Investigations into mechanisms of complement regulation and bacterial evasion of the innate immune response

Joseph J.E. Caesar



Sir William Dunn School of Pathology

and

Brasenose College

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity Term 2012

Investigations into mechanisms of complement regulation and bacterial evasion of the innate immune response

Joseph J.E. Caesar

Sir William Dunn School of Pathology and Brasenose College

Submitted for the degree of Doctor of Philosophy

Trinity Term 2012

Abstract

The complement system is a major mechanism of the innate immune system, providing a first line of defence against infection. Many pathogens have evolved mechanisms of evading the complement response by recruitment of complement components onto their surface. Presented here is characterisation of the interface between a complement regulator, factor H, and a protein from *Neisseria meningitidis*, factor H binding protein (fHbp), which increases complement evasion by the meningococcus. The affinity of the interaction has been measured and experiments have been performed showing that fHbp mimics the glycosaminoglycans found on the surface of host cells. The atomic resolution structure of a mutant form of fHbp that ablates factor H binding is also presented which is a potential vaccine candidate.

Borrelia burgdorferi evades complement by recruitment of factor H mediated by the complement regulator acquiring surface protein (CRASP) family of proteins. Investigations into a member of this family, CspA, has revealed that a dimeric form of the protein binds the complement regulator in a high-affinity interaction that utilises a binding site that has been mapped to domains five to seven of factor H. The atomic resolution structure of another member of the CRASP family, ErpC, has been determined that displays a different architecture to fHbp and CspA, yet experiments show that it is capable of binding factor H domains five through seven with an identical affinity. It is hypothesised that the multiple members of the CRASP family enable complement evasion across the range of reservoir hosts and throughout the different stages of the life-cycle of the zoonotic bacterium.

Complement requires tight regulation on order to prevent damage to host cells. The structure of the oligomerisation domains from a complement regulator, C4bp, and a homologue from *Gallus gallus*, are presented providing the first examples of homoheptameric coiled-coil formation found in nature. Rationalisation of the different C4bp isoforms has also been achieved suggesting that only 7α0β and 7α1β species exist. A new model for incorporation of the beta chain is presented.

The function of the complement factor H related (CFHR) family of proteins has never been reliably elucidated. Presented here is the structure of the N-terminus of CFHR-1 and data showing that CFHR-1, CFHR-2 and CFHR-5 form homo and heterodimeric species which behave as competitive antagonists towards factor H. This results in de-regulation of complement, increasing the levels of activation and helping to maintain the homeostatic balance of complement regulation on the surface of host-cells. Rationalisation of the links between complement related diseases and polymorphisms within the CFHR family of proteins is also provided.

Contents

Contents 1

Acknowledgments **13**

Publications **15**

Abbreviations **17**

Chapter One: Introduction **21**

1.1 The complement system 21

1.1.1 Activation: Classical pathway 23

1.1.2 Activation: Lectin pathway 25

1.1.3 Activation: Alternative pathway 26

1.1.4 The C3 and C5 convertases 27

1.1.5 Terminal cascade 28

1.2 Regulation of complement 29

1.3 Complement regulators: the factor H family 31

1.3.1 Factor H 32

1.3.2 Factor H like proteins 35

1.3.3 Factor H related proteins 35

1.4 Complement regulators: C4b binding protein 37

1.5 Complement related diseases 39

1.5.1 Age-related macular degeneration 39

1.5.2 Atypical haemolytic uraemic syndrome 40

1.5.3 Membraneoproliferative glomerulonephritis, type II 41

1.6 Pathogenic evasion of complement 41

1.6.2 *Borrelia burgdorferi* 42

1.6.3 *Neisseria meningitidis* 45

1.7 Aims of this thesis 47

Chapter Two: Binding of complement regulator factor H by fHbp of *Neisseria meningitidis* 49

2.1 Protein expression and purification 50

2.1.1 Details of plasmids pET-fHbp and pET-fH_{67} 50

2.1.2 fHbp expression and purification 52

2.1.3 fH_{67} Expression, Refolding and Purification 54

2.2 Investigation of the interaction between factor H and fHbp 56

2.2.1 Measurement of the affinity of fHbp for fH_{67} 56

2.2.2 Binding of full length factor H to fHbp 59

2.2.3 Residues involved in the binding interface 60

2.2.4 Heparin binding competition assays 63

2.3 Crystallisation, data collection and processing 65

2.3.1 Crystallisation of the fH_{67}:$fHbp_{E283A\ E304A}$ complex 65

2.3.2 Data collection for the fH_{67}:$fHbp_{E283A\ E304A}$ complex 66

2.3.3 Data Processing for the fH_{67}:$fHbp_{E283A\ E304A}$ complex 66

2.4 Structure determination and refinement 67

2.4.1 Structure solution of the fH_{67}:$fHbp_{E283A\ E304A}$ complex 67

2.4.2 Refinement of fH_{67}:$fHbp_{E283A\ E304A}$ complex 69

2.5 Structural analyses 70

2.6 Discussion 72

2.7 Summary 76

2.7.1 Future directions 76

Chapter Three: Binding of complement regulator factor H by CspA of *Borrelia burgdorferi* 77

3.1 Protein expression and purification 78

3.1.1 Details of plasmids pCspA, pET-fH$_{567}$ and pET-fH$_{678}$ 78

3.1.2 Multiple angle light scattering 81

3.1.3 CspA expression and purification 82

3.1.4 fH$_{567}$ and fH$_{678}$ expression and purification 84

3.2 Determination of the factor H domains involved in CspA binding 85

3.2.1 Surface plasmon resonance 85

3.2.2 Analytical gel filtration 87

3.2.3 Stoichiometry of the CspA:fH interaction 88

3.3 Crystallisation, data collection and processing 88

3.3.1 Crystallisation of the fH$_{567}$:CspA complex 88

3.3.2 Data collection of fH$_{567}$:CspA complex 89

2.3.3 Data processing of the fH$_{567}$:CspA complex 90

3.4 Structure determination and refinement 91

3.4.1 Refinement of PDB entry 1W33 91

3.4.2 Structure solution of fH$_{567}$:CspA complex 91

3.4.3 Refinement of CspA 93

3.5 Structural analyses 94

3.6 Discussion 97

3.7 Summary 101

3.7.1 Future directions 101

Chapter Four: Binding of complement regulator factor H by ErpC of *Borrelia burgdorferi* 103

4.1 Protein expression and purification 104

4.1.1 Details of plasmid pGEX-ErpC 104

4.1.2 ErpC expression and purification 105

4.1.3 Expression and purification of selenomethionine derivitised ErpC 107

4.2 Crystallisation, data collection and processing 108

4.2.1 Crystallisation of selenomethionine derivatized ErpC 108

4.2.2 Data Collection 109

4.2.3 Data Processing 110

4.3 Structure determination and refinement 111

4.3.1 Se-SAD phasing of ErpC 112

4.3.2 ErpC model building 115

4.3.3 ErpC Refinement and validation 116

4.4 Structural analyses of ErpC 117

4.5 Investigation of the oligomeric state of ErpC 120

4.6 Determination of the factor H domains involved in ErpC binding 120

4.7 Interactions of ErpC with the CFHRs 122

4.8 Discussion 122

4.9 Summary 124

4.9.1 Future directions 125

Chapter Five: Structure of the oligomerisation domain from a C4bp homologue in *Gallus gallus*..... 127

5.1 Protein expression and purification 128

5.1.1 Details of plasmid pIMX-315 128

5.1.2 Expression and purification of CRES oligomerisation domain 129

5.1.3 Expression and purification of selenomethionine derivatized CRES oligomerisation domain 131

5.2 Crystallisation, data collection and processing of CRES oligomerisation domain 132

5.2.1 Crystallisation of CRES 133

5.2.2 Data collection of CRES 134

5.2.3 Data processing of CRES 135

5.2.4 Attempted phasing of CRES from heavy atom soaks 136

5.4 Crystallisation, data collection and processing of $CRES_{SeMet}$ 137

5.4.1 Crystallisation of $CRES_{SeMet}$ 138

5.4.2 Data collection of $CRES_{SeMet}$ 138

5.4.3 Data processing of $CRES_{SeMet}$ 139

5.5 Structure determination and refinement of $CRES_{SeMet}$ 141

5.5.1 Se-SAD phasing of $CRES_{SeMet}$ 141

5.5.2 Se-MAD phase improvement of $CRES_{SeMet}$ 144

5.5.3 Initial model building of $CRES_{SeMet}$ 144

5.6 Structure determination and refinement of CRES II 145

5.6.1 Gd-SAD phasing of CRES 145

5.6.2 S-SAD phasing of CRES 147

5.6.3 Density modification 148

5.6.4 Model building 149

5.6.5 Refinement of CRES 150

5.6.6 Refinement of $CRES_{SeMet}$ 151

5.7 Structural analyses of CRES and $CRES_{SeMet}$ 153

5.8 Discussion 154

5.9 Summary 156

5.9.1 Future directions 156

Chapter Six: Macromolecular assembly of C4bp 157

6.1 Protein expression and purification 159

6.1.1 Details of plasmid pIMX-77 159

6.1.2 Generation of plasmid pET-$C4bp_{\beta}$ 160

6.1.3 Expression and purification of $C4bp_{\alpha}$ oligomerisation domain 162

6.1.4 Expression and purification of $C4bp_{\beta}$ oligomerisation domain 164

6.1.5 Purification of C4bp from plasma 167

6.2 Crystallisation, data collection and processing of the $C4bp_{\alpha}$ oligomerisation domain 169

6.2.1 Crystallisation of the $C4bp_{\alpha}$ oligomerisation domain 169

6.2.2 Data collection of the $C4bp_{\alpha}$ oligomerisation domain 171

6.2.3 Data processing of the $C4bp_{\alpha}$ oligomerisation domain 171

6.3 Structure determination and refinement of the $C4bp_{\alpha}$ oligomerisation domain .. 172

6.3.1 Structure determination of the $C4bp_{\alpha}$ oligomerisation domain 172

6.3.1 Refinement of the $C4bp_{\alpha}$ oligomerisation domain 175

6.4 Crystallisation of the $C4bp_{\beta}$ oligomerisation domain 176

6.5 Structural analyses 177

6.5.1 Assembly of natural heptameric coiled-coils 179

6.5.2 Comparison of heptameric coiled-coils: C4bp, CRES and PDB entry 2HY6 185

6.5.3 Disulphide-bond formation 187

6.6 Investigation into the stoichiometry of C4bp 189

6.6.1 Analysis of C4bp using multiple angle light scattering 190

6.6.2 Analysis of C4bp using native mass spectrometry 190

6.7 A model for C4bp assembly 193

6.8 Discussion 194

6.9 Summary 197

6.9.1 Future directions 198

Chapter Seven: A proposed mechanism of regulation by the CFHRs 201

7.1 Protein expression and purification 204

7.1.1 Generation of plasmids pKLAC-CFHR1 and pKLAC-CFHR1_{12} 204

7.1.2 Generation of plasmid pEF-CFHR1 207

7.1.3 Details of plasmids pET-CFHR1_{345} and pET-CFHR2_{34} 208

7.1.4 CFHR-1 expression and purification 211

7.1.5 CFHR-1_{12} expression and purification 214

7.1.6 CFHR-1_{345} and CFHR-2_{34} expression and purification 217

7.1.7 Generation and testing of a monoclonal antibody against CFHR-2_{34} 218

7.2 Crystallisation, data collection and processing of CFHR-1_{12} 219

7.2.1 Crystallisation of CFHR-1_{12} 220

7.2.2 Data collection of CFHR-1_{12} 220

7.2.3 Data processing of CFHR-1_{12} 221

7.3 Structure determination and refinement of CFHR-1_{12} 223

7.3.1 Structure solution of CFHR-1_{12} 223

7.3.2 S-SAD validation of the CFHR-1_{12} structure 224

7.3.3 Refinement of CFHR-1_{12} 225

7.4 Structural analysis of CFHR-1_{12} 226

7.5 Investigations into the role of dimerization by the CFHRs 229

7.5.1 MALS analysis of CFHR-1_{12} 229

7.5.2 MALS analysis of CFHR heterodimers 229

7.5.3 Design of dimerization knockout mutants 230

7.5.4 Effect of dimerization upon CFHR-5 binding to C3 *in vivo* 232

7.6 Investigation of the interactions of CFHR-1 and CFHR-2 with C3b or C5 234

7.6.1 Binding of C3b by CFHR-1 assessed using surface plasmon resonance 234

7.6.2 Affinity measurement of CFHR-1 and CFHR-2 binding to C3b 235

7.6.3 Binding of C5 by CFHR-1 assessed using surface plasmon resonance 237

7.7 Heparin binding properties of CFHR-1 and CFHR-2 238

7.8 Crystallisation, data collection and processing of CFHR-2_{34} 239

7.8.1 Crystallisation of CFHR-2_{34} 239

7.8.2 Data collection of CFHR-2_{34} 240

7.8.3 Data processing of CFHR-2_{34} 241

7.9 Structure determination and refinement of CFHR-2_{34} 241

7.9.1 Structure solution of CFHR-2_{34} 241

7.9.2 Refinement of CFHR-2_{34} 242

7.10 Structural analysis of CFHR-2_{34} 243

7.11 Crystallisation of CFHR-1_{345} 244

7.12 Regulatory activity assays 245

7.12.1 Serum titration 246

7.12.2 Factor H titration 247

7.12.3 Effect of the CFHRs upon complement activity 248

7.12.4 Effect of CFHR-1 upon the activity of complement regulators 249

7.12.5 Effect of disease-associated polymorphisms within the CFHRs upon activity 251

7.13 Discussion 252

7.14 Summary 261

7.14.1 Future directions 263

Bibliography 265

Appendix One: X-ray crystallography theory 303

A1.1 Bragg's law 303

A1.2 Lattice planes 304

A1.3 Structure factor equation 304

A1.4 The phase problem 306

A1.5 Symmetry 306

Appendix Two: X-ray crystallography practice 307

A2.1 Crystallisation using the sitting drop vapour diffusion method 307

A2.2 Data collection 307

A2.3 Data Processing 309

A2.4 Solving the phase problem I: Phasing by molecular replacement 310

A2.5 Solving the phase problem II: Phasing using anomalous dispersion 312

A2.6 Automated model building 314

A2.7 Refinement 315

Appendix Three: Surface plasmon resonance 317

A3.1 Theory 317

A3.2 Analysis of sensorgrams 318

Appendix Four: Multiple angle light scattering 321

Appendix Five: fH binding by all three fHbp variants 323

A5.1 Affinity of fH_{67} for fHbp variant 1 measured using isothermal titration calorimetry 323

A5.2 fH_{67} binding by fHbp variant 2 324

A5.3 fH_{678} binding by fHbp variant 2 325

A5.4 fH_{67} binding by fHbp variant 3 326

A5.5 fH_{678} binding by fHbp variant 3 327

Appendix Six: Investigation into fH binding by NGO0033 329

A6.1 Preparation of NGO0033 329

A6.2 fH_{67} binding by NGO0033 330

Appendix Seven: *Kluyveromyces lactis* media recipes 331

A7.1 YPGal 331

A7.2 MMGal (10x stock) 331

A7.3 YCB/agar with 5 mM acetamide 331

Appendix Eight: Native mass spectrometry spectra 333

A8.1 Masses of the dissociated C4bp assembly 333

A8.2 Masses of the intact C4b complex 334

Acknowledgments

I would like to thank my supervisor, Susan Lea, for giving me the chance to become part of her research group and her continued support and guidance throughout the course of my studies. I am also grateful to all of the members of the Lea group, past and present, for their help, support and friendship. I am particularly appreciative to Steven Johnson for discussions and analysis of MALS data; Pietro Roversi for help with crystallography; Matthew Colledge for work on the CFHRs; Imogen Breen for making CD55; and Rachel Everett for sparing 3C protease.

Much of this work would not have been possible without our collaborators. I would like to thank Lionel Tan and Christoph Tang for discussion and help with studies on fHbp and NGO0033; Reinhard Wallich, Peter Zipfel and Peter Kraiczy for help with studies on the CRASP family of proteins; Fergal Hill and Anna Blom for discussion and help with studies into CRES and C4bp; Deborah Hatherley for help with expression of CFHR-1 in mammalian cell culture; Elena Goicochea de Jorge and Matthew Pickering for discussion and help with the mechanism of the CFHRs; Paul Morgan and Claire Harris for discussion and help with creation of the CFHR-2 antibody and investigation into CFHR dimers; Julien Marcoux for performing native mass spectrometry analysis; and David Staunton for ES-TOF mass spectrometry.

For financial support, I must thank the Medical Research Council and the Sir William Dunn School of Pathology.

Publications

J.J.E. Caesar, S. Johnson, L. Tan, S. van der Veen, E. Goicoechea De Jorge, R. Everett, X. Bai, R.M. Exley, P.N. Ward, N. Ruivo, K. Trivedi, E. Cumber, R. Jones, L. Newham, D. Staunton, R. Ufret-Vincenty, R. Borrow, M.C. Pickering, S.M. Lea and C. Tang.

Design and evaluation of meningococcal vaccines through structure-based modification of host and pathogen molecules.

Plos Pathogens. **2012**.

Roversi, P., S. Johnson, J. J. E. Caesar, F. McLean, K. J. Leath, S. A. Tsiftsoglou, B. P. Morgan, C. L. Harris, R. B. Sim and S. M. Lea.

Structural basis for complement factor I control and its disease-associated sequence polymorphisms.

Proceedings of the National Academy of Sciences of the United States of America. **2011**.

Roversi, P., S. Johnson, J. J. E. Caesar, F. McLean, K. J. Leath, S. A. Tsiftsoglou, B. P. Morgan, C. L. Harris, R. B. Sim and S. M. Lea.

Structures of the rat complement regulator CrrY.

Acta Crystallographica Section F - Structural Biology and Crystallization Communications. **2011**.

Schneider, M. C., B. E. Prosser, J. J. E. Caesar, E. Kugelberg, S. Li, Q. Zhang, S. Quoraishi, J. E. Lovett, J. E. Deane, R. B. Sim, P. Roversi, S. Johnson, C. M. Tang and S. M. Lea.

Neisseria meningitidis **recruits factor H using protein mimicry of host carbohydrates.**

Nature. **2009**.

All papers in print are attached in Appendix 9.

Abbreviations

A_{280}	Absorbtion at 280nm
aHUS	Atypical haemolytic uraemic syndrome
AMD	Age-related macular degeneration
ASU	Asymmetric unit
AU	Arbitrary units
BME	β-mercapto ethanol
C4bp	C4b-binding protein
CCP	Complement control protein
CFHL	Complement factor H like
CFHR	Complement factor H related
CR1	Complement receptor 1
CRASP	Complement regulator aquiring surface protein
CRES	Complement regulatory secretory protein
CRP	C-reactive protein
CSS	Complexation significance score
CV	Column volumes
CVF	Cobra venom factor
DAF	Decay accelerating factor

DNA	Deoxyribonucleic acid
DTT	Dithiothreithol
E. coli	*Escherichia coli*
EDC	N-ethyl-N'-(3-dimethyl aminopropyl)-carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethyleneglycoltetraacetic acid
ESI-TOF	Electrospray ionisation time of flight mass spectrometry
ESRF	European Synchrotron Radiation Facility
Fc	Flow channel
fH	Factor H
fHbp	Factor H binding protein
FPLC	Fast performance liquid chromatography
GAG	Glycosaminoglycan
GPEs	Guinea pig erythrocytes
GST	Glutathione S-transferase
GWAS	Genome-wide association study
IPTG	Isopropyl thio-β-D-galactoside
LB	Luria-Bertani
MAC	Membrane attack complex

MAD	Multi-wavelength anomalous dispersion
MALS	Multiple angle light scattering
MASP	MBL-associated serine protease
MBL	Mannan-binding lectin
MCP	Membrane cofactor protein
MES	2-(N-morpholino)ethanesulfonic acid
MME	Mono methyl ether
MPD	2-Methyl-2,4-pentanediol
MPGN-II	Type II membraneoproliferative glomerulonephritis
MWCO	Molecular weight cut-off
NCS	Non-crystallographic symmetry
NHS	N-hydroxysuccinimide
NHS	Normal human serum
Ni-NTA	Nickel-nitriloacetic acid
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600nm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank

PEG	Polyethylene glycol
RCA	Regulators of complement activation
RMSD	Root-mean-square deviation
RPE	Retinal pigment epithelium
RPM	Revolutions per minute
RU	Response units
SAD	Single-wavelength anomalous dispersion
SAXS	Small angle X-ray scattering
SCIN	Staphylococcal complement inhibitor
SCR	Short consensus repeat
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis
SHBG	Sex-hormone binding globular
SPR	Surface plasmon resonance
TED	Thioester-containing domain
TLS	Translation/Libration/Screw
Tris	Tris(hydroxymethyl)-aminomethane

Chapter One: Introduction

The human body is constantly challenged by infectious agents that, if allowed to replicate unchecked inside the host, may result in tissue damage and lead to disease. Most microorganisms encountered on a daily basis are destroyed within hours of gaining access to the body by the innate immune system which provides an early barrier against infection and requires no prolonged induction period to elicit a response. The innate immune system is non-adaptive, relying on recognition of infectious agents by effectors that are broadly-specific for pathogens prior to clearance via one of several mechanisms. As a result, many pathogenic microorganisms have evolved mechanisms of evading the innate response; enabling rapid colonisation and leading to increased pathogenesis.

1.1 The complement system

The complement system is a major mechanism in the innate immune response, capable of recognition and clearance of infectious microorganisms as well as necrotic and apoptotic host cells, thromboses and immune complexes (reviewed in Sjoberg, Trouw *et al.* (2009)). Consisting of around thirty proteins found both in the plasma, and attached to the surface of bloodstream exposed cells, complement is an evolutionarily ancient system that forms around 5% of the globulin fraction [Walport 2001]. Upon activation, a proteolytic cascade is triggered which results in formation of a pore in the foreign cell surface, known as the membrane attack complex (MAC), which can lead to cell destruction (Figure 1.1). Other clearance mechanisms are initiated by small anaphylatoxins/chemotactic protein fragments that are released during the cascade which can trigger mast cell degranulation (reviewed in Marshall (2004)) and recruitment of phagocytes. Phagocyte recruitment is also triggered by complement components deposited on the surface of foreign cells resulting in their destruction via

opsonophagocytosis [Tofte, Peterson *et al.* 1980]. Finally, complement fragments are involved in the activation of B-cells via specific receptors CD21 and CD35 [Carroll 2004].

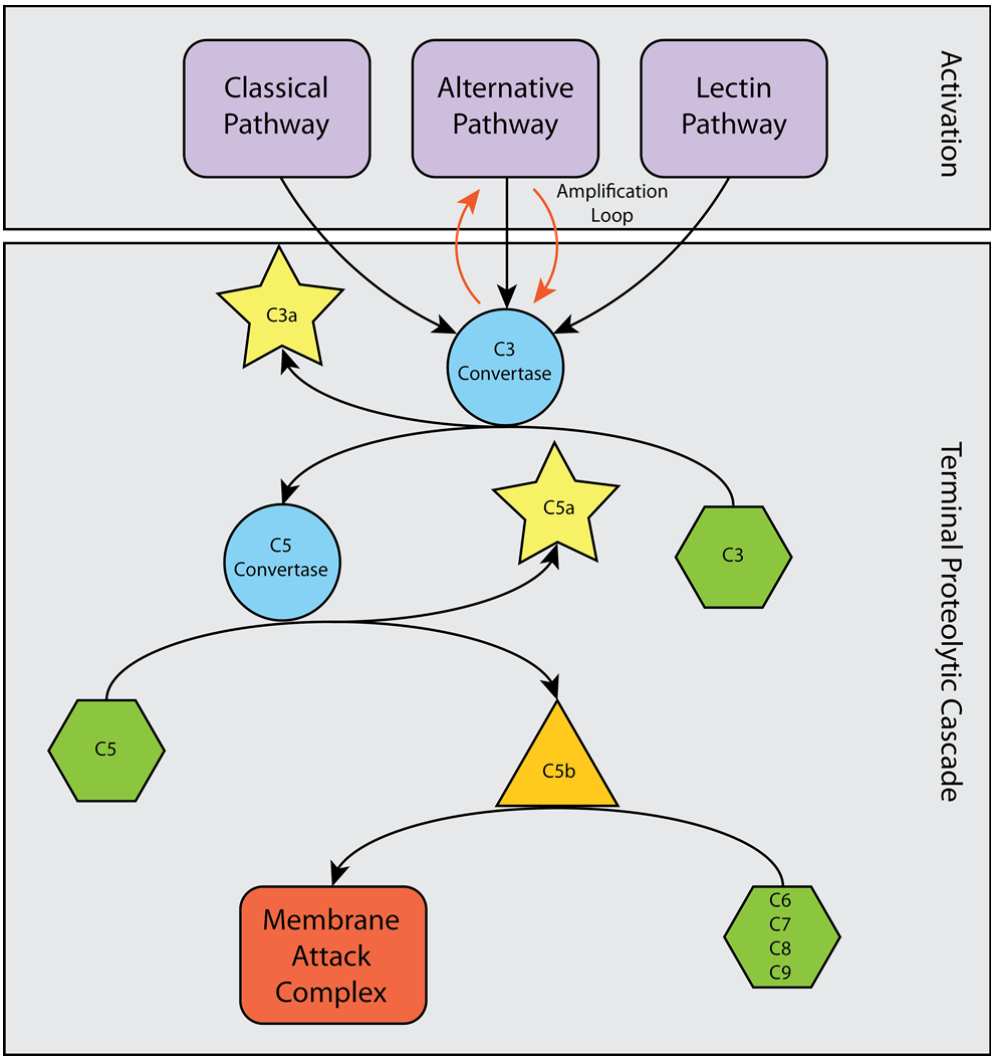


Figure 1.1 Simplified schematic diagram of the complement system detailing activation via three separate pathways which converge upon formation of a C3 convertase; triggering the terminal proteolytic cascade and resulting in clearance by formation of membrane attack complex. Enzymatic complexes are shown as blue circles, inactive complement components circulating in the bloodstream are shown as green hexagons and anaphylatoxins/chemotactic fragments as yellow stars. C5b is shown in a yellow triangle as the trigger for assembly of the membrane attack complex.

Complement can be activated via three distinct routes; the Classical, Alternative and Lectin pathways, all of which converge upon formation of a C3 convertase [Ikeda, Sannoh *et al.* 1987].

1.1.1 Activation: Classical pathway

Complement was named after its ability to be activated by antigen bound antibodies on a pathogenic cell surface, providing an effector mechanism to the adaptive immune system and "complementing" the response [Ehrlich 1890s]. This activation conduit is known as the classical pathway.

The classical pathway is activated when the C1 complex encounters either two IgG, or one IgM, antigen bound antibodies/antibody on a pathogenic cell surface [Porter and Reid 1979]. The C1 complex ($C1qr^2s^2$) is a Ca^{2+} dependent assembly which comprises two C1r, two C1s and one C1q molecule, the latter of which is itself a hexamer of trimers giving an overall arrangement not unlike a bunch of flowers (Figure 1.2, panels a and b) [Gaboriaud, Thielens *et al.* 2004]. Binding of the pro-enzymatic C1 complex to the Fc fragments invokes a conformational change within the assembly, activating an autocatalytic protease activity in the C1r

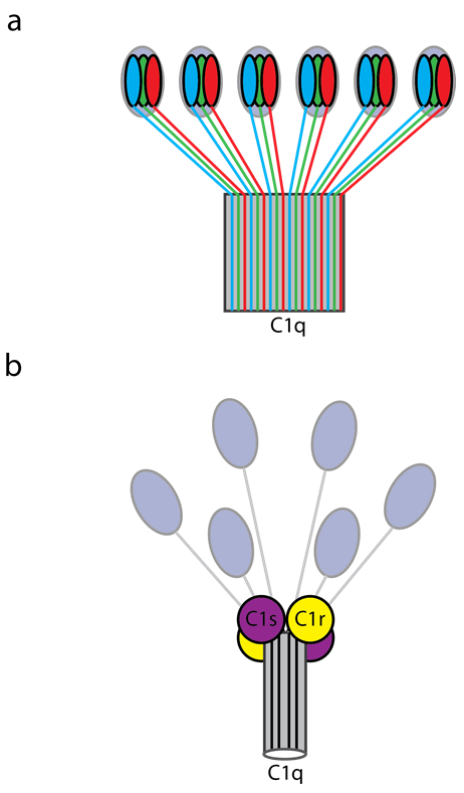


Figure 1.2 *Panel a*: Schematic representation of C1q showing formation of trimeric subunits between C-terminal domains which are multimerised into a hexamer of trimers via the N-terminus. *Panel b*: Cartoon representation of the active C1 complex with the six trimeric heads of C1q and two copies each of C1s and C1r bound.

subunits, which in turn cleave the C1s subunits to form an active serine protease ($C1r^2s^2$). $C1r^2s^2$ activates C4 by cleavage into C4a and C4b, allowing a previously buried thioester in the latter to covalently link to hydroxyl groups in surface carbohydrate molecules and to amine groups in target molecules present on the surface of all cells (Figure 1.3) [Law and Dodds 1990b; Law and Dodds 1997]. C2 can then bind to deposited C4b, forming C4b2, which is in turn cleaved into C2a and C2b by $C1r^2s^2$ to form the classical pathway C3 convertase C4b2a [Kerr 1980].

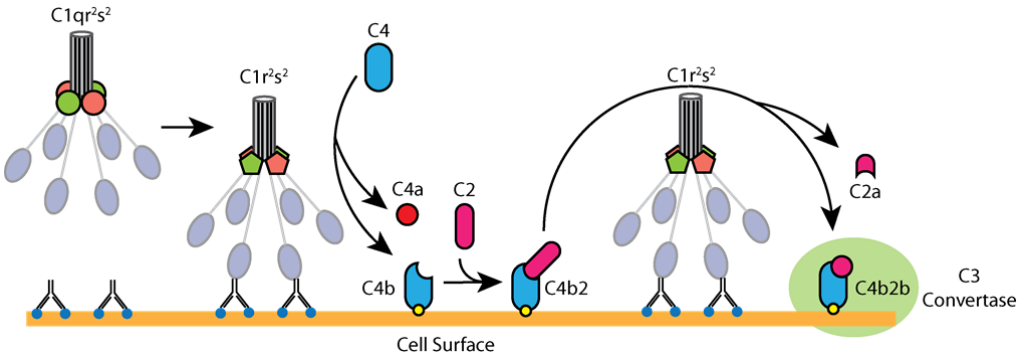


Figure 1.3 Cartoon of the classical activation pathway showing recognition of antigen-bound immunoglobulins by C1 complex leading to assembly of a C3 convertase. This mechanism constitutes the Classical pathway activation balloon shown in purple in Figure 1.1.

C1 complex can also activate the classical pathway in the absence of antibodies. The globular heads have been shown to bind to damaged neurons, myelin debris, amyloid fibrils and apoptotic cells, resulting in activation of the C1s serine protease and helping to provide the diverse range of clearance functions performed by the complement system (reviewed in Sontheimer, Racila *et al.* (2005)). An acute phase protein, C-reactive protein (CRP), also activates complement via the classical pathway. CRP binds to phosphocholine on the surface of necrotic cells where it binds C1q triggering activation [Thompson, Pepys *et al.* 1999].

1.1.2 Activation: Lectin pathway

The lectin pathway draws many parallels with the classical activation pathway. In place of the C1 complex, a complex between mannose-binding lectin (MBL), mannan-binding lectin-associated serine protease 1 (MASP-1) and MASP-2 take the places of C1q, C1r and C1s respectively (Figure 1.4) [Phillips, Toth *et al.* 2009].

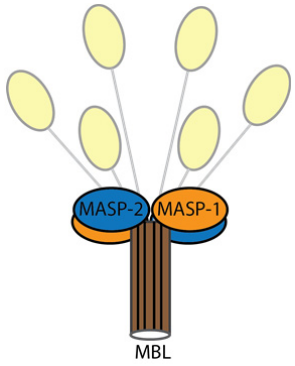


Figure 1.4 Schematic representation of the complex between MBL, MASP-1 and MASP-2 which is architecturally very similar to C1 complex. MBL is shown here with six heads but this number is variable.

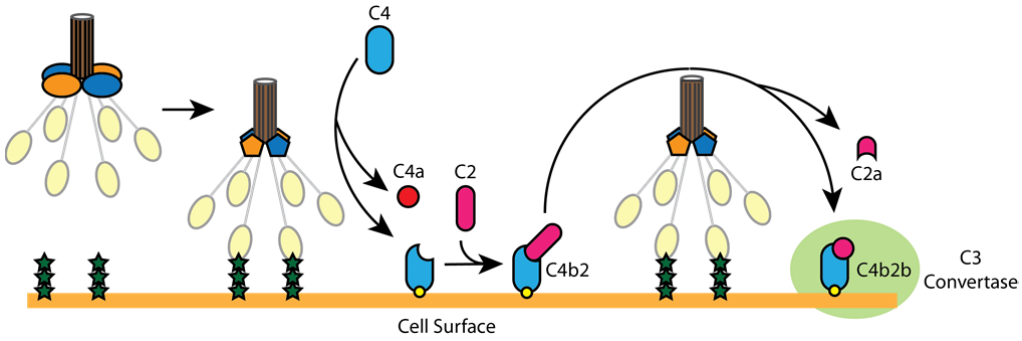


Figure 1.5 Cartoon of the lectin activation pathway showing recognition of surface carbohydrates by the complex between MBL, MASP-1 and MASP-2, leading to assembly of a C3 convertase. This mechanism constitutes the Lectin pathway activation balloon shown in purple in Figure 1.1.

In contrast to the classical pathway, the lectin pathway is not activated by antigen-bound antibodies. MBL is a multimeric assembly possessing between six and eighteen heads that recognise specific patterns of carbohydrate molecules on the surface of foreign cells, or attached to surface glycoproteins or glycolipids (reviewed in Thiel (2007)). Upon recognition, MASP-1 is activated which cleaves MASP-2 generating an active serine protease that can cleave C4 and generate C3 convertase C4b2a in an identical manner to the classical pathway (Figure 1.5) [Thiel, Vorup-Jensen *et al.* 1997]. As well as being activated by MBL, a member of the collectin family of proteins,

another family of carbohydrate recognising proteins, ficolins, can also activate the lectin pathway in a similar manner [Matsushita, Endo *et al.* 2001].

1.1.3 Activation: Alternative pathway

Similarly to C4, C3 is another central, highly abundant, complement protein (approximately 1 - 1.5 mg/ml [Kohler and Muller-Eberhard 1967]) containing a buried thioester linkage that is cleaved to form an active species [Isenman 1983]. Unlike C4, the thioester in C3 can spontaneously hydrolyse to form an active species, C3(H_2O), in a process known as tickover which underlies the alternative activation pathway [Pangburn and Muller-Eberhard 1983]. C3(H_2O) is able to bind another complement component, Factor B. The resulting pro-convertase, C3(H_2O)B, is cleaved into the alternative pathway fluid phase C3 convertase, C3(H_2O)Bb, in a Mg^{2+} dependant manner by a plasma protease, factor D, releasing Ba. C3(H_2O)Bb cleaves C3 into C3a and C3b, the latter of which contains the thioester bond which becomes activated upon cleavage resulting in deposition of C3b on cell surfaces via formation of an ester in an orthologous manner to C4b [Law and Dodds 1997]. The surface-bound C3 convertase, C3bBb, is then assembled around the deposited C3b in the same manner that C3(H_2O)Bb is generated from C3(H_2O). This mechanism is outlined in Figure 1.6.

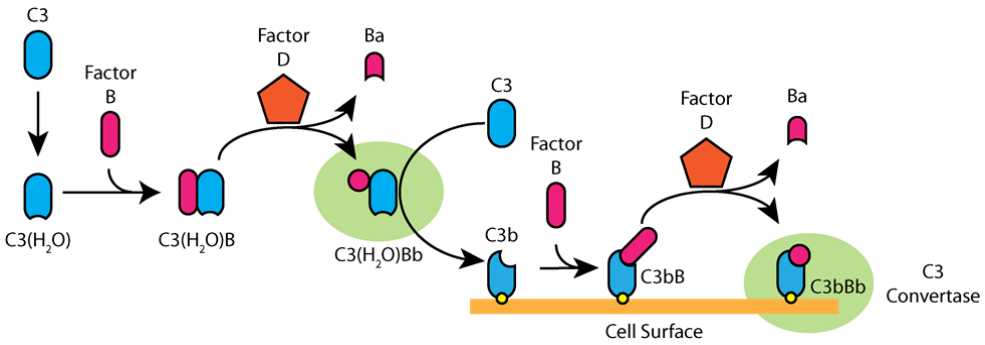


Figure 1.6 Cartoon of the alternative activation pathway showing activation of C3 by spontaneous hydrolysis leading to formation of a fluid phase C3 convertase. C3b is deposited on the cell surface leading to assembly of the C3 convertase. This mechanism constitutes the Classical pathway activation balloon shown in purple in Figure 1.1.

As well as providing its own activation conduit, the alternative pathway also provides an amplification loop to all three activation pathways [Lachmann 2009]. C3b activated by C4b2a is deposited on the activating cell surface where it can form additional C3 convertases, further catalysing the activation of more C3, amplifying the complement response. However, the alternative activation pathway has no built-in discrimination between host and foreign cell surfaces which can lead to amplification of complement on self cells resulting in tissue damage (reviewed in Zipfel and Skerka (2009)). Tight regulation of the alternative pathway is required to prevent this occurrence and ensure that the mechanism is targeted towards pathogen clearance.

1.1.4 The C3 and C5 convertases

Once assembled, both major C3 convertases of the complement system, C4b2a and C3bBb, catalyse cleavage of C3 into C3a and C3b. As previously discussed, nascent C3b generated can be deposited on the activating surface where it can form an additional C3 convertase, amplifying the response via the alternative pathway. C3b can also combine with either C4b2a, or C3bBb, to form a C5 convertase, C4b2aC3b and C3bBbC3b respectively, providing the next step in the complement cascade (reviewed in [Pangburn and Rawal 2002]).

Both the C3 and C5 convertases are highly active, yet unstable, complexes [Janssen, Gomes *et al.* 2009]. However, recent investigations into the mechanism of formation and function have yielded crystal structures that provide great insight into these processes [Rooijakkers, Wu *et al.* 2009; Forneris, Ricklin *et al.* 2010]. Utilising an inhibitor of the C3 convertase, Staphylococcal complement inhibitor (SCIN), from *Staphylococcus aureus* to stabilise C3bBb, the structure shows the precise arrangement of domains in the complex and hints at the C3b binding interface for cleavage and assembly of C3bBbC3b, the structure of which has still to be elucidated (Figure 1.7).

Currently, the structure of C4b2a is unknown, with only the separate structures of C4 (PDB entry 4FXK [Kidmose, Laursen *et al.* 2012]) and C2a (PDB entry 2I6Q [Milder, Raaijmakers *et al.* 2006]) having been determined. Once assembled, the C5 convertase catalyses proteolysis of C5 into C5a and C5b. C5b is deposited on the cell surface, triggering the terminal proteolytic cascade resulting in generation of membrane attack complex leading to clearance.

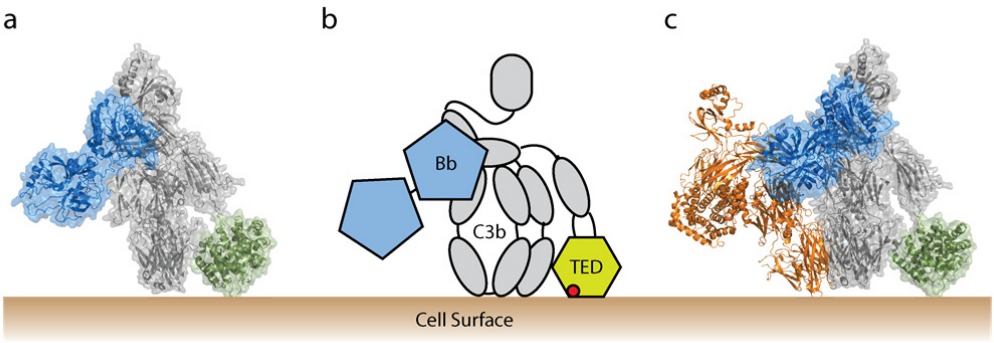


Figure 1.7 *Panel a*: Structure of C3bBb produced from PDB entry 2WIN [Rooijakkers, Wu *et al.* 2009] with the protein mainchain shown in cartoon representation underneath a transparent rendering of the surface. C3b is shown in grey with the thioester-containing domain (TED) in green. The Bb fragment is in blue. *Panel b*: Schematic representation of the C3 convertase. *Panel c*: The structure of the alternative pathway C3 convertase from PDB entry 2WIN with the neighbouring C3 copy shown in orange, suggesting the location of the substrate C3b binding site for cleavage by the C3 convertase.

C3a, and C5a, released by the convertases, and C4a from the classical activation pathway, are anaphylatoxins, triggering a local inflammatory response as well as recruitment of phagocytes to the site of activation via chemotaxis [Vallota and Muller-Eberhard 1973]. Deposited C3b and C4b on the cell surface also act as opsonins, labelling the cell for clearance through opsonophagocytosis [Hostetter, Krueger *et al.* 1984].

1.1.5 Terminal cascade

C5b released by the action of the C5 convertase is deposited on the cell surface via the same mechanism as C3b and C4b [Law and Dodds 1990a]. Following deposition, C5b can bind complement component C6 to form C5b6, which in turn recruits C7 generating

C5b67. Once assembled, the complex undergoes a conformational change allowing C7 to insert into the lipid bilayer constituting the cell membrane. Complex C8, consisting of two proteins, binds first to the C5b component of C5b67 using its $C8_{\beta}$ subunit prior to insertion of the $C8_{\alpha\text{-}\gamma}$ subunit into the membrane, promoting the insertion, and polymerisation, of ten to sixteen copies of C9 to form the membrane attack complex [Ware and Kolb 1981; Musingarimi, Plumb *et al.* 2002].

Fully formed membrane attack complex forms a channel through the cell membrane with an outer diameter of approximately 100Å. Water and ions can travel freely through this pore, disrupting the Ca^{2+} gradient across the cell membrane, giving rise to destruction via disruption of the homeostatic balance within the cell and osmolysis [Peitsch and Tschopp 1991].

1.2 Regulation of complement

The complement system requires tight regulation in order to prevent excessive damage to the host resulting from inappropriate activation. This is especially important within the alternative pathway due to the non-specific nature of C3 activation and leads to the requirement for regulators to distinguish between foreign and self-cell surfaces. The majority of regulation is directed towards the C3 and C5 convertases, being able to control the amplification of the response, deposition of opsonins, release of inflammatory mediators, and entry to the terminal cascade [Zipfel and Skerka 2009]. Regulators fall into two classes; positive and negative. Positive regulators act to stabilise the convertases on activating surfaces, underlining the instability of the complexes themselves which is critical for their regulation. An example is Properdin; a globular serum protein which stabilises the alternative pathway C3 convertase, C3bBb, on the surface of apoptotic cells resulting in clearance [Kemper and Hourcade 2008].

Negative regulators of complement are more common than positive regulators, acting by inhibiting the assembly of complexes or deactivation of the convertases, the latter of which has been observed to proceed via two different mechanisms; decay acceleration and proteolytic deactivation. Decay acceleration is a mechanism whereby a regulator promotes the dissociation of the convertase in an opposite manner to the stabilising effect of Properdin (Figure 1.8). An example is CD55 (or decay accelerating factor (DAF)) which possesses only decay accelerating activity towards both C3 convertases, C3bBb and C4b2a [Brodbeck, Mold *et al.* 2000].

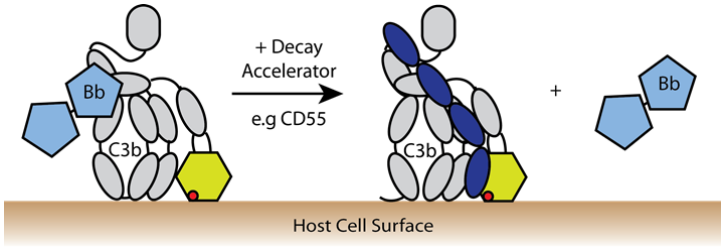


Figure 1.8 Schematic representation of the mechanism of decay acceleration. Binding of a regulator with decay accelerating properties to the C3 convertase, in this case C3bBb, displaces the Bb subunit, disassembling the convertase and terminating the cascade.

Factor I is a serine protease found in serum that is capable of cleaving surface deposited C3b into iC3b, a form that cannot assemble into an active convertase (reviewed in Morgan (1995)). In order for degradation to take place, a cofactor is required to bind to the C3b enabling recruitment of factor I and release of allosteric inhibition of the cleavage site [Roversi, Johnson *et al.* 2011a]. An example cofactor is CD46 (or membrane cofactor protein (MCP)) which possesses only cofactor activity for factor I mediated cleavage of C3b in the alternative pathway C3 convertase [Seya and Atkinson 1989]. Further degradation of iC3b to C3dg is achieved by Factor I in the presence of another cofactor; complement receptor 1 (CR1) [Lambris, Lao *et al.* 1996]. A schematic describing both cleavages is shown in Figure 1.9.

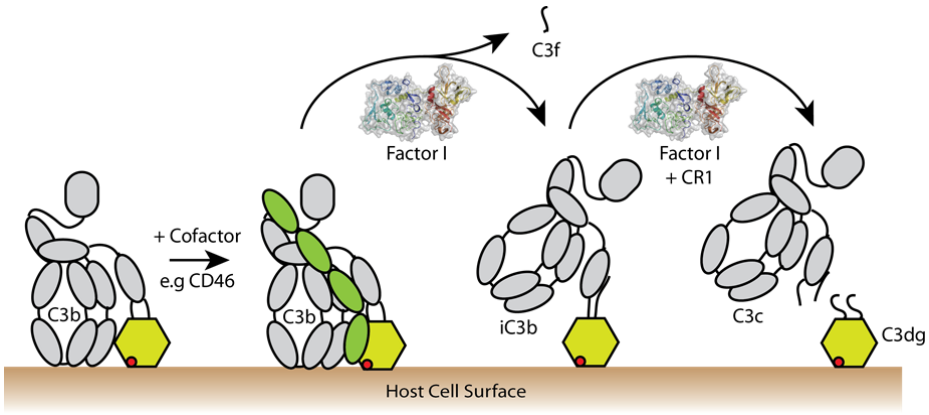


Figure 1.9 Schematic representation of C3b deactivation by Factor I. A regulator with cofactor activity binds to C3b allowing Factor I to cleave the membrane bound molecule into an inactive form, iC3b. In the presence of another complement regulator, CR1, Factor I can further degrade iC3b into C3dg. The structure of Factor I is a rendering based upon PDB entry 2XRC [Roversi, Johnson *et al.* 2011a].

Other regulators control the complement response by binding to critical components in an orientation that sterically occludes the interaction of another component, inhibiting the cascade. An example is CD59 (or Protectin) that inhibits membrane attack complex formation by binding to C5b678 and C9, preventing addition and polymerisation of the latter into the assembly [Kimberley, Sivasankar *et al.* 2007].

1.3 Complement regulators: the factor H family

Located in the regulators of complement activation (RCA) gene cluster on chromosome 1q32 (Figure 1.10, panel b), the factor H family of proteins are a set of serum proteins made up of differing numbers of short consensus repeat (SCR) domains (also known as complement control protein (CCP), or sushi, domains), resembling beads on a string [Hourcade, Holers *et al.* 1989]. Each small, independently-folded domain contains approximately sixty-one amino acid residues and is characterised by a conserved tryptophan residue coupled with a non-contiguous arrangement of the cysteine pairs in two disulphide bonds (Figure 1.10, panels a and c). The linkers between each domain vary between three and eight residues, giving rise to differing degrees of inter-domain

flexibility which is observed in small angle X-ray scattering (SAXS) models of fH [Aslam and Perkins 2001; Okemefuna, Nan *et al.* 2009].

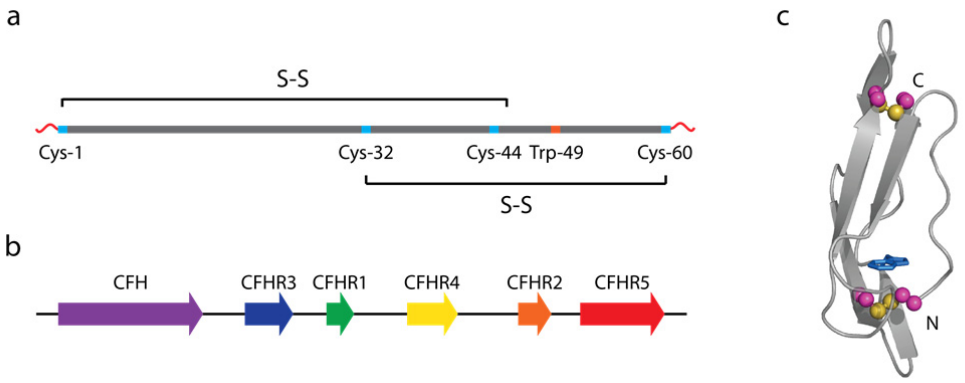


Figure 1.10 *Panel a*: Representation showing the non-contiguous arrangement of cysteine residues, with their approximate location, that form the two disulphide bonds found in an SCR domain. Also shown is the approximate location of the conserved tryptophan residue. *Panel b*: Relative arrangement of the Factor H family of proteins in the RCA gene cluster on chromosome1q32. *Panel c*: Cartoon representation of a typical SCR domain showing the low levels of secondary structure. Cysteine residues involved in disulphide bonding are shown in pink with gold sulphur atoms. The conserved tryptophan is in blue and is buried in the hydrophobic core. Rendering based on PDB entry 2XRB [Roversi, Johnson *et al.* 2011b]

1.3.1 Factor H

Factor H (fH) is the major negative regulator of the alternative pathway of complement activation. Consisting of twenty SCR domains [Kristensen, Wetsel *et al.* 1986; Ripoche, Day *et al.* 1988], the 155 kDa regulator is predominantly synthesised in the liver and has a high serum abundance with a concentration between 235 and 810 µg/ml [Vik, Munozcanoves *et al.* 1990; Neumann, Salzmann *et al.* 2003]. Complement regulatory activity towards the C3 and C5 convertases of the alternative pathway resides in the four N-terminal SCR domains which have been demonstrated to possess decay accelerating activity by displacement of the Bb fragment from C3bBb [Weiler, Daha *et al.* 1976]. These domains also act as a cofactor for factor I mediated inactivation of C3b by cleavage to iC3b [Whaley and Ruddy 1976; Kazatchkine, Fearon *et al.* 1979].

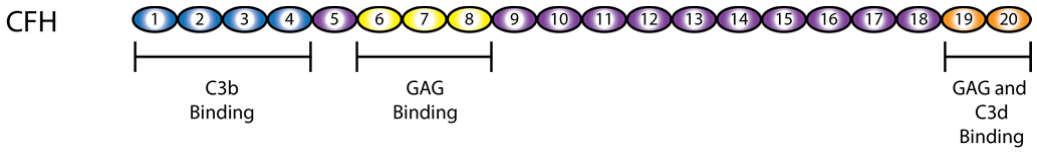


Figure 1.11 Cartoon representation of the structure of factor H, comprised of twenty SCR domains. Domains implicated in regulation and cell-surface localisation are highlighted.

In order to protect them against complement-mediated damage, factor H is localised on the surface of serum exposed host cells via interactions with glycosaminoglycans (GAGs), such as heparan sulphate, and sialic acid [Meri and Pangburn 1990]. This provides a mechanism of discrimination between foreign and self-cells as pathogenic cell surfaces lack these carbohydrates resulting in amplification of the complement response. Regions of fH involved in these binding events have been investigated using binding of heparin, a commonly used polyanionic analogue of the GAGs found on the surface of human cells, describing a major interacting region in domains nineteen and twenty and a lower affinity region in domains six to eight [Blackmore, Sadlon *et al.* 1996; Sharma and Pangburn 1996; Blackmore, Hellwage *et al.* 1998; Clark, Higman *et al.* 2006b; Herbert, Uhrin *et al.* 2006]. There is mixed evidence reporting the existence, and non-existence, of a third heparin binding region around domain thirteen, with a firm conclusion yet to be drawn [Blackmore, Sadlon *et al.* 1996; Sharma and Pangburn 1996].

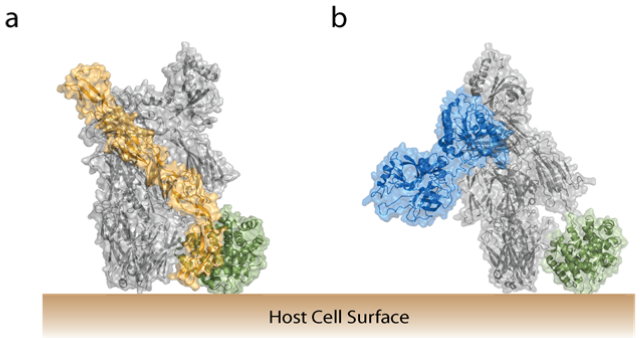


Figure 1.12 *Panel a*: Structure of fH_{1-4} in complex with C3b (PDB entry 2WII) shown with the mainchain in cartoon representation beneath a transparent surface. C3b is shown in grey with the TED domain in green. fH_{1-4} is in gold. *Panel b:* Duplicate of Figure 1.7, panel a, showing the structure of C3bBb.

The heparin binding region in domains six through eight has been described by solution of an X-ray crystal structure of these three SCRs in complex with another highly-sulphated GAG analogue, sucrose octasulphate (SOS) [Prosser, Johnson *et al.* 2007b], and shows utilisation of a positively charged groove across the surface of the protein. Nuclear magnetic resonance (NMR) studies have been coupled with cross-linking and mass spectrometry to characterise the interaction surface in domains nineteen and twenty, suggesting a similar distribution of interfacing residues across the domains [Blaum, Deakin *et al.* 2010].

