Structural and Functional Study of MACPF/CDC proteins



Tao Ni

Nuffield Department of Clinical Medicine

Division of Structural Biology

Magdalen College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

Abstract

The thesis mainly focuses on structural study of proteins in membrane attack complex-perforin/cholesterol-dependent cytolysins (MACPF/CDC) superfamily, in particular, Astrotactins from human and perforin-like proteins (PLPs) from *Toxoplasma* and *Plasmodium*. Both of these subfamily proteins are implicated in human diseases and structurally uncharacterised before.

Astrotactins (ASTNs) have been shown to play a crucial role in enabling neural migration along glial fibres. While ASTN1 directly forms contacts between the neuron and the fibre, ASTN2 is responsible for extracting ASTN1 from contacts at the lagging edge of the cell and recycling them to the leading edge. ASTN2 is associated with endosomes, with the majority of its structure (at the C-terminus) projecting into their lumen, anchored by a pair of transmembrane α-helices. Here we present the structure of this "endodomain" region of ASTN2, and find it to consist of a unique combination of polypeptide folds comprising a membrane attack complex/perforin (MACPF) domain, an EGF-like domain, a previously-unobserved form of fibronectin type III (Fn(III)) domain and an annexin-like domain. Taken together, the structural characterisation provides a framework for better understanding the mechanism of the ASTNs and related proteins in neural and other forms of vertebrate development.

Perforin-like proteins from *Toxoplasma* and *Plasmodium* (*Tg*PLP1 and *P*PLPs) are critical for normal life cycle progression of these parasites, and knockout out of any of them results in significant defects in their life cycle, entrapping the parasites within the host cells and thus limiting their ability to egress. Here we present the crystal structures of *Tg*PLP1 MACPF domain and C-terminal domain at 2.0 Å and 1.1 Å, respectively. We also presents the MACPF domain assembled in helical and hexameric ring form, which indicates the possibility of pore-formation by 6 subunits. This is the first structure of perforin-like protein from Apicomplexan parasites and provides a structural basis to elucidate the function of PLPs in toxoplasmosis and malaria pathogenesis.

The final section is a continuation of a previous study in our lab on pleckstrin homology (PH) domain of kindlins. We determined the crystal structure of kindlin-3 PH domain and characterised its lipid and membrane binding properties. Using nanodiscs incorporated with different lipids as model membrane system to study the interaction between inositol phosphate lipids and kindlin-3 PH domain, together

with molecular dynamic simulation studies (in collaboration with Mark Sansom's group, University of Oxford), we propose that a subset of PH domains is able to bind to multiple inositol phosphates simultaneously and so via an avidity effect have its interaction with target membranes strengthened.

Declaration of work

I hereby declare that this thesis is my own work and effort except in collaborative project, and that it has not been submitted anywhere for any award. The findings from collaborative project described here are clearly stated prior to the chapters. Where other sources of information have been used, they have been acknowledged.

signature
TAO NI (printed)

Acknowledgements

I will first thank my supervisor, Robert Gilbert, for his trust and guidance in every aspect of life and study in the last few years. I am mostly grateful for having been offered the opportunity to join Robert's group and have been enjoying the freedom to develop and test ideas. Thank Robert for offering the weekly tea and coffee during discussions for the last four years.

I also thank my lab members, especially Luke Yates for teaching me lab work skills during the early two years of my Dphil, and Ben Durrant, Antonio Biasutto, Monica Mantri, Kai Deng and Xiaoyan Wang for their assistance and stimulating discussions.

I would also like to thank all the members of the Division of Structural Biology, especially Tom Walter and Karl Harlos for their technical support in crystallisation and crystals manipulation, Jinshan Ren and Kamel Omari for sharing expertise in structure determination and model building, and Yuguang Zhao and Weixian Lu for assistance of tissue culture and protein expression. I also greatly appreciate the accompany of Hanwen Zhang next to me, Youxin Kong and Sai Li in the wetlab. More importantly, I am greateful for their encouragements and generosity.

Lastly, I thank my family, especially my father and mother for their deep love and support on me, and for their greatest efforts trying to earn a better life for their children even it is far beyond their scope, and for their life-long patience and expectations on us; I thank my elder brothers Haibo and Jiangfei for setting good models for me to pursue and thank my sister Yaning for bringing laughter to the whole family. I also thank my uncle Huanwen Chen for providing invaluable advices and supports whenever and wherever needed.

In the end, I sincerely thank my fiancée, Xiulian Yu for her persistent love and care; even being five hundred kilometres away across the British Channel. Her cheerful attitude towards life has always been inspiring me and it is my greatest honour to explore the world together with her.

Contents

Abstract i

Declaration of work iii

Acknowledgements iv

Contents v

List of Figures x

List of Tables xiii

Abbreviations xiv

1 Introduction 1

1.1 Summary . . . 2
1.2 General introduction to Pore-forming proteins . . . 3
1.3 MACPF/CDC superfamily proteins . . . 6
1.3.1 Distribution and general function of MACPF/CDC proteins . 7
1.3.2 Structure of Cholesterol-dependent cytolysins . . . 12
1.3.3 Structure of MACPF domain proteins . . . 15
1.3.4 Incomplete pore (toroidal pore) *vs* Complete pore . . . 19
1.3.5 Vertebrate-specific MACPF proteins . . . 22
1.4 Overview of the thesis . . . 29

2 Crystal Structure of ASTN2 endodomain 31

2.1 Summary . . . 32
2.2 Introduction . . . 33
2.2.1 Astrotacins in neuron development . . . 33
2.2.2 Astrotactins are linked with genetic diseases . . . 35
2.2.3 Domain and sequence analysis of Astrotactins . . . 36
2.3 Results and Discussion . . . 39
2.3.1 Overall description of the ASTN2 endodomain structure . . . 39
2.3.2 MACPF and EGF-like domain structures . . . 39
2.3.3 Structure of the MACPF domain alone at pH 5.5 and pH 4 . . 47
2.3.4 Fibronectin type III domain structure . . . 47
2.3.5 Details of the interface between the MACPF and Fn(III) domains 50

2.3.6 The annexin-like domain . . . 51
2.3.7 Inositol phosphate binding site . . . 55
2.4 Conclusion . . . 58
2.5 Methods and Materials . . . 59
2.5.1 Design of constructs, cloning and mutagenesis . . . 59
2.5.2 Protein expression and purification . . . 59
2.5.3 Crystallisation, data collection and processing . . . 60
2.5.4 Structure determination and refinement . . . 61
2.5.5 Surface plasmon resonance . . . 61
2.5.6 Thermofluor . . . 62
2.5.7 Small-angle X-ray scattering . . . 62
2.5.8 Analytical ultracentrifugation . . . 63
2.5.9 Structural phylogeny calculation . . . 64
2.5.10 Normal modes analysis . . . 64

3 Structure Determination of ASTN2 65
3.1 Summary . . . 66
3.2 Introduction . . . 67
3.3 Results and Discussion . . . 70
3.3.1 Selenium methionine SAD and sulphur SAD phasing trials . . . 70
3.3.2 Heavy-atom derivative screening and data analysis . . . 73
3.3.3 Data collection and analysis of platinum-derivative crystals . . . 74
3.3.4 SAD and SIRAS phasing with Pt derivative datasets . . . 75
3.3.5 Phase validation . . . 79
3.4 Conclusion and discussion . . . 81
3.5 Methods and Materials . . . 82
3.5.1 Cloning, expression and protein purification . . . 82
3.5.2 Protein crystallisation and optimisation . . . 82
3.5.3 Diffraction data collection and processing . . . 83

4 Functional and Biophysical Study of Astrotactins 84
4.1 Summary . . . 85
4.2 Introduction . . . 86
4.3 Results . . . 88
4.3.1 ASTN2 cytosolic domain is structurally ordered . . . 88
4.3.2 Complex of N-terminal domain and C-terminal regions through a disulphide bond . . . 89
4.3.3 pH-dependent Self-interaction of ASTN2 . . . 94
4.3.4 Controlling surface ASTN1 distribution by ASTN2 . . . 97

4.4 Discussion and Future Plan 97
4.4.1 ASTN2 cytosolic domain is structurally ordered 97
4.4.2 Interaction between ASTN1 and ASTN2 99
4.4.3 Self-interaction and oligomeric states of ASTN2 100
4.4.4 Linking ASTN2 with PCP pathway 101
4.5 Methods and Materials 101
4.5.1 Cloning, expression and purification of ASTN2 cytosolic domain 101
4.5.2 Purification of ASTN2 full-endodomain from HEK 293T cell lysate 102
4.5.3 Surface Plasmon Resonance 102
4.5.4 Circular dichroism 103
4.5.5 Flow cytometry 103

5 Structure of *Toxoplasma* Perforin-like Protein 1 105
5.1 Summary 106
5.2 Introduction 107
5.2.1 Brief introduction to *Toxoplasma* and *Plasmodium* 107
5.2.2 Brief introduction to the life cycle of *Toxoplasma gondii* and *Plasmodium falciparum* 109
5.2.3 Lytic cycle of *Toxoplasma gondii* 110
5.2.4 Implications of perforin-like proteins in *Toxoplasma* and *Plasmodium* life cycle 113
5.2.5 Domain architecture of perforin-like proteins from *Toxoplasma* and *Plasmodium* 115
5.3 Results 118
5.3.1 Overall structure of MACPF domain 118
5.3.2 Helical and ring structures formed by MACPF domain 123
5.3.3 Overall structure of the *Tg*PLP1 C-terminal domain 127
5.3.4 Solution structure of *Tg*PLP1 132
5.3.5 Proteolytic processing and oligomerisation of *Tg*PLP1 secreted into mammalian media 135
5.4 Discussion 138
5.4.1 Mechanism of pore formation by *Tg*PLP1 138
5.4.2 Proteolytic processing and dimerisation of *Tg*PLP1 140
5.4.3 Structure of MACPF/CDC superfamily proteins in pore-forming state at atomic resolution 141
5.5 Methods and materials 144
5.5.1 Small angle X-ray scattering 144

5.5.2 Analytical ultracentrifugation . . . 145
5.5.3 Reconstruction of MACPF domain helices in solution . . . 145

6 Protein Production and Structure Determination of ApiPLPs 146
6.1 Summary . . . 147
6.2 Expression screening of *Plasmodium* and *Toxoplasma* PLPs . . . 148
6.2.1 Expression and purification of PLPs in *E. coli* . . . 148
6.2.2 Protein expression and purification from yeast and mammalian cytosol . . . 149
6.2.3 Expression screening with mammalian secretion system . . . 151
6.2.4 Summary of ApiPLPs expression and purification . . . 152
6.3 Crystallisation of *Tg*PLP1 MACPF domain and C-terminal domain . 155
6.3.1 Initial crystallisation trials of *Tg*PLP1$_{462\text{-}1072}$ (MACPF domain and C-terminal domain) . . . 155
6.3.2 Crystallisation and structure determination of *Tg*PLP1 C-terminal domain . . . 159
6.3.3 Crystallisation and structure determination of *Tg*PLP1 MACPF domain . . . 161

7 Structural and Biophysical Study of Kindlin-3 PH Domain 170
7.1 Summary . . . 171
7.2 Introduction . . . 173
7.3 Results . . . 177
7.3.1 Overall structure of Kindlin-3 PH domain . . . 177
7.3.2 Molecular dynamics simulations reveal the membrane-bound state of kindlin-3 PH domain and its interaction with PIP lipids 182
7.3.3 Free energy of binding of the kindlin-3 PH domain to PIP(4,5)P_2 and PIP(3,4,5)P_3 . . . 191
7.3.4 SPR study of the kindlin-3 PH domain to multiple Phosphatidylinositol Phosphates . . . 192
7.3.5 Grp1 PH domain binds inositol phosphate with 1:1 specificity 197
7.4 Conclusion and perspective . . . 201
7.5 Methods and Materials . . . 205
7.5.1 Constructs design, cloning and mutagenesis . . . 205
7.5.2 Expression of kindlin-3 PH domain and Grp1 PH domain . . . 205
7.5.3 Protein purification and crystallisation . . . 206
7.5.4 Data collection, processing and structure determination . . . 206
7.5.5 Nanodiscs preparation . . . 207
7.5.6 Surface Plasmon Resonance . . . 208

7.5.7 Coarse-grained molecular dynamics (CG-MD) simulations . . 208
7.5.8 Potential of mean force (PMF) simulation 209

8 Conclusions and Perspectives **211**
8.1 Overview of Structural Biology 212
8.2 Conclusions and Future plans 214
8.2.1 Linking Astrotactins back to context 214
8.2.2 Pore formation by Apicomplexan Perforin-like proteins 215
8.2.3 Membrane binding property of Pleckstrin homology domains . 216
8.3 Final remarks 216

9 Bibliography **218**

List of Figures

1.1 Schematic view of pore formation by pore-forming toxins (PFTs) . . . 3
1.2 Structural architectures and mechanisms of pore formation by PFTs . 5
1.3 Sequence alignment and conservation of MACPF proteins 8
1.4 Distribution of Cholesterol-dependent cytolysins 11
1.5 Structure of CDCs and mechanism of pore formation 13
1.6 Structure of MACPF proteins and mechanism of pore formation . . . 16
1.7 Disulphide bond locking perforin-1 18
1.8 Incomplete pore (arcs) model 20
1.9 Schematic representation of Human MACPF proteins highlighting their domain architectures 25
1.10 Sequence alignment of human MACPF domains 27
2.1 Neuron migration and cycling model of Astrotactins 34
2.2 Phylogenetic tree of ASTN2 among vertebrate species 38
2.3 Overall structure of ASTN2 endodomain 40
2.4 SDS-PAGE analysis of ASTN2 crystals 41
2.5 Structural phylogeny of known MACPF domain structures 44
2.6 Comparison of MACPF domain and EGF-like domain 46
2.7 Normal modes analysis of the ASTN2 endodomain structure and ASTN2 at pH 5.5 and 4 48
2.8 The N-linked glycosylation at N732 49
2.9 Close-up of the MACPF domain-Fn(III) domain interface 51
2.10 Analytical ultracentrifugation (AUC) and SAXS analysis of ASTN2 . 52
2.11 Annexin-like domain homology and calcium binding 53
2.12 Superimposition of ASTN2 annexin-like domain and surface charge distribution comparison 54
2.13 Inositol phosphate binding to ASTN2 56
2.14 SPR experiments of ASTN1 and ASTN2 interactions with inositol phosphate species 57
3.1 ASTN2 selenium methionine incorporated crystal and fluorescent scan 70
3.2 Data analysis of SeMet-SAD, S-SAD and initial K_2PtCl_4 derivative screen 72
3.3 Comparison of Pt-derivative datasets for phasing 76
3.4 Superimposition of four copies of ASTN2 from one asymmetric unit . 78
3.5 Comparison of anomalous difference maps between Pt2 and Pt3 datasets 79
3.6 Phase validation with selenomethionine derivative dataset 80

4.1 Schematic diagram of planar cell polarity pathway 86
4.2 ASTN2 cytosolic domain 90
4.3 Biophysical characterisation of ASTN2 cytosolic domain 91
4.4 The endosomal domains are linked together through a disulphide bond 93
4.5 pH-dependent self-interaction of ASTN2 endodomain 96
4.6 Flow cytometry study showing that ASTN2 controls ASTN1 surface exposure 98
5.1 Diagram of *Toxoplasma* and *Plasmodium* 108
5.2 Lytic cycle of *Toxoplasma* and invasion process of *Toxoplasma* 112
5.3 Domain architecture of ApiPLPs 116
5.4 Overall structure of MACPF domain of *Tg*PLP1 119
5.5 Comparison of MACPF domains between *Tg*PLP1 and perforin-1 . . 121
5.6 Conservation of MACPF domain form ApiPLPs and identification of a conserved GGX_nW motif 122
5.7 Sequence alignment of MACPF domains from *Plasmodium* and *Toxoplasma* 124
5.8 Helical and ring arrangement of the *Tg*PLP1 MACPF domain 125
5.9 Oligomerisation of *Tg*PLP1 MACPF domain in DDM 127
5.10 Overall structure of *Tg*PLP1 C-terminal domain 128
5.11 Sequences alignment of C-terminal domain 129
5.12 Hydrophobic core of C-terminal domain and ion binding analysis . . . 131
5.13 AUC and SAXS study of *Tg*PLP1(MACPF-CTD) 132
5.14 SAXS modelling of *Tg*PLP1 (MACPF-CTD) structure 134
5.15 Expression and purification of *Tg*PLP1 constructs 136
6.1 Representative gels and gel filtration profiles showing the expression and purification of Apicomplexan PLPs from *E. coli* 150
6.2 Expression test of *Pb*PLP and *Pf*PLP constructs in a mammalian secretion system 153
6.3 Purification of $TgPLP1_{462\text{-}1072}$ (MACPF-CTD) and deglycosylation sites determination 156
6.4 Surface entropy reduction mutant of $TgPLP1_{462\text{-}1072}$ (MACPF-CTD) construct 158
6.5 Characterisation of $TgPLP1_{462\text{-}1072}$ crystals in SDS-PAGE 159
6.6 Expression, purification and crystallisation of *Tg*PLP1 C-terminal domain 160
6.7 Expression screening, purification and crystallisation of *Tg*PLP1 MACPF domain 165

6.8 *Tg*PLP1 MACPF domain crystals subject to limited proteolysis and electron density from SAD phasing . . . 167
7.1 Overall structure of kindlin-3 PH domain . . . 178
7.2 Phylogenic tree of representative PH domains from crystallography study . . . 180
7.3 Comparison of the canonical and non-canonical inositol phosphate binding sites of PH domain . . . 181
7.4 Hydrogen bond between Arg^{360} and Gln^{396} in simulation . . . 183
7.5 MD simulations of kindlin-3 PH interacting with PIP-incorporated lipids bilayer . . . 185
7.6 Contacts between Kindlin-3 PH domain and PIPs . . . 186
7.7 Orientation of kindlin-3 PH domain on the membrane bilayer . . . 189
7.8 Potential of mean force (PMF) calculations of Kindlin-3 PH domain with PIPs . . . 190
7.9 SPR study of kindlin-3 PH domain with immobilised PIP2 and PIP3 headgroups . . . 193
7.10 Kindlin-3 PH domain mutant dissociates from membrane during MD simulations . . . 194
7.11 Preparation and characterisation of nanodiscs . . . 195
7.12 SPR study of kindlin-3 PH domain with PtdIns(3,4,5)P_3 incorporated nanodiscs . . . 196
7.13 SPR study of interaction between Grp1PH domain and PtdIns(3,4,5)P_3 in two experimental setups . . . 198
7.14 Grp1 PH domain binding to PtdIns(3,5)P_2 . . . 199
7.15 Proposed model for lipid clustering in integrin activation . . . 203

List of Tables

1.1 List of Cholesterol-dependent cytolysins (selected) 10
1.2 List of members of vertebrate-specific MACPF subclass 23
2.1 Data collection of ASTN2 native crystals 42
2.2 Refinement statistics of ASTN2 crystals 42
3.1 Data collection for SeMet- SAD, S-SAD and initial Pt-SAD 71
3.2 Platinum-derivative data collection 75
6.1 Data collection of *Tg*PLP1 C-terminal 162
6.2 Refinement of *Tg*PLP1 C-terminal domain 163
6.3 Data collection and refinement of *Tg*PLP1 MACPF domain 168
6.4 Refinement of *Tg*PLP1 MACPF domain 169
7.1 Data collection statistics of kindlin-3 PH domain 179
7.2 Refinement statistics of kindlin-3 PH domain 179
7.3 Summary of simulations 188

Abbreviations

ApiPLPs	Apicomplexan Perforin-like Proteins
ASTN1/2	Astrotactin 1/2
ASU	Asymmetric Unit
AUC	Analytical Ultracentrifugation
BRINP1/2/3	Bone Morphogenetic Protein and Retinoic Acid Inducible Neural-specific Protein 1/2/3
C8α	Complememt Component 8 alpha subunit
C8β	Complememt Component 8 beta subunit
C9	Complememt Component 9
CDC	Cholesterol Dependent Cytolysin
CTD	C-Terminal Domain
FN(III)	Fibronectin Type III
Ins(3,4,5)P_3	Inositol 3,4,5-trisphosphate
Kd	Dissociation Constant
MACPF	Membrane Attack Complex Perforin Family
M6P	Maltose-6-phosphate
MD	Molecular Dynamics
MR	Molecular Replacement
NCS	Non-Crystallographic Symmetry
PCP	Planar Cell Polarity
***Pf*PLP1**	*Plasmodium falciparum* Perforin-like Protien 1
PFTs	Pore Forming Toxins
PH domain	Pleckstrin Homology Domain
PIP2	Phosphatidylinositol 4,5-bisphosphate
PIP3	Phosphatidylinositol 3,4,5-trisphosphate

PMF	Potential of Mean Force
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
***P*PLPs**	*Plasmodium* Perforin-like Proteins
PtdIns(3,4,5)P_3	Phosphatidylinositol 3,4,5-trisphosphate
PtdIns(4,5)P_2	Phosphatidylinositol 4,5-bisphosphate
***Pv*PLP1**	*Plasmodium vivax* Perforin-like Protein 1
RMSD	Root Mean Square Deviation
SAD	Single-wavelength Anomalous Dispersion
SAXS	Small-angle X-ray scattering
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
SPR	Surface Plasmon Resonance
***Tg*PLP1**	*Toxoplasma gondii* Perforin-like Protein 1
T_m	Melting Temperature

CHAPTER 1

Introduction

1.1 Summary

MACPF/CDC superfamily proteins have been identified in all forms of cellular life except, currently, *Archaebacteria* (Anderluh et al., 2014). They represent a sub-branch of the largest known family of pore-forming proteins/toxins, the membrane attack complex-perforin/cholesterol-dependent cytolysin (MACPF/CDC) family (Anderluh et al., 2014; Rosado et al., 2008). This chapter aims to present an overview of MACPF/CDC proteins. It starts with a brief overview of pore-forming proteins characterised to date, followed by introduction to the structural and functional study of cholesterol-dependent cytolysins (CDCs) and membrane attack complex/ perforin (MACPF) family proteins. Then it briefly discusses the mechanisms of pore formation by MACPF/CDC superfamily proteins. Subsequently, it briefly introduces our current knowledge on MACPF/CDC proteins in mammals, focusing on those not well-studied to date and that have not been observed to form a pore. Finally, the chapter will present an outline of the thesis as a whole.

1.2 General introduction to Pore-forming proteins

Pore-forming toxins (PFTs) are virulence factors produced by many pathogenic bacteria that have long been a fascinating family of proteins for structural biologists, microbiologists and immunologists. The most striking aspect of PFTs is that they initially fold into a water-soluble structure in solution and undergo dramatic conformational change to insert into membranes after attachment to their surface (**Figure 1.1**). PFTs are classified into two large groups (α-PFTs and β-PFTs) based on the secondary structures of the membrane-spanning elements (α-helices in α-PFTs and β-sheets in β-PFTs). Typically, pore formation can be viewed in three steps: (i) membrane binding: most PFTs require some specific receptor in order to attach

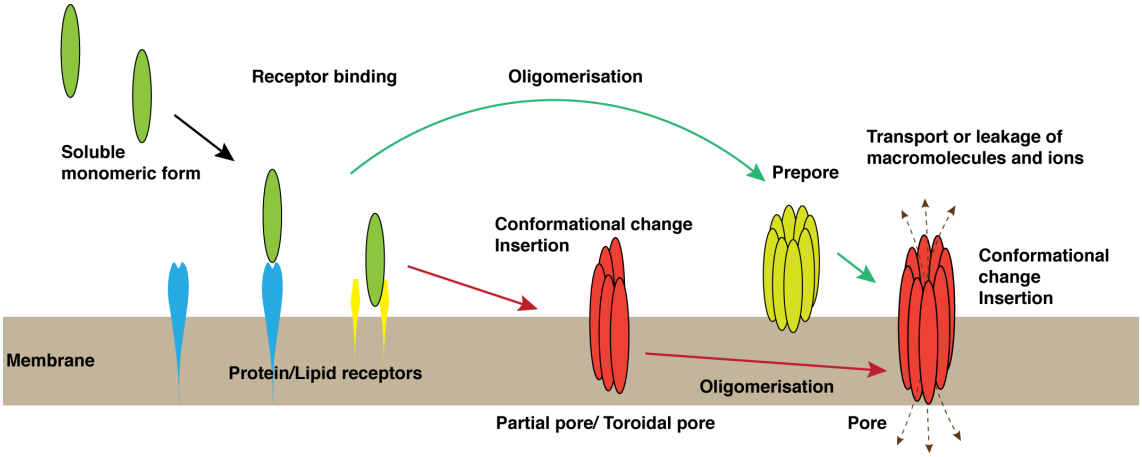

Figure 1.1: Schematic view of pore formation by pore-forming toxins (PFTs). PFTs are generally present in solution as monomers and become associated with a target membrane via protein receptors or lipid receptors (for example, sphingomyelin for actinporins and lysenin, and cholesterol for cholesterol-dependent cytolysins (CDCs)). Upon membrane binding, the toxins are concentrated to the membrane plane and initiate an oligomerisation process, which usually follows one of two pathways (green arrows and red arrows). Most β-PFTs oligomerise at the membrane surface first, resulting in an intermediate state known as prepore (green arrows). The prepore then undergoes conformational rearrangements that lead to concerted membrane insertion. In the pathway followed by most α-PFTs, PFT insertion into the membrane occurs simultaneously to sequential oligomerisation, which may lead to the formation of a partial pore or complete pore. Some CDCs form intermediate functional pores, presumably in a similar mechanism. The resulting pores from different PFTs differ significantly in architecture, stoichiometry, size and conduction features. The functions of pores are therefore distinct from each other.

to the target membrane; the receptors could be either membrane components (cholesterol for cholesterol-dependent cytolysins) or proteins that are anchored to/inserted in the membrane; (ii) oligomerisation (intermediate prepore): PFTs are concentrated on the membrane plane and form a prepore; (iii) membrane insertion and pore-formation: the prepore undergoes a conformational change and inserts into the membrane (**Figure 1.1**).

There are at least six families of PFT identified, with three α-PFTs and three β-PFTs. The overall structures of PFTs and their pore-forming activity have been thoroughly reviewed (Gilbert et al., 2014; Peraro and van der Goot, 2016). Here the diverse pore architecture and membrane insertion mechanisms are described, providing an overview about the structure of PFTs and their mechanisms of pore formation (**Figure 1.2**). α-PFTs generally preserve some α-helical structures in their soluble monomeric form and these α-helical motifs detach from the rest of the monomeric form to insert into the membrane, forming transmembrane structures formed by α-helices (**Figure 1.2A**). In contrast, β-PFTs are more diverse in terms of the motifs responsible for conformational change and membrane insertion. These motifs can either be β-strands, α-helices, loops or a combination of all three (**Figure 1.2B**). MACPF/CDC superfamily proteins represent one family of β-PFTs, characterised by their capacity to form pores with much larger diameters compared with other PFTs (**Figure 1.2B**).

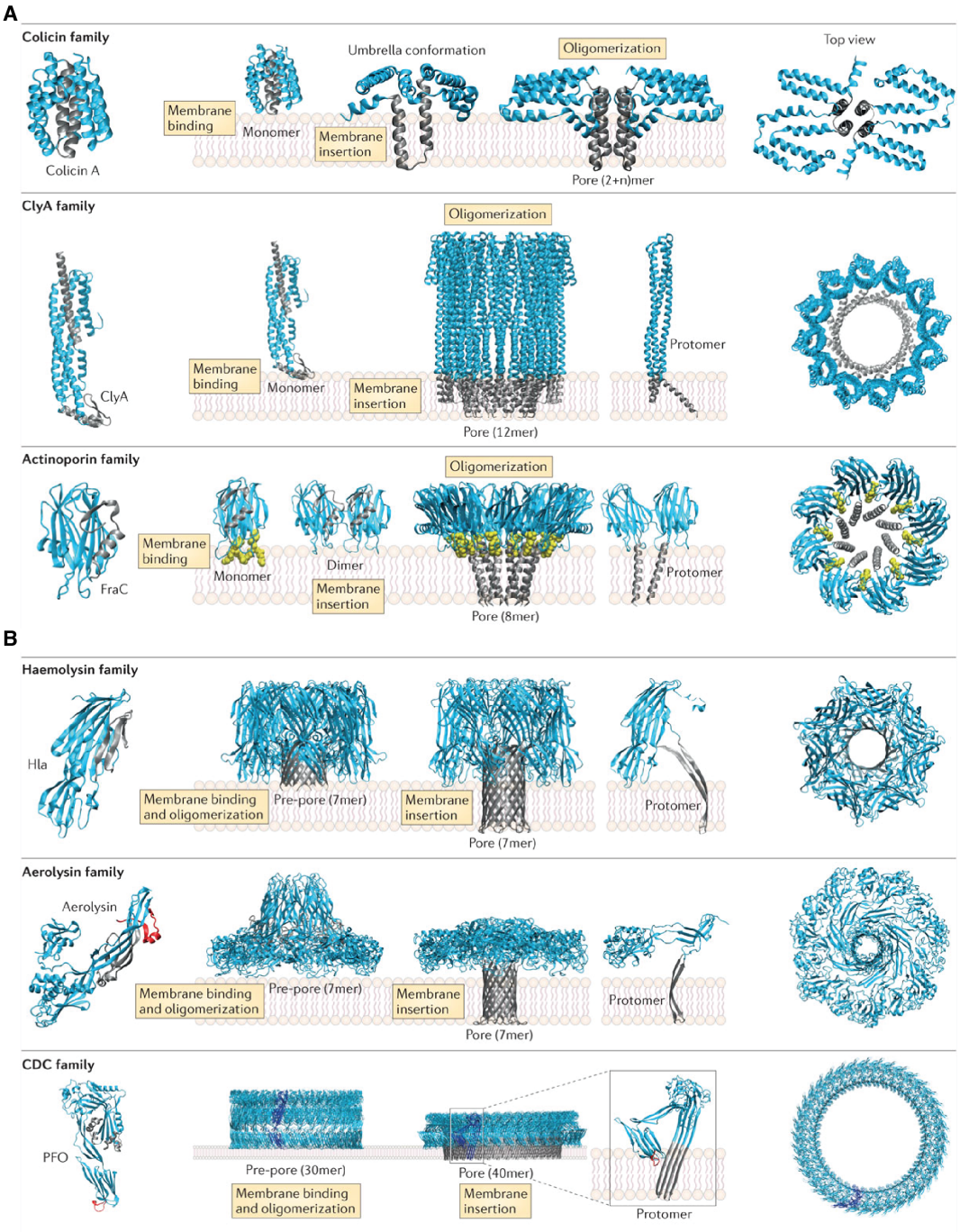

Figure 1.2: Structural architectures and mechanisms of pore formation by PFTs. The mechanism of pore formation by representative members of six PFTs. The structure motifs undergoing large conformational transitions are highlighted in grey. (A) Representative structures of monomeric soluble form and oligomeric pore form of α-PFTs. Three subfamilies of α-PFTs are shown here: Colicins, ClyA and Actinporins. *figure legend continued on next page.*

Figure 1.2 (*continued*) (i) Colicins: Colicin A structure (PDB entry 1COL) and proposed pore forming states; (ii) ClyA family: ClyA structure both in soluble monomeric form (PDB entry 1QOY) and mature dodecameric pore form (PDB entry 2WCD);(iii) Actinporin family: Structure of fragaceatoxin C (PDB entry FraC) in a lipid-bound monomeric form (PDB entry 4TSL), a dimeric intermediate (PDB entry 4TSN) and the mature octameric pore form (PDB entry 4TSY).(B) Representative structures of monomeric soluble forms and oligomeric pores for β-PFTs. Three subfamilies of β-PFTs are presented: Hemolysin, Aerolysin and Cholesterol-dependent cytolysin (CDCs). (i)Hemolysin family: α-Hemolysin structure in momomeric form (PDB entry 4IDJ) and heptameric oligomeric pore form (PDB entry 7AHL); (ii) Aerolysin family: soluble monomeric form (PDB entry 1PRE) and proposed oligomeric prepore and pore forms (based on 2d crystallography and computational modelling); Lysenin also belongs to the Aerolysin family and the structures of both monomeric and pore forms of lysenin have been determined (PDB entries 3ZXD and 5GAQ, 5EC5) - these suggest a structure for the pore somewhat different from that shown above for Aerolysin. (iii) CDCs: Structures of Perfringolysin (PFO) in soluble monomeric form, intermediate prepore and mature pore forms (proposed based on cryo-EM and biochemical characterisation). Note that pores formed by CDCs are significantly larger than others. Figure adapted from Peraro MD and van der Goot FG, 2016.

1.3 MACPF/CDC superfamily proteins

MACPF (membrane attack complex/perforin) and CDC (cholesterol-dependent cytolysin) family proteins were identified and studied separately until structures of a representative from each family were resolved. The solution of the structures of complement C8α and a bacterial MACPF protein revealed that MACPFs and CDCs belong to one homologous family of proteins (Hadders et al., 2007; Rosado et al., 2007). The only basis for MACPF/CDC activity properly established so far is the oligomerisation of many subunits in order to generate a pore-forming complex (Gilbert et al., 1999; Law et al., 2010; Metkar et al., 2015; Serna et al., 2016; Tilley et al., 2005). Binding to a target membrane serves the basic function of concentrating monomeric subunits on a planar substrate (Gilbert, 2016). Then, concentration-dependent oligomerisation occurs into a pre-pore complex before pore formation ensues (Gilbert et al., 2013; Tweten et al., 2015). However, it is worth to note that pore formation is not the only function of MACPF/CDC domain proteins.

In this section we firstly introduce the distribution of MACPF/CDC proteins and then give an overview about mechanisms of pore formation by MACPFs and CDCs, respectively. Subsequently, we will briefly discuss the toroidal pores formed by MACPF/CDC superfamily proteins, particularly CDCs, which have been for a long time underappreciated. Finally, the structures of MACPF proteins in vertebrates, and especially in humans, are reviewed, aiming to present the function diversity of MACPF proteins: they are not only involved in pore formation, but also probably a more fundamental role in development.

1.3.1 Distribution and general function of MACPF/CDC proteins

The MACPF-containing proteins are widespread in nature and can be found in all domains of life, especially metazoan. There are 1271 MACPF-containing proteins in SMART database, most of which are from metazoan (http://smart.embl-heidelberg.de/smart/). The distribution and conservation of MACPF and CDC proteins has reviewed thoroughly by Anderluh et al (Anderluh et al., 2014) and it has been shown that MACPF-containing proteins are involved in various aspects of biological processes. There is only a limited sequence conservation among MACPF family proteins, with less than 20% sequence identity (**Figure 1.3A**). Nevertheless, a signature motif was identified: Y(F)-G-T-H-X_7-G-G (**Figure 1.3B**). Here we will briefly discuss several key examples of MACPF/CDC superfamily proteins from different organisms.

There are 12 MACPF-containing proteins in the human genome classified into 4 groups: complement components, perforins, Astrotactins and BRINPs (**B**one morphogenetic protein and **R**etinoic acid **I**nducible **N**eural-specific **P**rotein), in which the first two groups are immune system proteins (pore formation) and the other two are involved in development (further details are given in the next section). In *Drosophila melanogaster*, Torso-like was firstly identified as associated with the localised activation of TOR (Target-of-rapamycin) (Savant-Bhonsale and Montell,

1993), and more recently it was shown that torso-like is translocated to the vitelline membrane space, controlling the polarisation of embryogenesis (Johnson et al., 2015; Mineo et al., 2015; Savant-Bhonsale and Montell, 1993). It is unknown which cellular pathway torso-like is involved in and its specific function is still unclear.

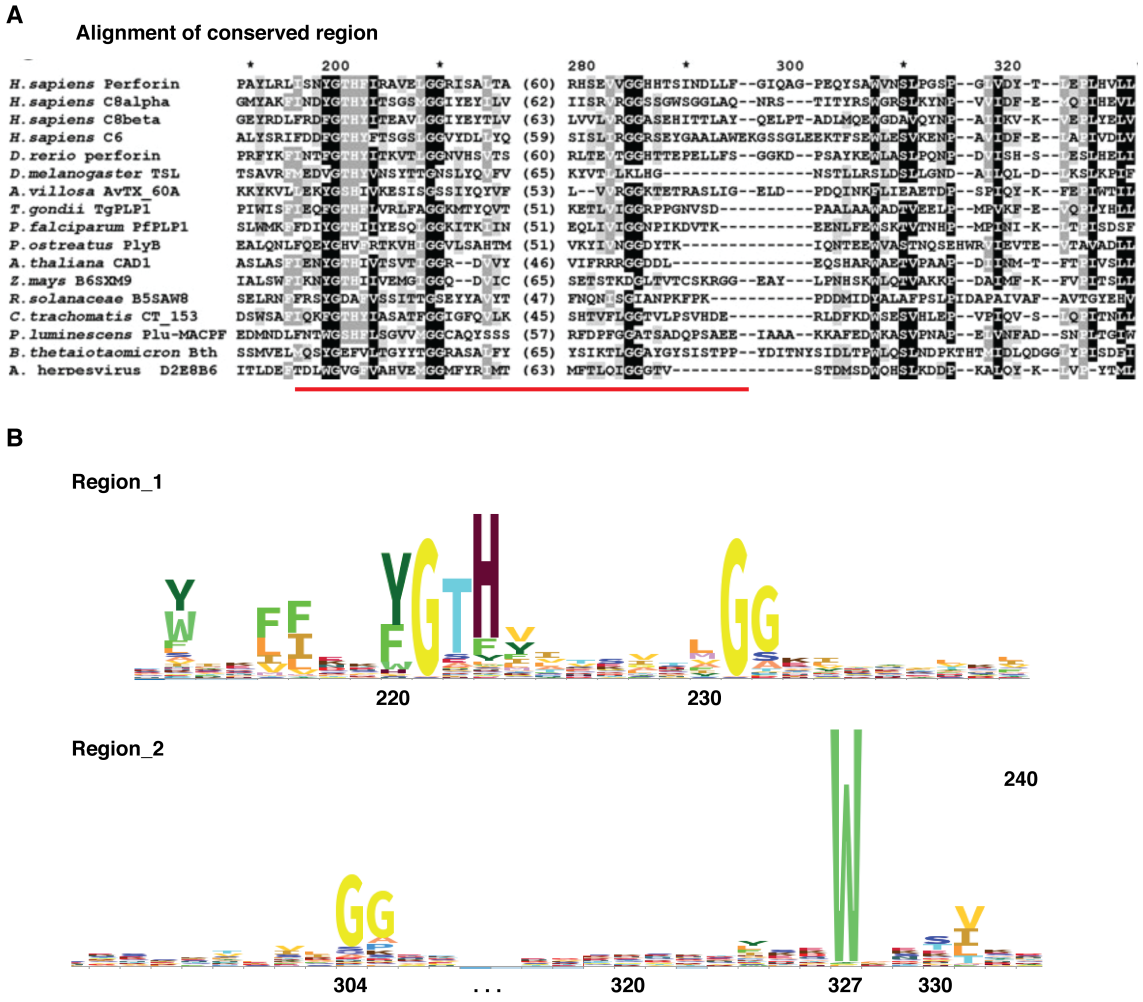

Figure 1.3: Sequence alignment and conservation of MACPF proteins. (A) Sequence alignment of the most conserved region in the MACPF domain of various representative proteins. The number in brackets indicates the length of the segments that are not conserved. The numbers on the right are according to the mature protein sequences. Figure adapted from Gilbert et al., 2013. (B) HMM logo representation of the MACPF signature motif. Region_1 corresponds to the region of sequences above the red bold line in (A). Region_2 highlights an extremely conserved tryptophan residue. This HMM logo was downloaded from the Pfam website (http://pfam.xfam.org/family/) and the numbering below the motifs is according to perforin-1 (Uniprot P14222).

MACPF proteins are abundant in Fungi as well, although only a little is known about their function. SpoC1-C1C from *Aspergillus nidulans* was shown to function

as a conidium-specific member of the SpoC1 gene cluster (Stephens et al., 1999). Pleurotolysin, a MACPF protein system consisting of two components, PlyA and PlyB, from *Pleurotus ostreatus* is the best-characterised fungal MACPF and forms a functional pore (Lukoyanova et al., 2015).

In the model plant *Arabidopsis thalianna*, four MACPF proteins have been identified in the genome and two of them were shown to be functional in the plant immune system. CAD1 (constitutively activated cell death 1) is preferentially expressed in newly developing organs, like newly grown leaves, implying its functional linkage to organ morphogenesis and *cad1* mutants showed a cell death phenotype (Morita-Yamamuro et al., 2005). Loss of NSL1 (necrotic spotted lesions 1) associates with cell death and defence (Noutoshi et al., 2006). The function of the other two *Arabidopsis* MACPF proteins have not been reported yet.

In the human pathogens *Plasmodium spp.* and *Toxoplasma gondii* for malaria and toxoplasmosis, respectively, the perforin-like proteins (PLPs) are involved in successful completion of their life cycles. Genetic disruption of any of them significantly affects their life cycle leaving the parasites trapped within their specific stage of infection, such as specific kinds of animal host cell, mainly due to the failure to disrupt the parasitophorous vacuole membrane barrier (Amino et al., 2008; Kafsack and Carruthers, 2010; Kafsack et al., 2009) (The structural and functional characterisation of these PLPs is one of the major pieces of work described in this thesis, in Chapters 5 and 6).

CDCs were originally identified in Gram-positive bacteria such as *Bacillus*, *Clostridium*, *Streptococcus* and *Listeria*, where they are involved in invasion and replication within animal hosts (Gilbert, 2005; Tweten, 2005) (**Table 1**). Phylogenetic analysis from Anderluh *et al* showed that CDCs are widely distributed, including Gram-negative bacteria and even within Diatoms (Anderluh et al., 2014) (**Figure 1.4**).

Table 1.1: List of Cholesterol-dependent cytolysins (selected)

Species	Toxin	Toxin (Abbreviation)	Implication in diseases
Arcanobacterium pyogenes	Pyolysin	PLO	Cytotoxic for murine peritoneal macrophage (Jost *et al.* 1999).
Bacillus anthracis	Anthrolysin O	ALO (3CQF)	Kills human neutrophils, monocytes and macrophages (Bourdeau et al. 2009).
Clostridium perfrgingens	Perfringolysin	PFO (1PFO)	Promotes dysfunctional human PMN/endothelial cell adhesion contacts and vascular leukostasis (Rossjohn et al., 1997).
Listeria monocytogenes	Listeriolysin	LLO (4CDB)	Pore formation within the listeria-infected vacuole and responsible for the releasing of listeria into host cytosol (Köster et al., 2013).
Streptococcus pneumoniae	Pneumolysin	PLY (5CR6)	Virulence factor responsible for pore formation (Lawrence et al., 2015).
Streptococcus intermedius	Intermedilysin	ILY (4BIK)	Binds to human complement receptor CD59 and make cytotoxic pores in human cells (Johnson et al. 2013).
Streptococcus pyogenes	Streptolysin O	SLO (4HSC)	Virulence factor of streptococci, preventing internalisation of the bacteria into lysosomes where they can be destroyed (Feil et al., 2014).
Streptococcus suis	Suilysin O	SLY (3HVN)	Virulence factor of *Streptococcus suis* (Xu et al. 2010).

[a]

[a] Values in parentheses are for highest-resolution shell.

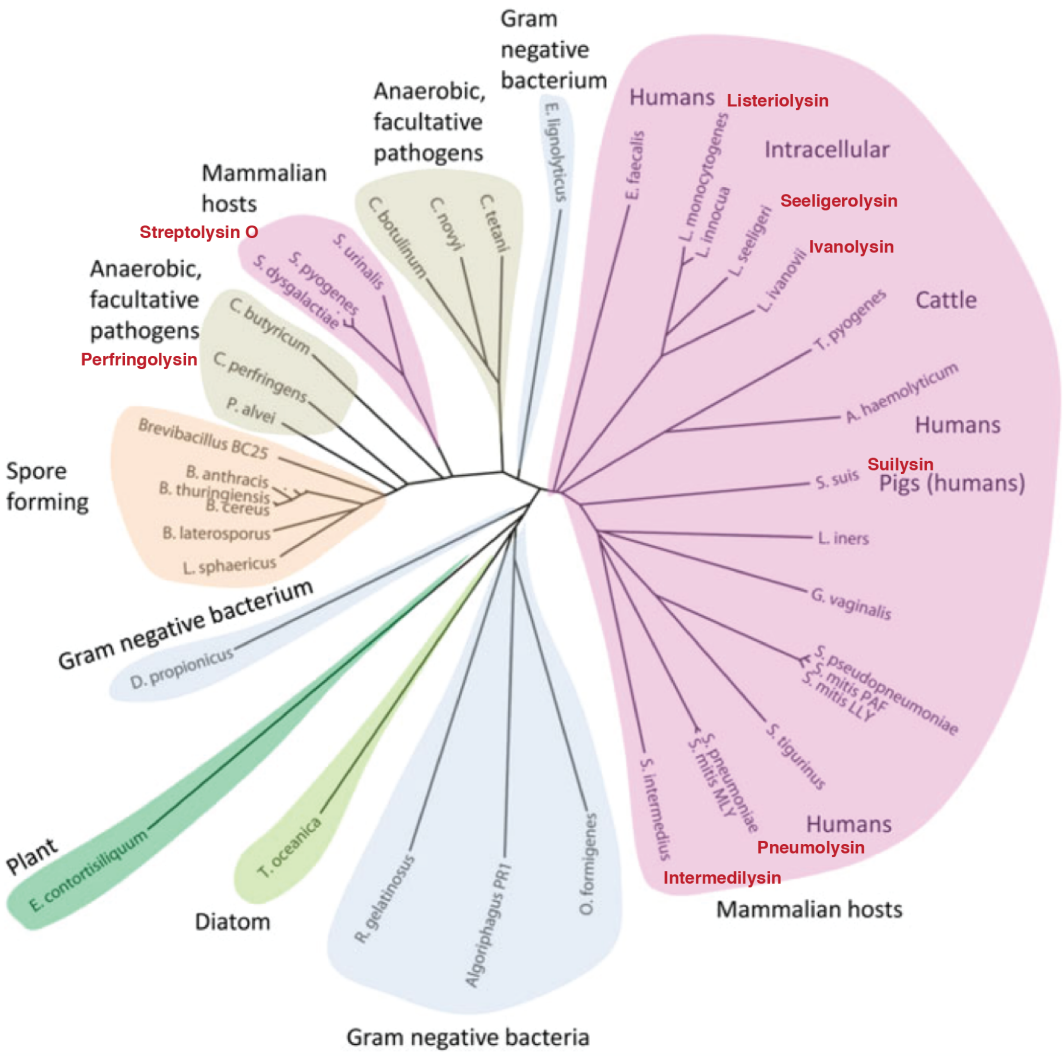

Figure 1.4: Distribution of Cholesterol-dependent cytolysins. Sequence-based phylogenetic tree of CDC proteins. Figure adapted from Anderluh et al. 2014. Representative CDCs that have been studied are highlighted in red.

1.3.2 Structure of Cholesterol-dependent cytolysins

The structures of CDCs and their mechanism of pore formation has been studied very thoroughly using a combination of structural and mechanistic approaches (Gilbert, 2005; Tweten, 2005). The first structure of a CDC, that of PFO (perfringolysin from *Clostridium perfringens*), was determined in 1997 (Rossjohn et al., 1997), revealing a flat molecule comprising of an N-terminal domain (originally annotated as three non-contiguous domains (I-III)) connected to a C-terminal Immunoglobulin (Ig) domain (domain IV) (**Figure 1.5A**). A number of CDC structures have been determined at an atomic resolution by x-ray crystallography since then (Bourdeau et al., 2009; Feil et al., 2014; Köster et al., 2014; Lawrence et al., 2015; Xu et al., 2010), which reveal the common architecture of the CDCs.

The N-terminal domain prominently features four central twisted β-strands with two sets of helices on either side of the β-strands. It contains three subdomains: domain I is involved in the initial oligomerisation and domain III provides the two sets of α-helices that transform into β-barrel (**Figure 1.5A**); domain II is a linking region providing the necessary flexibility required for the final pore-forming oligomer to be completed; and domain IV is responsible for membrane binding, providing a planar arrangement that allows CDCs to concentrate and oligomerise (Gilbert et al., 2013) (**Figure 1.5B** and **Figure 1.5C**).

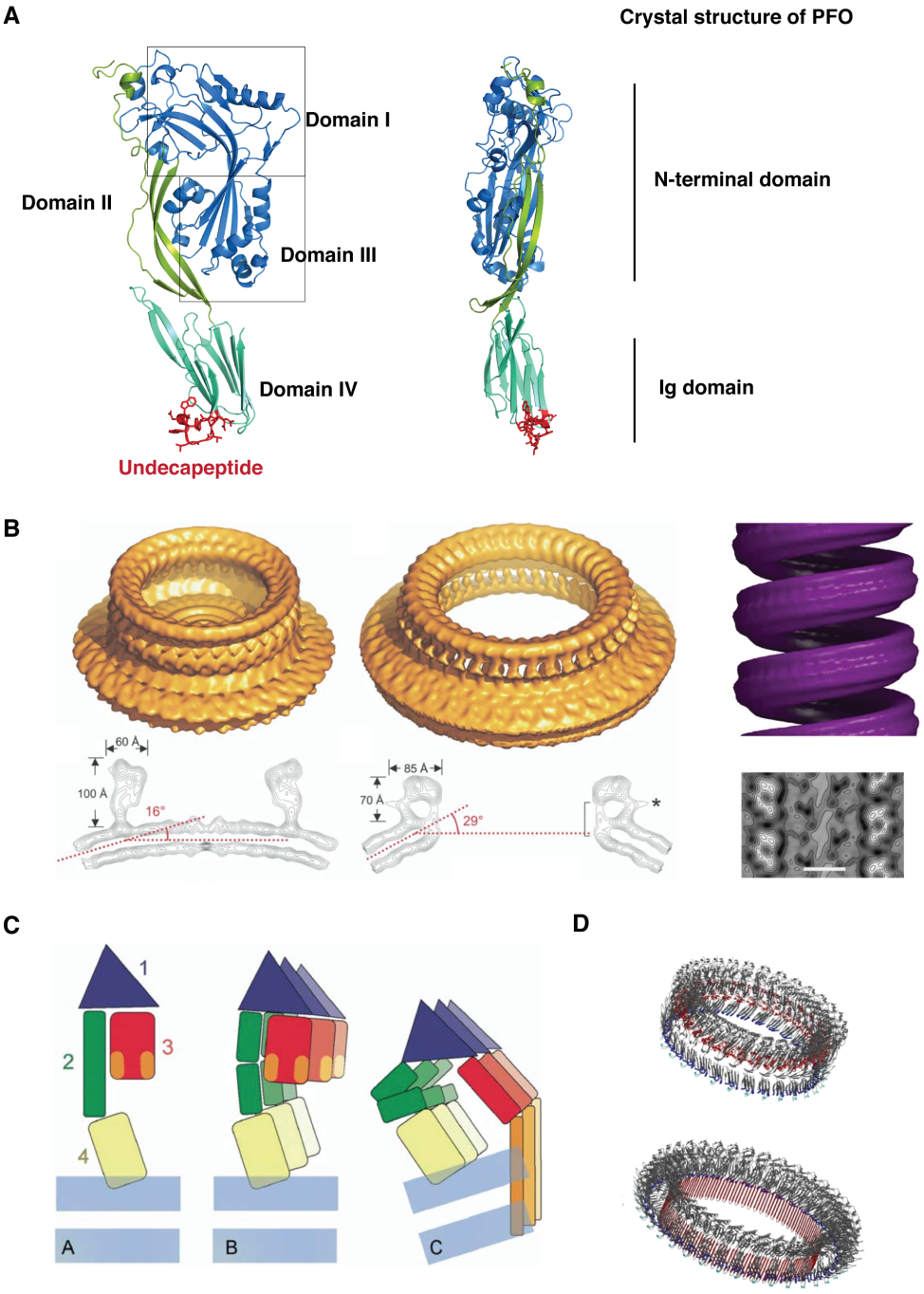

Figure 1.5: Structure of CDCs and mechanism of pore formation. (A) Crystal structure of the first cholesterol-dependent cytolysin to have its structure solved, perfringolysin (PFO). The domains are annotated as originally proposed (Rossjhon et al. 1997) with Domains I and III actually closely connected. The two sets of helices located on both sides of the Domain I β-strands are the regions responsible for membrane penetration. The C-terminal domain is an Ig-like domain in the bottom of the molecule responsible for initial membrane binding. The peptide coloured in red represents the undecapeptide responsible for initial membrane binding. (*Figure legend continued next page*)

Figure 1.5 (*continued*): (B) Cryo-EM reconstruction of pneumolysin prepore, pore and a helical oligomeric form arising in solution: Top, electron density map of reconstruction; Bottom, contoured cross-section of prepore, pore and helical form of pneumolysin. Figures adapted from Tilley et al., 2005 and Gilbert et al., 1999. (C) Cartoon showing the major stages in pore formation by CDCs: monomer binding, oligomerisation (assembly of prepore) and pore formation (conformational change and membrane penetration). Figure adapted from Tilley et al., 2005. (D) Proposed models of prepore and pore form of CDCs. The regions highlighted in red are the regions responsible for membrane penetration (two sets of helical structures in prepore forming β-hairpins in the mature pore).

The C-terminal domain (Ig domain) of CDCs mediates membrane recognition and contains 8 β-strands with a highly conserved tryptophan-rich undecapeptide (ECT-GLAWEWWR in PFO) in the tip (**Figure 1.5A**). As implied in its name, CDCs bind to cholesterol and this interaction is mediated by domain IV. The undecapeptide, as well as another short hydrophobic loop at the tip of domain IV has been shown to insert into the membrane, though not deeply. Together, these interactions anchor the CDCs to the cholesterol-incorporated membrane in a perpendicular orientation, which enables enough space for oligomerisation on both sides of the proteins (Dowd et al., 2012; Gilbert, 2005; Tweten, 2005). It should be noted that besides cholesterol, some CDCs require other binding partners for effective membrane binding. Intermedilysin (a CDC from *Streptococcus intermedius*), for example, requires human complement receptor CD59 (Johnson et al., 2013).

Mechanism of pore formation by CDCs generally follows that of β-PFTs with a well-defined series of events taking place. Via biophysical and biochemical studies with PFO and PLY (Gilbert et al., 1999; 1998; Shepard et al., 1998), the mechanism of pore formation became clear with the cryo-EM structure of a membrane-bound "pre-pore" form of pneumolysin, a solution-formed helical assembly (in which subunits adopt the pore-forming conformation) and a mature pore form (Gilbert, 2005; Gilbert et al., 1999; Tilley et al., 2005) (**Figure 1.5B**). Firstly, the soluble CDCs bind to the membrane through their C-terminal Ig domains, resulting in a concentration of toxins on the membrane. Secondly, these concentrated proteins are aligned together by the association between soluble subunits to form the prepore state. The self association

is mainly mediated by interfaces in N-terminal domain (specifically in subdomain I and III); the Ig domain does not form close inter-subunit contacts. The prepores consist of toxin molecules with an overall conformation similar to soluble proteins (Gilbert et al., 1999; Tilley et al., 2005) (**Figure 1.5B**). Thirdly, the pore formation is completed with a dramatic refolding of prepore to generate a transmembrane channel. This is mainly achieved by the insertion of domain III, in which two sets of helices substantially refold into two sets of extended β-strands and insert into the membrane (**Figure 1.5C** and **Figure 1.5D**).

1.3.3 Structure of MACPF domain proteins

Membrane attack complex/perforin family proteins were identified separately to cholesterol-dependent cytolysins, first as part of the complement membrane attack complex (MAC) and then in the form of perforin-1, which delivers granzymes from cytotoxic cells into target antigen presenting cells. The complement proteins C6-C9 and perforin share a common central domain of about 40 kDa known as the MACPF domain (Tschopp et al., 1986). The first crystal structure of the MACPF domain of a complement protein (C8α) was determined in 2007 (Hadders et al., 2007) , and most surprisingly the overall fold was found to share much similarity with the N-terminal region (domain I and III) of CDCs, with the central four twisted β-strands and two sets of helices (**Figure 1.5A** and **Figure 1.6A**). Meanwhile, the structure of a bacteria MACPF protein, Plu-MACPF from *Photorhabdus luminescens* was also determined and again closely resembles significantly a MACPF domain (Rosado et al., 2007) (**Figure 1.6A**). However, both C8α and Plu-MACPF do not form pore on membrane, despite their structural homology to CDCs.

The human genome encodes two perforin proteins, perforin-1 and perforin-2 (also known as Mpeg1 (Macrophage-expressed protein 1), functional studies will be briefly reviewed next section). Perforin-1 has been well-characterised alongside the MAC, due to its function in facilitating the delivery of granule-associated proteases (granzymes)

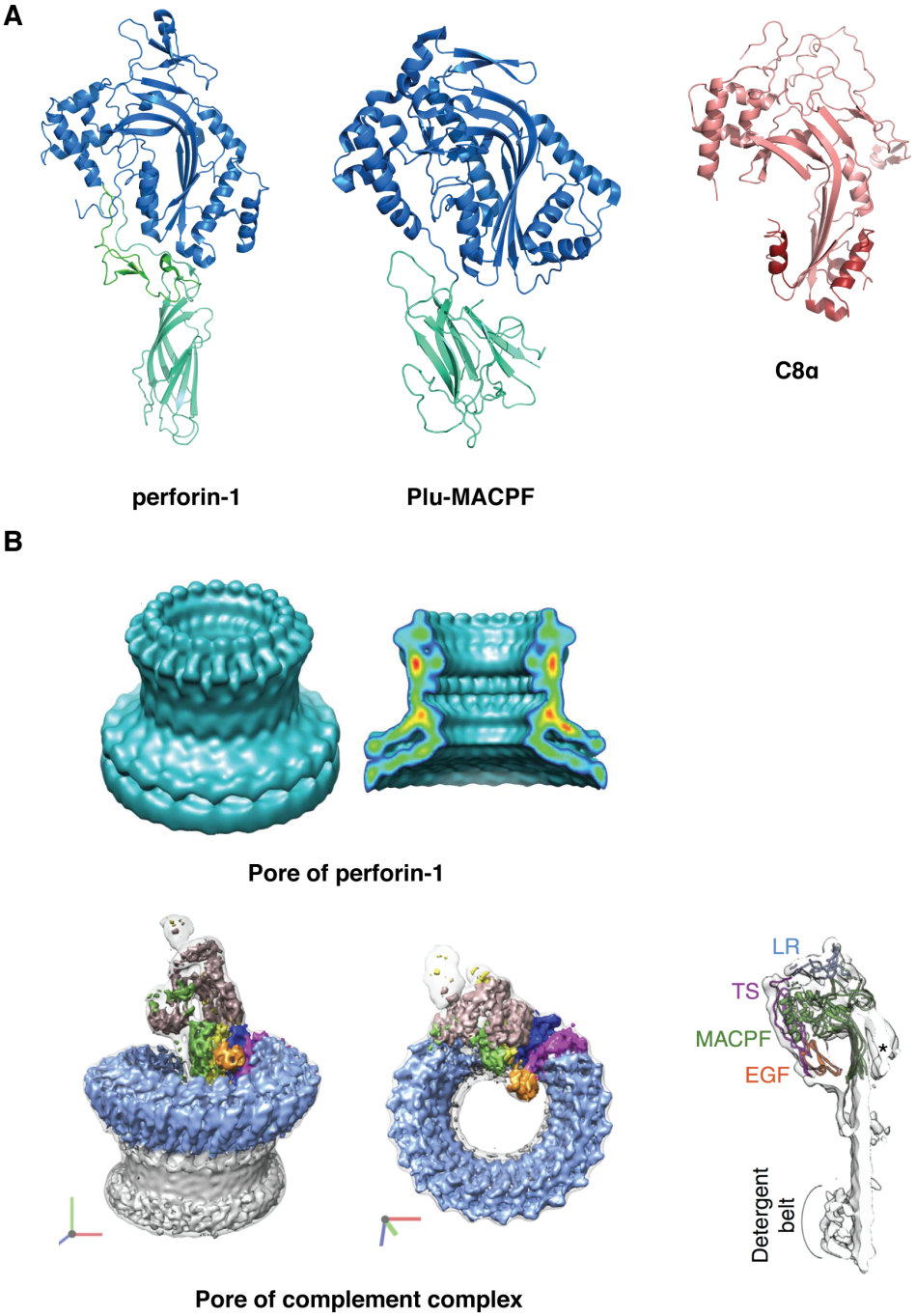

Figure 1.6: Structure of MACPF proteins and mechanism of pore formation. (A) Structures of representative MACPF domains: murine perforin-1 (PDB entry: 3NSJ), Plu-MACPF (PDB entry: 4V3N) of Photorhabdus luminescens (bacteria) and human complement C8α (PDB entry: 3OJY). (B) Cryo-EM reconstruction of complete pores: murine perforin-1 and membrane attack complex (MAC). The pore of perforin-1 is made up of perforin-1 only while the MAC is composed of several components (see further discussion below). The region coloured in blue in the MAC representation is the C9 molecule responsible for membrane penetration. Bottom right shows the cross-section of the mature MAC with one copy of C9 fitted in the electron density map. The electron density without structure fitted is where the β-strands are inserted into the membrane. Note the large conformational change before and after membrane insertion (C8α and C9 are homologous molecules).

to kill virus-infected and neoplastic cells. Unlike complement C9, perforin-1 is capable of forming pores without the assistance of other protein components or receptors. The delivery of granzyme can be achieved in two ways: (i) forming a full-ring or arc pore that allows the translocation of granzyme (Leung et al., 2014) (Metkar et al., 2015) (ii) or promoting membrane invagination followed by release of the granzyme from within the resulting endosome (Praper et al., 2011a; Thiery et al., 2011).

The mechanism of pore formation for MACPF domain proteins is in general similar to that of CDCs. The structure of monomeric perforin-1 was solved in 2010, which reveals a 'key'-shaped molecule encompassing a canonical MACPF domain, followed by an EGF-like domain and a C2-like domain in the C-terminal (Law et al., 2010) (**Figure 1.6A**). The C-terminal of perforin-1 is structurally similar to the Ig domain of CDCs, and contains a Ca^{2+} binding motif that mediates membrane binding. Perforin-1 binding to membrane does not require the presence of cholesterol, and its membrane binding is mainly induced by Ca^{2+} (Yagi et al., 2015).

Similar to CDCs, perforin-1 firstly assembles on the membrane and forms a prepore (Praper et al., 2011a; 2011b), and the prepore then undergoes conformational change to insert into the membrane to form a mature pore (Law et al., 2010). The pore formed by perforin-1 is smaller than that of CDCs, with about 19-24 subunits for most of the pore-forming rings, compared with 35-50 subunits in CDCs. However, there are significant differences in prepore to pore transition compared with that of CDCs. The perforin-1 prepore does not collapse to form the mature pore: the heights of prepore and mature pore are almost the same; while in contrast, pneumolysin exhibits significant difference of heights between prepore and pore form (Gilbert et al., 2013) (**Figure 1.5B** and **Figure 1.6B**).

Notably, the mechanism of pore formation by perforin-1 is similar to CDCs, and the orientation of MACPF domain in the pore is the same with that of CDCs (Gilbert et al., 2013), rather than as originally proposed by Law *et al.* (Law et al., 2010), and this has been thoroughly studied and comprehensively elaborated (Dudkina

et al., 2016; Gilbert et al., 2013; Praper et al., 2011b; Serna et al., 2016). While the overall mechanism of pore formation by perforin-1 is undoubtedly established, there are still more details remaining to be resolved. For example, the outer sets of helices which are supposed to undergo conformational change to convert into β-strands are covalently linked with the EGF-like domain within perforin-1. As a result, the bottom of the second, outer set of helices is stabilised and traps the whole first set of helices in the curvature of the central β-sheets (**Figure 1.7A**). The disulphide linkage is specifically identified within perforin-1 and well-conserved across all vertebrate species (**Figure 1.7B**). In solution, it is unlikely that the disulphide bond would automatically self-dissociate and then insert into membrane.

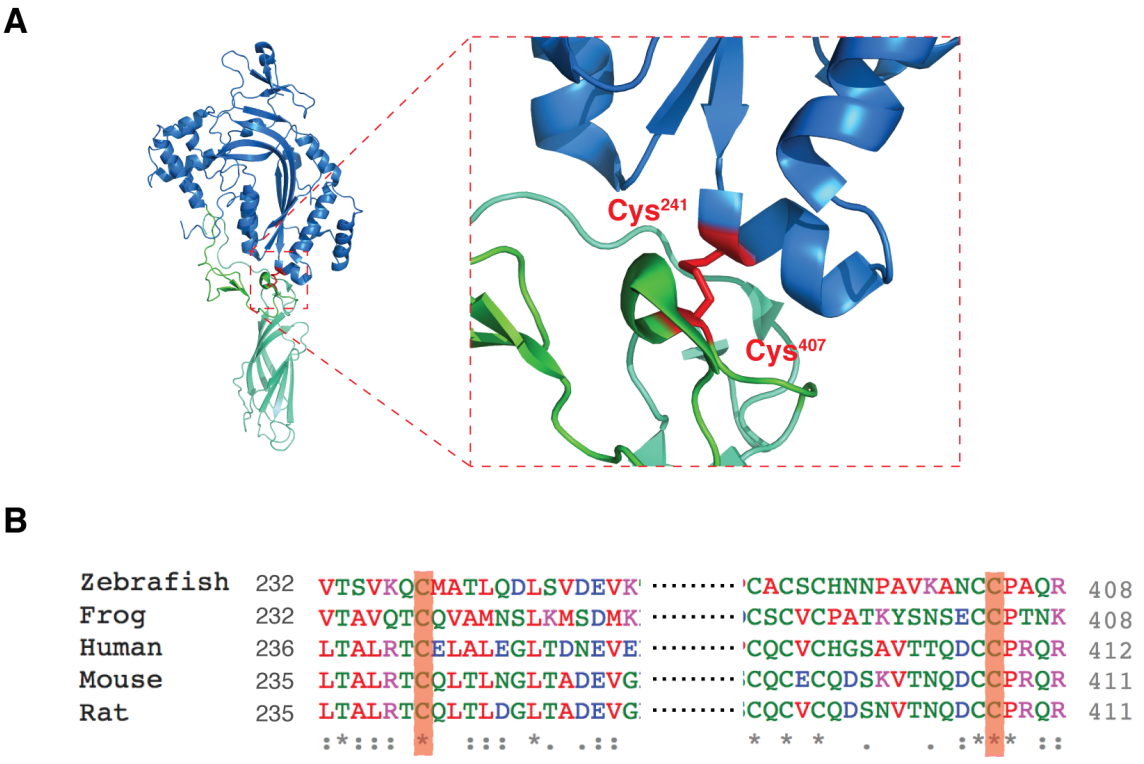

Figure 1.7: Disulphide bond locking perforin-1. (A) Disulphide bond (Cys^{241}-Cys^{407}) locking the transmembrane helices with EGF-like domain. (B) Sequence alignment of perforin-1 highlighting the conserved disulphide bond. The other conserved cysteins form disulphide bond within the EGF-like domain.

Complement is known as a tightly regulated protein network that plays an important role in host defence and inflammation through direct cell lysis and the mobilisation of innate and adaptive immunity. The activation of complement triggers assembly

of the MAC, a multiple protein pore that inserts into membrane of bacteria and directly lyses them. MAC is composed of multiple proteins including C5b, C6, C7, C8 (C8α, C8β and C8γ) and C9, most of which contain a central MACPF domain except C5b and C8γ (see **Figure 1.9** for the domains assembly of each protein). The structure of individual components has been resolved by x-ray crystallography for most of complement components except C7 and C9, although C7 is structurally close to C6 and C9 close to C8 (reviewed by Sonnen and Heneke, 2014) (**Figure 1.9**). The overall structure of MAC was recently revealed in a 'split-washer' configuration by cryo-EM at subnanometre resolution (Dudkina et al., 2016; Serna et al., 2016) (**Figure 1.6B**). In the final MAC, C9 is the main component that inserts into the membrane and forms functional pores, while other subunits provide a scaffold for its architecture.

1.3.4 Incomplete pore (toroidal pore) *vs* Complete pore

While the model of pore formation by a complete ring satisfies the human pursuit of beauty due to its appearance of a high-order symmetry, it does not agree with all of the experimental data for MACPF/CDC proteins. Undoubtedly, MACPF/CDC proteins undergo prepore (complete ring prepore)-to-pore transition to form a complete ring pore (**Figure 1.5B** and **Figure 1.6B**). However, it is ubiquitously observed that these proteins form arc-shaped assemblies on membranes as well, especially when their concentration is low (Palmer et al., 1998). Two questions were therefore brought about by this observation: on the one hand, does the prepore to pore transition absolutely require the completion of a ring by oligomerisation; on the other hand, do the incomplete rings (arcs) formed by MACPF/CDCs insert into membranes to form a functional pore, i.e., can the arcs form stable pores within the membrane bilayer interface?

It is only recently that it became clear that indeed individual types of MACPF/CDC protein can form functional pores using oligomers of variable size (Leung et al., 2014;

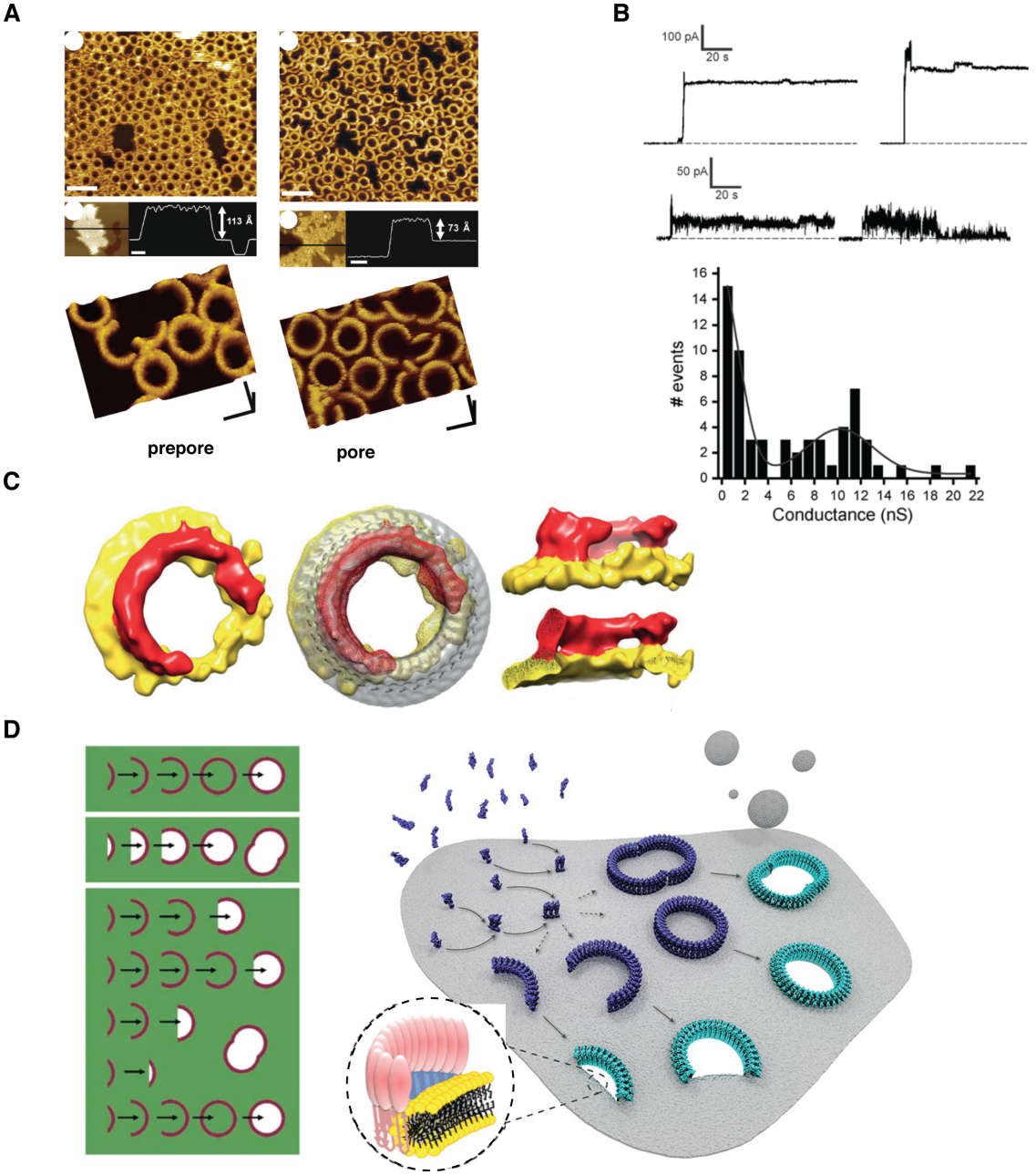

Figure 1.8: Incomplete pore (arcs) model. (A) Atomic force microscope (AFM) study of pneumolysin prepore and pore. Both arc form and ring forms of pneumolysin exist in prepore and pore states; the major difference as shown by AFM is that the prepore has a height of 113 Å while mature pores collapse to a height of 73 Å. Figures adapted from Czajkowsky et al., 2004. (B) Examples of diverse perforin-1 pores formed in POPC. Large pores with conductance more than 5 nS (above) and small with conductance below 1 nS (below) are shown. Figures adapted from Praper et al., 2011a. (C) Cryo-EM tomography study of pneumolysin arcs in the pore form with pneumolysin in red and cholesterol-containing membrane in yellow. Middle: superimposition of arc pore and complete ring pore (coloured in grey). Figure adapted from Sonnen et al., 2014. (D) Schematic diagram of models for pore formation by CDCs. Left: model proposed by Gilbert 2005; Right: model redrawn by Leung et al., 2014. The models highlight the functional implications of the arc form of CDCs: CDCs do not need to form a complete ring in order to insert into membrane; instead the arc form of CDCs can insert into a membrane and form a pore with a protein-membrane interface.

Metkar et al., 2015; Podobnik et al., 2015; Praper et al., 2011b; Sonnen et al., 2014). The same protein can generate functionally important lesions of different sizes in target membranes depending on prevailing conditions such as the concentration of protein but also, for example, pH (Birmingham et al., 2008; Czuczman et al., 2014; Metkar et al., 2015). At least three major observations support the idea that the incomplete pore (arcs) can insert into membranes and are functional (Gilbert et al., 2013; 2014). First of all, the gradual increase in ionic current with associated small pores has been observed for the MACPF/CDCs such as pneumolysin (El-Rachkidy et al., 2008; Korchev et al., 1992), listeriolysin (Podobnik et al., 2015) and perforin-1 (Praper et al., 2011b) (**Figure 1.8B**). Secondly, incomplete rings (arcs) were widely observed in various samples using electron microscopy. Recently convincing evidence was obtained by a cryo-electron microscopy sub-tomogram averaging study of pneumolysin pores within liposomal membranes (Gilbert et al., 2014; Sonnen et al., 2014), directly demonstrating the arc form assembled as both prepore (arcs in prepore conformation) and pore (arcs in pore conformation) (**Figure 1.8C**). Thirdly, a number of studies making use of atomic force microscopy to image pore formation by CDCs such as suilysin (*Streptococcus suis*) and listeriolysin (*Listeria monocytogenes*) (Leung et al., 2014; Mulvihill et al., 2015; Podobnik et al., 2015), showed that arcs were forming pore structures in membranes (**Figure 1.8A**). Much evidence suggests that perforin-1 has the same mechanism for pore formation by arcs or complete rings of subunits of perforin-1, including *in situ* imaging of pores formed by arcs of subunits on intact cells (Metkar et al., 2015).

So all in all, these data answer the question posed above directly: firstly, CDCs do not have to assemble into a complete ring in order to transit to a mature pore (Gilbert et al., 2013; 2014; Tweten et al., 2015); secondly, incomplete rings (arcs) of MACPF/CDC subunits can insert into membranes to form pores, with the size of any individual pore-forming assembly being determined by the availability of further monomers for incorporation into the pre-pore (Gilbert, 2002; 2005; Gilbert et al.,

2014; Metkar et al., 2015; Praper et al., 2011a). One might argue the biological significance of pore formation with a protein-lipid interface. If MACPF/CDC proteins can form pores using arcs of subunits, this would enable their producing organisms to tune the extent of host membrane damage in order to adapt to their stage of infection and life cycle progression. It would explain the capacity of listeriolysin from *Listeria monocytogenes* to form functional pores of differing sizes (Birmingham et al., 2008; Czuczman et al., 2014); and that of mammalian perforin-1 itself, among the MACPFs (Metkar et al., 2015; Praper et al., 2011a; 2011b).

