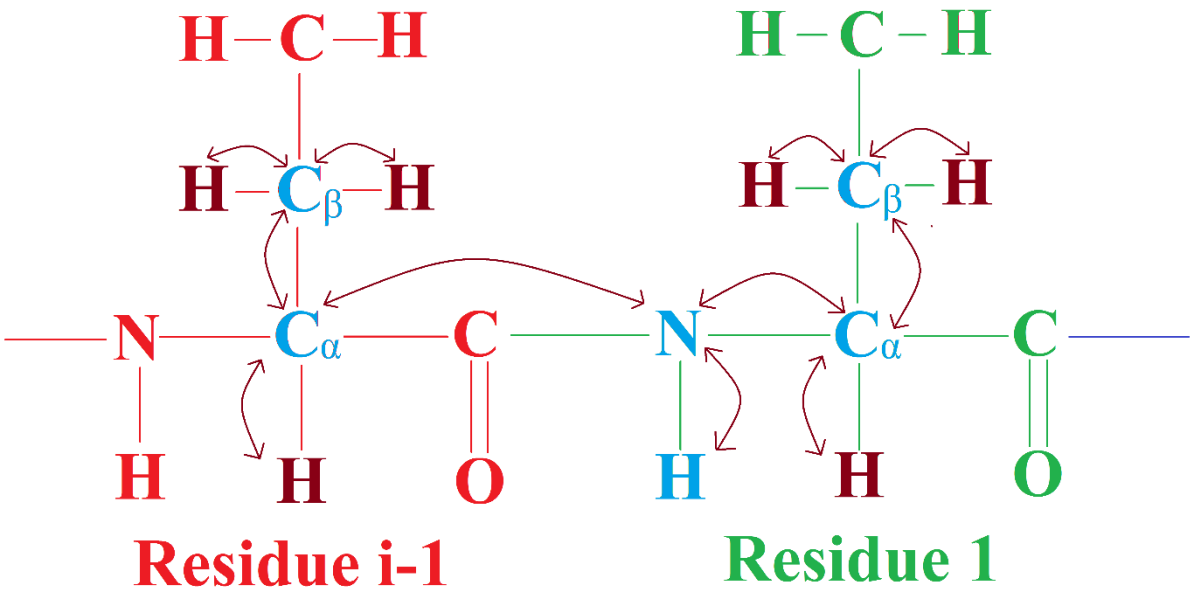


Figure 2.3.4a: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the CBCANH experiment. The chemical shifts of the brown atoms are not evolved whereas the chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

In the CBCA(CO)NH experiment, magnetisation is transferred from the H_α and H_β atoms of the previous residue to the C_α and C_β atoms respectively. The magnetisation is then transferred from the C_β atom to the C_α atom, relayed through the carbonyl ^{13}C to the amide ^{15}N and ^{1}H of the current residue as shown in figure 2.2.4b. Similar to the CBCANH, the chemical shift evolution of the ^{13}C atoms is performed simultaneously so that the two ^{13}C atoms are on the same dimension. Chemical shift evolution of the amide ^{15}N and ^{1}H atoms provide the other two dimensions of the spectra. Thus, every strip in the CBCA(CO)NH spectrum only contains two peaks – one peak that corresponds to the C_α atom of the previous residue and another peak that corresponds to the C_β atom of the previous residue. Using the CBCANH spectrum in conjunction with the CBCA(CO)NH spectra will allow us to determine which peaks correspond to the atoms in the previous residue and which peaks correspond to the atoms in the current residue. The C_α and C_β atoms in the protein sequence can be sequentially assigned using these spectra (Grzesiek & Bax, 1993).

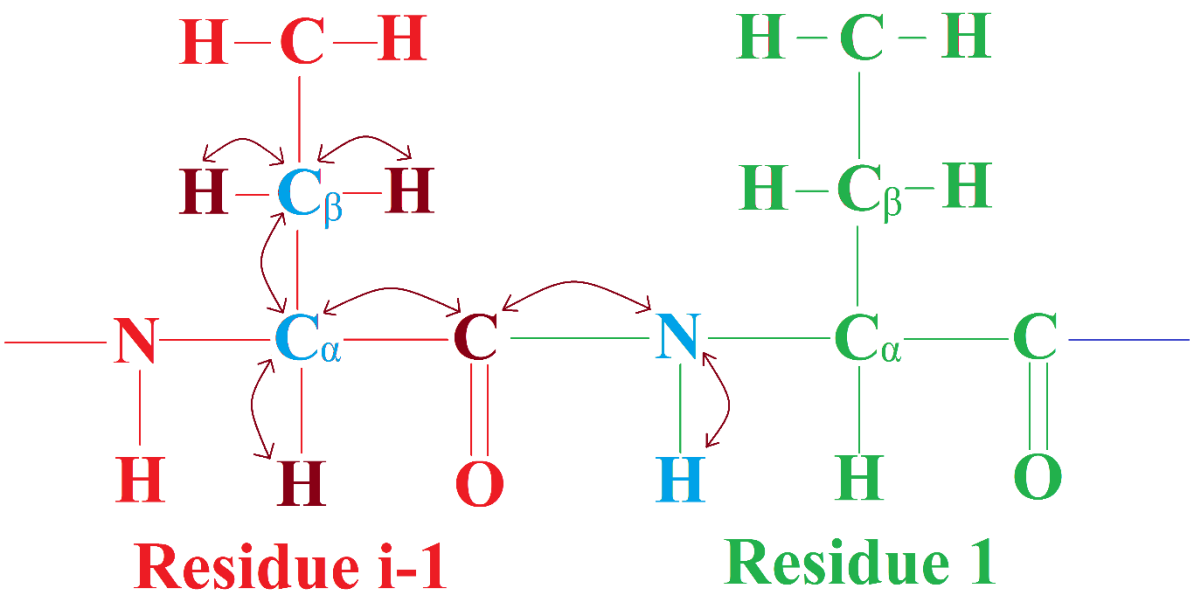


Figure 2.3.4b: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the CBCA(CO)NH experiment. The chemical shifts of the brown atoms are not evolved whereas the chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

HNCO and HN(CA)CO

In the HNCO experiment, magnetisation is transferred from the amide ^{1}H to the amide ^{15}N and then to the carbonyl ^{13}C of the previous residue. The magnetisation is then transferred from the carbonyl ^{13}C to the amide ^{15}N and ^{1}H, which is then detected as shown in figure 2.2.4c. Chemical shift evolution of the carbonyl ^{13}C, amide ^{15}N and ^{1}H nuclei is performed separately and they form the three dimensions of the spectrum. Every strip in the HNCO spectrum contains only one peak corresponding to the carbonyl ^{13}C of the previous residue (Grzesiek & Bax, 1993; Kay et al, 1990; Muhandiram & Kay, 1994). This spectrum can be used in conjunction with the HN(CA)CO spectrum to assign the carbonyl ^{13}C atoms in the protein backbone.

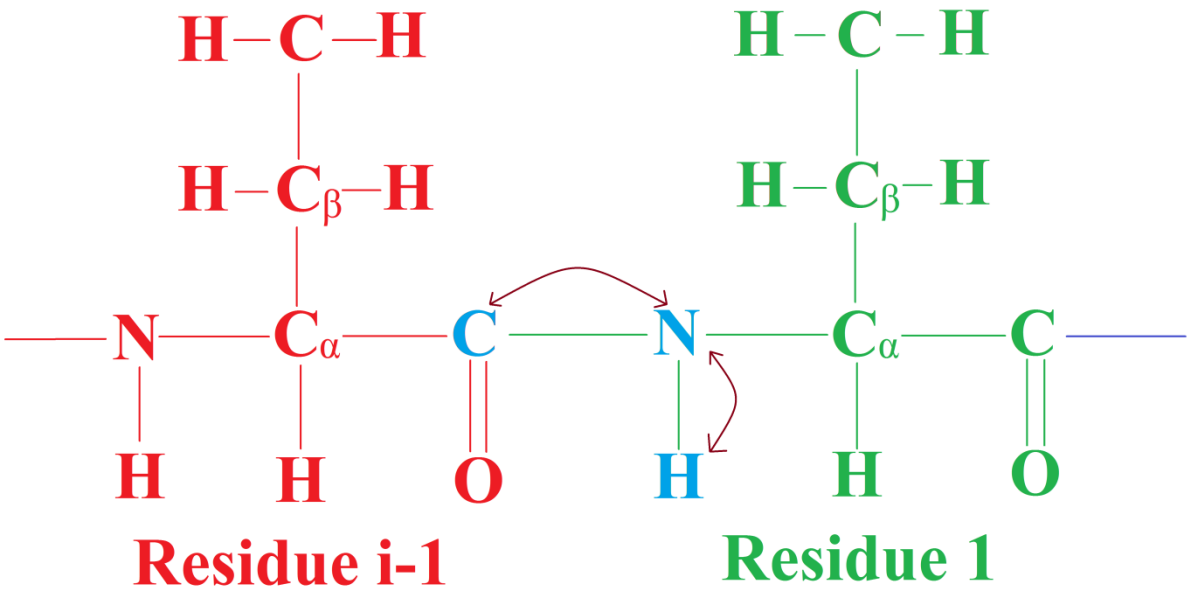


Figure 2.3.4c: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the HNCO experiment. The chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

In the HN(CA)CO experiment, magnetisation is transferred from the amide ^{1}H to the amide ^{15}N, which is then transferred to the C_{α} of the current and previous residue through scalar coupling. The magnetisation is then transferred from the C_{α} to the carbonyl ^{13}C through scalar coupling as shown in figure 2.2.4d. Chemical shift evolution of the C_{α} atom and the amide ^{15}N and ^{1}H atoms are performed separately, producing the three dimensions of the spectrum. Thus, every strip in the HN(CA)CO spectrum contains two peaks – one peak corresponds to the carbonyl ^{13}C of the current residue and the other peak corresponds to the carbonyl ^{13}C of the previous residue (Clubb et al, 1992). Using the HNCO and the HN(CA)CO spectra in conjunction will allow us to determine which peak corresponds to the carbonyl ^{13}C of the current residue and which peak corresponds to the carbonyl ^{13}C of the previous residue. The carbonyl ^{13}C can be assigned using the HNCO and HN(CA)CO spectra.

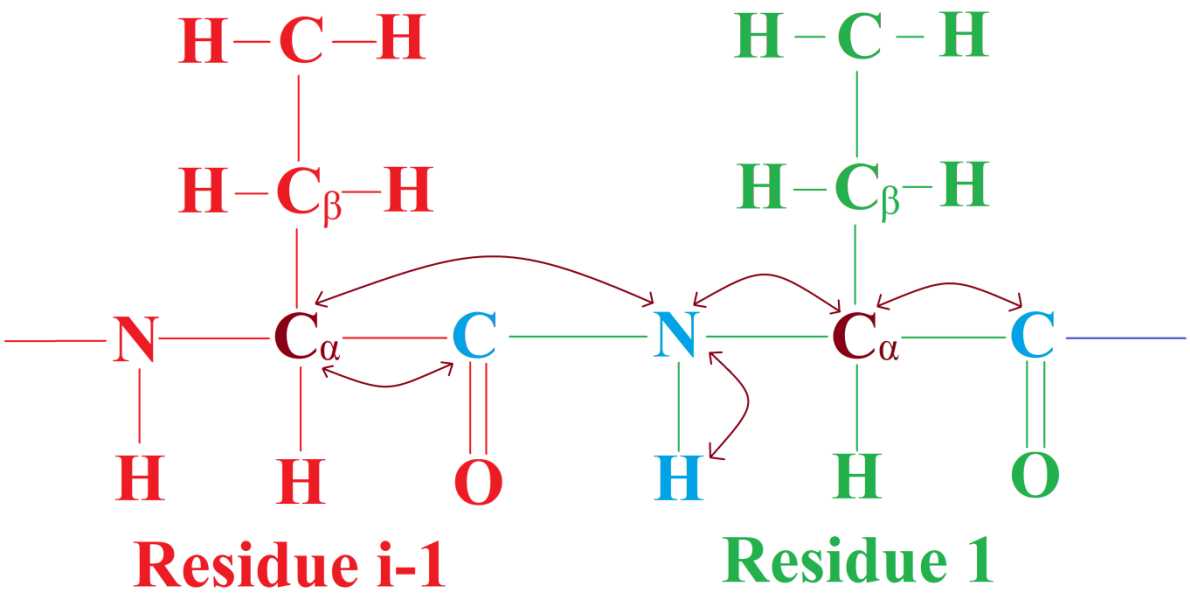


Figure 2.3.4d: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the HN(CA)CO experiment. The chemical shifts of the brown atoms are not evolved whereas the chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

HBHA(CBCACO)NH

In the HBHA(CBCACO)NH experiment, magnetisation is transferred from the H_α and H_β atoms of the previous residue to the C_α and C_β atoms respectively. The magnetisation is then transferred from the C_β atom to the C_α atom, relayed through the carbonyl ^{13}C to the amide ^{15}N and ^{1}H of the current residue, similar to the CBCA(CO)NH experiment as shown in figure 2.2.4e. However, in this experiment, chemical shift evolution of the two ^{1}H is performed simultaneously to obtain one dimension. Chemical shift evolution of the current amide ^{15}N and ^{1}H atoms yields two more dimensions (Grzesiek & Bax, 1993). Every strip in the HBHA(CBCACO)NH spectrum contains peaks corresponding to H_α and H_β of the previous residue.

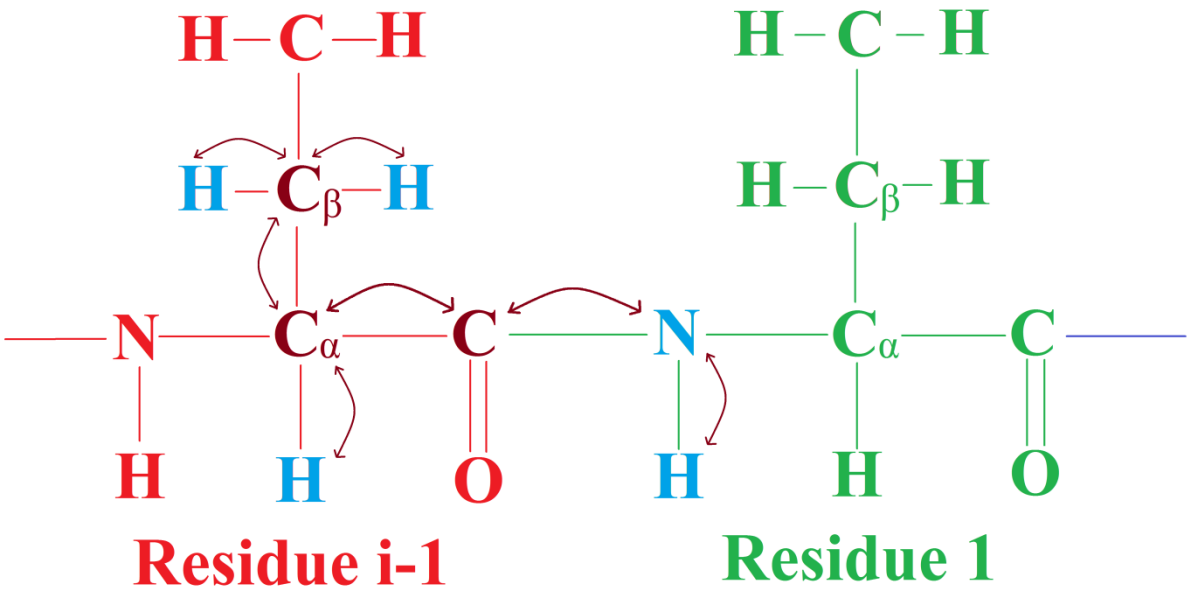


Figure 2.3.4e: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the HBHA(CBCACO)NH experiment. The chemical shifts of the brown atoms are not evolved whereas the chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

CC(CO)NH

In the CC(CO)NH experiment, magnetisation is transferred from all the side chain ^{1}H of the previous residue to the ^{13}C that they are bound to. Following this, isotropic ^{13}C mixing is performed to transfer magnetisation between the ^{13}C atoms. Magnetisation is then transferred to the carbonyl ^{13}C, then to the amide ^{15}N and ^{1}H of the current residue as shown in figure 2.2.4f. Chemical shift evolution of all the ^{13}C atoms of the previous residue except the carbonyl ^{13}C is performed creating one of the dimensions for the spectrum. Then, the chemical shifts of the amide ^{15}N and ^{1}H of the current residue are also performed to obtain the two other dimensions for the spectrum (Grzesiek & Bax, 1993). Every strip in the CC(CO)NH spectrum contains peaks corresponding to all the side chain ^{13}C of the previous residue.

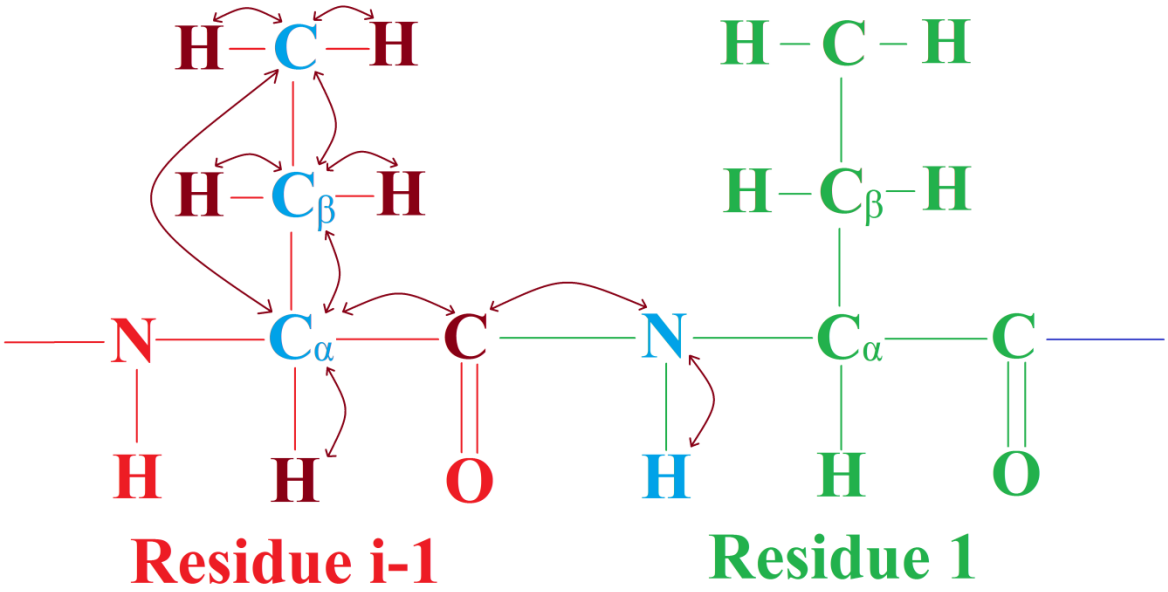


Figure 2.3.4f: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the CC(CO)NH experiment. The chemical shifts of the brown atoms are not evolved whereas the chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

HCCH-TOCSY

In the HCCH-TOCSY experiment, magnetisation is transferred from the side chain ^{1}H to the ^{13}C atoms they are bound to. Isotropic ^{13}C mixing is performed, following which magnetisation is transferred back to the side chain ^{1}H and detected as shown in figure 2.2.4g (Bax et al, 1990; Olejniczak et al). Every strip in the HCCH-TOCSY spectrum contains peaks corresponding to the side chain ^{1}H of the same residue.

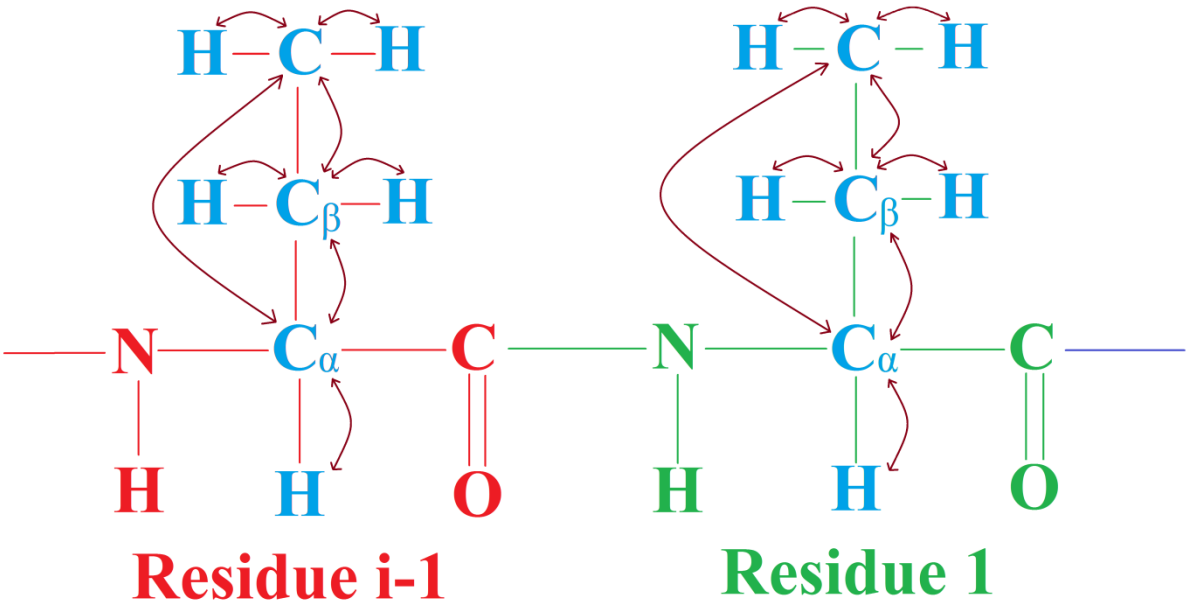


Figure 2.3.4g: Figure depicting the magnetisation transfer between the nuclei labelled by the brown arrows in the HCCH-TOCSY experiment. The chemical shifts of the blue atoms are evolved. Residue i-1 (red) and residue 1(green) are labelled in the figure.

NOESY-HSQC

Two types of NOESY-HSQCs are used for assigning the ^{1}H which are less than 5 Å away from the ^{1}H bound to the amide ^{15}N and the aliphatic/aromatic ^{13}C atoms. These are the ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC. The ^{15}N-NOESY-HSQC involves magnetisation transfer between all the neighbouring ^{1}H with the ^{1}H bound to the amide ^{15}N via NOE. The magnetisation is then transferred to the amide ^{15}N atom and back to the ^{1}H, which is then detected. Thus, every strip in the ^{15}N-NOESY-HSQC contains signal from all the ^{1}H which are less than 5 Å away from the ^{1}H bound to the amide ^{15}N atom. Similarly, the ^{13}C-NOESY-HSQC experiment involves magnetisation transfer between all the neighbouring ^{1}H via NOE. The magnetisation is then transferred either to the ^{1}H bound to the aliphatic ^{13}C in the protein backbone/side chain or to the aromatic ^{13}C atom in the side chain, depending on the atom targeted by the pulse sequence. The magnetisation is then transferred back to the ^{1}H for detection. Every strip in the ^{13}C-NOESY-HSQC spectrum contains signals from all ^{1}H atoms

which are less than 5 Å away from the ^{1}H bound to the targeted ^{13}C atom. These spectra provide information about ^{1}H nuclei which are close to the ^{1}H bound to the ^{15}N or ^{13}C atoms in space even though they may be far away from them in the sequence (Marion et al, 1989a; Marion et al, 1989b; Zuiderweg & Fesik, 1989). Thus, this provides us with information about the three dimensional structure of the protein.

2.4 ARIA – Ambiguous Restraints for Iterative Assignment

In order to perform a structure calculation, sequential chemical shift assignment is performed for the protein sequence using the data from the three dimensional NMR spectra. Following this, the peaks on the ^{13}C and ^{15}N NOESY spectra are picked to create two peak lists. Dihedral angle restraints are also generated from the chemical shift assignments, using TALOS+. TALOS+ uses the chemical shift assignments of the amide ^{1}H and ^{15}N, H_{α}, C_{β}, C_{α} and carbonyl ^{13}C atoms to predict the protein phi and psi backbone torsion angles. The limiting step in structure calculation by NMR is the assignment of NOE peaks. ARIA uses iterative assignments and ambiguous distance restraints (ADR) to assign the NOE peaks (Rieping et al, 2007). Structure calculation and refinement are performed by CNS (Crystallography and NMR System) using the simulated annealing refinement method (Brunger, 2007). The explanation of the concepts involved in the functioning of ARIA is from Nilges et al (1997) and Linge et al (2003).

ARIA was provided the manual chemical shift assignment, ^{13}C and ^{15}N NOESY peak lists and the dihedral angle restraints for the protein. ARIA separates the data provided into two parts – the first part consists of the manually assigned chemical shifts and the second part consists of the NOESY peak lists which are not assigned. As a first step, ARIA uses the manual assignments to generate an ensemble of structures (Nilges et al, 1997). This step is called the zero iteration and any ambiguous assignment is treated with ambiguous distance restraints (ADR) (Nilges, 1995). The dihedral angle restraints from TALOS+ provide

convergence in the protein structure in the initial stages (Linge et al, 2003). Following zero iteration, the structures in the ensemble are ordered based on their total energy. The structure which is of lowest energy is chosen as the reference.

In the next step, the NOESY peak lists are calibrated using the average distances calculated from the chosen reference structure. The unassigned peaks in the NOESY peak list are then compared to the chemical shift assignments and for every match, an ADR is added to the list. All the restraints that are extracted from the NOESY peaklists are then analysed to check for violations based on the reference structure. Any restraint that causes a systematic violation is removed from the list. There are many reasons that can contribute to the removal of a peak. This can occur when noise is picked as a peak, when a peak is not picked accurately due to overlap or when the picked peak is not associated with a peak in the spectrum.

The restraints from the list are then partially assigned. This procedure is performed for every spectrum that is provided to ARIA, and the lists generated from these spectra are merged to avoid duplicates. Following this, a new ensemble of structures is generated based on these restraints by CNS. This procedure is repeated through every iteration step with the creation of a new ensemble of structures, until the ensemble of structures and the data set do not change significantly (Nilges et al, 1997). This process is also shown in Figure 2.3.

The latest version of ARIA provides a graphic user interface (GUI) to set up the ARIA run. Once the chemical shifts, the NOESY peaklist files and the dihedral angle restraint files are converted into the .xml format, the run can be set up using the GUI. Following the initial run, ARIA provides violation files which provide a list of the NOE peaks which are causing violations. The NOESY peaklists can be analysed and any noise peaks and incorrectly picked peaks can be removed or corrected. Following this, a new run can be set up on ARIA with the altered NOESY peaklists and chemical shifts. ARIA ultimately provides an ensemble of

structures calculated by CNS, assigned distance restraints, distance restraints which were assigned but rejected because the method considers them to be artefacts and distance restraints which were rejected because they do not have any chemical shift that corresponds to them (Nilges et al, 1997). The structures obtained from ARIA can then be validated using various validation softwares to ensure that they are biologically relevant.

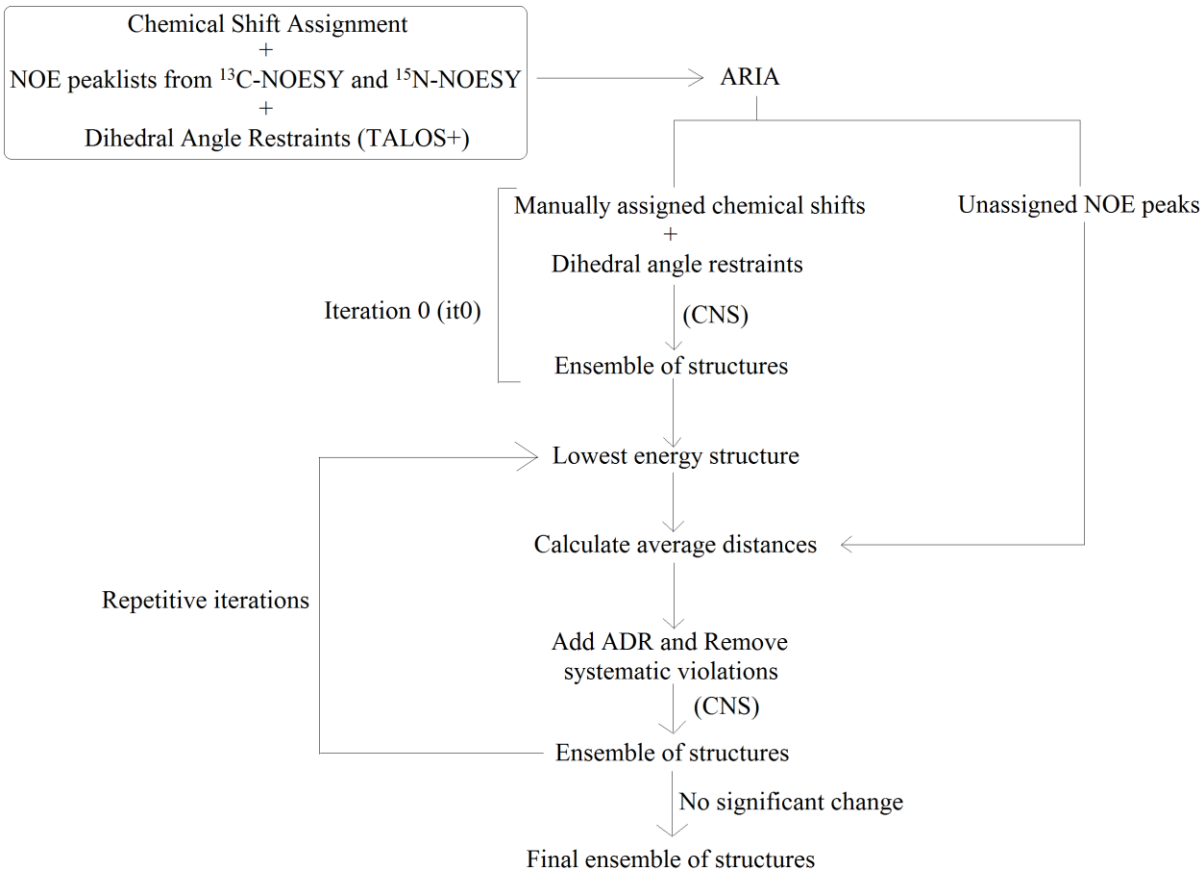


Figure 2.4: Figure showing a flowchart of the various stages followed by ARIA for protein structure calculation.

Chapter 3

Materials & Methods

3.1 *T. gondii* IMP1 homologue

3.1.1 PCR amplification

Primers were designed using GENtle (University of Cologne, Germany), for selective amplification of TgIMP1 from the pET32 Xa/LIC vector (Merck Millipore, US), with additional 5' sequences that allow LIC cloning (see Appendix Table 1). PCR reactions were performed following the Q5 HF Polymerase protocol (Fermentas, US) in Mastercycler gradient machine (Eppendorf, Germany) and conditions used for the different reactions are given in Tables 3.1 and 3.2. The PCR products were run on a 1% Agarose gel (Appendix Table 3) at 120 V for 60 min with a 100 bp Quickload DNA ladder (NEB, UK). The PCR products were purified for the positive reactions using a PCR purification kit (Fermentas, US).

Reaction Components (Initial concentration)	**Final concentrations**
Sense (5') Primer (10 µM)	1 µM
Anti-sense (3') Primer (10 µM)	1 µM
Template DNA (90 ng/µl)	50 ng
5x Q5 HF Buffer	1x
dNTP Mix, 2mM each	0.02 mM
Q5 HF DNA Polymerase	1 unit
Total Volume	50 µl

Table 3.1: Table showing the concentration of reaction components used for the two PCR reactions performed to selectively amplify the DNA of the TgIMP1 homologue from the T. gondii transcriptome.

Thermocycler: 30x
Polymerase Activation - 95°C for 2 min
Denaturation - 95°C for 20 s
Annealing - 60°C for 30 s
Extension - 72°C for 45 s
Repeat steps 2 to 4 for 30 cycles
Extension - 70°C for 30 s
End of reaction

Table 3.2: Table showing the PCR reaction conditions set for the thermocycler.

3.1.2 Screening for positive colonies

The amplified TgIMP1 gene was cloned into the pET46 Ek/LIC vector (Merck Millipore, US) following the LIC protocol. The plasmid containing the TgIMP1 encoding sequence and a carbenicillin resistance gene was transformed in DH5α cells (NEB, UK) and plated on LB-agar plate (Sigma-Aldrich, UK) containing 50 µg/ml carbenicillin (Sigma-Aldrich, UK). These colonies were duplicated on another LB-agar plate (Sigma-Aldrich, UK) and PCR amplified using the Ready-to-go PCR beads (GE Healthcare, UK). The PCR products were run on a 1% agarose gel (Sigma-Aldrich, UK). Gels were visualised and imaged in the UGenius Gel system (Syngene, UK).

One of the positive colonies detected by PCR was used to inoculate 10 ml LB (50 µg/ml carbenicillin) (Sigma-Aldrich, UK) and incubated overnight at 37°C with 220 rpm shaking. Plasmid DNA was purified from the overnight culture using the Plasmid miniprep kit (Fermentas, US). The purified DNA was sent for sequencing with T7 forward and reverse primers to Source Bioscience. The sequencing results were analysed using GENtle (University of Cologne, Germany) to verify the sequence of the sample.

3.1.3 Plasmid purification

Plasmid pET46 Ek/LIC (Merck Millipore, US) encoding TgIMP1 protein with a His tag at the N terminus and the ampicillin resistance gene was purified using the Fermentas Plasmid Miniprep kit. The concentration of DNA was measured using a Nanodrop Spectrophotometer (Thermofisher, UK). All plasmids were purified following this method.

3.1.4 Transformation of *E. coli*

Plasmid encoding the TgIMP1 protein and carbenicillin resistance gene was used to transform competent *E. coli* Rosetta 2 (DE3) cells (Merck Millipore, US) or DH5α cells (NEB, UK). 3 µl of plasmid DNA (~50 ng) was added to 50 µl of cells and left on ice for 30 min. The cells were then heat shocked at 42°C for 45 sec on a heat block and immediately transferred to ice for 2 min. 400 µl of LB media (Sigma-Aldrich, UK) was added and incubated at 37°C for 1 hour with shaking at 220 rpm. These were then plated on LB-Agar plates (Sigma-Aldrich, UK) with the correct antibiotic to select for transformants.

3.1.5 Expression test

For the expression, the sequenced plasmid containing the TgIMP1 encoding sequence was transformed into Rosetta 2 cells (DE3 *E. coli* cells which supply the tRNAs for seven rare codons which aid in translation, on a plasmid containing the chloramphenicol resistance gene) (Merck Millipore, US). They were plated on a LB-agar plate (Sigma-Aldrich, UK) containing 50 µg/ml carbenicillin (Sigma-Aldrich, UK) and 34 µg/ml chloramphenicol (Sigma-Aldrich, UK). 10 ml LB media (Sigma-Aldrich, UK) was inoculated with each of these colonies separately and grown at 37ºC with 220 rpm shaking till OD_{600} reached 0.6. OD_{600} was measured on a spectrophotometer (Jenway, UK). A 1 ml sample was taken from each culture and the culture induced with 0.1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) (Sigma-Aldrich, UK). The cultures were grown at 18°C and 25°C overnight with 220 rpm shaking for aeration. A 1 ml sample was taken after this. All samples were spun down at 12000g and the

pellet treated with BPER II (Thermo Scientific, USA) following the given protocol. All samples were then run on a 12% SDS-PAGE gel (Serva, Germany).

3.1.6 Expression

Expression was performed following the same protocol described in Section 3.1.5 with larger volumes of 1 L. The inoculated cultures were grown at 18°C with 220 rpm shaking overnight after induction. The cells were centrifuged at 4000g at 4°C for 15 min in a Sorvall centrifuge (ThermoFisher, UK).

3.1.7 Protein Purification

The cell pellet was resuspended in phosphate lysis buffer (Appendix Table 3) and then lysed in a cell disrupter (Constant systems, UK) at a pressure of 25 kpsi. The cell lysate was centrifuged at 30000 g for 45 min in a Sorvall centrifuge (ThermoFisher, UK). The supernatant was purified using pre-packed Nickel column, 5 ml Histrap FF (GE Healthcare, UK) in the AKTAprime (GE Healthcare, UK). The column was equilibrated with phosphate wash buffer (Appendix Table 3). The clear lysate was then loaded on to the column, the column equilibrated with the wash buffer until the UV dropped to zero and stabilised. Then, the protein eluted with an imidazole gradient that required 50 ml to reach 250 mM imidazole. The eluted protein was pooled and concentrated. The concentration was measured using a Nanodrop Spectrophotometer (Thermofisher, UK).

The protein was further purified by Gel filtration using a Superdex 75 Hiload 16/600 column (GE Healthcare, UK) with AKTAprime (GE Healthcare, UK). The column was first equilibrated with one column volume (120 ml) of phosphate GF buffer pH 8 (Appendix Table 3). The protein was injected into the column with a 2 ml syringe and 1.5 ml fractions were collected.

3.1.8 SDS-PAGE

Samples for SDS PAGE were mixed with 2x loading buffer (Serva, Germany), vortexed and heated at 100°C on a heat block for 5 min. Samples were run at 180 V for 60 min in a 12% SDS-PAGE gel (Serva, Germany) with 1X diluted 20X SERVA run buffer (Serva, Germany) and stained with Instant Blue (Expedeon, UK) for 15 min and destained with dH_2O for one hour. Gels were visualised and imaged in the UGenius Gel system (Syngene, UK).

3.1.9 Limited Proteolysis

50 µl of 10 µg/ml chymotrypsin (Sigma-Aldrich, UK) in Digestion buffer (Appendix Table 3) was incubated with 50 µl of 10 mg/ml protein at 37°C and 20 µl aliquots were taken at 5 min, 15 min, 1 hour, 2 hours and 1 day. 100 µl of 9 µg/ml Chymotrypsin was also incubated at 37°C and aliquots were taken after the same time periods as a control. Trypsin (Sigma-Aldrich, UK) digest was also performed following a similar protocol. Samples were run on a 12% gel by SDS PAGE. Gels were visualised and imaged in the UGenius Gel system (Syngene, UK).

3.1.10 Membrane transfer for N-terminal amino acid sequencing

The chymotrypsin digested protein from the gel was transferred to a PVDF membrane (Thermo Scientific, UK) following the given protocol. The membrane was stained using PVDF stain (Appendix Table 3) and destained with 50% Methanol. The membrane was washed with water and air dried. Sequencing was performed by PNAC facility, University of Cambridge.

3.1.11 NMR experiments

Deuteration of TgIMP1

10 ml LB media (50 µg/ml carbenicillin and 34 µg/ml chloramphenicol) (Sigma-Aldrich, UK) was inoculated with a single colony and incubated at 37°C with 220 rpm shaking until

OD_{600nm} reached 0.6. 10 ml of 50% D_2O M9 minimal media (50 µg/ml carbenicillin and 34 µg/ml chloramphenicol) was inoculated with 50 µl of LB culture and incubated overnight at 37°C at 220 rpm shaking. 1 L of 100% D_2O M9 minimal media (50 µg/ml carbenicillin and 34 µg/ml chloramphenicol) was inoculated with the 50% D_2O M9 minimal media overnight culture and incubated at 37°C with 220 rpm shaking. Cells were induced with 1 mM IPTG (Sigma-Aldrich, UK) once the OD_{600nm} of the samples reached around 0.6 and incubated at 18°C with 220 rpm shaking overnight. The cells were centrifuged at 4000 g at 4°C for 15 min in Sorvall centrifuge (ThermoFisher, UK) and purified as described in Section 2.7.

NMR sample preparation

Unlabelled, ^{15}N, $^{13}C/^{15}N$ and $^{13}C/^{15}N/^{2}H$ labelled samples were concentrated and buffer exchanged into phosphate NMR buffer (Appendix Table 3) by repeatedly washing the sample with the buffer in a 10 kDa MWCO (Molecular Weight cut off) concentrator (Sartorius, UK). 10% D_2O (Sigma-Aldrich, UK) was added to the sample (for the field-frequency lock) and the sample transferred to a 5 mm NMR tube (Sigma-Aldrich, UK) for NMR analysis.

NMR ^{1}H 1D and ^{15}N-^{1}H HSQC measurements

Data was recorded at 303 K on a Bruker Avancell 800 (800 MHz).

NMR backbone assignment

$^{13}C/^{15}N$ labelled TgIMP1 clone 7 (200 µM) was used to collect ^{15}N-HSQC, HNCACB and HNCOCACB NMR spectra at 303 K on a Bruker Avancell 800 (800 MHz) and backbone assignment was reached using standard double- and triple-resonance techniques. $^{13}C/^{15}N/^{2}H$ labelled TgIMP1 was used to collect ^{15}N-HSQC, HNCO, HNCACO, HNCACB and HNCOCACB NMR spectra at 303 K on a Bruker Avancell 800 (800 MHz).

3.1.12 Crystallisation trials

Crystallisation sample preparation

TgIMP1 Clone 6 and Clone 7 (Table 4.1) samples were concentrated and buffer exchanged into crystallisation buffer by repeatedly washing the sample with the buffer in a 10 kDa MWCO concentrator (Sartorius, UK).

96 – well Crystallisation plates

Two 100 nl drops of TgIMP1 clone 6 at 25 mg/ml were added to the two wells for each condition in the 14 commercially available screens pre-dispensed into 96-well MRC plates (provided by the Centre for Structural Biology at Imperial College London) using the Mosquito nano-litre high throughput robot (TTP labtech, UK). One drop was mixed with 100 nl of the corresponding reservoir solution and the second drop was mixed with 200 nl of the same reservoir solution. Two plates – plates 1 and 2 from the commercially available screens were set up with TgIMP1 Clone 7 at 50 mg/ml the same way as with TgIMP1 Clone 6.

3.2 *Eimeria tenella* IMP1 homologue

3.2.1 PCR Amplification

The *E. tenella* cDNA library produced from merozoite stage parasites was provided by Dr. Damer Blake. EtIMP1 clone 2 was amplified from the *E. tenella* cDNA library, with additional 5' sequences for LIC cloning (Appendix Table 1). PCR Amplification of the EtIMP1 clone 2 gene was performed the same way as described in section 3.2.1, except that the Tm was 59°C for EtIMP1 clone 2 and the elongation time was adjusted to 20 seconds.

3.2.2 Screening for positive colonies

The amplified EtIMP1 clone 2 gene was cloned into the pET30 Xa/LIC vector (Merck Millipore, UK) following the LIC protocol. The plasmid containing the EtIMP1 clone 2

encoding sequence and kanamycin resistance gene was transformed in DH5α cells (NEB, UK) and plated on a LB-agar plate (Sigma-Aldrich, UK) containing 50 μg/ml kanamycin (Sigma-Aldrich, UK).

Colony screening, transformation, protein expression and purification, SDS-PAGE and limited proteolysis were performed following the same protocol described for TgIMP1.

3.2.3 NMR experiments

NMR sample preparation

500 μl of ^{15}N labelled EtIMP1 clone 2 in phosphate NMR buffer pH 8 (Appendix Table 3) was prepared. 10% D_2O (Sigma-Aldrich, UK) was added to this sample and the sample transferred to a 5 mm NMR tube (Sigma-Aldrich, UK) for NMR analysis.

NMR ^{1}H 1D and ^{15}N-^{1}H HSQC measurements

Data was recorded at 303 K on a Bruker Avancell 600 (600 MHz).

3.3 *P. falciparum* IMP2 homologue

3.3.1 PCR Amplification

The IMP1-like homologue from *P. falciparum* (Accession number: XP_001349200) was codon optimised and synthesised (GeneArt, ThermoFisher). Various primers were designed with additional 5' sequences for LIC cloning using GENtle (University of Cologne, Germany) to clone the PfIMP2 gene into different vectors (Appendix Table 1) – pET30 Xa/LIC (Merck Millipore, US), pET32 Xa/LIC (Merck Millipore, US) and pNIC-ZB (Genbank GU452710). pNIC-ZB was a gift from Opher Gileadi (Addgene plasmid # 26107). The pNIC-ZB vector contains a His tag and Z-tag, followed by a TEV protease cleavage site. The Z-tag is sequence derived from protein A, modified to have a highly positive surface charge.

For ectopic expression of PfIMP2 in the *P. falciparum* 3D7 parasite, the PfIMP2 gene was synthesised without codon optimisation (GeneArt, ThermoFisher). PCR reactions were performed following the Q5 HF Polymerase protocol (Fermentas, UK) in Mastercycler gradient machine (Eppendorf, Germany) as described in section 3.1.1. The PfIMP2 gene was cloned into the pRM2_DD-3HA-GFP vector and transformed into Stellar DE3 cells following the Infusion HD Cloning Kit (Clontech, UK). The pRM2_DD-3HA-GFP vector contains the strong, schizont-specific msp2 promoter and a triple HA and GFP tag at the C-terminus (de Azevedo et al, 2012). Stellar DE3 cells lack the gene cluster for degrading methylated DNA.

PCR amplification was performed the same way as described in section 3.1.1 in both cases except that the Tm given in Appendix Table 1 were used for the relevant PCR runs and the elongation time was decreased to 20 seconds. Transformation and colony screening were performed as described in section 3.1.2 for the codon optimised construct.

3.3.2 Plasmid purification

Plasmid purification was performed as described in section 3.1.3 for the codon optimised construct. The pRM2_DD-3HA-GFP vector containing the PfIMP2 gene was purified using the GeneJet Plasmid Midiprep Kit (ThermoFisher, UK). The concentration of DNA was measured using a Nanodrop Spectrophotometer (Thermofisher, UK).

The codon optimised PfIMP2 plasmids were transformed into BL21 cells (NEB, UK), Rosetta 2 cells (Merck Millipore, US) and ArcticExpress cells (Agilent technologies, US) following the recommended protocols. Expression tests and large scale protein expression was performed as described in section 3.1.5.

3.3.3 Protein Purification

The cell pellet was resuspended in hepes lysis buffer (Appendix Table 3) and 1 µg/ml PMSF and then lysed in a Sonopuls HD 2070 sonicator (Bandelin, Germany) at 60% power for 6

cycles in 6 minutes. The cell lysate was centrifuged at 30000 g for 45 min in a Sorvall centrifuge (ThermoFisher, UK). The supernatant was purified using pre-packed Nickel column, 5 ml Histrap FF (GE Healthcare, UK) in the AKTAprime (GE Healthcare, UK). The column was equilibrated with Hepes wash buffer (Appendix Table 3). The clear lysate was then loaded on to the column, the column equilibrated with the wash buffer until the UV dropped to zero and stabilised. Then, the protein eluted with an imidazole gradient that required 50 ml to reach 250 mM imidazole. The eluted protein was pooled and concentrated. The concentration was measured using a Nanodrop Spectrophotometer (Thermofisher, UK).

The protein was further purified by Gel filtration using a Superdex 75 Hiload 16/600 column (GE Healthcare, UK) with AKTAprime (GE Healthcare, UK). The column was first equilibrated with one column volume (120 ml) of Hepes GF buffer 10 mM DTT. The protein was injected into the column with a 2 ml syringe and 1.5 ml fractions were collected.

3.3.4 Factor Xa Cleavage

10 µg of PfIMP2 (pET32 Xa/LIC) in 50 mM Tris pH 8 150 mM NaCl was cleaved with 0.5 U of Factor Xa protease (NEB, UK) at room temperature following the given protocol. This was then scaled up to cleave 10 mg of protein.

3.3.5 DSF Buffer Screen

PfIMP2 (pNIC-ZB) construct in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer was used to perform a buffer screen. 40 µM PfIMP2 was used for each condition in the screen with 1x SYPRO Orange dye (Sigma-Aldrich, UK). The screens were set up in a 96-well thin wall PCR plate (Bio Rad, US) and heated from 26°C to 97°C using the Stratagene Mx3005P QPCR system (Agilent technologies, US). Fluorescence was monitored at excitation wavelength of 380 nM and emission wavelength of 440 nM. The conditions used for the buffer screen are given below.

Buffer composition	**pH**			
50mM NaH_2PO_4 100mM NaCl	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 100mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 150mM NaCl (**CONTROL**)	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 150mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 200mM NaCl	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 200mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 250mM NaCl	pH4	pH5	pH7	pH8
50mM NaH_2PO_4 250mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM Hepes 100mM NaCl	pH4	pH5	pH7	pH8
50mM Hepes 100mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM Hepes 150mM NaCl	pH4	pH5	pH7	pH8
50mM Hepes 150mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM Hepes 200mM NaCl	pH4	pH5	pH7	pH8
50mM Hepes 200mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM Hepes 250mM NaCl	pH4	pH5	pH7	pH8
50mM Hepes 250mM NaCl 10mM DTT	pH4	pH5	pH7	pH8
50mM Tris 100mM NaCl	pH7		pH8	
50mM Tris 100mM NaCl 10mM DTT	pH7		pH8	
50mM Tris 150mM NaCl	pH7		pH8	
50mM Tris 150mM NaCl 10mM DTT	pH7		pH8	
50mM Tris 200mM NaCl	pH7		pH8	
50mM Tris 200mM NaCl 10mM DTT	pH7		pH8	
50mM Tris 250mM NaCl	pH7		pH8	

50mM Tris 250mM NaCl 10mM DTT	pH7	pH8

Table 3.3: Table showing the different conditions used for the buffer screen performed for PfIMP2 (pNIC-ZB).

3.3.6 Western blot

Western blots were performed on recombinant PfIMP2 expressed in *E. coli* using the recommended protocol (GE Healthcare, UK) and was probed with anti-His antibody (GE Healthcare, UK). Western blot performed on the ectopically expressed PfIMP2 in *P. falciparum* 3D7 strain was probed with 0.1 µg/ml anti-HA (rat) primary antibody and 0.1 µg/ml anti-rat secondary antibody (ThermoFisher, UK).

3.3.7 TEV cleavage

1 mg TEV protein was incubated with 20 ml of PfIMP2 (1 mg/ml) in 50 mM Tris pH 8, 250 mM NaCl 200 mM Imidazole 10 mM DTT in dialysis tubing (ThermoFisher, UK). This reaction was incubated at 4°C overnight and dialysed against TEV buffer (Appendix Table 3).

3.3.8 Site-directed mutagenesis

The PfIMP2 C71S mutant was generated using Q5 Quickchange Mutagenesis Kit (NEB, UK). Once the plasmid was mutated, the plasmid was transformed into DH5α cells (NEB, UK) as described in section 3.1.2.

3.4 NMR

3.4.1 NMR Sample preparation

Unlabelled and $^{13}C/^{15}N$ labelled samples were concentrated and buffer exchanged into Hepes GF buffer 10 mM DTT (Appendix Table 3) by repeatedly washing the sample with the buffer in a 10 kDa MWCO concentrator (Sartorius, UK). 10% D_2O (Sigma-Aldrich, UK) was added

to the sample and the sample transferred to a 5 mm NMR tube (Sigma-Aldrich, UK) for NMR analysis.

3.4.2 NMR experiments

^{1}H 1D and ^{15}N-HSQC spectra were recorded at 303 K on a Bruker Avancell 600 (600 MHz). ^{13}C/^{15}N labelled PfIMP2 (1.5 mM), with and without His tag, were used to collect ^{15}N-HSQC, HNCACB and HNCOCACB NMR spectra at 303 K on a Bruker Avancell 600 (600 MHz) and backbone assignment was reached using standard double- and triple-resonance techniques. The same samples were used to collect HBHA(CBCACO)NH, HCCH-TOCSY, (H)CC(CO)NH-TOCSY and ^{1}H-C^{13}NOESY-HMQC spectra at 303 K on Bruker Avancell 600. Side chain assignment of aliphatic and aromatic ^{1}H and ^{13}C atoms were reached using the data from these experiments.

3.5 ARIA

ARIA calculations were set up following successful backbone and side chain assignments using NMRView. TALOS+ was used to generate the dihedral angle restraints based on the chemical shift assignment. CCpnmr Analysis was used to create the NOESY peaklists. ARIA runs were set up under the following tolerance levels given in table 3.4.

13C-NOESY (800 MHz)		**15N-NOESY (800 MHz)**	
Direct ^{1}H frequency window (ppm)	0.04	Direct ^{1}H frequency window (ppm)	0.04
Direct ^{13}C frequency window (ppm)	0.6	Direct ^{15}N frequency window (ppm)	0.5
Indirect 1H frequency window (ppm)	0.06	Indirect 1H frequency window (ppm)	0.06
Indirect ^{13}C frequency window (ppm)	0.5	Indirect ^{15}N frequency window (ppm)	0.5

Chemical shift difference between ^{13}C planes (ppm)	0.25	Chemical shift difference between ^{15}N planes (ppm)	0.195

Table 3.4: Table showing the tolerance levels set for the 13C- and15N- NOESY datasets recorded on the Bruker Avancell 800 (800 MHz)

3.6 In vitro Experiments

3.6.1 Culturing *P. falciparum* parasites

P. falciparum 3D7 strain transfected with the pRM2_DD-3HA-GFP plasmid containing the PfIMP2 gene was cultured in 10 ml of complete media (Appendix Table 3) with 20 nM WR99210 (Sigma-Aldrich, UK). The parasites were cultured in complete media with 4% RBC, gassed with 5% CO_2, 5% O_2 and incubated at 37°C.

3.6.2 Sorbitol synchronisation

Parasites were synchronised when ring parasitemia was high in the culture. Synchronisation was performed following the standard sorbitol treatment protocol with 5% sorbitol (Lambros & Vanderberg, 1979).

3.6.3 Immunofluorescence assay

Samples were prepared in suspension and as fixed samples for IFA. The samples were fixed with 4% paraformaldehyde/ 0.01% gluteraldehyde/PBS (Sigma-Aldrich, UK). The samples were prepared following standard protocol. Both fixed and suspension samples were incubated with Anti-HA (rat) and Anti-GAP45 (rabbit) primary antibodies (ThermoFisher, UK) and with Anti-GFP (mouse) and Anti-GAP45 (rabbit) primary antibodies (ThermoFisher, UK). Following this, they were incubated with anti-rat 488 and anti-rabbit 594 secondary antibodies (ThermoFisher, UK) and anti-mouse 488 and anti-rabbit 594 secondary antibodies (ThermoFisher, UK), respectively with DAPI (ThermoFisher, UK).

The suspension samples were spread on a slide and air dried. Coverslips were mounted on both suspension and fixed sample slides with Prolong Gold (ThermoFisher, UK). The slides were allowed to dry flat overnight and observed under the Leica TCS SP5 II confocal microscope using the Leica LAS AF software (Imperial College London).