A recent study has presented the structure of the factor H regulatory domains, SCRs one through four, in complex with C3b (Figure 1.11, panel a. PDB entry 2WII [Wu, Wu *et al.* 2009]). The resulting molecular model for regulation involves displacement of Bb by the binding of the two N-terminal domains, explaining the mechanism of decay acceleration. Rearrangements within the C3b domains also occur that release the allosteric inhibition and allow cleavage by factor I, describing the cofactor activity of factor H. Comparison of this structure with that of C3bBb (Figure 1.12, panels a and b. PDB entry 2WIN [Rooijakkers, Wu *et al.* 2009]) demonstrates both of these mechanisms and also highlights how fH bound to C3b prevents formation of the convertase.

Factor H domains nineteen and twenty interact with the TED domain of C3b, further complicating the mechanism of action of fH. These domains do not possess any cofactor or decay-accelerating properties and most likely further localise fH on self cell surfaces with deposited C3b, regulating complement and preventing host cell damage (reviewed in [Schmidt, Herbert *et al.* 2008a]). The crystal structure of the two C-terminal domains of factor H in complex with the C3b TED domain has been published

showing two separate binding sites, the reason for which is unknown [Kajander, Lehtinen *et al.* 2011].

1.3.2 Factor H like proteins

Complement factor H like protein one (CFHL-1) is an alternatively spliced product of the factor H gene consisting of domains one to seven of factor H followed by four unique amino acids (Figure 1.13) [Schwaeble, Zwirner *et al.* 1987; Fontaine, Demares *et al.* 1989; Sim, Kolble *et al.* 1993]. Circulating at a concentration of approximately 30 µg/ml [Hellwage, Kuhn *et al.* 1997], CFHL-1 domains one to four possess the same regulatory functions as those in the full length molecule [Kuhn, Skerka *et al.* 1995]. However, due to the lack of the C-terminal domains from factor H, CFHL-1 has a lowered affinity for C3b resulting in greatly reduced regulatory ability and a functional role for CFHL-1 has yet to be proposed [Kuhn and Zipfel 1996]. It is still unknown whether there are any other members of this sub-family of proteins, with the possibility that other alternatively spliced factor H transcripts could yet be discovered.

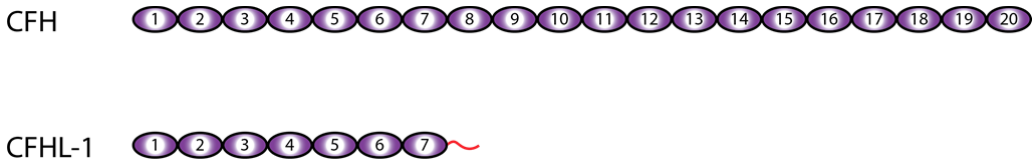


Figure 1.13 Cartoon representation of fH and CFHL-1. The domains are aligned showing that CFHL-1 is identical to fH_{1-7}. The four unique C-terminal residues on the alternatively spliced transcript are shown in red.

1.3.3 Factor H related proteins

In contrast to factor H like protein one, the factor H related (CFHR) sub-family have arisen as a result of a gene duplication events and lie downstream of the full length protein in the RCA gene cluster (Figure 1.10). There are five CFHRs with differing numbers of SCR domains; described in Figure 1.14 with each domain aligned to its most similar counterpart in fH [McRae, Cowan *et al.* 2001]. Genome-wide association

studies (GWAS) have linked polymorphisms within the CFHR family of proteins with enhanced susceptibility to development of complement related diseases suggesting that these molecules are functionally important [Hughes, Orr *et al.* 2006; Gharavi, Kiryluk *et al.* 2011; Zhao, Wu *et al.* 2011]. However, the role of the CFHR proteins within the complement system is still unknown.

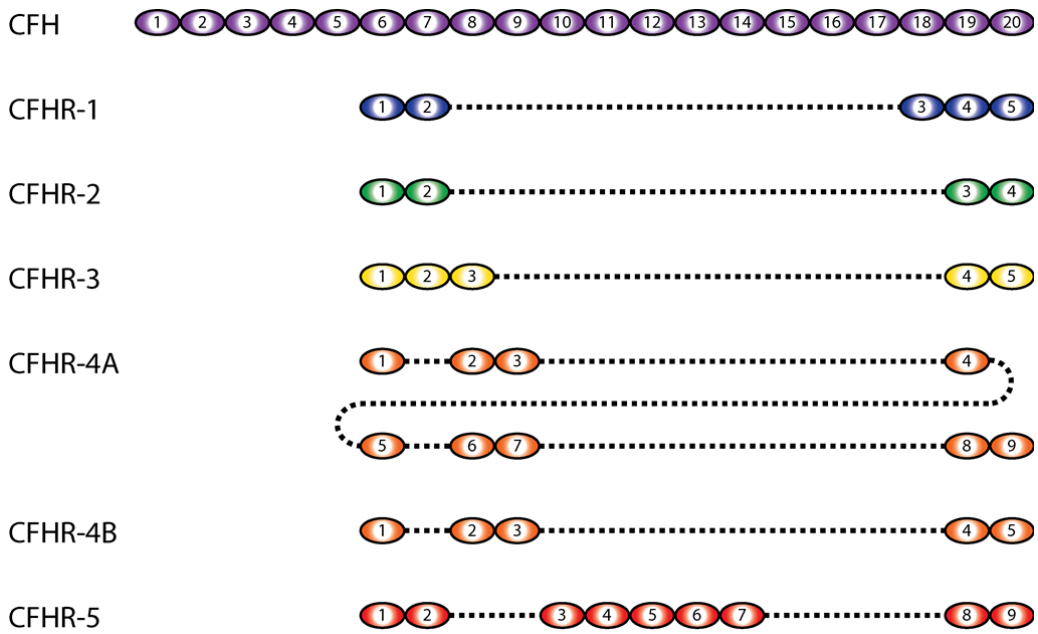


Figure 1.14 Alignment of the individual domains in each member of the CFHR sub-family with their most similar in factor H.

Studies have suggested that CFHR-1 is a negative regulator of the alternative pathway C5 convertase, preventing the terminal cascade [Heinen, Hartmann *et al.* 2009]. Activity was found to reside in the two N-terminal domains. However, a mechanism for how regulation occurs has yet to be proposed. Other data suggests that CFHR-3 can inhibit the C3 convertase whilst CFHR-5 inhibits the C3 convertase and possesses cofactor activity towards factor I mediated inactivation of C3b [McRae, Duthy *et al.* 2005]. However, conflicting reports suggest that CFHR-4 and CFHR-5 do not exhibit cofactor activity [Gale, de Jorge *et al.* 2010; Hebecker and Jozsi 2012]. These mixed

findings mean that the function of the CFHRs, and an associated mechanism for their action, has still to be elucidated.

The serum concentrations of the CFHRs are also the subject of debate due to the reports that CFHR-1, CFHR-2, CFHR-4 and CFHR-5 associate with lipoprotein particles suggesting that serum levels are not a true representation of their overall concentration [Park and Wright 1996; Skerka, Hellwage *et al.* 1997]. To date reliable concentration measurements have only been made for CFHR-1 (70 – 100 µg/ml. [Heinen, Hartmann *et al.* 2009]) and CFHR-5 (3 – 6 µg/ml. [McRae, Duthy *et al.* 2005]).

1.4 Complement regulators: C4b binding protein

C4b binding protein (C4bp) is a negative regulator of the classical and lectin pathway C3 and C5 convertases, functioning as a cofactor for factor I mediated inactivation of C3b and possessing decay accelerating activity towards C4b2a [Scharfstein, Ferreira *et al.* 1978; Gigli, Fujita *et al.* 1979; Rawal, Rajagopalan *et al.* 2009]. Synthesised in the liver, C4bp is an acute-phase protein which has differential expression levels but has a plasma concentration of approximately 150 µg/ml in a healthy individual [Saeki, Hirose *et al.* 1989; Barnum and Dahlback 1990; Griffin, Gruber *et al.* 1992]. C4bp is a disulphide-linked oligomeric protein assembled from two distinct α and β chains, with a major isoform consisting of 7α and 1β chain (7α1β), although the existence of other isoforms, 7α0β and 6α1β, has been reported [Dahlback, Smith *et al.* 1983; Hillarp and Dahlback 1988; Hillarp, Hessing *et al.* 1989].

The C4bp α chain consists of eight SCR domains with a C-terminal oligomerisation domain that is necessary for assembly of the macromolecule, whereas the β chain has three unique SCR domains followed by a different C-terminal oligomerisation domain that interacts with its counterpart in the α chain [Rodriguez de Cordoba, Sanchez-Corral

et al. 1991; Hillarp, Pardo-Manuel *et al.* 1993]. Assembly of the oligomeric protein occurs via formation of intermolecular disulphide bridges between the oligomeristaion domains [Hillarp and Dahlback 1988]. However, the precise arrangement of disulphides within the assembled complex is unknown. Assembly in this manner confers C4bp with an overall architecture that resembles a bunch of flowers (Figure 1.15).

Complement regulatory activity resides in the two N-terminal domains of the α chain, of which the NMR solution structure has been published [Blom, Webb *et al.* 1999; Blom, Zadura *et al.* 2000; Jenkins, Mark *et al.* 2006]. These domains also exhibit GAG binding properties, which will be greatly magnified by the avidity effect resulting from binding of multiple α chains, ensuing a greater local concentration of the surface of plasma exposed cells [Hessing, Vlooswijk *et al.* 1990; Blom, Webb *et al.* 1999]. Avidity will have a great affect upon the affinity of C4bp for C4b2a, resulting in increased regulation at lower levels of C4b deposition, providing a rationale for evolution of the multimeric assembly versus a monomeric regulator such as factor H.

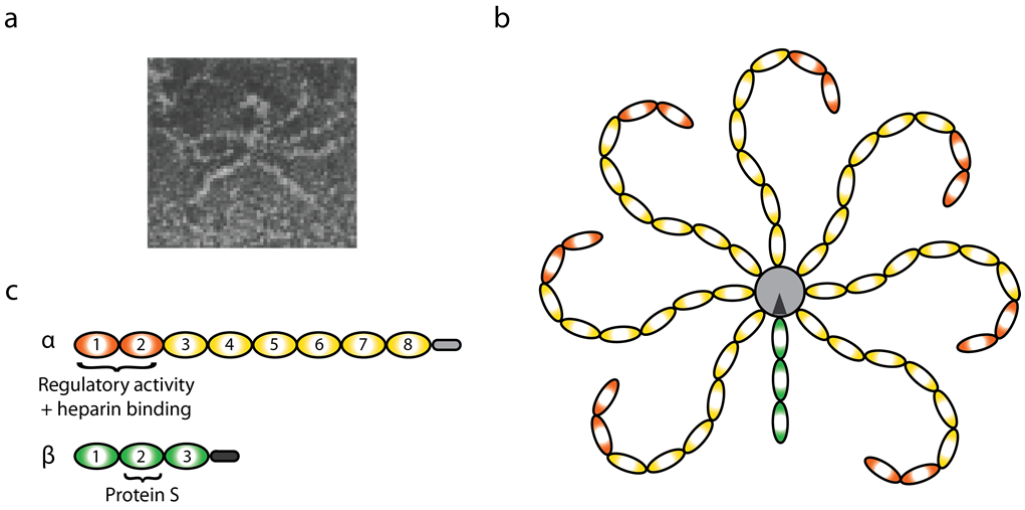


Figure 1.15 *Panel a*: The spider-like structure of C4bp as seen by electron microscopy. Image taken from [Dahlback, Smith *et al.* 1983]. *Panel b*: Cartoon representation of C4b showing the most abundant 7α1β species with the individual chains assembled via oligomerisation of their C-terminal domains to form a central body. *Panel c*: Cartoon of the individual α and β chains showing their multiple SCR domain srchitecture with the domains implicated in regulation, cell-surface localisation and protein S binding shown.

C4bp exists predominantly in complex with protein S; a cofactor in the anti-coagulation pathway [Dahlback and Stenflo 1981]. This interaction has been reported to be mediated through the second SCR domain of the C4bp beta chain and the sex-hormone binding globular (SHBG) like domain of protein S, resulting in inactivation of the cofactor [Evenas, Garcia De Frutos *et al.* 1999; van de Poel, Meijers *et al.* 1999]. The high interaction affinity (100 - 600 pM [van de Poel, Meijers *et al.* 1999]) suggests that C4bp may have a secondary function in control of the coagulation pathway, but a mechanism has yet to be proposed.

1.5 Complement related diseases

There are a wide variety of pathologies associated with deficiencies or polymorphisms in complement components. These result in inappropriate complement activation, altered regulation and increased susceptibility to infection. Complement related diseases often centre around C3 and the C3 convertase as increased C3b activation will result in host tissue damage, whereas decreased activation will result in greater susceptibility to infection and other forms of pathogenesis.

1.5.1 Age-related macular degeneration

Age-related macular degeneration (AMD) is the leading cause of blindness in the aging population of the Western world [Friedman, O'Colmain *et al.* 2004]. Symptoms often present around age fifty with a loss of central vision that steadily worsens with time making everyday tasks increasingly difficult. AMD is associated with the build up of drusen; extra-cellular deposits of protein and lipids between the retinal pigment epithelium (RPE) and Bruch's membrane of the retina [Anderson, Mullins *et al.* 2002]. Drusen contain complement components suggesting that they could result from failed clearance of immunological debris leading to accumulation of proteins coupled with localised inflammation, but a mechanism has yet to be elucidated.

A polymorphism in Factor H domain seven, Tyr402His, has been associated with an increased risk of developing AMD [Haines, Hauser *et al.* 2005]. This mutation occurs within the heparin binding groove previously described (Chapter 1.3.1) suggesting a mechanism for amplified susceptibility for disease that centres around altered GAG binding [Prosser, Johnson *et al.* 2007b]. Currently, it is unclear whether the Tyr402His polymorphism results in an altered affinity for heparin, as several studies have presented conflicting results, requiring further investigation [Clark, Higman *et al.* 2006a; Herbert, Deakin *et al.* 2007; Skerka, Lauer *et al.* 2007]. Decreased risk of developing AMD has also been linked to a simultaneous genomic deletion of CFHR-1 and CFHR-3 [Sinilnikova, McKay *et al.* 2007], again requiring further investigation into a possible mechanism.

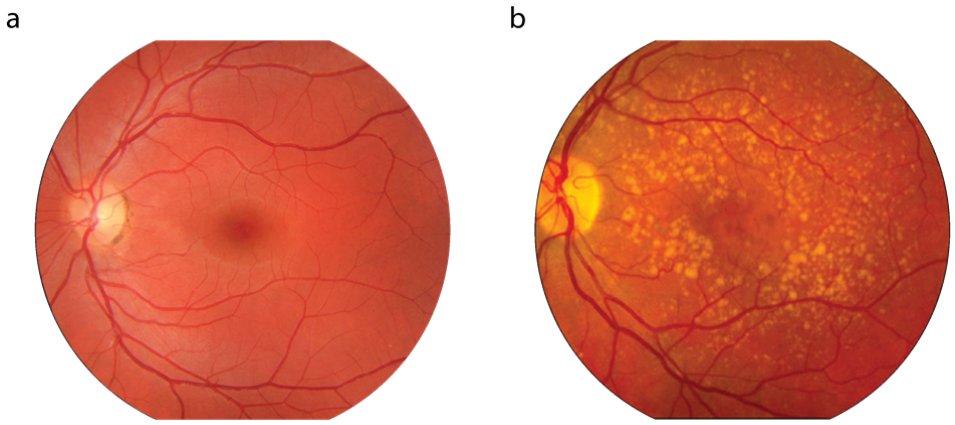


Figure 1.16 *Panel a*: Image of the fundus from a healthy eye. Reproduced under the creative commons CC0 1.0 universal public domain dedication. *Panel b*: Image of the fundus from a patient suffering from AMD showing build-up of drusen. This image was placed in the public domain by the National Institutes of Health (USA).

1.5.2 Atypical haemolytic uraemic syndrome

Atypical haemolytic uraemic syndrome (aHUS) is a renal disorder resulting from insufficient regulation of complement (reviewed in [Waters and Licht 2011]. Accounting for 5 - 10% of HUS cases, aHUS is triggered by minor epithelial damage initiating a largely uncontrolled complement response resulting in microangiopathic

haemolytic anaemia and thrombocytopenia with subsequent thrombosis [Noris and Remuzzi 2005]. The kidney is particularly susceptible to damage due to the necessity for high blood flow through its microvasculature. aHUS results in restricted blood flow giving rise to uraemia and has a high rate of recurrence, often becoming chronic leading to the requirement for kidney transplantation.

aHUS has been associated associated with mutations, or deficiency, in factor H, factor I, factor B and CD46 resulting in reduced regulation of the alternative pathway C3 convertase [Dragon-Durey, Fremeaux-Bacchi *et al.* 2004; Fremeaux-Bacchi, Dragon-Durey *et al.* 2004; Fremeaux-Bacchi, Moulton *et al.* 2006; Goicoechea de Jorge, Harris *et al.* 2007].

1.5.3 Membraneoproliferative glomerulonephritis, type II

Type II membraneoproliferative glomerulonephritis (MPGN-II or dense deposits disease) is a form of glomerulonephritis resulting from complement deposition, via the alternative pathway, in the kidney glomerular mesangium [Appel, Cook *et al.* 2005]. Inflammation mediated by the complement response results in thickening of the mesangium and basement membrane causing damage to the glomeruli, often leading to renal disease.

1.6 Pathogenic evasion of complement

In order to promote infection, pathogens have evolved to bypass the innate immune system and particularly the non-specific complement response. A common bacterial mechanism of complement evasion is recruitment of soluble complement regulators onto their surface where they inhibit activation and amplification of the response. Bacterial species that have been demonstrated to utilise this evasion tactic include *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Yersinia enterocolitica*, *Streptococcus*

pneumoniae, *Fusobacterium necrophorum* and *Borrelia burgdorferi* (reviewed in [Zipfel, Wurzner *et al.* 2007].

1.6.2 *Borrelia burgdorferi*

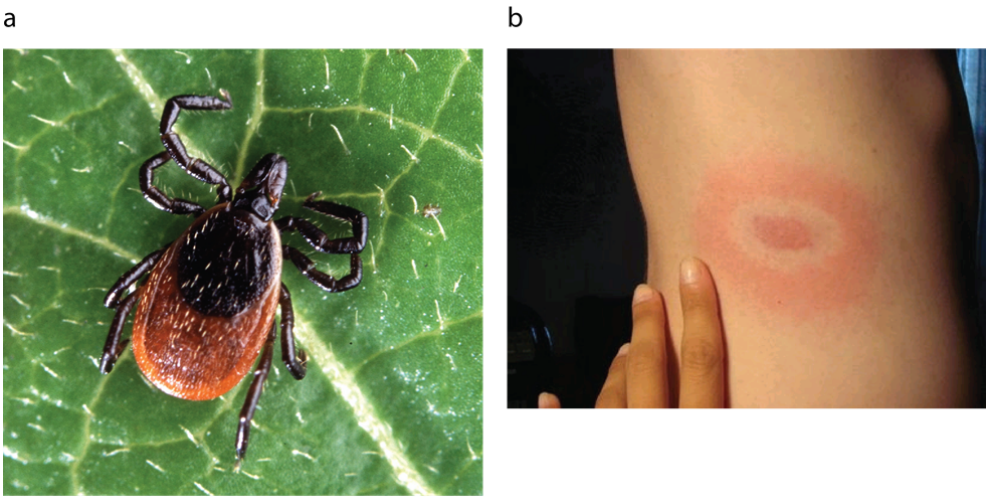


Figure 1.17 *Panel a*: An tick from the *Ixodes* genus; the vector for infection by the zoonotic bacteria *Borrelia burgdorferi.* Image placed in th epublic domain by the Agricultural Research Service (USA). *Panel b:* Patient displaying the charachteristic bulls-eye rash (erythema migrans) associated with borrelial infection. Reproduced under the creative commons attribution-share alike 2.5 generic license.

Borrelia burgdorferi is a Gram-negative spirochete and the causative agent of the most commonly occurring vector-borne disease in Europe and North America, Lyme borreliosis [Barbour and Hayes 1986; Steere 1989; 2007]. Following transmission into the dermis during feeding of infected ticks of the *Ixodes* genus, the predominant indication of infection is a spontaneously resolving skin rash (erythema migrans. Figure 1.17) often accompanied by other symptoms including headache and fever [Steere 1989; Stanek and Strle 2003]. If the infection is not immediately cleared by host immunity or antibiotic treatment, spirochetes can spread to major organs within the host causing a chronic, multisystemic disorder [Steere 1989].

Borrelia species have developed many strategies for evading the different immune systems across its range of reservoir hosts including capture and presentation of host

complement regulators [Embers, Ramamoorthy *et al.* 2004]. Resistance of distinct borrelial species against the complement response upon exposure to human serum has been linked to the binding of major alternative pathway regulators factor H and factor H-like protein-1 (FHL-1) by a family of molecules termed complement regulator-acquiring surface proteins (CRASPs) [Kraiczy, Skerka *et al.* 2001a; Kraiczy, Skerka *et al.* 2001c; Kraiczy, Skerka *et al.* 2001d; Stevenson, El-Hage *et al.* 2002]. There are five members of the family, CspA, CspZ, ErpP, ErpC and ErpA, which posses no sequence similarities between them and are spread throughout the multiple linear and circular replicons which make up the borrelial genome (Figure 1.18) [Bykowski, Woodman *et al.* 2008].

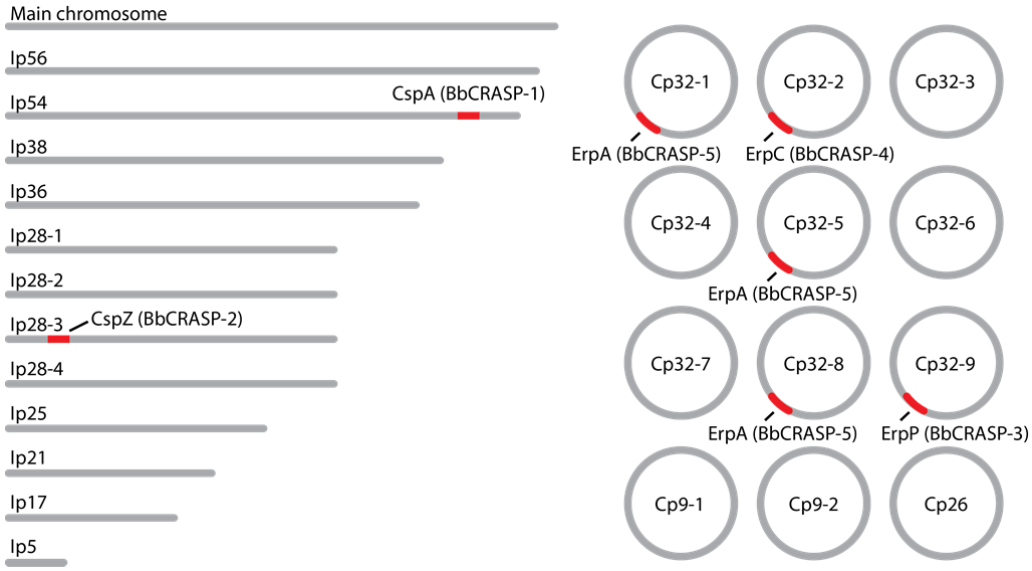


Figure 1.18 Representation of the location of multiple copies of the genes encoding the CRASP family of proteins throughout the borrelial genome.

CspA (also referred to as BbCRASP-1 or BBA68 [McDowell, Harlin *et al.* 2005]) is expressed on the surface of *B. burgdorferi* and binds factor H and FHL-1 in the region of domains five to seven with an affinity measured in the 10 - 30 nM range [Kraiczy, Hellwage *et al.* 2004]. The atomic structure of CspA is known and consists of seven alpha-helices joined by short loops (Figure 1.19) [Cordes, Roversi *et al.* 2005]. The

crystal structure shows formation of a homodimeric species possessing a cleft between the two subunits (Figure 1.19, panel b) within which a putative factor H binding site has been proposed following *in vitro* mutagenesis studies [Cordes, Kraiczy *et al.* 2006; Kraiczy, Hanssen-Hubner *et al.* 2009].

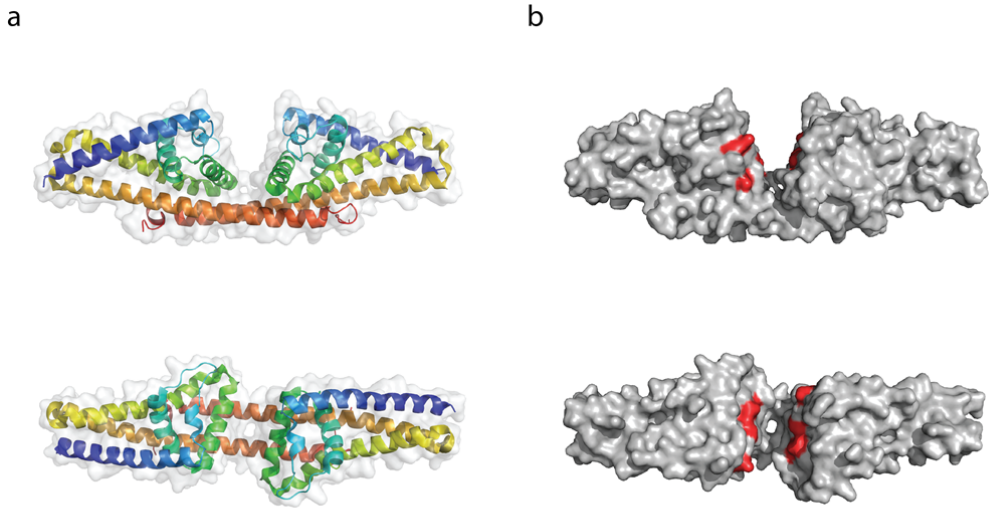


Figure 1.19 *Panel a*: The structure of the CspA dimer from PDB entry 1W33 [Cordes, Roversi *et al.* 2005] from two orientations. A cartoon representation of the mainchain is shown coloured from blue at the N-terminus to red at the C-terminus. *Panel b*: Surface representation of the CspA dimer with the residues within the cleft that have been implicated in fH binding (Lys136, Lys141, Glu147 and Leu146).

Another member of the CRASP family is ErpC (or BbCRASP-4). Whilst retaining fH and CFHL-1 binding properties, ErpC has been implicated in recruitment of CFHR-1, CFHR-2 and CFHR-5, suggesting that *B.burgdorferi* has developed a complex mechanism of complement evasion that involves binding of multiple complement proteins [Kraiczy, Skerka *et al.* 2001a; Hammerschmidt, Hallstrom *et al.* 2012]. Studies have shown that different CRASP proteins are expressed during different stages of the tick life-cycle indicating the importance of subtly different evasion strategies throughout infection of a mammalian host [Bykowski, Woodman *et al.* 2008].

1.6.3 *Neisseria meningitidis*



Figure 1.20 Example of the purpuric rash often associated with meningococcal infection. Reproduced under the creative commons attribution-share alike 3.0 unported license.

Neisseria meningitidis is a Gram negative bacterium found as part of the normal flora in the nasopharynx in up to 15% of adults [Stephens, Greenwood *et al.* 2007]. Spread through exchange of saliva and other respiratory secretions, meningococcal infection is a major cause of mortality and morbidity in the developed world, whilst causing severe epidemics in the developing world. Upon crossing into the bloodstream, infection can result in septicaemia (often termed meningococcemia), or, upon gaining access to the meningeal space, lead to bacterial meningitis, carrying high mortality rates of 50% and 10% respectively [Riedo, Plikaytis *et al.* 1995]. Early symptoms of meningococcal infection can include headache, fever, fatigue and neck-stiffness, often accompanied by the classic purpuric rash (Figure 1.20).

In order to promote infection, *N.meningitidis* has evolved a variety of mechanisms to evade immune clearance including its polysaccharide capsule which inhibits phagocytosis (reviewed in [Lo, Tang *et al.* 2009]). Differences in the molecular makeup of the capsule leads to the major meningococcal serogroups, A, B, C, W135, X and Y. Evasion of phagocytosis combined with being Gram-negative makes the meningococcus particularly susceptible to complement clearance by generation of membrane attack complex. In order to prevent activation and amplification of complement on its surface, *N.meningitidis* recruits factor H using a protein known as factor H binding protein (fHbp) [Madico, Welsch *et al.* 2006].

fHbp (also known as GNA1870 [Masignani, Comanducci *et al.* 2003] or LP2086 [Fletcher, Bernfield *et al.* 2004]) is a 27 kDa surface lipoprotein that is a receptor for factor H and confers serum resistance on the bacterium [Schneider, Exley *et al.* 2006]. The interaction site has previously been determined to involve factor H domains six and seven, with an affinity (approximately 5 nM) that has allowed solution of the X-ray crystal structure of these domains in complex with fHbp [Schneider, Prosser *et al.* 2009]. Observation of the structure shows fHbp to consist of two beta barrels with the complement regulator binding surface across the top (Figure 1.21). fHbp is understood to be anchored to the Neisserial surface via lipidation of a cysteine residue in a predicted lipobox motif at the N-terminus in an analogous manner to transferrin-binding protein B (TbpB) [Masignani, Comanducci *et al.* 2003].

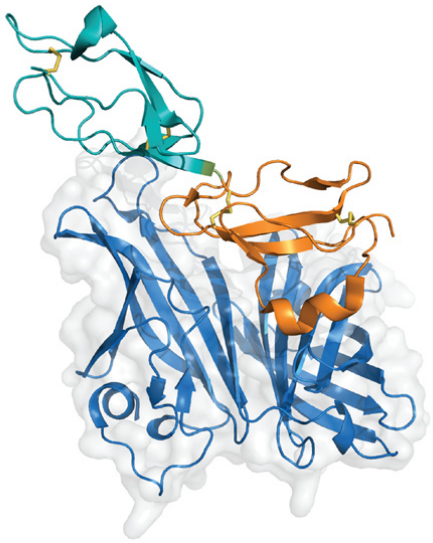


Figure 1.21 Structure of fHbp in complex with factor H domains six and seven from PDB entry 2W81 [Schneider, Prosser *et al.* 2009]. Shown in cartoon representation of the secondary structure with a semi-transparent surface around fHbp (blue). fH domain six is in orange and domain seven in green with disulphides displayed in gold.

Analysis of fHbp sequences from different strains of *N.meningitidis* allows classification into three variant groups based on their primary structure [Masignani, Comanducci *et al.* 2003]. Sequence conservation within each group is between 91.6% and 100%, where as conservation between each variant group can be as low as 62.8%. Most characterisation, and solution of the crystal structure, has been performed using fHbp variant one.

1.7 Aims of this thesis

It is hoped that further investigation into the molecular machinery behind both innate immunity, and its evasion, will result in increased knowledge of the molecular processes involved, possibly leading to the development of new therapeutics for use in the treatment and prevention disease. This thesis aims to further our knowledge of complement evasion by pathogenic bacteria from the *Neisseria* and *Borrelia* genera which scavenge factor H from serum in order to present it on the bacterial surface in an analogous mechanism to host cell protection. Research will focus on characterisation of the molecular mechanisms of factor H binding to fHbp on the surface of *Neisseria meningitidis* and the CRASP family of proteins on *Borrelia burgdorferi* requiring determination of atomic resolution structures of the proteins involved, measurement of the affinities of interactions and elucidation of the underlying interactions that mediate complement evasion.

Mechanisms of complement regulation in order to protect host tissues are not fully understood whilst improper regulation can lead to development of, or increased susceptibility to, disease. Investigation of the currently unknown function of the CFHR group of proteins will be performed in the hope that their role in the complement system can be determined. Complement regulation by C4b-binding protein will also be explored, centring on the macromolecular assembly of the alpha and beta chains. It is hoped that this may shed some mechanistic insights into regulation of both complement and the anti-coagulation pathway by this protein. Increased understanding of these mechanisms may go some way to describing the links to disease and could lead to development of new treatments against complement related diseases.

Chapter Two: Binding of complement regulator factor H by fHbp of *Neisseria meningitidis*

Factor H binding protein (fHbp) has been identified as a receptor for factor H on the surface of *Neisseria meningitidis*, recruiting the complement regulator from serum and presenting it on the bacterial cell surface [Schneider, Exley *et al.* 2006]. This mechanism is analogous to that used to protect host cells from complement-mediated damage and aids the ability of the meningococcus to evade immune clearance by inhibition of the alternative pathway C3 convertase [Madico, Welsch *et al.* 2006]. If *N. meningitidis* is allowed to proliferate, infection may continue, possibly resulting in bacterial meningitis or septicaemia [Stephens, Greenwood *et al.* 2007].

Currently, there is no universal vaccine against *N.menigitidis*, owing to the large variations between different strains [Feavers 2000; Gray, Trotter *et al.* 2006]. There are proven vaccines against meningococcal serogroups A, C, W135 and Y, but no effective inoculation has been developed against serogroups B and X, prompting the search for novel antigens [Harrison 2006]. fHbp has been identified as a promising new vaccine candidate and demonstrated to elicit protective bactericidal antibodies [Fletcher, Bernfield *et al.* 2004]. As fHbp is found on the surface of all meningococcal serogroups it may offer a universal vaccine against all strains [Masignani, Comanducci *et al.* 2003], especially addressing the lack of prophylaxis against serogroup B, the leading cause of bacterial meningitis in Europe and North America [Kvalsvig and Unsworth 2003].

Data are presented here that further describe the interaction between fHbp and factor H in light of the crystal structure of the complex. Quantitation of the affinity of fHbp for factor H has been determined, describing a very tight interaction, whilst analysis of the structure of the complex suggests that fHbp has evolved as a protein mimic of the

GAGs found on host-cell surfaces. Experiments showing that GAGs and fHbp share the same binding site on fH are shown alongside mutagenesis experiments proving that the same fH residues are involved in both interactions. Mutations have also been introduced into fHbp to ablate two salt bridges which mimic carbohydrate binding to fH. These resulted in a dramatic decrease in the affinity of fHbp for fH and is a potential vaccine candidate against *Neisseria meningitidis*. In order to confirm that mutagenesis has not affected the folding of the mutant fHbp, its atomic resolution structure has been solved by X-ray crystallography confirming correct folding.

2.1 Protein expression and purification

In order to express and purify large quantities of fHbp, recombinant prokaryotic expression was chosen as a controllable, high yield system that also allows expression of target proteins as a fusion with a purification tag. Factor H domains six and seven (fH_{67}) were also expressed in a prokaryotic expression system, but form inclusion bodies due to the lack of bacterial machinery to form disulphide bonds. Protein with the correct pattern of disulphide bonds was refolded from purified and solubilised inclusion bodies resulting in homogeneous material which is free from glycosylation, and other post-translational modifications which may occur in eukaryotic expression systems, which is ideal for structural and biophysical studies.

2.1.1 Details of plasmids pET-fHbp and pET-fH_{67}

Collaborators in the group of Professor Christoph Tang (Imperial College, London) cloned the DNA sequence encoding fHbp residues 71 to 320 under an inducible T7 promoter in a pET-21b vector (Novagen, USA), incorporating a C-terminal poly-histidine tag, to create pET-fHbp (Outlined in Figure 2.1 and described in Schneider, Exley *et al.* (2006)). Residues one through sixty-nine have been truncated as these residues are predicted to contain a lipobox lipidation motif and flexible linker which is

likely to be largely unstructured and has been shown to be unnecessary for factor H binding. Forms of pET-fHbp expressing fHbp containing single point mutants E283A and E304A, and double point mutant E283A/E304A, were also generated Schneider, Prosser *et al.* (2009).

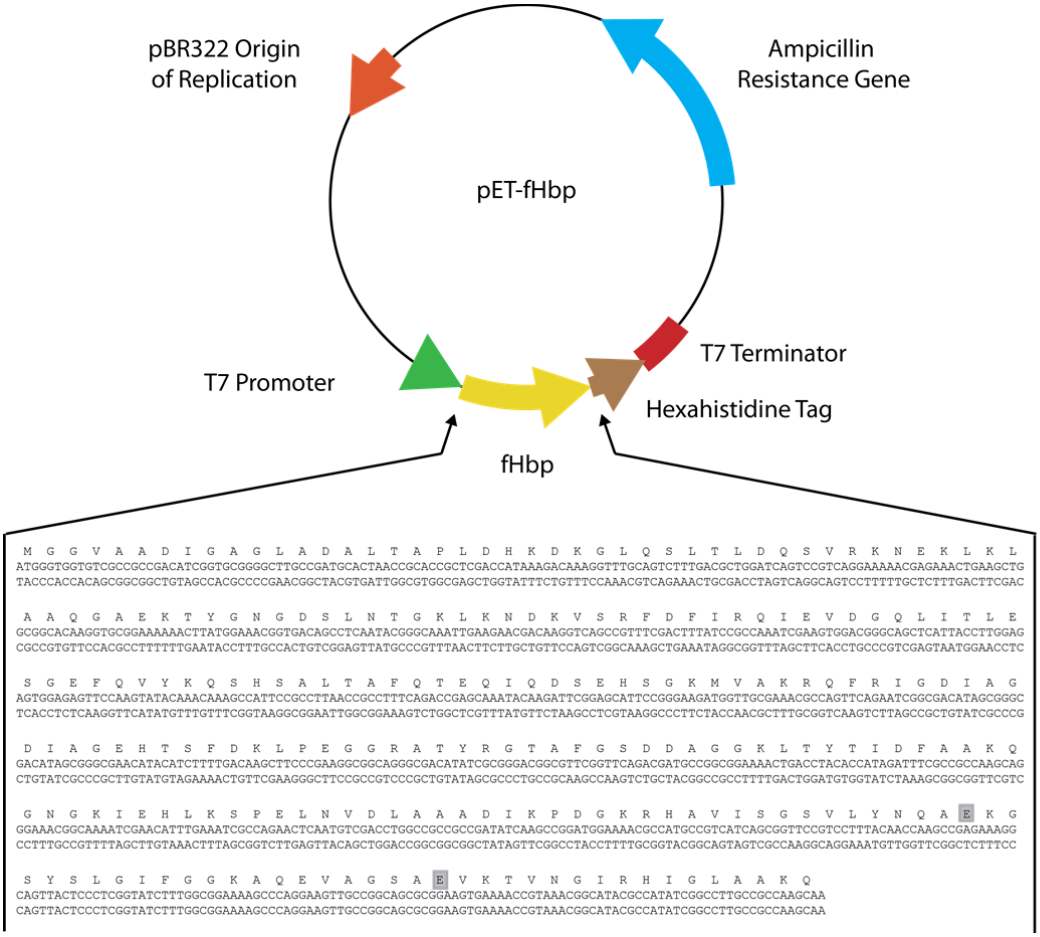


Figure 2.1 Schematic representation of pET-fHbp. fHbp is expressed, with a C-terminal hexahistidine purification tag, under the control of an inducible T7 promoter. The vector was produced by cloning of the fH construct into pET-14b (Novagen, USA) between the NcoI and BamHI restriction sites.

The production of pET-fH$_{67}$ is detailed in Prosser, Johnson *et al.* (2007b) and contains factor H domains six through seven under the control of a T7 promoter (Figure 2.2). A mutated form of pET-FH$_{67}$ expressing fH$_{67}$ with point mutations H337A and R341A has also been previously generated as described in Schneider, Prosser *et al.* (2009).

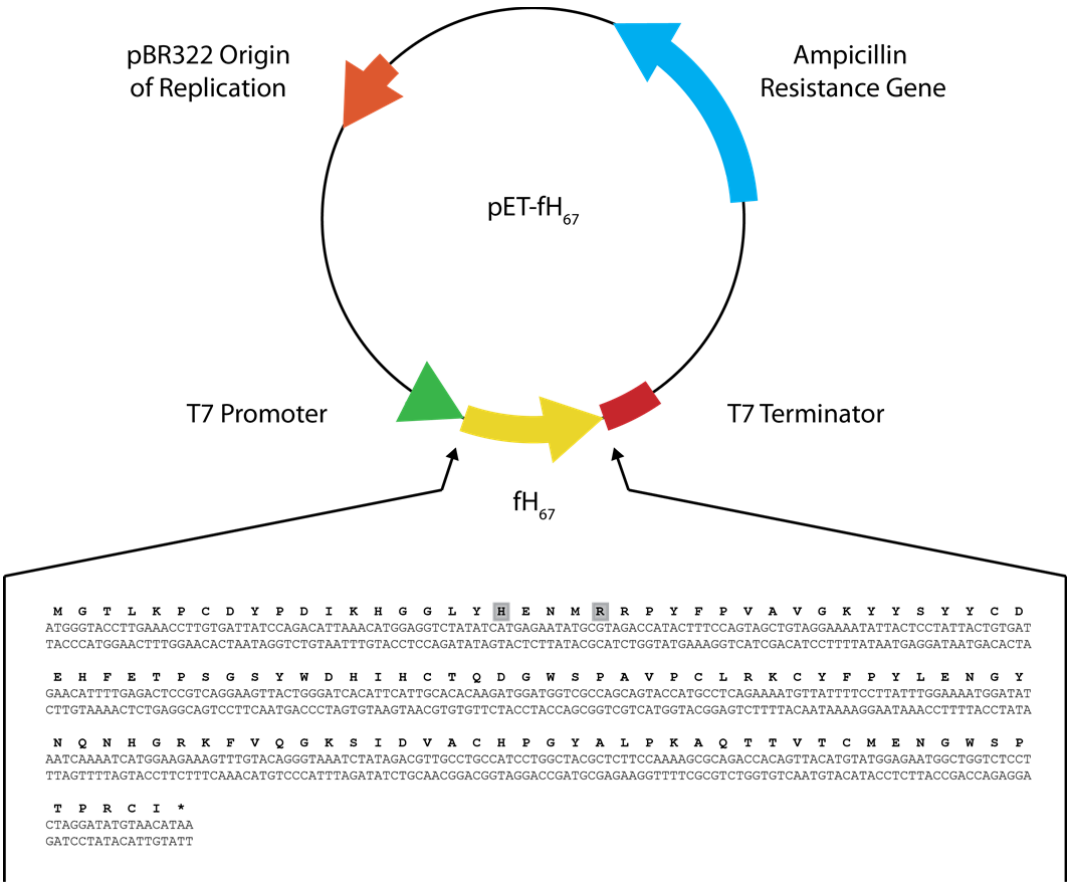


Figure 2.2 Schematic representation of pET-fH$_{67}$. fH$_{67}$ is expressed, without a purification tag, under the control of an inducible T7 promoter. The vector was produced by cloning of the fH construct into pET-14b (Novagen, USA) between the NcoI and BamHI restriction sites. Residues H337 and R341 are highlighted showing the points of mutation in fH$_{67\ H337A\ R241A}$.

2.1.2 fHbp expression and purification

fHbp was expressed and purified in an identical manner to that described in Schneider, Prosser *et al.* (2009). *Escherichia coli* strain BL21(DE3) transformed with the pET-fHbp plasmid was grown overnight at 37^{o}C with 180 RPM shaking in 10 ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 2 × 1L Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37^{o}C with 180 RPM shaking. Cell proliferation was monitored using absorbance at 600 nm (OD$_{600}$) and expression of fHbp was induced in mid-log phase

(OD_{600} = 0.6) by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by continued incubation at 37°C with 180 RPM shaking for three hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4°C) before the pellets were combined and stored at -20°C until required.

Defrosted cell pellets were resuspended in 40 ml of 20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5 and lysed by five passes through an Emulsiflex C5 homogeniser (Avestin, Canada). The crude lysate was centrifuged (48500 g, 20 minutes, 4°C) prior to the supernatant being loaded on a pre-equilibrated 5 ml Ni-NTA column (Qiagen, USA) at 1ml/min. After washing with 50 ml of 20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5 the protein was eluted on an AKTA FPLC (GE Healthcare, USA) using a 20 – 300 mM gradient of imidazole over 150 ml. The fractions containing protein were pooled and concentrated to 5 ml in a 10000 MWCO spin concentrator (Millipore, USA) prior to a final purification step using an S75 16/60 size exclusion column (GE Healthcare, USA) equilibrated in 50 mM Tris, 150 mM NaCl, 1 mM ethylenediamine tetraacetic acid (EDTA), pH 7.5. fHbp elutes at 0.55 CV typically yielding 8 mg per litre of culture assuming ε_{280} = 7762 M^{-1} cm^{-1} (calculated using ProtParam at ExPASy [Gasteiger E. 2005]). The purity was assessed by SDS-PAGE (Figure 2.3) and ES-TOF mass spectrometry returned a mass of 27523 Da corresponding to the expected mass of 27524 Da. An SDS-PAGE gel charting the purification and UV traces from the Ni-NTA elution and size exclusion chromatography steps are shown in Figure 2.3.

fHbp single point mutants E283A and E304A, and double point mutant E283A/E304A, were expressed and purified in an identical manner to the wild type protein.

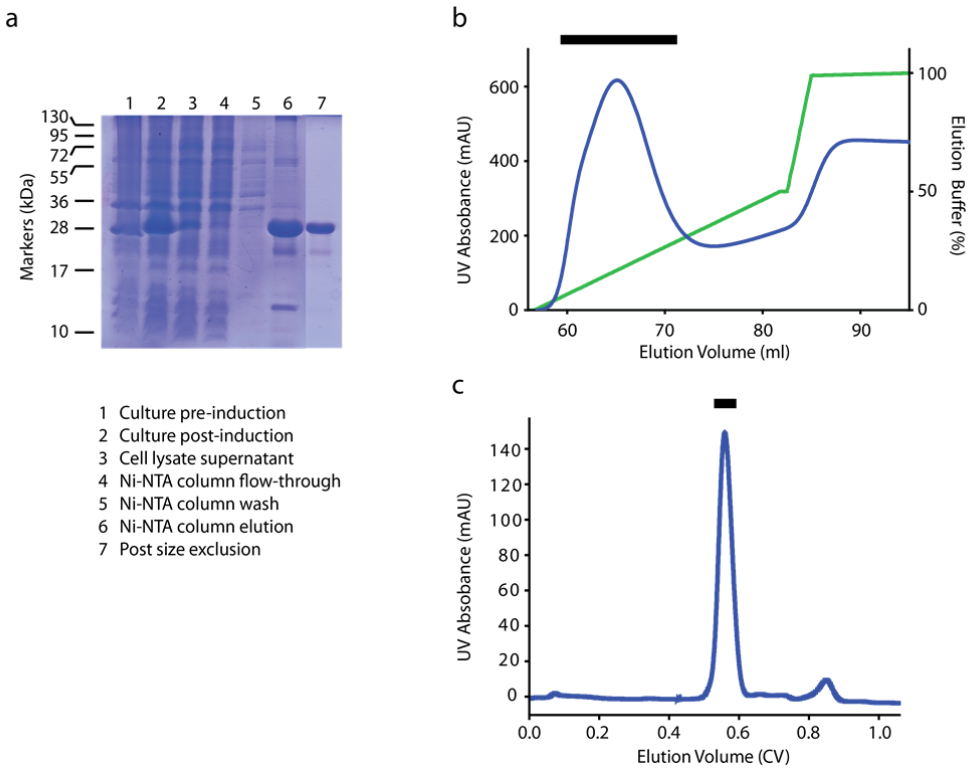


Figure 2.3 *Panel a*: 15% (w/v) reducing SDS-PAGE gel charting the course of the fHbp purification. *Panel b*: Chromatogram detailing the Ni-NTA column elution. The UV_{280} absorbance is shown in blue and the imidazole elution gradient in green. fHbp eluted at approximately 200 mM imidazole and the peak highlighted with a black bar was collected. *Panel c*: Chromatogram detailing the elution profile of fHbp from an S75 16/60 column. The peak containing fHbp elutes at approximately 0.55 CV and the fractions collected are marked with a black bar.

2.1.3 fH_{67} Expression, Refolding and Purification

The protocol for refolding fH_{67} from inclusion bodies was adapted from White, Lukacik *et al.* (2004) and Prosser, Johnson *et al.* (2007a). B834(DE3) *Escherichia coli* cells transformed with pET-fH_{67} were grown overnight in 4 × 10 ml Luria-Bertani broth supplemented with 100 µg/ml ampicillin at 37°C and 180 RPM shaking. Saturated cultures were used to inoculate four litres of LB broth containing 100 µg/ml ampicillin which were grown at 37°C with 180 RPM shaking. The cells were induced with 1 mM IPTG when OD_{600} reached 0.6 and incubated for a further three hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4°C) and the cell pellets stored at -20°C until required.

In a variation to the published protocols, thawed cell pellets were resuspended in 40ml of 50 mM Tris, 1 mM EDTA, 150 mM NaCl, pH 8.0 in place of phosphate buffered saline (PBS) to keep the buffering agent, Tris, constant throughout the preparation. Homogenisation, using an Emulsiflex C5 homogeniser (Avestin, Canada), was subsequently performed without prior addition of lysozyme and exonuclease which was deemed unnecessary for efficient lysis. Another significant deviation from the pubished protocols was the addition of TWEEN-20 to the lysate to a final concentration of 0.05% which was then rocked at 4^{o}C for 20 minutes (as opposed to overnight in the published protocols as there was no significant improvement in purity of the inclusion bodies following a longer wash) prior to centrifugation (48500 g, 20 minutes, 4^{o}C). The purpose of the detergent addition at this stage is to solubilise as much of the bacterial cell membrane in order to remove as much as possible in the initial centrifugation step, saving performing an additional wash of the inclusion bodies. The supernatant was discarded and the pellet washed by resuspension in 40 ml of 50 mM Tris, 1 mM EDTA, 150 mM NaCl, 0.05% TWEEN-20, pH 8.0.

Following centrifugation (48500 g, 20 minutes, 4^{o}C) the pellet was solubilised in 20 ml of 8 M urea, 1 mM EDTA, 100 mM Tris, 25 mM DTT, pH 8.0 prior to acidification to pH 3-4. Reducing agent was removed by dialysis against 6 M urea, pH 3-4 and then the denatured protein was refolded by dilution into two litres of 500 mM arginine, 1 mM cysteine, 2 mM cystine, 20 mM ethanolamine, 1 mM EDTA, pH 11.0 at 4^{o}C overnight. The solution was concentrated to 5 ml and then run on an S75 16/60 size exclusion column (GE Healthcare, USA) equilibrated in 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.5. Correctly folded fH_{67} appears to interact with the Superdex resin in the column and elutes with an elution volume of approximately 1.38 CV with a typical yield of 1 mg per litre of culture (assuming ε_{280} = 32591 M^{-1} cm^{-1} (calculated using

ProtParam)). The purity was assessed by SDS-PAGE (Figure 2.4) and ES-TOF mass spectrometry returned a mass of 14254 ± 6 Da corresponding well to the expected mass of 14252 Da. An SDS-PAGE gel charting the purification and UV trace from the size exclusion chromatography step are shown in Figure 2.4.

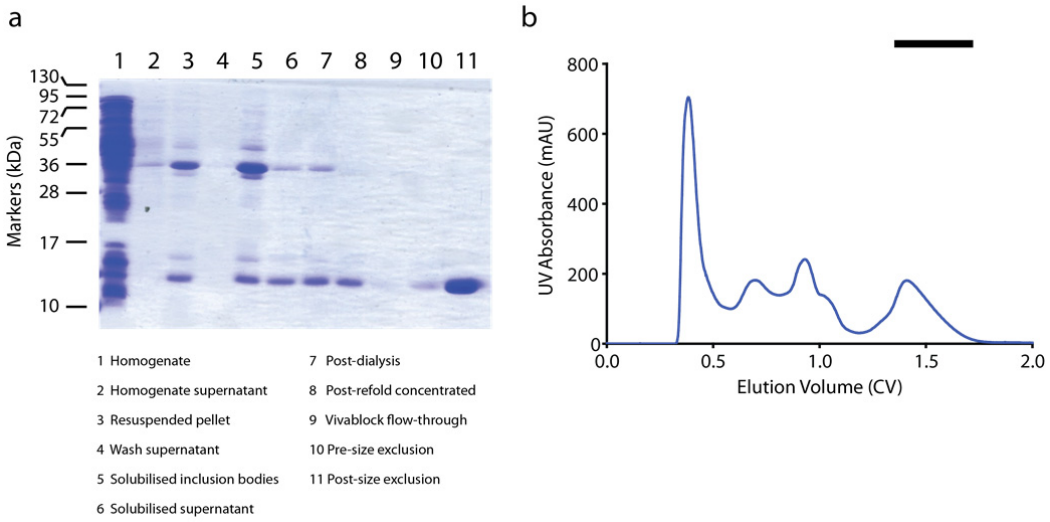


Figure 2.4 *Panel a*: 15% reducing SDS-PAGE gel charting the course of the fH_{67} purification. *Panel b*: Chromatogram detailing the elution profile of fH_{67} from an S75 16/60 column. The peak containing correctly folded fH_{67} elutes at approximately 1.4 CV. Fractions collected are marked with a black bar.

2.2 Investigation of the interaction between factor H and fHbp

Previous work has shown that factor H domains six and seven are both required and necessary for binding of the complement regulator by fHbp and the X-ray crystal structure of the complex between the two is known [Schneider, Prosser *et al.* 2009]. In order to further probe the interaction, analysis of the affinity of complex formation was performed alongside competition assays and mutagenesis studies.

2.2.1 Measurement of the affinity of fHbp for fH_{67}

The affinity of the interaction between fHbp and Factor H domains six and seven (fH_{67}) was investigated using surface plasmon resonance (SPR. Described in Appendix 3). A CM5 sensor chip (GE Healthcare, USA) was docked into a Biacore 3000 machine (GE

Healthcare, USA) and primed into 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P-20, pH 7.5 (HBS-EP) prior to activation with 40 µl of a 1:1 (v/v) mixture of 400 mM ethyl(dimethylaminopropyl) carbodiimide (EDC) and 100 mM *N*-hydroxysuccinimide (NHS) at 5µl/min. fHbp, diluted to 1 µM in 10 mM sodium acetate pH 4.0 in order to increase the efficiency of coupling, was injected over flowchannels 1, 2 and 3 at 5 µl/min to achieve an immobilisation level of between 500 and 1500 RU before all flowchannels were deactivated using 100 µl of 1 M ethanolamine. Flowchannel 4 is a reference channel that has been mock activated-deactivated. Figure 2.5 shows an example sensorgram of fHbp coupling to a chip using this procedure.