1.3.5 Vertebrate-specific MACPF proteins

As mentioned earlier, four groups of proteins containing MACPF domains have been identified in humans, including Membrane attack complex of Complement (C6, C7, C8α, C8β and C9), Perforins (perforin-1, perforin-2 (also known as Mpeg1 (Macrophage-expressed protein 1)), Astrotactins (ASTN1 and ASTN2) and BRINPs (BRINP1, BRINP2 and BRINP3). Complement proteins and the Perforin family proteins are well-known for their pore-forming capacity, while Astrotactins and BRINPs are MACPF proteins involved in neuronal development, which are more likely to not be pore-forming proteins. However, the exact structure and function of of developmentally-involved MACPF proteins remains unknown. This section will focus on the human MACPF proteins, with a particular emphasis on the ones that have not been well-studied. It aims to analyse the structures of these proteins on the basis of their sequences and provides some clues for a better understanding about their functions. The function and structure of perforin-1 and MAC have been described in the previous section and therefore not to be included here.

Perforin-2

Perforin-2, initially named as Mpeg1 (Macrophage expressed protein 1), was identified as a gene with expression tightly restricted to macrophages and has subsequently

Table 1.2: List of members of vertebrate-specific MACPF subclass

MACPF subclasses	Common names	Expression pattern and function
Complements	C6	Vertebrate membrane attack complex (MAC), secreted scaffold protein (Serna M et al 2016).
	C7	anchoring MAC to the target membrane, secreted scaffold protein.
	C8α and C8β	anchoring MAC to the target membrane, secreted scaffold protein.
	C9	membrane insertion, pore formation and lytic activity, secreted protein.
Perforins	perforin-1	Released from granules from natural killer cells and cytotoxic T lymphocytes, responsible for granzyme delivery, secreted protein.
	perforin-2	effector of the innate immune system, limiting the proliferation and spread of Gram-negative, -positive bacteria; type I transmembrane protein (McCormack et al. 2015).
Astrotactins	Astrotactin-1	Neuron receptor mediating glial-guided neuron migration; double transmembrane protein (Zheng et al. 1996)
	Astrotactin-2	Regulate the trafficking of ASTN1 during glial-guided neuron migration; endosomal protein double transmembrane protein (Wilson et al., 2010).
BRINPs	BRINP1/ DBRCC1	Mainly expressed in nervous system and involved in neuron development (Kawano *et al.* 2004); Deleted in bladder cancer region-1, tumor suppressor, and over-expression of BRINP1 suppress tumour cell growth (Habuchi *et al.* 1998; Motomiya *et al.* 2007)
	BRINP2	Mainly expressed in nervous system, unknown function
	BRINP3/ FAM5C	Mainly expressed in nervous system, associated with gonadotrope cell pituitary tumorgenesis (Shortcary *et al.* 2007)

[a]

[a] Values in parentheses are for highest-resolution shell.

been used as a marker for this specific cell type (Ellett et al., 2011; Zakrzewska et al., 2010). Sequence analysis shows that perforin-2 is perhaps the origin of other MACPF domain proteins, evidenced by the wide distribution of perforin-2 (Mpeg1) in sponge, invertebrate and vertebrates (D'Angelo et al., 2012; McCormack and Podack, 2015). Unlike complement components or perforin-1 which are secreted soluble proteins, perforin-2 is a type I transmembrane protein, projecting its main portion (MACPF domain followed by a β-strand rich domain) into an endoluminal compartment, while leaving a short tail on the cytoplasmic side (**Figure 1.9**). Recently it has been shown that perforin-2 is essential for intracellular defence of parenchymal cells (McCormack and Podack, 2015; McCormack et al., 2015a; 2015b). and forms a pore with a diameter about 10 nm (smaller than the complete pore of perforin-1 (15 - 20 nm)) on bacterial membranes (McCormack et al., 2015a). Pore formation can be induced by trypsin digestion of over-expressed perforin-2 on cell membranes (McCormack et al., 2015a). Very recently it was also shown that perforin-2 restricts the proliferation of the intracytosolic pathogen *Listeria monocytogenes* (McCormack et al., 2016).

Astrotactins

There are two Astrotactins in humans, Astrotactin-1 (ASTN1) and Astrotactin-2 (ASTN2), and both of them are integral transmembrane proteins with a molecular weight of about 150 kDa. Overall, these two proteins share more than 50% sequence identity and contain the same domain architecture: ASTN1 contains two transmembrane helices, spanning a large portion of proteins (a small N-terminal domain and large C-terminal domain) in the extracellular region or endosomal compartments, and a cytosolic domain in the cytosol (**Figure 1.9**). The C-terminal domain contains three EGF-like domains, a MACPF domain, fibronectin type III (Fn(III)) domain and annexin-like domain (identified from this thesis). There are no obviously homologous domains with N-terminal domain and cytosolic domain.

Both ASTNs are mainly expressed in neuronal cells in both the cerebellum and the cerebral cortex (Edmondson et al., 1988; Wilson et al., 2010; Zheng et al., 1996).

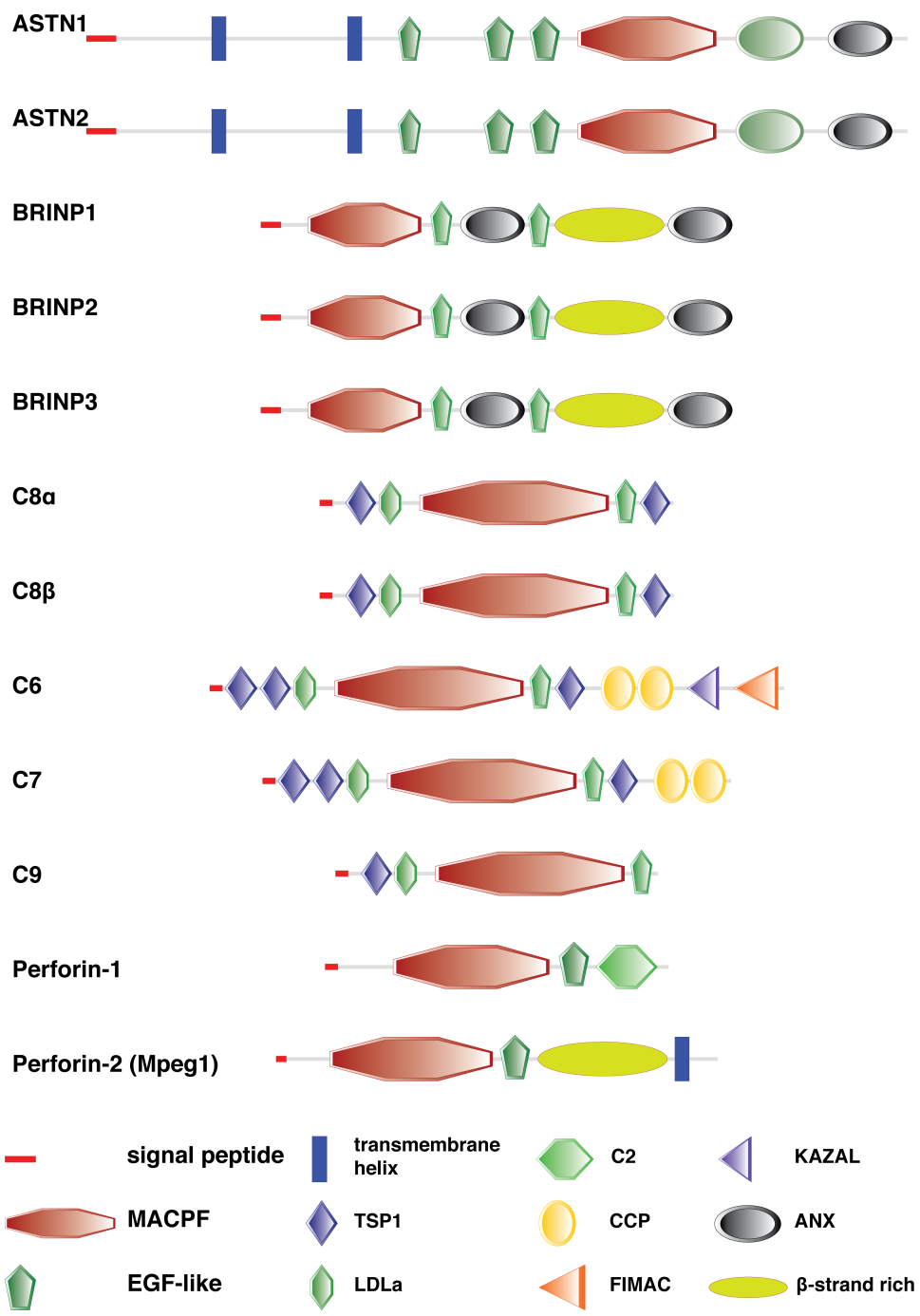

Figure 1.9: Schematic representation of Human MACPF proteins highlighting their domain architectures. Neural-specific MACPF proteins have abbreviated MACPF domains and their MACPF domains are smaller than those of perforins and MAC. The annexin-like domains of the BRINPs are predicted based on their secondary structures and comparison to the structure of ASTN2. The ?-strand rich domains of BRINPs and perforin-2 (ellipses) do not fall into any known domain families. ASTNs and perforin-2 are integral membrane proteins, and their transmembrane helices are labelled with dark blue rectangles. Other domains are depicted following colour scheme and representations using the default settings of SMART protein database. The cartoon representation was downloaded from SMART protein database (http://smart.embl-heidelberg.de) and modified. Abbreviations: FN3, Fibronectin type III; ANX, annexin-like domain; TSP1: TSP type-1 domain; LDLa, LDL receptor class A domain; CCP: Sushi (CCP/CCR) domain; KAZAL: Kazal-like domain; FIMAC: Factor I Module MAC domain.

ASTN1 is directly responsible for establishing contacts between neuronal cells and glial cells. Antibody binding to surface-exposed ASTN1 retards the migration of isolated neuronal cells along glial fibres. ASTN1 is localised to both endosomal compartments and the cell membrane, while ASTN2 mainly resides in endosomes. ASTNs interact with each other, although the details of the interaction have not been thoroughly characterised (Wilson et al., 2010). Mice with deleted ASTN1 have smaller sizes of cerebella, and poorer balance and coordination (Adams et al., 2002). ASTN2 has been identified as a genetic modifier that regulates the global orientation of mammalian hair follicles (Chang et al., 2015). Genetic studies have identified ASTN2 as a risk factor in autism, attention deficit hyperactivity disorder and other neurodevelopmental disorders (Lesch et al., 2008; Lionel et al., 2014; Wang et al., 2014).

BRINPs

The human genome encodes three BRINP family proteins, and all are predicted to contain a signal peptide, which suggest its localisation in membrane compartments. BRINP proteins are smaller in molecular weight (90kDa) compared with Astrotactins (150 kDa). Bioinformatics studies showed that they contain a MACPF domain followed by an EGF-like domain (a cysteine-rich domain) in the N-terminal, and the structure of C-terminal region is unknown. Compared with MACPF domains from MAC and perforins, MACPF domain in BRINPs are much smaller: the inner set of helices bundles are very likely missing, although the outer ones remain the similar size to those of perforins (**Figure 1.9** and **Figure 1.10**). Secondary structure prediction shows that the C-terminal region contains two short sections of sequence predicted to be rich in helical bundles, matching the secondary structure of the annexin-like domain in Astrotactins, which suggests that BRINPs might contain an annexin-like domain as well (**Figure 1.9**). BRINPs are highly conserved in vertebrates: zebrafish and humans share $> 90\%$ similarity for BRINP1, $> 80\%$ similarity for BRINP2 and approximately 80% for BRINP3.

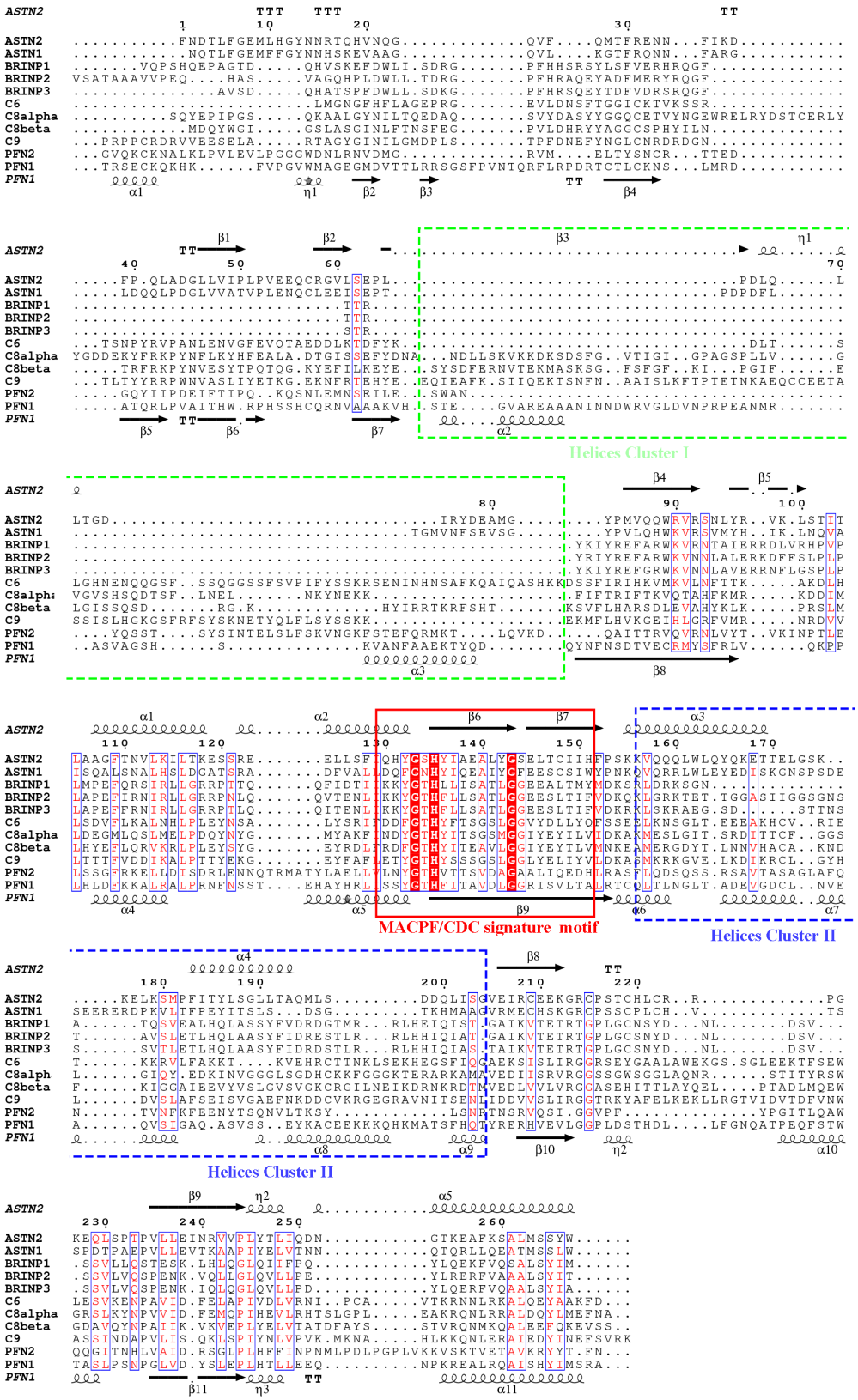

Figure 1.10: Sequence alignment of human MACPF domains. MACPF/CDC signature motif is highlighted in red bracket with conserved The major difference lies in Helices cluster I: in neuronal MACPF proteins, this loop is almost missing; while perforin and MAC have an intact helical cluster, matching the size of Helices cluster II. The Helices cluster I and II refer to inner and outer set of helices mentioned above, respectively.

BRINP1, bone morphogenetic protein and retinoic acid inducible neural-specific protein 1, as indicated by its name, was initially identified as a tumor suppressor protein in human bladder cancer with a different name called DBCCR1 (Deleted in bladder cancer chromosome region 1), in an effort to locate tumor suppressor genes in bladder cancer cells (H et al., 2001; Habuchi et al., 1998). Subsequently, it was observed that ectopically expressing BRINP1 caused cell cycle arrest in cultured cell lines such as bladder cancer cells and the NIH 3T3 cell line (H et al., 2001; Kawano et al., 2004; Wright et al., 2004), or even mouse embryonic stem cell-derived neural stem cells (Terashima et al., 2010). Independently, Kawano et al identified that BRINP1 was predominantly expressed in the central nervous system from developmental stage to adulthood, and more importantly, that the expression level of these proteins were greatly induced upon stimulation with bone morphogenetic protein or retinoic acid, from which study the name of the proteins was proposed. They also observed that the two orthologues of BRINP1 (BRINP2 and BRINP3) were expressed predominantly in the nervous system, although the expression pattern of individual BRINP genes differed from each other (Kawano et al., 2004). The absence of BRINP1 in mice causes an increase in hippocampal neurogenesis and behavioural alterations possibly relevant to human psychiatric disorders (Kobayashi et al., 2014). BRINP3, however, has mainly been studied in non-neuronal cells, and is also known as a family with sequence similarity 5 member C (FAM5C). Genetic studies have linked BRINP3 to myocardial infarction. It was also shown that BRINP3 was expressed in tongue squamous cell carcinoma and gonadotrope cell pituitary adenomas (Shorts-Cary et al., 2007), implying its potential role in tumorigenesis. Intriguingly, BRINP3 was observed to localise to the mitochondria of gonodotrope cells, although its exact function remains to be determined. A recent study of expression pattern of zebrafish BRINPs during development also observed that they were broadly expressed in the developing nervous system at early stages and then relocated to specific structures, indicating their conserved functions in neuronal development (Giousoh et al., 2015).

1.4 Overview of the thesis

This thesis mainly focuses on structural and functional studies of MACPF/CDC superfamily proteins, in particular Astrotactins from humans and perforin-like proteins (PLPs) from *Toxoplasma* and *Plasmodium*. Both of these subfamilies of proteins are implicated in human diseases and structurally uncharacterised before. The results are presented in three sections, with two sections on MACPF/CDC superfamily proteins and one section on pleckstrin homology domain of kindlins.

The first section (Chapter 2, Chapter 3 and Chapter 4) describes studies of human Astrotactins. Chapter 2 describes the crystal structure of the carboxy-terminal of Astrotactin-2 (80kDa), and describes the novel protein architecture possessed by ASTN2, with a combination of a canonical MACPF domain with EGF-like, Fibronectin-type (III) and Annexin-like domains. Furthermore, it discusses the membrane binding properties of Astrotactin-2. The structure of ASTN2 represents the first integral membrane MACPF protein family member to have a structure solved. Chapter 3 presents details of structure determination of the ASTN2 endodomain whose structure is described in Chapter 2. The use of experimental phasing using Platinum-derivative single isomorphous replacement with anomalous scattering (Pt-SIRAS) in spacegroup P1 at low resolution is discussed. Chapter 4 concerns the characterisation of full-length Astrotacin-2 and its individual domains.

The next section (Chapter 5 and Chapter 6) presents work on Perforin-like proteins from *Toxoplasma gondii* and *Plasmodium spp.* (*Tg*PLP1 and *P*PLPs), with an emphasis on *Tg*PLP1. Chapter 5 describes the structures of *Tg*PLP1 MACPF domain and C-terminal domain determined at a resolution of 2.0 Å and 1.1 Å, respectively. It also presents structures of the MACPF domain arranged in a helical and hexameric ring form, which indicates the potential for pore-formation by 6 subunits. This is the first structure of a perforin-like protein from an Apicomplexan parasite and provides a structural basis for understanding the function of the PLPs

in toxoplasmosis and malaria pathogenesis. Chapter 6 describes details of protein expression screening of *Tg*PLP1 and the *P*PLPs and efforts in crystallising *Tg*PLP1 which eventually resulted in its structure determination.

The final section is a continuation of a previous study in our lab on the pleckstrin homology (PH) domains of kindlins. We determined the crystal structure of kindlin-3 PH domain and use it as the base for molecular dynamic simulations (in collaboration with Mark Sansom's group, Department of Biochemistry, University of Oxford) to study its lipid and membrane binding properties. Using nanodiscs with different lipids incorporated as a model membrane system the interaction between inositol phosphate lipids and kindlin-3 PH domain was studied. A model is proposed that one subset of PH domains might be able to bind to multiple inositol phosphates simultaneously and thus strengthen their interaction with target membranes.

CHAPTER 2

Crystal Structure of ASTN2 endodomain

2.1 Summary

Astrotactin-2 (ASTN2) is one of a pair of proteins which have been shown to play a crucial role in enabling neural migration along glial fibres. While ASTN1 directly forms contacts between the neuron and the fibre, ASTN2 is responsible for extracting ASTN1 from contacts at the lagging edge of the cell and recycling them to the leading edge. ASTN2 is associated with endosomes, with the majority of its structure (at the C-terminus) projecting into their lumen, anchored by a pair of transmembrane α-helices. This chapter presents the structure of this "endodomain" region of ASTN2, and find it to consist of a unique combination of polypeptide folds comprising a membrane attack complex/perforin (MACPF) domain, an EGF-like domain, a previously-unobserved form of fibronectin type III (Fn(III)) domain and an annexin-like domain. The MACPF domain differs from other proteins of the perforin-like family in ways which suggest it does not function via pore formation, unlike other related proteins. The Fn(III) domain is modified by a long β-hairpin from the C-terminus of the polypeptide folded across its own β-sheet sandwich and associated with an inositol phosphate moiety also bound to the EGF-like domain. The annexin-like domain shows no sequence homology to human annexin but is closely related in structure, and also binds calcium. Taken together, our structural characterisation provides a framework for better understanding the mechanism of the ASTNs and related proteins in neural and other forms of vertebrate development.

The structure determination procedure of the structures described in this chapter is presented in Chapter 3 in detail .

2.2 Introduction

2.2.1 Astrotacins in neuron development

MACPF proteins are involved in a wide range of biological processes, including not only immunity (MAC, perforin-1 and perforin-2) (McCormack et al., 2013; McCormack et al., 2015a; Sonnen et al., 2014) but also cell invasion and egress by apicomplexan parasites (Amino et al., 2008; Kafsack et al., 2009; Tavares et al., 2014) and organismal development (Edmondson et al., 1988; Johnson et al., 2015; Kawano et al., 2004; Kobayashi et al., 2014; Solecki, 2012; Wilson et al., 2010). Our focus here is on a group of MACPF proteins especially identified as being involved in neurodevelopment, which includes astrotacin-1 (ASTN1) (Edmondson et al., 1988) and ASTN2 (Solecki, 2012; Wilson et al., 2010). However, in addition to its role in neural migration control, a recent report has indicated that ASTN2 is also part of the mechanism of planar cell polarity (PCP) determination *via* the activation of non-canonical Wnt signalling Frizzled-6 receptors (Chang et al., 2015).

ASTN1 was first discovered because antibodies against it block neuron-glial interactions *in vitro* (Edmondson et al., 1988); it was subsequently shown to be expressed in post-mitotic neuronal precursors of the cerebellum, hippocampus, cerebrum and olfactory bulb with a role in the establishment of laminar structures (Zheng et al., 1996) . ASTN1 is directly responsible for the formation of neuron-glial fibre contacts in the cerebellum (Adams et al., 2002; Fishell and Hatten, 1991; Stitt and Hatten, 1990; Zheng et al., 1996b) and the discovery of ASTN2 as its intracellular counterpart showed how they might be plucked up and put down to enable the forward migration of neuronal cells (Wilson et al., 2010) (**Figure 2.1A**).

ASTN2 is most highly expressed in the cerebellum but is also found in the cortex, olfactory bulb, hindbrain, and the dentate gyrus of the hippocampus at murine P6, P10 and adult stages (Wilson et al., 2010). The expression of ASTN1 and

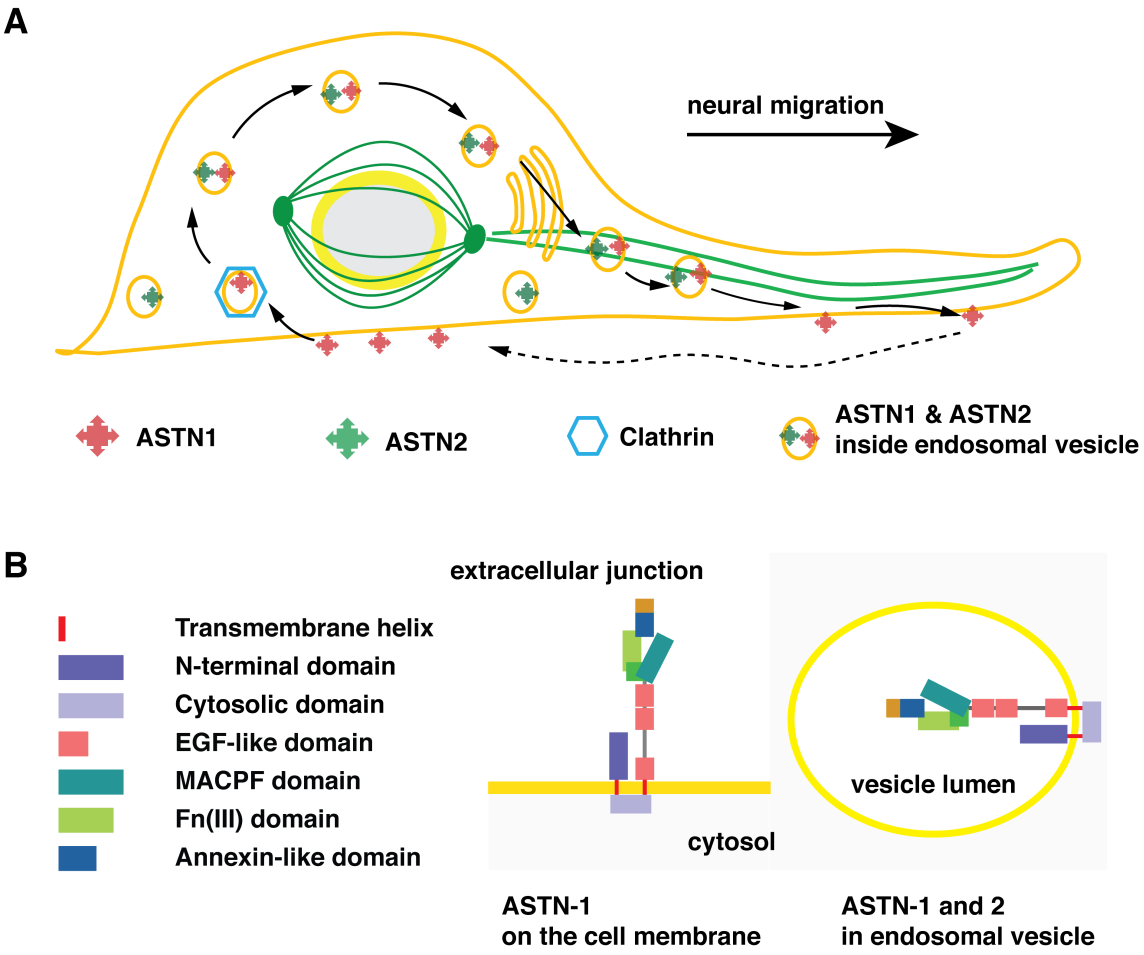

Figure 2.1: Neuron migration and cycling model of Astrotactins. (A) Schematic of roles played by ASTNs 1 and 2. ASTN1-mediated adhesions undergo endocytosis into clathrin-coated vesicles dependent on ASTN2 (1); the vesicles cycle through early and recycling endosomes (2-4) and undergo microtubular migration (5-7) until the ASTN1 is re-deposited towards the leading process (7) to form a new adhesion (8) which will be recycled again (9) in step with the cell migration. (B) Outline of ASTNs domain architectures and intracellular localisation. Both ASTNs are integral membrane proteins, with two transmembrane helices projecting a large C-terminal domain into the extracellular junction (ASTN1) or endosomal vesicle lumen (ASTN1 and ASTN2), while exposing cytosolic domains on the other side of membrane. Panel (A) is modified from Wilson et al.2010.

2 is differentially regulated in terms of cellular location (Wilson et al., 2010) and developmental stage (Lionel et al., 2014). While ASTN1 is expressed on the neural cell surface to form glial fibre contacts *via* its C-terminal ectodomain, ASTN2 is not expressed on the cell surface but is localised instead to vesicles inside the cell, giving rise to a punctate antibody staining pattern in the soma and along neuronal processes, and with particular accumulation at one pole. In agreement, co-expression of ASTN1 and fluorescently-labelled ASTN2 in HEK293T cells shows them to be colocalised to a subset of $RhoB^+$ endosomes (as well as in non-$RhoB^+$ vesicles) (Wilson et al., 2010). As cerebellar granule neurons grow, ASTN1 and ASTN2 are found within clathrin-coated and dynamin-dependent vesicles at the base of the leading process and opposite an interstitial junction with the glial fibre (**Figure 2.1A**). Intracellular imaging reveals the ASTN2-dependent cycling of vesicles bearing ASTN1 from the anterior pole of the neuronal soma and the base of the leading process, into the cell, and down the leading process to form a new neuron-glial fibre junction towards its tip (Solecki, 2012; Wilson et al., 2010)(**Figure 2.1A**). In this way, ASTN1-mediated contacts between the migrating neuron and glial fibre are released, recycled and reformed.

2.2.2 Astrotactins are linked with genetic diseases

As might be expected given their role in neurodevelopment, data suggest that ASTN1 and ASTN2 are key to many aspects of basic mammalian neurobiology. For example, ASTN1 is increased in rat brain following hippocampal injury, implying a role in repair processes (Price et al., 2003) while ASTN1 knockout mice display poorer balance and coordination than wild type (Adams et al., 2002). In humans, a comparative genomic hybridisation study of a Russian cohort with intellectual disability identified an individual with a duplication of ASTN1 showing multiple neurodevelopmental defects and delays as well as a number of non-neuronal phenotypic effects (Kashevarova et al., 2014) which seem to map at least in part onto the PCP pathway (Chang et al.,

2015; Wang and Nathans, 2007).

Linkage of ASTN2 to the development of the mammalian CNS is even stronger than that of ASTN1. ASTN2 has been implicated *via* genome wide association studies in attention deficit hyperactivity disorder (ADHD) (Lesch et al., 2008), by copy number variant (CNV) analysis of a large human cohort in autism spectrum disorders (ASDs) (Glessner et al., 2009) and in another CNV study in schizophrenia (Vrijenhoek et al., 2008). In one recent paper, a large set of neurodevelopmental disorder (NDD) subjects were compared with a set of population-based controls to identify 46 deletions and 12 duplications affecting ASTN2; the NDD subjects demonstrated a variety of phenotypes including ASDs, ADHD, speech delay, anxiety and obsessive compulsive disorder (Lionel et al., 2014). In the same study, analysis of the spatiotemporal expression patterns of ASTNs -1 and -2 in human brain samples from individuals of different ages showed ASTN1 expression at consistently high levels whereas ASTN2 expression peaked in the early embryonic neocortex and postnatal cerebellar cortex (Lionel et al., 2014). Another recent paper identified a clinical link between ASTN2-associated polymorphisms and the onset of Alzheimer's disease $\sim$5 years earlier than the median (Wang et al., 2014).

2.2.3 Domain and sequence analysis of Astrotactins

ASTN1 and ASTN2 have been shown to interact with each other *in vitro* in a calcium-independent way with multiple regions of each protein apparently contributing to their interface. Deletions of individual domains within the full-length ASTNs does not impair the interaction (Wilson et al., 2010). Both ASTN1 and ASTN2 are integral membrane proteins in which a large C-terminal domain (extracellular for ASTN1, endosome luminal for ASTN2; **Figure 2.1B**) is complemented by an N-terminal cytosolic domain suspended between two transmembrane α-helices. The bulk of ASTN2 projects into the lumen of endosomes on which it is expressed, with two transmembrane α-helices providing a firm anchor in the membrane bilayer(**Figure**

2.1B and **Figure 2.3A**). Thus, the endosomal regions of ASTN2 comprise a smaller N-terminal domain and a larger C-terminal domain whose structure we report here and which we refer to as its *endodomain*. The ASTN2 endodomain is exactly equivalent to the cell surface-exposed ectodomain of ASTN1; and it is the first in a class of integral membrane PLPs (which includes perforin-2) to have its structure determined, though perforin-2 differs from ASTN2 (and ASTN1) in having only a single transmembrane helix (McCormack et al., 2015a, 2015b).

In describing here the structure of the endodomain of ASTN2 we provide a framework for understanding its role in cell migration and tissue development. We describe its unique combination of a canonical MACPF/CDC domain with a minimal EGF-like repeat (EGF-4), a previously-unobserved type of fibronectin type III (Fn(III)) domain in which an additional two β-strands are folded across the core, and an annexin-like domain. We have determined structures of the MACPF/CDC domain at pH 7, pH5.5 and pH 4 and find it to be remarkably insensitive to acidification. We have shown that the ASTN2 endodomain binds inositol phosphates at a binding pocket located between the EGF-4 and Fn(III) domains and that the annexin-like domain binds calcium, as might be expected. ASTN2 appears to be unique to vertebrates and its polypeptide sequence is extremely conserved, with 53% identity between jawless fishes and humans (last common ancestor, 485 million years ago) (**Figure 2.2**). Sequences are much more diverse among fish species than among terrestrial vertebrates, as befits their extreme antiquity as a group; in fact the sequences of land-dwelling vertebrates are remarkably constrained. It is tempting to suggest that the distinctive challenges of terrestrial life, and the fundamental role apparently played by astrotactin molecules in vertebrate nervous systems, has particularly constrained diversity on land. In any case, the isolation of astrotactins to vertebrates suggests they do play a key role specifically in the development of the vertebrate CNS, which provided the basis for the evolution of mammals leading, ultimately, to humans.

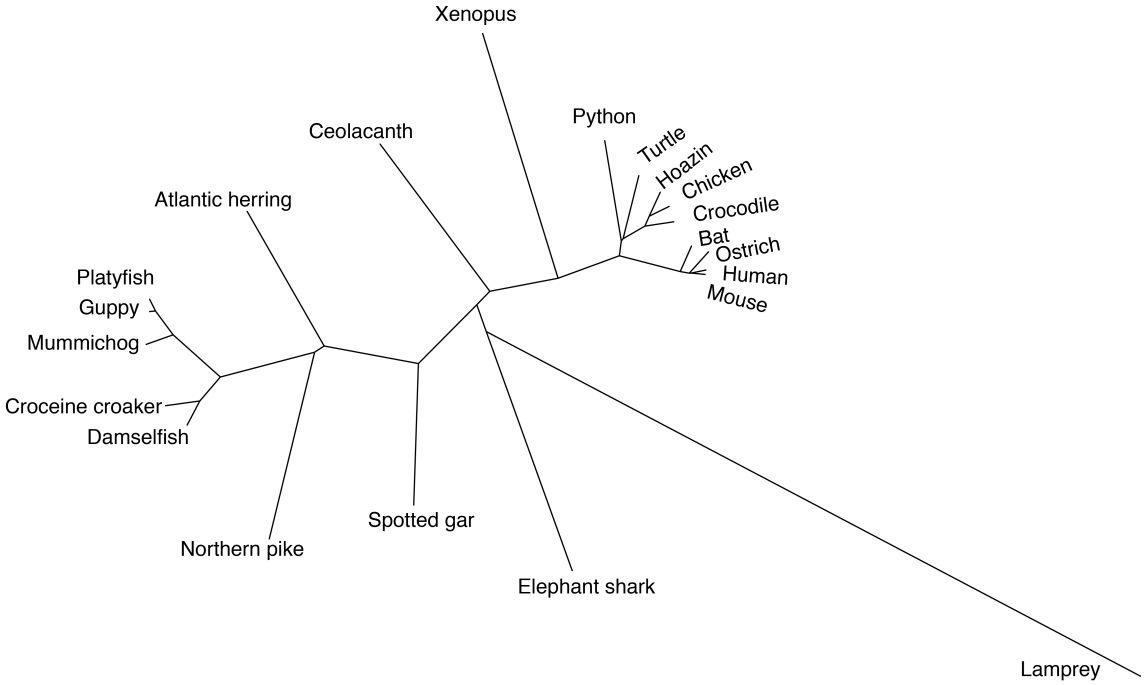

Figure 2.2: Phylogenetic tree of ASTN2 among vertebrate species. Protein amino acid sequences of ASTN2 were retrieved from NCBI where the full-length sequences were available. Sequences of ASTN2 from crocodile, elephant shark and lamprey were assembled from genome shotgun sequences. ASTN2 from lamprey appears to be the most diverse protein among the vertebrate species above, although only one version of the ASTNs was found in its genome while other species usually contain two (ASTN1 and ASTN2), probably indicating that ASTNs were split into two copies early on in vertebrate evolution. Sequence alignment was carried out in Clustal Omega and the phylogenetic tree was generated using FITCH and DRAWTREE as part of PHYLIP package (Felsenstein J. 1997).

2.3 Results and Discussion

2.3.1 Overall description of the ASTN2 endodomain structure

Figure 2.3 reports the overall structure of the ASTN2 endodomain at pH 7.5 and comprising the MACPF/CDC, EGF-4, Fn(III) and annexin-like domains. The structure was solved using single isomorphous replacement with anomalous scattering (SIRAS) with a platinum-soaked derivative (Pt derivative phasing is described detail in **Chapter 3**). The protein were purified as monomeric species and the analytical ultracentrifugation study further confirmed this property (**Figure 2.10A**). Although two further EGF repeats (EGF-2 and EGF-3) were present in the construct and in the crystallised protein (**Figure 2.4**), they were not resolved in the structure, presumably due to disorder, though small-angle X-ray scattering data indicate that the domains are folded (**Figure 2.10B** and **Figure 2.10C**). We also solved the structure of the MACPF domain alone at pH5.5 and pH4 (**Figure 2.7**); these structures are compared to the neutral pH endodomain structure below.

2.3.2 MACPF and EGF-like domain structures

The N-terminus of the endodomain structure is formed by the canonical MACPF domain consisting of a central, broken, antiparallel 4-stranded β-sheet that possesses a distinctive central $\sim$90° bend. **Figure 2.5A** shows a structure-based phylogenetic comparison of the ASTN2 MACPF domain and other known equivalent structures from the MACPF/CDC superfamily. The most distinctive features of MACPF/CDC domains are two loops containing α helices which link the core domain β strands and which undergo refolding in pore-forming family members to generate a partial or complete β-barrel inserted in the membrane (Lukoyanova et al., 2015; Tilley et al., 2005)(**Figure 2.5**). In ASTN2, the first such loop (loop 1, linking strands β1-β3 and β4-β5), comprises 19 residues and contains a single-turn 3_{10}-helix (η1) and which

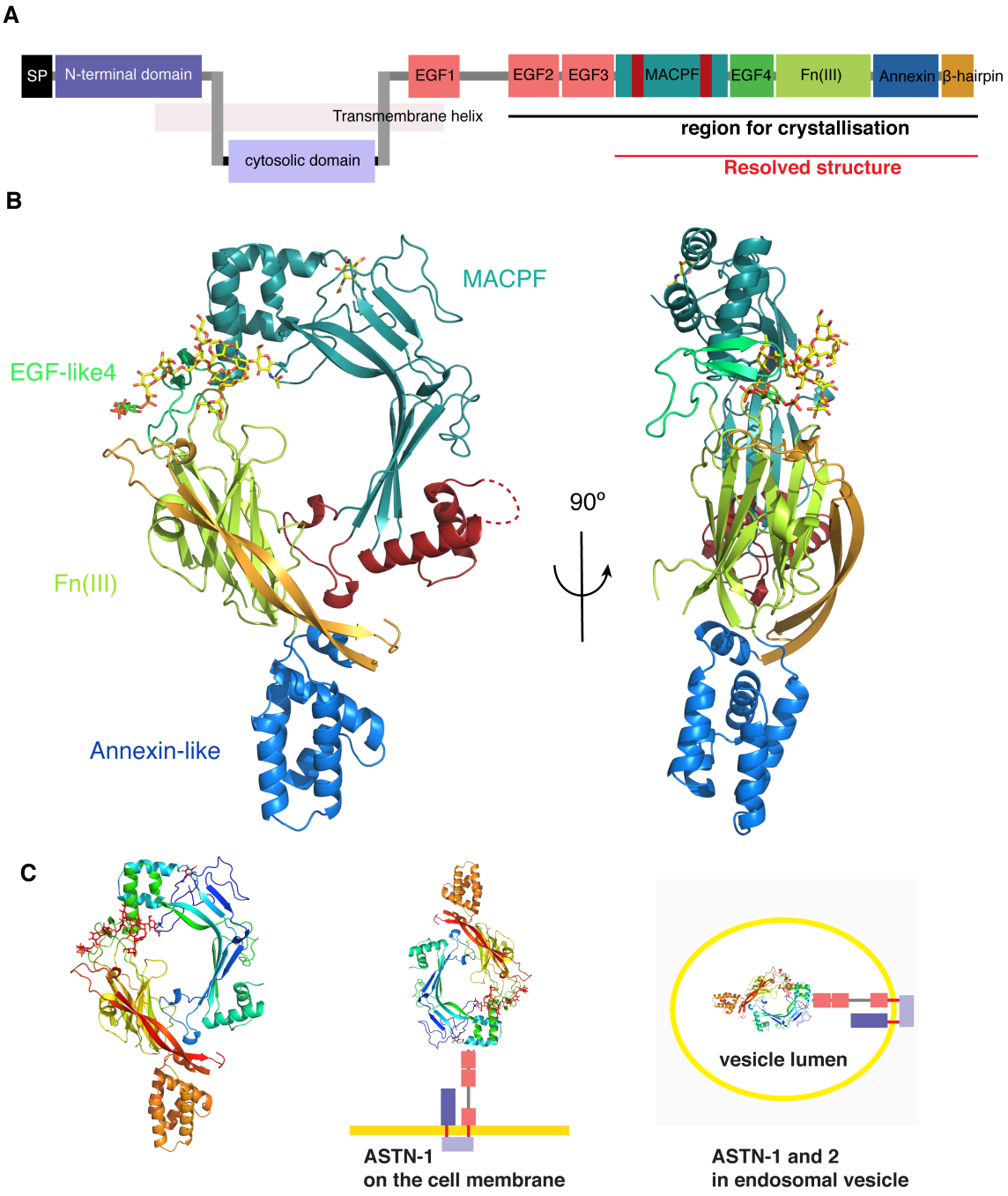

Figure 2.3: Overall structure of ASTN2 endodomain. (A) Schematic diagram of the domain organisation of ASTN2 including its two transmembrane helices and the endodomain resolved here. The black line indicates the regions of the protein expression constructs and the red line shows the region that are resolved in the final structure. (B) Two views of the ASTN2 endodomain structure related by a 90°rotation, as shown. The different sub-domains (MACPF, EGF-like (EGF-4), Fn(III) and annexin-like) are highlighted and labelled directly; loop 1 and loop 2 are shown in red and the C-terminal β-hairpin packed back across the Fn(III) domain in yellow. The dashed line indicates the amino acid residues that are disordered and not resolved in the electron density and are drawn arbitrarily. (C) The same structure as shown in B is coloured in rainbow from N-terminal to C-terminal (blue to red) to highlight the protein sequence presentation onto the structure. On the right two panels, the full-length proteins with a proposed orientation of resolved structures are shown both on cell membrane and endosomal vesicles.

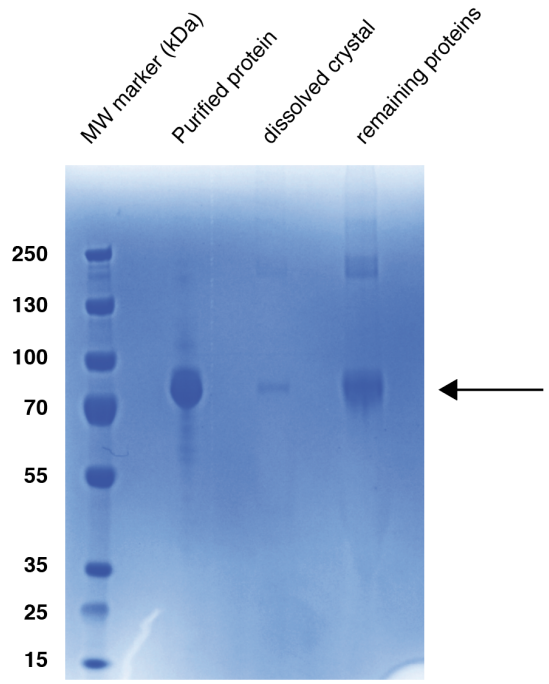

Figure 2.4: SDS-PAGE analysis of ASTN2 crystals. A relatively large crystal (20 x 40 x 80 μm) was dissolved in SDS-PAGE running buffer picked up from the crystallisation drop. Proteins remaining in crystallisation drops collected from 5 wells were prepared in the same way. The molecular weights from purified proteins and dissolved crystals were shown to be the same. The arrow indicates the protein bands with expected size.

Table 2.1: Data collection of ASTN2 native crystals

Data collection	$ASTN2_{601\text{-}1288}$	$ASTN2_{649\text{-}1288}$	$ASTN2_{649\text{-}984}$
Spacegroup	P1	$P4_12_12_1$	C2
Cell dimensions			
a, b, c (Å)	101.80, 106.03, 112.96	103.9, 103.9, 304.2	98.7, 86.6, 108.36
α, β, γ? (°)	87.21, 84.76, 65.87	90, 90, 90	90, 111.74, 90
Total observations	716701 (51576)	164862 (10473)	475699 (4647)
Total unique	18547 (1388)	6915 (490)	9653 (688)
Resolution	112.48 - 5.00 (5.13 - 5.00)	85.8 - 5.22 (5.36 - 5.22)	60.77 - 3.63 (3.70 - 3.63)
R_{merge}	0.183 (1.626)	0.12 (3.047)	0.132 (1.571)
R_{pim}	0.034 (0.272)	0.027 (0.643)	0.061 (0.711)
I/σI	16.4 (4.2)	17.3 (2.0)	8.7 (1.2)
Completeness (%)	99.9 (99.9)	100 (100)	98.8 (98.7)
Redundancy	38.6 (37.2)	23.8 (21.4)	6.6 (6.8)
CC_half	0.995 (0.904)	0.999 (0.798)	0.998 (0.509)

[a]

[a] Values in parentheses are for highest-resolution shell.

Table 2.2: Refinement statistics of ASTN2 crystals

Data collection	$ASTN2_{601\text{-}1288}$	$ASTN2_{649\text{-}984}$	$ASTN2_{649\text{-}1288}$
Spacegroup	P1	C2	$P4_12_12_1$
Resolution (Å)	98.58 - 3.16	60.77 - 3.63	85.8 - 5.22
No. reflections	627041 (18219)	475699 (4647)	164862 (10473)
Unique reflections	73326 (5382)	9653 (688)	6915 (490)
R_{work}/R_{free} (%)	23.3 (25.9)	27.7 (29.9)	32.8 (35.9)
No. of atoms			
Protein	17842	4034	4456
Ligand/ion	568	56	154
Average ***B***-factors			
Protein	135.9	196	335
Ligand/ion	224.1	217	348
RMS deviations			
Bond length (Å)	0.008	0.003	0.005
Bond angles (°)	1.18	0.67	0.85

[a]

[a] Values in parentheses are for highest-resolution shell.

packs against the Fn(III) domain interface (see below; see Supplementary Fig. 5 for sequence alignment and secondary structure). The equivalent region in perforin-1 contains 55 residues and two substantial α-helices; and in perfringolysin (and all CDCs) ~33 residues and three short multi-turn helices (**Figure 2.6** A). The loop 1 region of the ASTN2 structure is therefore an outlier with respect to MACPF/CDC proteins as a whole, due to its brevity. The second loop (loop 2; linking strands β6-β7 and β8-β9) consists of 54 residues and contains two substantial α-helices (α3 and α4). Loop 2 is solvent exposed in contrast to the packing of loop 1 against the Fn(III) domain. The equivalent region to ASTN2 loop 2 in perforin-1 contains 59 residues (4 α-helices, two short, two long) and in perfringolysin 30 residues (3 helices). Thus while loop 1 of the ASTN2 MACPF/CDC domain is unusually short, loop 2 is similar in length to that of perforin-1 and quite a bit longer than that of the CDCs. Most critically, however, the two MACPF domain loops are mismatched in length.

Two factors suggest that ASTN2 loops 1 and 2 may not function in pore formation or that, if they do, the mechanism of pore formation by ASTN2 differs significantly from that of the pore-forming members of the MACPF/CDC superfamily (e.g. perforin-1, C8α, C6, pleurotolysin, perfringolysin etc.; **Figure 2.5**). Firstly, the shortened ASTN2 loop 1, packed against the Fn(III) repeat, is too short to span a bilayer within a β-barrel structure such as the CDCs, perforin and some PLPs are known to form (Gilbert et al., 2013; Law et al., 2010; Lukoyanova et al., 2015; Tilley et al., 2005). Furthermore, while both perforin-1 and perfringolysin show an amphipathic sequence pattern in their respective loops 1 and 2, in ASTN2 this pattern is lost (**Figure 2.6A**). This is most striking in loop 1, but even loop 2 is more hydrophobic than the equivalent regions of perforin-1 and perfringolysin, and rather than a hydrophobic-hydrophilic pattern alternating every residue, the alteration seems to be two-by-two. If loop 2 inserts into a membrane to form a pore this suggests is could do so as a helical structure rather than in the form of a β-hairpin - but whether ASTN2

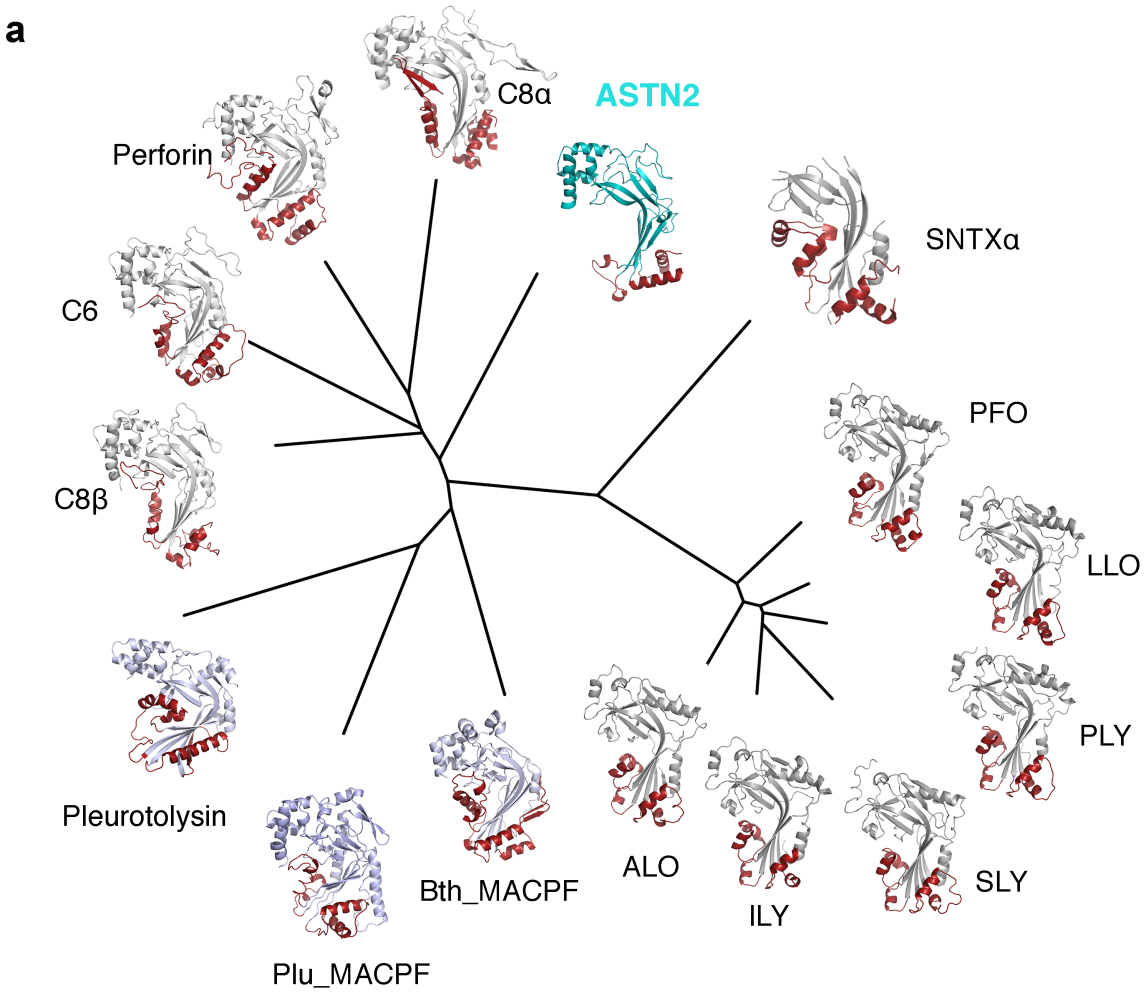

Figure 2.5: Structural phylogeny of known MACPF domain structures. The ASTN2 MACPF domain core bent sheet is coloured cyan, with loop 1 and loop 2 in red; in all other cases the core sheet is coloured grey. From ASTN2 clockwise the proteins shown are as follows: stonustoxin from Stonefish (PDB ID 4WVM); perfringolysin from *Clostridium perfringens* (1PFO); listeriolysin from *Listeria monocytogenes* (4CDB); pneumolysin from *Streptococcus pneumoniae* (4QQA); suilysin from *Streptococcus suis* (3HVN); intermedilysin from *Streptococcus intermedius* (1S3R); anthrolysin from *Bacillus anthracis* (3CQF); *Bacteriodes thetaiotamicron* MACPF protein (3KK7); *Photorhabdus luminescens* MACPF protein (2QP2); pleurotolysin from fungus *Pleurotus osteatus*; complement C8β from human (3OJY); complement C6 from human (3T5O); perforin from mouse (3NSJ); and complement C8α from human (3OJY).

really forms membrane pores is unknown. Also, while the exposed, unabbreviated ASTN2 loop 2 is similar in length to the equivalent region of perforin-1, its large mismatch in length with loop 1 (ASTN2 loop 1 is >30 residues shorter than loop 2) means that they could not pair in β-barrel formation. Yet for each other member of the MACPF/CDC superfamily, the two equivalent loops are similar in length to each other; this mismatch suggests that ASTN2 may not be a pore-forming protein, at least not in the way other MACPF/CDC protein family members are.

The MACPF domain is completed by a single-turn 3_{10} helical turn (η2) and a final α-helix (α5; 4 turns) which are paired structurally with α1 and α2 to form a head region from which the MACPF domain β-strands extend. This head is linked to the Fn(III) domain by the EGF-4 domain module, which contains β10 and β11 and matches best to the integrin β_3-subunit PSI domain EGF-like folds according to the SCOP database (family g.3.11.6) (**Figure 2.6B**); the two disulphide bridges in the EGF-4 domain superpose perfectly with those in the integrin β_3 subunit PSI domain (Xiong et al., 2004). EGF repeats are also found acting as linker domains in perforin-1 itself and in complement C8α and C8β, at points which constitute hinges during the pre-pore to pore transition. To provide for comparison of the EGF repeats found in perforin, other MACPF proteins and other proteins identified as having similarly structured regions to the ASTN2 EGF-4 repeat, we performed a structural phylogenetic analysis (**Figure 2.6B**). The EGF-like repeats compared were those from perforin-1, C8α and C8β; the PSI domain of integrin subunit β_3, and an EGF repeat from P-selectin as well as EGF-4 from ASTN2. As shown, it seems that the EGF repeat found in perforin-1 itself and EGF-4 in ASTN2 share a common line of ancestry, separate from the EGF repeats found in other immune system MACPFs (C8α and C8β). ASTN2 also seems to be closer to the common point of origin of itself and perforin-1 and thus perhaps to represent a more ancient structural form; indeed the perforin-1 EGF repeat is rather an extreme structure compared to ASTN2 EGF-4 and the EGF repeats from C8α and C8β.

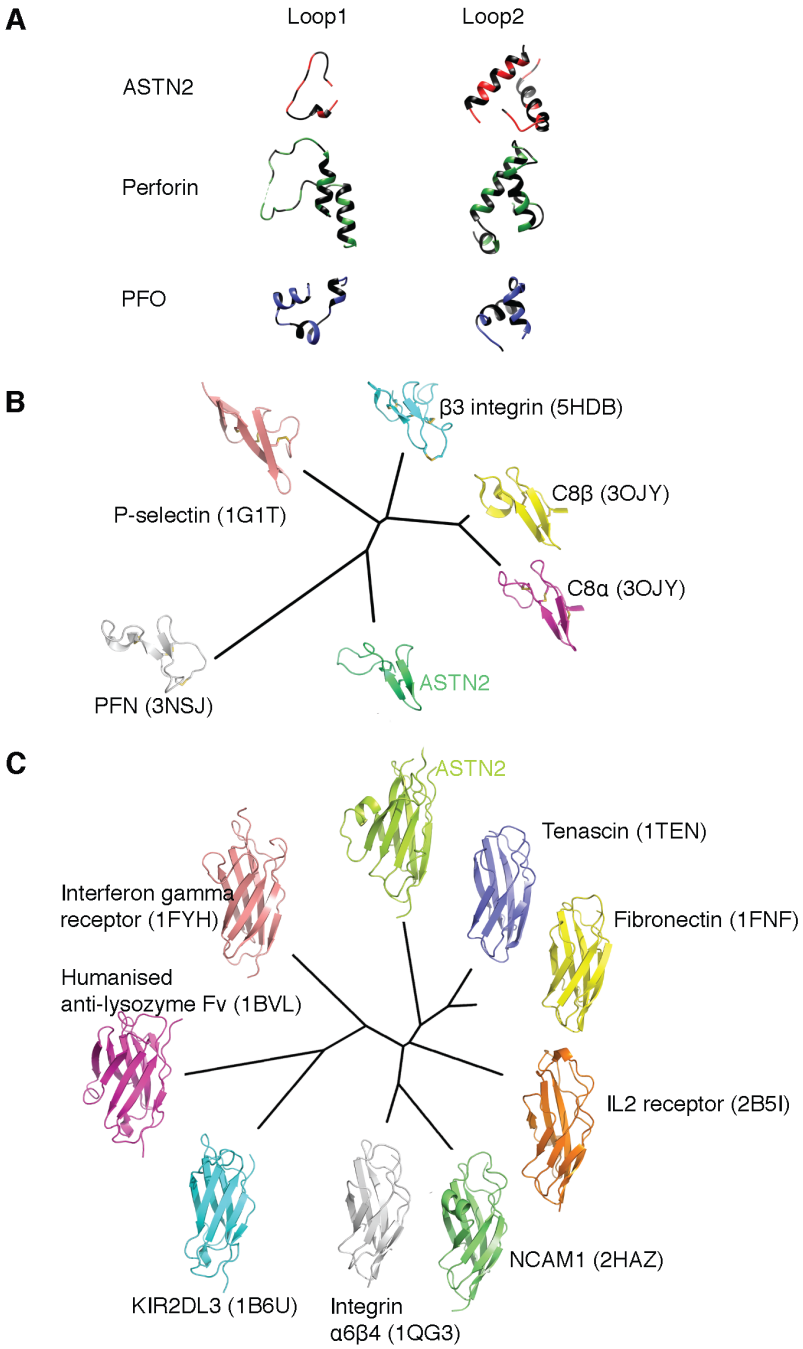

Figure 2.6: Comparison of MACPF domain and EGF-like domain. Structural comparison of ASTN2 loops and phylogenetic analysis of EGF-4 and Fn(III). (A) Loop1 and loop2 regions from ASTN2, perforin-1 and PFO showing the abbreviated loop1 of ASTN2. (B) Phylogenetic analysis of ASTN2 EGF4 domain. (C) Phylogenetic analysis of ASTN2 Fn(III) domain among other Fn(III) domains. PDB codes are shown in brackets.

The junction between the EGF-4 and Fn(III) domains contains density which is not part of the ASTN2 polypeptide and which can plausibly be modelled as an inositol triphosphate bound within a surface cavity. The binding mode of this species and mutational and binding data which support its identification with a phosphorylated inositol are described below.

2.3.3 Structure of the MACPF domain alone at pH 5.5 and pH 4

The movement of ASTN2 through the endosomal system suggests that its activity might be regulated by pH. In an attempt to investigate the effect of reduced pH on the ASTN2 structure we obtained crystals of the MACPF domain with just one N-terminal EGF-like repeat (EGF-3), at pH5.5 and pH4. EGF-3 was again not resolved in the resulting structures, and the conformation of the lower pH forms of the MACPF domain are otherwise very similar to that determined in the context of the whole endodomain at neutral pH (**Figure 2.7B** and **Figure 2.7C**) (RMSD 0.23 Å (pH 5.5) and 1.52 Å (pH 4)). Within these limited differences the region of greatest variation is loop 1 between β3 and β4, which as discussed is shortened compared to the equivalent region of other MACPF/CDC proteins of known structure, and which is packed against the Fn(III) domain. The main movement here is a shift of $\sim$2 Å by 3_{10} helix η1 towards the Fn(III) domain. Yet overall the MACPF domain seems remarkably insensitive to pH which suggests that, if pH is a triggering factor in ASTN2 activity as it moves through the endosomal system, then the sensor is not in this region.

2.3.4 Fibronectin type III domain structure

The Fn(III) domain (β12-β15, π1, β16-β19 with the start of β16 and the end of β19 tied together by a disulphide bond) follows directly on from EGF-4 and has the unexpected feature that it is complemented by a 54-residue β-hairpin from the

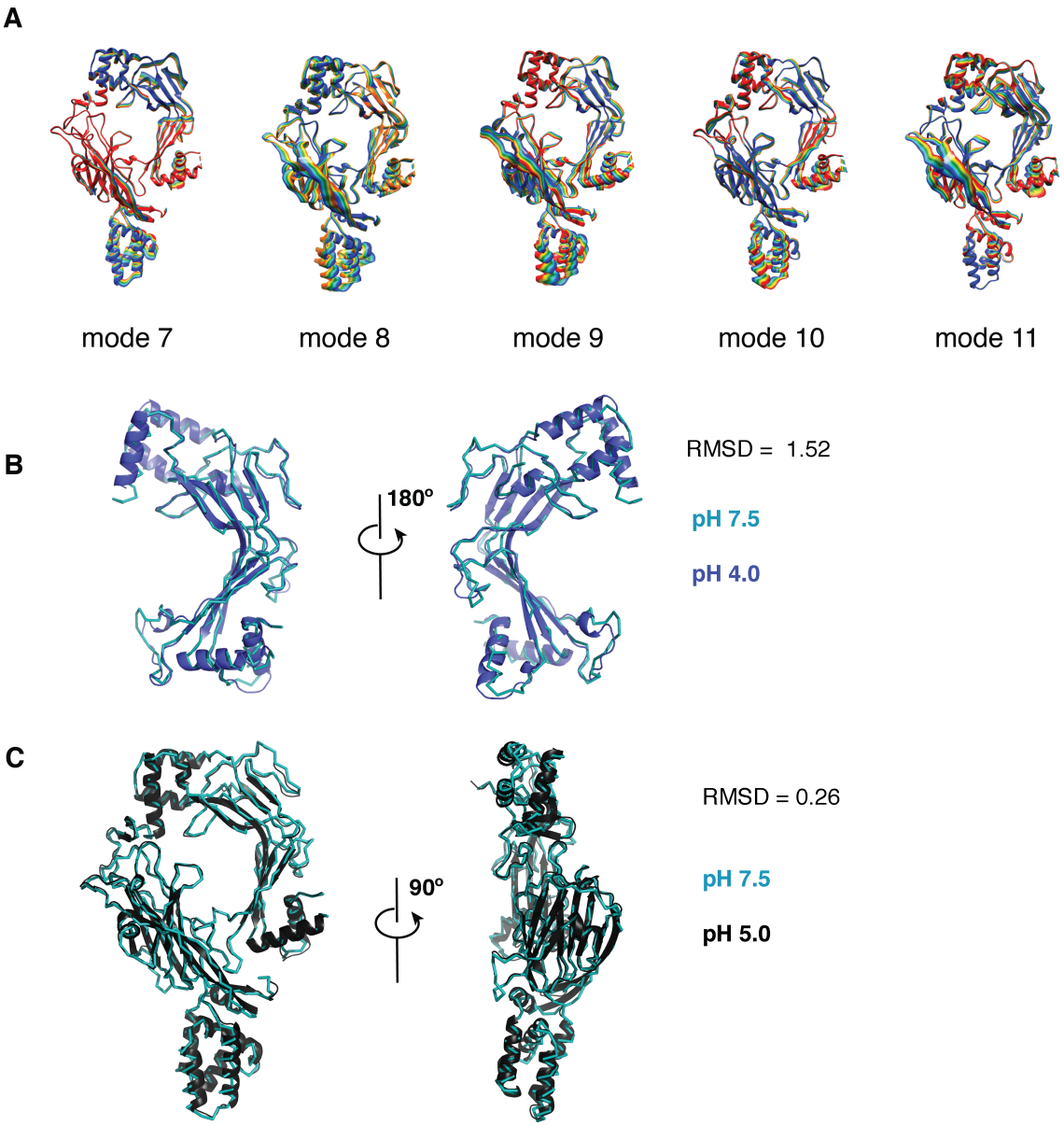

Figure 2.7: Normal modes analysis of the ASTN2 endodomain structure and ASTN2 at pH 5.5 and 4. (A) Normal modes were computed using the El Nemo web server (Suhre K and Sanejouand YH. 2004). (B) ASTN2 MACPF domain at pH 4 (blue cartoon) in superimposition with itself crystallised at pH 7.5 (cyan ribbon). (C) $ASTN2_{649\text{-}1288}$ structure at pH 5.0 (black cartoon) superimposed with itself at pH 7.5 (cyan ribbon). EGF3 domain is not resolved in the structure.

ASTN2 C-terminus (β20-β21) that folds across the Fn(III) β sandwich, clasping the edges of both sheets. As such it seems to lock the C-terminus back to the centre of the endodomain, rigidifying the structure. The loop between the hairpin strands projects towards EGF-4 and towards the inositol phosphate binding site, suggesting that this point in the structure may be a key point in its conformational regulation.

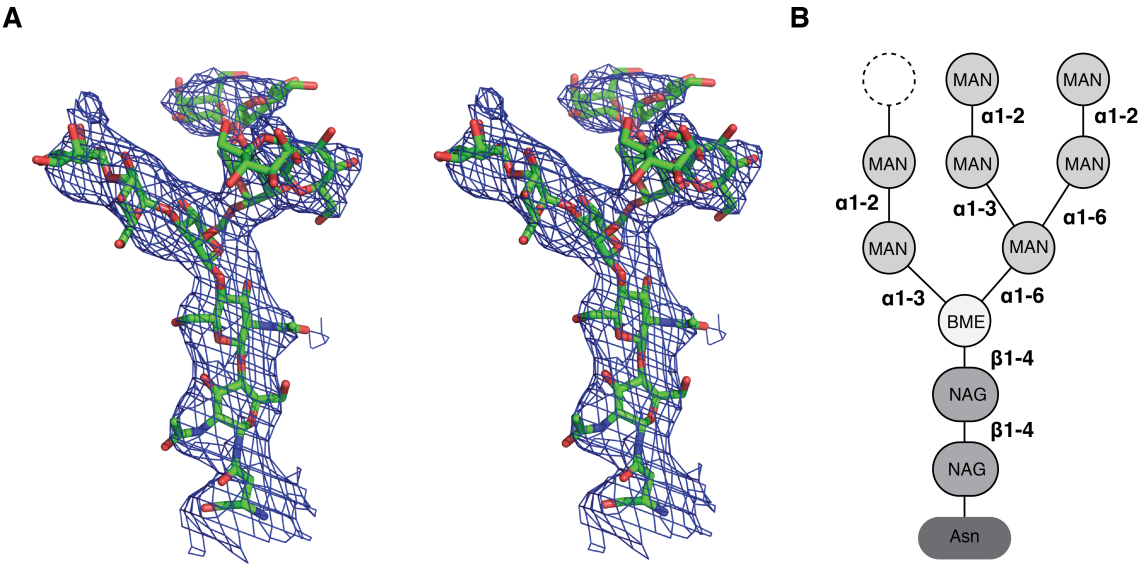

Figure 2.8: The N-linked glycosylation at N732. (A) Stereo diagram showing the ordered electron density of N-linked glycosylation at N732. The N732 residue is shown in the bottom of the electron density. (B) Diagrammatic representation of N-linked glycosylation. There are three branches of oligosaccharides, with 9 mannoses in total. The dashed circle in the cartoon represents the missing (disordered) α-mannose which is not clearly resolved in electron density. MAN: α-mannnose; BME: β-mannose; NAG: N-acetylglucosamine.

The MACPF domain contains two glycosylation sites; one at the top of the domain (N719) was trimmed back by EndoF1 (see Methods) to a single N-acetyl galactosamine residue. The other (N732) is occupied by a high-mannose glycan sidechain protected from glucosidase trimming that appears to assist in the stabilisation of the MACPF/EGF-4/Fn(III) domains cassette (**Figure 2.3B** and **Figure 2.8**). The principal contacts made by this sugar sidechain with EGF-4 and the Fn(III) surfaces are formed by the two N-acetyl galactosamine moieties and terminal mannoses of the glycan sidechain branches. The role of the glycan sidechain in stabilising the MACPF/EGF-4/Fn(III) domains cassette is indicated by the effects of a N>Q mutation knocking

out its sequon. Such a construct is not secreted during mammalian cell expression, suggesting it is misfolded.

2.3.5 Details of the interface between the MACPF and Fn(III) domains

The interface between the MACPF and Fn(III) domains is formed by the truncated loop 1 between strands β3 and β4 of the MACPF domain (**Figure 2.3B**). The residues involved in the interactions are mostly charged (e.g. Arg^{793}, Asp^{791}, Glu^{796}) though Phe^{800} projects towards the domain interface (**Figure 2.9**). Despite the involvement of charged sidechains there do not seem to be direct electrostatic contacts formed between the domain surfaces, rather the presenting faces of the MACPF and the Fn(III) domains have a shape complementarity. This suggests the possibility of breathing movements (supported by normal modes 8 and 9; **Figure 2.7A**) and even of a significant conformational change caused by movement of the MACPF away from the Fn(III) domain which might be enabled if the C-terminal β-hairpin ungrasps the Fn(III) domain. In an attempt to probe the functional significance of the MACPF/Fn(III) interface we introduced a disulphide bond to lock the two together (Asp791Cys and Thr1134Cys mutations); this resulted in a form of the protein which secreted poorly from HEK293T cells, although small-angle X-ray scattering (SAXS) data suggest that the protein that is secreted is folded like wild-type ASTN2 (**Figure 2.10B**).

Structural phylogenetic analysis of the Fn(III) domain excluding the long C-terminal β-hairpin (β20-β21) indicates that this domain is closest in structure to Fn(III) repeat 8 from human fibronectin itself, and occupies a single branch with this domain and a Fn(III) domain from tenascin (**Figure 2.6C**). It is certainly therefore a *bona fide* Fn(III) domain despite its unique complementation with the β20-β21 hairpin, for which there are no currently described structural parallels.

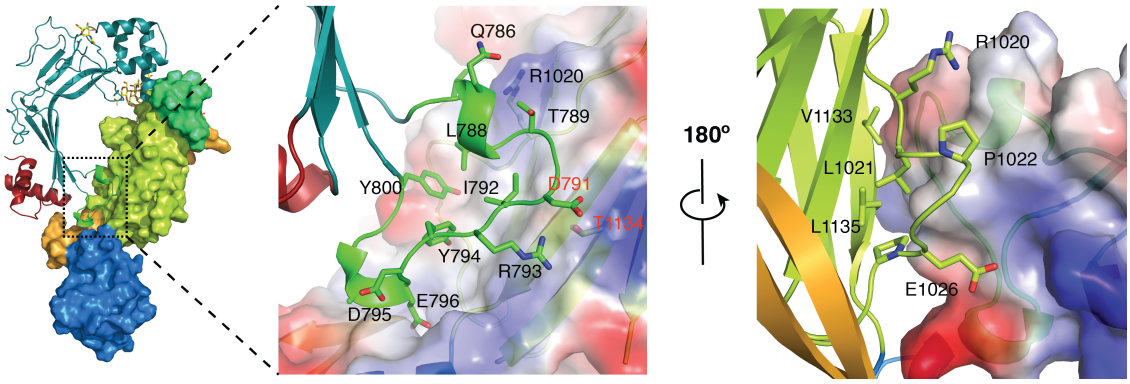

Figure 2.9: Close-up of the MACPF domain-Fn(III) domain interface. On the left is an overview in which the MACPF domain is shown in ribbon format and the EGF-4, Fn(III) and annexin-like domains with rendered surfaces. The domains are coloured as in **Figure 2.2**. In the centre is a close-up view of the MACPF domain-Fn(III) interface in the same orientation with the Fn(III) surface rendered by charge from -5 kT/e (red) to +5 (blue). On the right is an equivalent view but rotated by 180°as shown, bringing the C-terminal β-hairpin packing across the Fn(III) domain into view.

2.3.6 The annexin-like domain

The annexin-like domain is positioned between the Fn(III) domain and the long β-hairpin folded across it (β20-β21) but, functionally, forms the tip of the ASTN2 endodomain. An annexin-like domain is an unexpected feature of the ASTN2 structure (α6-α11) which despite undetectable sequence homology with annexin itself superimposes very well on an annexin single repeat (**Figure 2.11A**; see **Figure 2.12A** for a superimposition). Annexin itself consists of 4 tandem repeats of a common fold (Gerke and Moss, 2002) and it is striking that the ASTN2 annexin-like domain is closer in structure to human annexin repeat 3 (RMSD 1.45 Å) than human annexin repeat 3 is to repeat 1 (RMSD 1.58 Å). Thus, despite no sequence homology ASTN2 and human annexin have very similar structures; their charge distributions however are very different (**Figure 2.12B**). For example, with the Annexin V domains and the ASTN2 annexin-like domain all oriented equivalently, the most-similar Annexin V domain 3 presents a highly positively charged face whereas ASTN2 presents a mixed face with regions of positive and negative charge.

Like annexin itself, the ASTN2 annexin-like domain binds calcium, resulting in a

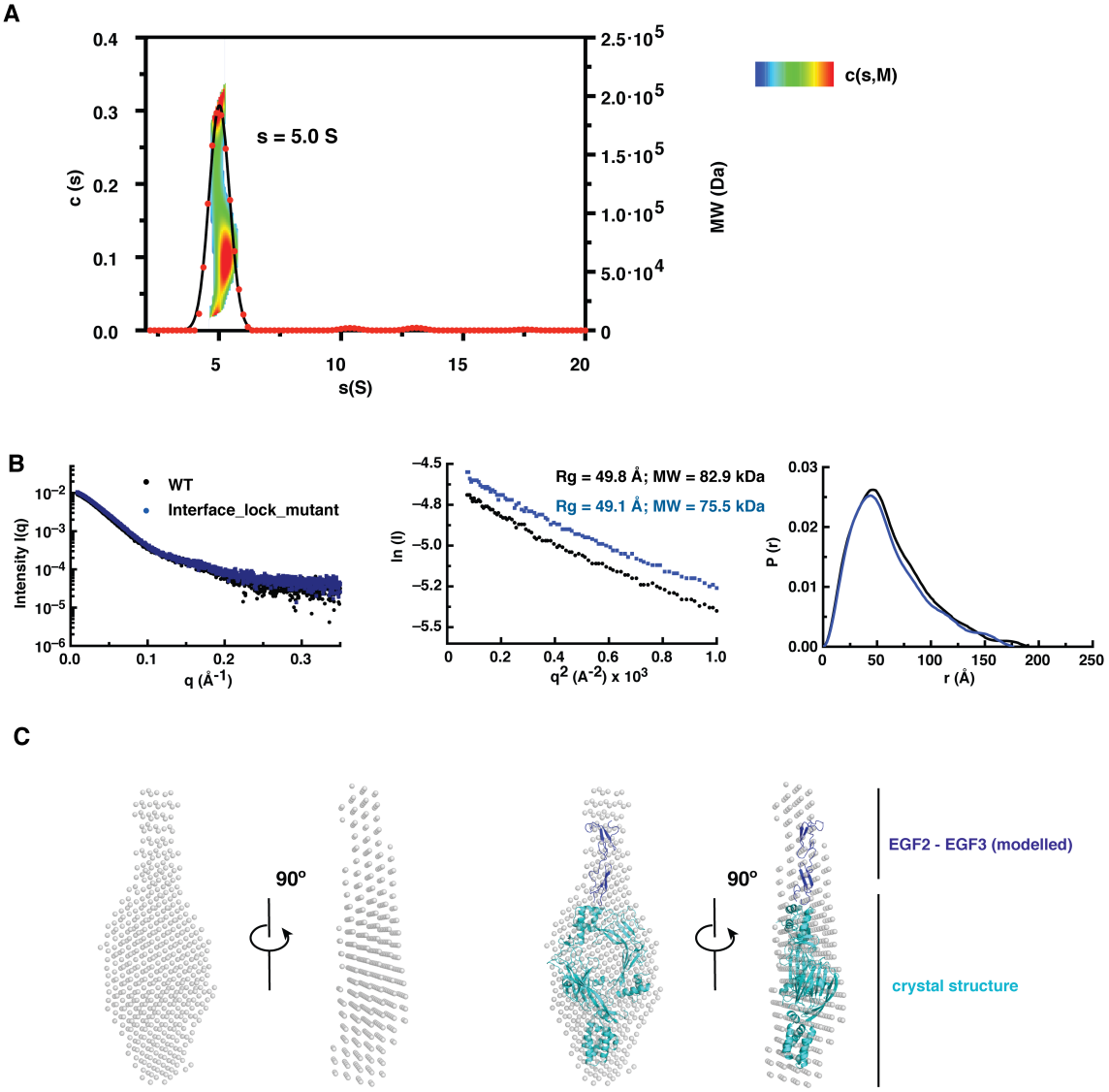

Figure 2.10: Analytical ultracentrifugation (AUC) and SAXS analysis of ASTN2. (A) AUC analysis of $ASTN2_{601\text{-}1288}$ WT showing the monomeric property of the protein in solution, with sedimentation coefficient of 5.0 s. The contour map represents the distribution of species at different sedimentation coefficients and molecular weights and the red centre shows that the majority of the proteins have a sedimentation coefficient of 5 S and calculated molecular weight around 75 kDa. (B) SAXS analysis of $ASTN2_{601\text{-}1288}$ WT and interface locking mutant. The scattering curves of protein samples at the same concentration (1mg/ml) are shown here, with a similar calculated Rg value and distance distribution function. (C) *ab initio* model of ASTN2 WT calculated from the P(r) function in damfilt. The P(r) function was calculated from merged data including low-resolution data from a high-concentration sample (4.92mg/ml) and high-resolution data from a low-concentration one (1.27mg/ml).