3.6.4 Pull down Assay

Wild type 3D7 strain parasites and parasites transfected with the pRM2_DD-3HA-GFP plasmid with the PfIMP2 gene were harvested at 6 % parasitemia and resuspended in parasite lysis buffer (Appendix Table 3). Following protein lysis, the samples were incubated with 500 µl of Pierce anti-HA magnetic beads (ThermoFisher, UK) in consecutive batches of 500 µl each. After each incubation period, the beads were separated on the magnetic stand and the supernatant was removed. Two washes were performed with the lysis buffer and a final elution was performed with parasite elution buffer (Appendix Table 3). The supernatant was analysed following the final elution by SDS-PAGE on a 12% SDS-PAGE gel (Serva, Germany).

3.6.5 Silver staining

The gel was silver stained using the SilverQuest Silver Stain kit (ThermoFisher, UK) following the protocol given in the kit.

Chapter 4

Characterisation of coccidian IMP-like homologues

Three IMP1 homologues were initially chosen for structural studies – TgIMP1, EtIMP1 and PfIMP2. The results obtained from cloning, expression, purification and preliminary NMR studies of the coccidian homologues (EtIMP1 and TgIMP1) have been discussed in this chapter.

4.1 Characterisation of IMP1-like homologue in *E. tenella*

In this study, the IMP1-like homologue gene from *E. tenella* (GenBank: JN987546) was amplified and cloned for recombinant expression in bacteria.

4.1.1 PCR Amplification

A short construct of the *E. tenella* IMP1 homologue was cloned into pET30 Xa/LIC (NEB, UK). The amino acid sequence of EtIMP1 clone 2 is given in Figure 4.1.1. The start point of EtIMP1 clone 2 is homologous to the start point of TgIMP1 clone 7 (optimised construct). The EtIMP1 amino acid sequence contains alanine and glutamine repeats and has the same domain architecture as TgIMP1.

MGGACGKSRGTAAAAEPPVSAADKAAEAAASAASQAEKAQEAAAAA
AAAAANGAAAAAALTGEDRSEREPANASDGKDTNPTNPTTAEQQQQ
QEQEQQQQEQQQQEQQQQEQQQEAAAAQQPASPVVALSAADAELL
AAAQQQAAAAAAAGSNANLPHAYLFYAAELNEGSLILQWTAAAMQQ
QEMQDKKLLLLASFVPPKYKTVTKSKLQQNGGITFLLQEMKYKWDI
WNKAQRQAYFQGWAKFLKAADEMEASLLLHPFELPAPPATVFLLHT
GPIENKVVPVKLGEPIGISMFGFAAVAPPPAPYKAGANITPKRFGELAT
QAGGAYIQLSRRGGDAAFSEADVVKWLAEDGLEIQQGNGITLDSTGA
YERRSDKKGGNVAAAT

Figure 4.1.1: Figure showing the amino acid sequence of full length EtIMP1 and the short construct (clone 2) produced for structural studies in brown. The sequence has numerous Ala (green) and Gln (purple) repeats. The plasma membrane localisation signal is labelled blue, predicted myristoylation site labelled grey, predicted palmitoylation site labelled black, linker labelled red and the IMP domain labelled brown.

4.1.2 Colony screening

Following cloning of EtIMP1 clone 2 into pET30 Xa/LIC vector and transformation in DH5α cells, the PCR products from colony PCR were run in a 1% agarose gel and the result is

shown in Figure 4.1.2. The pET30 Xa/LIC vector contains C-terminal His and S tags. Sequencing results confirmed that the positive colonies contain the plasmid encoding EtIMP1 clone 2.

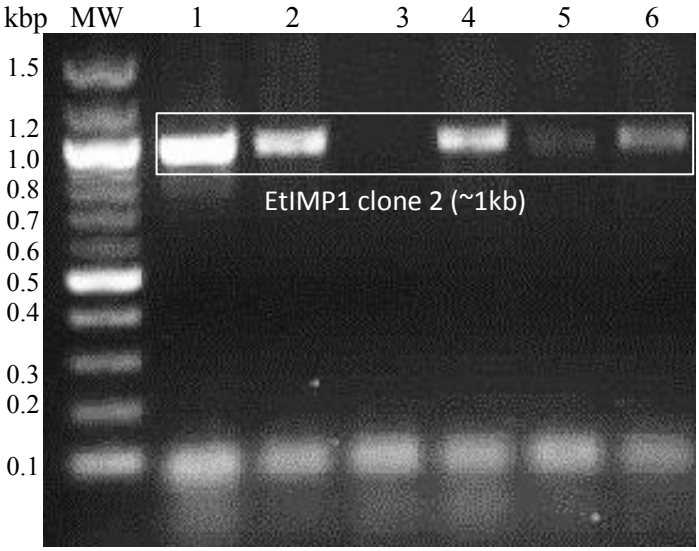


Figure 4.1.2: Figure showing the PCR products obtained from colony PCR of EtIMP1 clone 2 (pET30 Xa/LIC). Six colonies were tested out of which five were positive (white box). The band is around 1.0 kb which is the expected size of EtIMP1 clone 2. The six labelled lanes refer to colonies 1 to 6.

4.2.3 Expression test

EtIMP1 clone 2 (pET30 Xa/LIC) was transformed into Rosetta2 DE3 cells and expression was tested on small scale to check for protein overexpression and solubility. The molecular weight of the protein (including the His and S tags) is 28.3 kDa. The SDS PAGE results are shown in figure 4.1.3.

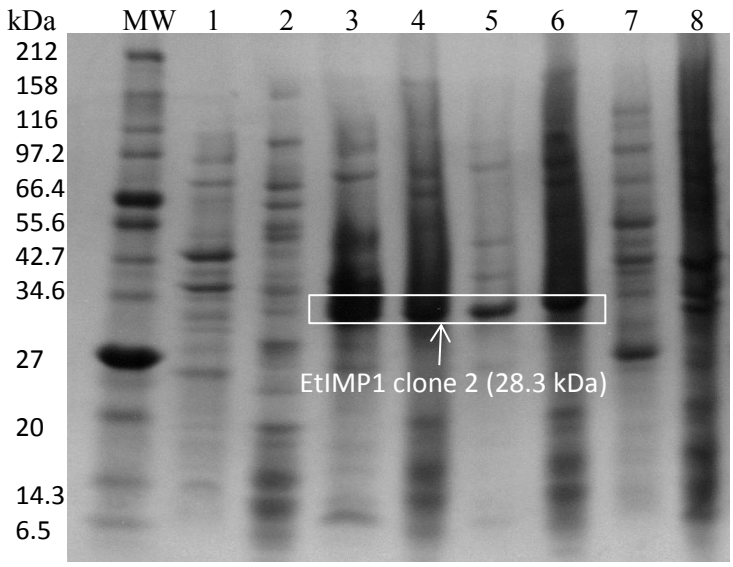


Figure 4.1.3: Figure showing the expression test results obtained from EtIMP1 clone 2. The white box shows the bands that correspond to EtIMP1 clone 2. This confirms soluble overexpression of the protein. The lanes in the gel are (1) Pre-induction (soluble fraction) (2) Pre-induction (insoluble fraction) (3) 24 hours after induction at 18°C (soluble fraction) (4) 24 hours after induction at 18°C (insoluble fraction) (5) 24 hours after induction at 22°C (soluble fraction) (6) 24 hours after induction at 22°C (insoluble fraction) (7) 24 hours after induction at 37°C (soluble fraction) (8) 24 hours after induction at 37°C (insoluble fraction).

4.1.4 Protein expression and purification

The average mass of cells, once harvested, was 5 g per litre of culture (with a standard deviation of ±0.5 g) for EtIMP1 clone 2 expression. The concentration of protein after Nickel affinity chromatography was 51.33 mg/ml in 1 ml. The SDS PAGE gel obtained from running the fractions from Nickel affinity chromatography for EtIMP1 clone 2 (pET30 Xa/LIC expressed in Rosetta 2 cells) shows that the protein is pure following purification (Figure 4.1.4).

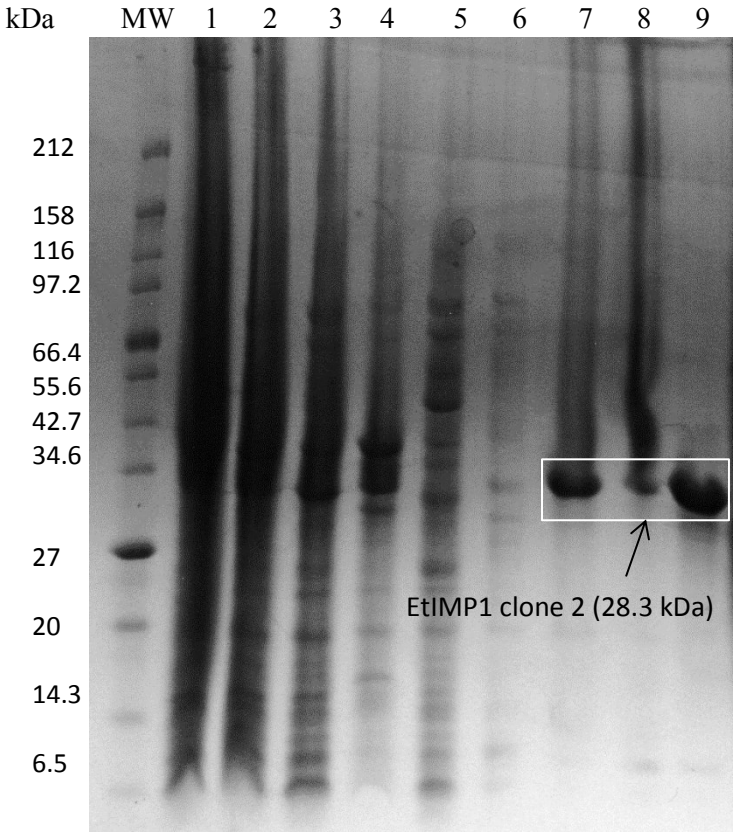


Figure 4.1.4: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of EtIMP1 clone 2. The white box shows the bands that correspond to EtIMP1 clone 2 at the expected molecular weight of 28.3 kDa. The lanes in the gel are (1) Pre-lysis sample (2) Post-lysis sample (3) Clear Lysate (4) Insoluble pellet (5) Flow through following Nickel affinity chromatography (6) Wash (7) Eluent 1 (8) Eluent 2 (9) Eluent 3

4.1.5 Limited proteolysis

Trypsin digestion of EtIMP1 clone 2 was not successful and no cleavage was observed. Chymotrypsin digestion of EtIMP1 clone 2 was performed to identify any stable fragments of EtIMP1 clone 2 which may be suitable for NMR studies (Figure 4.1.5a). Two stable bands were observed in an SDS-PAGE gel and sequence analysis of this band following Western blotting showed chymotrypsin cleavage sites as shown in Figure 4.1.5b. However, these protein fragments are too small compared to the full length EtIMP1 protein suggesting that the folded domain is being broken down by Chymotrypsin.

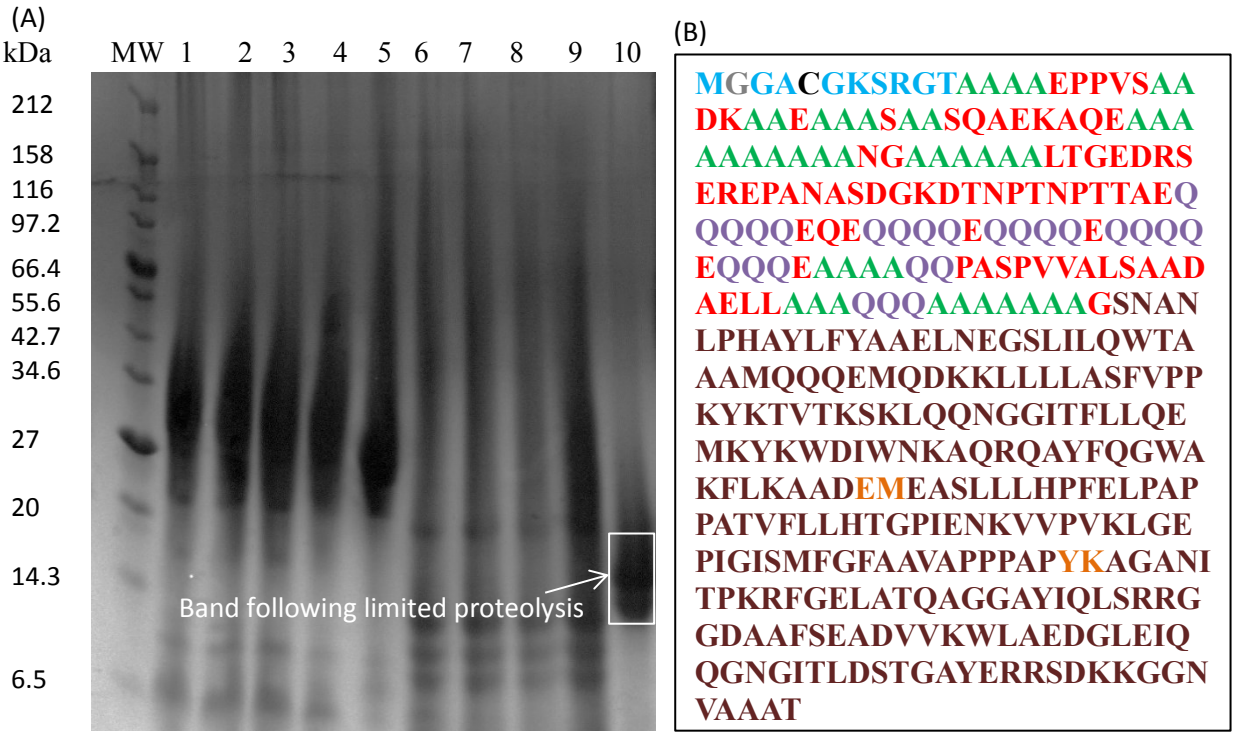


Figure 4.1.5: Figure showing (A) the SDS PAGE gel of the samples obtained during limited proteolysis of EtIMP1 clone 2 using chymotrypsin. The white box shows the bands that correspond to the bands produced by limited chymotrypsin digestion. The lanes in the gel are (1) 5 minutes after addition of trypsin to EtIMP1 clone 2 (2) 15 minutes after trypsin addition (3) 30 minutes after trypsin addition (4)1 hour after trypsin addition (5) 1 day after trypsin addition (6) 5 minutes after addition of chymotrypsin to EtIMP1 clone 2 (7) 15 minutes after chymotrypsin addition (8) 30 minutes after chymotrypsin addition (9) 1 hour after chymotrypsin addition (10) 1 day after chymotrypsin addition.

(B) Sequence of full length EtIMP1. The Ala (green) and Gln (purple) repeats are labelled in the sequence. Two Chymotrypsin digestion sites were found and chymotrypsin cleaves between the residues labelled orange. The plasma membrane localisation signal is labelled blue, predicted myristoylation site labelled grey, predicted palmitoylation site labelled black, linker labelled red and the IMP domain labelled brown.

4.1.6 ^{1}H 1D NMR

^{1}H 1D NMR experiment was performed on the pET30 Xa/LIC construct of EtIMP1 clone 2 expressed in Rosetta 2 cells. The ^{1}H 1D NMR spectrum of this construct shows that it is folded (Figure 4.1.6). It is evident from figure 4.1.6 that the signal from the unfolded domain of the protein (0.8 ppm) is much more intense than the signal from the folded domain as observed in the methyl region of the spectrum.

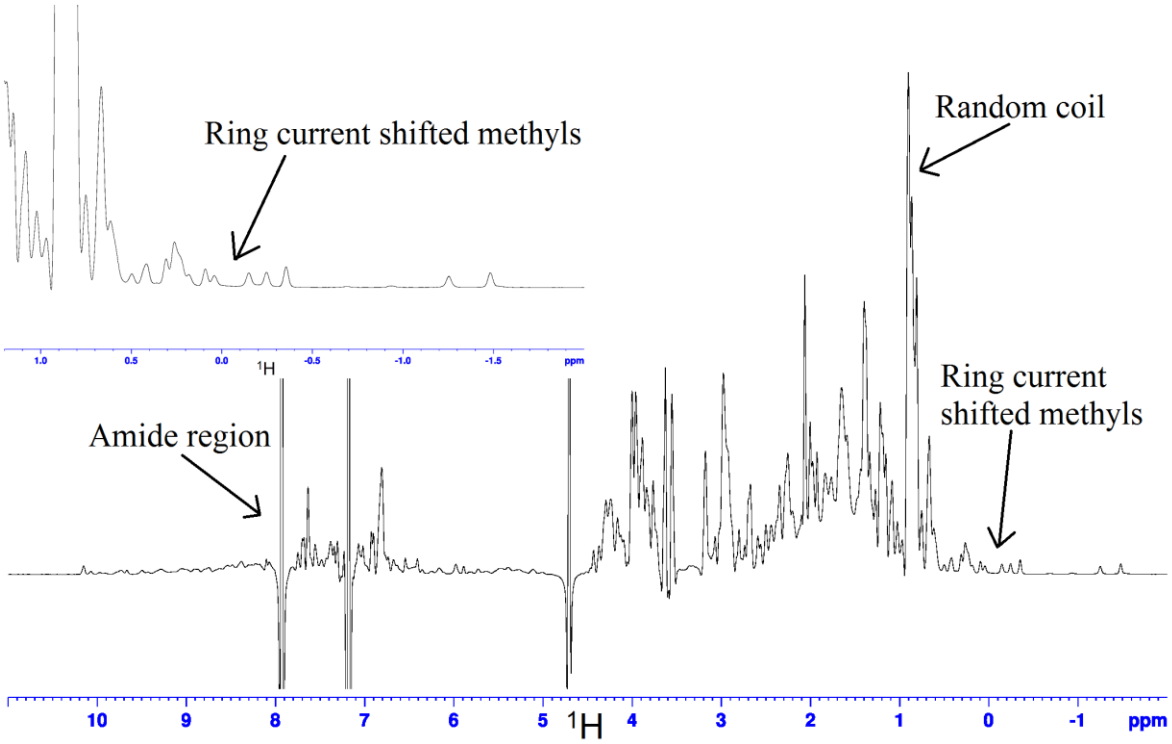


Figure 4.1.6: Figure showing the ^{1}H 1D spectrum of the EtIMP1 clone 2 (pET30 Xa/LIC). On the top left, a zoom in of the methyl region of the spectrum between 1.5 ppm and -2.00 ppm is shown. The presence of ring current shifted peaks in the methyl region shows that the protein is folded. However, the large signal from the unfolded domain of the protein at 0.8 ppm is masking the signal from the folded domain.

4.1.7 ^{15}N TROSY-HSQC

Double labelled ($^{13}C/^{15}N$) EtIMP1 clone 2 sample in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer was used to record ^{15}N TROSY-HSQC spectrum (Figure 4.1.7). Although this spectrum shows that the protein is folded, there are ~170 peaks in the spectrum which is less than the number of amino acids (236 amino acids) in the protein suggesting that the data is not complete. The molecular weight of this construct is 28.3 kDa which is higher than the optimum sample size for NMR studies.

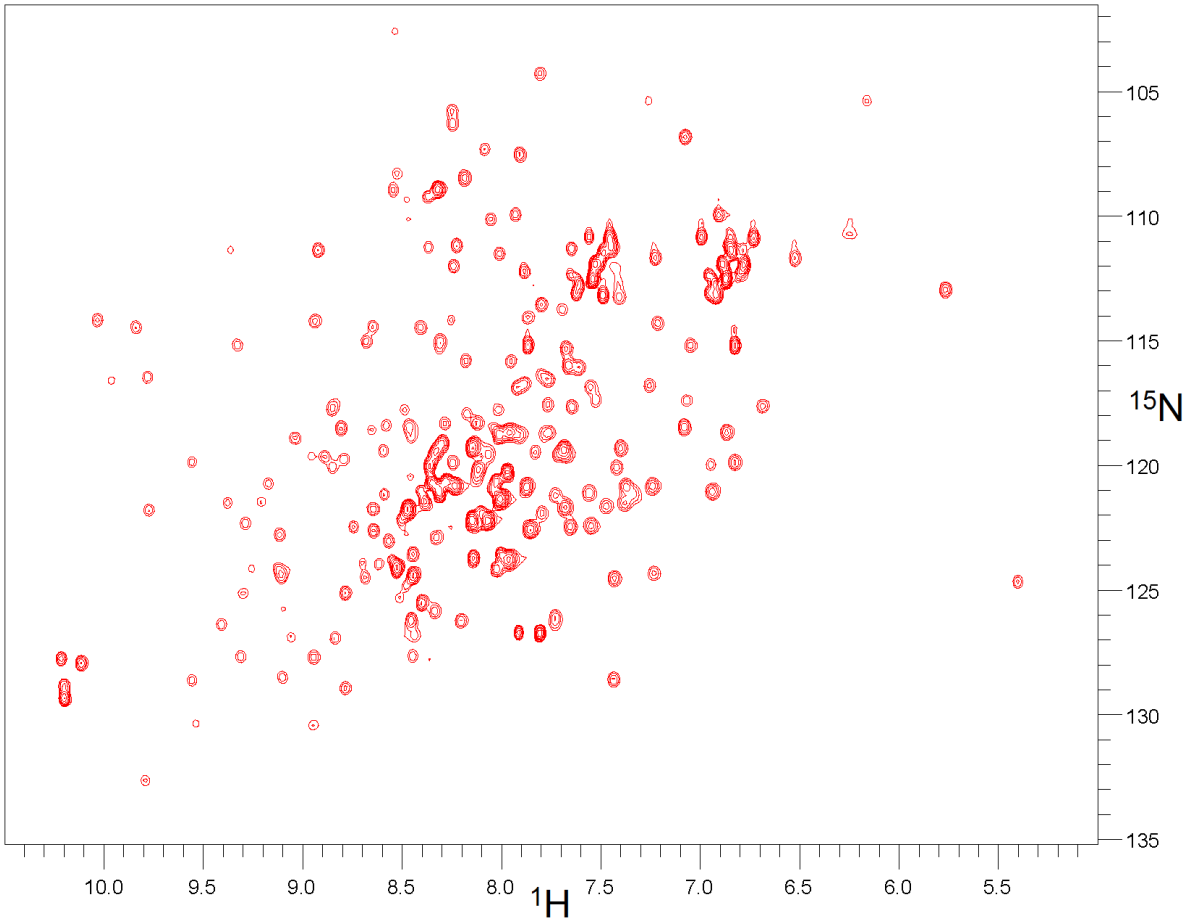


Figure 4.1.7: Figure showing the 2D ^{15}N TROSY- HSQC spectrum of double labelled EtIMP1 clone 2. The number of peaks observed in the spectrum is ~170 whereas there are 236 amino acids in the protein sequence suggesting that the data observed in the ^{15}N HSQC spectrum is incomplete.

4.2 Characterisation of IMP1-like homologue in *T. gondii*

In this study, the IMP1-like homologue gene from *T. gondii* ME49 (GenBank JN657189) was amplified and cloned for recombinant expression in bacteria.

4.2.1 Bioinformatics analysis

A BLAST search of the TgIMP1 amino acid sequence showed that TgIMP1 protein has homologues only within apicomplexans. The first six hits obtained from BLAST are shown in Figure 4.2.1a (Altschul et al, 1990). None of the IMP1 homologues have a known structure or function. Secondary structure predictions of TgIMP1 were performed using Phyre (Bennett-Lovsey et al, 2008) and JPred (Cuff et al, 1998) using the TgIMP1 amino acid sequence. Results from both prediction servers suggested that TgIMP1 has a disordered N-terminal region. Sequence alignment was performed using ESPript 3.0 (Robert & Gouet, 2014). The sequence identity scores calculated based on the sequence alignments of these homologues is also shown in Figure 4.2.1b.

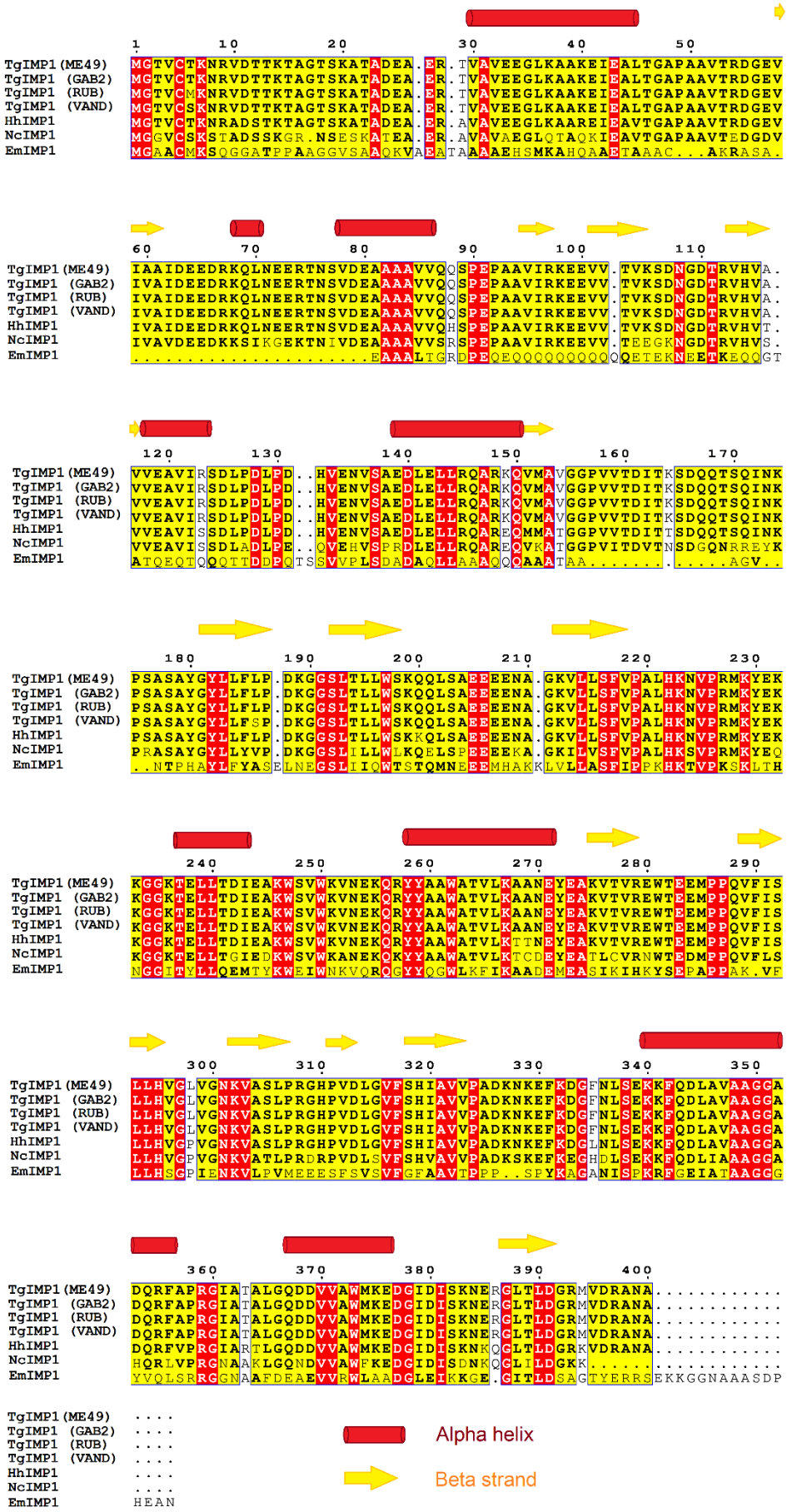

TgIMP1 (ME49)
TgIMP1 (GAB2)
TgIMP1 (RUB)
TgIMP1 (VAND)
HhIMP1
NcIMP1
EmIMP1
Alpha helix
Beta strand

Figure 4.2.1a: Figure showing the top six results from the BLAST search performed with the Toxoplasma gondii ME49 IMP1 homologue sequence (Robert & Gouet, 2014). Residues highlighted red are identical and those highlighted yellow are conserved. The secondary structure prediction of the TgIMP1 (ME49) protein generated by JPred is shown above the sequences (Cuff et al, 1998).

The host organisms that the homologues are from are – TgIMP1 ME49 is T. gondii ME49 strain, TgIMP1 (GAB2) is T. gondii GAB2-2007-GAL-DOM2, TgIMP1 (RUB) is T. gondii RUB strain, TgIMP1 (VAND) is T. gondii VAND strain, HhIMP1 is Hammondia hammondi, NcIMP1 is Neospora caninum Liverpool strain and EmIMP1 is Eimeria mitis.

TgIMP1 (ME49)	100%						
TgIMP1 (GAB2)	99.75%	100%					
TgIMP1 (RUB)	99.5%	99.75%	100%				
TgIMP1 (VAND)	99.25%	99.5%	99.5%	100%			
HhIMP	94%	94.25%	94%	93.75%	100%		
NcIMP1	72.26%	72.51%	72.51%	72.77%	75.31%	100%	
EmIMP1	28.64%	28.64%	28.91%	28.64%	28.38%	28.64%	100%
	TgIMP1 (ME49)	**TgIMP1 (GAB2)**	**TgIMP1 (RUB)**	**TgIMP1 (VAND)**	**HhIMP1**	**NcIMP1**	**EmIMP1**

Figure 4.2.1b: Figure showing the sequence identity scores calculated using the BLOSUM62 scoring matrix based on the sequence alignments of the homologues shown in Figure 4.2.1a (Henikoff & Henikoff, 1992).

The host organisms that the homologues are from are – TgIMP1 ME49 is T. gondii ME49 strain, TgIMP1 (GAB2) is T. gondii GAB2-2007-GAL-DOM2, TgIMP1 (RUB) is T. gondii RUB strain, TgIMP1 (VAND) is T. gondii VAND strain, HhIMP1 is Hammondia hammondi, NcIMP1 is Neospora caninum Liverpool strain and EmIMP1 is Eimeria mitis.

4.2.2 Optimisation of TgIMP1 construct

The TgIMP1 gene was amplified from the cDNA library and cloned into the pET46 Ek/LIC vector (NEB, UK). Seven constructs of TgIMP1 were designed based on the secondary structure prediction. Based on the secondary structure prediction, the N-terminal start point of two homologues were set at amino acids 90 and 138, to remove disordered regions from the constructs. The C-terminal truncation was defined at residue 350 as it is the natural stop point for some homologous sequences (not shown). The list of the constructs and their molecular weights are given in Table 4.1.

Construct Name	**Amino acids**	**Molecular Weight (kDa)**
TgIMP1 Clone 1 (Full length)	1 to 400	43.5
TgIMP1 Clone 2	1 to 350	38.1
TgIMP1 Clone 3	91 to 350	28.7
TgIMP1 Clone 4	91 to 400	34.1
TgIMP1 Clone 5	138 to 350	23.6
TgIMP1 Clone 6	138 to 400	29.0
TgIMP1 Clone 7	196 to 400	27.0

Table 4.1: Table showing a list of the TgIMP1 constructs, the amino acids in these constructs and their molecular weights in kDa.

4.2.3 Expression tests

Once the TgIMP1 constructs were cloned into pET46 Ek/LIC vector, *E. coli* was transformed and plated on Agar plates. Cells were then checked for expression, as this is the first attempt to express TgIMP1 in *E. coli*. The results from the expression test showed bands around the

expected molecular weight confirming the overexpression of TgIMP1. A figure showing the expression test performed for full length TgIMP1 is shown in Figure 4.2.2.

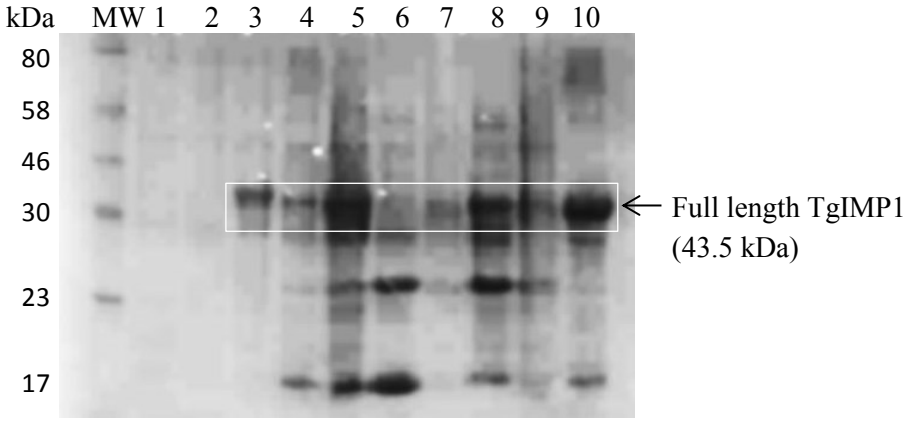


Figure 4.2.2: Figure showing the expression test results obtained for full length TgIMP1 expressed in Rosetta 2 cells. The white box shows that the bands are around 43.5 kDa which corresponds to the molecular weight of full length TgIMP1 protein. The lanes in the gel are (1) Pre-induction (insoluble fraction) (2) Pre-induction (soluble fraction) (3) 2 hours after induction at 37°C (insoluble fraction) (4) 2 hours after induction at 37°C (soluble fraction) (5) 24 hours after induction at 37°C (insoluble fraction) (6) 24 hours after induction at 37°C (soluble fraction) (7) 24 hours after induction at 22°C (insoluble fraction) (8) 24 hours after induction at 22°C (soluble fraction) (9) 24 hours after induction at 18°C (insoluble fraction) (10) 24 hours after induction at 18°C (soluble fraction)

4.2.4 ^{1}H 1D NMR

^{1}H 1D NMR experiments were performed on all the TgIMP1 constructs produced to check if the protein is folded and if the construct is suitable for further NMR studies. A comparison of the intensity of signal from the folded and unfolded regions in the protein would indicate whether or not the construct is suitable for structural studies. The signal from the random coil regions in the protein can be observed around 0.8 ppm. The ^{1}H 1D NMR spectrum of the full length TgIMP1 protein is shown in Figure 4.2.3. All the ^{1}H 1D experiments were performed on samples in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer at 303 K. The full length protein was found to have a long, unfolded region which was masking the signal from the folded domain (Figure 4.2.3).

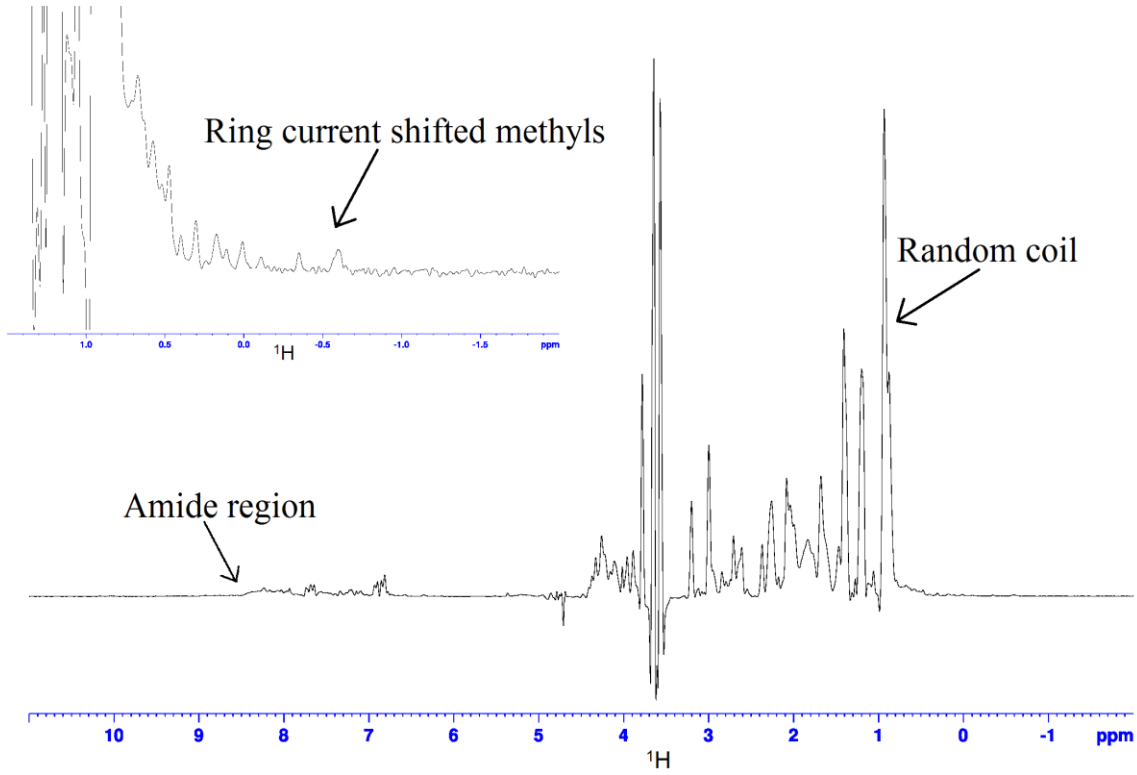


Figure 4.2.3: Figure showing the ^{1}H 1D spectrum of the full length TgIMP1 protein. On the top left, a zoom in of the methyl region of the spectrum between 1.5 ppm and -2.00 ppm is shown. The presence of ring current shifted peaks in the methyl region shows that the protein is folded. However, the large signal from the random coil regions in the protein at 0.8 ppm is masking the signal from the folded domain.

Following successful cloning, expression and purification of TgIMP1 clone 6, a ^{1}H 1D spectrum was recorded as shown in Figure 4.2.4. This spectrum shows an improvement compared to the full length TgIMP1 as the intensity of signal from the unfolded region is lower and the signal from the folded domain appears more resolved.

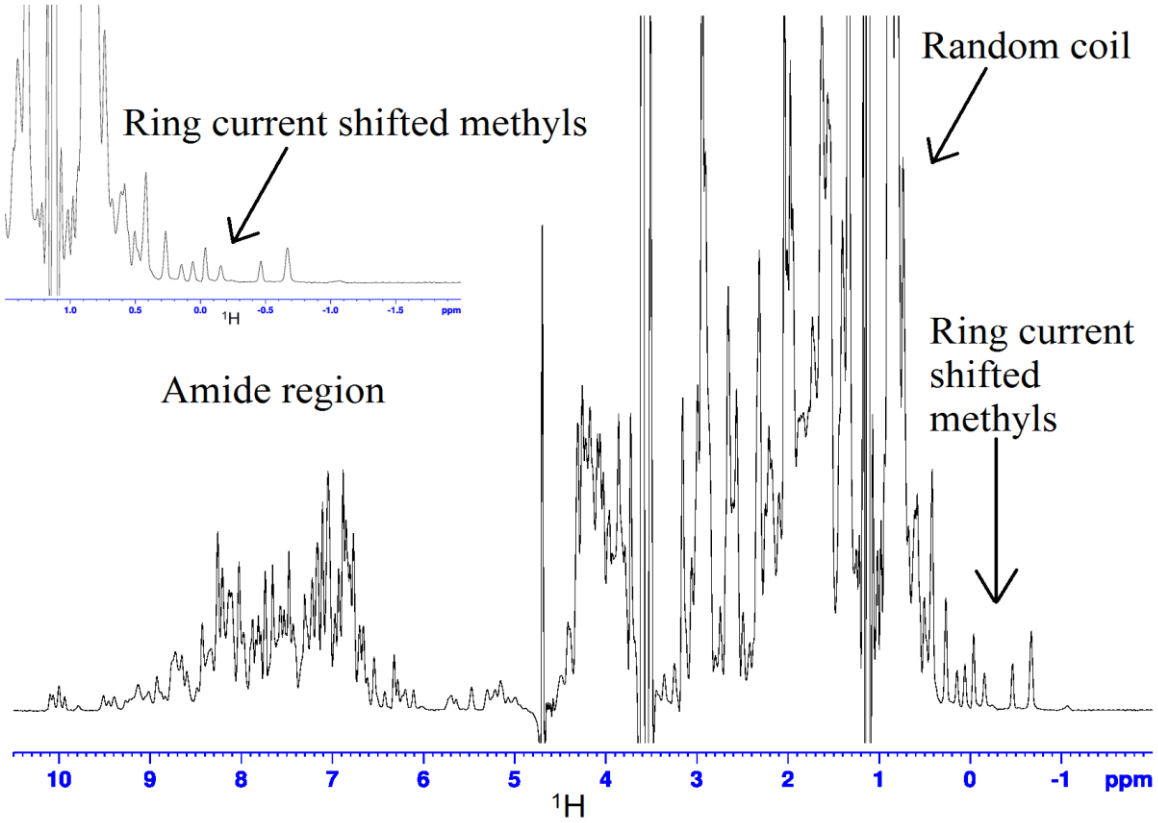
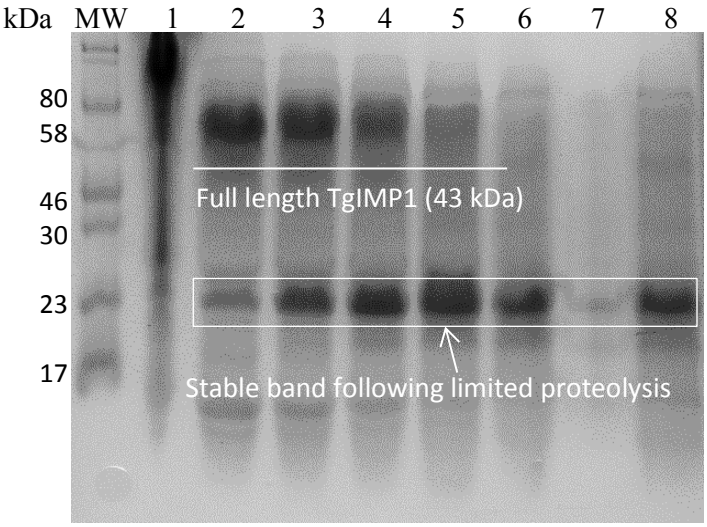


Figure 4.2.4: Figure showing the ^{1}H 1D spectrum of TgIMP1 clone 6. On the top left, a zoom in of the methyl region of the spectrum between 1.5 ppm and -2.00 ppm is shown. The presence of ring current shifted peaks in the methyl region shows that the protein is folded. The signal from the unfolded domain of the protein at 0.8 ppm is smaller than for full length TgIMP1.

4.2.5 Limited proteolysis

Chymotrypsin digestion of full length TgIMP1 was performed to identify any stable fragments of TgIMP1 that may be suitable for NMR studies. A short stable band was observed on an SDS-PAGE gel (Figure 4.2.5a) and sequence analysis of the band following Western blotting showed three chymotrypsin cleavage sites which are shown in Figure 4.2.5b.

(a)



(b)

MGTVCTKNRVDTTKTAGTSKATADEAERTVAVEEGLKAAKEIEALTGAPAAVTR
DGEVIAAIDEEDRKQLNEERTNSVDEAAAAVVQQSPEPAAVIRKEEVVTVKSDNG
DTRVHVAVVEAVIRSDLPDLPDHVENVSAEDLELLRQARKQVMAVGGPVVTDIT
KSDQQTSQINKPSASAYGYLLFLPDKGGSLTLLWSKQQLSAEEEENAGKVLLSFVP
ALHKNVPRMKYEKKGGKTELLTDIEAKWSVWKVNEKQRYYAAWATVLKAANE
YEAKVTVREWTEEMPPQVFISLLHVGLVGNKVASLPRGHPVDLGVFSHIAVVPAD
KNKEFKDGFNLSEKKFQDLAVAAGGADQRFAPRGIATALGQDDVVAWMKEDGI
DISKNERGLTLDGRMVDRANA

Figure 4.2.5: Figure showing (a) SDS-PAGE analysis of samples obtained following limited proteolysis of full length TgIMP1 with Chymotrypsin. The lanes in the gel are (1) Full length TgIMP1 (2) 15 min following Chymotrypsin digestion (3) 30 min following Chymotrypsin digestion (4) After 1 hour (5) After 2 hours (6) After 3 hours (7) Empty lane and (8) After 1 day. The full length TgIMP1 protein ran at a higher molecular weight of ~70 kDa (expected molecular weight = 43 kDa).

(b) Sequence of full length TgIMP1. Three Chymotrypsin digestion sites were found and chymotrypsin cleaves between the amino acids labelled green. The plasma membrane localisation signal is labelled blue, predicted myristoylation site labelled grey, predicted palmitoylation site labelled black, linker labelled red and the IMP domain labelled brown.

4.2.6 Optimised TgIMP1 construct

The start point for TgIMP1 clone 7 was decided to be S176 based on the results from limited proteolysis of TgIMP1 (Figure 4.2.5b) in conjunction with the secondary structure predictions. TgIMP1 clone 7 was cloned into pET46-Ek/LIC vector (NEB, UK). This construct starts at S176 and ends at A400, which is the C-terminal residue in full length TgIMP1. The molecular weight of this construct is 27 kDa including the N-terminal His tag.

Following cloning and transformation into DH5α cells, the PCR products from colony PCR were run on a 1% agarose gel. The result for TgIMP1 clone 7 is shown in Figure 4.2.6.

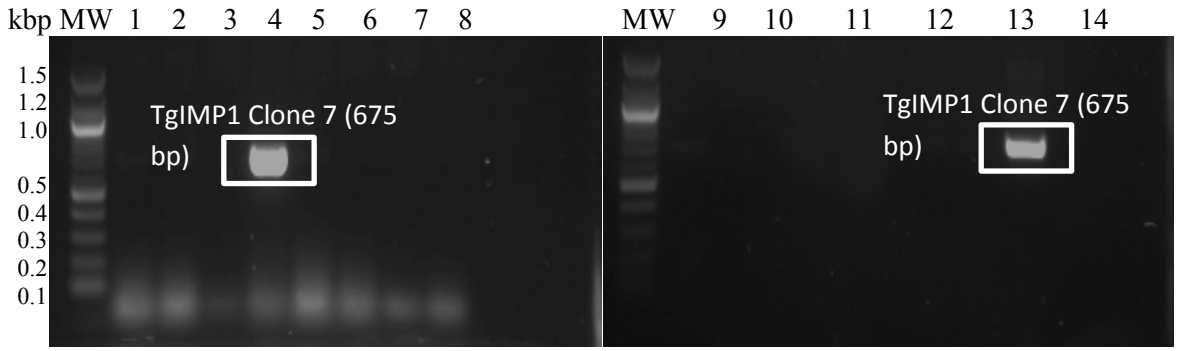


Figure 4.2.6: Figure showing the PCR products obtained from colony PCR of TgIMP1 clone 7. 15 colonies were tested but only two colonies were positive (white box). The band lies between 0.5 kb and 1.0 kb which is the expected range for TgIMP1 clone 7. The 15 lanes labelled refer to colonies 1 to 15.

Sequencing results confirmed that the positive colonies contain the plasmid encoding TgIMP1 clone 7.

4.2.7 Protein Expression and Purification

The average mass of cells, once harvested, was 5 g per litre of culture for all clones of TgIMP1. The concentration of TgIMP1 clone 7 after nickel affinity chromatography was 41.6 mg/ml in 2 ml and following size exclusion chromatography was 51.9 mg/ml in 500 µl. The SDS PAGE gels obtained from running the fractions from nickel purification and gel filtration for TgIMP1 clone 7 are shown in Figure 4.2.7a and Figure 4.2.7b, respectively. The results from SDS-PAGE following purification by size exclusion chromatography shows a single intense band at 27 kDa which shows that the protein is pure.

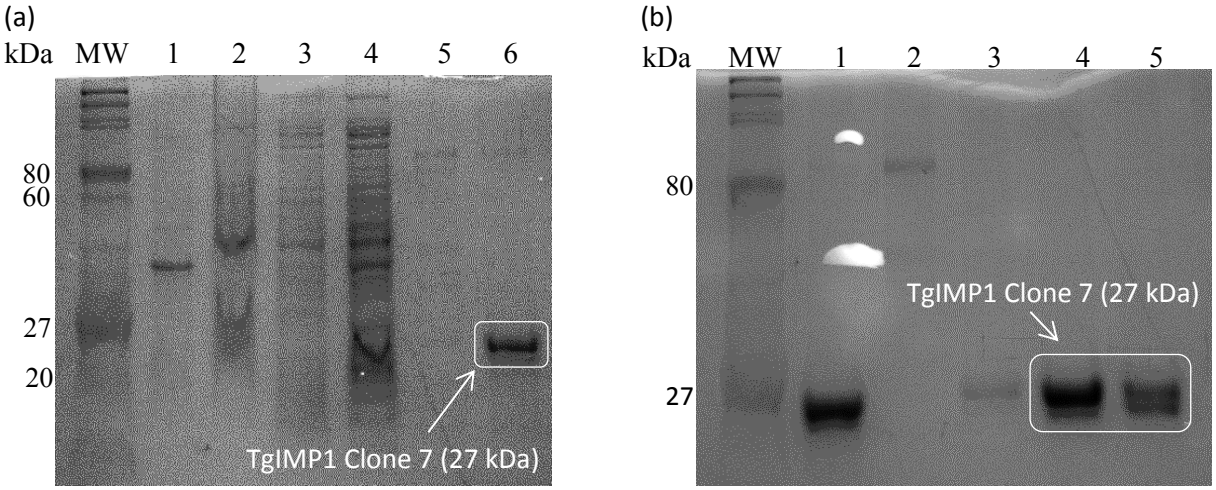


Figure 4.2.7: Figure showing (a) SDS-PAGE analysis of samples obtained during Nickel affinity chromatography of TgIMP1 clone 7. The white box shows the band around 27 kDa which corresponds to the molecular weight of TgIMP1 clone 7. The lanes in the gel are (1) Insoluble pellet (2) Clear lysate (3) Wash following Nickel purification (4) Flow through following Nickel purification (5) Eluent 1 and (6) Eluent 2

(b) SDS-PAGE analysis of samples obtained during Size exclusion chromatography of TgIMP1 clone 7. The white box shows the bands around 27 kDa which correspond to the molecular weight of TgIMP1 clone 7. The lanes in the gel are (1) Eluent from Nickel purification (2) Void (3) Peak 1 from size exclusion chromatography (4) Middle of Peak 2 and (5) End of Peak 2

4.2.8 ^{1}H 1D NMR

A ^{1}H 1D spectrum of ^{13}C/^{15}N labelled TgIMP1 clone 7 in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer was recorded at 303 K (Figure 4.2.8). This showed that the signal from the unfolded region of the protein was lower than that observed for Clone 6 (Figure 4.2.4). The signal from the folded domain is also more resolved in the TgIMP1 clone 7 spectrum.

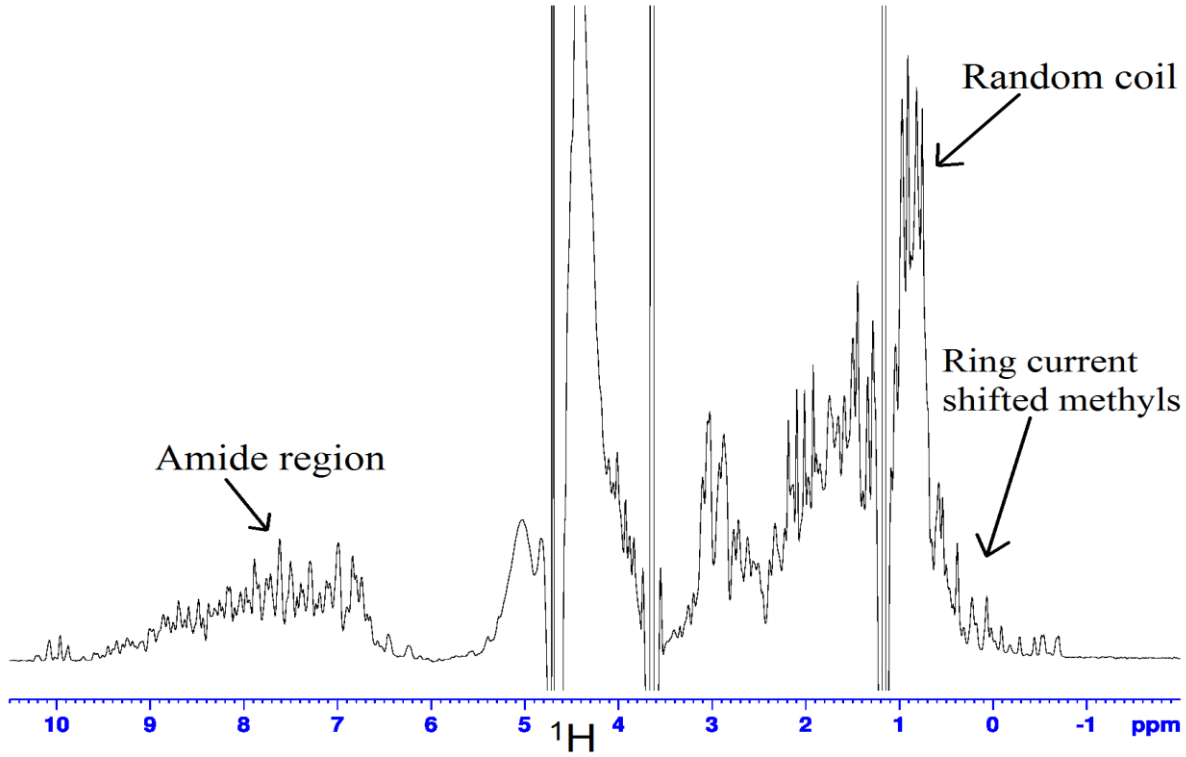


Figure 4.2.8: Figure showing the ^{1}H 1D spectrum of the TgIMP1 clone 7 ($^{13}C/^{15}N$ labelled). The presence of ring current shifted peaks in the methyl region (1 ppm to -1 ppm) shows that the protein is folded. The signal from the unfolded domain of the protein at 0.8 ppm is smaller than for TgIMP1 clone 6.

4.2.9 ^{15}N TROSY-HSQC

Double labelled ($^{13}C/^{15}N$) TgIMP1 clone 7 sample in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer was used to record ^{15}N TROSY-HSQC spectrum (Figure 4.2.9). The peaks were dispersed across the ^{15}N TROSY-HSQC spectrum which showed that the protein is folded and suitable for further structural studies.