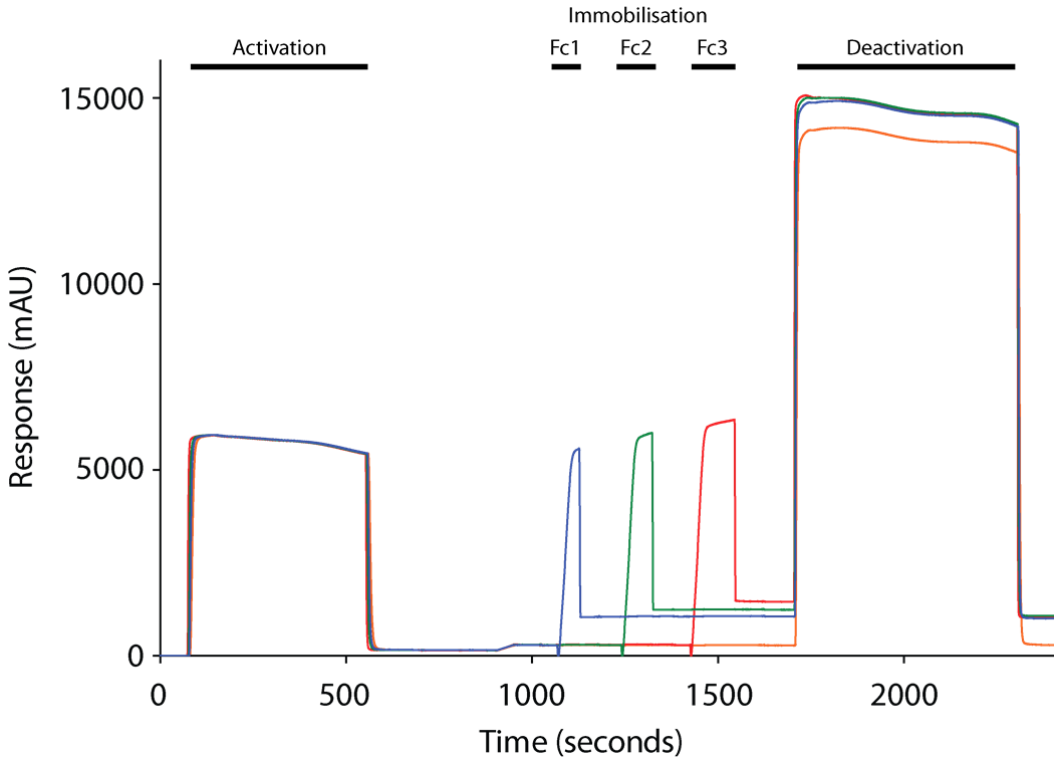


Figure 2.5 Sensorgram detailing an example immobilisation of fHbp on the surface of a CM5 chip. Each stage is labelled and marked with a black bar. Flowchannel (Fc) 1 is shown in blue, flowchannel 2 in green, flowchannel 3 in red and the mock activated-deactivated reference flowchannel 4 in orange.

Dilution series of fH_{67} between 200 nm and 0.2 nM in HBS-EP were injected (200 µl of each) over the chip at 40 µl/min using the KINJECT function with a dissociation time of 1200 seconds. Between each injection any remaining fH_{67} was removed using a 50 µl regeneration step with 10 mM glycine pH 3.0. All injections were performed at 25°C. Analysis was performed using BiaEvaluation 4.0.1. Firstly the reference channel was subtracted from each curve for each fH_{67} dilution. A 1:1 Langmuir fit was then performed to all the curves from each flowchannel, globally fitting k_{on} and k_{off}. Quality of fit was ensured by χ^2 values less than 2 and analysis of residuals showed none greater than 5. Example curves and their corresponding Langmuir fits are shown in Figure 2.6. Measurements were repeated four times using two different chips, and fH_{67} from different preparations, giving a mean K_d of 2 nM (standard deviation = 0.1 nM, n = 4) and a mean χ^2 of 1 (standard deviation = 0.02, n = 4).

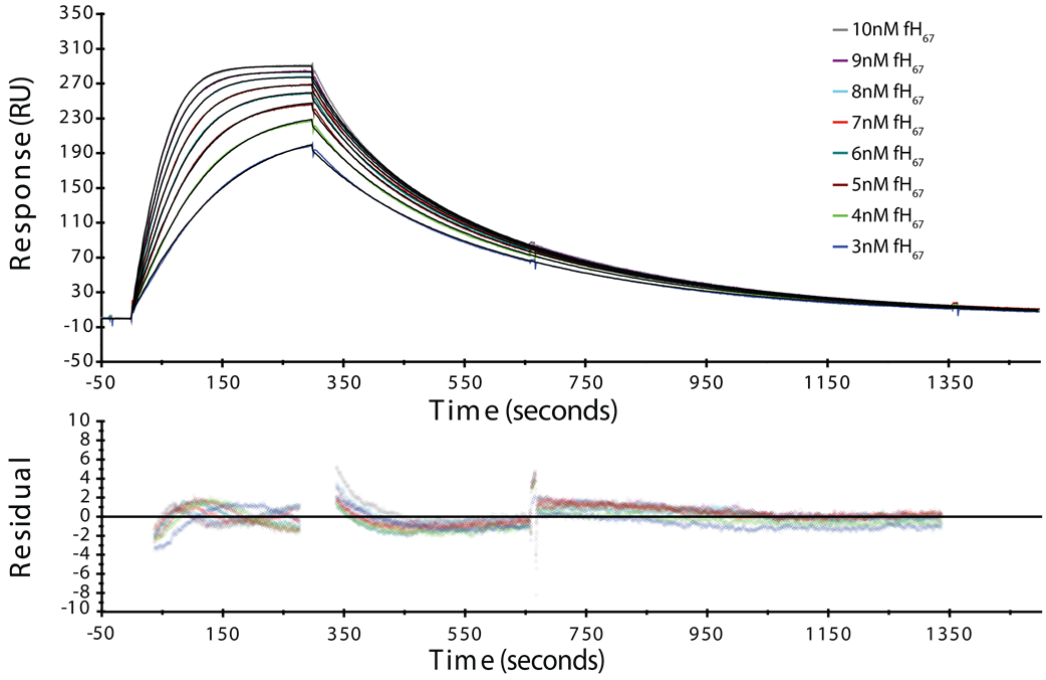


Figure 2.6 Sensorgrams for a dilution series of fH_{67} concentrations flowed over immobilised fHbp. Global 1:1 Langmuir fits have been made to each curve (shown in black) and the residual differences between the experimental curves and their corresponding fits are shown below.

An attempt to measure the kinetics of the interaction in the other orientation was made by flowing fHbp dilutions over fH_{67} immobilised on the surface. These experiments yielded curves that gave a K_d of 14 nM but with a higher χ^2 of 3.9 and with fits that were visibly poor. This discrepancy is likely due to the sugar-binding properties of fH_{67}. During immobilisation the factor H construct may interact with the carboxymethyl dextran layer resulting in occlusion of the interaction surface. In order for fHbp to bind it must compete for the binding site resulting in complex kinetics which can not be satisfactorily fit by a 1:1 Langmuir model.

The affinity of fHbp variants two and three for factor H have also been measured using a Protein XPR36 (Bio-Rad, USA) surface plasmon resonance instrument (Appendix 5) returning K_D values of 2 nM ($\chi^2 = 1.4$) and 3 nM ($\chi^2 = 1.5$) respectively. Only a single replicate of these measurements was taken. However, these results have been repeated and confirmed by other members of Professor Lea's research group.

Isothermal titration calorimetry was employed to confirm the SPR measurements by determining the affinity between fHbp and fH_{67} in solution (Appendix 5). However, due to experimental limitations caused by the tightness of the binding, the accuracy of the measurement was low but sufficient to support an affinity in the low nanomolar range, suggesting that the SPR measurements reliably reflect the solution affinity between fHbp and fH_{67}.

2.2.2 Binding of full length factor H to fHbp

Full length factor H purified from human serum was kindly provided by Dr. Pietro Roversi (Sir William Dunn School of Pathology, University of Oxford). Dilutions of factor H ranging between 10 nM and 50 nM were injected over a chip with fHbp coupled in an identical manner to that outlined in section 2.2.1 (Figure 2.7). 1:1

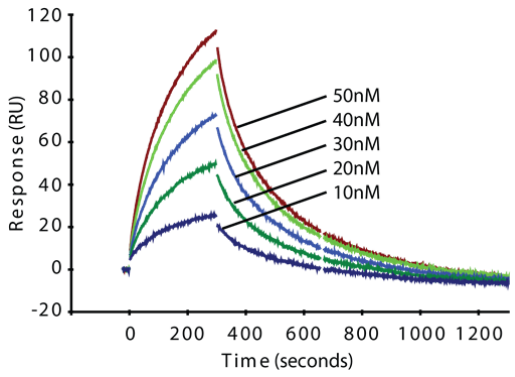


Figure 2.7 Sensorgrams of a dilution series of full length fH over immobilised fHbp. Although the curves show clear dose dependence, 1:1 Langmuir fits were poor, suggesting more complex kinetics.

Langmuir fits to these data are poor which could be due to large size and flexibility of factor H complicating kinetics, rate limiting diffusion through the carboxymethyl dextran, possible self interactions and non-specific interactions with the chip surface. However, the K_Ds obtained from these fits are broadly constant with an affinity in the 1 – 10 nM range as determined in Chapter 2.2.1.

2.2.3 Residues involved in the binding interface

Analysis of the crystal structure of fH_{67} in complex with fHbp (PDB id: 2W81) [Schneider, Prosser *et al.* 2009] using the protein interfaces, surfaces and assemblies service (PISA) at the European Bioinformatics Institute [Krissinel and Henrick 2007] (interfaces summarised in Figure 2.8), highlights two amino acids in fHbp, E283 and E304, that form salt bridges with residues in fH_{67} (Figure 2.9, panel a). Also involved in the interaction are Factor H residues H337 and R341 which form a hydrogen bond and salt bridge to fHbp respectively (Figure 2.9, panel b). In order to assess the contribution of these residues to the overall binding interface, mutant proteins $fH_{67\ H337A\ R341A}$, $fHbp_{E283A}$, $fHbp_{E304A}$ and $fHbp_{E283A\ E304A}$ were generated, replacing the charged, hydrophilic residues with hydrophobic alanine.

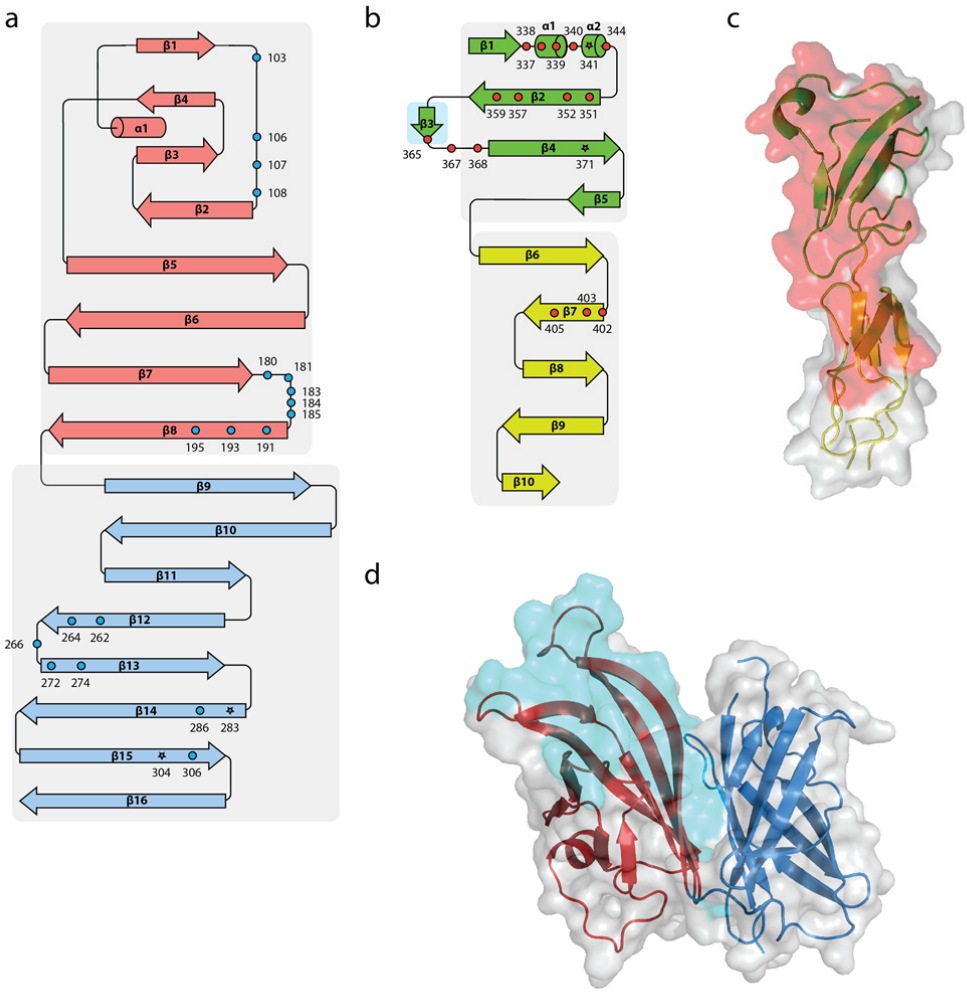


Figure 2.8 *Panel a*: Topology diagram of fHbp with the residues involved in the interface with fH_{67}, as determined using PISA, highlighted. Residues involved in hydrogen bonding interactions are shown with a circle. Residues forming salt bridges are shown with a star. *Panel b*: Topology diagram of fH_{67} showing the interfacing residues in the same scheme as Panel a. *Panel c*: Structure of fH_{67} from PDB entry 2W81 [Schneider, Prosser *et al.* 2009] shown in a cartoon representation of the secondary structure within a transparent surface. fH domains six and seven are coloured green and yellow respectively. The residues involved in the interface are coloured red on the surface. *Panel d*: Structure of fHbp from PDB entry 2W81. The N and C-terminal barrels are coloured red and blue respectively. Residues involved in the interaction interface are coloured blue on the surface.

In order to assess whether binding occurs, 800 RU fHbp and 1500 RU $fHbp_{E283A\ R304A}$ were coupled via primary amine coupling to separate flow channels of a CM5 chip. 200 µl of 50 nM fH_{67} and $fH_{67\ H337A\ R341A}$ were injected over the chip using KINJECT at 40 µl/min. Reference subtracted curves are shown in Figure 2.10, panel a. All mutant

proteins show reduced binding, with no interaction observable between $fH_{67\ H337A\ R341A}$ and $fHbp_{E283A\ R304A}$ under these conditions, supporting the PISA analysis that the mutated residues are highly involved in the interaction interface. In order to probe this further, the affinity of $fH_{67\ H337A\ R341A}$ for fHbp, and fH_{67} for $fHbp_{E283A}$, $fHbp_{E304A}$ and $fHbp_{E283A\ E304A}$, were measured using SPR. Either fHbp or $fHbp_{E283A\ E304A}$ were immobilised on the chip and analyte concentrations between 5 nM and 900 nM were flowed over. Example curves with fits are shown in Figure 2.10 with average K_d and χ^2 values from n repeats. As expected, all three mutant proteins display reduced affinity for fH_{67} with an approximately 10-fold decrease in the K_D for $fHbp_{E283A}$, 100-fold for $fHbp_{E304A}$ and 1000-fold for $fHbp_{E283A\ E304A}$.

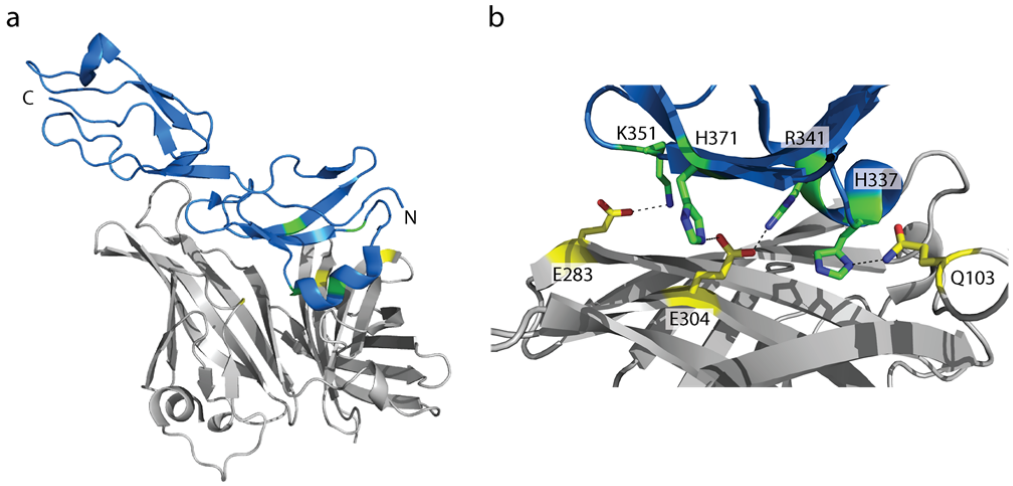


Figure 2.9 *Panel a*: Cartoon representation of the fHbp:fH_{67} structure showning the residues involved in salt bridges. fHbp residues Glu283, Glu304 and Gln103 have their side chains shown and are coloured yellow. fH residues Lys351, His371, Arg341 and His337 are shown in green. *Panel b*: Close up view of the salt-bridging residues.

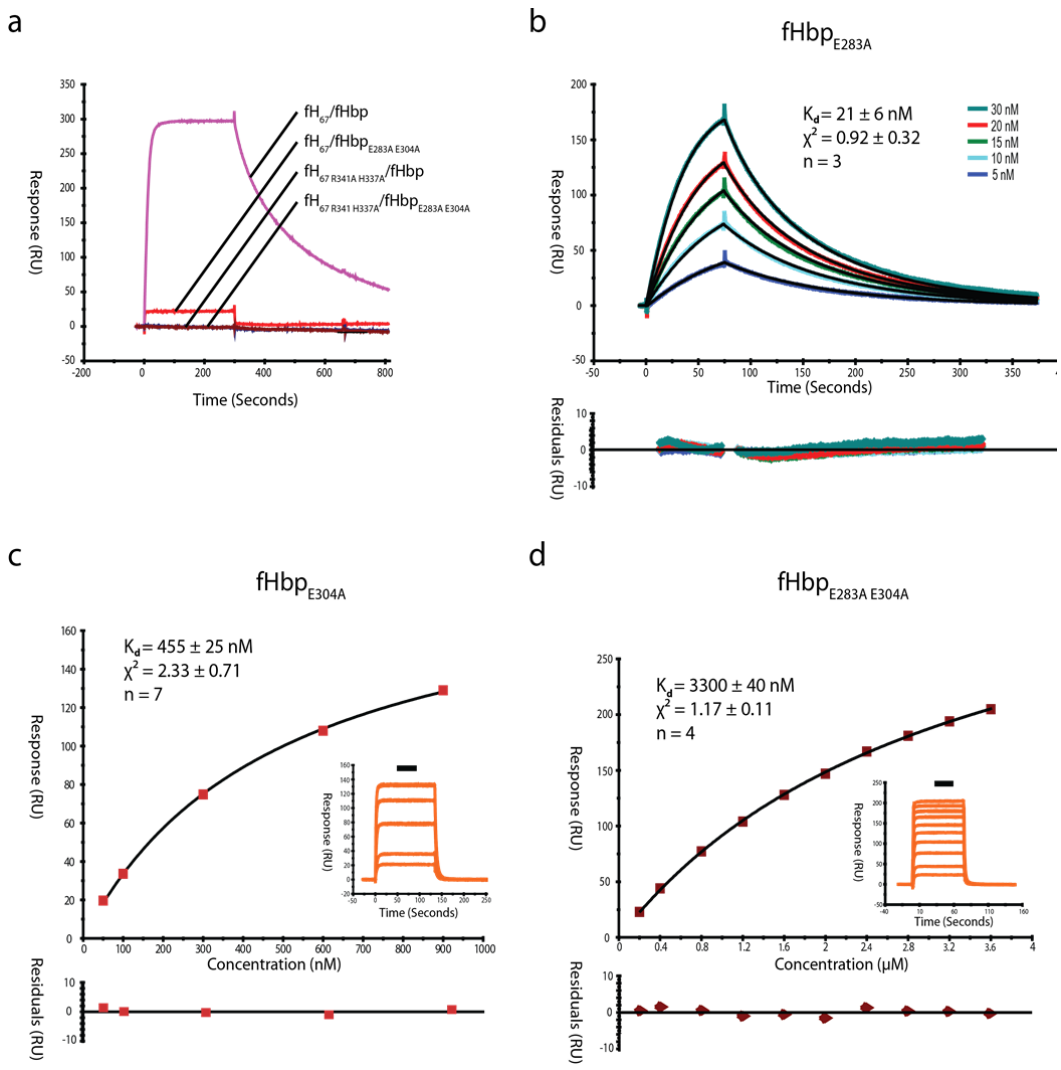


Figure 2.10 *Panel a*: Sensorgrams from 50nM fH_{67} and $fH_{67\ H337A\ R341A}$ flowed over immobilised fHbp and $fHbp_{E283A\ E304A}$.. *Panel b*: Kinetic measurements the affinity of $fHbp_{E283A}$ for fH_{67} The mean K_d and χ^2 are reported from three separate experiments with plus/minus the standard deviation. *Panel c*: Measurement of the affinity of $fHbp_{E304A}$ for fH_{67} using a steady state fit. Example curves are inset showing equilibrium was reached at each concentration and the region over which as average response was measured. Values reported come from seven separate experiments. *Panel d*: Measurement of the affinity of $fHbp_{E304A}$ for fH_{67}. Values reported come from four experiments.

2.2.4 Heparin binding competition assays

Analysis of the complex structure shows that binding of fHbp Variant 1 occurs in the GAG binding site on fH_{67} categorised by [Prosser, Johnson *et al.* 2007b] (Figure 2.11). In order to assess whether *N.meningitidis* has evolved to use fHbp as a mimic for the glycosaminoglycans found on host cell surface, competition assays for fH_{67} binding were performed using heparin.

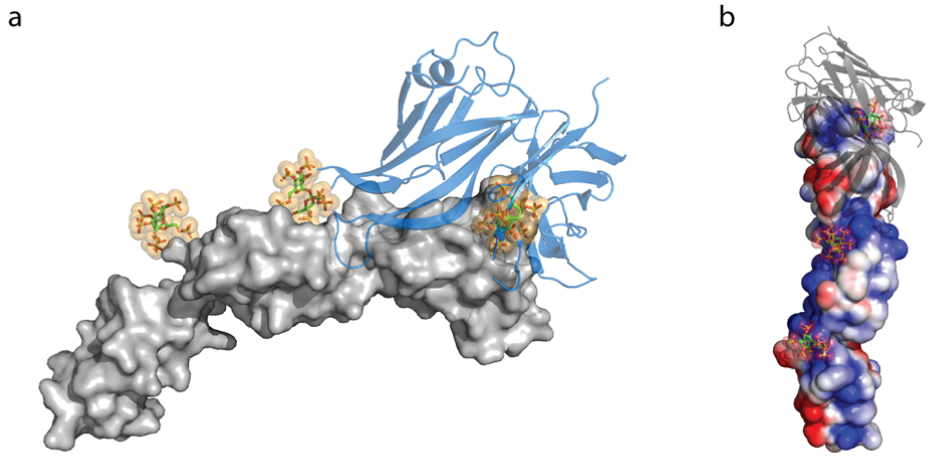


Figure 2.11 *Panel a*: Structure of fH_{678} in complex with SOS, from PDB entry 2UWN [Prosser, Johnson *et al.* 2007b], with domains six and seven overlaid on their equivalents in the fH_{67}:fHbp structure. fH is shown as a grey surface, fHbp as a blue cartoon and SOS as a gold surface around a stick representation. *Panel b*: Electrostatic calculation of fH_{678} using APBS [Baker, Sept *et al.* 2001]. The heparin binding groove can be seen as a blue contiguous patch into which the SOS molecules and fHbp bind.

A CM5 chip with fHbp immobilised in an identical manner to that described in section 2.2.1 had fH_{67} flowed over at 2 nM in the absence and presence of heparin purified from porcine intestinal mucosa (Sigma Aldrich, UK) at concentrations of 0.1, 1 and 5 mM (assuming a mean mass of 18 kDa for heparin chains). Responses were blank subtracted and control subtracted for heparin alone leaving the fH_{67} contribution. This contribution was expressed as a percentage of the uninhibited fH67 signal in the absence of the GAG mimic. All experiments were performed in triplicate and are represented in Figure 2.12 showing dose dependence of fH_{67} binding on heparin concentration confirming that fHbp utilises the same GAG binding site in fH_{67} used for localisation on host cell surfaces.

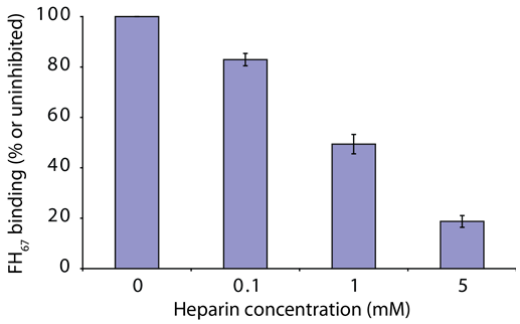


Figure 2.12 Competition for fH_{67} binding by heparin.

2.3 Crystallisation, data collection and processing

In order to ascertain whether the decreased binding observed by the mutations could be caused by disruption of the protein structure, an attempt was made to crystallize the mutant proteins in order to obtain their atomic structures for direct comparison to the wild type sequences

2.3.1 Crystallisation of the fH_{67}:$fHbp_{E283A\ E304A}$ complex

Due to the lower affinity of $fHbp_{E283A\ E304A}$ for fH67, and $fH_{67\ H337A\ R341A}$ for fHbp, compared to the wild type proteins, purification of a complex using size exclusion chromatography is not possible. In order to generate complexes, fHbp, or $fHbp_{E283A\ E304A}$, and $fH_{67\ H337A\ R341A}$, or fH_{67}, were mixed in an equimolar ratio prior to concentration in a 10,000 MWCO spin concentrator (Millipore, UK) until A_{280} = 10.4. Crystallisation trials were set up using an Oryx Nano robot (Douglas Instruments, UK) in 96-well MRC plates (Molecular Dimensions, UK) containing 80 µl mother liquor with a drop size of 400 nl containing a 1:1 (v/v) protein to precipitant ratio. The mother liquor used was 20% PEG 6000, 0.1M Bicine pH 9.0 which is identical to that used by Schneider, Prosser *et al.*

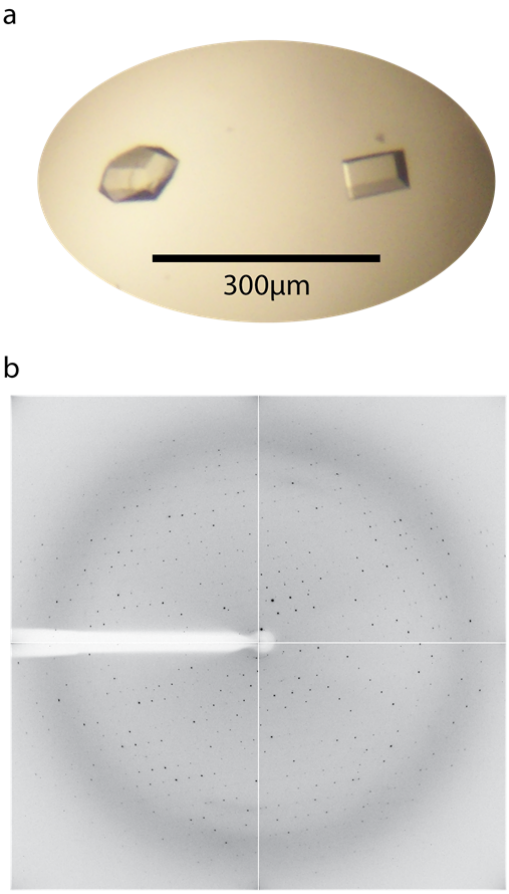


Figure 2.13 *Panel a*: Crystals of fH_{67}:$fHbp_{E283A\ E304A}$ complex. *Panel b*: Example diffraction image from the complex crystals. Figure generated with DIFF2JPEG [Bailey 1994].

(2009) to grow crystals of the complex between wild type proteins. Trays were sealed with Starseal C covers (Star Labs, Germany) and stored in an incubator at 21°C. Crystals grew in the drops containing fH_{67} with $fHbp_{E283A\ E304A}$ after approximately 21 days (Figure 2.13, panel a). These were mounted in cryoloops (Hampton Research, UK) prior to cryoprotection with 20% ethylene glycol and flash cooling in liquid nitrogen. Unfortunately, no crystals grew in any of the drops containing fHbp and $fH_{67\ H337A\ R341A}$.

2.3.2 Data collection for the fH_{67}:$fHbp_{E283A\ E304A}$ complex

	Dataset 1
Diffraction source	ESRF ID14-1
Detector	ADSC Q210
Energy (eV)	13283
Wavelength (Å)	0.9334
Temperature (K)	120
Detector resolution (Å)	2.5
Exposure per image (s)	1
Φ_{start}	28
Φ_{end}	208
$\Delta\Phi$	1
Number Images	180

Table 2.1 Data collection parameters

X-ray data were collected on beamline ID14-1 at the European Synchrotron Radiation Facility (ESRF). Diffracting crystals were indexed using IMOSFLM [Leslie and Powell 2007; Battye, Kontogiannis *et al.* 2011] in a monoclinic unit cell with C2 symmetry. The optimal strategy for data collection was calculated for each crystal prior to collection of a full dataset using the rotation method. Data collection parameters for the crystal that diffracted to the highest resolution, and was used in subsequent operations, are presented in Table 2.1.

2.3.3 Data Processing for the fH_{67}:$fHbp_{E283A\ E304A}$ complex

Data were processed in an monoclinic cell (cell dimensions a = 187.4 Å, b = 53.3 Å, c = 130.1 Å, $\alpha = \gamma = 90.0^{o}$, $\beta = 117.7^{o}$) with C2 symmetry using autoPROC [Vonrhein, Flensburg *et al.* 2011]. The resolution was truncated to 2.8 Å, above which the completeness of the data drops dramatically. autoPROC is an automated package for data processing that was run in its default mode which uses XDS [Kabsch 2010] for

spot-finding, indexing and integration followed by scaling, and merging, with SCALA [Evans 2006]. Intensities are converted to amplitudes, with anisotropy correction, by CTRUNCATE [Winn, Ballard *et al.* 2011] and spacegroup evaluation, followed by reindexing if required, by POINTLESS [Evans 2006]. Data processing statistics for the highest resolution dataset that was used in subsequent structure solution and refinement are shown in Table 2.2.

	Dataset 1
Spacegroup	C2
Unit-cell dimensions (Å)	a=187.4, b=53.3, c=130.1
Unit-cell angles (°)	α=90, β=117.7, γ=90.0
Wilson B factor	63.7
Resolution (Å)	90.01-2.80 (2.96-2.80)
R_{merge}	0.043 (0.170)
R_{pim}	0.035 (0.153)
<I/σ(I)>	16.7 (4.3)
Completeness (%)	94.1 (70.5)
Multiplicity	2.4 (2.3)

Table 2.2 Data processing statistics.

2.4 Structure determination and refinement

Measurement of diffraction data gives us intensities, from which only the magnitude of the structure factors can be determined (Appendix 1). In order to accurately describe the structure within the crystal, phase information is needed to combine with the amplitudes (calculated from the measured intensities) in the structure factor calculation. Molecular replacement (Appendix 2) was employed to obtain the phases starting with the published structure of fHbp in complex with fH_{67} (PDB entry: 2W81 [Schneider, Prosser *et al.* 2009]) as an initial model.

2.4.1 Structure solution of the fH_{67}:$fHbp_{E283A\ E304A}$ complex

Comparison of the unit cells from PDB entry 2W81 (C2 symmetry, a = 185.5 Å, b = 52.2 Å, c = 128.8 Å, $\alpha = \gamma = 90.0^{o}$, $\beta = 118.2^{o}$) and the crystals of fH_{67}:$fHbp_{E283A\ E304A}$ (C2 symmetry, a = 187.4 Å, b = 53.3 Å, c = 130.1 Å, $\alpha = \gamma = 90.0^{o}$, $\beta = 117.7^{o}$) suggests that the crystal packing is highly similar. Analysis of the unit cell content using MATTHEWS [Matthews 1968] suggests the presence of three copies of the complex in

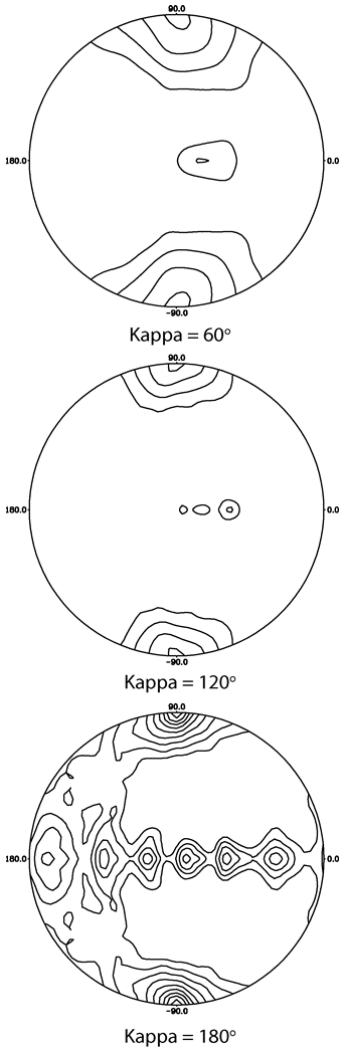


Figure 2.14 Self-rotation functions generated from dataset 1 using POLARRFN [Kabsch 1976].

the asymmetric unit (solvent content = 49.3%) which is supported by significant peaks (approximately 20% of the origin peak) in the chi = 120° self-rotation function (Figure 2.14) suggesting a three-fold rotational non-crystallographic symmetry (NCS). There are also significant peaks in the chi = 60° self-rotation function which probably result from a combination of the three-fold NCS with the crystallographic two-fold symmetry. Combination of all of these observations, plus the lack of significant peaks in the native Patterson (calculated with FFT [Read and Schierbeek 1988]) signifying the absence of a pseudo-translation, suggests that the fH_{67}:$fHbp_{E283A\ E304A}$ complex has crystallised with identical packing to the fH_{67}:fHbp complex in PDB entry 2W81 with three copies in the asymmetric unit related by a three fold, non-crystallographic rotation axis.

Due to the expected high similarity between the structures of the fH_{67}:$fHbp_{E283A\ E304A}$ and fH_{67}:fHbp complexes (99.4% sequence identity), a molecular replacement model consisting of a single copy of the complex was generated from chains A and D from PDB entry

	RFZ	TFZ	LLG
Copy 1	19.6	17.4	456
Copy 2	19.2	44.7	1788
Copy 3	19.2	46.8	3571

Table 2.3 Rotation function (RFZ) Z-scores, translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of three copies of the model in the data.

2W81. Three copies of this model were successfully placed into the fH_{67}:$fHbp_{E283A\ E304A}$ data using PHASER [McCoy, Grosse-Kunstleve *et al.* 2007] with the scores described in Table 2.3. The model produced was rigid body refined using AUTOBUSTER [Blanc, Roversi *et al.* 2004; Bricogne, Blanc *et al.* 2011] with each factor H domain defined as a separate body in order to fit any difference in inter-domain angle between the initial model or NCS related copies. The model was then subjected to 3 big cycles of restrained refinement in AUTOBUSTER, using the autoncs flag to invoke usage of NCS restraints [Smart, Brandl *et al.* 2008], resulting in an R/R_{free} of 0.2531/0.2783. Observation of the resulting $2F_o$-F_c and F_o-F_c maps (Figure 2.15) shows no density and negative residuals around the Glu283 and Glu304 side chain. Prior to refinement mutation of these residues to alanine was performed on each of the three copies using COOT [Emsley, Lohkamp *et al.* 2010].

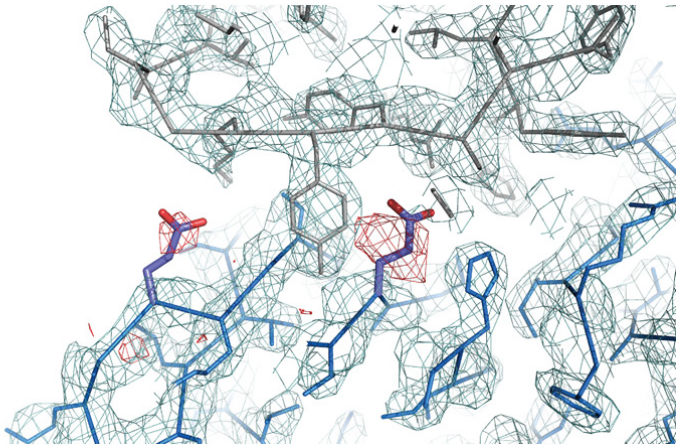


Figure 2.15 2Fo-Fc (teal) and Fo-Fc (red) maps contoured at 1.5σ and -3σ respectively after AUTOBUSTER refinement. Residues E283 and E304 ahow negative residuals confirming that they have been mutated.

2.4.2 Refinement of fH_{67}:$fHbp_{E283A\ E304A}$ complex

The structure was refined iteratively with cycling between rebuilding in COOT and refinement in AUTOBUSTER using the autoncs and noWAT flags to invoke NCS restraints and not update waters respectively. Waters and ethylene glycol molecules

	fH_{67}:$fHbp_{E283A\ E304A}$
Resolution (A)	2.8-57.57 (2.80-2.91)
Number of reflections	26588 (2026)
Number of reflections for R_{free}	1347 (105)
R_{work}/R_{free}	0.2068/0.2258 (0.2626/0.3451)
Number of atoms	
Protein	8444
Ligand/Ion	210 Ethylene glycol
Water	134
R.m.s deviations	
Bond lengths (A)	0.007
Bond angles (°)	1.02
Ramachandran plot analysis	
Residues in favoured regions (%)	96.94
Residues in disallowed regions (%)	0.0
Molprobity analysis	
Molprobity score	1.63
Molprobity score percentile	100th
Resolution range (A)	2.55-3.05

Table 2.4 Refinement and validation statistics

were added in COOT. The accuracy of refinement was scrutinised by comparison of the R factor with the R_{free} which was calculated from the same test set that the refinement of 2W81 was validated against. Throughout refinement the accuracy and validity of the model was assessed using MOLPROBITY [Chen, Arendall *et al.* 2010] resulting in a final model that has refinement and validation statistics displayed in Table 2.4. Final coordinates were deposited in the PDB with accession code 4AYE.

2.5 Structural analyses

The structure of fH_{67} in complex with $fHbp_{E283A\ E304A}$ is unchanged from that previously published, except at the sites of mutation (Figure 2.16, panel a). Deviations between all copies in both structures are within the deviations between NCS related copies in each structure (Figure 2.16, panels c and d). The crystal packing is also identical between the two structures (Figure 2.16, panel b).

	ω	φ	χ
Copy 1 (AD) - Copy 2 (BC)	87.8	91.3	118.3
Copy 1 (AD) - Copy 3 (EF)	89.3	-82.4	123.1
Copy 2 (CD) - Copy 3 (EF)	83.9	97.1	119.4

Table 2.5 Rotational transformations between NCS copies of the fH_{67}:$fHbp_{E283A\ E304A}$ complex.

Calculation of the rotational transformations in spherical polar coordinates (Table 2.5) that relate the three copies of the fH_{67}:$fHbp_{E283A\ E304A}$ complex in the

asymmetric unit describe the peaks in the chi = 120° self-rotation function and account for their broadness. This three-fold non-crystallographic axis combines with the crystallographic two-fold axis to produce the peaks in the chi = 60° self-rotation function. The small peaks in the chi = 180° plot probably result from mapping the individual SCR domains, and fHbp barrels, onto each other.

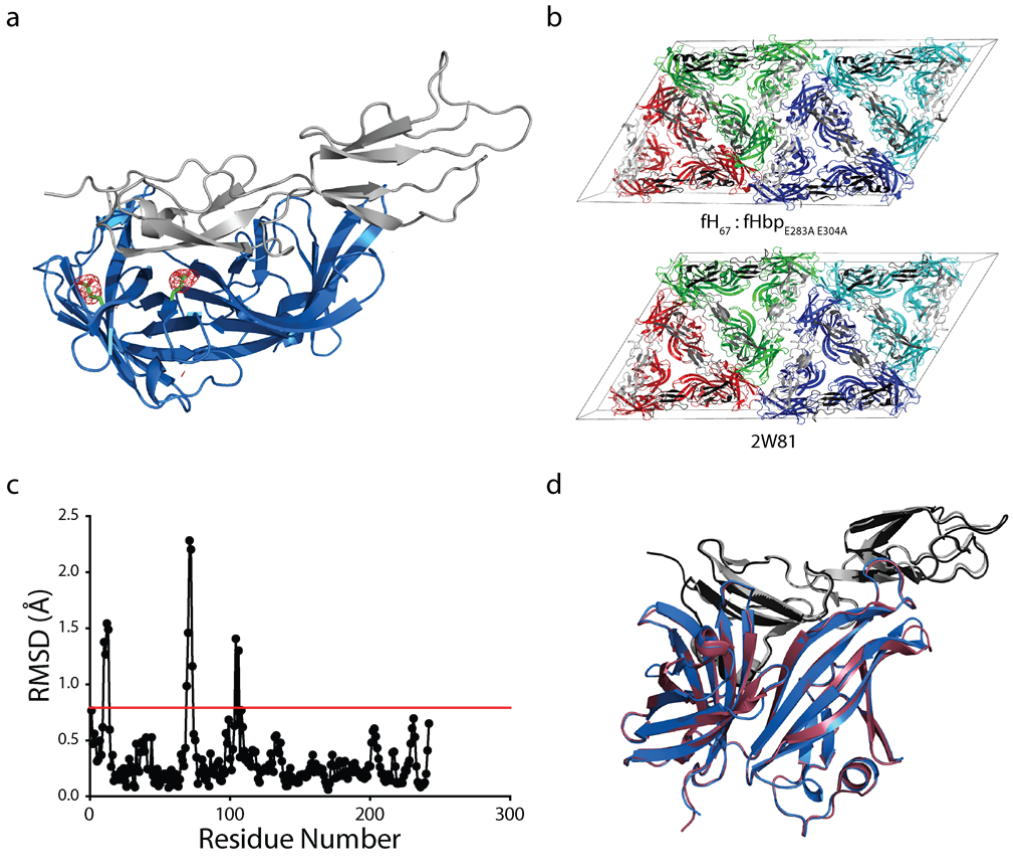


Figure 2.16 *Panel a*: Cartoon structure of the fH_{67}:fHbp structure with an F_O-F_C map calculated from the the fH_{67}:$fHbp_{E283A\ E304A}$ data contoured at -3 sigma. The only differences between the structures is the absence of the Glu283 and Glu304 sidechains. *Panel b*: Identical crystal packing between the two complexes, coloured by the contents of the asymmetric unit. *Panel c*: Average per residue RMSD between both structures. The average deviation between NCS related copies is shown in red. *Panel d*: Superposition of both structures.

Analysis of the binding interface for both complexes using the protein interfaces, surfaces and assemblies service (PISA) at the European Bioinformatics Institute is shown in Table 2.6. The fH_{67}:$fHbp_{E283A\ E304A}$ complex has a slightly increased surface area which is probably due to a slightly altered conformation caused by the loss of two

major single interactions. However, the complexation significance score (CSS) is dramatically reduced and the free energy of solvation (Δ^iG) has also reduced as expected due to the loss of two major interfacing interactions. There has also been an increase in the probability that the interface has occurred by random (signified by the Δ^iG p-value), reflecting the drastically reduced affinity between fH_{67} and $fHbp_{E283A\ E304A}$. Although the affinity of the fH_{67}:$fHbp_{E283A\ E304A}$ complex is significantly lower (3 μM) than that with the native fHbp (2 nM), the concentration of the crystallisation stock protein solution is approximately 200 μM driving the complex together and allowing crystallisation in the same condition as the original structure.

	4AYE - AD	4AYE - BC	4AYE - EF	4AYE-Average	2W81-Average
CSS	0.163	0.163	0.163	0.163	1.0
Surface area (Å²)	1475.3	1314.4	1415.0	1401.6	1335.6
Δ^iG (kcal/mol)	-4.7	-5.5	-5.1	-5.1	-5.5
Δ^iG P-value	0.466	0.416	0.424	0.435	0.38

Table 2.6 Pisa analysis of each copy of the fH_{67}:$fHbp_{E283A\ E304A}$ complex in the asymmetric unit. Also shown is the average values for each structure.

2.6 Discussion

Neisseria meningitidis binds complement regulator factor H from serum and presents it on its surface in a manner akin to host-cells resulting in evasion of complement-mediated killing allowing increased serum survival resulting in sustained infection. Sequestration of factor H is orchestrated by a surface protein, fHbp, that binds the complement regulator in the region of SCR domains six and seven. The affinity of this interaction has been measured giving a K_D in the region of 2 nM. This is by far the tightest interaction yet measured with factor H, the affinity of whose interactions normally reflect its micromolar concentration in the bloodstream, suggesting that every fHbp molecule presented on the meningococcal surface will be in complex with the complement regulator. This could have profound consequences upon meningococcal

entry to the vasculature as the local serum concentration of factor H may be depleted resulting in increased complement-mediated damage of host cells which could contribute to the petechial rash often indicative of meningococcal infection.

Scrutiny of the structure of fH_{67} in complex with fHbp shows that the interface occurs over the glycosaminoglycan binding groove in fH_{67}. Competition assays with heparin, a commonly used GAG analogue, show decreased fHbp binding to fH_{67} confirming that they share the same binding site. Mutation of factor H residues implicated in GAG binding, R341 and H337, to alanine results in a dramatic decrease in fHbp binding indicating that these residues are important to both interactions and that fHbp uses charged residues to mimic the charged sugars that localise Factor H on host cell surfaces, but are absent on most bacteria.

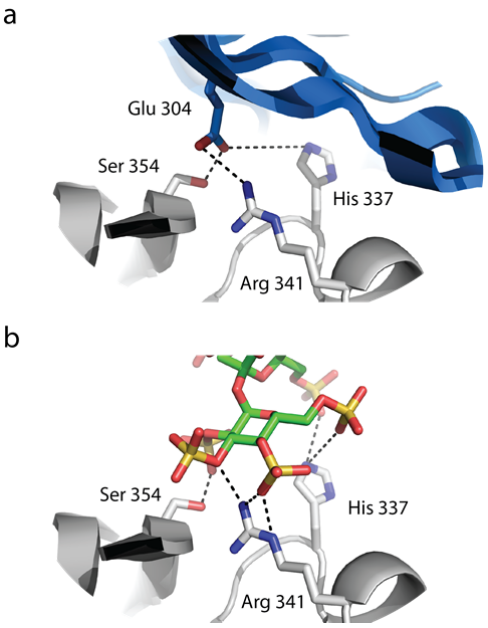


Figure 2.17 *Panel a*: fHbp Glu304 interacts with three residues on factor H. *Panel b*: The same three residues on factor H interact with charged sugars.

That fHbp is a protein mimic of host glycosaminoglycans is supported by a decrease in affinity for factor H upon mutation of charged residues, E283 and E304, that form intermolecular hydrogen bonds, or salt bridges, to hydrophobic alanine. Structural analyses of these residues show that they bind to the same fH residues as SOS, another GAG analogue, in the crystal structure of factor H domains six through eight (Figure 2.17). The X-ray crystal structure of $fHbp_{E283A\ E304A}$ in complex with fH_{67} has been

solved in order to verify that introduction of the mutations has not affected the structure of fHbp.

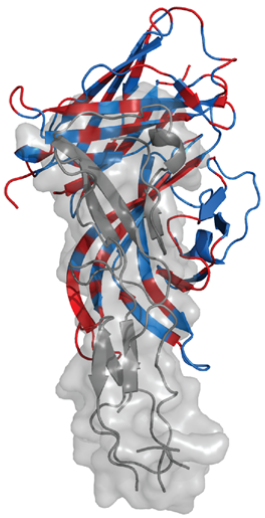


Figure 2.18 Residues conserved between the fHbp variants are in blue, those that differ are in red. fH_{67} is shown in grey.

All conclusions drawn so far have been made using fHbp variant one. Measurement of the affinity between fH_{67} and fHbp variants two and three are reported in Appendix 5. All three fHbp variants bind Factor H with a similar affinity (~2 nM), despite considerable sequence differences within the binding interface described in Figure 2.18. Conservation of this affinity reinforces its importance to the underlying mechanism of complement evasion and advocates that meningococcal strains expressing each fHbp variant are equally capable of scavenging the complement regulator from serum. This suggests that variation in the serum survival ability between different meningococcal strains is probably due to differing expression levels of fHbp on the bacterial cell surfaces. Of note is a protein from *Neisseria gonorroheae*, NGO0033, which is 90% identical to fHbp variant 3, yet has an affinity for factor H of 5.6 ± 0.1 μM (see Appendix 6), suggesting that it doesn't participate in the same complement evasion mechanism in the gonococcus and may supplement another mechanism, or provide another function altogether. To this end, work is currently underway to see whether NGO0033 is presented on the gonococcal surface.

fHbp shows promising potential as a viable vaccine candidate against all strains of meningococcus, eliciting bactericidal antibodies. In light of the high affinity of fHbp for factor H, vaccination with the wild type protein could decrease serum factor H levels

resulting in increased host-cell damage, or dis-regulatory effects. Vaccination with fHbp may also lead to increased numbers of autoantibodies against factor H as binding of the Neisserial protein may act as an adjuvant for eliciting α-fH antibodies. Autoantibodies against factor H have been associated with a variety of diseases [Dhillon, Wright *et al.* 2010; Strobel, Hoyer *et al.* 2010; Foltyn Zadura, Zipfel *et al.* 2012] and would not be an acceptable side effect of vaccination. $fHbp_{E283A\ E304A}$ has the potential to be a new vaccine candidate as the crystal structure shows that it is correctly folded and demonstrates decreased factor H binding, reducing the risk of factor H depletion and generation of autoantibodies. Currently, mapping of known antibody recognition epitopes onto the fHbp structure shows them clustering away from the factor H binding site. Vaccination with $fHbp_{E283A\ E304A}$ may result in antibodies elicited against the previously obscured interface which may have higher bactericidal activity due to blockage of factor H binding leading to increased complement-mediated lysis.

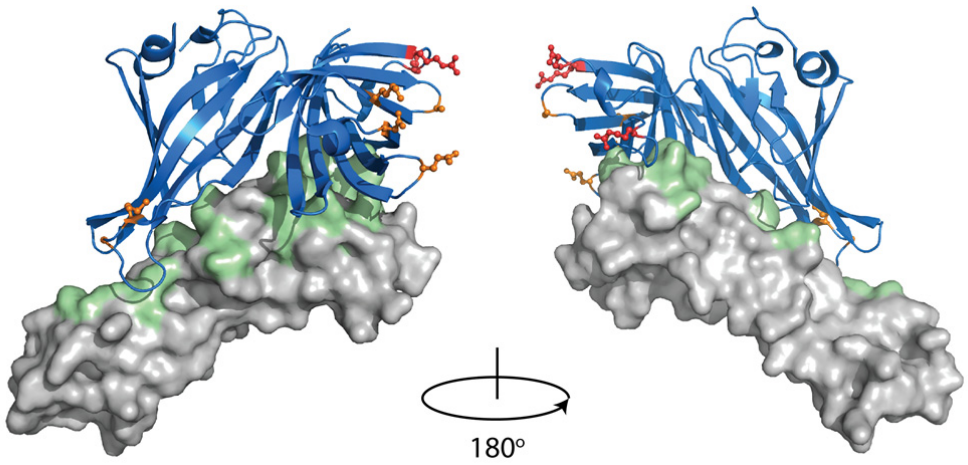


Figure 2.19 Antibody epitopes mapped on fHbp. fH_{67} is shown as a grey surface with the fHbp interaction interface in green. Residues eliciting bactericidal antibodies are shown in red CPK [Giuliani, Santini *et al.* 2005; Cantini, Savino *et al.* 2006]. Epitopes eliciting antibodies that block factor H binding are shown in orange CPK[Beernink and Granoff 2008].

2.7 Summary

The work presented in this chapter represents a significant advancement in knowledge of how *Neisseria meningitidis* evades the complement system by recruitment of factor H using fHbp. Robust measurement of the affinity between fHbp and fH has been presented, describing the tightest interaction with factor H yet observed with a K_D of approximately 2 nM. Data are also presented showing that fHbp mimics the glycosaminoglycans present on host cells surfaces, describing the full mechanism of complement evasion by replicating the interactions that protect host cells. The major fHbp residues involved in this mimicry have been mutated and the X-ray crystal structure of the mutant protein obtained. This mutant protein has a vastly reduced affinity for factor H and may be a viable vaccine candidate against bacterial meningitis.

2.7.1 Future directions

Future work should be undertaken to fully characterise the fHbp variant two and three interactions with factor H alongside large scale mutagenesis of the residues in the interface to observe if any other mutations ablate binding. Work should also be carried out to investigate the vaccine potential of the fHbp variants, and their interface knockout mutants, for generating a universal vaccine against all strains of *Neisseria meningitidis*. NGO0033 should also be investigated further to determine whether it is present on the Gonococcal surface, what its role is and whether it could be a possible candidate for a vaccine against *Neisseria gonorrhoea.*

Chapter Three: Binding of complement regulator factor H by CspA of *Borrelia burgdorferi*

Ticks of the *Ixodes* genus possess a post-embryonic life-cycle that progresses from larvae through nymph to adult with each stage requiring one blood meal from a vertebrate host [Bykowski, Woodman *et al.* 2008]. *Borrelia burgdorferi* can be acquired by feeding larvae from an infected host and colonises the tick midgut through development into the nymph stage. The bacteria can then be transferred back into a mammalian reservoir during feeding of nymphs, completing the zoonotic cycle [Stanek and Strle 2003]. When an infected *Ixodes* nymph feeds on a human host, the resulting infection may lead to presentation of Lyme's disease [Wormser 2006].