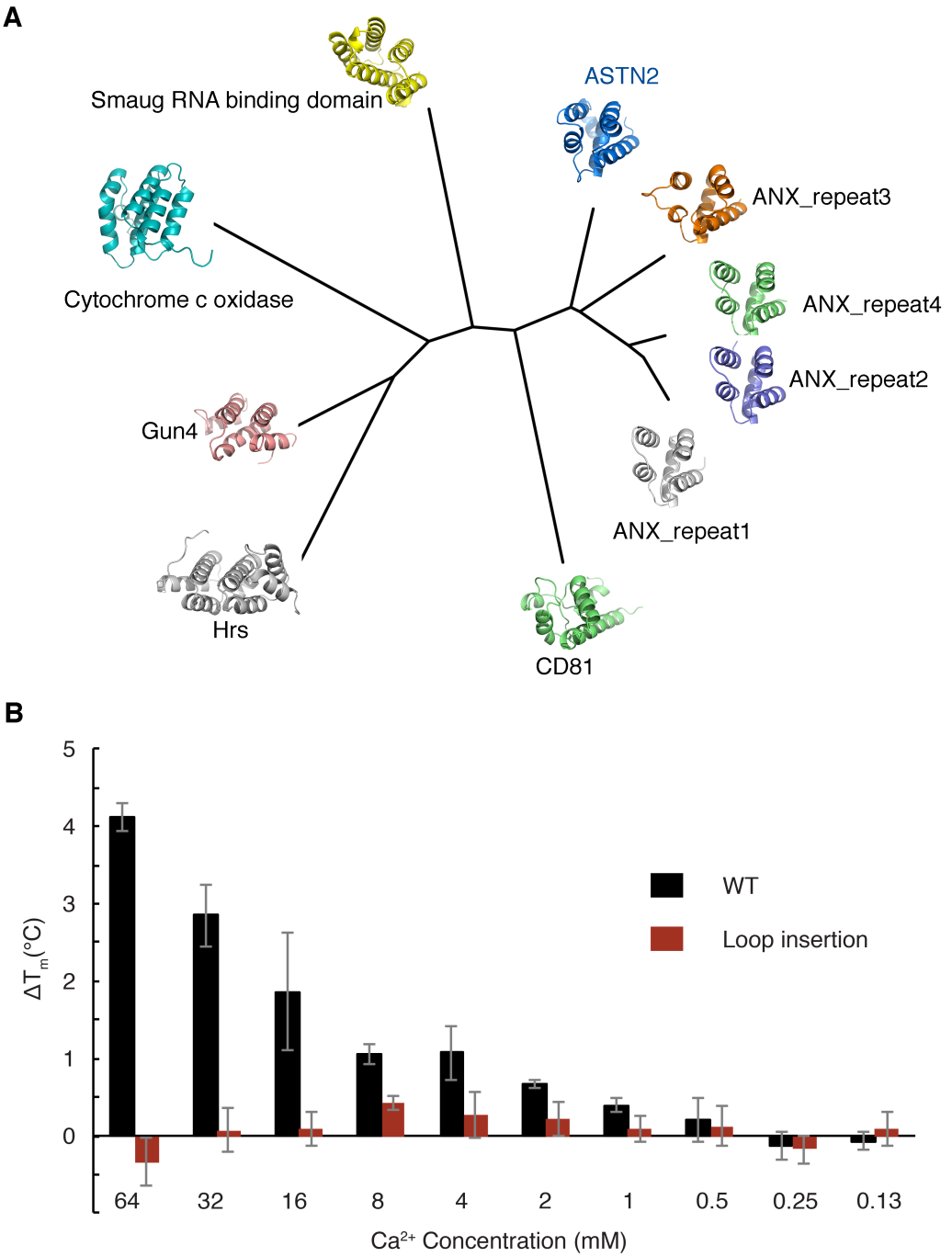

Figure 2.11: Annexin-like domain homology and calcium binding. (A) A structural phylogeny, constructed as described in the Materials and Methods, comparing the annexin- like domain of ASTN2 with close structural homologues. Clockwise from the ASTN2 domain as labelled, the structures shown are the repeats 3, 4, 2 and 1 from human annexin V (PDB ID 1AVH); the extracellular domain from human CD81 (1G8Q); a domain from Hrs1 from Drosophila melanogaster (1DVP); from Gun4 of Thermosynechococcus elongates (1Z3X); from bovine cytochrome C oxidase (1OCR); and from Smaug RNA-binding protein from Drosophila melanogaster (1OXJ).(B) Melting temperature shifts for the $ASTN2_{601\text{-}1288}$, both as wild-type sequence and with the calcium-binding motif disrupted (loop inserted). The melting temperature of wild-type protein in absence of calcium was used as the baseline. The difference of melting temperatures was calculated by subtracting the T_m of wild-type protein from that observed from each experimental conditions. Experiments were performed in triplicate and the mean ± SD is reported.

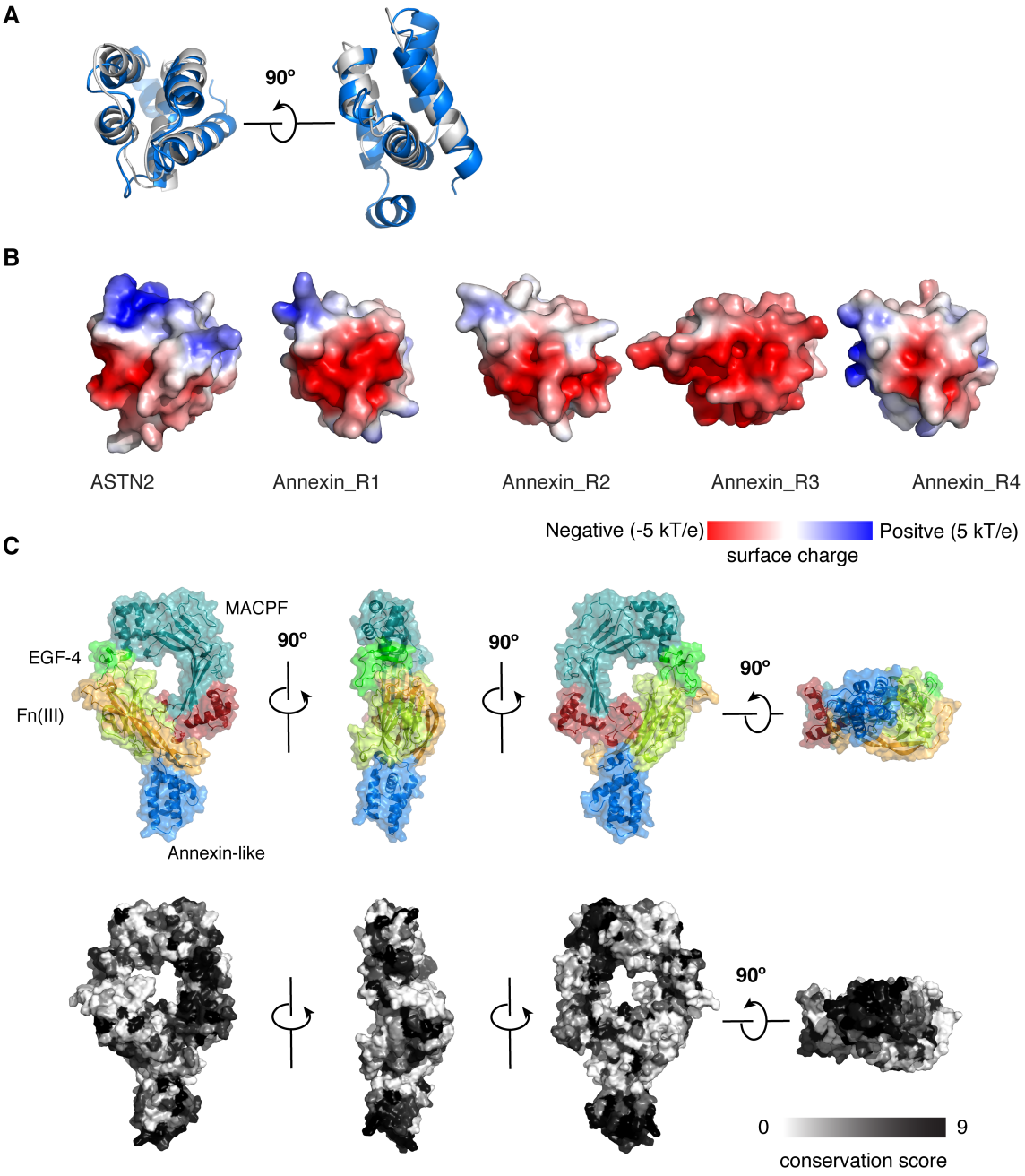

Figure 2.12: Superimposition of ASTN2 annexin-like domain and surface charge distribution comparison. (a) ASTN2 annexin-like domain in superimposition with Annexin V (PDB: 1AVR) repeat 1 domain. (b) Comparison of ASTN2 annexin-like domain surface charge with each repeat from Annexin V; each domain is shown in the same equivalent orientation as the others. The surface electrostatic potential was calculated in APBS using default settings. (C) Domain conservation calculated from sequences of ASTN2 in species the same of that in **Figure 2.2** using Consurf server (http://consurf.tau.ac.il). The conservation score was plotted from 0 to 9, with colour gradient from white to black.

stabilisation of its structure as evidenced by an increased temperature of melting (**Figure 2.11B**). This effect is confirmed by mutation of the putative calcium-binding residues within the annexin-like domain, which knock out the stabilising effect of the presence of calcium (**Figure 2.11B**). Annexin-like domains are known for their capacity to remodel membranes, triggered by calcium binding, and have also been suggested to be involved in the formation of pores in membranes (Lemmon, 2008) - both are possible biological roles of the ASTN2 annexin-like domain. The MACPF/ EGF-4/Fn(III) domains cassette is further stabilised by a disulphide bridge that caps the whole annexin-like fold, covalently linking its start (a linker region just before α6) to its end (a linker region in α11 which leads into β20).

2.3.7 Inositol phosphate binding site

As mentioned above, the crystal structure contains density consistent with an inositol triphosphate species being bound in a shallow pocket formed by loops from EGF-4 and the β-hairpin slung across the Fn(III) domain. Initially, we identified difference density of unknown origin in this area which could be modelled as an inositol triphosphate, but faced a need to demonstrate its identity. We were assisted here by the observation that although ASTN2 interacts with Ins(3,4,5)P_3 and Ins(4,5)P_2 in a surface plasmon resonance (SPR)-based experiment (K_d = 30 μM) (**Figure 2.13** and **Figure 2.14A**), ASTN1 does not (**Figure 2.14B**). A comparison of the sequences of ASTN2 and ASTN1 in this region indicated that an arginine apparently interacting with Ins(3,4,5)P_3 in the ASTN2 crystal structure is replaced by a threonine in ASTN1. Mutation of ASTN2 Arg1124 to Thr resulted in a form of the protein which, like ASTN1, did not bind Ins(4,5)P_2 and had much-reduced affinity for Ins(3,4,5)P_3 (**Figure 2.13**). It is possible that the binding of inositol phosphates is non-specific, for example that ASTN2 binds specifically to mannose-6-phosphate instead. We therefore performed a competition assay (**Figure 2.13C** and **Figure 2.14C**) to demonstrate that ASTN2 makes a specific choice

of Ins(3,4,5)P_3 over mannose-6-phosphate in our hands. We do not however know the biological significance of this interaction - which could be either that ASTN2 binds phosphorylated phosphatidylinositol lipid species, or that it binds to inositol phosphate second messengers, or to another phosphorylated or similarly modified carbon ring system.

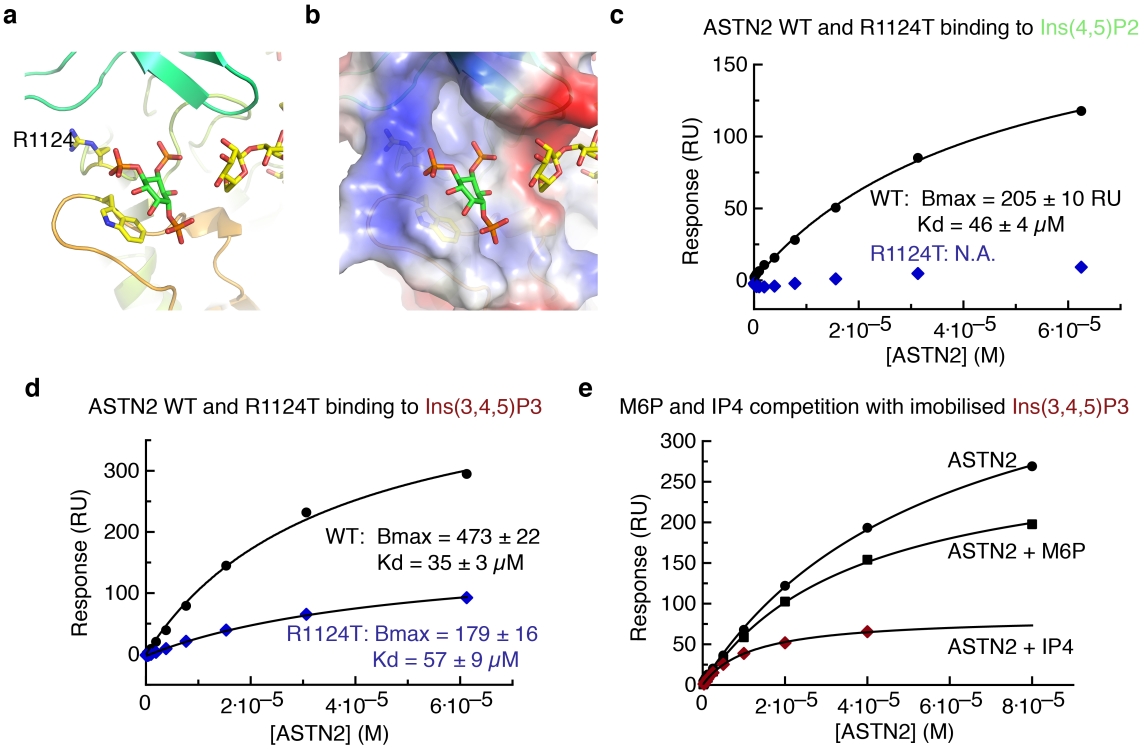

Figure 2.13: Inositol phosphate binding to ASTN2. (A) Close-up of the inositol triphosphate modelled into extra density, one phosphate in electrostatic interaction with arginine 1124, a position occupied by a threonine in ASTN1 and the inositol ring π-stacked with tryptophan 1259. (B) Surface charge representation of the inositol triphosphate binding pocket. (C) Wild-type ASTN2 and an R1124T mutant interacting with immobilised inositol diphosphates, as measured by surface plasmon resonance (SPR). (D) Wild-type ASTN2 and an R1124T mutant binding to immobilised inositol triphosphate, again measured by SPR. (E) A competition assay for the binding of ASTN2 to the triphosphorylated inositide alone and with competition from Ins(1,3,4,5)P_4 and mannose-6-phosphate in solution. The affinity of ASTN2 for the immobilised Ins(3,4,5)P_3 is much less affected by the presence of mannose-6-phosphate in solution than by the presence of free Ins(1,3,4,5)P_4.

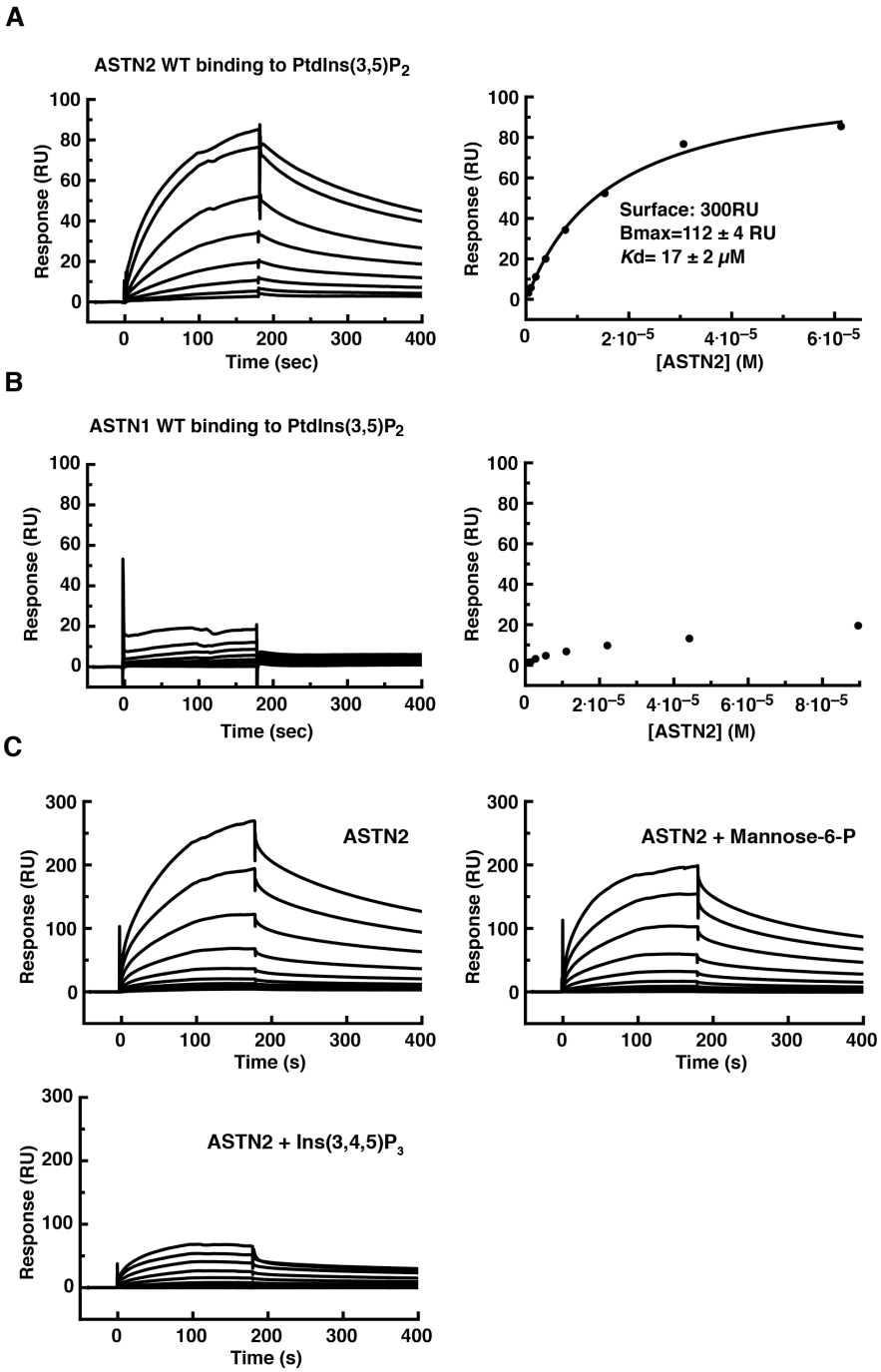

Figure 2.14: SPR experiments of ASTN1 and ASTN2 interactions with inositol phosphate species. (A) SPR experiment showing interaction of PtdIns(3,5)P_2 with ASTN2$_{601\text{-}1288}$; left: SPR sensogram, right: fitted data. (B) SPR study of PtdIns(3,5)P2 with ASTN1. The ASTN1 construct here was composed of the equivalent regions of ASTN2 construct. (C) Sensograms of SPR competition experiments showing Ins(3,4,5)P_3 headgroup competes with immobilised PtdIns(3,4,5)P_3 binding to ASTN2 (residues 701 - 1288) while mannose-6-phosphate doesn't. Removing the EGF2-EGF3 domains (residues 601 - 700) does not affect ASTN2 binding to PtdIns(3,4,5)P_3.

2.4 Conclusion

Here we have described the structure of a major portion of ASTN2, the first structure from an integral membrane MACPF/CDC domain-containing protein. ASTN2 is closely related to not only ASTN1 but also the BRINP proteins which, similarly, function in neural cell migration and guidance. Other integral membrane MACPF/CDC proteins include perforin-2, a macrophage-specific form of perforin involved in bacterial killing (McCormack et al., 2015a, 2015b), and torso-like, which plays a role in early development of the *Drosophila* embryo (Johnson et al., 2015). The structure we describe is of relevance for understanding ways in which both closely-related proteins (ASTN1 and the BRINPs) and also those more distantly related might function biologically. For example, we have shown that the MACPF/CDC domain of ASTN2 does not equip it to form pores by a mechanism similar that of perforin-1 and the complement MAC, or bacterial members of the family. We have also shown how additional domains associated with the MACPF/CDC homologous region of ASTN2 contribute to its structure and the basis on which it might act, by binding calcium (the annexin-like domain) and phosphorylated inositol species. Our structure reveals how different functions, such as membrane binding via an annexin-like domain, have been co-opted to enable the functioning of a modified MACPF/CDC domain in a novel context, as well as currently uniquely described mechanisms for bracing a protein structure by the folding of a long hairpin across the cleft of a Fn(III) domain. The stage is now set for further experiments investigating the details of ASTN2 functioning in neuronal and other cells, including its role in planar cell polarity determination, and also for the determination of structures for full-length ASTN2, for ASTN1 and the BRINPs which, like ASTN2, have modified MACPF/CDC domains without the capacity to form pores like other MACPF/CDC proteins. Without an easy analogy to pore formation by perforin/MAC or the CDCs it is impossible to say why a MACPF/CDC domain is of advantage for the activity of ASTNs and BRINPs but our structural studies provide an essential stepping stone along the way to an

answer. One possibility is that ASTN2 has a membrane fusion-type activity (Gilbert, 2016).

2.5 Methods and Materials

2.5.1 Design of constructs, cloning and mutagenesis

The full-length cDNA clone of Human ASTN2 (UniprotKB/Swiss-prot O75129, isoform 2) was obtained from Source Bioscience (UK) and the expression constructs reported here are all based on the pHLsec vector (Aricescu et al., 2006). For production of the larger portion of ASTN2's C-terminal endodomain (residues 601-1288), the construct was cloned in-frame with a Human Rhinovirus (HRV)-3C protease cleavage site, followed by a monoVenus and 8xhis tag. Other constructs (residues 701-1288 and residues 701-984) used in this study were all also cloned in-frame with histidine$_6$. Site-directed mutagenesis was introduced by overlapping PCR (Heckman and Pease, 2007). Construct sequences were verified by DNA sequencing (Source Bioscience, UK). Sequence alignments were performed using CLUSTAL Omega and displayed using Espript 3 (Robert and Gouet, 2014). Conservation of sequence was calculated using ConSurf (Ashkenazy et al., 2010) for plotting on the surface of the ASTN2 structure (**Figure 2.12**).

2.5.2 Protein expression and purification

The expression constructs were transiently transfected into HEK293T cell using polyethylenimine (PEI) as described before (Aricescu et al., 2006). All the recombinant proteins used in this study were expressed, secreted from HEK293T cells and purified from the media. For crystallisation, the proteins were produced in the presence of 5 μM of the class I α-mannosidase inhibitor, kifunensin, explaining the mannose-rich (unedited) glycan (Nettleship et al., 2008). For all other experiments,

the proteins were produced without kifunensin.

The media containing the secreted recombinant proteins were harvested 5 to 6 days after transfection, then filtered through 0.22 μm filters and dialysed against phosphate-buffered saline (PBS, pH 7.4) overnight before loading onto pre-equilibrated HisTrap HP column (GE Healthcare, 5 ml) in a cold room. The column with proteins bound were washed with 10 column volumes of Tris buffer (20 mM Tris-HCl pH 7.5 500 mM NaCl) before being eluted with 10-500mM imidazole with a gradient concentration in the same Tris buffer. GST-3C was added into the eluates to remove the C-terminal monoVenus-8xhis where necessary. The glycan chains of proteins for crystallisation were trimmed using Endo-F1. Subsequently, the eluates were buffer-exchanged into low salt buffer (20 mM Tris-HCl, pH 8.5, 50 mM NaCl) and loaded onto a pre-equilibrated Q column (GE Healthcare, 5ml) and eluted with the same buffer containing a higher concentration of NaCl. This step efficiently removed albumin contamination. Finally, proteins eluted from the Q column were applied to a size-exclusion chromatography column (Superdex 200 16/600, GE Healthcare) and homogeneous proteins were pooled together, concentrated to about 10mg/ml, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

2.5.3 Crystallisation, data collection and processing

Crystallisation screening was carried out using commercially available crystallisation reagents by vapour diffusion methods in sitting drop format in 96-well plates. For crystallisation of $ASTN2_{601\text{-}1288}$, 200 nl sitting drops of protein solutions (12mg/ml in 10 mM HEPES pH7.5, 150mM NaCl) were mixed with 100 nl of precipitant (12% PEG 20000, 150 mM KSCN, 0.1 M Bis-tris propane pH7.5 - 8.5) and equilibrated against 90 μl of crystallisation condition. The small needle-like crystal usually appeared after a week and microseeding with the needle-like crystals yielded reasonable sized crystals. For crystallisation of $ASTN2_{649\text{-}984}$, the protein was concentrated to 15mg/ml and crystallised in 0.1 M Citric buffer pH 4.0 - 4.5 20% PEG 1000. $ASTN2_{649\text{-}1288}$

was crystallised in 0.1 M MES, pH 5.0 - 6.0, 5% PEG 6000. Platinum derivative crystals of $ASTN2_{601\text{-}1288}$ were prepared by soaking the crystals with concentrated K_2PtCl_4 for 1 hour. The concentrated K_2PtCl_4 was prepared by dissolving a granule amount of K_2PtCl_4 into 10 μl of crystallisation reservoir. 25% (vol/vol) glycerol prepared in mother liquor was used as cryoprotectant for both native and platinum-soaked crystals, which were subsequently flash-frozen in liquid nitrogen. Diffraction data were collected at 100K at Diamond Light Source (Didcot, UK) on beamline I02, I03, I04, I04-1 and I24. Although most of the crystals diffracted poorly with high resolution limit up to 3.6 Å, a native dataset with higher resolution (3.16 Å) was eventually achieved by screening a number of crystals. An inverse-beam data collection strategy was employed to accurately record the anomalous signal from platinum derivative crystals. Diffraction data were indexed, integrated, and scaled using the Xia2 pipeline (Winter, 2009).

2.5.4 Structure determination and refinement

The structure of ASTN2 endodomain in spacegroup P1 is described in detail in **Chapter 3**. The crystal structure of MACPF domain in pH 4.5 was solved by molecular replacement in PHASER and refinement was carried out in PHENIX (Adams et al., 2010). The crystallographic statistics are listed in Table 1 and all models were validated with Molprobity (Chen et al., 2009). Data collection and structure determination will be described in detail in **Chapter 3**.

2.5.5 Surface plasmon resonance

SPR experiments were performed using a Biacore T200 machine (GE Healthcare Life Sciences) at room temperature in 10 mM HEPES, pH 7.5, 150 mM NaCl, PtdIns(3,5)P_2 (C-35B6a), PtdIns(4,5)P_2 (C-45B6a), PtdIns(3,4,5)P_3 (C-39B6a) and Ins(1,3,4,5)P_4(Q-1345) were purchased from Echelon Biosciences. To immobilize

the biotinylated inositol phosphate onto the sensor chip, a BIAcore CM5 chip (GE Healthcare Life Sciences) was first derivatized with streptavidin following the manufacturer's instructions, and inositol phosphates were then injected on to channel 2 and 4 of the biosensor surface, leaving channel 1 and 3 as the empty control. The analyte with 2-fold serial dilutions was applied at a flow-rate of 20 μl/min for 180 s followed by 300 s of dissociation time. The biosensor chip was regenerated by 0.1 % SDS after each running cycle. To perform the competition assay, the analyte ($ASTN2_{701\text{-}1288}$) was incubated with 10-fold molar concentration of Ins(1,3,4,5)P_4 and mannose-6-phosphate for 45 min, respectively, before serially diluted in running buffer. The data were fit with 1:1 Langmuir adsorption model ($B = B_{max}C/(K_d + C)$, where B is the response of bound analyte and C is the concentration of the analyte in the sample) to calculate the dissociation constant (K_d) using BIAcore BIAanalysis software.

2.5.6 Thermofluor

Thermofluor experiments were conducted using real-time PCR machine (Nettleship et al., 2008). The same amount of protein samples (4μg) with SYPRO orange stain (Life Technologies S-6650, final concentration 3x) was mixed with serial concentrations of $CaCl_2$ (0.13μM - 64 μM) in 10 mM HEPES, pH 7.5, 150 mM NaCl. Fluorescence measurements were recorded from 298 K to 372 K with a 1 K temperature increase each cycle. The melting temperature T_m (the midpoint of the unfolding transition) was calculated from the melting curve.

2.5.7 Small-angle X-ray scattering

SAXS data were collected on B21 at the Diamond Light Source (Didcot, UK). The measurements were carried out at 293K in 10 mM HEPES, pH 7.5, 150mM NaCl buffer with a momentum transfer range of 0.004 $Å^{-1} < q < 0.45$ $Å^{-1}$, where q =

$4\pi\sin(\theta)/\lambda$ and 2θ is the scattering angle. SAXS datasets for WT ASTN2 were collected at 4 concentrations: 4.92 mg/ml, 2.49 mg/ml, 1.27 mg/ml and 1 mg/ml, and the disulphide bond locking mutant was measured in 1 mg/ml. 18 frames of measurements were recorded and the frames without radiation damage were averaged. The scattering intensity from buffer alone was subtracted from the averaged data to obtain the protein scattering in solution. Data reduction were carried out using ATSAS package (Petoukhov et al., 2012). For WT ASTN2, a merged dataset was obtained by combining the the low-angle part of the low-concentration dataset with the high-angle part of high-concentration dataset. The radius of gyration (R_g) was calculated from a Guinier plot using AutoRg (Petoukhov et al., 2012). Particle distance distribution function P(r) was calculated in GNOM (Svergun and IUCr, 1992) using low-resolution range of dataset (0.01 $Å^{-1}$ < q < 0.15 $Å^{-1}$). An *ab initio* model was then calculated from the P(r) function using dammif and dammin (Franke and Svergun, 2009). EGF-2 and EGF-3 domain structures were predicted using Phyre2 using tandem EGF domains of Del-1 as a threading model (PDB code: 4D90, (Schürpf et al., 2012)). The crystal structure of ASTN2 was then fitted manually in Pymol.

2.5.8 Analytical ultracentrifugation

Sedimentation velocity experiments were performed using a Beckman Optima XL-I analytical ultracentrifuge equipped with absorbance and interference optics. Double-sector 3mm centrepieces were used with protein samples at 1 mg/ml and using absorbance measurements at 280nm. Experiments were performed at 20 °C, taking sample distribution scans every 6 minutes, with a total of 120 scans taken. The absorbance data were analysed using SEDFIT software using the $c(s,f/f_0)$ method of interpretation to generate sample distributions in s (sedimentation coefficient) without assuming a particular number of species (Brown and Schuck, 2006). The interference data was also analysed and shown similar results and not further used in

this study. The resulting distributions were curve fit in ProFit assuming that there is one single species with a gaussian distribution (QuantumSoft, Uetikon am See, CH).

2.5.9 Structural phylogeny calculation

Superimposition of homologous protein structures was performed using SHP (Stuart et al., 1979) as previously reported (Riffel et al., 2002). The phylogenetic tree was calculated using a pairwise evolutionary distance matrix determined from the superimposed domains. The tree representation was generated using the programs FITCH and DRAWTREE as part of the PHYLIP package (Felsenstein, 1997).

2.5.10 Normal modes analysis

Normal modes analysis was performed using the online server elNémo (Suhre and Sanejouand, 2004). The vibrational movements of the structure were visualised by superimposition of models within each of the five lowest frequency non-trivial normal modes in Chimera (Pettersen et al., 2004).

CHAPTER 3

Structure Determination of Astrotactin-2 Endodomain

3.1 Summary

Single-wavelength anomalous dispersion is one of the most commonly-used methods for *de novo* macromolecule structure determination. Recording Bijvoet pairs accurately within a reasonable resolution range is necessary for phasing and typically it requires datasets with high-redundancy to achieve high signal-to-noise ratio. However, achieving a dataset with strong anomalous signal for phasing from a single crystal is usually difficult, especially for low-resolution datasets with a relatively large cell dimension in spacegroup P1. Here we describe detailed procedures for the structure determination of ASTN2 endodomain (residues 601-1288) using platinum-derivative single isomorphous anomalous scattering (Pt-SIRAS) in spacegroup P1 at low resolution. By combining datasets collected from multiple crystals, we obtained the initial phase solution at a resolution of 6 Å using single-wavelength anomalous dispersion of Pt atoms (Pt-SAD). Density modification with non-crystallographic symmetry (NCS) and phase extension to 3.16 Å resolution resulted in an interpretable electron density map with excellent connectivity. This chapter describes the details of experimental procedures and discussions towards uncovering the crystal structure of ASTN2. It shows how phasing with a low-resolution derivative dataset is possible for the P1 spacegroup, with the help of isomorphous replacement and phase extension with native data.

3.2 Introduction

Structure determination using macromolecular X-ray crystallography requires reconstruction of phases, which are lost when recording the diffraction pattern of crystals. Generally, such information can be obtained using two approaches: molecular replacement (MR) and experimental phasing. MR phasing utilises known proteins that are structurally homologous to the proteins in of the crystals whose structure is to be determined, and in practice the success of MR phasing largely relies on the degree of similarity between the putatively homologous phasing model and the target protein. Experimental phasing with anomalous dispersion uses the anomalous scattering signal from atoms either intrinsic to the macromolecules or externally introduced by selenomethionine (SeMet) incorporation or chemical derivatisation. Such methods have become essential to the determination of the structures of novel proteins, as well as for those where an MR solution is not achievable for various reasons. The discussion below refers to single or multiple isomorphous replacement with anomalous scattering (MIRAS or SIRAS) and single-wavelength anomalous dispersion (SAD).

SAD using selenomethionine incorporated crystals is nowadays the workhorse approach for the *de novo* phasing of macromolecular crystals. However, incorporation of selenomethionine into proteins can be very challenging in mammalian cell lines due to its toxicity to mammalian cells. As a result, the yield of recombinant proteins turns out to be very low. Derivatisation of crystals with heavy atom compounds is another way to introduce a basis for anomalous signal measurement, in which crystals are exposed to heavy atom derivatives in crystallisation solutions for a certain amount of time so that the compounds can diffuse into crystals and bind to their preferential sites. For example, platinum prefers to bind to methionine, histidine and free cysteins. In practice, success of phasing by heavy-atom derivatisation relies heavily on trial and error (Joyce et al., 2010), since it is difficult to predict the

binding capacity of heavy-atom compounds to crystals, and even if some of them do bind to proteins, the binding event usually compromises the diffraction capacity of crystal by introducing distortions. Therefore, in practice, it is routine to optimise the compounds concentration and soaking time to achieve crystals diffracting to a reasonable resolution as well as containing anomalous signal strong enough for phasing. For crystals diffracting to a reasonable resolution range (d $<$ 3 Å), phasing by SAD with heavy-atom derivatised crystals is feasible. However, for native crystals that diffract at relatively low resolution ($d_{min} \geqq 3.0$ Å), heavy-atom derivatisation can significantly compromise the diffraction capacity or even completely destroys the crystals, making SAD phasing very challenging, or impossible.

With the rapid development in synchrotron beamline instrumentation, strategies of data collection and processing, and advances in software, it is now possible to phase structures which were otherwise extremely difficult (Garman, 2014; Liu and Hendrickson, 2015). In particular, it is now possible to perform *de novo* native-SAD phasing using the extremely weak anomalous signal from sulphur or other light atoms in crystals at low resolution (Akey et al., 2014; Liu et al., 2012; Omari et al., 2014). The anomalous signal could only be observed by accurately recording the Friedel mates, (*hkl*) and (-*h*, -*k*, -*l*), which is essential for successful SAD phasing, and it is very difficult for crystals that diffract poorly since the signal-to-noise ratio is rather low under these circumstances. Equipped with modern hybrid pixel array X-ray detectors that operate in single-photon-counting mode (Henrich et al., 2009), data collection with fine φ-slicing substantially improves the signal-to-noise ratio, as well as the accuracy of the anomalous signal measured (Mueller et al., 2012). Moreover, microfocus x-ray beam can effectively reduce the unnecessary background noise with the same amount of x-ray dosage (Zeldin et al., 2013). Radiation damage is another major source of difficult that hinder the accurate recording of Friedel mates. An inverse-beam data collection strategy has therefore been introduced to record the Friedel mates before radiation damage draws back the accuracy with which they

can be measured (Hendrickson et al., 1989), with the Friedel mates being recorded close in time in small wedges (5°) at φ and $\varphi + 180^{o}$. One should note that data collection in inverse beam mode takes significantly longer time compared with a normal data collection strategy because of the need to repeatedly switch back and forth by 180° as segments of data are gathered. The other approach underlying the same principle aligns the prospective mirror planes of the crystal relative to the beam so that Bijvoet mates can be recorded from the same diffraction patterns, which was made feasible with the development of goniometer-controlling software (Brockhauser et al., 2013). A high signal-to-noise ratio could also be achieved from datasets with high redundancy (Liu et al., 2011). This can generally be done by either collecting data from a single crystal from different locations with ideally microfocus beam, or combining isomorphous datasets collected from different crystals. The later approach is particularly important for crystals in spacegroups with low symmetry, in which it is very difficult to achieve high-redundancy dataset from a single crystal.

Given advances in x-ray crystallography, both in hardware and software, most *de novo* structure determination cases are feasible by SAD with heavy-atom derivatisation or selenomethione incorporation, but the challenge remains for crystals with poor diffraction. Here we present the procedure for structure determination of the ASTN2 endodomain in spacegroup P1 at moderate resolution (3.16 Å), and discuss its application to structure determination cases encountering similar difficulties. Briefly, an inverse beam data collection strategy was applied to collect diffraction patterns of Pt-derivative crystals at very low resolution. By merging data from multiple isomorphous crystals, SAD phasing was successfully conducted. Density modification and phase extension taking the advantage of high solvent content, as well as non-crystallographic symmetry, significantly improved the electron density map, even at a moderate resolution.

3.3 Results and Discussion

3.3.1 Selenium methionine SAD and sulphur SAD phasing trials

We firstly produced selenium-methionine incorporated crystals. Successful incorporation of selenium methionine in proteins was confirmed by a strong absorption at a wavelength of 0.9793 Å of X-ray (K edge of selenium) (**Figure 3.1** and **Figure 3.2A** and **Table 3.1**). These crystals generally diffracted to about 4 Å, and a weak anomalous signal was observed after merging 12 datasets collected from 3 crystals (multiplicity: 14.6). This is most likely due to the low-efficiency of selenium-methionine incorporation during protein production in mammalian HEK 293T cell line. Protein yield in the presence of selenomethionine was only about 0.1 mg from 3 L of media. An alternative explanation is that these methionines are disordered in the crystal.

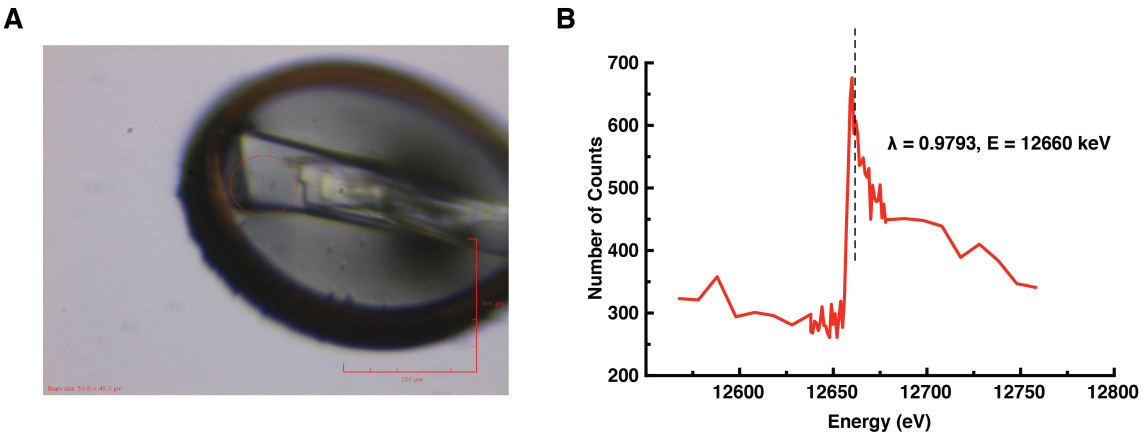

Figure 3.1: ASTN2 selenium methionine incorporated crystal and fluorescent scan. (A) Se-methinione incorporated crystal in cryo-loop with x-ray beam size attenuated to 50 x 45 μm shown in red circle. (B) Fluorescence scanning profile of Se-met crystal in beamline. The wavelength near the K-edge used in data collection was highlighted in dashed line.

We next explored the possibility of a sulphur-SAD approach to solve the phase problem encouraged by a recent report on the structure determination of HCV E1 (Omari et al., 2014a, 2014b) and NS1 (Akey et al., 2014). By combining data collected from 32 crystals, Omari *et al.* were able to solve the structure of N-terminal domain

of the ectodomain of HCV E1 using the weak anomalous signal of sulphur up to 7 Å despite the fact that the native crystals could only diffract to 3.5 Å. We conducted sulphur-SAD experiments following the same protocol (Omari et al., 2014a). The crystal-detector distance was set to cover diffraction up to 5 Å and data with total oscillations ranging from 360° to 720° were collected from each crystal in inverse beam mode. Weak anomalous signal up to 10 Å was observed after merging data collected from the two best crystals with a data multiplicity of 6.6 (**Table 3.1** and **Figure 3.2B**). Combining more datasets collected from other crystals did not significantly improve anomalous signal, very likely due to a lack of isomorphism among crystals. High-redundancy is critical for structure determination by sulphur-SAD (Liu et al., 2012, 2013), especially for datasets at low resolution. However, this is particularly difficult to achieve in spacegroup P1 where data collection with total oscillation of 360° could only result in a multiplicity of about 3.3. Therefore, a significant amount of crystals and beamtime would be required for S-SAD phasing, and not further pursued.

Table 3.1: Data collection for SeMet- SAD, S-SAD and initial Pt-SAD

Data collection	Se-met	S-SAD	K_2PtCl_4 - soak (Initial screening)
Spacegroup	P1	P1	P1
Cell dimensions			
a, b, c (Å)	101.27, 107.91, 111.14	101.18, 107.9, 111.45	101.17, 110.40, 113.14
α, β, γ (°)	87.23, 84.34, 65.73	87.23, 84.51, 65.65	87.51, 84.62, 65.10
Resolution	98.37 - 3.92 (4.02 - 3.92)	98.37 - 5.00 (5.13 - 5.00)	112.64 -7.00 (7.18 - 7.00)
R_{merge}	0.158 (0.365)	0.081 (0.165)	0.242 (0.923)
R_{pim}	0.042 (0.223)	0.034 (0.069)	0.081 (0.281)
I/σI	20.0 (3.0)	19.1 (12.1)	12.8 (3.2)
Completeness (%)	99.9 (99.4)	98.8 (97.5)	99.9 (100.0)
Redundancy	14.6 (4.2)	6.6 (6.4)	13.9 (12.8)
CC_half	0.995 (0.880)	0.998 (0.983)	0.985 (0.835)

[a]

[a] Values in parentheses are for highest-resolution shell.

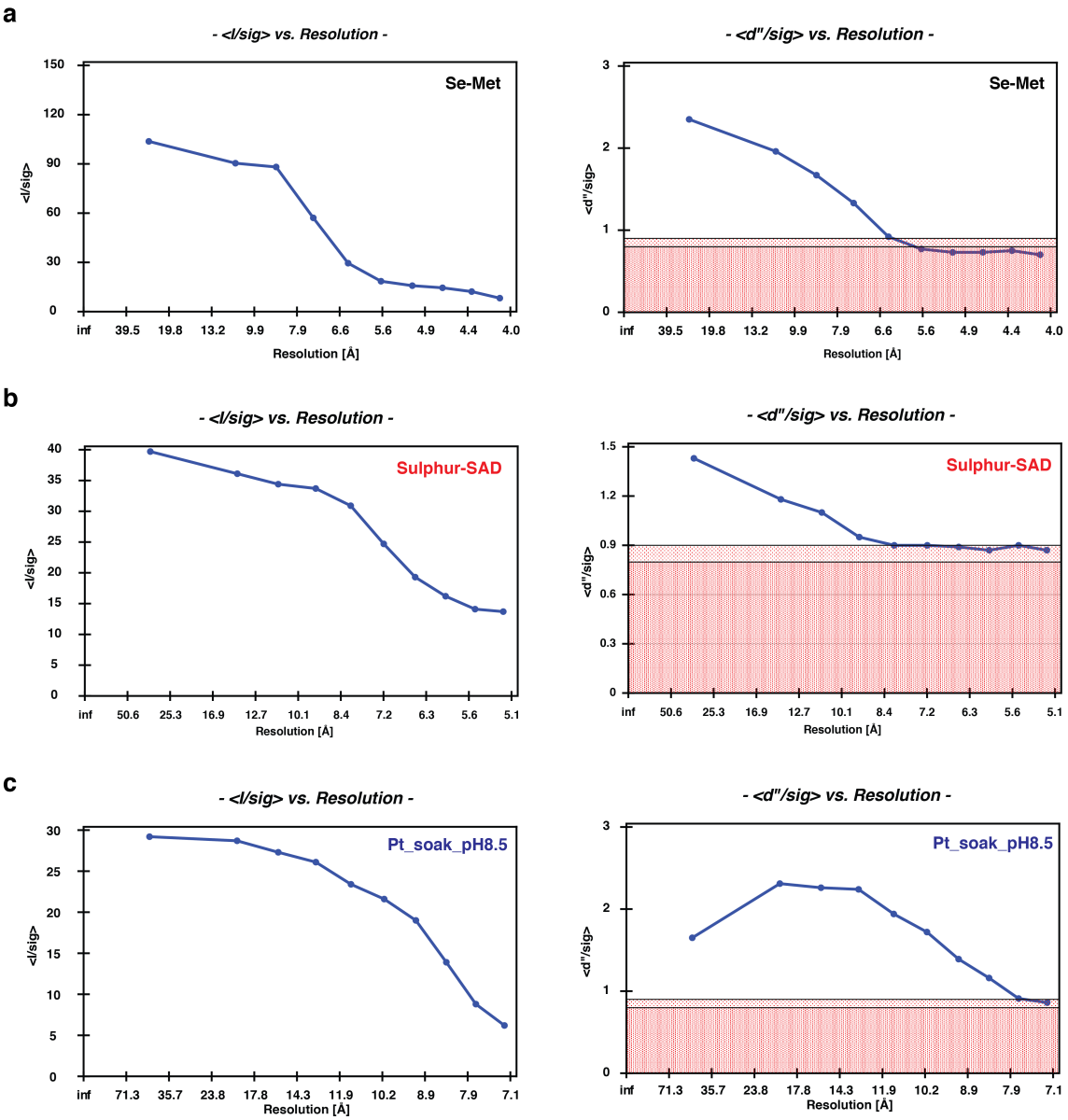

Figure 3.2: Data analysis of SeMet-SAD, S-SAD and initial heavy-atom derivative screen. (A) Overall dataset quality indicated by diffraction signal in each resolution shell. (B) Anomalous signal of merged dataset in each resolution shell. Anomalous signal from the resolution shells extending into the red part of the graph was considered to be non-existent.

3.3.2 Heavy-atom derivative screening and data analysis

A series of derivative soaking experiments were conducted. The crystals were initially soaked with concentrated K_2PtCl_4, K_2PtCl_6, $AuCl_3$, $HgCl_2$ and PbMethyI$_3$ compounds in the crystallisation mother liquor for 30 min to 2 hours. The crystal-to-detector distance were set to cover diffraction to 4 Å resolution at the detector edge since the diffraction capacity of all these crystals were compromised after heavy-atom derivative soaking. The wavelengths for each derivative data collection was tuned to theoretical L-III edge of each heavy-atom (Pt: 1.071 Å, Au: 1.040 Å, Hg: 1.009 Å and Pb: 0.951 Å). The beam transmission was set to 20 % - 50 % of total beam intensity with an exposure time of 0.2 s per frame. Alternatively, 100 % transmission power was applied with an exposure time of 0.02 - 0.05 s per frame. X-ray beam transmission and exposure time per frame were adjusted according to the size of each crystal and the recommended parameters from each beamline. The radiation damage was monitored by observing the diffraction resolution changes along with time of exposure. Typically, the crystals were able to endure X-ray exposure within 360° of data collection.

Au, Hg and Pb derivative datasets showed no anomalous signal, regardless of the number of datasets combined together (data not shown). Indeed, these crystals diffract to similar resolution to that of native (typically around 3.5 - 4 Å), which may indicate that these compounds did not bind strongly to ASTN2 crystals. In contrast, Pt derivatives could only diffract to about 7 Å. Despite its poor diffraction, a weak anomalous signal was observed after combining datasets collected from several crystals with inverse-beam data collection strategy, although the signal was not good enough for successful phasing of the structure (**Figure 1.2C**).

3.3.3 Data collection and analysis of platinum-derivative crystals

As discussed above, platinum-derivative, especially K_2PtCl_4-derivative datasets showed some anomalous signal and seemed to be the most promising heavy-atom derivative for phasing. To obtain crystals with some anomalous signal while still remaining in a reasonable resolution range, we shortened the crystal soaking time with K_2PtCl_4. We also shifted the pH of mother liquor from 8.5 to pH 6.0, considering that pH might affect the binding efficiency of platinum to proteins. In order to improve the signal-to-noise ratio of the anomalous signal, datasets from multiple crystals were merged together, resulting in datasets with high redundancy. Crystals with differences of cell dimension within 1 Å (about 1% of each axis of the unit cell) were considered to be isomorphous. However, overall, decisions on whether to include an individual dataset were based on whether the anomalous signal would be improved by including additional dataset or not. Three high-redundancy datasets were eventually obtained (**Table 3.2**), each by combining 10, 7 and 10 crystals, respectively, but only dataset Pt2 led to successful phasing. The rationale for data collection and comparison of these three datasets are discussed below.

Typically, we were able to obtain useful data with a total oscillation of 360° from one platinum-soaked crystal with optimal parameters. The first Pt-derivative dataset (Pt1: 79350 frames with a multiplicity of 38.6) was obtained by combining 44 wedges from 10 crystals soaked with the same concentration of K_2PtCl_4 for 3 hours, while the second dataset (Pt2: 60400 frames with a multiplicity of 26.3) was achieved by combining 21 wedges selected from from 7 crystals soaked with lower concentration of K_2PtCl_4 and shorter period of time (2 hours). For the third Pt derivative dataset (Pt3: 80200 frames with a multiplicity of 41.3), 23 wedges from 10 K_2PtCl_6-soaked crystals were used. All three datasets showed reasonable anomalous signal, despite the fact that their resolution could only extend to $\sim$ 5 Å (**Table 3.2** and **Figure 3.3B**). The Pt1 dataset exhibited a similar level of anomalous signal (d"/sig) compared with

Table 3.2: Platinum-derivative data collection

Data collection	Pt1 (K_2PtCl_4-1)	Pt2 (K_2PtCl_4-2)	Pt3(K_2PtCl_6-1)
Spacegroup	P1	P1	P1
Cell dimensions			
a, b, c (Å)	101.80, 106.03, 112.96	101.59, 106.93, 112.65	102.07, 105.11, 113.02
α, β, γ (°)	87.21, 84.76, 65.87	87.29, 84.68, 65.66	87.82, 85.01, 65.63
Total observations	716701 (51576)	489683 (36643)	761796 (55223)
Total unique	18547 (1388)	18616 (1439)	18432 (1377)
Resolution	112.48 - 5.00 (5.13 - 5.00)	98.37 - 5.00 (5.13 - 5.00)	112.59 - 5.00 (5.13 - 5.00)
R_{merge}	0.183 (1.626)	0.209 (1.111)	0.199 (0.966)
R_{pim}	0.034 (0.272)	0.046 (0.234)	0.034 (0.156)
I/σI	16.4 (4.2)	17.1 (5.5)	20.1 (6.2)
Completeness (%)	99.9 (99.9)	99.9 (99.8)	100.0 (100.0)
Redundancy	38.6 (37.2)	26.3 (25.5)	40.1 (40.1)
CC_half	0.995 (0.904)	0.969 (0.936)	0.928 (0.942)

[redacted]

[a]

[a] Values in parentheses are for highest-resolution shell.

that of Pt3 dataset, although the data strength (I/σ) is weaker. The Pt2 dataset showed the weakest data strength as well as anomalous signal, but it contained a higher resolution anomalous signal (extended to $\sim$ 5.5 Å compared with $\sim$ 6.5 Å in the Pt1 and Pt3 datasets).

3.3.4 SAD and SIRAS phasing with Pt derivative datasets

SAD phasing was firstly attempted using the high-redundancy Pt1 dataset, in which 25 heavy-atom sites were successfully identified with SHELXD (Sheldrick and IUCr, 2010) in HKL2MAP (Pape et al., 2004) at 7 Å resolution (20% success rate) (**Figure 3.3C**). A SAD experiment in *phenix.autosol* (Adams et al., 2010) revealed good electron density map with local RMS density correlation coefficient of 0.83. Four copies of molecules could be identified from the electron density per asymmetric unit (ASU) with careful inspection, especially from the helix-rich regions (annexin-

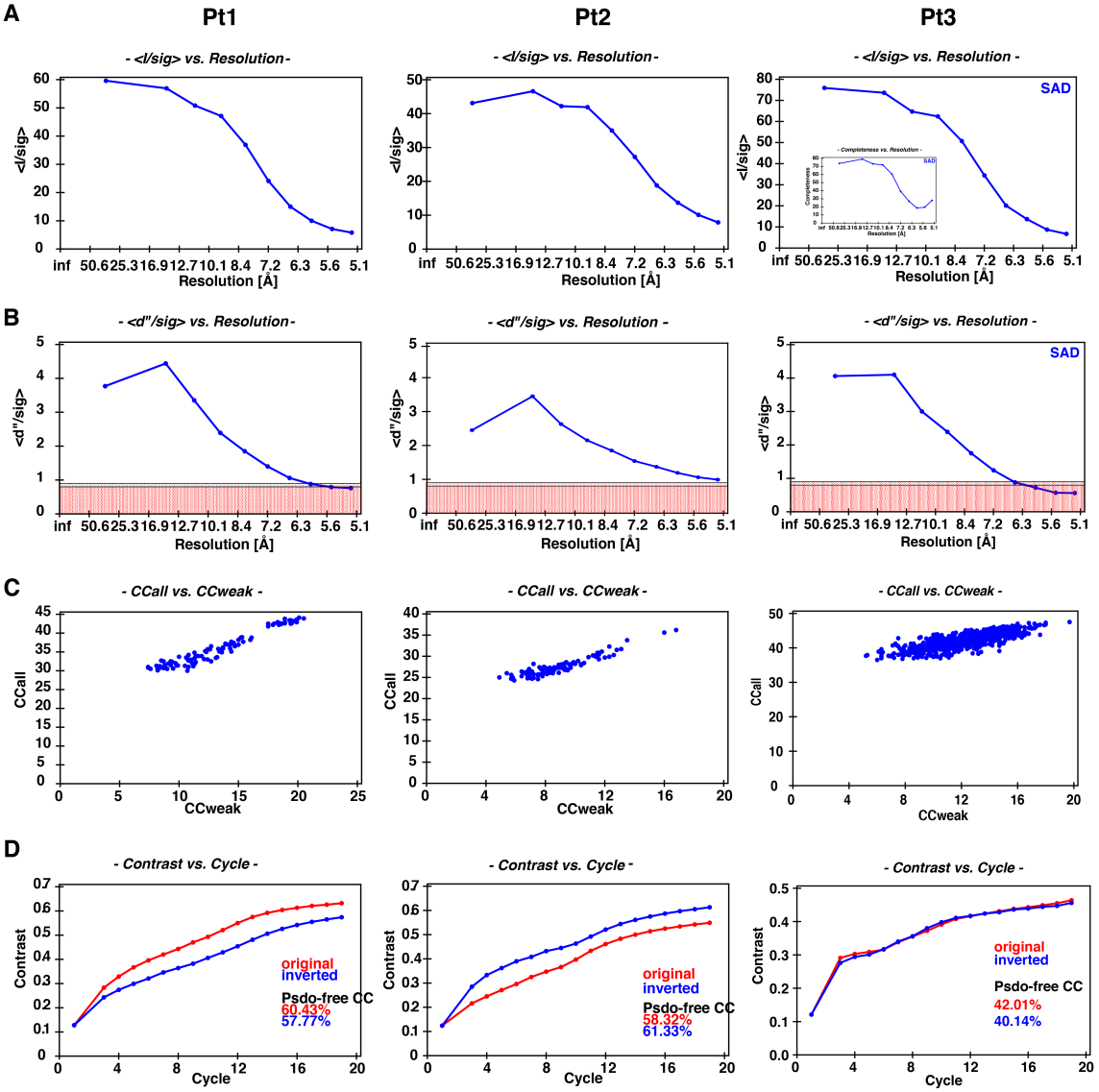

Figure 3.3: Comparison of Pt-derivative datasets for phasing. (A) Overall data quality assessed by the signal -to-noise ratio of diffraction intensity. Each dataset is combined from multiple crystals. (B) Profile of <d"/sig> as in different resolution shells. The anomalous signal showed that the Pt1 and Pt3 dataset contain higher signal-to-noise ratio. (C) Profiles of correlation coefficient between observed and calculated Bijveot differences using *SHELXD*. (D) Contrast between the electron density in the protein region and solvent region as a function of cycle number with phases calculated based on the original and inverted substructure from *SHELXE*.

like domain), although the noncrystallographic symmetry (NCS) operator was not determined correctly. However, we were not able to obtain an interpretable map using single isomorphous replacement with anomalous scattering (SIRAS) with a high-resolution native dataset, presumably due to anisomorphism between the derivative dataset and native dataset.

Pt2 dataset had lower anomalous signal, but a solution with 19 heavy-atom sites at 6.5 Å resolution cutoff (about 2% success rate) was obtained in SHELXD. SHELXE showed a clear separation in the map contrast between the two hands (0.55 vs 0.61) (**Figure 3.3D**). The heavy-atom sites were refined in *phenix.autosol* in a SAD experiment, and then isomorphous replacement with anomalous scattering (SIRAS) was performed in *phenix.autosol* using the refined heavy-atom sites. The same as observed in the Pt1 dataset, with 4 molecules per ASU, but only a 3-fold NCS operator was correctly determined initially. Density modification with the 3-fold NCS averaging and phase extension to full resolution of the native dataset resulted in an interpretable map by RESOLVE (Terwilliger, 2004). The electron density showed excellent connectivity within each molecule, although it was not possible to fit residues within it due to the limited resolution. Therefore, a partial polyalanine model of one molecule was built into the electron density map manually in Coot (Emsley et al., 2004), and translocated into the other three copies. Difference in domains arrangement was clearly observed and each domain was rigid-body fitted into the electron density, followed by rigid body refinement in *phenix.refine* (Adams et al., 2010) with external (Pt-SAD) phase restraints. The electron density was better resolved after two cycles of model building and refinement. Further density modification in *phenix.autobuild* with 4-fold NCS calculated from the 4 copies of the polyalanine model resulted in an excellent electron density map, which could then be fed into *phenix.autobuild*. The model was completed by manual rebuilding in Coot and refinement in *refmac5* and *PHENIX*. One copy of ASTN2 in the final refined structure differs from the other three, with the annexin-like domain positioned in

a way rotated about 10° compared with the other three (RMSD between copy A and D: 0.93 Å) (**Figure 3.4**), explaining the difficulty in calculating the right NCS operator at the beginning of model refinement. The crystallographic statistics are listed in **Table 2.1** and **Table 2.2**, and the model was validated with Molprobity (Chen et al., 2009).

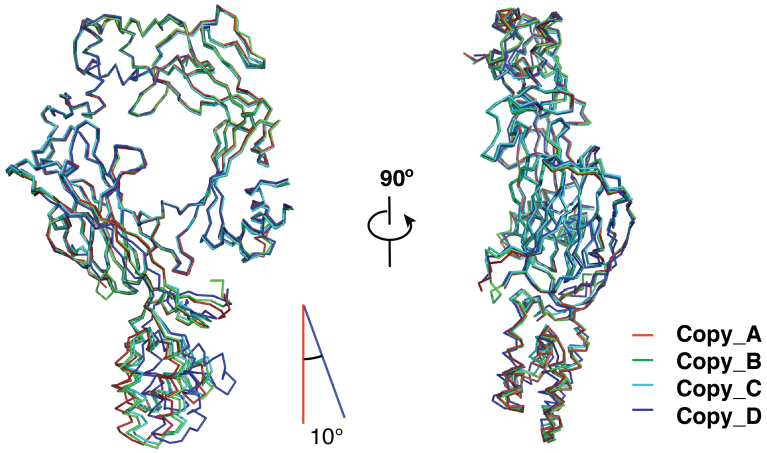

Figure 3.4: Superimposition of four copies of ASTN2 from one asymmetric unit. The C-terminal Annexin-like domain of copy_D reorientates about 10° compared with other copies. The rearrangement mainly occurs in one direction.

We tried merging the derivative dataset Pt1 and Pt2, but the merged data did not improve the anomalous signal, nor did it result in successful phase determination in a SAD experiment. This is very likely due to the anisomorphism among crystals soaked in Pt compound for differing periods of time, in which both the occupancy of Pt and cell dimension are not the same. We also investigated why Pt3 dataset did not result in successful phasing since it showed equivalent strength of anomalous signal. The Pt3 dataset was collected from crystals soaked with K_2PtCl_6 instead of K_2PtCl_4. It was unexpected that the anomalous completeness was only about 70%, compared with more than 99% completeness for the other two datasets (**Figure 3.3**). An anomalous difference map calculated with the completed structural model revealed that there is in fact stronger anomalous signal contributed by K_2PtCl_6 compound, as indicated in the electron density of anomalous difference map from Pt2 and Pt3 datasets (**Figure 3.5**).

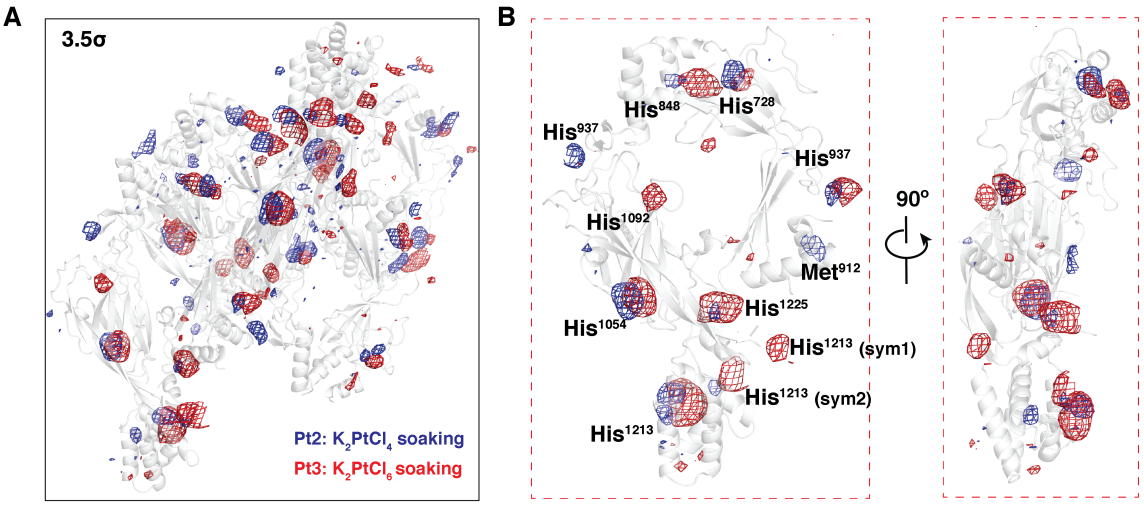

Figure 3.5: Comparison of anomalous difference maps between Pt2 and Pt3 datasets. (A) Overall view of anomalous difference map showing the positions of Pt. Electron density from one asymmetric unit was shown with four copies of molecules and coloured with blue and red for Pt2 and Pt3 dataset, respectively. (B) Close-up view of anomalous difference map from one copy of molecules, with blue from Pt2 and red from Pt3 dataset. K_2PtCl_4 and K_2PtCl_6 prefer to bind histidine residues on protein surface with different occupancies. The anomalous difference maps were generated using a final-refined model from native crystal dataset with phenix.maps and contoured at 3.5σ.

3.3.5 Phase validation

We further validated the phase determination using the datasets collected for S-SAD and SeMet-SAD (**Table 3.1** and **Figure 3.6**), respectively. Although in both datasets the anomalous signal is very low, weak electron density was still observed for some anomalous scatters (selenomethionines), which cross-validated the final refined phases for the Pt-based structure determination.

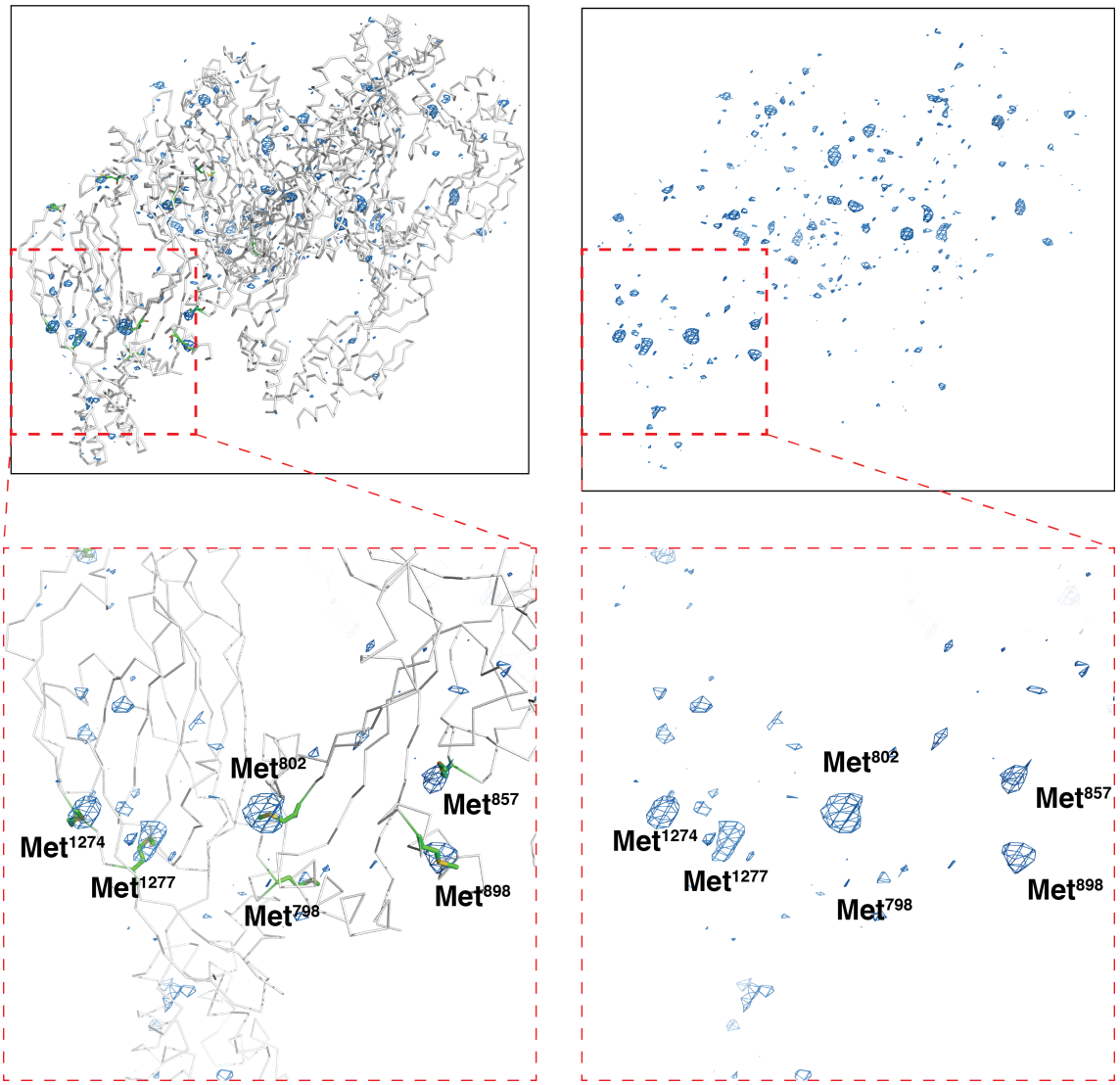

Figure 3.6: Phase validation with selenomethionine derivative dataset. The side chain of methionine residues in the final refined model is shown in stick representation, highlighting sulphur atoms in yellow. The blue mesh represents electron density of seleniums in methionine sidechains in the anomalous difference map. The anomalous difference map was calculated with the final-refined model of the native dataset. The electron density map was contoured at 3σ.

3.4 Conclusion and discussion

The phase problem remains one of major difficulties in structure determination by X-ray crystallography, especially at low resolution. Achieving a high redundancy dataset by combining data collected from multiple crystals proved to be a very useful approach for structure determination (Liu et al., 2011). In our case, even for datasets with low crystallographic symmetry (spacegroup P1) and at moderate resolutions, it was still possible to record the anomalous signal accurately and improve the signal-to-noise ratio by carefully selecting and combining data collected from multiple isomorphous crystals. Phase extension to high-resolution data was achieved by a 4-fold noncrystallographic symmetry and high solvent content.

The structure determination procedure presented in this chapter showed several key points in experimental phasing at low resolution. (i) Low-resolution anomalous signal could be useful for phasing with careful optimisation. (ii) Combining datasets collected from multiple isomorphous crystals significantly improved the accuracy of the anomalous signal, as observed in other studies (Liu et al., 2011, 2012). Importantly, only isomorphous datasets should be combined together and combining anisomorphous datasets would only deteriorate overall data quality. (iii) In practice, for preparation of heavy-atom soaked crystals, it is necessary to carefully optimise experimental conditions like total amount of soaking time and derivative compound concentration. In our case, it is preferable to soak crystals with high-concentrated compounds for a short period of time (30 min to 2 hours). Crystals were transferred into the same well for soaking with heavy-atom derivative, minimising the likelihood of anisomorphism caused by differences in crystal manipulation. (iv) Efforts in data collection trying to maximise the accuracy of recording diffraction was applied wherever possible: the distance-to-detector was set to cover diffraction up to 5 Å to minimise background, an inverse beam data collection strategy allowed the recording of Friedel mates as accurately as possible before radiation damage emerged and fine

φ-slicing benefitted from modern detector use.

3.5 Methods and Materials

3.5.1 Cloning, expression and protein purification

Human Astrotactin-2 (residues 601-1288) was cloned into mammalian transient expression vector pHLsec-monoVenus between AgeI and KpnI restriction cutting sites, resulting in a recombinant protein secreted into media with a cleavable C-terminal mVenus and His_8 tag (Aricescu et al., 2006). The construct was transiently transfected into HEK 293T cells with PEI in presence of 5 μM N-glycosylation inhibitor kifunensine (Toronto Research Chemicals, North York, ON, Canada). Recombinant ASTN2 secreted in the media was purified using the three-step purification protocol: nickel affinity chromatography, anion exchange chromatography using Q column and size-exclusion chromatography and the yield after the final purification step was about 3 mg from 3 L of cell culture media. The proteins were deglycosylated with endo_F1 before crystallisation.

3.5.2 Protein crystallisation and optimisation

Initial crystallization screen was carried out using sitting-drop vapour-diffusion methods in CrystalQuick 96-well plates (Greiner Bio-One) at 10 - 15 mg/ml by mixing 100 nL protein solution in 10 mM HEPES pH 7.5, 150 mM NaCl and 100 nL precipitant at 293 K. Crystals appeared within a week and continued to grow to full size in three weeks in 20 % PEG 3350, 100 mM KSCN, 0.1 M Bis-tris propane pH 8.5, but diffracted poorly (about 7 Å). In order to obtain better diffracting crystals, we first replaced PEG 3350 with other PEG molecules (from low-molecular weight PEG 200 to high-molecular weight PEG 20000), each with serial dilutions. A grid screen varying the concentration of PEG 20000, KSCN and the pH of Bis-tris propane

resulted in a best diffracted crystal (3.16 Å) in 12% PEG 20000, 150 mM KSCN, 0.1 M Bis-tris propane pH 7.5 - 8.5, although other crystals grown in the same condition only diffracted to 3.4 Å. To produce a good quantity of isomorphous crystals for phasing, 2 μl protein sample and 2 μl precipitant were mixed and equilibrated against reservoir solution in the presence of micro crystal seeds (Bergfors, 2003; D'Arcy et al., 2007).

3.5.3 Diffraction data collection and processing

All the crystals for data collection were cryoprotected with 25% glycerol in mother liquor prior to flash-freezing in liquid nitrogen. After screening a number of crystals, the best native data were collected on beamline i03 at Diamond Light Source (Didcot, UK) at a wavelength of 0.9796 Å, with the crystal-detector distance of 623.5 mm to cover the diffraction to 3 Å resolution. Three sweeps of data from the same crystal were merged together to obtain the final native dataset in space group P1 with relatively large unit-cell parameters. Efforts seeking other crystal forms by rescreening crystallisation conditions using shorter protein construct (residues 649 - 1288) or surface entropy reduced protein constructs predicted by SERp server (Goldschmidt et al., 2007) either resulted in crystals in the same form in similar crystallisation conditions to that in which the previous construct (residues 601 -1288) was crystallised, or in no crystals at all.

An inverse-beam data collection strategy was applied (Hendrickson et al., 1989) for data collection from all crystals for phasing. A very small rotation angles of 0.05° or 0.1° per image was used to all the data collection procedures. The percentage of transmission varied among different beamlines, but in all cases a minimum amount of exposure was applied to crystals in order to minimise radiation damage. Diffraction data were indexed, integrated, and scaled using the Xia2 pipeline (Winter, 2009).

CHAPTER 4

Functional and Biophysical Study of Astrotactins

4.1 Summary

Astrotactins (ASTNs) are integral membrane perforin-like proteins conserved across all vertebrates species. ASTN1 is localised on both cell membrane and endosomal vesicles, functioning as a neuron receptor responsible for neuron migration along a glial fibre, while ASTN2 helps recycle surface-exposed ASTN1 receptors into the endosomal system. A naturally-occurring exon5 deletion mutant of ASTN2 in a Fz6 knockout mouse has presented a ridge phenotype which recapitulated a defect involved in the non-canonical Wnt/Planar cell polarity (PCP) pathway. However, the detailed role which ASTN2 plays in the PCP pathway and how the single exon deletion affects this pathway remains unknown. Therefore, this study aims to provide structural information on ASTN2 by dissecting the full-length protein into individual domains. The cytosolic domain and the whole endosomal domain were expressed and purified from *E. coli* and human HEK 293T cells, respectively. This chapter describes the functional and biophysical characterisation of ASTNs.