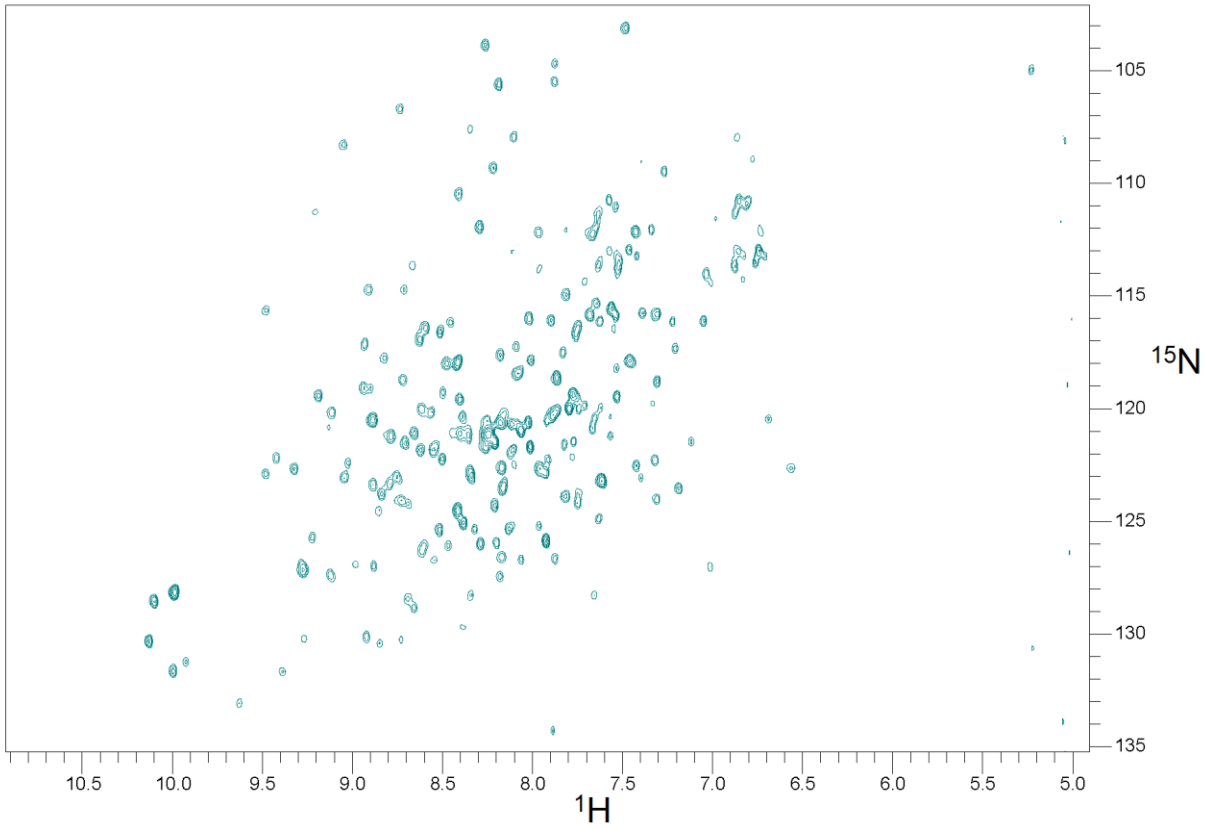


Figure 4.2.9: Figure showing the ^{15}N TROSY-HSQC spectrum of the TgIMP1 clone 7 ($^{13}C/^{15}N$ labelled). The presence of dispersed peaks shows that the protein is folded.

4.2.10 First attempt at backbone assignment

To assign the backbone resonances of TgIMP1 clone 7, a 2D $^{1}H/^{15}N$ HSQC spectrum was recorded under the optimized conditions (303K, pH 7). A series of 3D experiments were then recorded in order to assign the specific residues in the ^{15}N HSQC: HNCO, HNCACO, HNCACB and HNCOCACB. Backbone assignment was not feasible as majority of the protein lacked 3D data.

4.3 Further NMR studies of TgIMP1

Preliminary studies of both coccidian homologues of IMP1 show that TgIMP1 is preferable for further NMR studies due to the lower molecular weight of the construct and the well dispersed nature of the ^{15}N TROSY-HSQC. However, lack of complete 3D data for this construct raised the need for deuteration of the protein.

4.3.1 Deuteration of TgIMP1 clone 7

Expression tests were performed to test for soluble protein overexpression of TgIMP1 clone 7 by Rosetta 2 cells in D_2O minimal medial. SDS-PAGE results, following Nickel affinity chromatography, show that soluble overexpression of TgIMP1 clone 7 is obtained in D_2O minimal media.

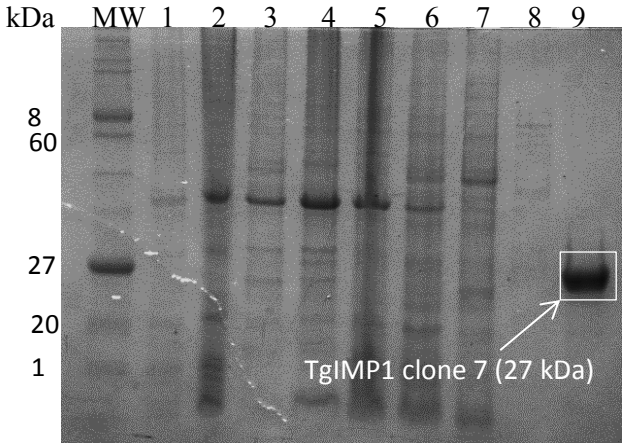


Figure 4.3.1a: Figure showing the expression test results obtained for TgIMP1 clone 7 expression in Rosetta 2 cells in D_2O minimal media. The white box shows the band is slightly lower than the expected molecular weight of 27 kDa for TgIMP1 clone 7. The lanes in the gel are (1) Pre-induction (soluble fraction) (2) Pre-induction (insoluble fraction) (3) 24 hours after induction at 18°C (soluble fraction) (4) 24 hours after induction at 18°C (insoluble fraction) (5) Clear lysate (6) Flow through following Nickel purification (7) Wash (8) Eluent 1 (9) Eluent 2

^1H 1D NMR of TgIMP1 clone 7 shows that deuteration was successful (Figure 4.3.1b). The intensity of the methyl peaks is lower in the deuterated sample which shows that the signal from the protons has decreased owing to the replacement of ^1H with ^2H atoms.

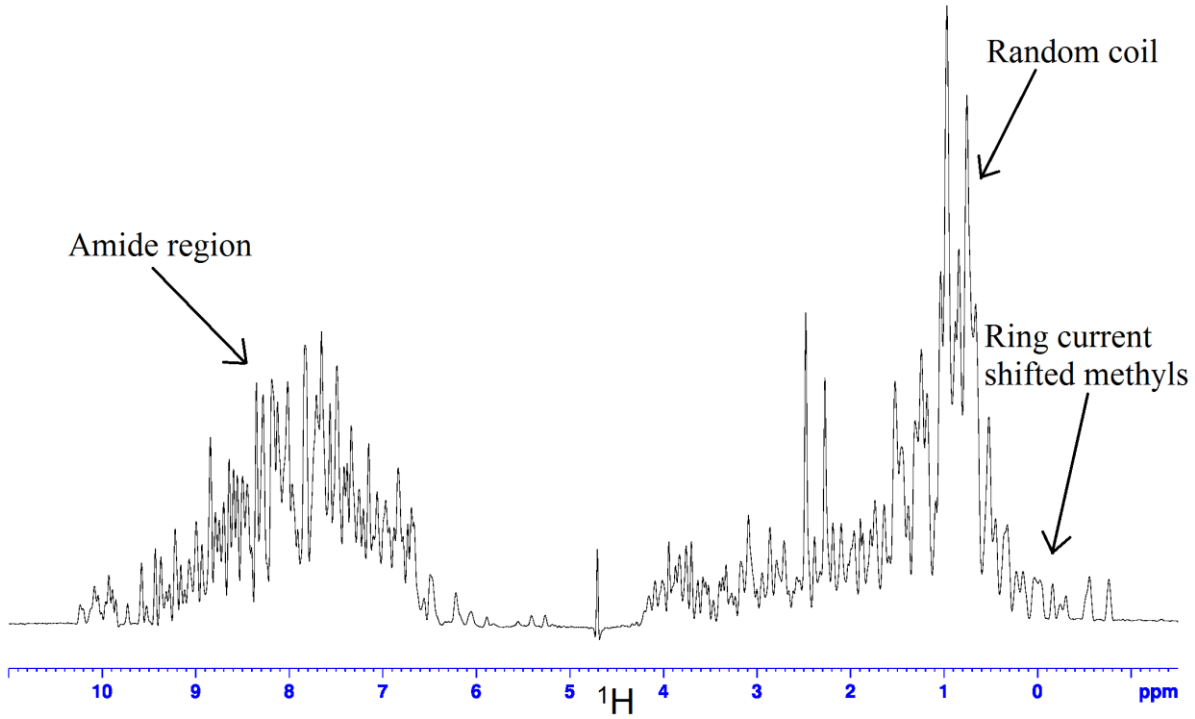


Figure 4.3.1b: Figure showing the ^{1}H 1D spectrum of the TgIMP1 clone 7 ($^{13}C/^{15}N/^{2}H$ labelled). The intensity of the methyl peaks in the deuterated spectrum is lower in comparison to the intensity of methyl peaks in the protonated spectrum (Figure 4.1.8).

4.3.2 ^{15}N TROSY-HSQC

Triple labelled ($^{13}C/^{15}N/^{2}H$) TgIMP1 clone 7 sample in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer was used to record ^{15}N TROSY-HSQC spectra. Comparison of the double and triple labelled spectra shows that some peaks are missing in the spectrum of the deuterated sample possibly due to lack of proton exchange in some regions of the deuterated protein (Figure 4.3.2).

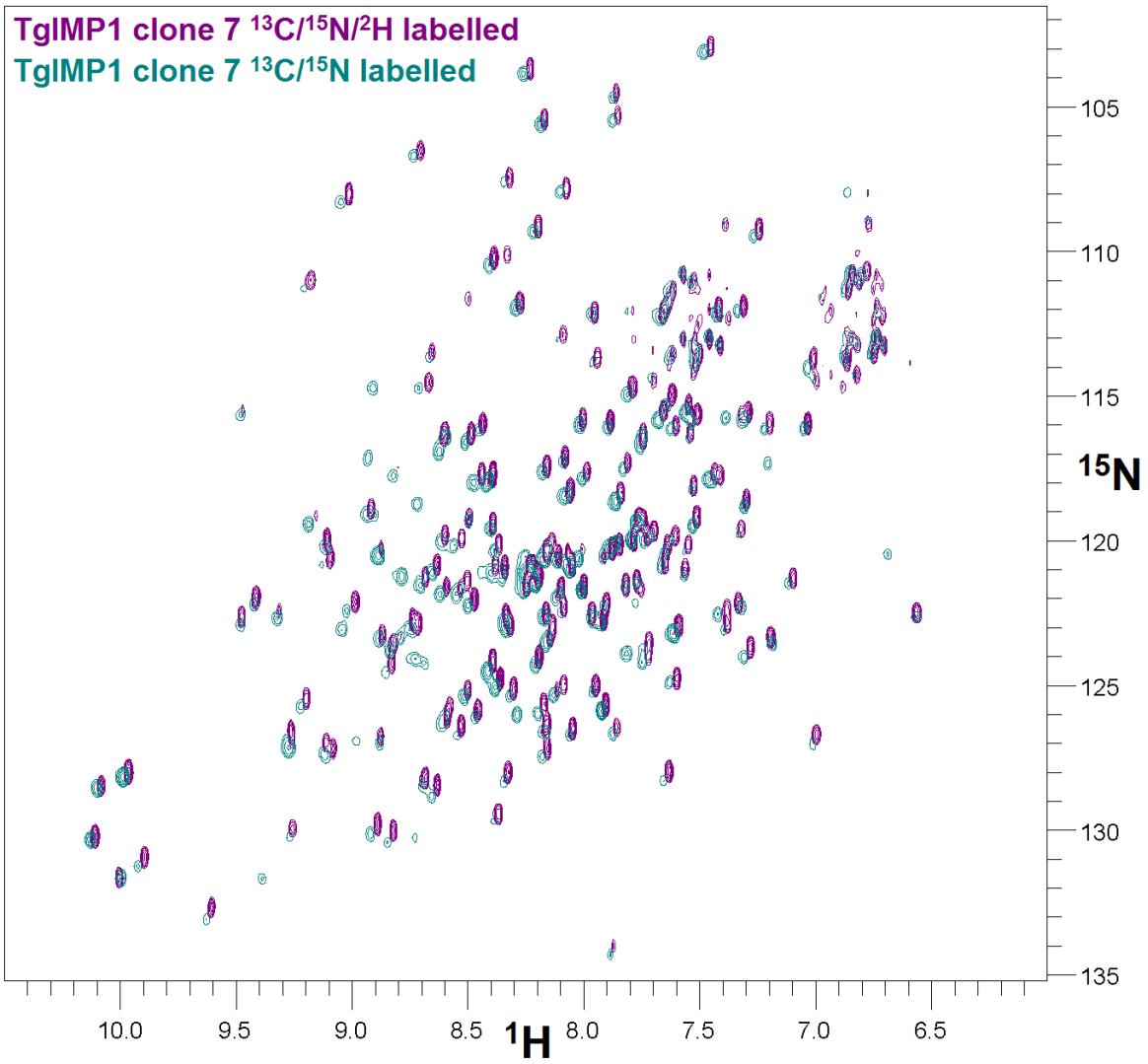


Figure 4.3.2: Figure showing an overlay of the 2D ^{15}N HSQC spectra of the double (green) and triple (purple) labelled TgIMP1 clone 7 protein. Some of the peaks present in the double labelled spectrum are missing in the triple labelled spectrum due to incomplete proton exchange of the deuterated sample.

4.3.3 Backbone assignment of TgIMP1 clone 7

To assign the backbone resonances of TgIMP1 clone 7, a 2D ^{15}N HSQC spectrum was recorded under the optimized conditions (303K, pH 7) with the $^{13}C/^{15}N/^{2}H$ TgIMP1 clone 7 sample. A series of 3D experiments were then recorded in order to assign the specific residues in the ^{15}N HSQC: HNCO, HNCACO, HNCACB and HNCOCACB. The chemical shifts were assigned using a modified version of the program NMRview (Marchant et al, 2008) and the program MARS (Jung & Zweckstetter, 2004) was used to fit the individual

residues into the sequence, based on matching the chemical shift values between residues. Attempts to fully assign the sequence were not successful as a third of the protein sequence still lacked 3D data following deuteration. Due to lack of data, less than 50% of the assignments made by MARS were 100% confident. The peaks assigned with 100% confidence have been labelled in Figure 4.3.3a. The strips of a stretch of assigned residues K100 to R104 are also shown below in Figure 4.3.3b.

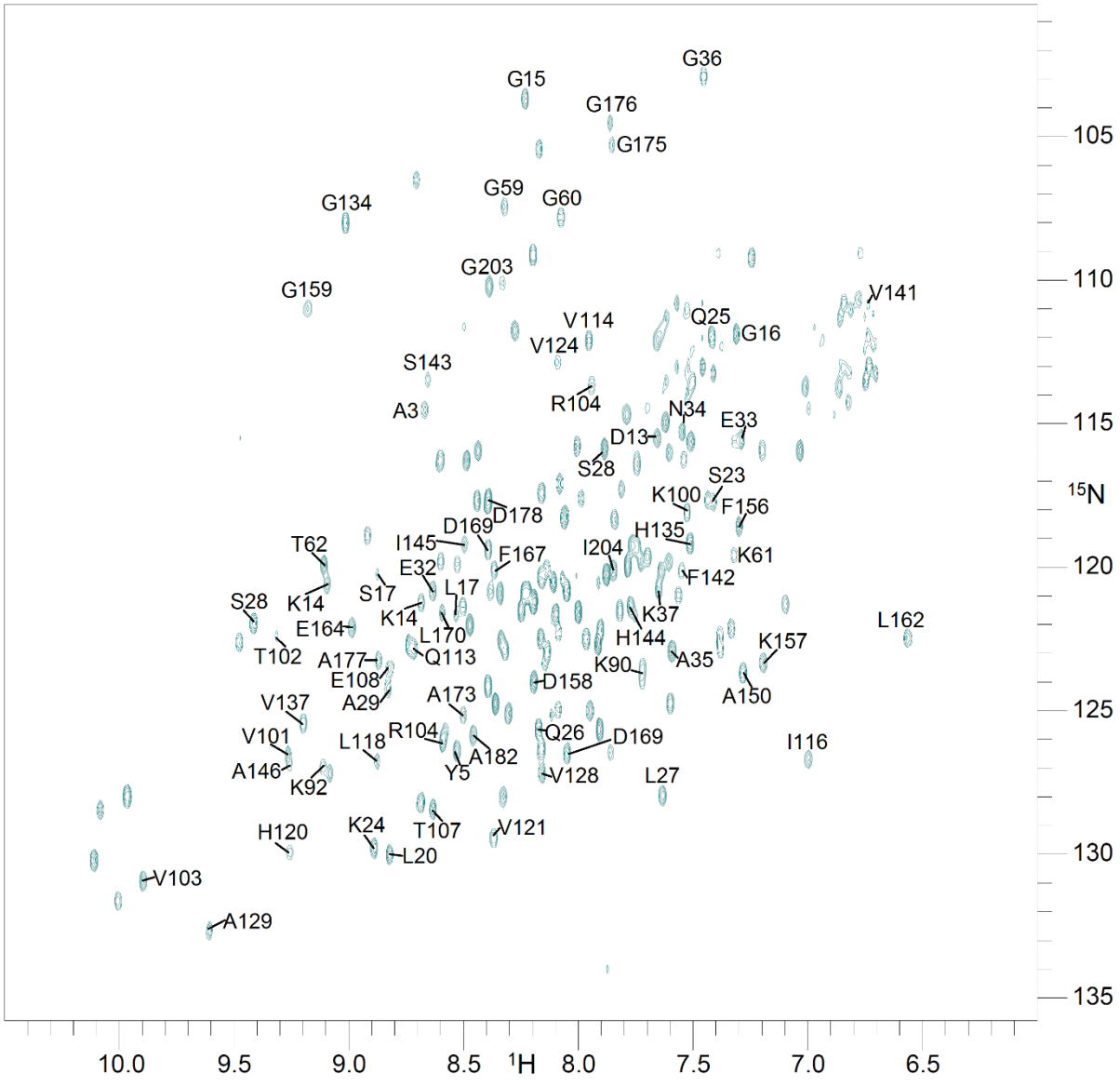


Figure 4.3.3a: Figure showing the 2D ^{15}N TROSY-HSQC spectrum of double ($^{13}C/^{15}N$) labelled TgIMP1 clone 7. The peaks are labelled with single letter amino acid code followed by their position in the recombinant TgIMP1 clone 7 sequence. Less than 50% of the peaks were assigned with 100% confidence.

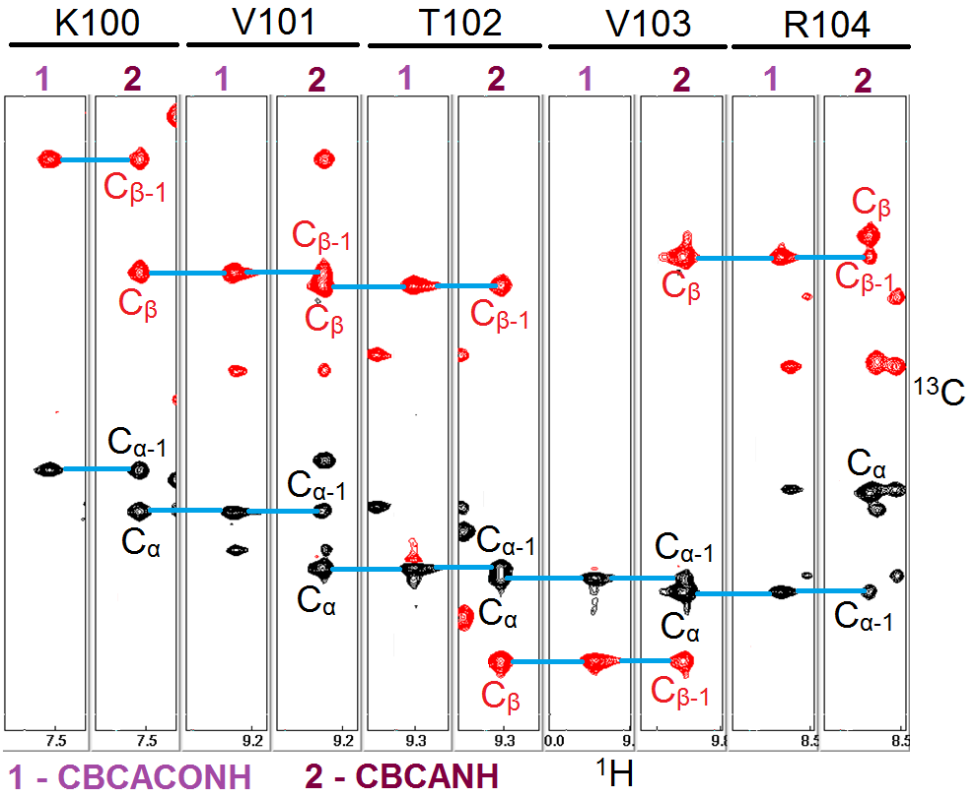


Figure 4.3.3b: Figure showing the strips that were obtained from the CBCACONH (1) and CBCACNH (2) experiments for triple labelled ($^{13}C/^{15}N/^{2}H$) TgIMP1 clone 7. Data is shown for residues K100 to R104. The blue lines connect the peaks that were used to assign the backbone atoms of the labelled residues. The atoms corresponding to the peaks are labelled in the figure.

4.3.4 Crystallisation trials of TgIMP1 clone 7

Crystals were not obtained in any of the conditions present in the screens that were set up. Clear drops were predominantly present (around 70%) in the plates set up with TgIMP1 Clone 7 at 50 mg/ml which suggests that this protein concentration was not high enough for the protein to crystallise. An example of clear drops and precipitation observed in crystal plates is shown in Figure 4.3.4.

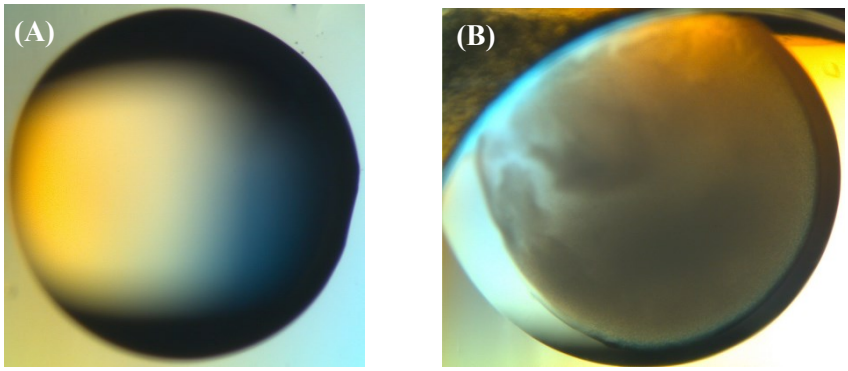


Figure 4.3.4: Figure showing (A) clear drop in ICL1-A1 and (B) precipitation in ICL1-E2 observed in the crystal trial set up with 50 mg/ml of TgIMP1 Clone 7

4.4 Summary

Characterisation of TgIMP1 and EtIMP1 shows that both proteins have a flexible N-terminal domain which is highly dynamic. Limited proteolysis of TgIMP1 revealed the presence of a globular C-terminal domain which appears to be structurally independent. Recent studies by Jia et al (Appendix 1.4) have shown that the first 11 amino acids are necessary for the localisation of the TgIMP1 protein to the plasma membrane in the parasite.

The ^{1}H 1D NMR data for the long and short constructs of TgIMP1 have confirmed the presence of a long flexible domain at the N-terminus of the protein. Truncation of the N-terminus of TgIMP1 showed a decrease in the random coil signal. The NMR data of TgIMP1 clone 7 and EtIMP1 clone 2 also suggest that the IMP domains of both homologues are globular and suitable for structural studies. Truncation of the N-terminal region does not seem to affect the folding of the IMP domain which further suggests that there is a linker sequence between the plasma membrane localisation signal and the IMP domain.

Recent studies of EtIMP1 have shown that a fragment of EtIMP1 (EtIMP1-C), AA151 to AA387, is sufficient to raise immunity against *Eimeria* infections. Further studies have confirmed that the efficacy of EtIMP1-C is the same as the efficacy of the full length EtIMP1 protein (Nolan et al, 2015; Yin et al, 2014). This study confirms that EtIMP1-C is the immune-dominant region in the protein. EtIMP1 clone 2 consists of residues AA152 to AA387 of full length EtIMP1. This confirms that the C-terminal IMP domain is the immune-dominant region in IMP1 proteins.

Limited proteolysis of EtIMP1 clone 2 was not successful as this construct only contains the IMP domain of EtIMP1. Cleavage within this globular domain could have led to complete degradation of the construct as observed in the SDS PAGE gel and the sequencing results following limited proteolysis.

Further structural studies of both TgIMP1 and EtIMP1 constructs were not possible as the 3D data collected for both samples were incomplete. This could be because the IMP domains of both homologues still contain intrinsic flexible regions which are dynamic. Dynamic regions in the protein cannot be observed in the NMR spectrum leading to gaps in the NMR data.

BLAST results (not shown) show that there are two classes of IMP proteins. The first class of proteins contain the plasma membrane localisation signal followed by a linker and the IMP domain and will be referred to as Class I throughout this thesis. The second class of proteins lack the plasma membrane localisation signal and the linker and only contain the IMP domain and will be referred to as Class II throughout this thesis.

The localisation of Class II proteins is unknown as they lack the plasma membrane localisation domain. However, most apicomplexan parasites contain both classes of IMP proteins which shows that the IMP family of proteins is redundant in apicomplexan parasites. The redundancy of this protein suggests that IMP proteins play an important role in the parasite, which is yet to be deduced. Attempts to crystallise TgIMP1 clone 7 were also unsuccessful, possibly due to the highly dynamic nature of the protein.

Since structural studies of Class I IMP1 homologues from *T. gondii* and *E. tenella* were unsuccessful due to the flexible nature of both proteins, characterisation of Class II IMP homologue from *Plasmodium falciparum* was attempted.

Chapter 5

Characterisation of *Plasmodium falciparum* homologue of IMP (Class II)

5.1 Cloning, Protein expression and purification

In this study, the IMP1-like homologue from *P. falciparum* (Accession number: XP_001349200) was codon optimised, cloned and transformed for recombinant expression in bacteria.

5.1.1 PCR Amplification

Full length PfIMP2 was codon optimised for expression in *E. coli* and synthesised in pMA-T vector (GeneArt, Life Technologies). The PfIMP2 sequence was then cloned into various vectors – pET32 Xa/LIC (NEB, UK), pET30 Xa/LIC (NEB, UK) and pNIC-ZB. pNIC-ZB was a gift from Opher Gileadi (Addgene plasmid # 26107).

5.1.2 Colony screening

Following cloning of PfIMP2 into pNIC-ZB vector and transformation in DH5α cells, the PCR products from colony PCR were run in a 1% agarose gel and the result is shown in Figure 5.1.1.

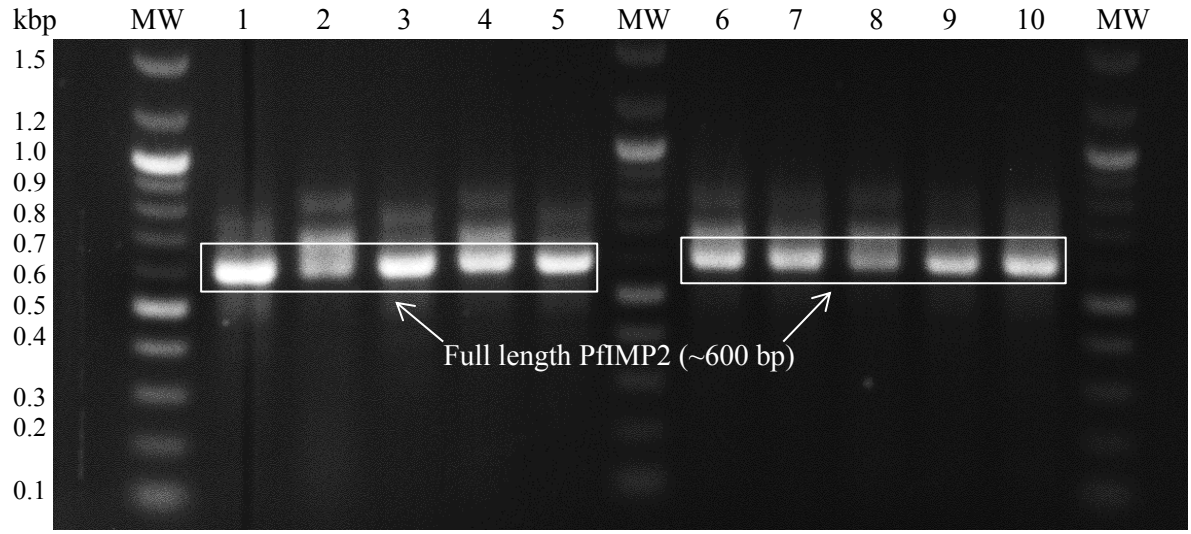


Figure 5.1.1: Figure showing the PCR products obtained from colony PCR of full length PfIMP2 cloned into pNIC-ZB vector. Ten colonies were tested, all of which were positive (white box). The bands are around 0.6 kb which is the expected molecular weight of full length PfIMP2. The ten lanes labelled refer to colonies 1 to 10.

The sequencing results confirmed that the positive colonies contain plasmid encoding PfIMP2.

5.1.3 Expression test

Following cloning of PfIMP2 gene into various vectors, different strains of *E. coli* (BL21, Rosetta 2 and ArcticExpress) were transformed to express PfIMP2 and plated on Agar plates. Then, the cells were checked for expression as this is the first attempt of expressing PfIMP2 in *E. coli*. The results from the expression test showed bands around the expected molecular weight (~35 kDa) confirming the soluble overexpression of PfIMP2 in pET32 Xa/LIC vector only in Rosetta 2 cells. A figure showing the expression test performed for PfIMP2 (pET32 Xa/LIC vector) is shown in Figure 5.1.2. Expression tests were also performed for PfIMP2 in pNIC-ZB vector but overexpression was not observed (Figure 5.1.3).

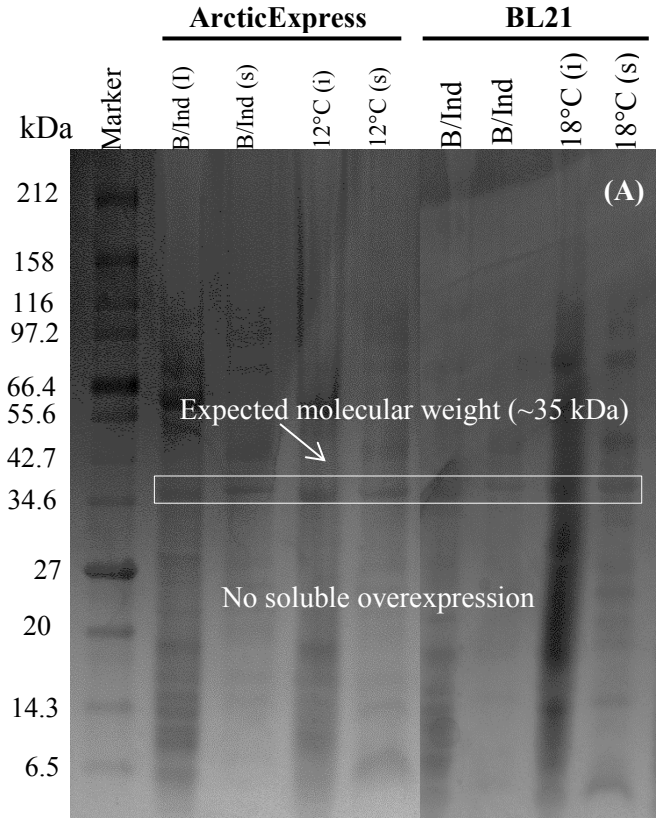


Figure 5.1.2a: Figure showing the expression test results obtained for full length PfIMP2 (pET32 Xa/LIC) expressed in ArcticExpress and BL21 cells. No soluble overexpression is observed.

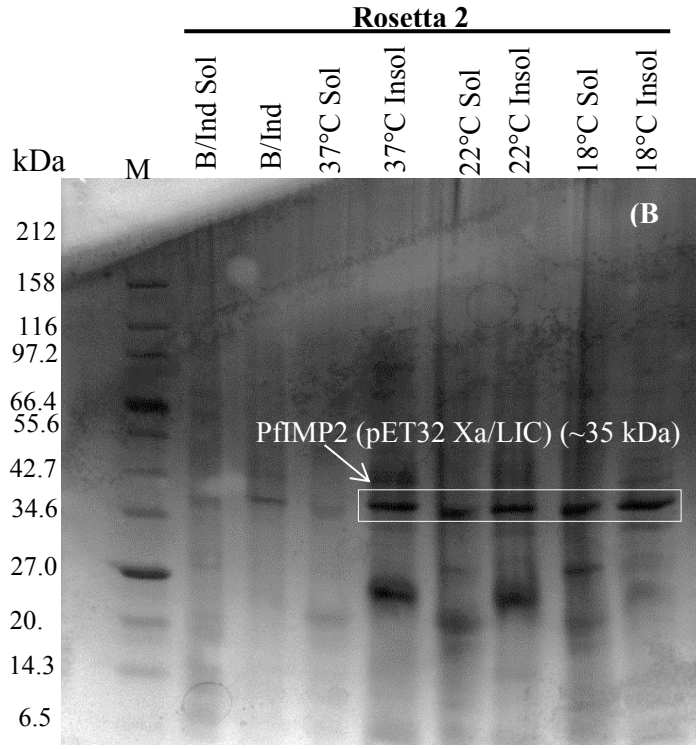


Figure 5.1.2b: Figure showing the expression test results obtained for full length PfIMP2 (pET32 Xa/LIC) expressed in Rosetta 2 cells. The white box shows the bands that correspond to the molecular weight of full length PfIMP2 protein. Soluble overexpression is only observed in Rosetta 2 cells.

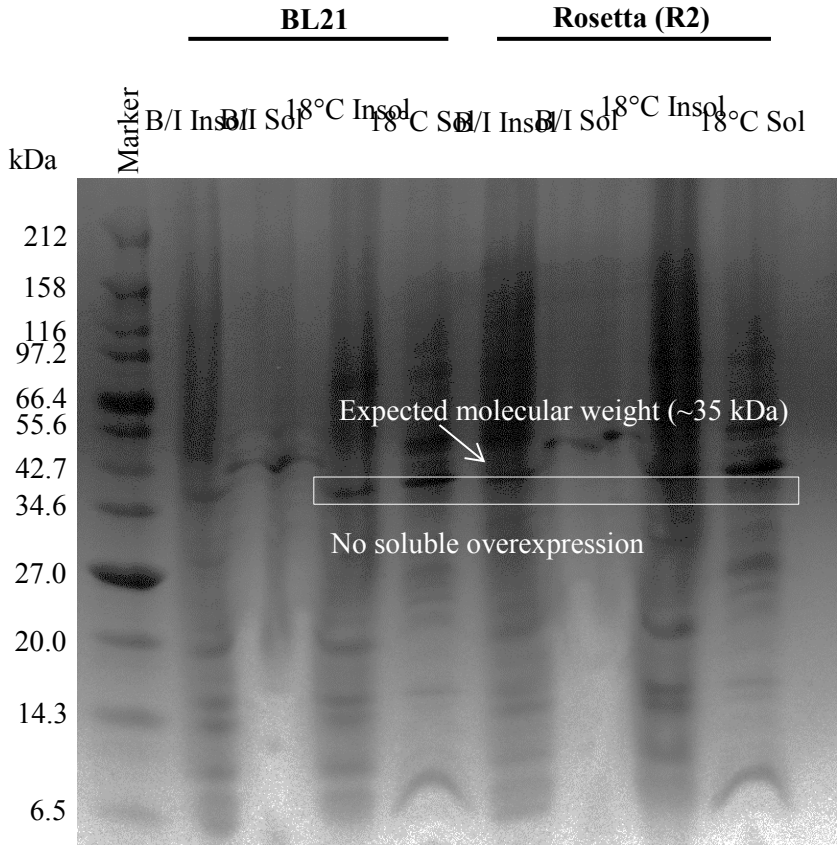


Figure 5.1.3: Figure showing the expression test results obtained for full length PfIMP2 (pNIC-ZB) in BL21 and Rosetta2 cells. Overexpression of PfIMP2 is not observed.

5.2 pET32Xa/LIC construct

5.2.1 Protein expression and purification

The average mass of cells, once harvested, was 5 g per litre of culture for PfIMP2. The concentration of protein after Nickel affinity chromatography was 10.08 mg/ml in 2 ml. The SDS PAGE gel obtained from running the fractions from Nickel affinity chromatography for PfIMP2 (pET32 Xa/LIC expressed in Rosetta2 cells) shows that the molecular weight of the overexpressed protein sample is lower than expected (Figure 5.2.1). The expected molecular weight of PfIMP2 (with N-terminal His, S and thioredoxin tags) is 32 kDa whereas the purified protein sample runs at an apparent molecular weight of 20 to 27 kDa.

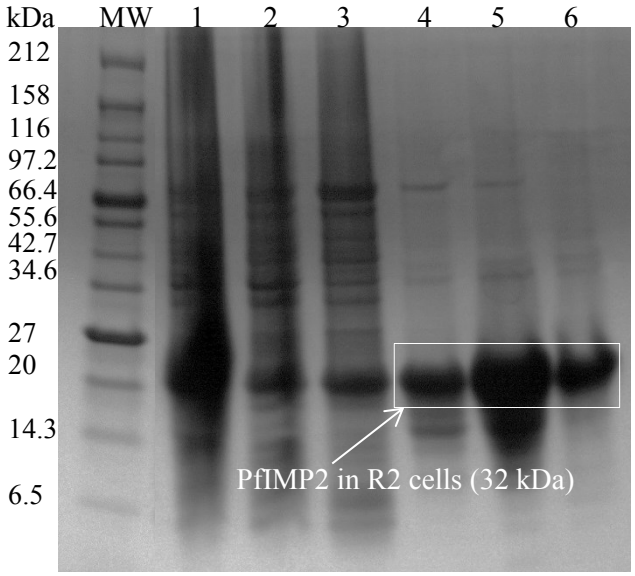


Figure 5.2.1a: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of full length PfIMP2 (pET32 Xa/LIC) expressed in Rosetta 2 cells. The white box shows the bands that correspond to full length PfIMP2 with the thioredoxin tag at a lower molecular weight (~28 kDa) than expected (32 kDa). The lower molecular weight of the protein could be due to degradation of improperly folded protein. The lanes in the gel are (1) Clear Lysate (2 Flow through following Nickel affinity chromatography (3) Wash (4) Eluent 1 (5) Eluent 2 (6) Eluent 3

Expression and purification of PfIMP2 (pET32 Xa/LIC) in BL21 cells also gave soluble protein of a lower molecular weight (Figure 5.1.4b).

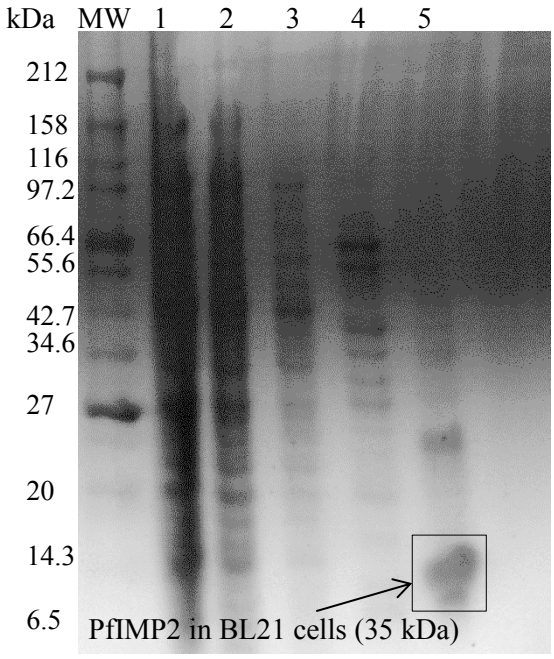


Figure 5.2.1b: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of full length PfIMP2 (pET32 Xa/LIC) expressed in BL21 cells. The black box shows the band that corresponds to full length PfIMP2 with the thioredoxin tag between 6.5 kDa to 14.3 kDa which is lower than the expected molecular weight of 35 kDa. The band is at a lower molecular weight than expected possibly due to degradation of improperly folded protein. The lanes in the gel are (1) Clear Lysate (2) Flow through following Nickel affinity chromatography (3) Wash (4) Eluent 1 (5) Eluent 2

5.2.2 ^{1}H 1D NMR

^{1}H 1D NMR experiments were performed on all the PfIMP2 constructs produced to check if the protein is folded. ^{1}H 1D NMR was performed on pET32 Xa/LIC-PfIMP2 construct that was expressed in BL21 cells and the spectrum for this protein (Figure 5.2.2a) showed that this construct is unfolded.

^{1}H 1D NMR was performed on pET32 Xa/LIC-PfIMP2 construct that was expressed in Rosetta 2 cells (Figure 5.2.2b).

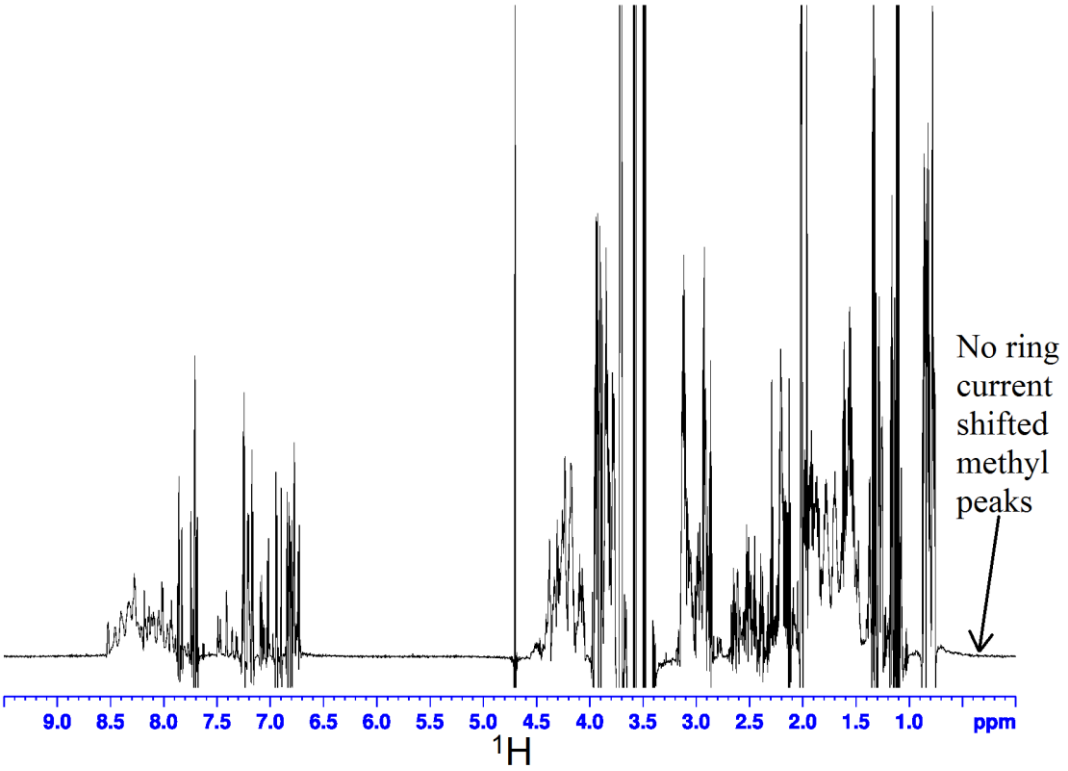


Figure 5.2.2a: Figure showing the ^{1}H 1D spectrum of PfIMP2 (pET32 Xa/LIC construct) expressed in BL21 cells. The lack of ring current shifted methyl peaks beyond 0.5 ppm shows that the protein is unfolded.

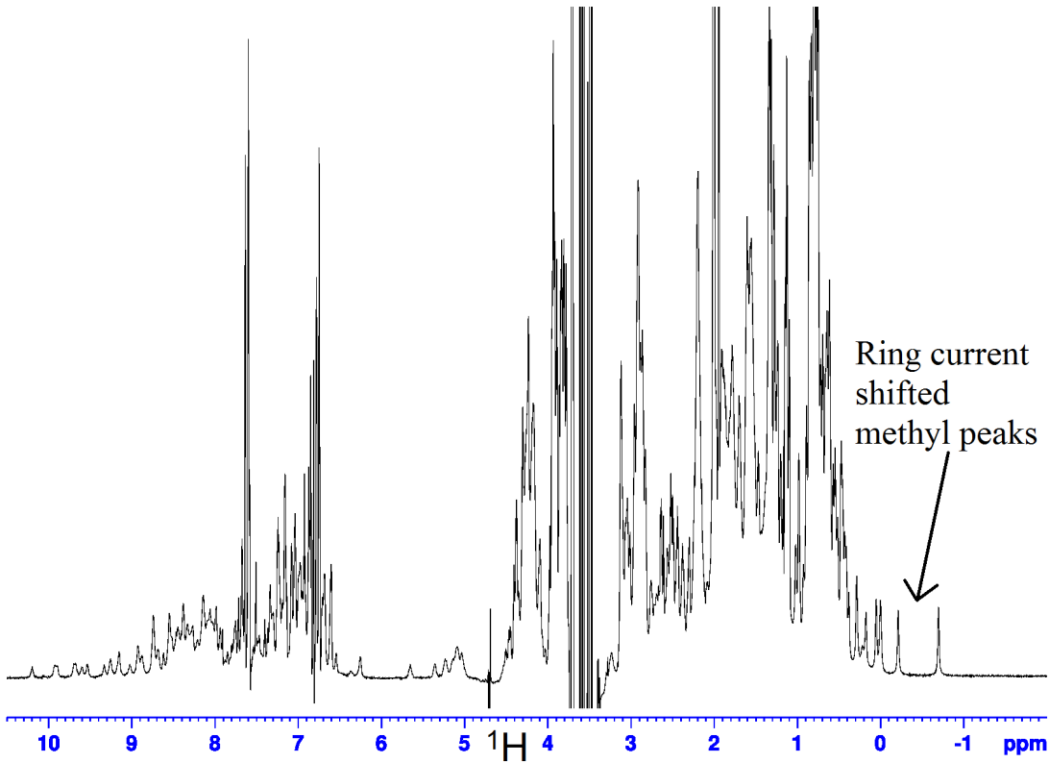


Figure 5.2.2b: Figure showing the ^{1}H 1D spectrum of PfIMP2 (pET32 Xa/LIC construct) expressed in Rosetta2 cells. This construct appears to be folded as it has two ring current shifted methyl peaks beyond 0 ppm.

The PfIMP2 (pET32 Xa/LIC) construct, expressed in Rosetta 2 cells, has ring-current shifted methyl peaks which suggests that it is folded but an overlay of a ^{1}H 1D spectrum of 100% thioredoxin showed that these peaks may be from the thioredoxin tag rather than from the protein (Figure 5.2.2c). The presence of an amide peak around 10 ppm shows that the protein under study has a tryptophan which is absent in thioredoxin. This shows that this signal is from the protein. However, the lack of additional ring current shifted methyl peaks other than the ones from thioredoxin suggests that although the protein is present in the sample, it is not folded.

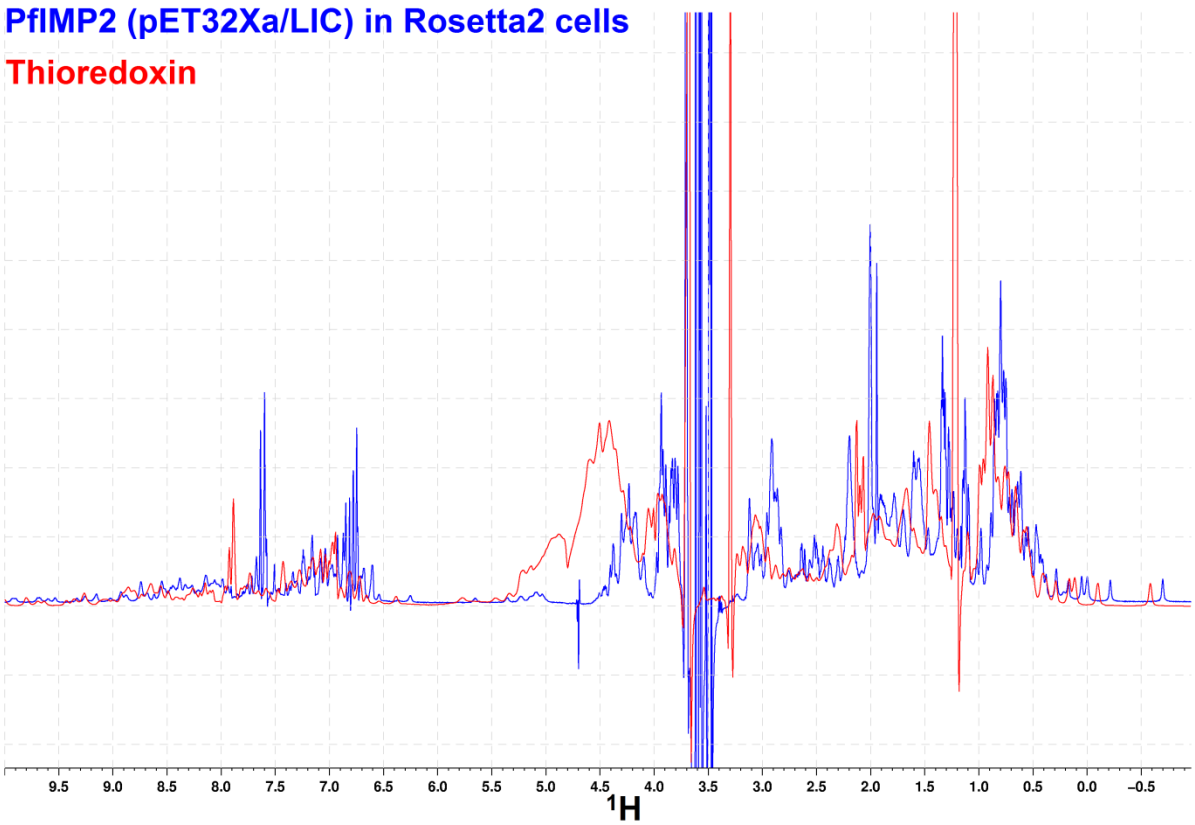


Figure 5.2.2c: Figure showing the overlay of ^{1}H 1D spectra of PfIMP2 (pET32 Xa/LIC construct) expressed in Rosetta2 cells (blue) and thioredoxin (red). The ring current shifted methyl peaks of PfIMP2 (pET32 Xa/LIC) and thioredoxin are the same suggesting that there is very little signal from the protein.

5.3 pET30Xa/LIC construct

5.3.1 Protein expression and purification

Expression of ^{15}N labelled PfIMP2 (pET30 Xa/LIC) was performed in Rosetta 2 cells. The SDS PAGE gel obtained from running the fractions from nickel purification shows that there is overexpression but the size of the protein is much lower than expected (Figure 5.3.1).

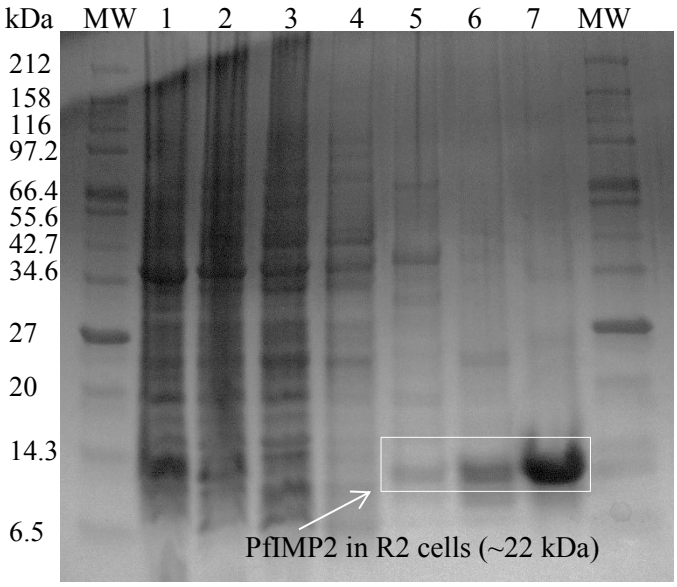


Figure 5.3.1: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of full length PfIMP2 (pET30 Xa/LIC) expressed in Rosetta 2 cells. The white box shows the band that corresponds to full length PfIMP2 with the His and S tags. The band is at a lower molecular weight than expected possibly due to degradation of improperly folded protein. The lanes in the gel are (1) Lysate (2) Clear Lysate (3) Flow through following Nickel affinity chromatography (4) Wash (5) Eluent 1 (6) Eluent 2 (7) Eluent 3

5.3.2 ^{1}H 1D NMR

The ^{1}H 1D NMR spectrum of the full length PfIMP2 (pET30 Xa/LIC) protein expressed in Rosetta 2 cells shows that this construct is unfolded (Figure 5.3.2).