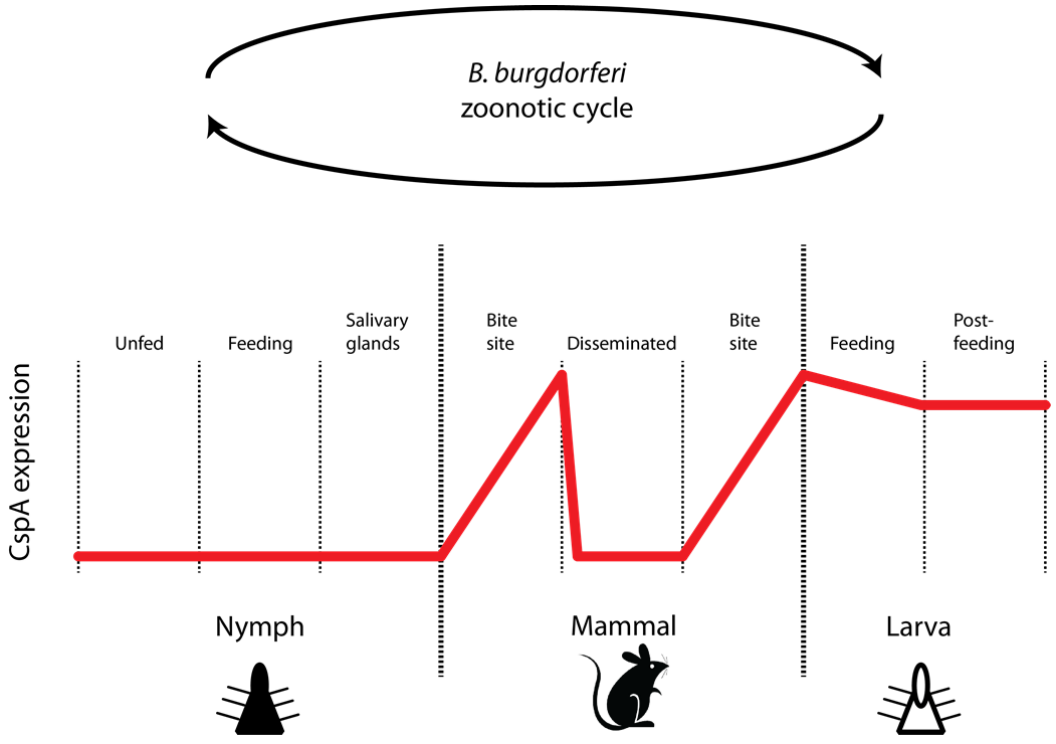


Figure 3.1 Expression profile of CspA throughout the zoonotic cycle of *Borrelia burgdorferi.*

CspA is expressed on the surface of *Borrelia burgdorferi* during transmission from tick to mammal and *vice versa* (Figure 3.1) [Bykowski, Woodman *et al.* 2008], aiding evasion of the complement response via recruitment of factor H in a similar manner to fHbp. Although the structure of CspA has been solved and the factor H binding site mapped to the cleft between the two monomer subunits [Cordes, Roversi *et al.* 2005; Cordes, Kraiczy *et al.* 2006], the symmetry of the proposed binding interface questions whether one or two molecules of factor H are bound.

Presented here is biophysical analysis of CspA confirming it forms a dimer and determination that factor H domain five is required for high affinity binding to the Borrelial protein. Measurement of the affinity between CspA and fH_{567} has been performed, showing great similarity to that determined between the Neisserial protein fHbp and fH with a K_D of 2 nM. Attempts have been made to obtain the atomic resolution structure of the complex between CspA and fH_{567}, yielding a new X-ray crystal structure of the CspA dimer showing flexibility between the monomer subunits which could be implicated in fH recruitment.

3.1 Protein expression and purification

3.1.1 Details of plasmids pCspA, pET-fH_{567} and pET-fH_{678}

Vector pCspA was kindly supplied by Professor Reinhard Wallich (University of Heidelberg, Germany) and encodes residues 70 - 250 of CspA isolated from *Borrelia burgdorferi* strain ZS7 expressed as a fusion protein with an N-terminal glutathione-S-transferase under the control of an inducible P_{tac} promoter (Figure 3.2). The deleted residues (1-69) contain the cell surface adhesion motif and are not required for factor H binding [Kraiczy, Hellwage *et al.* 2004]. These residues are also unstructured, with no electron density observed in the crystal structure presented by Cordes *et al.* (2005).

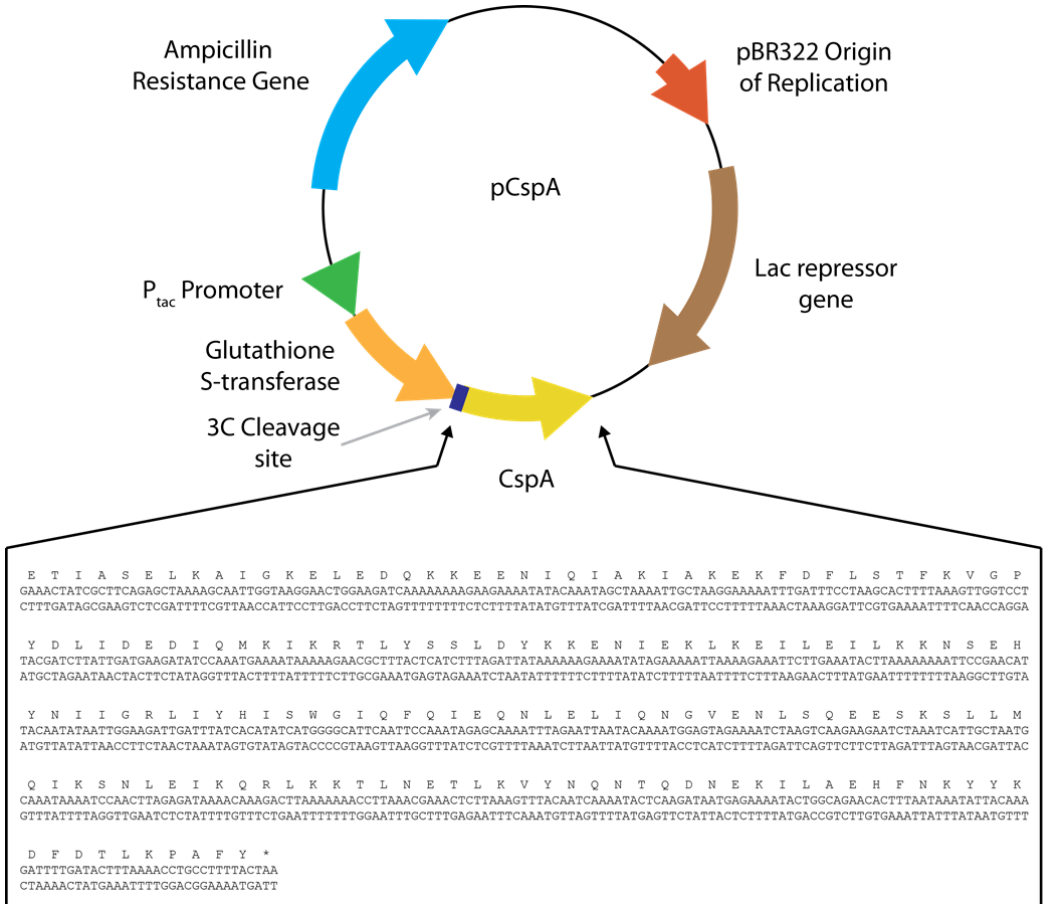


Figure 3.2 Schematic representation of pCspA. CspA (lacking residues 1-69) is expressed as an N-terminal GST fusion protein which can be removed using thrombin digestion. The vector was produced by cloning of the CspA construct into pGEX-2T (GE Healthcare) between the BamHI and EcoRI restriction sites.

Vectors pET-fH$_{567}$ and pET-fH$_{678}$ were produced by former members of Professor Lea's research group (detailed in Prosser, Johnson *et al.* (2007b) and Clark, Higman *et al.* (2006b)) and contain domains five through seven and six through eight respectively under the control of a T7 promoter (Figure 3.3 and Figure 3.4).

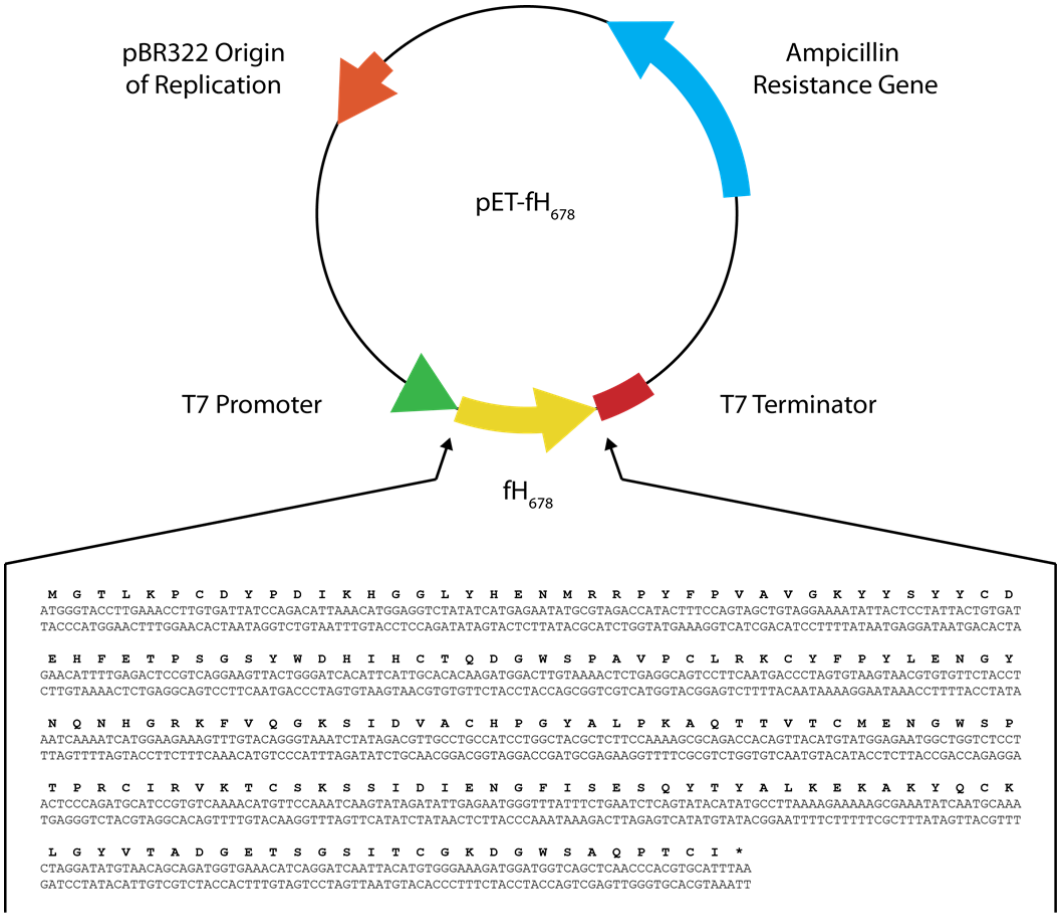


Figure 3.3 Schematic representation of pET-fH$_{678}$. The gene encoding factor H domains six to eight is under the control of an uninducible T7 promoter. The vector was produced by cloning of the fH$_{678}$ construct into pET-14b (Novagen) between the NcoI and BamHI restriction sites.

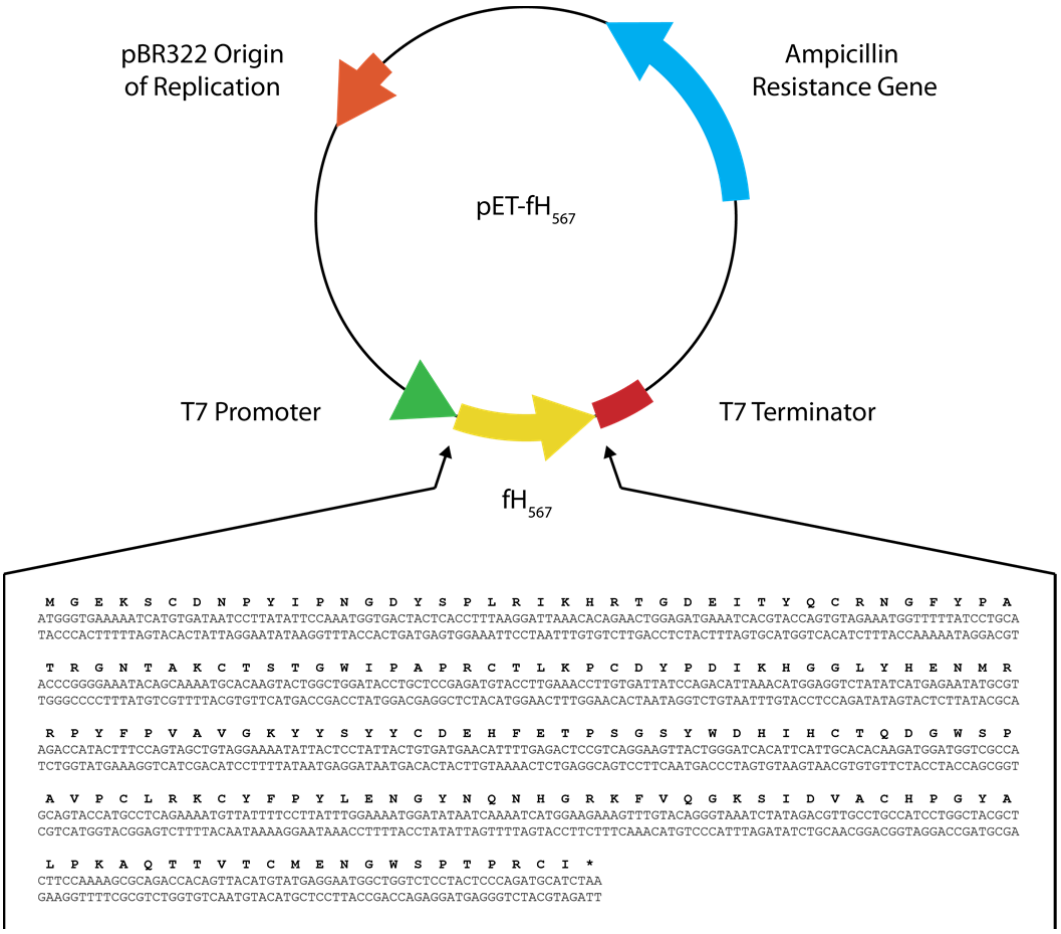


Figure 3.4 Schematic representation of pET-fH$_{567}$. The gene encoding factor H domains five to seven is under the control of an uninducible T7 promoter. The vector was produced by cloning of the fH$_{567}$ construct into pET-14b (Novagen) between the NcoI and BamHI restriction sites.

3.1.2 Multiple angle light scattering

Multiple angle light scattering (MALS) is a technique that can be used to determine the average molecular weight of a protein. The intensity of monochromatic laser light scattered by a sample is measured at multiple angles, allowing calculation of the root mean square hydrodynamic radius. This radius, coupled with an accurately measured refractive index, can then be used to obtain the molecular weight in an orientation independent manner.

MALS can also be used with a constant flow-cell, allowing measurement of the mass distribution across the elution profile of an analytical size exclusion chromatography column. This is the setup used throughout this thesis. 100 µl of sample (1 mg/ml) was injected onto an S200 16/60 column (GE Healthcare, USA) and the elution monitored using a Dawn Helios II (Wyatt Technology, USA) and an Optilab T-rEX (Wyatt Technology, USA) to measure the scattering and refractive index respectively. All data were analysed using ASTRA (Wyatt Technology, USA). Within this thesis, all samples were kindly run and analysed with the assistance of Dr. Steven Johnson.

3.1.3 CspA expression and purification

Escherichia coli strain BL21(DE3) transformed with the pCspA plasmid was grown overnight at 37^{o}C with 180 RPM shaking in 10 ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 2 × 1 L Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37^{o}C with 180 RPM shaking. Cell proliferation was monitored using absorbance at 600 nm and expression of CspA was induced in mid-log phase (OD_{600} = 0.6) by addition of 1 mM IPTG followed by continued incubation at 37^{o}C with 180 RPM shaking for four hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4^{o}C) before the pellets were combined and stored at -20^{o}C until required.

Defrosted cells were resuspended in 40 ml of 100 mM Tris, 300 mM NaCl, pH 8.0 and one EDTA-free protease inhibitor cocktail tablet (Roche, Germany) added. Cells were lysed by five passes through an Emulsiflex C5 homogeniser (Avestin, Canada) and the lysate cleared by centrifugation (48000 g, 20 minutes, 4^{o}C). Filtered supernatant was loaded on a 1ml GSTrap FF column (GE Heathcare, USA) pre-equilibrated with 50ml of 100 mM Tris, 300 mM NaCl, pH 8.0. The loaded GSTrap was washed with 10 ml of 100 mM Tris, 300 mM NaCl, pH 8.0 and 10 ml of 100 mM Tris, 1 M NaCl, pH 8.0 to

remove non-specifically bound contaminants before reequilibration with 20 ml of 50 mM Tris, 150 mM NaCl, 2.5 mM $CaCl_2$, 0.1% (v/v) BME, pH 7.5. 1 ml Thrombin solution (80 U/ml in 50 mM Tris, 150 mM NaCl, 2.5 mM $CaCl_2$, 0.1% (v/v) BME, pH 7.5) was loaded onto the column and incubated at 21°C overnight.

A 1 ml Benzamidine FF column (GE Healthcare, USA) pre-equilibrated with 50 ml of 100 mM Tris, 1 M NaCl, pH 8.0 was connected in series following the GSTrap and cleaved CspA eluted with 20 ml of 100 mM Tris, 1 M NaCl, pH 8.0. The eluate was concentrated to 5 ml using a 3000 MWCO PES spin concentrator (Millipore, UK) prior to a final purification step on an S75 16/60 size exclusion chromatography column (GE Healthcare, USA) pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 7.5 at 1 ml/minute connected to an AKTA purifier FPLC system (GE Healthcare, USA). An example A_{280nm} trace is shown in Figure 3.5 alongside an SDS-PAGE gel charting the course of the purification.

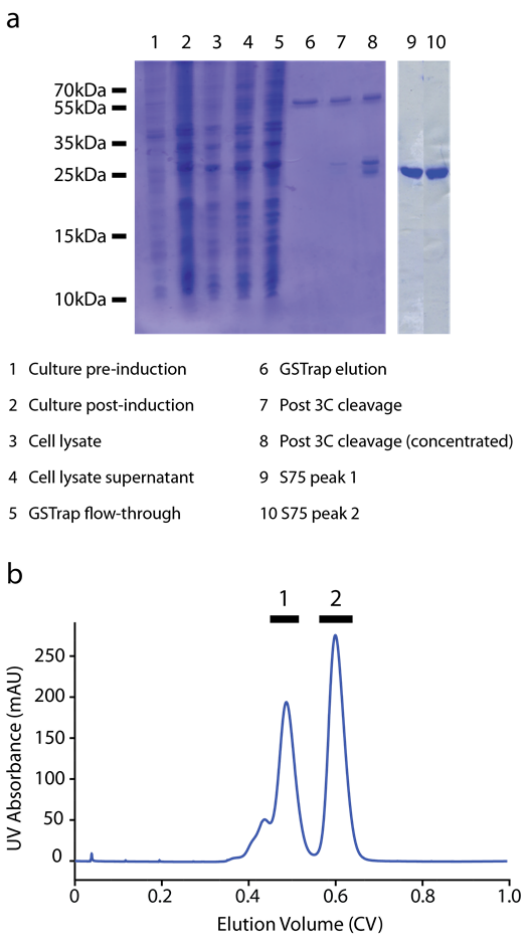


Figure 3.5 *Panel* a: 15% reducing SDS-PAGE gel charting the course of the CspA purification. *Panel* b: Chromatogram detailing the elution profile of CspA from an S75 16/60 column. The target protein elutes in two peaks at 0.45 CV and 0.6 CV which are labelled, detailing the fractions collected with a black bar.

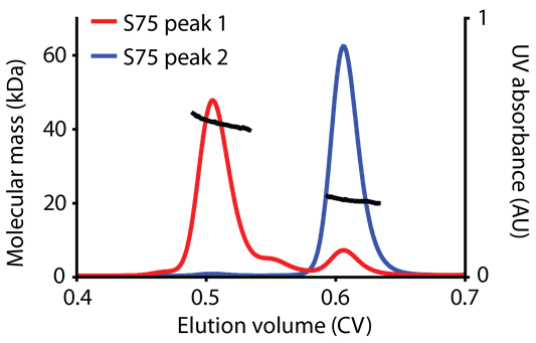


Figure 3.6 MALS analysis of each peak containing CspA from from Figure 3.5, panel b. The UV trace for each sample is shown in red or blue with the mass distribution across each peak in black.

Both major peaks from the S75 16/60 (marked in Figure 3.5, panel b) contain CspA (Figure 3.5, panel a). Analysis of peaks at 0.45 CV and 0.6 CV using multi-angle laser light scattering reveals that they contain CspA in dimeric and monomeric forms respectively, confirming the findings of Cordes *et al.* (2005) that CspA forms dimers in solution under these conditions (Figure 3.6). The published study also suggested that the monomeric and dimeric forms of CspA do not readily interconvert which is supported by the lack of observation of an equilibrium between monomer and dimer in each peak from the size exclusion column when analysed using MALS.

3.1.4 fH_{567} and fH_{678} expression and purification

fH_{567} and fH_{678} were expressed in *Escherichia coli* and refolded from inclusion bodies in an identical manner to fH_{67} detailed in Chapter 2.2. Example size exclusion traces and SDS-PAGE gels charting the purifications are presented in Figure 3.7. Typical yields of fH_{567} and fH_{678} were 0.1 mg/l and 1 mg/l of bacterial culture respectively. The low yield of fH_{567} is likely due to low solubility of the refolded material. During concentration precipitation occurs (demonstrated by the loss of protein between lanes 10 and 11 in Figure 3.7, panel a) which is non-reversible and results in loss of protein. Attempts to reduce this loss were unsuccessful. In a similar manner to fH_{67}, both fH_{567} and fH_{678} interact with the Superdex resin during the size exclusion chromatography purification, resulting in elution volumes that are greater than 1 CV (Figure 3.7, panels b and c).

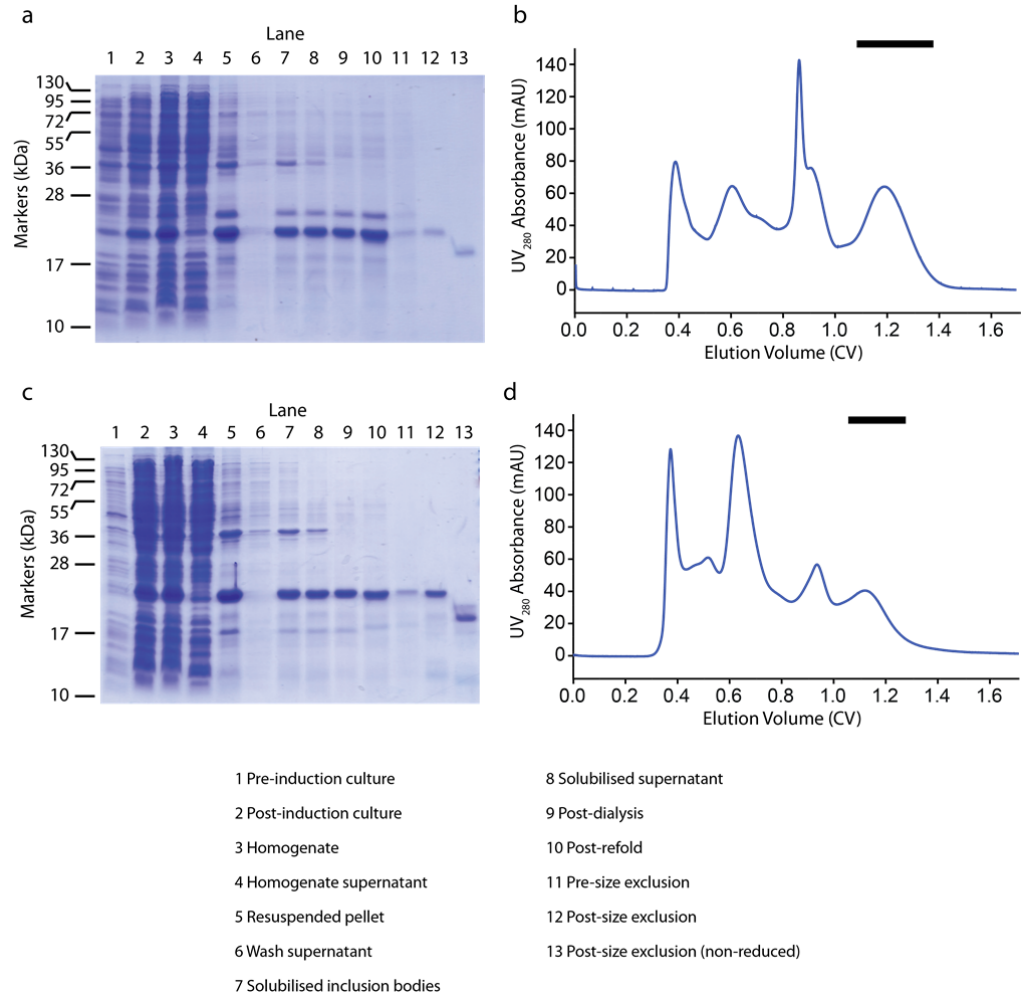


Figure 3.7 *Panel a*: 15% reducing SDS-PAGE gel charting the preparation of fH_{567}. *Panel b*: Chromatogram detailing the elution profile of fH_{567} from an S75 16/60 column. The target protein elutes at approximately 1.2 CV in the peak highlighted with a black bar. *Panel c*: 15% reducing SDS-PAGE gel charting the preparation of fH_{678}. *Panel d*: Chromatogram detailing the elution profile of fH_{678} from an S75 16/60 column. The target protein elutes at approximately 1.15 CV in the peak highlighted with a black bar

3.2 Determination of the factor H domains involved in CspA binding

3.2.1 Surface plasmon resonance

Surface plasmon resonance was employed in order to assess the binding of fH_{567} and fH_{678} to both the monomeric and dimeric forms of CspA. A Biacore 3000 (GE Healthcare, USA) instrument was used to covalently immobilise CspA to a CM5 sensor chip (GE Healthcare, USA) via primary amine coupling. The CM5 chip was activated with 50 µl EDC:NHS at 10 µl/min prior to immobilisation of each form of CspA, diluted into 10 mM sodium acetate pH 4.0 to a final concentration of 50 µg/ml, on

separate flow channels. The surface was deactivated by injection of 80 µl ethanolamine at 10 µl/min. A reference channel was mock activated and deactivated in an identical manner. Immobilisation resulted in immobilisation of approximately 1200 RU of CspA monomer and 1500 RU of CspA dimer.

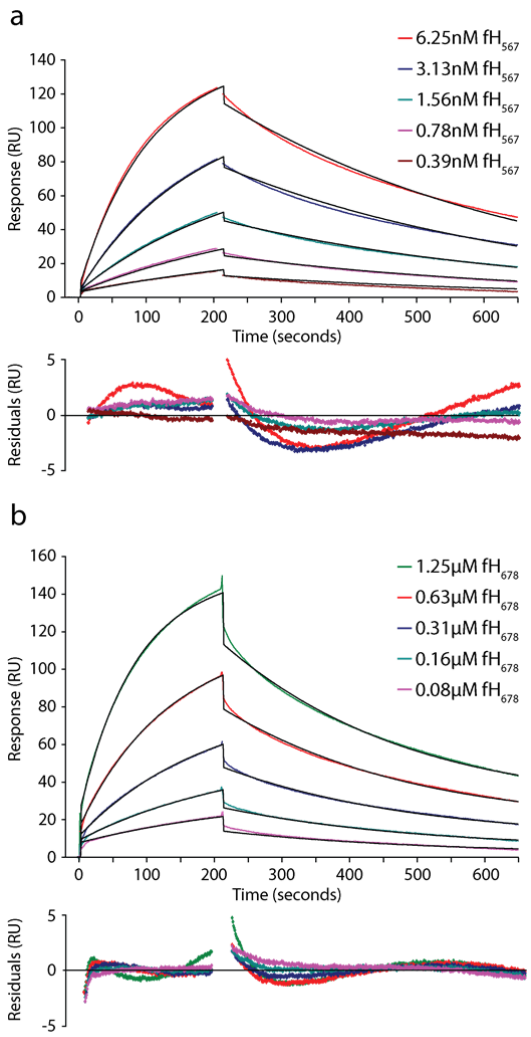


Figure 3.8 *Panel a*: Example SPR curves and 1:1 Langmuir fits allowing determination of the binding affinity of fH_{567} for the CspA dimer. *Panel b*: Measurement of the affinity of fH_{678} for the CspA dimer

Two separate dilution series of fH_{567} ranging between 0.2 nM and 100 nM were flowed over the chip using the KINJECT command with a flow rate of 40 µl/min, a dissociation time of 600 seconds and temperature of 25°C. In between each analyte injection the surface was regenerated using a 5 µl injection of 10 mM glycine pH 3.0. A further two dilution series of fH_{678} ranging between 0.02 µM and 10 µM were flowed over the chip in an identical manner. All the sensorgrams were blank subtracted and fit with a global Langmuir binding model assuming a 1:1 stoichiometry. Example curves with their corresponding fits are shown in Figure 3.8 alongside Table 3.1 detailing the measured affinities.

The higher affinities of CspA for fH_{567} over fH_{678} suggests that factor H domain five plays an important role in the interaction, but that the binding surface is not limited to

this domain and is spread onto at least domain six. There appears to be no observable difference in the affinity for either fH_{567} or fH_{678} between the monomeric and dimeric forms of CspA. This allows averaging of the affinities measured in all four dilution series giving K_Ds of 1.6 ± 0.1 nM and 376 ± 40 nM for the interaction of CspA with fH_{567} and fH_{678} respectively. The lack of any difference could be due to the state of the CspA on the chip surface. Unfortunately this is unknown as the low pH of the coupling and regeneration buffers may promote dissociation to the monomeric form, or the higher surface concentration achieved by coupling at low pH may encourage dimerization. Previous studies have measured the affinity of full length factor H to CspA to be between 10 nM and 30 nM, whereas no interaction was detected at all in the absence of dimerization, suggesting that the affinities measured are to the dimeric form of CspA.

	CspA Monomer		CspA Dimer	
	a	b	a	b
fH_{567}				
K_D (nM)	1.71	1.56	1.77	1.55
k_{on} ($M^{-1}s^{-1}$)	1.25×10^{6}	1.36×10^{6}	1.21×10^{6}	1.42×10^{6}
k_{off} (s^{-1})	2.13×10^{-3}	2.12×10^{-3}	2.14×10^{-3}	2.19×10^{-3}
X^2	2.01	1.02	1.53	1.13
fH_{678}				
K_D (nM)	409	338	410	345
k_{on} ($M^{-1}s^{-1}$)	7.24×10^{3}	6.64×10^{3}	7.72×10^{3}	6.77×10^{3}
k_{off} (s^{-1})	2.96×10^{-3}	2.25×10^{-3}	3.17×10^{-3}	2.34×10^{-3}
X^2	0.20	0.68	0.346	0.71

Table 3.1 Details of the affinity measurements of fH_{567} and fH_{678} for both monomeric and dimeric CspA from two separate dilution series labelled a and b.

3.2.2 Analytical gel filtration

The high affinity interaction between CspA and fH_{567} suggests that co-purification of a complex may be possible. Figure 3.9 shows elution profiles from an S200 10/300 size exclusion chromatography column (GE Healthcare, USA) for samples of fH_{567}, CspA, and an equimolar mixture of the two. An extra peak appears in the traces from the mixture which contains both species signifying that co-purification of a complex is achievable. However, attempts to scale up the purification were hindered by the

insolubility of the complex, resulting in loss of 90% of material during concentration and not providing a viable way to produce enough protein for crystallisation trials.

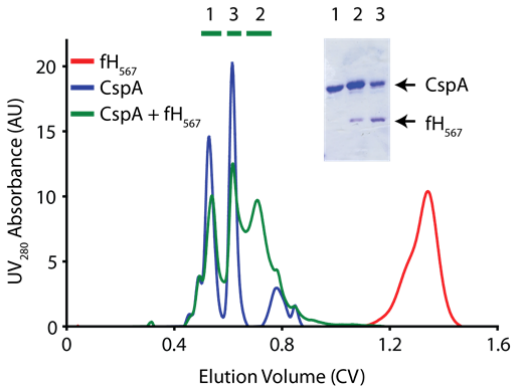


Figure 3.9 *Panel a*: Overlay of the elution profiles of fH_{567}, CspA and a mixture of the two from an S200 10/300 column with an inset 15% reducing SDS-PAGE gel of each peak from the mixture.

3.2.3 Stoichiometry of the CspA:fH interaction

Although a complex can be co-purified, initial attempts to determine the stoichiometry of the interaction between CspA and fH_{567} using MALS were unsuccessful. This was due to fH_{567} binding to the column matrix, both in the presence and absence of CspA, making a mass measurement impossible.

3.3 Crystallisation, data collection and processing

In order to further investigate the high-affinity binding interface between fH_{567} and CspA, an attempt was made to co-crystallise a complex between the interacting partners.

3.3.1 Crystallisation of the fH_{567}:CspA complex

Unfortunately, large-scale formation and purification of the CspA:fH_{567} complex was hindered by aggregation of the protein during concentration following size exclusion, a

property probably attributable to fH_{567} which also possesses limited solubility. In order to minimise losses CspA and fH_{567} were mixed in a 2:1 molar ratio respectively and concentrated to A_{280} = 5.20. Crystallisation trials were performed at 294 K using vapour diffusion in 400 nL sitting drops produced by an Oryx Nano crystallisation robot (Douglas Instruments, UK) within MRC crystallisation plates. Stura, Macrosol, PACT Premier, JCSG and MD 1+2 screens were used, with two drops set up for each condition containing protein to precipitant ratios of 1:1 and 1:3. Small trigonal-pyramidal crystals grew in a 1:1 mixture of 18% (w/v) PEG 8000, 5 mM zinc acetate, 100 mM sodium cacodylate pH 6.5 to protein solution (Figure 3.10, panel a). Crystals were mounted in Cryoloops (Hampton Research) prior to cryoprotection using 15% ethylene glycol and flash cooling in liquid nitrogen.

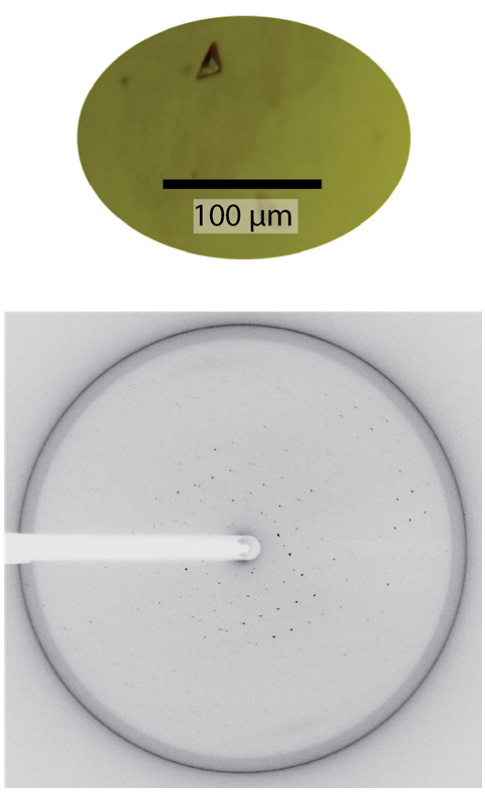


Figure 3.10 *Panel a*: Crystals grown of the fH_{567}:CspA complex. *Panel b*: Example diffraction image. The dark circle is caused by diffraction from ice crystals that have formed due to insufficient cryoprotection.

3.3.2 Data collection of fH_{567}:CspA complex

Data were collected on beamline ID23-2 at the ESRF (Grenoble, France) using the rotation method (Figure 3.10, panel b). Diffracting crystals were indexed in a C centred monoclinic cell (cell dimensions a = 115.3 Å, b = 44.2 Å, c = 186.5 Å, $\alpha = \gamma = 90.0^{o}$, $\beta = 90.7^{o}$) with C2 symmetry using IMOSFLM [Battye, Kontogiannis *et al.* 2011], which

was then used to calculate the most efficient data collection strategy. Data collection parameters from two crystals, which are subsequently used for structure solution, are presented in Table 3.2.

	Dataset 1	Dataset 2
Diffraction source	ESRF ID23-2	ESRF ID23-2
Detector	MAR 225	MAR 225
Energy (eV)	14209	14209
Wavelength (Å)	0.8726	0.8726
Temperature (K)	120	120
Detector resolution (Å)	3.7	3.5
Exposure per image (s)	2	1.5
Φ_{start}	50	64
Φ_{end}	140	153
$\Delta\Phi$	1	1
Number Images	90	90

Table 3.2 Data collection parameters from two separate crystals.

2.3.3 Data processing of the fH_{567}:CspA complex

	Combined datasets
Spacegroup	C2
Unit-cell dimensions (Å)	a=115.3, b=44.2, c=186.5
Unit-cell angles (°)	α=90, β=90.7, γ=90.0
Wilson B factor	72.9
Resolution (Å)	93.26-4.06 (4.16-4.06)
R_{merge}	0.194 (0.596)
R_{pim}	0.109(0.333)
<I/σ(I)>	5.2 (2.7)
Completeness (%)	95.9 (98.6)
Multiplicity	3.7 (3.7)

Table 3.3 Data processing statistics for a combined dataset of diffraction data from both crystals collected.

Data from two crystals (Dataset 1 and Dataset 2 in Table 3.2) were processed using XIA2 [Winter 2010] with invocation of the 3dii option which uses programs XDS [Kabsch 2010] for spot-finding, indexing and integration followed by scaling using XSCALE [Kabsch 2010]. Due to the limited resolution, and signs that the data quality reduced as exposure increased, it was found that merging the data from both crystals (done automatically using XIA2) resulted in a dataset with the best merging statistics and highest resolution of 4.1Å

(Table 3.3). The data quality reduction upon increased exposure could be attributed to arsenic in the cacodylate buffer which has a high X-ray absorption cross-section leading to increased radiation damage.

3.4 Structure determination and refinement

3.4.1 Refinement of PDB entry 1W33

Prior to commencement of molecular replacement using the published structure of CspA (PDB entry: 1W33 [Cordes, Roversi *et al.* 2005]), the deposited structure was analysed using MOLPROBITY [Chen, Arendall *et al.* 2010] and found to be only in the 14th percentile (see Table 3.5), probably due to the development of refinement packages and generation of rotamer libraries since publication. In order to produce the most accurate model for molecular replacement, 1W33 was re-refined against the structure factors given in the published mmCIF file (converted into MTZ format using CIF2MTZ [Winn, Ballard *et al.* 2011]). After three iterative cycles of rebuilding using COOT [Emsley, Lohkamp *et al.* 2010] and refinement in AUTOBUSTER [Blanc, Roversi *et al.* 2004; Bricogne, Blanc *et al.* 2011] using the autoncs flag, the model was improved to the 100th percentile (Table 3.5).

3.4.2 Structure solution of fH_{567}:CspA complex

Calculation of the Matthews coefficients for increasing numbers of fH_{567}:CspA complexes in the asymmetric unit using MATTHEWS [Matthews 1968] suggests the presence of two copies of the assembly between fH_{567} and a CspA dimer (solvent content = 55.3%). Analysis of the native Patterson does not contain any significant, non-origin peaks suggesting the absence of translational non-crystallographic symmetry. However, the self-rotation function contains peaks in the chi = 180^{o} section in addition to the crystallographic symmetry operators suggesting that the two copies

are related by a two-fold rotational non-crystallographic symmetry whose axis is orthogonal to the crystallographic two-fold.

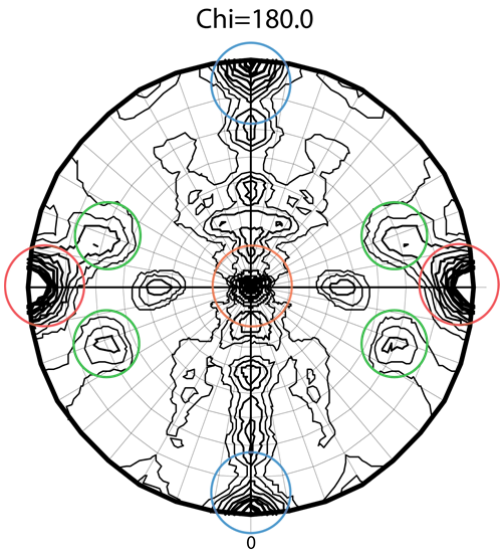


Figure 3.11 Chi = 180° self-rotation function calculated from the processed dataset using MOLREP [Vagin and Teplyakov 1997]. Peaks resulting from the crystallographic two-fold are highlighted in red. Other peaks are discussed later in Chapter 3.

Molecular replacement using chain A from the re-refined CspA model and the structure of factor H domains 6 and 7 (PDB entry: 2W81 [Schneider, Prosser *et al.* 2009]) was performed using PHASER [McCoy, Grosse-Kunstleve *et al.* 2007]. Four copies of CspA were successfully placed in the asymmetric unit (Table 3.4), but no placement of factor H domains was successful suggesting their absence from the crystal which was confirmed by visual analysis of the packing which shows two CspA dimers similar to that in PDB entry 1W33 in the asymmetric unit. The PHASER packing constraint had to be relaxed in order to place the third and fourth CspA copies which was explained by observation of the C-terminal residues clashing with the second monomer subunit. An improved model was produced from the re-refined CspA structure consisting of chain A residues 70 - 230 and chain B residues 231 - 249. This model was superimposed onto each chain prior to refinement.

	RFZ	TFZ	LLG
Copy 1	5.7	4.1	62
Copy 2	4.8	8.8	160
Copy 3	5.1	11.4	279
Copy 4	4.5	13.0	286

Table 3.4 Rotation function (RFZ) Z-scores, translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of four copies of CspA in the data.

3.4.3 Refinement of CspA

The new structure of CspA was refined iteratively, cycling between rounds of rebuilding in COOT and refinement using AUTOBUSTER with secondary structure target restraints [Smart, Brandl *et al.* 2008] to the re-refined CspA structure. Due to the low resolution, individual B-factor refinement would be unstable so the B-factors were not refined and were set to the Wilson B (72.9 $Å^2$). Refinement and validation statistics for the final model were calculated using MOLPROBITY and are shown in Table 3.5. The final model main chain, alongside representative electron density, is shown in Figure 3.12 and was deposited in the PDB with accession code 4ATR.

	1W33	1W33 Rerefined	New CspA structure
Resolution (A)	2.7-19.88	2.7-19.87 (2.70-2.89)	4.06-93.26 (4.06-4.54)
Number of reflections	15962	15961 (2916)	7626 (2160)
Number of reflections for R_{free}	810	810 (135)	353 (112)
R_{work}/R_{free}	0.224/0.262	0.2197/0.2556 (0.2594/0.2714)	0.2605/0.2699 (0.2655/0.2767)
Number of atoms			
Protein	3001	3001	6004
Ligand/Ion	12 glycerol	53 glycerol	0
Water	41	41	0
R.m.s deviations			
Bond lengths (A)	0.011	0.010	0.009
Bond angles (°)	0.958	1.17	1.13
Ramachandran plot analysis			
Residues in favoured regions (%)	93.56	97.20	97.75
Residues in disallowed regions (%)	1.12	0.00	0.00
Molprobity analysis			
Molprobity score	3.64	1.61	1.30
Molprobity score percentile	14th	100th	100th
Resolution range (A)	2.45 - 2.95	2.45 - 2.95	3.25 - 4.31

Table 3.5 Refinement and validation statistics for the deposited CspA structure (PDB entry: 1W33), the rerefined CspA structure, and the new structure presented here.

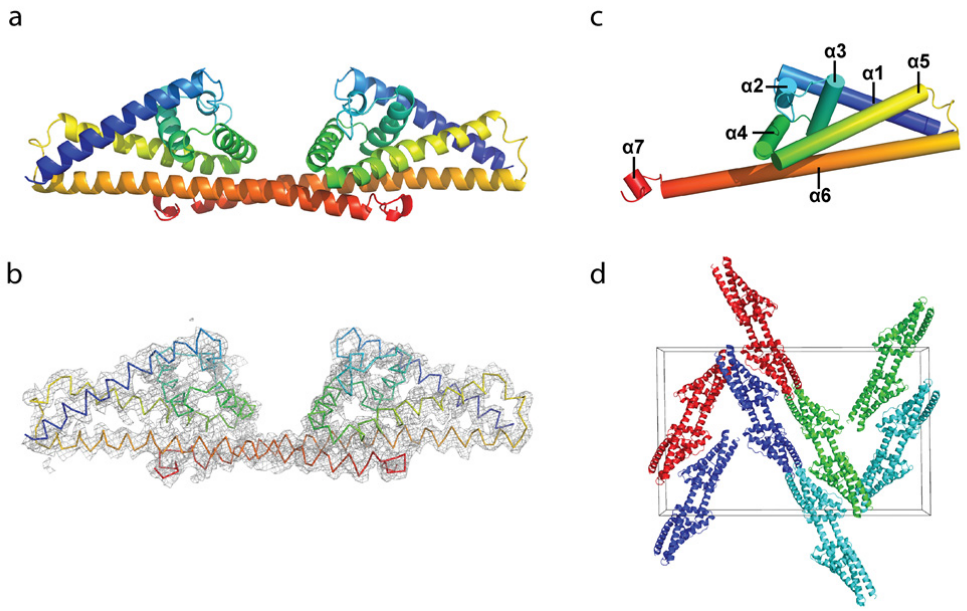


Figure 3.12 *Panel a*: Cartoon representation of the CspA dimer found in the new structure. Colours run from blue at the N-terminus to red at the C-terminus. *Panel b*: Ribbon tracing of the main chain with representative electron density contoured at 1.5σ. *Panel c*: Labelling of the individual helices found in the CspA monomer. *Panel d*: Packing in the unit cell, coloured by the contents of the asymmetric unit.

3.5 Structural analyses

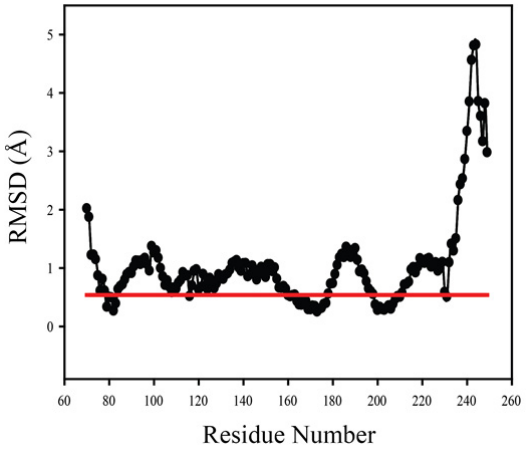


Figure 3.13 Mean RMSD per residue between the new structure of CspA and PDB entry 1W33 showing deviation at the C-terminus. The mean RMSD between the four copies of CspA in the asymmetric unit is shown in red.

Although the crystals did not contain the CspA:fH$_{567}$ complex, they have yielded a second structure of CspA in a new crystal form. The new structure again shows formation of a dimer with each monomer subunit being highly similar to the original structure except at the C-terminus (Figure 3.13). Analysis using DYNDOM [Hayward and Berendsen 1998] shows that helix six is bent between residues 225 and 227 (Figure 3.14, panel d) which manifests itself as

an increase in the angle between monomer subunits; leaving the structure of the ten C-terminal residues (240 - 250) unaffected (Figure 3.14, panel b). Measurement of the difference in inter-domain angles between the new structure and PDB entry 1W33 was performed using secondary structure matching over residues 70-220 with LSQKAB [Kabsch 1976] showing an average increase in inter-subunit angle of 16.8^{o} which is illustrated in Figure 3.14, panel f.

	4ATR - AB	4ATR - CD	4ATR - Average	1W33 -Average
CSS	1.0	1.0	1.0	0.945
Surface area (Å^2)	2283.5	2279.8	2281.7	2266.8
Δ^{i}G (kcal/mol)	-36.6	-37.5	-37.0	-31.4
Δ^{i}G P-value	0.015	0.011	0.013	0.044

Table 3.6 Pisa analysis of each copy of the CspA dimer in the asymmetric unit. Also shown are the average values for both the new structure and PDB entry 1W33.

Analysis of the dimerization interfaces in 4ATR and 1W33 was performed using the protein interfaces, surfaces and assemblies service (PISA) at the European Bioinformatics Institute [Krissinel and Henrick 2007] and is presented in Table 3.6. However, due to the low resolution of the new CspA structure (PDB entry: 4ATR) the positions of the interfacing side chains may not be exactly representative of their correct positions, perturbing the calculated values. Thus comparison of minor differences in the calculated values from PDB entries 4ATR and 1W33 should not be counted as significant.

	ω	φ	χ
Chain A - Chain B	100.7	69.1	180.0
Chain A - Chain C	87.8	0.0	170.5
Chain A - Chain D	162.8	-90.5	139.6
Chain C - Chain D	68.7	-69.0	179.1

Table 3.7 Rotational transformations between copies of the CspA monomer in the asymmetric unit.

Calculation of the rotational transformations, in spherical polar coordinates, between all four CspA copies in the asymmetric unit using SUPERPOSE [Krissinel and Henrick 2004] (shown in Table 3.7), allows interpretation of

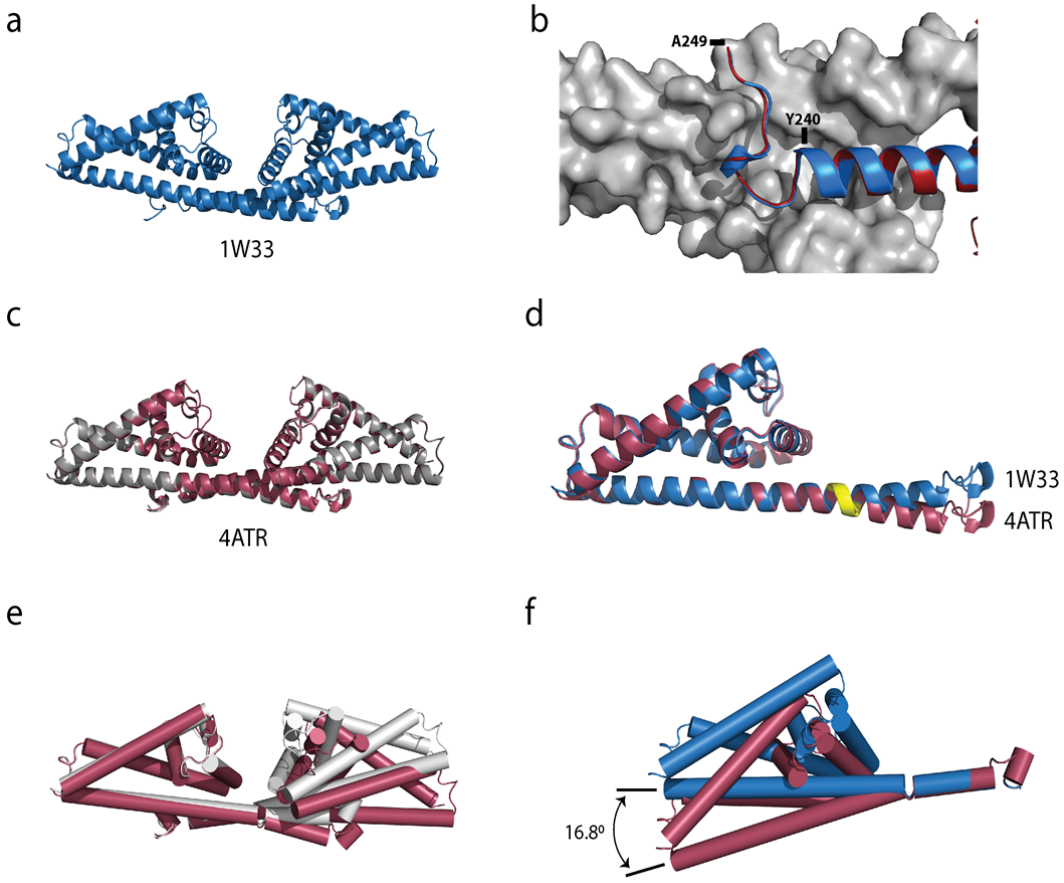


Figure 3.14 *Panel* a: Cartoon representation of the structure of the CspA dimer found in PDB entry 1W33. *Panel* b: Superposition (RMSD = 0.138 Å) of C-terminal residues 230-250 from chain A in both the new structure (red) and 1W33 (blue) shown as a cartoon against chain B rendered as a surface (grey). *Panel c*: Cartoon representation of the structures of both copies of the CspA dimer found in the asymmetric unit of the new structure. The dimer formed by chains C and D (grey) has been superimposed on that formed by chains A and B (red) using secondary structure matching with an RMSD of 0.264 Å. *Panel d*: 1W33 (blue) and 4ATR (red) monomer subunits superposed showing deflection of helix F between residues 225 and 227 (shown in yellow). *Panel e*: Overlay of dimer assemblies found in 1W33 and 4ATR showing difference in inter-monomer angle. Chain A from 1W33 (grey) and 4ATR (red) were superposed using secondary structure matching over residues 70-220. The average RMSD of superposition of each chain in 1W33 onto each chain in 4ATR over this residue range is 0.539 Å. *Panel f*: Superposition of C-terminal residues from chain A of 1W33 and 4ATR showing an average increase in inter-monomer angle of 16.8° averaged over both copies of the dimer in the 4ATR asymmetric unit (individual angles for the chain AB and CD dimers are 17.0° and 16.7° respectively).

the self-rotation function in Figure 3.11. The non-crystallographic two-fold peaks at omega ~ 90, phi = 0 (shown in blue) are generated by mapping chain A onto chain C, and transforming the resulting peak by the crystallographic two-fold rotation. The peaks in green are generated by mapping chain A onto B, and C onto D, and then transforming through the crystallographic two-fold axis. The peak at omega = 0, phi = 0 (shown in

orange) probably arises from superposition of the separate helices onto themselves and each other.