4.2 Introduction

Planar cell polarity (PCP) refers to the polarisation of cells within the same cell plane and this signalling controls the global orientation of surface structures, such as cuticular plates and bristles in invertebrates and hairs or feathers in vertebrates (Chang et al., 2015; Walck-Shannon and Hardin, 2014). PCP plays a significant role in many areas of body pattern formation including convergent extension (body pattern elongation during embryogenesis), neural tube closure, axon growth and guidance, neuronal migration within the hindbrain (e.g. cerebellum), and dendritic arborisation (Gray et al., 2011; Wang and Nathans, 2007; Zallen, 2007). Genetic disruptions to PCP cause severe developmental abnormalities in vertebrates, notably neural tube defects and ciliopathies, which suggests the essential role of PCP in development (Kibar et al., 2007; Song et al., 2010; Torban et al., 2008; Wang et al., 2006a, 2006b).

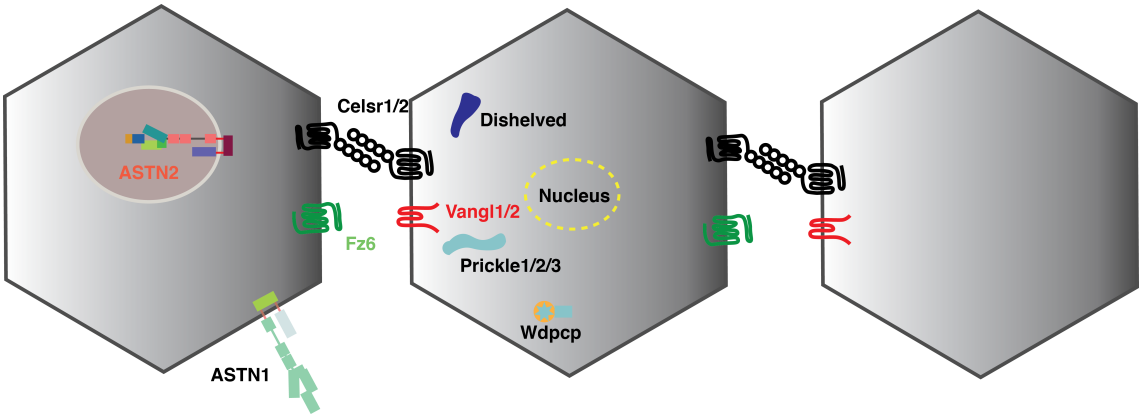

Figure 4.1: Schematic diagram of planar cell polarity pathway. Major components of PCP pathway include Frizzled, Celsr1/2, Dishelved, Vangl, etc.. Note the polarised distribution of Vangl1/2 and Frizzled; while Cersr1/2 forming junctions between cells is localised on both sides of cell. It is unknown how astrotactins are related to PCP components.

Several proteins in the core PCP pathway have been identified, including Frizzled3/6 (Fz3/6) (Hua et al., 2014; Wang et al., 2006a, 2010), Celsr1 and Van Gogh-like2 (Vangl2), as well as Dishevelled (Butler and Wallingford, 2015; Wang et al., 2006a), Prickled (Butler and Wallingford, 2015) and Wdpcp (Butler and Wallingford, 2015;

Cui et al., 2013), the latter of which are also involved in the canonical Wnt pathway (Niehrs, 2012). PCP components are initially uniformly distributed in the cell membrane or cytoplasm, and these core factors typically localise in opposite subcellular regions, generating the molecular basis for establishing morphological and functional polarisation during development (**Figure 4.1**). Asymmetric distribution of Vangl2 is one indicator of the PCP pathway and is observed ubiquitously in different tissues where PCP is established (Hemler, 2014; Merte et al., 2010; Roszko et al., 2015; Torban et al., 2008). Deletion of any one core PCP component resultes in a loss of the asymmetric distribution of all of the others, indicating a mutual interdependence among core PCP components. So it is very likely that the establishment of polarity is the result of feedback amplification of an initial directional bias, rather than a defined hierarchical cascade (Devenport, 2014).

ASTNs are integral membrane proteins in which an small N-terminal domain and a large C-terminal ectodomain (extracellular for ASTN1, endoluminal for ASTN2) are complemented by a central cytosolic domain suspended between two transmembrane α-helices (Wilson et al., 2010). ASTN2 was identified to be associated with planar cell polarity (PCP) determination in a mouse model by Jeremy Nathans and Hao Chang (Chang et al., 2015). In *Frizzled6*$^{-/-}$ (*Fz6*$^{-/-}$) mice, hair follicle orientations on the head and back are nearly random at birth, but reorient during early postnatal development and eventually form a nearly parallel anterior-to-posterior pattern. In a spontaneously-occurring mutation of a Fz6$^{-/-}$ mouse, Chang *et al* observed an unusual hair pattern, where all the follicles on the posterior half of the back had a reversed orientation, forming a hair *ridge* in the middle of the back. It was later discovered by next-generation sequencing that the phenotype was due to exon5 deletion of ASTN2 and the same phenotype could be recapitulated in a *Fz*$^{-6-}$*/ASTN2*$^{\text{exon5-/exon5-}}$ mouse model (Chang et al., 2015) .

Deciphering protein interaction networks at the molecular level is key to understanding the genetic interaction observed between ASTN2 and the PCP pathway.

ASTN1 and ASTN2 have been shown to interact with each other *in vitro* in a calcium-independent way with multiple regions of each protein contributing to their interface (Wilson et al., 2010) . The Nathans lab found that ASTN2 binds to itself (with and without exon 5 present) using co-immunoprecipitation in HEK 293T cells, and has been testing the possibility of a direct interaction between ASTN2 and the PCP-determining proteins Fz6 and Celsr1 (Hao Chang and Jeremy Nathans, personal communication). However, the detail of these interactions remains to be elucidated: it is not clear whether the two ASTNs interact directly with each other, how the self-interaction of ASTN2 occurs, and whether there exists a direct interaction between ASTN2 and the PCP pathway proteins.

This part of the study focused on biophysical characterisation of ASTNs, both the cytosolic domains and the endoluminal domains. Both the cytosolic domain and the whole endoluminal domain (endodomain) of ASTN2 were purified, and biophysical characterisation revealed that the cytosolic domain is structurally compact and the whole endodomain (comprising a smaller N-terminal region and a larger C-terminal region, separated by transmembrane segments and cytosolic domain) is linked together via a disulphide bond. Moreover, we observed a potential pH-driven self-interaction of ASTN2, which is currently under investigation.

4.3 Results

4.3.1 ASTN2 cytosolic domain is structurally ordered

A bioinformatic study of the ASTN2 cytosolic domain (residues 230 -383) predicted that this domain is largely disordered, with no predicted α-helix or β-sheet; no homologous solved protein structures could be identified. Exon deletion could potentially disintegrate the structure of a given protein. Exon 5 (residues 340 - 374) is within the cytosolic domain, immediately in front of the second transmembrane helix (**Figure 3.2A**) and exon 5 deletion is unlikely to affect protein topology

spanning the membrane. We thus wanted to know whether the absence of exon 5 would disintegrate the cytosolic domain, or lead to an incorrect distribution of ASTN2.

To express and purify the cytosolic domain, a series of constructs varying the domain boundary was prepared and transformed into *E. coli* B834 cells. Interestingly, all the constructs in a small-scale expression test showed two major protein bands (**Figure 4.2B**). Two constructs, C1 (residues 229 - 339, excluding exon5) and C2 (residues 229 - 383) were selected for large-scale expression and purification. The His_6-GST tag was removed during protein purification. In agreement with the observation from the small-scale expression screening, size-exclusion chromatography (SEC) resolved two separated peaks (**Figure 4.2C**). Native mass spectrometry (MS) showed that these two fragments have a molecular weight of 12978 Da and 8782 Da, corresponding to the theoretical molecular weights of the clone (12957 Da, residues 229 - 339) and a proteolytic fragment (8782 Da, residues 229 - 302) (**Figure 4.2D**).

We further asked whether these two fragments are structurally ordered. A circular dichroism (CD) experiment with purified proteins revealed that both fragments did not have typical secondary structures (**Figure 4.3A**). As a control, the pleckstrin homology (PH) domain of kindlin-3 showed a clear signal for β-sheets and α-helices, which matches the structure of a PH domain. Analytical ultracentrifugation study showed that the protein behaved as a typical globular protein, with a sedimentation coefficient of 1.27 S, in agreement with the monomeric property observed by size-exclusion chromatography (**Figure 4.2B**).

4.3.2 Complex of N-terminal domain and C-terminal regions through a disulphide bond

The whole endodomain of ASTN2 is composed of a small N-terminal domain (NTD, residues 50 - 205) and a large C-terminal domain (CTD, residues 401 - 1288). The

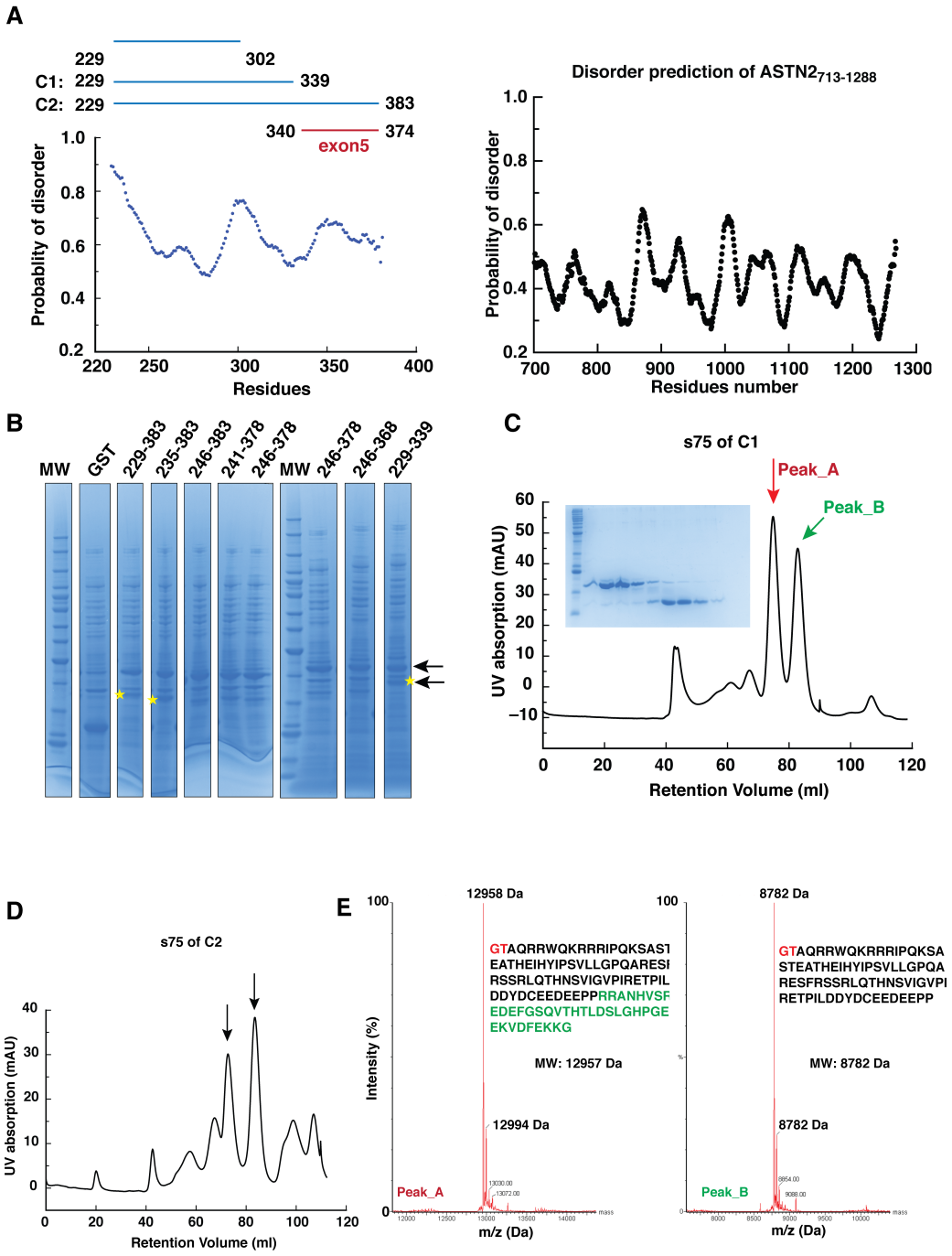

Figure 4.2: ASTN2 cytosolic domain. (A) Disorder prediction of the ASTN2 cytosolic domain (residues 229 - 383) by RONN (https://www.strubi.ox.ac.uk/RONN). Exon 5 (residues 340 - 374) is highlighted with a red bar and two constructs for further study is labelled C1 and C2, respectively. On the right panel, disorder prediction for the C-terminal domains, the structures of which were presented in Chapter 2, is shown. (B) Small-scale expression test of ASTN2 cytosolic domain constructs. The domain boundary of constructs is based on secondary structure prediction and disorder prediction and the proteolytic fragment was observed in all the constructs highlighted with a yellow asterisk. (C and D) Size-exclusion chromatography profile of two constructs C1 and C2. Two peaks corresponding to the intact construct and a proteolytic fragment are shown with two arrows. (E) Native mass spectrum of intact protein and proteolytic fragment of construct C1. The theoretical molecular weight calculated from protein sequences matches with the experimental data.

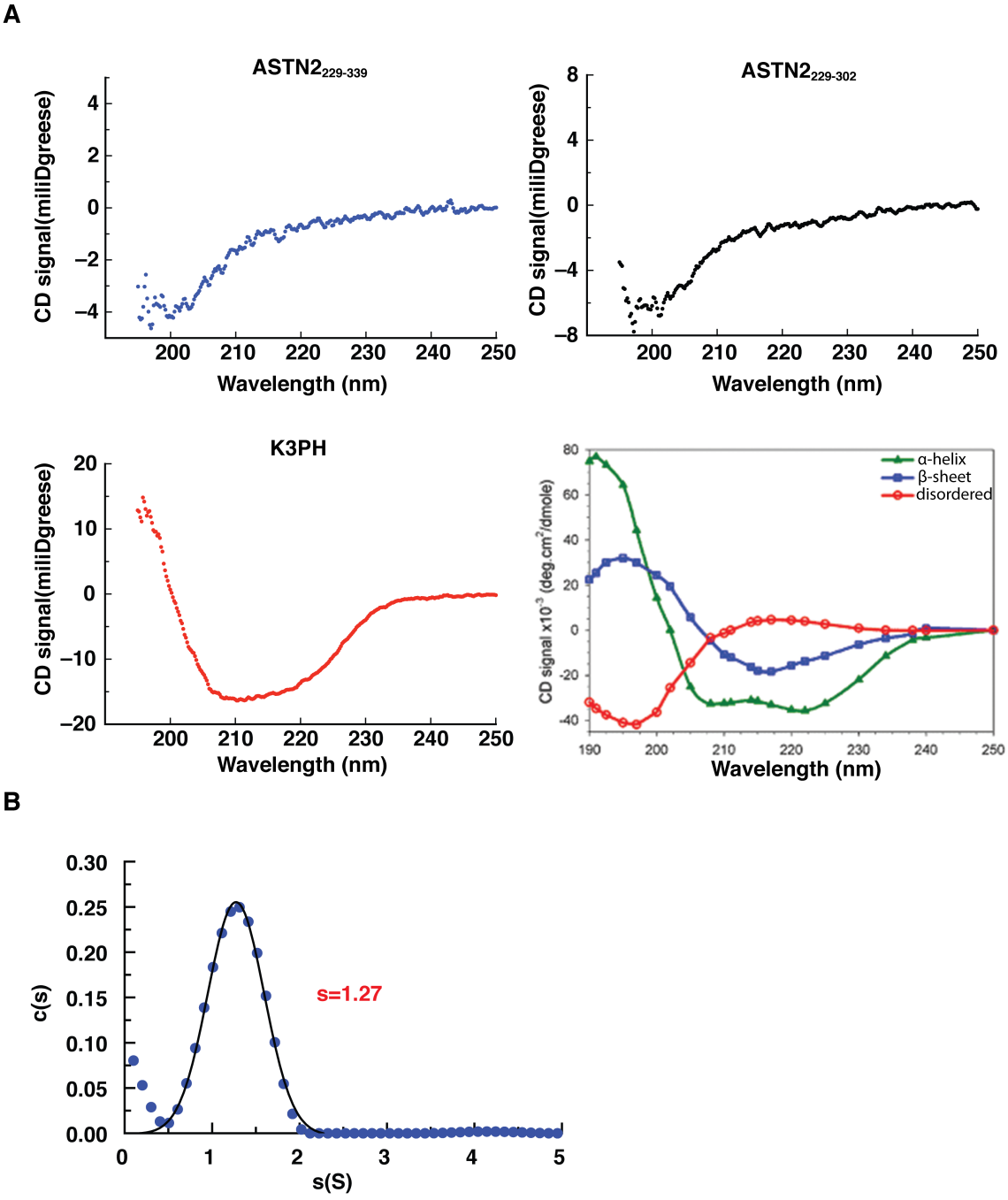

Figure 4.3: Biophysical characterisation of ASTN2 cytosolic domain. (A) Circular dichroism spectrum of ASTN2 cytosolic domain and Kindlin-3 pleckstrin homology domain at 0.1 mg/ml. The CD spectrum of standard reference from proteins with pure secondary structures was shown on the bottom right, adapted from http://www.fbs.leeds.ac.uk/facilities/cd/. (B) Sedimentation velocity analytical ultracentrifugation study of ASTN2 cytosolic domain without exon 5 at 2 mg/ml at 293 K showing a stable mono-dispersed species in solution.

NTD is predicted to be a β-sheet rich domain sharing no structural homology to known structures. The CTD comprises of a series of domains: three tandem EGF-like, MACPF, FN(III) and annexin-like domain. Besides the domains already identified, a short region (residues 401 - 600), with the first EGF-like domain inserted, is predicted to be structurally ordered although no homology model is available.. The annexin-like domain was only identified via our work, due to the solution of its X-ray crystal structure, as described in detail in Chapter 2.

Nathan's lab identified a site-specific proteolytic region in the C-terminal region of mouse ASTN2 immediately following the second transmembrane helix (data not shown), although the biological relevance is not clear yet. More interestingly, the N-terminal region and C-terminal region were linked together by a disulphide bond (**Figure 4.4A**) (unpublished data from the Nathan's lab, personal communications). Our results confirmed that the cleavage between the second transmembrane helix and the C-terminal domain also occurs for human ASTN2 (**Figure 4.4B**). In agreement with their results, we observed that the full-length human $ASTN2_{1\text{-}1288}$ construct migrated as two separated bands on a western blot, the molecular weight of each matching that of full-length and C-terminal region, respectively (**Figure 4.4B**). A shorter construct used by Wilson et al., which lacks the N-terminal domain is also shown for comparison and will be discussed later.

To better characterise the whole endodomain, we tried to co-express NTD and CTD in the HEK 293T secretion pathway although no secretion was detected in the media. One possibility for this failure of secretion is that the co-expressed endodomain is misfolded and thus entrapped in the endoplasmic reticulum of the cells. However, considering that full-length ASTN2 is found normally in endosomal vesicles while ASTN1 cycles between the endosomal system and cell surface, it is also possible that some other factor associated with its biological function impede ASTN2 secretion since the spatial regulation of these ASTNs remains unclear. If this is the case, failure of secretion does not mean that the protein is misfolded. To test these possibilities,

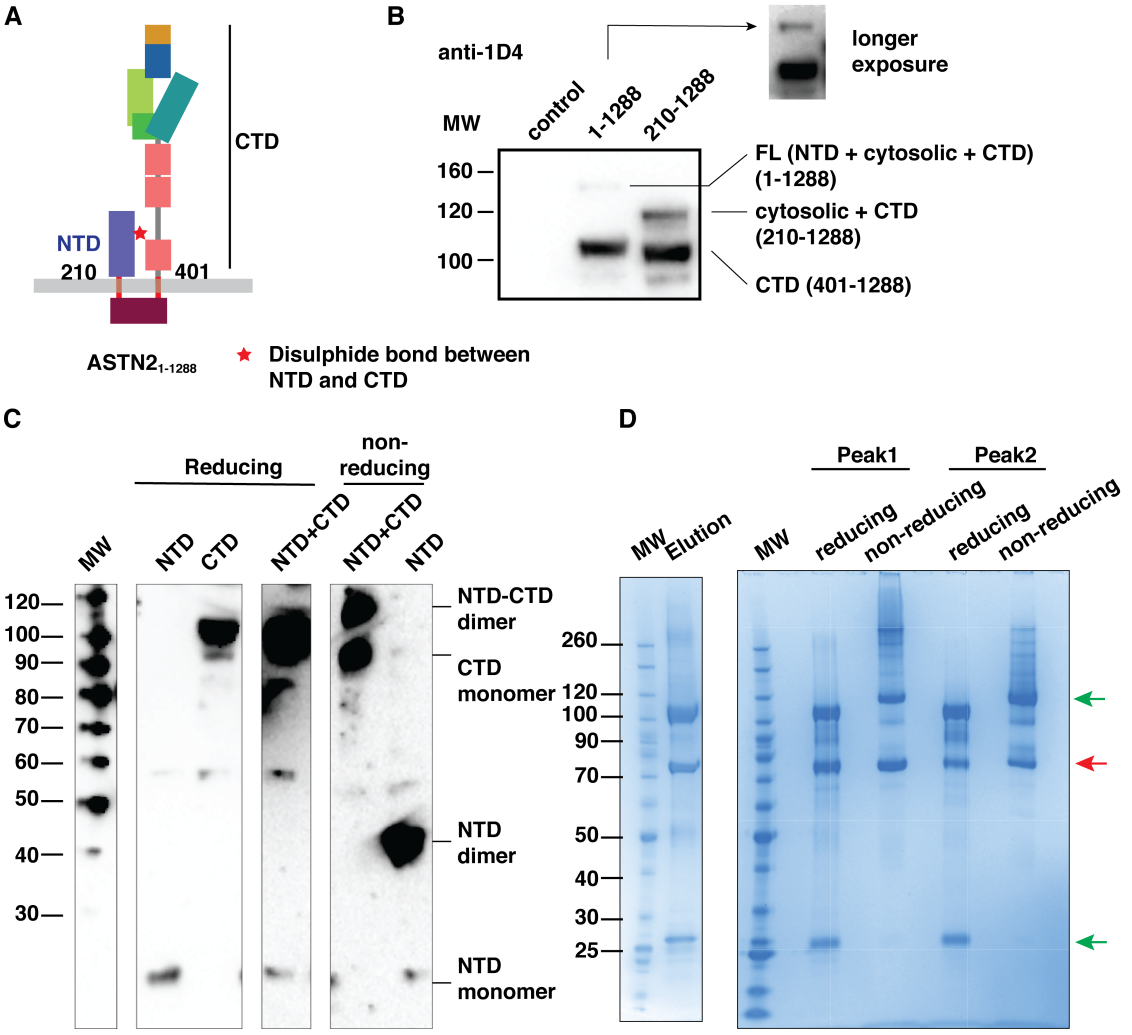

Figure 4.4: The endosomal domains are linked together through a disulphide bond. (A) Schematic of full-length ASTN2. The disulphide bond linking NTD and CTD is shown here, but the exact position of disulphide bond had not been specified yet. The residue numbers 210 and 401 specify the alternative translation initiation site and the end of the secondary transmembrane helix (the start of C-terminal domain). (B) Western blot of ASTN2 constructs showing the proteolytic processing between cytosolic domain and CTD. Full-length ASTN2 typically showed two bands on SDS-PAGE due to partial cleavage between the cytosolic domain and CTD. Longer exposure image is also shown for the middle lane, highlighting the size of the full-length protein. (C) ASTN2 NTD and CTD showing a heterodimeric form of endosomal domain linking NTD and CTD. Co-expression of ASTN2 NTD and CTD revealed a heterodimeric species formed between them in non-reducing conditions, which could be disrupted in reducing conditions. The isolated NTD also formed a dimer in non-reducing conditions. (D) Co-expression and purification of ASTN2 endosomal domain. The left panel shows the protein eluted from 1D4 conjugated agarose beads and the right shows the status of protein from size-exclusion chromatography under reducing and non-reducing conditions. The red arrow indicates the presence of unknown protein bound to ASTN2, and the green arrows NTD and CTD, respectively.

we tried purifying the protein directly form the cell lysate. Western blot analysis of cell lysate revealed that NTD and CTD bridged each other through a disulphide bond (**Figure 4.4C**), in agreement with data from the Nathan's lab. The cysteines that potentially form the disulphide bond could easily be identified since: i) there is only one cysteine (Cys^{63}) in the NTD; ii) for the CTD, excluding the cysteines identified in our crystal structure (see **Chapter 2**) and the conserved ones in EGF-like domains, there are only two potential sites (Cys^{406} and Cys^{565}). Intriguingly, the NTD itself was found to form a homodimer through a disulphide bond as well when the NTD construct alone was transfected into HEK 293T cells (**Figure 4.4C**), although we do not know if this is purely artefactual or might relate to a biological role for such a linkage.

Under reducing conditions, two bands corresponding to monomeric NTD and CTD were clearly resolved in SDS-PAGE, with one additional unknown protein bands with a molecular weight of about 75 kDa (**Figure 4.4C** and **Figure 4.4D**). In attempting to remove this contamination, the eluate was applied to size-exclusion chromatography (s200, 26/600) and two major peaks were resolved, both with a broad spectrum, potentially indicating a dynamic oligomerisation status. SDS-PAGE under reducing conditions showed that both peaks were composed of the same proteins, although the unknown band was present in both peaks, implying that this unidentified protein binds to ASTN2. Under non-reducing conditions, however, a larger molecular weight band estimated at 250-300 kDa was observed, indicating the presence of dimeric or oligomeric species (**Figure 4.4D**)

4.3.3 pH-dependent Self-interaction of ASTN2

ASTN2 is localised in endosomal vesicles and recycles in the endosomal system. Therefore it is tempting to ask whether the protein would undergo conformational change at different pHs or whether a pH-dependent self-interaction could occur. In our study, ASTN2 endodomain ($ASTN2_{601\text{ - }1288}$) is a monomeric protein in solution

at neutral pH, as observed in size-exclusion chromatography, small-angle x-ray scattering and analytical ultracentrifugation studies (**Figure 2.10**, see **Chapter 2**). In acidic pH, there are several different stages: at pH 5.5, ASTN2 started to associate with itself, with a dissociation constant (Kd) of 1mM (**Figure 4.5A**); at pH 5.0, ASTN2 formed a large aggregate although the aggregate could be recovered by raising the pH back to neutral (**Figure 4.5A**). So it is tempting to speculate that ASTN2 underwent a pH-dependent self-interaction. In contrast, ASTN1, its close homolog is monomeric in both pH 7.5 and pH 5.5 and even at lower pHs (data not shown).

To explore whether ASTN2 underwent a pH-dependent conformational change, a crystal was obtained at pH 5.0, and the structure was solved by molecular replacement at 5.5 Å resolution (PDB code: 5J68). However, there was essentially no difference in the structure of ASTN2 between neutral and acidic pHs (**Figure 4.5B**), although the low-pH crystal form revealed a new intermolecular interface mainly formed between annexin-like domain (**Figure 4.5D**). Sequence alignment of ASTN1 and ASTN2 identified two conserved histidines in ASTN2, both of which are absent in ASTN1 (**Figure 4.5C**), suggesting their potential role in mediating the pH-dependent self-interaction. However, mutating these two histidine residues back to ASTN1 did not completely eliminate the self-interaction. Interestingly, a construct (residues 700 - 1288) which lacks the EGF2-EGF3 domain was monomeric at both pH 7.5 and pH 4.5, which strongly suggests that the tandem EGF domains play a determining role in conferring pH-dependent self-interaction, rather than these conserved histidines (data not shown). The EGF2-EGF3 domain is disordered in our structure (as shown in chapter 2), so it is not yet possible to explain the detailed mechanism for this pH-sensitive switch. Further structural study of the tandem EGF domain would help uncover its role in dimerisation or oligomerisation of the full-length ASTN2.

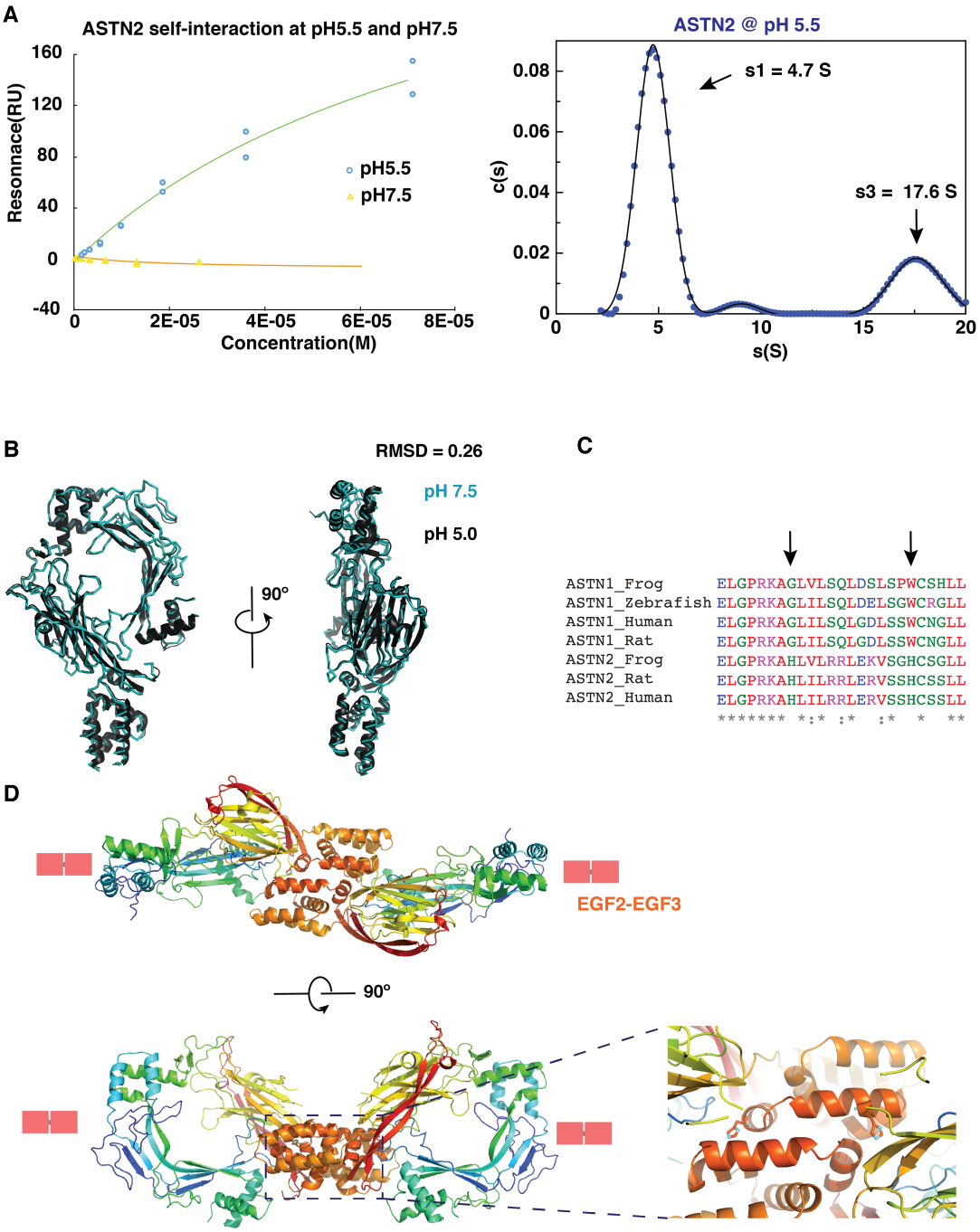

Figure 4.5: pH-dependent self-interaction of ASTN2 endodomain. (A) Surface plasmon resonance measuring self-interaction of $ASTN2_{601\text{-}1288}$ at pH 7.5 and pH 5.5. The plot shows the equilibrium data presented in the similar way to the right panel of Figure 2.14, excluding the original sensorgrams. (B) Analytical ultracentrifugation study of $ASTN2_{601\text{-}1288}$ at pH 5.5, showing the presence of high-molecular weight species. (C) Crystal structure of $ASTN2_{649\text{-}1288}$ at pH 5.0. No major conformational transition is observed in acid environment. ASTN2 structures are presented as cyan ribbon in pH 7.5 and black cartoon in pH 5.0. (D) Sequence alignment of ASTN1 and ASTN2 from several species showing the conservation of His^{1213} and His^{1225} in ASTN2. (E) New interface of ASTN2 structure at pH 5.0, highlight the two conserved histidines with the interface. The tandem EGF2-EGF3 domains are presented as pink blocks to indicate the relative protein orientation.

4.3.4 Controlling surface ASTN1 distribution by ASTN2

ASTN2 has been shown to control the surface exposure of ASTN1 on the cell membrane (Wilson et al., 2010). To dissect the contribution from ASTN2 individual domains to the surface distribution of ASTN1, we conducted flow cytometry experiments with a similar setup as Wilson *et al.* (Wilson et al., 2010) (**Figure 4.6B**). It was noticed that Wilson et al claimed that the translation initiation point starts from the first transmembrane helix (Met^{210} in our full-length construct), instead of the signal peptide identified by us and by Jeremy Nathans and colleagues (**Figure 4.6A**). This is probably due to their observation that under reducing conditions the full-length ASTN2 migrated on the SDS-PAGE gel with a molecular weight of about 100 kDa since the protein was proteolytically cleaved at Gln^{403}, instead of 142.5 kDa for the full-length without cleavage (**Figure 4.4B** and **Figure 4.6A**). In agreement with data from Wilson et al, expression of $ASTN2_{210\text{-}1288}$ lead to internalisation of ASTN1, revealed by a dramatic decrease in ASTN1 on the cell surface in a flow cytometry experiment (**Figure 4.6C**). However, expression of the full-length ASTN2 did not alter the level of surface-exposed ASTN1, which indicates a function of N-terminal domain in ASTN2. It is unknown how the N-terminal domain makes the difference and we will further characterise it. The domain deletion mutants, lacking individual domains (tandem EGF domain, MACPF domain, FN(III) domain and annexin-like domains) of ASTN2 are being prepared for future flow cytometry experiments.

4.4 Discussion and Future Plan

4.4.1 ASTN2 cytosolic domain is structurally ordered

Our biophysical characterisation showed that the ASTN2 cytosolic domain is a globular monomeric protein though with no presence of α-helix or β-sheet, in

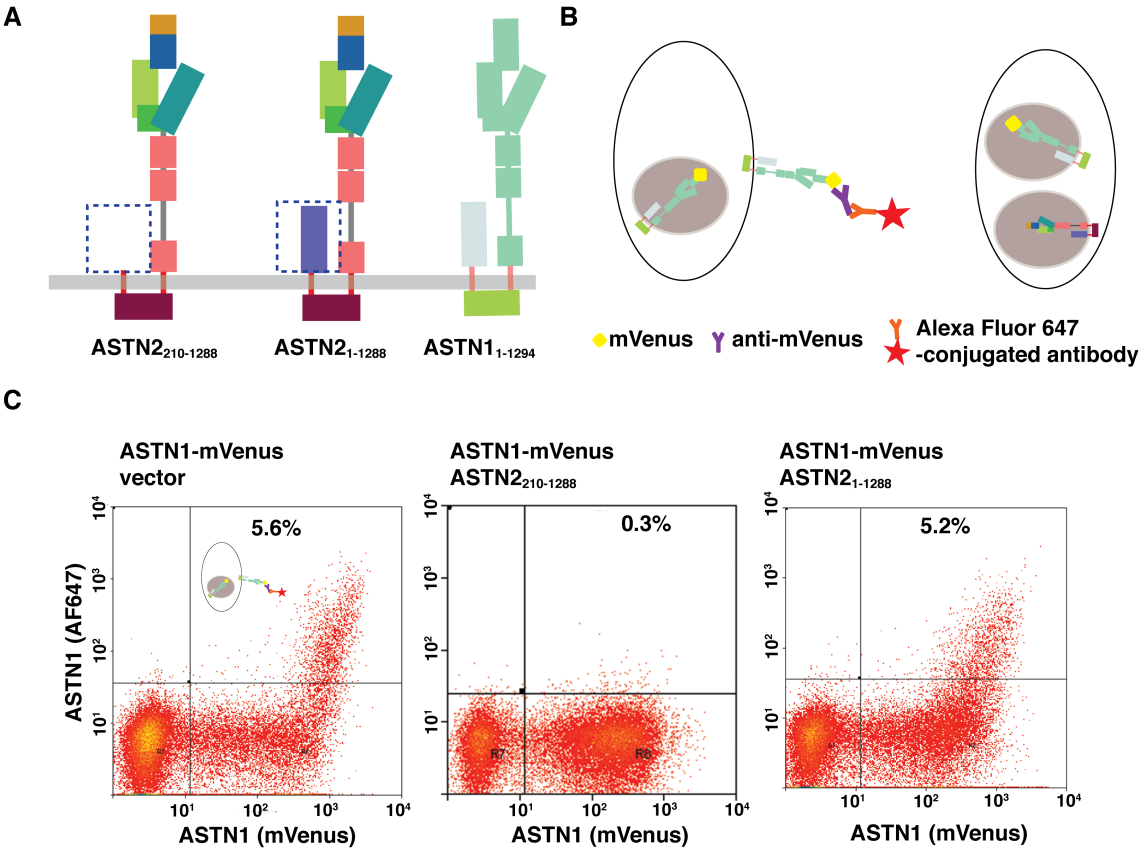

Figure 4.6: Flow cytometry study showing that different ASTN2 constructs affect ASTN1 surface exposure in a different manner. (A) Schematics of ASTN2 constructs. The constructs used from Wilson M et al, 2010 (left) and the one used in this study (middle) are compared here, highlighting the missing NTD in their case. (B) Schematics of flow cytometry labels used in this study. ASTN1-mVenus was transfected into HEK 293T cells. Cell surface-exposed ASTN1-mVenus was first recognised by anti-mVenus antibody, which then bound to Alexa Fluor 647 conjugated secondary antibody. The fluorescent signal of mVenus indicates successful transfection and AF647 signal shows the presence of surface ASTN1. Only surface-exposed ASTN1-mVenus could be detected by AF647 antibody (left panel). The intracellular ASTN1-mVenus is not accessible to the extracellular AF647 and therefore not be detected (right panel). (C) Surface exposure of ASTN1 in a flow cytometry experiment. The percentage indicated in the upper-right panel indicate the percentage of cells that shows surface-exposed ASTN1-mVenus. HEK 293T cells were co-transfected with the same amount of ASTN1-mVenus and ASTN2 constructs (vector, $ASTN2_{210\text{-}1288}$ and $ASTN2_{1\text{-}1288}$). Each data point represents the fluorescent signals of AF647 and mVenus from a single cell. The successful transfection of ASTN2 constructs was confirmed by western blot as shown in Figure 4.4B. The full-length ASTN2 does not recruit surface ASTN1 into the endosomal system.

agreement with the secondary structure prediction. The cytosolic domain could be expressed and purified with or without exon 5 and both version of the domain are stable in solution. Exon 5 deletion seems to make little impact on folding and stability of the cytosolic domain. Further CD experiments measuring melting temperature of the domain will better characterise it, and it is possible that absence of exon5 will result in a conformational change or variation in thermal stability. Furthermore, experiments including NMR would help to solve the structure of cytosolic domain and elucidate the effects of exon deletion on protein folding. It is possible that the cytosolic domains works as an adaptor mediating the interaction between the ASTN2-bearing endosome and cytoskeleton, and that exon5 deletion compromises its ability to bind to these proteins. Live-cell imaging comparing the intracellular distribution and trafficking of ASTN2 with and without exon5 will help to provide a clue for an answer to this hypothesis.

4.4.2 Interaction between ASTN1 and ASTN2

Understanding how ASTNs interact with each other is key to answering how ASTN2 controls the recycling of ASTN1 from the cell surface. The interaction of between ASTN1 and ASTN2 has been observed *in vivo* in a pull-down experiment (Wilson et al., 2010), and domain deletions in the C-terminal region did not impair the interaction. Indeed, in our SPR study, there is no interaction between the C-terminal regions of ASTNs (data not shown). It is plausible to hypothesise that the interaction was contributed by the other domains, such as the NTD and cytosolic domain, or the CTD region (residues 401 - 600) containing the EGF1 domain. However, one should note that there is no evidence confirming the direct interaction *in vivo* using purified proteins, and the apparent interaction observed from pull-down experiment (Wilson et al., 2010) could be due to indirect interactions bridged by other proteins.

We are currently working on further studies of the interaction between ASTNs by examining the interaction among equivalent domains. (i) Interaction between

cytosolic domains, if exists, could be easily confirmed by pull-down experiment using purified proteins. The cytosolic domain of ASTN1 has already been cloned and well-expressed in a small-scale expression trial in *E. coli* and further expression and purification in large-scale will follow. (ii) On the other hand, if the NTD dominates this interaction, then purified NTD domains could be used for pull-down experiments. (iii) We even suspect that the interaction between the ASTNs could be directed by the two sets of transmembrane helices, as shown in the interaction between transmembrane (TM) domains of the integrin α_{IIb} and β_3 subunits (Critchley, 2009; Luo et al., 2005).

4.4.3 Self-interaction and oligomeric states of ASTN2

The other key question we next wish to address is the potential oligomeric status of ASTNs. It is not yet clear whether the full-length ASTNs exist on membranes as a monomer, dimer or oligomer. An *in vivo* pull-down study using HEK 293T cells co-transfected with full-length ASTN2-3HA and ASTN2-RIM by Hao Chang and Jeremy Nathans (personal communication) revealed that ASTN2 can undergo self-interactions. Together with our characterisation of the ASTN2 endodomain which identified the presence of multiple peaks in size-exclusion chromatography and distinct bands with different molecular weights in SDS-PAGE, it is possible that ASTN2 exists in an oligomeric state, or in an equilibrium state between both monomeric and oligomeric species, in the cellular environment.

Further characterisation of ASTN2 using various techniques including size-exclusion chromatography coupled with multiple-angle light scattering (SEC-MALS), small-angle X-ray scattering and electron microscopy, are necessary to resolve this issue. We are for now focusing on structural characterisation of ASTN2 endodomain and carrying out the expression and purification trials of full-length ASTNs.

4.4.4 Linking ASTN2 with PCP pathway

The exact functions of ASTNs remain unclear. ASTN1 has been shown to be a neuronal receptor for neuron migration along glial fibres, but its ligand has not yet been identified. ASTN2, as the homolog of ASTN1, is localised to endosomal vesicles and controls the surface exposure of ASTN1 on cell membrane (Wilson et al., 2010). With little information about the protein interaction network of ASTNs, it is difficult to infer their molecular mechanism. Therefore a key step would be to identify the interacting partners to put ASTNs back context.

Establishment of PCP pathway requires the polarisation of several key proteins, including Fz3/Fz6, Vangl and cytoplasmic effectors such as Prickled1 and Dishelved1. It is important that these proteins are delivered to the right pole of cells through the endosomal recycling pathway. The involvement of microtubule filaments in the establishment and maintenance of asymmetrical distribution of PCP components implies that vesicles transporting PCP cargos travel along microtubule, providing a scaffold for cell polarisation. It will be very insightful to check the distribution of PCP proteins in the absence of ASTNs and whether they colocalise with ASTNs *in vivo*.

4.5 Methods and Materials

4.5.1 Cloning, expression and purification of ASTN2 cytosolic domain

Regions encoding different lengths of ASTN2 cytosolic domain were cloned into pOPINJ vector encoding an N-terminal GST tag followed by an HRV 3C protease cutting site, facilitating its purification using Glutathione agarose beads. Protein expression and purification procedures were the same as with the kindlin-3 Pleckstrin homology (PH) domain detailed in chapter 7.

4.5.2 Purification of ASTN2 full-endodomain from HEK 293T cell lysate

The yield of the complex was very low and could only be detected in western blot using anti-pentahistidine antibody, while purification using Ni-NTA column failed (data not shown). In order to purify the protein, the CTD was recloned into a pHLsec-1D4 vector, replacing the C-terminal polyhistidine tag with a Rho-1D4 tag instead, so that a Rho-1D4 antibody could be used for protein purification. NTD and CTD vectors with His_6 and Rho-1D4 tag, respectively were co-transfected into HEK 293T cell line. Cells from 12 roller bottles were harvested by mechanically disrupting cell attachment from the roller bottle surface 2 days after transfection,. The cells were washed with ice-cold PBS (phosphate-buffered saline) twice and lysed with 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM EDTA, with a standard protease inhibitor cocktail (Roch) and 1% Tween-20 with gentle sonication. The lysate was subsequently cleaned by centrifugation at 48,000 g for 2 hours, and the supernatant was then passed through 0.22 μm filter before incubated with 300 μl of agarose beads conjugated with monoclonal anti-1D4 antibody for 4 hours at 4 ℃. The agarose beads with protein bound were washed extensively and then eluted by incubating with 1D4 peptide at 1 mg/ml in 20 mM HEPES, pH 7.5, 500 mM NaCl overnight. All the procedures were conducted on ice or cold room (4 ℃). The eluate was pure except the unknown protein band at 75 kDa.

4.5.3 Surface Plasmon Resonance

SPR was conducted using the same procedure as described in **Chapter 2**. In order to immobilise ASTNs onto streptavidin-derivatised SPR CM5 chip, ASTNs with C-terminal Avitag-His_8 tag were prepared as follows. Briefly, ASTNs ($ASTN1_{601\text{-}1294}$ and $ASTN2_{601\text{-}1288}$) were cloned in frame with Avitag-His_8 (pHlsec-Avitag-His_8 vector) and transiently expressed and secreted in 293T cells (2 roller bottle for each construct). The secreted proteins were one-step purified using Nickel chelating agarose beads

and eluted with 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 500 mM imidazole. The proteins were *in vitro* biotinylated by BirA (expressed and purified from the lab) in the biotinylation buffer (10 mM Tirs-HCl pH 8.0, 7.5 mM $MgCl_2$, 5 mM NaCl, 5 mM ATP and 0.1 mM biotin), and the success of biotinylation was confirmed by their efficient binding to streptavidin magnetic beads. To perform SPR at low pH conditions, the proteins were serially diluted into pH 5.5 buffer (10 mM MES pH 5.5, 150 mM NaCl) and the running buffer in the system was also changed to the same buffer. The protein samples at low pH were injected twice, and the responses at different dilutions of injected analytes at the equilibrium states were plotted in the figure.

4.5.4 Circular dichroism

Circular dichroism experiments were conducted using a Jasco J-815 Spectropolarimeter in the Molecular Biophysics Suite at the Department of Biochemistry, University of Oxford. CD spectra were taken at a concentration of 0.1 mg/ml of the protein diluted in PBS (10 mM phosphate pH 7.5, 150 mM NaCl) using a cuvette with 1 mm path length at 293 K.

4.5.5 Flow cytometry

Flow cytometry experiments were conducted following the protocol described by Wilson and colleagues (Wilson et al., 2010). Briefly, full-length ASTN1 with its original signal peptide was cloned into pHLsec-mVenus-His_8 vector (restriction enzyme sites: EcoRI and KpnI) and full-length ASTN2 into a pHLsec-Rho1D4 vector (restriction enzyme sites: EcoRI and KpnI). HEK 293T cells were cultured to reach about 50% confluence in a 6-well plate before being co-transfected with 1 μg of ASTN1 and the same amount of the ASTN2 constructs, or empty backbone plasmids (pHLsec-Rho1D4). 48 hrs after transfection, the cells were harvested by trypsin

digestion or EDTA-treatment, and washed with ice-cold PBS supplemented with 2% FBS (fetal bovine serum, Invitrogen) twice. The cell suspension was then incubated with anti-GFP antibody (1:5000) (Sigma) for 60 min and washed with PBS-FBS solution twice before further incubation with secondary antibody (1:5000, Alexa Fluor®647 conjugate) for 30 min. The unbound secondary antibody was washed away and propidium iodine (1 μg/ml at final concentration) was added to cell suspensions for 10 min before flow cytometry experiments were conducted. The health status of the cell line is critical for obtaining reliable results and is indicated by the lack of propidium iodide signal for uncompromised cells. The flow cytometry experiments were conducted using a Beckman Coulter CyAn ADP in the Flow Cytometry Suite at the NDM Research Building with the assistance of Dr Andrew Worth.

CHAPTER 5

Structural Study of *Toxoplasma* Perforin-like Protein 1 (*Tg*PLP1)

5.1 Summary

Perforin-like proteins (PLPs) have been identified to be involved in pathogenesis of Apicomplexan parasites, particulaly *Toxoplasma* and *Plasmodium*, which are causative agents of toxoplasmosis and malaria, respectively. *Toxoplasma* encodes two PLPs (*Tg*PLP1 and *Tg*PLP2) and knockout out of *Tg*PLP1 resulted in significant defects in its life cycle, entrapping most of the parasites within the cells and limiting their ability to egress (Kafsack et al., 2009; Roiko and Carruthers, 2013; Roiko et al., 2014). *Plasmodium* contains five PLPs and each was shown to express at a specific stage of *Plasmodium* life cycle and to be indispensable at relevant stages for parasites to pass through different membranes (Deligianni et al., 2013; Garg et al., 2013; Kaiser, 2004; Risco-Castillo et al., 2015; Wirth et al., 2015). However, the exact roles PLPs play in Apicomplexan species pathogenesis remain unclear. Apicomplexan perforin-like proteins (ApiPLPs) share a significant structural homology with a pore-forming protein superfamily, the membrane attack complex-perforin/cholesterol-dependent cytolysin (MACPF/CDC) family (Gilbert, 2014; Kafsack and Carruthers, 2010), most of which are implicated in membrane remodelling and pore formation. Given the structural similarity and the general role of perforin-like proteins, it was speculated that ApiPLPs might function in the parasites through a similar mechanism. Our structural and biophysical study aimed to understand the structure and function of these proteins in detail and in this chapter we describe the crystal structure of *Toxoplasma* perforin-like protein 1 (*Tg*PLP1), which reveals key features of the MACPF domain distinct to other members of MACPF/CDC superfamily, and uncovers the novel fold of the C-terminal domain of ApiPLPs with 3-fold pseudo-symmetry unique to the Apicomplexan parasites. Most strikingly, we observed a helical form and ring form of MACPF domain packed in crystals, which indicates that *Tg*PLP1 might form a hexameric ring. Our structure of *Tg*PLP1 represents the first structure of ApiPLPs at an atomic resolution, providing clues on how these proteins might facilitate the progression of parasitic life cycles.

5.2 Introduction

5.2.1 Brief introduction to *Toxoplasma* and *Plasmodium*

Toxoplasma and *Plasmodium* are obligate intracellular, parasitic protozoans of major biomedical significance. *Toxoplasma gondii,* which is the causative agent of toxoplasmosis, can lead to serious problems in pregnant women and people with compromised immune systems, such as those who have had an organ transplant, or been infected with HIV, or those having chemotherapy (Grant et al., 1990; Tenter et al., 2000). For most adult humans, however, infection with *Toxoplasma* usually causes no symptoms. *Plasmodium spp.* is relatively well-known for being the causative agent of malaria, which is responsible for millions of deaths each year worldwide (World Health Organization, 2015). The species that infect humans include *Plasmodium falciparum* (the casue of malignant tertian malaria), *P. vivax* and *P. ovale* (the two causes of benign tertian malaria), *P. malariae* and *P knowlesi* (the causes of benign and severe quotidian malaria). *Plasmodium bergei*, *P. chabaudi* and *P. yoelli* infect rodents and have been used to study malarial infections and disease in the laboratory.

In terms of phylogeny, *Toxoplasma* and *Plasmodia* are Apicomplexa, defined by their unique complement of organelles including a specific apicoplast (a type of plastid) and an apical complex structure (responsible for initial gliding and attachment of parasites to host cell membrane) (**Figure 5.1**). Together with the ciliates and dinoflagellates, the Apicomplexa make up the superphylum Alveolata, which share a peripheral alveolar membrane system (Blader et al., 2015). In Apicomplexan parasites, the alveolar system is called the inner membrane complex (IMC). Other members of the Apicomplexa include *Cryptosporidium parvum*, *Eimeria spp.* and *Theileria spp.*, the latter two of which have considerable agricultural importance (Attar, 2015; Mans et al., 2015; Ryan, 2010).

The organelles unique to Apicomplexan parasites (the apical complex and apicoplast)

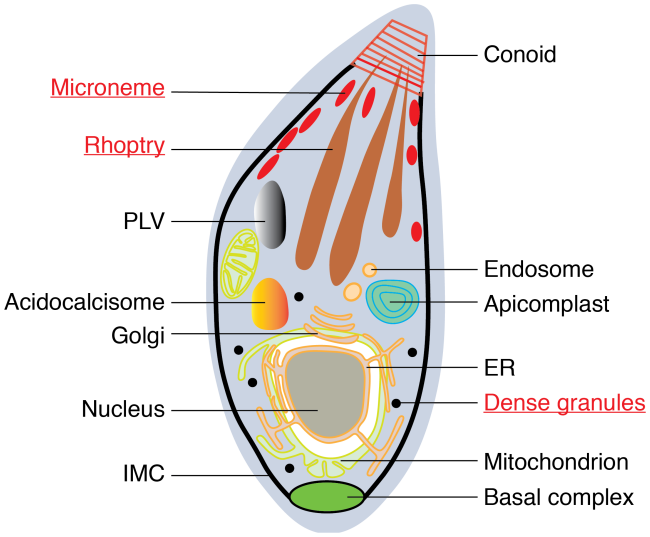

Figure 5.1: Diagram of *Toxoplasma* and *Plasmodium*. Basic tachyzoite organisation highlighting secretory pathway and apicomplexan-specific organelles. Abbreviations: ER, endoplasmic reticulum; IMC, inner membrane complex; PLV, plant-like vacuole. The organelles that contain proteins secreted to host cells are underlined (red). Figure adapted from Blader et al. 2015.

strongly indicate the parasites' adaptations to their host environments (Dubey et al., 1998). The apical complex is composed of a collection of unique organelles, including rhoptries, the micronemes, the apical polar ring, and the conoid (Hu et al., 2002, 2006). Rhoptries and micronemes are secretory organelles (similar to endosomes) that contain proteins required for motility, attachment to host cell membranes, cell invasion and the establishment of the parasitophorous vacuole within parasitised cells (Aikawa et al., 1988; Morrison, 2009; Scholtyseck and Mehlhorn, 1970). The enrichment of organelles responsible for secretion implies that proteins are secreted by Apicomplexa into targeted host cells in their life cycle. The conoid is a small cone-shaped structure composed of spiral filaments (Hu et al., 2002) that undergoes extension and retraction during host cell invasion, suggesting its mechanical role during membrane attachment or penetration, or both (Mondragon and Frixione, 1996). The apicoplast, a non-photosynthetic choloroplast-like organelle, similar to other dark-active plastids in function, acts as an anabolic hub and does not engage in photosynthesis. The importance of apicoplast-produced metabolites varies with parasites and their life stages and reflects adaptation to host environments (Sheiner

and Striepen, 2014).

5.2.2 Brief introduction to the life cycle of *Toxoplasma gondii* and *Plasmodium falciparum*

The heteroxenous life cycle of *Toxoplasma gondii* incorporates both asexual and sexual stages. The sexual stages of *Toxoplasma gondii* are exclusive to cats, while the asexual, or vegetative stages (tachyzoites and bradyzoites) have been encountered in virtually every mammal and bird tested (Black and Boothroyd, 2000; Blader et al., 2015; Dubey, 2009). The sexual stage occurs in the gut of their definitive host (feline) resulting in the formation of small spores (oocytes), each of which further develops into two sporocytes and eventually eight sporozoites. The typical route of horizontal infection by the asexual parasites startes from the oocysts in the faeces of the definitive host (feline). After ingestion into an intermediate host, the oocysts sporulate into sporozoites and attach to the small intestine and infect epithelial cells. Subsequently, the sporozoites develop into tachyzoites that replicate inside the parasitophorous vacuole (PV), which is established upon invasion. Following replication in the PV, the tachyzoites will egress from the host cells, often destroying host cells in this process. These tachyzoites are very active and spread throughout all tissues in the body unless controlled by the host immune system. In most circumstances, these tachyzoites are in fact cleared by the immune system, but some tachyzoites are able to develop into bradyzoites, the persistent stage of *Toxoplasma* infection. Consumption of animal tissues containing bradyzoites is the second route of horizontal transmission. The bradyzoites can differentiate back to tachyzoites, which is the major cause of *Toxoplasma*-associated diseases in immunocompromised humans. Finally, tachyzoites of *Toxoplasma* are able to cross the placenta and infect the developing fetus, causing congenital toxoplasmosis (Torgerson and Mastroiacovo, 2013).

In contrast to *Toxoplasma*, the *Plasmodium* life cycle does not necessitate two warm-

blooded hosts (cat and another) but mosquito and mammalian hosts (Amino et al., 2008b; Bray and Garnham, 1982; Flemming, 2015; Kappe et al., 2010; Spillman et al., 2015). Similar to the invasion process of *Toxoplasma*, mammalian infection starts with the injection of sporozoites by a mosquito into the host dermis, through which they cross a blood vessel wall into the blood stream. Sporozoites are then transported to the liver via the blood stream, and infect hepatocytes to initial liver-stage infection. These sporozoites usually pass through several liver cells without damaging the cells (a phenotype called traversal) (Amino et al., 2008b) before ultimately infecting another liver cell. The sporozoites invaginate to form, again, a parasitophorous vacuole and replicate within it to form mature merozoites, which are then released into the blood stream. These merozoites either migrate back to the liver and infect further hepatocytes or invade erythrocytic cells by invagination to form another parasitophorous vacuole and replicate within it. During the erythrocytic stage, some merozoites form sexual-stage gametocytes after invasion, and these can be taken up by a mosquito during a blood meal. Gametes mate to form a zygote and develop into ookinete that infect the mosquito midgut. The resulting oocysts produce sporozoites, which invade the salivary gland, ready to transmit to the next host through a mosquito's bite.

5.2.3 Lytic cycle of *Toxoplasma gondii* (invasion, replication and egress)

The lytic cycle of *Toxoplasma gondii* has been reviewed in great detail (Black and Boothroyd, 2000; Blader et al., 2015). Here we briefly recapitulate the key steps of lytic cycle and present a simplified model in order to understand the roles of perforin-like proteins (**Figure 5.2A**). Invasion of Apicomplexan parasites into host cells generally takes several steps, which are conserved among different parasites, although the detailed mechanisms are different. For simplicity, only the invasion process of toxoplasma is described here (**Figure 5.2B**).

Before *Toxoplasma* tachyzoites invade host cells, these parasites usually glide on the

surface of host cell membrane (Blader et al., 2015). Invasion by tachyzoites is an extraordinarily well-orchestrated event, which involves multiple secretory compartments, each with its own specialised products and role. The invasion procedure starts with the attachment of parasite to host cell membrane through glycosylphophatidylinositol (GPI)-anchored surface antigens (GAGs) from the parasite side. The evenly distributed GAGs recognise sulfated proteolygcans on the host cells (He et al., 2002). Following the initial recognition by GAGs, the parasite reorientates itself to use its apical end to penetrate into the target cell membrane. Micronemal proteins (MICs, mainly type I transmembrane proteins) are now secreted in a Ca^{2+}-dependent manner and presented onto the surface of the parasite. The most well-known member of the MICs is the apical membrane antigen 1 (AMA1), which forms a direct and tight contact with the transmembrane protein RON2 (Baum et al., 2008; Besteiro et al., 2011; Sharma and Chitnis, 2013). RON proteins are secreted into host cells from the rhoptry neck (RON), after MICs been presented on the parasite cell surface. The tight interactions of RON proteins and MIC proteins establish the moving junctions (MJ), providing a molecular basis for host plasma membrane invagination, which leads to the formation of parasitophorous vacuole.

The parasitophorous vacuole is a membrane-enclosed compartment that is refractory to fusion with the host cell membrane trafficking machinery (Mordue et al., 1999), providing a safe microenvironment for parasites to replicate. A series of a proton ATPase, H^+/K^+ ATPase and P-type ATPase was discovered to localise on the membrane of the parasitophorous vacuole, preventing the acidification of the PV during parasite replication (Bouchot et al., 2001; Sibley et al., 1985). Residing inside the parasitophorous vacuole, the parasite manipulates and redirects pathways of host cells, inducing global changes in cell physiology (Blader et al., 2001). *Plasmodium spp.* were shown to export their proteins into the cytoplasm of parasite-infected erythrocytes through a potential translocon, PTEX (Coffey et al., 2016; Elsworth et al., 2014; Spillman et al., 2015). It is not yet known whether *Toxoplasma* utilises a

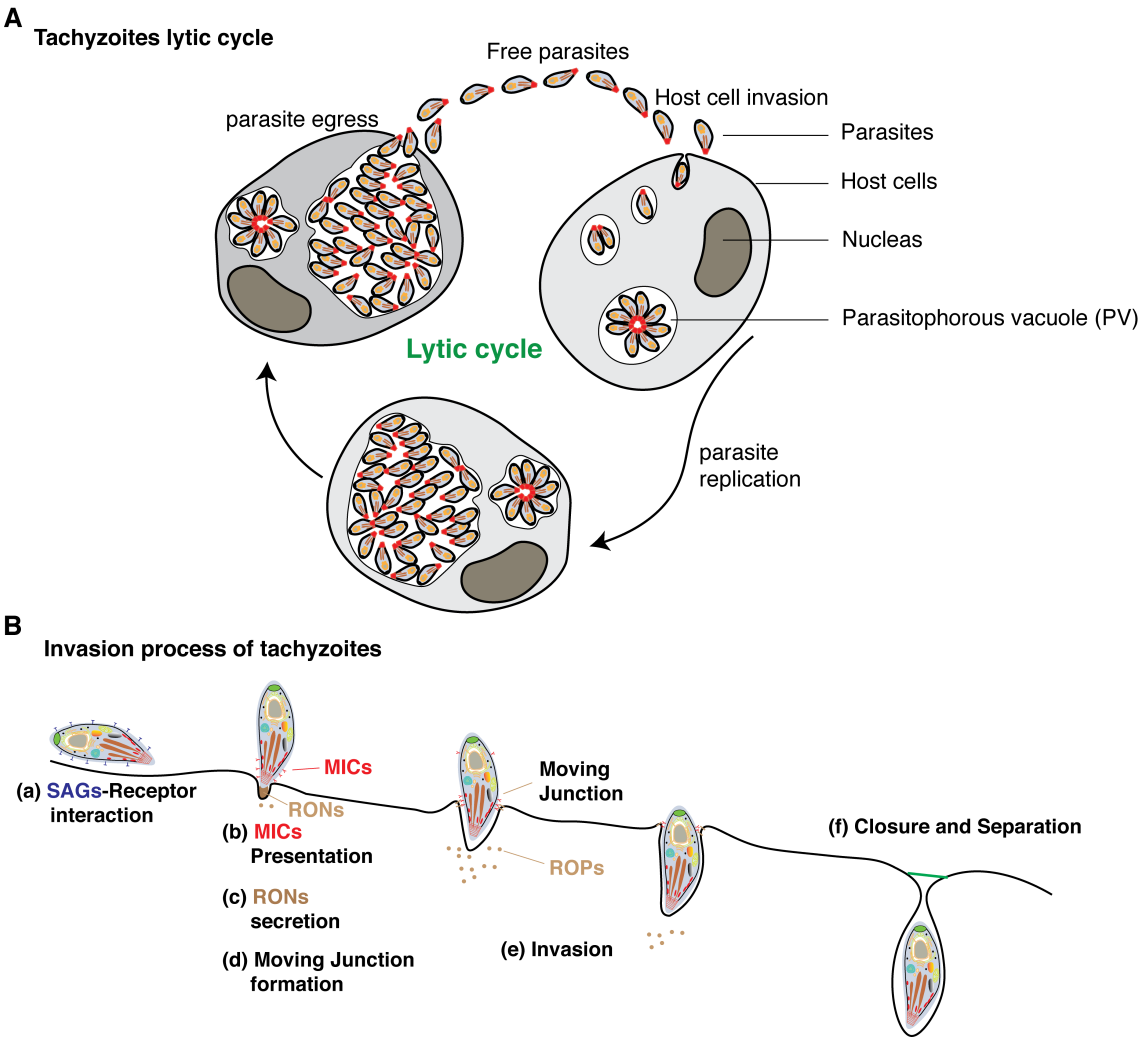

Figure 5.2: Lytic cycle of *Toxoplasma* and invasion process of *Toxoplasma*. (A) The lytic cycle of invasion ,replication and egress. (B) Working model of *Toxoplasma* invasion: (a) Initial attachment of parasites to host cell surface receptors by SAGs; (b) Reorientation of parasite and secretion and presentation of microneme proteins (the microneme-derived *Tg*AMA1 are presented on the parasite membrane surface and concentrated in to apical region; (c and d) Release of RON proteins (brown). RONs protein interact strongly with *Tg*AMA1 and form moving junction; (e) Rhoptry contents (abbreviated as ROPs) are injected into host cells. The invasion process is accompanied by the movement of the moving junction from apical to basal end, pulling host membrane together to form the parasitophorous vacuole (note that the moving junction forms a ring structure accommodating the entire parasite passing through); (f) Closure and separation of parasitophorous vacuole. Closure involves fission of the parasitophorous vacuole membrane and host plasma membrane. The diagram is a simplified overview and adapted from Carruthers et al. 2007.

similar machinery for protein export into host cell cytoplasm, although a common motif (PEXEL motif in *Plasmodium* protein and TEXEL motif in *Toxoplasma* protein) has been identified in the one subset of proteins that are targeted to dense granules and subsequently secreted to host cells (Coffey et al., 2016).

Egress of *Toxoplasma* from the parasitophorous vacuole (PV) is not as well studied as invasion. In order to egress, the parasites need to pass through two membrane barriers: the parasitophorous vacuole membrane and host cell membrane. It has been observed that the parasites can escape either by rupturing the whole host cells (thus completely destroying the host cells), or simply by squeezing through the host cell membrane, presenting a phenotype similar to invasion but from the opposite direction: the formation of moving junction inside of host cells. The exact signal transduction cascade that controls *Toxoplasma* egress remains to be established, but it has been confirmed that Ca^{2+} is able to trigger the egress of *Toxoplasma* (Carruthers and Sibley, 1999; Wetzel et al., 2004). Although the pH of the PV is usually neutral and there is no direct evidence demonstrating that the PV undergoes acidification before parasite egress, experimental acidification (pH 5.4) was able to enhance *Toxoplasma* motility and egress by inducing the secretion of micronemal proteins, as further discussed below (Endo et al., 1987; Roiko et al., 2014).

5.2.4 Implications of perforin-like proteins in *Toxoplasma* and *Plasmodium* life cycle

Although the genome of *Toxoplasma gondii* encodes two perforin-like proteins, *Tg*PLP1 and *Tg*PLP2, only *Tg*PLP1 was observed to express in blood stage tachyzoites. *Tg*PLP1 was first identified as being responsible for enabling rapid egress of *Toxoplasma* tachyzoites from infected cells (Kafsack et al., 2009; Roiko and Carruthers, 2009). It is localised to micronemes and secreted to host cells in a Ca^{2+}-dependent manner, similar to other micronemal proteins. Knockout of *Tg*PLP1 significantly hampered parasite egress though the effect on invasion was less clear.

Most tachyzoites remained imprisoned within the parasitophorous vacuole where they reproduced after invasion in the absence of *Tg*PLP1. Compared with WT tachyzoites, egress of *Tg*PLP1-knockout tachyzoiteswas significantly delayed upon induction by a calcium ionophore. Although some of these tachyzoites were still able to escape, primarily by motility-based mechanical rupture of the PV membrane, they turned out to be defective at subsequent invasion (Kafsack et al., 2009; Roiko et al., 2014), maybe because of exhaustion of the relevant secreted proteins stocks to drive it. It was observed from the same group that acidification of the parasitophorous vacuole triggered the release of microneme proteins (including *Tg*PLP1) and resulted in an enhancement of tachyzoites' motility. Echoing these findings *in vivo*, recombinant *Tg*PLP1 produced in bacteria was reported to show a 7x enhancement in pore-forming activity at pH 5.4 compared to pH 7.4 *in vitro* (Roiko et al., 2014). Taken together, these findings concluded that *Toxoplasma* egress was triggered by acidification: acidification activates the tachyzhoites and triggers the pH-dependent pore-forming activity of *Tg*PLP1, enabling the exit of tachyzoites from the PV. Recently, *Tg*PLP1 has been reported to be present in the cyst-forming *Toxoplasma* bradyzoite cells as well as the tachyzoites (Shan et al., 2015), although its exact role in this stage remains unknown.

In contrast to *Toxoplasma, Plasmodium spp.* encode five perforin-like proteins and all of them have been shown to be implicated in a specific stage of their life cycle (Kafsack and Carruthers, 2010; Tavares et al., 2014; Wade and Tweten, 2015). Briefly, *P*PLP1 (*Pf*PLP1, also known as SPECT2) is expressed at the sporozoite warm-blooded host invasion stage and is required for parasite traversal through the dermis, burrowing into the liver (Amino et al., 2008b; Ishino et al., 2005). It has recently been shown that the *Plasmodium yoelii* cell traversal phenotype is dependent on transient vacuole (not parasitophorous vacuole, PV) formation followed by pH-dependent vacuole release caused by *Py*PLP1 (Risco-Castillo et al., 2015). This seems to be a very similar process to the PV release enabled by *Tg*PLP1 (Kafsack et al., 2009; Roiko et

al., 2014). In blood-stage malaria, *P*PLP1 is required for merozoite exit from PVs and erythrocytes membranes in a calcium-dependent manner, due to oligomerisation and formation of transmembrane pores by *P*PLP1 (Garg et al., 2013). *P*PLP2 (*Pb*PLP2 and *Pf*PLP2) was shown to express in blood stage schizonts and to be crucial for egress of mature gametocytes during blood stage malaria, but dispensable for merozoites in the asexual blood replicative cycle (Deligianni et al., 2013; Garg et al., 2013; Wirth et al., 2014). *P*PLPs 3-5 are expressed in ookinetes and involved in the mosquito stage of the *Plasmodium*, specifically in midgut invasion. Deletion of any of these three *P*PLPs resulted in a complete loss of the capacity to pass through the epithelium of the midgut, indicating that the three *P*PLPs have synergistic genetic interaction with each other and are involved in the same cellular process or signalling pathway (Chatzidaki-Livanis et al., 2014; Deligianni et al., 2013; Ecker et al., 2007; Kadota et al., 2004; Risco-Castillo et al., 2015). The degrees of complexity in life cycles of *Toxoplasma* and *Plasmodium* seems to correlate with the diversity of PLPs in each parasite: *Toxoplasma* completes its life cycle within two animals (feline and the other warm-blooded animal) while *Plasmodium spp.* require both mosquito and a warm-blooded vertebrate animal host. *P*PLP3-5 are mostly likely the result of the adaptation of *P*PLP1 and/or *P*PLP2 for invasion of epithelium cells of the mosquito midgut, which are completely different type of cells compared with those from warm-blooded animals.

5.2.5 Domain architecture of perforin-like proteins from *Toxoplasma* and *Plasmodium*

The architecture of perforin-like proteins in *Toxoplasma* and *Plasmodium* features with a central MACPF domain followed by a conserved β-sheet rich C-terminal domain (CTD). The N-terminal domains of these PLPs vary considerably both in length and amino acid sequence and are mainly predicted to be disordered (Kafsack and Carruthers, 2010) (**Figure 5.3A**). Repeats of polar residues (such as poly-

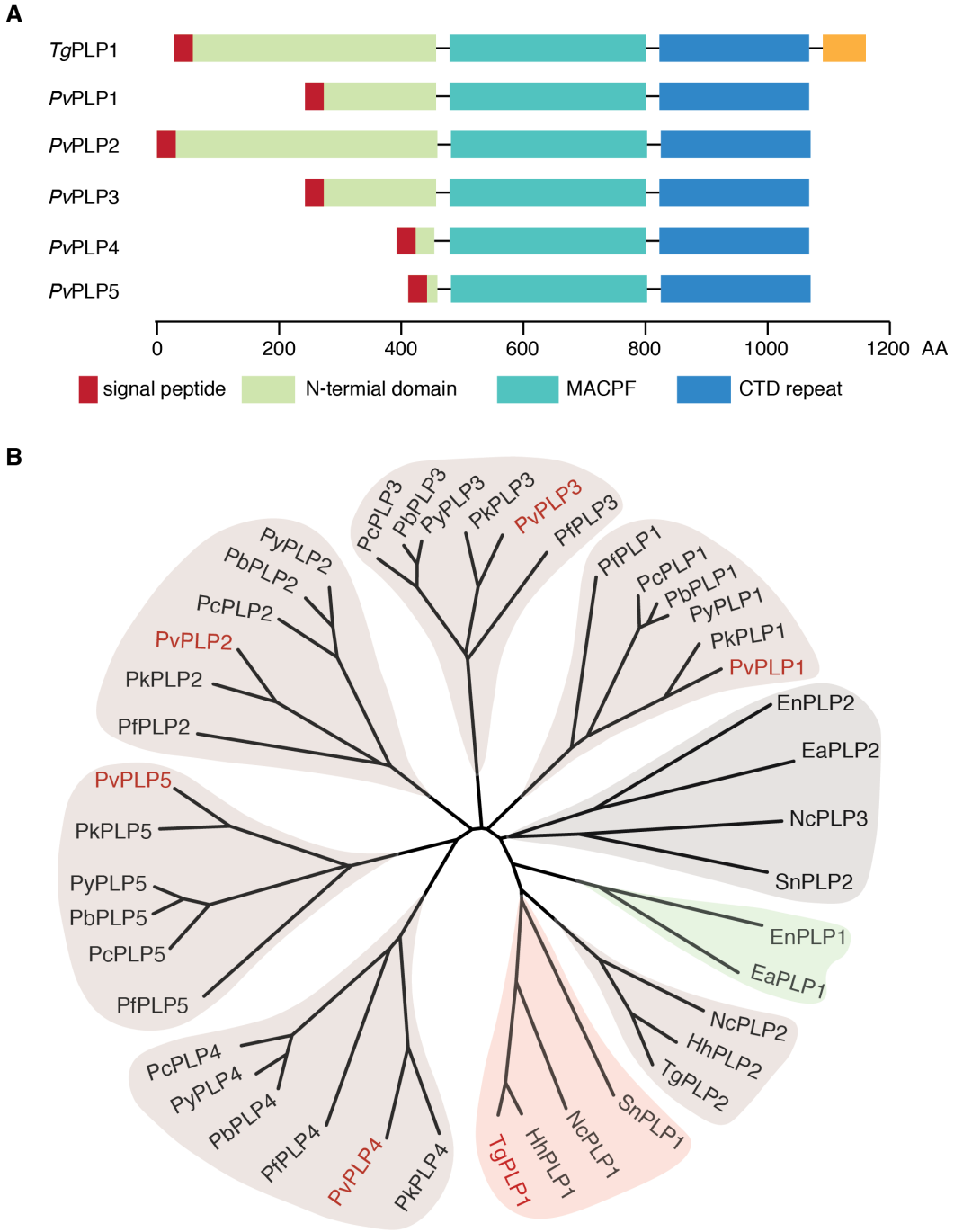

Figure 5.3: Domain architecture of ApiPLPs. (A)Schematic of domain architecture of PLPs from *Toxoplasma gondii* and *Plasmodium vivax*. PLPs share conserved MACPFs domain and C-terminal domains, with large variance in the N-terminal domains. *Tg*PLP1 has extra about 80 amino acids in the very C-terminal end, which is depicted in yellow. (B) Phylogeny of ApiPLPs. For simplicity, only PLPs from *Plasmodium spp.* and representative *Toxoplasma spp.* were used in construction of this phylogenetic tree. It shows that ApiPLPs diverged early in evolution. Abbreviations: Pf: *Plasmodium falciparum*, Pk: *Plasmodium knowlesi*, Pv: *Plasmodium vivax, Pb: Plasmodium berghei, Pc: Plasmodium chabaudi,* Py: *Plasmodium yoelli.*

arginine and poly-glutamine) are readily identified within the N-terminal domain. The C-terminal domain is predicted to contain three repeats, each of which is composed of several β-sheets. The paring of the MACPF domain and the same kind of the C-terminal domain is found across the phylum, indicating that this union occurred early in evolution (Kafsack and Carruthers, 2010).