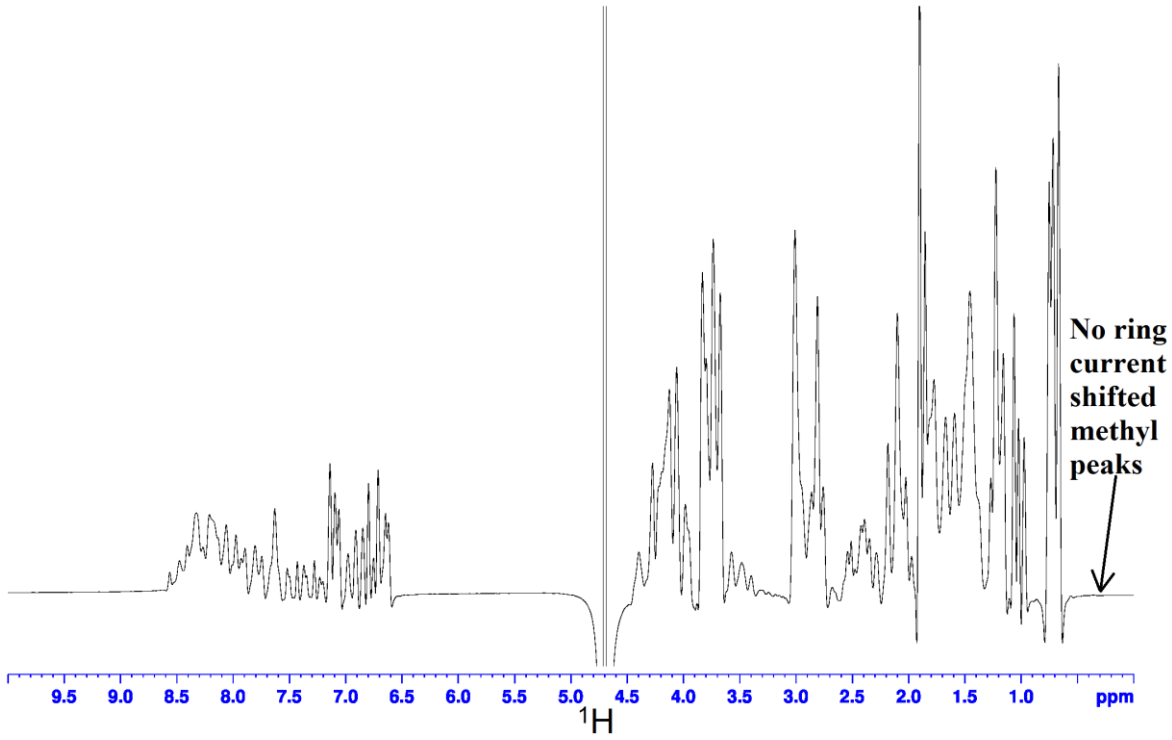


Figure 5.3.2: Figure showing the ^{1}H 1D spectrum of PfIMP2 (pET30 Xa/LIC construct) expressed in Rosetta 2 cells. The lack of ring current shifted methyl peaks beyond 0.5 ppm shows that the protein is unfolded.

5.3.3 ^{15}N TROSY-HSQC

Single labelled (^{15}N) PfIMP2 (pET30 Xa/LIC) sample expressed in Rosetta 2 cells was used to record ^{15}N TROSY-HSQC spectrum (Figure 5.3.3). The ^{15}N HSQC further confirms that the protein is unfolded.

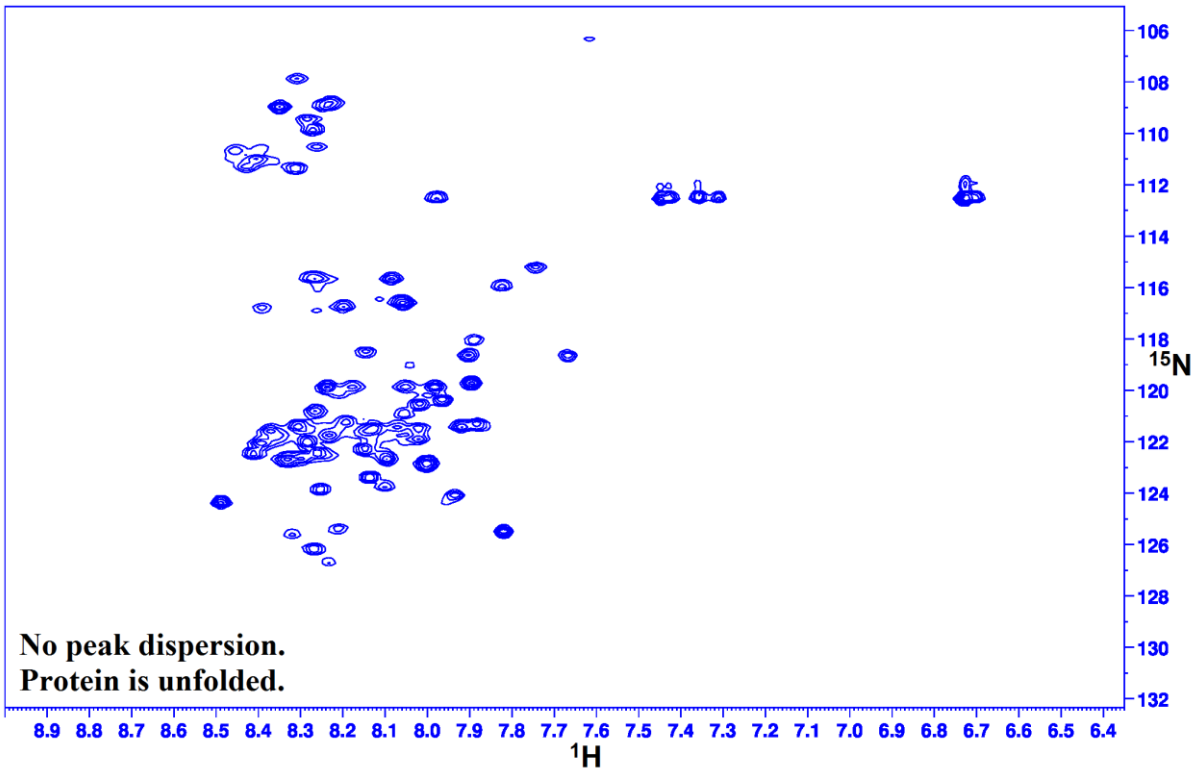


Figure 5.3.3: Figure showing the 2D ^{15}N TROSY-HSQC spectrum of single (^{15}N) labelled PfIMP2 (pET30 Xa/LIC). The clustering of peaks around 8 ppm on the proton scale shows that the protein is unfolded as this signal corresponds to random coil.

5.4 pNIC-ZB construct

5.4.1 Protein expression and purification

Expression of PfIMP2 (pNIC-ZB) in Rosetta 2 cells was performed. The SDS PAGE gel obtained from running the fractions from nickel purification for PfIMP2 (pNIC-ZB expressed in Rosetta 2 cells) shows that the protein is overexpressed (Figure 5.4.1a).

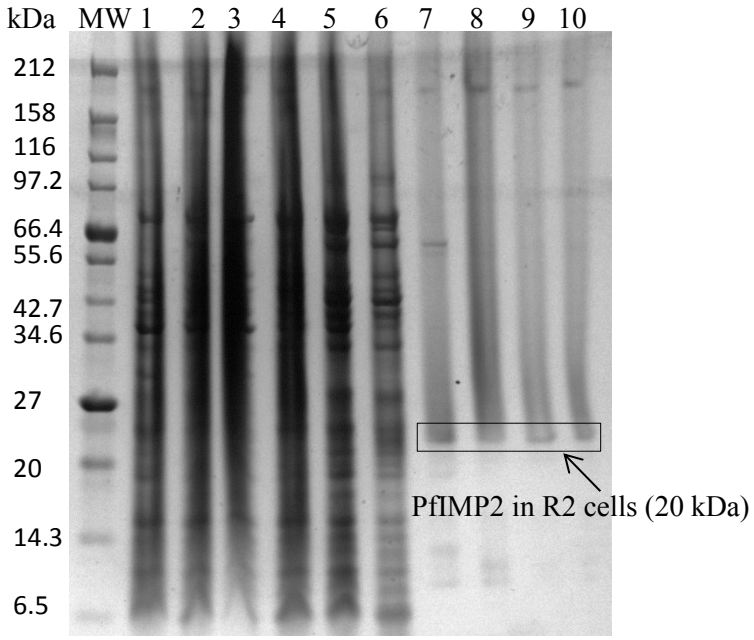


Figure 5.4.1a: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of full length PfIMP2 (pNIC-ZB) expressed in Rosetta 2 cells. The black box shows the band that corresponds to full length PfIMP2 with the His tag. The band is a little higher than the expected molecular weight of 20 kDa. The lanes in the gel are (1) Pre-lysis (2) Post-lysis (3) Insoluble pellet (4) Clear lysate (5) Wash during Nickel affinity chromatography (6) Flow through following Nickel affinity chromatography (7) Eluent 1 (8) Eluent 2 (9) Eluent 3 (10) Pooled eluents

Following nickel affinity chromatography, unlabelled and $^{13}C/^{15}N$ labelled PfIMP2 (pNIC-ZB) samples were purified by size exclusion chromatography. The SDS PAGE gel obtained from running the fractions from size exclusion chromatography are shown in figure 5.4.1b. The results show that the protein is pure. Since the protein runs as a smear, the PfIMP2 sample was run on the gel at various concentrations from 20 mg/ml to 0.5 mg/ml (Figure 5.1.4c). The protein ran as a smear under all these concentrations.

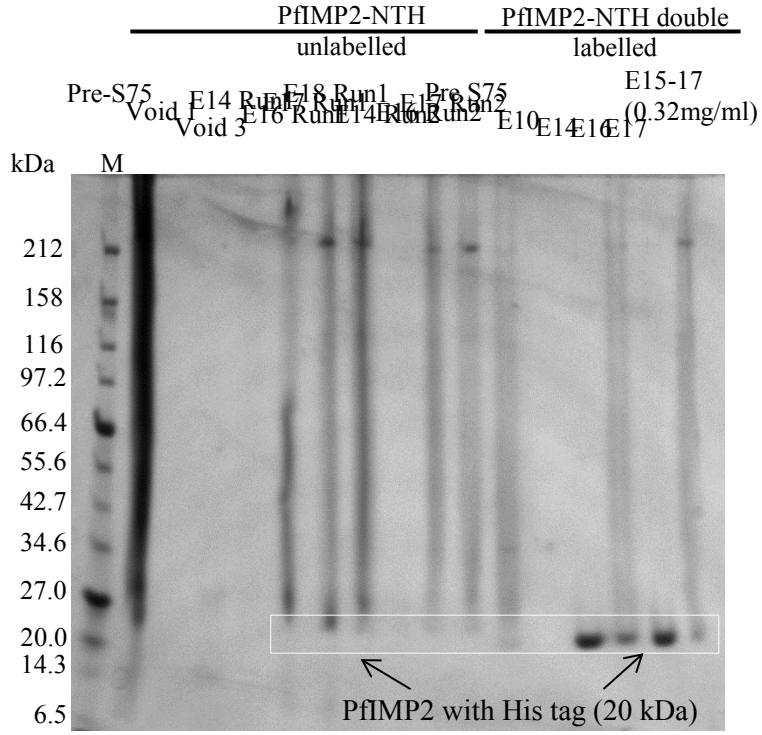


Figure 5.4.1b: Figure showing the SDS PAGE gel of the samples obtained during size exclusion chromatography of unlabelled and $^{13}C/^{15}N$ labelled full length PfIMP2 (pNIC-ZB) expressed in Rosetta 2 cells. The white box shows the band that corresponds to full length PfIMP2 with the His tag. The band is at the expected molecular weight of 20 kDa.

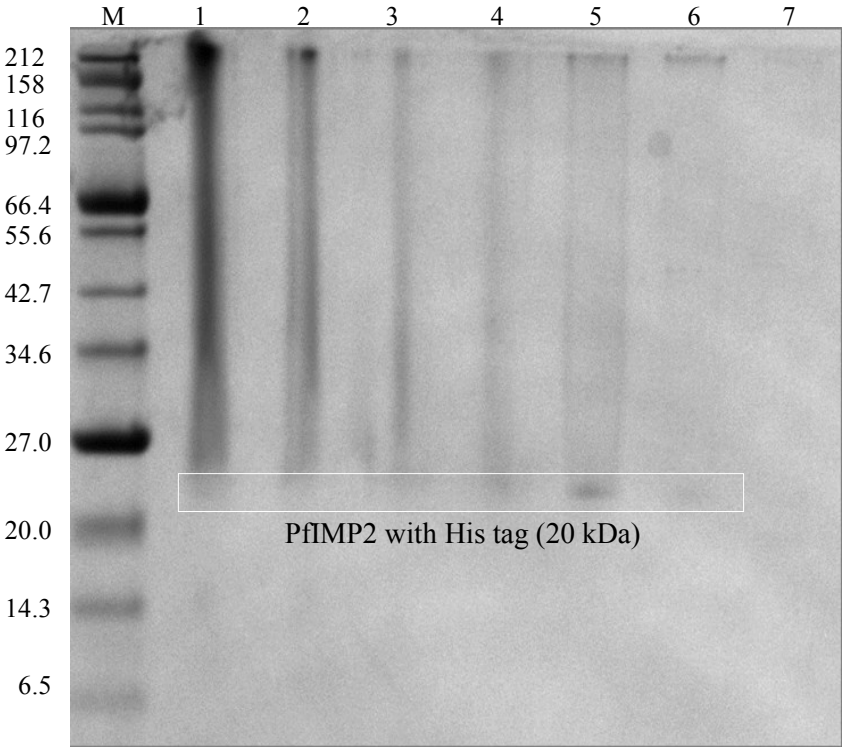


Figure 5.4.1c: Figure showing SDS-PAGE gel with various concentrations of PfIMP2 sample. The lanes in the gel are – M is the marker, 1 is 20 mg/ml (1 mM) PfIMP2, 2 is 15 mg/ml (0.75 mM) PfIMP2, 3 is 10 mg/ml (0.5 mM) PfIMP2, 4 is 5 mg/ml (0.25 mM) PfIMP2, 5 is 2 mg/ml (0.1 mM) PfIMP2, 6 is 1 mg/ml (0.05 mM) PfIMP2 and 7 is 0.5 mg/ml (0.025 mM) PfIMP2.

5.4.2 Western Blot analysis

As PfIMP2 runs as a smear on the SDS PAGE gel, western blot was performed to confirm the identity of the band. The results shown in Figure 5.4.2 show that the smear is the His tagged PfIMP2 protein.

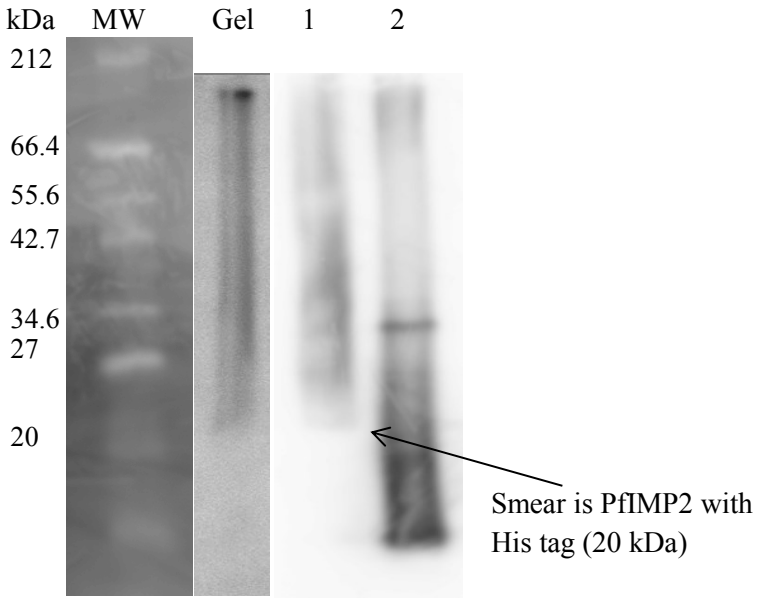


Figure 5.4.2: Figure showing the Western blot of full length PfIMP2 (pNIC-ZB) with the His tag. The entire smear shows positive results which shows that it corresponds to the His tagged PfIMP2 protein. The lanes in the figure are MW – protein molecular weight marker, Gel showing PfIMP2, (1) His tagged PfIMP2 and (2) Positive control

5.4.3 Differential Scanning Fluorimetry (DSF)

The PfIMP2-NTH construct was initially in 50 mM NaH_2PO_4 150 mM NaCl pH 7 buffer. The protein precipitated rapidly at room temperature in this buffer. DSF was used to perform a buffer screen to determine the condition which is optimum for protein stability. The screen showed that the protein was most stable in 50 mM Hepes, 150 mM NaCl 10 mM DTT pH 8.

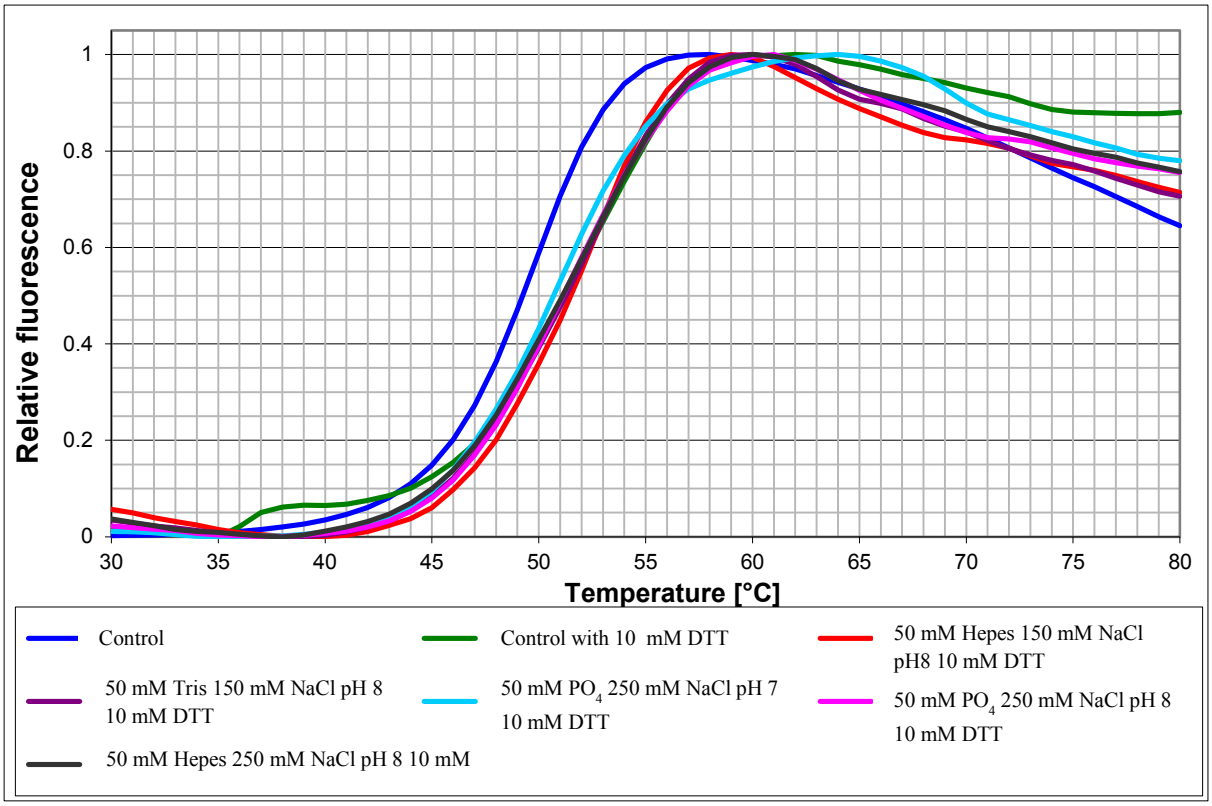


Figure 5.4.3: Figure showing the buffer screen results using DSF for PfIMP2 (pNIC-ZB). The control is 50 mM NaH_2PO_4 150 mM NaCl pH 7. The five conditions which gave the highest Tm have been shown in the figure. Each curve has been labelled in the figure legend. The best condition is 50 mM Hepes pH8 150 mM NaCl 10 mM DTT.

5.4.4 ^{1}H 1D NMR

^{1}H 1D NMR was performed on pNIC-ZB-PfIMP2 construct that was expressed in Rosetta 2 cells. The spectrum of this protein shows that this construct is folded (Figure 5.4.4).

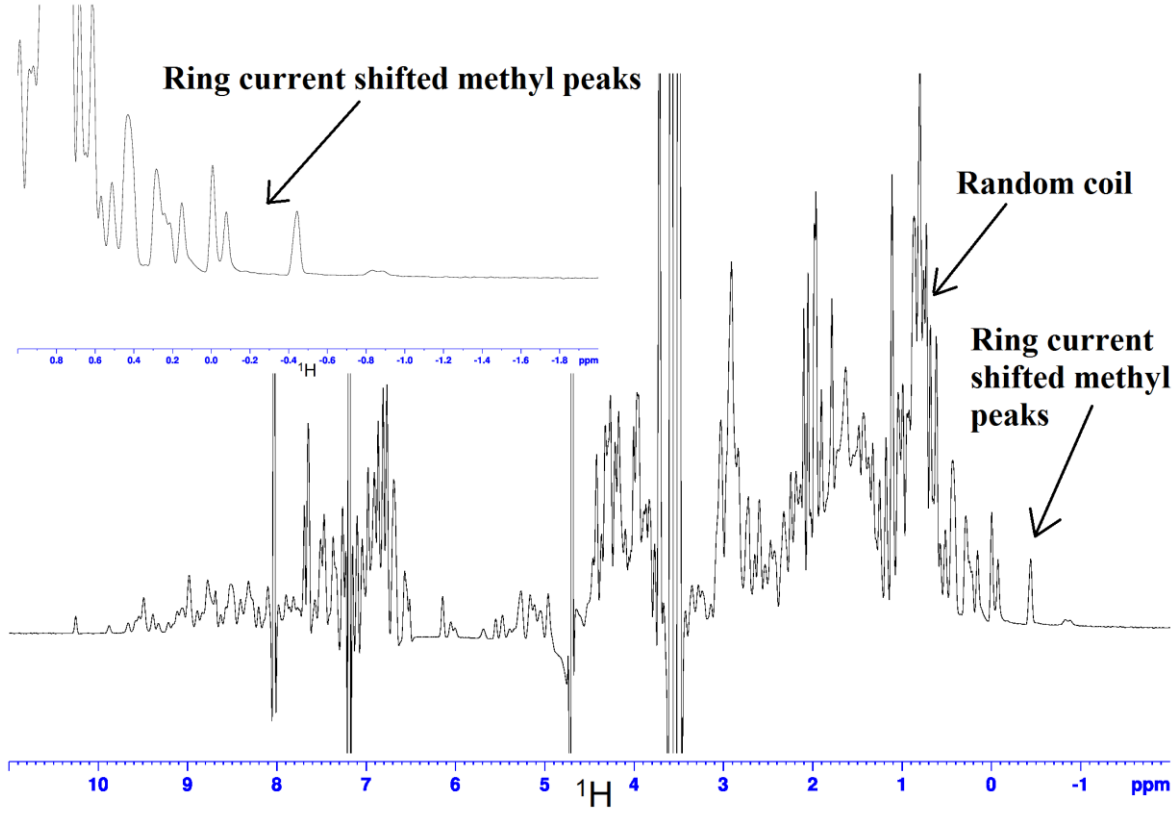


Figure 5.4.4: Figure showing the ^{1}H 1D spectrum of the PfIMP2 (pNIC-ZB). On the top left, a zoom in of the methyl region of the spectrum between 1.5 ppm and -2.00 ppm is shown. The presence of ring current shifted peaks in the methyl region shows that the protein is folded.

5.4.5 ^{15}N TROSY-HSQC of PfIMP2 with the His tag

Double labelled ($^{13}C/^{15}N$) PfIMP2 (pNIC-ZB) sample expressed in Rosetta 2 cells was used to record ^{15}N TROSY-HSQC. The ^{15}N HSQC has dispersed peaks which shows that the protein is folded and suitable for further structural studies by NMR.

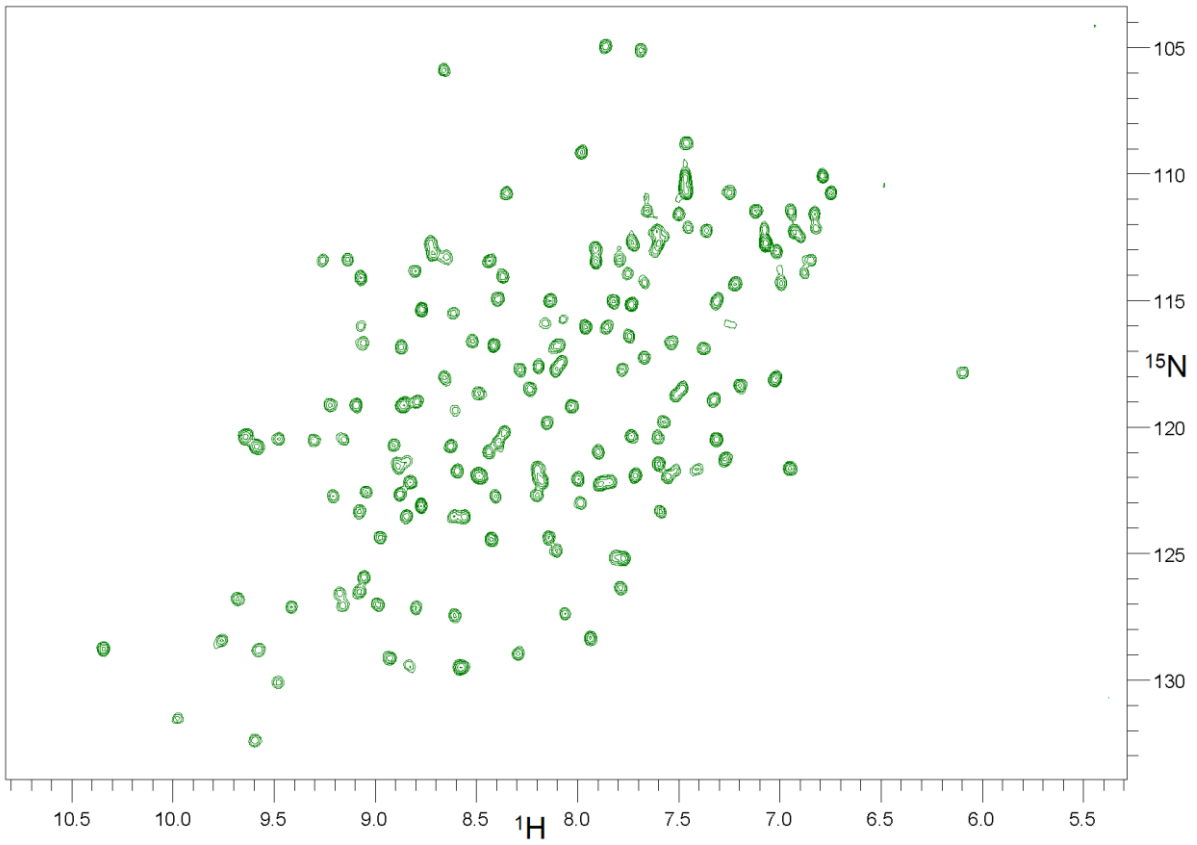


Figure 5.4.5: Figure showing the 2D ^{15}N TROSY-HSQC spectrum of double labelled PfIMP2 (pNIC-ZB) protein with the N-terminal His tag. The dispersed peaks in the spectrum show that the protein is folded.

Backbone assignment could not be completed for this construct as the data from the 3D experiments was not complete. This could have been due to interference from the flexible N-terminal His tag.

5.4.6 TEV cleavage

Following Nickel affinity chromatography, the N terminal His tag was cleaved off from the protein by TEV cleavage. Following TEV cleavage, the protein was further purified by size exclusion chromatography. The SDS PAGE gel obtained from running the fractions from the TEV cleavage and size exclusion chromatography are shown in Figure 5.4.6.

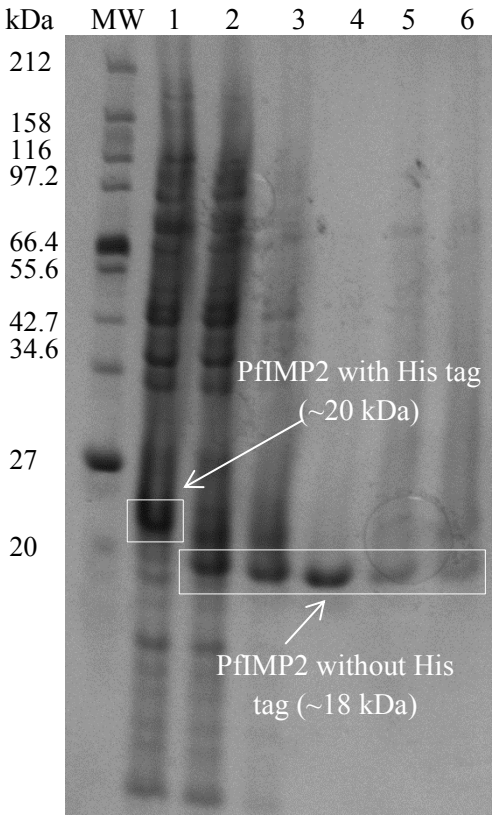


Figure 5.4.6: Figure showing the SDS PAGE gel of the samples obtained following TEV cleavage and size exclusion chromatography of $^{13}C/^{15}N$ labelled full length PfIMP2 (pNIC-ZB) expressed in Rosetta 2 cells. The white box shows the band that corresponds to full length PfIMP2 without the His tag. The band is at the expected molecular weight of 18 kDa. The lanes in the gel are (1) Pooled eluents from Nickel affinity chromatography (2) Following overnight incubation with TEV (3) to (6) Purified fractions following size exclusion chromatography

5.4.7 ^{15}N TROSY HSQC of PfIMP2 without the His tag

^{15}N TROSY-HSQC was recorded following TEV cleavage of PfIMP2 to remove the His tag. Comparison of the ^{15}N HSQC recorded with and without the tag shows that some peaks are missing in the ^{15}N HSQC recorded without the tag as shown in Figure 5.4.7. The spectrum shows that the peaks are dispersed which shows that the protein is folded following removal of the His tag.

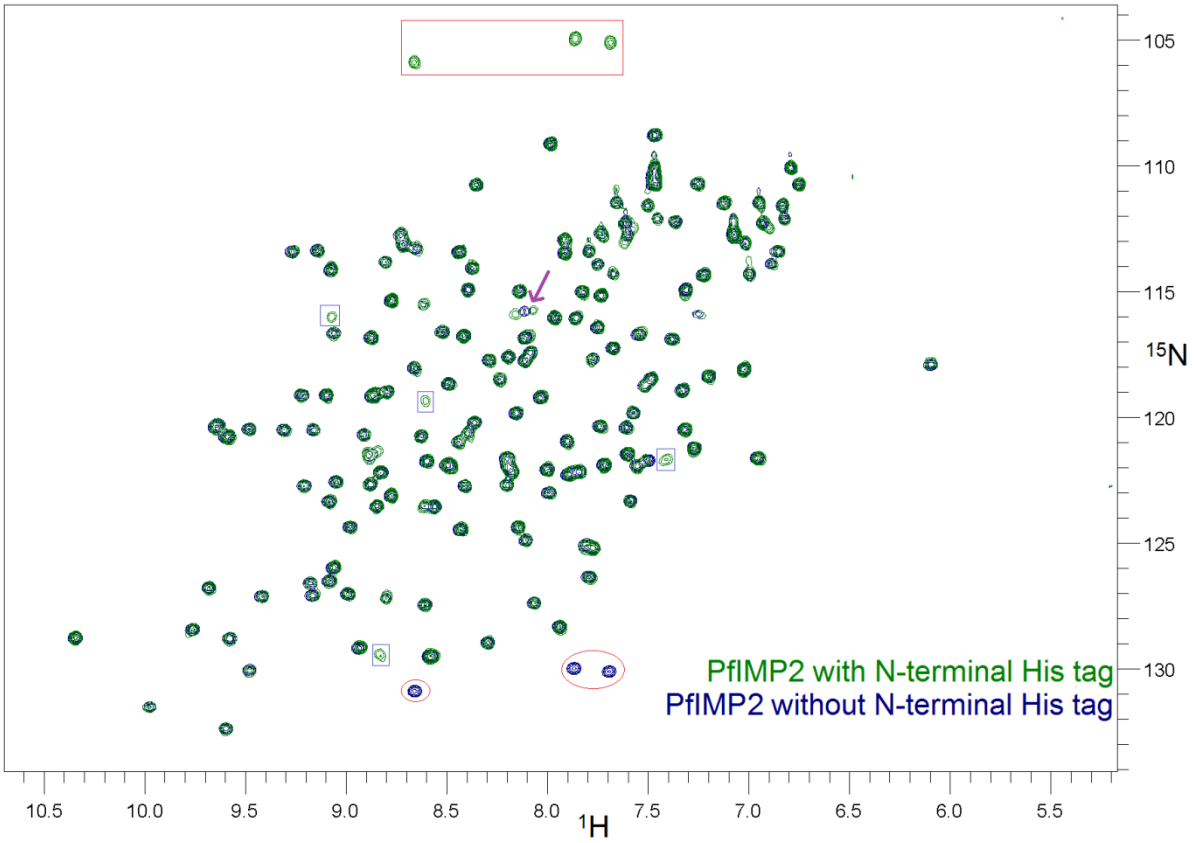


Figure 5.4.7: Figure showing the 2D ^{15}N TROSY-HSQC spectrum of double labelled PfIMP2 (pNIC-ZB) protein with (spectrum 1) and without (spectrum 2) the N-terminal His tag, as shown in the figure. Unaliased peaks in spectrum 1 are shown with the red box and the corresponding aliased peaks in spectrum 2 are shown with red circles. Two peaks in spectrum 1 have merged in spectrum 2 and this is highlighted by the purple arrow. The peaks in spectrum 1 which are missing in spectrum 2 are shown by the blue boxes.

5.4.8 PfIMP2 C71S mutant

The PfIMP2 sequence has only one cysteine residue, C71. This residue was mutated to a serine and the C71S mutant (pNIC-ZB) was expressed in Rosetta 2 cells and purified using Nickel affinity chromatography. The SDS PAGE results obtained from running the Nickel affinity chromatography fractions of the wildtype and the C71S mutant is shown in figure 5.4.8. Overexpression of soluble protein was not observed for the C71S mutant.

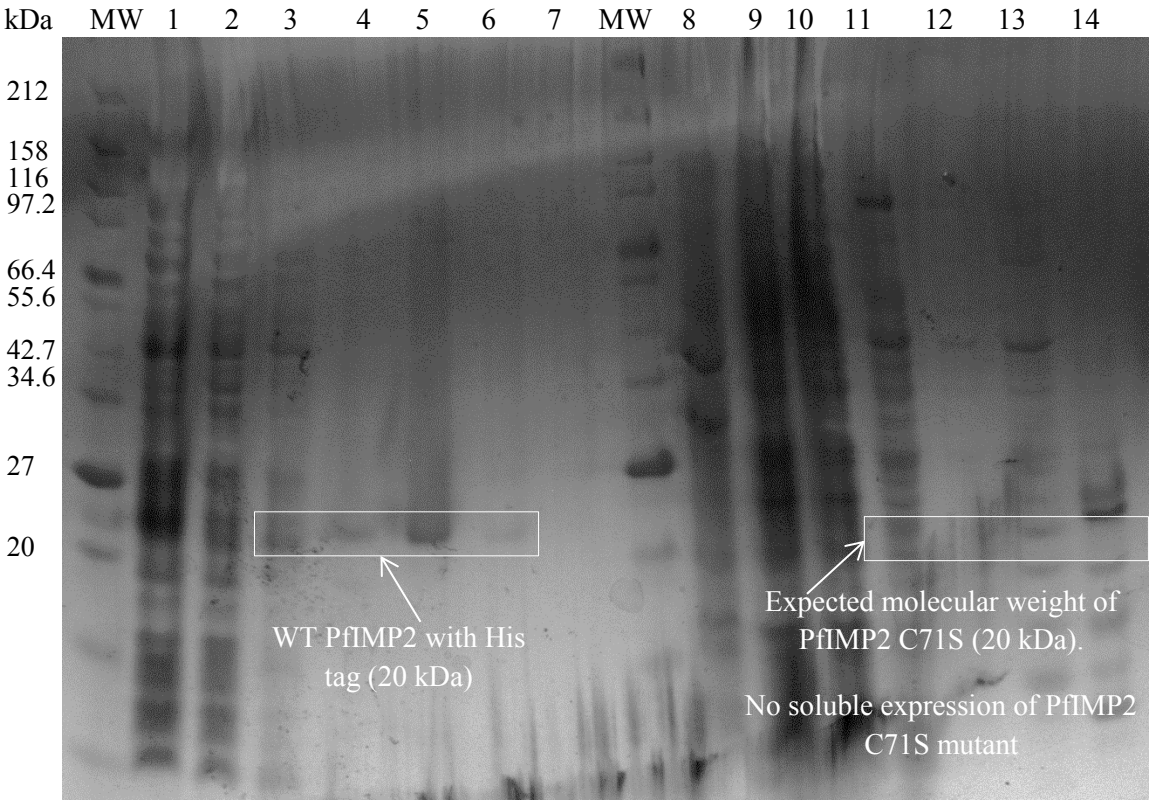


Figure 5.4.8: Figure showing the SDS PAGE gel of the samples obtained during Nickel affinity chromatography of full length PfIMP2 (pNIC-ZB) wild type and C71S mutant expressed in Rosetta 2 cells. The white box shows the band that corresponds to full length WT PfIMP2 with the His tag. The band is a little higher than the expected molecular weight of 20 kDa. Lanes (1) to (7) correspond to WT PfIMP2 and lanes (8) to (14) correspond to C71S mutant of PfIMP2. The lanes in the gel are (1) Clear lysate (2) Flow through following Nickel affinity chromatography (3) Wash during Nickel affinity chromatography (4) Eluent 1 (5) Eluent 2 (6) Eluent 3 (7) Blank lane (8) Insoluble pellet (9) Clear lysate (10) Flow through following Nickel affinity chromatography (11) Wash during Nickel affinity chromatography (12) Eluent 1 (13) Eluent 2 (14) Eluent 3. No overexpression is observed for the C71S mutant.

5.5 Summary

Class II IMP homologue from *P. falciparum* has shown promising results which suggest that this protein is suitable for structural studies. Various vectors were used to express PfIMP2 but soluble expression of folded PfIMP2 was only obtained with the pNIC-ZB vector. Although soluble overexpression was observed with other vectors, the NMR data confirmed that all the constructs were unfolded except for the pNIC-ZB construct. The apparent lower molecular weight of these unfolded constructs could have been due to degradation of these unfolded proteins.

Following expression using the pNIC-ZB vector, PfIMP2 was found to run as a smear on SDS-PAGE gels at concentrations of 0.5 mg/ml and above (Figure 5.1.4c). However, at lower concentrations of 0.2 mg/ml and below, PfIMP2 can be seen as a clear band on the gel (Figure 5.1.4b). At concentrations above 0.2 mg/ml, the protein is observed as smears of various intensities. This behaviour of PfIMP2 could be due to possible self-association of the protein via the single cysteine residue, C71, in the sequence. Another unusual feature of PfIMP2 is that the pI of this protein is 9.34 due to the high number of lysines in the protein (13%). Due to the amino acid composition of the protein, the protein is positively charged overall. The possible self association of the protein via the cysteine residue and the overall positive charge of the protein are possible reasons why this protein runs as a smear at high concentrations.

BLAST searches of PfIMP2 sequence (class II) against apicomplexan proteins have shown that there are other homologues of Class II IMP proteins within apicomplexans (Figure 5.5). Analysis of the PfIMP2 sequence shows that the protein has only one cysteine residue (C71) which appears to be conserved across the Class II IMP homologues (figure 5.5).

During purification of PfIMP2, precipitation was observed in the absence of a reducing agent. This suggests that C71 might be forming intermolecular disulphide bonds

leading to aggregation of the protein. Addition of a reducing agent (10 mM DTT) to the sample eliminated precipitation which confirms this theory. The optimum buffer for this protein was found to be 50 mM Hepes 150 mM NaCl 10 mM DTT pH 8 using DSF. PfIMP2 C71S mutant was not expressed solubly. This suggests that C71 is necessary for proper folding of the protein.

The 3D NMR data recorded for PfIMP2 with the His tag had a few gaps which could have been due to the flexible nature of the N-terminal His tag. In order to overcome this issue, the N-terminal His tag was removed by TEV cleavage. The 3D data for PfIMP2 without the His tag was complete and complete backbone assignment was possible.

The lack of homologous structures of IMP suggests that the protein has a novel fold. Further studies into the structure of PfIMP2 might help us to understand the mechanism by which this protein induces immunogenic responses in the host against the parasite.

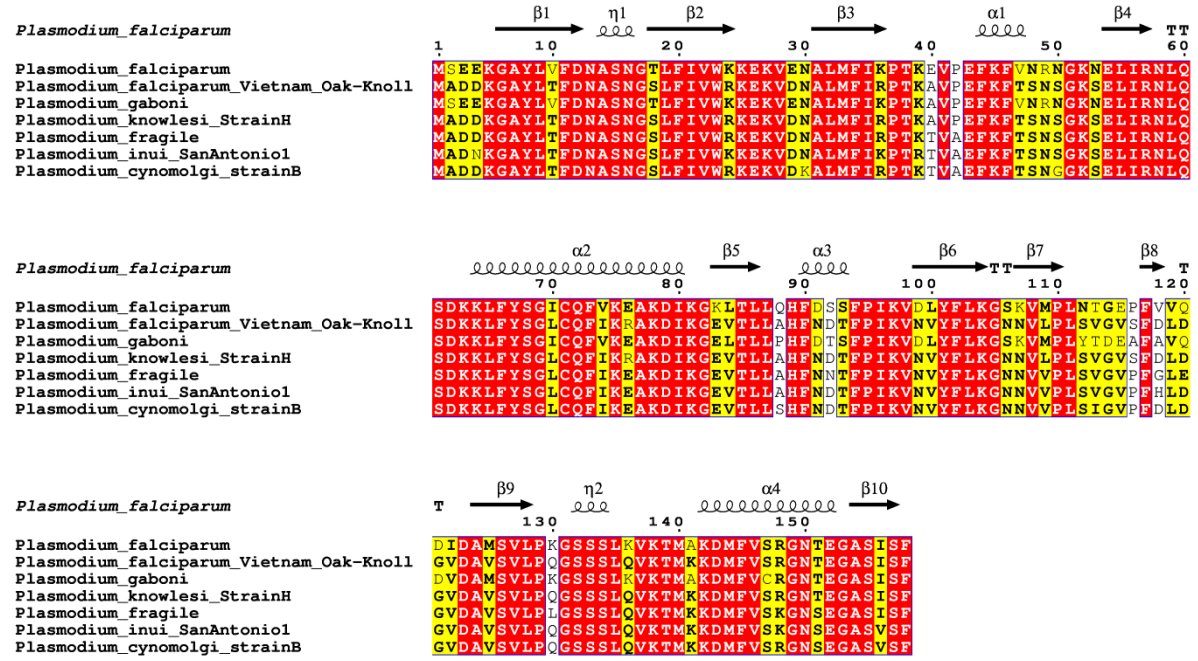


Figure 5.5: Figure showing the sequence alignment of PfIMP2 against the top six sequences from the BLAST search. The residues highlighted red are identical and the ones highlighted yellow are conserved. The secondary structure of the sequence based on the PfIMP2 structure is shown above the sequences and the secondary structure elements are labelled above the schematic. TT refers to beta turns, η refers to 3_{10} helix, helices are shown as squiggles and beta strands are shown as arrows. The C71 residue appears to be conserved across all seven homologues.

Chapter 6

NMR assignments and Structure Determination of PfIMP2

6.1 NMR assignment

Double labelled ($^{13}C/^{15}N$) full length PfIMP2 without the His tag in 50 mM Hepes pH 7.5 150 mM NaCl 10 mM DTT was used to record NMR data. Following TEV cleavage, the PfIMP2 construct contains a first serine residue, followed by the PfIMP2 sequence. The molecular weight of PfIMP2 without the His tag is 17.89 kDa. The final construct used for NMR studies contains 159 residues but the methionine residue that follows the first serine residue (left over from the TEV cleavage) will be referred to as residue 1 throughout this thesis.

6.1.1 Backbone assignment

A 2D $^{1}H/^{15}N$ HSQC spectrum was recorded under the optimized conditions (303K, pH 7.5) to assign the backbone atoms of PfIMP2. A series of 3D experiments – HNCO, HNCACO, HNCACB and HNCOCACB – were then recorded to assign the specific residues in the ^{15}N HSQC. Backbone assignment of PfIMP2 was performed semi-automatically using MARS and then confirmed and completed manually. Figure 6.1.1a shows the assigned $^{1}H/^{15}N$ HSQC spectrum (Benjamin et al, 2015).

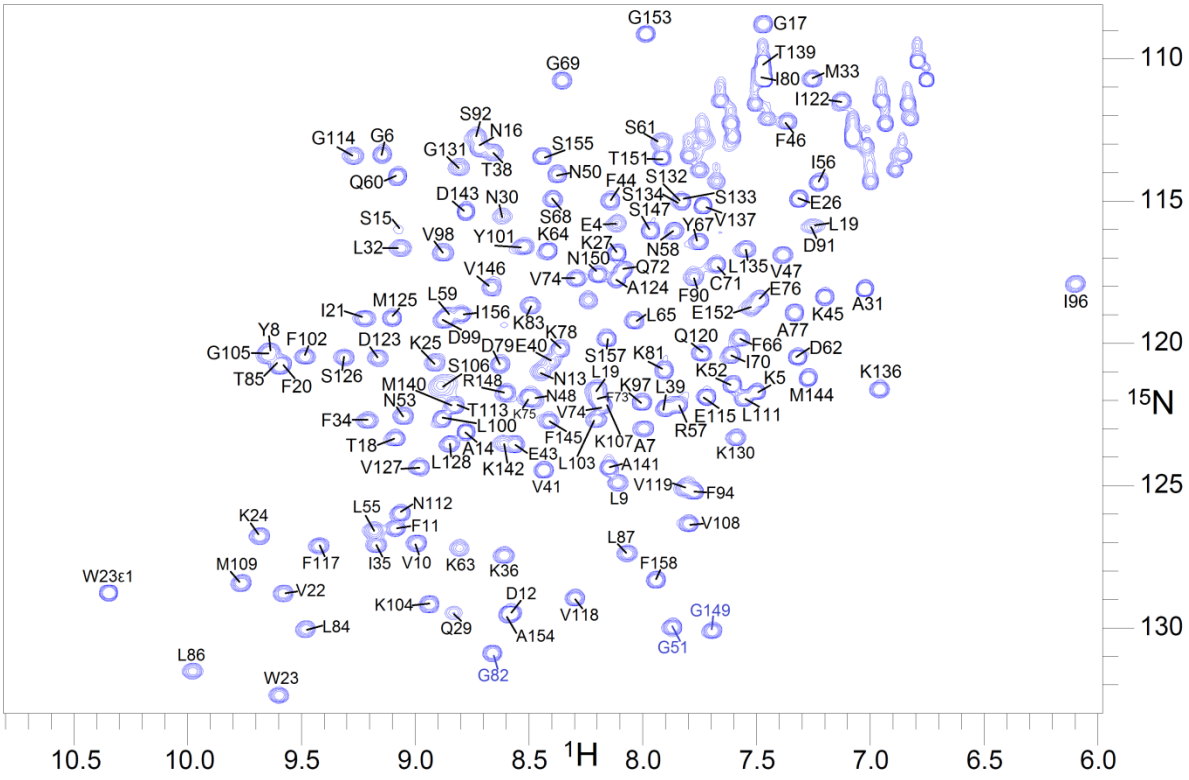


Figure 6.1.1a: Figure showing assigned $^{1}H/^{15}N$ HSQC spectrum of recombinant PfIMP2 without the His tag. The peaks are labelled with single letter amino acid code followed by their position in the recombinant PfIMP2 sequence. The underlined residues are aliased signals (G82, G51 and G149).

13% of the residues in the PfIMP2 sequence are lysines. This led to challenges in backbone assignment as the data from these lysine residues were heavily overlapped. ~96 % of all possible backbone atoms were assigned. The strips of CBCACONH and CBCANH and those of HNCO and HN(CA)CO for a stretch of assigned residues K100 to R104 are shown below in Figure 6.1.1b (Benjamin et al, 2015).

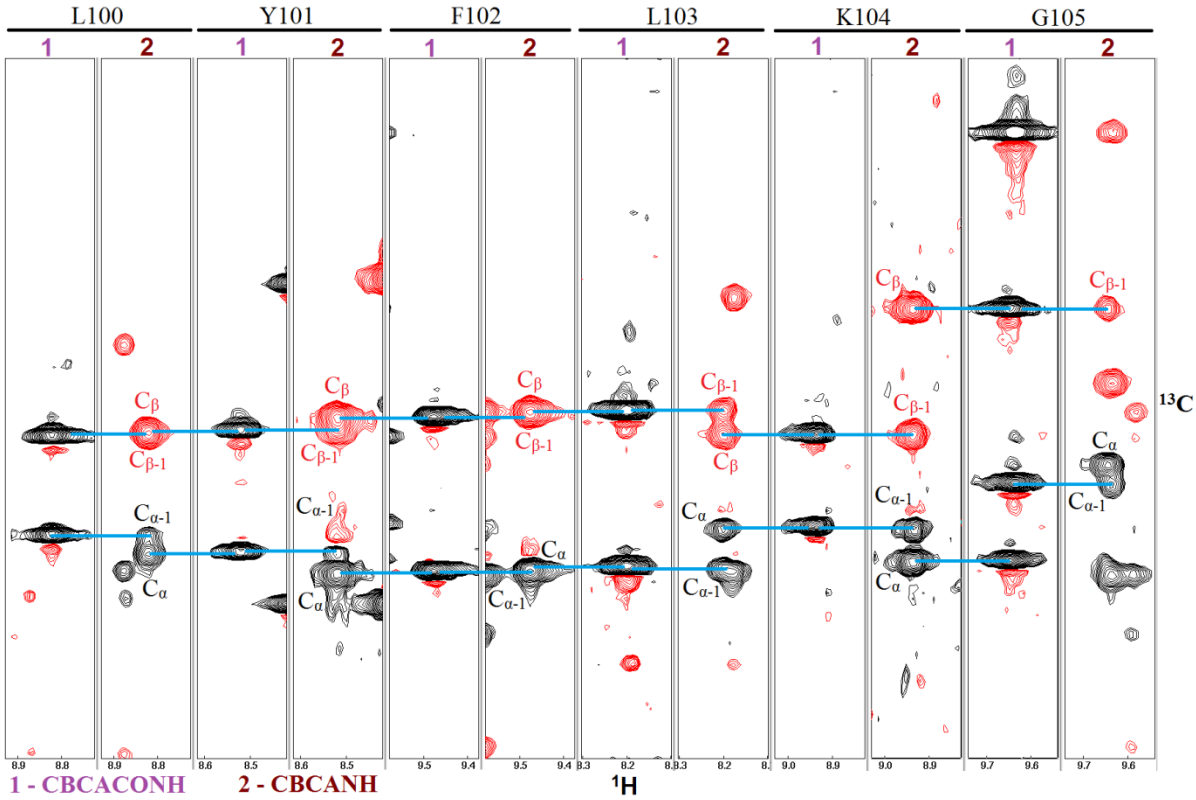


Figure 6.1.1b: Figure showing the strips that were obtained from the CBCACONH (1) and CBCACNH (2) experiments for double labelled ($^{13}C/^{15}N$) full length PfIMP2. Data is shown for residues L100 to G105. The blue lines connect the peaks that were used to assign the backbone atoms of the labelled residues. The atoms corresponding to the peaks are labelled in the figure.

6.1.2 Side chain assignments

Following backbone assignment, assignment of side chain atoms of PfIMP2 were performed using data from side chain experiments – HBHA(CBCACO)NH, (H)CC(CO)NH-TOCSY, HCCH-TOCSY and ^{1}H-C^{13}NOESY-HMQC spectra. ~94% of the amino acid side chain atoms were assigned manually using a modified version of the program NMRview (Marchant et al, 2008). The ^{13}C NOESY and ^{15}N NOESY strips for residue I35 with the assigned NOE peaks are shown in figure 6.1.2.