3.6 Discussion

Borrelia burgdorferi employs a mechanism of complement evasion that is orthologous to factor H recruitment by *Neisseria meningitidis* using fHbp. The spirochete binds complement regulator factor H from serum and presents it on its surface using a family of proteins termed complement regulator acquiring surface proteins (CRASPs) [Kraiczy, Skerka *et al.* 2001c]. It has been hypothesised that multiple members of this family are required in order to provide enhanced protection against the different immune systems across the range of reservoir hosts of the bacterium [Wallich, Pattathu *et al.* 2005]. To date, the structure of only one of the members of this family, CspA, has been determined and shows formation of a dimeric species, within which the factor H binding interface has been mapped to the cleft between the subunits [Cordes, Roversi *et al.* 2005; Cordes, Kraiczy *et al.* 2006].

Although CspA and fHbp differ greatly at the sequence and structural level, they both bind factor H around the minor GAG binding region located between SCR domains five and seven with an affinity of approximately 2 nM. Evolution of separate molecules for factor H binding with the same affinity in two diverse species underlines the importance of the magnitude of the interaction upon this mechanism of complement evasion. The high serum concentration of factor H (approximately 1.5 - 3 μM) suggests that a weaker interaction could still provide a similar function, whereas mutagenesis studies on fHbp has shown that the interaction can be tightened by mutation of a single residue [Johnson *et al*. 2012]. These findings suggest that an affinity of ~2 nM is critical for effective scavenging of factor H, yet is not too tight, perhaps to allow removal of factor H under conditions when it is not advantageous to present it on the surface. As CspA is only

expressed during transmission to and from mammalian hosts and not during mammalian dissemination (Figure 3.1), it is possible that the role of CspA is to aid in bacterial evasion of the increased complement response around the tick bite site. This is supported by the finding that the efficacy of borrelial infection does not appear to be changed in a fH knockout mouse upon intravenous administration of the infectious agent [Woodman, Cooley *et al.* 2007].

Binding factor H to a cell surface between domains five and seven is critical to the mechanism of prevention of C3b deposition on both host and pathogenic cell surfaces. On host cells, the region under scrutiny is bound to the cell surface via interactions with glycosaminoglycans leaving the four N-terminal SCR domains free to regulate active C3 convertases but within a much closer radius of the surface than if bound by the C-terminus resulting in more efficient regulation. *N. meningitidis* and *B. burgdorferi* have mimicked this system, binding in the same region to give maximal evasion of C3b deposition at the surface. It has already been shown that fHbp is a GAG mimic, but it is still unknown whether CspA utilises the same charged sugar-binding residues, or has evolved a distinct binding interface elsewhere in domains five through seven. Future work to continue attempts at crystallisation of a complex, as well as heparin competition assays should be performed to try to elucidate more details of this mechanism.

Comparison of the new CspA structure with 1W33 suggests that the dimeric assembly possesses a degree of flexibility between the monomer subunits resulting in a widening of the cleft compared to the original structure. Opening of the cleft may result in increased access to the putative binding site suggested by the mutagenesis data presented by Zipfel *et a.l* (Figure 3.15). It may also be possible for the CspA dimer to "clamp" around a factor H molecule bound within the cleft, contributing to the high affinity of the interaction

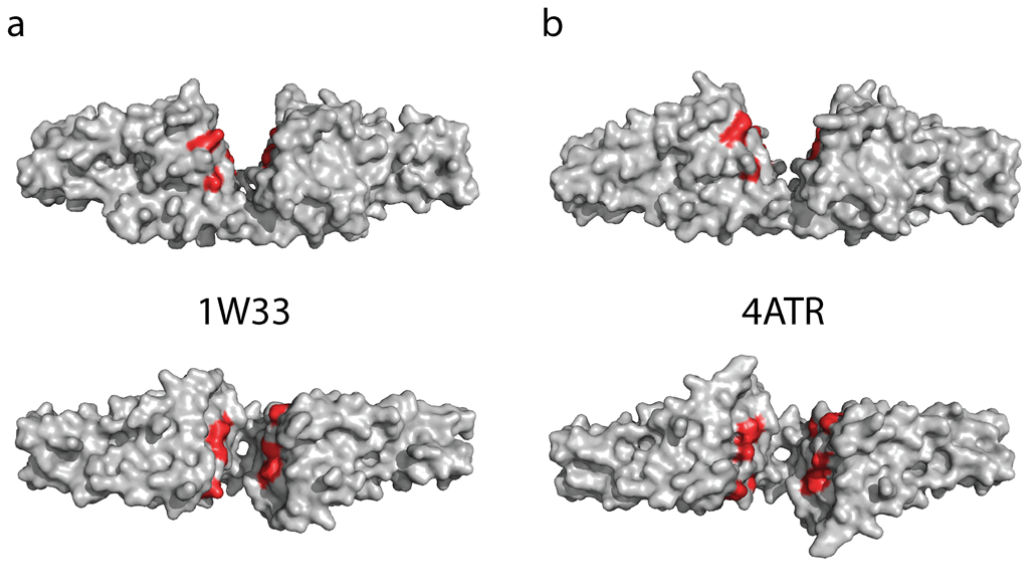


Figure 3.15 *Panel a*: Structure of the CspA dimer in PDB entry 1W33. Shown as a surface with the residues in the cleft implicated in factor H binding (Lys136, Lys141, Glu147 and Leu146) in red. *Panel b*: Identical representation of the CspA dimer in the new structure showing increased accessibility to the potential factor H binding site.

The structure of the ten C-terminal residues of one monomer subunit, bound in a cleft on the second subunit, is completely conserved. This supports reported evidence that these residues are required for assembly of a stable dimer and adds further weight to the hypothesis that the dimeric species is essential for function as the "lock" to the second monomer subunit is unaffected. Analysis of the dimerization interface in both structures shows an increased surface area of the dimerization interface in the "open" structure associated with a slightly more negative free energy of solvation (Table 3.6) which could suggest that the two conformers can interconvert with a possible propensity for the more open form. Flexibility allows alteration of the binding site confirmation which could enable binding to the factor H homologues in each of the different reservoir hosts, including hares, shrews, hedgehogs, mice, squirrels and a variety of small birds [Gern, Estrada-Pena *et al.* 1998], perhaps providing a key role of CspA in enabling the zoonotic bacteria *B. burgdorferi* to evade complement mediated killing across different hosts.

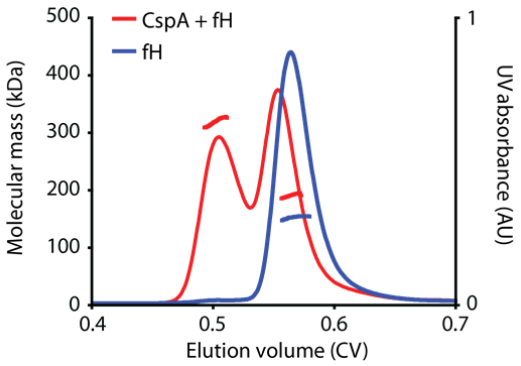


Figure 3.16 MALS analysis of the complex between full length fH and CspA. Factor H alone returned a mass of 155 kDa which corresponds exactly to the expected mass. Addition of CspA results in two peaks with masses of 330 kDa and 195 kDa. The 195 kDa peak corresponds to one fH (155 kDa) in complex with two CspA molecules (21.6 kDa each). The 330 kDa peak appears to contain two factor H molecules and one CspA. However, the error on this last measurement is high due to the peaks proximity to the void and the associated background scattering.

Unfortunately, MALS could not be used to confirm that CspA binds fH_{567} in a 2:1 ratio and show that the CspA dimer is both responsible, and required, for factor H binding. This was due to complications arising from interaction of fH_{567} with the size exclusion column matrix. However, preliminary MALS results from CspA with full length factor H suggest that the system could be even more complex (Figure 3.16). CspA is observed forming a complex with factor H with a mass consistent with a dimer binding the complement regulator. In addition to this species there is another that has a mass consistent with two factor H molecules and CspA, but the oligomeric state of the latter could not be determined. This finding has profound implications to the interaction interface as it may suggest two binding surfaces per dimer, or per monomer, challenging the putative binding site in the cleft. It may also arise from two separate binding sites; one in the cleft for domains five through seven and another elsewhere for other fH domains allowing one fH to bind in both sites, or to bind two separate molecules. A third possibility is that binding of CspA promotes dimerization of factor H. It has been suggested that fH can form dimers and higher oligomers [Perkins, Nealis *et al.* 1991]. Binding by CspA may promote or stabilise formation of a fH dimer, resulting in increased complement evasion due to higher surface concentrations of the regulator.

In light of these findings, the putative binding site within the cleft is called into question, and raises the possibility that presentation of factor H dimers on the surface of a pathogen may provide improved complement resistance. Further investigation is required in order to try to understand how factor H is bound by CspA and the exact mechanism of how complement resistance is conferred to *Borrelia burgdorferi* by this molecule.

3.7 Summary

The findings presented in this chapter show that CspA binds factor H with a high affinity (~2 nM) in the region of SCR domain five and confirms that CspA forms dimeric species in solution. MALS experiments also show that CspA forms a complex with fH in a 2:1 molar ratio confirming that the CspA dimer binds the complement regulator. A new atomic resolution structure of CspA has also been determined showing that there is a degree of flexibility between the monomer subunits which may have implications in recruitment of factor H from serum and could play a role in complement evasion across the different reservoir hosts of *Borrelia burgdorferi*.

3.7.1 Future directions

Further efforts to obtain the atomic resolution structure of factor H in complex with CspA should be performed to fully understand the nature and position of the interface. Other experiments could include heparin competition assays to ascertain whether CspA is mimicking host-cell surface GAGs in an analogous manner to fHbp. Work could also be undertaken to assay binding of factor H analogues from other reservoir hosts in order to assess whether CspA is specific for the human protein or if this mechanism of complement evasion is maintained in other animals.

Chapter Four: Binding of complement regulator factor H by ErpC of *Borrelia burgdorferi*

ErpC, or BbCRASP-4, is another member of the CRASP family from *Borrelia burgdorferi* that is expressed throughout colonisation of a mammalian host during the zoonotic cycle of the bacterium (Figure 4.1), underlining its importance in aiding complement evasion by binding and presenting factor [Kraiczy, Hellwage *et al.* 2004]. Scavenging of CFHR-1, CFHR-2 and CFHR-5, but not FHL-1, has also been observed, although the functional implications of this ability to bind multiple complement proteins has yet to be fully understood [Kraiczy, Skerka *et al.* 2001b; Haupt, Kraiczy *et al.* 2007; Hammerschmidt, Hallstrom *et al.* 2012]. Alignments and secondary structure predictions of ErpC suggest that there is no structural similarity to CspA, the former being predicted to be predominantly beta sheets whilst the latter is all alpha helices (Figure 4.2).

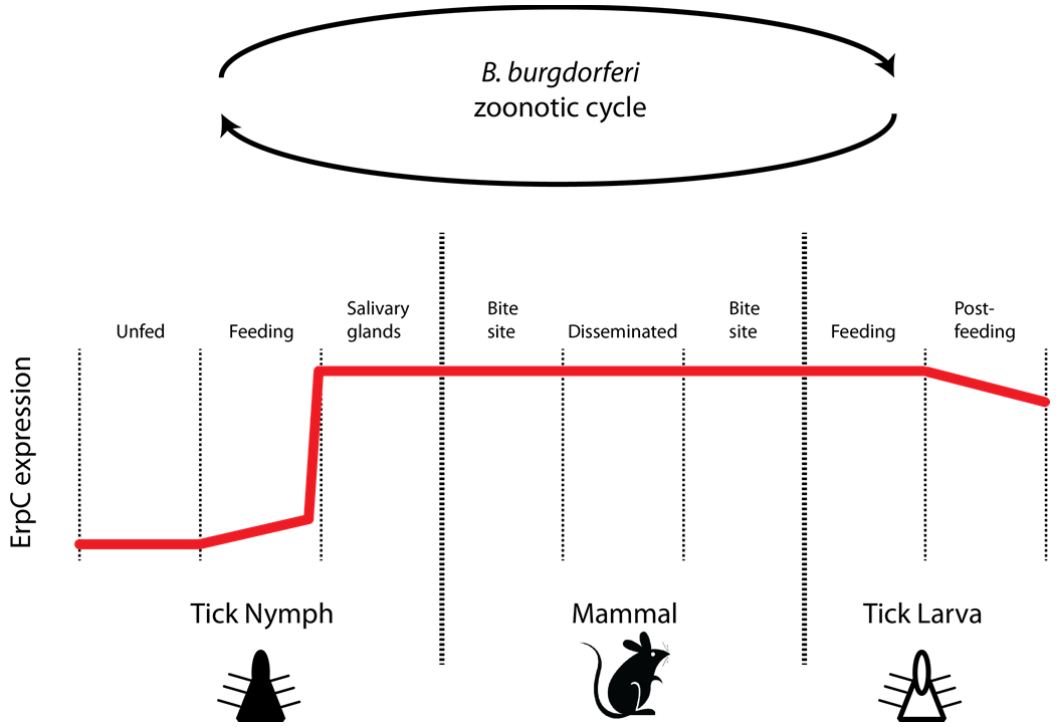


Figure 4.1 Expression profile of ErpC throughout the zoonotic cycle of *Borrelia burgdorferi.*

Presented here are binding data that show that ErpC binds factor H around domain five in a high affinity interaction that is similar in magnitude to fHbp and CspA. In contrast to ErpC, biophysical characterisation shows ErpC to be monomeric whilst the atomic resolution structure of ErpC has been solved, demonstrating that the architecture of ErpC is different to that of other fH binding proteins.

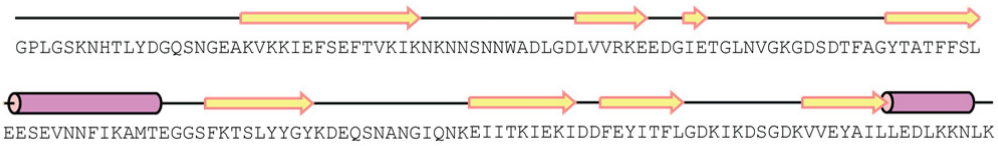


Figure 4.2 Secondary structure prediction of ErpC calculated using the PSIPRED server [Jones 1999; Buchan, Ward *et al.* 2010].

4.1 Protein expression and purification

ErpC was expressed in *Escherichia coli* as a fusion protein with glutathione-S-transferase (GST) as previously published [Haupt, Kraiczy *et al.* 2007]. The purification tag is then cleaved off by proteolytic cleavage to obtain ErpC alone. Expression of protein in this manner using a bacterial strain that is a methionine auxotroph also enables incorporation of selenomethionine which may be used for phasing the structure.

4.1.1 Details of plasmid pGEX-ErpC

Plasmid pGEX-ErpC was kindly provided by Professor Peter Zipfel (Friedrich Schiller University of Jena, Germany) and encodes residues 21 to 179 of ErpC isolated from *Borrelia burgdorferi* strain ZS7 expressed as a fusion protein with an N-terminal GST affinity tag under the control of an inducible P_{tac} promoter (Figure 4.3). Residues 1 to 20 were not included as they constitute the signal peptide and lipidation motif required for membrane adhesion, and are not involved in complement protein binding [Haupt, Kraiczy *et al.* 2007].

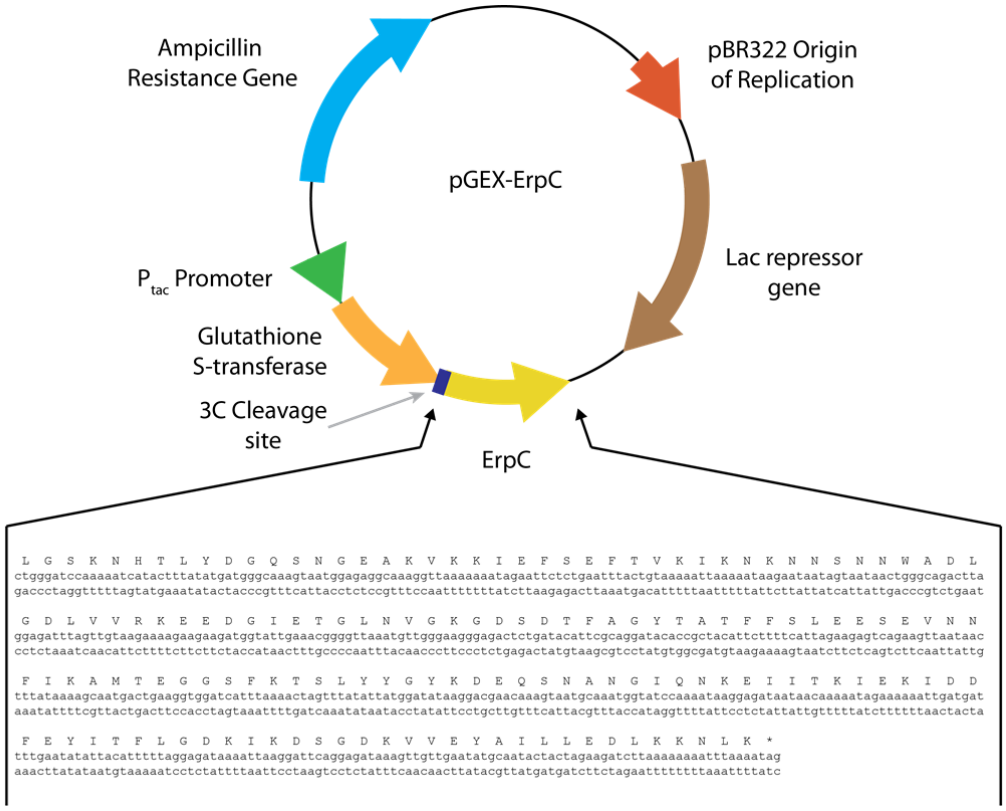


Figure 4.3 Schematic representation of pErpC. CspA (lacking residues 1-20) is expressed as an N-terminal GST fusion protein which can be removed by digestion with 3C protease. The vector was produced by cloning of the ErpC construct into pGEX-6P-1 (GE Healthcare, USA) between the BamHI and XhoI restriction sites.

4.1.2 ErpC expression and purification

Escherichia coli strain B834(DE3) transformed with the pGEX-ErpC plasmid was grown overnight at 37°C with 180 RPM shaking in 2 × 10ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 4 × 1L Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37°C with 180 RPM shaking. Cell proliferation was monitored using absorbance at 600 nm and expression of ErpC was induced during mid-log phase (OD_{600} = 0.6) by addition of 1 mM IPTG followed by continued incubation at 21°C with 180 RPM shaking overnight. Cells were harvested by centrifugation (4000 g, 20 minutes, 4°C) before the pellets were combined and stored at -20°C until required.

Defrosted cells were resuspended in 40 ml of 100 mM Tris, 300 mM NaCl, 10 mM DTT, 1mM EDTA, pH 8.0 and one EDTA-free protease inhibitor cocktail tablet (Roche, USA) was added. Cells were lysed by five passes through an Emulsiflex C5 homogeniser (Avestin, Canada) and the lysate cleared by centrifugation (48400 g, 20 minutes, 4^{o}C). Filtered supernatant was loaded on a 1 ml GSTrap FF column (GE Healthcare, USA) pre-equilibrated with 100 mM Tris, 300 mM NaCl, 10 mM DTT, 1 mM EDTA, pH 8.0 at 0.2 ml/min. The loaded GSTrap was washed with 100 ml of 100 mM Tris, 300 mM NaCl, 10 mM DTT, 1 mM EDTA, pH 8.0 prior to elution with 20 ml of 100 mM Tris, 300 mM NaCl, 10 mM DTT, 1 mM EDTA, 10 mM reduced glutathione, pH 8.0. The amount of ErpC in the elution was estimated assuming ε_{280} = 58790 M^{-1} cm^{-1} (calculated using ProtParam [Gasteiger E. 2005] with the sequence of the GST-ErpC fusion). GST tagged 3C protease (A kind gift from Miss. Rachel Everett) was added in a 1:20 ratio by mass and the sample dialysed against 50 mM Tris, 150 mM NaCl, 10 mM BME, 1 mM EDTA, pH 7.5 in 3000 MWCO Snakeskin (Pierce, USA) at 21^{o}C overnight.

The digested sample was re-loaded on a 1 ml GSTrap pre-equilibrated in 50 mM Tris, 150 mM NaCl, 10 mM BME, 1 mM EDTA, pH 7.5 at 0.2 ml/min and the flowthrough collected. Any ErpC non-specifically bound to the resin was removed with a 20 ml wash with 50 mM Tris, 1 M NaCl, 10 mM BME, 1 mM EDTA, pH 7.5 and the pooled flowthroughs were concentrated to 5 ml using a 3000 MWCO PES spin concentrator (Millipore, UK) prior to a final purification step on an S75 16/60 size exclusion chromatography column (GE Healthcare, USA) pre-equilibrated in 50 mM Tris, 150 mM NaCl, 10 mM BME, pH 7.5 connected to an AKTA purifier FPLC system (GE Healthcare, USA). An example UV_{280nm} trace is shown in Figure 4.4 alongside an SDS-PAGE gel charting the course of the purification. Typical yields using this purification

regime are 1 mg ErpC per litre of culture (assuming ε_{280} = 15930 M^{-1} cm^{-1}) and the purified protein was checked for identity and purity using ES-TOF mass spectrometry returning a mass of 18317.3 ± 2 Da which corresponds well to the expected mass of 18316 Da.

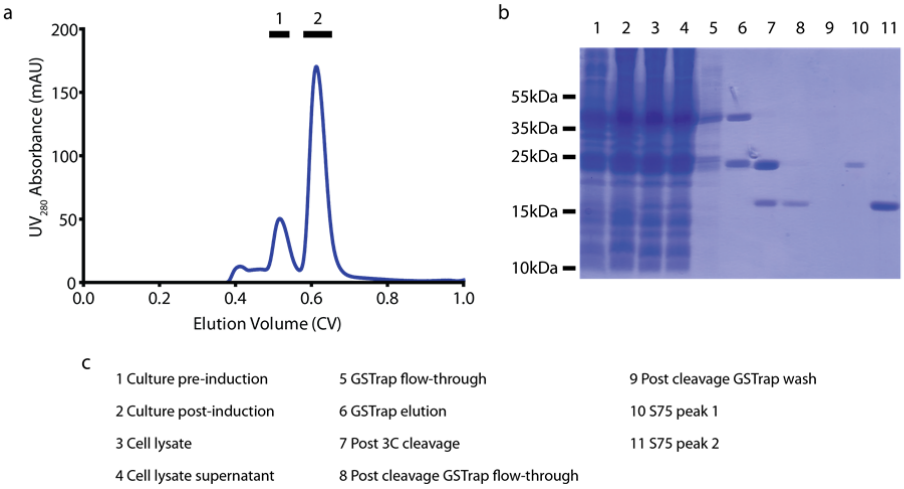


Figure 4.4 *Panel a*: Elution profile of ErpC from an S75 16/60 size exclusion column. Fractions collected from two peaks are marked with black bars. *Panel b*: 15% (w/v) reducing SDS-PAGE gel charting the course of the ErpC purification. *Panel c*: Legend describing the samples on each lane of the gel in panel b.

4.1.3 Expression and purification of selenomethionine derivitised ErpC

Methionine auxotrophic *Escherichia coli* strain B834(DE3) transformed with the pGEX-ErpC plasmid was grown overnight at 37°C with 180 RPM shaking in 10 ml Luria-Bertani broth containing 100 µg/ml ampicillin. 1 ml of the saturated culture was used to inoculate 100 ml Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37°C with 180 RPM shaking. Cell proliferation was monitored until OD_{600} = 0.6. Cells were harvested by centrifugation (4000 g, 10 minutes, 4°C) and washed twice by resuspension in 60 ml of PBS followed by centrifugation (4000 g, 10 minutes, 4°C). The final cell pellet was resuspended in 4 ml of PBS and then used to inoculate 4 × 1L SelenoMet medium (Molecular Dimensions, UK) supplemented with 40 mg/l L-selenomethionine (Acros Organics, Belgium). Cell proliferation was monitored until

OD_{600} = 0.6 when ErpC expression was induced by addition of 1 mM IPTG followed by continued incubation at 21°C with 180 RPM shaking overnight. Cells were harvested by centrifugation (4000 g, 20 minutes, 4°C) before the pellets were combined and stored at -20°C until required. Further purification of selenomethionine derivatised ErpC proceeded via the same protocol described in section 4.1.2 with a yield of approximately 1 mg per litre of culture (assuming ε_{280} = 15930 M^{-1} cm^{-1}). Incorporation of selenomethionine was assessed using ES-TOF mass spectrometry returning a mass of 18365 ± 0.5 Da (expected mass 18363 Da) indicating 100% integration.

4.2 Crystallisation, data collection and processing

4.2.1 Crystallisation of selenomethionine derivatized ErpC

Selenomethionine derivatized ErpC was concentrated to A_{280} = 10.5 and crystallisation trials were performed at 294K using vapour diffusion in 200 nL sitting drops produced by an Oryx Nano crystallisation robot (Douglas Instruments, UK) within MRC crystallisation plates (Molecular Dimensions, UK). Stura, Macrosol, Pact Premier, JCSG, Structure 1, Structure 2, Proplex and Morpheus crystallisation screens were used (Molecular Dimensions, UK), with two drops set up for each condition containing protein to precipitant ratios of 1:1 and 1:3. Small thin, multiple crystals nucleated after approximately three hours in 36% (w/v) PEG 2000 MME, 0.1 M sodium cacodylate pH 6.5 and after 24 hours in 27% (w/v) PEG 2000 MME, 0.1 M sodium cacodylate pH 6.5. Optimisation of crystals grown in the latter condition was achieved by slowing crystal growth using a larger drop size (800 nL) which slows the rate of equilibration and thus the crystal nucleation and growth speed, resulting in larger, more single crystals (Figure 4.5, panel b). The cacodylate anion contains arsenic which has a large X-ray absorption cross-section which can lead to amplified radiation damage due to the increased energy absorption. In order to minimise radiation damage, crystals were also grown in 27%

(w/v) PEG 2000 MME, 0.1 M MES pH 6.5 with substitution of the sodium cacodylate buffer for 2-(*N*-morpholino)ethanesulfonic acid (MES).

4.2.2 Data Collection

Crystals were mounted in Cryoloops (Hampton Research) prior to cryoprotection in 20% ethylene glycol, 27% (w/v) PEG 2000 MME, 0.1 M MES pH 6.5 and flash cooling in liquid nitrogen. The MES containing condition was used in order to backsoak the cacodylate containing crystals and remove as much of the arsenic containing compound as possible to increase the maximum X-ray exposure by minimising radiation damage. Prior to data collection an X-ray stimulated fluorescence scan was performed over the L3 absorption edge of elemental selenium and the peak and inflection energies, with their coherent and anomalous scattering factors, reported using CHOOCH [Evans and Pettifer 2001] (Figure 4.5, panel a).

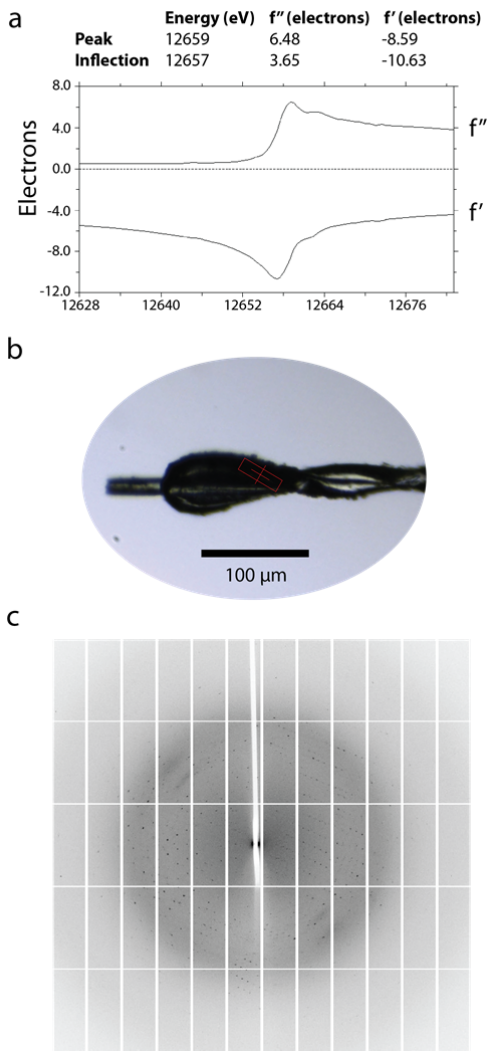


Figure 4.5 *Panel a*: Plot of the variation of *f'* and *f''* over the L3 absorption edge of selenium. Peak and inflection energies, with corresponding *f'* and *f''* values, have been calculated. *Panel b*: ErpC crystal mounted in a Cryoloop. *Panel c*: Example diffraction image from a cacodylate buffered crystal.

A SAD and two MAD datasets were collected for both the cacodylate and MES buffered crystals using the rotation method at the wavelengths derived from the fluorescence scan and at a high energy remote wavelength (data collection parameters

for the cacodylate buffered crystals are detailed in Table 4.1). For each dataset collection the optimal data collection strategy for anomalous data was calculated using EDNA [Incardona, Bourenkov *et al.* 2009].

	Dataset 1 Peak	Dataset 2 Inflection	Dataset 3 Remote	Dataset 4 Inflection	Dataset 5 Remote	Dataset 6 Peak
Diffraction source	Diamond I03	Diamond I03	Diamond I03	Diamond I03	Diamond I03	Diamond I03
Detector	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F
Energy (eV)	12659	12657	12664	12657	12664	12659
Wavelength (Å)	0.9794	0.9796	0.979	0.9796	0.979	0.9794
Temperature (K)	120	120	120	120	120	120
Detector resolution (Å)	2.2	2.2	2.2	2.2	2.2	1.8
Exposure per image (s)	0.2	0.2	0.2	0.15	0.15	0.2
Φ_{start}	60	60	60	126	126	21
Φ_{end}	240	240	240	276	276	201
$\Delta\Phi$	0.15	0.15	0.15	0.15	0.15	0.15
Number Images	1200	1200	1200	1000	1000	1200

Table 4.1 Data collection parameters for collection of two MAD datasets and one SAD dataset from three separate crystals grown in the cacodylate buffered condition.

4.2.3 Data Processing

All data were processed using XIA2 invoking the 3da flag which enforces usage of XDS [Kabsch 2010] for autoindexing and integration followed by scaling with AIMLESS [Evans 2006]. Cacodylate buffered crystals were indexed in a primitive orthorhombic cell with approximate dimensions of a = 62 Å, b = 68 Å, c = 75 Å, $\alpha = \beta = \gamma = 90^{o}$. MES buffered crystals were processed in primitive tetragonal cell with approximate dimensions of a = 65 Å, b = 65 Å, c = 78 Å, $\alpha = \beta = \gamma = 90^{o}$. However, the latter crystals were found to be twinned (twin fraction = 0.1. Figure 4.6, panel d) and were disregarded in the presence of non-twinned crystals that diffracted to a similar resolution. Analysis of the systematic absences in the orthorhombic data confirm screw axes along all three two-fold symmetry axes (Figure 4.6 panels a, b and c) making the

likely spacegroup $P2_12_12_1$. There are also anomalous correlations between Bijvoet pairs confirming the presence of anomalous signal. Data processing statistics are presented in Table 4.2.

	Dataset 1 Peak	Dataset 2 Inflection	Dataset 3 Remote
Spacegroup	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell dimensions (Å)	a=62.4, b=68.0, c=75.4	a=62.4, b=68.0, c=75.4	a=62.4, b=68.0, c=75.4
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	48.6	70.5	72.8
Resolution (Å)	2.32-50.51 (2.32-2.38)	2.68-48.09 (2.68-2.75)	3.13-48.07 (3.13-3.21)
R_{merge}	0.157 (1.745)	0.223 (2.196)	0.265 (1.485)
R_{pim} (I+/I-)	0.098 (1.076)	0.132 (1.331)	0.167 (0.959)
<I/σ(I)>	13.4 (3.5)	9.9 (2.5)	7.2 (3.1)
Anomalous completeness (%)	98.9 (99.8)	98.7 (99.3)	98.3 (99.0)
Anomalous multiplicity	3.4 (3.5)	3.4 (3.5)	3.4 (3.3)
Anomalous correlations	0.25 (-0.01)	0.04 (-0.04)	0.08 (0.06)

	Dataset 4 Inflection	Dataset 5 Remote	Dataset 6 Peak
Spacegroup	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell dimensions (Å)	a=62.7, b=68.2, c=76.0	a=62.8, b=68.2, c=75.8	a=62.6, b=68.3, c=75.1
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	68.5	78.5	61.2
Resolution (Å)	2.56-25.37 (2.56-2.67)	2.74-26.71 (2.74-2.87)	2.64-48.08 (2.64-2.71)
R_{merge}	0.067 (0.668)	0.076 (0.753)	0.205 (1.809)
R_{pim} (I+/I-)	0.047 (0.459)	0.053 (0.517)	0.13 (1.108)
<I/σ(I)>	15.5 (2.8)	14.9 (2.7)	12.3 (3.8)
Anomalous completeness (%)	97.5 (98.7)	97.3 (98.5)	98.9 (98.4)
Anomalous multiplicity	2.8 (2.8)	2.8 (2.9)	3.4 (3.6)
Anomalous correlations	0.13 (0.00)	0.18 (0.00)	0.23 (-0.04)

Table 4.2 Data processing statistics for each dataset.

4.3 Structure determination and refinement

Molecular replacement is a powerful technique for solving X-ray structures that works well with a good model for the target protein. However, when a model is not available, perhaps due to the lack of structural homologues or failure to predict a structural similarity based upon sequence, the protein must be experimentally phased.

Incorporation of selenomethionine into ErpC allows structure solution using single and multiple wavelength anomalous diffraction methods (SAD/MAD) to obtain phases.

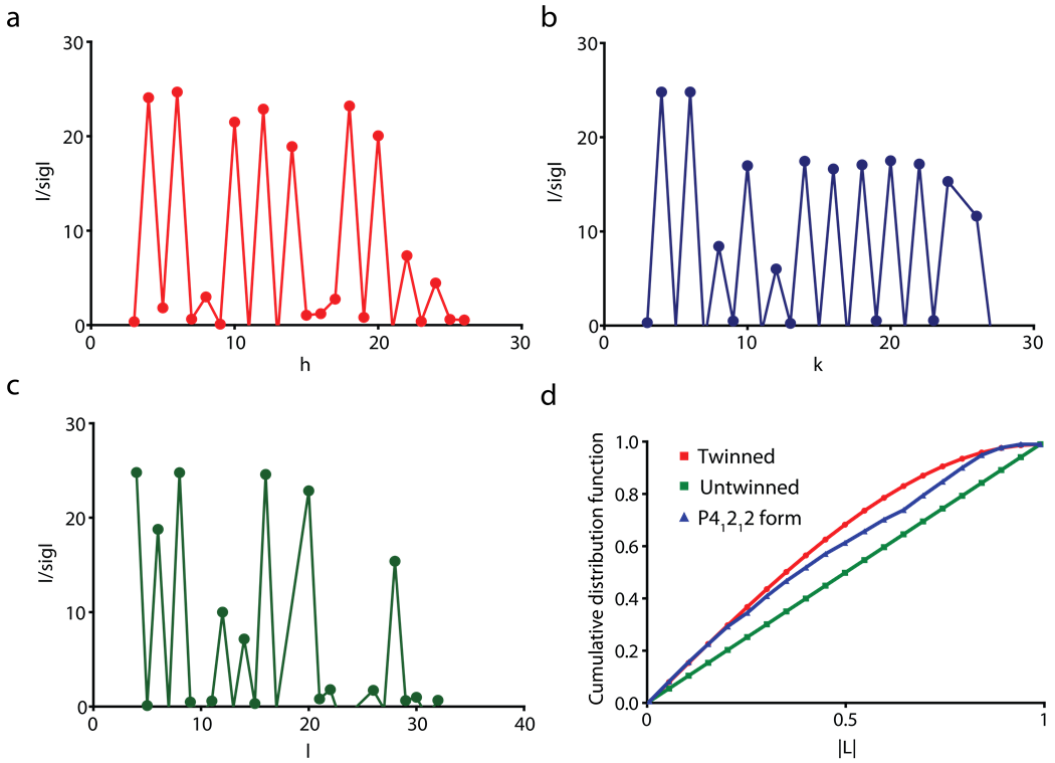


Figure 4.6 *Panel a*: Systematic absences amongst reflections with a given value of h. The lack of odd values of h suggest a screw axis parallel to a. *Panel b*: Screw axis parallel to b. *Panel c*: Screw axis parallel to c. *Panel d*: L-test for twinning for the $P4_12_12$ crystals.

4.3.1 Se-SAD phasing of ErpC

Analysis of the content of the $P2_12_12_1$ asymmetric unit and the self rotation function, using MATTHEWS [Matthews 1968] and MOLREP [Vagin and Teplyakov 1997] respectively, suggests the presence of two copies of ErpC that are related by a two-fold rotational NCS, the axis of which is parallel to the crystallographic two-fold rotation along c creating a four-fold rotation resulting in the peak observed in the chi = 90° section (Figure 4.7, panel a). The corresponding two fold peaks in the chi = 180° function make the crystals appear to be pseudo-tetragonal. All six datasets were fed into a single AUTOSHARP [Vonrhein, Blanc *et al.* 2007] job which was tasked with finding

two selenium sites using the SHELX [Sheldrick 2008] suite of programs. After scaling all six datasets together and outlier rejection using SCALEIT [Howell and Smith 1992], AUTOSHARP cycles through all dataset combinations, both MAD and SAD, using SHELXC [Sheldrick 2010] to determine the resolution cut off for anomalous signal prior to site finding with SHELXD [Sheldrick 2010]. Sites found for each combination are scored by the correlation between the observed structure factors and those calculated from the heavy atom locations. The highest correlations (CC = 0.35) were calculated from a SAD phasing job using dataset 1, which had an anomalous signal cutoff at 3.7 Å (determined by SHELXC when d''/sigd'' = 1.3 (Figure 4.7, panel b)) and successfully found two selenium sites (Table 4.4) after 180 trials. Peaks that correspond to these sites may be observed in the Z = 1/2 Harker section of the anomalous difference Patterson map shown in Figure 4.8.

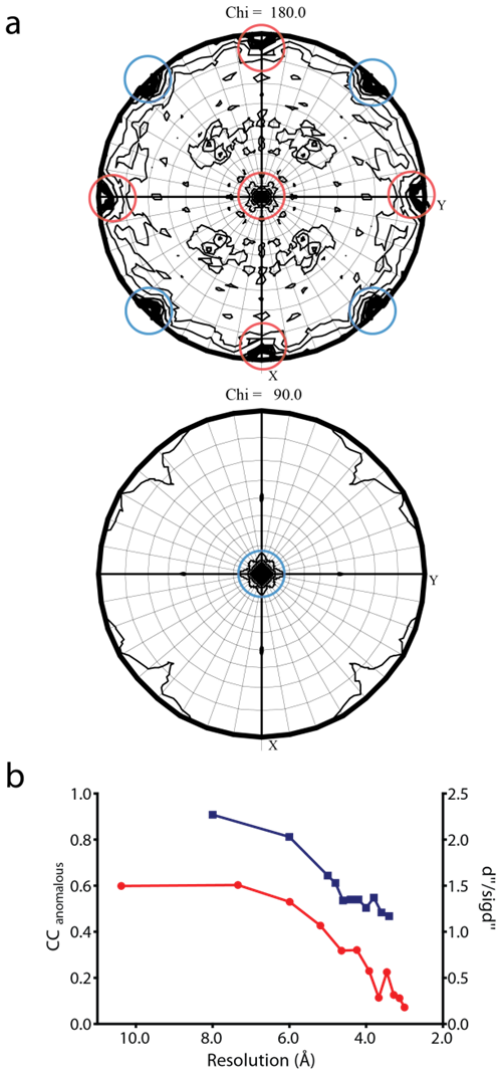


Figure 4.7 *Panel a*: Self-rotation function of $P2_12_12_1$ crystal form of ErpC clsculated using MOLREP. Crystallographic symmetry axes are highlighted in red. Peaks resulting from rotational NCS are marked in blue. *Panel b*: Anomalous correlations (red) and d/sigd (blue) within peak wavelength dataset 1.

The selenium site positions and occupancies were refined against all six datasets using SHARP [Bricogne, Vonrhein *et al.* 2003] given the experimentally derived f' and f'', or those calculated for the high energy remote wavelength using CROSSEC [Cromer

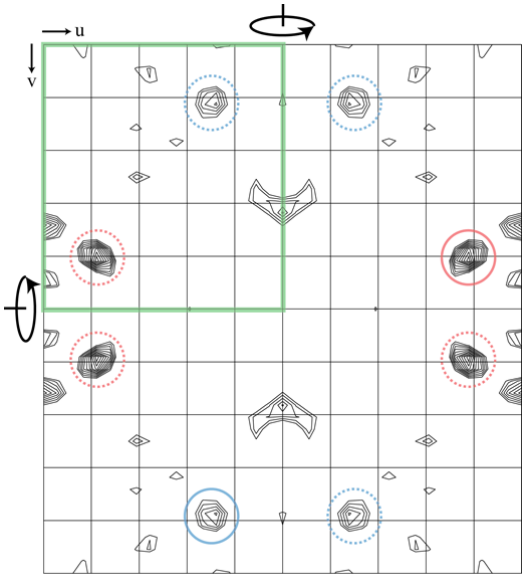


Figure 4.8 Z = 0.5 Harker section of the anomalous Patterson map. Se site 1and its symmetry mates are shown in red, Se site 2 in blue. The asymmetric unit containing two Se sites is highlighted in green.

1983] (f' = -9.8034 electrons, f' '= 3.8438 electrons). Convergence was reached after 17 cycles of refinement and the resulting residual maps analysed for outliers, of which none were significant, unsurprisingly given that there are only two selenium atoms in the asymmetric unit from the single methionine residue in each of the two copies of ErpC (there is no N-formyl methionine as this was cleaved off during removal of the GST affinity tag). The occupancies were inflated from the expected value of 1 to 1.3 and 1.8 which is due to the their being dependent on the overall scale factor, which is inaccurate. Final overall figure-of-merit was 0.48 for acentric and 0.13 for centric reflections with the phasing power, and corresponding resolution cutoff, shown in Table 4.5. Phases were calculated from the refined sites in both the original and inverted hands, the maps from which were then subjected to solvent flattening in SOLOMON [Abrahams and Leslie 1996] with an assumed solvent content of 43.5% resulting from two copies of ErpC in the asymmetric unit.

	x	y	z	u	v	w
Se site 1	0.3071	0.7012	0.0642	-0.1142	1.4024	0.5
Se site 2	0.0799	0.4431	0.1844	0.3402	0.8862	0.5

Table 4.3 Locations of the Se atoms in the unit cell and their locations is the Z = 0.5 Harker section of the anomalous difference Patterson.

Symmetry operations	Harker vector
x, y, z	0, 0, 0
½-x, -y, ½+z	½-2x, -2y, ½
-x, ½+y, ½-z	-2x, ½, ½-2z
½+x, ½-y, -z	½, ½-2y, -2z

Table 4.4 Symmetry operations and their corresponding Harker vectors for spacegroup $P2_12_12_1$. The vector that gave Z=0.5 was used to claclulate the heavy atom locations in the Z = 0.5 Harker section (Table 4.3).

Comparing the correlation coefficients between the structure factor amplitudes from the observed data and the solvent flattened map showed that the inverted hand has the best correlation coefficient of 0.28, versus 0.16 for the original hand, which was indicative of a correct solution. Optimisation of the solvent content in the inverted hand increased the correlation coefficient to 0.7962 at 49.4%, with secondary structure elements clearly visible by inspection of the resulting solvent flattened electron density map (Figure 4.9, panel b).

	Dataset 1 Peak	Dataset 1 Inflection	Dataset 1 Remote	Dataset 2 Inflection	Dataset 2 Remote	Dataset 3 Peak
Phasing power						
Overall	1.17	0.343	0.211	0.581	0.647	0.727
Drops below 1.0 (Å)	3.92	6.75	50.51	6.05	5.13	5.13

Table 4.5 Overall phasing power of anomalous differences for each wavelength in each ErpC dataset in $P2_12_12_1$. Also shown is the corresponding resolution at which the phasing power drops below 1.0.

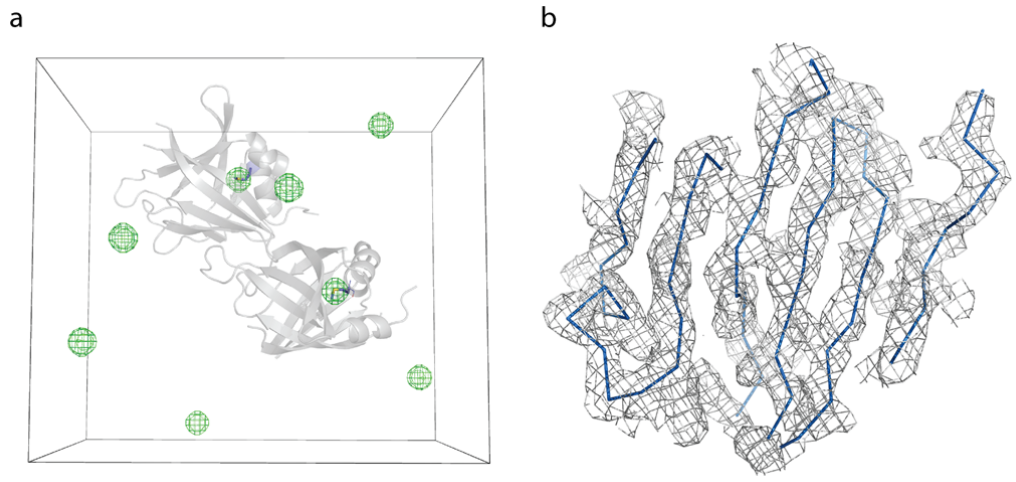


Figure 4.9 *Panel a*: Location of the Se sites in the unit cell. Both copies of ErpC in one of the assymetric units are shown with the selenomethionine residues in stick form. *Panel b*: Example of the initial solvent flattened electron density map contoured at 1.5σ. The main chain of the final structure is shown as a ribbon for clarity.

4.3.2 ErpC model building

BUCCANEER [Cowtan 2006] was used to build an initial model in the solvent flattened electron density map from AUTOSHARP described by structure factor

amplitudes FBshasol, and their corresponding SIGFBshasol, and phases given as Hendrickson-Lattman coefficients, HLAshasol, HLBshasol, HLCshasol and HLDshasol. Three cycles of chain extension, growing and sequencing were performed between refinement using REFMAC5 [Murshudov, Skubak *et al.* 2011] resulting in a model with 260 residues, of which 203 were sequenced, built in 9 chains with R/R_{free} of 0.3925/0.4972. COOT [Emsley, Lohkamp *et al.* 2010] was used to combine the chains, and symmetry copies where necessary, into the two separate copies of ErpC before unknown and missing residues were built in by hand.

4.3.3 ErpC Refinement and validation

	ErpC
Resolution (A)	2.50-24.15 (2.50-2.74)
Number of reflections	11676 (2755)
Number of reflections for R_{free}	561 (148)
R_{work}/R_{free}	0.1928/0.2360 (0.2170/0.2553)
Number of atoms	
Protein	2095
Ligand/Ion	58 Ethylene glycol
Water	18
R.m.s deviations	
Bond lengths (A)	0.01
Bond angles (°)	1.24
Ramachandran plot analysis	
Residues in favoured regions (%)	96.05
Residues in disallowed regions (%)	0.4
Molprobity analysis	
Molprobity score	4.53
Molprobity score percentile	99th
Resolution range (A)	2.25 - 2.75

Table 4.6 Refinement and validation statistics of the final ErpC model.

ErpC was refined iteratively between rounds of refinement in AUTOBUSTER [Blanc, Roversi *et al.* 2004; Bricogne, Blanc *et al.* 2011] and rebuilding in COOT. Restraints from the non-crystallographic symmetry between the two ErpC molecules in the asymmetric unit were invoked during AUTOBUSTER refinement using the autoncs flag [Smart, Brandl *et al.* 2008]. Throughout model building, MOLPROBITY [Chen, Arendall *et al.* 2010] was used to analyse the structure and validate the model. Final refinement and validation statistics are presented in Table 4.6. Residues 66 to 73 could not be built in either copy of the molecule in the asymmetric unit due to the lack of electron density in this loop region.

This is most likely due to disorder of this region in the crystal suggesting that the loop is conformationally labile. The refined structure is shown in Figure 4.10, panels a and b.

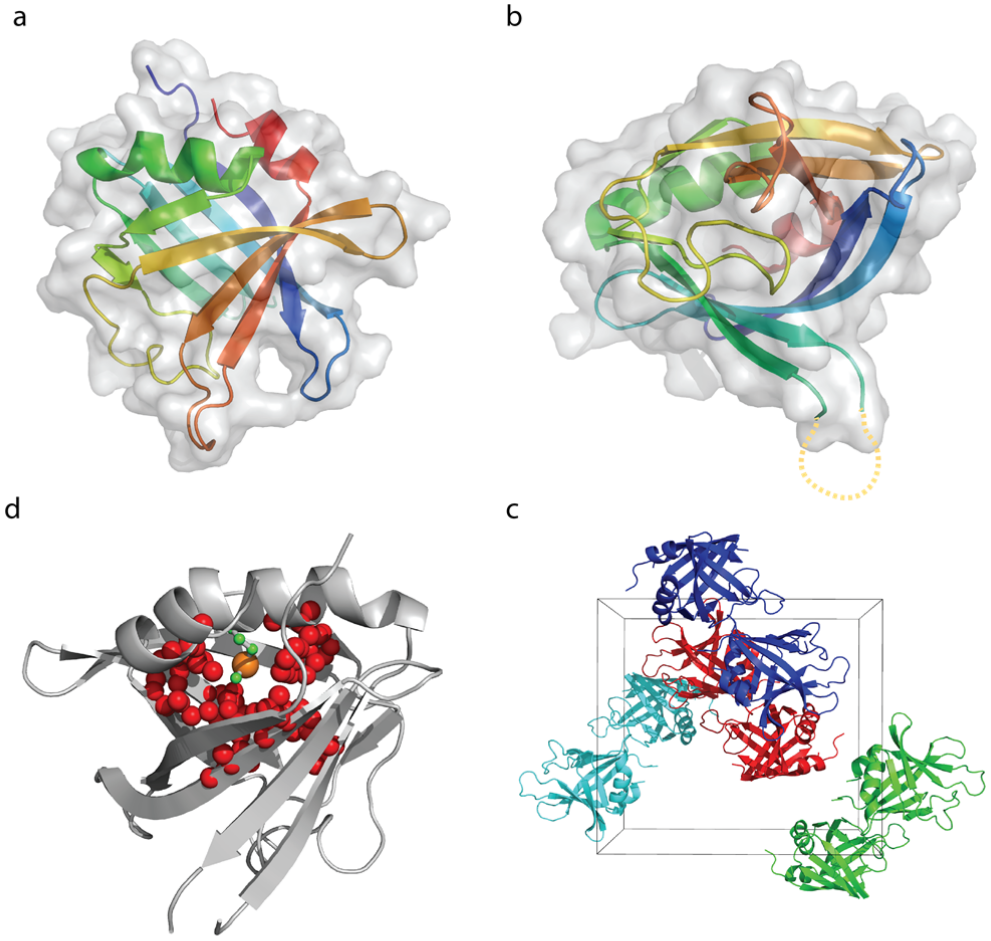


Figure 4.10 *Panel a*: Cartoon representation of the structure of ErpC within a transparent surface rendering. *Panel b*: A different view of the structure of ErpC. The loop that could not be built due to disorder can be seen to the lower right. *Panel c*: Cartoon representation of ErpC with the selenomethionine sidechain displayed in green and the selenium atom in orange. The residues forming the hydrophobic pocket around the selenmethionine are shown as red spheres. *Panel d*: Crystal packing in the unit cell coloured by the contents of each asymmetric unit.

4.4 Structural analyses of ErpC

The structure of ErpC consists of ten beta strands forming a barrel capped by an alpha helix (Figure 4.10). Searching for structurally similar proteins using the DALI webserver [Holm and Rosenstrom 2010] shows a similarity to a sporulation-specific cell

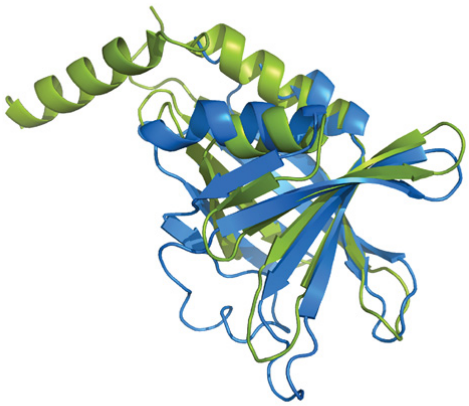


Figure 4.11 Superposition of the structure of ErpC (blue) with PDB entry 3CM1 (green) showing a common architecture.

division protein from *Thermobifida fusca* (PDB entry 3CM1 [Xu, Traag *et al.* 2009]), with a 91% match of secondary structure elements and an RMSD of 2.34 Å over 137 residues (Figure 4.11). However, despite the similarity, the sequence identity between ErpC and PDB entry 3CM1 is only 7%. Whilst the structural similarity is striking and it cannot be ruled out that these proteins share a common ancestry, the low sequence conservation suggests that they could have evolved separately from each other and are not functionally related.

The angular relationship between the two copies of ErpC in the asymmetric unit is not quite 90° (Table 4.7), explaining the peak in the chi = 90° self-rotation function and removing the possibility of a four-fold symmetry axis along c. The result of this is that the cacodylate buffered crystals have a lower symmetry (orthorhombic) than those buffered with MES (tetragonal) which has a beneficial effect as the orthorhombic crystals were not twinned.