Compared with other ApiPLPs, *Tg*PLP1 has been studied in most detail and depth. It has three domains: an amino-terminal domain of unknown structure lacking detectable homology to other proteins; a central canonical MACPF domain; and a carboxy-terminal domain dominated by a repeating β-sheeted structure and essential for initiation of *Tg*PLP1 activity via membrane binding (Kafsack et al., 2009; Roiko and Carruthers, 2013; Roiko et al., 2014). Data suggests that both the NTD and CTD contributed to membrane binding. The CTD's role was dominant and pH-sensitive, whereas loss of the NTD reduced the activity of *Tg*PLP1 but not tachyzoites pathogenesis, and its binding to membranes was pH-insensitive (Roiko and Carruthers, 2013; Roiko et al., 2014). The MACPF domain was shown to be necessary for *Tg*PLP1 oligomerisation (Roiko et al., 2014) , and there is evidence that the *Toxoplasma* subtilisin-like protease 1 (*Tg*SUB1) is responsible for a proteolytic processing of *Tg*PLP1 between the NTD and MACPF domains (Lagal et al., 2010; Roiko et al., 2014). Both *Tg*SUB1 and *Tg*PLP1 were found in micronemes, secretory vesicles used by the parasite to enable movement into and egress from host cells.

5.3 Results

This section describes the structural and functional analysis of *Tg*PLP1. The MACPF and C-terminal domain of *Tg*PLP1 were expressed, purified and crystallised separately. Briefly, three crystal forms were obtained for the MACPF domain, $P6_5$, $C222_1$, and C121. Crystals with spacegroups $P6_5$ (2.03 Å) and C2 (5.0 Å) were generated by *in situ* trypsin limited proteolysis and the other one was from native protein without typsin treatment (3.11 Å). Crystals with spacegroup $P6_5$ were used for Pt-derivative phasing. For the C-terminal domain, several crystal forms were obtained, all within an atomic resolution, and the one which diffracted to highest resolution (1.1 Å) was used for Pt-derivative phasing. The details for crystal preparation and structure determination are described in **Chapter 6**.

5.3.1 Overall structure of MACPF domain

The MACPF domain is generally presented as a thin plate-like or planar structure featured with four central β strands making it bent by about 90 degrees, with two sets of helices on either sides of the β sheets. The overall architecture of *Tg*PLP1 MACPF domain revealed typical features shared with other MACPF family proteins (**Figure 5.4**), with RMSDs of 1.61 Å, 1.75 Å and 1.46 Å to murine perforin-1, and complement components C8α and C8β, respectively. However, the overall shape of *Tg*PLP1 MACPF domain is not as flat as perforin-1 (**Figure 5.4B** and **Figure 5.5**), mainly due to a shift of a particular helix (α9) away from the plane of the molecule. Despite the overall similarity with previously-determined structures, phasing by molecular replacement using these MACPF domains was not successful, very likely due to the differences in regions specific to each MACPF domain protein, such as the curvature of the central twisted β-sheet and the postion of helices mentioned above.

Compared with other MACPF domains, the first significant difference shown by

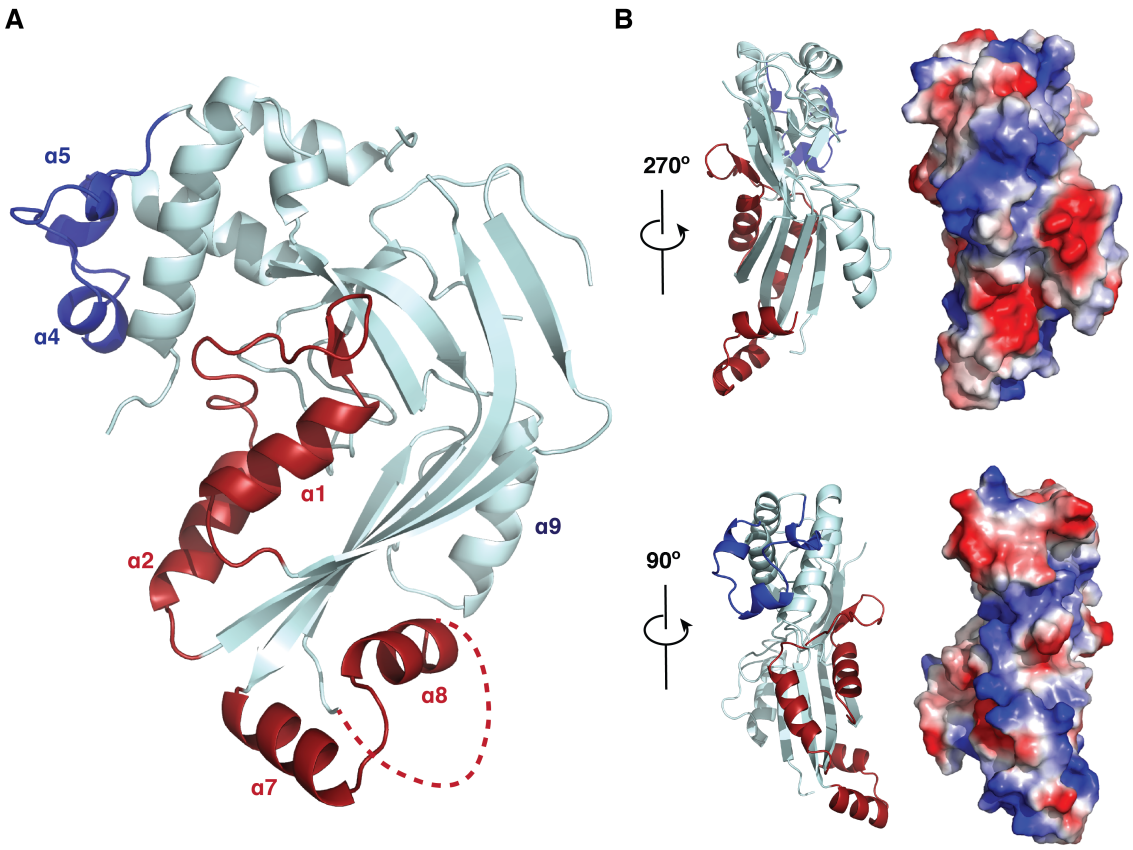

Figure 5.4: Overall structure of MACPF domain of *Tg*PLP1. (A) Overall structure of MACPF domain: two sets of α-helices are highlighted in red (α1-α2, and α7-α8). The extension of helices (α4-α5) are coloured in blue. The disordered loop (20 amino acid residues) within helices set 2 is arbitrarily draw in red dashed-line to highlight its position. (B) Two side-views of MACPF domain with surface electrostatics coloured.

the *Tg*PLP1 MACPF domain is a loop extension between two helices (α4 and α5, about 30 amino acids) (**Figure 5.4**), within which a disulphide bond is formed. This loop insertion is structurally ordered and folded together with the rest of the MACPF domain, which is well-conserved across all the MACPF domain proteins from Apicomplexan parasites (Kafsack and Carruthers, 2010). Initial protein expression trials deleting this loop completely blocked protein secretion from mammalian HEK 293T cells, presumably due to misfolding without this loop being present. The other key difference found in *Tg*PLP1 lies in the two helical regions (α1-α2 and α7-α8, which in pore-forming MACPF/CDC proteins characterised to date are responsible for the membrane insertion and the formation the beta-barrel pore by helix-to-sheet transition. Structurally, the inner set of helices (α1 and α2) is more conserved than the outer: the overall structure of the inner helices is rather similar to those of perforin-1 and complement components (**Figure 5.5A**); in contrast, the outer set (α7 and α8) is much more flexible, which is expected since it is exposed to solvent, rather than forming extensive contact with the rest of the domain as the other set does. Indeed, in two crystal forms which resulted from *in situ* limited proteolysis by trypsin, the outer set of helices is absent. Most strikingly, the absence of the outer set of helices enabled the rest of the MACPF domain to pack into either a helical form or a ring form, with 6 molecules each turn/ring, as will be discussed in detail below.

The length of these two (inner and outer) sets of helices is of particular interest in studying the mechanism of membrane insertion by MACPF/CDC proteins because of their role in pore formation. In the MACPF domain of *Tg*PLP1, these two sets of helical regions consists of 44 and 48 amino acids residues, respectively, compared with 52 - 55 and 56 - 60 in perforin and complement components. The difference in lengths of helices sets might reflect their properties in pore formation, such as the diameter and height of pores, and the stoichiometry of pore-forming complexes.

The MACPF/CDC superfamily features a large degree of sequence variance among

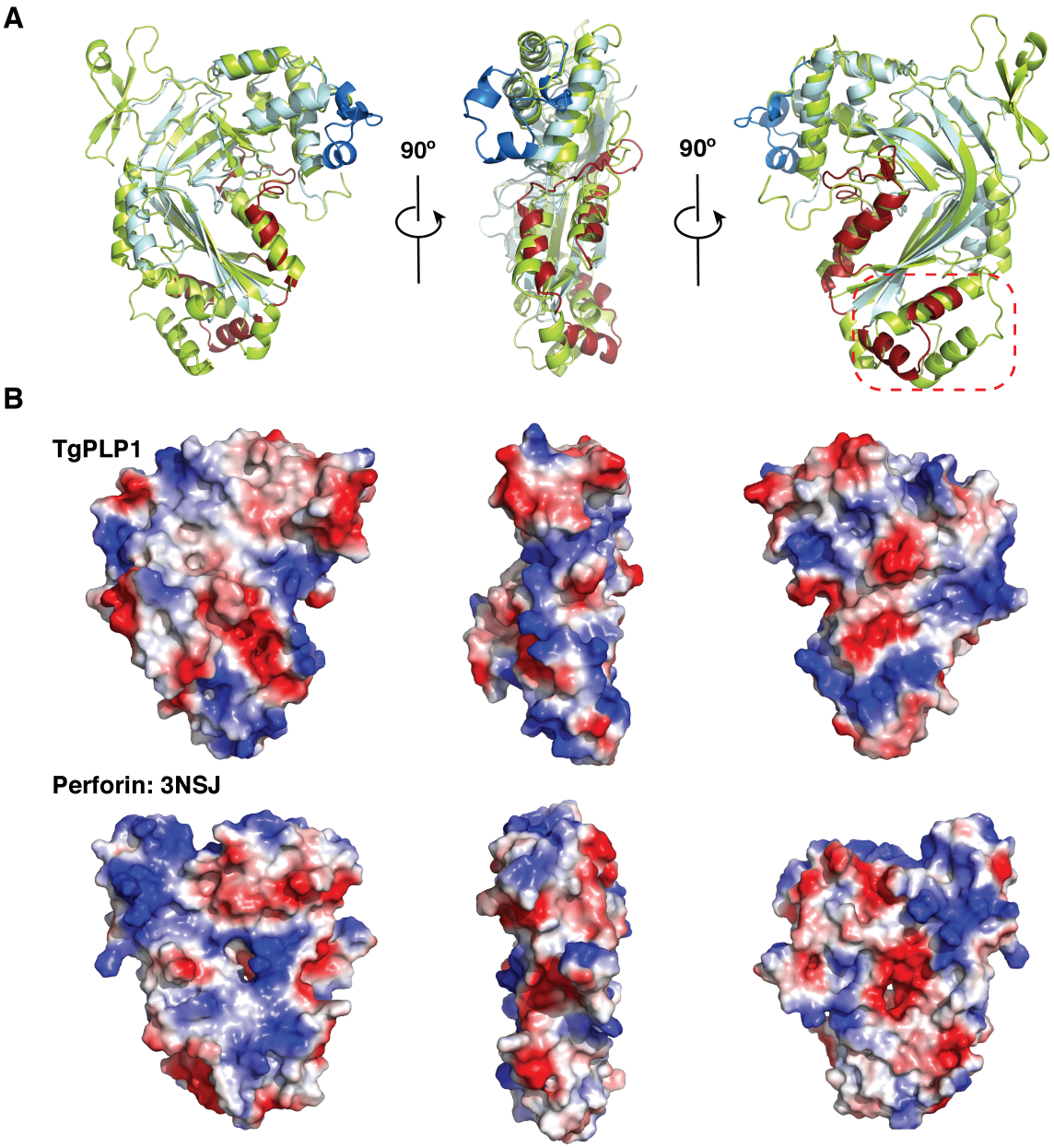

Figure 5.5: Comparison of MACPF domains between *Tg*PLP1 and perforin-1. (A) Superimposition of MACPF domains of *Tg*PLP1 (light blue) and perforin-1 (light green) presented in three different orientations. Overall the central twisted β-sheets of both domains are in the same conformation.The large differences found are the helices set 2 (highlighted within a red frame) and extension with two addtional helices (α4-α5). (B) Surface charge distribution and overall morphology of MACPF domains of *Tg*PLP1 and perforin-1. No much similarity about surface charge is observed. While the overall triangular morphology are shared between two domains, *Tg*PLP1 MACPF domain is not as flat as MACPF domain of perforin-1 (middle: side view). The domain orientations in (B) are the same to these of (A).

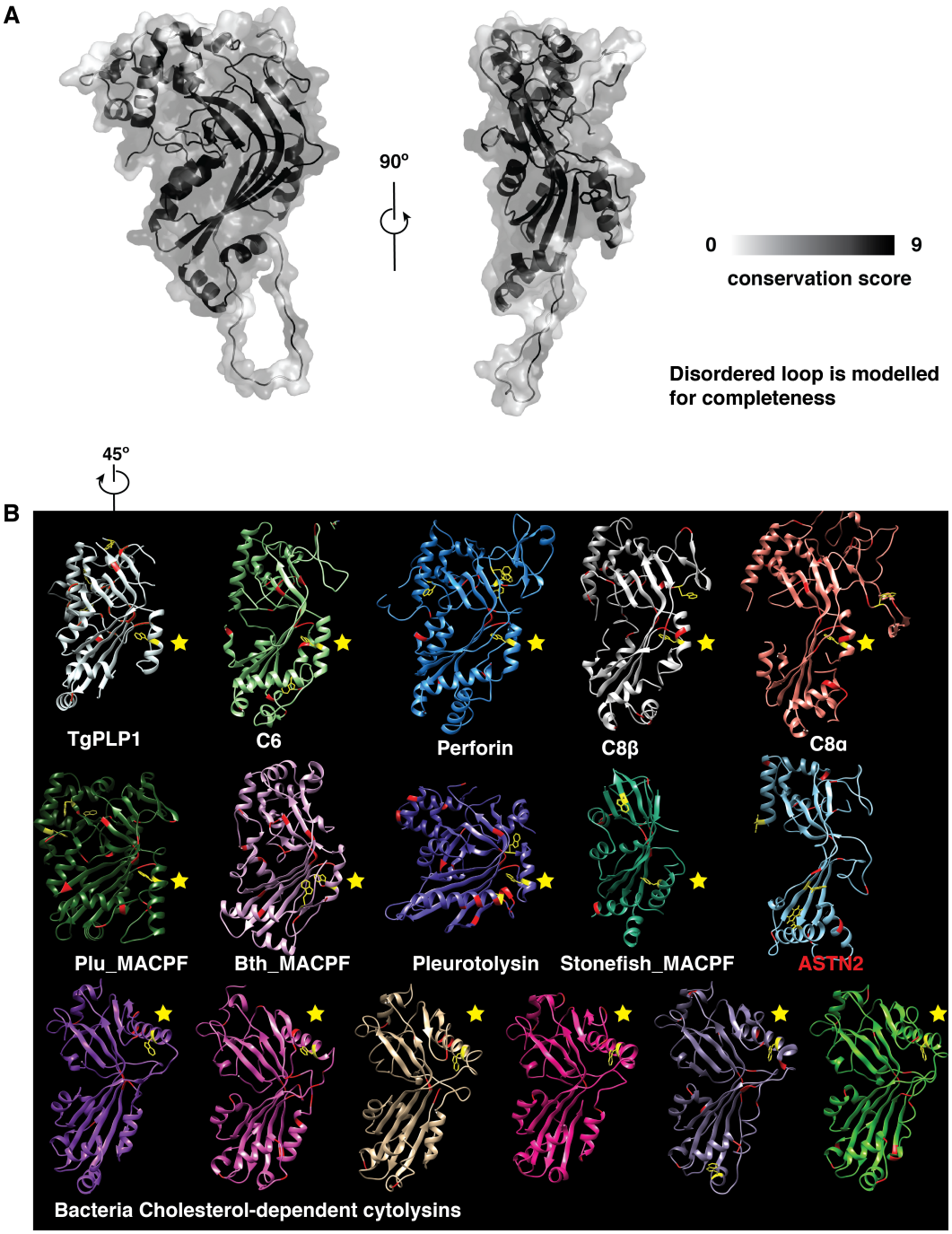

Figure 5.6: Conservation of MACPF domain form ApiPLPs and identification of a conserved GGX_nW motif. (A) Conservation of MACPF domain among the ApiPLPs. The conservation score was calculated using sequences of all ApiPLPs available in ToxoDB (www.toxodb.org). The disordered loop in helices set 2 was modelled in MODELLER. The conservation score was plotted from 0 to 9, with colour gradient from white to black. (B) Conserved GGX_nW motif among MACPF/CDC superfamily proteins. All glycine residues (red) and tryptophan residues (yellow) are shown within each protein. The yellow star indicates the position of the conserved tryptophan.

family members while sharing a single type of polypeptide fold. A signature motif F(Y/H/W)-G-S(T)-H-F(Y)-X_6-GG was identified the greatest sequence similarity within the MACPF/CDC superfamily (Hadders et al., 2007; Slade et al., 2008). The MACPF domains of ApiPLPs themselves showed the same pattern of conservation, with less than 30% sequence identity (**Figure 5.6A** and **Figure 5.7**). Mapping the conserved residues across the ApiPLPs revealed that the majority of the conserved residues are localised within the core of the MACPF domain, rather on the surface (**Figure 5.6A**). Apart from the conserved MACPF domain motif, of particular note is a GGX_nW motif with a tryptophan in a helix (α9 in *Tg*PLP1 MACPF domain) following a glycine in a loop (**Figure 5.6B** and **Figure 5.7**). It appeared that this feature is highly conserved among MACPF proteins (except ASTN2). In CDCs, a conserved tryptophan was also identified, although it was positioned in a different region (**Figure 5.6B**). The function of this conserved region remains to be identified; electron microscopy reconstruction of the complete membrane attachment complex pore and polyC9 structure showed that this region is structurally ordered and positioned inside of the pore (Dudkina et al., 2016; Serna et al., 2016), with a small conformational adaptation after pore formation.

5.3.2 Helical and ring structures formed by MACPF domain

As briefly mentioned earlier, by using *in situ* limited proteolysis, we obtained two crystal forms: $P6_5$ with one molecule per asymmetric unit, and C121 with 6 molecules per asymmetric unit and forming a ring. Both crystal forms revealed the same protein fragments and the same conformation. Reconstruction of the MACPF domain in the $P6_5$ symmetry showed a helical structure (**Figure 5.8A**), with a diameter of 130 Å and a pitch of 52 Å (6 molecules each turn). The ring form of MACPF domain were composed of 6 copies as well, with the overall structure positioned in the same orientation as that in the helical form (**Figure 5.8B**). The diameter of the ring (139 Å) was larger than that of the helical form (130 Å), due presumably to the

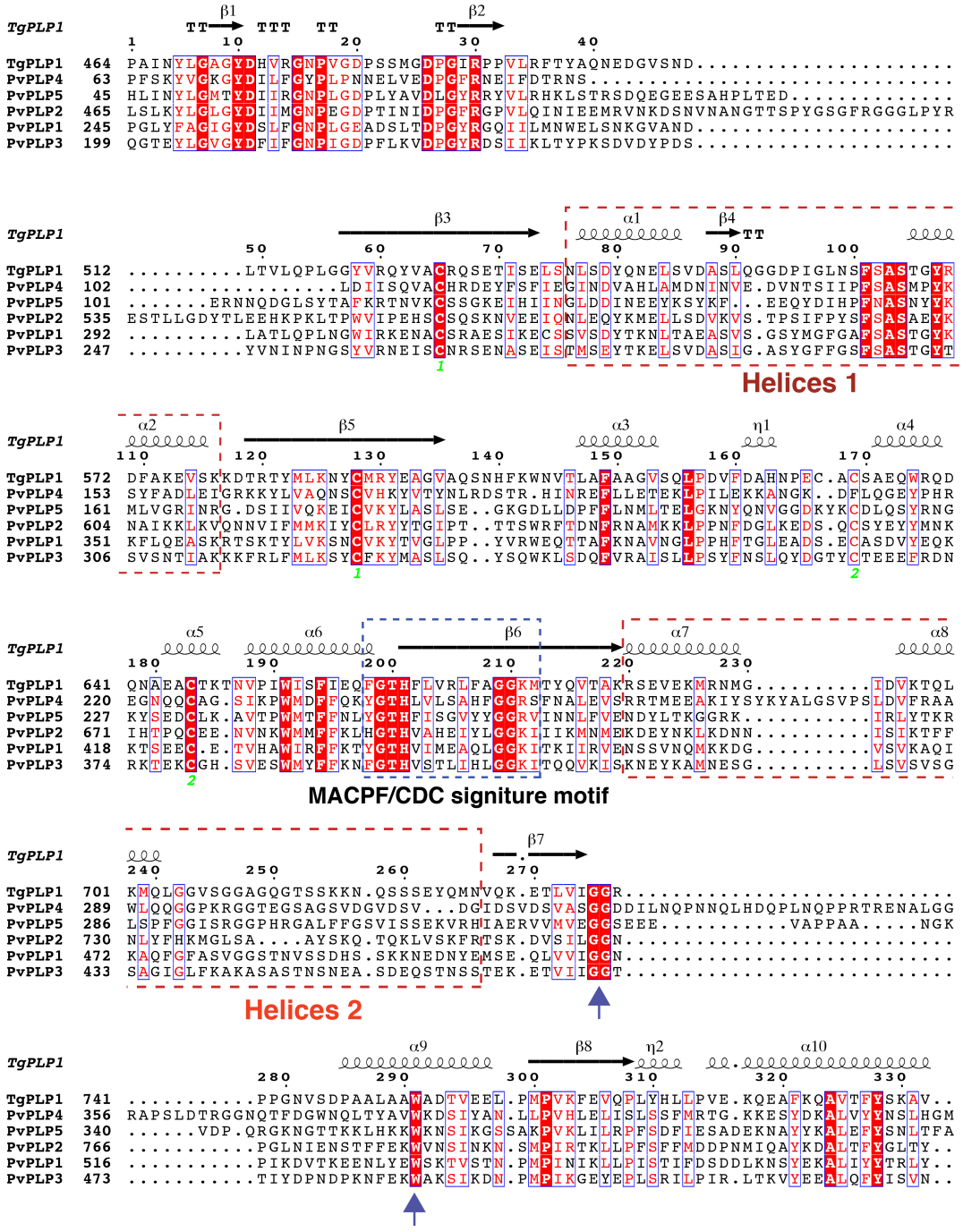

Figure 5.7: Sequence alignment of MACPF domains from *Plasmodium* and *Toxoplasma*. The MACPF/CDC signature motif F(Y/H/W)-G-S(T)-H-F(Y)-X_6-GG is highlighted with a blue frame and the two sets of helices for membrane insertion are labelled with red frames. The conserved GGXnW motif around $\alpha 9$ is highlighted with blue arrows.

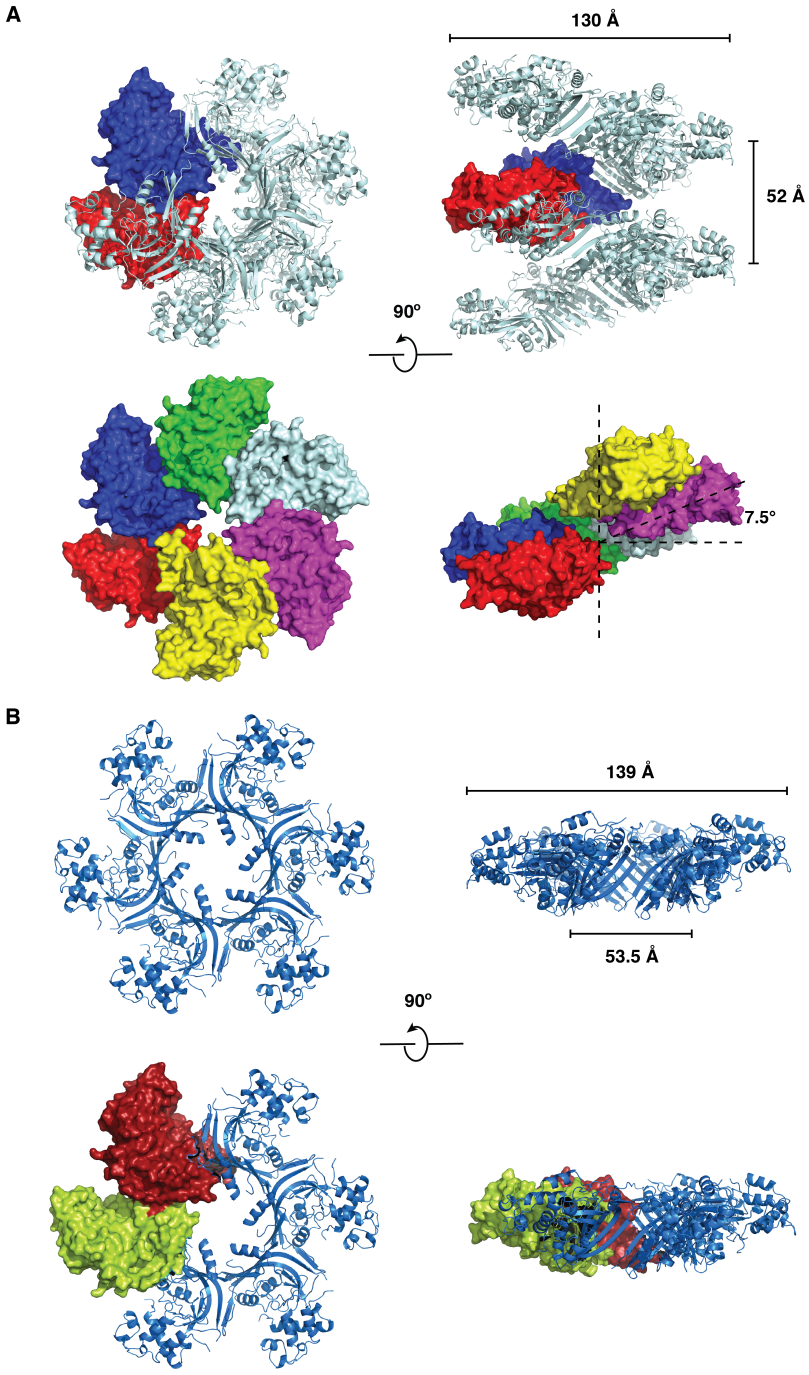

Figure 5.8: Helical and ring arrangement of the *Tg*PLP1 MACPF domain. (A) Helical arrangement of MACPF domain in crystal structure (spacegroup $P6_5$). The helical form of the MACPF domain has a diameter of 130 Å and pitch of 52 Å (6 molecules each turn). (B) Ring form of the MACPF domain in another crystal structure (spacegroup C121). The diameter of the inner ring formed by the central β-sheets is 53.5 Å, while the diameter of the outer ring is 139 Å.

slightly tighter turn imposed by the helical screw. The pore formed by the central β-sheets had a diameter of 53.5 Å. It was of note that helices set 2 was missing (due to the *in situ* proteolysis) in both structures, leaving enough space to enable the MACPF domain packing into such ways. Given that the pore formation mechanism of MACPF/CDC proteins involves the "loss" of this helical region as β-sheet this suggests that the hexameric ring assembly formed by the *Tg*PLP1 MACPF domain represents the pore-forming state of the molecule.

We wondered whether it was possible to reconstruct the helical/ring form of MACPF domain in solution. Like many other perforin-like proteins, *Tg*PLP1 existed in solution in a monomeric form before inserting into membrane. Forming a ring or helical form requires conformational adaptations. It was observed that some pore-forming proteins were able to transit from monomeric form to oligomeric form in solution in presence of detergent, for example, lysenin (Benke et al., 2015; De Colibus et al., 2012; Mueller et al., 2009). Pneumolysin was even able to self-associate into a helical form without detergent, mimicking its conformation in its pore form (Gilbert et al., 1998, 1999). To explore whether *Tg*PLP1 could undergo conformational change and oligomerisation in the presence of detergent, we incubated *Tg*PLP1 MACPF domain with n-Dodecyl β-D-maltoside (DDM) overnight at 37°C. Heavy precipitation was observed after incubation and SDS-PAGE analysis of the precipitants revealed multiple species with different apparent molecular weight matching to different oligomeric states (dimer, trimer, etc.) (**Figure 5.9**). The heavy precipitation was soluble in presence of 0.1% SDS and behaved as oligomers with different apparent molecular weights in size-exclusion chromatography (data not shown).

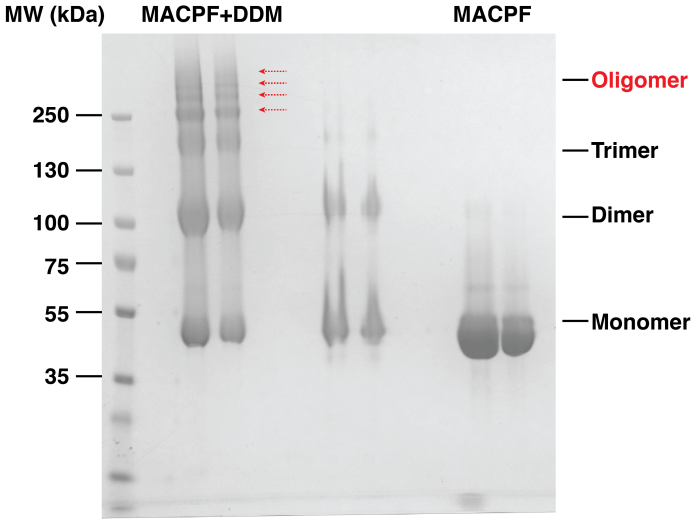

Figure 5.9: Oligomerisation of *Tg*PLP1 MACPF domain in DDM. The oligomer was induced by incubation of MACPF domain with DDM at 37°C (left lane) and 4°C (middle lane). The red arrows indicate the oligomer species with high molecular weights.

5.3.3 Overall structure of the *Tg*PLP1 C-terminal domain

The structure of the C-terminal of *Tg*PLP1 showed that this domain was mainly composed by β-sheets, forming an overall fold with three repeats (RMSD between repeats: 1.28 - 1.56 Å). The three repeats, each with about 80 amino acid residues, presented themselves with three-fold pseudo-symmetry (**Figure 5.10A**). Within each repeat, there were six β-strands, with four of them in one sheet and two on the outside (**Figure 5.10A** and **Figure 5.10B**). These two planes were linked together by two highly-conserved disulphide bonds. The core formed by the three-fold repeats in the centre of the domain was dominated with aromatic residues, forming an extremely hydrophobic core (**Figure 5.12A**), which greatly stabilised the whole domain. A conservation study with sequences of all available ApiPLPs showed that sequences in the inner plane (four β-strands) facing towards the core are more conserved than in the outer set (two β-strands), and that the disulphide bonds are highly conserved (**Figure 5.11**), which can be treated as a signature motif of this domain (Kaiser et al., 2004).

Sequence analysis had indicated that the structure of the C-terminal domain would

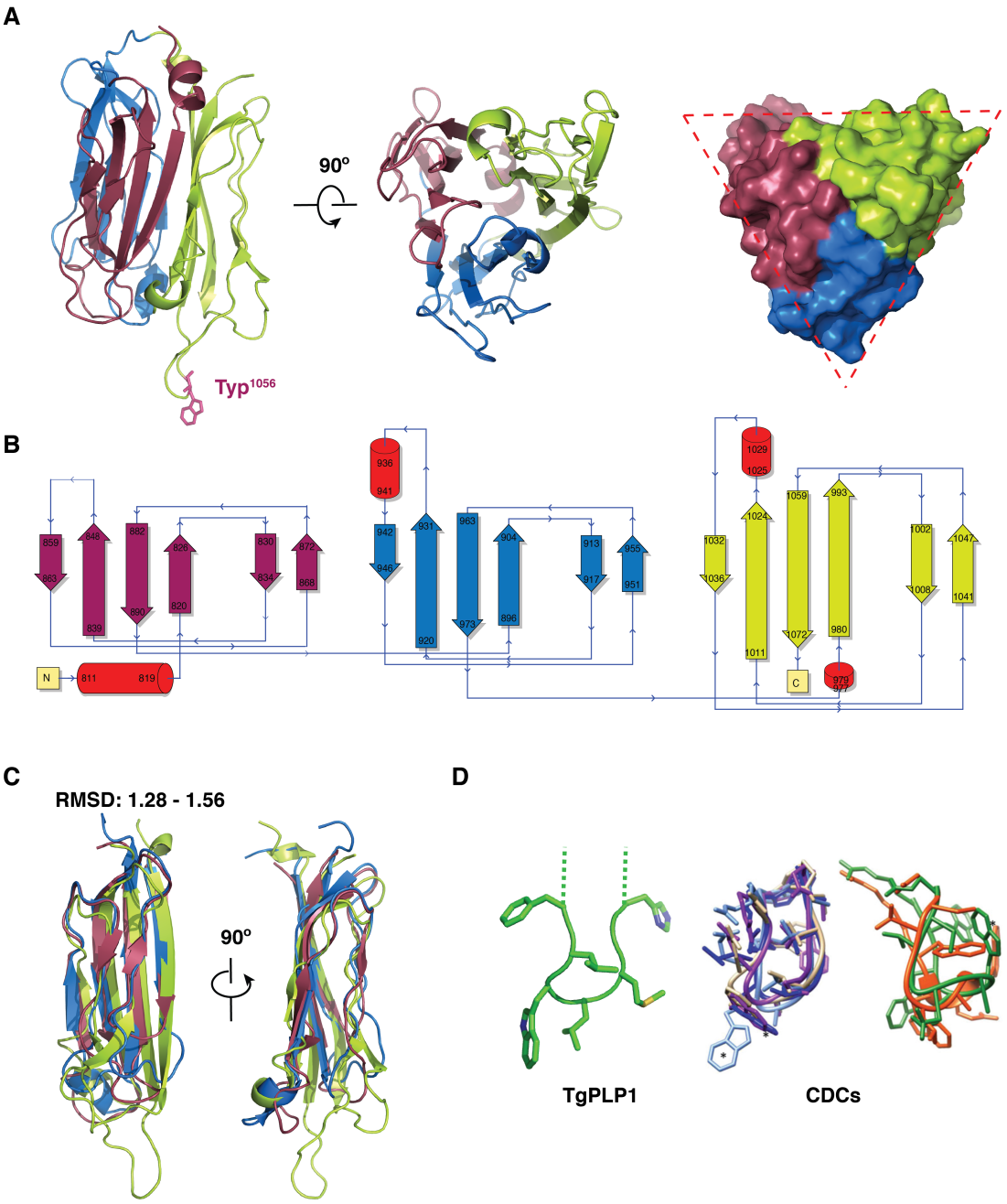

Figure 5.10: Overall structure of *Tg*PLP1 C-terminal domain. (A) Overall structure of the C-terminal domain in two orientations. The three repeats are coloured with blue, lemon and dark red, respectively. The 3-fold pseudo symmetry is readily identifiable from the top view (middle). The overall triangular shape of the C-terminal domain is highlighted with a dashed-line triangle. (B) Topology diagram of the C-terminal domain showing the arrangement of the β-strands in each domain. The three domains are coloured in the same way as in (A). (C) Superimposition of the three repeats. (D) Closeup view of the extended loop of C-terminal domain and comparison with the undecapeptides of cholesterol-dependent cytolysins. The image of the undecapeptides was adapted from Gilbert 2014.

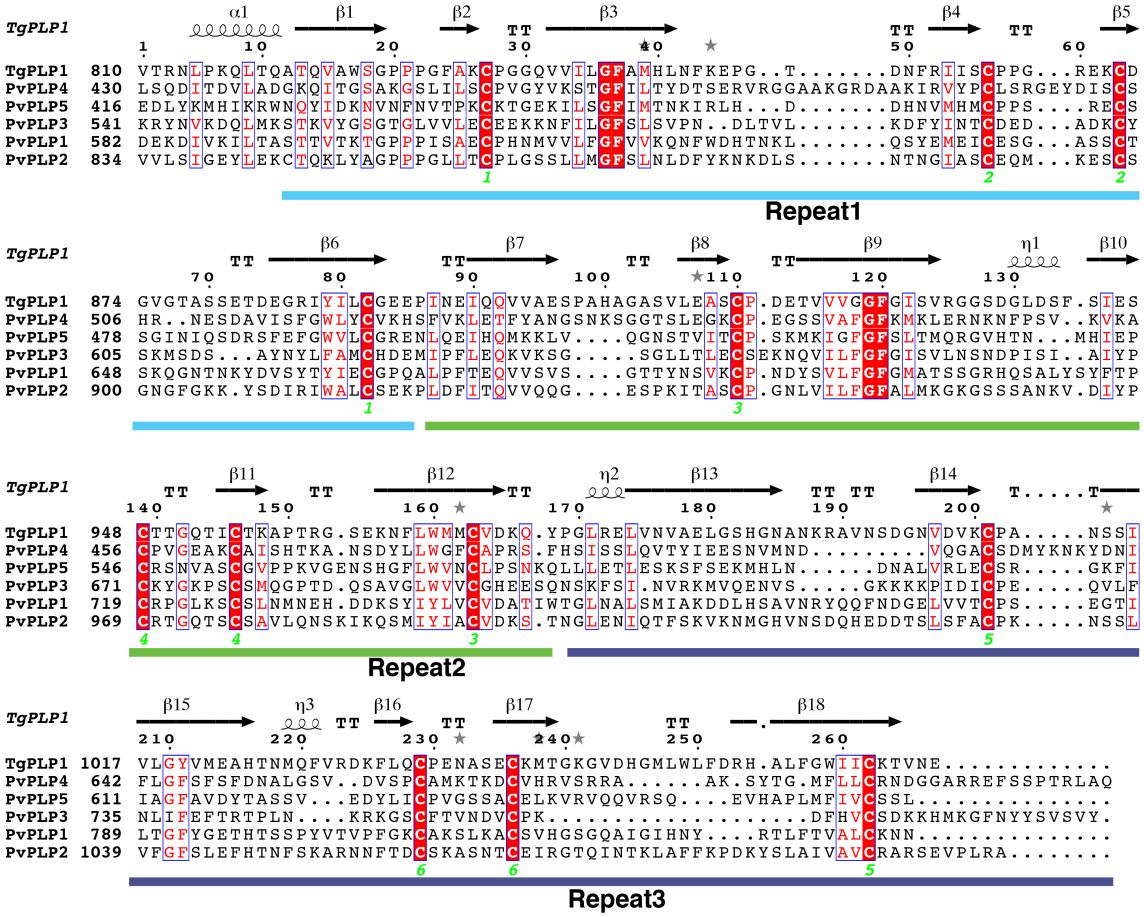

Figure 5.11: Sequences alignment of C-terminal domain. The C-terminal domains of PLPs from Toxoplasma gondii and Plasmodium vivax vary in sequence but are predicted to share the same fold. The cysteine are identified to be the most conserved residues, forming two disulphide bonds within each repeat. 18 β-sheets are equally distributed into three repeats of the C-terminal domain.

not fall into any of known structural family. Nevertheless, solution of the actual structure has identified structures with similarity. The repeats of *Tg*PLP1 C-terminal domain showed a certain degree of similarity to a lectin (PDB entry: 3WMP, RMSD: 2.77 - 3.15 Å), which is involved in sugars-mediated membrane binding. Lectin-like domains generally contain six β-strands and some of them formed a homo-trimer which in the *Tg*PLP1 CTD has obviously been varied by gene duplication and fusion to make a single polypeptide of similar appearance. Different to *Tg*PLP1 C-terminal domain, however, the lectin domains consist of two sheets with three β-strands in each. Another protein sharing some degree of similarity is the fibronectin domain type III (Fn(III)) of Roundabout homolog 1 (PDB entry code: 3WIH, RMSD between the third repeat of PLP1 C-terminal domain and Fn(III) is 2.7 Å), which contains seven β-strands with four in one sheet (equivalent to the inner set in *Tg*PLP1 CTD) and three in the other. The overall artchitecture with a 3-fold pseudo-symmetry was also observed with Vitelline membrane outer layer protein 1 (VMO-1, PDB entry: 1VMO), an eggshell protein with unknown function (Shimizu et al., 1994). Each repeat of VMO-1 contains four β-strands instead of six, and a single disulphide bond is identified within each repeat. Interestingly, both of the C-terminal domains of Plu-MACPF from *Photorhabus luminescens* (Rosado et al., 2007) and Pleurotolysin from fungi (Lukoyanova et al., 2015) are β-strands structures with 3-fold pseudo-symmetry, similarly to that of *Tg*PLP1. Thus overall, the domains discussed here for comparison with the *Tg*PLP1 C-terminus show a rather limited similarity to it; if they are homologous and share a common ancestor in a meaningful sense then their divergence would appear to be very ancient. Alternatively, they may be analogous structures and lack a meaningful common ancestor.

A loop containing several hydrophobic residues including a tryptophan, was identified at the base of the C-terminal domain. The sequence and structure of this loop resembled those of the undecapeptides from cholesterol-dependent cytolysins (Gilbert, 2014; Tweten, 2005), which were shown to be responsible for initial membrane

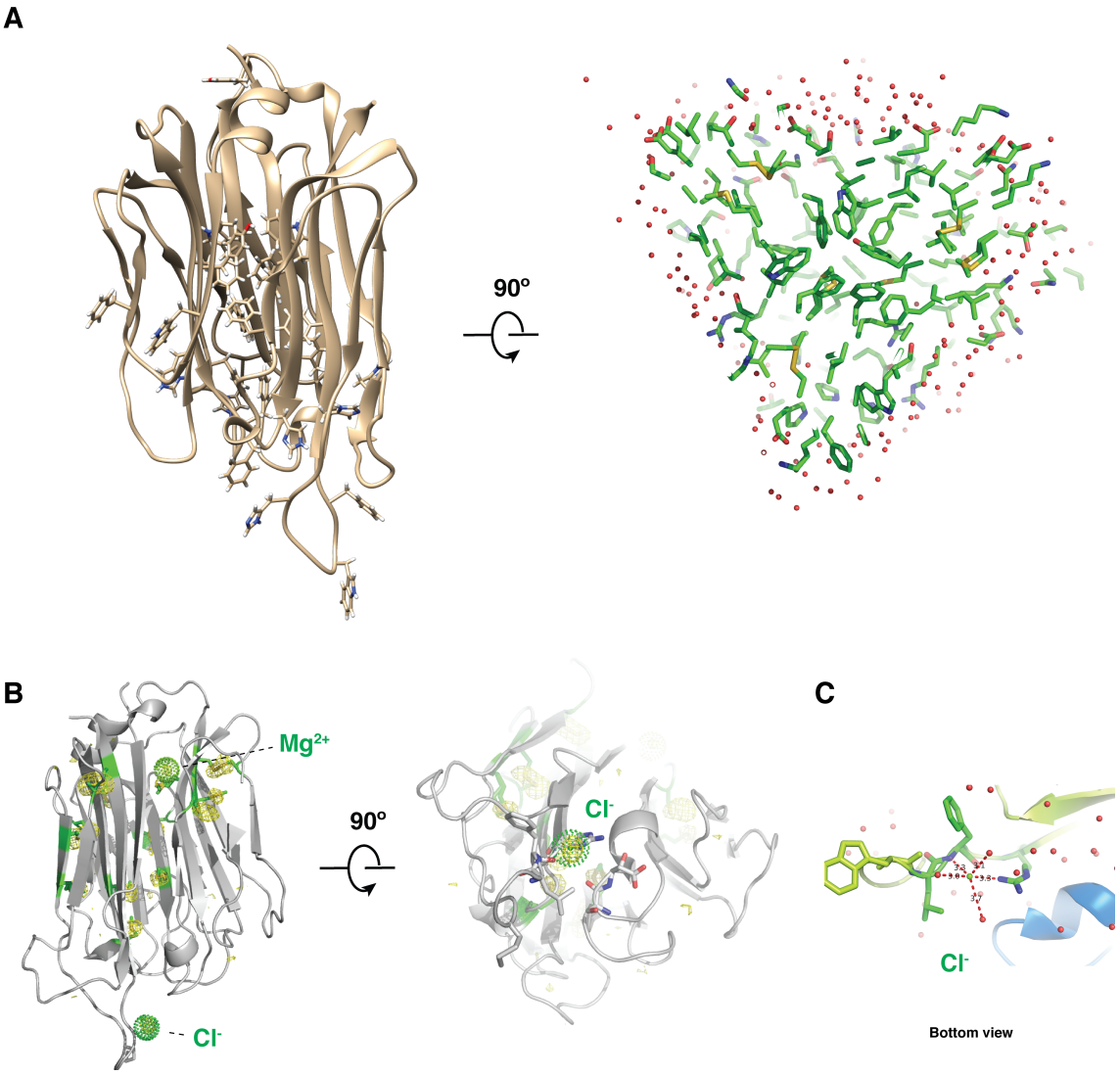

Figure 5.12: Hydrophobic core of C-terminal domain and ion binding analysis. (A) Overview of aromatic residues in the C-terminal domain (left) and an intersection view from "beneath" (with the Trp-containing loop at the base of the domain towards the viewer). Water molecules are rendered as red dots and are shown only on the surface of the molecule. (B) Anomalous difference map highlighting the positions of 'heavy-atoms' including sulphurs, magnesium and chloride ions. The electron density is shown in yellow and contoured at 3σ. Magnesium and chloride ions are highlighted with green spheres. (C) Close-up view of a chloride ion bound to the bottom of the C-terminal domain.

attachment of the proteins in this family. This observation led us to consider whether the C-terminal domain could bind to membrane in a cholesterol-dependent way, which is currently under investigation.

5.3.4 Solution structure of *Tg*PLP1

MACPF/CDC superfamily proteins often present as soluble monomers or dimers before being targeted to membranes to form an oligomeric pore-forming states. *Tg*PLP1 is monomeric in solution, as shown by being a single species in analytical ultracentrifugation (AUC) with a sedimentations coefficient of 3.54 S (**Figure 5.13A**). SAXS measured at several different concentrations indicated some concentration-dependent self-interaction, with a stepwise increase of radius of gyration (R_g) from 33.9 Å to 37.5 Å (**Figure 5.13B**).

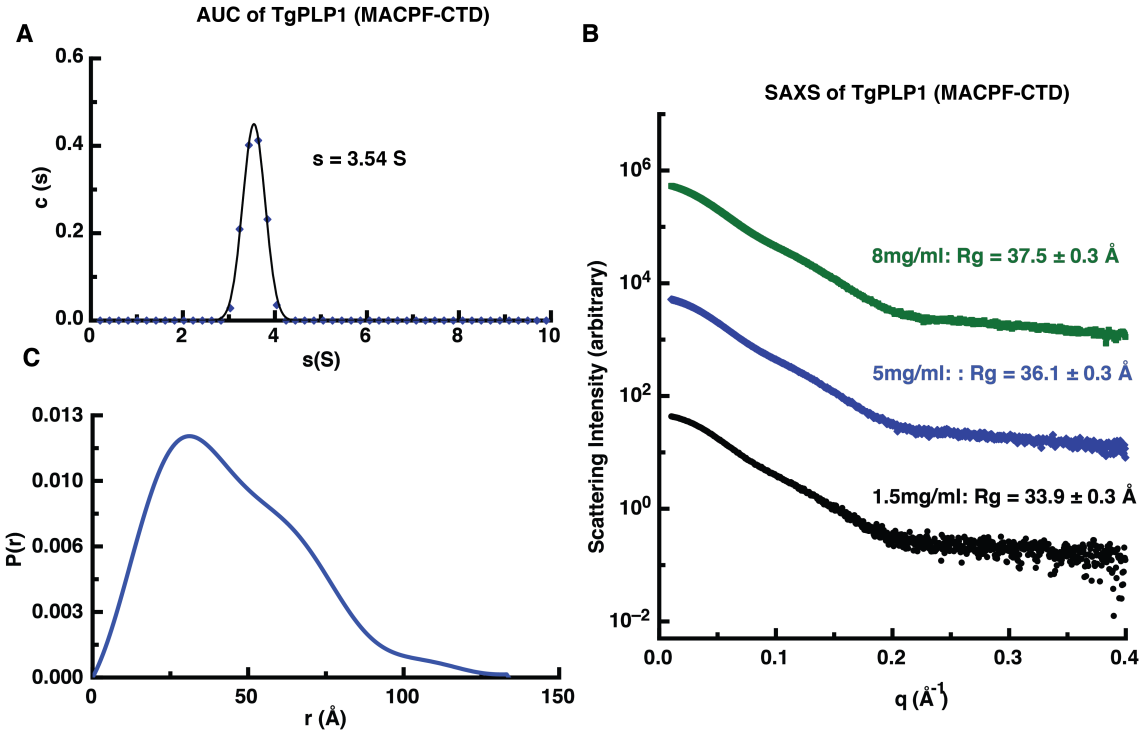

Figure 5.13: AUC and SAXS study of *Tg*PLP1(MACPF-CTD). (A) AUC analysis of $TgPLP1_{462\text{-}1072}$ at 1 mg/ml showing the monomeric property of the protein in solution, with sedimentation coefficient of 3.54 S. (B) SAXS analysis of $TgPLP1_{462\text{-}1072}$ at three different concentrations: 1.5, 5 and 8 mg/ml. The calculated Rg values are shown as well. (C) Distance distribution function (P(r)) of $TgPLP1_{462\text{-}1072}$ at 5 mg/ml.

Topologically, the MACPF/CDC pore-forming domain (here, the "MACPF" domain) is usually closely accompanied by a membrane binding domain, such as the C2-like domain in perforin-1 (Law et al., 2010b), Ig domain in pneumolysin (Lawrence et al., 2015b) and annexin-like domain in astrotactins. The overall architecture of these proteins are usually very rigid. For example, perforin-1 contains a EGF-like domain closely packed with MACPF domain (specifically with a disulphide bond), followed by the C2 domain in the bottom. Astrotactins are composed of a MACPF-EGF-FN(III)-Annexin cassette. The MACPF domain and C-terminal domain of *Tg*PLP1 are linked together with a 10-residue linker, the structure of which is not resolved in the structure of either domain.

To explore whether the rigidity observed with other MACPF/CDC proteins applies to *Tg*PLP1 as well, we modelled the structure of the MACPF domain and C-terminal domain against SAXS data using the online server software Multi-FoXS (Schneidman-Duhovny et al., 2013). About 10,000 models were generated, treating both MACPF domain and C-terminal domain as rigid bodies, and the 10 amino acid residues between these two domains as a flexible linker. These models are fitted against the SAXS data using either single conformation (one-state model) or combination of several conformations (two-state model, three-state model, etc).

As shown in **Figure 5.14**, the relative orientation of the MACPF domain and C-terminal domain was not fixed, instead at least two dominant species were resolved. Both of the two conformations revealed that the C-terminal domain resided on the same side of the central twisted β-strands of the MACPF domain (near α1 and α2). The one-state model suggested that a single species was able to fit the SAXS data very well ($\chi = 0.89$ for the best one). Unexpectedly, superimposition of the top 10 models (all with $\chi < 1$) revealed that these were divided into two major groups, with the C-terminal domain placed in two different orientations with respect to the MACPF domain. The two-state model fits slightly better with the scattering data than the one-state model ($\chi = 0.85$). Echoing with our observation using a one-state

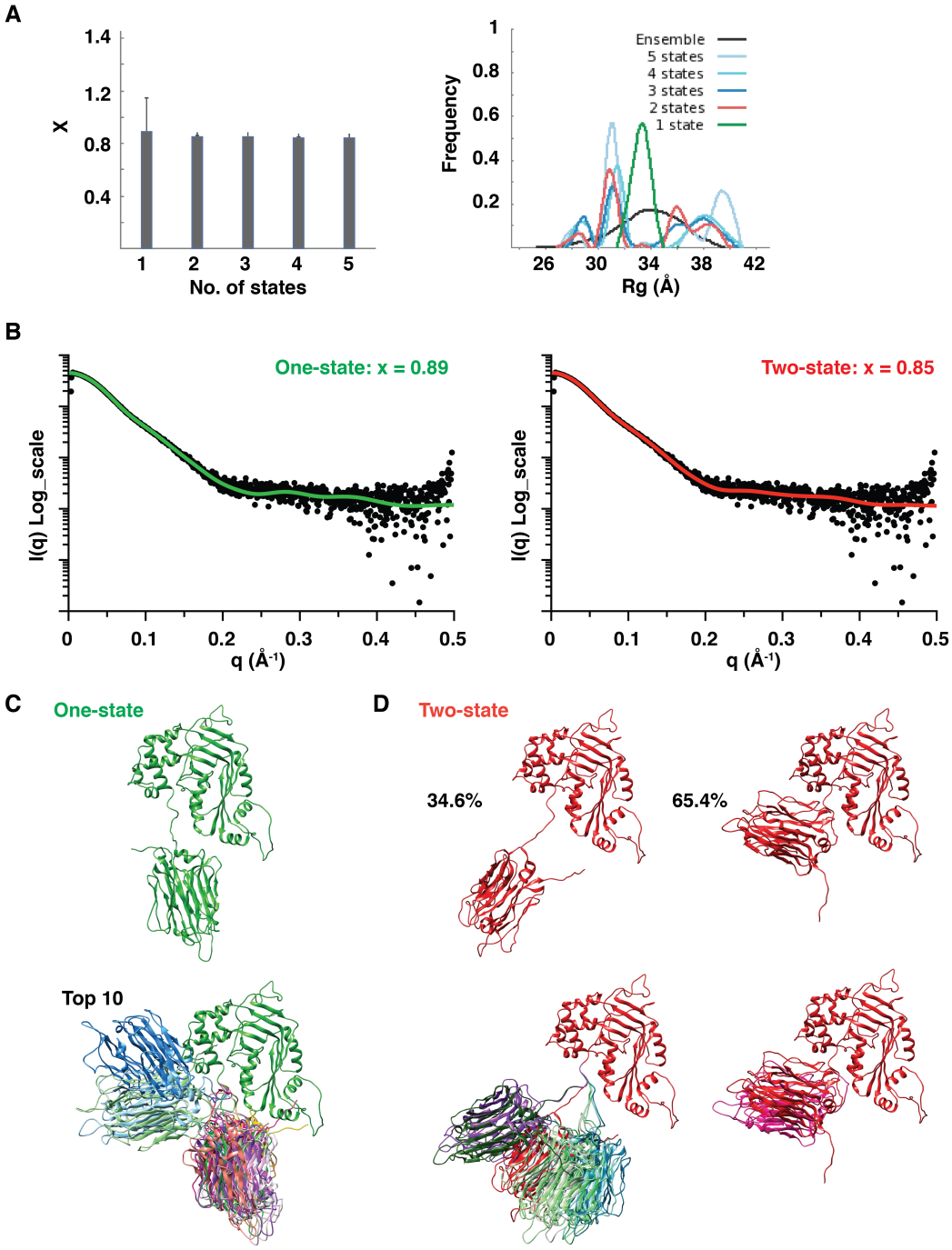

Figure 5.14: SAXS modelling of *Tg*PLP1 (MACPF-CTD) structure. (A) (left): The lowest χ scores for each of the N-state models (N = 1-5) are shown with an error bar indicating the range of χ values for the top 1000 models; (right): Rg distributions for the entire ensemble of the 10,000 conformations and the top 1000 N-state models. (B) Comparison of the experimental SAXS profiles (black dots) with computed SAXS profiles for the top-scoring single-state (left) and two-state (right) models. (C) Conformations of the top-scoring single-state and two-state models (B) are shown. These structures are aligned with the MACPF domain.

models, the two conformations from each of the two-state models resemble with these two conformations found in the top 10 one-state models. There was no significant improvement of χ score for models with three or more states.

Overall, the SAXS models indicated that it is very likely that the *Tg*PLP1 (MACPF-CTD) exists in the form of at least two conformations in solution, rather than the rigid architecture observed with other proteins in MACPF/CDC superfamily. This also explains our failure in attempts to crystallising the larger construct (MACPF-CTD), as described in **Chapter 6**.

5.3.5 Proteolytic processing and oligomerisation of *Tg*PLP1 secreted into mammalian media

The full-length *Tg*PLP1 secreted from mammalian cell lines migrated on SDS-PAGE gel as three distinct bands, with molecular weights corresponding to an oligomeric form (larger than 300 kDa), full-length protein (about 130 kDa) and a fragment (80 kDa) composed of MACPF and C-terminal (**Figure 5.15A**). Purification of the full-length construct with a C-terminal His_6-tag resolved two distinct bands (with molecular weights of about 130 kDa and 80 kDa, respectively) (**Figure 5.15B**), indicating that the cleavage site is between the N-terminal and MACPF domain. We were not able to isolate the dimeric/oligomeric species by His-trap, presumably because the His_6-tag was hidden in these species. Constructs containing the N-terminal domain and MACPF domain (residues 1 - 792) also showed cleavage between the N-terminal and MACPF during secretion (data not shown).

Interestingly, the proteolytic processing and dimerisation/oligomerisation of *Tg*PLP1 were not observed for samples in cell lysate, implying that the proteolytic processing and dimerisation/oligomerisation occurred in the media after secretion. Moreover, the dimer/oligomer species were absent in constructs without N-terminal region, which suggests that N-terminal region is indispensable for the dimerisation or oligomeri-

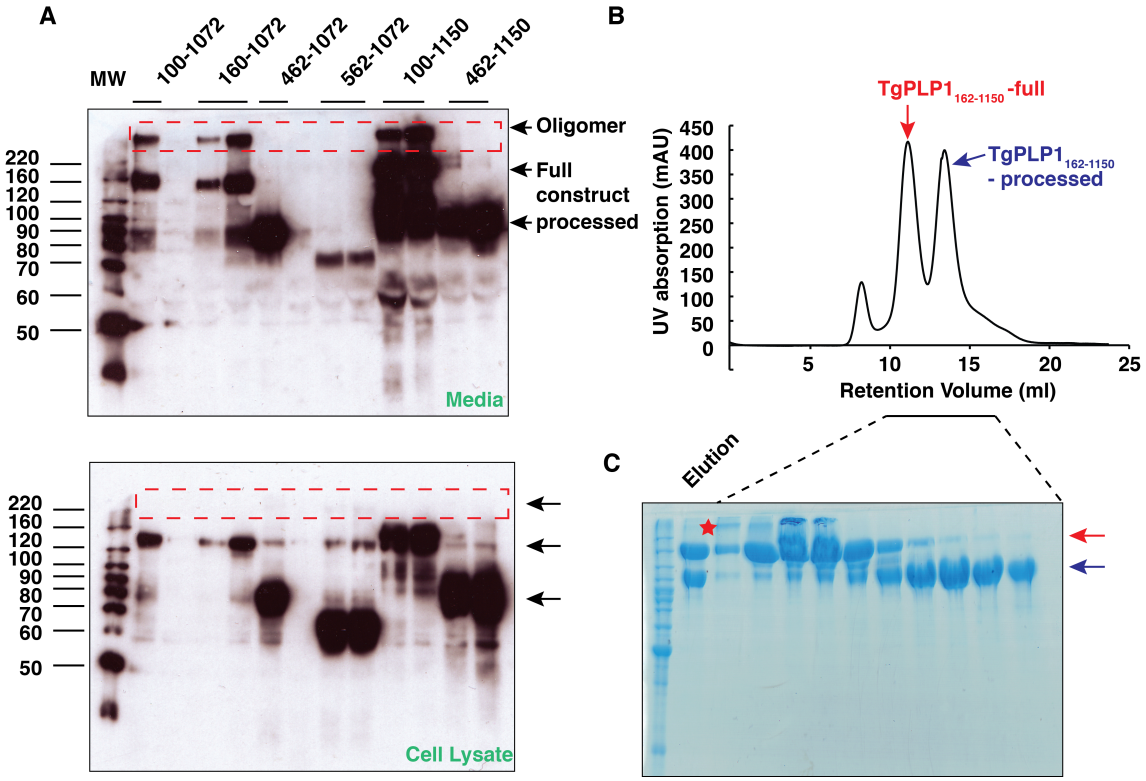

Figure 5.15: Expression and purification of *Tg*PLP1 constructs. (A) Representative results of small-scale expression tests of *Tg*PLP1 constructs. Both the media and cell lysates were subject to western blot using anti-pentahistidine antibody. Note the presence of high molecular weight oligomers of near full-length proteins in the media. The oligomers were absent without the N-terminal domain and were only observed in the media in contrast to cell lysate, highlighted by dashed red lines. The proteolytic processing of near full-length *Tg*PLP1 was also observed in several constructs. (B) Size-exclusion chromatography profile of $TgPLP1_{162\text{-}1150}$ after elution from Nickel-NTA column. There are three peaks from the size-exclusion chromatography profile, with the first one as aggregates eluted at the void volume of the column and two latter peaks corresponding to the intact and processed fragment. (C) SDS-PAGE of $TgPLP1_{162\text{-}1150}$ of the samples from size-exclusion chromatography in (B). The very left of the gel is the molecular weight marker and the second lane is the sample before size-exclusion chromatography. The latter samples are all from the size-exclusion chromatography corresponding to the two peaks in (B). The red and dark blue arrows indicate intact and processed proteins and the red star highlights the oligomer in solution.

sation during secretion. However, it remains unknown whether this dimerisation/ oligomerisation occurrs during secretion is relevant to its pore-formation capacity. Further characterisation with purified material will resolve these issues.

5.4 Discussion

5.4.1 Mechanism of pore formation by *Tg*PLP1

The structure of the *Tg*PLP1 MACPF domain revealed that it is equipped with all the essential elements for pore formation in a way similar to other members of the MACPF/CDC superfamily of proteins (**Figure 5.4**), together with a C-terminal domain potentially responsible for initial membrane attachment. It is very likely that *Tg*PLP1 is able to form a pore on the membrane, but the details await further discovery.

We were surprised that MACPF domain packed itself in crystals in either a hexameric form or a helical form with six subunits per turn, which is a stoichiometry of oligomerisation not reported previously for other members of MACPF/CDC superfamily proteins. Although we cannot rule out the possibility that both of our ring form and helical form of MACPF domain structures are artefacts due to crystallisation packing, there are several features implying that both forms are physiologically relevant in uncovering the mechanism of pore formation. Firstly, the orientation of MACPF domain in the pore showed that the central twisted β-sheets faced to the centre of the ring/helix, which matches the observation with other MACPF/CDC proteins, such as the pneumolysin prepore (Gilbert et al., 1999), MAC pore (Serna et al., 2016) and poly-C9 (Dudkina et al., 2016). Secondly, the helical form of the MACPF domain found in the $P6_5$ crystal form recapitulates the observation of a helical form of pneumolysin in solution which is structurally related to the ring-form pores of pneumolysin (Gilbert et al., 1999). The pneumolysin helix contains about 41 subunits each turn, matching the size of its pore form with about 40 to 50 subunits in a ring; the subunits in the helix and ring are packed similarly and the conformation of the helical subunit is equivalent to the pore (but not the pre-pore) form (Gilbert, 2005). Thirdly, both of these crystals were obtained using limited *in situ* proteolysis which removed the outer set of helices to leave space for the packing together of

the MACPF domains, and of course these helices unfurl during pore formation in pore-forming MACPF/CDC proteins. Fourthly, the edge-on interaction of the β-sheets between adjacent MACPF domains in the hexameric helix and ring formed by the *Tg*PLP1 MACPF domain are the interactions expected from cryo-EM studies of pores formed by the MAC (Serna et al. 2016) and pleurotolysin (Lukoyanova et al., 2015). Taken together these considerations indicate that both the ring form and helical form are mimicking transient stages towards pore formation.

The size of pores formed by MACPF/CDC proteins are very diverse, ranging from 45 nm (pneumolysin) to 8 nm (Pleurotolysin). Both the helical and ring form structures reconstructed from crystal packing point to the high possibility of pore formation by hexameric *Tg*PLP1 (5.35 nm in diameter) (**Figure 5.8B**). This is significantly smaller compared with those of other MACPF/CDC structures and such a tight curvature is accomplished by the more bulky shape of the *Tg*PLP1 MACPF domain (it is less planar than others, as discussed above) and by tilting of the subunits from the vertical orientation normally observed for side-by-side MACPF/CDC subunit oligomerisation. However, without direct evidence of the hexameric form in solution or on a membrane, the conclusion that *Tg*PLP1 forms pores using a hexameric assembly is not yet validated. Further experiments investigating the pore-forming mechanism in solution and on a membrane are under way. For example, an initial negative staining electron microscopy analysis of these detergent-induced oligomers would be helpful to understand the nature of these oligomers: whether they form hexameric species or even larger species? In addition, in a liposome based membrane systems, it would be insightful to know whether *Tg*PLP1 has any preference to specific lipids.

It was proposed that ApiPLPs bound to Ca^{2+}, similar to the C2-like domain of perforin-1 (Conroy et al., 2015), and both *Pf*PLP1 and *Pf*PLP2 were shown to have a Ca^{2+}-dependent pore-forming activity using *Pf*PLPs recombinantly expressed in mammalian cells (Garg et al., 2013). Unfortunately, we were unable to repeat

their results since we failed to express these two proteins even following the same protocol as claimed (Garg et al., 2013), except that the genes in their study might be codon-optimised in a different way. In fact, there is no direct evidence showing the interaction of Ca^{2+} with any of these ApiPLPs.

We are keen to know whether Ca^{2+} binds to the C-terminal domain of *Tg*PLP1, or in general, whether it binds to the C-terminal domain of ApiPLPs. Thermofluor study of the *Tg*PLP1 C-terminal domain showed that stability of the protein was slightly compromised in the presence of high-concentrations Ca^{2+}, indicating a minor negative effect upon Ca^{2+} binding (data not shown). We have not yet obtained a crystal of the C-terminal domain in the presence of calcium. The crystallisation conditions which resulted in our highest resolution structure (1.1 Å) contained $MgCl_2$, and both ions (Mg^{2+} and Cl^-) were shown to bind to this domain in crystal structure. In an attempt to solve the structure by sulphur-SAD, we collected an X-ray diffraction data at a wavelength of 1.77 Å. The anomalous difference map showed strong intensity of electron density from anomalous scatters, in which disulphide bonds and methionines could be unequivocally located. Moreover, it was feasible to identify other 'heavy' atoms, such as Cl^- and Mg^{2+} ions (**Figure 5.12B**). We are now considering soaking the crystals with Ca^{2+} in an attempt to clarify the puzzle of whether calcium does bind this protein.

5.4.2 Proteolytic processing and dimerisation of *Tg*PLP1

ApiPLPs contain regions with conserved MACPF domains and C-terminal domains, while the N-terminal domain is rather diverse both in protein sequence and length (**Figure 5.3A**). Identifying the exact cleavage site between N-terminal region and MACPF domain and assessing its biological function (if any) is a remaining challenge. In fact, *Tg*PLP1 is not the only perforin-like protein that was observed to undergo proteolytic processing, *Py*PLP1 (*Plasmodium yoelii* perforin-like protein 1) was also detected as two distinct species with molecular weights of about 80kDa and 30 kDa,

respectively, with an antibody specifically binding to the N-terminal region (Kaiser et al., 2004). The cleavage between N-terminal domain and MACPF domain was also observed for *Pf*PLP1 (SPECT2) and *Pf*PLP2 (Garg et al., 2013; Ishino et al., 2005). It is possible that the N-terminal region is present for tuning the function of ApiPLPs, for example, mediating the equilibrium between monomeric and oligomeric forms of proteins.

It is not known whether the dimerisation/oligomerisation during secretion is specific to *Tg*PLP1 and whether it is relevant to the pore-formation capacity. The N-terminal regions of ApiPLPs are diverse both in sequences of amino acid residues and size: *Pv*PLP1, *Pv*PLP2 and *Pv*PLP3 contain N-terminal regions with about 250 to 500 amino acid residues; while *Pv*PLP4 and *Pv*PLP5 do not contain an N-terminal region at all (**Figure 5.3A**). We are now proceeding to characterise dimer/oligomers of the full-length *Tg*PLP1, and a comparison with *Pv*PLPs will be very useful to check whether this property is unique to *Tg*PLP1.

5.4.3 Structure of MACPF/CDC superfamily proteins in pore-forming state at atomic resolution

Recent years witnessed structure determination of new MACPF/CDC superfamily proteins at an atomic resolution by X-ray crystallography. These structures provided more details of structural features that are unique to these proteins, and also demonstrated the high degree of structural conservation within this superfamily. The mechanism of pore-formation elucidated by cryo-EM studies of pneumolysin (Gilbert et al., 1999, 2014; Tilley et al., 2005), perforin-1 (Law et al., 2010b) and pleurotolysin (Lukoyanova et al., 2015) revealed large conformational changes of MACPF/CDC domains during pore formation. These cryo-EM studies provided strong and direct evidence demonstrating that the two sets of helices from either side of the central MACPF domain β-sheet undergo a dramatic conformational transition from α-helices to anti-parallel amphipathic β-hairpins. In line with this it

was proposed that the hydrophilic residues were exposed facing into the centre of the pore and hydrophobic ones formed contact with lipid tails of membrane bilayer after pore formation (Shaturksy et al, 2000). Other regions of MACPF/CDCs also showed large conformational changes. However, without a structure of MACPF/CDC proteins in the pore form at an atomic resolution, uncovering the detailed mechanism of pore formation remains difficult.

Efforts in obtaining a structure of MACPF/CDC proteins in mature pore form at an atomic resolution have not been successful. Major challenges lie in the instability of pores and large variance in stoichiometry of pores. Pneumolysin, for example, formed a ring using 40 to 50 copies of its monomers (Gilbert et al., 1999; Tilley et al., 2005), while rings of perforin-1 were much smaller (about 16 nm, with 18 to 22 copies per ring) (Law et al., 2010b). Earlier cryo-EM study of these pores by either single particle reconstruction or subtomogram averaging could only achieve a resolution of about 25 Å, limited by both the technique itself and the samples available. The helical form of pneumolysin spontaneously formed in solution at high concentration (or promoted by use of a D_2O-based buffer instead of an H_2O buffer (Gilbert et al., 1998) has been a promising target to obtain a higher resolution structure, due to the inherent properties of a helix, although it required a significant amount of work to prepare enough images for helical reconstruction attained previously.

Recent development in structural techniques, revolutionised by advances in cryo-EM (Cheng et al., 2015), now show the greatest promise in deciphering the structure of these MACPF/CDC pores at an atomic resolution. Considering the heterogeneous property of pores formed by membrane attack complex and by poly-C9 alone, it was of great structural value to resolve both structures at a resolution of about 8 Å (Dudkina et al., 2016; Serna et al., 2016), to reveal more details of the overall architecture of MAC. In particular, the positions and orientations of individual components (such as C5b, C6, C7, C8 and C9) were resolved within the structure of whole MAC (Serna et al., 2016), settling once and for all confusion about the mechanism of pore-formation

by perforin, which was originally interpreted as having a distinct mechanism from the CDCs (Law et al., 2010b) (see Gilbert et al., 2013 for discussion). These studies also revealed a helical region (equivalent to α9 of *Tg*PLP1 MACPF) which contained a GGX_nW motif in the electron density map, indicating that this region was rigid during pore formation (Serna et al., 2016). The EM structure of the poly-C9 ring provided similar structural information backing up the findings of the MAC structure, although the poly-C9 alone is not a physiologically relevant pore-forming structure. Pores of pleurotolysin are mostly composed of 13 heterodimeric subunits in each ring (13 PlyA and 13 PlyB) with a diameter of 100 Å and height of 80 Å, which is relatively small compared with those of other MACPF/CDC members. Single-particle cryo-EM might be suitable to achieve a higher resolution structure for the pleurotolysin pore than currently reported (Lukoyanova et al., 2015).

Crystallising the mature pores of MACPF/CDC proteins is still a very difficult challenge. It is not verified yet whether the *Tg*PLP1 MACPF domain does assemble into a hexameric ring or helical structure in solution. If it does, it is very likely that other ApiPLPs assemble in a similar way (considering the high degree of conservation among them), and therefore that we have captured a structure representing the pore-forming oligomeric state of the ApiPLPs. Compared with pores assembled with a large and variable number of subunits (such as pores of the complement MAC and CDCs), smaller pores such as the ApiPLPs may apparently form are more likely to be structurally determined at a high resolution due to their stability and likelihood of homogeneity. Technically speaking, the hexameric pore form of the *Tg*PLP1 MACPF domain (or pores of other ApiPLPs) probably stands the best chance of providing a high resolution structure of a pore form by X-ray crystallography, and this is one of our targets in the next stage.

5.5 Methods and materials

5.5.1 Small angle X-ray scattering

Small-angle X-ray scattering (SAXS) experiment were carried out in 10 mM Tris pH 7.5, 150 mM NaCl with deglycosylated *Tg*PLP1 (residues 462 - 1072, MACPF and C-terminal domain) at 3 different concentrations: 1.5, 5, and 8 mg/ml. SAXS data were collected were collected on BM29 in at the ESRF with a momentum transfer range of 0.004 $Å^{-1} < q < 0.45$ $Å^{-1}$, where $q = 4\pi\sin(\theta)/\lambda$ and 2θ is the scattering angle. The scattering intensity from buffer alone was subtracted from the averaged data to obtain the protein scattering in solution. The radius of gyration (R_g) was calculated from a Guinier plot using AutoRg (Petoukhov et al., 2012). The particle distance distribution function P(r) was calculated in GNOM (Svergun and IUCr, 1992) using a low-resolution data range (0.01 $Å^{-1} < q < 0.15$ $Å^{-1}$). Data reduction was carried out using the ATSAS package (Petoukhov et al., 2012).

The combination of the MACPF and C-terminal domain was modelled into the SAXS data using the online server MultiFoXS (http://modbase.compbio.ucsf.edu/multifoxs/). Briefly, the disordered loops within the MACPF domain and a polyhistidine tag at the C-terminus of CTD were firstly modelled in MODELLER. Then, ten animo acid residues (residues 800 - 809), which were absent in both structures, were treated as a flexible linker between the MACPF domain and C-terminal domain acting as rigid bodies. In total, about 10,000 models were generated (Raveh et al., 2009), and the SAXS profile is then calculated for each conformation using FoXS method (Schneidman-Duhovny et al., 2010). The multi-state models were enumerated in MultiFoXS using the branch-and-bound method. In order to characterise the range of conformations consistent with the SAXS data, we analysed distributions of R_g for the entire ensemble of 10,000 models as well as the 1000 best-scoring N-state models ($N = 1 \ldots 5$).

5.5.2 Analytical ultracentrifugation

Sedimentation velocity experiments were performed using a Beckman Optima XL-I analytical ultracentrifuge equipped with absorbance and interference optics. Double-sector 3 mm centrepieces were used with protein samples at 1 mg/ml and using absorbance measurements at 280 nm. Experiments were performed at 20°C, taking sample distribution scans every 6 minutes. Data were analysed using SEDFIT software using the $c(s,f/f_0)$ method of interpretation to generate sample distributions in s (sedimentation coefficient) without assuming a particular number of species (Brown and Schuck, 2006b). The resulting distributions were curve fit in ProFit (QuantumSoft, Uetikon am See, CH).

5.5.3 Reconstruction of MACPF domain helices in solution

MACPF domain (residues 462 - 805) at a concentration of 40 mg/ml in 10 mM HEPES pH 7.5, 150 mM NaCl was mixed with 5% DDM in H_2O at 1:1 volume ration and incubated at 37℃ overnight. Heavy precipitation was observed overnight and these precipitants could be dissolved in presence of 0.5% SDS. To check the oligomerisation of protein in solution, the precipitated proteins were mixed with SDS-PAGE loading dye without further treatment (no cross-linking reagents were added) and a normal SDS-PAGE was performed. It was not explored yet whether oligomerisation of the MACPF domain could be induced at low concentration with DDM.

CHAPTER 6

Expression, Purification, Crystallisation and Structure Determination of Apicomplexan Perforin-like Proteins.

6.1 Summary

This chapter is focused on expression, purification and crystallisation of perforin-like proteins (PLPs) from Apicomplexan species (*Toxoplasma gondii* and *Plasmodium spp.*). Protein expression screening of plasmodia *(Plasmodium falciparum*, *Plasmodium berghei* and *Plasmodium vivax*) perforin-like proteins (*Pf*PLPs, *Pb*PLPs and *Pv*PLPs) carried out in different expression systems is presented and one successful expression protocol is discussed in detail here. Efforts leading to successful crystallisation and structure determination of the *Tg*PLP1 MACPF domain and C-terminal domain are described in detail.