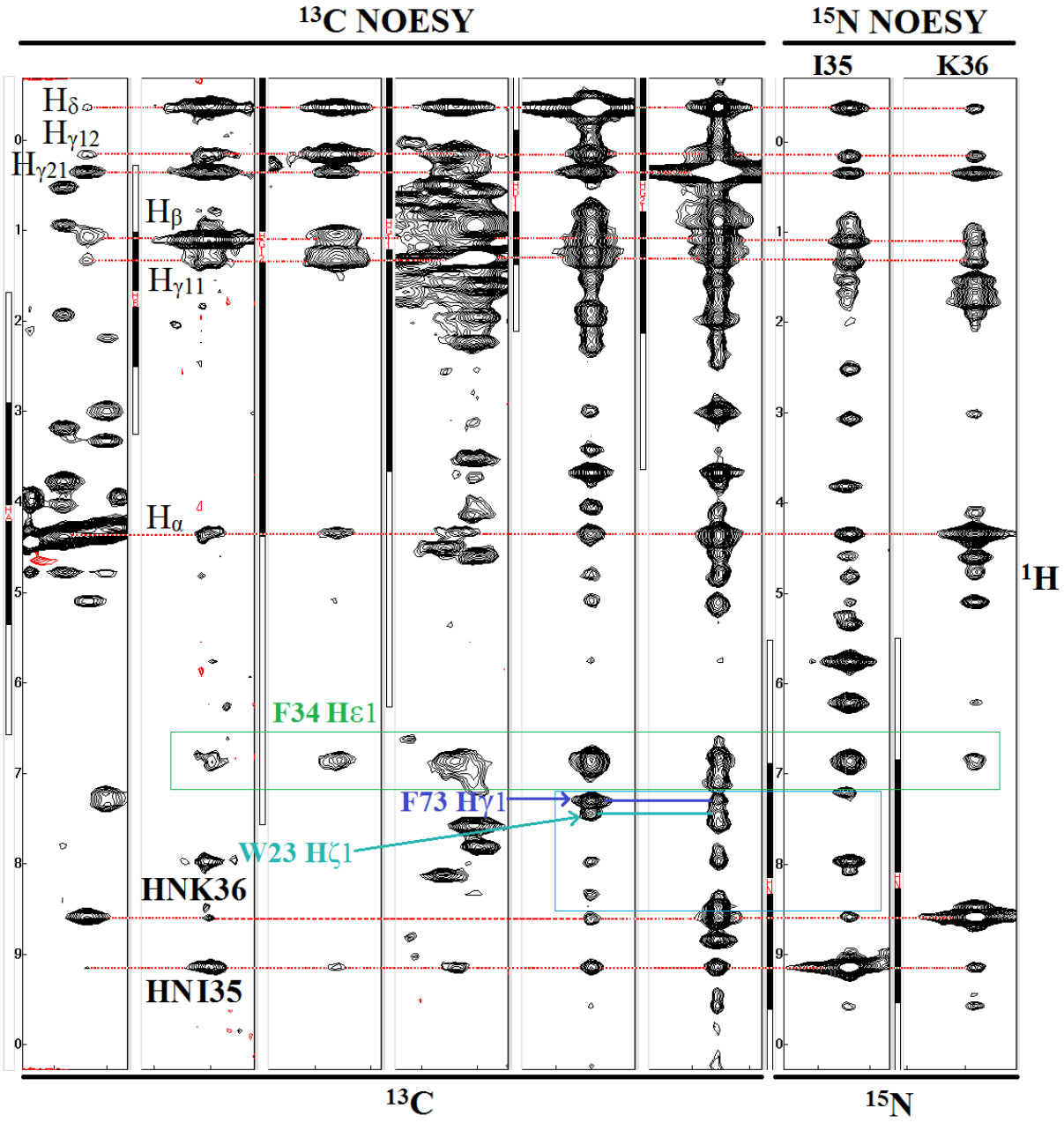
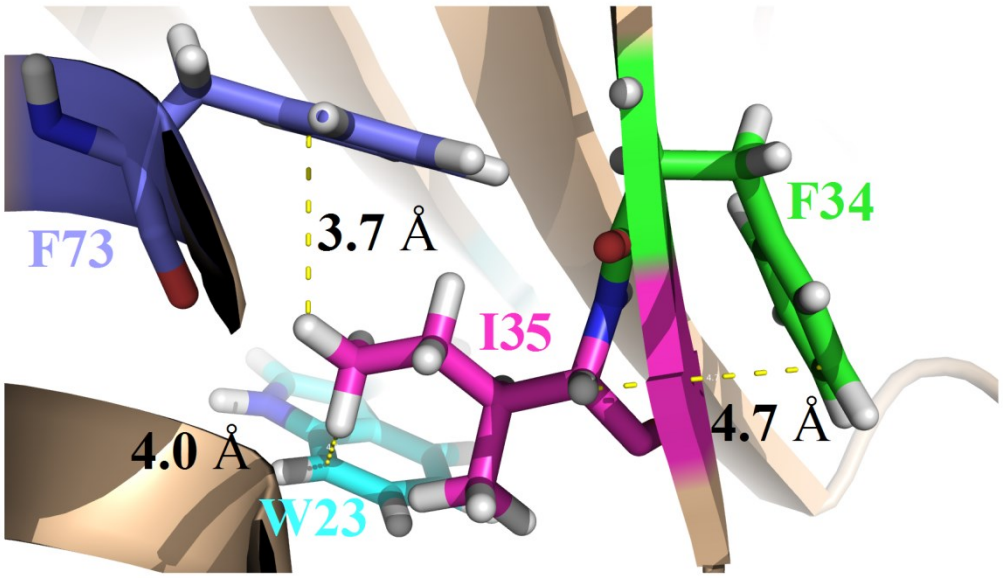


Figure 6.1.2a: Figure showing the ^{13}C-NOESY and ^{15}N-NOESY strips for residue I35 of double labelled (^{13}C/^{15}N) full length PfIMP2. The red dotted lines connect the peaks that correspond to the side chain atoms of I35 and the (HN) amide ^{1}H atom of the next residue, K36, across the NOESY strips as labelled in the figure. The NOE peaks labelled in the green box correspond to short range NOE from F34 Hε1 and the NOE peaks labelled in the blue box correspond to long range NOEs from other aromatic residues. The blue line shows the NOE peaks from F73 Hγ1 and the cyan line shows the NOE peaks from W23 Hζ1.

(A)



(B)

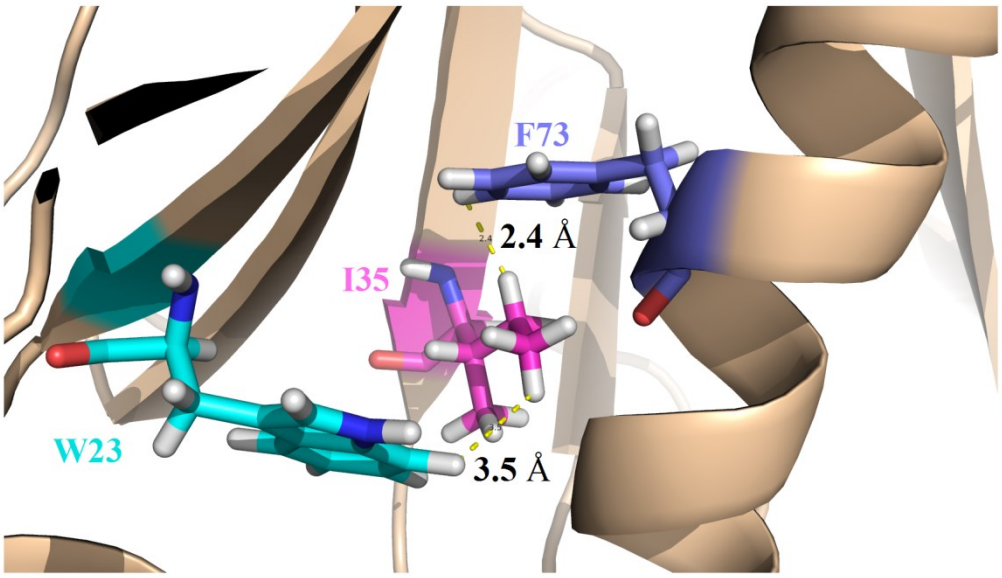


Figure 6.1.2b: Figure showing (A) the short range NOE between the F34 and I35 residues in the PfIMP2 structure. The distance between the atoms connected by the yellow dashed line is labelled in the figure. (B) the long range NOEs between the I35 (pink), W23 (cyan) and F73 (blue) residues in the PfIMP2 structure. The side chains of these residues are represented as sticks. ^{1}H atoms are coloured white, oxygen atoms are coloured red and ^{15}N atoms are coloured blue.

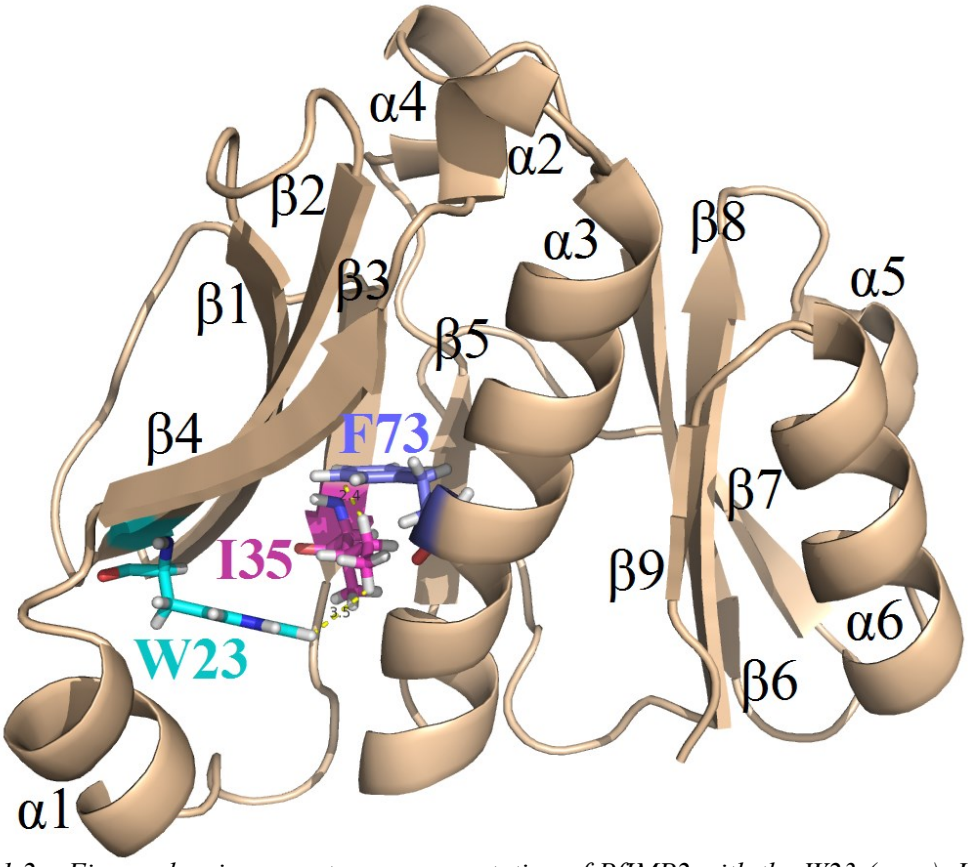


Figure 6.1.2c: Figure showing a cartoon representation of PfIMP2 with the W23 (cyan), I35 (pink) and F73 (blue) residues represented as sticks. W23 is present in the second beta sheet, I35 is present in the third beta sheet and F73 is present in the third alpha helix.

The peaks corresponding to the side chain ^{1}H and ^{13}C of the I35 residue in PfIMP2 are labelled in figure 6.1.2a. Figure 6.1.2c shows that the orientation of I35 is such that the side chain is protruding into the hydrophobic core of the protein. I35 is located in the third beta sheet in the protein and W23 and F73 are spacially proximal aromatic residues located in the second beta sheet and third alpha helix respectively. The distance between the H_β of I35 and the Hε1 of F34 is 4.7 Å, the Hγ of I35 and the Hζ of W23 is 4.0 Å and the Hγ of I35 and the Hγ of F73 is 3.7 Å (Figure 6.1.2b). The NOE between F34 and I35 is a short range NOE and that between I35 and W23 and between I35 and F73 are long range NOEs. The NOESY strips in figure 6.1.2a show that there are additional long range NOEs to I35.

6.2 Structure calculation

Structure calculation was performed through ARIA/CNS using the NOE peaklists, the backbone and side chain atom assignments and the dihedral angle restraints (Table 6.1). A total of 2319 NOE peaks were used for structure calculation, out of which 924 were intra-residual NOE peaks and 1395 were inter-residual NOE peaks. After several rounds of iteration, an ensemble of structures was obtained for PfIMP2 (figure 6.2). The structural statistics for the final structure of PfIMP2 are also given in Table 6.1. The average pairwise RMSD value for the heavy atoms shows that the resolution of the final structure is equivalent to a 0.55 Å crystal structure.

NMR restraints	
Distance constraints	
Total NOE	2319
Intraresidue	924
Interresidue	1395
Sequential (|i-j|)=1)	505
Short (2< (|i-j|)<3)	193
Medium range (4<(|i-j|)<5)	68
Long range (|i-j|)>5)	629
Total Dihedral angle Restraints	
Φ	127
Ψ	127
Structural Statistics	
Violations (mean and SD)	
Distance constraints (Å)	0.03943382 ± 0.003690475
Dihedral angle constraints (°)	1.0489
Maximum dihedral angle violation (°)	0.781025
Maximum distance constraint violation (Å)	0.3 ± 0.458258
Deviations from idealized geometry	

Bond length (Å)	0.004144955 ± 0.0001096
Bond angle (°)	0.583356 ± 0.009882051
Impropers (°)	1.59978 ± 0.05995317
Average Pairwise RMSD (Å)	
Heavy	0.55379 ± 0.07666714
Backbone	0.2475 ± 0.05153124

Table 6.1: Table showing the NMR restraints used for structure calculation and the structural statistics for the final structure of PfIMP2.

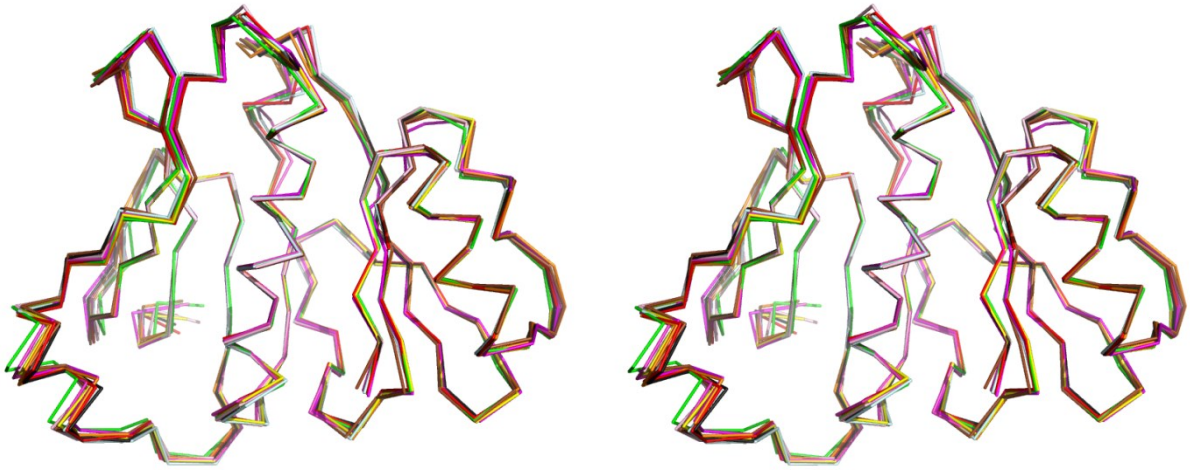
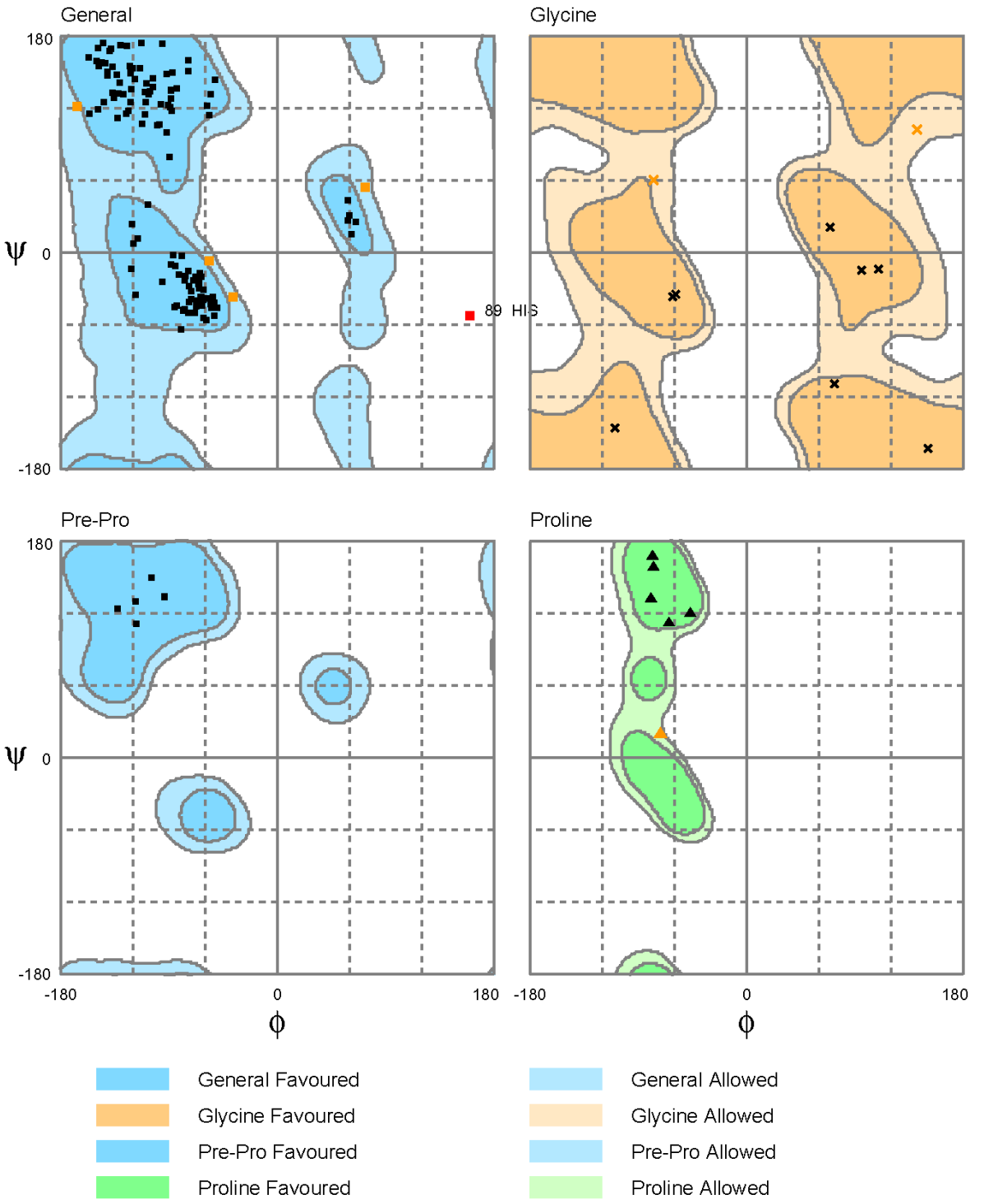


Figure 6.2: Figure showing a stereo representation of the ensemble of the 10 NMR structures obtained following the final ARIA run.

6.3 Ramachandran plot

The Ramachandran plot for the PfIMP2 structure (figure 6.3a) shows that there is one residue which falls in the outlier region – H89. This residue has been highlighted in the structure as shown in Figure 6.3b.



Number of residues in favoured region (~98.0% expected)	:	145 (94.8%)
Number of residues in allowed region (~2.0% expected)	:	7 (4.6%)
Number of residues in outlier region	:	1 (0.7%)

Figure 6.3a: Figure showing the Ramachandran plot for the PfIMP2 structure solved by solution state NMR. One residue falls in the outlier region – H89.

The Ramachandran plot shows that 99.3% of residues in the PfIMP2 structure fall in favoured or allowed regions with the exception of one outlier, H89. The structure shows that H89 is present in a loop region which is likely to be flexible and dynamic. The flexible nature of this region could possibly give rise to an unusual chemical shift for H89, causing it to appear as an outlier on the Ramachandran plot.

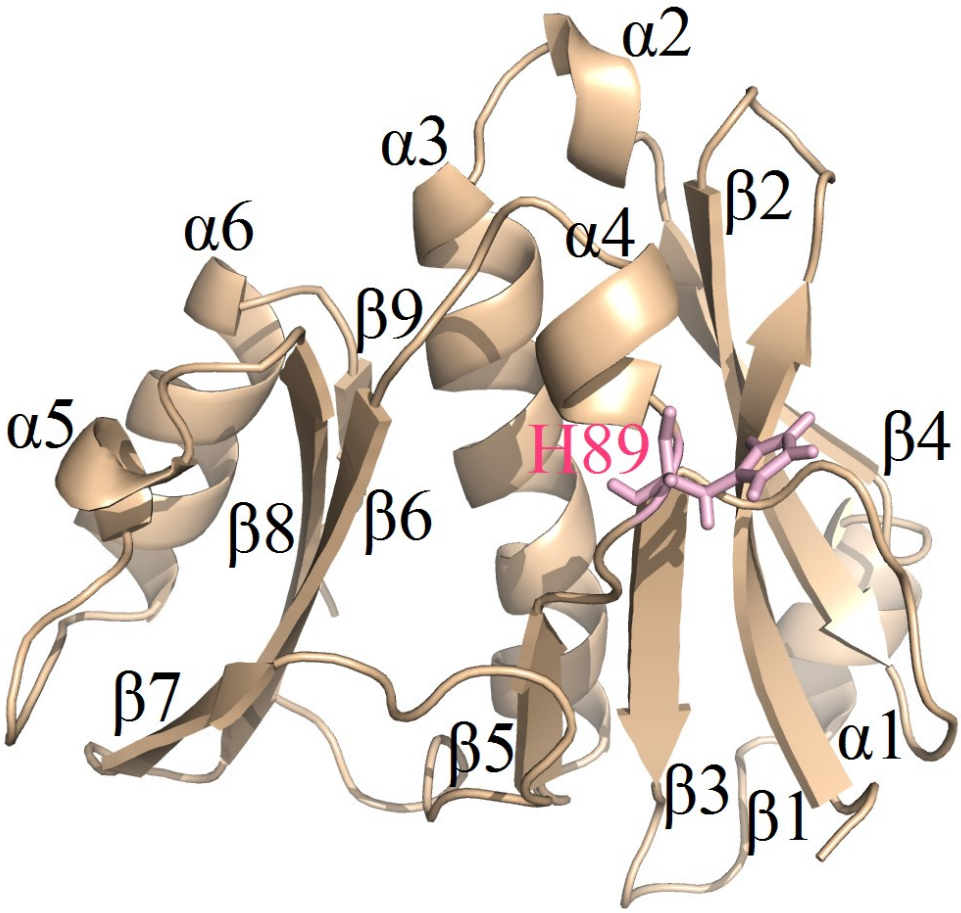


Figure 6.3b: Figure showing a cartoon representation of the structure of PfIMP2 with secondary structure elements labelled and the outlier residue, H89, represented as sticks and coloured pink. The H89 residue lies in the loop region between β5 and α4.

6.4 Solution state structure of PfIMP2

The secondary structure of PfIMP2 consists of six alpha helices – between residues 42-48, 58-60, 62-79, 89-92, 130-133 and 140-150. There are nine beta sheets – between residues 5-12, 16-23, 31-35, 51-56, 82-85, 97-102, 105-108, 121-127 and 152-154. The secondary structure elements of PfIMP2 have been labelled in figure 6.4a.

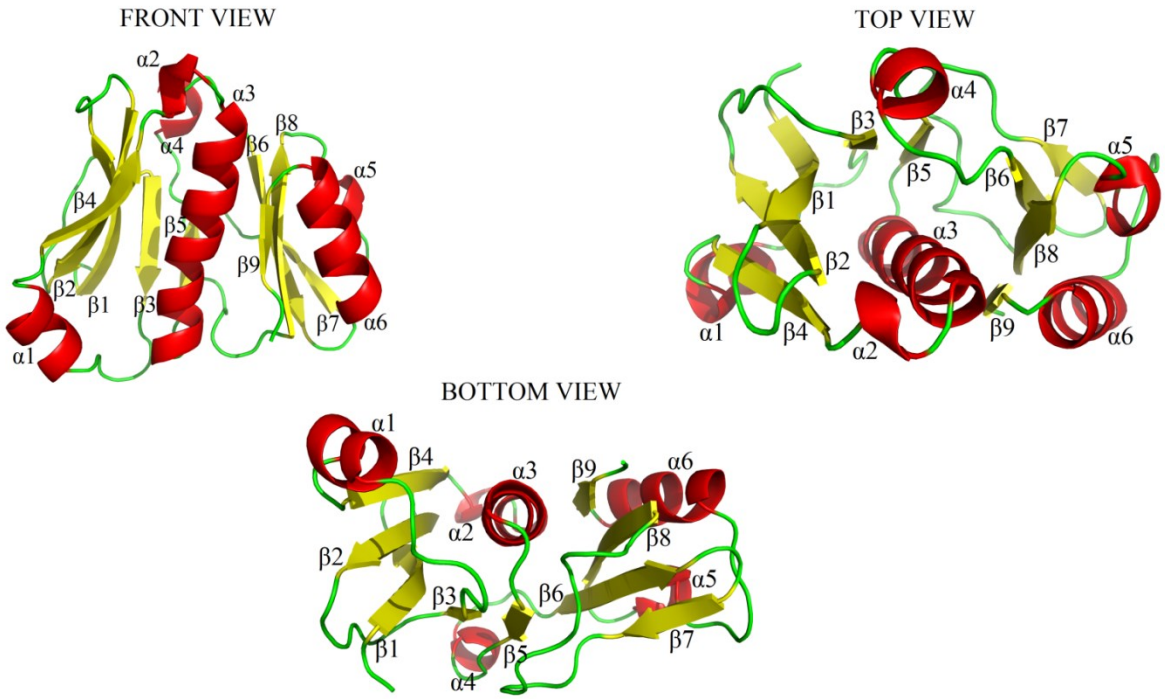


Figure 6.4a: Figure showing three views of the cartoon representation of the structure of PfIMP2 with the labelled secondary structure elements. Alpha helices are coloured red, beta sheets coloured yellow and loops coloured green.

The structure of PfIMP2 consists of a long alpha helix surrounded by nine beta sheets and five alpha helices which appear to form half a beta barrel. This structure is a novel fold which does not seem to have any homologues within the apicomplexan family excluding other Class I and Class II IMP proteins. The structure was analysed using DALI to verify that the protein contains a novel fold (Holm & Rosenström, 2010). The results shows that a MDB1 protein (PDB: 4s3h-C) had a Z-score of 3.8. Although Z-scores above 2 are considered to be significant, the sequence identity of these two sequences is 9%. In addition to this, the

structure of MDB1 shows that it is visually not similar to PfIMP2 as shown in figure 6.4b. The top three hits in DALI were to MDB1 protein homologues, all of which had a Z-score above 3.5. However, analysis of the structure of MDB1 confirms that MDB1 does not have a similar fold to PfIMP2 and PfIMP2 has a novel fold.

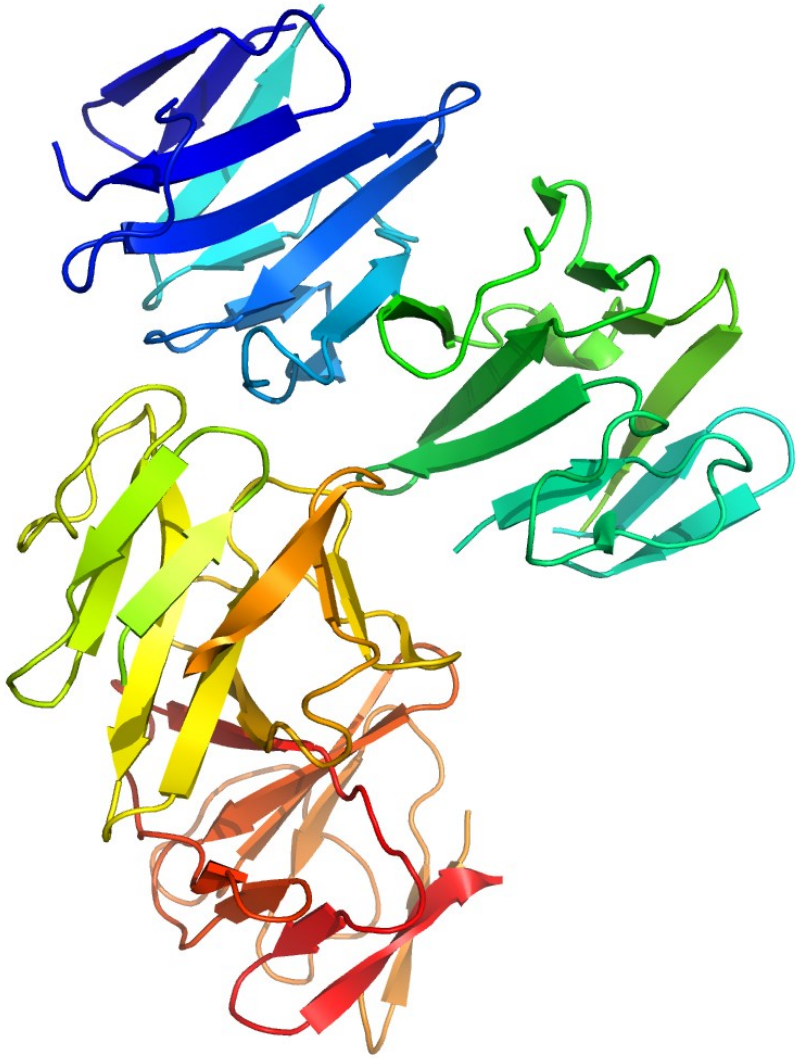


Figure 6.4b: Figure showing a cartoon representation of the structure of MDB1 coloured in rainbow colours from the N-terminus to the C-terminus. Although this structure has a Z-score of 3.8 on DALI, it is clear that MDB1 does not have a similar fold to PfIMP2.

6.5 Structural features of PfIMP2

Analysis of the PfIMP2 sequence shows that there are two predicted phosphorylation sites – S92 and S93 in the protein (figure 6.5a) (Florens et al, 2004). Both these residues are exposed

on the surface and present on the fourth α helix. Various proteomic studies by Floren et al (2004) have not identified any other post-translational modification sites in the protein during the schizont and merozoite stages of the parasite.

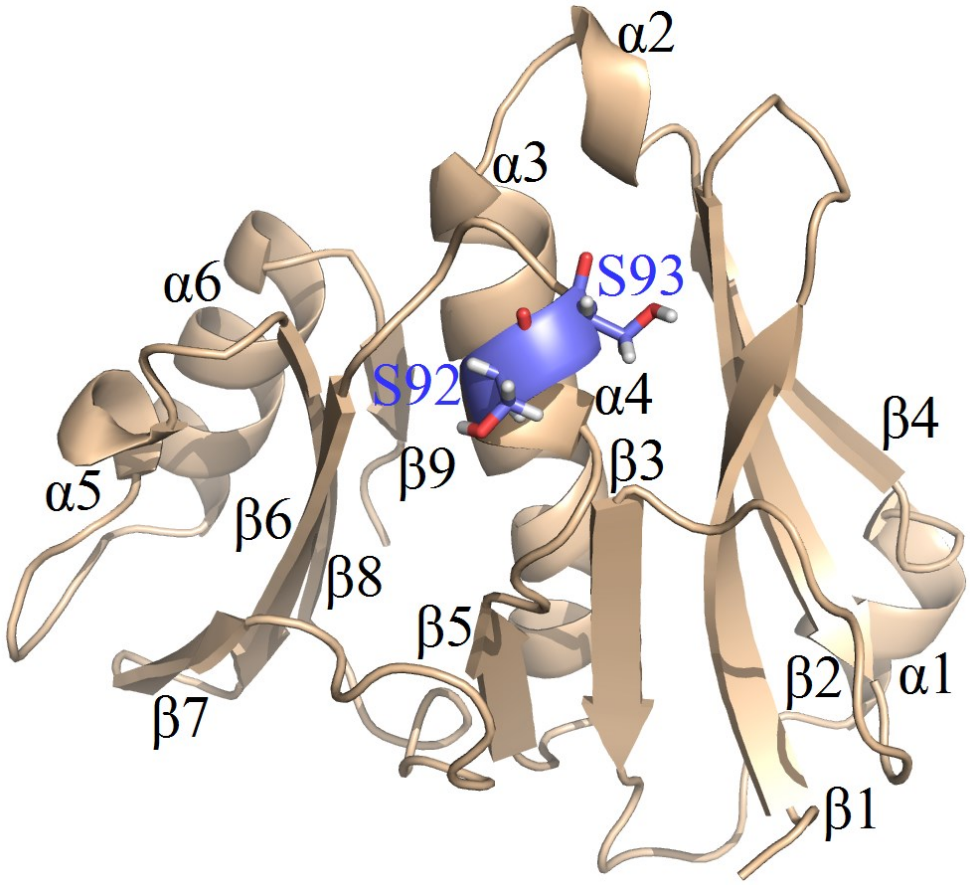


Figure 6.5a: Figure showing a cartoon representation of the structure of PfIMP2. The two predicted phosphorylation sites – S92 and S93 – are represented as sticks and coloured blue.

The only cysteine residue, C71, in the entire PfIMP2 amino acid sequence is present in the third α helix. The structure of PfIMP2 shows that the C71 residue is surface exposed which would allow it form inter-molecular disulphide bonds leading to protein aggregation which was observed following protein purification. Mutation of C71 to S71 was attempted. However, no soluble overexpression was observed. This suggests that C71 is necessary for proper folding and expression of PfIMP2. This cysteine residue is also conserved in other Class II IMP homologues.

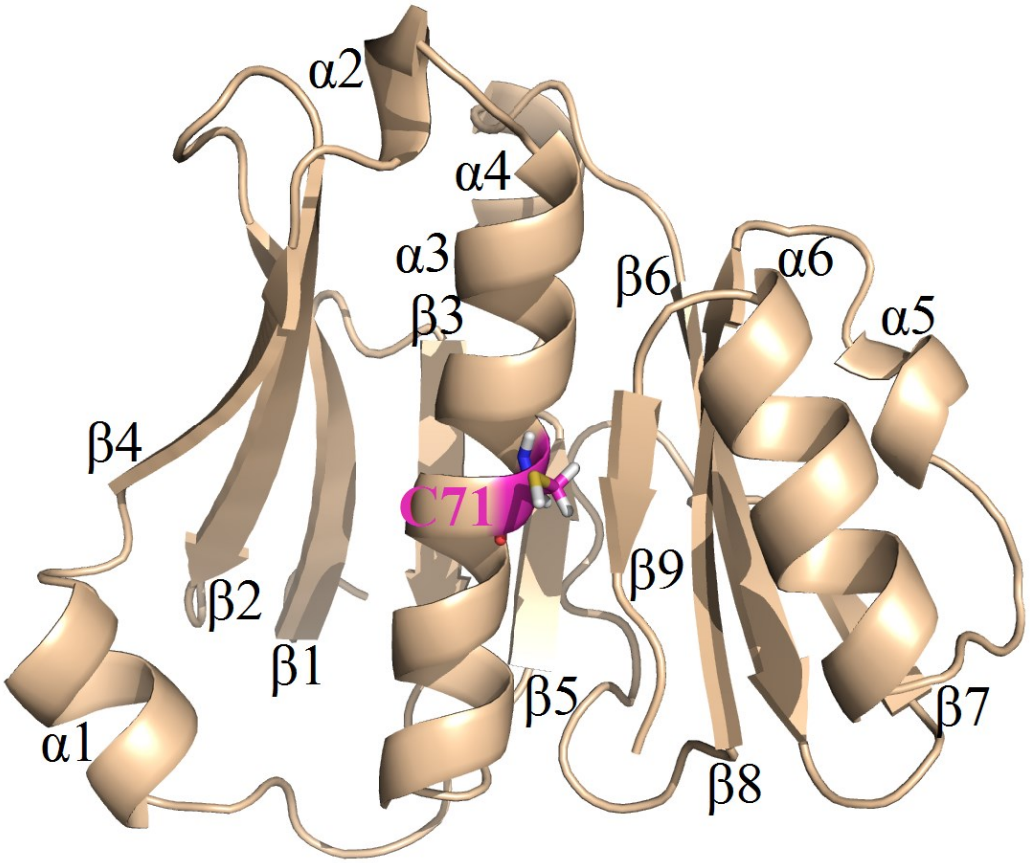


Figure 6.5b: Figure showing a cartoon representation of the structure of PfIMP2. The only cysteine residue in the PfIMP2 amino acid sequence, C71, is represented as sticks and coloured pink.

Analysis of the hydrophobicity of the surface of PfIMP2 shows the presence of a hydrophobic groove formed by the first three beta strands in the sequence. This groove could potentially form a binding pocket or an interface for the protein to interact with binding partners.

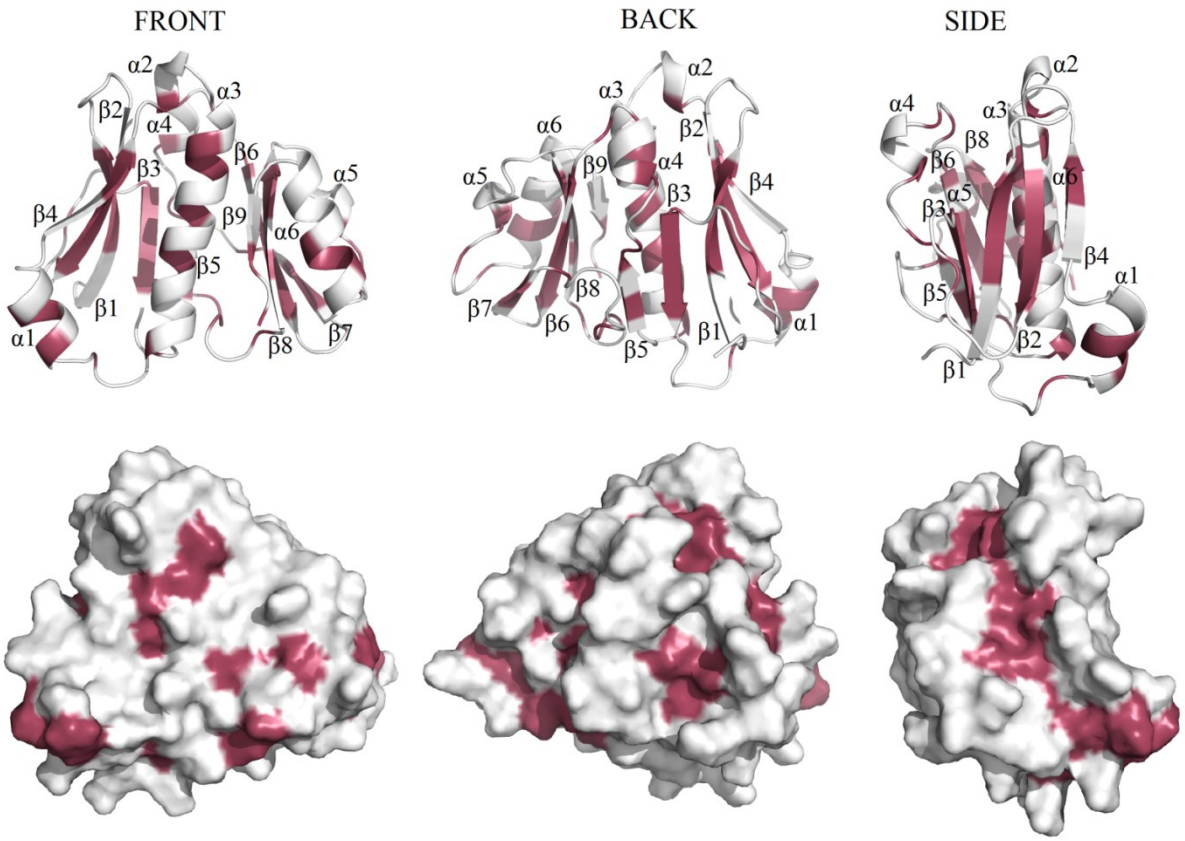


Figure 6.5c: Figure showing a cartoon representation of the structure of PfIMP2 in three views (top) and the hydrophobic surface representation of PfIMP2 in the same orientations (bottom). Hydrophobic residues (I, L, V, W, Y, M and F) are coloured purple and the rest of the residues are coloured white.

Analysis of the electrostatic surface representation of the protein shows that the protein is more positively charged overall due to the high amount of lysines in the sequence (13%). There is an extended patch of positive residues observed on the protein surface in the back view. This could potentially play a role in interacting with charged binding partners.

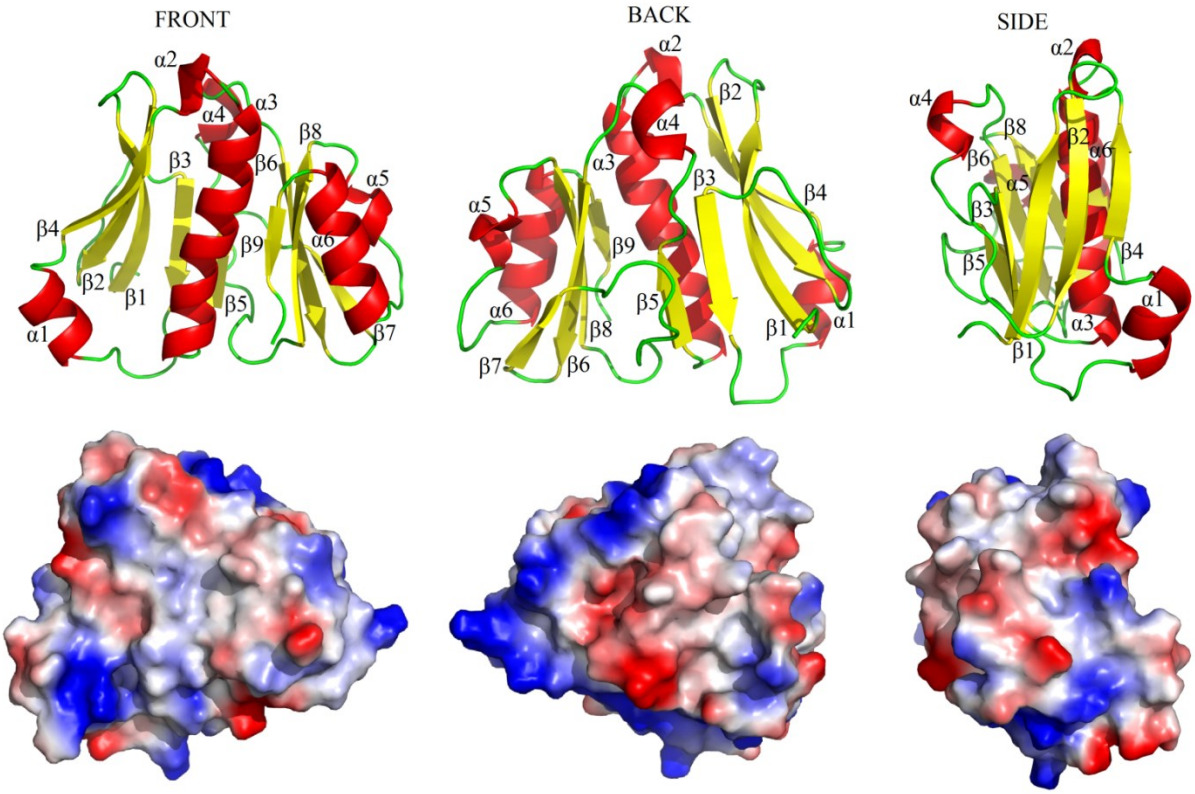


Figure 6.5d: Figure showing a cartoon representation of the structure of PfIMP2 in three views (top) with the alpha helices coloured red, beta sheets coloured yellow and loops coloured green and the electrostatic surface representation of PfIMP2 in the same orientations (bottom). Positive residues are coloured blue and negative residues are coloured red. An extended stretch of positive residues is observed on the protein surface in the back view.

Further analysis of the conservation of the residues in the PfIMP2 structure shows that the hydrophobic groove is conserved to some extent. This further confirms that this groove may have a functional role.

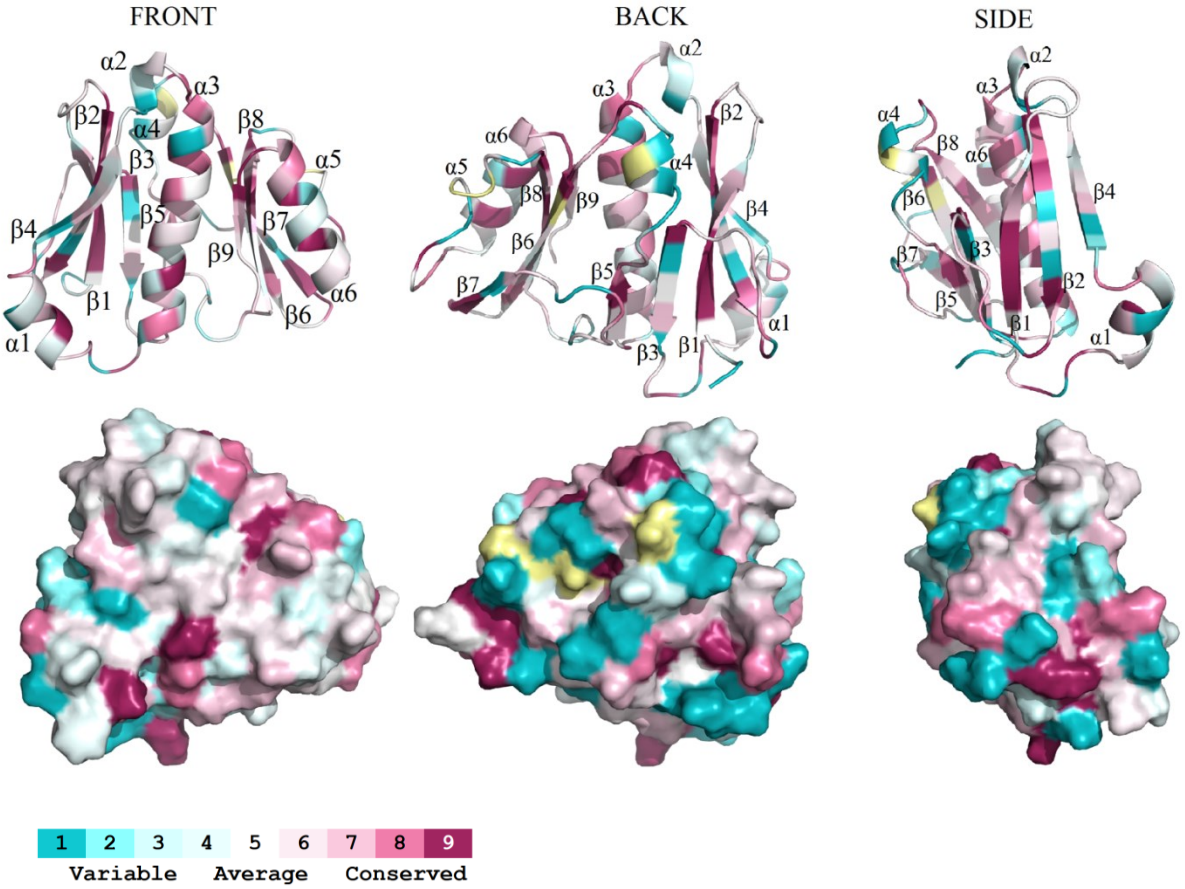


Figure 6.5e: Figure showing a cartoon representation of the structure of PfIMP2 in three views (top) and the level of conservation of the amino acid residues in PfIMP2 in the same orientations (bottom). The figure legend shows the conservation grades represented by the residue colours. The residues coloured yellow had insufficient data to calculate a reliable conservation grade.

6.6 Modelled structures of homologues – TgIMP1 and EtIMP1

The structures of the IMP domains of TgIMP1 and EtIMP1 were modelled based on the structure of full length PfIMP2 using Phyre2. The sequence alignment of Class II PfIMP2 with Class I TgIMP1 and EtIMP1 proteins shows that there are a few insertions in the Class I IMP sequences which are absent in the Class II sequence. These insertions appear to be present in similar positions in the TgIMP1 and EtIMP1 structures (Figure 6.6a and b).

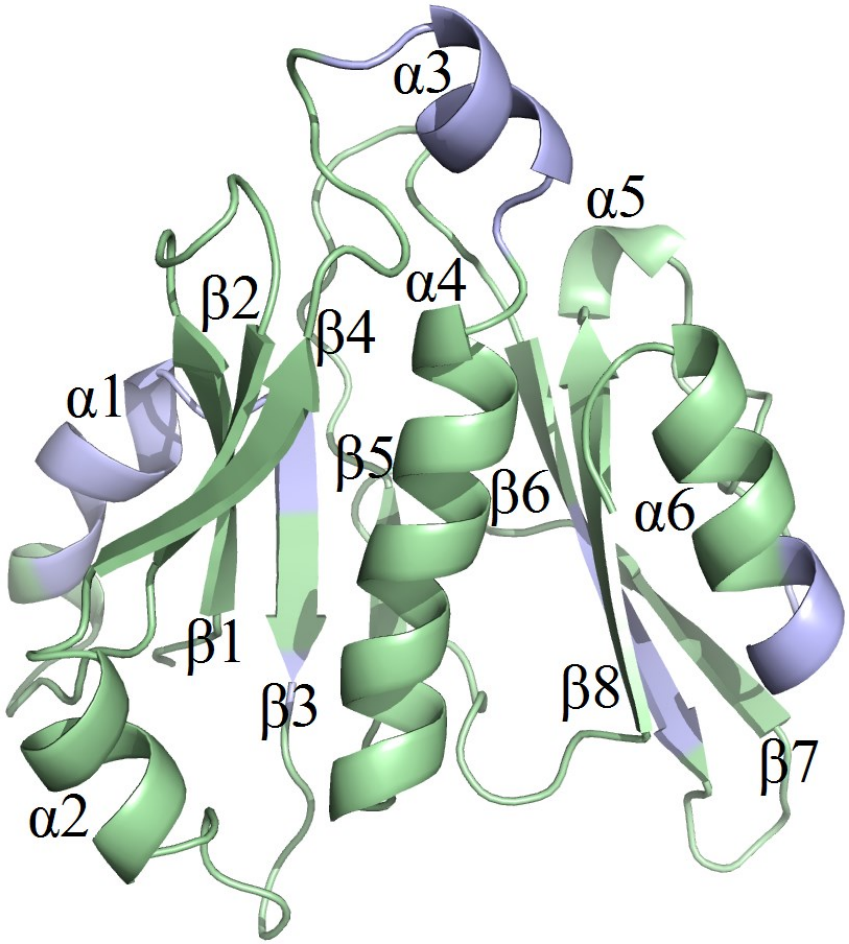


Figure 6.6a: Figure showing a cartoon representation of the modelled structure of the IMP domain of TgIMP1 (green) with the secondary structure elements labelled. The insertions in the structure are coloured blue.

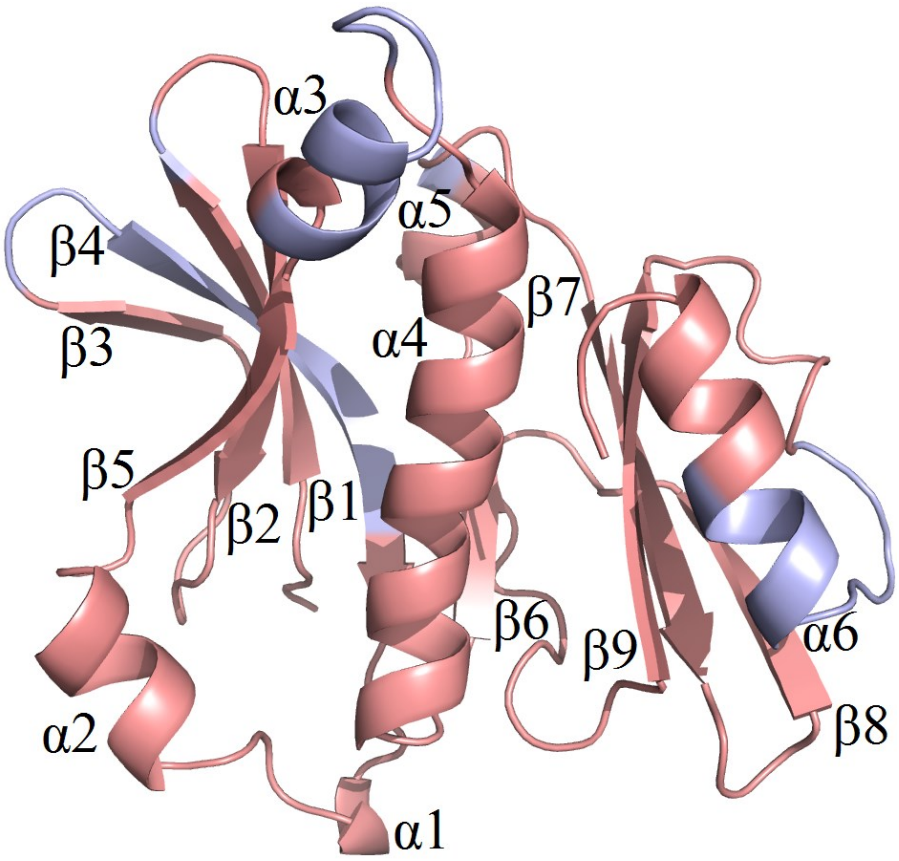


Figure 6.6b: Figure showing a cartoon representation of the modelled structure of the IMP domain of EtIMP1 with the secondary structure elements labelled. The insertions in the structure are coloured blue.

The conserved position of these insertions in the modelled structures of the Class I IMP homologues suggests that these insertions may have a functional role in the protein. It can also be observed that the IMP domain of TgIMP1 contains the hydrophobic groove similar to Class II PfIMP2 (Figure 6.6c). Although the IMP domain of EtIMP1 contains the conserved hydrophobic residues, the hydrophobic groove is not accessible for protein interaction (Figure 6.6c). Finally, the cysteine residue is not conserved in TgIMP1 or EtIMP1 but appears to be conserved across other Class II IMP proteins (Figure 5.5). This suggests that this cysteine residue may have a functional role which is specific to Class II IMP proteins.

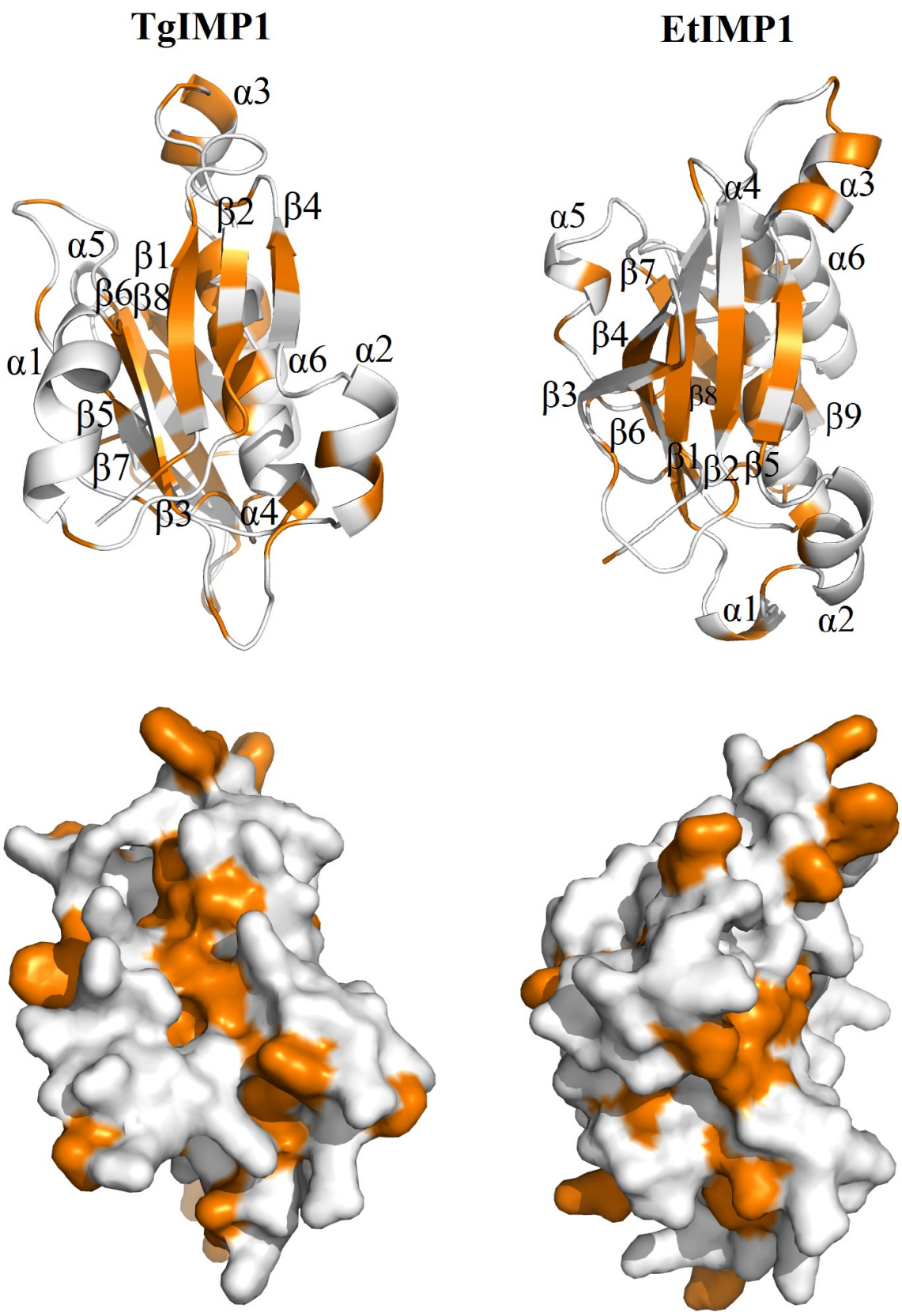


Figure 6.6c: Figure showing (top) a cartoon representation of the modelled structure of the IMP domain of TgIMP1 and EtIMP1 with the secondary structure elements labelled and (bottom) a hydrophobic surface representation of the same. The hydrophobic residues (I, L, V, W, Y, M and F) are coloured orange. The first three beta strands in the IMP domain of TgIMP1 appear to form a hydrophobic groove similar to PfIMP2.