	ω	φ	χ
Chain A - Chain B	1.0	-4.6	90.9

Table 4.7 Rotational transformations, in spherical polar coordinates, between both copies of ErpC in the asymmetric unit.

The selenomethionine residue is very highly ordered which contributed to the strength of the anomalous signal and the quality of the calculated phases. Located at the C-terminal end of the capping helix, the side chain points into the hydrophobic core of the protein where the selenium atom resides between hydrophilic side chains (Figure 4.10, panel d). Proteins on average contain approximately 2.2% methionine residues [Doolittle 1989] where as ErpC contains only one methionine (0.6%). The natural

abundance of methionine means that most X-ray structures phased from selenomethionine derivatization contain significantly more selenium atoms than ErpC and no other proteins could be found that were phased using Se-SAD/Se-MAD and contain less than 1% selenomethionine. Structure solution of ErpC in this manner is significant, demonstrating that the phasing ability of a single selenium atom is stronger than previously observed and selenomethionine derivitization should not be discounted in cases of low methionine abundance.

The two ErpC molecules in the asymmetric unit do not appear to be contacting via any physiologically relevant interaction surfaces (Table 4.8) suggesting that ErpC is monomeric on the Borrelial cell surface. Another interesting feature is that the N and C termini meet in the same region of space, suggesting that it is this side of the molecule that packs against the cell surface, perhaps mediated by the two helices, α_1 and α_2. This makes it highly likely that the fH binding site is formed by the flexible loops on the opposite side of the molecule in a similar manner to fHbp (Figure 4.12).

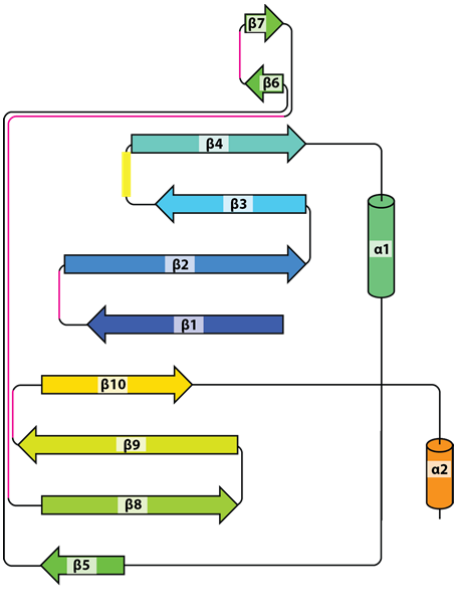


Figure 4.12 Topology diagram of the secondary structure of ErpC. The missing loop is marked in red, whilst the loops presumed to be involved in binding complement proteins are in yellow.

	Interface A	Interface B
CSS	0.0	0.0
Surface area (Å²)	460.3	167.5
ΔⁱG (kcal/mol)	4.0	-0.2
ΔⁱG P-value	0.789	0.426

Table 4.8 PISA analysis of the interfaces between ErpC molecules in the crystal. The only interfaces are non-physiological crystal contacts.

4.5 Investigation of the oligomeric state of ErpC

Analysis of the crystal structure suggests that ErpC is monomeric, highlighting another difference between ErpC and its CRASP family relative CspA. MALS was employed to confirm this finding. As expected, the mass derived from the peak definitively shows that ErpC doesn't form oligomers in solution under these conditions (Figure 4.13).

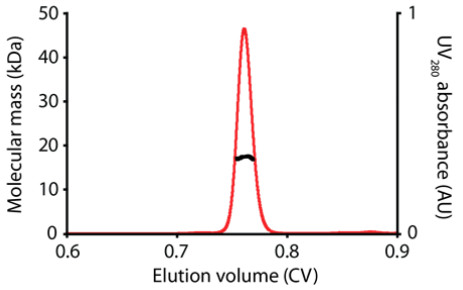


Figure 4.13 MALS analysis of ErpC. The determined mass of 17 kDa corresponds exactly to the mass from mass spectrometry showing that it is purely monomeric in solution.

4.6 Determination of the factor H domains involved in ErpC binding

Surface plasmon resonance was employed in order to assess the binding of fH_{567} and fH_{678} (Chapter 3.4) to ErpC. A Biacore 3000 (GE Healthcare, USA) instrument was used to covalently immobilise ErpC to a CM5 sensor chip (GE Healthcare, USA) via primary amine coupling. The CM5 chip was activated with 50 µl of EDC:NHS at 10 µl/min prior to immobilisation of each form of ErpC, diluted into 10 mM sodium acetate pH 4.0 to a final concentration of 50 µg/ml, on separate flow channels. The surface was deactivated by injection of 80 µl ethanolamine at 10 µl/min. A reference channel was mock activated and deactivated in an identical manner. Coupling resulted in immobilisation of approximately 1200 RU of ErpC.

Two separate dilution series of fH_{567} ranging from 0.2 nM to 100 nM were flowed over the chip using the KINJECT command with a flow rate of 40 µl/min, dissociation time of 600 seconds and temperature of 25^{o}C. In between each analyte injection the surface

was regenerated using a 5µl injection of 10 mM glycine pH 3.0. A further two dilution series of fH_{678} ranging from 0.02 µM to 10 µM were flowed over the chip in an identical manner. All the sensorgrams were blank subtracted and fits performed to the equilibrium responses using a steady state model. Example curves with their corresponding fits are shown in Figure 4.14.

The higher affinity of ErpC for fH_{567} over fH_{678} suggests that factor H domain five plays an important role in the interaction, but binding of fH_{678} suggests that the interface is not limited to domain five and is spread onto at least domain six.

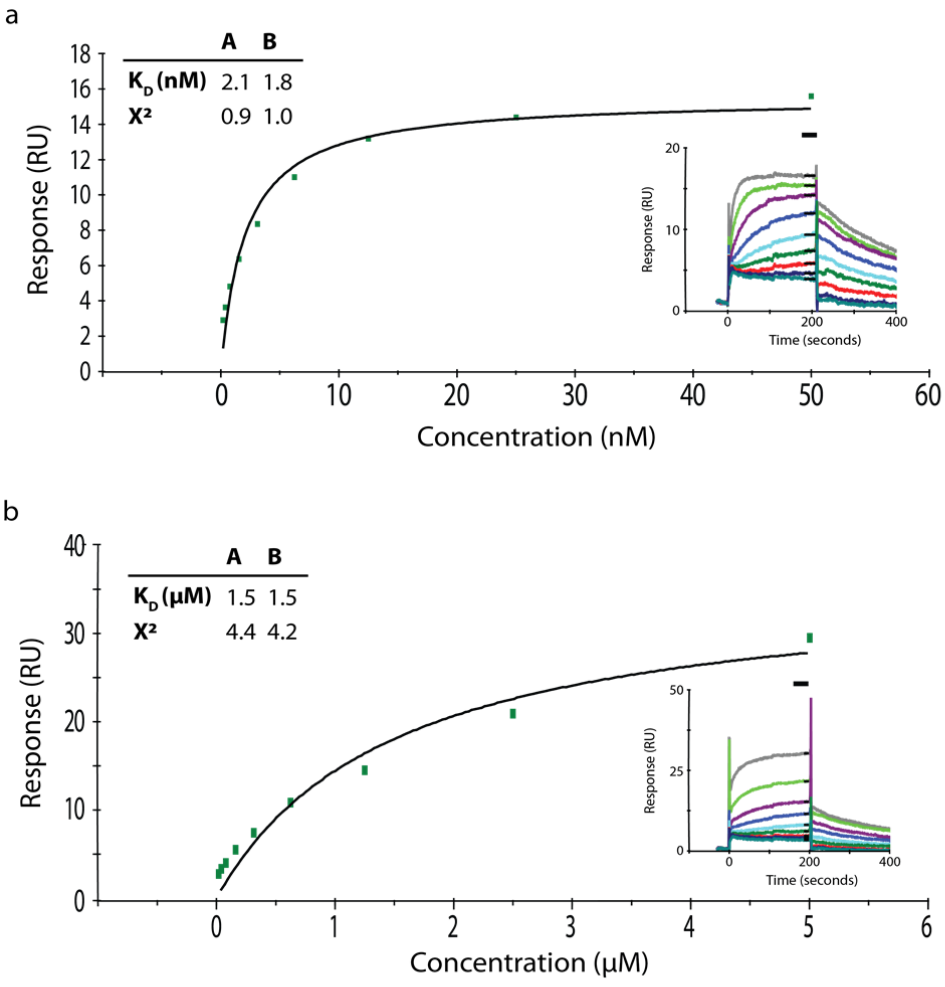


Figure 4.14 *Panel a*: Measurement of the affinity of ErpC for fH_{567} using a steady state fit to equilibrium data from a dilution series (inset). Values reported are from two separate experiments labelled a and b. *Panel b*: Measurement of the affinity of ErpC for fH_{678}. Values quoted are from two separate experiments.

4.7 Interactions of ErpC with the CFHRs

In a preliminary screen, ErpC was immobilised on a Biacore chip in an identical manner to that described in Chapter 4.6. 10 µl of 10 µM CFHR-1_{12}, CFHR-1_{345} and CFHR-2_{34} (Chapter 7.1) was flowed over the chip at 10 µl/min. Initial findings suggest that only CFHR-2_{34} interacts with ErpC (data not shown). However, further experiments are required to confirm this observation, and measure the affinity in order to suggest whether it is physiologically relevant.

4.8 Discussion

ErpC is another member of the CRASP family of complement regulator binding proteins which aid complement evasion by *Borrelia burgdorferi*. Both ErpC and CspA bind factor H within domains five to seven, yet they possess completely different architectures. As discussed in Chapter two, CspA is all alpha helical and forms a dimer which is required for fH binding. However, ErpC possesses a novel structure that forms a beta barrel capped with a single alpha helix and does not form a dimer or higher oligomer. Differences also exist between the putative fH binding sites. The interaction interface in CspA has been mapped to the cleft between the two monomer subunits (Chapter 3) in a structurally well ordered region, suggesting a well defined binding site. However, it is likely that the fH interface on ErpC includes the flexible loops between each beta strand in a similar manner to fHbp (Chapter 2), especially those loops on the side of the molecule away from both termini, as this face will be closer to the cell surface due to the proximity of the lipid anchor. Some of these loops were not observed in the crystal structure suggesting that the binding site could lack order. This added flexibility could also explain the reports that ErpC is also able to bind CFHR-1, CFHR-2 and CFHR-5 [Hammerschmidt, Hallstrom *et al.* 2012] as well as factor H [Kraiczy, Skerka *et al.* 2001a].

Although the mechanism by which fH is bound differs between ErpC and CspA, the affinity for fH_{567} is unaltered at approximately 2 nM. This is a fifth example of a bacterial factor H binding protein that binds this region of fH with a very similar affinity to CspA and fHbp variants one, two and three. As ErpC, CspA and fHbp have all evolved separately to have this affinity, these findings further support the hypothesis presented in Chapter 3.10 that the binding of fH cannot be too tight as there may be instances where it is beneficial to the bacterium to not scavenge fH too readily.

ErpC has also been found to bind CFHR-2 via its two C-terminal domains. However, contrary to other published studies, no interactions could be observed with CFHR-1. Further work is required to confirm these results and measure the affinity of CFHR-2 to determine whether the interaction is specific and physiologically relevant. The ability of ErpC to bind CFHRs may be linked to its expression profile. Contrary to CspA, ErpC is expressed throughout colonisation of a mammalian host, suggesting that the role of ErpC is slightly different to that of CspA, and is required throughout infection rather than just at transferral from tick to mammalian host and *vice versa*.

An interesting observation from published studies is that ErpC does not bind CFHL-1. This is surprising given that domains five through seven of CFHL-1 are identical to those of factor H, which is bound by ErpC. It is likely that CFHL-1 is scavenged by ErpC, but due to its low serum concentration, it is competed away by the binding of factor H and the CFHRs which are much more abundant, making it difficult to detect. This argument could also explain recent reports that *Borrelia burgdorferi* strains that only express ErpC do not appear to bind factor H [Hammerschmidt, Hallstrom *et al.* 2012] in the presence of an excess of CFHRs [Hammerschmidt, Hallstrom *et al.* 2012].

Due to the high affinity between ErpC and fH_{567} further work should be undertaken to obtain the X-ray crystal structure of the complex between the two. This would allow proper elucidation of the binding site, and show whether ErpC is mimicking GAGs in a similar manner to fHbp, or is binding in a different manner. Heparin competition and mutagenesis studies should also be undertaken to investigate this possibility. Both CspA and ErpC have potential as vaccine candidates to provide prophylaxis against Borrelial infection and hence prevent its development into Lyme's disease. Knowledge of the atomic resolution structures of these molecules may enable structure based vaccine design, which could also be extended to knock-out factor H binding in order to elicit antibodies that may either by bactericidal or block the interactions. A structure of the complex of the CRASP proteins in complex with complement regulators could also enable design of small molecule inhibitors of the interaction which could be used for treatment of established infection.

4.9 Summary

Similarly to CspA, ErpC has been shown to bind factor H in the region of SCR domain five with a similar affinity (K_D = 2 nM) leading to a theory that interactions of this precise affinity are necessary for the mechanism of complement evasion. In contrast to the dimeric functional form of CspA, ErpC binds factor H in a monomeric form. The atomic resolution of ErpC has been determined using X-ray crystallography showing the protein to have a completely different architecture to that of other factor H binding proteins. Some flexible loops are not observed in the structure suggesting the location of the binding interface and possibly accounting for the reported ability of ErpC to bind multiple complement regulators.

4.9.1 Future directions

The structure of the complex between ErpC and the C-terminal domains of CFHR-2 should be pursued in order to determine the location and nature of the binding interface. Attempts to obtain the structure of the complex between ErpC and factor H domains five and six should also be undertaken. In order to determine whether ErpC is a GAG mimic in a similar manner to fHbp, competition assays with heparin could also be performed.

Chapter Five: Structure of the oligomerisation domain from a C4bp homologue in *Gallus gallus*

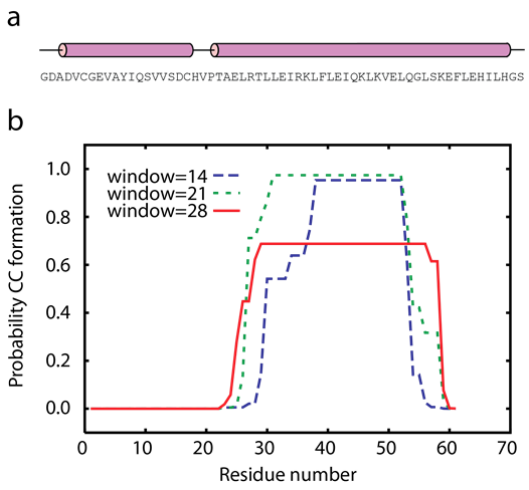


Figure 5.1 *Panel a*: Secondary structure prediction on the CRES oligomerisation domain performed by the PSIPRED server [Buchan, Ward *et al.* 2010]. *Panel b*: Prediction of coiled-coil formation by the C-terminal helix of the CRES oligomerisation domain by the COILS server.

Complement regulatory secretory protein (CRES) has been identified as a homologue of the human C4bp alpha chain found in the equivalent of the RCA locus in *Gallus gallus* [Oshiumi, Shida *et al.* 2005]. Possessing ten SCR domains, compared to seven in C4bp, CRES has been shown to have complement regulatory activity and multimerises into a disulphide linked heptameric species via an oligomerisation domain analogous to that found in the human protein [Oshiumi, Shida *et al.* 2005; Ogun, Dumon-Seignovert *et al.* 2008]. Secondary structure prediction of the CRES oligomerisation domain (Figure 5.1, panel a) suggests that the monomer consists of two alpha helices joined by a short linker. The COILS server [Lupas, Van Dyke *et al.* 1991] predicts that the C-terminal helix is involved in coiled-coil formation (Figure 5.1, panel b), suggesting a mechanism of heptamerisation. However, this hypothesis has yet to be confirmed and no model is available for the involvement of the N-terminal helix, the nature of the coiled-coil (parallel or antiparallel) or the pattern of disulphide bonds.

The CRES oligomerisation domain has been shown to display vaccine adjuvant-like properties, conferring enhanced immunogenicity towards antigens fused to either end of the protein [Ogun, Dumon-Seignovert *et al.* 2008]. Whilst multimerisation of the antigen is a technique often used to improve a vaccine's efficacy, the authors also saw a

significant increase in T helper type 1 (T_H1) associated antigen-specific IgGs, not accounted for by heptamerisation alone, suggesting that there is an increased T-cell response possibly due to the incidental presence of T-cell epitopes in the oligomerisation domain [Draper, Moore *et al.* 2008; Ogun, Dumon-Seignovert *et al.* 2008].

In order to investigate how CRES assembles, structural studies and biophysical characterisation of the oligomerisation domain will be performed in order to confirm the formation of a heptameric coiled-coil and ascertain the pattern of disulphide bonding. Knowledge of the multimerisation mechanism may also enable insight into the assembly of C4bp and the adjuvant-like properties associated with the molecule.

5.1 Protein expression and purification

CRES is expressed as a recombinant protein in *Escherichia coli* prior to purification and *in vitro* disulphide bond formation. This expression system gives high yields, allows straightforward mutation of constructs and has the potential for selective labelling of proteins via incorporation of selenomethionine or selenocysteine residues.

5.1.1 Details of plasmid pIMX-315

Plasmid pIMX-315 was kindly provided by Dr. Fergal Hill (Imaxio SA, France) and contains the oligomerisation domain of CRES expressed under the control of an inducible T7 promoter (Figure 5.3). A mutated form of pIMX-315, pIMX-478, was also generated containing a Val443Met mutation (discussed in Chapter 5.4). The numbering used throughout this thesis is consistent with that in the CRES sequence deposited in the UNIPROT database with accession identifier E1C2T5 [Wallis, Aerts *et al.* 2004].

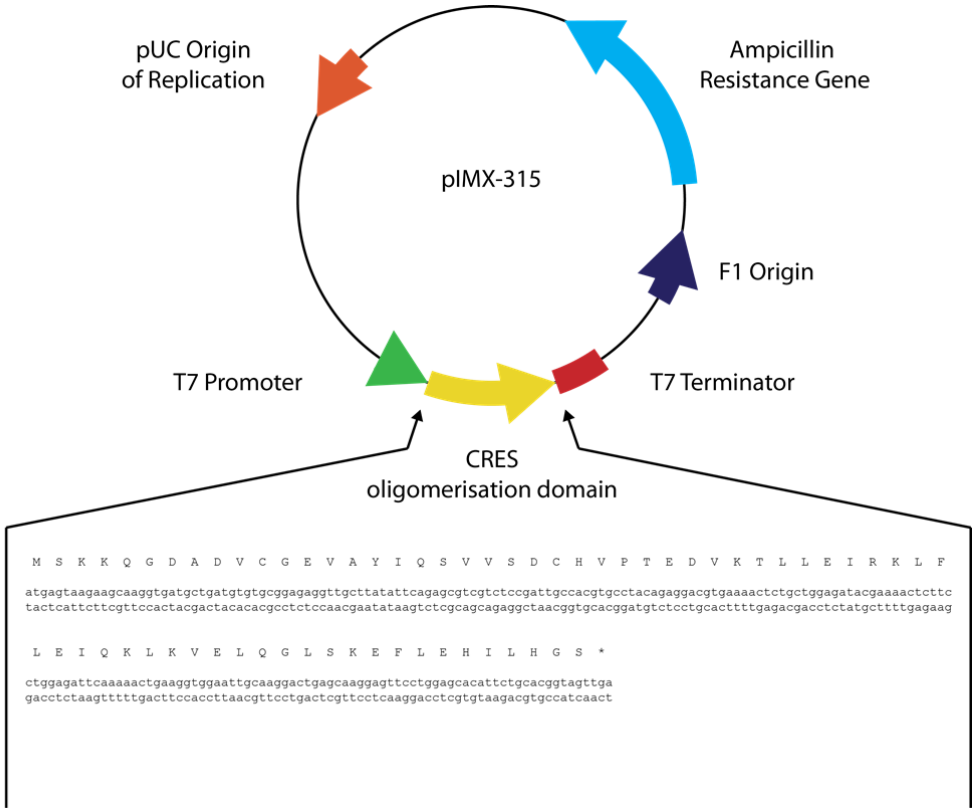


Figure 5.2 Schematic representation of pIMX-315. The oligomerisation domain from CRES is expressed without an affinity tag from a vector that was produced by cloning the CRES fragment into a pRSET vector (Invitrogen, USA) between the XbaI and HindIII restriction sites.

5.1.2 Expression and purification of CRES oligomerisation domain

CRES was purified using a protocol based on one supplied by Dr. Fergal Hill. The over-expressed protein is heat stable, allowing initial purification by thermal precipitation of contaminants. Further purification is achieved using a Ni-NTA column. Although lacking a poly-histidine tag, CRES binds to the immobilised Ni^{2+}, presumably mediated through the C-terminal histidine residues with an affinity that is increased by the avidity effect as a result of multimerisation. The intermolecular disulphide bonds form whilst the protein is captured on the column, possibly encouraged by the increased concentration and the redox potential of the Ni^{2+}.

Escherichia coli strain BL21(DE3) transformed with the pIMX-315 plasmid was grown overnight at 37°C with 180 RPM shaking in 2 × 5 ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 2 × 1 L Luria-Bertani

broth supplemented with 100 μg/ml ampicillin prior to incubation at 37^{o}C with 180 RPM shaking. Cell proliferation was monitored until OD_{600} = 0.6 when CRES expression was induced by addition of 1 mM IPTG followed by continued incubation for three hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4^{o}C) before the pellets were combined and stored at -20^{o}C until required.

Defrosted cell pellets were resuspended in 40 ml of 50 mM $NaPO_4$, 250 mM NaCl, 20 mM Imidazole, pH 7.5 with 1 EDTA-free protease inhibitor cocktail tablet (Roche, Germany) prior to homogenisation in an Emulsiflex C5 (Avestin, Canada). Cell lysate was cleared by centrifugation (39200 g, 15 minutes, 4^{o}C) and the supernatant incubated at 75^{o}C for 20 minutes in a water bath. Precipitate was removed by centrifugation (39200 g, 20 minutes, 4^{o}C) and the filtered supernatant loaded on a 5 ml Ni-NTA column (Qiagen, USA), pre-equilibrated in 50 mM $NaPO_4$, 250 mM NaCl, 20 mM imidazole, pH 7.5, at 0.5 ml/min. The loaded column was washed with 10 CV of 250 mM NaCl, 20 mM imidazole, pH 7.5 at 5 ml/min on an AKTA fplc system (GE Healthcare, UK) prior to elution with a 20 – 500 mM gradient of imidazole over 10 column volumes (Figure 5.3, panel b). Fractions containing CRES were pooled and 1 mM EDTA added to complex any Ni^{2+} leached from the column. The sample was concentrated to 5 ml using a 30000 MWCO PES spin concentrator (Millipore, UK) prior to a final size exclusion purification step using an S75 26/60 column (GE Healthcare, UK) pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 7.4 (Figure 5.3, panel c). The peak containing the target protein elutes at approximately 0.44 CV with a typical yield of 20 mg/l culture.

The course of the purification was charted using SDS-PAGE showing formation of a disulphide liked multimer (Figure 5.3, panel a). MALS was also employed to assess the solution stoichiometry of the complex, showing that the CRES oligomerisation domain

is heptameric with a mass of 50 kDa (Figure 5.3, panel d), exactly seven times that of the monomer (7293 Da).

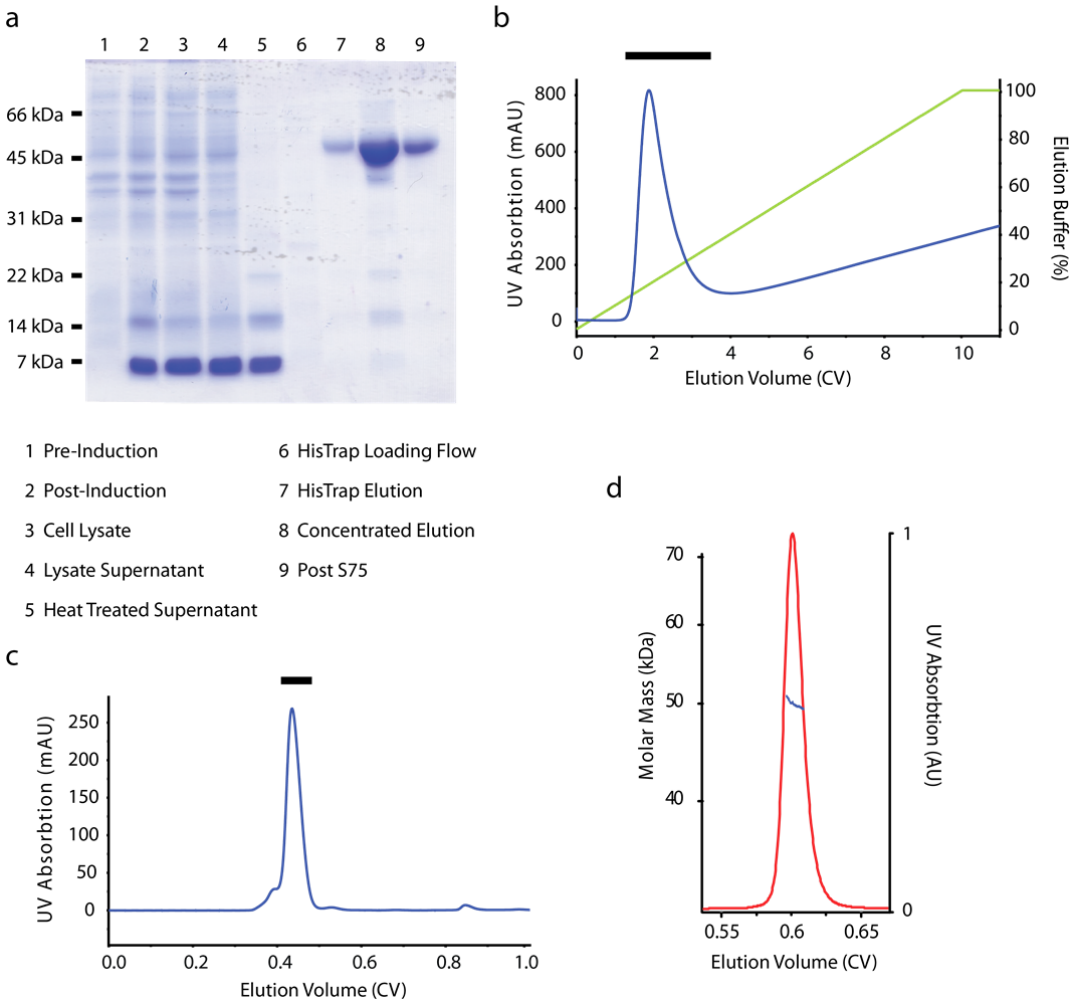


Figure 5.3 *Panel a*: 15% non-reducing SDS-PAGE gel charting the protein purification. *Panel b*: Elution profile from the Ni-NTA column. *Panel c*: Elution profile from an S75 16/60 column. *Panel d*: MALS analysis of purified material.

5.1.3 Expression and purification of selenomethionine derivatized CRES oligomerisation domain

Methionine auxotrophic *Escherichia coli* strain B834(DE3) transformed with the pIMX-478 plasmid was grown overnight at 37°C with 180 RPM shaking in 5 ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 100 ml Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37°C with

180 RPM shaking. Cell proliferation was monitored until OD_{600} = 0.6 when cells were harvested by centrifugation (4000 g, 10 minutes, 4°C) and washed twice by resuspension in 60 ml of PBS followed by centrifugation (4000 g, 10 minutes, 4°C). The final cell pellet was resuspended in 10 ml of PBS and then used to inoculate 2 × 1 L SelenoMet medium (Molecular Dimensions, UK) supplemented with 40 mg/l L-Selenomethionine (Molecular Dimensions, UK). Cell proliferation was monitored until OD_{600} = 0.6 when ErpC expression was induced by addition of 1 mM IPTG followed by continued incubation at 21°C with 180 RPM shaking overnight. Cells were harvested by centrifugation (4000 g, 20 minutes, 4°C) before the pellets were combined and stored at -20°C until required. Further purification of selenomethionine derivatised CRES, $CRES_{semet}$, proceeded via the same protocol described in section 5.1.2 with a yield of approximately 20 mg per litre of culture (assuming ε_{280} = 1615 M^{-1} cm^{-1}). Incorporation of selenomethionine was assessed using ES-TOF mass spectrometry returning a mass of 51686.3 Da indicating 100% integration.

5.2 Crystallisation, data collection and processing of CRES oligomerisation domain

The structure of the oligomerisation domain from CRES is expected to contain a heptameric coiled-coil. There is only one published structure of a homo-heptameric coiled-coil structure, PDB entry 2HY6 [Liu, Zheng *et al.* 2006], which is an engineered version of a leucine-zipper protein and is not expected to be an appropriate model for molecular replacement. In order to solve the structure, experimental phases will have to be obtained.

5.2.1 Crystallisation of CRES

CRES was concentrated to A_{280} = 0.19 and initial crystallisation trials were performed at 294K using vapour diffusion in 400 nL sitting drops produced by an Oryx Nano crystallisation robot (Douglas Instruments, UK) within MRC crystallisation plates (Molecular Dimensions, UK). Stura, Macrosol, Pact Premier, JCSG, Structure 1, Structure 2, PGA-LM, Morpheus and Proplex screens (Molecular Dimensions, UK) were used, with two drops set up for each condition containing protein to precipitant ratios of 1:1 and 1:3.

Multiple, small, thin, plate-like crystals grew overnight in approximately 30% of conditions screened (Figure 5.4, panel a) with the largest of these growing in 0.2 M ammonium acetate, 0.1 M sodium acetate, 8% (w/v) PEG 3350, pH 4.5. Many attempts at optimisation of these crystals were attempted including batch seeding, micro seeding, poisoning with glycerol and MPD, and addition of additives from the Silver Bullets and Silver Bullets Bio screens (Hampton Research, USA). Large, single crystals were finally obtained using a mother liquor made using an aged batch of PEG 3350 and a drop size of 800 nl (Figure 5.4, panel b). Crystals were mounted in Cryoloops (Hampton Research, USA) prior to cryoprotection using 15% ethylene glycol and flash cooling in liquid nitrogen.

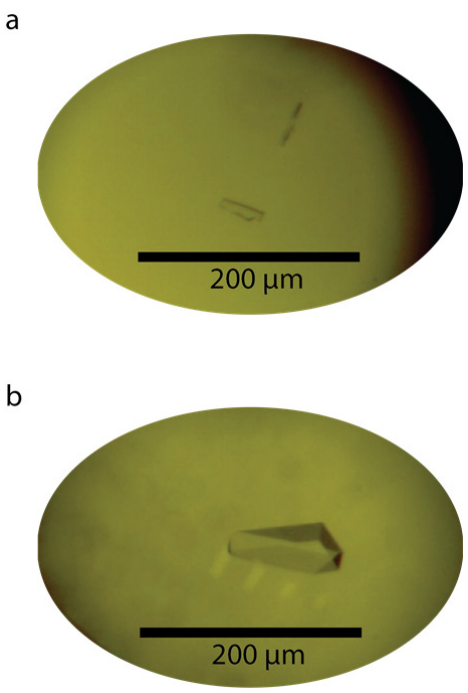


Figure 5.4 *Panel a*: Crystals of the CRES oligomerisation domain prior to optimisation. *Panel b*: Optimised crystals of the oligomerisation domain from CRES.

In order to obtain experimental phases, crystals were soaked for varying times between 30 seconds and 15 minutes in 10 mM heavy atom solutions (made up in mother liquor) prior to backsoaking in the cryoprotectant and flash cooling. Heavy atom compounds used in this manner included potassium tetrachloroplatinate, mercury (II) chloride, samarium (III) chloride and gadolinium (III) chloride.

5.2.2 Data collection of CRES

	Native	Gd-SAD	Sm-SAD	Pt-SAD	S-SAD
Diffraction source	Diamond I03	ESRF BM14	Diamond I04	ESRF BM14	ESRF ID29
Detector	ADSC Q315r	MAR 225	ADSC Q315r	MAR 225	ADSC Q315
Energy (eV)	12659	7251	7797	11921	6525
Wavelength (Å)	0.9794	1.7099	1.5900	1.0400	1.9000
Temperature (K)	120	120	120	120	120
Detector resolution (Å)	2.5	2.8	3.5	3.0	3.5
Exposure per image (s)	1.0	150	1.0	150	0.5
Φ_{start}	300	278	20	27	350
Φ_{end}	60	374	220	117	105
$\Delta\Phi$	1.0	1.0	1.0	1.0	1.0
Number Images	120	96	200	90	835
f′	-	-9.22	-7.57	-9.85	0.32
f″	-	28.68	12.92	9.68	0.52

Table 5.1 Data collection parameters for the native, Gd-SAD and S-SAD datasets used in subsequent structure solution and refinement. The parameters for the other SAD data collections are exemplars from multiple datasets collected, but were found to lack anomalous correlations.

Data were collected at the ESRF and Diamond light source using the rotation method as outlined in Table 5.1. Diffracting crystals were indexed and strategized in a base centred orthorhombic cell (average cell dimensions a = 83.9 Å, b = 140.2Å, c = 85.5Å, $\alpha = \beta = \gamma = 90.0^{o}$) with C222 symmetry using IMOSFLM [Battye, Kontogiannis *et al.* 2011]. Prior to data collection from the heavy atom soaks, an X-ray stimulated fluorescence scan was collected from each derivatized crystal over the L3 absorbtion edge of the heavy atom. The spectra were analysed with CHOOCH [Evans and Pettifer 2001] to

determine the peak wavelength and the corresponding f' and f'' values (Table 5.1). Data were then collected at the peak to complement any isomorphous signal with anomalous scattering.

5.2.3 Data processing of CRES

	Native	Gd-SAD	Sm-SAD
Spacegroup	$C222_1$	$C222_1$	$C222_1$
Unit-cell dimensions (Å)	a=83.9, b=140.2 c=85.5	a=90.2 b=137.6, c=85.6	a=90.6, b=138.7, c=85.7
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	35.3	70.6	24.0
Resolution (Å)	85.46-2.24 (2.36-2.24)	40.88-3.15 (3.32-3.15)	138.7-3.66 (3.85-3.66)
R_{merge}	0.058 (0.493)	0.111 (0.483)	0.170 (0.478)
R_{pim} (I+/I-)	0.046 (0.432)	0.084 (0.343)	0.200 (0.289)
<I/σ(I)>	13.5 (2.2)	6.5 (2.2)	10.8 (7.4)
Anomalous completeness (%)	96.0 (79.6)*	93.6 (91.9)	98.9 (99.8)
Anomalous multiplicity	4.3 (2.8)*	1.8 (1.8)	3.6 (3.7)
Anomalous correlations	-	0.289 (0.028)	0.023 (0.0)

	Pt-SAD	S-SAD
Spacegroup	$C222_1$	$C222_1$
Unit-cell dimensions (Å)	a=87.2, b=138.2, c=86.0	a=86.4, b=139.4, c=85.6
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	68.5	63.0
Resolution (Å)	61.23-2.65 (2.72-2.65)	139.4-3.11 (3.27-3.11)
R_{merge}	0.070 (0.512)	0.179 (0.292)
R_{pim} (I+/I-)	0.041 (0.240)	0.183 (0.235)
<I/σ(I)>	11.4 (2.0)	20.9 (2.0)
Anomalous completeness (%)	84.5 (42.7)	90.3 (51.9)
Anomalous multiplicity	1.9 (1.6)	19.0 (1.7)
Anomalous correlations	0.063 (0.076)	0.0 (0.0)

Table 5.2 Data processing statistics for each dataset. Values marked with a star are the respective non-anomalous statistics.

All six datasets were processed using AUTOPROC invoking usage of MOSFLM [Leslie and Powell 2007] for indexing, XDS [Kabsch 2010] for integration and SCALA [Evans 2006] for merging and scaling. Analysis of the reflections using POINTLESS

[Evans 2006] highlighted a two-fold screw axis parallel to c due to the absence of any odd reflections along l, making the spacegroup $C222_1$ (Figure 5.5, panel b). Inspection of the anomalous correlations in each dataset suggested that only the Gd soak and long wavelength S-SAD experiment contained any anomalous signal indicative of a well ordered heavy atom site (Figure 5.5, panel a). Data processing statistics are displayed in Table 5.2.

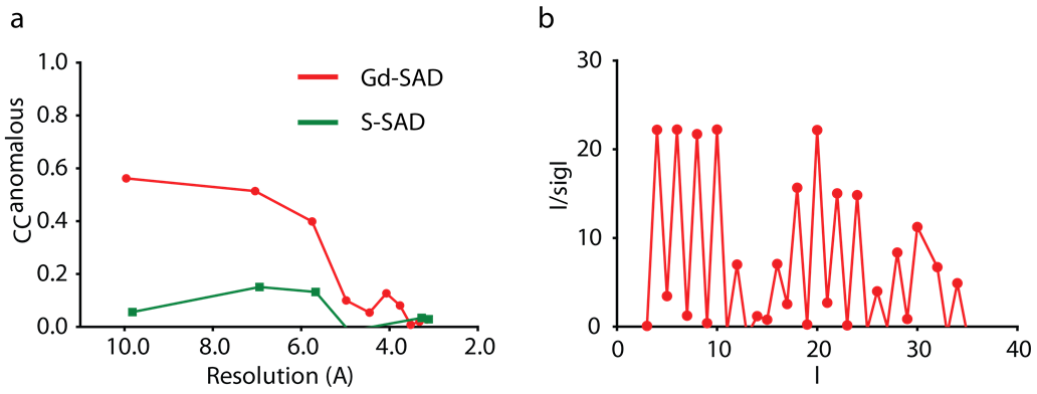


Figure 5.5 *Panel a*: Anomalous correlations within the Gd-SAD and S-SAD datasets. *Panel b*: Systematic absences along I describing a screw axis along c.

5.2.4 Attempted phasing of CRES from heavy atom soaks

Analysis of the contents of the $C222_1$ unit cell using MATTHEWS [Matthews 1968] suggests the presence of seven copies of the CRES oligomerisation domain in the asymmetric unit (solvent content = 50.3%). Subsequent scrutiny of the self-rotation function calculated using MOLREP [Vagin and Teplyakov 1997] (Figure 5.6) reveals a peak in the chi = 52° plot suggesting the presence of a seven-fold rotational non-crystallographic symmetry parallel to b. Combination of these observations suggests that the crystal comprises the heptameric protein, which itself is seven-fold symmetric.

Initial attempts to solve the structure using the isomorphous and anomalous signal from the Gd soak and the S-SAD data were unsuccessful. AutoSHARP [Vonrhein, Blanc *et al.* 2007], HYSS [Grosse-Kunstleve and Adams 2003], CRANK [Bailey 1994] and HKL2MAP [Pape and Schneider 2004] were used to locate heavy atom sites and generate phases, but no reliable heavy atom sites could be found.

In order to obtain phases, selenomethionine derivatisation was used to incorporate electron dense selenium atoms in a controlled manner that can be verified using mass spectrometry.

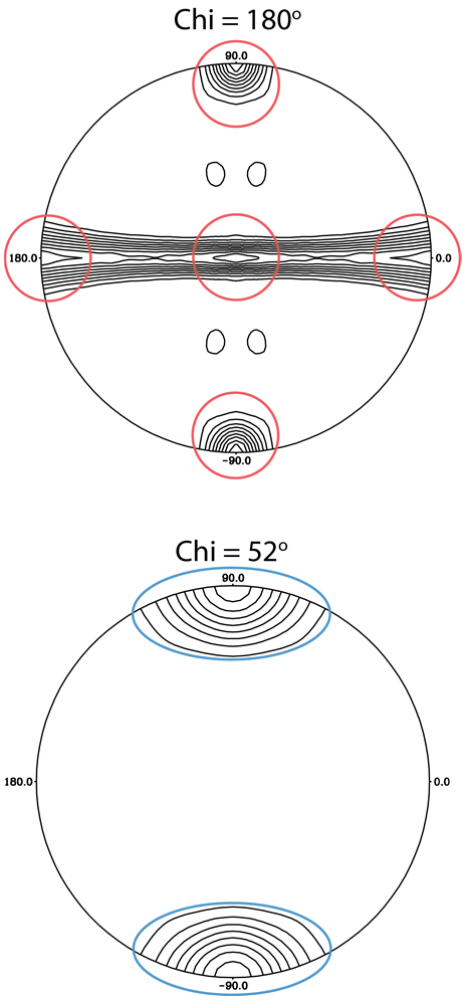


Figure 5.6 Self-rotation functions calculated from the native dataset. Peaks arising from crystallogrpahic symmetryare highlighted in red. Other significant peaks are highlighted in blue.

5.4 Crystallisation, data collection and processing of $CRES_{SeMet}$

Unfortunately, the oligomerisation domain from CRES contains no intrinsic methionine residues except for the N-terminal formyl-methionine which is often lost upon expression or disordered in crystal structures. In order to introduce a methionine residue, sequence alignments were produced highlighting a methionine residue in the predicted second helix found in the zebrafinch. This mutation, Val44Met, was introduced and selenomethionine derivatized protein expressed in strain B834(DE3) of *Escherichia coli* as described in Chapter 5.1.3.

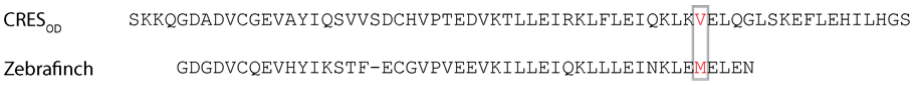


Figure 5.7 Alignments showing the introduction of a methionine residue into the CRES oligomerisation domain.

5.4.1 Crystallisation of CRES$_{SeMet}$

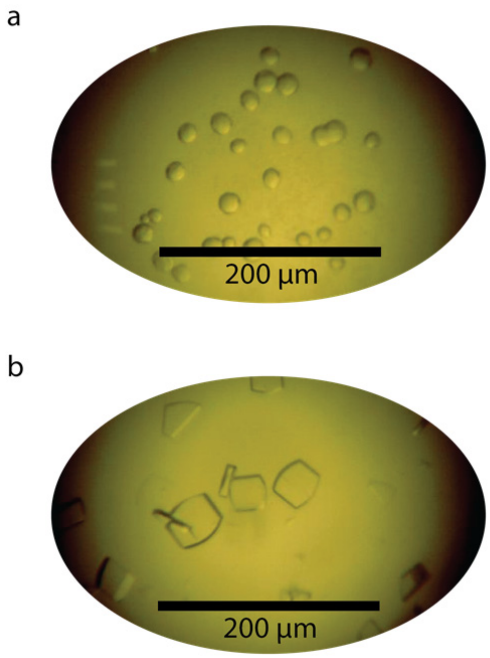


Figure 5.8 *Panel a*: Crystals of the CRES oligomerisation domain grown in condition 1. *Panel b*: Crystals grown in condition 2.

CRES$_{SeMet}$ was concentrated to A_{280} = 6.8 prior to crystallisation trials performed in an identical manner to the wild type protein (Chapter 5.2.1). Crystals grew after approximately six weeks in 0.5 M ammonium sulphate, 10% (w/v) PEG 4000 (condition 1. Figure 5.8, panel a) and 0.2 M lithium sulphate, 20% (w/v) PEG 1000, 0.1 M phosphate/citrate buffer pH 4.2 (condition 2, Figure 5.8, panel b). Samples were mounted in Cryoloops (Hampton Research, USA) prior to cryoprotection with 20% ethylene glycol and flash cooling in liquid nitrogen.

5.4.2 Data collection of CRES$_{SeMet}$

Data were collected on beamline ID23-1 at the ESRF using the rotation method. CRES$_{SeMet}$ crystals grown in condition 1 were indexed in a face centred orthorhombic cell (cell dimensions a = 111.8 Å, b = 122.8 Å, c = 156.0 Å, $\alpha = \beta = \gamma = 90.0^{o}$) with F222 symmetry prior to strategizing for anomalous data collection in IMOSFLM [Battye, Kontogiannis *et al.* 2011]. Collection of a single two wavelength MAD dataset

used in subsequent attempts at structure solution is detailed in Table 5.3, using peak and inflection wavelengths calculated by CHOOCH from a fluorescence scan over the L3 edge of selenium. Crystals from condition 2 were indexed in a body centred orthorhombic cell (cell dimensions a = 63.8 Å, b = 111.2 Å, c = 154.6 Å, α = β = γ = 90.0^{o}) with I222 symmetry and collection of two, two wavelength MAD datasets is detailed in Table 5.3.

	Condition 1		Condition 2 (Dataset 1)		Condition 2 (Dataset 2)	
	Peak	**Remote**	**Peak**	**Remote**	**Inflection**	**Remote**
Diffraction source	ESRF ID23-1	ESRF ID23-1	ESRF ID23-1	ESRF ID23-1	ESRF ID23-1	ESRF ID23-1
Detector	ADSC Q315r	ADSC Q315r	ADSC Q315r	ADSC Q315r	ADSC Q315r	ADSC Q315r
Energy (eV)	12664	12700	12664	12700	12660	12677
Wavelength (Å)	0.9790	0.9762	0.9790	0.9762	0.9793	0.9780
Temperature (K)	120	120	120	120	120	120
Detector resolution (Å)	2.5	2.5	3.0	2.5	2.5	2.5
Exposure per image (s)	0.2	0.2	0.2	0.2	0.2	0.2
Φ_{start}	210	210	278	278	71	71
Φ_{end}	30	30	98	98	251	251
$\Delta\Phi$	1.0	1.0	1.0	1.0	1.0	1.0
Number Images	180	180	180	180	180	180
f' (electrons)	-6.44	-4.9	-6.44	-4.9	-10.14	-5.5
f'' (electrons)	6.33	4.0	6.33	4.0	4.1	4.2

Table 5.3 Data collection parameters for crystals of selenomethionine derivatised CRES grown in both conditions. f' and f'' values for peak and inflection wavelengths were experimentally derived from fluorescence scans, those for remote wavelengths were calculated using CROSSEC [Cromer 1983].

5.4.3 Data processing of $CRES_{SeMet}$

All $CRES_{SeMet}$ data were processed using autoPROC, invoking usage of MOSFLM for indexing, XDS for integration, SCALA for merging and scaling, and POINTLESS for spacegroup determination. Data processing statistics are presented in Table 5.4. Anomalous correlations from the selenium atoms are present in all datasets (Figure 5.9).

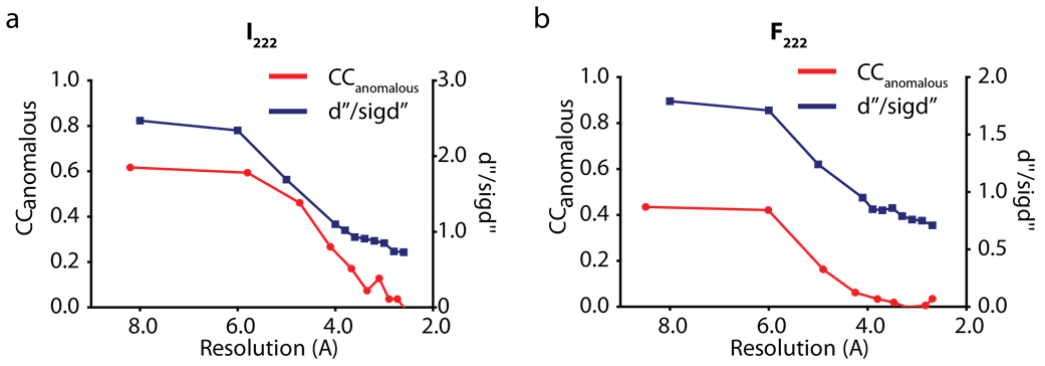


Figure 5.9 *Panel a*: Plots of the anomalous correlations and strength of the anomalous differences in the I222 crystal form. *Panel b*: Identical plots for the F222 crystal form.

	Condition 1 Dataset 1 Peak	**Condition 1 Dataset 2 Remote**	**Condition 2 Dataset 1 Peak**
Spacegroup	F222	F222	I222
Unit-cell dimensions (Å)	a=109.5, b=122.9 c=156.3	a=109.5, b=122.9, c=156.3	a=63.8, b=111.2, c=154.6
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	62.3	57.3	62.5
Resolution (Å)	78.27-2.69 (2.83-2.69)	78.27-2.83 (3.00-2.83)	90.27-2.59 (2.73-2.59)
R_{merge}	0.067 (0.476)	0.085 (0.62)	0.055 (0.493)
R_{pim} (I+/I-)	0.080 (0.202)	0.100 (0.266)	0.070 (0.310)
<I/σ(I)>	18.8 (4.3)	17.9 (3.4)	19.8 (2.3)
Anomalous completeness (%)	99.6 (99.9)	99.8 (100.0)	93.3 (69.8)
Anomalous multiplicity	3.7 (3.8)	3.8 (3.8)	3.5 (2.3)
Anomalous correlations	0.29 (0.04)	0.18 (0.13)	0.49 (0.0)

	Condition 2 Dataset 2 Remote	**Condition 2 Dataset 3 Inflection**	**Condition 2 Dataset 4 Remote**
Spacegroup	I222	I222	I222
Unit-cell dimensions (Å)	a=63.8, b=111.2, c=154.6	a=63.8, b=111.0, c=153.6	a=63.8, b=111.0, c=153.6
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	65.0	59.5	59.6
Resolution (Å)	90.27-2.59 (2.73-2.59)	89.99-2.77 (2.92-2.77)	89.99-2.77 (2.92-2.77)
R_{merge}	0.061 (0.617)	0.075 (0.610)	0.069 (0.477)
R_{pim} (I+/I-)	0.073 (0.261)	0.090 (0.258)	0.083 (0.201)
<I/σ(I)>	19.6 (3.2)	18.4 (3.3)	18.1 (3.9)
Anomalous completeness (%)	99.8 (99.7)	99.4 (99.0)	99.3 (98.9)
Anomalous multiplicity	3.8 (3.8)	3.8 (3.8)	3.8 (3.8)
Anomalous correlations	0.21 (0.0)	0.23 (0.0)	0.28 (0.0)

Table 5.4 Data processing statistics for each wavelength in each dataset from crystals grown in both conditions.

5.5 Structure determination and refinement of $CRES_{SeMet}$

CRES and $CRES_{SeMet}$ have crystallised in different cells and spacegroups; ruling out combination of isomorphous and anomalous signal in a SIRAS phasing experiment. Structure solution was achieved using SAD and MAD approaches.

5.5.1 Se-SAD phasing of $CRES_{SeMet}$

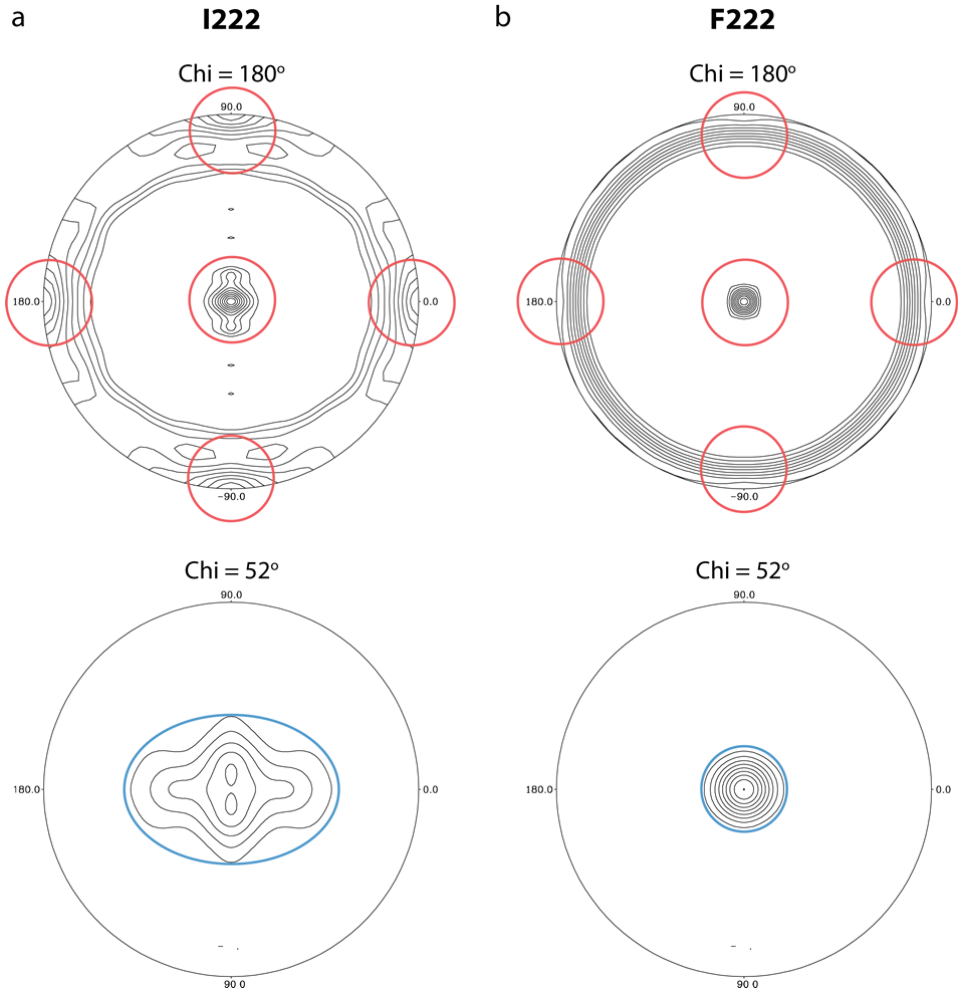


Figure 5.10 *Panel a*: Self-rotation functions calculated from the I222 peak dataset using POLARRFN [Kabsch 1976]. Peaks resulting from crystallographic symmetry are shown in red. Other major peaks are highlighted in blue. *Panel b*: Self-rotation functions from the F222 peak wavelength dataset.