6.2 Expression screening of *Plasmodium* and *Toxoplasma* PLPs

The *Plasmodium falciparum* genome is notable for its exceptionally low GC content of under 20% and PCR amplification of PLPs using genomic DNA proved to be very difficult (Christina Mayer, Andreas Sonnen and Robert Gilbert, unpublished work). Thus, codon-optimised genes for *Pf*PLP1, *Pf*PLP2 and *Pb*PLP1 (synthesised by Geneart Ltd) were used in the study here. There has been hardly any report of successful recombinant expression of Apicomplexan perforin-like proteins (PLPs) when the project was started, so great efforts were devoted to protein expression and purification of these proteins at its beginning and are thus described here.

6.2.1 Expression and purification of PLPs in *E. coli*

Given the fact proteins from the same MACPF/CDC superfamily, such as the minimal region of mammalian C8α (Hadders et al., 2007) and bacteria cytolysins, could be successfully purified from *E. coli* and further crystallised, we firstly carried out our protein expression trials in *E. coli*, due to its ease of manipulation and time and cost-efficiency in general. We designed a good number of constructs covering domains of PLPs including *Pf*PLP1, *Pf*PLP2, *Pf*PLP3 and *Tg*PLP1 with different tags (MBP, GST and Histidine$_6$ in OPPF pOPIN vectors) in the N-terminal or C-terminal of the proteins. Preparation of constructs in pOPIN vectors allows the same vector be used in three expression systems (*E. coli*, *Sf9* insect cells and HEK 293T mammalian cells), without the requirement of recloning the same DNA fragment among vectors specific to each expression system (Berrow et al., 2007).

The constructs were transformed into B834 or BL21 Star^{TM} (DE3) cells (ThermoFisher). Among the constructs tested, many could express proteins and a few could be purified (**Figure 6.1A** and **Figure 6.1B**). However, all the proteins pu-

rified from *E coli.* were not correctly folded: firstly, the retention volume of these proteins in size-exclusion chromatography were all the same and within the void volume of the columns; secondly, dynamic light scattering showed the particle size to be about 50 nm, which is much larger than expected for these proteins; lastly, the thermofluor stability assay showed that the proteins bound to hydrophobic dyes at room temperature (25 ℃) and did not have a characteristic protein denaturation curve. Refolding of proteins using several different additives (NDSBs, glycerol, arginine and glutamate etc.) was pursued without success.

Considering the localisation of *Tg*PLP1 and *Pf*PLP1 in pathogens (Kafsack et al., 2009)(Amino et al., 2008a) and the conserved cysteine pairs in the C-terminal regions of PLPs, we then tried Rosseta-gami cells (Cat. No.71432, Merck Millipore), which allowed disulphide bond formation in the cytosol. Initial small-scale expression test revealed that only one of these constructs showed expression and could be purified (*Tg*PLP1$_{461\text{-}792}$-pOPINF encoding the MACPF domain with an N-terminal His$_6$ tag). The expression was scaled up and about 1mg of protein was obtained from 10 L of LB overnight culture. Size exclusion chromatography showed one peak corresponding to the monomer size of *Tg*PLP1$_{461\text{-}792}$ (**Figure 6.1C**). The purified protein was applied for crystallisation screening at 5 mg/ml but unfortunately resulted in no crystals.

6.2.2 Protein expression and purification from yeast and mammalian cytosol

We thought that failure of expressing correctly folded protein was due to the simplicity of *E. coli* expression system and the proteins could be potentially obtained in eukaryotic cells. To test this possibility, constructs encoding the same regions of *Pf*PLP1 were made into the pRS423GPD vector (LaCount et al., 2005; 2009), with modified His$_8$ and Flag-tag in the C-terminus, and transformed into a temperature-sensitive yeast *Saccharomyces cerevisiae* strain (both the pRS423GPD vector and

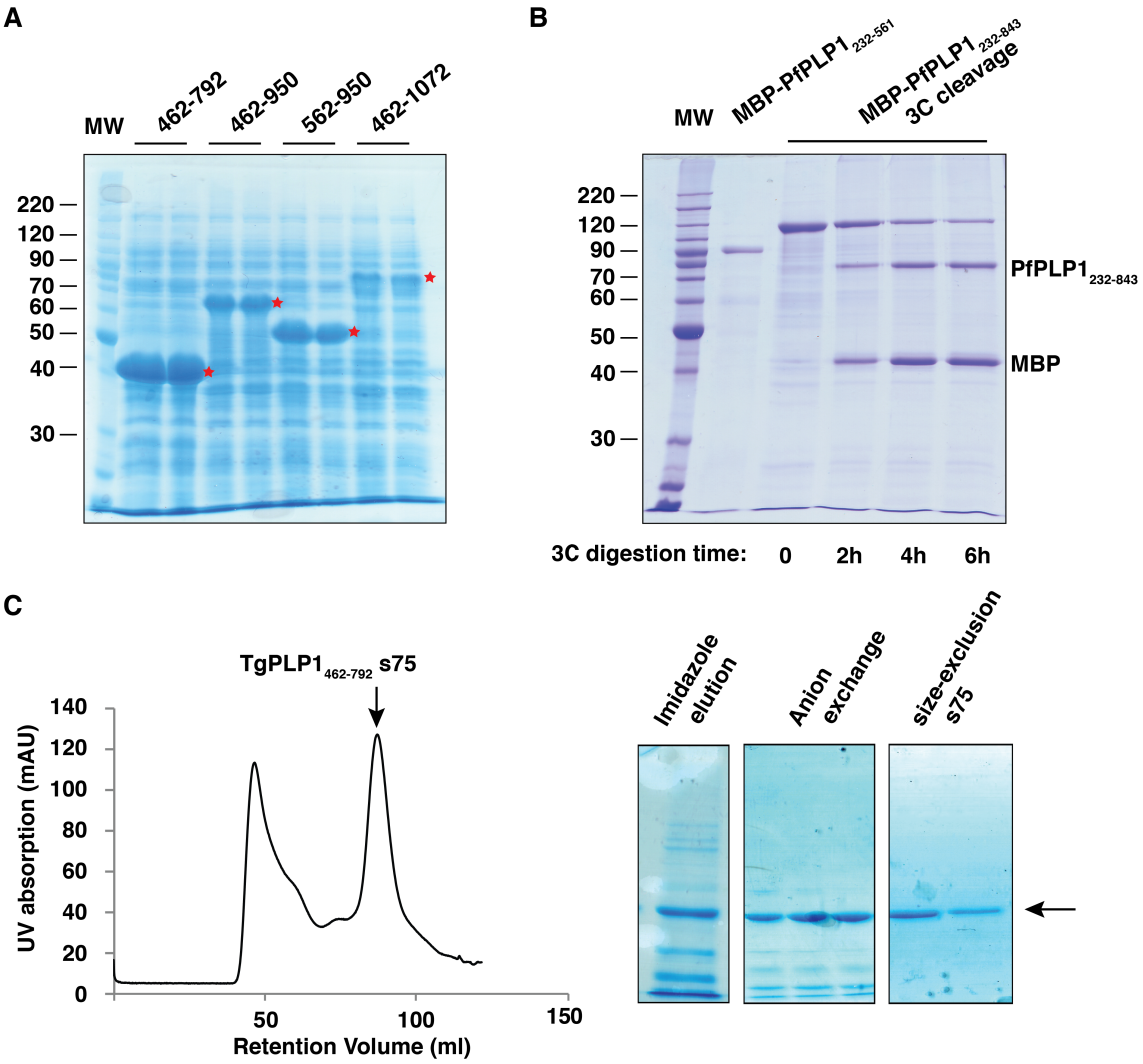

Figure 6.1: Representative gels and gel filtration profiles showing the expression and purification of Apicomplexan PLPs from *E. coli*. (A) Representative SDS-PAGE gel showing the expression of *Tg*PLP1 in *E. coli*. The red stars shows the high expression level of *Tg*PLP1 constructs. (B) Purification of MBP-*Pf*PLP1 from *E. coli* B21. Two constructs encoding MACPF and MACPF-CTD of *Pf*PLP1 were purified from *E. coli* cell lysates using amylose resin. MBP was successfully removed with 3C cleavage, but proteins without MBP are not stable. (C) Purification of *Tg*PLP1 MACPF domain (residues 462 - 792) from Rosetta *gami* cells. A size-exclusion chromatography profile with the expected molecular weight is shown on the left and representative gels from the three-step purification is shown on the right. The peak eluted around 50 ml corresponds to aggregated MACPF proteins and other contaminants. Pure *Tg*PLP1 MACPF protein peak is indicated with a black arrow. 10 L *E. coli* with transformed *Tg*PLP1 MACPF domain construct were harvested and lysed with sonication in 50 mM Tris pH 7.5, 500 mM NaCl. The cleared lysates were applied to Ni-NTA beads and eluted with 50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole and buffer-exchanged into 20 mM Tris pH 8.5, 50 mM NaCl to load onto anion exchange column Q. The eluted proteins were applied to superdex S75 in 20 mM HEPES pH 7.5, 150 mM NaCl.

yeast strains were kind gifts from Dr. Douglas La Count, University of Washington). All the constructs could be expressed in yeast cells, although the expression level was too low to purify using nickel-chelating sepharose beads. We then tried expressing these proteins in mammalian cell line, and again protein expression was confirmed in western blot, although the yield is far from that required for a biophysical study. Due to a lack of success with this strategy these data are not shown.

6.2.3 Expression screening with mammalian secretion system

Taking account of failures of protein expression and purification in *E. coli*, yeast and mammalian cytosol, and the little but encouraging success of *Tg*PLP1 MACPF protein expression in *E. coli* Rosetta *gami* cells, we started to consider protein expression in mammalian secretion system. The mammalian secretion system has several obvious advantages here: (i) the mammalian cell line has complex quality control system in endoplasmic reticulum (ER), and misfolded proteins would be largely degraded; (ii) given the structural and functional similarity between MACPF domain of PLPs and mammal perforin-1, and that these proteins are all localised in vesicles and secreted to exert their functions, secretion system would mimic the pathways of protein maturation; (iii) the C-terminal region of all apicomplexan PLPs, including these from *Plasmodia* and *Toxoplasma*, comprises 6 pairs of conserved cysteines forming 3 tandem repeats, which could potentially form disulphide bonds.

Domains of PLPs (*Pf*PLP1, *Pf*PLP2 and *Tg*PLP1), the same as cloned in pOPIN vectors for *E. coli* expression system and modified pRS423GPD vector for yeast expression system were inserted into mammalian expression vector pHlsec (Aricescu et al., 2006) between the restriction cutting sites AgeI and KpnI. Small-scale expression test revealed that almost all constructs could be detected in cell lysate, confirming a successful expression of these proteins, although only *Tg*PLP1 constructs resulted in successful secretion into media (see **Figure 5.15** in **Chapter 5**), which became the major focus of this project and the details of purification will be described later.

To maximise the success of PLPs expression and secretion, we continued to try other PLPs from other *Plasmodium spp.* Considering the fact that the *P. falciparum* and *P. vivax* caused the vast majority of deaths in malaria pathogenesis, and that GC-content of *P. vivax* genome is about 50% instead of 25-30% in *P. falciparum* (Carlton et al., 2008; Hall et al., 2005), we cloned different regions of *Pv*PLP2 and *Pv*PLP3 from the genome of *P. vivax* (a kind gift from Dr. Arturo Reyes-Sandoval, Jenner Institute, Oxford) into mammalian expression vectors. Full-length *Pv*PLP1 was codon-optimised and synthesised because gene encoding *Pv*PLP1 is composed of eight exons, instead of a single exon as in the case of the other PLPs. To facilitate better visualisation during purification steps, all the constructs were tagged with C-terminal monoVenus followed by His_8. A small-scale expression test revealed several constructs (*Pv*PLP1 and *Pv*PLP2) could be successfully secreted and purified (**Figure 6.2**). Interestingly, the appropriate domain boundaries of constructs appeared to be very important, and even several amino acid residues difference had a significant impact on protein secretion efficiency (**Figure 6.2B**). Crystallisation experiments were carried out with *Pv*PLP1 and *Pv*PLP2 constructs and resulted in no crystals.

6.2.4 Summary of ApiPLPs expression and purification

Expression and purification of perforin-like proteins (PLPs) has been one of the major obstacles to study proteins of this family. We screened PLP expression in different expression systems and the result revealed that these proteins could be successfully expressed and secreted using mammalian expression system. The protocol established here could apply to the expression of other PLPs. Indeed, protein expression screening has shown that *Pv*PLP1 and *Pv*PLP2 could be successfully secreted and further study of these proteins will provide better understanding of this family of proteins, and in turn give more information in expressing these proteins.

It is a mystery that some PLPs (*Pb*PLP3 and *Pv*PLP3), with more that 30% sequence identity to *Pv*PLP1 could not be secreted in mammalian expression system.

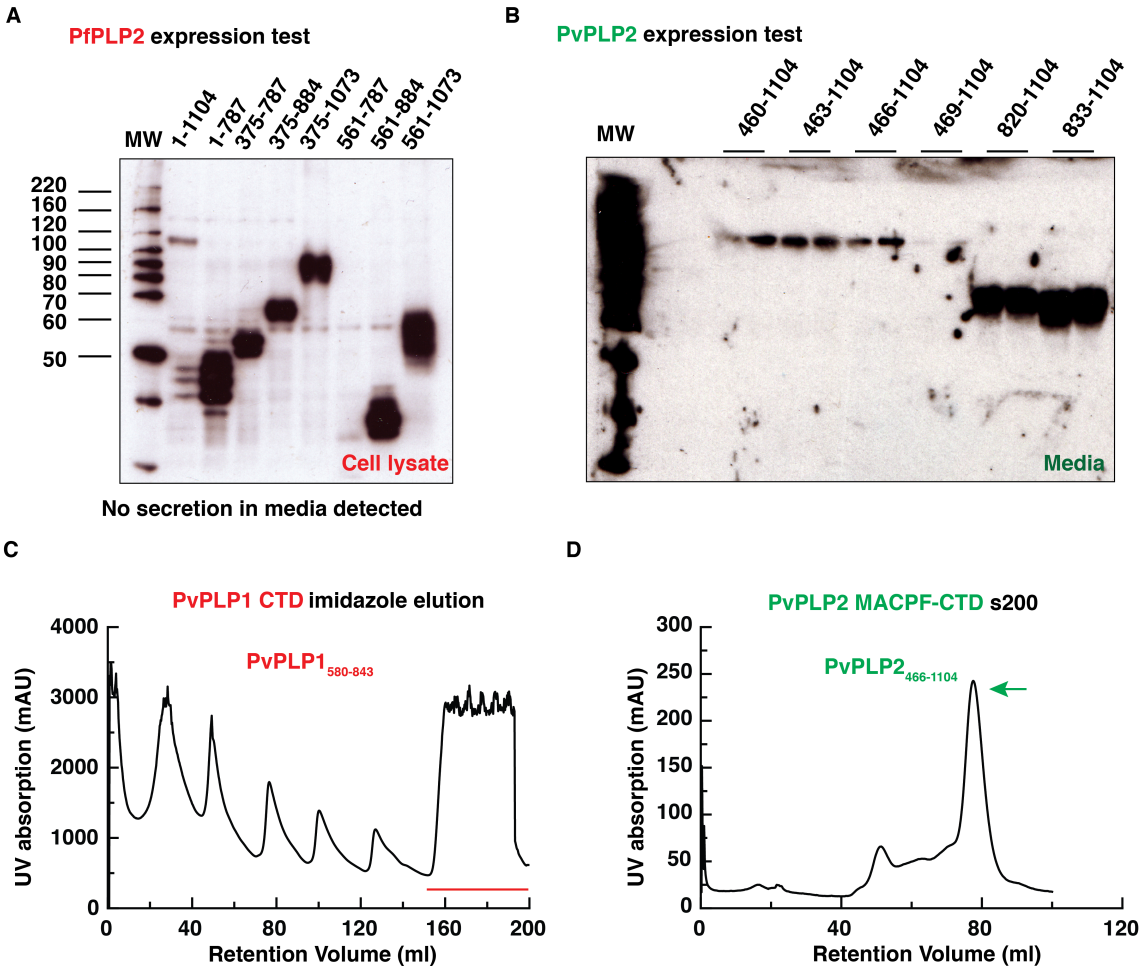

Figure 6.2: Expression test of *Pb*PLP and *Pf*PLP constructs in a mammalian secretion system. (A) Small-scale expression trial of *Pf*PLP2 constructs. Constructs encoding different regions of *Pf*PLP2 were cloned into pHlsec vector for the mammalian secretion pathway. These constructs showed high levels of expression but all failed to secret into media. The western blot of the media fraction is not shown here since it only contains the signal of the protein standards ladder. (B) Small-scale expression test of *Pv*PLP2 in mammalian HEK 293T cells. Constructs encoding MACPF and C-terminal domains of *Pv*PLP2 with 3 amino acid residues truncated from the MACPF domain were screened. *Pv*PLP2 (residues 466 - 1104) is the minimal construct for successful secretion. (C) Elution of *Pv*PLP1 C-terminal domain (residues 580 - 843) from a His-trap column. The red bar indicates the position of eluted proteins and the total yield of this construct is about 100 mg. The other peaks before the main peak (red bar) are contamination proteins, which are eluted with stepwise increased concentration of imidazole buffer. (D) Size-exclusion chromatography profile of *Pv*PLP2 MACPF-CTD (residues 466 - 1104) construct purification. The green arrow points to the peak corresponding to the monomeric species of *Pv*PLP2 and the peaks before the major target protein peak are contaminant protein co-eluted from imidazole buffer elution in the previous Histrap purification step.

Particularly in the cases of *Pf*PLP1 and *Pf*PLP2, they share more than 75% amino acid sequence identity to *Pv*PLP1 and *Pv*PLP2, respectively, but no secretion was detected for *Pf*PLPs even a codon-optimised (humanised) genes was used. Codon-usage bias has been observed in almost all genomes and regulates protein translation efficiency, accuracy and correct folding. For example, non-optimal codon usage across the entire open reading frame was shown to affect the expression, structure and function of circadian clock protein FRQ (Zhou et al., 2013). In practice for heterologous protein expression, harmonizing the codon usage frequencies of the *Plasmodium falciparum* genes with those of *E. coli* has been shown to greatly enhance their expression level (Angov et al., 2008). It is not clear whether the humanized genes we synthesised via Geneart follow the same principles of codon-usage frequency harmonization. We are currently looking into this in more depth and if a pattern can be identified further protein expression trials using codon-harmonized genes could be a potential approach for the expression of *Pf*PLPs.

6.3 Crystallisation of *Tg*PLP1 MACPF domain and C-terminal domain

6.3.1 Initial crystallisation trials of *Tg*PLP1$_{462\text{-}1072}$ (MACPF domain and C-terminal domain)

Initial crystallisation trials were carried out using a larger construct *Tg*PLP1$_{462\text{-}1072}$ covering the conserved MACPF domain and C-terminal domain conserved together among apicomplexan species. Protein glycosylation was considered to be one of the major issues in crystallisation due to the surface flexibility and variability it introduces. To facilitate crystallisation, the protein was expressed, purified and processed according to several protocols, resulting in proteins with different glycosylation status (**Figure 6.3A** and **Figure 6.3B**). Compared with the complexed glycan chain from protein secreted from HEK 293T cell line, glycans from the HEK 293S cell were much simpler and more homogeneous. The type II glycosylation inhibitor, kifunensine, can also inhibit complex glycosylation and make it possible for Endo_F1 endoglycosidase to remove the glycan during protein purification, leaving only one sugar attached to an asparagine sidechain. *Tg*PLP1$_{462\text{-}1072}$ could be secreted into media with great efficiency and more than 10 mg of protein could be purified from a single batch of cells (3 L). Representative gel filtration profiles and SDS-PAGE gels of *Tg*PLP1$_{462\text{-}1072}$ are shown in **Figure 6.3A**. Crystallisation screening revealed plate-like crystals in 2-6 months of time, and a dataset in spacegroup P1 was collected, which was severely anisotropic (data not shown).

Molecular replacement with models of MACPF domains available were tried without success, so we pursued experimental phasing for this crystal. Unfortunately, these crystals could not be reproduced easily, usually taking 3 months to grow and the resulting crystals lacking isomorphism, and even appearing with different spacegroups (data not shown).

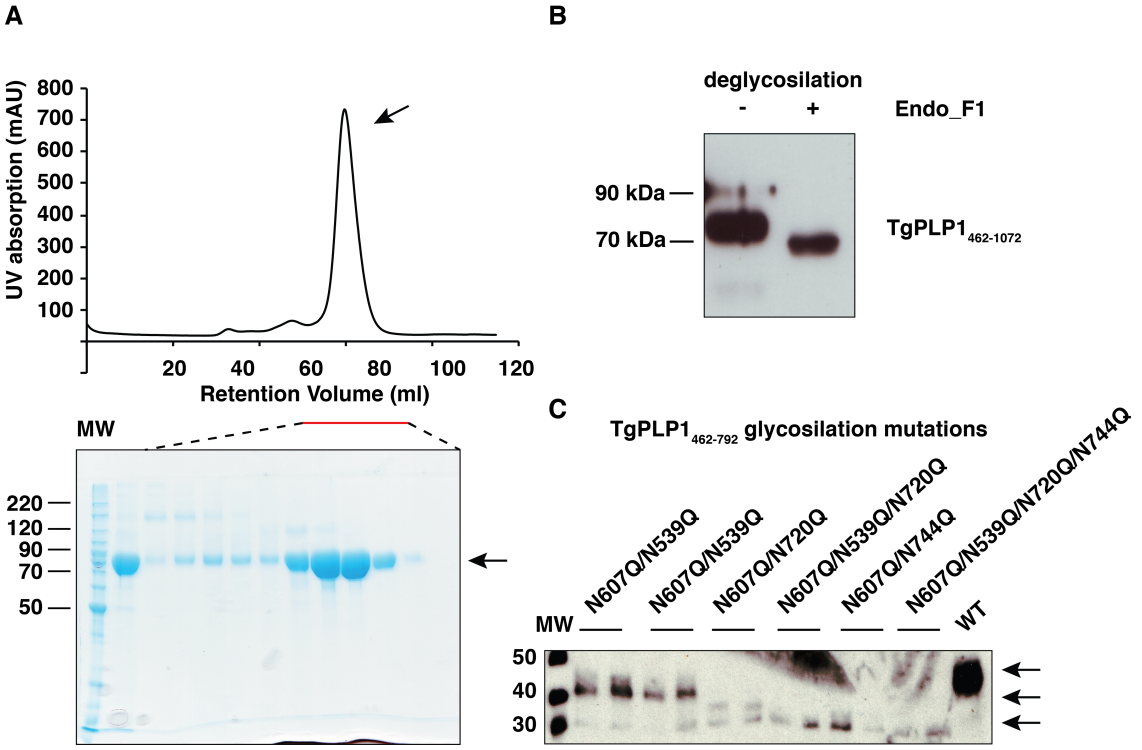

Figure 6.3: Purification of *Tg*PLP1$_{462\text{-}1072}$ (MACPF-CTD) and glycosylation sites determination. (A) Representative size-exclusion chromatography profile and SDS-PAGE gel of *Tg*PLP1 MACPF-CTD construct. (B) Western blot of *Tg*PLP1 MACPF-CTD subjected to deglycosylation treatment. Clear size shifting was observed before and after deglycosylation with Endo F1. (C) Identification of N-linked glycosylation sites in *Tg*PLP1 MACPF domain (residues 462 - 792) by site-directed mutagenesis. The potential N-linked glycosylation sites were mutated to glutamine with different combinations of mutations. N607 and N744 are the most likely glycosylated asparagine during secretion from mammalian cells. Three arrows indicate at least three different glycosylation statuses were identified.

To seek better crystals for structure determination, several potential approaches were tried. Surface entropy reduction (SER) by mutating the surface-exposed high-entropy residues (such as arginine, lysine and glutamine) to alanine has been shown to be an effective approach to facilitate crystallisation (Goldschmidt et al., 2007; Yates et al., 2015). *Tg*PLP1$_{462\text{-}1072}$ constructs carrying SER mutations were tested here and all could be expressed and secreted in HEK 293T cells with small variances in secretion level (**Figure 6.4**). The construct carrying 3 clusters of mutations was selected for further experiments and concentrated to about 40 mg/ml for crystallisation screening - but no crystal appeared. Another approach based on similar principle using *in vitro* lysine methylation to reduce the solubility of proteins was also tried using a protocol previously reported (Walter et al., 2006) without success.

Meanwhile, we suspected that proteins (*Tg*PLP1$_{462\text{-}1072}$) in crystallisation drop were degrading since it took so long for crystallisation to occur and the crystallisations result varied from batch to batch, although this protein had been shown to be very stable in solution (no significant degradation was observed after leaving the purified protein in 4 ℃ or 20 ℃ for 6 months, data not shown). Several small crystals from the same crystallisation conditions were collected and analysed by SDS-PAGE and this showed that the crystals were actually composed of a degradation fragment of *Tg*PLP1, instead of the full-length protein (**Figure 6.5**). Interestingly, proteins left in the same crystallisation drop remained largely intact, which indicated that this larger construct is not ideal for crystallisation, probably due to structural flexibility between MACPF domain and C-terminal domain. Mass spectrum fingerprinting showed that the fragment is mainly a C-terminal region of *Tg*PLP1 (**Figure 6.5**). Therefore, the C-terminal domain was prioritised for recloning and crystallisation trials, to be followed by the MACPF domain, as described in detail below.

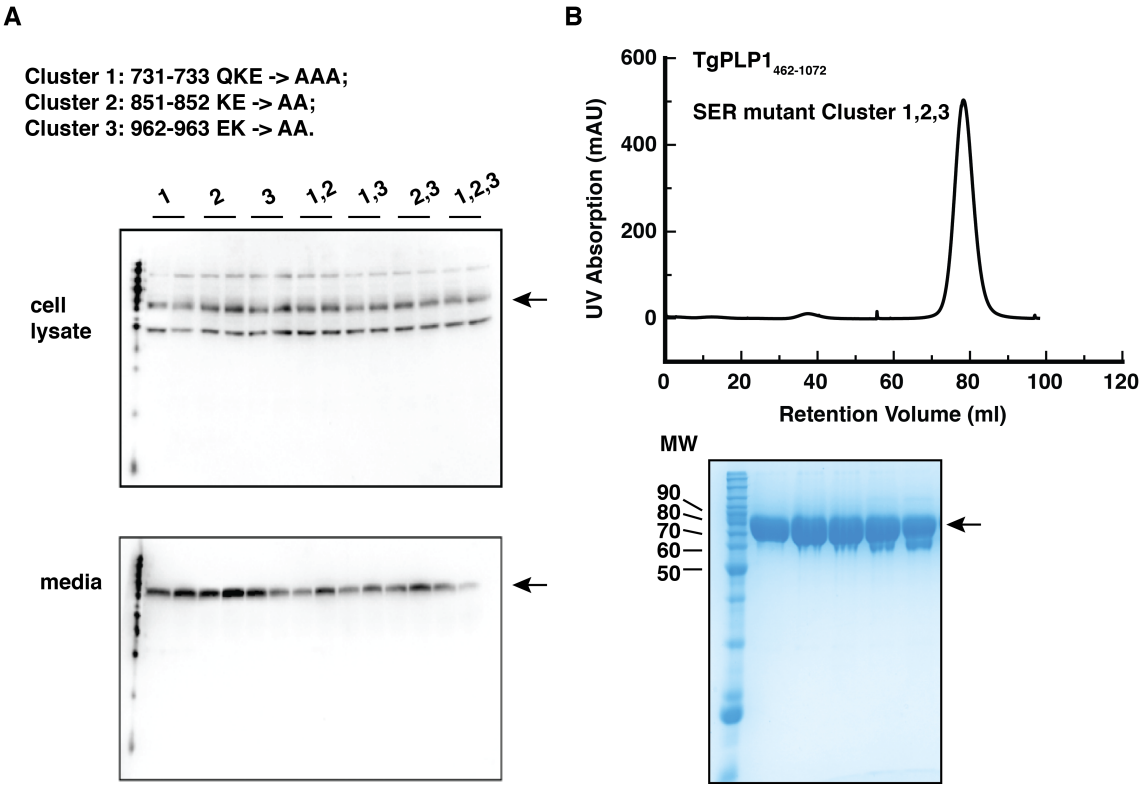

Figure 6.4: Surface entropy reduction mutant of *Tg*PLP1$_{462\text{-}1072}$ (MACPF-CTD) construct. (A) Small-scale expression test of *Tg*PLP1 MACPF-CTD construct introduced with different sets of mutations. The clusters of amino acid residues predicted to have high surface entropy were mutated into alanines. All constructs can be successfully expressed and secreted at a high-level. The other two bands in the cell lysate are proteins that interact with anti-pentaHis antibody non-specifically in the upper panel. (B) Representative size-exclusion chromatography profile and SDS-PAGE of *Tg*PLP1$_{462\text{-}1072}$ surface entropy reduction mutant (Cluster 1,2,3). The small amount of protein underneath the major band is the contaminant BSA.

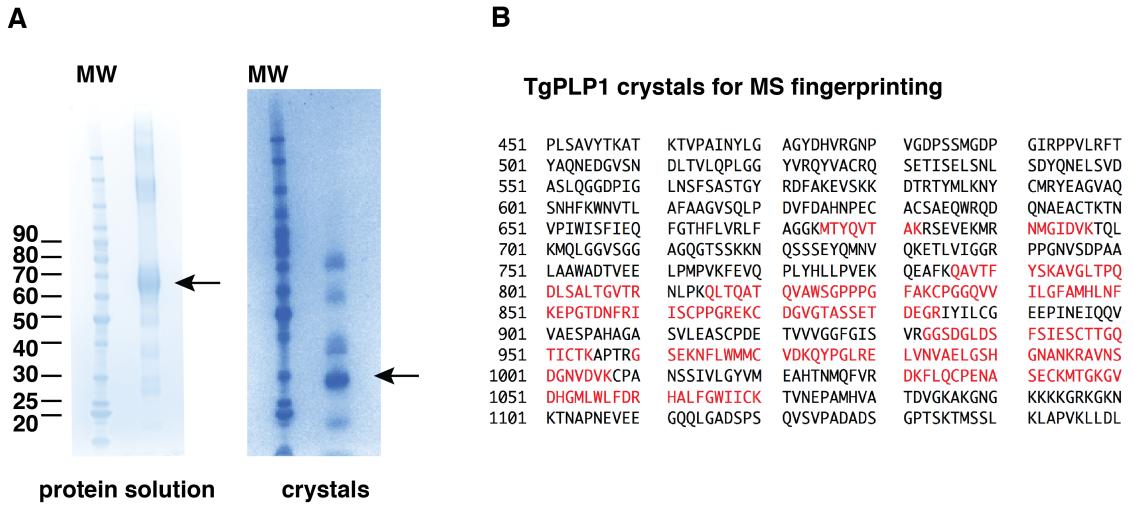

Figure 6.5: Characterisation of *Tg*PLP1$_{462\text{-}1072}$ crystals in SDS-PAGE. (A) SDS-PAGE analysis of *Tg*PLP1$_{462\text{-}1072}$ samples from crystallisation trays with the same crystallisation conditions: left, crystallisation drop without crystals; right, crystallisation drop with crystals. (B) Mass spectrum fingerprinting identifying the crystal contents. The red regions are successfully identified as mainly C-terminal region of *Tg*PLP1, although some fragments of the MACPF domain were also identified.

6.3.2 Crystallisation and structure determination of *Tg*PLP1 C-terminal domain

Expression, purification and crystallisation of *Tg*PLP1 C-terminal domain

Given the information on mass spectrum fingerprinting of the crystals, several constructs covering the C-terminal domain with different domain boundaries were designed and prepared. Small-scale expression in 293T cells revealed that construct comprising of residues 810 - 1072 was the minimal integral domain required for successful secretion (**Figure 6.6A**), which was used for further large-scale expression. The expression level of this construct is unexpectedly high, with about 80 mg of purified proteins from 3 L of media (**Figure 6.6B**). The proteins were concentrated to $\sim$ 30 mg/ml for crystallisation screening and crystals appeared overnight (**Figure 6.6C**) from several conditions.

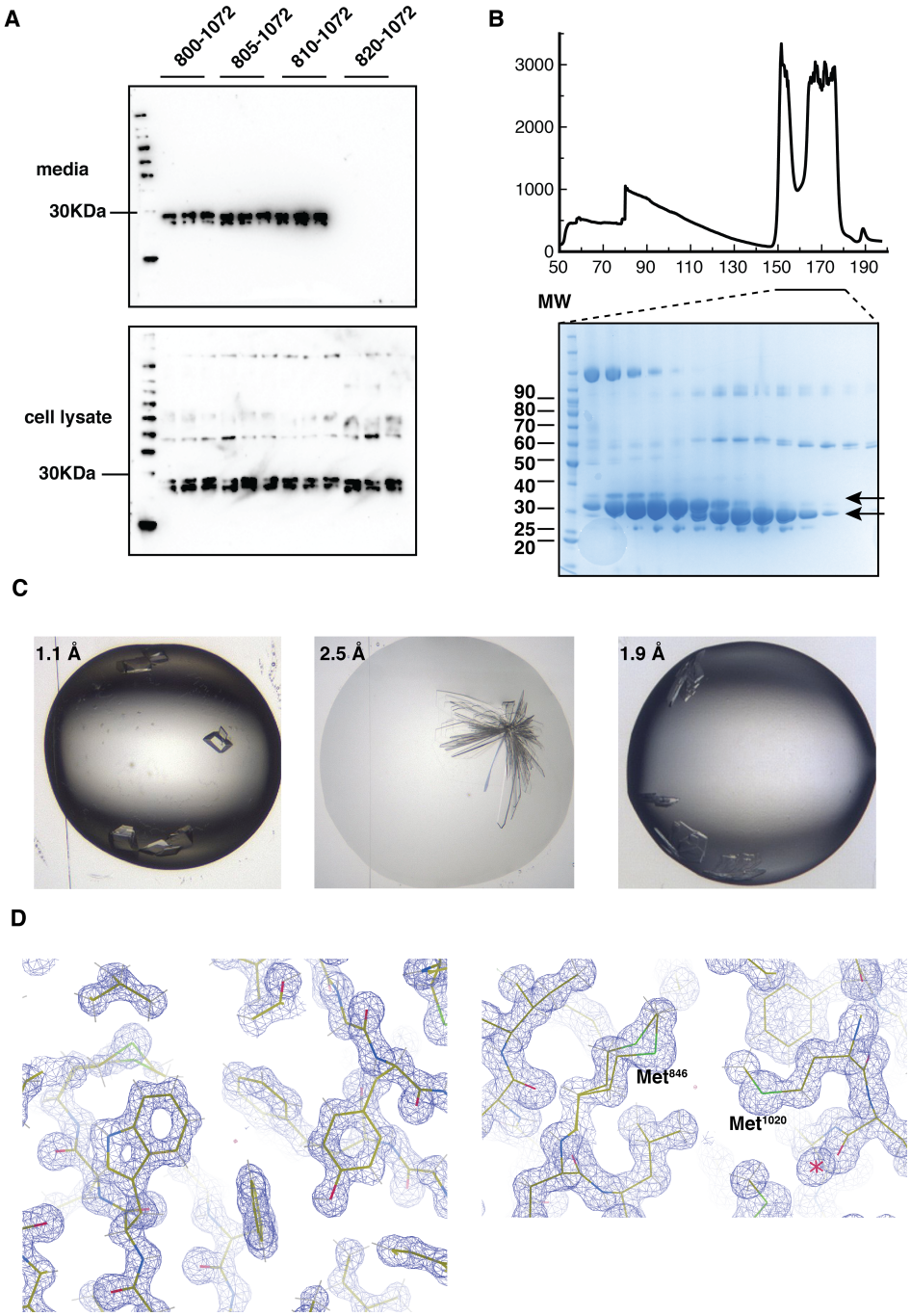

Figure 6.6: Expression, purification and crystallisation of *Tg*PLP1 C-terminal domain. (A) Expression test of *Tg*PLP1 C-terminal domain with different domain boundaries. A construct encoding the minimal C-terminal region (residues 810 - 1072) was used for further study. (B) Elution profile of *Tg*PLP1 C-terminal domain from a His-Trap column. Further purification using Superdex 75 26/600 was carried out, resulting in pure protein with a total yield of about 50 mg. (C) Gallery of *Tg*PLP1 CTD crystals from different conditions, with their diffraction resolution indicated top left. (D) Representative electron density map of the high-resolution (1.1 Å) structure: left, electron density of aromatic residues in the core of the domain; right, double conformation of some methionine residues (Met^{846}). The electron density was contoured at 2σ and 1.5σ, respectively.

Data collection, processing and structure determination

Diffraction data were collected on I03 and I24 in Diamond Light Source (Didcot, UK) (**Table 6.1**). High-resolution native data (1.1 Å) were obtained in spacegroup C121, alongside with datasets with moderate resolution (ranging from 1.9 Å to 3.0 Å) in other spacegroups from different crystallisation conditions. Experimental phasing was pursued here due to the absence of known structural homology to a previously-solved protein structure for molecular replacement. A Pt-soaked derivative dataset was obtained at 1.5 Å resolution. The heavy-atom derivatisation lead to the change of spacegroup from C121 to P1, with two copies of proteins in each asymmetric unit. The phasing was successfully solved by Pt-SAD and resulted in excellent electron density, which was then fed into *phenix.autobuild* for automatic model building. One single copy were then taken to solve the high-resolution data in spacegroup C121 at 1.1 Å. The resulting structure model was manually modified in coot and further refinement carried out in *phenix.refine*. Double conformation of side chains of some amino acid residues (e.g., methionine) were clearly observed in the electron density (**Figure 6.6D**). Individual anisotropic B-factors were refined for the model, resulting in a final R_{work} and R_{free} of 14.92% and 16.67%, respectively (**Table 6.2**). Structures in other spacegroups were determined by molecular replacement using the refined high-resolution model.

6.3.3 Crystallisation and structure determination of *Tg*PLP1 MACPF domain

Expression and purification of *Tg*PLP1 MACPF domain

Following the success in the C-terminal domain of *Tg*PLP1, constructs with every five residues truncated from the C-terminal of MACPF domain were cloned into the pHlsec vector with a C-terminal His_6 tag. All the constructs showed good expression and secretion level (**Figure 6.7**A) and the construct covering the MACPF domain

Table 6.1: Data collection of *Tg*PLP1 C-terminal

Data collection	Native_1	Derivative	Native_2	Native_3	Native_4	Native_5
Spacegroup	C121	P1	$P2_12_12_1$	P1	$P12_11$	C121
Cell dimensions						
a, b, c (Å)	100.76, 50.21, 52.28	51.01, 52.88, 56.89	73.20, 104.0, 107.04	72.25, 72.46, 97.06	77.7, 73.7, 101.23	221.35, 61.43, 42.23
α, β,γ (°)	90, 90.96, 90	89.62, 63.34, 89.97	90, 90, 90	100.44, 94.39, 118.93	90, 112.5, 90	90, 98.99, 90
Resolution (Å)	50.27 - 1.11	52.88 - 1.51	104 - 1.99	93.7 - 2.24	93.5 - 2.41	59.14 - 2.4
R_{merge}	0.065 (0.672)	0.052 (0.431)	0.238 (1.709)	0.106 (0.932)	0.224 (1.272)	0.154 - 1.759
R_{pim}						
I/σI	11.4 (1.5)	30.5 (3.0)	9.0 (1.6)	7.5 (1.2)	6.7 (1.3)	8.5 (1.0)
Completeness (%)	96.8 (76.3)	60.4 (3.7)	100.0 (100.0)	98.2 (97.3)	99.8 (99.8)	99.5 (99.3)
Redundancy	6.1 (4.1)	8.6 (4.6)	13.4 (13.3)	3.5 (3.5)	6.7 (6.0)	6.8 (7.0)
CC_half	0.999 (0.568)	0.999 (0.845)	0.996 (0.576)	0.993 (0.471)	0.984 (0.459)	0.996 (0.552)

[a]

[a] Values in parentheses are for highest-resolution shell.

Table 6.2: Refinement of *Tg*PLP1 C-terminal domain

Data collection	Native_1	Native_2	Native_3	Native_4	Native_5
Spacegroup	C121	P212121	P1	P1211	C121
Resolution	50.37 - 1.11	93.7 - 2.24	93.7 - 2.24	93.5 - 2.41	59.14 - 2.4
No. reflections	605707 (33178)	758547 (74957)	273074 27046	275548 (24732)	150555 (15313)
Unique reflection	99521 (7980)	56716 (5587)	78566 (33178)	40962 (4073)	22017 (2205)
R_{work}/R_{free} (%)	14.93 / 16.67	18.44 / 21.57	22.60 /26. 43	21.58 / 25.11	22.94 / 25.39
No. of atoms					
Protein	2039	6000	12090	8060	4021
Ion/Ligand	437	632	327	162	7
Average B-factors					
Protein	16.27	31.34	61.54	33.55	65
Solvent	33.93	29.26	50.81	31.89	55.4
RMS deviations					
Bond length (Å)	0.006	0.003	0.005	0.005	0.004
Bond angles (°)	1.08	0.78	1.03	1.02	0.87

[a]

[a] Values in parentheses are for highest-resolution shell.

and liner region (residues 462 - 805) was then used for large-scale protein expression and purification. Two potential glycosylation sites were mutated (**N720Q/N744Q**) in these constructs.

The MACPF domain was purified following a two-step purification protocol. Briefly, 3 L of media were passed through HisTrap HP column (GE Healthcare Life Sciences) and eluted with imidazole gradient (10 - 500 mM) in 20 mM Tris-HCl, pH 7.5, 500 mM NaCl (**Figure 6.7B**). The eluates were subsequently concentrated to about 2.5 ml and loaded into a size-exclusion chromatography column (Superdex 200 26/600) and a peak corresponding to monomeric species of *Tg*PLP1 MACPF domain was collected (**Figure 6.7B**) and concentrated to 35 mg/ml in 10 mM HEPES, pH 7.5, 150 mM NaCl. Then proteins were flash-frozen in liquid nitrogen and stored at -80 ℃ until further use.

Crystallisation and optimisation of *Tg*PLP1 MACPF domain

Initial screen for the *Tg*PLP1 MACPF domain resulted in some crystals grown in PEG conditions, however, these crystal only diffracted to about 3.4 Å despite several rounds of attempted optimisation. Truncating flexible loops in the proteins was shown to be an effective approach for some proteins which were resistant to crystallisation (Dong et al., 2007). Particularly in our case, structural information about the linker region between MACPF and C-terminal domain (residues 796 -809, in the crystal structure) is unknown. Therefore, *in situ* limited proteolysis were performed before crystallisation. Briefly, a mixture of *Tg*PLP1 MACPF domain (8 mg) and trypsin (5 μg, Prote-Ace, Hampton Research) were incubated on ice and rescreened against all the crystallisation conditions available in our lab. Crystals appeared overnight from several different conditions (**Figure 6.8A**) and one of them diffracted to 2.0 Å. Importantly, crystals from these conditions can be repeated following the same protocol, ensuring that a good number of crystals were prepared for phasing.

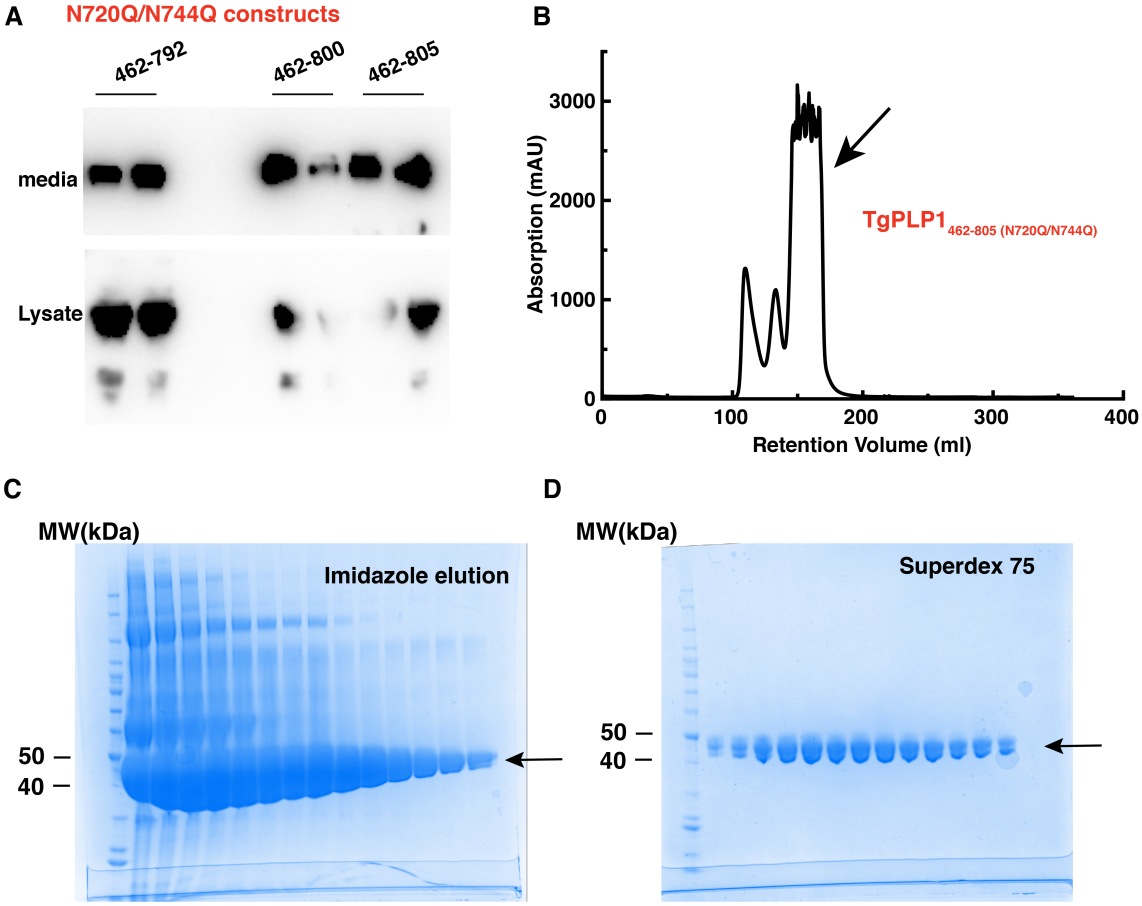

Figure 6.7: Expression screening, purification and crystallisation of *Tg*PLP1 MACPF domain. (A) Expression screening of *Tg*PLP1 MACPF domain. Small-scale expression screening showing that all the constructs could be expressed and secreted. All the constructs were prepared in pHLsec-His_6 vector using the same restriction cutting sites: AgeI and KpnI, which resulted protein coding sequence in frame with a C-terminal His_6 tag. The small-scale expression test were carried out in 293T cells seeded in 12-well plate. Two days after transfection, the cell media and cell lysates were harvested and subjected to western blot analysis using anti-pentaHis antibody. (B) Size-exclusion chromatography profile of *Tg*PLP1 MACPF domain (residues 462 - 805). The black arrow points to the monomeric species with a total yield of about 80 mg from 3 L of media. The other two peaks are contaminants from an earlier purification step. (C) SDS-PAGE gel showing the two-step purification of *Tg*PLP1 MACPF domain. The doublet bands of protein are probably due to heterogeneous glycosylation. The protein were not deglycosylated for crystallisation. There are still one potential glycosylation sites which was not mutated for this constructs

Data collection and processing

Diffraction data of intact MACPF domain crystals were collected on i24, Diamond Light Source (Didcot, UK) and best diffracted to 3.11 Å (**Table 6.3** for data collection). The dataset was achieved by combining data from 6 crystals. For crystals from the MACPF domain subject to limited proteolysis, the native datasets were collected on i24 and Platinum-soaked derivative datasets were recorded in i04 (**Table 6.3**). To prepare the Pt-soaked crystals, the K_2PtCl_4 powder was dissolved with a mixture of native protein and crystallisation conditions and incubated with crystals for 40min or 70min. Pt-soaked crystals still diffracted to 2.3 Å resolution and showed strong anomalous signal. All the diffraction data were indexed, integrated and scaled using the Xia2 pipeline (Winter, 2009) .

Structure determination of *Tg*PLP1 MACPF domain

Molecular replacement was initially tried with high-resolution native dataset, using MACPF domains of Perforin-1 (PDB: 3NSJ), Complement component 8α (C8α) and β (C8β) (PDB: 3OJY) and Astrotactin-2, and all failed to give a solution. For experimental phasing, the Pt-derivative dataset alone resulted in a solution with Pt-single anomalous dispersion, benefiting from reasonable resolution and a spacegroup with high symmetry, so that high-redundancy dataset could be achieved from a single crystal. The structure solution from the Pt-derivative dataset gave electron density of good quality (Correlation of local RMS density: 0.84) (**Figure 6.8B**), and side chains could be assigned into the electron density without ambiguity. With phase extension to 2.0 Å from the native dataset, the autobuild tracing of structure was successful and nearly 80% complete. There is one molecule per asymmetric unit and the model was completed in *Coot* (Emsley et al., 2004) and further refinement was carried out in *phenix.refine* (Adams et al., 2010), with a final Rwork and Rfree of 19% and 23.5% (**Table 6.4** for refinement statistics), respectively. The final-refined model

was then taken for molecular replacement to solve the structure of datasets from intact (untrypsinised) protein crystals. Extra electron density was resolved in these maps despite their low resolution limit. To facilitate model building, multiple dataset averaging with noncrystallographyic symmetry (NCS) was applied, and resulted in clear electron density in which the missing residues could be assigned.

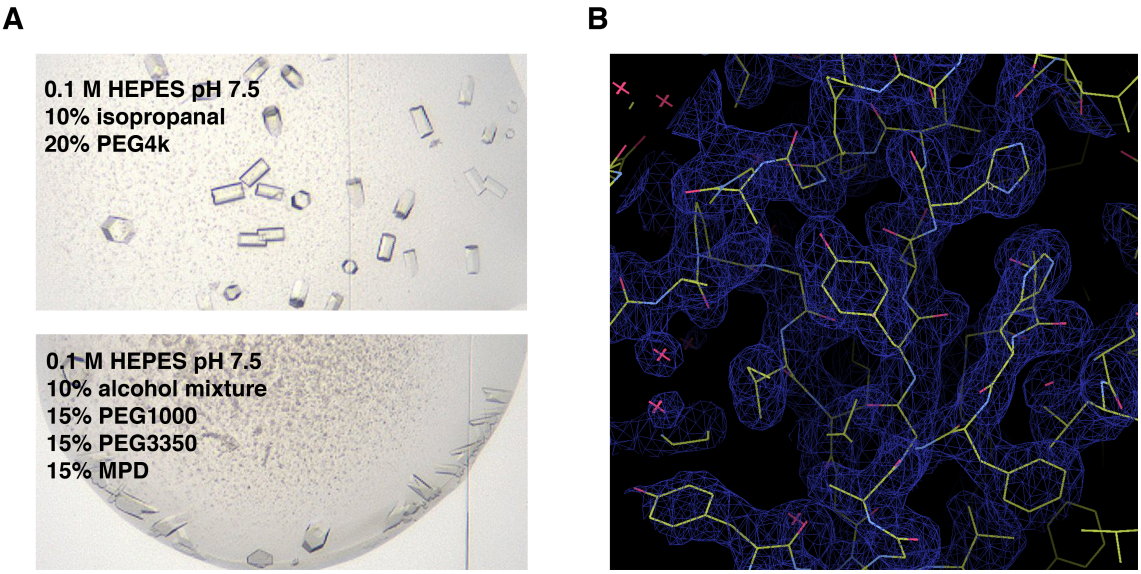

Figure 6.8: *Tg*PLP1 MACPF domain crystals subject to limited proteolysis and electron density from SAD phasing. (A) Gallery of *Tg*PLP1 MACPF crystals subject to limited proteolysis. The alcohol additive is necessary for crystallisation in these conditions.(B) Electron density of Pt-derivative single anomalous dispersion experiment in Phenix.autosol. A partial model was built in automatically. The electron density was contoured at 1σ.

The structure model solved from trypsin-treated protein crystals does not contain all the residues of the MACPF domain. Notably, no electron density was observed for the loop 2 regions of MACPF domain (residues 684 - 726), which forms two helices in the solution state, a highly conserved structural feature shared by all the MACPF/CDC superfamily proteins. In contrast, the electron density of the same region was clearly resolved in intact protein crystals, with two sets of helices. The fact that this region can be either trimmed off from the compact domain or exist as an ordered motif within the domain indicates its possession of a certain degrees of flexibility, which is essential for the dramatic conformational change (from helix to sheet) observed in members of this superfamily (as discussed in **Chapter 5**).

Table 6.3: Data collection and refinement of *Tg*PLP1 MACPF domain

Data collection	Native_1	Pt-derivative	Native_2 (Non-trypsin treated)	Native_3
Spacegroup	$P6_5$	$P6_5$	C 2 2 2_1	C121
Cell dimensions				
a, b, c (Å)	104.86, 104.86, 52.01	105.53, 105.53, 52.72	95.23, 206.58, 82.74	159.49185.74 104.19
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90 121.48 90
Resolution (Å)	90.81 - 2.03 (2.08 - 2.03)	52.76 -2.34 (2.40 - 2.34)	64.58 - 3.11 (3.16 - 3.11)	56.35 - 5.11 5.29 - 5.11
R_{merge}	0.171 (1.994)	0.115 (1.901)	0.278 (2.132)	0.131 (1.044)
R_{pim}	0.041 (0.464)	0.021 (0.341)	0.049 (0.596)	0.056 (0.473)
I/σI	13.4 (1.6)	24.0 (2.1)	12.9 (1.1)	8.2 (1.5)
Completeness (%)	100.0 (20.1)	99.3 (93.2)	100 (100)	99.7 (100)
Redundancy	13.4 (1.6)	39.6 (31.3)	31.9 (13.4)	6.7 (6.7)
CC_half	0.999 (0.514)	1.000 (0.444)	0.997 (0.582)	0.995 (0.702)
[redacted]				

[a]

[a] Values in parentheses are for highest-resolution shell.

Table 6.4: Refinement of *Tg*PLP1 MACPF domain

Data collection	Native_1	Native_2	Native_3
Spacegroup	$P6_5$	$C222_1$	C121
Resolution (Å)	98.37 - 2.03	64.58 - 3.11	56.35 - 5.11
No. of reflections	418830 (42406)	482085 (20145)	
Unique reflections	21264 (2104)	15081 (1489)	
R_{work}/R_{free} (%)	19.11 (23.25)	27.04 (29.02)	28.60 29.73
No. of atoms			
Macromolecules	2144	4552	12864
Ligands/Ion	14	22	84
Average B-factors			
Macromolecules	55.34	94.34	330.77
Ligand/Ion	51.27	173.69	363.24
RMS deviations			
Bond length (Å)	0.002	0.004	0.003
Bond angles (°)	0.58	0.79	0.65

[a]

[a] Values in parentheses are for highest-resolution shell.

Summary of crystallisation and structure determination

Crystallisation of both MACPF and C-terminal domain was very difficult and a major achievement of this study. We started with crystallisation screening with a construct containing both MACPF domain and C-terminal domain and found that only a proteolytic C-terminal fragment would crystallise. So the C-terminal domain alone was recloned, crystallised and its structure was solved at 1.1 Å. The MACPF domain with its glycosylation sites mutated was crystallised with *in situ* limited proteolysis and the structure was obtained at 2.0 Å. Both structures were determined with experimental phasing using platinum-soaked derivative single anomalous dispersion. The structures of the MACPF domain and C-terminal domain not only provide insights about their biological function, but also present a rationale for constructs design of other Apicomplexan perforin-like proteins.

CHAPTER 7

Structural and Biophysical Study of Kindlin-3 Pleckstrin Homology Domain

7.1 Summary

Kindlins are essential co-activators of integrins alongside talin. They consist, like the talin head, of a FERM domain comprising F0-F3 subdomains, but with a pleckstrin homology (PH) domain intersected into the F2 subdomain. The PH domain is required for kindlins to mediate integrin activation, at least partly by mediating their membrane association. As part of a structural and biophysical analysis of the role of PH domains in kindlin function, we present the crystal structure of the murine kindlin-3 PH domain and characterise its lipid binding using biophysical and computational approaches. Molecular dynamics (MD) simulations of the PH domain crystal structure in solution suggest that the PH domain retains its overall fold during the simulations but there is a degree of flexibility in the loops connecting the β-strands. Simulations of the PH domain with PtdInsP-containing bilayers reveal that the PH domain associates with PIP molecules mainly via the positively charged region within the β1-β2 loop and that it binds with somewhat higher affinity to PtdIns(3,4,5)P_3 molecules compared to PtdIns(4,5)P_2. Surface plasmon resonance (SPR) studies in which lipid headgroups are immobilised and the PH domain is the analyte in solution indicate affinities of 300μM for PtdIns(3,4,5)P_3 and 1mM for PtdIns(4,5)P_2. In contrast, SPR studies with immobilised PH domain as ligand and lipid nanodiscs as the analyte in solution show affinities of 0.40 μM for PtdIns(3,4,5)P_3 and no affinity for PtdIns(4,5)P_2 when the inositol phosphate constitutes 5% of the total lipids in nanodiscs ($\sim$5 molecules per nanodisc). If the inositol phosphate composition is reduced to 1% then no affinity can be measured even for a membrane containing PtdIns(3,4,5)P_3. Site-directed mutagenesis of two lysines within the β1-β2 loop abrogates binding to PtdIns(3,4,5)P_3. Comparison to interactions made by a canonical PtdIns(3,4,5)P_3-binding PH domain Grp1 indicates that its measured affinity is not similarly influenced by SPR experimental design. Taken together these data suggest a role for PtdIns(3,4,5)P_3 clustering in the binding of some types of PH domain and not others, highlighting the importance of allowing for lipid mobility

and clustering while assessing protein-membrane interactions biophysically.

Declaration

This project is in collaboration with Antreas C. Kalli, Fiona B. Naughton and Mark S. P. Sansom in Department of Biochemistry, University of Oxford. The molecular dynamic simulation studies were conducted by Sansom's group. Figures resulted from molecular dynamic simulation study were prepared by Sansom's group and clearly indicated in this chapter.

7.2 Introduction

Integrin-mediated adhesion between cells and the extracellular matrix is essential for the development of multicellular organisms. Integrins are heterodimeric (α and β) transmembrane receptors comprising a large extracellular domain for ligand binding (such as fibronectin, collagen and vitronectin, *etc.*) and short cytoplasmic tails (Harburger and Calderwood, 2009; Humphries et al., 2006) , and are responsible for bi-directional signal transduction across the cell membrane. The cytoplasmic tail of the β subunit, despite its short length, is critical for integrin activation and has been shown to interact with several regulatory proteins. Talin was the first essential intracellular activator of integrins to be identified, and it contains an amino-terminal FERM domain complemented with a long rod-like domain at its carboxy-terminal end (Calderwood et al., 2013; Critchley, 2009). The later identification of the kindlin family of proteins identified them as key co-activators of integrins, alongside talin. Talin and kindlin interact directly with the cytoplasmic tails of integrin β subunits, with kindlin binding to a membrane-distal NPxY motif and talin a membrane-proximal NPxY motif (Brahme and Calderwood, 2012; Moser et al., 2009; Shattil et al., 2010; Yates et al., 2012a). Biophysical data indicate that talin and kindlins are able to bind simultaneously to their respective NPxY sites (Yates et al., 2012a) while kindlins may be earlier recruited to integrins during activation than talin is (Bachir et al., 2014).

In humans, three isoforms of kindlin proteins, kindlins -1, -2 and -3, have been identified and shown to adopt tissue-specific patterns of expression. Kindlin-1 is expressed in the epithelial cells of the skin and gut, kindlin-2 is widely expressed but most notably in striated and smooth muscle, and kindlin-3 is expressed particularly in hematopoietic tissues but also in endothelial cells (Bialkowska et al., 2010; Ussar et al., 2006). Abnormal expression of, or mutations in, kindlins can result in severe disease. Kindlin-1 mutations are associated with Kindler syndrome (Siegel et al.,

2003) and kindlin-3 with leukocyte adhesion deficiency III (LAD-III) (Malinin et al., 2009; Moser et al., 2008; Schmidt et al., 2011), while mutant forms of kindlin-2 have not been identified in any live births, underscoring its ubiquitous expression and developmental significance. Kindler syndrome is a rare genodermatitis which shows transient skin blistering, skin atrophy and variegated hyper-pigmentation, and which can in rare cases lead to squamous cell carcinoma (Ussar et al., 2006). Enhanced wild-type kindlin-1 expression has been shown to be associated with metastasis of colon, bladder and breast cancers to the lungs, and with primary lung and colon carcinoma (Sin et al., 2011; Weinstein et al., 2003) . Kindlin-2 is more strongly linked to cancers of various kinds, having been described as playing a key role at the invading edge of mestheliomas (An et al., 2010), and as over-expressed in uterine leiomyomas (Kato et al., 2004). It is also shown to cause cisplatin resistance in prostate cancer (Gong et al., 2010) whereas its knockdown is synergistic with cyclopamine treatment in reducing the viability of prostate cancer cells in culture (Gao et al., 2013). Furthermore, kindlin-2 has also been found to promote stromal invasion and lymph node metastasis of gastric cancers (Shen et al., 2012) and to be expressed in bladder cancers although it is not normally expressed in the urothelium at all (Talaat et al., 2011). Kindlin-2 possesses a nuclear localization signal and is capable of bilocating between the plasma membrane and the cell nucleus. Its effect seems especially pathological in breast cancer, where it promotes anchorage-independent cell growth and the appearance of chromosomal abnormalities including genome region amplifications and deletions, and changes in chromosome copy number (Zhao et al., 2013).

A nuclear role has not been shown for kindlin-3 but it is over-expressed in chronic lymphocytic leukaemia (which led to its discovery) (Boyd et al., 2003) and when over-expressed in breast cancer may play a key angiogenetic role (Sossey-Alaoui et al., 2014). The importance of kindlin-3 in blood clot formation is however clear - for example mice carrying a QW>AA substitution in the integrin-binding F3 subdomain

have compromised clotting ability (Xu et al., 2014) because kindlin-3 activity is essential for platelet activation (Moser et al., 2008). As detailed above, nuclear localisation of kindlin-2 is oncogenic (Yu et al., 2012; 2013; Zhao et al., 2013) and another recent intriguing insight is that kindlin-2 is the preferred binding partner of β_1 integrins over kindlin-3 (Huet-Calderwood et al., 2014). This suggests one possible mechanism linking the two is that over-expression of kindlin-3 could dislodge kindlin-2 from an integrin-bound state, leading to nuclear relocalisation.

In any case, membrane binding by kindlins is thus not only an important factor in integrin activation but may also determine their cell signalling impact, and even the cell's fate. The membrane binding capacity of kindlins has been proved to be essential for integrin activation and is dependent on their PH domains (Calderwood et al., 2013; Hart et al., 2013; Liu et al., 2012; Yates et al., 2012b) and on a long loop within the F1 FERM subdomain which is conserved and similar to an equivalent loop in talin (Bouaouina et al., 2012). Deletion of the kindlin-3 PH domain eliminates its ability to participate in adhesion and migration of B cells mediated by the leukocyte integrin LFA-1(Mory et al., 2008).

The previous work on the kindlin PH domains focused on kindlin-1, where we solved its crystal structure to reveal an isoform-specific salt bridge occluding the canonical inositol phosphate binding site (Yates et al., 2012b). Molecular dynamics simulations indicated that the salt bridge is dynamic and led us to make a mutant lacking it which had altered inositol phosphate binding properties. We showed however that in either case the affinity for inositol phosphate ligands is relatively weak (10^{-4} M), although with a specific preference of PtdIns(3,4,5)P_3. We also showed that the apparent affinity was influenced by the buffer choice for an experiment: phosphate-based buffers seem to interfere with the apparent affinity measured, suggesting a lack of specificity in the PtdIns(3,4,5)P_3 interactions taking place (Yates et al., 2012a).

In this chapter we describe the crystal structure of the kindlin-3 PH domain together with molecular dynamics-based and biophysical characterisation of its inositol

phosphate lipid binding properties. We show that the kindlin-3 PH domain has a hydrogen bond-based occlusion to its PtdInsP binding cleft, and that like the kindlin-1 PH domain it binds surface-immobilised inositol phosphates with rather low (10^{-4} M) affinity. Using lipid nanodiscs as a model membrane system and molecular dynamics simulations to study the interaction of the kindlin-3 PH domain with different combinations of lipid species including inositol phosphate lipids, we show that lipid clustering is likely to be a significant factor in its binding of PtdIns(3,4,5)P_3 in bilayer membranes. On this basis we propose that a subset of PH domains is able to bind to multiple inositol phosphates simultaneously and so via an avidity effect have its interaction with target membranes strengthened. This would explain how an apparently low affinity for inositol phosphates can be biologically relevant to the localisation of kindlins to the plasma membrane and their subsequent activation of integrins.

7.3 Results

7.3.1 Overall structure of Kindlin-3 PH domain

The kindlin-3 PH domain exhibits an archetypal pleckstrin homology (PH) superfamily fold at its core (Lemmon et al., 2001), with a partly open seven-stranded β-barrel capped at one end by a C-terminal α-helix. Similar to that of kindlin-1 and 2 (Liu et al., 2012; Yates et al., 2012b), the PH domain of kindlin-3 contains an additional C-terminal amphipathic helix extension, which is not present in other PH domains (**Figure 7.1A**). A salt bridge in kindlin-1 PH domain (Arg^{380}-Glu^{416}) has been shown to occlude its canonical phosphoinositol binding site (Yates et al., 2012b), and in kindlin-3 PH domain this salt bridge is replaced by a hydrogen bond (Arg^{360}-Gln^{396}) in the equivalent position **(Figure 7.1A)**. As shown in Figure 1B, the kindlin-3 PH domain has a mixed surface charge distribution with a pronounced concentration of positive charge at the canonical inositol phosphate binding site.

The overall fold of the kindlin PH domain appears remarkably conserved. The kindlin-3 PH domain is closest in structure to the PH domain of kindlin-2 (RMSD = 0.69Å), with the major conformational difference being in the β1-β2 loop. The close relationship between the kindlin PH domain structures is reflected in their position within a structural phylogenetic tree we constructed for a set of PH domains of different kinds of function and (where relevant) with different lipid binding properties **(Figure 7.2)**. Of note, a previously-constructed structural phylogeny places the previously-determined kindlin-3 PH domain NMR structure in a different relationship to that of kindlins -1 and -2 to the crystal structure reported here (Yates et al., 2012a). This highlights the sensitivity of structural phylogenetic comparisons to the quality of the models used.

PH domains with canonical inositol phosphate binding sites typically exhibit a highly positively-charged core at the open end of the PH domain, whereby they interact with

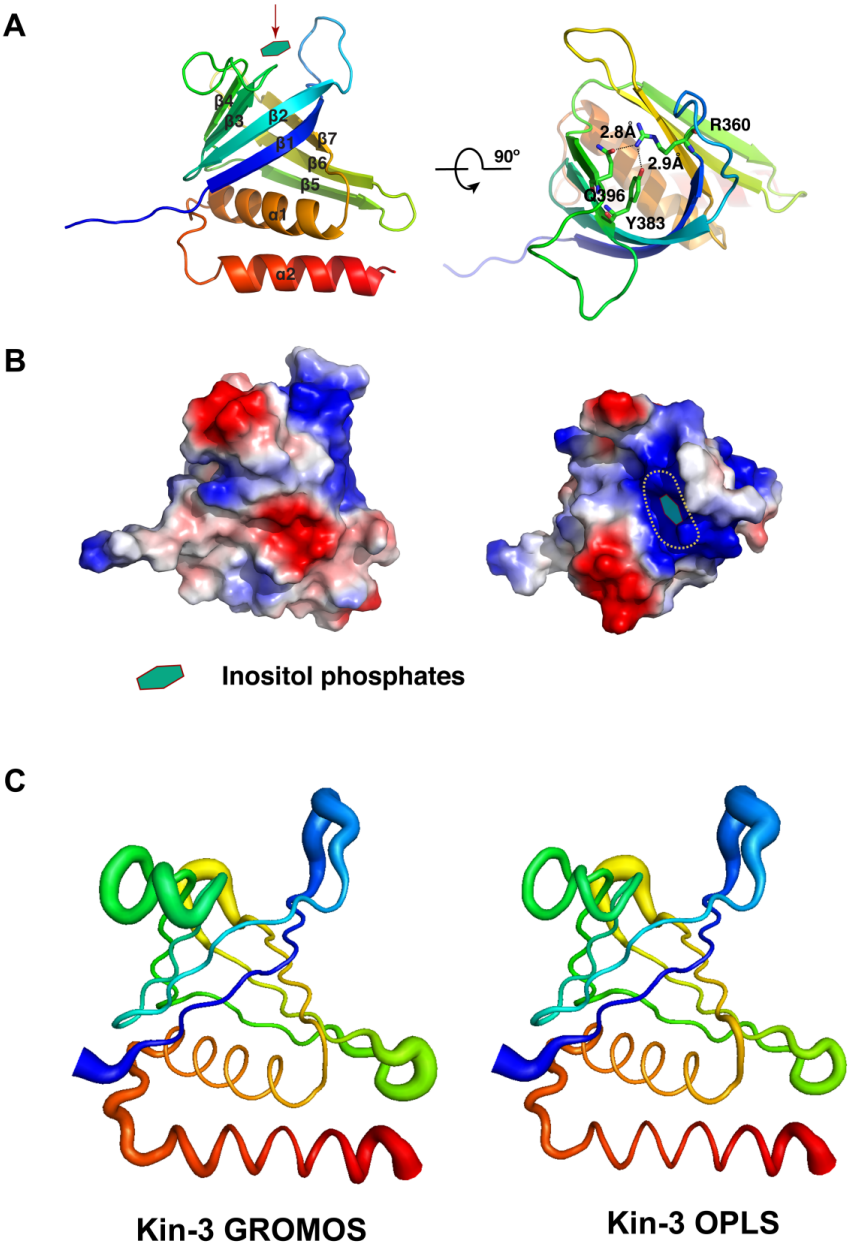

Figure 7.1: Overall structure of kindlin-3 PH domain. (A) Crystal structure of kindlin-3 PH domain. The protein is coloured with rainbow and the canonical inositol phosphate binding site is highlighted with arrow (left). The hydrogen bonds within the open end of PH domain are shown (right). (B) Surface electrostatic charge of kindlin-3 PH domain. The surface electrostatic was calculated in Pymol with vacuum surface electrostatics. The surface electrostatics calculated using Pymol and APBS show similar distribution of surface charge. The colour scheme here from red to blue matches the surface potential on solvent accessible calculated from APBS within in a range of -2.5 kT/e to 2.5 kT/e. (C) Average Root Mean Square Fluctuation (RMSF) of the kindlin-3 PH domain during the simulations in water (in the absence of a bilayer) for two different forcefields (i.e. GROMOS and OPLS). The magnitude of the average RMSF of each residue was mapped on the kindlin-3 PH domain crystal structure. **Figure (C) was prepared by Kalli AC., Naughton FB. and Sansom MSP.**

Table 7.1: Data collection statistics of kindlin-3 PH domain

Data collection	
Spacegroup	$P2_12_12$
Cell dimensions	
a, b, c (Å)	132.04, 36.19, 52,69
α, β, γ (°)	90, 90, 90
Resolution (Å)	31.73 - 2.23 (2.31 - 2.23)
R_{merge}	0.049 (0.432)
I/σI	23.7 (2.3)
Completeness (%)	96 (93)
Redundancy	3.1 (2.6)
$CC_{1/2}$	0.998 (0.771)

[a]

[a] Values in parentheses are for highest-resolution shell.

Table 7.2: Refinement statistics of kindlin-3 PH domain

Refinement	
Resolution (Å)	31.73 - 2.23 (2.31 - 2.23)
No. reflections	38032 (3014)
Unique reflections	12441 (1155)
R_{work} /R_{free} (%)	21.94 (27.07)
No. atoms	
Protein	2123
Ligand/ion	1
B-factors	
Protein	84.92
Ligand/ion	13.06
Water	61.66
R.m.s deviatioins	
Bond lengths (Å)	0.007
Bond angles (°)	1.11

[a]

[a] Values in parentheses are for highest-resolution shell.

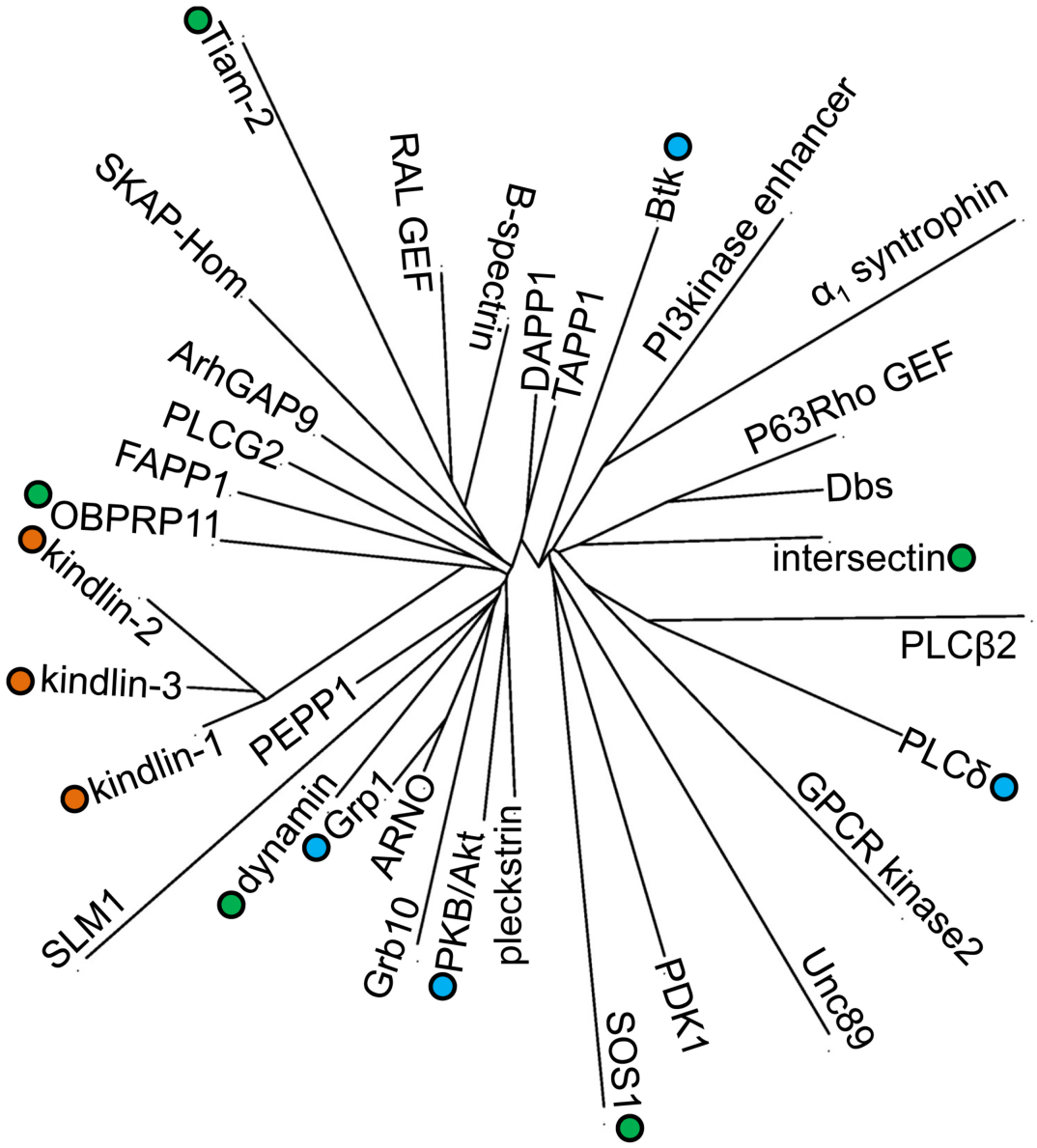

Figure 7.2: Phylogenic tree of representative PH domains from crystallography study. Some PH domains are highlighted: orange, kindlins; blue, high-affinity specific PtdInsP binders; green, low affinity or promiscuous PtdInsP binders. The following proteins are used to generate the phylogenic tree: Btk (1B55); PI3Kinase enhancer (2RLO); α1 syntrophin (2ADZ); P63Rho GEF (2RGN); Dbs (1KZ7); intersectin (1KI1); PLCβ2 (2FJU); PLCδ (1MAI); GPCR kinase 2 (1BAK); Unc89 (1FHO); PDK1 (1W1H); SOS1 (1AWE); pleckstrin (2I5F); PKB/Akt (1UNQ); Grb10 (3HK0); ARNO (1U27); Grp1 (1FGY); dynamin (3ZVR); SLM1 (3NSU); PEPP1 (1UPQ); kindlin-1(4BBK); kindlin-2 (4F7H); kindlin-3 (PDB code); OBPRP1 (2D9X); FAPP1 (3RCP); PLCG2 (2W2W); ArhGAP9 (2P0D); SKAP-Hom (1U5E); Tiam-2 (3A8P); RAL GEF (2DTC); B-spectrin (1BTN). The structural phylogeny calculation was carried out in the similar way to that described in Chapter 2.

their ligand, as seen in PH domains of Grp1 (Lai et al., 2013) and PLCδ1 (Ferguson et al., 1995). Non-canonical PH domains, such as Slm1(Anand et al., 2012), Tiam1 and ArhGAP9 (Ceccarelli et al., 2007), lack the same kind of positive

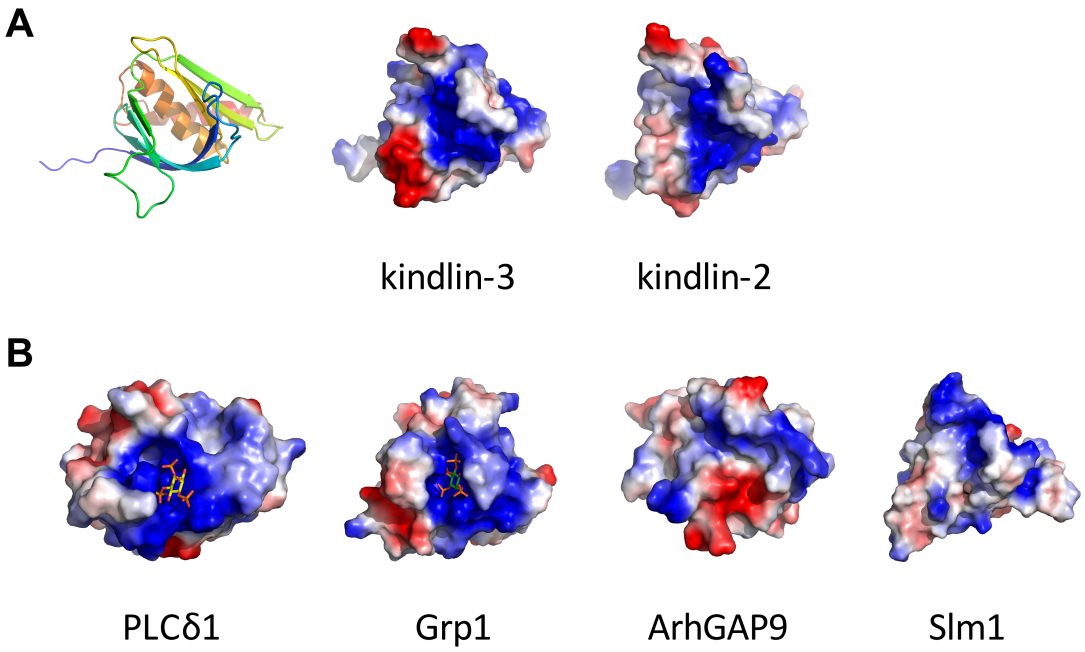

Figure 7.3: Comparison of the canonical and non-canonical inositol phosphate binding sites of PH domain. (A) Surface electrostatic charge distribution of kindlin-2 and kindlin-3 PH domains. Kindlin-2 and -3 are placed in the same orientation with open barrel on the top. (B) Surface electrostatic charge distribution of PH domains with canonical inositol phosphate binding pocket (PLCδ1-PH and Grp1-PH) and non-canonical binding site (ArhGAP9 and Slm1). These molecules are placed in the same orientation as in (A). PH domains with a canonical inositol phosphate binding site exhibit a typical positive-charged open barrel.

charge distribution, and instead have their inositol phosphate binding site on the side of the barrel mainly between the β1-β2 loop and β5-β6 loop **(Figure 7.3)**. The PH domains of kindlin-3 and kindlin-2 have similar surface charge distribution to the PH domains with canonical inositol phosphate binding sites (**Figure 7.1B** and **Figure 7.3A**), indicating that inositol phosphate might interact in a canonical way through the β1-β2 loop. The previously-reported crystal structure of the kindlin-1 PH domain is not as complete as either of the available crystallographic kindlin-2 or kindlin-3 models, lacking seven residues within the β1-β2 loop; its surface charge characteristics

however seem to be similar to those of the other kindlin PH domains.