6.7 Summary

The high prevalence of lysines (13%) in the PfIMP2 sequence led to overlap in the NMR data. However, 94% of all possible backbone atoms were successfully assigned. The data for side chain experiments was incomplete when the sample contained the His tag. Removal of the His tag solved this issue and allowed assignment of 96% of the side chain atoms. Structure calculation was performed using 924 intra-residual NOE peaks and 1395 inter-residual NOE peaks. 505 peaks of the inter-residual NOE peaks were sequential and 629 NOE peaks were long range. Following assignment and structure calculation, it was found that PfIMP2 contains a novel fold. The resolution of the final structure of PfIMP2 is equivalent to a 0.55 Å crystal structure. Structure validation shows that the structure has one outlier in the Ramachandran plot – H89. However, it has been found that this residue lies in a loop region in the structure. The high flexibility of the loop regions can cause the residues in these regions to have unusual chemical shifts. This can lead to these residues to be classified as outliers in the Ramachandran plot.

The structure of PfIMP2 appears to have a long alpha helix surrounded by half a beta barrel. Class II IMP proteins appear to have a conserved cysteine residue in the long alpha helix in the structure. This C71 residue is surface exposed and could potentially play a role in protein 173dhesion173ati by forming inter-residual disulphide bonds. However, the biological relevance of this potential dimer has not been explored. Bioinformatic analysis of the PfIMP2 amino acid sequence showed that there are two possible phosphorylation sites – S92 and S93. Analysis of the PfIMP2 structure shows that both S92 and S93 are present in the fourth alpha helix in the structure and are surface exposed. This suggests that the protein could potentially undergo post-translational modification in the parasite. However, lack of post-translational phosphorylation does not seem to affect protein folding. The structure also reveals that PfIMP2 contains a hydrophobic groove formed by the first three beta helices. This

hydrophobic groove is also conserved in the IMP domain of the TgIMP1 homologue. This suggests that this hydrophobic groove may play a role in protein interaction or act as a binding pocket. However, the structure of the IMP domain of the EtIMP1 homologue shows that although the hydrophobic residues are conserved, the hydrophobic groove is not accessible for interaction with other proteins. This suggests that the hydrophobic groove may also play a different role to protein interaction, which is yet to be determined.

Comparison of the modelled structures of the IMP domain of the TgIMP1 and EtIMP1 homologues to the PfIMP2 structure shows that the Class I homologues contain a few insertions in the amino acid sequences. Mapping these insertions on the TgIMP1 and EtIMP1 structures shows that these insertions are present in similar regions in the structure. This suggests that these insertions may have a functional significance, which is specific to the Class I homologues. It is also likely that these insertions are responsible for the species-specific antigenicity of these proteins.

Structural analysis of PfIMP2 has revealed many features of this protein which were previously unknown. Both TgIMP1 and EtIMP1 have been found to have immunogenic effects on the hosts in a species-specific manner. Analysis of the amino acid sequences and the structures of TgIMP1, EtIMP1 and PfIMP2 homologues has shown that the IMP domain is conserved across all three proteins. This suggests that the IMP domain is the functional domain in IMP homologues. Recent studies by Jia et al (Appendix 1.4) have shown that the TgIMP1 contains a plasma membrane localisation signal at the N-terminus which localises the TgIMP1 protein to the inner leaflet of the plasma membrane. Sequence alignment shows that this plasma membrane localisation signal is conserved across Class I IMP homologues. However, Class II IMP homologues including PfIMP2 lack this sequence which raises questions about the localisation of Class II IMP homologues, particularly PfIMP2.

Chapter 7

Final Remarks

7.1 Immune Mapped Protein 1 (IMP1) – a novel immunogenic protein

IMP1 is an immunogenic protein which was first identified in *Eimeria maxina (Blake et al, 2011)*. Bioinformatic analysis of this protein has shown that IMP1 has homologues across most apicomplexan parasites. This conservation of the protein across the apicomplexan family suggests that it plays an important role in the parasite. Further studies have shown that mice vaccinated with a DNA vaccine of TgIMP1 had a longer life span and increased immunity against *T. gondii* infections (Cui et al, 2012).

Studies by another group have also shown that vaccination of chickens with a C-terminal fragment of EtIMP1 (EtIMP1-C) is sufficient to protect chickens against *Eimeria* infections. The same group has gone on to show that the efficacy of the EtIMP1-C vaccine is the same as the efficacy of the full length EtIMP1 vaccine. This has shown that the C-terminal domain is the immune-dominant region in IMP1 proteins (Nolan et al, 2015; Yin et al, 2014). All these studies have shown that IMP1 is an immunogenic protein that raises immunity against the parasite in a species-specific manner.

The immunogenic nature of IMP1 protein made it an interesting candidate to study. IMP1 proteins do not share sequence similarities with any proteins other than with other IMP homologues. None of these homologues have a known structure.

7.2 Domain architecture of IMP1

TgIMP1 and EtIMP1 were initially chosen for structural studies. ^{1}H 1D NMR of full length TgIMP1 showed that the protein has a large disordered domain which was masking the signal from the folded domain. The intensity and resolution of the folded signal improved upon truncation of the N-terminal region of the protein. Limited proteolysis of TgIMP1 showed that the folded globular domain of TgIMP1 starts close to S176. Bioinformatic analysis revealed that S176 was the start point for some IMP1 homologues. ^{1}H 1D NMR of this

optimised construct of TgIMP1 (clone 7) revealed that the protein was more compact. The 2D $^{1}H/^{15}N$ HSQC further confirmed that this construct is suitable for structural studies by NMR. However, the data recorded from 3D experiments with this sample were incomplete. Since the molecular weight of TgIMP1 clone 7 is about 26 kDa including the His tag, it was assumed that the gaps in the data were caused by the large molecular weight of the sample. Since the molecular weight was above 25 kDa, the protein was deuterated. 3D NMR experiments were performed on the deuterated sample. However, the data was still incomplete and less than 50% of the backbone atoms could be assigned with 100% confidence.

EtIMP1 was also studied and it was found that a shorter construct of EtIMP1 starting from S152 consisted of the folded domain similar to TgIMP1. Limited proteolysis of the EtIMP1 short construct with chymotrypsin led to protein degradation, whereas limited proteolysis with trypsin did not affect the protein. These results suggested that the short construct does not have any disordered region at the N-terminus. Analysis of the $^{1}H/^{15}N$ TROSY-HSQC spectrum of EtIMP1 clone 2 revealed that the number of peaks observed in the spectrum was less than the total number of amino acids in the sequence. Thus, the NMR data for this construct was incomplete. This showed that EtIMP1 clone 2 was also not suitable for NMR studies.

Crystallisation trials of TgIMP1 clone 7 were unsuccessful even at concentrations of 190 μM. This suggested that TgIMP1 clone 7 still contained dynamic regions within the folded domain which were unfavourable for crystal formation. Although structural studies of both these homologues were unsuccessful, these studies revealed that the folded domain in the IMP1 protein is present at the C-terminus and the N-terminal region of the protein is largely unfolded. However, recent studies by Jia et al (Appendix 1.4) have revealed that the first 11 amino acids in the TgIMP1 sequence are necessary for the localisation of the protein

to the plasma membrane. The study also reveals that TgIMP1 is localised to the inner membrane leaflet of the plasma membrane. Recent studies have also revealed that the immune dominant region in EtIMP1 is the same as EtIMP1 clone 2. This result confirms that the globular domain that was isolated for structural studies contained the IMP domain of the protein. IMP1 proteins, therefore, contain a plasma membrane localisation signal at the N-terminus, followed by a long linker sequence and the C-terminal folded domain, which we refer to as the IMP domain.

7.3 Two classes of IMP proteins

BLAST searches against the TgIMP1 protein sequence revealed a homologue in *P. falciparum*. However, this homologue was much shorter than TgIMP1 and EtIMP1. The PfIMP2 homologue was homologous to the IMP domain of TgIMP1 and EtIMP1. Further 178dhesion178atics analysis of this sequence revealed there are two classes of IMP homologues – Class I and Class II.

The Class I IMP homologues (IMP1) contain a plasma membrane localisation signal at the N-terminus, followed by a linker and the IMP domain. The Class II IMP homologues (IMP2) lack the N-terminal plasma membrane localisation signal and the linker and contain only the IMP domain. PfIMP2 which is a Class II IMP homologue was studied to determine the structure of the IMP domain. As the Class II homologue lacks the localisation signal, the localisation of this homologue within the parasite is not known.

7.4 PfIMP2 has a novel fold

PfIMP2 was cloned into pNIC-ZB vector (with a TEV cleavable N-terminal His tag) and expression was tested in Rosetta2 DE3 cells. Overexpression of PfIMP2 was successful and ^{1}H 1D NMR revealed that the protein is folded. However, constant precipitation was observed in the protein sample following purification. The PfIMP2 protein sequence contains a single

cysteine residue, C71, which could potentially form inter-molecular disulphide bridges. A buffer screen revealed that the optimum buffer for this protein is 50 mM Hepes pH 7.5 150 mM NaCl 10 mM DTT. The use of a reducing agent eliminated protein precipitation. It was also observed that the protein runs as a smear on the gel at protein concentrations of 0.5 mg/ml and above. The theoretical pI of this protein is 9.34, which is unusually high. The high pI is due to the high amount of lysines (13%) in the protein sequence. The overall positive charge of the protein and the possible self-association of the protein via the single cysteine residue are possible explanations for the unusual behaviour of the protein on the SDS-PAGE gel. The $^{1}H/^{15}N$ TROSY-HSQC spectrum of PfIMP2 showed that the protein is suitable for structural studies by NMR as the peaks in the spectrum were dispersed.

The sample used for recording preliminary NMR data contained a N-terminal His tag, followed by a 22-amino acid long linker which was followed by the PfIMP2 protein. The 3D experiments recorded with this sample had gaps in the data, possibly due to the flexible nature of the His tag. The 3D experiments recorded following removal of the His tag did not have any gaps in the data. ~94% of possible backbone residues and 96% of side chain atoms were successfully assigned (Benjamin et al, 2015).

Structure calculation was performed through ARIA/CNS by providing the chemical shifts of the atoms, the NOESY peaklists and dihedral angle restraints. 2319 NOE peaks were provided for the structure calculation, out of which, 924 were intra-residual NOE peaks and 1395 were inter-residual NOE peaks. ARIA produced an ensemble of 10 NMR structures with a resolution equivalent to a 0.55 Å crystal structure. The Ramachandran plot showed that H89 was an outlier in the PfIMP2 structure. This residue is present in a loop region between the fifth beta strand and the fourth alpha helix. Loop regions in proteins are highly dynamic and this dynamic nature could be the reason for the H89 atoms to have unusual chemical shifts and to be classified as an outlier.

Further analysis of the structure of PfIMP2 showed that the protein has a long alpha helix surrounded by a semi-beta barrel. PfIMP2 is predicted to have two phosphorylation sites – S92 and S93. These residues are present on the fourth alpha helix in the structure. It is evident that protein folding is not compromised by lack of phosphorylation as overexpression of soluble protein was achieved in bacteria.

The single cysteine residue in the PfIMP2 sequence, C71 is present in the third alpha helix in the structure. Although this residue was reduced when performing structural studies, it is present on the surface such that it can form inter-molecular disulphide bridges. Depending on the localisation of this protein, PfIMP2 can potentially form dimers. Mutagenesis studies have shown that when C71 was mutated to S71, no soluble overexpression was observed. This shows that this C71 residue is necessary for proper folding of the protein. Bioinformatic analysis shows that the C71 residue is conserved across other Class II IMP homologues. The structure also reveals that the protein has a hydrophobic groove formed by the first three beta sheets in the structure. This groove could play a role in protein-protein interaction.

7.5 Comparison of Class I and Class II IMP structures

The structures of the IMP domains of the TgIMP1 and EtIMP1 homologues were modelled based on the structure of PfIMP2 using Phyre. Sequence alignment of the three proteins reveals that the Class I homologues contain a few insertions in the IMP domain which are missing in the Class II homologue. When these insertions were mapped on the structure, it was found that the localisation of these insertions is conserved between the two Class I homologues. This suggests that these insertions may have a functional role. It is also likely that the insertions may be responsible for the species-specific antigenicity of IMP1 homologues.

Bioinformatic analysis has revealed that the hydrophobic groove found in PfIMP2 is conserved among other Class II IMP homologues. Analysis of the modelled structures of the IMP domains of TgIMP1 and EtIMP1 shows that although the hydrophobic residues are conserved in the first four beta strands of the protein, the groove is exposed in the TgIMP1 homologue but not in the EtIMP1 homologue. This suggests that although the hydrophobic groove may be functionally significant in TgIMP1 and PfIMP2, it does not appear to be so in the EtIMP1 homologue.

7.6 Localisation and Pull down studies

PfIMP2, with C-terminal GFP and triple HA tags, was ectopically expressed in the *P. falciparum* 3D7 strain (Appendix II). The expression profile of PfIMP2 shows that the protein is expressed mainly during the schizont and early ring stages. This suggests that the protein might play a role in invasion. IFA was performed to study the localisation of the protein within the parasite. Since the expression profile revealed that the protein may be involved in the invasion mechanism, it was hypothesised that PfIMP2 might be an IMC protein. In order to confirm this hypothesis, GAP45 was used as a marker for IFA. However, the IFA results proved to be inconclusive.

The structure of PfIMP2 also revealed the presence of a hydrophobic groove which could potentially form protein-protein interactions. A pull down was performed to check for any binding partners of PfIMP2 within the parasite. The pull down results did not reveal any differences between the control sample and the PfIMP2 strain. A western blot was performed to check for the presence of the HA-tagged PfIMP2 protein in the pull down sample, which revealed that PfIMP2 was expressed in very low amounts in the parasite. Therefore, the results obtained for the pull down assay were not conclusive due to the low concentration of PfIMP2 expressed in the parasite.

Due to time limitations and problems with ectopic overexpression of the PfIMP2 protein, these experiments could not be pursued further.

7.7 Challenges and further work

Although we now possess knowledge about the structure of PfIMP2, because the structure is a novel fold, it does not provide any insight about the functional role of this protein within the parasite. Identifying the localisation of this protein within the parasite might allow us to draw closer to the role played by this protein within the parasite. Class I TgIMP1 has been found to be localised to the inner leaflet of the plasma membrane. It is likely that the Class II homologues are also localised in the same way. However, due to the lack of the localisation signal in Class II homologues, the mechanism of localisation is not the same and is yet to be identified.

Various studies have shown that numerous apicomplexan proteins function as complexes such that one of the proteins in the complex anchors the complex to the parasite plasma membrane whereas the other proteins serve as a signalling molecule or an 182dhesion. Since Class I IMP homologues are localised to the plasma membrane, Class II IMP homologues can be localised to the plasma membrane by a chaperone protein. However, the localisation of Class II IMP proteins and potential binding partners are yet to the identified.

Endogenous expression of tagged PfIMP2 might allow us to identify the localisation of the protein by IFA and identify any binding partners or chaperones that the protein may interact with. Since the localisation of the Class I TgIMP1 homologue is now known, a plasma membrane marker can be used instead of GAP45 in order to confirm if PfIMP2 is also localised to the plasma membrane.

Although IMP1 has been found to be localised to the inner leaflet of the plasma membrane, various studies have shown that vaccination of hosts with IMP1 raises immunity against the parasite in a species-specific manner. This suggests that at some stage of the parasite life cycle, IMP1 might be presented on the parasite surface such that it is recognised by the host immune system. Although there is no evidence for this so far, it is worth studying the localisation of the protein in parasites at various stages in order to analyse whether the localisation of the protein remains the same or if it changes.

Recent studies by Jia et al (Appendix 1.4) have shown that the acylation of specific N-terminal residue in TgIMP1 are necessary for plasma membrane localisation of this protein. PfIMP2 is also predicted to be phosphorylated at one of two possible residues – S92 and S93, at different stages of the parasite life cycle (Florens et al, 2004). It will be interesting to generate mutants which lack these phosphorylation sites to study if lack of phosphorylation affects localisation of the protein.

PfIMP2 also contains a single cysteine residue, C71, which appears to be necessary for protein folding. This residue is also capable of forming inter-molecular disulphide bridges leading to dimer formation. However, the functional relevance of these dimers is unknown. The cysteine residues were reduced for structural studies and this did not affect the structure. Depending on the localisation of the protein, if the protein is oxidised, the protein can form dimers due to the exposed nature of this residue in the structure. Endogenous overexpression of tagged PfIMP2 coupled with pull down assays will allow us to check for the presence or absence of dimers in the parasite.

Bioinformatic analysis shows that most apicomplexan parasites contain homologues of both Class I and Class II IMP proteins. This shows that this protein is highly redundant within the parasite which further highlights the importance of this protein. Generating

knockouts of this protein would not give us a clear indication regarding protein function as the function would be complemented by one of the other homologues within the parasite. Thus, the redundancy of the protein within the parasite makes function determination even more challenging.

Recent studies by Yin et al (2014) and Nolan et al (2015) have shown that the IMP domain of EtIMP1 is the immune dominant domain in the protein. Now that the structure of the IMP domain is known, it is possible to perform NMR titrations to identify which residues and structural regions in the protein interact with antibodies. Once these residues are identified, mutagenesis studies can be performed to further confirm the core residues which are involved in raising immunity against the parasite.

The NMR structure of PfIMP2 and the modelled structure of the IMP domains of TgIMP1 and EtIMP1 show that although there are variations in the amino acid sequences of the proteins, they are structurally very similar. This suggests that the minor variations in the structure of these proteins and the insertions they possess or lack may be the reason for the species-specific immune response raised by these proteins.

Even though performing structural studies of IMP1 proteins was challenging, solving the structure of PfIMP2 was only a small step towards identifying a potential vaccine candidate. Understanding the function of this protein within the parasite and the mechanism by which it raises immunity against the parasite will give us insight about the viability of developing this protein into a potential vaccine candidate.

Bibliography

1. Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. (2000) Automation of X-ray crystallography. *Nat Struct Biol*, 7 Suppl, 973-7.
2. Aguilar, R., Magallon-Tejada, A., Achtman, A. H., Moraleda, C., Joice, R., Cisteró, P., Li Wai Suen, C. S., Nhabomba, A., Macete, E., Mueller, I., Marti, M., Alonso, P. L., Menéndez, C., Schofield, L. & Mayor, A. (2014) Molecular evidence for the localization of Plasmodium falciparum immature gametocytes in bone marrow. *Blood*, 123(7), 959-66.
3. Alexander, D. L., Mital, J., Ward, G. E., Bradley, P. & Boothroyd, J. C. (2005) Identification of the moving junction complex of Toxoplasma gondii: a collaboration between distinct secretory organelles. *PLoS Pathog*, 1(2), e17.
4. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol*, 215(3), 403-10.
5. Bai, X. C., McMullan, G. & Scheres, S. H. (2015) How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci*, 40(1), 49-57.
6. Bax, A., Clore, G. M. & Gronenborn, A. M. (1990) 1H□1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *Journal of Magnetic Resonance (1969)*, 88(2), 425-431.
7. Beghetto, E., Nielsen, H. V., Del Porto, P., Buffolano, W., Guglietta, S., Felici, F., Petersen, E. & Gargano, N. (2005) A combination of antigenic regions of Toxoplasma gondii microneme proteins induces protective immunity against oral infection with parasite cysts. *J Infect Dis*, 191(4), 637-45.
8. Benjamin, S., Williams, F., Kerry, L. & Matthews, S. (2015) NMR assignment of the immune mapped protein 1 (IMP1) homologue from Plasmodium falciparum. *Biomol NMR Assign*, 9(2), 393-5.
9. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J. & Kelley, L. A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, 70(3), 611-25.
10. Blake, D., Billington, K., Copestake, S. & Oakes, R. (2011) Genetic mapping identifies novel highly protective antigens for an apicomplexan parasite. *PLoS pathogens* 7(2).

11. Bonenfant, C., Dimier-Poisson, I., Velge-Roussel, F., Buzoni-Gatel, D., Del Giudice, G., Rappuoli, R. & Bout, D. (2001) Intranasal immunization with SAG1 and nontoxic mutant heat-labile enterotoxins protects mice against Toxoplasma gondii. *Infect Immun*, 69(3), 1605-12.
12. Boucher, L. E. & Bosch, J. (2015) The apicomplexan glideosome and adhesins - Structures and function. *J Struct Biol*, 190(2), 93-114.
13. Bousema, T., Okell, L., Felger, I. & Drakeley, C. (2014) Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat Rev Microbiol*, 12(12), 833-40.
14. Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J. & DeRisi, J. L. (2003) The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol*, 1(1), E5.
15. Brown, C. R. & McLeod, R. (1990) Class I MHC genes and CD8+ T cells determine cyst number in Toxoplasma gondii infection. *J Immunol*, 145(10), 3438-41.
16. Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, 2(11), 2728-33.
17. Bullen, H. E., Tonkin, C. J., O'Donnell, R. A., Tham, W. H., Papenfuss, A. T., Gould, S., Cowman, A. F., Crabb, B. S. & Gilson, P. R. (2009) A novel family of Apicomplexan glideosome-associated proteins with an inner membrane-anchoring role. *J Biol Chem*, 284(37), 25353-63.
18. Buscaglia, C. A., Coppens, I., Hol, W. G. & Nussenzweig, V. (2003) Sites of interaction between aldolase and thrombospondin-related anonymous protein in plasmodium. *Mol Biol Cell*, 14(12), 4947-57.
19. Buxton, D. & Innes, E. A. (1995) A commercial vaccine for ovine toxoplasmosis. *Parasitology*, 110 Suppl, S11-6.
20. Buxton, D., Thomson, K., Maley, S., Wright, S. & Bos, H. J. (1991) Vaccination of sheep with a live incomplete strain (S48) of Toxoplasma gondii and their immunity to challenge when pregnant. *Vet Rec*, 129(5), 89-93.
21. Cao, A., Liu, Y., Wang, J., Li, X., Wang, S., Zhao, Q., Cong, H., He, S. & Zhou, H. (2015) Toxoplasma gondii: Vaccination with a DNA vaccine encoding T- and B-cell epitopes of SAG1, GRA2, GRA7 and ROP16 elicits protection against acute toxoplasmosis in mice. *Vaccine*, 33(48), 6757-62.
22. Carruthers, V. & Boothroyd, J. C. (2007) Pulling together: an integrated model of Toxoplasma cell invasion. *Curr Opin Microbiol*, 10(1), 83-9.

23. Carruthers, V. B. (2002) Host cell invasion by the opportunistic pathogen Toxoplasma gondii. *Acta Trop*, 81(2), 111-22.
24. Carruthers, V. B. & Sibley, L. D. (1997) Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. *Eur J Cell Biol*, 73(2), 114-23.
25. Charron, A. J. & Sibley, L. D. (2004) Molecular partitioning during host cell penetration by Toxoplasma gondii. *Traffic*, 5(11), 855-67.
26. Clubb, R. T., Thanabal, V. & Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins. *Journal of Magnetic Resonance (1969)*, 97(1), 213-217.
27. Cody, V., Luft, J. R., Pangborn, W., Gangjee, A. & Queener, S. F. (2004) Structure determination of tetrahydroquinazoline antifolates in complex with human and Pneumocystis carinii dihydrofolate reductase: correlations between enzyme selectivity and stereochemistry. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 4), 646-55.
28. Crawford, J., Tonkin, M. L., Grujic, O. & Boulanger, M. J. (2010) Structural characterization of apical membrane antigen 1 (AMA1) from Toxoplasma gondii. *J Biol Chem*, 285(20), 15644-52.
29. Crutcher, J. M. & Hoffman, S. L. (1996) Malaria, in Baron, S. (ed), *Medical Microbiology*, 4 edition. Galveston.
30. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics*, 14(10), 892-3.
31. Cui, X., Lei, T., Yang, D., Hao, P., Li, B. & Liu, Q. (2012) *Toxoplasma gondii* immune mapped protein – 1 (TgIMP1) is a novel vaccine candidate against toxoplasmosis. *Vaccine*, 30(13), 2282-2287.
32. da Fonseca, P. C. & Morris, E. P. (2015) Cryo-EM reveals the conformation of a substrate analogue in the human 20S proteasome core. *Nat Commun*, 6, 7573.
33. de Azevedo, M. F., Gilson, P. R., Gabriel, H. B., Simões, R. F., Angrisano, F., Baum, J., Crabb, B. S. & Wunderlich, G. (2012) Systematic analysis of FKBP inducible degradation domain tagging strategies for the human malaria parasite Plasmodium falciparum. *PLoS One*, 7(7), e40981.
34. Debard, N., Buzoni-Gatel, D. & Bout, D. (1996) Intranasal immunization with SAG1 protein of Toxoplasma gondii in association with cholera toxin dramatically reduces development of cerebral cysts after oral infection. *Infect Immun*, 64(6), 2158-66.

35. Drenth, J. (2010) *Principles of protein X-ray crystallography*, 3rd edition. edition. New York, NY: New York, NY : Springer.
36. Dzierszinski, F., Popescu, O., Toursel, C., Slomianny, C., Yahiaoui, B. & Tomavo, S. (1999) The protozoan parasite Toxoplasma gondii expresses two functional plant-like glycolytic enzymes. Implications for evolutionary origin of apicomplexans. *J Biol Chem*, 274(35), 24888-95.
37. Ferguson, D. J. & Hutchison, W. M. (1987) An ultrastructural study of the early development and tissue cyst formation of Toxoplasma gondii in the brains of mice. *Parasitol Res*, 73(6), 483-91.
38. Fernando, A. M. (1990) Eimeria: infections of the intestine, *Coccidiosis of man and domestic animals*. Boca Raton: CRC Press, 63-75.
39. Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D. J., Yates, J. R. & Wu, Y. (2004) Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol Biochem Parasitol*, 135(1), 1-11.
40. Friedrich, N., Santos, J. M., Liu, Y., Palma, A. S., Leon, E., Saouros, S., Kiso, M., Blackman, M. J., Matthews, S., Feizi, T. & Soldati-Favre, D. (2010) Members of a novel protein family containing microneme adhesive repeat domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites. *J Biol Chem*, 285(3), 2064-76.
41. Garcia-Réguet, N., Lebrun, M., Fourmaux, M. N., Mercereau-Puijalon, O., Mann, T., Beckers, C. J., Samyn, B., Van Beeumen, J., Bout, D. & Dubremetz, J. F. (2000) The microneme protein MIC3 of Toxoplasma gondii is a secretory adhesin that binds to both the surface of the host cells and the surface of the parasite. *Cell Microbiol*, 2(4), 353-64.
42. Garnett, J. A., Liu, Y., Leon, E., Allman, S. A., Friedrich, N., Saouros, S., Curry, S., Soldati-Favre, D., Davis, B. G., Feizi, T. & Matthews, S. (2009) Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of Toxoplasma gondii. *Protein Sci*, 18(9), 1935-47.
43. Gazzinelli, R. T., Hakim, F. T., Hieny, S., Shearer, G. M. & Sher, A. (1991) Synergistic role of CD4+ and CD8+ T lymphocytes in IFN-gamma production and protective immunity induced by an attenuated Toxoplasma gondii vaccine. *J Immunol*, 146(1), 286-92.
44. Gerhold, R. W. (2014) *The Merck Veterinary Manual*, 2014. Available online: [Accessed.

45. Gruszczyk, J., Lim, N. T., Arnott, A., He, W. Q., Nguitragool, W., Roobsoong, W., Mok, Y. F., Murphy, J. M., Smith, K. R., Lee, S., Bahlo, M., Mueller, I., Barry, A. E. & Tham, W. H. (2016) Structurally conserved erythrocyte-binding domain in Plasmodium provides a versatile scaffold for alternate receptor engagement. *Proc Natl Acad Sci U S A*, 113(2), E191-200.
46. Grzesiek, S. & Bax, A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR*, 3(2), 185-204.
47. Gunther, H. (2013) *NMR Spectroscopy: Basic Principles, concepts and applications in chemistry*, 3 edition. John Wiley & Sons Ltd.
48. Harper, J. M., Huynh, M. H., Coppens, I., Parussini, F., Moreno, S. & Carruthers, V. B. (2006) A cleavable propeptide influences Toxoplasma infection by facilitating the trafficking and secretion of the TgMIC2-M2AP invasion complex. *Mol Biol Cell*, 17(10), 4551-63.
49. He, X. L., Grigg, M. E., Boothroyd, J. C. & Garcia, K. C. (2002) Structure of the immunodominant surface antigen from the Toxoplasma gondii SRS superfamily. *Nat Struct Biol*, 9(8), 606-11.
50. Henikoff, S. & Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*, 89(22), 10915-9.
51. HiszczyŃSka-Sawicka, E., Gatkowska, J. M., Grzybowski, M. M. & DŁUgoŃSka, H. (2014) Veterinary vaccines against toxoplasmosis. *Parasitology*, 141(11), 1365-1378.
52. Hoff, F. (2015) *How to Prepare Your Specimen for Immunofluorescence Microscopy*, 2015. Available online: [Accessed.
53. Holm, L. & Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res*, 38(Web Server issue), W545-9.
54. Huynh, M. H. & Carruthers, V. B. (2006) Toxoplasma MIC2 is a major determinant of invasion and virulence. *PLoS Pathog*, 2(8), e84.
55. Huynh, M. H., Liu, B., Henry, M., Liew, L., Matthews, S. J. & Carruthers, V. B. (2015) Structural Basis of Toxoplasma gondii MIC2-associated Protein Interaction with MIC2. *J Biol Chem*, 290(3), 1432-41.
56. Huynh, M. H., Rabenau, K. E., Harper, J. M., Beatty, W. L., Sibley, L. D. & Carruthers, V. B. (2003) Rapid invasion of host cells by Toxoplasma requires secretion of the MIC2-M2AP adhesive protein complex. *EMBO J*, 22(9), 2082-90.

57. Jackson, M. H. & Hutchison, W. M. (1989) The prevalence and source of Toxoplasma infection in the environment. *Adv Parasitol*, 28, 55-105.
58. Jung, C., Lee, C. Y. & Grigg, M. E. (2004) The SRS superfamily of Toxoplasma surface proteins. *Int J Parasitol*, 34(3), 285-96.
59. Jung, Y. S. & Zweckstetter, M. (2004) Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR*, 30(1), 11-23.
60. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 89(3), 496-514.
61. Keeler, J. (2010) *Understanding NMR Spectroscopy*, 2 edition. Cambridge, UK: John Wiley & Sons Ltd.
62. Kensil, C. R. (1996) Saponins as vaccine adjuvants. *Crit Rev Ther Drug Carrier Syst*, 13(1-2), 1-55.
63. Khan, I. A., Ely, K. H. & Kasper, L. H. (1991) A purified parasite antigen (p30) mediates CD8+ T cell immunity against fatal Toxoplasma gondii infection in mice. *J Immunol*, 147(10), 3501-6.
64. Lagal, V., Binder, E. M., Huynh, M. H., Kafsack, B. F., Harris, P. K., Diez, R., Chen, D., Cole, R. N., Carruthers, V. B. & Kim, K. (2010) Toxoplasma gondii protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites. *Cell Microbiol*, 12(12), 1792-808.
65. Lambros, C. & Vanderberg, J. P. (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J Parasitol*, 65(3), 418-20.
66. Lebrun, M., Michelin, A., El Hajj, H., Poncet, J., Bradley, P. J., Vial, H. & Dubremetz, J. F. (2005) The rhoptry neck protein RON4 re-localizes at the moving junction during Toxoplasma gondii invasion. *Cell Microbiol*, 7(12), 1823-33.
67. Leriche, M. A. & Dubremetz, J. F. (1990) Exocytosis of Toxoplasma gondii dense granules into the parasitophorous vacuole after host cell invasion. *Parasitol Res*, 76(7), 559-62.
68. Li, H., O'Donoghue, A. J., van der Linden, W. A., Xie, S. C., Yoo, E., Foe, I. T., Tilley, L., Craik, C. S., da Fonseca, P. C. & Bogyo, M. (2016) Structure- and function-based design of Plasmodium-selective proteasome inhibitors. *Nature*, 530(7589), 233-6.

69. Li, Z. Y., Chen, J., Petersen, E., Zhou, D. H., Huang, S. Y., Song, H. Q. & Zhu, X. Q. (2014) Synergy of mIL-21 and mIL-15 in enhancing DNA vaccine efficacy against acute and chronic Toxoplasma gondii infection in mice. *Vaccine*, 32(25), 3058-65.
70. Liesenfeld (2007) Toxoplasmosis, in L, G. & D, A. (eds), 23 edition. Philadelphia: Saunders Elsevier.
71. Lim, S. S. & Othman, R. Y. (2014) Recent advances in Toxoplasma gondii immunotherapeutics. *Korean J Parasitol*, 52(6), 581-93.
72. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, 19(2), 315-6.
73. Liu, M. M., Yuan, Z. G., Peng, G. H., Zhou, D. H., He, X. H., Yan, C., Yin, C. C., He, Y., Lin, R. Q., Song, H. Q. & Zhu, X. Q. (2010) Toxoplasma gondii microneme protein 8 (MIC8) is a potential vaccine candidate against toxoplasmosis. *Parasitol Res*, 106(5), 1079-84.
74. Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T. & DeRisi, J. L. (2006) Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res*, 34(4), 1166-73.
75. Louis, W. & Kami, K. (2007) ***Toxoplasma Gondii: The Model Apicomplexan. Perspectives and Methods***. London: Elsevier Ltd.
76. Lovett, J. L., Marchesini, N., Moreno, S. N. & Sibley, L. D. (2002) Toxoplasma gondii microneme secretion involves intracellular Ca(2+) release from inositol 1,4,5-triphosphate (IP(3))/ryanodine-sensitive stores. *J Biol Chem*, 277(29), 25870-6.
77. Luft, B. J. & Remington, J. S. (1992) Toxoplasmic encephalitis in AIDS. *Clin Infect Dis*, 15(2), 211-22.
78. Marchant, J., Cowper, B., Liu, Y., Lai, L., Pinzan, C., Marq, J. B., Friedrich, N., Sawmynaden, K., Liew, L., Chai, W., Childs, R. A., Saouros, S., Simpson, P., Roque Barreira, M. C., Feizi, T., Soldati-Favre, D. & Matthews, S. (2012) Galactose recognition by the apicomplexan parasite Toxoplasma gondii. *J Biol Chem*, 287(20), 16720-33.
79. Marchant, J., Sawmynaden, K., Saouros, S., Simpson, P. & Matthews, S. (2008) Complete resonance assignment of the first and second apple domains of MIC4 from Toxoplasma gondii, using a new NMRView-based assignment aid. *Biomol NMR Assign*, 2(2), 119-21.
80. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989a) Overcoming the overlap problem in the assignment of proton

NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta. *Biochemistry*, 28(15), 6150-6156.

81. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society*, 111(4), 1515-1517.
82. Meissner, M., Reiss, M., Viebig, N., Carruthers, V. B., Toursel, C., Tomavo, S., Ajioka, J. W. & Soldati, D. (2002) A family of transmembrane microneme proteins of Toxoplasma gondii contain EGF-like domains and function as escorters. *J Cell Sci*, 115(Pt 3), 563-74.
83. Mondragon, R. & Frixione, E. (1996) Ca(2+)-dependence of conoid extrusion in Toxoplasma gondii tachyzoites. *J Eukaryot Microbiol*, 43(2), 120-7.
84. Mordue, D. G., Desai, N., Dustin, M. & Sibley, L. D. (1999) Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring. *J Exp Med*, 190(12), 1783-92.
85. Morrissette, N. S. & Sibley, L. D. (2002) Cytoskeleton of apicomplexan parasites. *Microbiol Mol Biol Rev*, 66(1), 21-38; table of contents.
86. Muhandiram, D. R. & Kay, L. E. (1994) Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B*, 103(3), 203-216.
87. Ménard, R. (2013) *Malaria : methods and protocols*, 2nd edition. [Totowa, N.J.]: Humana.
88. Nilges, M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol*, 245(5), 645-60.
89. Nilges, M., Macias, M. J., O'Donoghue, S. I. & Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol*, 269(3), 408-22.
90. Nolan, S. J., Romano, J. D., Luechtefeld, T. & Coppens, I. (2015) Neospora caninum recruits host cell structures to its parasitophorous vacuole and salvages lipids from organelles. *Eukaryot Cell*.

91. Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A 4D HCCH-TOCSY experiment for assigning the side chain1H and13C resonances of proteins. *Journal of Biomolecular NMR*, 2(6), 655-659.
92. Payne, R. O., Milne, K. H., Elias, S. C., Edwards, N. J., Douglas, A. D., Brown, R. E., Silk, S. E., Biswas, S., Miura, K., Roberts, R., Rampling, T. W., Venkatraman, N., Hodgson, S. H., Labbé, G. M., Halstead, F. D., Poulton, I. D., Nugent, F. L., de Graaf, H., Sukhtankar, P., Williams, N. C., Ockenhouse, C. F., Kathcart, A. K., Qabar, A. N., Waters, N. C., Soisson, L. A., Birkett, A. J., Cooke, G. S., Faust, S. N., Woods, C., Ivinson, K., McCarthy, J. S., Diggs, C. L., Vekemans, J., Long, C. A., Hill, A. V., Lawrie, A. M., Dutta, S. & Draper, S. J. (2016) Demonstration of the Blood-Stage Plasmodium falciparum Controlled Human Malaria Infection Model to Assess Efficacy of the P. falciparum Apical Membrane Antigen 1 Vaccine, FMP2.1/AS01. *J Infect Dis*, 213(11), 1743-51.
93. Petri, W. A., Markell, E. K. & Voge, M. (2006) *Markell and Voge's Medical parasitology*, 9th edition. Philadelphia, Pa.: Elsevier Saunders.
94. Pintilie, G. D. (2009) *Segmentation and Registration of Molecular Components in 3-Dimensional Density Maps from Cryo Electron Microscopy*. Doctor of Philosophy MASSACHUSETTS INSTITUTE OF TECHNOLOGY
95. Pinzan, C. F., Sardinha-Silva, A., Almeida, F., Lai, L., Lopes, C. D., Lourenço, E. V., Panunto-Castelo, A., Matthews, S. & Roque-Barreira, M. C. (2015) Vaccination with Recombinant Microneme Proteins Confers Protection against Experimental Toxoplasmosis in Mice. *PLoS One*, 10(11), e0143087.
96. Reiss, M., Viebig, N., Brecht, S., Fourmaux, M. N., Soete, M., Di Cristina, M., Dubremetz, J. F. & Soldati, D. (2001) Identification and characterization of an escorter for two secretory adhesins in Toxoplasma gondii. *J Cell Biol*, 152(3), 563-78.
97. Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E. & Nilges, M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3), 381-2.
98. Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, 42(Web Server issue), W320-4.
99. Roos, D. S., Crawford, M. J., Donald, R. G., Fohl, L. M., Hager, K. M., Kissinger, J. C., Reynolds, M. G., Striepen, B. & Sullivan, W. J. (1999) Transport and trafficking: Toxoplasma as a model for Plasmodium. *Novartis Found Symp*, 226, 176-95; discussion 195-8.

100. Sagara, I., Dicko, A., Ellis, R. D., Fay, M. P., Diawara, S. I., Assadou, M. H., Sissoko, M. S., Kone, M., Diallo, A. I., Saye, R., Guindo, M. A., Kante, O., Niambele, M. B., Miura, K., Mullen, G. E., Pierce, M., Martin, L. B., Dolo, A., Diallo, D. A., Doumbo, O. K., Miller, L. H. & Saul, A. (2009) A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine*, 27(23), 3090-8.

101. Saouros, S., Dou, Z., Henry, M., Marchant, J., Carruthers, V. B. & Matthews, S. (2012) Microneme protein 5 regulates the activity of Toxoplasma subtilisin 1 by mimicking a subtilisin prodomain. *J Biol Chem*, 287(43), 36029-40.

102. Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T. J., Jäkle, U., Ramboarina, S., Shivarattan, T., Matthews, S. & Soldati-Favre, D. (2005) A novel galectin-like domain from Toxoplasma gondii micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex. *J Biol Chem*, 280(46), 38583-91.

103. Sawmynaden, K., Saouros, S., Friedrich, N., Marchant, J., Simpson, P., Bleijlevens, B., Blackman, M. J., Soldati-Favre, D. & Matthews, S. (2008) Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition. *EMBO Rep*, 9(11), 1149-55.

104. Shen, B., Buguliskis, J. S., Lee, T. D. & Sibley, L. D. (2014) Functional analysis of rhomboid proteases during Toxoplasma invasion. *MBio*, 5(5), e01795-14.

105. Shen, B. & Sibley, L. D. (2014) Toxoplasma aldolase is required for metabolism but dispensable for host-cell invasion. *Proc Natl Acad Sci U S A*, 111(9), 3567-72.

106. Soldati, D., Dubremetz, J. F. & Lebrun, M. (2001) Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite Toxoplasma gondii. *Int J Parasitol*, 31(12), 1293-302.

107. Straub, K. W., Peng, E. D., Hajagos, B. E., Tyler, J. S. & Bradley, P. J. (2011) The moving junction protein RON8 facilitates firm attachment and host cell invasion in Toxoplasma gondii. *PLoS Pathog*, 7(3), e1002007.

108. Suzuki, Y., Israelski, D. M., Dannemann, B. R., Stepick-Biek, P., Thulliez, P. & Remington, J. S. (1988) Diagnosis of toxoplasmic encephalitis in patients with acquired immunodeficiency syndrome by using a new serologic method. *J Clin Microbiol*, 26(12), 2541-3.

109. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods*, 100, 3-15.

110. Tomas, J. (2010) *A structural and functional study of the periplasmic loop P2 of MalF in the maltose transporter of Escherichia coli*. PhD Humboldt-Universität zu Berlin.

111. Tonkin, M. L., Roques, M., Lamarque, M. H., Pugnière, M., Douguet, D., Crawford, J., Lebrun, M. & Boulanger, M. J. (2011) Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. *Science*, 333(6041), 463-7.

112. Tyler, J. S. & Boothroyd, J. C. (2011) The C-terminus of Toxoplasma RON2 provides the crucial link between AMA1 and the host-associated invasion complex. *PLoS Pathog*, 7(2), e1001282.

113. Voges, D., Zwickl, P. & Baumeister, W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem*, 68, 1015-68.

114. Wang, Y., Fang, R., Yuan, Y., Hu, M., Zhou, Y. & Zhao, J. (2014) Identification of host proteins interacting with the integrin-like A domain of Toxoplasma gondii micronemal protein MIC2 by yeast-two-hybrid screening. *Parasit Vectors*, 7(1), 543.

115. WHO (2015) *Fact Sheet: World Malaria Report 2015*, 2015. Available online: [Accessed.

116. Williams, M. J., Alonso, H., Enciso, M., Egarter, S., Sheiner, L., Meissner, M., Striepen, B., Smith, B. J. & Tonkin, C. J. (2015) Two Essential Light Chains Regulate the MyoA Lever Arm To Promote Toxoplasma Gliding Motility. *MBio*, 6(5), e00845-15.

117. Woodison, G. & Smith, J. E. (1990) Identification of the dominant cyst antigens of Toxoplasma gondii. *Parasitology*, 100 Pt 3, 389-92.

118. Yin, G., Lin, Q., Wei, W., Qin, M., Liu, X., Suo, X. & Huang, Z. (2014) Protective immunity against Eimeria tenella infection in chickens induced by immunization with a recombinant C-terminal derivative of EtIMP1. *Vet Immunol Immunopathol*, 162(3-4), 117-21.

119. Yin, G., Qin, M., Liu, X., Suo, J., Tang, X., Tao, G., Han, Q., Suo, X. & Wu, W. (2013) An Eimeria vaccine candidate based on Eimeria tenella immune mapped

protein 1 and the TLR-5 agonist Salmonella typhimurium FliC flagellin. *Biochem Biophys Res Commun*, 440(3), 437-42.

120. Zhang, G., Huong, V. T., Battur, B., Zhou, J., Zhang, H., Liao, M., Kawase, O., Lee, E. G., Dautu, G., Igarashi, M., Nishikawa, Y. & Xuan, X. (2007) A heterologous prime-boost vaccination regime using DNA and a vaccinia virus, both expressing GRA4, induced protective immunity against Toxoplasma gondii infection in mice. *Parasitology*, 134(Pt 10), 1339-46.

121. Zhang, T. E., Yin, L. T., Li, R. H., Wang, H. L., Meng, X. L. & Yin, G. R. (2015) Protective immunity induced by peptides of AMA1, RON2 and RON4 containing T-and B-cell epitopes via an intranasal route against toxoplasmosis in mice. *Parasit Vectors*, 8, 15.

122. Zuiderweg, E. R. P. & Fesik, S. W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry*, 28(6), 2387-2391.

Appendix I

Protein	Primer names	Sequence	Tm (°C)	PCR (°C)
TgIMP1 clone 1 (full length)	pET46Ek/LIC forward	GAC GAC GAC AAGATGGGAACCGTTTGCACG AAG	67	63
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA CTACGCATTTGCTCTGTCCACC AT	68	
TgIMP1 clone 2 (1 to 350)	pET46Ek/LIC forward	GAC GAC GAC AAGATGGGAACCGTTTGCACG AAG	67	64
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA GCCTGCTGCCACTGCTAAGT	69	
TgIMP1 clone 3 (91 to 350)	pET46Ek/LIC forward	GAC GAC GAC AAG ATG GAACCGGCTGCGGTAATTCG	67	64
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA GCCTGCTGCCACTGCTAAGT	69	
TgIMP1 clone 4 (91 to 400)	pET46Ek/LIC forward	GAC GAC GAC AAG ATG GAACCGGCTGCGGTAATTCG	67	63
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA CTACGCATTTGCTCTGTCCACC AT	68	
TgIMP1 clone 5 (138 to 350)	pET46Ek/LIC forward	GAC GAC GAC AAG ATG GCTGAGGACCTGGAACTGCTTC G	67	64
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA GCCTGCTGCCACTGCTAAGT	69	
TgIMP1 clone 6 (138 to 400)	pET46Ek/LIC forward	GAC GAC GAC AAG ATG GCTGAGGACCTGGAACTGCTTC G	67	63
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA CTACGCATTTGCTCTGTCCACC AT	68	
TgIMP1 clone 7 (176 to 400)	pET46Ek/LIC forward	GAC GAC GAC AAGATGAGTGCGTCGGCTTAC GG	66	63
	pET46Ek/LIC reverse	GA GGA GAA GCC CGG TTA CTACGCATTTGCTCTGTCCACC AT	68	
EtIMP1 clone 2 (152 to 367)	pET30Xa/LIC forward	GGTATTGAGGGTCGCATG AGCAACGCGAACCTGC	64	59
	pET30Xa/LIC reverse	AGA GGA GAG TTA GAG CCT TTAATCCAAAGTAATTCCATTT CCTTGTTGAATTTC	64	
PfIMP2 (codon optimised)	Xa/LIC forward	GGT ATT GAG GGT CGC ATGAGCGAAGAAAAAGGT	56	59
	Xa/LIC reverse	AGA GGA GAG TTA GAG CCT	59	

		TTATTTCGATTACGAACGTGG		
PfIMP2 (codon optimised)	pNIC-ZB forward	TACTTCCAATCCATGAGCGAAG AAAAAGGT	68	63
	pNIC-ZB reverse	TATCCACCTTTACTGTCAAAAG CTAATGCT	66	
PfIMP2 (full length)	pRM2_DD-3HA-GFP forward	ATATATCTCGAGATGTCAGAA GAAAAG	60	63
	pRM2_DD-3HA-GFP reverse	ATTATTGGTACCAAAGGATATA CTAGC	60	

Appendix Table 1.1: Table showing the primer sequences for various constructs designed during the course of the project, the vector that the gene was cloned into, the Tm of the primers and the annealing temperature used for PCR amplification (PCR).

Vector	**Resistance**	**Tags**	**Homologue**
pET46 Ek/LIC	Carbenicillin	N-terminal His tag	TgIMP1
pET32 Xa/LIC	Kanamycin	N-terminal His, S and thioredoxin tags	PfIMP2
pET30 Xa/LIC	Kanamycin	N-terminal His and S tags	EtIMP1 & PfIMP2
pNIC-ZB	Kanamycin	N-terminal His tag	PfIMP2
pRM2_DD-3HA-GFP	WR99210	C-terminal triple HA and GFP tags	PfIMP2

Appendix Table 1.2: Table showing the resistance gene and tags present in the various vectors used in the project and the gene sequences that were cloned into them.

Appendix 1.3: Paper describing the backbone assignment of PfIMP2 (referred to as PfIMP1 throughout the manuscript). Reference: Benjamin, S., Williams, F., Kerry, L. & Matthews, S. (2015) NMR assignment of the immune mapped protein 1 (IMP1) homologue from Plasmodium falciparum. Biomol NMR Assign, 9(2), 393-5.

NMR assignment of the Immune Mapped Protein 1 homologue (IMP1) in *Plasmodium falciparum*

Stefi Benjamin, Felix Williams, Louise Kerry and Steve Matthews.

Abbreviations

IMP1	Immune Mapped Protein 1
IPTG	Isopropyl -D-thiogalactopyranoside
LIC	Ligation independent cloning
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect enhancement spectroscopy
TOCSY	Total correlation spectroscopy

Abstract

Plasmodium falciparum is responsible for causing cerebral malaria in humans. IMP1 is an immunogenic protein, present in the parasite, which has been shown to induce an immune response against apicomplexan parasites in a species-specific manner. Here, we report the complete NMR assignments of PfIMP1.

Key word

IMP1, antigenic protein, *Plasmodium falciparum*, NMR

Background

Immune mapped protein 1 (IMP1) is an antigenic protein that is conserved across most apicomplexan parasites (Blake et al, 2011). It was first identified in *Eimeria maxima* and was found to raise protective immune response in hosts against the parasite (Blake et al, 2011). Cui X et al (2012) have shown that mice immunised with a DNA vaccine of *Toxoplasma gondii* IMP1 (TgIMP1) have a prolonged lifespan in comparison to control mice upon infection with *T. gondii*. Studies by Yin G et al (2013) have also shown that a chimeric subunit vaccine developed using truncated IMP1 from *Eimeria tenella* (EtIMP1) and a molecular adjuvant raises protective immunity against *E. tenella* infection in chickens. These studies show that IMP1 raises an antigenic response in hosts against parasites in a species specific manner.

The localisation of IMP1 within apicomplexan parasites is ambiguous. Cui X et al (2012) have shown that TgIMP1 is localised to the parasite membrane. Yin G et al (2013) suggest that EtIMP1 could be a membrane protein based on localisation studies and bioinformatics analysis. Although the function and localisation of IMP1 within the parasite is unknown, the antigenic response raised by the protein in hosts makes it an interesting candidate for structural studies. Hence, we perform an NMR analysis on an IMP1-like protein from *P. falciparum* (PfIMP1) to understand the three dimensional structure of the protein in solution. Here, we present the complete NMR assignments for IMP1 for PfIMP1.

Materials and Methods

PfIMP1 sequence (Accession number: XP_001349200) was synthesised following codon optimisation for recombinant expression in *E. coli.* The codon optimised PfIMP1 sequence was then amplified using primer set PfIMP1-NTH-F (5' -

TACTTCCAATCCATGAGCGAAGAAAAAGGT - 3') and PfIMP1-NTH-R (5' – TATCCACCTTTACTGTCAAAAGCTAATGCT - 3'). The PCR product was purified and cloned into pET28-NTH vector by LIC cloning. The resulting plasmid pET28-NTH-PfIMP1 was transformed into Rosetta2 cells (NEB, UK). The recombinant strains were grown in M9 minimal media containing 0.07% $^{15}NH_4Cl$ and 0.2% $^{13}C_6$-glucose (Sigma) for $^{15}N/^{13}C$-labeling of PfIMP1. Once OD_{600} = 0.6, expression was induced using 1mM IPTG. Following overnight incubation at 18°C, the cells were harvested by centrifugation at 4500 rpm for 15 minutes. The cells were lysed by sonication and the lysate was clarified by centrifugation at 16000 rpm for 40 minutes. The clarified lysate was purified by affinity chromatography using pre-packed Nickel column, 5 ml Histrap FF (GE Healthcare, UK) in the AKTAprime (GE Healthcare, UK). The purified protein was cleaved with TEV protease to remove the N-terminal His tag and further purified by Gel filtration using a Superdex 75 Hiload 16/600 column (GE Healthcare, UK) with AKTAprime (GE Healthcare, UK). The purified protein was then concentrated and dialysed into 50 mM Hepes, 250 mM NaCl and 10mM DTT, pH7.5. The final construct that was used to record NMR experiments contained a serine residue left from the TEV cleavage, followed by the complete amino acid sequence of PfIMP1. The molecular weight of this protein is 17.89 kDa. There are 159 residues in this construct but the methionine following the first serine residue is labelled as residue 1 throughout this article.