Analysis of the content of the contents of the asymmetric units of the F222 and I222 crystal forms using MATTHEWS suggests the presence of one copy of the heptamer in each asymmetric unit, whilst analysis of the chi = 52° self rotation functions hints at the

presence of a seven-fold non-crystallographic rotational symmetry axis parallel to c in both crystals (Figure 5.10). The circular patterns in both chi = 180° self-rotation functions suggest that the molecule is highly symmetric, perhaps having a cylindrical appearance, with the axis parallel to c.

	x	y	z	u	v	w
Se site 1	0.5302	0.8256	0.0397	0.0	-1.6512	-0.0794
Se site 2	0.8145	0.6739	0.0748	0.0	-1.3478	-0.1496
Se site 3	0.4699	0.5489	0.0909	0.0	-1.0978	-0.1818
Se site 4	0.3161	0.6356	0.0682	0.0	-1.2712	-0.1364
Se site 5	0.6917	0.5614	0.0952	0.0	-1.1228	-0.1904
Se site 6	0.2455	0.1938	0.0536	0.0	-0.3876	-0.1072
Se site 7	0.1720	0.0837	0.5820	0.0	-0.1674	-1.1640

Table 5.5 Locations of the Se atoms in the I222 unit cell and their locations in the X = 0 Harker section of the anomalous difference Patterson, calculated using the Harker vector in Table 5.5 giving X = 0.

Symmetry operations	Harker vector
x, y, z	0, 0, 0
-x, -y, z	-2x, -2y, 0
-x, y, -z	-2x, 0, -2z
x, -y, -z	0, -2y, -2z

Table 5.6 Symmetry operations in the I222 spacegroup and their corresponding Harker vectors.

AutoSHARP was employed to find the expected seven selenium sites in each unit cell of each crystal form using anomalous signal in the peak datasets. Analysis of the anomalous correlations against resolution using SHELXC [Sheldrick 2010] suggests a high resolution cut off of 5 Å for the signal in both datasets based upon the strength of the Bijvoet differences. Sites were then found from the selenium SAD signal using SHELXD [Sheldrick 2010]. Seven selenium sites were found in the I222 data with details of the sites, and corresponding peaks in the respective Harker sections of the anomalous Patterson, shown in Tables 5.5, Table 5.6 and Figure 5.11. However, no reliable sites were found in the F222 data and phasing with these data was not pursued further.

The seven selenium sites found in the I222 data were subjected to seven cycles of refinement in SHARP [Bricogne, Vonrhein *et al.* 2003] prior to phase calculation giving overall figures of merit of 0.175 and 0.052 for acentric and centric reflections

respectively. Overall phasing power and corresponding resolution cutoffs are shown in Table 5.7. Solvent flattening of the phases derived from both hands using SOLOMON [Abrahams and Leslie 1996] with an estimated solvent content of 53.5% gave correlation coefficients of 0.1383 and 0.1091 to the original and inverted hands respectively. Although these values are low, the difference between them is significant suggesting that the original hand is the correct solution. Optimisation of the solvent content in SOLOMON gave an optimal solvent content of 62%, with a correlation coefficient of 0.60, and the resulting map clearly showed helical secondary structure (Figure 5.12, panel a).

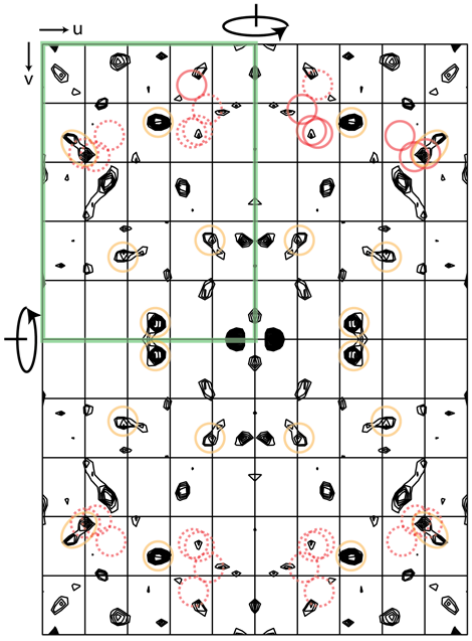


Figure 5.11 X = 0 Harker section from the anomalous difference Patterson map calculated from the I222 crystal form showing the location of the selenium sites in red, and cross peaks in orange. The asymmetric unit is highlighted in green.

	Dataset 1 Peak	Dataset 1 Inflection	Dataset 2 Peak	Dataset 2 Inflection
Phasing power				
Overall	0.731	0.541	0.571	0.510
Drops below 1.0 (Å)	4.10	5.20	5.30	5.15

Table 5.7 Overall phasing power and corresponding resolution cutoff for each wavelength in each dataset.

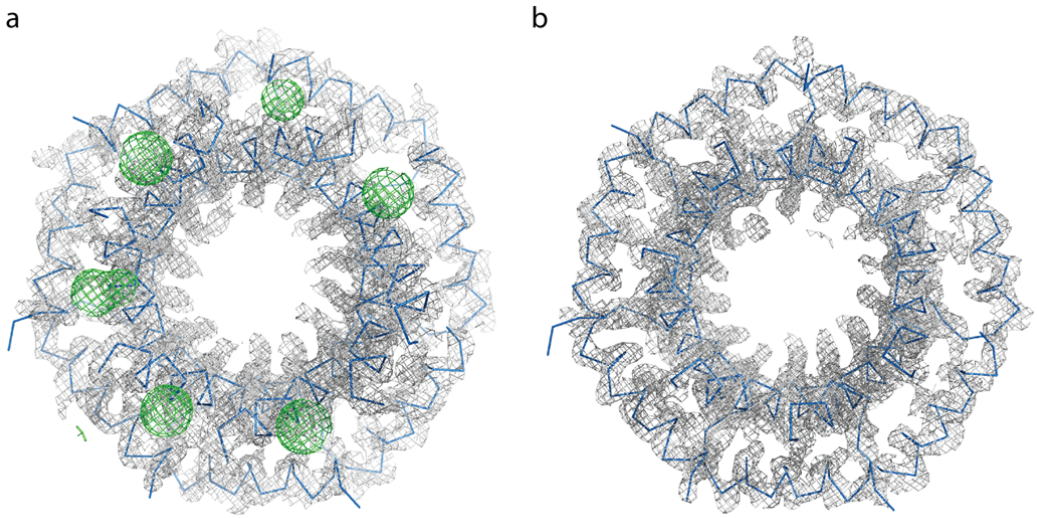


Figure 5.12 *Panel a*: Electron density map calculated from experimental phases from the Se-SAD experiment, contoured at 1.5 σ (grey). The refined model is shown as a blue ribbon for clarity. The selenium sites are shown in green (contoured at 5 σ), displaying seven sites, one of which is duplicated and may suggest disorder, which may also explain the missing site. *Panel b*: Electron density map (contoured at 1.5 σ) calculated using the improved phases following the inclusion of the Se-MAD data.

5.5.2 Se-MAD phase improvement of $CRES_{SeMet}$

Prior to building an initial model of $CRES_{SeMet}$ the phases were improved by combination of all four wavelengths from both MAD I222 experiments. SHARP was used to refine the selenium sites in the correct hand found in the Se-SAD experiment (Chapter 5.5.1) against these datasets; giving figures of merit of 0.342 and 0.318, for acentric and centric reflections respectively (phasing power and resolution cutoffs are given in Table 5.8), after six cycles of maximum likelihood refinement. Solvent flattening using SOLOMON, given the already proven optimal solvent content of 62%, resulted in a man that was much cleaner than that previously calculated; containing a greater proportion of connective density (Figure 5.12, panel b).

5.5.3 Initial model building of $CRES_{SeMet}$

BUCCANEER [Cowtan 2006] was used to build $CRES_{SeMet}$ in the I222 solvent flattened map. The resulting model had 383 residues in 19 chains with an associated

R/R_{free} of 0.3336/0.3607. Although this model could have been improved and refined in I222, model building and refinement was performed in the high-resolution $C222_1$ crystal form in order to obtain the most accurate model prior to back solving the I222 and F222 crystal forms.

5.6 Structure determination and refinement of CRES II

The partial structure from the I222 crystal form (Chapter 5.5.3) was located into the high resolution $C222_1$ dataset using PHASER [McCoy, Grosse-Kunstleve *et al.* 2007] and the resulting phases were used in a new BUCCANEER autobuilding cycle. The resulting model was an improvement on that built from the I222 selenomethionine data with 349 residues built in 9 chains with an R/R_{free} of 0.2986/0.3643, but was still incomplete. In order to build a more complete model, phases were improved using the Gd-SAD and S-SAD data collected in Chapter 5.2.2.

5.6.1 Gd-SAD phasing of CRES

Experimental location of the Gd sites in the Gd-SAD dataset failed (Chapter 5.2.4). However, using known phases from the partial model, it is possible to locate the gadolinium atoms which can in turn be used to improve the phase information. PHASER was used to place the partial model into the Gd-SAD dataset prior to ten cycles of maximum likelihood restrained refinement in REFMAC [Murshudov, Skubak *et al.* 2011]. HLTOFOM [Winn, Ballard *et al.* 2011] was employed to convert the REFMAC phases from phi/fom to Hendrickson-Latmann coefficients which were combined with the original Gd-SAD dataset with CAD [Winn, Ballard *et al.* 2011]. SHARP was used to calculate anomalous residual maps in the Gd-SAD data in the presence of the external REFMAC phases.

Symmetry Operations	Subtracted
x, y, z	0, 0, 0
-x, -y, ½+z	-2x, -2y, ½
-x, y, ½-z	-2x, 0, ½-2z
x, -y, -z	0, -2y, -2z
½-x, ½-y, ½+z	½-2x, ½-2y, ½
½-x, ½+y, ½-z	½-2x, ½, ½-2z
½+x, ½-y, -z	0, ½-2y, -2z

Table 5.9 Crystallographic symmetry operations and their corresponding Harker vectors for spacegroup $C222_1$.

	x	y	z	u	v	w
Gd site 1	0.3858	0.0484	0.3695	0	0.9032	0.2610
Gd site 2	0.3855	0.0693	0.6392	0	0.8614	0.2772

Table 5.8 Gadolinum sites in the $C222_1$ unit cell and their locations in the X = 0 Harker section of the anomalous difference Patterson, calculated using the corresponding Harker vector in table 5.10.

	Gd-SAD Peak
Phasing power	
Overall	0.371
Drops below 1.0 (Å)	N/A

Table 5.10 Overall phasing power and corresponding resolution cutoff for the Gd-SAD data.

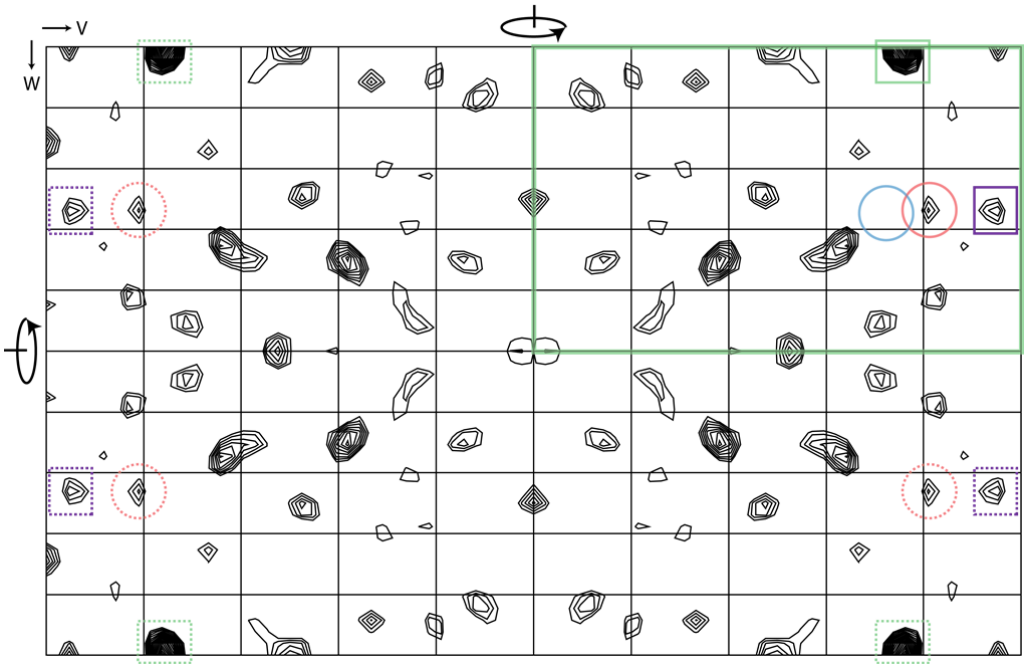


Figure 5.13 X = 0 Harker section of the anomalous difference Patterson map calculated from the Gd-SAD data. Site one is denoted in red, and site two in blue. Cross peaks are shown in purple and green squares. The asymmetric unit is highlighted in green.

The resulting residual map (Figure 5.14) shows two unique gadolinium sites which where subsequently refined using SHARP in the absence of the external phases to reduce bias. Phase calculation from the refined sites gave figures of merit of 0.145 and

0.035 for centric and acentric reflections respectively (overall phasing power and resolution cutoff is presented in Table 5.10). SOLOMON was used to solvent flatten the resulting map with the optimal solvent content of 52.2%, calculated in Chapter 5.3.2. The initial model from REFMAC was also included in the first ten cycles.

Discovery of the gadolinium atom sites allows assignment of the corresponding peaks in the anomalous Patterson (Figure 5.13 and Tables 5.8 and 5.9). Only site 1 is visible in the map, with a very weak peak, whilst the other site is not visible at this contour level. This is reflected in the low occupancies of the sites, being 0.16 and 0.05 respectively. Of particular interest are the cross peaks (denoted in rectangular boxes in Figure 5.13), one of which (highlighted in blue) appears stronger than the heavy atom peaks themselves. This is because the cross peak coincides with a translation vector between the chains that is visible in the native Patterson.

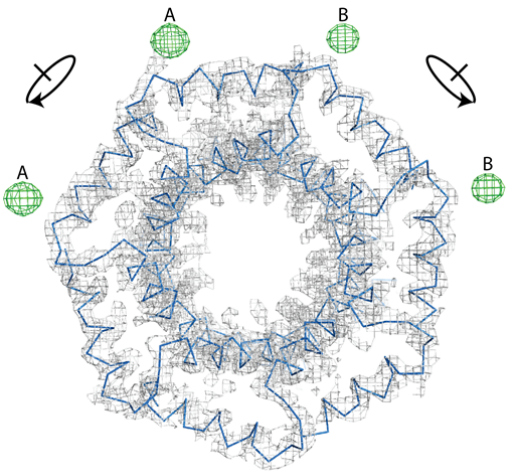


Figure 5.14 Electron density map calculated using the Gd-SAD phases (grey. Contoured at 1.5 σ) with the reisdual map denoting the locations of the heavy atoms (gree. Contoured at 5 σ). The final refined model is shown as a blue ribbon for clarity.

5.6.2 S-SAD phasing of CRES

The S-SAD data were treated in an identical manner to the Gd-SAD data; the same partial model was molecular replaced into the data prior to refinement and conversion of the phases into Hendrickson-Latmann coefficients. An anomalous residual map calculated from these phases in SHARP shows seven unique S sites (Figure 5.15, panel a) which were refined in the absence of external

	S-SAD Peak
Phasing power	
Overall	0.228
Drops below 1.0 (Å)	N/A

Table 5.11 Overall phasing power and corresponding resolution cutoff for the S-SAD data

phases prior to phase calculation with figures of merit of 0.2240 and 0.1130 for acentric and centric reflections respectively (phasing power and corresponding resolution cut off is presented in Table 5.11). The resulting map was solvent flattened using SOLOMON given the optimal solvent content of 52.2% and the presence of the partial model in the first ten cycles (Figure 5.15, panel b).

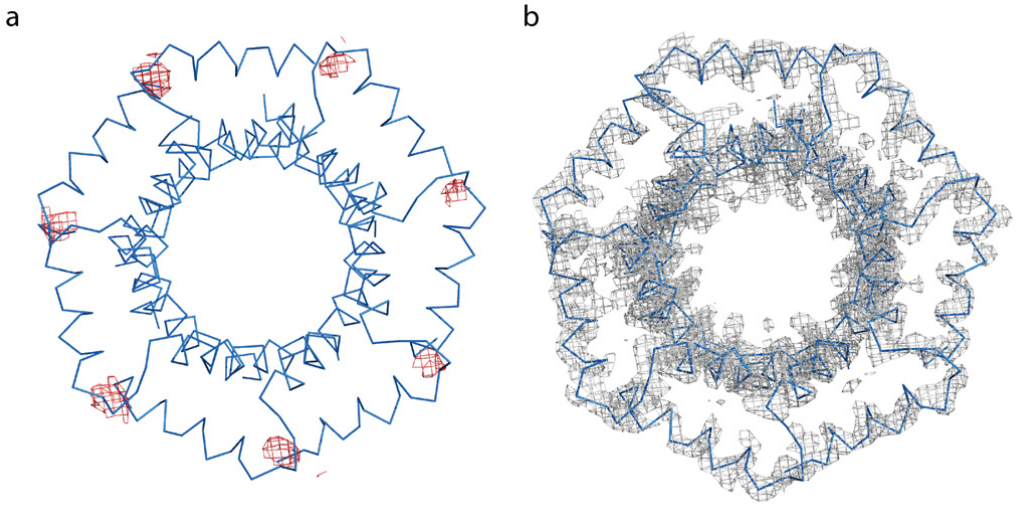


Figure 5.15 *Panel a*: Residuals around the sulphur sites are shown in red (contoured at 3 σ), displaying seven sites which correspond to the disulphide bonds. The refined model is shown as a blue ribbon for clarity. *Panel b*: Electron density map (contoured at 1.5 σ) calculated using the S-SAD data.

5.6.3 Density modification

In order to generate the best possible map to extend the partial model in, the phases calculated from the Gd-SAD and S-SAD data were combined and extended into the high resolution native data in a density modification job. SUPERPOSE was used to calculate the translation and rotation (in Eulerian angles) operators from the partial model in the high resolution native dataset to the PHASER placement of the same model in the Gd-SAD and S-SAD datasets collected and processed in Chapters 5.6.1 and 5.6.2 respectively. A mask around the partial model in the high resolution dataset was produced using NCSMASK [Winn, Ballard *et al.* 2011] at a radius of 3 Å prior to averaging in DMMULTI [Cowtan 2010]; performing 100 cycles of phase extension

starting at 2.6 Å resolution with solvent flattening and histogram matching enabled in each step. The final correlations between the maps were 0.860, 0.924 and 0.910 for the native, Gd-SAD and S-SAD respectively. Observation of the resulting map showed it to contain the greatest amount of contiguous density of any so far, with obvious density for side chains (Figure 5.16, panel a).

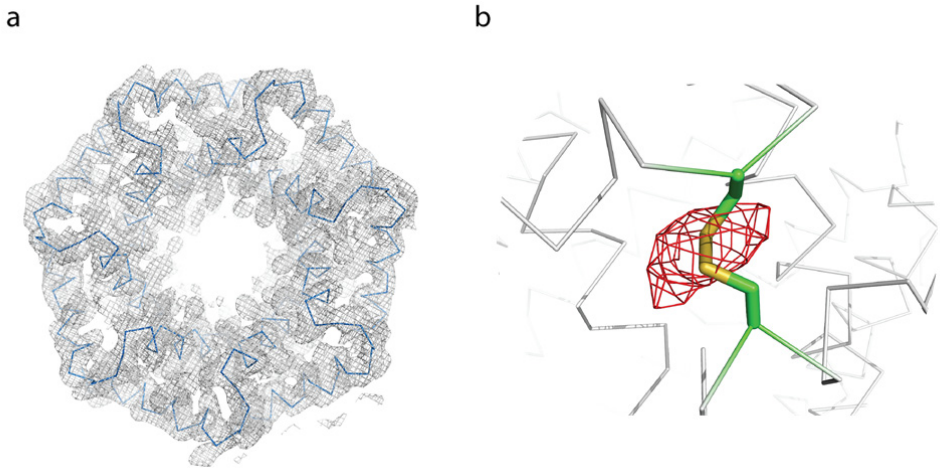


Figure 5.16 *Panel a*: Averaged and phase extended electron density map, contoured at 1.5 σ (grey). The final model is shown as a blue ribbon for clarity. *Panel b*: Residual in the S-SAD map confirming the placement of the disulphide bond.

5.6.4 Model building

The averaged and phase extended map was the highest quality map produced so far; having the most contiguous connective density and clear density for side chains. BUCCANEER was employed to autobuild in this new map, resulting in a model comprising 354 residues in seven chains, the longest of which was 58 residues long, with an R/R_{free} of 0.3160/0.3629.

The model contained obvious evidence for existence of disulphide bonds. Not only were there two cysteine thiol groups within disulphide bonding distance, but they were also within the residuals calculated from the S-SAD data (Chapter 5.6.2), validating their position in the model (Figure 5.16, panel b).

5.6.5 Refinement of CRES

	$C222_1$
Resolution (A)	2.24-55.06 (2.24-2.34)
Number of reflections	23563 (2287)
Number of reflections for R_{free}	1199 (113)
R_{work}/R_{free}	0.2202/0.2357 (0.2530/0.2406)
Number of atoms	
Protein	2882
Ligand/Ion	51 Acetate ions 76 Ethylene glycol
Water	72
R.m.s deviations	
Bond lengths (A)	0.009
Bond angles (°)	0.99
Ramachandran plot analysis	
Residues in favoured regions (%)	99.14
Residues in disallowed regions (%)	0.0
Molprobity analysis	
Molprobity score	1.48
Molprobity score percentile	99th
Resolution range (A)	1.99-2.49

Table 5.12 Refinement and validation statistics for the final model of the oligomerisation domain from CRES.

Initial rounds of refinement against the high resolution $C222_1$ dataset were performed by cycling between refinement in REFMAC and rebuilding in COOT [Emsley, Lohkamp *et al.* 2010]; resulting in a model with an R/R_{free} of 0.23/0.27. Subsequent iterative rounds of rebuilding in COOT and refinement in AUTOBUSTER [Bricogne, Blanc *et al.* 2011], using the autoncs_noprune flag to enforce strict NCS, resulted in a final model with an R/R_{free} of 0.2202/0.2357 (Figure 5.17, panel a). MOLPROBITY [Chen, Arendall *et al.* 2010] was used to validate the final model and its statistics are presented alongside those from refinement in Table 5.12. Residues before Asp403 and after Gln446 could not be built in most chains due to the lack of electron density. However, those chains with a C-terminal helix that is involved in forming crystal contacts appear to be stabilised with clear electron density allowing model building of this region as viewed in Figure 5.17, panel a, which shows a representation of the fully refined structure.

Prior to deposition in the Protein Data Bank, the residues were renumbered to ensure that they match those in the full length CRES molecule (UNIPROT accession E1C2T5). The finalised coordinates were deposited in the PDB with accession identifier 2YF2.

5.6.6 Refinement of $CRES_{SeMet}$

	F222	I222
Resolution (A)	2.69-23.96 (2.69-2.90)	2.59-24.28 (2.59-2.77)
Number of reflections	14727 (3003)	16451 (2305)
Number of reflections for R_{free}	991 (215)	1016 (142)
R_{work}/R_{free}	0.2448/0.2619 (0.2571/0.2739)	0.2573/0.2682 (0.2477/0.2815)
Number of atoms		
Protein	2610	2584
Ligand/Ion	0	0
Water	0	0
R.m.s deviations		
Bond lengths (A)	0.010	0.009
Bond angles (°)	1.10	1.09
Ramachandran plot analysis		
Residues in favoured regions (%)	99.37	99.68
Residues in disallowed regions (%)	0.0	0.0
Molprobity analysis		
Molprobity score	1.01	1.49
Molprobity score percentile	100th	100th
Resolution range (A)	2.44-2.94	2.34-2.84

Table 5.13 Refinement and validation statistics for the final model of the oligomerisation domain from $CRES_{SeMet}$ in both the I222 and F222 crystal forms.

The refined structure of CRES was molecular replaced into the F222 and I222 $CRES_{SeMet}$ datasets using PHASER. COOT was used to tweak the structures, and convert the methionine residues into selenomethionine, prior to refinement in AUTOBUSTER using the autoncs_noprune option to enforce usage of NCS over the entire chain length of each monomer. Final models were generated by cycling between refinement in AUTOBUSTER and rebuilding in COOT giving final R/R_{free} values of 0.2448/0.2619 and 0.2573/0.2682 for the F222 and I222 crystal forms respectively. The models were validated using the MOLPROBITY server and the results are presented alongside refinement statistics in Table 5.13. Representations of the structure from both crystal forms are shown in Figure 5.17. A flow diagram charting the entire course of the structure solution is also presented in Figure 5.18.

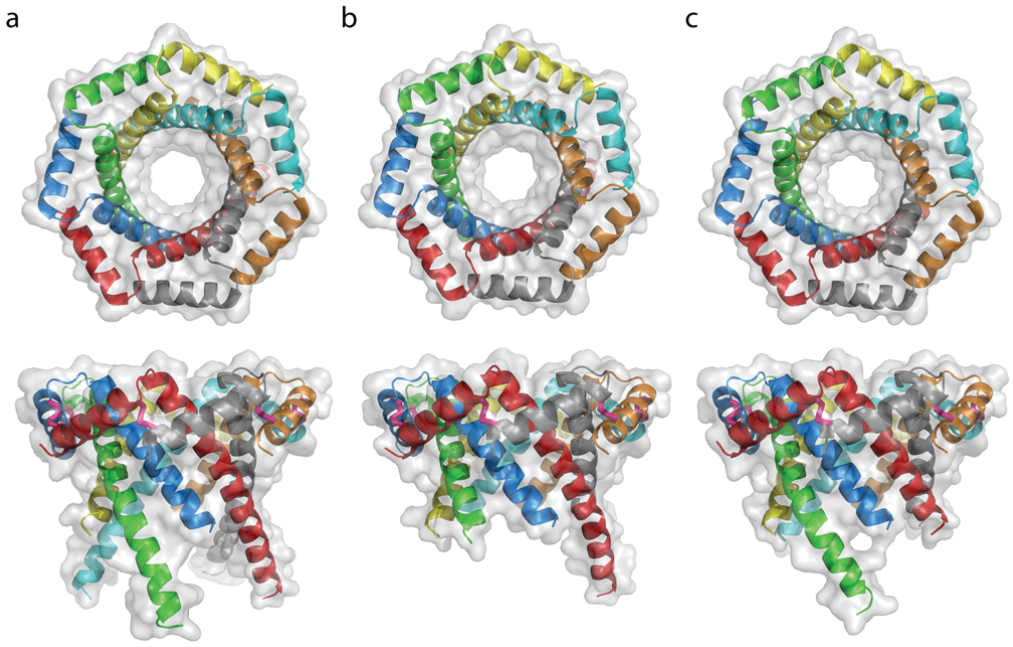


Figure 5.17 *Panel a*: Final model of the CRES oligomerisation domain in the $C222_1$ crystal form. Shown as a cartoon of the secondary structure coloured by chain within a transparent surface rendering. *Panel b*: Final model in the I222 crystal form. *Panel c*: Final model in the F222 crystal form.

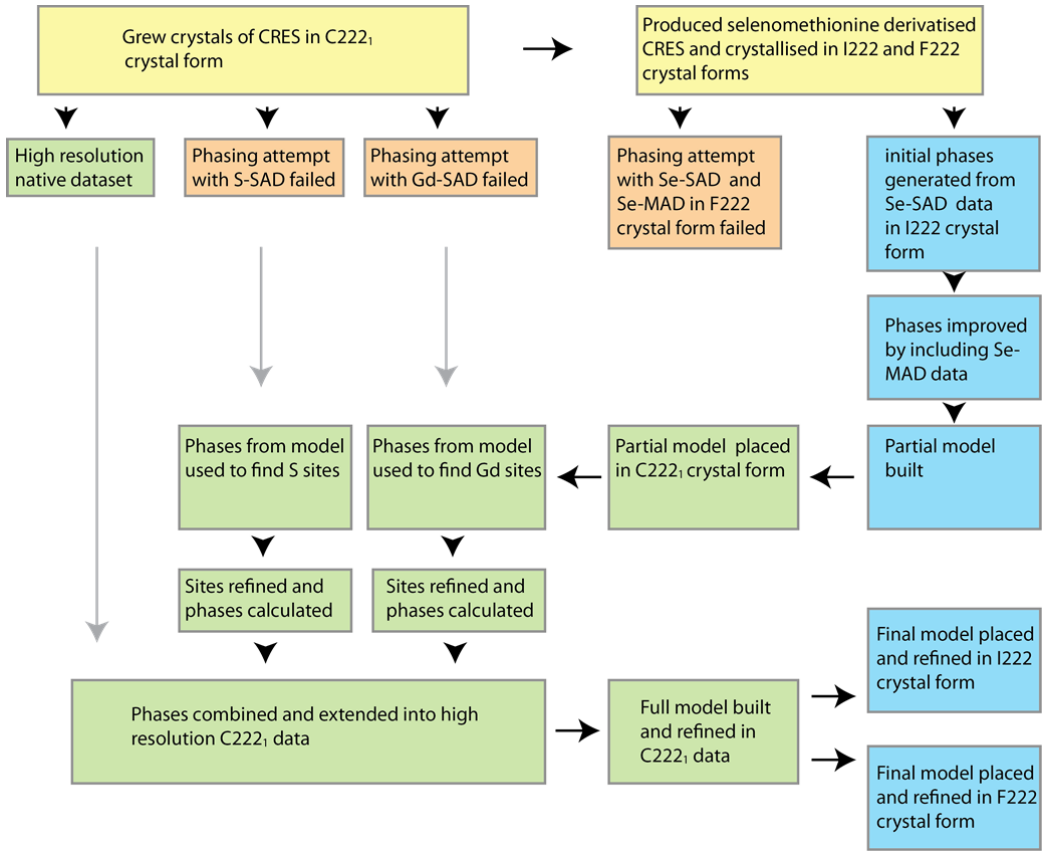


Figure 5.18 Flow diagram charting the structure solution of CRES. Crystallisations are shown in yellow and failed phasing attempts in orange. Steps using selenomethionine derivatised CRES are shown in blue whilst operations involving the native protein are in green.

5.7 Structural analyses of CRES and $CRES_{SeMet}$

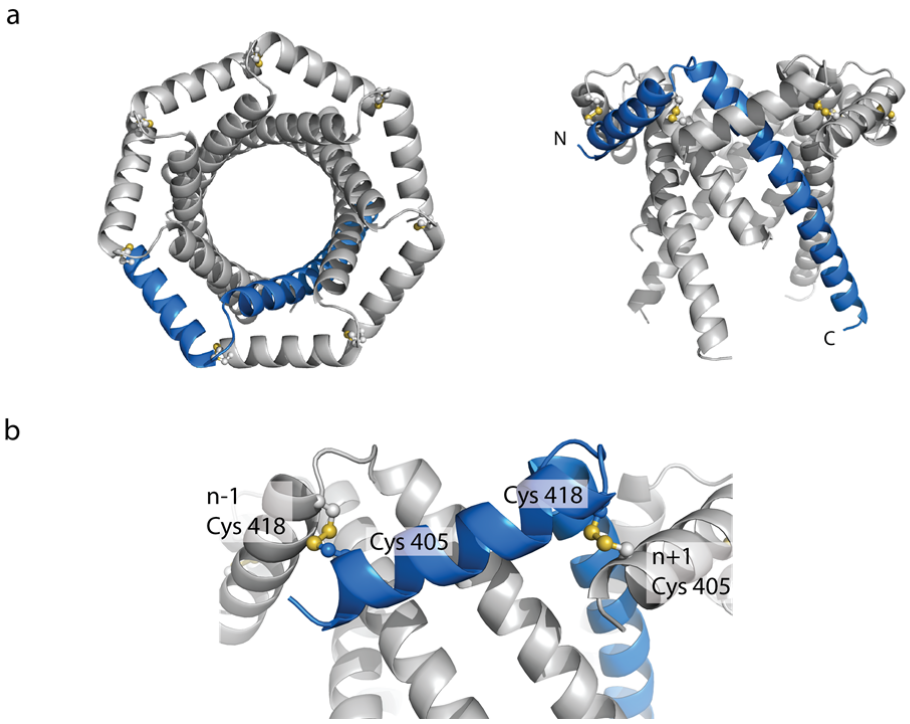


Figure 5.19 *Panel a*: Cartoon representation of the oligomerisation domain from CRES highlighting a single chain demonstrating the helix-hinge-helix structure. The C-terminal helix forms the heptameric coiled-coil where as the cysteine residues as either end of the N-terminal helix form intermolecular disulphide bonds to the neighbouring chains (shown in gold). *Panel b*: Zoom of the pattern of disulphide bond formation between a chain and its n+1 and n-1 neighbours.

As expected from the MALS characterisation, the crystal structure of CRES shows formation of a homoheptameric species with each CRES monomer consisting of two alpha helices joined by a flexible linker in a helix-hinge-helix motif (Figure 5.19, panel a). Multimerisation is mediated through formation of a heptameric, parallel coiled-coil between the C-terminal helices whilst the N-terminal helices pack back against this central core. Both cysteine residues are located at opposing ends of the N-terminal helix and form intermolecular disulphide bonds with the n+1 and n-1 monomer (Figure 5.19, panel b), locking the assembly together in a manner resembling a napkin ring around a folded napkin. A full structural analysis, and discussion of possible immunological implications, is presented in Chapter 6 alongside a comparison of the structure of CRES

with C4bp and the only other homoheptameric coiled-coil structure in the Protein Data Bank.

Electron density is absent for most of the C-terminal helices beyond residue Gln406, suggesting that the chains are flexible beyond this point resulting in the breakdown of the coiled-coil. Only the chains forming crystal contacts are stabilised enough to be able to build the model up to the C-terminus. Similarly, the N-terminal residues before the first cysteine could not be built in most of the chains suggesting that the linker to the last SCR domain is flexible.

5.8 Discussion

The structure of the oligomerisation domain of CRES is now known. As predicted the monomer consists of two helices, the C-terminal of which mediates heptamerisation via formation of a coiled-coil, complementing the accuracy of prediction algorithms. However, what could not be predicted was the parallel nature of the coiled-coil and the disulphide bridging between the N-terminal helices, forming a ring around the central assembly, requiring the atomic resolution structure to be determined. This structure describes the first heptameric coiled-coil found in nature and the tight geometry of its formation, coupled with that of the disulphide bond formation, suggests that evolution of a heptamer versus a hexamer, or octamer, is physiologically important to the function of CRES; an argument that is discussed further, and in context of C4bp, in Chapter 6.

Although the structure of the oligomerisation domain does not directly explain its adjuvant-like properties, it provides a framework for further investigations into a possible mechanism. Structure guided mutations can be made that do not affect assembly of the coiled-coil but may affect the immune response to a fused antigen. This could lead to a final determination of the presence of a T-cell epitope, or another

immunological feature. Due to the reduced order of the C-terminal helix, it may also be possible to truncate the C-terminus

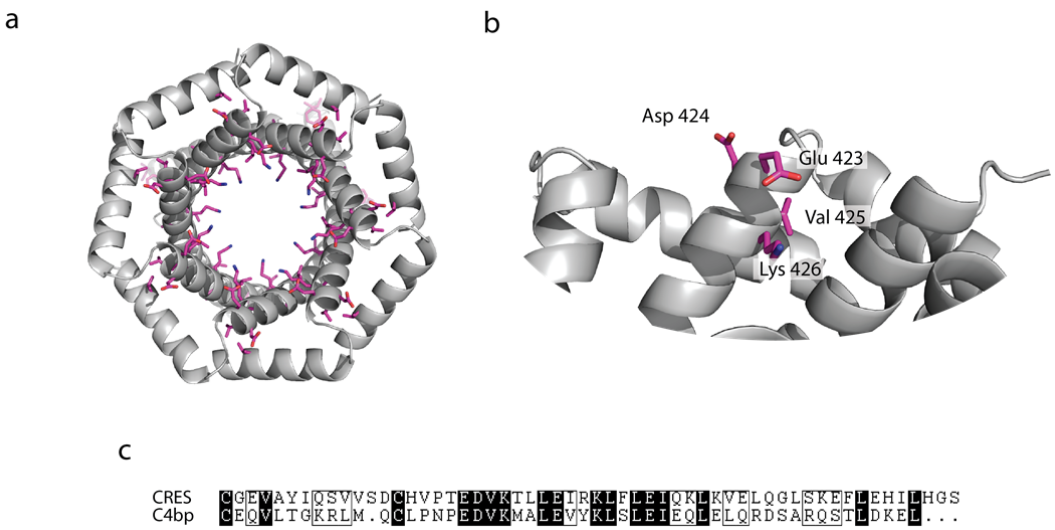


Figure 5.20 *Panel a*: Side-chains of residues conserved between CRES and C4bp are shown in pink. *Panel b*: Most conserved residues are buried except those around the hinge region. *Panel c*: Alignment of the CRES and C4bp oligomerisation domains starting from the first cysteine.

In light of the structure is may also be possible to modify the oligomerisation domain in order to increase its efficacy as a vaccine. There are regions of sequence identity between the homologous CRES and C4bp oligomerisation domains (Figure 5.20, panels a and c) which could result in generation of cross reactive antibodies upon vaccination leading to a reduced response against the boost phase of a prime-boost vaccination strategy. Analysis of the structure shows that most of these similarities lie buried within the assembly and are thus unlikely to elicit a response. There is a region of similarity in the hinge region between the helices which could be mutated to a less homologous sequence (Figure 5.20, panel b). The disorder displayed in the C-terminal helix suggests that the construct could be truncated to residue Gln446 and still heptamerise, giving a more compact assembly that may increase the viability of peptide synthesis of a vaccine. A similar approach may be applied to the N-terminus where there appears to be

little structure before the first cysteine residue, suggesting that it may also be truncated to Asp403 or perhaps even to the first cysteine.

The linkers between the oligomerisation domain assembly and a fused antigen, either N or C terminal, should also be investigated. Using the truncated form previously discussed, linkers can be designed using the structure geometry to pack the antigens tightly, or loosely, around the heptamer and assess whether the immune response is significantly altered, perhaps leading to an optimised construct. Crystallisation trials are being performed on the oligomerisation domain of CRES with a variety of antigens fused to either end of the molecule in an attempt to visualise how they are displayed.

5.9 Summary

The atomic resolution structure of the oligomerisaton domain from CRES, a homologue of C4bp from *Gallus gallus*, has been determined using X-ray crystallography which required experimental phasing using anomalous dispersion methods. In the crystal structure the oliogomerisation domains form a disulphide-linked heptameric species mediated by formation of a homoheptameric coiled-coil and is the first example of this assembly found in nature. It has also been shown that the oligomerisation domains form a heptameric species in solution.

5.9.1 Future directions

The adjuvant-like properties of the CRES oligomerisation domain cannot be fully explained by multimerisation. In order to investigates this further, structure guided mutagenesis should be used to resolve if there is an epitope present that provokes an increased immune response. Further research should also be performed in order to determine the optimal length of the chains required for assembly and the role of the linker sequences between the oligomerisation domain and the antigen.

Chapter Six: Macromolecular assembly of C4bp

C4b-binding protein (C4bp) is a major negative regulator of the C3 and C5 convertases common to the classical and lectin activation pathways, possessing both decay-accelerating activity against the convertases and displaying cofactor activity for factor I mediated inactivation of C4b [Scharfstein, Ferreira *et al.* 1978; Gigli, Fujita *et al.* 1979; Rawal, Rajagopalan *et al.* 2009]. C4bp is a disulphide-bonded macromolecular assembly of α and β chains with the most common isoform, approximately 70 - 75% of that found in plasma, regarded to be 7α1β (Figure 6.1, panel a) [Dahlback, Smith *et al.* 1983; Hillarp, Hessing *et al.* 1989]. However, the existence of other isoforms, 7α0β and 6α1β, has been reported [Hillarp, Hessing *et al.* 1989].

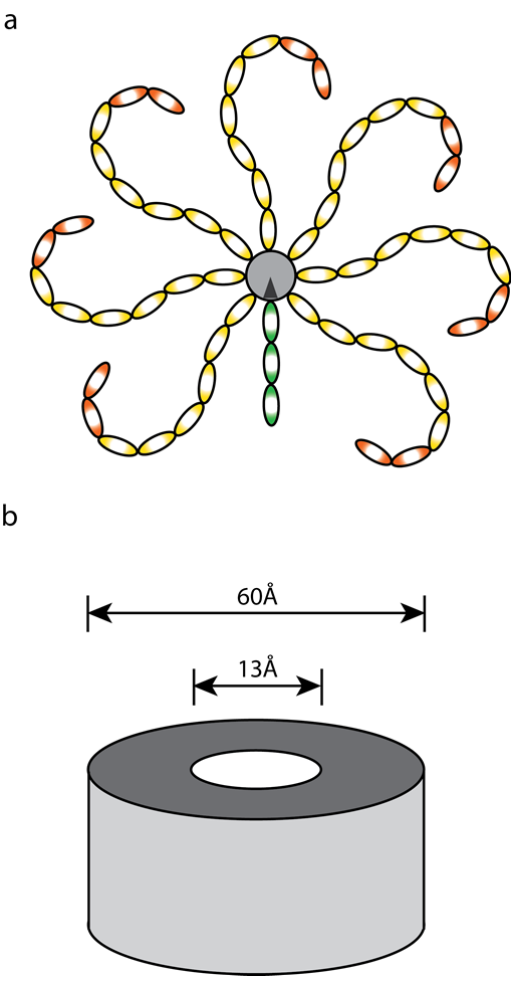


Figure 6.1 *Panel a*: Schematic diagram of the assembly of C4bp showing the most prevalent 7α1β isoform. Alpha chains are coloured in yellow with the two N-terminal domains, where regulatory activity resides, in orange The beta chain is shown in green and the central body assembled from the C-terminal oligomerisation domains in grey/black. *Panel b*: Approximate dimensions of the central body of C4bp as determined by electron microscopy.

The story is further complicated by the fact that the beta chain binds protein S, a cofactor in the anti-coagulation pathway [Sanchez-Corral, Criado Garcia *et al.* 1995]. Thus, by expression of more $C4bp_{\alpha}$, resulting in greater serum levels of 7α0β, the complement regulatory function of C4bp can be boosted without increased inactivation of Protein S by binding to $C4bp_{\beta}$, a condition linked with increased susceptibility to

thromboembolic disorders [Comp, Nixon *et al.* 1984]. This theory is supported by evidence that the presence of acute phase and inflammatory cytokines result in modulation of the levels of the different C4bp isoforms [Criado Garcia, Sanchez-Corral *et al.* 1995].

The C4bp α chain consists of eight SCR domains with a C-terminal oligomerisation domain that is necessary for assembly of the macromolecule, where as the β chain has three unique SCR domains followed by a different C-terminal oligomerisation domain that interacts with its counterpart in the α chain [Rodriguez de Cordoba, Sanchez-Corral *et al.* 1991; Hillarp, Pardo-Manuel *et al.* 1993]. Complement regulatory activity resides in the two N-terminal domains of the α chain, of which the NMR solution structure has been published [Blom, Webb *et al.* 1999; Blom, Zadura *et al.* 2000; Jenkins, Mark *et al.* 2006].

The mechanism by which the C4bp chains oligomerise via their C-terminal domains and the pattern of disulphide bond formation is currently unknown. However, upon chymotryptic cleavage, C4bp forms two major species; a 48 kDa fragment possessing complement regulatory properties corresponding to the SCR domains of the alpha chain, and a 160 kDa fragment composed of 25 kDa subunits corresponding to the oligomerisation domains with part of the eighth SCR domain still attached [Nagasawa, Mizuguchi *et al.* 1982]. Electron microscopy studies of the 160 kDa fragment show it to have a ring-like structure with inner and outer diameters of 13 Å and 60 Å respectively (Figure 6.1, panel b) [Dahlback and Mullereberhard 1984].

Presented here is the atomic resolution structure of the oligomerisation domain from the C4bp α chain, allowing insight into the macromolecular assembly of C4bp and describing how the alpha chains form a heptameric assembly in the 7α0β isoform.

Renewed investigation into the existence of multiple C4bp isoforms is performed using native mass spectrometry which could not detect any species other than 7α0β and 7α1β assemblies. Research into how the beta chain is incorporated into the 7α1β assembly is also presented, challenging the accepted model and suggesting that the beta chain is not disulphide linked to the alpha. A new model for the macromolecular assembly of C4bp is suggested and it is hoped that these realisations will lead to increased understanding of the processes of complement regulation by C4bp and its possible secondary role in regulation of the coagulation cascade.

6.1 Protein expression and purification

Oligomerisation domains from $C4bp_{\alpha}$ and $C4bp_{\beta}$ were expressed in a prokaryotic expression system, prior to *in vitro* formation of the intermolecular disulphide bonds. Attempts were made using multiple approaches to assemble the α and β oligomerisation domains into a physiological complex for further characterisation. Intact C4bp was also purified from human plasma by barium citrate precipitation for investigation into the stoichiometry of assembly.

6.1.1 Details of plasmid pIMX-77

Plasmid pIMX-77 was kindly provided by Dr. Fergal Hill (Imaxio SA, France) and contains the oligomerisation domain of $C4bp_{\alpha}$ expressed under the control of an inducible T7 promoter as a C-terminal fusion with a transcription enhancing downstream box sequence from *Escherichia coli* [Sprengart, Fuchs *et al.* 1996]. A schematic diagram of the plasmid is displayed in Figure 6.2.

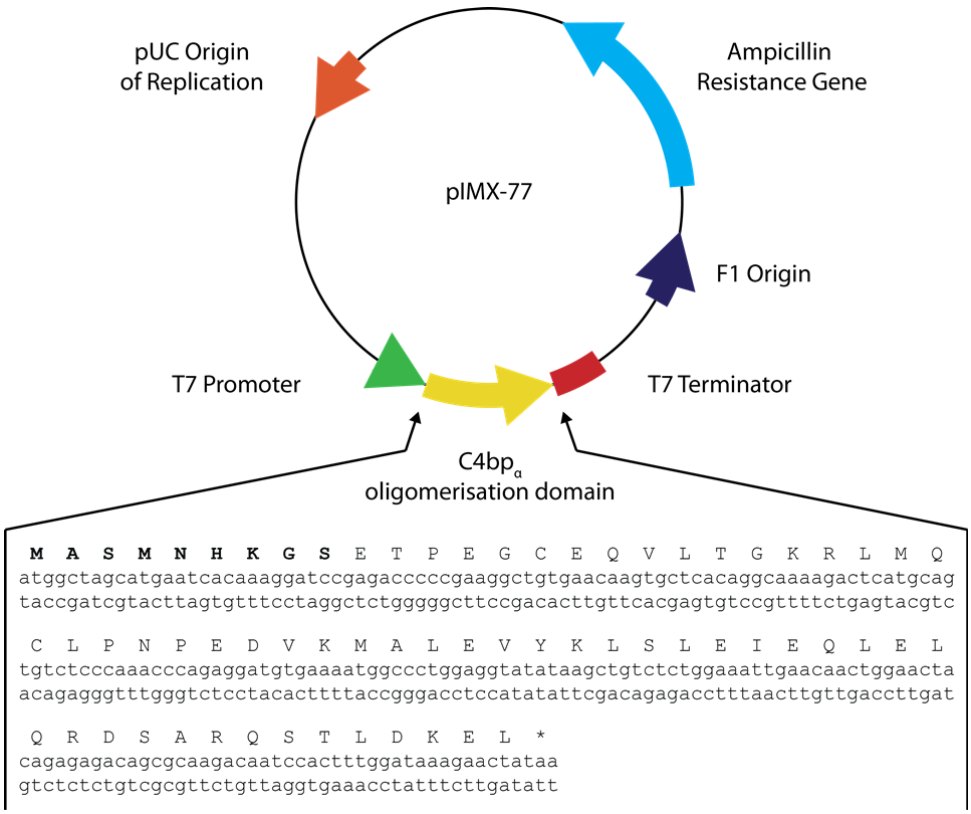


Figure 6.2 Schematic representation of pIMX-77. The oligomerisation domain from the α chain of C4bp is expressed without an affinity tag from a vector that was produced by cloning the C4bp fragment into a pRSET vector (Invitrogen, USA) between the XbaI and HindIII restriction sites. An N-terminal downstream box (shown in bold) has been added to enhance transcription and increase yields.

6.1.2 Generation of plasmid pET-C4bp$_\beta$

In order to generate a construct of the C4bp$_\beta$ oligomerisation domain that has a high yield, is easily purified and is suitable for crystallisation trials, it was decided to express the protein with an N-terminal hexahistidine tagged glutathione S-transferase (GST) affinity tag that can be cleaved using 3C protease. This construct will also be useful for attempting to form a complex between the α and β chains as it

C4bp_β_For
5′-TTCCCAGGGGCCCCTGATCCAGGAAGCGCCGAAAC-3′
C4bp_β_Rev
5′-GGTGGTGCTCGATTAGAGCAGCTTAGCTTTCAG-3′

Table 6.1 Oligonucleotide primers used for PCR amplification of the C4bp$_\beta$ oligomerisation domain insert.

will enable pull-down experiments with glutathione coated beads, and may also allow stoichiometry determination using MALS as incorporation of a beta chain will increase the mass of the complex by approximately 30 kDa giving an unambiguous result.

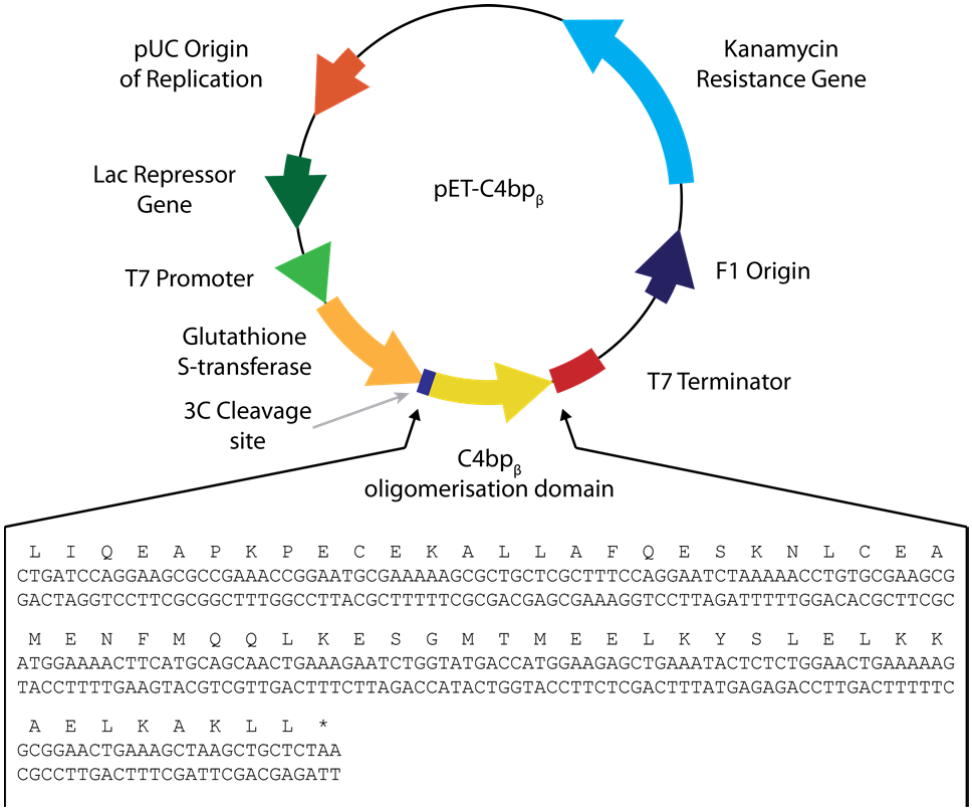


Figure 6.3 Schematic representation of pET-C4bp$_\beta$. The oligomerisation domain from the β chain of C4bp is expressed as a C-terminal GST fusion from a vector that was produced by cloning the C4bp fragment into a pET-M33 vector between the NcoI and XhoI restriction sites.

An insert encoding the sequence of the oligomerisation domain from C4bp$_\beta$ was PCR amplified from a vector containing the C4bp$_\beta$ gene, kindly provided by Dr. Fergal Hill (Imaxio SA, France), using the oligonucleotide primers described in Table 6.1 and KOD dna polymerase (Merck, USA) according to the manufacturers instructions. The insert was ligated into vector pET-M33 (EMBL, Heidelberg), pre-digested with NcoI and XhoI (Fermentas, USA) restriction endonucleases as per standard protocols, using an InFusion Dry-Down ligation system (Clontech, USA) as per the manufacturers instructions. *Escherichia coli* strain Novablue (Novagen, USA) was transformed with

the ligation reaction mixture and grown on LB/agar plates containing 30 µg/ml kanamycin. Positive colonies were selected using colony PCR (using GoTaq polymerase (Promega, USA)), as per standard protocols, and the sequence verified by Sanger sequencing using T7 forward and reverse primers. A map detailing the final vector, pET-C4bp$_\beta$, is shown in Figure 6.3

6.1.3 Expression and purification of C4bp$_\alpha$ oligomerisation domain

C4bp$_\alpha$ oligomerisation domain was purified in the absence of an affinity tag using a protocol based on one used by Dr. Fergal Hill (Imaxio SA, France). Both anion and cation exchange were used to purify the sample prior to *in vitro* formation of the disulphide bonds and a final size-exclusion chromatography purification step.

Escherichia coli strain C41(DE3) transformed with the pIMX-77 plasmid was grown overnight at 37^{o}C with 180 RPM shaking in 10 ml Luria-Bertani broth containing 100 µg/ml ampicillin. The saturated culture was used to inoculate 2 × 1L Luria-Bertani broth containing 100 µg/ml ampicillin prior to incubation at 37^{o}C with 180 RPM shaking. Cell proliferation was monitored using absorbance at 600nm (OD_{600}) and expression of the target protein was induced in mid-log phase (OD_{600} = 0.6) by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by continued incubation at 37^{o}C with 180 RPM shaking for three hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4^{o}C) before the pellets were combined and stored at -20^{o}C until required.