Molecular dynamics simulation of the kindlin-3 PH crystal structure in solution (i.e. in the absence of any lipid bilayer) with two different force-fields showed that the overall fold of the PH domain is retained during the simulations. The seven-stranded β-barrel in the core of the PH domain is very stable, whereas the unstructured regions connecting the β-strands are somewhat more flexible **(Figure 7.1C)**. In particular, the β1-β2 loop is very flexible with both simulation force-fields, in good agreement with the crystal structure **(Figure 7.1A)**. At the end of the simulation the side chains of the positive residues on the β1-β2 loop remain exposed towards the open end of the barrel, presumably providing a site for inositol phosphate binding. With the exception of one simulation, the Arg^{360}-Gln^{396} hydrogen bond identified in the crystal structure was generally preserved during the simulations although it breaks transiently before reforming **(Figure 7.4A** and **7.4B)**.

7.3.2 Molecular dynamics simulations reveal the membrane-bound state of kindlin-3 PH domain and its interaction with PIP lipids

PH domains have been shown to bind phosphatidyl inositol phosphates (PtdInsPs) with different degrees of specificity (Moravcevic et al., 2012) and the PH domains of the kindlins have a limited affinity for them (Yates et al., 2012a), although the membrane interaction mediated by the PH domain is essential for integrin activation (Yates et al., 2012a). Phosphatidylinositol lipids are present in membranes at different concentrations, which may lead to variable activity and roles in membrane structure and membrane recognition. PtdIns(4,5)P_2 (PIP_2) has been reported to be the most abundant phosphatidylinositol lipid in cells (van Meer et al., 2008), and can undergo phosphorylation to a PtdIns(3,4,5)P_3 (PIP_3) state; both PIP_2 and PIP_3 are found at the plasma membrane. Monophosphorylated inositide lipids are, by contrast, associated with compartments such as the Golgi and endoplasmic reticulum (van Meer et al., 2008).

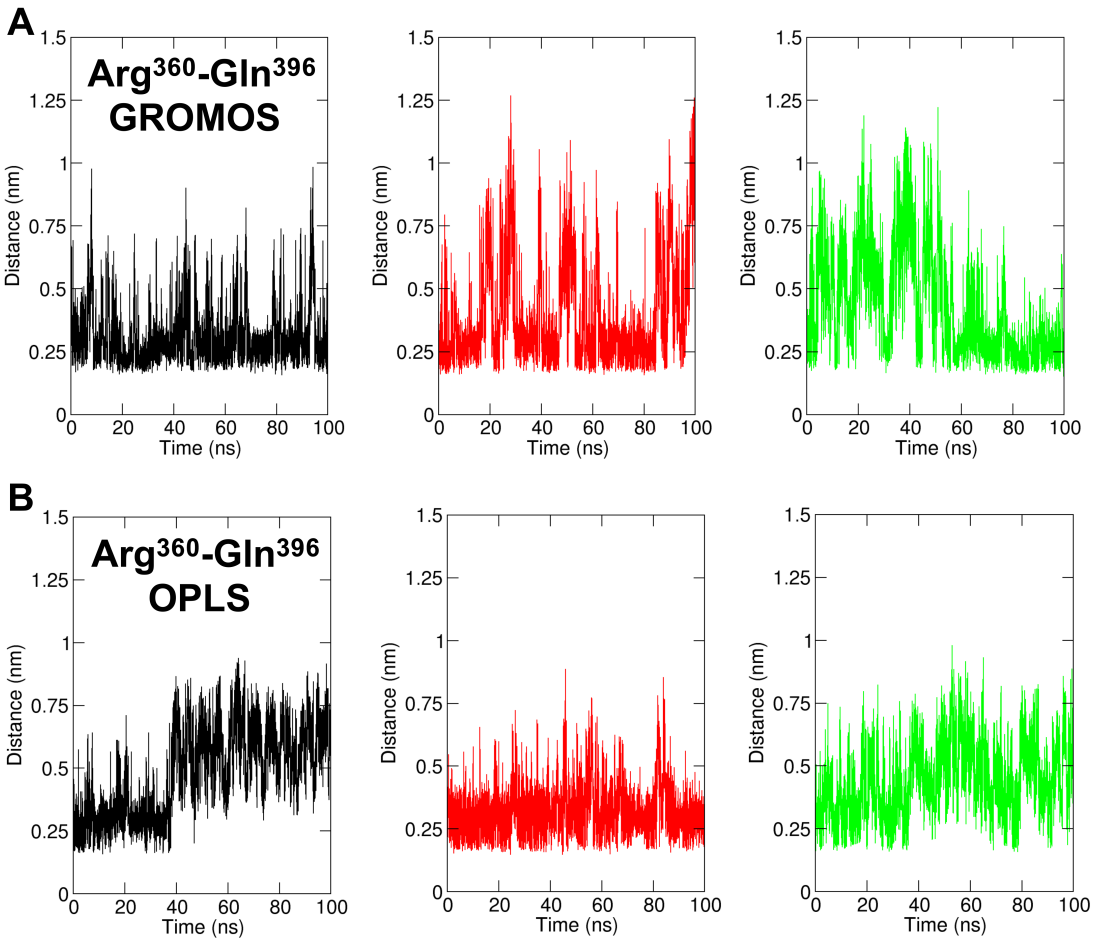

Figure 7.4: Hydrogen bond between Arg^{360} and Gln^{396} in simulation. (A,B) Minimum distance between residues Arg^{360} and Gln^{396} during the simulations with the kindlin-3 PH domain in solution. The distance is shown separately for the three repeat simulations with the GROMOS (A) and the OPLS (B) force-fields. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

MD simulations can be used to identify the molecular mechanism of peripheral membrane proteins interacting with model membranes incorporated with anionic e.g. PIP lipids (Kalli and Sansom, 2014). In particular, MD simulations have been used to study the interaction of a number of PH domains e.g. Dok7 PH and Grp1 PH with PIP lipids at the molecular level (study from Mark Samson's group, unpublished). A recent study has also demonstrated that MD simulations can be used to study the free energy of the interaction of Grp1 PH domain with PIP_2 and PIP_3 molecules (Naughton et al., 2016).

To examine the association of the new crystal structure of the kindlin-3 PH domain with model membranes, the PH domain was displaced away from a preformed bilayer containing 4 PIP_3 molecules in each leaflet (concentration of ∼1.5 % that mimics the *in vivo* concentration of PIP_3 lipids in the plasma membrane) and an ensemble of 20 simulations of 1.5 μs each was performed (**Figure 7.5A**). Calculation of the orientation of the PH domain relative to the bilayer (by calculating the R_{zz} component of its rotational matrix) suggests that the PH domain had a preferred orientation relative to the bilayer (**Figure 7.5B** and **7.5C**). In this orientation residues 360 to 370 (i.e. the positively charged loop between strands $\beta1/\beta2$) made the largest number of contacts with the PIP_3 molecules. Regions 410-415 and 475-478 (i.e. end of the second helix) also face towards the bilayer (**Figure 7.5**) and interact with PIP_3 molecules. This orientation is rather different from e.g. the canonical orientation of Grp1-PH domain relative to a bilayer (Lai et al., 2013; Lumb et al., 2011). Indeed, the helix at the end of the Grp1-PH domain was shown to be in a parallel orientation relative to the bilayer. We note, however, that in our ensemble we also observe secondary binding modes in which the kindlin-3 PH domain adopts an orientation that is similar to that of the Grp1 PH domain (**Figure 7.6E**). Atomistic simulations starting from the preferred orientation of the PH domain with the PIP_3-containing bilayer, reveal that the PH domain retains its orientation relative to the bilayer.

To examine whether the concentration of PIP3 in membrane bilayer affects the

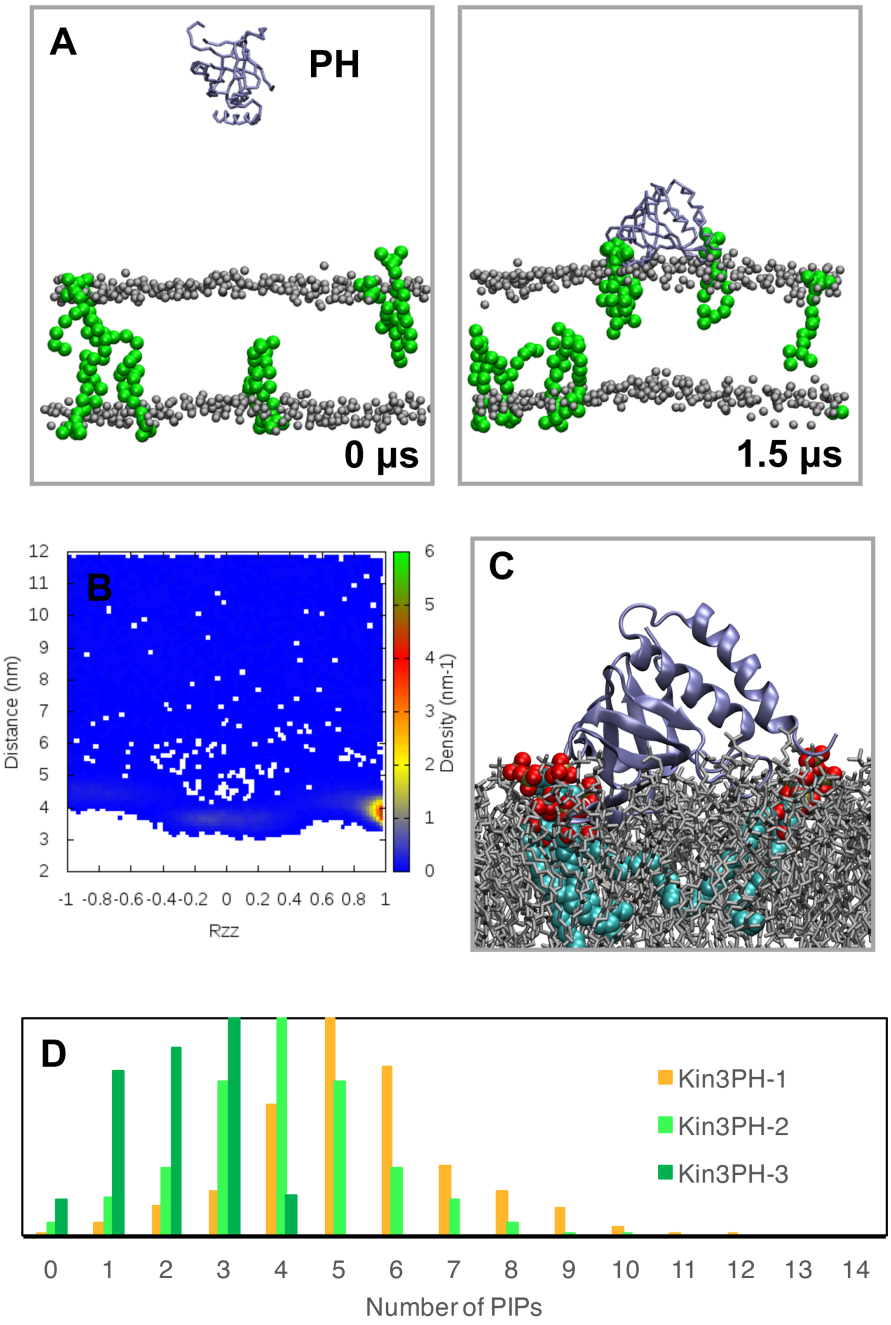

Figure 7.5: MD simulations of kindlin-3 PH interacting with PIP-incorporated lipids bilayer. (A) Snapshots at the beginning and at the end of simulation with 4 PIP_3 molecules in each leaflet (Kin3PH-3 in Table 2). PIP_3 molecules are shown in green, POPC and POPS phosphate atoms are shown in grey and the kindlin-3 PH domain is shown in iceblue.(B) Rotation of the kindlin-3 PH domain (shown as the R_{zz} component of its rotational matrix) relative to the distance between the centre of mass of the the protein and the centre of mass of the bilayer. (C) Final snapshot from the atomism simulation of one of the simulations in (A). (D) Number of PIP molecules interacting with the kindlin-3 PH domain. The number of PIP lipids is shown for Kin3PH-1, Kin3PH-2 and Kin3PH-3 simulations (**Table 7.3**) in orange, green and light green, respectively. For this calculation a cutoff distance of 1 nm was used to define the PIP lipids that were close to the protein. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

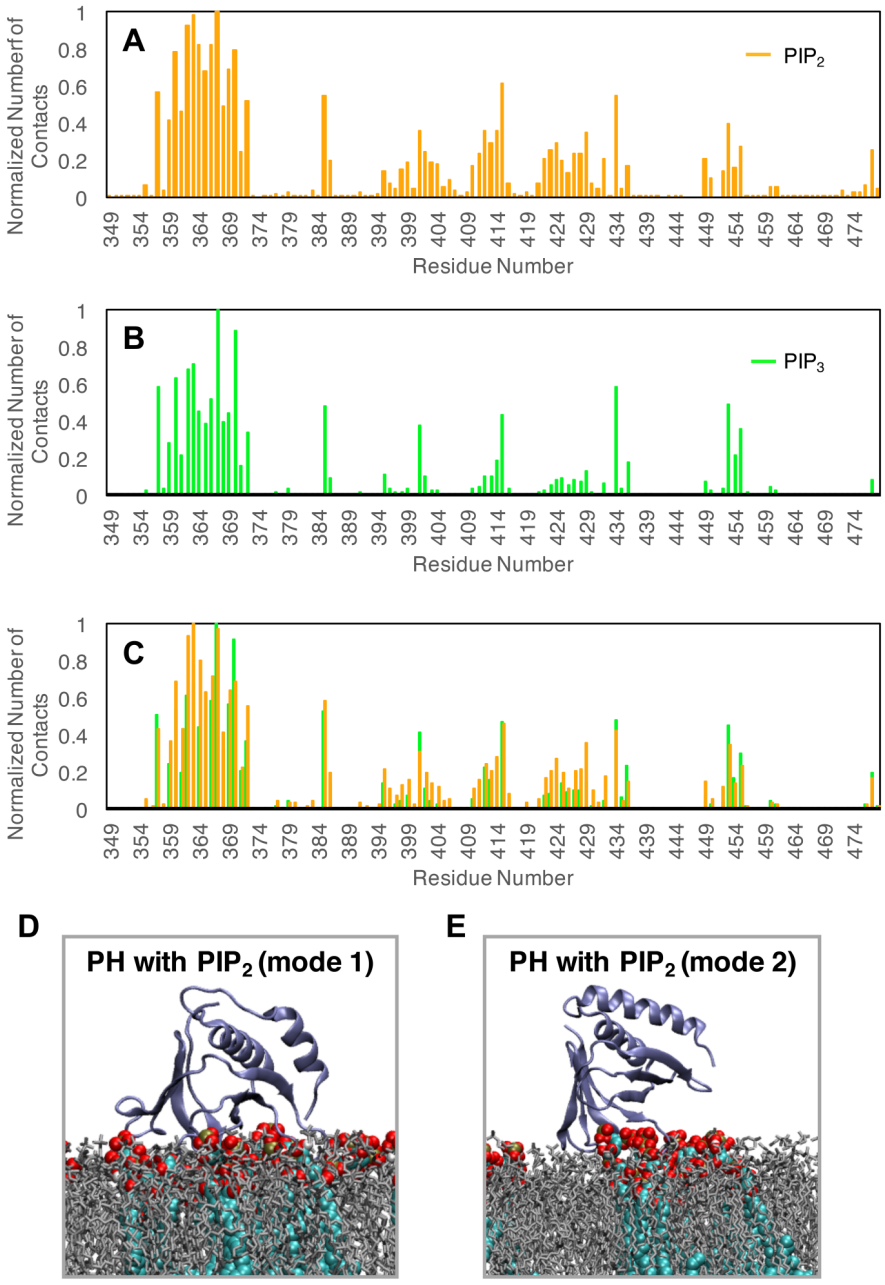

Figure 7.6: Contacts between Kindlin-3 PH domain and PIPs. (A,B,C,) Normalised average number of contacts of the interactions of the kindln-3 PH domain with PIP_2 or PIP_3 molecules. The contacts were calculated using a cutoff distance of 0.5 nm. The interactions with PIP_2 molecules (A) were calculated from the Kin3PH-1 simulation. The interactions of kindlin-3 with the PIP_3 molecules are shown as the average over all the simulations with PIP_3 in the bilayer (B; simulations Kin3PH-2 to 5) and the interactions with PIP_2 and PIP_3 molecules in (C) were calculated from Kin3PH-6 system. (D,E) Final snapshots from the atomistic simulations with the PIP_2-containing bilayer showing the two binding modes of kindlin-3 PH domain to the bilayer. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

association of the kindlin-3 PH domain to membrane, MD simulations were then carried out with various concentrations of PIP_3 lipids in the bilayer (5% PIP_3 or 2 PIP_3 in each leaflet or 1 PIP_3 in each leaflet; see **Table 7.3**). Despite that fact that when we reduce the number of PIP_3 molecules in the bilayer the kindlin-3 PH domain associates with the bilayer less frequently, in all cases it adopts a preferred orientation which is similar to the preferred orientation described above (**Figure 7.7**). The PH domain adopt the same orientation when we ran simulations with 5% PIP_2 lipids or with both PIP_3 and PIP_2 lipids in the bilayer (Kin3PH-1 and Kin3PH-6 in **Table 7.3**, respectively). Interestingly, atomistic simulations starting from four different snapshots of the kindlin-3 PH domain bound to PIP_2-containing bilayers at the preferred orientation (6 simulations overall; **Table 7.3**) resulted in three simulations in which the kindlin-3 PH domain retained its orientation relative to the bilayer and in three simulations in which the PH domain adopted the secondary binding mode at the end of the atomistic simulations (**Figure 7.6**).

Analysis of the contacts between the PIP lipids and the PH domain during the simulations suggested that residues R362, K363 and K367 and residues K363, K367 and R370 made the largest number of contacts with PIP_2 and PIP_3 lipids, respectively. Our analysis also revealed that the PH domain induces clustering of the PIP lipids when it binds to the bilayer (**Figure 7.5**). The clustering of PIP_2 lipids is somewhat higher compared to PIP_3 lipids. Note that the clustering of the PIP lipids occurs only in the leaflet in which the protein is bound. In the simulations with 5% of PIP_2 or 5% PIP_3 lipids in the bilayer most of the time we observe 4 PIP_3 and 5 PIP_2 lipids respectively in an annulus of 1 nm from the protein surface (**Figure 7.7**). Given that recent studies have suggested an important role for anionic lipids and PIPs in integrin activation (Orłowski et al., 2015; Schmidt et al., 2015), the change in the local lipid environment by the binding of the PH domain may be have some functional role.

Table 7.3: Summary of simulations

Simulations	Bilayer composition	Duration
coarse-grained		
Kin3PH-1	POPC:POPS:PIP_2 (75:20:5)	20 x 1.5 μs
Kin3PH-2	POPC:POPS:PIP_3 (75:20:5)	20 x 1.5 μs
Kin3PH-3	POPC:POPS (75:25) + 4 PIP_3 molecule per leaflet	20 x 1.5 μs
Kin3PH-4	POPC:POPS (75:25) + 2 PIP_3 molecule per leaflet	20 x 1.5 μs
Kin3PH-5	POPC:POPS (75:25) + 1 PIP_3 molecule per leaflet	20 x 1.5 μs
Kin3PH-6	POPC:POPS (75:20) + 5% PIP_2 and 4 PIP_3 molecule per leaflet	20 x 1.5 μs
Kin3PHmut-1	POPC:POPS (75:25) + 4 PIP_2 and PIP_3 molecule per leaflet - R363A, K367A mutation	20 x 1.5 μs
Atomistic		
Kin3PH _AT-1	No bilayer - GROMOS force-field	3 x 100 ns
Kin3PH _AT-2	No bilayer - OPLS force-field	3 x 100 ns
Kin3PH _AT-3	POPC:POPS:PIP_2 (75:20:5)	6 x 100 ns
Kin3PH _AT-4	POPC:POPS (75:25) + 4 PIP_3 molecule per leaflet	4 x 100 ns

[a]

[a] **Table prepared by Kalli AC., Naughton FB. and Sansom MSP.**

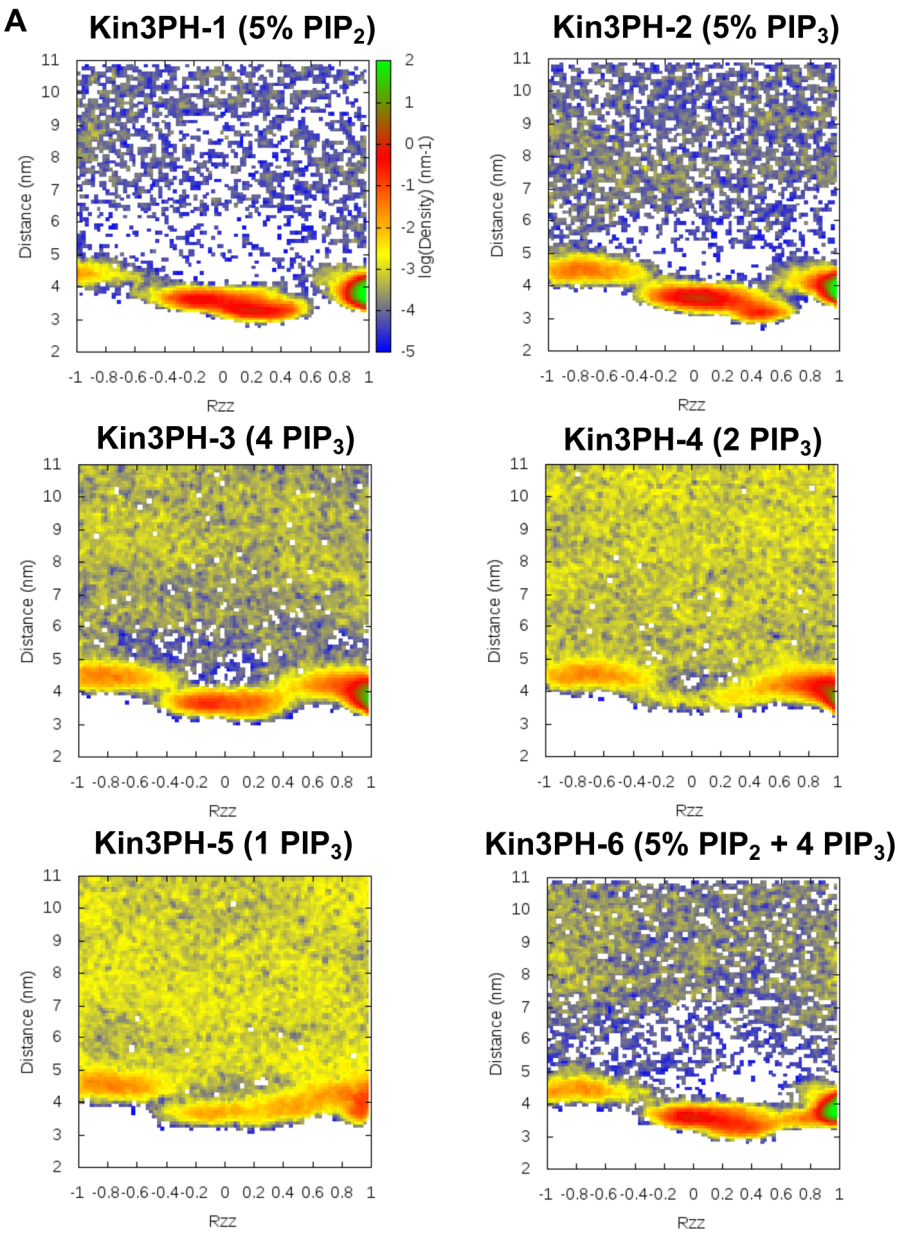

Figure 7.7: Orientation of kindlin-3 PH domain on the membrane bilayer. Orientation of the kindlin-3 PH domain relative to the bilayer shown as the R_{zz} component of the rotational matrix to align the kindlin-3 PH domain to a reference structure versus the distance between the centres of mass of the protein and the bilayer. The reference structure was in the same orientation as in **Figure 7.1C**. In the figure we show the logarithm of the density to aid the comparison between the different systems. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

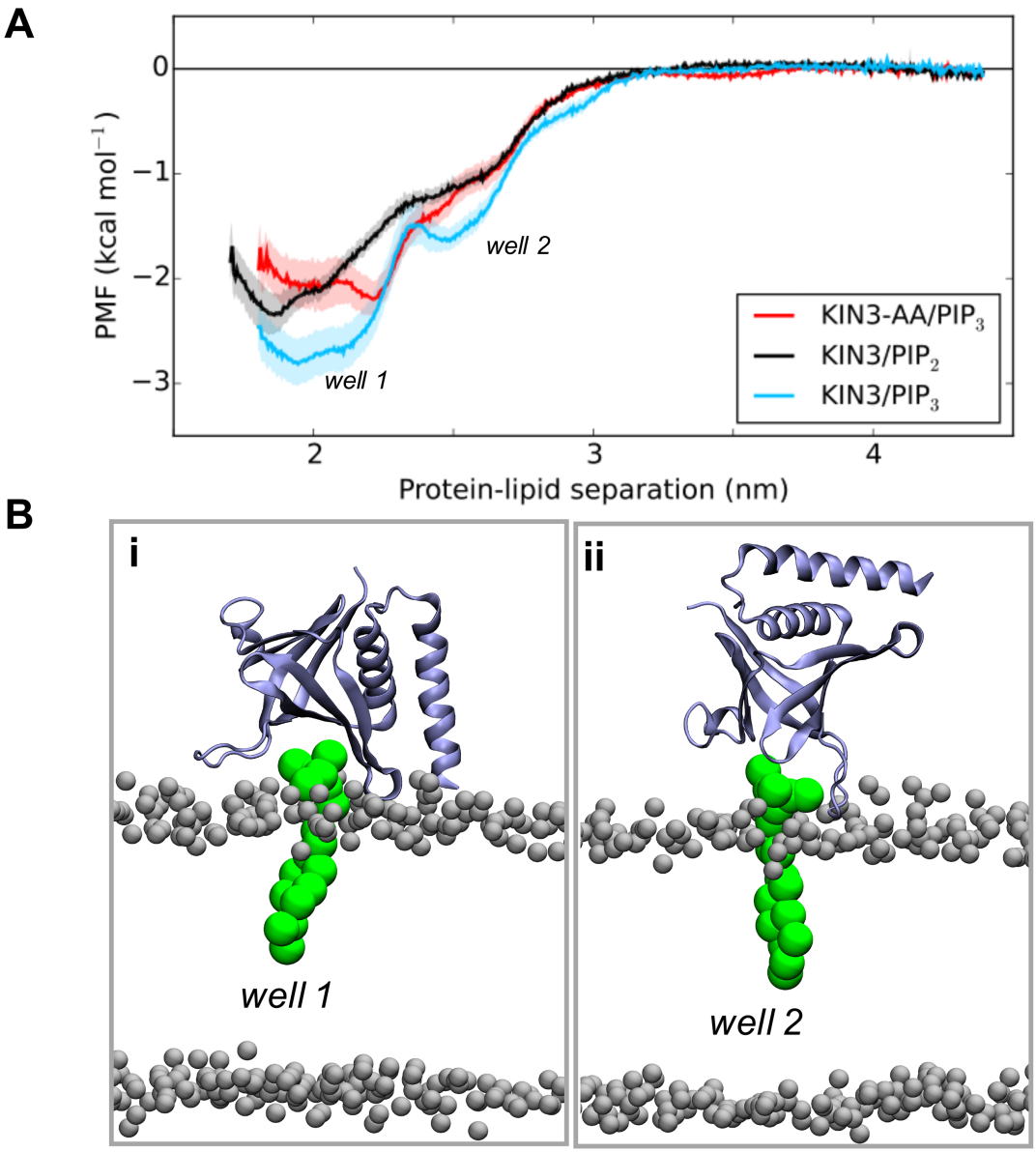

Figure 7.8: Potential of mean force (PMF) calculations of Kindlin-3 PH domain with PIPs. (A) PMF calculations for the binding of the wild type and the mutated form of kindlin-3 PH domain to PIP_2 (black) and PIP_3 (cyan and red) molecules. (B) Kindlin-3 PH/bilayer complexes corresponding to wells 1 and 2 in (A). The PIP_3 molecules are shown in green and the PH domain in iceblue. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

7.3.3 Free energy of binding of the kindlin-3 PH domain to PIP(4,5)P_2 and PIP(3,4,5)P_3

Our analysis above revealed the molecular mechanism of the association of the kindlin-3 PH domain with the membrane. To examine the strength of interaction of the PH domain with PIP_2 and PIP_3 molecules we performed potential of mean force (PMF) calculations using a protocol developed recently (Naughton et al., 2016). PMF profiles for PIP_3 and PIP_2 lipids have a global minimum with a well depth of ∼-3 kcal/mol for PIP_3 and a well of ∼-2.5 kcal/mol for PIP_2 lipids. This suggests that the protein associates somewhat more strongly with PIP_3 lipids (**Figure 7.8**). The first well in the profile of the PIP_3 molecules was followed by a second swallower well at a distance of ∼2.5 nm. Examination of the interactions of the kindlin-3 PH domain with the lipids showed that in the windows covering ∼1.5 to ∼2.3 nm protein-lipid separation, the protein is in a similar orientation to the preferred orientation (mode 1) described above in which the protein interacts with the bilayer mainly via the $\beta1/\beta2$ unstructured loop. Interestingly, in the second well the protein adopts an orientation that is similar to the secondary binding mode also observed in the simulations above. In contrast, the profiles for PIP_2 lipids and for the mutated form of kindlin 3-PH did not have a second well. We note that the depth of the well for the binding of the Grp1 PH domain to PIP lipids using the same method was -5.3 kcal/mol for PIP_3 and -3.8 kcal/mol for PIP_2 suggesting that the binding of kindlin-3 PH domain to the aforementioned lipids is weaker compared to Grp1 PH domain binding a single lipid (Naughton et al., 2016). Disruption of the PH/PIP interactions by mutating the K363 and K367 residues which were identified to form a large number of contacts with the PIP lipids resulted in a reduced binding to PIP_3 molecules by -1 kcal/mol (**Figure 7.8**).

7.3.4 SPR study of the kindlin-3 PH domain to multiple Phosphatidylinositol Phosphates

To verify the potential interaction between clustered phosphatidylinositol phosphates and the kindlin-3 PH domain observed in our MD simulations, we conducted surface plasmon resonance (SPR) experiments. Firstly, a biotinylated inositol phosphate was immobilised onto a streptavidin-derivatized sensor chip, as in previous studies (Yates et al., 2012). The use of streptavidin-mediated immobilization meant that the distance separating individual lipids would be larger than the size of the tetrameric streptavidin, and therefore than the PH domain itself meaning that one PH domain would be able to bind no more than one lipid head group simultaneously, imposing one-to-one binding in the interaction (**Figure 7.9A**). In agreement with the MD simulation results, the kindlin-3 PH domain displays a preference for PIP_3, though with a dissociation constant (K_D) of about 300 μM, in agreement with previous measurements (Yates et al., 2012a). The apparent binding capacity for PIP_2 was much lower (K_D = 1100 μM). Introduction of two point mutations (R363A, K367A) effectively abolished the binding capacity with both inositol phosphate moieties, indicating that the interaction is specific even if it is low affinity (**Figure 7.9**). We note that in the MD simulation with the mutated form of the kindlin-3 PH domain, the protein is able to bind to the bilayer in a similar orientation to the wild type however it quickly dissociates from the bilayer (**Figure 7.10**). A naturally-occurred IPRR insertion in β1-β2 loop of kindlin-3 has been reported to disable integrin activation and thus lymphocyte adhesion and migration (Hart et al., 2013). In agreement with results obtained by others using a pulldown assay with IP_3 and IP_4-coated beads, our SPR study *in vitro* revealed that the IPRR insertion directly disrupted the interaction between the PH domain and phosphoinositol moieties (**Figure 7.9B**), presumably because of the steric effect of an elongated in β1-β2 loop.

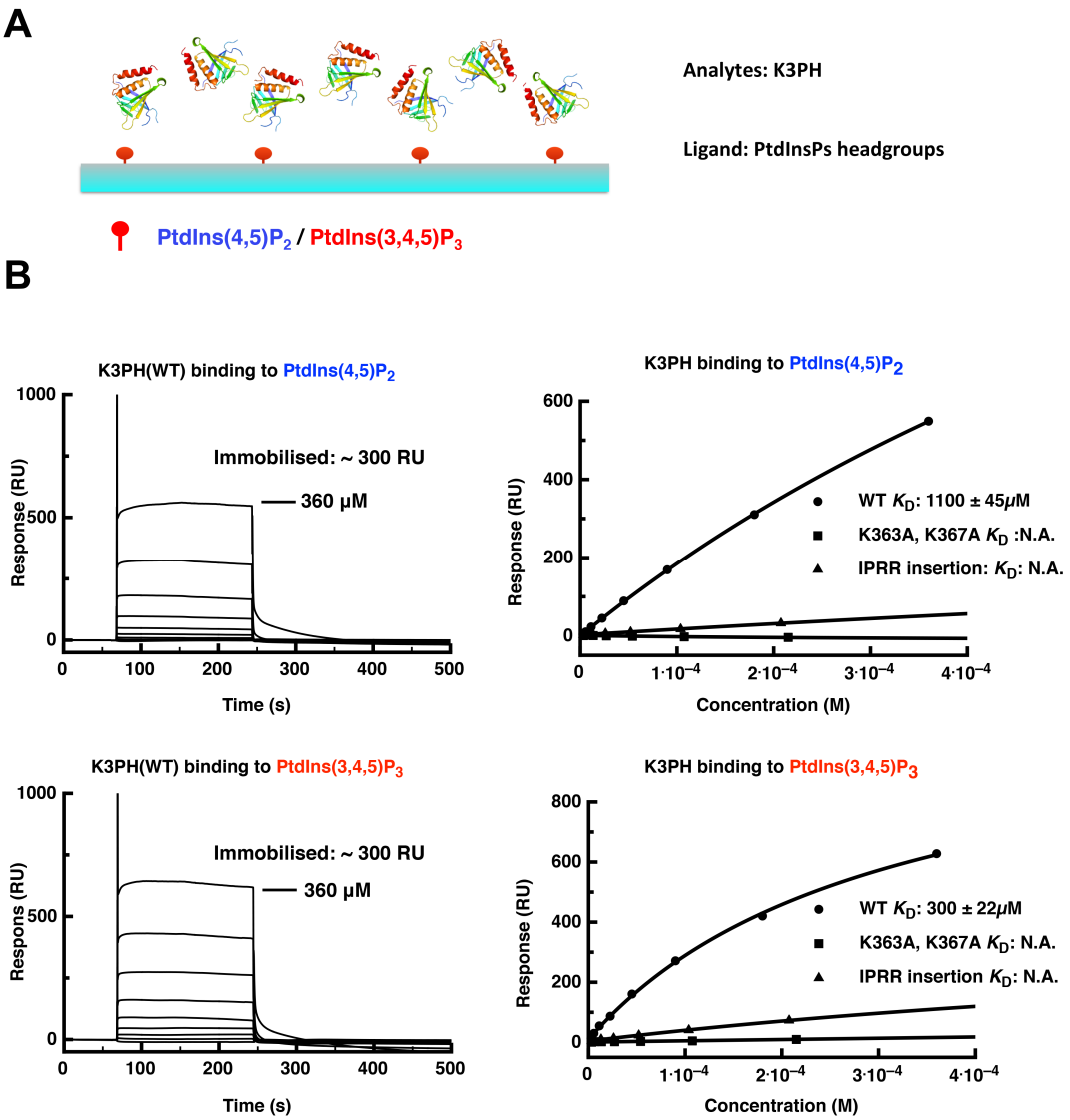

Figure 7.9: SPR study of kindlin-3 PH domain with immobilised PIP2 and PIP3 headgroups. (A) Schematic representation of SPR experiments with immobilised lipid head group as ligand and PH domain as analyte. The biotinlyated lipid head groups were bound to streptavidin derivertised CM5 chip. The CM5 chip was activated following the protocol of the manufacturer and streptavidin was covalently coupled to the activated surface until saturation. The indicated amount of lipid head groups were bound to the chip. (B) SPR measurements of kindlin-3 PH domain WT and mutants with immobilised PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3, respectively. On the left in each case we show the background-subtracted raw sensogram data for the wild type protein, while on the right we show binding saturation curves for wild type, K363A K367A and IPRR insertion mutants. The highest concentration used is given on each set of sensogram curves; subsequent binding profiles derive from a series of twofold dilutions.

The observation in MD simulation that multiple phosphoinositol molecules bind to one single PH domain (**Figure 7.5**) led us to attempt further verification of this mode of interaction experimentally. Nanodiscs have been shown to be an effective model membrane system due to their stability and capacity to mimic the membrane bilayer (Borch and Hamann, 2009; Wang et al., 2015) , including the fluidity of lipids though within a limited bilayer area. To study the effect of phosphatidylinositol lipid concentrations on their binding by PH domains, we performed SPR experiments in an inverted setup. The biotinylated PH domain was immobilised onto the sensor chip and the nanodiscs were prepared as the analyte (**Figure 7.12A**).

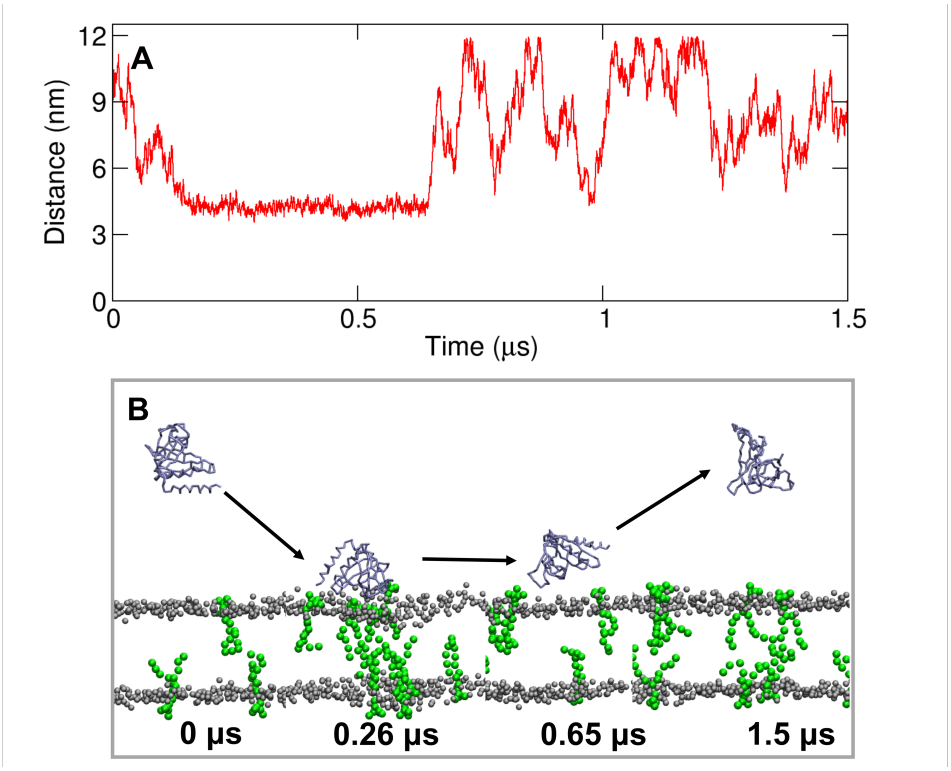

Figure 7.10: Kindlin-3 PH domain mutant dissociates from membrane during MD simulations. (A) Distance between the centre of mass of the protein and the centre of mass of the bilayer for one of the simulations with the mutated form of the protein (Kin3mutPH-1). (B) Snapshots from the same system shown at 0, 0.26, 0.65 and 1.5 μs. PIP_3 molecules are shown in green, POPC and POPS phosphate atoms are shown in grey and the kindlin-3 PH domain is shown in iceblue. **Figure prepared by Kalli AC., Naughton FB. and Sansom MSP.**

Nanodiscs incorporated with different components were stable in solution (**Figure 7.11A** and **Figure 7.11D**), with an estimated molecular weight of 120 kDa; they

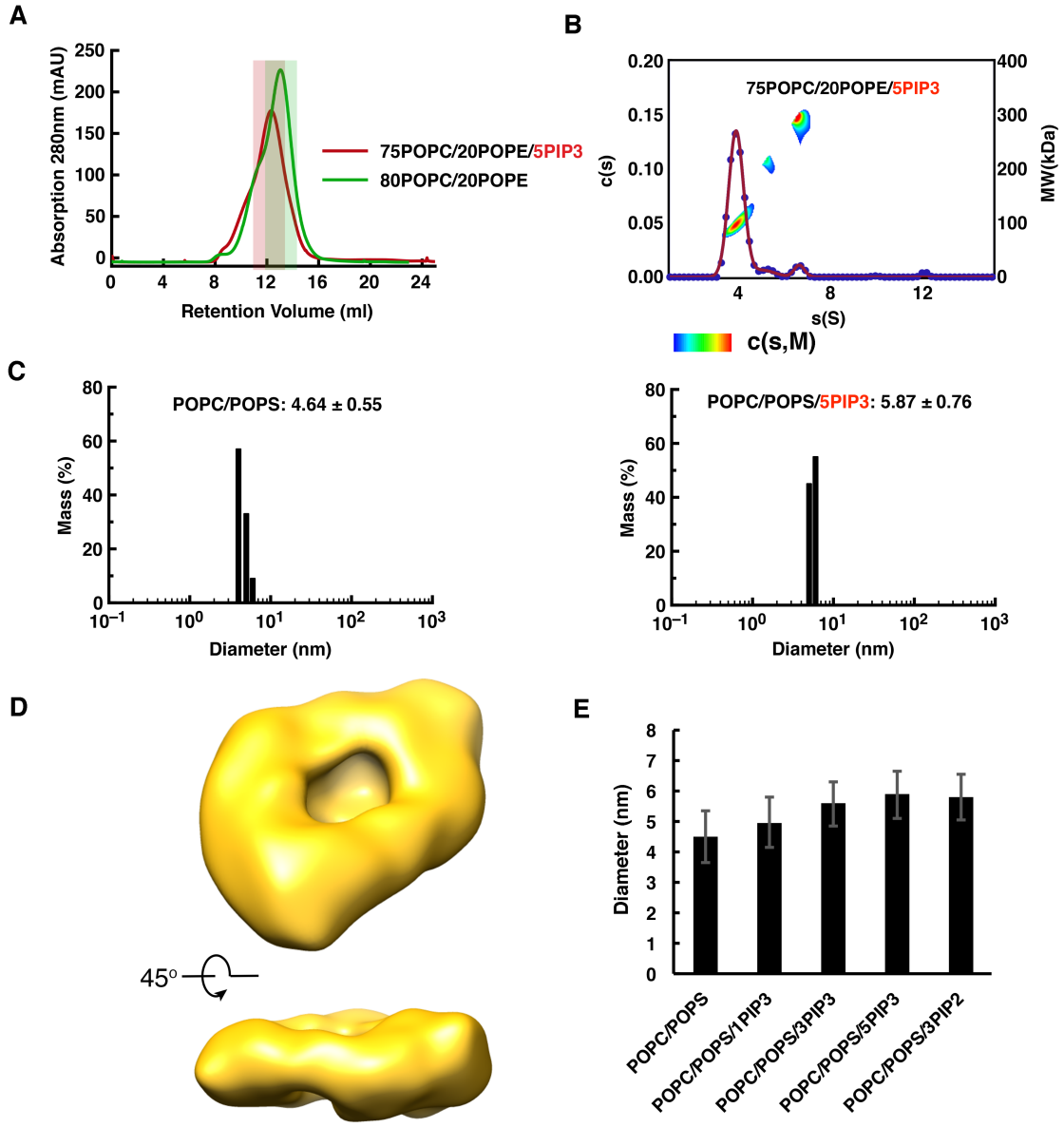

Figure 7.11: Preparation and characterisation of nanodiscs. (A) Size-exclusion chromatography (Superdex 200 10/30) profile of nanodiscs purification. Two representative gel-filtration profiles for the nanodiscs with the indicated lipids compositions.The regions shed in light pink and green represent the peak fractions used for this study. (B) Analytical ultracentrifugation study of nanodiscs showing its homogeneous distribution in solution. The AUC study was conducted at 0.1 mg/ml at 293K in a sedimentation velocity experiment. The majority of the nanodiscs are distributed with a sedimentation coefficient around 4S, with an estimated molecular weight of 120 kDa. The higher molecular weight species only takes a small quantity of the total amount. The scale bar underneath the main figure indicates the distribution of sedimentation coefficient and molecular weight. (C) Dynamic light scattering measurement of two representative nanodiscs species. Two nanodiscs with different composition of lipids were measured. (D) 3D reconstruction of the negative-stained POPC/POPS nanodiscs from cryo-electron microscopy. (E) Size distribution of nanodiscs with various lipids composition measured by dynamic light scattering. The diameter was calculated from three consecutive fractions during size-exclusion chromatography within the right half of the peak. Including PIP2 or PIP3 apparently increased the overall size of the nanodiscs.

behaved as single species with a sedimentation coefficient of 4 S as assessed by analytical ultracentrifugation (**Figure 7.11B**). It appears that the diameter of the nanodiscs correlates with the molar ratio of phosphatidylinositides incorporated in their membranes: nanodiscs with 5% PtdIns(3,4,5)P_3 (d = 5.87 nm) were bigger than those with 3% PtdIns(3,4,5)P_3 (d = 4.64 nm) (**Figure 7.11C** and **Figure 7.11E**).

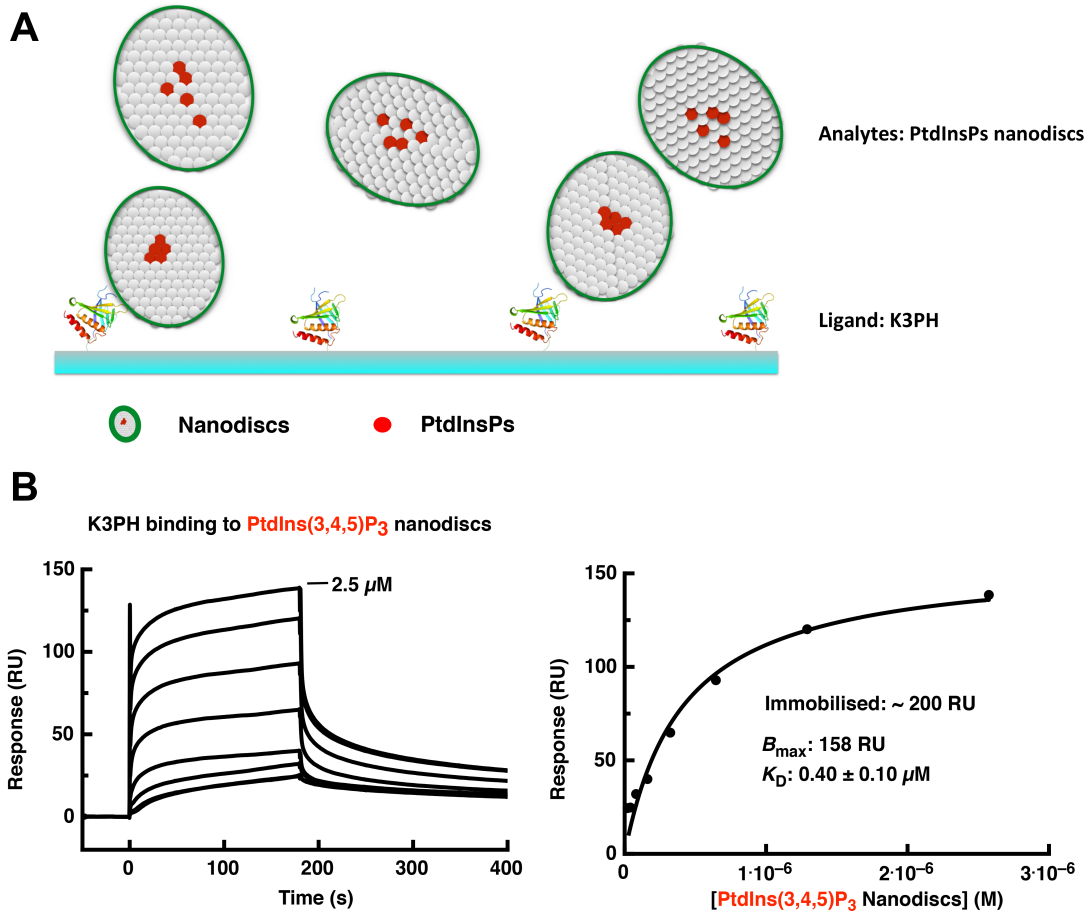

Figure 7.12: SPR study of kindlin-3 PH domain with PtdIns(3,4,5)P_3 incorporated nanodiscs. (A) Schematic representation of SPR experiments with immobilised PH domain as ligand and lipid nanodiscs as analyte in solution. The recombinant PH domains with a C-terminal Avitag was *in vivo* biotinylated and coupled to the streptavidin derived CM5 sensor chip. (B) SPR measurement between WT kindlin-3 PH domain and nanodiscs with 5% PtdIns(3,4,5)P_3. The background-subtracted raw sensogram data for the wild type protein is shown on the left and the resulting binding saturation curve on the right. The highest concentration used is given on the sensogram graph; subsequent binding profiles derive from a series of two-fold dilutions. See Materials and Methods for experimental details.

The SPR results showed that 5% PIP_3-incoporated nanodiscs bind to kindlin-3 PH

domain strongly, with a much-enhanced affinity (K_D = 0.4 μM) (**Figure 7.12B**) compared to that shown by the PH domain towards immobilised inositol phosphates. In contrast, nanodiscs with 5% PIP_2 showed no binding to kindlin-3 PH domain, demonstrating that PIP_3 is its real ligand and suggesting that the apparent binding to immobilised PIP_2 is due to a non-specific interaction. We also performed the same experiments with nanodiscs without incorporated phosphatidylinositol moieties and observed no binding in SPR experiments.

There is evidence for PIP_2 clustering within lipid bilayers (Ellenbroek et al., 2011) while other data suggest protein-driven PIP_2 clustering in signalling complex formation (van den Bogaart et al., 2011). While a study of yeast proteins has revealed that the interaction between one set of PH domains and membranes can be enhanced by cooperative interaction between PIPs and other lipids (such as POPS) (Vonkova et al., 2015a), the interaction between kindlin-3 PH domain and PIP_3 seems to more rely on the concentration of PIP_3 in the membrane, i.e., the cooperativity among copies of the same lipid.

7.3.5 Grp1 PH domain binds inositol phosphate with 1:1 specificity

We further asked whether our observations are specific to kindlin-3 PH domain (and similar domains) or could also be shown for PH domains which are already known to bind to inositol phosphate species in a classical, 1:1 interaction mode. In this case, the observed binding affinity should not be affected by the experimental setup. The crystal structure of the Grp1-PH domain in complex with Ins(1,3,4,5)P_4 (equivalent to the PtdIns(3,4,5)P_3 headgroup) reveals the ligand well embedded in a pocket formed by positively-charged residues (Ferguson et al., 2000). In line with this, we showed by SPR that the analyte/ligand design of our experiment does not significantly affect the apparent affinity of Grp1 for PIP_3 compared to the effect it has on the kindlin-3 PIP_3 binding observed. Grp1-PH domain was shown to bind to immobilised PIP_3 headgroups with a K_D of 2.8 $\pm$ 0.4 μM (**Figure 7.12A**), and to

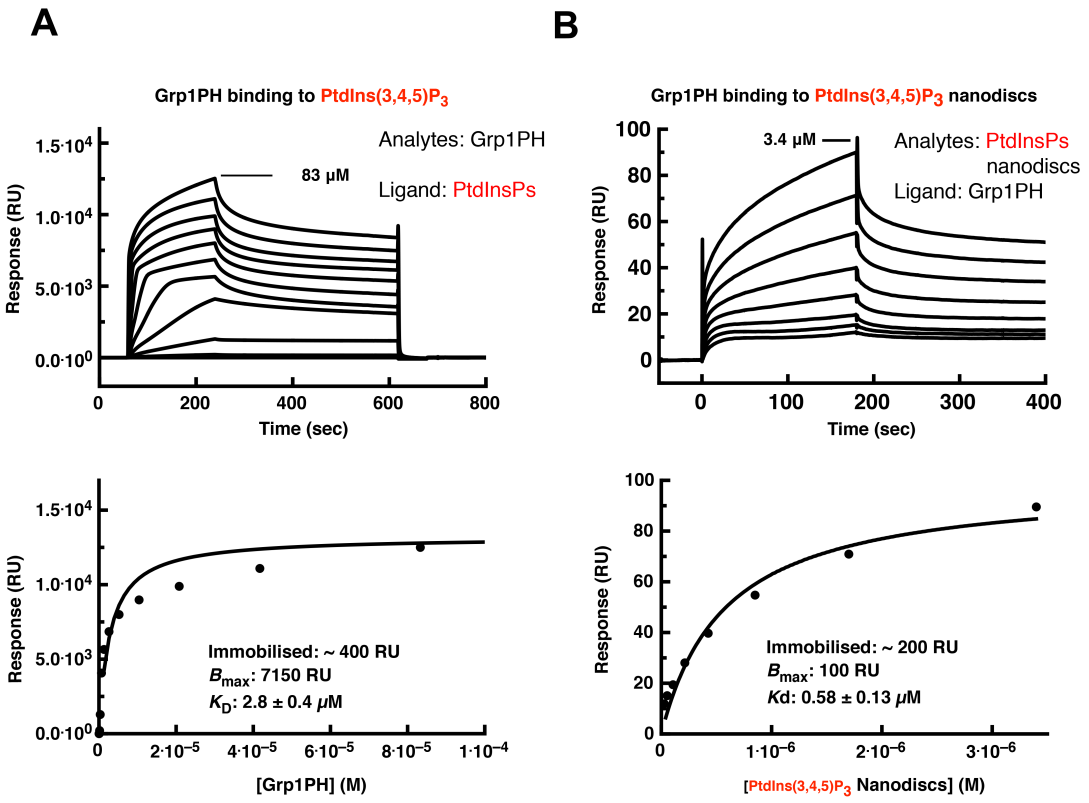

Figure 7.13: SPR study of interaction between Grp1PH domain and PtdIns(3,4,5)P_3 in two experimental setups. (A) Immobilised PtdIns(3,4,5)P_3 headgroups as ligand and PH domain as analyte. (B) Immobilised Grp1 PH domain as ligand and lipid nanodiscs as analyte.The upper panel in each case gives the background-subtracted raw sensogram data and the lower panel the resulting binding saturation curve. The highest concentration used is given on the sensogram graphs (83 μM Grp1 PH domain in A; 3.4 μM nanodiscs in B); subsequent binding profiles derive from a series of twofold dilutions. See Materials and Methods for experimental details.

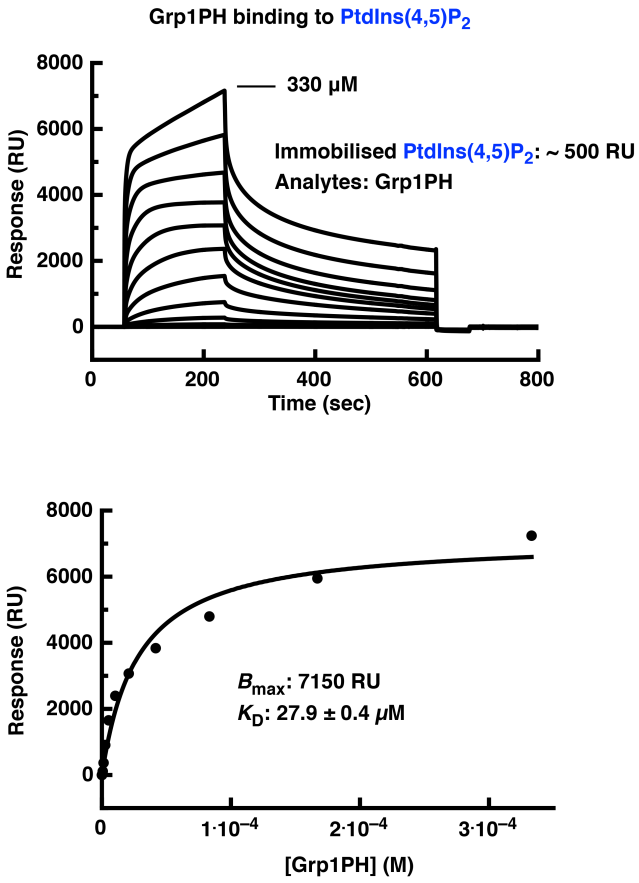

Figure 7.14: Grp1 PH domain binding to PtdIns(3,5)P_2. SPR study with immobilised PtdIns(3,5)P_2 on sensor chip and Grp1 PH domain as analyte. The experimental setup was the same with that in **Figure 7.13A**.

immobilised PIP_2 headgroups with a K_D ten times smaller (**Figure 7.13**), consistent with a previous study though with a lower absolute value due to methodological differences (Lai et al., 2013; Pilling et al., 2011) . In contrast, the binding affinity to PIP_3 within nanodiscs (5% of constituent lipids) has a K_D of 0.58 μM (**Figure 7.12B**). So, in the case of Grp1, there is only a 5 times difference in affinity between immobilised lipid headgroups and whole lipids in a lipid nanodisc (rather than the $\sim$100x enhancement observed with the kindlin-3 PH domain). This suggests that the Grp1-PH interaction with PIP_3 at high affinity does not rely on the binding of multiple PIP_3, though it does indicate a non-specific enhancement of lipid binding - most likely by a simple electrostatic attraction - in a membrane-like context (Vonkova et al., 2015a).

7.4 Conclusion and perspective

The cytoplasmic surface of the plasma membrane serves as a platform for many critical biological processes. The recruitment of peripheral membrane proteins from the cytoplasm to the plasma membrane surface is spatiotemporally and accurately regulated. The PH domain is one of a selection of commonly-used folds directing protein-membrane association and has been shown to confer specificity to different inositol phosphates on protein-membrane interactions. In the present study, we firstly solved the crystal structure of kindlin-3 PH domain, and then used the crystal structure as the basis for a molecular dynamics simulation study. MD simulation revealed that kindlin-3 PH domain interact with PIP lipids primarily via the loop connecting the β1 and β2 strands and that it has a somewhat higher affinity for binding PtdIns(3,4,5)P_3 rather than PtdIns(4,5)P_2 in a membrane-like system. Moreover, one kindlin-3 PH domain appears to bind multiple inositol phosphate lipids, by means of which the affinity of interaction between the PH domain and the membrane is greatly enhanced by an avidity effect. In contrast, with a PH domain (from Grp1) previously shown to have canonical and specific one-to-one binding to inositol phosphate lipid, this one-to-more binding mode was not observed. These observations from MD simulation were supported by our surface plasmon resonance experiments in two complementary setups. The first one involves immobilisation of biotinylated inositol phosphates on the sensor chip and use of the PH domain (WT and mutant) as the analyte, to mimic an isolated one-to-one binding mode. In this case, each PH domain could only bind to no more than one phospholipid due to the distance restraints imposed by streptavidin-mediated immobilisation. The other setup attempted to mimic the bilayer plasma membrane, by immobilising the PH domain on the chip, and using soluble membranes (nanodiscs) as the analyte. The affinity of interaction measured between the PH domain and inositol phosphate lipids were remarkably affected by experimental design.

Compared with liposomes, nanodiscs have obvious advantages for protein-lipid interaction studies in a membrane-like environment: they are more stable and have a limited number of lipids per disc. Nanodiscs made of MSP1E3D1 are estimated to accommodate about 125 POPC lipids, with a diameter of about 13 nm (Bayburt and Sligar, 2010); this is in agreement with our cryo-EM map of the nanodisc which is fitted well (length ∼14nm, width ∼9nm) with the elongated and twisted conformation of the human apolipoprotein A1(Borhani et al., 1997). In the SPR experiments here, the small and homogeneous size of the nanodisc membrane is critical for 1:1 model fitting and interpretation of the SPR results. Liposomes, with inherently larger diameters (e.g., 50 - 100 nm), would create multiple possible contact sites for PH domains, confounding binding affinity analysis. Nanodiscs however are able to bind no more than one immobilised PH domain due their limited size and the size and chip surface density of the immobilised PH domains. It should be noted that in this study all the CM5 sensor chips were coated with a similar level of streptavidin. The activated CM5 chips were allowed to bind streptavidin until saturation, when the total response caused by streptavidin derivatisation no longer increased (The total responses were about 3000RU). In such cases, the chip surface would only be able to bind to a similar amount of ligand even all the streptavidin were occupied with ligands. In our study, all the CM5 sensor chips were coated with a similar amount of ligand in terms of response unit (about 200 RU). This level of ligand immobilisation did not reach the saturation level, since the measured responses were still almost in a linear binding curve instead of reaching a plateau when the immobilisation procedure was finished. Moreover, considered the molecular weight difference of lipid head groups and PH domains (about 1kDa for lipid head groups *vs* 18 kDa for PH domains), the PH domain would result in a similar level of response to that of lipid head groups even with only about 5 percent quantity of lipid head groups. Therefore, the PH domain immobilised on the chip were far from saturation and would only occupy a limited number of coated streptavidin on the chip. As a result, a single nanodisc would not be able to bind to two (or more) adjacent PH domain

simultaneously, and a single immobilised PH domain would not be able to bind more that one nanodisc.

There is increasing evidence showing that PH domains can interact with multiple lipid species simultaneously. A recent crystal structure of the ASAP1-PH domain suggested that there were two anionic lipid binding sites on the ASAP1-PH domain (a canonical site and an atypical site) (Jian et al., 2015). Moreover, Vonkova *et al.* studied 91 different PH domains and showed that lipid cooperativity is essential for their recruitment to the membrane (Vonkova et al., 2015b). Clustering of lipids by the kindlin-3 PH domain is expected to change the local lipid environment. Given that kindlins act in synergy with talin, which was also shown to interact with anionic lipids (Elliott et al., 2010; Kalli et al., 2013), this may provide a mechanism for how talin and kindlins in turn associate with the integrin. Focal adhesion kinase has also been shown to bind to inositol phosphates, inducing clustering of a protein-mediated kind (Goñi et al., 2014).

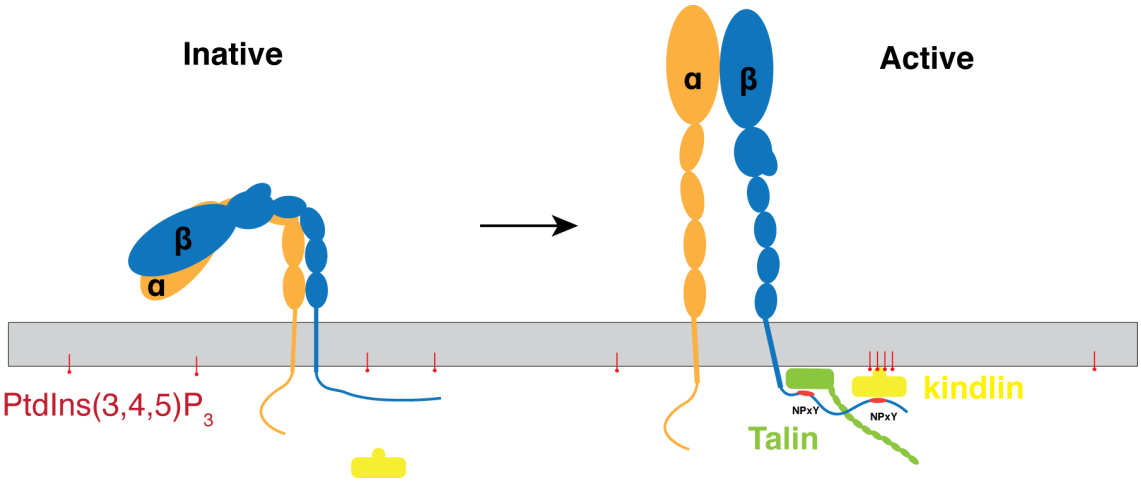

Figure 7.15: Proposed model for lipid clustering in integrin activation. The activating conversion of integrins to an upright conformation is promoted by the binding of talin to the membrane-proximal NPxY and of kindlin to the membrane-distal NPxY on an integrin beta-subunit cytoplasmic tail. Clustered PtdIns(3,4,5)P_3 interacts and recruits kindlin-3 from cytosol to membrane, thus facilitating the formation of integrin-talin-kindlin tertiary complex.

The orientation of the kindlin-3 PH domain on a target membrane indicated by this study, and the fact that kindlins have a rather long unstructured loop in the F1

domain, allow us to form a hypothesis for how the kindlin head domain may interact with the membrane and with the integrin receptor. In particular, the orientation of the PH domain that we find, in combination with the short linker regions plugging it into the F2 subdomain, suggest that it acts as a buffer or spacer between the kindlin and the membrane surface, and thereby allows the F3 subdomain to interact with the membrane distal NPxY integrin β-subunit tail motif rather than the membrane proximal NPxY bound by talin (Harburger and Calderwood, 2009; Yates et al., 2012a; Calderwood et al., 1999) (**Figure 7.15**). Thus in combination with the N-terminal F0 domain (Goult et al., 2009) and long F1 loop (Bouaouina et al., 2012) the PH domain provides kindlins with a specific location and binding mode for integrin activation.

There are no high resolution structural data for the rest of kindlin-3, however it has been shown to have a similar structure to the talin head domain (Yates et al., 2012b). Therefore, further studies will focus on structure determination for kindlins and use structural and biophysical methods to test the hypothesis outlined above that binding of clustered phosphatidyl inositol phosphate lipids is a mechanism employed by some proteins and specifically a subset of PH domains to enable the binding of membranes with an enhanced affinity.

7.5 Methods and Materials

7.5.1 Constructs design, cloning and mutagenesis

Mouse *FERMT3* (UniProtKB: Q8K1B8) encoding kindlin-3 PH domain (residues 344 - 479) and Grp1 (General receptor of phosphoinositides 1, UniProtKB: O43739) PH domain (residues 264 - 392) were cloned into a pOPINJ vector, which contains a tandem His_6 and GST (Glutathione S-transferase) tag followed by a Human Rhinovirus (HRV)-3C protease cleavage site in the N-terminus. The successful cloning was confirmed by DNA sequencing (Source Bioscience, UK). Mutagenesis and insertions were achieved by overlapping PCR following the standard protocol (Heckman and Pease, 2007). A DNA fragment encoding 15-residue AviTag^{TM} was inserted between the 3C protease cleavage site and PH domain by PCR to enable *in vivo* protein biotinylation (Fairhead and Howarth, 2014) using *E. coli* strain for surface plasmon resonance experiments.

7.5.2 Expression of kindlin-3 PH domain and Grp1 PH domain

The construct encoding recombinant kindlin-3 PH domain was transformed into *E. coli* B834 strain. A single colony was picked from fresh transformants on LB-agar-carbenicillin plate and cultured in 100ml of LB supplemented with 50 μg/ml of carbenicillin overnight. The pre-culture was diluted 50 times into LB media with antibiotics and incubated at 37°with vigorous shaking at 200 rpm until OD_{600} reached about 0.6. The expression of protein was induced by 1mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) at 18 ℃ overnight. Cells were harvested by centrifugation at 6000 g for 20 min and pellets were frozen in -80 ℃ freezer until further use. Other proteins used in this study were expressed with the same protocol.

7.5.3 Protein purification and crystallisation

The frozen cell pellet of kindlin-3 PH domain (about 10 g of cell pellet from 2 l of overnight culture) was resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl and 0.1% Tween-20) and lysed with sonication and the debris was separated by centrifugation at 40,000 g for 1hr. The resulting supernatant was incubated with glutathione sepharose equilibrated with lysis buffer for 2 hrs. The glutathione sepharose with protein bound was then washed with lysis buffer extensively. GST-3C protease was added to cleave the kindlin3-PH domain from sepharose overnight at 4 ℃ with gentle shaking. Supernatant containing PH domain were concentrated and applied to size exclusion chromatography (Superdex 75 16/60, GE Healthcare Life Sciences) in 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.7 mM TCEP (*tris*(2-carboxyethyl)phosphine) buffer to obtain the purified proteins. The purification of kindlin-3 PH domain mutants and Grp1-PH domain and biotinylated proteins was carried out in the same way. Successful *in vivo* biotinylation was verified by efficient binding to streptavidin magnetic beads (Sigma-49532).

Crystallisation screening was performed by means of vapour diffusion methods. Crystals appeared within 4 days and continued growing until 2 weeks in 4 M Sodium Nitrate, 0.1 M Sodium Acetate, pH 4.6. Crystals were flash-cooled in a cryoprotectant with 25% (v/v) glycerol in mother liquor.

7.5.4 Data collection, processing and structure determination

Diffraction data were collected at beamline i24 in Diamond Light Source (Didcot, UK) and then indexed, integrated and scaled using Xia2 pipeline(Winter, 2009). The structure of kindlin-3 PH domain was solved by molecular replacement (MR) using PHASER (Adams et al., 2010) with a polyalanine model of kindlin-1 PH domain (PDB: 4BBK). The resulting model from MR was further built by *Buccanneer* (CowtanIUCr, 2006) and corrected manually in *Coot* (Emsley et al., 2004). Refinement was

accomplished by using *phenix.refine* together with manual correction in *coot*. Molecular graphs were prepared with *PyMOL*. Attempts in either co-crystallising kindlin-3 PH domain with Ins(1,3,4,5)P_4 or soaking the crystal with highly concentrated Ins(1,3,4,5)P_4 resulted in PH domain structure without ligand bound, presumably due to low affinity between Ins(1,3,4,5)P_4 and kindlin-3 PH domain.

7.5.5 Nanodiscs preparation

POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) and POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) were purchased from Avanti Polar Lipids and PtdIns(4,5)P_2 (P-4516), PtdIns(3,4,5)P_3 (P-3924)and Ins(1,3,4,5)P_4 (Q-1345) were purchased from Echelon Biosciences. Nanodiscs were prepared following the standard protocols (Borch and Hamann, 2009).Briefly, to prepare nandiscs containing 5% (mol/mol) PtdIns(3,4,5)P_3, a mixture of 200 μL of 10 mg/ml POPC, 200 μL of 10 mg/ml POPE together with 670 μL of 0.5mg/ml PtdIns(3,4,5)P_3 in chloroform were dried on clean pyrex tube under argon. Residual chloroform was then removed by desiccation in a desiccator attached on a VARIO-SP diaphragm pump (Vacuubrand). The lipid film was subsequently hydrated by addition of 1 ml solubilisation buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 17 mM Na- cholate) followed by vigorous votexing. The hydrated lipid film was then sonicated for 10 min in a water bath sonicator before 250 μl of MSP1E1 (7 mg/ml) proteins were added to the lipid mixture. Nanodiscs were self-assembled upon removal of Na-Cholate by dialysing against 20 mM HEPES, pH 7.5, 150 mM NaCl buffer overnight. The aggregate was removed by size-exclusion chromatography (Superdex 200 10/300, GE Healthcare Life Sciences) and the peak fractions corresponding to size of monomeric nanodiscs species were collected (Fig S7A) and used immediately for biophysical characterisation and SPR experiments.

7.5.6 Surface Plasmon Resonance

SPR experiments were performed using a Biacore T200 machine (GE Healthcare Life Sciences) at 20 ℃ in 10 mM HEPES, pH 7.5, 150 mM NaCl, PtdIns(4,5)P_2 (C-45B6a) and PtdIns(3,4,5)P_3 (C-39B6a) were purchased from Echelon Biosciences. The BIAcore CM5 chip (GE Healthcare Life Sciences) was firstly covered with streptavidin following the manufacturer's instructions before the biotinylated PIPs or biotinylated proteins were immobilised. The analyte with 2-fold serial dilutions was applied at a flow-rate of 20 μl/min for 180 s followed by 600 s of dissociation time. The biosensor chip was regenerated after each sample injection cycle with different buffer in the two experiment setups: in the case of immobilised PIPs on sensor chip, 0.1 % SDS was used; on the other hand, the surface was regenerated by 2 M $MgCl_2$. The data were fit with 1:1 Langmuir adsorption model ($B = B_{max}C/(K_{\mathrm{D}} + C$), where B is the response of bound analyte and C is the concentration of the analyte in the sample) to calculate the dissociation constant (K_{D}) using BIAanalysis software. The molecular weight of nanodiscs for SPR analysis was estimated to be 120 kDa.