NMR spectra were recorded at 303K on Bruker DRX600 and DRX800 spectrometers equipped with cryo-probes. The Chemical shifts of ^{1}HN, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$ and ^{13}CO cross peaks were assigned using CBCA(CO)NH, HNCACB, HNCO and HN(CA)CO. The aliphatic and aromatic side chain ^{1}H and ^{13}C assignments were obtained by using HBHA(CBCACO)NH, HCCH-TOCSY, (H)CC(CO)NH-TOCSY and ^{1}H-C^{13}NOESY-HMQC spectra.

NMR Assignments for PfIMP1

Backbone assignment of PfIMP1 was initially performed semi-automatically using MARS (Jung and Zweckstetter, 2004) then subsequently confirmed and completed manually. Due to the high number of lysines (13%) present in the PfIMP1 sequence, it was challenging to assign them as these side chain resonances were heavily overlapped. In total, ~96 % of all possible backbone atoms were assigned. Figure 1 shows the assigned ^{15}N, ^{1}H – HSQC spectrum. ~94% of the amino acid side chain atoms have also been assigned. The secondary structure prediction (Figure 2) was obtained from Talos N (Shen and Bax, 2013). The amide proton of I96 resonates at an unusual upfield-shifted position compared to the other structured amides. It is located in a loop region between secondary structure elements and is immediately preceded by an FP sequence. The proline residue within this loop likely configures the backbone conformation such that amide proton of I96 experiences a significant shielding ring current effect from F94.

References

BLAKE, D., BILLINGTON, K., COPESTAKE, S. & OAKES, R. 2011. Genetic mapping identifies novel highly protective antigens for an apicomplexan parasite. *PLoS pathogens* 7.

CUI, X., LEI, T., YANG, D., HAO, P., LI, B. & LIU, Q. 2012. *Toxoplasma gondii* immune mapped protein – 1 (TgIMP1) is a novel vaccine candidate against toxoplasmosis. *Vaccine,* 30, 2282-2287.

JUNG, Y. S. & ZWECKSTETTER, M. 2004. Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR,* 30, 11-23.

YIN, G., QIN, M., LIU, X., SUO, J., TANG, X., TAO, G., HAN, Q., SUO, X. & WU, W. 2013. An Eimeria vaccine candidate based on Eimeria tenella immune mapped protein 1 and the TLR-5 agonist Salmonella typhimurium FliC flagellin. *Biochem Biophys Res Commun,* 440, 437-42.

SHEN, Y. & BAX, A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR, 56, 227-41.

Figure Captions

Fig.1 Assigned ^{15}N, ^{1}H – HSQC spectrum of recombinant PfIMP1. The peaks are labelled with single letter amino acid code followed by their position in the recombinant PfIMP1 sequence. The underlined residues indicate aliased amide signals (namely G82, G51 and G149).

Fig. 2 Secondary structure determination using TALOS-N (Shen and Bax, 2013). The predicted secondary structure of this protein consists of four α-helices and eleven β strands. Helices are located between residues 43-49, 58-80, 88-92 and 142-151. β strands are located between residues 5-12, 17-24, 32-36, 53-57, 83-86, 96-104, 107-112, 115-118, 124-128, 139-140 and 154-157.

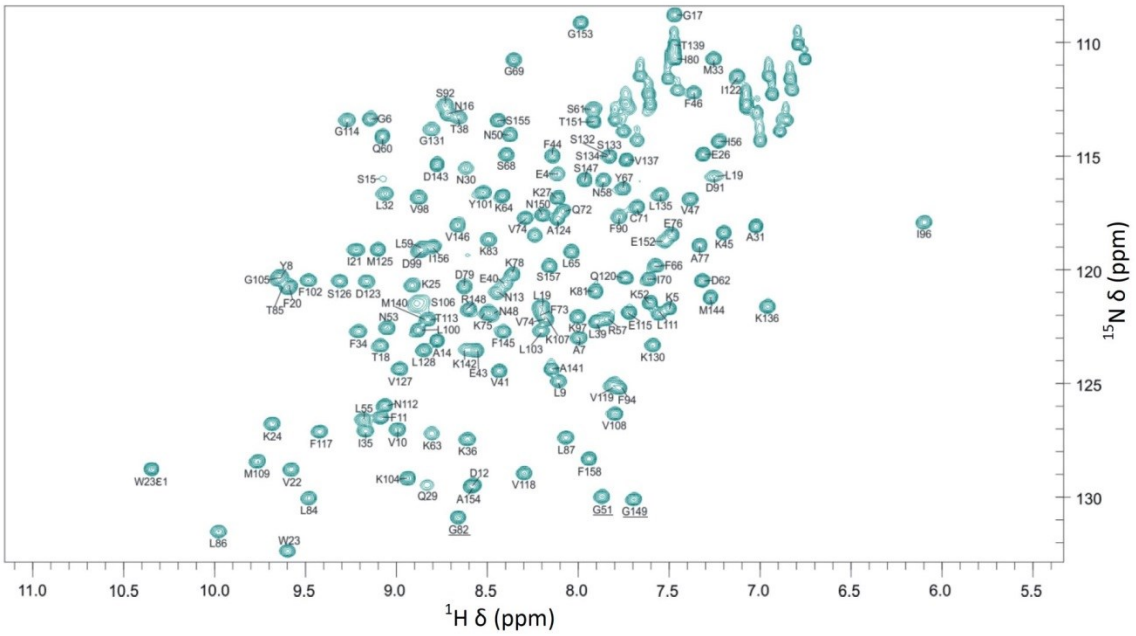


Figure 1

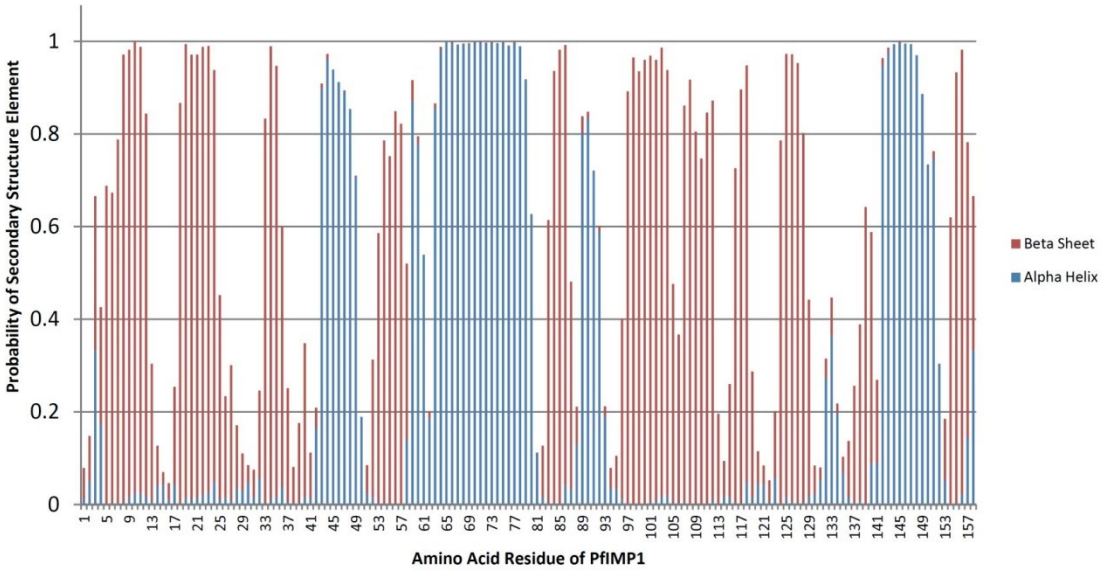


Figure 2

Appendix 1.4: Unpublished manuscript of the paper describing the structure of PfIMP2 and various functional studies performed for TgIMP1. (Yonggen et al, 2016)

BBA - Proteins and Proteomics xxx (2016) xxx-xxx

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com

Toxoplasma gondii immune mapped protein 1 is anchored to the inner leaflet of the plasma membrane and adopts a novel protein fold

Yonggen Jia [a, 1], Stefi Benjamin [b, 1], Qun Liu [c], Yingqi Xu [b], Sunil Kumar Dogga [a], Jing Liu [c], Stephen Matthews [b, *], Dominique Soldati-Favre [a, *]

[a] Department of Microbiology and Molecular Medicine, CMU, University of Geneva, Rue Michel-Servet 1, CH-1211 Geneva, Switzerland
[b] Department of Life Sciences, Imperial College London, South Kensington, London SW7 2AZ, UK
[c] Key Laboratory of Animal Epidemiology and Zoonosis, Ministry of Agriculture, College of Veterinary Medicine, China Agricultural University, Beijing 100193, China

ARTICLE INFO

Article history:
Received 23 July 2016
Received in revised form 14 November 2016
Accepted 21 November 2016
Available online xxx

Keywords:
Toxoplasma gondii
Acylation
Invasion
Vaccine
NMR
Protein structure

ABSTRACT

The immune mapped protein 1 (IMP1) was first identified as a protective antigen in *Eimeria maxima* and described as vaccine candidate and invasion factor in *Toxoplasma gondii*. We show here that TgIMP1 localizes to the inner leaflet of plasma membrane (PM) via dual acylation. Mutations either in the N-terminal myristoylation or palmitoylation sites (G2 and C5) cause relocalization of TgIMP1 to the cytosol. The first 11 amino acids are sufficient for PM targeting and the presence of lysine (K7) is critical. Disruption of *TgIMP1* gene by double homologous recombination revealed no invasion defect or any measurable alteration in the lytic cycle of tachyzoites. Following immunization with TgIMP1 DNA vaccine, mice challenged with either wild type or IMP1-ko parasites showed no significant difference in protection. The sequence analysis identified a structured C-terminal domain that is present in a broader family of IMP1-like proteins conserved across the members of Apicomplexa. We present the solution structure of this domain determined from NMR data and describe a new protein fold not seen before.

© 2016 Published by Elsevier Ltd.

1. Introduction

The phylum of Apicomplexa comprises medical and veterinary important protozoan parasites including *Plasmodium*, *Toxoplasma*, *Eimeria*, *Neospora*, *Sarcocystis*, *Babesia*, *Theileria* and *Cryptosporidium* species. The limited repertoire of efficacious drugs and the emergence of drug resistance have considerably hampered the control of these parasites and boosted the research towards the development of vaccines [1]. In this context, an immune mapped protein 1 (IMP1) was identified as an antigen eliciting protective immunity against chicken coccidian *E. maxima*, by combining parasite genetics and selective barriers with population-based genetic fingerprinting [2]. More recent studies in *Toxoplasma gondii* and *Neospora caninum* reported that TgIMP1 and NcIMP1 provided protection following immunization with DNA vaccine in the mouse model [3,4]. Antibodies raised against NcIMP1 localized it to the plasma membrane (PM) and were shown to have inhibitory effects on host cell invasion [4].

IMP1 lacks conserved domains with known function except for the predicted myristoylation and palmitoylation sites at the N-terminus, which are presumed to confer membrane association to the protein. Protein myristoylation is a process of co-translational attachment of myristic acid through an amide bond (C14:0) to N-terminal glycine residues and palmitoylation is the reversible posttranslational attachment of palmitate (C16:0) via a thioester linkage to cysteine. The lipid acylation of proteins, especially those implicated in invasion or as structural components of inner membrane complexes, contribute to pathogenesis in *T. gondii* and *Plasmodium* species [5–7].

In this study, we show that the acylation sites within the N-terminus of TgIMP1 are involved in anchoring the protein to the cytoplasmic face of the parasite plasma membrane (PM). Given this internal localization, we have revisited the reported roles of IMP1 in invasion and as a vaccine candidate by generating a mutant parasite lacking *TgIMP1*. Furthermore, we have identified a conserved globular domain at the C-terminus of TgIMP2.1, which represents a broader protein family present in all Apicomplexa. We also elucidate its high-resolution structure by solution state NMR and reveal a new protein fold.

2. Results

2.1. IMP1 possesses a globular C-terminal domain conserved across the Apicomplexa

Bioinformatic analysis indicate that IMP1 is restricted to the members of the coccidian-subgroup of the Apicomplexa. It also reveals that IMP1 belongs to a broader family of proteins, including closely related IMP1-like proteins that group into six distinct phylogenetic clusters (IMP1 to IMP6 Fig. 1). The protein sequence alignment of all IMP1 and IMP1-like proteins used for the phylogeny is

* Corresponding authors.
Email addresses: s.j.matthews@imperial.ac.uk (S. Matthews); dominique.soldati-Favre@unige.ch (D. Soldati-Favre)
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.bbapap.2016.11.010
1570-9639/© 2016 Published by Elsevier Ltd.

UNCORRECTED PROOF

presented in Supplemental Fig. S1. TgIMP1 and other members of the IMP1 cluster possess putative N-terminal myristoylation and palmitoylation sites followed by an extended linker region and a conserved C-terminal domain (termed the IMP1-like domain). In contrast, members in other groups lack the residues predicted for N-terminal acylation. While the proteins clustering in IMP1, IMP5 and IMP6 are found restricted to the coccidians, the members of IMP4 are present in all Apicomplexa except *Cryptosporidium species*. The conservation of this globular domain across the phylum defines this broader family of IMP1-like proteins that may share conserved or similar functions.

2.2. IMP1 is localized to the inner leaflet of the plasma membrane in Toxoplasma gondii

In eukaryotic cells, protein post-translational modification by dual acylation such as myristoylation and palmitoylation, confers membrane association for those proteins lacking a transmembrane domain. The conserved glycine followed by cysteine residues (MGXXCS/T) at the N-terminus of IMP1 implies that this protein is membrane-bound through such dual acylation. To determine the subcellular distribution of TgIMP1 in tachyzoites, a second copy coding for the C-terminally epitope-tagged TgIMP1 was expressed stably under the control of a tubulin promoter (TubTgIMP1-Ty). Additionally, to detect the expression and localization of endogenous TgIMP1, a parasite line was generated in which a triple Ty-tag was inserted at the C-terminus of the protein via single homologous recombination at the endogenous *TgIMP1* locus in the RH ku80-ko strain. Expression of endogenous TgIMP1 or TgIMP1-3Ty-ki (knock-in) and second copy TubTgIMP1-Ty were assessed by western blot analysis using either rabbit polyclonal anti-IMP1 [3] or mouse monoclonal anti-Ty antibodies (Fig. 2A). On SDS-PAGE, TgIMP1, TgIMP1-Ty and endogenous TgIMP1-3Ty migrate considerably slower than the predicted molecular mass (43.49 kDa, 44.9 kDa and 47.8 kDa, respectively), possibly due to the post-modifications and/or to the abundance of negatively charged amino acids [8]. Indirect immunofluorescence assay (IFA) showed that TgIMP1-Ty predominantly localized to the parasite pellicle (Fig. 2B). Despite a low level of endogenous expression, TgIMP1 and TgIMP1-3Ty also localized to the pellicle (Fig. 2B middle and lower panel, respectively).

The group Apicomplexa belongs to the clade of alveolates that possess a pellicle composed of a set of flattened cisternae forming the inner membrane complex (IMC) underneath the plasma membrane (PM) [9,10]. Given that TgIMP1 was not detected in the nascent IMC of growing daughter cells during parasite division, the protein is presumably anchored into the PM [11]. To confirm that TgIMP1 is associated to the PM rather than the IMC, fresh egressed parasites were treated with aerolysin. This toxin damages the membrane, leading on osmotic shock which separates the PM from the underlying IMC, enabling protein localization within these closely apposed membrane systems [12,13]. Following aerolysin treatment, TgIMP1-Ty separated from the gliding associated protein 45 (GAP45) and co-localized with the GPI-anchored surface antigen 1 (SAG1), two markers that are used for IMC and PM, respectively (Fig. 2C). A sequential extraction procedure with PBS, TX-100/PBS and TX100/TBS was performed to assess the nature of TgIMP1 association to membrane (Fig. 2D). As anticipated for an acylated protein, TgIMP1 was fully soluble in the presence of TX100/PBS. TgGAP45, a component of the glideosome, acylated between the IMC and PM, showed a slightly more complex fractionation profile due to its interaction with MLC1, the myosin light chain of TgMyoA [12].

To determine if TgIMP1 is facing the extracellular or cytosolic face of the PM, parasites freshly egressed from host cells were incubated with proteinase K (PK) with or without permeabilization with digitonin (Dig.). On intact parasites (in absence of Dig.) PK can only access and digest polypeptides exposed to the surface, such as TgSAG1. Under these conditions TgIMP1 was fully protected, confirming that this protein is anchored to inner surface of the PM (Fig. 2E).

2.3. N-terminal acylation is required for plasma membrane localization of TgIMP1

All the coccidian IMP1 homologues harbor putative myristoylation and palmitoylation motifs at their N-termini (Fig. 3A). According to the CSS-Palm tool (version 4.0) an S-palmitoylation site is predicted on cysteine 5 (C5), with a score of 38.63 (above the high threshold cutoff score of 4.22). To assess the importance of acylation for TgIMP1 localization to the pellicle, both the glycine (G2) and C5 were mutated to alanine residues and expressed as second copy (TubTgIMP1(GC-AA)-Ty) in the parasites. Western blot analysis indicated that the apparent molecular mass of TgIMP1(GC-AA)-Ty was slightly smaller than the wild type protein' which is likely due to the absence of post-translational lipid modification (Fig. 3B). TgIMP1(GC-AA)-Ty was cytoplasmic and thus confirmed that the dual acylation is critical for TgIMP1 localization to the PM (Fig. 3C, upper panel).

2.4. The first eleven amino acids of TgIMP1 are sufficient for plasma membrane targeting

An alignment of the N-terminus amino acid sequences of the IMP1 homologues highlighted a conserved lysine residue at position seven (K7) in addition to the acylated residues G2 and C5 (Fig. 3A). To delineate the sequence determinant conferring localization of TgIMP1, a C-terminal mCherry-Ty fusion to the first 11 amino acids of TgIMP1 was shown to be sufficient to target the chimeric protein to the PM (Fig. 3C, middle panel). Since mCherry-Ty lacks cysteine residues, no additional palmitoylation could participate in its PM localization. Subsequently, individual mutation of C5 and K7 residues was shown to perturb PM targeting. The two mutants are excluded from the PM and nucleus, but appeared enriched in intracellular membranes possibly endoplasmic reticulum (Fig. 3C). While C5A mutation abrogates palmitoylation, myristoylation at the G2 still occurred and conferred enhanced affinity to membranes. The membrane association of the two mutants (C5A and K7A) was further assessed by fractionation. The K7A mutant was entirely soluble in the presence of 1% Triton X-100, suggesting that the conserved K7 residue plays an important role in the N-terminus myristoylation membrane binding of TgIMP1 (Fig. 3D). Collectively, these results indicate that the first 11 amino acids of TgIMP1 are sufficient to confer PM targeting via acylation.

2.5. The solution structure of the globular IMP1-like domain reveals a novel fold

The N-terminal myristoylation and palmitoylation sites identified in this study as crucial for the PM localization of IMP1 are absent in IMP2. Furthermore, IMP1 possesses an extended N-terminus that is predicted to a flexible linker connecting to a conserved C-terminus. This C-terminal region is present across both IMP1 and IMP2 families suggesting that it contains a globular domain that may contribute to a shared function. We termed this region the 'IMP1-like domain' and to explore it further we expressed recombinant TgIMP1, subjected it to proteolytic digestion and analyzed the products by SDS-PAGE (Fig. 4A). A stable C-terminal fragment of approxi-

UNCORRECTED PROOF

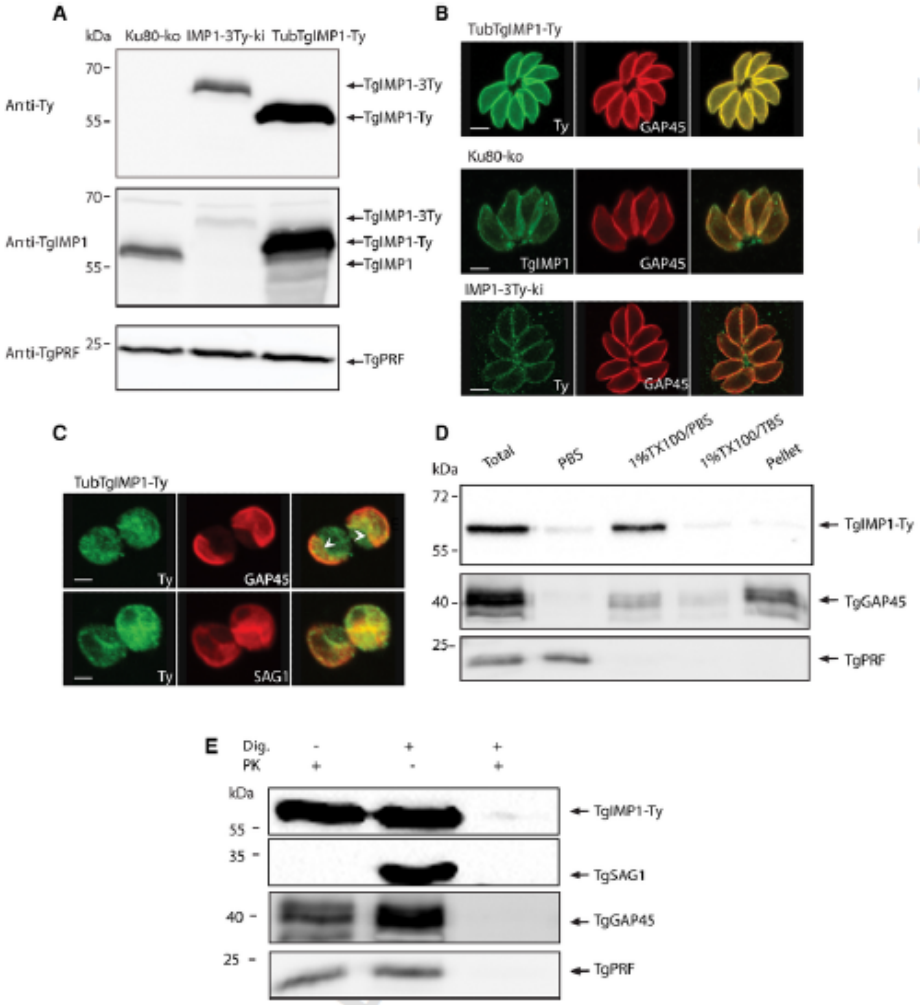


Fig. 2. Localization of TgIMP1 in tachyzoites. (A) Western blotting analysis of Ku80-ko, IMP1-3Ty-ki and TubTgIMP1-Ty using anti-Ty or anti-TgIMP1, anti-profilin (PRF) was used as loading control. (B) Immunofluorescence microscopic analysis of parasites expressing a second copy of TgIMP1-Ty using anti-Ty and the endogenous TgIMP1 using anti-TgIMP1, respectively (upper and middle panel). The pellicle localization of TgIMP1 was confirmed in the stain of IMP1-3Ty-ki, using anti-Ty (lower panel). (C) Extracellular tachyzoites expressing IMP1-Ty were incubated 30 min with aerolysin. After aerolysin treatment, the plasma membrane dramatically separated from the underlying IMC and the staining of TgIMP1 is co-localized with the plasma membrane marker SAG1. (D) Freshly released parasites were incubated with proteinase K (PK) with or without Digitonin (Dig.) permeabilization. Unlike TgSAG1 localized to the outer surface of the plasma membrane, TgIMP1 could be digested only after digitonin treatment. (E) Freshly released parasites were subjected to detergent extraction in PBS, 1%TX100 in PBS or 1%TX100 IN TBS. Proteins in the total lysate (Total), supernatants (PBS, 1%TX100/PBS or 1%TX100/TBS), and pellet fractions were assessed by western blot analysis.

mately 25 kDa was observed, which we predicted to represent the conserved IMP1-like domain. To characterize the foldedness of TgIMP1 further we recorded a 1D ^{1}H NMR spectrum. The presence of high-field shifted methyl groups between 0.5 and −0.5 ppm (Fig. 4B and inset) indicates an ordered, globular domain, however the spectrum is dominated by a large highly disordered region, represented by the sharp aliphatic peaks between 0.5 and 3.5 ppm (Fig. 4B). This region was attributed to the N-terminus carrying the acylation sites. Truncation of the N-terminus and subsequent evaluation by NMR localized the folded domain globular within TgIMP1 residues encompassing S176 to A400, which corresponds to the predicted IMP1-like domain identified from the sequence alignment (Fig. 4C). We then embarked on structural studies using multidimensional NMR. Attempts to complete a full backbone resonance assignment of a $^{13}C/^{15}N/^{2}H$-labeled sample of the TgIMP1 IMP1-like domain were unsuccessful as several amide resonance were absent in the spectrum, which indicates significant conformation exchange that would preclude a high resolution structure determination (Fig. 4D).

To circumvent this, we explored our NMR studies on several homologues of TgIMP1, in which the IMP1-like domain is present and conserved. One such example that yielded dramatically improved NMR spectra was IMP2 from *Plasmodium falciparum* (PfIMP2).

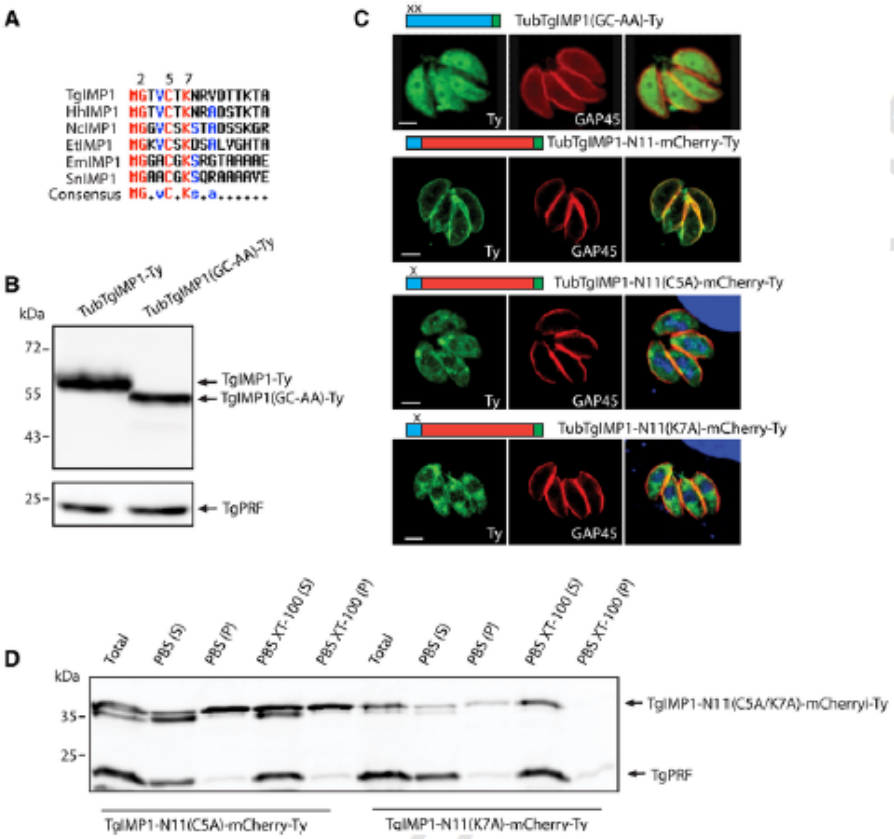


Fig. 3. The N-terminal 11 amino acids contain the minimal signals for TgIMP1 plasma membrane localization. (A) The first 16 amino acids of TgIMP1 was aligned with its homologues in coccidian: *Neospora caninum* (NCLIV_000430), *Hammondia hammondi* (HHA_293470), *Sarcocystis neurona* (SN3_00701045), *Eimeria tenella* (ETH_00030475) and *Eimeria maxima* (EMWEY_00011000). All of the homologues have a conserved glycine at position two predicted to be myristoylated, a cysteine at position five predicted to be palmitoylated and a lysine at position seven. (B) Mutation of the residues predicted for myristoylation (G2) and palmitoylation (C5) to alanine were generated by point mutation of the construct pTubTgIMP1-Ty-HX. After the mutation of the dual acylation sites, a size shift was observed in the mutant by western blot, due to lost dual acylation. (C) The TgIMP1 (GC-AA)-Ty mutation altered localization from plasma membrane (PM) to cytoplasm, indicating a loss of PM interaction due to the absence of acylation. (C middle) The N-terminus 11 amino acids of TgIMP1 were fused with mCherry-Ty and the chimera is still localized to the PM. TgIMP1 lost PM localization when C5A and K7A were introduced. (D) The solubility of TgIMP1n11mCherry-Ty mutants was analyzed by fractionation assay and C5A and K7A mutants were more soluble in PBS compared to TgIMP1.

PfIMP2 is expressed predominantly in the late schizont stages of the intra-erythrocytic development, suggesting that the IMPs could participate in egress or invasion [14]. In *T. gondii*, TgIMP2.1 was epitope-tagged and shown to be cytosolic (Supplemental Fig. S2A) with an apparent molecular weight of ~ 26 kDa (Supplemental Fig. S2B).

The quality of the NMR spectra for PfIMP2 facilitated the complete sequential assignment using heteronuclear multidimensional NMR methods [15]. Subsequent side-chain assignment facilitated an automated analysis of the nuclear Overhauser data and the solution structure determination using the ARIA protocol [16]. The final structural statistics are summarized in Table 1. The 10 lowest energy-minimized structures exhibited a root-mean-square deviation (RMSDs) from the average structure of 0.25 Å for all backbone atoms and about 0.58 Å for all non-hydrogen atoms (Fig. 5A).

The structure of PfIMP2 contains a central long helix (α3) that is sandwiched between two β-sheets; one comprising five β-strands (β1–β5) and the other four (β6–β9; Fig. 5B). Two other major helices, α1 and α5, are located within inter-strand loops β3–β4 and β8–β9, respectively. Two short helices are present at the end of the large helix α3 and at the C-terminus α4. Interestingly, the structure displays a striking internal symmetry. The protein sequence contains a single cysteine residue (C71) located in the second α-helix, which is conserved across the IMP2 family (Fig. 5C). Searches for protein structure similarities using available servers [17], principally DALI and PDBeFold [18,19], failed to reveal any statistically significant similarity to protein domains of known structure, thus confirming the identification of a new fold. Predicted phosphorylation sites residues include S92 and S93, which are located near the sixth beta strand (Fig. 5C). Based on the structure of PfIMP2, S92 does not seem to be available for phosphorylation as it points to the interior of the structure, whereas S93 is solvent accessible. The structure of the IMP1-like domain of TgIMP1 was readily homology modeled based on the structure of PfIMP2. A comparison of the structures of TgIMP1 and PfIMP2 showed that TgIMP1 has three insertions that are absent from PfIMP2 (Fig. 5D and Supplemental Fig. S3). These insertions occur within loop regions of the IMP1-like domain and are predicted to possess some helical secondary structure. The homology model predictions are consistent with available NMR chemical shifts

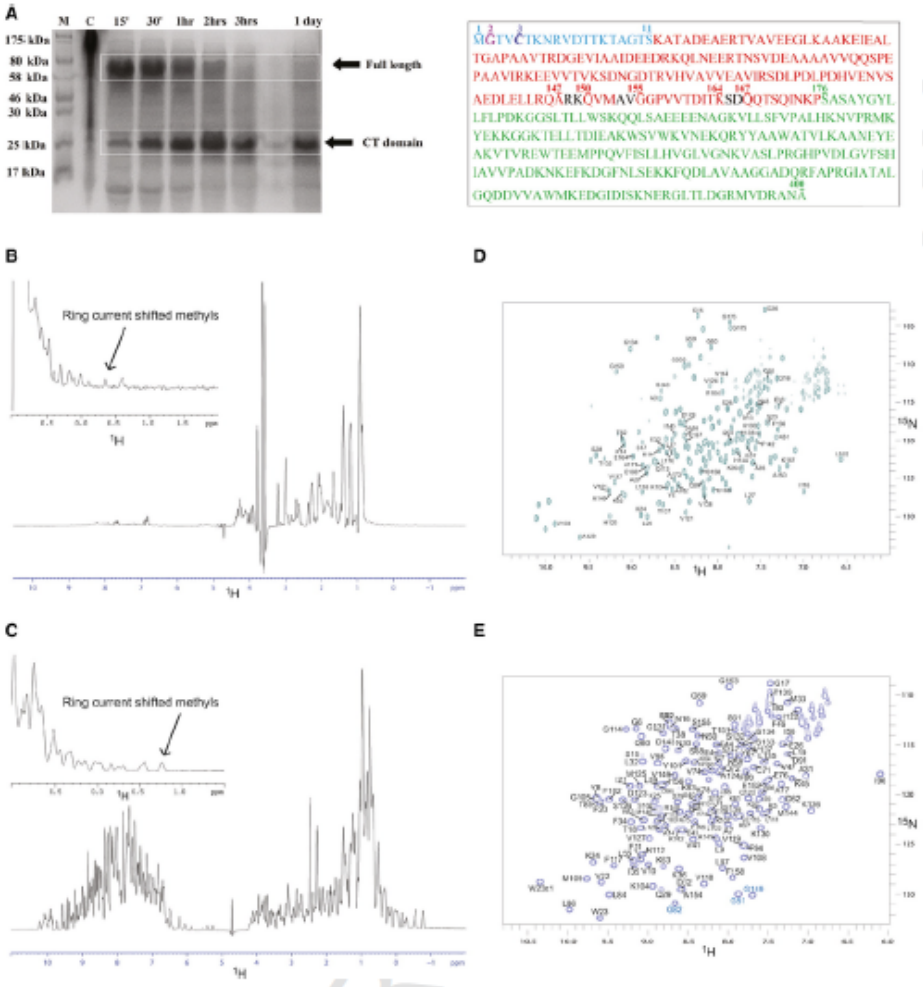


Fig. 4. Toward the solution structure of the globular IMP1-like domain of TgIMP1. (A) SDS-PAGE gel obtained from running the samples from limited proteolysis of TgIMP1 IMP1 domain with Chymotrypsin. The lanes labeled M and C correspond to marker and control samples (TgIMP1 IMP1 domain), respectively. The time points at which samples were taken are labeled for the remaining lanes. Sequence showing of the locations of the domains and motifs in TgIMP1. The N-terminal plasma membrane localization signal is colored blue, the myristoylation site is colored purple, palmitoylation site colored navy blue, the linker region colored red and the IMP1 domain colored green. The cleavage sites of chymotrypsin, determined by sequencing, are colored black. (B) Figure showing the 1D ^{1}H NMR spectrum of full length TgIMP1. The methyl region of the spectrum is zoomed in and shown on the top left. The spectrum shows that the protein has a large unfolded domain, which is masking the signal from the folded domain. (C) 1D ^{1}H NMR spectrum of the isolated globular domain from TgIMP1, termed the IMP1-like domain. The high chemical shifts dispersion in the spectrum with peaks at < 0.5 ppm and > 8.5 ppm shows that this construct contains the ordered domain and the flexible region seen in full length TgIMP1 protein is absent. (D) 2D ^{1}H-^{15}N HSQC spectrum of the IMP1-like domain of TgIMP1 (^{13}C/^{15}N/^{2}H labeled). The 3D data was incomplete with between 5 and 10% of amide peaks absent, therefore a comprehensive assignment with 100% confidence was not possible. Approximately 50% of the sequence was assigned confidently as shown. (E) 2D ^{1}H-^{15}N HSQC spectrum of the IMP domain of TgIMP1 (^{13}C/^{15}N/^{2}H labeled). The residues that were assigned with 100% confidence are labeled in the spectrum. As the 3D data was incomplete, less than 50% of the sequence was assigned confidently.

and are conserved across equivalent coccidian proteins suggesting that they may have some functional significance. The analysis of the solvent-exposed surface properties of PfMP2 reveals a prominent surface comprising exposed hydrophobic residues delineated by significant surface charge (Fig. 6A–C; middle and right). Although this observation would suggest a role for this region in a functional interaction, presumably with another parasite component, the surface conservation is limited in places that may suggest different target ligands (Fig. 6D).

2.6. The disruption of TgIMP1 does not impact on tachyzoite fitness

To assess its importance of TgIMP1 for parasite survival, a gene replacement strategy based on double homologous recombination was designed using HXGPRT as selectable marker and Ku80-ko as recipient strain (Fig. 7A). Parasites resistant to mycophenolic acid were cloned and *TgIMP1* knockouts (IMP1-ko) were identified based on the loss of its N-terminal coding region by genomic PCR analysis (Fig. 7B). The absence of TgIMP1 was confirmed by western blot,

Table 1
NMR structure calculation statistics data for PfIMP2.

NMR restraints	
Distance constraints	
Total NOE	1683
Intraresidue	698
Interresidue	985
Sequential (| *i-j* | = 1)	443
Medium range (| *i-j* | < 4)	94
Long range (| *i-j* | > 5)	348
Total dihedral angle restraints	
Φ	136
Ψ	136
Structural statistics	
Violations (mean and SD)	
Distance constraints (Å)	0.033 ± 0.004
Dihedral angle constraints (°)	1.40
Maximum dihedral angle violation (°)	0.67
Maximum distance constraint violation (Å)	0.54
Deviations from idealized geometry	
Bond length (Å)	0.004 ± 0.001
Bond angle (°)	0.52 ± 0.014
Impropers (°)	1474 ± 0.080
Average pairwise rmsd*(Å)	
Heavy	1.966
Backbone	2.317

using anti-TgIMP1 antibodies (Fig. 7C). The phenotypic consequence of TgIMP1 deletion was first examined by a plaque assay, which recapitulates all the steps of the lytic cycle including host cell attachment, invasion, intracellular proliferation and egress from infected cells. No obvious difference in the size of the plaques was observed between parental parasites (Ku80-ko) and IMP1-ko (Fig. 7D). Quantitative assays also failed to detect any significant defect in intracellular growth and replication (Fig. 7E). Furthermore, and in contrast to a previous report from the literature [4], IMP1-ko showed no impairment in host cell invasion (Fig. 7F).

2.7. TgIMP1 confers no significant immune protection

Parasites lacking TgIMP1 were first examined for possible alteration of virulence in the mouse model of infection. The Ku80-ko RH strain, used as the background to disrupt TgIMP1, is a type I strain that exhibits a severe virulence phenotype in mice with a very low inoculum. The intraperitoneal injection of 500 parental (Ku80-ko) or IMP1-ko parasites led to the development of severe symptoms in all the mice of the two groups and the animals were sacrificed on day 8 following infection (Fig. 8A).

An earlier study indicated a benefit of TgIMP1 DNA as vaccine that could elicit protective immunity against *T. gondii* tachyzoites challenge [3]. The IMP1-ko parasites were used here to determine the specificity of this DNA vaccine toward TgIMP1. When comparing the protection efficacy of vaccinated mice (as previously described) challenged with either parental stain (Ku80-ko) or IMP1-ko parasites, no difference in immune protection could be observed (Fig. 8B, C).

3. Discussion

This study aimed to explore the role and importance TgIMP1 in *T. gondii*, a conserved protein across the coccidians which was previously identified as a key immunity protective protein in *E. maxima* [2]. TgIMP1 is shown here to localize to the parasite pellicle and more precisely on the inner leaflet of the PM, most likely via dual acylation. Presumably TgIMP1 is co-translationally modified by N-myristoylation at G2 post-translationally by S-acylation where palmitic acid is added at C5. This interpretation is further supported by the identification of TgIMP1 in the recently reported palmitome of *T. gondii* [20,21]. The dual acylation in the N-terminal of TgIMP1 is essential for its PM localization. Besides the two conserved residues for myristoylation (G2) and palmitoylation (C5), the K7 is also highly conserved. Relevantly, K7 has previously been reported as a critical determinant for myristoylation in viral and mammalian proteins [22]. Myristoylation or other features that enhance affinity to membrane are a prerequisite for palmitoylation, which drives the initial membrane interaction. Expectedly, mutation of K7/A in the first N11 amino acid sequence of TgIMP1 abolished the PM localization of TgIMP1, while retaining its membrane-binding properties, indicating the rules of substrate specificity vary in different organism [23]. The protein S-acyl transferase (PAT) implicated in the anchoring of TgIMP1 to the PM is not known, however, the repertoire of PAT in *T. gondii* has been fully characterized and it is plausible that PAT located at the PM such as TgDHHC4 or TgDHHC13 could be involved [24].

NMR data showed that the N-terminal region of TgIMP1 is highly dynamic and attached to a structured C-terminal domain of ~ 25 kDa in size. Although the apparent disordered nature of this region could be due to the lack of myristoylation and palmitoylation in recombinant bacterial expression, it is more likely a genuine property and enables projection of the C-terminal domain away from the inner leaflet of the plasma membrane. The inherent flexibility of the linker would allow the C-terminal domain in IMP1 to sample various orientations for optimal interaction with a binding partner. This domain is conserved in a wider family of apicomplexan proteins including IMP1 and IMP2. The structure of this domain reveals a fold not previously observed in protein structures determined to date. Although this new structure fails to yield clear functional clues through similarity, the domain architecture provides a valuable addition to template libraries used in protein structure prediction. Furthermore, our structure provides some immediate suggestions for mutagenesis, once the specific cellular roles of IMP1 and IMP2 have been deduced.

Proteomics studies indicated that expression of IMP1 and IMP2 are elevated during invasion and therefore might participate directly host cell invasion and gliding mobility. Furthermore, a previous study showed that antisera against NcIMP1 inhibited *N. caninum* invasion in tissue culture [4]. The disruption of TgIMP1 in this work established the dispensable nature of the gene, and egress and invasion assays with IMP1-ko parasites showed no significant invasion defect when compared to the parental Ku80-ko strain.

Previous immunization experiments using TgIMP1 as a DNA vaccine induced modest but statistically significant immune protection against *T. gondii* [3]. However, the use of parasites lacking TgIMP1 failed to confirm that the previously reported protection was specific to TgIMP1. Importantly, studies on *T. gondii* and *E. tenella* have demonstrated that the efficiency and potency of the host immune response elicited by a certain antigen is considerably influenced by its subcellular localization [25,26]. There are few examples of myristoylated and palmitoylated proteins that could be exported to the outer surface of the PM, like for example phosphinositide-specific phospholipase C (PLC) in *Trypanosoma cruzi* [27], the protection of TgIMP1 from PK degradation in intact extracellular parasites showed that the protein is anchored in the cytoplasmic face of PM. Since intracellular *Toxoplasma* proteins have been reported to induce protection [28] it is possible that IMP1 serves as an adjuvant or as stimulator of innate immune responses.

T. gondii isolates have been classified into three genetic types (I, II, and III) that differ markedly in their virulence [29]. Although amino acid polymorphism of IMP1 occurs in two *E. maxima* strains, the comparison of the amino acid sequences of TgIMP1 from the three different types revealed no polymorphism in *T. gondii*. RH parasite lacking TgIMP1 have shown no alteration of virulence, how-

UNCORRECTED PROOF

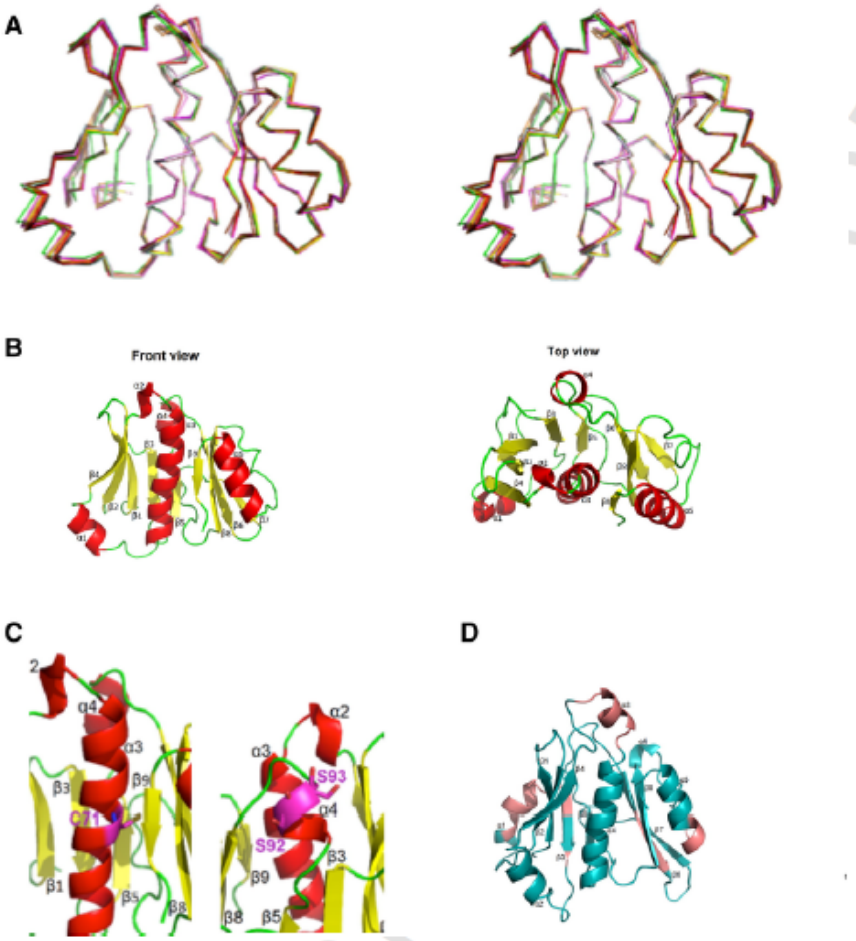


Fig. 5. Solution structure of PfIMP2 determined from NMR data. (A) Stereo representation of PfIMP2 NMR family of structures. (B) Front and top views of cartoon representation of PfIMP2. Secondary structure elements showing in red for α-helices, yellow for β-strands and green for loops, and are labeled according to sequence order. (C) Left – Cartoon representation of PfIMP2 showing the location of the single cysteine residue (C71 – pink stick representation). Right Cartoon representation of PfIMP2 showing the two predicted phosphorylation sites (S92 and S93) in pink. (D) Cartoon representation of TgIMP1 structure, which was homology modeled based on the structure of PfIMP2 using Phyre3. The insertions present in TgIMP1 but absent in PfIMP2 are shown in light pink.

ever, it is not possible to extrapolate this result to the other strain types.

TgIMP1 is expressed in all life stages of *T. gondii* except in merozoites [30] and the raw data from transcript database indicate two expression peaks appearing in sporulated oocysts and during the second generation of the merozoites in *E. tenella* [31]. Although disruption of the *TgIMP1* in tachyzoites showed no loss of parasite fitness, a role in other stages, as well as some redundancy between members of this family of proteins, cannot be excluded.

4. Materials and methods

4.1. Cell culture and transfection of Toxoplasma gondii

T. gondii tachyzoites (RH-hxgprt-ko or RH-Ku80-ko) were maintained by serial passage in human foreskin fibroblasts (HFF) monolayer cells. Transfections were undertaken by electroporation as described preciously [32]. Selection of transgenic parasites was performed with mycophenolic acid (MPA) and xanthine for HXGPRT gene selection [33] and stable clones were isolated by limiting dilution in 96 well plates.

4.2. Phylogenetic tree construction

Sequences of Apicomplexan orthologs of IMP related proteins were procured from EuPathDB and aligned using MUSCLE sequence alignment software [34]. The resulting sequence alignment was manually curated with BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) to select for the IMP region and to remove uninformative alignments. The curated MUSCLE alignment was fed into PhyML [35] for phylogenetic tree construction, using LG model of

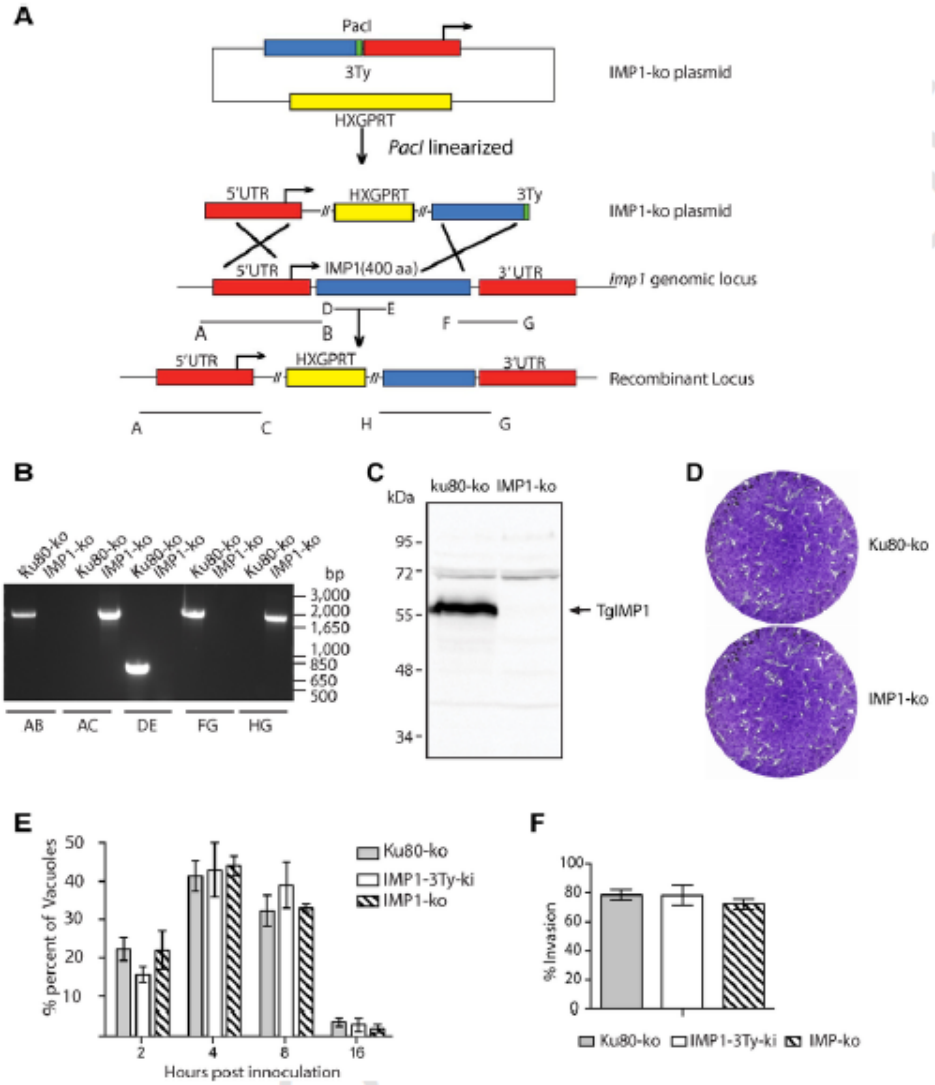


Fig. 7. Generation of a TgIMP1 knockout. (A) Schematic representation of the strategy used to delete the endogenous *TgIMP1* N-terminus region by double homologous recombination. The knockout TgIMP1 construct contains 1481 bp of the 5' un-translated region (5'UTR) and the 1629 bp of its C-terminus coding sequence. Arrows represent the primers used for PCR reaction and the lengths of the PCR products are marked. (B) PCR analysis was performed on the IMP1-ko stain, showing that double homologous recombination occurs (primer combination AB, AC for the 5' region and FG, HG for 3' region) and the N-terminus of *TgIMP1* was deleted (primer combination DE). (C) Western blotting using antisera against recombinant TgIMP1 indicated the lost TgIMP1 expression in the IMP1-ko strain. (D) Plaque assay performed on an HFF monolayer inoculated with Ku80-ko or IMP1-ko parasites. Eight days after inoculation, the HFF was fixed and strained with crystal violet. (E) Freshly released parasites from Ku80-ko or IMP1-ko strain invade new HFF cell monolayer. Numbers of parasites per vacuoles were counted at 24 h after HFF cell infection. Values are means standard deviations (± SD) for three independent experiments. (F) Parasites lacking IMP1 were not impaired in invasion.

amino acids substitution with NNI topology search. The sequences used in the phylogenetic analysis were listed in Supplemental Fig. S1.

4.3. Toxoplasma constructs

The primers for the TgIMPs were designed based on the TgIMP1 (TGME49_293470) and TgIMP2.1 (TGME49_276930) sequences obtained from the ToxoDB database and all the primers used in this study were listed in the Supplemental Table 1. The vector pTub8T-gIMP1-Ty-HX was obtained by cloning of the TgIMP1 cDNA fragment into the *Eco*RI and *Nsi*I sites in the pTubmy-cGFP-PfmyoATail-Ty-HX. For the full length C-terminal triple Ty-tagging constructs pKI-TgIMP1-3Ty-HX and pKI-TgIMP2.1-3Ty-HX both around 1.5 kb genomic fragments were amplified, digested with *Kpn*I and *Sbf*I and cloned into *Kpn*I and *Nsi*I sites of the pTub8MIC13-3Ty-HX construct [36]. Prior to transfection, plasmids were linearized in the middle of the PCR-amplified genomic fragments.

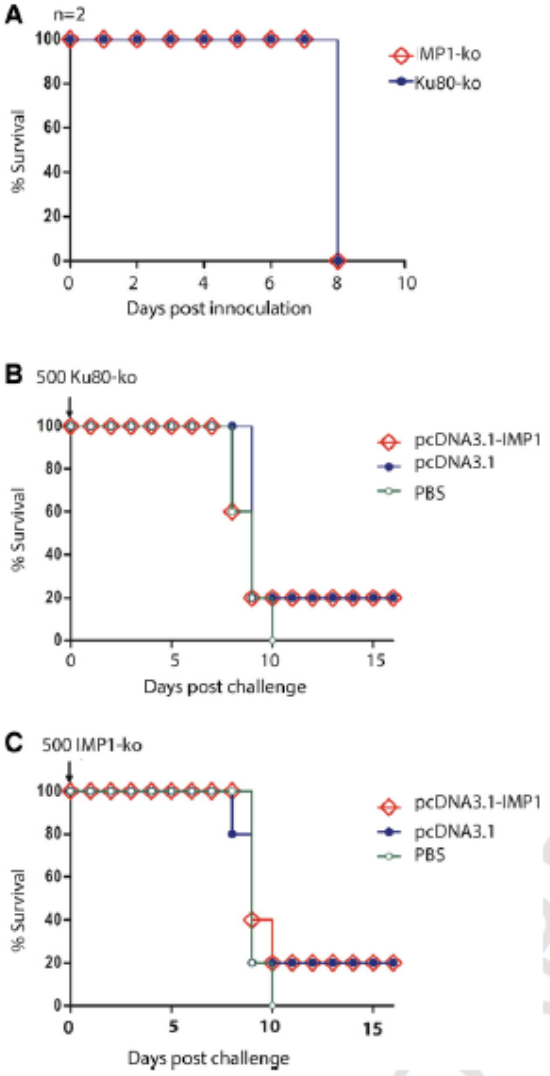


Fig. 8. Protection of BALB/c mice against Ku80-ko and IMP1-ko challenge. A) Virulence comparison between Ku80-ko and IMP1-ko. All mice were sacrificed on day 8 following peritoneal injection of 500 parental (Ku80-ko) or IMP1-ko parasites. B) Survival curves of immunized BALB/c mice after lethal challenge with 500 tachyzoites of Ku80-ko and IMP1-ko, 2 weeks after the last immunization.