7.5.7 Coarse-grained molecular dynamics (CG-MD) simulations

The CG-MD simulations were performed using GROMACS 4.5.5 (Hess et al., 2008) with the Martini 2.1 force field (Monticelli et al., 2008; Siewert J Marrink et al., 2007) . The simulation set up is shown in **Figure 7.5**. A summary of all simulation performed is shown in **Table 7.3**. Prior to the production simulations all systems were energy minimized for 300 steps and equilibrated for 500 ns with the protein backbone particle restrained. For each repeat simulation within an ensemble, the protein was rotated around the x, y, and z axes to form a different initial configuration. An ensemble of 20 simulations of 1.5 μs each were run with a time step of 20 fs. An elastic network model (ENM) was applied to all backbone particles with a cut-off

distance of 0.7 nm (Atilgan et al., 2001). The LINCS algorithm was used to constrain to equilibrium bond lengths (Hess et al., 1997). Lennard-Jones interactions were shifted to zero between 0.9 nm and 1.2 nm. Coulombic interactions were shifted to zero between 0 and 1.2 nm. The pressure was 1 bar and the temperature was 323 K. A Berendsen's algorithm (Berendsen et al., 1984) was used to control the pressure and the temperature with a coupling time of 1 ps.

7.5.8 Potential of mean force (PMF) simulation

Initial structures of wild type or mutant kindlin-3 PH domain bound to PIP_3 or PIP_2 in a 80:20 POPC:POPS bilayer were generated by alignment with an existing Grp1-PH/PIP_3 complex (see (Naughton et al., 2016)). The system was converted to a coarse-grained representation. The PH domain was pulled from the bound PIP lipid along the membrane normal (z-axis) by fixing the center of mass to a reference point which was moved at a rate of 0.001 nm ps^{-1} for a total distance of 2.6 nm, while the PIP phosphate atom was restrained to its initial position. Snapshots with protein-lipid separations (measured from the PH domain center-of-mass to the phosphate atom of the PIP) in 0.1 nm intervals for the first 2 nm and 0.2 nm thereafter were extracted and used as initial structures for umbrella sampling simulations. In each simulation, the phosphate atom was again restrained to its initial position and the z-distance of the PH center-of-mass was restrained relative to the phosphate atom with a force constant of 1000 kJ mol^{-1} nm^{-2}. To reduce time required for adequate sampling, the PH domain center-of-mass was also restrained in the x and y directions, by a force constant of 100 kJ mol^{-1} nm^{-1}. Each window was simulated for 1μs (mutant), 1.400 μs (WT/PIP_3) or 1.5 μs (WT/PIP_2), with the distance between PH center-of-mass and PIP atom bead along the z-axis recorded every 20 ps. PMF profiles were calculated using the GROMACS *g_wham* tool . Time taken to reach convergence was judged from profiles generated for sequential time intervals; data before this time was discarded as equilibration before a final

profile was calculated. Errors were obtained from bootstrap analysis. All profiles were shifted so the value of the PMF in bulk was 0 kcal mol^{-1}.

CHAPTER 8

Conclusions and Perspectives

8.1 Overview of Structural Biology

Structural biology has greatly broadened our scientific horizons and transformed our understanding of life. It will continue to elucidate the details that build the blocks of life, ranging from the very fine and simple motif to the large and sophisticated complex. The development of new tools and methods in structural biology and other areas are now revolutionising our approach to a scientific understanding of life.

X-ray crystallography remains the major tool for structure determination at the moment. On one hand, x-ray synchrotron facilities continue to improve, with brighter, tunable and more flexibly modified (e.g. attenuated) x-ray beams, more sensitive and faster detectors and an automated data collection and processing pipeline. These advances have enabled *de novo* phasing using the weak anomalous signal of 'light' atoms (such as sulphur and phosphate) intrinsic to the proteins forming macromolecule crystals. In particular, an ambitiously-scoped beamline (I23 in Diamond Light Source, UK) has been built specifically for long-wavelength de novo phasing. On the other hand, the development of x-ray free-electron lasers with an extremely brilliant beam enables small crystals to be exposed to x-rays for picosecond exposure times before radiation damage takes over. This will facilitate the assessment of smaller and poorer diffracting crystals, especially for the crystals of membrane proteins.

The recent breakthroughs in cryo-EM (both in software and hardware) make it possible to solve the structures of macromolecules at near atomic resolution without the need for crystallisation (Cheng et al., 2015). Cryo-EM provided the first direct evidence demonstrating that CDCs undergo large conformational rearrangements during pore formation and clearly it will continue to contribute to determining the high-resolution structure of pores assembled by MACPF/CDC proteins in the future. The lower size limit for structure determination by cryo-EM remains above 100kDa, largely due to the difficulty of assigning accurate alignments to such small unstained

objects imaged using an electron microscope. Lastly, it should be noted that although x-ray crystallography and cyro-EM provide high-resolution structures, providing a basis to infer the mechanisms of pore formation, they are largely snapshots of a specific static state, lacking the time dimension to visualise the structural transition among different states. Although not utilised in this study, a combination of techniques including atomic force microscopy (AFM), electrophysiology and super-resolution single-molecule fluorescent microscopy have enabled more details of pore formation to be captured. Together, these techniques enabled the argument about the function and structure of the incomplete rings (arcs) formed by MACPF/CDC superfamily proteins to be solved. So all in all, how can we make use of these opportunities to further study MACPF/CDC superfamily proteins?

8.2 Conclusions and Future plans

8.2.1 Linking Astrotactins back to context

We have determined the structure of a large portion of Astrotacin-2 and demonstrated its unique combination of an abbreviated MACPF domain with other domains (Fn(III) and annexin-like domain). The biophysical properties of this large portion of the protein were also investigated such as its interaction with inositol phosphate and Ca^{2+}. We have also shown that this portion of ASTN2 is very likely to undergo a pH-dependent self-interaction, although the details of the self-interaction remain unknown and we are not yet aware of its biological significance. Another effort we have made is to obtain a full-length endodomain of ASTN2, and optimization of co-expression and purification of such endodomains is ongoing.

Looking forward, it is of particular interest to know the oligomerisation status of full-length ASTN2 (including the transmembrane regions and cytosolic domain). ASTN2 cycling in the endosomal system points out the possibility of it encountering different pHs during its activity, and it would be interesting to see how its oligomeric state changes in its native membrane environment along with a change in pH. As part of the experiments towards this goal, we have employed live-cell FRET-FLIM (Fluorescence Resonance Energy Transfer (FRET) and Fluorescence Lifetime Imaging (FLIM)) experiments to check the self-interaction of ASTN-2 *in vivo.* The constructs of full-length ASTNs with both cytosolic and C-terminal regions labelled with fluorescent tags have now been cloned and have shown expression in HEK 293T cells. Moreover, we are now trying to identify the protein partners that associate with both Astrotactin-2 and -1, to put them back into a cellular context.

8.2.2 Pore formation by Apicomplexan Perforin-like proteins

We have determined the structure of *Tg*PLP1 in this study. Our structure of the MACPF domain and C-terminal domain of *Tg*PLP1 will help to elucidate the mechanism of pore formation by this family of perforin-like proteins. The observation that the MACPF domain of *Tg*PLP1 assembles as a hexameric ring in crystallisation conditions is extremely pleasing and we are now energetically pursuing an assessment of the actual pore oligomeric state to verify the biological significance of our observations there. In biological context we with collaborators will now turn our attention also to a molecular-detail investigation of the roles played by the PLPs during the life cycles of *Plasmodium* and *Toxoplasma*.

It should be noted that our successes with *Tg*PLP1 are very likely transferrable to other ApiPLPs. Indeed, aside from the expression and structure solution of *Tg*PLP1, we have also been able to express several other ApiPLPs, particularly the C-terminal domain of *Pv*PLP1 and a large portion of *Pv*PLP2. There are large variances in expression levels of these proteins, and some constructs did not show any secretion at all. To express these proteins, one potential approach might be to use an expression system mimicking the native environment of these proteins, for example, LEXSY (eukaryotic protein expression in *Leishmania tarentolae*), since the Apicomplexas are closer to Leishmania in phylogeny than to any other expression host we are now working with (yeast, *E.coli*, insect cell and mammalian cell).

ApiPLPs of *Plasmodium spp.* share about 30% identity with each other and each of them is expressed and functional at specific stages in the life cycle of *Plasmodium*. The overall structure of these ApiPLPs are very likely very similar, but their detailed mechanisms of membrane binding and pore formation could still be quite or very different. Cholesterol-dependent cytolysins (CDCs) characterised today share more than 50% identity among each other and showed a very similar crystal structure, but the mechanisms of membrane attachment and biophysical properties are still

distinct. While all of the CDCs require cholesterol as a receptor, intermedilysin requires an additional GPI-anchored protein (CD59) for efficient membrane binding and pore formation; and listeriolysin showed a pH-dependent activity while others do not. It would not be surprising that *Api*PLPs bind to membrane using different mechanisms considering the relatively larger variance in sequence compared with that of the CDCs. So a thorough biophysical and structural characterisation of these proteins is clearly an important goal direction.

8.2.3 Membrane binding property of Pleckstrin homology domains

Pleckstrin homology domains (PH) are one of the most studied membrane-binding domains and are known for their specific binding to phosphatidylinositol phosphates. However, even though the overall fold of PH domains are similar to each other, some of them bind to inositolphosphtes very strongly, with sub-micromolar or nanomolar affinity, while others bind very weakly or not at all. Kindlin-3 PH domain is one PH domain that binds to inositolphosphates rather weakly if at all (about 300 μM in one SPR study). However, this extremely weak interaction is known to be biologically relevant. Our characterisation of the kindlin-3 PH domain points out the possibility that they might bind to several inositolphosphates in a co-operative manner, and this has lead us to further explore whether other PH domains bind to inositolphosphates in the same manner. We are grateful that the BBSRC has funded this project to enable the details to be further investigated.

8.3 Final remarks

Great progress has been achieved towards understanding the mechanism of pore formation by MACPF/CDC superfamily proteins, but there are still much more details to fill in and we are continuing to devote our efforts to the characterision of proteins from this family.

Studies on perforins and the membrane attack complex reveal that these MACPF proteins are undoubtedly of great significance in mammalian immune defence. Considering the ubiquitous distribution of MACPF domain proteins among the various kingdoms of life, and the high-degree conservation of the individual proteins among vertebrate species, it is tempting to speculate that these MACPF proteins undertake fundamental roles. MAC and perforin-1 have been well-studied structurally as paradigms of pore formation for MACPF proteins, while the structure and function of others remain unclear. While a simple model of pore formation explains the functions observed for membrane attack complex and perforins, it doesn't explain that of BRINPs and Astrotactins in the human brain.

Technically, structural biology alone will not directly tell us how these neuron-specific MACPF proteins work in the brain, and so it needs to be combined with various techniques using *in vivo* systems. We are very fortunate and grateful that continuation of the work on Astrotactins in this thesis has been supported by the UK Medical Research Council so that these further details can be investigated. In the end, why are MACPF proteins expressed specifically in brain? Do they form pores in the brain and what is that significance of that?

We don't know, but we are on the way to discover it ...

CHAPTER 9

Bibliography

Adams, N.C., Tomoda, T., Cooper, M., Dietz, G., and Hatten, M.E. (2002). Mice that lack astrotactin have slowed neuronal migration. Development (Cambridge, England) *129*, 965–972.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica. Section D, Biological Crystallography *66*, 213–221.

Aikawa, M., Wernsdorfer, W.H., and McGregor, I. (1988). Fine structure of malaria parasites in the various stages of development.

Akey, D.L., Brown, W.C., Konwerski, J.R., Ogata, C.M., Smith, J.L., and IUCr (2014). Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1. Acta Crystallographica. Section D, Biological Crystallography *70*, 2719–2729.

Amino, R., Giovannini, D., Thiberge, S., Gueirard, P., Boisson, B., Dubremetz, J.-F., Prévost, M.-C., Ishino, T., Yuda, M., and Ménard, R. (2008a). Host cell traversal is important for progression of the malaria parasite through the dermis to the liver. Cell Host & Microbe *3*, 88–96.

An, Z., Dobra, K., Lock, J.G., Strömblad, S., Hjerpe, A., and Zhang, H. (2010). Kindlin-2 is expressed in malignant mesothelioma and is required for tumor cell adhesion and migration. International Journal of Cancer *127*, 1999-2008.

Anand, K., Maeda, K., and Gavin, A.-C. (2012). Structural Analyses of the Slm1-PH Domain Demonstrate Ligand Binding in the Non-Canonical Site. PLoS ONE *7*, e36526.

Anderluh, G., Kisovec, M., Kraševec, N., and Gilbert, R.J.C. (2014). Distribution of MACPF/CDC proteins. Subcell. Biochem. *80*, 7-30.

Angov, E., Hillier, C.J., Kincaid, R.L., and Lyon, J.A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PloS One *3*, e2189.

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallographica. Section D, Biological Crystallography *62*, 1243–1250.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research *38*, NaN – NaN.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. Biophys. J. *80*, 505-515.

Attar, N. (2015). Parasite biology: PINning down Theileria. Nature Reviews Microbiology *13*, 126–127.

Bachir, A.I., Zareno, J., Moissoglu, K., Plow, E.F., Gratton, E., and Horwitz, A.R. (2014). Integrin-Associated Complexes Form Hierarchically with Variable Stoichiometry in Nascent Adhesions. Current Biology *24*, 1845-1853.

Baum, J., Gilberger, T.-W., Frischknecht, F., and Meissner, M. (2008). Host-cell invasion by malaria parasites: insights from Plasmodium and Toxoplasma. Trends in Parasitology *24*, 557–563.

Bayburt, T.H., and Sligar, S.G. (2010). Membrane protein assembly into Nanodiscs. FEBS Letters *584*, 1721-1727.

Benke, S., Roderer, D., Wunderlich, B., Nettels, D., Glockshuber, R., and Schuler, B. (2015). The assembly dynamics of the cytolytic pore toxin ClyA. Nature Communications *6*, 6198.

Bergfors, T. (2003). Seeds to crystals. Journal of Structural Biology *142*, 66–76.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics *81*, 3684-3690.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Research *35*, e45.

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

Black, M.W., and Boothroyd, J.C. (2000). Lytic cycle of Toxoplasma gondii. Microbiology and Molecular Biology ....

Blader, I.J., Manger, I.D., and Boothroyd, J.C. (2001). Microarray analysis reveals previously

unknown changes in Toxoplasma gondii-infected human cells. Journal of Biological Chemistry *276*, 24223–24231.

Bialkowska, K., Ma, Y.-Q., Bledzka, K., Sossey-Alaoui, K., Izem, L., Zhang, X., Malinin, N., Qin, J., Byzova, T., and Plow, E.F. (2010). The integrin co-activator Kindlin-3 is expressed and functional in a non-hematopoietic cell, the endothelial cell. J. Biol. Chem. *285*, 18640-18649.

Birmingham, C.L., Canadien, V., Kaniuk, N.A., Steinberg, B.E., Higgins, D.E., and Brumell, J.H. (2008). Listeriolysin O allows Listeria monocytogenes replication in macrophage vacuoles. Nature *451*, 350-354.

Bouchot, A., Millot, J.M., Charpentier, S., Bonhomme, A., Villena, I., Aubert, D., and Pinon, J.M. (2001). Membrane potential changes after infection of monocytes by Toxoplasma gondii. International Journal for Parasitology *31*, 1114–1120.

Borch, J., and Hamann, T. (2009). The nanodisc: a novel tool for membrane protein studies. Biological Chemistry *390*, 805-814.

Borhani, D.W., Rogers, D.P., and Engler, J.A. (1997). Crystal structure of truncated human apolipoprotein AI suggests a lipid-bound conformation (Proceedings of the . . .).

Bouaouina, M., Goult, B.T., Huet-Calderwood, C., Bate, N., Brahme, N.N., Barsukov, I.L., Critchley, D.R., and Calderwood, D.A. (2012). A conserved lipid-binding loop in the kindlin FERM F1 domain is required for kindlin-mediated αIIbβ3 integrin coactivation. J. Biol. Chem. *287*, 6979-6990.

Bourdeau, R.W., Malito, E., Chenal, A., Bishop, B.L., Musch, M.W., Villereal, M.L., Chang, E.B., Mosser, E.M., Rest, R.F., and Tang, W.-J. (2009). Cellular functions and X-ray structure of anthrolysin O, a cholesterol-dependent cytolysin secreted by Bacillus anthracis. Journal of Biological Chemistry *284*, 14645-14656.

Boyd, R.S., Adam, P.J., Patel, S., Loader, J.A., Berry, J., Redpath, N.T., Poyser, H.R., Fletcher, G.C., Burgess, N.A., Stamps, A.C., et al. (2003). Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B. Leukemia *17*, 1605-1612.

Brahme, N.N., and Calderwood, D.A. (2012). Cell Adhesion: A FERM Grasp of the Tail Sorts Out Integrins. Current Biology *22*, R692-R694.

Bray, R.S., and Garnham, P.C.C. (1982). THE LIFE-CYCLE OF PRIMATE MALARIA PARASITES. British Medical Bulletin *38*, 117–122.

Brockhauser, S., Ravelli, R.B.G., McCarthy, A.A., and IUCr (2013). The use of a mini-κ goniometer head in macromolecular crystallography diffraction experiments. Acta Crystallographica. Section D, Biological Crystallography *69*, 1241–1251.

Brown, P.H., and Schuck, P. (2006a). Macromolecular Size-and-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation. Biophysical Journal *90*, 4651–4661.

Butler, M.T., and Wallingford, J.B. (2015). Control of vertebrate core planar cell polarity protein localization and dynamics by Prickle 2. Development (Cambridge, England) *142*, 3429–3439.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. The Journal of Chemical Physics *126*, 014101.

Calderwood, D.A., Zent, R., Grant, R., Rees, D.J., Hynes, R.O., and Ginsberg, M.H. (1999). The Talin head domain binds to integrin beta subunit cytoplasmic tails and regulates integrin activation. Journal of Biological Chemistry *274*, 28071-28074.

Calderwood, D.A., Campbell, I.D., and Critchley, D.R. (2013). Talins and kindlins: partners in integrin-mediated adhesion. Nat. Rev. Mol. Cell Biol. *14*, 503-517.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., et al. (2008). Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature *455*, 757–763.

Carruthers, V.B., and Sibley, L.D. (1999). Mobilization of intracellular calcium stimulates microneme discharge in Toxoplasma gondii. Molecular Microbiology.

Ceccarelli, D.F.J., Blasutig, I.M., Goudreault, M., Li, Z., Ruston, J., Pawson, T., and Sicheri, F. (2007). Non-canonical interaction of phosphoinositides with pleckstrin homology domains of Tiam1 and ArhGAP9. Journal of Biological Chemistry *282*, 13864-13874.

Chang, H., Cahill, H., Smallwood, P.M., Wang, Y., and Nathans, J. (2015). Identification of Astrotactin2 as a Genetic Modifier That Regulates the Global Orientation of Mammalian Hair Follicles. PLoS Genetics *11*, e1005532.

Chatzidaki-Livanis, M., Coyne, M.J., and Comstock, L.E. (2014). An antimicrobial protein of the gut symbiont Bacteroides fragilis with a MACPF domain of host immune proteins. Molecular Microbiology *94*, 1361–1374.

Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica. Section D, Biological Crystallography *66*, 12–21.

Cheng, Y., Grigorieff, N., Penczek, P.A., and Walz, T. (2015). A Primer to Single-Particle Cryo-Electron Microscopy. Cell *161*, 438–449.

Coffey, M.J., Jennison, C., Tonkin, C.J., and Boddey, J.A. (2016). Role of the ER and Golgi in protein export by Apicomplexa. Current Opinion in Cell Biology *41*, 18–24.

Cowtan, K., IUCr (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002-1011.

Critchley, D.R. (2009). Biochemical and Structural Properties of the Integrin-Associated Cytoskeletal Protein Talin. Http://Dx.Doi.org/10.1146/Annurev.Biophys.050708.133744 *38*, 235-254.

Czuczman, M.A., Fattouh, R., van Rijn, J.M., Canadien, V., Osborne, S., Muise, A.M., Kuchroo, V.K., Higgins, D.E., and Brumell, J.H. (2014). Listeria monocytogenes exploits efferocytosis to promote cell-to-cell spread. Nature *509*, 230-234.

De Colibus, L., Sonnen, A.F.-P., Morris, K.J., Siebert, C.A., Abrusci, P., Plitzko, J., Hodnik, V., Leippe, M., Volpi, E., Anderluh, G., et al. (2012). Structures of Lysenin Reveal a Shared Evolutionary Origin for Pore-Forming Proteins And Its Mode of Sphingomyelin Recognition. Structure (London, England : 1993) *20*, 1498–1507.

Conroy, P.J., Yagi, H., Whisstock, J.C., and Norton, R.S. (2015). Structural basis for Ca2 -mediated interaction of the perforin C2 domain with lipid membranes.

Critchley, D.R. (2009). Biochemical and Structural Properties of the Integrin-Associated Cytoskeletal Protein Talin. Dx.doi.org *38*, 235–254.

Cui, C., Chatterjee, B., Lozito, T.P., Zhang, Z., Francis, R.J., Yagi, H., Swanhart, L.M., Sanker, S., Francis, D., Yu, Q., et al. (2013). Wdpcp, a PCP Protein Required for Ciliogenesis, Regulates Directional Cell Migration and Cell Polarity by Direct Modulation of the Actin Cytoskeleton. PLoS Biology *11*, e1001720.

D'Angelo, M.E., Dunstone, M.A., Whisstock, J.C., Trapani, J.A., and Bird, P.I. (2012). Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. BMC Evolutionary Biology 2012 12:1 *12*, 1.

D'Arcy, A., Villard, F., Marsh, M., and IUCr (2007). An automated microseed matrix-screening method for protein crystallization. Acta Crystallographica. Section D, Biological Crystallography *63*, 550–554.

Deligianni, E., Morgan, R.N., Bertuccini, L., Wirth, C.C., Silmon de Monerri, N.C., Spanos, L., Blackman, M.J., Louis, C., Pradel, G., and Siden-Kiamos, I. (2013). A perforin-like protein mediates disruption of the erythrocyte membrane during egress of Plasmodium berghei male gametocytes. Cellular Microbiology *15*, 1438–1455.

Devenport, D. (2014). The cell biology of planar cell polarity. The Journal of Cell Biology *207*, 171–179.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., et al. (2007). In situ proteolysis for protein crystallization and structure determination. Nature Methods *4*, 1019–1021.

Dowd, K.J., Farrand, A.J., and Tweten, R.K. (2012). The cholesterol-dependent cytolysin signature motif: a critical element in the allosteric pathway that couples membrane binding to pore assembly. PLoS Pathog. *8*, e1002787.

Dubey, J.P. (2009). History of the discovery of the life cycle of Toxoplasma gondii. International Journal for Parasitology *39*, 877–882.

Dubey, J.P., Lindsay, D.S., and Speer, C.A. (1998). Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clinical Microbiology Reviews *11*, 267–299.

Dudkina, N.V., Spicer, B.A., Reboul, C.F., Conroy, P.J., Lukoyanova, N., Elmlund, H., Law, R.H.P., Ekkel, S.M., Kondos, S.C., Goode, R.J.A., et al. (2016). Structure of the poly-C9 compo-

nent of the complement membrane attack complex. Nature Communications *7*, 10588.

Ecker, A., Pinto, S.B., Baker, K.W., Kafatos, F.C., and Sinden, R.E. (2007). Plasmodium berghei: Plasmodium perforin-like protein 5 is required for mosquito midgut invasion in Anopheles stephensi. Experimental Parasitology *116*, 504–508.

Edmondson, J.C., Liem, R.K., Kuster, J.E., and Hatten, M.E. (1988). Astrotactin: a novel neuronal cell surface antigen that mediates neuron-astroglial interactions in cerebellar microcultures. The Journal of Cell Biology *106*, 505–517.

Ellett, F., Pase, L., Hayman, J.W., Andrianopoulos, A., and Lieschke, G.J. (2011). mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. Blood *117*, e49-e56.

Ellenbroek, W.G., Wang, Y.-H., Christian, D.A., Discher, D.E., Janmey, P.A., and Liu, A.J. (2011). Divalent Cation-Dependent Formation of Electrostatic PIP2 Clusters in Lipid Monolayers. Biophys. J. *101*, 2178-2184.

Elliott, P.R., Goult, B.T., Kopp, P.M., Bate, N., Grossmann, J.G., Roberts, G.C.K., Critchley, D.R., and Barsukov, I.L. (2010). The Structure of the Talin Head Reveals a Novel Extended Conformation of the FERM Domain. Structure *18*, 1289-1299.

Elsworth, B., Matthews, K., Nie, C.Q., Kalanon, M., Charnaud, S.C., Sanders, P.R., Chisholm, S.A., Counihan, N.A., Shaw, P.J., Pino, P., et al. (2014). PTEX is an essential nexus for protein export in malaria parasites. Nature *511*, 587–591.

El-Rachkidy, R.G., Davies, N.W., and Andrew, P.W. (2008). Pneumolysin generates multiple conductance pores in the membrane of nucleated cells. Biochem. Biophys. Res. Commun. *368*, 786-792.

Emsley, P., Cowtan, K., and IUCr (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica. Section D, Biological Crystallography *60*, 2126–2132.

Endo, T., Tokuda, H., Yagita, K., and Koyama, T. (1987). Effects of Extracellular Potassium on Acid Release and Motility Initiation in Toxoplasma gondii. Journal of Eukaryotic Microbiology *34*, 291–295.

Fairhead, M., and Howarth, M. (2014). Site-Specific Biotinylation of Purified Proteins Using BirA. In Electron Crystallography of Soluble and Membrane Proteins, I. Schmidt-Krey, and Y. Cheng, eds. (New York, NY: Springer New York), pp. 171-184.

Feil, S.C., Ascher, D.B., Kuiper, M.J., Tweten, R.K., and Parker, M.W. (2014). Structural studies of Streptococcus pyogenes streptolysin O provide insights into the early steps of membrane penetration. Journal of Molecular Biology *426*, 785–792.

Felsenstein, J. (1997). An Alternating Least Squares Approach to Inferring Phylogenies from Pairwise Distances. Systematic Biology *46*, 101–111.

Ferguson, K.M., Kavran, J.M., Sankaran, V.G., Fournier, E., Isakoff, S.J., Skolnik, E.Y., and Lemmon, M.A. (2000). Structural Basis for Discrimination of 3-Phosphoinositides by Pleckstrin Homology Domains. Molecular Cell *6*, 373-384.

Ferguson, K.M., Lemmon, M.A., Schlessinger, J., and Sigler, P.B. (1995). Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. Cell *83*, 1037-1046.

Fishell, G., and Hatten, M.E. (1991). Astrotactin provides a receptor system for CNS neuronal migration. Development (Cambridge, England) *113*, 755–765.

Flemming, A. (2015). Malaria: Hitting all stages of the parasite life cycle. Nature Reviews Microbiology *13*, 458–459.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342–346.

Gao, J., Khan, A.A., Shimokawa, T., Zhan, J., Strömblad, S., Fang, W., and Zhang, H. (2013). A feedback regulation between Kindlin-2 and GLI1 in prostate cancer cells. FEBS Letters *587*, 631-638.

Garg, S., Agarwal, S., Kumar, S., Yazdani, S.S., Chitnis, C.E., and Singh, S. (2013). Calcium-dependent permeabilization of erythrocytes by a perforin-like protein during egress of malaria parasites. Nature Communications *4*, 1736.

Garman, E.F. (2014). Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. Science (New York, N.Y.) *343*, 1102–1108.

Gerke, V., and Moss, S.E. (2002). Annexins:

from structure to function. Physiological Reviews *82*, 331–371.

Gilbert, R.J.C. (2002). Pore-forming toxins. Cell. Mol. Life Sci. *59*, 832-844.

Gilbert, R. (2014). Structural features of cholesterol dependent cytolysins and comparison to other MACPF-domain containing proteins. Sub-Cellular Biochemistry *80*, 47–62.

Gilbert, R.J.C. (2005). Inactivation and Activity of Cholesterol-Dependent Cytolysins: What Structural Studies Tell Us. Structure (London, England : 1993) *13*, 1097–1106.

Gilbert, R.J.C. (2016). Protein-lipid interactions and non-lamellar lipidic structures in membrane pore formation and membrane fusion. Biochimica et Biophysica Acta *1858*, 487–499.

Gilbert, R.J.C., Rossjohn, J., Parker, M.W., Tweten, R.K., Morgan, P.J., Mitchell, T.J., Errington, N., Rowe, A.J., Andrew, P.W., and Byron, O. (1998). Self-interaction of pneumolysin, the pore-forming protein toxin of Streptococcus pneumoniae. Journal of Molecular Biology *284*, 1223–1237.

Gilbert, R.J.C., Jiménez, J.L., Chen, S., Tickle, I.J., Rossjohn, J., Parker, M., Andrew, P.W., and Saibil, H.R. (1999). Two Structural Transitions in Membrane Pore Formation by Pneumolysin, the Pore-Forming Toxin of Streptococcus pneumoniae. Cell *97*, 647–655.

Gilbert, R.J.C., Mikelj, M., Serra, M.D., Froelich, C.J., and Anderluh, G. (2013). Effects of MACPF/CDC proteins on lipid membranes. Cellular and Molecular Life Sciences *70*, 2083–2098.

Gilbert, R.J.C., Dalla Serra, M., Froelich, C.J., Wallace, M.I., and Anderluh, G. (2014). Membrane pore formation at protein-lipid interfaces. Trends in Biochemical Sciences *39*, 510–516.

Giousoh, A., Vaz, R., Bryson-Richardson, R.J., Whisstock, J.C., Verkade, H., and Bird, P.I. (2015). Bone morphogenetic protein/retinoic acid inducible neural-specific protein (brinp) expression during Danio rerio development. Gene Expr. Patterns.

Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P., et al. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature *459*, 569–573.

Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., and Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Science *16*, 1569–1576.

Gong, X., An, Z., Wang, Y., Guan, L., Fang, W., Strömblad, S., Jiang, Y., and Zhang, H. (2010). Kindlin-2 controls sensitivity of prostate cancer cells to cisplatin-induced cell death. Cancer Letters *299*, 54-62.

Goñi, G.M., Epifano, C., Boskovic, J., Camacho-Artacho, M., Zhou, J., Bronowska, A., Martín, M.T., Eck, M.J., Kremer, L., Gräter, F., et al. (2014). Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes. Proc. Natl. Acad. Sci. U.S.a. *111*, E3177-E3186.

Goult, B.T., Bouaouina, M., Harburger, D.S., Bate, N., Patel, B., Anthis, N.J., Campbell, I.D., Calderwood, D.A., Barsukov, I.L., Roberts, G.C., et al. (2009). The Structure of the N-Terminus of Kindlin-1: A Domain Important for αIIbβ3 Integrin Activation. J. Mol. Biol. *394*, 944-956.

Gozgit, J.M., Pentecost, B.T., Marconi, S.A., Otis, C.N., Wu, C., and Arcaro, K.F. (2006). Use of an aggressive MCF-7 cell line variant, TMX2-28, to study cell invasion in breast cancer. Mol Cancer Res *4*, 905-913.

Grant, I.H., Gold, J.W.M., Rosenblum, M., Niedzwiecki, D., and Armstrong, D. (1990). Toxoplasma gondii serology in HIV-infected patients: the development of central nervous system toxoplasmosis in AIDS. Aids *4*, 519.

Gray, R.S., Roszko, I., and Solnica-Krezel, L. (2011). Planar Cell Polarity: Coordinating Morphogenetic Cell Behaviors with Embryonic Polarity. Developmental Cell *21*, 120–133.

H, N., JH, G., E, P., W, K., and MA, K. (2001). Negative regulation of G(1)/S transition by the candidate bladder tumour suppressor gene DBCCR1. Oncogene *20*, 2956-2964.

Habuchi, T., Luscombe, M., Elder, P.A., and Knowles, M.A. (1998). Structure and Methylation-Based Silencing of a Gene (DBCCR1) within a Candidate Bladder Cancer Tumor Suppressor Region at 9q32-q33. Genomics *48*, 277-288.

Hadders, M.A., Beringer, D.X., and Gros, P. (2007). Structure of C8 -MACPF Reveals Mechanism of Membrane Attack in Complement Immune Defense. Science (New York, N.Y.) *317*,

1552–1554.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W.A., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., et al. (2005). A Comprehensive Survey of the Plasmodium Life Cycle by Genomic, Transcriptomic, and Proteomic Analyses. Science (New York, N.Y.) *307*, 82–86.

Harburger, D.S., and Calderwood, D.A. (2009). Integrin signalling at a glance. J. Cell. Sci. *122*, 159-163.

Hart, R., Stanley, P., Chakravarty, P., and Hogg, N. (2013). The Kindlin 3 Pleckstrin Homology Domain Has an Essential Role in Lymphocyte Function-associated Antigen 1 (LFA-1) Integrin-mediated B Cell Adhesion and Migration. Journal of Biological Chemistry *288*, 14852-14862.

He, X., Grigg, M.E., Boothroyd, J.C., and Garcia, K.C. (2002). Structure of the immunodominant surface antigen from the Toxoplasma gondii SRS superfamily. Nature Structural & Molecular Biology *9*, 606–611.

Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nature Protocols *2*, 924–932.

Hess, B., Bekker, H., and Berendsen, H. (1997). LINCS: a linear constraint solver for molecular simulations. Journal of Computational Chemistry *18*, 1463-1472.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. *4*, 435-447.

Hemler, M.E. (2014). Tetraspanin proteins promote multiple cancer stages. Nature Reviews. Cancer *14*, 49–60.

Hendrickson, W.A., Pahler, A., Smith, J.L., Satow, Y., Merritt, E.A., and Phizackerley, R.P. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. Proceedings of the National Academy of Sciences *86*, 2190–2194.

Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E.F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A., and Schmitt, B. (2009). PILATUS: A single photon counting pixel detector for X-ray applications. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment *607*, 247–249.

Hu, K., Roos, D.S., and Murray, J.M. (2002). A novel polymer of tubulin forms the conoid of Toxoplasma gondii. The Journal of Cell Biology *156*, 1039–1050.

Hu, K., Johnson, J., Florens, L., Fraunholz, M., Suravajjala, S., DiLullo, C., Yates, J., Roos, D.S., and Murray, J.M. (2006). Cytoskeletal Components of an Invasion Machine—The Apical Complex of Toxoplasma gondii. PLoS Pathogens *2*, e13.

Hua, Z.L., Chang, H., Wang, Y., Smallwood, P.M., and Nathans, J. (2014). Partial interchangeability of Fz3 and Fz6 in tissue polarity signaling for epithelial orientation and axon growth and guidance. Development (Cambridge, England) *141*, 3944–3954.

Huet-Calderwood, C., Brahme, N.N., Kumar, N., Stiegler, A.L., Raghavan, S., Boggon, T.J., and Calderwood, D.A. (2014). Differences in binding to the ILK complex determines kindlin isoform adhesion localization and integrin activation. J. Cell. Sci. *127*, 4308-4321.

Humphries, J.D., Byron, A., and Humphries, M.J. (2006). Integrin ligands at a glance. J. Cell. Sci. *119*, 3901-3903.

Ishino, T., Chinzei, Y., and Yuda, M. (2005). A Plasmodium sporozoite protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection†. Cellular Microbiology *7*, 199–208.

Jian, X., Tang, W.-K., Zhai, P., Roy, N.S., Luo, R., Gruschus, J.M., Yohe, M.E., Chen, P.-W., Li, Y., Byrd, R.A., et al. (2015). Molecular Basis for Cooperative Binding of Anionic Phospholipids to the PH Domain of the Arf GAP ASAP1. Structure *23*, 1977-1988.

Johnson, S., Brooks, N.J., Smith, R.A.G., Lea, S.M., and Bubeck, D. (2013). Structural Basis for Recognition of the Pore-Forming Toxin Intermedilysin by Human Complement Receptor CD59. Cell Rep *3*, 1369-1377.

Johnson, T.K., Henstridge, M.A., Herr, A., Moore, K.A., Whisstock, J.C., and Warr, C.G. (2015). Torso-like mediates extracellular accumulation of Furin-cleaved Trunk to pattern the Drosophila embryo termini. Nature Communications *6*, 8759.

Joyce, M.G., Radaev, S., and Sun, P.D. (2010). A rational approach to heavy-atom derivative screening. Acta Crystallographica. Section D, Biological Crystallography *66*, 358–365.

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

Kafsack, B.F.C., and Carruthers, V.B. (2010). Apicomplexan perforin-like proteins. Communicative & Integrative Biology *3*, 18–23.

Kafsack, B.F.C., Pena, J.D.O., Coppens, I., Ravindran, S., Boothroyd, J.C., and Carruthers, V.B. (2009). Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science (New York, N.Y.) *323*, 530–533.

Kaiser, K. (2004). A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Molecular and Biochemical Parasitology *133*, 15–26.

Kaiser, K., Camargo, N., Coppens, I., Morrisey, J.M., Vaidya, AB, and Kappe, S.H.I. (2004). A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Molecular and Biochemical Parasitology *133*, 15–26.

Kappe, S.H.I., Vaughan, A.M., Boddey, J.A., and Cowman, A.F. (2010). That Was Then But This Is Now: Malaria Research in the Time of an Eradication Agenda. Science (New York, N.Y.) *328*, 862–866.

Kashevarova, A.A., Nazarenko, L.P., Skryabin, N.A., Salyukova, O.A., Chechetkina, N.N., Tolmacheva, E.N., Sazhenova, E.A., Magini, P., Graziano, C., Romeo, G., et al. (2014). Array CGH analysis of a cohort of Russian patients with intellectual disability. Gene *536*, 145–150.

Kawano, H., Nakatani, T., Mori, T., Ueno, S., Fukaya, M., Abe, A., Kobayashi, M., Toda, F., Watanabe, M., and Matsuoka, I. (2004). Identification and characterization of novel developmentally regulated neural-specific proteins, BRINP family. Molecular Brain Research *125*, 60–75.

Kibar, Z., Torban, E., McDearmid, J.R., Reynolds, A., Berghout, J., Mathieu, M., Kirillova, I., De Marco, P., Merello, E., Hayes, J.M., et al. (2007). Mutations in VANGL1Associated with Neural-Tube Defects. New England Journal of Medicine *356*, 1432–1437.

Kalli, A.C., and Sansom, M.S.P. (2014). Interactions of peripheral proteins with model membranes as viewed by molecular dynamics simulations. Biochemical Society Transactions *42*, 1418-1424.

Kalli, A.C., Campbell, I.D., and Sansom, M.S.P. (2013). Conformational Changes in Talin on Binding to Anionic Phospholipid Membranes Facilitate Signaling by Integrin Transmembrane Helices. PLOS Comput Biol *9*, e1003316.

Kato, K., Shiozawa, T., Mitsushita, J., Toda, A., Horiuchi, A., Nikaido, T., Fujii, S., and Konishi, I. (2004). Expression of the mitogen-inducible gene-2 (mig-2) is elevated in human uterine leiomyomas but not in leiomyosarcomas. Human Pathology *35*, 55-60.

Kobayashi, M., Nakatani, T., Koda, T., Matsumoto, K.-I., Ozaki, R., Mochida, N., Takao, K., Miyakawa, T., and Matsuoka, I. (2014). Absence of BRINP1 in mice causes increase of hippocampal neurogenesis and behavioral alterations relevant to human psychiatric disorders. Molecular Brain *7*, 12.

Korchev, Y.E., Bashford, C.L., and Pasternak, C.A. (1992). Differential sensitivity of pneumolysin-induced channels to gating by divalent cations. J. Membarin Biol. *127*, 195-203.

Köster, S., Hudel, M., Chakraborty, T., and Yildiz, Ö. (2013). Crystallization and X-ray crystallographic analysis of the cholesterol-dependent cytolysin listeriolysin O from Listeria monocytogenes. Acta Crystallographica. Section F, Structural Biology and Crystallization Communications *69*, 1212–1215.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., et al. (2005). A protein interaction network of the malaria parasite Plasmodium falciparum. Nature *438*, 103–107.

LaCount, D.J., Schoenfeld, L.W., and Fields, S. (2009). Selection of yeast strains with enhanced expression of Plasmodium falciparum proteins. Molecular and Biochemical Parasitology *163*, 119–122.

Lagal, V., Binder, E.M., Huynh, M.H., Kafsack, B.F.C., Harris, P.K., Diez, R., Chen, D., Cole, R.N., Carruthers, V.B., and Kim, K. (2010). Toxoplasma gondii protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites. Cellular Microbiology *12*, 1792–1808.

Lai, C.-L., Srivastava, A., Pilling, C., Chase,

A.R., Falke, J.J., and Voth, G.A. (2013). Molecular mechanism of membrane binding of the Grp1 PH domain. J. Mol. Biol. *425*, 3073-3090.

Law, R.H.P., Lukoyanova, N., Voskoboinik, I., Caradoc-Davies, T.T., Baran, K., Dunstone, M.A., D'Angelo, M.E., Orlova, E.V., Coulibaly, F., Verschoor, S., et al. (2010a). The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature *468*, 447–451.

Law, R.H.P., Lukoyanova, N., Voskoboinik, I., Caradoc-Davies, T.T., Baran, K., Dunstone, M.A., D'Angelo, M.E., Orlova, E.V., Coulibaly, F., Verschoor, S., et al. (2010b). The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature *468*, 447–451.

Lawrence, S.L., Feil, S.C., Morton, C.J., Farrand, A.J., Mulhern, T.D., Gorman, M.A., Wade, K.R., Tweten, R.K., and Parker, M.W. (2015a). Crystal structure of Streptococcus pneumoniae pneumolysin provides key insights into early steps of pore formation. Scientific Reports *5*, 14352.

Lemmon, M.A. (2008). Membrane recognition by phospholipid-binding domains. Nature Reviews. Molecular Cell Biology *9*, 99–111.

Lemmon, M.A., Ferguson, K.M., and Abrams, C.S. (2001). Pleckstrin homology domains and the cytoskeleton. FEBS Letters *513*, 71-76.

Lesch, K.-P., Timmesfeld, N., Renner, T.J., Halperin, R., Röser, C., Nguyen, T.T., Craig, D.W., Romanos, J., Heine, M., Meyer, J., et al. (2008). Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. Journal of Neural Transmission (Vienna, Austria : 1996) *115*, 1573–1585.

Leung, C., Dudkina, N.V., Lukoyanova, N., Hodel, A.W., Farabella, I., Pandurangan, A.P., Jahan, N., Pires Damaso, M., Osmanović, D., Reboul, C.F., et al. (2014). Stepwise visualization of membrane pore formation by suilysin, a bacterial cholesterol-dependent cytolysin. Elife *3*, e04247.

Lionel, A.C., Tammimies, K., Vaags, A.K., Rosenfeld, J.A., Ahn, J.W., Merico, D., Noor, A., Runke, C.K., Pillalamarri, V.K., Carter, M.T., et al. (2014). Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Human Molecular Genetics *23*, 2752–2768.

Liu, Q., and Hendrickson, W.A. (2015). Crystallographic phasing from weak anomalous signals. Current Opinion in Structural Biology *34*, 99–107.

Liu, Q., Zhang, Z., and Hendrickson, W.A. (2011). Multi-crystal anomalous diffraction for low-resolution macromolecular phasing. Acta Crystallographica. Section D, Biological Crystallography *67*, 45–59.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F., and Hendrickson, W.A. (2012). Structures from Anomalous Diffraction of Native Biological Macromolecules. Science (New York, N.Y.) *336*, 1033–1037.

Liu, Q., Liu, Q., and Hendrickson, W.A. (2013). Robust structural analysis of native biological macromolecules from multi-crystal anomalous diffraction data. Acta Crystallographica. Section D, Biological Crystallography *69*, 1314–1332.

Liu, Y., Zhu, Y., Ye, S., and Zhang, R. (2012). Crystal structure of kindlin-2 PH domain reveals a conformational transition for its membrane anchoring and regulation of integrin activation. Protein Cell *3*, 434-440.

Lukoyanova, N., Kondos, S.C., Farabella, I., Law, R.H.P., Reboul, C.F., Caradoc-Davies, T.T., Spicer, B.A., Kleifeld, O., Traore, D.A.K., Ekkel, S.M., et al. (2015). Conformational changes during pore formation by the perforin-related protein pleurotolysin. PLoS Biology *13*, e1002049.

Lumb, C.N., He, J., Xue, Y., Stansfeld, P.J., Stahelin, R.V., Kutateladze, T.G., and Sansom, M.S.P. (2011). Biophysical and Computational Studies of Membrane Penetration by the Grp1 Pleckstrin Homology Domain. Structure *19*, 1338-1346.

Luo, B.-H., Carman, C.V., Takagi, J., and Springer, T.A. (2005). Disrupting integrin transmembrane domain heterodimerization increases ligand binding affinity, not valency or clustering. Proceedings of the National Academy of Sciences *102*, 3679–3684.

Malinin, N.L., Zhang, L., Choi, J., Ciocea, A., Razorenova, O., Ma, Y.-Q., Podrez, E.A., Tosi, M., Lennon, D.P., Caplan, A.I., et al. (2009). A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans. Nature Medicine *15*, 313-318.

Mans, B.J., Pienaar, R., and Latif, A.A. (2015). A review of Theileria diagnostics and epidemiology. International Journal for Parasitology: Parasites and Wildlife *4*, 104–118.

McCormack, R., and Podack, E.R. (2015). Perforin-2/Mpeg1 and other pore-forming proteins throughout evolution. J. Leukoc. Biol. *98*, 761-768.

McCormack, R.M., Lyapichev, K., Olsson, M.L., Podack, E.R., and Munson, G.P. (2015a). Enteric pathogens deploy cell cycle inhibiting factors to block the bactericidal activity of Perforin-2. eLife *4*, 263.

McCormack, R.M., de Armas, L.R., Shiratsuchi, M., Fiorentino, D.G., Olsson, M.L., Lichtenheld, M.G., Morales, A., Lyapichev, K., Gonzalez, L.E., Strbo, N., et al. (2015b). Perforin-2 is essential for intracellular defense of parenchymal cells and phagocytes against pathogenic bacteria. eLife *4*, e06508.

Merte, J., Jensen, D., Wright, K., Sarsfield, S., Wang, Y., Schekman, R., and Ginty, D.D. (2010). Sec24b selectively sorts Vangl2 to regulate planar cell polarity during neural tube closure. Nature Cell Biology *12*, 41.

Metkar, S.S., Marchioretto, M., Antonini, V., Lunelli, L., Wang, B., Gilbert, R.J., Anderluh, G., Roth, R., Pooga, M., Pardo, J., et al. (2015). Perforin oligomers form arcs in cellular membranes: a locus for intracellular delivery of granzymes. Cell Death & Differentiation *22*, 74-85.

Mineo, A., Furriols, M., and Casanova, J. (2015). Accumulation of the Drosophila Torso-like protein at the blastoderm plasma membrane suggests that it translocates from the eggshell. Development *142*, 1299-1304.

Mondragon, R., and Frixione, E. (1996). Ca2+-Dependence of Conoid Extrusion in Toxoplasma gondii Tachyzoites. Journal of Eukaryotic Microbiology *43*, 120–127.

Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., and Marrink, S.-J. (2008). The MARTINI Coarse-Grained Force Field: Extension to Proteins. J. Chem. Theory Comput. *4*, 819-834.

Moravcevic, K., Oxley, C.L., and Lemmon, M.A. (2012). Conditional Peripheral Membrane Proteins: Facing up to Limited Specificity. Structure *20*, 15-27.

Mordue, D.G., Håkansson, S., Niesman, I., and David Sibley, L. (1999). Toxoplasma gondii Resides in a Vacuole That Avoids Fusion with Host Cell Endocytic and Exocytic Vesicular Trafficking Pathways. Experimental Parasitology *92*, 87–99.

Morita-Yamamuro, C., Tsutsui, T., Sato, M., Yoshioka, H., Tamaoki, M., Ogawa, D., Matsuura, H., Yoshihara, T., Ikeda, A., Uyeda, I., et al. (2005). The Arabidopsis gene CAD1 controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. Plant Cell Physiol. *46*, 902-912.

Morrison, D.A. (2009). Evolution of the Apicomplexa: where are we now? Trends in Parasitology *25*, 375–382.

Mory, A., Feigelson, S.W., Yarali, N., Kilic, S.S., Bayhan, G.I., Gershoni-Baruch, R., Etzioni, A., and Alon, R. (2008). Kindlin-3: a new gene involved in the pathogenesis of LAD-III. *112*, 2591-2591.

Moser, M., Legate, K.R., Zent, R., and Fässler, R. (2009). The Tail of Integrins, Talin, and Kindlins. Science *324*, 895-899.

Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M., and Fässler, R. (2008). Kindlin-3 is essential for integrin activation and platelet aggregation. Nature Medicine *14*, 325-330.

Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R., and Ban, N. (2009). The structure of a cytolytic |[agr]|-helical toxin pore reveals its assembly mechanism. Nature *459*, 726–730.

Mueller, M., Wang, M., and Schulze-Briese, C. (2012). Optimal fine φ-slicing for single-photon-counting pixel detectors. Acta Crystallographica. Section D, Biological Crystallography *68*, 42–56.

Mulvihill, E., van Pee, K., Mari, S.A., Müller, D.J., and Yildiz, Ö. (2015). Directly Observing the Lipid-Dependent Self-Assembly and Pore-Forming Mechanism of the Cytolytic Toxin Listeriolysin O. Nano Lett. *15*, 6965-6973.

Naughton, F.B., Kalli, A.C., and Sansom, M.S.P. (2016). Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of Grp1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers. J. Phys. Chem. Lett. *7*, 1219-1224.

Nettleship, J.E., Brown, J., Groves, M.R., and Geerlof, A. (2008). Methods for Protein Characterization by Mass Spectrometry, Thermal Shift (ThermoFluor) Assay, and Multiangle or Static Light Scattering. Electron Crystallography of Soluble and Membrane Proteins *426*, 299–318.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nature Reviews. Molecular Cell Biology *13*, 767–779.

Noutoshi, Y., Kuromori, T., Wada, T., Hirayama, T., Kamiya, A., Imura, Y., Yasuda, M., Nakashita, H., Shirasu, K., and Shinozaki, K. (2006). Loss of Necrotic Spotted Lesions 1 associates with cell death and defense responses in Arabidopsis thaliana. *62*, 29-42.

Omari, K. El, Iourin, O., Kadlec, J., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., and Stuart, D.I. (2014a). Pushing the limits of sulfur SAD phasing: de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. Acta Crystallographica. Section D, Biological Crystallography *70*, 2197–2203.

Omari, K. El, Iourin, O., Kadlec, J., Sutton, G., Harlos, K., Grimes, J.M., and Stuart, D.I. (2014b). Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. Nature Communications *5*, 4874.

Orłowski, A., Kukkurainen, S., Pöyry, A., Rissanen, S., Vattulainen, I., Hytönen, V.P., and Róg, T. (2015). PIP2 and Talin Join Forces to Activate Integrin. J. Phys. Chem. B *119*, 12381-12389.

Palmer, M., Harris, R., Freytag, C., Kehoe, M., Jensen, J.T., and Bhakdi, S. (1998). Assembly mechanism of the oligomeric streptolysin O pore: the early membrane lesion is lined by a free edge of the lipid membrane and is extended gradually during oligomerization. Embo J. *17*, 1598-1605.

Pape, T., Schneider, T.R., and IUCr (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. Journal of Applied Crystallography *37*, 843–844.

Peraro, M.D., and van der Goot, F.G. (2016). Pore-forming toxins: ancient, but never really out of fashion. Nature Reviews Microbiology *14*, 77-92.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics *52*, 7182-7190.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., Svergun, D.I., et al. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography *45*, 342–350.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *25*, 1605–1612.

Pilling, C., Landgraf, K.E., and Falke, J.J. (2011). The Grp1 PH domain, like the AKT1 PH domain, possesses a sentry glutamate residue essential for specific targeting to plasma membrane PI(3,4,5)P(3). Biochemistry *50*, 9845-9856.

Podobnik, M., Marchioretto, M., Zanetti, M., Bavdek, A., Kisovec, M., Cajnko, M.M., Lunelli, L., Serra, M.D., and Anderluh, G. (2015). Plasticity of Listeriolysin O Pores and its Regulation by pH and Unique Histidine. Sci Rep *5*, 9623.

Praper, T., Sonnen, A.F.-P., Kladnik, A., Andrighetti, A.O., Viero, G., Morris, K.J., Volpi, E., Lunelli, L., Dalla Serra, M., Froelich, C.J., et al. (2011a). Perforin activity at membranes leads to invaginations and vesicle formation. Proc. Natl. Acad. Sci. U.S.a. *108*, 21016-21021.

Praper, T., Sonnen, A., Viero, G., Kladnik, A., Froelich, C.J., Anderluh, G., Dalla Serra, M., and Gilbert, R.J.C. (2011b). Human perforin employs different avenues to damage membranes. J. Biol. Chem. *286*, 2946-2955.

Price, M., Lang, M.G., Frank, A.T., Goetting-Minesky, M.P., Patel, S.P., Silviera, M.L., Krady, J.K., Milner, R.J., Ewing, A.G., and Day, J.R. (2003). Seven cDNAs enriched following hippocampal lesion: possible roles in neuronal responses to injury. Brain Research. Molecular Brain Research *117*, 58–67.

Raveh, B., Enosh, A., Schueler-Furman, O., and Halperin, D. (2009). Rapid Sampling of Molecular Motions with Prior Information Constraints. PLoS Comput Biol *5*, e1000295.

Riffel, N., Harlos, K., Iourin, O., Rao, Z., Kingsman, A., Stuart, D., and Fry, E. (2002). Atomic Resolution Structure of Moloney Murine Leukemia Virus Matrix Protein and Its Relationship to Other Retroviral Matrix Proteins. Structure (London, England : 1993) *10*, 1627–1636.

Risco-Castillo, V., Topçu, S., Marinach, C., Manzoni, G., Bigorgne, A.E., Briquet, S., Baudin, X., Lebrun, M., Dubremetz, J.-F., and Silvie, O. (2015). Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. Cell Host & Microbe *18*, 593–603.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new

ENDscript server. Nucleic Acids Research *42*, NaN – NaN.

Roiko, M.S., and Carruthers, V.B. (2009). New roles for perforins and proteases in apicomplexan egress. Cellular Microbiology *11*, 1444–1452.

Roiko, M.S., and Carruthers, V.B. (2013). Functional dissection of Toxoplasma gondii perforin-like protein 1 reveals a dual domain mode of membrane binding for cytolysis and parasite egress. The Journal of Biological Chemistry *288*, 8712–8725.

Roiko, M.S., Svezhova, N., and Carruthers, V.B. (2014). Acidification Activates Toxoplasma gondii Motility and Egress by Enhancing Protein Secretion and Cytolytic Activity. PLoS Pathogens *10*, e1004488.

Rosado, C.J., Buckle, A.M., Law, R.H.P., Butcher, R.E., Kan, W.-T., Bird, C.H., Ung, K., Browne, K.A., Baran, K., Bashtannyk-Puhalovich, T.A., et al. (2007). A common fold mediates vertebrate defense and bacterial attack. Science *317*, 1548-1551.

Rosado, C.J., Kondos, S., Bull, T.E., Kuiper, M.J., Law, R.H.P., Buckle, A.M., Voskoboinik, I., Bird, P.I., Trapani, J.A., Whisstock, J.C., et al. (2008). The MACPF/CDC family of pore-forming toxins. Cell. Microbiol. *10*, 1765-1774.

Rossjohn, J., Feil, S.C., McKinstry, W.J., Tweten, R.K., and Parker, M.W. (1997). Structure of a Cholesterol-Binding, Thiol-Activated Cytolysin and a Model of Its Membrane Form. Cell *89*, 685–692.

Roszko, I., S Sepich, D., Jessen, J.R., Chandrasekhar, A., and Solnica-Krezel, L. (2015). A dynamic intracellular distribution of Vangl2 accompanies cell polarization during zebrafish gastrulation. Development (Cambridge, England) *142*, 2508–2520.

Ryan, U. (2010). Cryptosporidium in birds, fish and amphibians. Experimental Parasitology *124*, 113–120.

Savant-Bhonsale, S., and Montell, D.J. (1993). torso-like encodes the localized determinant of Drosophila terminal pattern formation. Genes Dev. *7*, 2548-2555.

Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Research *38*, NaN – NaN.

Serna, M., Giles, J.L., Morgan, B.P., and Bubeck, D. (2016). Structural basis of complement membrane attack complex formation. Nat Commun *7*, 10587.

Schmidt, S., Nakchbandi, I., Ruppert, R., Kawelke, N., Hess, M.W., Pfaller, K., Jurdic, P., Fässler, R., and Moser, M. (2011). Kindlin-3-mediated signaling from multiple integrin classes is required for osteoclast-mediated bone resorption. J. Cell Biol. *192*, 883-897.

Schmidt, T., Suk, J.-E., Ye, F., Situ, A.J., Mazumder, P., Ginsberg, M.H., and Ulmer, T.S. (2015). Annular anionic lipids stabilize the integrin αIIbβ3 transmembrane complex. J. Biol. Chem. *290*, 8283-8293.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. Biophysical Journal *105*, 962–974.

Scholtyseck, E., and Mehlhorn, H. (1970). Ultrastructural study of characteristic organelles (paired organelles, micronemes, micropores) of sporozoa and related organisms. Zeitschrift Für Parasitenkunde *34*, 97–127.

Schürpf, T., Chen, Q., Liu, J., Wang, R., Springer, T.A., and Wang, J. (2012). The RGD finger of Del-1 is a unique structural feature critical for integrin binding. FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology *26*, 3412–3420.

Serna, M., Giles, J.L., Morgan, B.P., and Bubeck, D. (2016). Structural basis of complement membrane attack complex formation. Nature Communications *7*, 10587.

Shan, D., Qian, W., Liu, J., Liu, R., and Liu, Q. (2015). Identification and co-localization of perforin-like (TgPLP1) protein in Toxoplasma gondii bradyzoites. Experimental Parasitology *153*, 39–44.

Sharma, P., and Chitnis, C.E. (2013). Key molecular events during host cell invasion by Apicomplexan pathogens. Current Opinion in Microbiology *16*, 432–437.

Shattil, S.J., Kim, C., and Ginsberg, M.H. (2010). The final steps of integrin activation: the end game. Nat. Rev. Mol. Cell Biol. *11*, 288-300.

Sheiner, L., and Striepen, B. (2014). The Apicoplast: A Parasite's Symbiont. Plastid Biology 209–238.

Sheldrick, G.M., and IUCr (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallographica. Section D, Biological Crystallography *66*, 479–485.

Shen, Z., Ye, Y., Dong, L., Vainionpää, S., Mustonen, H., Puolakkainen, P., and Wang, S. (2012). Kindlin-2: a novel adhesion protein related to tumor invasion, lymph node metastasis, and patient outcome in gastric cancer. The American Journal of Surgery *203*, 222-229.

Shepard, L.A., Heuck, A.P., Hamman, B.D., Rossjohn, J., Parker, M.W., Ryan, K.R., Johnson, A.E., and Tweten, R.K. (1998). Identification of a Membrane-Spanning Domain of the Thiol-Activated Pore-Forming Toxin Clostridium perfringensPerfringolysin O: An α-Helical to β-Sheet Transition Identified by Fluorescence Spectroscopy †. Biochemistry *37*, 14563-14574.

Shorts-Cary, L., Xu, M., Ertel, J., Kleinschmidt-Demasters, B.K., Lillehei, K., Matsuoka, I., Nielsen-Preiss, S., and Wierman, M.E. (2007). Bone morphogenetic protein and retinoic acid-inducible neural specific protein-3 is expressed in gonadotrope cell pituitary adenomas and induces proliferation, migration, and invasion. Endocrinology *148*, 967-975.

Shimizu, T., Vassylyev, D.G., Kido, S., Doi, Y., and Morikawa, K. (1994). Crystal structure of vitelline membrane outer layer protein I (VMO-I): a folding motif with homologous Greek key structures related by an internal three-fold symmetry. The EMBO Journal *13*, 1003–1010.

Sibley, L.D., Weidner, E., and Krahenbuhl, J.L. (1985). Phagosome acidification blocked by intracellular Toxoplasma gondii. , Published Online: 30 May 1985; | doi:10.1038/315416a0 *315*, 416–419.

Siegel, D.H., Ashton, G.H.S., Penagos, H.G., Lee, J.V., Feiler, H.S., Wilhelmsen, K.C., South, A.P., Smith, F.J.D., Prescott, A.R., Wessagowit, V., et al. (2003). Loss of Kindlin-1, a Human Homolog of the Caenorhabditis elegans Actin-Extracellular-Matrix Linker Protein UNC-112, Causes Kindler Syndrome. The American Journal of Human Genetics *73*, 174-187.

Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, A., and de Vries, A.H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations (American Chemical Society).

Sin, S., Bonin, F., Petit, V., Meseure, D., Lallemand, F., Bièche, I., Bellahcène, A., Castronovo, V., de Wever, O., Gespach, C., et al. (2011). Role of the Focal Adhesion Protein Kindlin-1 in Breast Cancer Growth and Lung Metastasis. JNCI J Natl Cancer Inst *103*, 1323-1337.

Slade, D.J., Lovelace, L.L., Chruszcz, M., Minor, W., Lebioda, L., and Sodetz, J.M. (2008). Crystal Structure of the MACPF Domain of Human Complement Protein C8α in Complex with the C8γ Subunit. Journal of Molecular Biology *379*, 331–342.

Solecki, D.J. (2012). Sticky situations: recent advances in control of cell adhesion during neuronal migration. Current Opinion in Neurobiology *22*, 791–798.

Song, H., Hu, J., Chen, W., Elliott, G., Andre, P., Gao, B., and Yang, Y. (2010). Planar cell polarity breaks bilateral symmetry by controlling ciliary positioning. Nature *466*, 378–382.

Song, S.J., Poliseno, L., Song, M.S., Ala, U., Webster, K., Ng, C., Beringer, G., Brikbak, N.J., Yuan, X., Cantley, L.C., et al. (2013). MicroRNA-Antagonism Regulates Breast Cancer Stemness and Metastasis via TET-Family-Dependent Chromatin Remodeling. Cell *154*, 311-324.

Sonnen, A.F.-P., and Henneke, P. (2014). Structural Biology of the Membrane Attack Complex. In MACPF/CDC Proteins - Agents of Defence, Attack and Invasion, G. Anderluh, and R. Gilbert, eds. (Dordrecht: Springer Netherlands), pp. 83-116.

Sonnen, A.F.-P., Plitzko, J.M., and Gilbert, R.J.C. (2014). Incomplete pneumolysin oligomers form membrane pores. Open Biology *4*, 140044-140044.

Sossey-Alaoui, K., Pluskota, E., Davuluri, G., Bialkowska, K., Das, M., Szpak, D., Lindner, D.J., Downs-Kelly, E., Thompson, C.L., and Plow, E.F. (2014). Kindlin-3 enhances breast cancer progression and metastasis by activating Twist-mediated angiogenesis. Faseb J. *28*, 2260-2271.

Spillman, N.J., Beck, J.R., and Goldberg, D.E. (2015). Protein Export into Malaria Parasite–Infected Erythrocytes: Mechanisms and Functional Consequences. Dx.doi.org *84*, 813–841.

Stansfeld, P.J., and Sansom, M.S.P. (2011). From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. J. Chem. Theory Comput. *7*, 1157-1166.

Stitt, T.N., and Hatten, M.E. (1990). Antibodies that recognize astrotactin block granule neuron

binding to astroglia. Neuron *5*, 639–649.

Stephens, K.E., Miller, K.Y., and Miller, B.L. (1999). Functional Analysis of DNA Sequences Required for Conidium-Specific Expression of the SpoC1-C1C Gene of Aspergillus nidulans. Fungal Genetics and Biology *27*, 231-242.

Stuart, D.I., Levine, M., Muirhead, H., and Stammers, D.K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. Journal of Molecular Biology *134*, 109–142.

Suhre, K., and Sanejouand, Y.-H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Research *32*, NaN – NaN.

Svergun, D.I., and IUCr (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography *25*, 495–503.

Talaat, S., Somji, S., Toni, C., Garrett, S.H., Zhou, X.D., Sens, M.A., and Sens, D.A. (2011). Kindlin-2 Expression in Arsenite- and Cadmium-transformed Bladder Cancer Cell Lines and in Archival Specimens of Human Bladder Cancer. Urology *77*, 1507.e1-1507.e7.

Tavares, J., Amino, R., and Ménard, R. (2014). The role of MACPF proteins in the biology of malaria and other apicomplexan parasites. Sub-Cellular Biochemistry *80*, 241–253.

Tenter, A.M., Heckeroth, A.R., and Weiss, L.M. (2000). Toxoplasma gondii: from animals to humans. International Journal for Parasitology *30*, 1217–1258.

Terashima, M., Kobayashi, M., Motomiya, M., Inoue, N., Yoshida, T., Okano, H., Iwasaki, N., Minami, A., and Matsuoka, I. (2010). Analysis of the expression and function of BRINP family genes during neuronal differentiation in mouse embryonic stem cell-derived neural stem cells. J. Neurosci. Res. *88*, 1387-1393.

Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. Journal of Synchrotron Radiation.

Thiery, J., Keefe, D., Boulant, S., Boucrot, E., Walch, M., Martinvalet, D., Goping, I.S., Bleackley, R.C., Kirchhausen, T., and Lieberman, J. (2011). Perforin pores in the endosomal membrane trigger the release of endocytosed granzyme B into the cytosol of target cells. Nat. Immunol. *12*, 770-777.

Tilley, S.J., Orlova, E.V., Gilbert, R.J.C., Andrew, P.W., and Saibil, H.R. (2005). Structural basis of pore formation by the bacterial toxin pneumolysin. Cell *121*, 247–256.

Torban, E., Patenaude, A.-M., Leclerc, S., Rakowiecki, S., Gauthier, S., Andelfinger, G., Epstein, D.J., and Gros, P. (2008). Genetic interaction between members of the Vangl family causes neural tube defects in mice. Proceedings of the National Academy of Sciences of the United States of America *105*, 3449–3454.

Torgerson, P.R., and Mastroiacovo, P. (2013). The global burden of congenital toxoplasmosis: a systematic review. Bulletin of the World Health Organization *91*, 501–508.

Tschopp, J., Masson, D., and Stanley, K.K. (1986). Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature *322*, 831-834.

Tweten, R.K. (2005). Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. Infection and ... *73*, 6199–6209.

Tweten, R.K., Hotze, E.M., and Wade, K.R. (2015). The Unique Molecular Choreography of Giant Pore Formation by the Cholesterol-Dependent Cytolysins of Gram-Positive Bacteria. Annu. Rev. Microbiol. 2015. 69:323-40.

Ussar, S., Wang, H.-V., Linder, S., Fässler, R., and Moser, M. (2006). The Kindlins: Subcellular localization and expression during murine development. Experimental Cell Research *312*, 3142-3151.

van den Bogaart, G., Meyenberg, K., Risselada, H.J., Amin, H., Willig, K.I., Hubrich, B.E., Dier, M., Hell, S.W., Grubmüller, H., Diederichsen, U., et al. (2011). Membrane protein sequestering by ionic protein-lipid interactions. Nature *479*, 552-555.

van Meer, G., Voelker, D.R., and Feigenson, G.W. (2008). Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. *9*, 112-124.

Vonkova, I., Saliba, A.-E., Deghou, S., Anand, K., Ceschia, S., Doerks, T., Galih, A., Kugler, K.G., Maeda, K., Rybin, V., et al. (2015b). Lipid Cooperativity as a General Membrane-Recruitment Principle for PH Domains. Cell Rep *12*, 1519-1530.

Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A.,

and Veltman, J.A. (2008). Recurrent CNVs Disrupt Three Candidate Genes in Schizophrenia Patients. The American Journal of Human Genetics *83*, 504–510.

Wade, K.R., and Tweten, R.K. (2015). The Apicomplexan CDC/MACPF-like pore-forming proteins. Current Opinion in Microbiology *26*, 48–52.

Walck-Shannon, E., and Hardin, J. (2014). Cell intercalation from top to bottom. Nature Reviews. Molecular Cell Biology *15*, 34–48.

Walter R P Scott, Philippe H Hünenberger, Ilario G Tironi, Alan E Mark, Salomon R Billeter, Jens Fennen, Andrew E Torda, Thomas Huber, Peter Krüger, A., and van Gunsteren, W.F. (1999). The GROMOS Biomolecular Simulation Program Package (American Chemical Society).

Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. Structure (London, England : 1993) *14*, 1617–1622.

Wang, J., Hamblet, N.S., Mark, S., Dickinson, M.E., Brinkman, B.C., Segil, N., Fraser, S.E., Chen, P., Wallingford, J.B., and Wynshaw-Boris, A. (2006a). Dishevelled genes mediate a conserved mammalian PCP pathway to regulate convergent extension during neurulation. Development (Cambridge, England) *133*, 1767–1778.

Wang, K.-S., Tonarelli, S., Luo, X., Wang, L., Su, B., Zuo, L., Mao, C., Rubin, L., Briones, D., and Xu, C. (2014). Polymorphisms within ASTN2 gene are associated with age at onset of Alzheimer's disease. J Neural Transm.

Wang, X., Mu, Z., Li, Y., Bi, Y., and Wang, Y. (2015). Smaller Nanodiscs are Suitable for Studying Protein Lipid Interactions by Solution NMR. Protein J *34*, 205-211.

Wang, Y., and Nathans, J. (2007). Tissue/planar cell polarity in vertebrates: new insights and new questions. Development (Cambridge, England) *134*, 647–658.

Wang, Y., Chang, H., and Nathans, J. (2010). When whorls collide: the development of hair patterns in frizzled 6 mutant mice. Development (Cambridge, England) *137*, 4091–4099.

Wang, Y., Guo, N., and Nathans, J. (2006b). The role of Frizzled3 and Frizzled6 in neural tube closure and in the planar polarity of inner-ear sensory hair cells. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience *26*, 2147–2156.

Weinstein, E.J., Bourner, M., Head, R., Zakeri, H., Bauer, C., and Mazzarella, R. (2003). URP1: a member of a novel family of PH and FERM domain-containing membrane-associated proteins is significantly over-expressed in lung and colon carcinomas. Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease *1637*, 207-216.

Wetzel, D.M., Chen, L.A., Ruiz, F.A., Moreno, S.N.J., and Sibley, L.D. (2004). Calcium-mediated protein secretion potentiates motility in Toxoplasma gondii. Journal of Cell Science *117*, 5739–5748.

Wilson, P.M., Fryer, R.H., Fang, Y., and Hatten, M.E. (2010). Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience *30*, 8529–8540.

Winter, G. (2009). xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography *43*, 186–190.

Wirth, C.C., Glushakova, S., Scheuermayer, M., Repnik, U., Garg, S., Schaack, D., Kachman, M.M., Weißbach, T., Zimmerberg, J., Dandekar, T., et al. (2014). Perforin-like protein PPLP2 permeabilizes the red blood cell membrane during egress of Plasmodium falciparum gametocytes. Cellular Microbiology *16*, 709–733.

Wirth, C.C., Bennink, S., Scheuermayer, M., Fischer, R., and Pradel, G. (2015). Perforin-like protein PPLP4 is crucial for mosquito midgut infection by Plasmodium falciparum. Molecular and Biochemical Parasitology *201*, 90–99.

World Health Organization (2015). World malaria report 2014. 2014.

Wright, K.O., Messing, E.M., and Reeder, J.E. (2004). DBCCR1 mediates death in cultured bladder tumor cells. Oncogene *23*, 82-90.

Xiong, J.-P., Stehle, T., Goodman, S.L., and Arnaout, M.A. (2004). A novel adaptation of the integrin PSI domain revealed from its crystal structure. Journal of Biological Chemistry *279*, 40252–40254.

Xu, L., Huang, B., Du, H., Zhang, X.C., Xu, J., Li, X., and Rao, Z. (2010). Crystal structure of cytotoxin protein suilysin from Streptococcus

suis. Protein Cell *1*, 96-105.

Xu, Z., Chen, X., Zhi, H., Gao, J., Bialkowska, K., Byzova, T.V., Pluskota, E., White, G.C., Liu, J., Plow, E.F., et al. (2014). Direct interaction of kindlin-3 with integrin αIIbβ3 in platelets is required for supporting arterial thrombosis in mice. Arterioscler. Thromb. Vasc. Biol. *34*, 1961-1967.

Yagi, H., Conroy, P.J., Leung, E.W.W., Law, R.H.P., Trapani, J.A., Voskoboinik, I., Whisstock, J.C., and Norton, R.S. (2015b). Structural Basis for Ca2+-mediated Interaction of the Perforin C2 Domain with Lipid Membranes. J. Biol. Chem. *290*, 25213-25226.

Yates, L.A., Durrant, B.P., Barber, M., Harlos, K., Fleurdépine, S., Norbury, C.J., Gilbert, R.J.C., and IUCr (2015). Improved crystallization and diffraction of caffeine-induced death suppressor protein 1 (Cid1). Acta Crystallographica. Section F, Structural Biology and Crystallization Communications *71*, 346–353.

Yates, L.A., Füzéry, A.K., Bonet, R., Campbell, I.D., and Gilbert, R.J.C. (2012a). Biophysical analysis of Kindlin-3 reveals an elongated conformation and maps integrin binding to the membrane-distal β-subunit NPXY motif. J. Biol. Chem. *287*, 37715-37731.

Yates, L.A., Füzéry, A.K., Bonet, R., Campbell, I.D., and Gilbert, R.J.C. (2012b). Biophysical analysis of Kindlin-3 reveals an elongated conformation and maps integrin binding to the membrane-distal β-subunit NPXY motif. J. Biol. Chem. *287*, 37715-37731.

Yates, L.A., Lumb, C.N., Brahme, N.N., Zalyte, R., Bird, L.E., De Colibus, L., Owens, R.J., Calderwood, D.A., Sansom, M.S.P., and Gilbert, R.J.C. (2012c). Structural and functional characterization of the kindlin-1 pleckstrin homology domain. J. Biol. Chem. *287*, 43246-43261.

Yates, L.A., Norbury, C.J., and Gilbert, R.J.C. (2013). The Long and Short of MicroRNA. Cell *153*, 516-519.

Yook, J.I., Li, X.-Y., Ota, I., Hu, C., Kim, H.S., Kim, N.H., Cha, S.Y., Ryu, J.K., Choi, Y.J., Kim, J., et al. (2006). A Wnt-Axin2-GSK3β cascade regulates Snail1 activity in breast cancer cells. Nature Cell Biology *8*, 1398-1406.

Yu, Y., Wu, J., Guan, L., Qi, L., Tang, Y., Ma, B., Zhan, J., Wang, Y., Fang, W., and Zhang, H. (2013). Kindlin 2 promotes breast cancer invasion via epigenetic silencing of the microRNA200 gene family. International Journal of Cancer *133*, 1368-1379.

Yu, Y., Wu, J., Wang, Y., Zhao, T., Ma, B., Liu, Y., Fang, W., Zhu, W.G., and Zhang, H. (2012). Kindlin 2 forms a transcriptional complex with β-catenin and TCF4 to enhance Wnt signalling. EMBO Reports *13*, 750-758.

Zakrzewska, A., Cui, C., Stockhammer, O.W., Benard, E.L., Spaink, H.P., and Meijer, A.H. (2010). Macrophage-specific gene functions in Spi1-directed innate immunity. Blood *116*, e1-e11.

Zallen, J.A. (2007). Planar Polarity and Tissue Morphogenesis. Cell *129*, 1051–1063.

Zeldin, O.B., Brockhauser, S., Bremridge, J., Holton, J.M., and Garman, E.F. (2013). Predicting the X-ray lifetime of protein crystals. Proceedings of the National Academy of Sciences of the United States of America *110*, 20551–20556.

Zhang, H.-F., Zhang, K., Liao, L.-D., Li, L.-Y., Du, Z.-P., Wu, B.-L., Wu, J.-Y., Xu, X.-E., Zeng, F.-M., Chen, B., et al. (2014). miR-200b suppresses invasiveness and modulates the cytoskeletal and adhesive machinery in esophageal squamous cell carcinoma cells via targeting Kindlin-2. Carcinogenesis *35*, 292-301.

Zhao, T., Guan, L., Yu, Y., Pei, X., Zhan, J., Han, L., Tang, Y., Li, F., Fang, W., and Zhang, H. (2013). Kindlin-2 promotes genome instability in breast cancer cells. Cancer Letters *330*, 208-216.

Zheng, C., Heintz, N., and Hatten, M.E. (1996b). CNS Gene Encoding Astrotactin, Which Supports Neuronal Migration Along Glial Fibers. Science (New York, N.Y.) *272*, 417–419.

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature *495*, 111–115.