The construct pTubTgIMP1(GC-AA) with mutations at the myristoylation and palmitoylation sites was generated by PCR mutagenesis (primers listed in Supplemental Table 1). The mCherry-Ty fusion construct mutants were verified by DNA sequencing. For the IMP1-ko vector, a genomic fragment of 1481 bp corresponding of the 5′ UTR of the *TgIMP1* gene was amplified by PCR and inserted into the *Pac*I site of pKI TgIMP1-3Ty-HX. The direction of insertion and sequence was checked by restriction digestion and sequencing. The final IMP1-ko construct was linearized with *Pac*I prior to transfection.

4.4. Western blot

Freshly released tachyzoites were harvested, washed once in PBS, solubilized directly in SDS-loading buffer, separated by electrophoresis in 12% polyacrylamide gels and transferred. Western blots were incubated with respective mouse/rabbit antibodies in PBS, 0.05% Tween 20 and 5% skimmed milk power. After washes, the membrane was incubated with a peroxidase-conjugated goat anti-mouse or anti-rabbit antibody. Bound antibodies were visualized using the ECL plus system (GE Healthcare Bio-Sciences). Rabbit monoclonal anti-profilin (1:3000) was used as loading control.

4.5. Recombinant protein production

A DNA fragment encoding PfIMP2 (GenBank XP_001349200, PF3D7_0730400) was codon optimized for bacterial expression and synthesized (GeneArt, Invitrogen). This PfIMP2 sequence was subcloned into pNIC-ZB vector to express PfIMP2 as a fusion protein with TEV cleavable N-terminal His6 tag. pNIC-ZB was a gift from Opher Gileadi (Addgene plasmid # 26107). For protein expression, this vector was transformed into *Escherichia coli* Rosetta™2(DE3) (NEB, UK).

Transformants were grown overnight at 37 °C in 100 ml M9 minimal medium with 50 ng/ml kanamycin and 35 ng/ml chloramphenicol. This culture was clarified by centrifugation at 4500 rpm and the cell pellet resuspended in 2 l M9 minimal medium with 0.07% $^{15}NH_4Cl$ and 0.2% $^{13}C_6$-glucose (Sigma) and kanamycin and chloramphenicol for $^{15}N/^{13}C$-labeling of PfIMP2. Cell cultures at OD600 0.5–0.7 were induced with 1 mM IPTG and left to grow at 18 °C overnight. Cells were then harvested by centrifugation, resuspended in ice-cold 50 mM Hepes buffer pH 8, 300 mM NaCl and lysed by sonication. The lysate was centrifuged at 16,000 rpm for 40 min to obtain supernatant. The supernatant was purified by affinity chromatography using pre-packed Nickel column, 5 ml Histrap FF (GE Healthcare, UK) in the AKTAprime (GE Healthcare, UK). The purified protein was cleaved with TEV protease to remove the N-terminal His tag and further purified by Gel filtration using a Superdex 75 Hiload16/600 column (GE Healthcare, UK) with AKTAprime (GE Healthcare, UK). The purified protein was then concentrated and dialyzed into 50 mM Hepes, 250 mM NaCl and 10 mM DTT, pH 7.5. The final construct that was used to record NMR experiments contained a serine residue left from the TEV cleavage, followed by the complete amino acid sequence of PfIMP2. The molecular weight of this protein is 17.89 kDa. There are 159 residues in this construct but the methionine following the first serine residue is labeled as residue 1 throughout this article.

4.6. NMR assignment and solution structure calculation

NMR spectra were recorded at 303 K on Bruker DRX600 and DRX800 spectrometers equipped with cryo-probes. The chemical shifts of ^{1}HN, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$ and ^{13}CO cross peaks were assigned using CBCA(CO)NH, HNCACB, HNCO and HN(CA)CO experiments [15,37]. The aliphatic and aromatic side chain ^{1}H and ^{13}C assignments were obtained from HBHA (CBCACO)NH, HCCH-TOCSY, (H)CC(CO)NHTOCSY and ^{1}H–^{13}C NOESY-HMQC spectra. Three-dimensional ^{1}H-$^{15}N/^{13}C$ NOESY-HMQC (mixing time 100 ms at 600 and 800 MHz) experiments provided the distance restraints for final structure calculation. The ARIA protocol (Linge et al., 2003) was used for expansion of the NOE assignment and final structure calculation. Dihedral angle restraints were derived from TA-

LOS (Shen and Bax, 2013). The frequency window tolerance for assigning NOEs was ± 0.05 and ± 0.07 ppm for direct and indirect proton dimensions and ± 0.7 and ± 1.2 ppm for nitrogen and carbon dimensions, respectively. The ARIA parameters were set to default values. The structure is deposited under PDB accession number is 5LG9 and the BRMB ID for the assignment is 26552.

4.7. Antibodies

Anti-GAP45, anti-Ty tag, anti-Profilin (PRF), anti-SAG1, anti-Catalase and anti-IMP1 antibodies were previously described [3,12]. Immunoblots were visualized using a chemiluminescent substrate (Amersham, GE Healthcare). Anti-PRF antibodies were used to detect Profilin as loading control.

4.8. Indirect immunofluorescence assay (IFA) and microscopy

Parasites infected HFF cells were fixed with 4% paraformaldehyde and 0.05% Glu in phosphate-buffered saline (PBS) for 15 min, followed by 5 min of neutralization with 0.1 M glycine in PBS. Fixed cells were permeabilized with 0.2% Triton X-100 in PBS (PBS-TX) for 20 min and then blocked with 2% bovine serum albumin (BSA). Intracellular growth and plaque assay were performed as previously described. Confocal images were taken with a Leica laser scanning confocal microscope (TCS-NT DM/IRB and SP2) using a 100 × Plan-APO objective.

4.9. Aerolysin assay and Proteinase K protection assay

Separation between IMC and MP was achieved by aerolysin treatment as previously described [38]. To further refine the localization of TgIMP1 at the pellicle, freshly released parasites were harvested, washed in PBS once and resuspended in 1.5 ml of cold SoTE (0.6 M sorbitol, 20 mM Tris–HCl pH 7.5 and 2 mM EDTA) prior to division into three tubes (0.5 ml each). Cold SoTE was added to tube 1 and used as a control. Parasites in tubes 2 and 3 were permeabilized by addition of 0.5 ml cold 0.01% digitonin (Sigma) in SoTE. Samples were carefully mixed by inversion and incubated on ice/10 min prior to centrifugation (2000 g/4 °C/10 min). Supernatant was discarded and 0.5 ml cold SoTE was added to tube 1, while 2.25 μl of cold Proteinase K (Sigma, 20 mg/ml) in SoTE was added to tube 3. All tubes were gently mixed by inversion and incubated on ice/30 min. Proteinase K was inactivated by addition of cold trichloroacetic acid (TCA) to a final concentration of 10% and further incubation on ice/30 min. All samples were centrifuged at 14,000 rpm for 20 min, washed twice with acetone, air-dried briefly and subsequently resuspended in TE buffer prior to SDS–PAGE analyses.

4.10. Differential detergent extraction and cell fractionation

The differential detergent extraction of proteins was performed as described previously [39]. Briefly, the freshly released tachyzoites were harvested, washed once with PBS and then were extracted for 10 min on ice into PBS with 1% Triton X-100 (TX100/PBS). Soluble material and pellet were separated by centrifugation at 14,000 rpm for 30 min at 4 °C. The pellet was subsequently extracted for 10 min on ice in TBS with 1% Triton X-100 (TX100/TBS) and the pellet and soluble fraction were separated by centrifugation as described above.

Purified tachyzoites were resuspended into PBS or PBS with 1% Triton-X-100. Parasites were lysed by freeze and thaw three times followed by sonication on ice. Pellet and soluble fractions were separated by centrifugation at 14,000 rpm for 30 min at 4 °C. The final pellet was washed once with PBS and resuspended in 100 μl PBS containing the insoluble fraction. Equal amounts of all supernatants were analyzed by immunoblotting. Proteins were detected using anti-Ty antibodies, the cytosolic profilin (TgPRF) and the pellicular TgGAP45 were used as control.

4.11. TgIMP1 DNA vaccine protection experiments

Female Kunming mice (5–6 weeks old) were purchased from Academy of Military Medical Sciences Laboratory Animal Center (Beijing, China). All mice were maintained under standard conditions and provided with rodent feed and water *ad libitum*. Prior to experiment, mice were acclimatized for one week. Animal experiments were conducted according to institutional guidelines for animal ethics. Mice (10/group) were injected intramuscularly with 100 μg of pcDNA-IMP1 plasmid DNA suspended in 100 μl sterile PBS in the thigh skeletal muscle. Control mice received PBS alone or empty pcDNA 3.1(+) vector for the first injection. The second and third injections were carried out in 2 and 4 weeks post-primary injection with the same dosage. One week after the final injection, the mice of all groups were challenged intraperitoneally with 500 tachyzoites of the Ku80-ko or IMP1-ko strain.

Supplementary data to this article can be found online at doi:10.1016/j.bbapap.2016.11.010.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgments

This work was supported by the Swiss National Foundation (FN3100A0-116722 to DSF, National Natural Science Foundation of China (31372424) to QL and in part by the Wellcome Trust (Senior Investigator Award 100280 and multiuser equipment grant 104833 to SJM) DSF is HHMI senior international research scholar.

References

[1] E.A. Innes, A.N. Vermeulen, Vaccination as a control strategy against the coccidial parasites Eimeria, Toxoplasma and Neospora, Parasitology 133 (Suppl) (2006) S145–S168.
[2] D.P. Blake, K.J. Billington, S.L. Copestake, R.D. Oakes, M.A. Quail, K.L. Wan, M.W. Shirley, A.L. Smith, Genetic mapping identifies novel highly protective antigens for an apicomplexan parasite, PLoS Pathog. 7 (2011).
[3] X. Cui, T. Lei, D. Yang, P. Hao, B. Li, Q. Liu, Toxoplasma gondii immune mapped protein-1 (TgIMP1) is a novel vaccine candidate against toxoplasmosis, Vaccine 30 (2012) 2282–2287.
[4] X. Cui, T. Lei, D.Y. Yang, P. Hao, Q. Liu, Identification and characterization of a novel Neospora caninum immune mapped protein 1, Parasitology 139 (2012) 998–1004.
[5] M.M. Corvi, L.G. Berthiaume, M.G. De Napoli, Protein palmitoylation in protozoan parasites, Front. Biosci. 3 (2011) 1067–1079.
[6] M.M. Corvi, A.M. Alonso, M.C. Caballero, Protein palmitoylation and pathogenesis in apicomplexan parasites, J. Biomed. Biotechnol. 2012 (2012) 483969.
[7] K. Frenal, L.E. Kemp, D. Soldati-Favre, Emerging roles for protein S-palmitoylation in Toxoplasma biology, Int. J. Parasitol. 44 (2014) 121–131.
[8] Y. Shi, R.A. Mowery, J. Ashley, M. Hentz, A.J. Ramirez, B. Bilgicer, H. Slunt-Brown, D.R. Borchelt, B.F. Shaw, Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding, Protein Sci. 21 (2012) 1197–1209.
[9] N. Ogino, C. Yoneda, The fine structure and mode of division of Toxoplasma gondii, Arch. Ophthalmol. 75 (1966) 218–227.
[10] J.F. Dubremetz, G. Torpier, Freeze fracture study of the pellicle of an eimerian sporozoite (Protozoa, Coccidia), J. Ultrastruct. Res. 62 (1978) 94–109.
[11] D.T. Ouologuem, D.S. Roos, Dynamics of the Toxoplasma gondii inner membrane complex, J. Cell Sci. 127 (2014) 3320–3330.

UNCORRECTED PROOF

[12] K. Frenal, V. Polonais, J.B. Marq, R. Stratmann, J. Limenitakis, D. Soldati-Favre, Functional dissection of the apicomplexan glideosome molecular architecture, Cell Host Microbe 8 (2010) 343–357.
[13] L. Abrami, M. Fivaz, P.E. Glauser, R.G. Parton, F.G. van der Goot, A pore-forming toxin interacts with a GPI-anchored protein and causes vacuolation of the endoplasmic reticulum, J. Cell Biol. 140 (1998) 525–540.
[14] C. Aurrecoechea, J. Brestelli, B.P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O.S. Harb, M. Heiges, F. Innamorato, J. Iodice, J.C. Kissinger, E. Kraemer, W. Li, J.A. Miller, V. Nayak, C. Pennington, D.F. Pinney, D.S. Roos, C. Ross, C.J. Stoeckert Jr., C. Treatman, H. Wang, PlasmoDB: a functional genomic database for malaria parasites, Nucleic Acids Res. 37 (2009) D539–D543.
[15] S. Benjamin, F. Williams, L. Kerry, S. Matthews, NMR assignment of the immune mapped protein 1 (IMP1) homologue from Plasmodium falciparum, Biomol. NMR Assign. 9 (2015) 393–395.
[16] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, ARIA2: automated NOE assignment and data integration in NMR structure calculation, Bioinformatics 23 (2007) 381–382.
[17] G. Mayr, F.S. Domingues, P. Lackner, Comparative analysis of protein structure alignments, BMC Struct. Biol. 7 (2007) 1–15.
[18] L. Holm, C. Sander, Dali: a network tool for protein structure comparison, Trends Biochem. Sci. 20 (1995) 478–480.
[19] E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2256–2268.
[20] I.T. Foe, M.A. Child, J.D. Majumdar, S. Krishnamurthy, W.A. van der Linden, G.E. Ward, B.R. Martin, M. Bogyo, Global analysis of palmitoylated proteins in Toxoplasma gondii, Cell Host Microbe 18 (2015) 501–511.
[21] M.C. Caballero, A.M. Alonso, B. Deng, M. Attias, W. de Souza, M.M. Corvi, Identification of new palmitoylated proteins in Toxoplasma gondii, Biochim. Biophys. Acta 1864 (2016) 400–408.
[22] J.M. Kaplan, G. Mardon, J.M. Bishop, H.E. Varmus, The first seven amino acids encoded by the v-src oncogene act as a myristylation signal: lysine 7 is a critical determinant, Mol. Cell. Biol. 8 (1988) 2435–2441.
[23] J.A. Traverso, C. Giglione, T. Meinnel, High-throughput profiling of N-myristoylation substrate specificity across species including pathogens, Proteomics 13 (2013) 25–36.
[24] K. Frenal, C.L. Tay, C. Mueller, E.S. Bushell, Y. Jia, A. Graindorge, O. Billker, J.C. Rayner, D. Soldati-Favre, Global analysis of apicomplexan protein S-acyl transferases reveals an enzyme essential for invasion, Traffic 14 (2013) 895–911.
[25] B. Gregg, F. Dzierszinski, E. Tait, K.A. Jordan, C.A. Hunter, D.S. Roos, Subcellular antigen location influences T-cell activation during acute infection with Toxoplasma gondii, PLoS One 6 (2011), e22936.
[26] X. Huang, J. Zou, H. Xu, Y. Ding, G. Yin, X. Liu, X. Suo, Transgenic Eimeria tenella expressing enhanced yellow fluorescent protein targeted to different cellular compartments stimulated dichotomic immune responses in chickens, J. Immunol. 187 (2011) 3595–3602.
[27] V.D. Martins, M. Okura, D. Maric, D.M. Engman, M. Vieira, R. Docampo, S.N.J. Moreno, Acylation-dependent export of Trypanosoma cruzi phosphoinositide-specific phospholipase C to the outer surface of amastigotes, J. Biol. Chem. 285 (2010) 30906–30917.
[28] L.D. Sibley, S.D. Sharma, Ultrastructural localization of an intracellular Toxoplasma protein that induces protection in mice, Infect. Immun. 55 (1987) 2137–2141.
[29] D.K. Howe, L.D. Sibley, Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease, J. Infect. Dis. 172 (1995) 1561–1566.
[30] B. Gajria, A. Bahl, J. Brestelli, J. Dommer, S. Fischer, X. Gao, M. Heiges, J. Iodice, J.C. Kissinger, A.J. Mackey, D.F. Pinney, D.S. Roos, C.J. Stoeckert Jr., H. Wang, B.P. Brunk, ToxoDB: an integrated Toxoplasma gondii database resource, Nucleic Acids Res. 36 (2008) D553–D556.
[31] L.T. Rangel, J. Novaes, A.M. Durham, A.M. Madeira, A. Gruber, The Eimeria transcript DB: an integrated resource for annotated transcripts of protozoan parasites of the genus Eimeria, Database (Oxford) 2013 (2013), bat006.
[32] D. Soldati, J.C. Boothroyd, Transient transfection and expression in the obligate intracellular parasite Toxoplasma gondii, Science 260 (1993) 349–352.
[33] R.G. Donald, D. Carter, B. Ullman, D.S. Roos, Insertional tagging, cloning, and expression of the Toxoplasma gondii hypoxanthine-xanthine-guanine phosphoribosyltransferase gene. Use as a selectable marker for stable transformation, J. Biol. Chem. 271 (1996) 14010–14019.
[34] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[35] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321.
[36] N. Friedrich, J.M. Santos, Y. Liu, A.S. Palma, E. Leon, S. Saouros, M. Kiso, M.J. Blackman, S. Matthews, T. Feizi, D. Soldati-Favre, Members of a novel protein family containing microneme adhesive repeat domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites, J. Biol. Chem. 285 (2010) 2064–2076.
[37] J. Marchant, K. Sawmynaden, S. Saouros, P. Simpson, S. Matthews, Complete resonance assignment of the first and second apple domains of MIC4 from Toxoplasma gondii, using a new NMRView-based assignment aid, Biomol. NMR Assign. 2 (2008) 119–121.
[38] M.J. Wichroski, J.A. Melton, C.G. Donahue, R.K. Tweten, G.E. Ward, Clostridium septicum alpha-toxin is active against the parasitic protozoan Toxoplasma gondii and targets members of the SAG family of glycosylphosphatidylinositol-anchored surface proteins, Infect. Immun. 70 (2002) 4353–4361.
[39] T.M. Johnson, Z. Rajfur, K. Jacobson, C.J. Beckers, Immobilization of the type XIV myosin complex in Toxoplasma gondii, Mol. Biol. Cell 18 (2007) 3039–3046.

UNCORRECTED PROOF

Appendix II

In vitro studies of PfIMP2 isolated from *Plasmodium falciparum* 3D7 strain

A2.1 Culturing of parasites

The structure of PfIMP2 reveals a variety of interesting features about the protein which were previously not known. However, none of this gives us a definite indication regarding the function or the localisation of the protein in the parasite. *In vivo* studies were performed in Dr Jake Baum's lab. Parasites were only studied in the erythrocytic stages in this study.

A2.1.1 Transfection of parasites

The PfIMP2 gene sequence was cloned into pRM-2 vector for ectopic expression of PfIMP2 in *P. falciparum* 3D7 strain. The vector contains a C-terminal GFP and triple HA tag which are necessary for localisation and pull down assays. Transfection of 3D7 strain with this vector was performed by Dr. Chwen Ling Tay from the Baum lab (Imperial College London). The vector also contains a resistance gene to the WR99210 drug.

WR99210 is a drug that inhibits the activity of dihydrofolate reductase in the parasite. Dihydrofolate reductase is an enzyme which reduces dihydrofolate into tetrahydrofolate in a NADPH dependent manner. This step acts as a methyl group shuttle which plays a major role in de novo synthesis of certain amino acids, purines and thymidylic acid (Cody et al, 2004). Therefore, disruption in the activity of this enzyme would affect the survival of the parasite. Due to the pivotal role of this protein in the parasite, it is a drug target for antimalarial drugs. Resistance to these drugs is on the rise as this protein undergoes mutations which evade the drug without affecting enzyme activity. The structure of the *P. falciparum* dihydrofolate reductase/thymidylate synthase enzyme has been solved bound to WR99210 (Figure A2.1.1) (Cody et al, 2004).

Increase in the concentration of WR99210 in the media would lead to increased expression of the PfIMP2 protein. The optimal drug concentration for successful ectopic expression of PfIMP2 was found to be 20 nM by Dr. Chwen Ling Tay. Following successful transfection of the parasite, the transgenic line had to be maintained and synchronised regularly in order to achieve the desired parasitemia for the respective experiments.

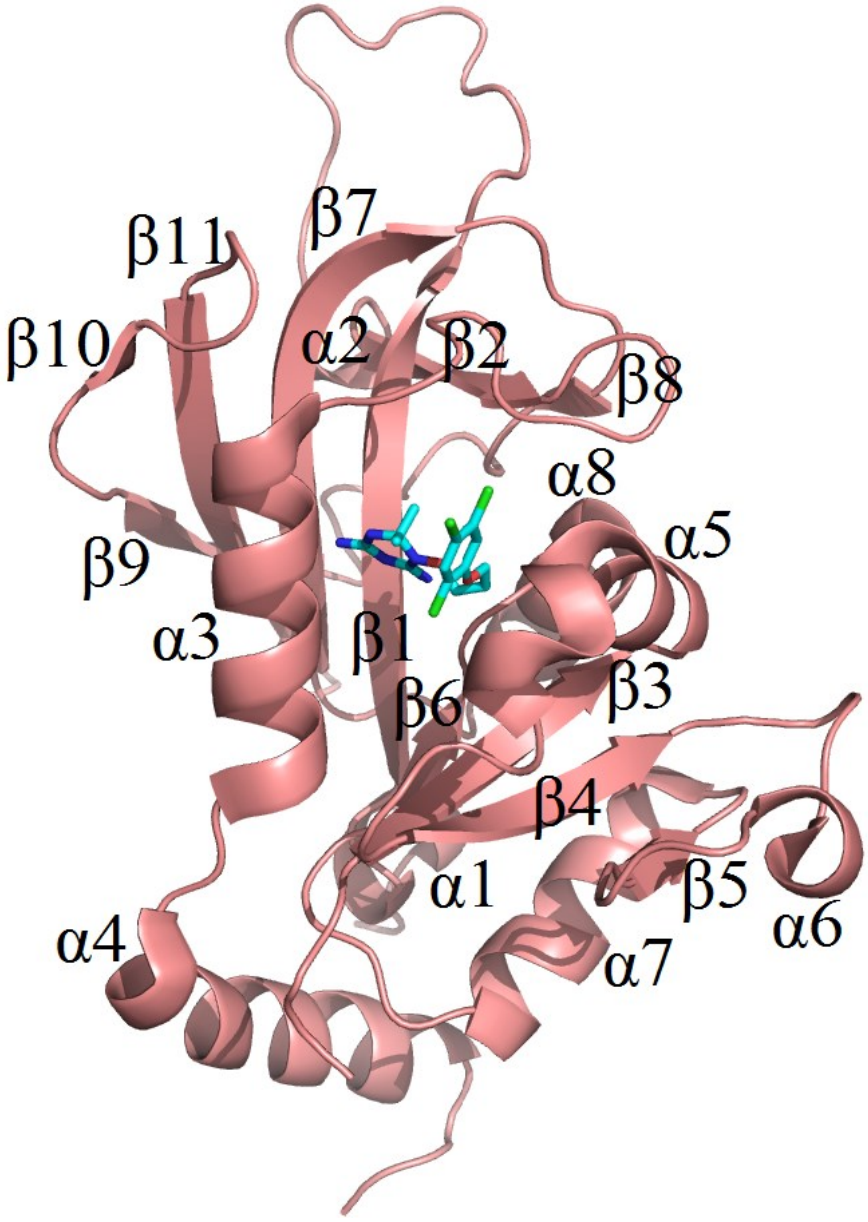


Figure A2.1.1: Figure showing a cartoon representation of the structure of the dihydrofolate reductase domain (pink) of P. falciparum dihydrofolate reductase-thymidylate synthase protein bound to WR99210. Secondary structure elements have been labelled. WR99210 is represented as sticks and carbon atoms are coloured cyan, nitrogen atoms coloured blue, chlorine atoms coloured green and oxygen atoms coloured red. (Cody et al, 2004)

A2.1.2 Sorbitol synchronisation

Maintaining the parasites is an important aspect of parasite culturing which involves changing the media regularly and diluting the culture to prevent them from crashing. Another important step in maintaining the parasites is keeping them synchronised. In order to study the parasites, all of the parasites in the culture need to be in the same stage in the erythrocytic cycle. In order to synchronise the parasites, the culture is regularly treated with 5% sorbitol when parasitemia of the ring stage parasites is high. Sorbitol synchronisation results in the survival of only ring stage parasites in the culture (Lambros & Vanderberg, 1979).

A2.1.3 Immunofluorescence Assay (IFA)

IFA is a commonly used technique that allows us to selectively visualise an antigen within the host organism. In this project, the antigens that were targeted for IFA were the GFP and triple HA tags present at the C-terminus of the PfIMP2 protein. Visualising the location of these tags will allow us to deduce the localisation of PfIMP2. There are two methods of IFA – direct and indirect (Figure A2.1.2). Direct IFA refers to the technique where an antibody with a bound chromophore is used to bind to the target antigen. Indirect IFA refers to the technique where a primary antibody is used to bind to the target antigen and a secondary antibody with a bound chromophore is used to bind to the primary antibody (Hoff, 2015). In this project, indirect IFA was used to deduce the localisation of PfIMP2 within the parasite.

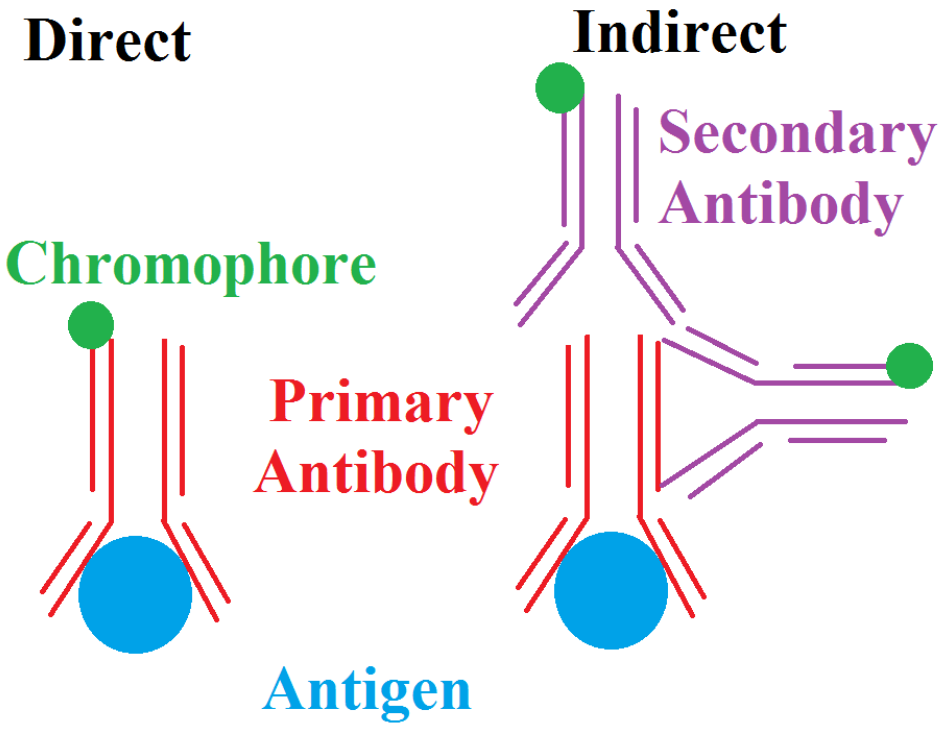


Figure A2.1.2: Figure showing a schematic depicting the direct (left) and indirect (right) IFA techniques. The various parts of the figure are labelled in a colour-coded manner.

A2.1.4 Confocal Microscopy

Once the IFA samples were prepared, they were visualised using a confocal microscope with a photon-multiplier tube. Confocal microscopy allows us to obtain higher resolution images than those obtained when using light microscopes. A laser is used to generate the light source in confocal microscopy which allows the generation of more than one wavelength of light, within the range of visible light, simultaneously. This allows the excitation of more than one chromophore in the sample simultaneously which reduces the exposure time of the sample (Hoff, 2015). However, the laser cannot produce UV radiation which is required to excite DAPI, which is a stain used to stain nuclei. DAPI excitation is usually done as a second step.

In light microscopy, the sample is flooded with light to illuminate the sample and all of the emitted light is detected. The disadvantage of this method is that the detector receives signal from the sample and from the background. The background signal drastically reduces image resolution.

Confocal microscopy relies on sample excitation at single points by the incident light. The emitted light from the sample then reaches the photon multiplier through an aperture. This aperture is positioned in such a way that only the emitted light originating from the focal plane of the sample reaches the photon multiplier and appears in the final image. This is referred to as in-focus emission light ray. The light emitted from the background does not pass through the aperture and does not appear in the final image. This is referred to as out-of-focus emission light ray (Figure A2.1.3). The elimination of background noise in confocal microscopy improves the resolution of the images. A photon multiplier tube is usually used as a detector to magnify the intensity of the signal, which in turn improves image resolution further (Hoff, 2015).

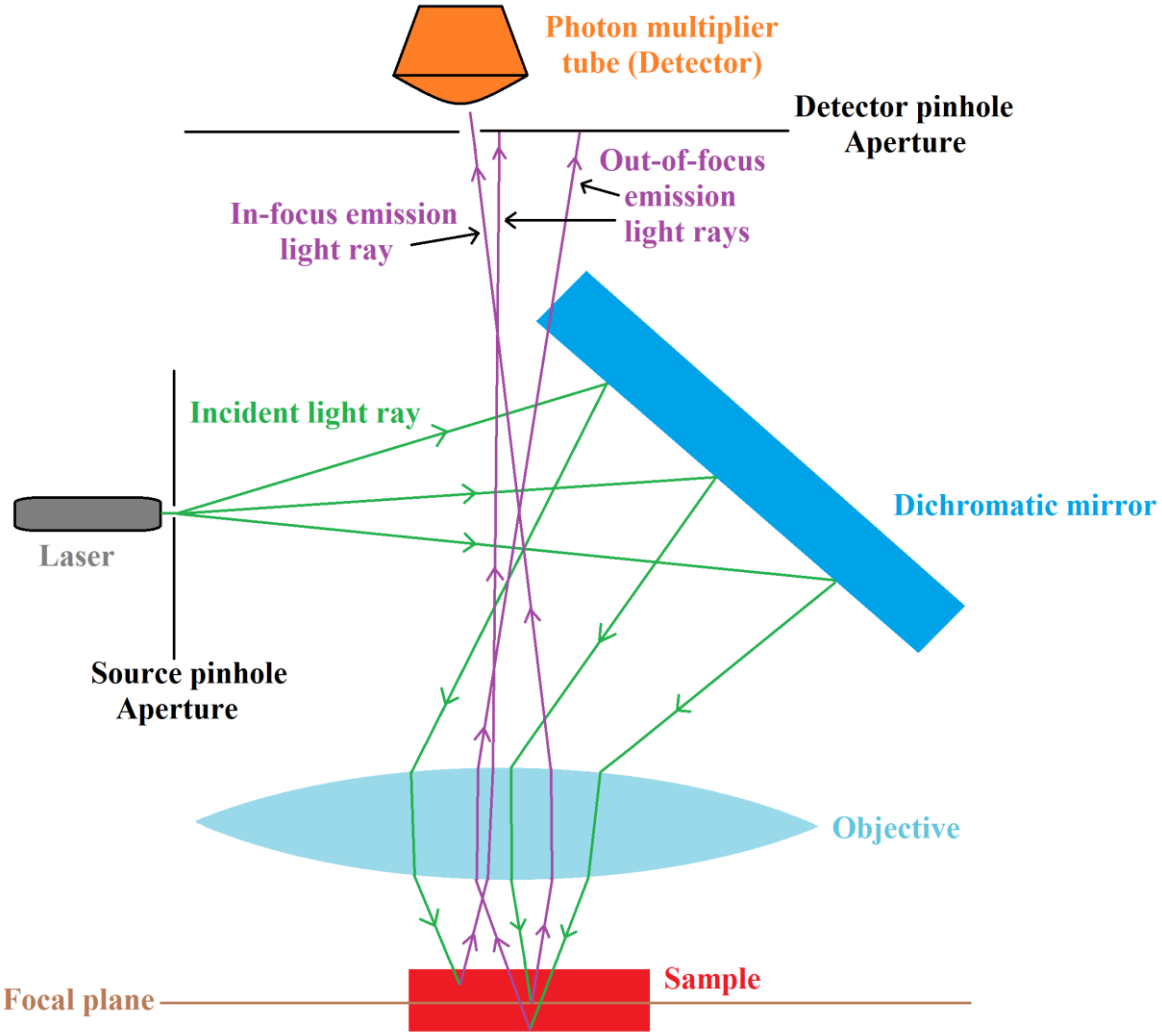


Figure A2.1.3: Figure showing a schematic representation of the principle behind confocal microscopy. The various components in the figure are labelled in a colour-coded manner.

A2.1.5 Pull down Assay

Pull down assay is a technique used to identify binding partners for a particular protein. In this technique, the protein of interest is tagged such that it can bind a column containing an affinity tag specific to that tag. In this project, anti-HA magnetic beads were used to bind the triple HA tag in the PfIMP2 construct. Following this, a protein source containing the ligand protein is passed through the column containing the bound tagged protein. The protein of interest is then eluted and any bound complex analysed. In this project, the *P. falciparum* 3D7 strain expressing PfIMP2 ectopically was lysed and the lysate was incubated with the anti-HA

magnetic beads to allow PfIMP2 to bind the beads. Following this, the protein was eluted and analysed on a SDS-PAGE gel. A schematic depicting the technique is shown in Figure A2.1.4.

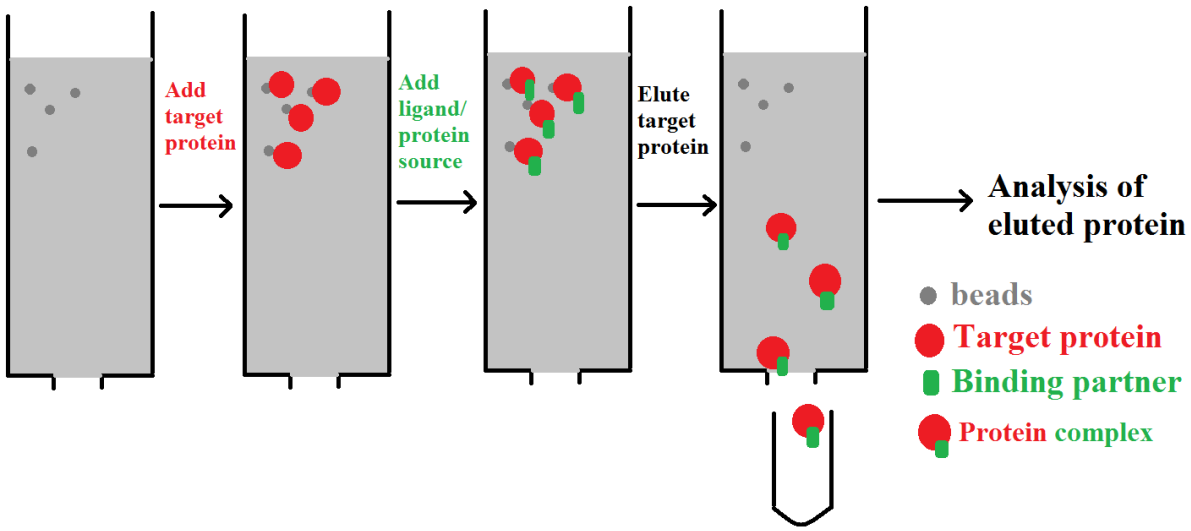


Figure A2.1.4: Figure showing a schematic representation of the steps involved in a pull down assay

A2.2 Localisation and pull down studies

A2.2.1 PfIMP2 expression profile

Previous work on PfIMP2 by the Derisi lab has shown that the expression of PfIMP2 is high around 8 hours and 42 hours post-erythrocytic invasion (Figure A2.2.1). This data suggests that PfIMP2 is expressed mainly in the schizont stage of the parasite life cycle (Bozdech et al, 2003; Llinás et al, 2006). This further suggests that PfIMP2 might play a role in erythrocyte invasion. Further insight about the localisation of the protein in the parasite and possible binding partners would allow us to identify the role of this protein in this parasite.

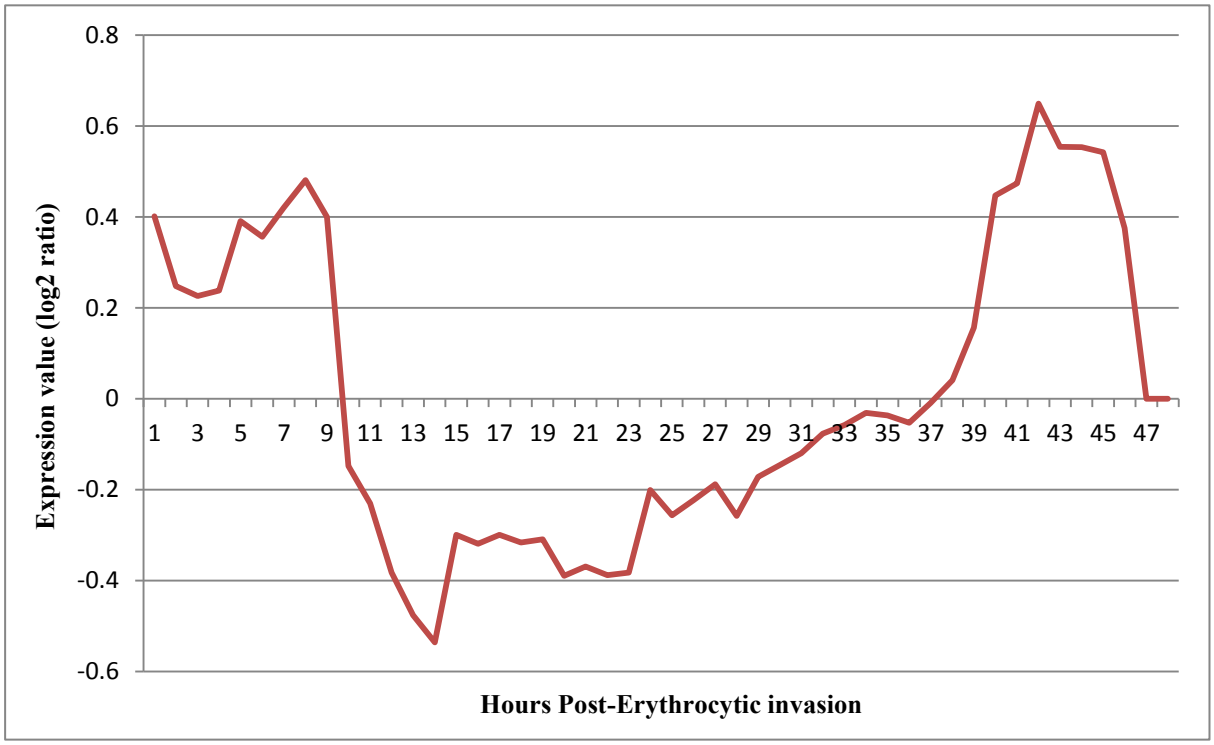


Figure A2.2.1: Figure showing the expression of PfIMP2 in P. falciparum 3D7 at various time points following erythrocytic invasion (Bozdech et al, 2003; Llinás et al, 2006)

A2.2.2 Localisation studies of PfIMP2

It was initially hypothesised that PfIMP2 might be localised to the IMC. In order to validate this hypothesis, IFA was used to study the localisation of PfIMP2 within *P. falciparum*. Indirect IFA was performed against the GFP and the triple HA tags in the ectopically expressed protein, using both fixed and suspended samples. The samples were also treated with anti-GAP45 antibody as GAP45 is known to localise to the IMC. However, fluorescence was not observed for PfIMP2 in the samples treated with the anti-GFP antibody. Both fixed and suspended samples treated with anti-HA antibodies showed fluorescence for PfIMP2 (Figure A2.2.2). The results obtained from this study were inconclusive on their own and further experiments are required to conclusively report the exact localisation of the protein.

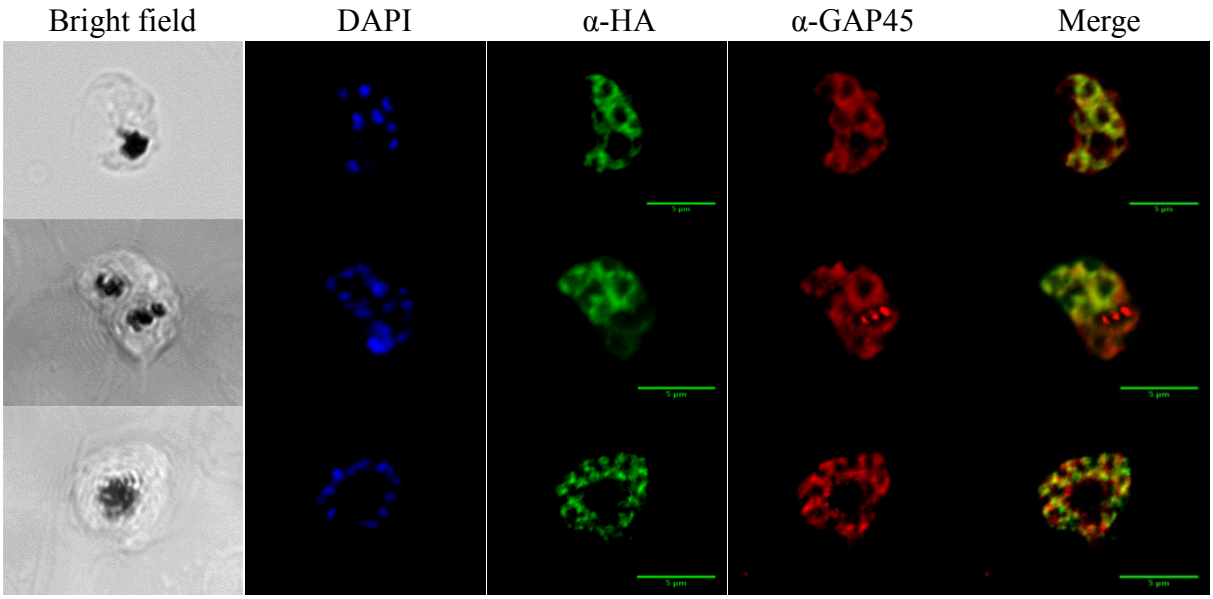


Figure A2.2.2: Figure showing (left to right) the bright field view, DAPI stain, anti-HA antibody, anti-GAP45 antibody and a merge of the anti-HA and anti-GAP45 fluorescence for three RBCs. A scale (5µm) is also shown on the anti-HA and merge windows. Data was recorded on a confocal microscope. Data and figure provided by Dr. Chwen Ling Tay (Baum lab).

A2.2.3 Identification of PfIMP2 binding partners

The structure of PfIMP2 reveals the presence of a hydrophobic groove which could potentially be involved in protein-protein interactions. A pull down assay was performed in order to identify potential binding partners of PfIMP2. A small scale pull down was performed, on a wild type *P. falciparum* 3D7 strain (control) and the strain expressing PfIMP2 ectopically, using anti-HA magnetic beads. SDS PAGE of the pull down samples was performed and the gel was silver stained (Figure A2.2.3). However, no notable difference is observed between the wild type and the strain expressing PfIMP2 ectopically. A band is expected around 48 kDa in the PfIMP2 strain, which would correspond to the ectopically over-expressed PfIMP2 protein. This band is not observed on the gel.

Western blot of the gel was also performed, against the triple HA tag, to further confirm the presence or absence of the HA tagged-PfIMP2 protein. Although the silver stained gel does not show any difference between the two samples, the western blot shows two faint bands around 50 kDa (shown by the black box) which are more intense in the PfIMP2 sample. These bands could correspond to overexpressed PfIMP2. However, the concentration of the overexpressed protein appears to be very low. It is also worth noting that the wild type contains additional bands which are absent in the PfIMP2 sample.

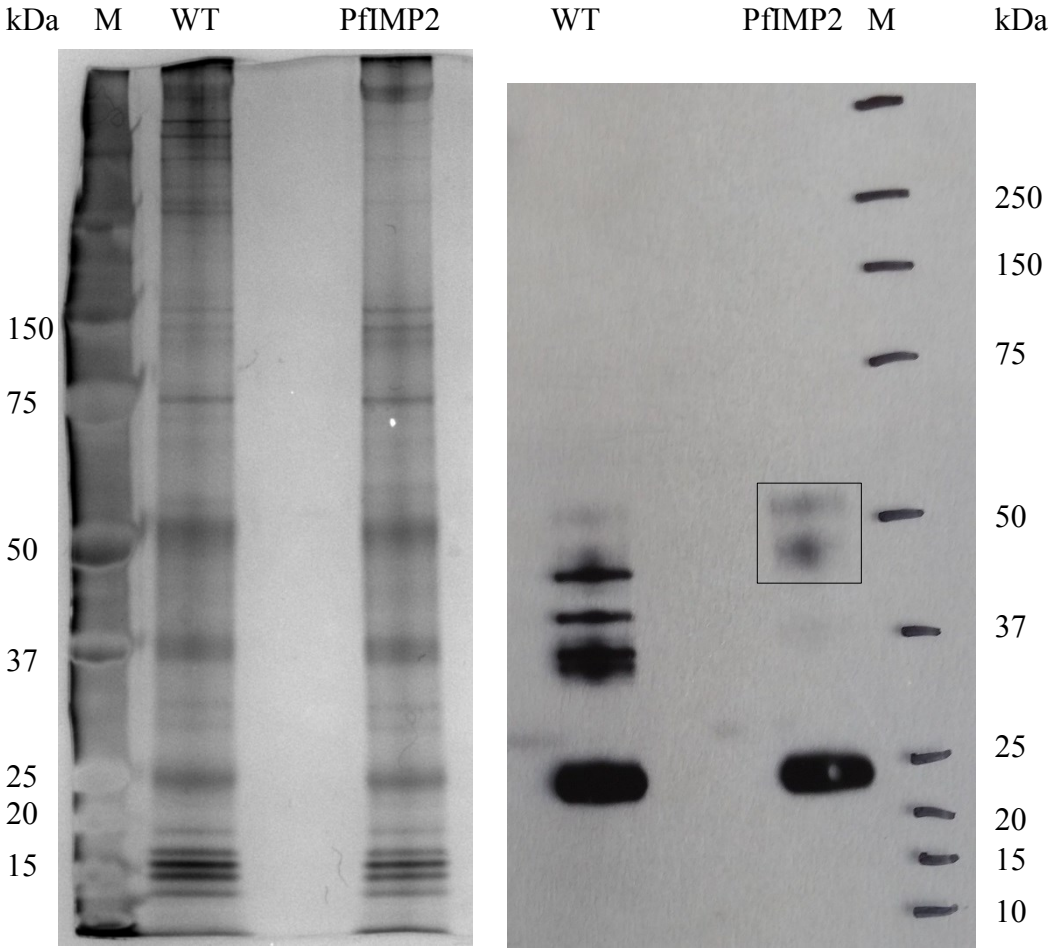


Figure A2.2.3: Figure showing the silver stained SDS-PAGE gel (left) and Western blot (right) of the pull down samples from the WT and PfIMP2 strains. No difference is observed between the WT and PfIMP2 strains in the silver stained gel. The Western blot shows two faint bands around 50 kDa which could correspond to the tagged PfIMP2 protein (shown by the black box). The lane labelled M corresponds to the protein molecular weight marker. The lane containing the wild type sample contains additional bands which are absent in the PfIMP2 sample.

A2.3 Summary

PfIMP2 was ectopically expressed as a fusion protein with triple HA and GFP tags at the C-terminus. Protein expression was controlled by altering the concentration of the WR99210 drug. The optimal concentration of WR99210 for successful overexpression of PfIMP2 was found to be 20 nM.

Parasites were grown to a high concentration and the parasitemia of the culture was monitored. Once the culture reached the desired parasitemia, samples were taken for IFA. IFA of PfIMP2 has given inconclusive results as the fluorescence obtained from PfIMP2 does not match the fluorescence from GAP45. Although this data is inconclusive, different markers can be used to identify the exact localisation of the protein in the parasite.

The structure of PfIMP2 revealed that the protein has a hydrophobic groove which could potentially act as a binding pocket for ligand binding. In order to test this hypothesis, a pull down assay was performed using a wild type *P. falciparum* 3D7 strain as a control. The wild type strain and the strain expressing PfIMP2 ectopically were grown simultaneously and once the required parasitemia was achieved, samples were taken for pull down assays from both cultures. However, the pull down assay also proved to be inconclusive. There is no difference between the result observed for the wild type and for the PfIMP2 strain. It is also worth noting that the PfIMP2 lane does not have any prominent band corresponding to the molecular weight of the ectopically overexpressed PfIMP2 construct (~48 kDa).

In order to check for the presence of the HA-tagged PfIMP2 protein, a western blot of the pull down samples was also performed. The western blot revealed a faint band in the PfIMP2 sample around 50 kDa which could correspond to the PfIMP2 protein. Although PfIMP2 is overexpressed, the concentration of the PfIMP2 protein appears to be much lower than expected. Due to the low concentration of the protein, it is not possible to observe any

further bands that might correspond to binding partners of this protein. Additional bands can be observed in the wild type sample which are absent in the PfIMP2 sample. The identities of these bands were not tested. Mass spectrometry would reveal the identities of these proteins and it is likely that they contain the HA amino acid sequence, which would have allowed them to bind the affinity column.

Western blot has revealed that the concentration of the overexpressed PfIMP2 protein was much lower than expected. The IFA results further confirm that the protein is expressed in the parasite. WR99210 was used at 20 nM concentration to drive protein overexpression. This concentration is high and elevating it further would affect the viability of the parasite.

Further IFA studies of PfIMP2 with different markers that localise in different compartments of the cell would allow us to conclusively determine PfIMP2 localisation within the parasite. Pull down of PfIMP2 could be repeated with a new transgenic line of *P. falciparum* 3D7 parasites with the tagged sequence of PfIMP2 present endogenously. This would also allow us to track the expression profile and localisation of the protein within the parasite by IFA.

The expression profile of PfIMP2 shows that the protein is expressed in the schizont stage. This suggests that PfIMP2 may be involved in invasion of red blood cells, in which case it is likely that the protein forms interactions with the host proteins. However, there is not enough evidence to support this hypothesis and further studies are necessary to confirm or disprove the same.

Appendix III

Table A3: Table showing recipes for buffers used throughout the project

Buffer Name	Ingredients	Company
TAE buffer	40 mM Tris pH7.6	Sigma-Aldrich, UK
	20 mM Acetic acid	
	1 mM EDTA	
1% Agarose gel	1% w/v Agarose power	Sigma-Aldrich, UK
	TAE buffer	
Phosphate Lysis buffer	50 mM NaH_2PO_4 pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	10 mM Imidazole	
	1 µg/ml PMSF	
Phosphate wash buffer	50 mM NaH_2PO_4 pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	20 mM Imidazole	
Phosphate elution buffer	50 mM NaH_2PO_4 pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	250 mM Imidazole	
Phosphate GF buffer	50 mM NaH_2PO_4 pH 8	Sigma-Aldrich, UK
	200 mM NaCl	
Phosphate NMR buffer	50 mM NaH_2PO_4 pH 8	
	150 mM NaCl	
Crystallisation buffer	50 mM Tris pH 8	Sigma-Aldrich, UK
	150 mM NaCl	
Chymotrypsin digestion buffer	100 mM Tris-HCl pH 8	Sigma-Aldrich, UK
	10 mM $CaCl_2$	
PVDF stain	0.1% Coomassie Blue R	Sigma-Aldrich, UK
	50% Methanol	
	1% acetic acid	
M9 Minimal media (per litre)	6 g Na_2HPO_4	Sigma-Aldrich, UK
	3 g KH_2PO_4	
	0.5 g NaCl	
	0.7 g $^{15}NH_4Cl$	
	pH adjusted to 7-7.4 and autoclaved	
	2 ml 1M $MgSO_4$	Sigma-Aldrich, UK
	10 µl 1M $CaCl_2$	
	2 g D-Glucose-$^{13}C_6$	
	1 ml Vitamins solution	
	1 ml micronutrients	
Micronutrients (1000x)	3 x 10^{-6}M Ammonium Molybdate	Sigma-Aldrich, UK
	4 x 10^{-4}M H_3BO_3	
	3 x 10^{-5}M $CoCl_2$	
	1 x 10^{-5}M $CuSO_4$	
	8 x 10^{-5}M $MnCl_2$	
	1 x 10^{-5}M $ZnSO_4$	
Vitamins (1000x)	1 g riboflavin	Sigma-Aldrich, UK
	1g niacinamide	

	1 g pyridoxine monohydrate	
	1 g thiamine	
Hepes lysis buffer	50 mM Hepes pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	10 mM Imidazole	
Hepes wash buffer	50 mM Hepes pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	20 mM Imidazole	
Hepes elution buffer	50 mM Hepes pH 8	Sigma-Aldrich, UK
	300 mM NaCl	
	250 mM Imidazole	
Hepes GF buffer	50 mM Hepes pH 7.5	Sigma-Aldrich, UK
	150 mM NaCl	
TEV buffer	50 mM Tris pH 8	Sigma-Aldrich, UK
	250 mM NaCl	
	10 mM DTT	
Complete media	5% Albumax	Sigma-Aldrich, UK
	7.5 % sodium bicarbonate	
	Incomplete media	
Incomplete media	RPMI	Sigma-Aldrich, UK
	30 mM Hepes pH 7.2	
	50 mg Hypoxanthine per litre	
	25 mg Gentamicin per litre	
	2 g Glucose per litre	
Parasite lysis buffer	1 mM EGTA	Sigma-Aldrich, UK
	0.2 M NaCl	
	20 mM Tris pH 7.5	
	5 mM $MgCl_2$	
	5% glycerol	
	2 mM DTT	
Parasite elution buffer	1 mM EGTA	Sigma-Aldrich, UK
	0.2 M NaCl	
	20 mM Tris pH 7.5	
	5 mM $MgCl_2$	
	5% glycerol	
	2 mM DTT	
	2 mg/ml Influenza Hemagglutinin peptide	