Defrosted cell pellets were resuspended in 30 ml of 20 mM bis-Tris-propane, 10 mM NaCl, 10 mM dithiothreitol (DTT), 1 mM ethylenediaminetetraacetic acid (EDTA), pH 7.0 and lysed by five passes through an Emulsiflex C5 homogeniser (Avestin, Canada). The crude lysate was centrifuged (48500 g, 20 minutes, 4^{o}C) prior to the supernatant

being loaded on a DEAE 16/10 anion exchange column (GE Healthcare, USA) pre-equilibrated with 20mM bis-Tris-propane, 10 mM DTT, 1 mM EDTA, 10 mM NaCl, pH 7.0. After washing with 150 ml of buffer, the column was eluted with a 10 – 500 mM gradient of NaCl over 150 ml (Figure 6.4, panel a). Fractions containing the target protein were identified using SDS-PAGE (Figure 6.4, panel d) and were acidified (pH 4.5) using acetic acid.

The acidified sample was loaded on an SP 26/5 cation exchange column (GE Healthcare, USA) pre-equilibrated with 20 mM sodium acetate, 10 mM NaCl, 1 mM EDTA, pH 4.5 and unbound material washed out with 100 ml buffer. The column was eluted using a 10 – 1000 mM gradient of NaCl over 200 ml and the fractions containing the target protein identified using SDS-PAGE (Figure 6.4, panels b and d).

In a significant deviation from the supplied protocol, fractions containing the C4bp$_\alpha$ oligomerization domain were dialysed against 50 mM Tris, 4 mM cystine, 2 mM cysteine, pH 9.0 for 72 hours using 3000 MWCO Snakeskin dialysis tubing (Pierce, USA) prior to a final purification step using size exclusion chromatography on a Superdex S75 26/60 column (GE Healthcare, USA) pre-equilibrated using 50 mM Tris, 150 mM NaCl, pH 7.5 (Figure 6.4, panel c). The presence of a redox pairing between cysteine/cystine encourages formation of the disulphide bonds, resulting in a greater percentage of homogenous material.

The purity of the protein was assessed by SDS-PAGE showing a doublet (Figure 6.4, panel d). However, MALS was employed to confirm the oligomerisation state in solution, returning a consistent mass of 51.3 kDa which corresponds very well to the expected mass of a heptameric assembly of the C4bp$_\alpha$ oligomerisation domain, 51.5 kDa (7 × 7630 Da) under these conditions (Figure 6.4, panel e). Typical yields were

approximately 20 mg/l of bacterial culture, assuming ε_{280} = 1615 M^{-1} cm^{-1} (calculated using ProtParam [Gasteiger E. 2005]).

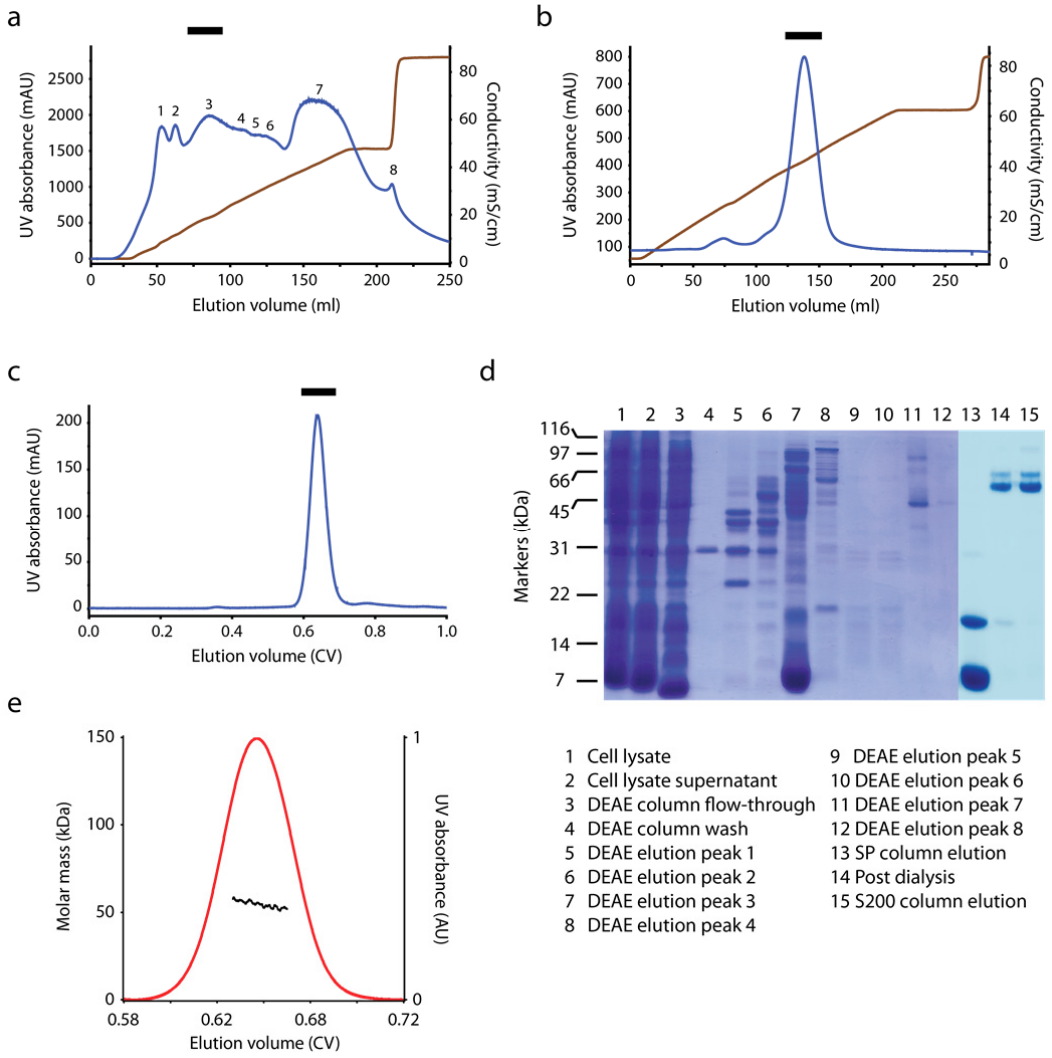


Figure 6.4 *Panel a*: Chromatogram charting the elution of the C4bp$_\alpha$ oligomerisation domain from a DEAE 16/10 ion exchange column (GE Healthcare, USA). The UV absorbance profile is shown in blue and the conductivity trace in brown. Fractions collected containing the target protein are highlighted with a black bar. *Panel b*: Elution profile from an SP 26/5 column (GE Healthcare, USA) showing the collected fractions. *Panel c*:Elution profile from an S75 26/60 column (GE Healthcare, USA) showing the collected fractions. *Panel d*: 15% (w/v) non-reducing SDS-PAGE gel charting the course of the purificiation. *Panel e*: MALS analysis of the purified material. The UV absorbance is shown in red with the mass distribution across the peak shown in black.

6.1.4 Expression and purification of C4bp$_\beta$ oligomerisation domain

Escherichia coli strain C41(DE3) transformed with the pET-C4bp$_\beta$ plasmid was grown overnight at 37^{o}C with 180 RPM shaking in 10 ml Luria-Bertani broth containing 30

μg/ml kanamycin. The saturated culture was used to inoculate 2 × 1L Luria-Bertani broth containing 30 μg/ml kanamycin prior to incubation at 37^{o}C with 180 RPM shaking. Cell proliferation was monitored using absorbance at 600 nm and expression of C4bp$_{\beta}$ was induced in mid-log phase (OD_{600} = 0.6) by addition of 1 mM IPTG followed by continued incubation at 21^{o}C, with 180 RPM shaking, overnight. Cells were harvested by centrifugation (4000 g, 20 minutes, 4^{o}C) before the pellets were combined and stored at -20^{o}C until required.

Defrosted cells were resuspended in 40 ml of 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.5 and one EDTA-free protease inhibitor cocktail tablet (Roche, Germany) added. Cells were lysed by five passes through an Emulsiflex C5 homogeniser (Avestin, Canada) and the lysate cleared by centrifugation (48000 g, 20 minutes, 4^{o}C). Filtered supernatant was loaded on a 5 ml GSTrap FF column (GE Heathcare, USA) pre-equilibrated with 50 ml of 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.5. The loaded GSTrap was washed with 100 ml of 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 8.0 to remove non-specifically bound contaminants prior to elution with 20 ml of 50 mM Tris, 150 mM NaCl, 10 mM reduced glutathione, 1 mM EDTA, pH 7.5. Size exclusion chromatography was employed to further purify the fusion protein using an S200 16/60 column (GE Healthcare, USA) pre-equilibrated using 50 mM Tris, 150 mM NaCl, pH 7.5 (Figure 6.5, panel a).

The amount of the C4bp$_{\beta}$:GST fusion was estimated assuming ε_{280} = 44350 M^{-1} cm^{-1} (calculated using ProtParam) and 1 unit of hexahistidine tagged 3C-HRV protease (Millipore, USA) was added per 200 μg protein. Following incubation at 4^{o}C for 90 hours, the digestion reaction was flowed over a 5 ml Ni-NTA column (Qiagen, USA), pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 7.5, in order to remove the hexahistidine tagged protease and GST. The flow through containing C4bp$_{\beta}$ was

collected during loading and washing with 40 ml of 50 mM Tris, 150 mM NaCl, pH 7.5 prior to a final purification step using an S75 16/60 size exclusion chromatography column (GE Healthcare, USA) pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 7.5 (Figure 6.5, panel b).

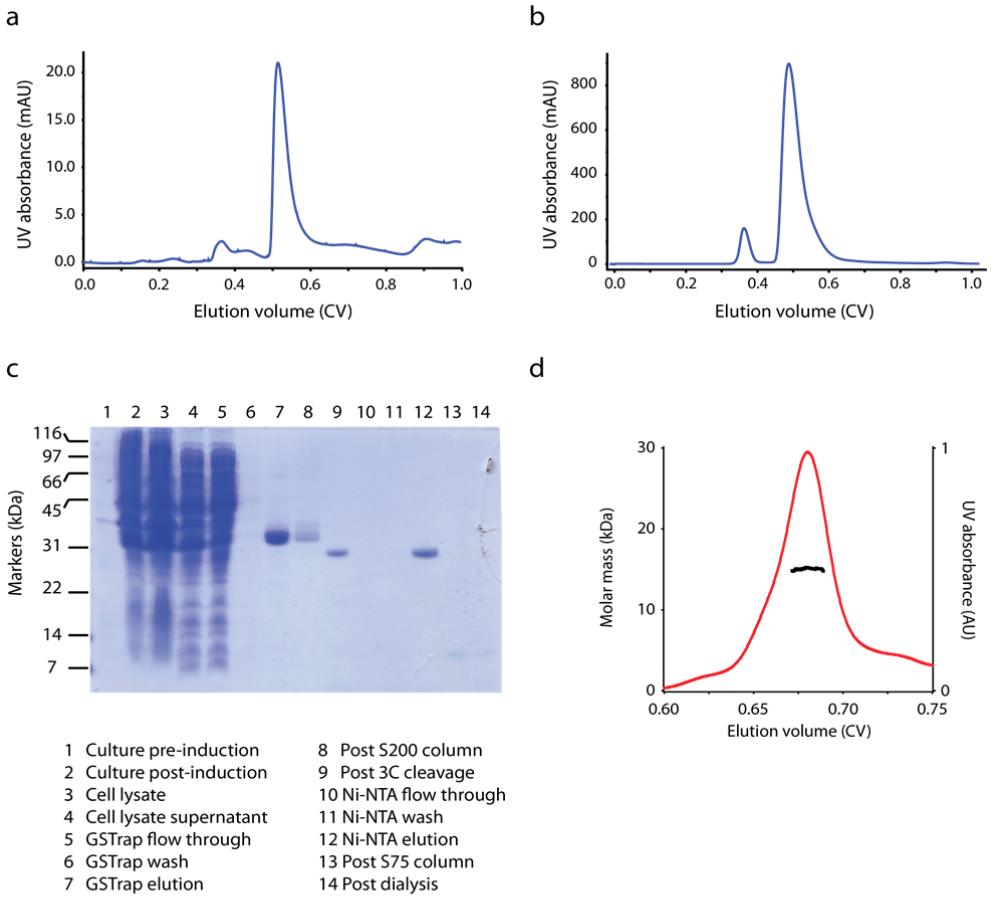


Figure 6.5 *Panel a*: Elution profile of the C4bp$_\beta$:GST fusion from an S200 16/60 size exclusion column (GE Healthcare, USA). The fractions collected containing the target protein are highlighted with a black bar. *Panel b*: Elution profile of C4bp$_\beta$ following cleavage of the GST affinity tag from an S75 16/60 size exclusion column (GE Healthcare, USA) with the collected fractions shown. *Panel c*: 15% (w/v) non-reducing SDS-PAGE gel charting the progress of the purification. *Panel d*: MALS analysis of the purified C4bp$_\beta$ oligomerisation domain.

SDS-PAGE analysis over the course of the preparation and of the material eluted from the column suggests that the latter contains a mixture of monomeric and disulphide-bonded dimeric species (Figure 6.5, panel c). However, MALS analysis shows that all

of the C4bp$_\beta$ has an average mass of 14 kDa under these conditions (Figure 6.5, panel d) which is consistent with formation of a dimeric species. In order to generate 100% disulphide bonded dimers (as assessed by SDS-PAGE) the protein was dialysed against 50 mM Tris, 4 mM cystine, 2 mM cysteine, pH 9.0 for 72 hours (Figure 6.5, panel c). ES-TOF analysis of the final material returned a mass of 14148 ± 4 Da which is consistent with a disulphide bonded dimer of the C4bp$_\beta$ oligomerisation domain (2 × 7073 Da). Typical yields from this protocol are approximately 2 mg/l of bacterial culture assuming that the ε_{280} of the cleaved material is 1400 M^{-1} cm^{-1} (calculated using ProtParam).

6.1.5 Purification of C4bp from plasma

C4bp was purified from human plasma using a hybrid protocol based on those previously published by [Nagasawa and Stroud 1980] and [Dahlback 1983]. Barium chloride is added to citrated plasma resulting in barium citrate precipitation which co-precipitates the C4bp:Protein S complex. This enables purification of material that contains the beta chain as the isoforms lacking this protein are unable to bind Protein S and therefore are not precipitated [Dahlback and Stenflo 1981].

2L of mixed pool citrated human plasma (TCS Biosciences, UK) were cleared by two rounds of centrifugation (4000 g, 4^{o}C, 20 minutes) before 160 ml of 1 M $BaCl_2$ was added dropwise with stirring at 4^{o}C. Following 1 hour of continued stirring, the precipitate was harvested by centrifugation (4000 g, 4^{o}C, 20 minutes) and the pellet washed twice to remove contaminants by resuspension in 200 ml of 50 mM Tris, 150 mM NaCl, pH 7.5 (in place of 0.9% NaCl described by [Nagasawa and Stroud 1980]) followed by centrifugation (4000 g, 4^{o}C, 20 minutes). The precipitate was then solubilised in 400 ml of 200 mM EDTA, pH 7.5 to which four protease inhibitor cocktail tablets (Roche, Germany) had been added in place of benzamidine

hydrochloride, di-isopropylphosphorofluoridate, phenylmethanesulphonylfluoride and soya-bean trypsin inhibitor described by [Dahlback 1983].

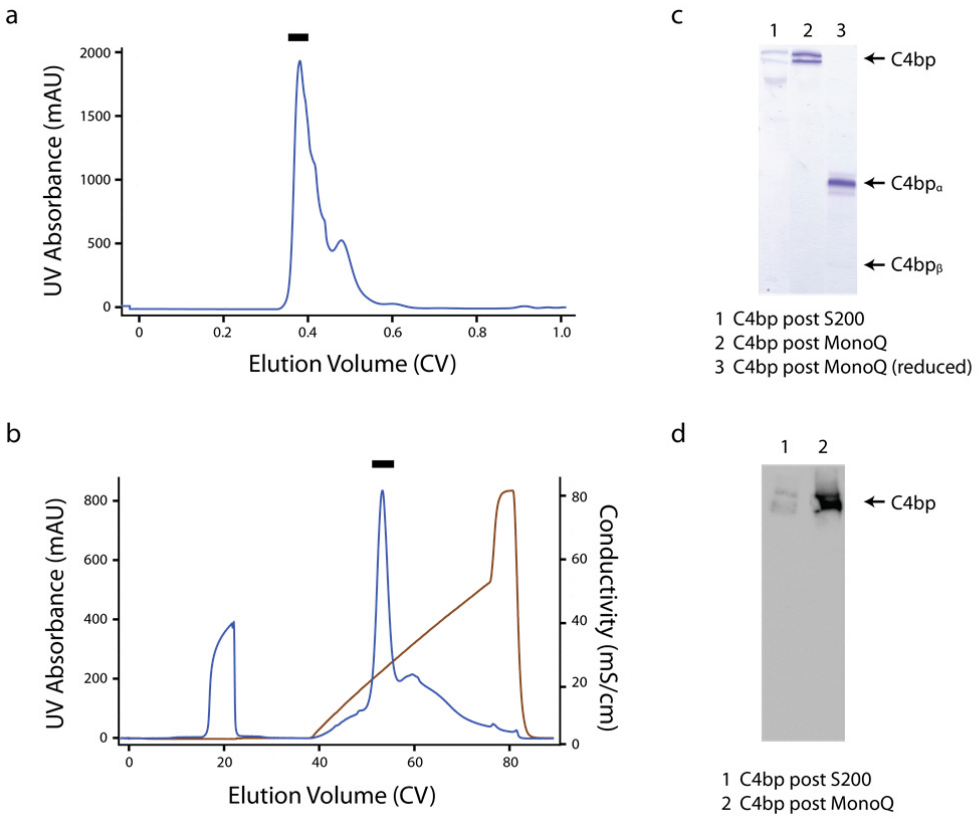


Figure 6.6 *Panel a*: Elution profile of C4bp purified by barium citrate precipitation on an S200 26/60 column (GE Healthcare, USA). Fractions collected are highlighted with a black bar. *Panel b*: Elution profile of C4bp from a MonoQ 5/50 ion exchange column (GE Healthcare, USA). The UV and conductivity profiles are shown in blue and brown respectively and the collected fractions are highlighted. *Panel c*: 4-20% (w/v) gradient non-reducing SDS-PAGE gel analysing the elutions from the S200 size exclusion and MonoQ ion exchange column. *Panel d*: Western blot of the eluted C4bp fractions detected with an α-C4bp$_\beta$ monoclonal antibody.

Due to advances in purification technologies since the original papers were published, further purification of C4bp was performed using a new protocol. [Nukatsuka and Nagasawa 1987] reported that Protein S could be dissociated from C4bp, without affecting the function of the complement regulator, using 3M guanidinium chloride. The solubilised sample was made 3 M with respect to guanidinium chloride and dialysed against 3 M guanidinium chloride, pH 7.5 overnight. The sample was then purified on

an S200 26/60 size exclusion column (GE Healthcare, USA) pre-equilibrated in 3 M guanidinium chloride, pH 7.5 (Figure 6.6, panel a). Due to the large size of C4bp (~520 kDa) it elutes in the void peak (elution volume approximately 0.4 CV). However, Protein S (~70 kDa. Elution volume ~ 0.48 CV) and other contaminants are resolved allowing them to be removed. Fractions containing C4bp were pooled and dialysed against 50 mM Tris, 10 mM NaCl, pH 8.5 overnight prior to loading on a MonoQ 5/50 anion exchange column (GE Healthcare, USA) pre-equilibrated in 50 mM Tris, 10 mM NaCl, pH 8.5. The column was eluted using a 10 – 1000 mM gradient of NaCl over 40 column volumes and the fractions containing C4bp collected as shown in Figure 6.6, panel b.

This purification protocol resulted in approximately 40% recovery of C4bp yielding around 80 mg of C4bp per litre of plasma (using the calculated A_{280} (1 mg/ml)= 1.41 [Garcia de Frutos, Alim *et al.* 1994]). The identity of C4bp was confirmed by SDS-PAGE and Western blotting with antibodies against the beta chain (Figure 6.6, panels c and d), a kind gift from Dr. Anna Blom (Lund University, Sweden).

6.2 Crystallisation, data collection and processing of the C4bp$_\alpha$ oligomerisation domain

In order to assess the structural similarity between the oligomerisation domains of C4bp$_\alpha$ and CRES (Chapter 5) attempts were made to obtain the X-ray crystal structures of the human proteins.

6.2.1 Crystallisation of the C4bp$_\alpha$ oligomerisation domain

C4bp$_\alpha$ oligomerisation domain was concentrated to A_{280} = 5.20 prior to crystallisation trials being performed at 294K using the sitting drop vapour diffusion in 400 nL sitting drops produced by an Oryx Nano crystallisation robot (Douglas Instruments, UK)

within MRC crystallisation plates (Molecular Dimensions, UK). Stura, Macrosol, PACT Premier, JCSG, Structure 1, Structure 2, Morpheus and Proplex screens (Molecular Dimensions, UK) were used, with two drops set up for each condition containing protein to precipitant ratios of 1:1 and 1:3 in one set of trays, and 3:1 and 9:1 in the second set.

Small, plate-like crystals grew after approximately two weeks in a condition containing 2 M ammonium sulphate, 0.1 M sodium acetate, pH 4.6 (Figure 6.7, panel a). Attempts were made to optimise these crystals by varying the protein concentration, crystallisation temperature, differing drop sizes to effect nucleation rate, and incremental adjustment of pH and precipitant concentration. Larger crystals grew in 0.8 ul drops at 21^{o}C from the original condition. These crystals were mounted in Cryoloops (Hampton Research, USA) prior to cryoprotection in 20% ethylene glycol and flash freezing in liquid nitrogen.

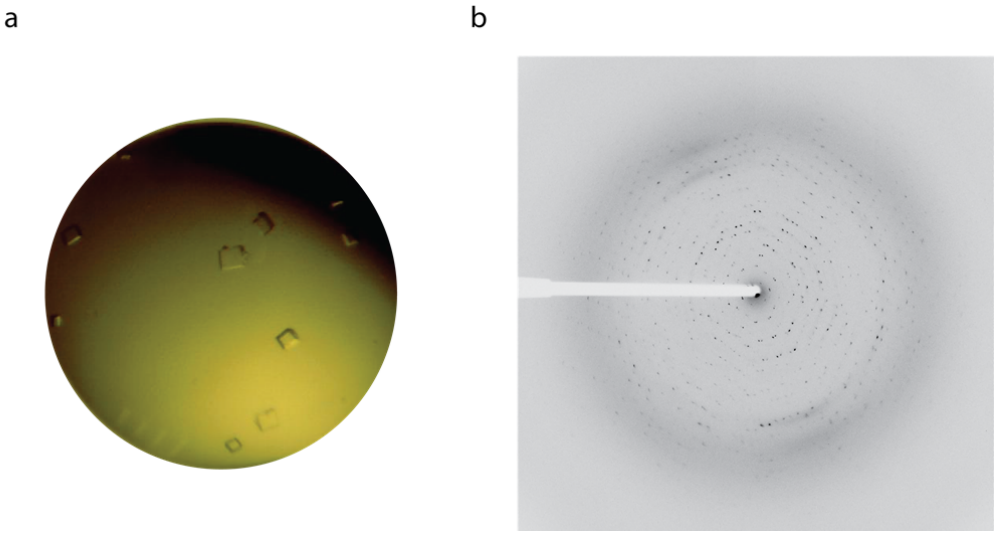


Figure 6.7 *Panel a*: Crystals of the C4bp$_{\alpha}$ oligomerisation domain. *Panel b*: Example diffraction image. Figure generated with DIFF2JPEG [Bailey 1994].

6.2.2 Data collection of the C4bp$_\alpha$ oligomerisation domain

Data were collected on beamline I04-1 at DIAMOND (Harwell Science and Innovation Campus, UK) using the rotation method (Figure 6.7, panel b). Diffracting crystals were indexed in a triclinic cell (cell dimensions a = 53.1 Å, b = 53.2 Å, c = 71.4 Å, α = 106.4°, β = 94.5°, γ = 92.5°) with no rotational symmetry (spacegroup P1) using IMOSFLM, which was then used to calculate the most efficient data collection strategy. Data collection parameters are presented in Table 6.2.

	Dataset 1
Diffraction source	DIAMOND I04-1
Detector	MAR Mosaic
Energy (eV)	13516
Wavelength (Å)	0.9173
Temperature (K)	120
Detector resolution (Å)	2.5
Exposure per image (s)	1
Φ_{start}	0
Φ_{end}	180
$\Delta\Phi$	1
Number Images	180

Table 6.2 Data collection parameters.

6.2.3 Data processing of the C4bp$_\alpha$ oligomerisation domain

	Dataset 1
Spacegroup	P1
Unit-cell dimensions (Å)	a=53.1, b=53.2, c=71.4
Unit-cell angles (°)	α=106.4, β=94.5, γ=92.5
Wilson B factor	54.9
Resolution (Å)	2.80-47.9 (2.80-2.95)
R_{merge}	0.085 (0.343)
R_{pim}	0.085 (0.343)
<I/σ(I)>	7.1 (2.1)
Completeness (%)	97.3 (96.5)
Multiplicity	2.0 (1.9)

Table 6.3 Data processing statistics.

Diffraction data were processed using XIA2 [Winter 2010] invoking the 3dii option to enforce integration with XDS [Kabsch 2010] and scaling with XSCALE [Kabsch 2010]. Analysis of the data using POINTLESS [Evans 2006] confirmed the spacegroup as P1. Data processing statistics are presented in Table 6.3.

6.3 Structure determination and refinement of the C4bp$_\alpha$ oligomerisation domain

Due to a 35% level of sequence identity between the oligomerisation domains of C4bp$_\alpha$ and CRES, molecular replacement was attempted with a model derived from the atomic resolution structure of the oligomerisation domain from CRES (Chapter 5).

6.3.1 Structure determination of the C4bp$_\alpha$ oligomerisation domain

Analysis of the content of the asymmetric unit and the self rotation function, using MATTHEWS [Matthews 1968] and POLARRFN [Kabsch 1976] respectively, suggests the presence of two copies of the heptameric C4bp$_\alpha$ oligomerisation domain (solvent content = 37.3). Both heptamers are related by a two-fold rotational NCS, the axis of which is not quite parallel to c, resulting in the series of peaks observed in the chi = 180^o section relating the monomers between each assembly (shown in red in Figure 6.8).

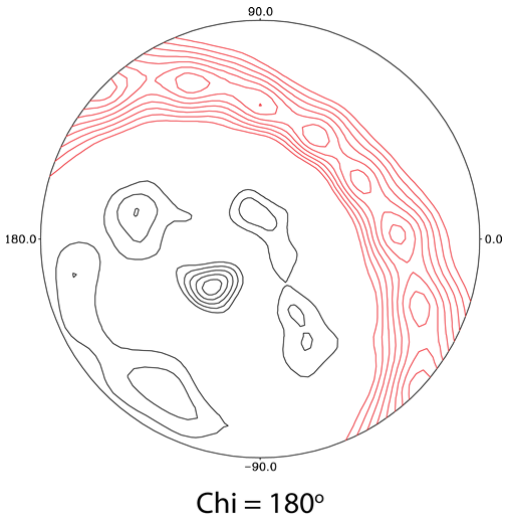


Figure 6.8 Self-rotation function generated using POLARRFN [Kabsch 1976].

Initial attempts to solve the structure using molecular replacement with the oligomerisation domain from CRES failed due to a high number of clashes during the packing test in PHASER [McCoy, Grosse-Kunstleve *et al.* 2007]. A revised model of CRES was generated by pruning the N-termini up to the first cysteine (residue 405) and the C-termini from the breakdown of the coiled coil (residue 446). The model was further modified by trimming the side chains of the residues in CRES

back to their common atom with those in C4bp using CHAINSAW [Stein 2008]. This also removed Ser416 as there is an extra residue between the cysteine residues in the N-terminal helix of CRES compared to C4bp (Figure 6.13). Molecular replacement using PHASER with this model was successful, placing two copies in the asymmetric unit with the scores reported in Table 6.4. Examination of the contents of the unit cell confirmed that the two heptamers are related by a two-fold rotation that is not parallel to the principle axis (Figure 6.9, panel b).

	RFZ	TFZ	LLG
Copy 1	9.0	100.0	67
Copy 2	7.0	13.0	223

Table 6.4 Rotation function (RFZ) Z-scores, translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of two copies of the model in the data

The resulting map had good density for the C-terminal helices in the coiled-coil but the density around the N-terminal helix was poor (Figure 6.9, panel a). Due to the reduced number of residues between the cysteines in the N-terminal helix, its orientation could be altered in C4bp compared to CRES. In order to assess this possibility, the residues comprising the hinge between the helices were deleted and the resulting model subjected to a single cycle of rigid body refinement in AUTOBUSTER [Blanc, Roversi *et al.* 2004; Bricogne, Blanc *et al.* 2011], defining the two helices as separate domains. The N-terminal helix was in much stronger density in the resulting map (Figure 6.9, panel c), suggesting its correct placement. However, the density for the deleted hinge residues was still poor.

In order to generate the most detailed map possible for building these missing residues, NCS averaging was employed to average the density over the C4bp oligomerisation domain monomer from all 14 copies (2 × heptamers) in the asymmetric unit. DM [Cowtan 1999] was employed to generate the averaged density from all 14 chains within a solvent mask generated around chain A at a 3 Å radius using NCSMASK [Winn, Ballard *et al.* 2011]. The translation and rotation operators from chain A to each

of the other chains, in orthogonal and Eulerian coordinates respectively, were calculated using SUPERPOSE [Krissinel and Henrick 2004] and the transformations defined as a single domain in DM. The average correlation between the mask and all 14 transformed copies was 0.86 ± 0.02, suggesting a high similarity leading to successful averaging. This proved correct as the averaged map displays much more detailed and connected density (Figure 6.9, panel d).

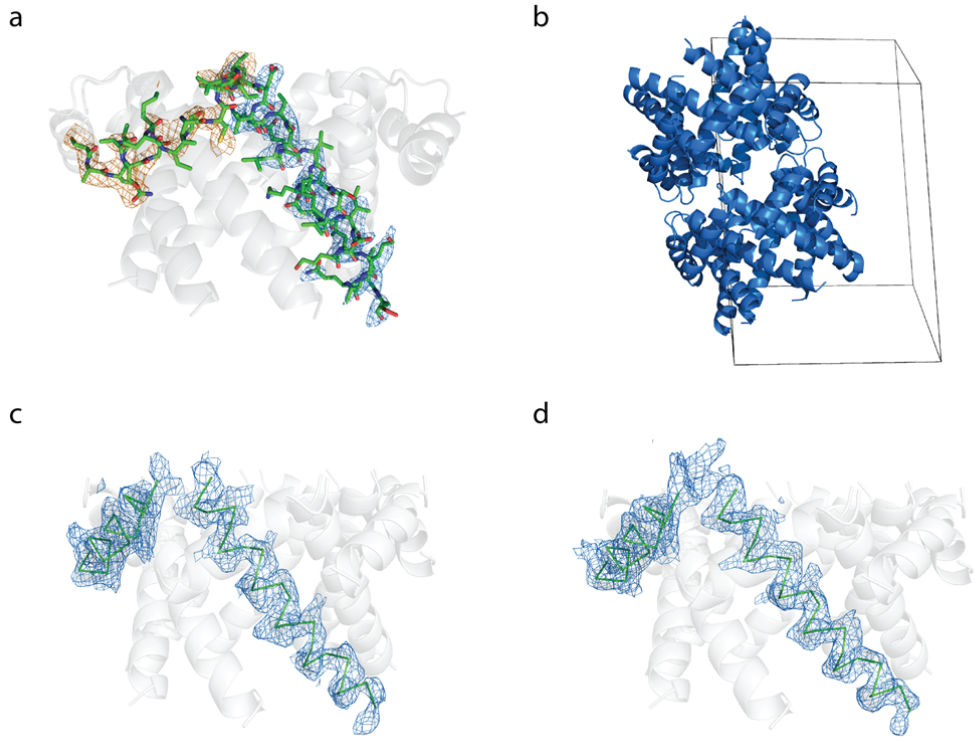


Figure 6.9 *Panel a*: Initial Fo-Fc map shown around a single chain of the molecular replacement model for clarity. Density (contoured at 1.2 σ) is shown in orange around the N-terminal helix and blue around the C-terminal helix. *Panel b*: Cartoon representation of both copies of the heptameric assembly in the asymmetric unit. *Panel c*: Electron density around a single chain following rigid body refinement. *Panel d*: NCS averaged electron density within the solvent mask around chain a.

BUCCANEER [Cowtan 2006] was employed to autobuild in the NCS averaged map, with the initial model to place the resulting in 27 chains, the longest of which was 77 residues long, and an R/R_{free} of 0.3082/0.4476. A complete model of the monomer was created by combining two N and C-terminal helices that had been fully built, but as separate chains. The hinge was built in by hand using COOT. This full monomer model

was superimposed over every chain in the rigid body refined model resulting in a complete model of both heptamers in the asymmetric unit.

6.3.1 Refinement of the C4bp$_\alpha$ oligomerisation domain

The C4bp$_\alpha$ oligomerisation domain was refined iteratively between rounds of refinement in AUTOBUSTER and rebuilding in COOT. Restraints from the non-crystallographic symmetry between the two heptameric assemblies in the asymmetric unit were invoked during AUTOBUSTER refinement using the autoncs flag [Smart, Brandl *et al.* 2008]. Throughout model building, MOLPROBITY [Chen, Arendall *et al.* 2010] was used to analyse the structure and validate the model. Final refinement and validation statistics are presented in Table 6.5. The fully refined structure is shown in Figure 6.10 and the chains have been numbered in accordance with UNIPROT accession P04003 [Gerhard, Wagner *et al.* 2004]. Residues preceding Pro533 and following Ala597 could not be built due to lack of electron density.

	C4bp$_\alpha$ oligomerisation domain
Resolution (A)	2.8-43.8 (2.80-2.97)
Number of reflections	17826 (2855)
Number of reflections for R_{free}	915 (152)
R_{work}/R_{free}	0.2064/0.2575 (0.2249/0.3277)
Number of atoms	
Protein	4972
Ligand/Ion	166 Ethylene glycol
Water	166
R.m.s deviations	
Bond lengths (A)	0.01
Bond angles (^{o})	1.15
Ramachandran plot analysis	
Residues in favoured regions (%)	99.17
Residues in disallowed regions (%)	0.17
Molprobity analysis	
Molprobity score	1.85
Molprobity score percentile	99th
Resolution range (A)	2.55-3.05

Table 6.5 Refinement and validation statistics.

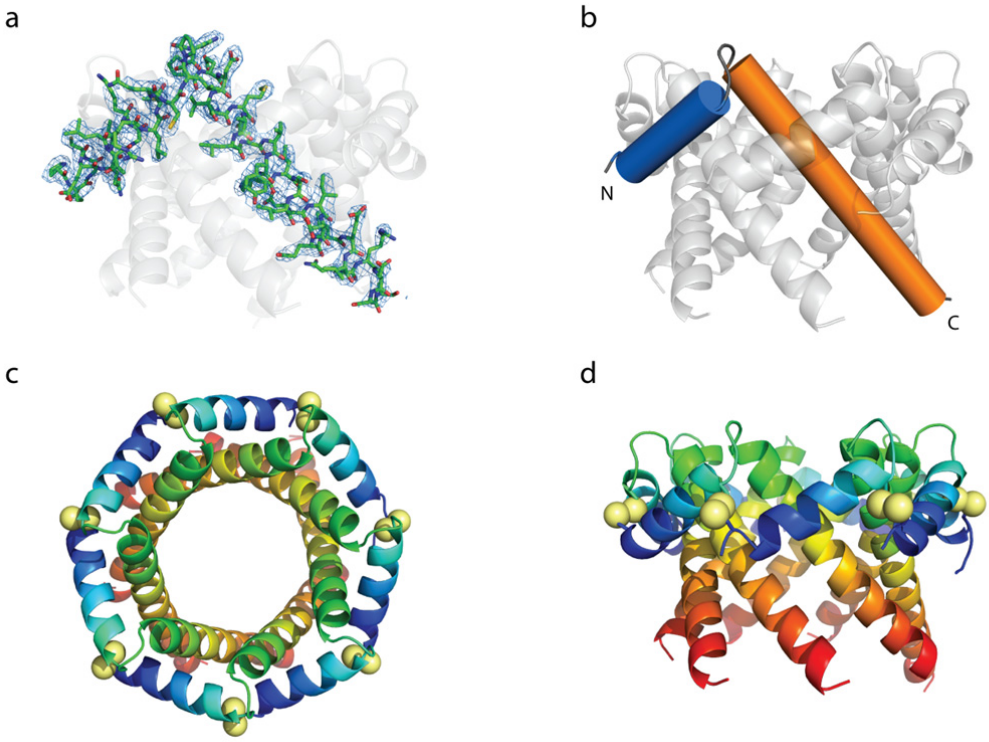


Figure 6.10 *Panel a*: Fully refined structure of the C4bp$_\alpha$ oligomerisation domain shown in cartoon representation with a single chain shown in stick format. Representative electron density (2Fo-Fc map contoured at 1.2 σ) is shown for this chain. *Panel b*: Representation of the helix-hinge-helix motif of each oligomerisation domain. The N-terminal helices (blue) form a collar around a heptameric coiled-coil made up of the C-terminal helices (orange). *Panel c*: Top-down view of the heptameric structure. Each chain is shown in cartoon format coloured from the N-terminus (blue) to the C-terminus (red). Sulphur atoms within cysteine residues forming intermolecular disulphide bonds are shown as gold spheres. *Panel d*: Side view of the assembly.

6.4 Crystallisation of the C4bp$_\beta$ oligomerisation domain

Crystallisation trials of the C4bp$_\beta$ oligomerisation domain were carried out from a protein stock with A_{280} = 0.45 in an identical manner to that described in Chapter 6.2.1. Large, single crystals grew after approximately 2 weeks in a condition containing 0.5 M ammonium sulphate, 10% (w/v) PEG 4000 (Figure 6.11, panel a). These initial crystals diffracted poorly, to approximately 8 Å resolution, and indexing was unreliable. Attempts to optimize the crystals were performed and increased the resolution to approximately 6 Å (Figure 6.11, panel b). However, the quality of the data was not sufficient to reliably perform structure determination.

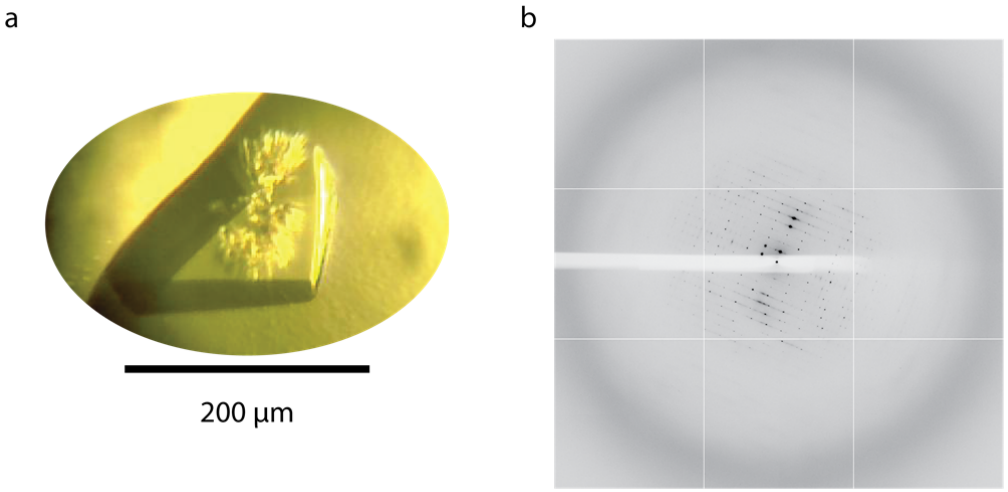


Figure 6.11 *Panel* a:Crystals of the C4bp$_{\beta}$ oligomerisation domain. *Panel b*: Example diffraction image.

6.5 Structural analyses

The atomic resolution structure of the oligomerisation domain from the C4bp α chain has been determined showing formation of heptameric species with 7-fold symmetry between monomers, each of which comprises of two alpha helices joined by a linker in a helix-hinge-helix arrangement (see Figure 6.10, panel b). Assembly into a mutimeric species occurs via formation of a parallel coiled-coil between C-terminal helices, whilst the N-terminal helices form a collar around the coiled-coil and are held in place by formation of intermolecular disulphide bonds between two cysteine residues at either end of the N-terminal helix. This gives the structure an overall shape that resembles a champagne cork (Figure 6.10, panels c and d). Analysis of the heptameric assembly shows the body to have an outer diameter of 50

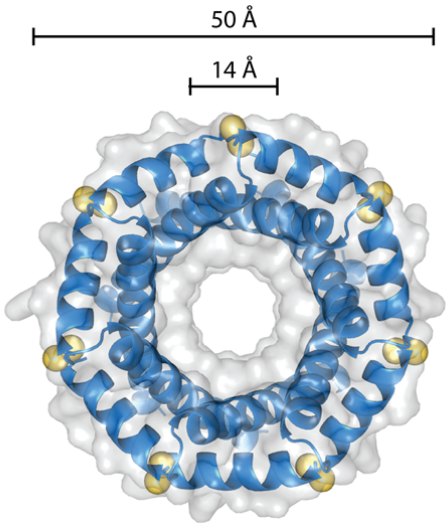


Figure 6.12 Dimensions of the heptameric assembly made up of oligomerisation domains from C4bp$_{\alpha}$.

Å and a diameter across the central channel of 14 Å (Figure 6.12), correlating well with those measured using electron microscopy, 60 Å and 13 Å respectively, published by Dahlback and Muller-Eberhard (1984. Figure 6.1, panel b). The material analysed in the latter study was generated by chymotryptic digestion of C4bp and possibly still has a portion of the eight SCR domain still attached accounting for its slightly larger outer radius compared to the new structure.

The disulphide-bonded heptameric assembly in C4bp is homologous to that found in the structure of the oligomerisation domain of CRES and are the first examples of homo-heptameric coiled-coils to be found in nature, forming left handed superhelices. Comparison of these structures (Figure 6.13) shows that the overall architecture is conserved with very high conservation of the coiled-coil, but the geometry of the collar formed by the N-terminal helices differs, being tighter in C4bp than in CRES. The effect of this is to pull the disulphides closer to the coiled coil in the former, making C4bp appear round from above, whilst CRES looks like a heptagon (Figure 6.13, panel a). Within CRES the latter portion of the C-terminal helix is labile, only being observed when stabilised by involvement in crystal contacts (Chapter 5.7). A similar observation is made in C4bp as no electron density was present for any residues following Ala597, and is described by the increase in RMSD between C4bp and CRES as the C-terminal helices become more mobile as one proceeds along the helix (Figure 6.13, panel c).

The only other structure of a homoheptameric coiled-coil was presented by [Liu, Zheng *et al.* 2006] and consisted of a mutated form of a leucine zipper protein (PDB entry: 2HY6). Comparison of all three structures yields clues to the specificity of heptameric coiled-coil formation and describes the specific pattern of interactions required for their formation.

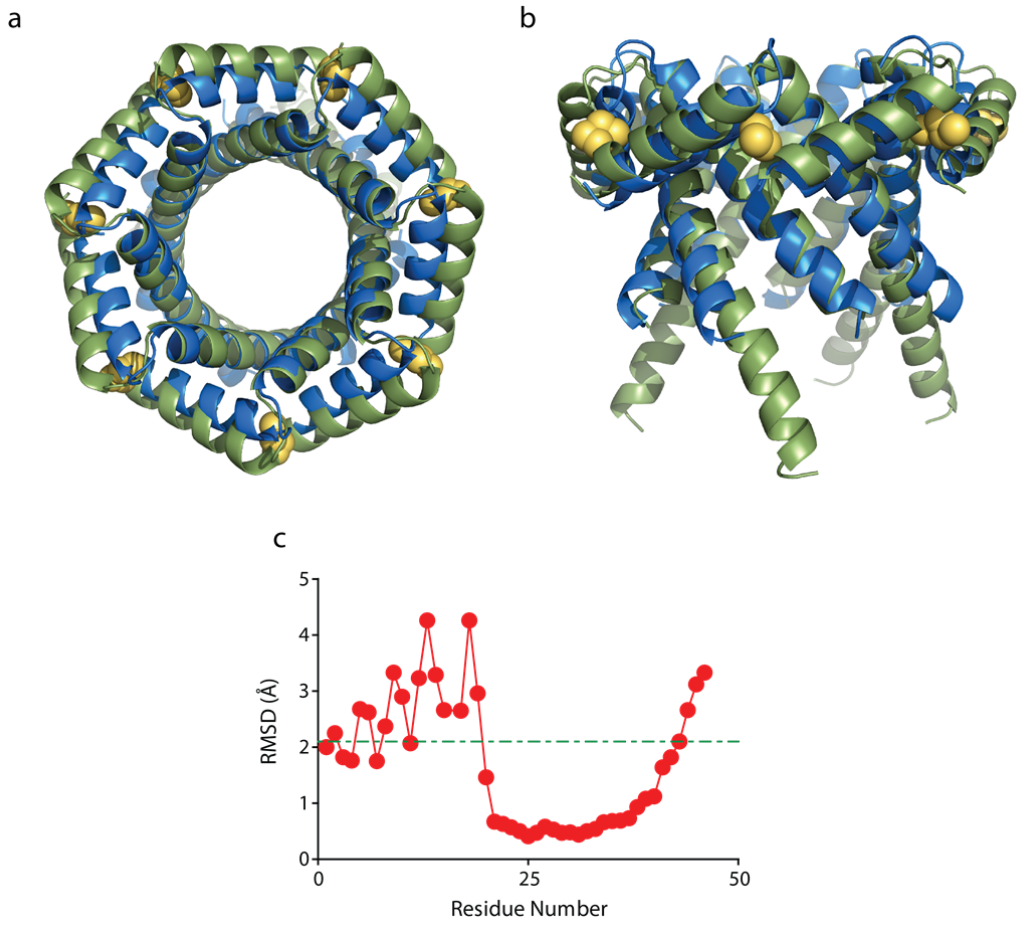


Figure 6.13 *Panel a*: Superposition of the heptameric structures of the oligomerisation domains from C4bp$_\alpha$ (blue) and CRES (green) viewed from above. Sulphur atoms in cysteine residues forming intermolecular disulphide bridges are shown as gold spheres. *Panel b*: Side view of the superposition. *Panel c*: Graph of the average RMSD per residue of the superposition. The overall RMSD (2.1 Å) is marked in green.

6.5.1 Assembly of natural heptameric coiled-coils

Coiled-coils can form between alpha helices when their side chains can pack in a "knobs into holes" fashion first proposed by [Crick 1953] resulting in a supercoil. Assembly in this manner distorts the helices, reducing the number of residues per turn from 3.6, found in regular helices, to nearer 3.5, causing a heptad periodic repeat within every two helical turns [Mason and Arndt 2004]. The "Peptide Velcro" hypothesis describes a set of rules upon the properties of residues, denoted **a-b-c-d-e-f-g**, within the heptad repeat required for formation of a specific coiled-coil [Arndt, Pelletier *et al.*

2002]. Residues **a** and **d** must be hydrophobic, mediating hydrophobic and Van der Waals interactions, whilst **e** and **g** must be charged, forming interhelical electrostatic interactions. **b**, **c** and **f** must be hydrophilic as they are solvent exposed. Specificity for the number of strands in the coiled-coil and the parallel, or anti-parallel, arrangement of the helices is determined by the identity of the residues at each position throughout the assembly.

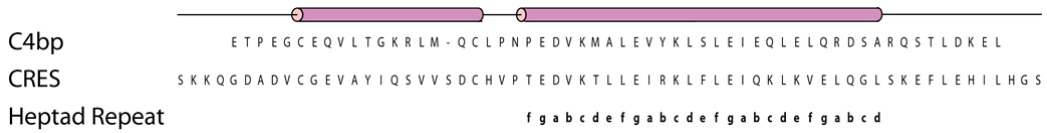


Figure 6.14 Alignment of the oligomerisation domains from C4bp$_\alpha$ and CRES with the heptad assignment of the heptameric coiled-coil formed by the C-terminal helices.

The heptad repeats of the heptameric coiled-coils in C4bp and CRES have been assigned with the aid of the program SOCKET [Walshaw and Woolfson 2001] and are displayed in Figure 6.14. Both coiled-coils are generated by hydrophobic residues at the **f** position inserting into hydrophobic pockets on the clockwise monomer formed between two **a**'s and a **d** residue (Figure 6.15, panel a). Hydrophilic residues at the **c** and **g** positions form a hydrogen-bonding network within the central channel of the assembly (Figure 6.15, panel c and d). Residues in the **b** and **e** positions are situated on the outer surface of the assembly and are generally hydrophilic, shielding the hydrophobic surface from solvent thus aiding solubility (Figure 6.15, panel b). This pattern is in contrast to the "Peptide Velcro" rules previously discussed, possessing a hydrophobic residue at position **f** and may be a contributing factor to the specificity for heptameric coiled-coil formation.

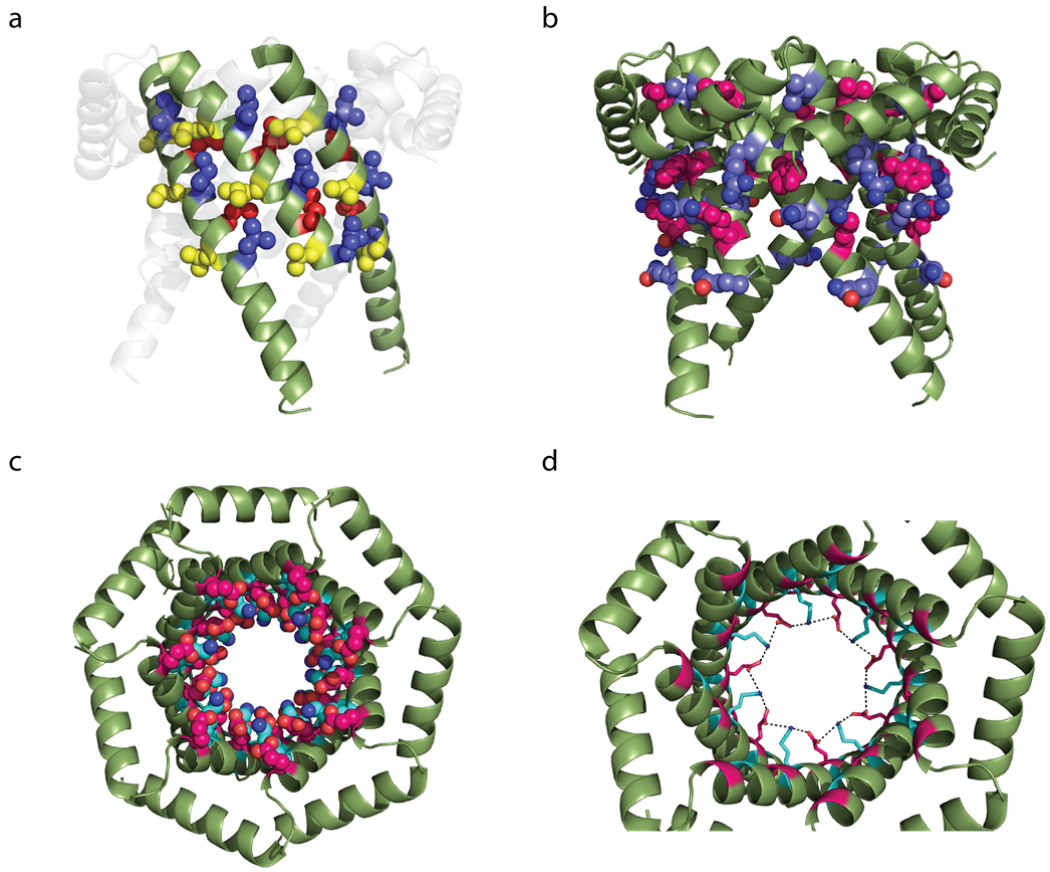


Figure 6.15 *Panel a*: Cartoon representation of the oligomerisation domain from CRES showing residues in the **a** (blue), **d** (red) and **f** (yellow) positions of the heptad repeat on three neighgouring chains in the homoheptameric coiled-coil. Each hydrophobic residue in an **f** position inserts into a pocket formed by two **a** residues and a **d** on the neighbouring chain. *Panel b*: Hydrophilic residues in the **b** (blue) and **e** (pink) positions are found on the outer, solvent exposed surface of the assembly. *Panel c*: Hydrophilic residues in the **c** (cyan) and **g** (pink) positions line the central channel, forming an extensive hydrogen bonding network along the length of the coiled-coil. *Panel d*: Example section of the hydrogen bonding network formed within the central channel. Shown here are the side chains of Lys433 (cyan) and Glu 437 (pink) forming a ring of intermolecular hydrogen bonds.

The manner in which the **f** residues insert into the pockets between **a** and **d** residues would result in a phase offset in the heptad repeat after a supercoil turn. However, analysis of the supercoil properties of C4bp and CRES using TWISTER [Strelkov and Burkhard 2002] shows that they have a rise per supercoil turn of approximately 1.0, resulting in no phase shift and this produces a molecule with a flat top and rotational symmetry about the supercoil axis. This effect is created by the radius of curvature which ensures that the interactions between neighbouring heptad repeats remains in

	C4bp	CRES	2HY6
Superhelix properties			
Radius (Å)	17.6	18.3	11.4
Rise per turn (heptads)	1.0	1.0	1.4
Pitch (Å)	356.0	370.5	202.2
Radius of curvature (Å)	17.9	18.4	137.0
α-helix properties			
Radius (Å)	2.26	2.24	2.29
Rise per turn (Å)	1.64	1.66	3.61
Residues per turn	3.5	3.5	3.6

Table 6.6 Properties of the heptameric coiled-coils from C4bp$_\alpha$, CRES and PDB entry 2HY6.

phase. Formation of the coiled-coils in C4bp and CRES break down after residue Ala578 and Leu448 respectively as the helices diverge and become labile with electron density only being present for those chains stabilised by involvement in crystal packing. The pattern of hydrophobic and hydrophilic side chains in the heptad repeat also breaks down after these residues.

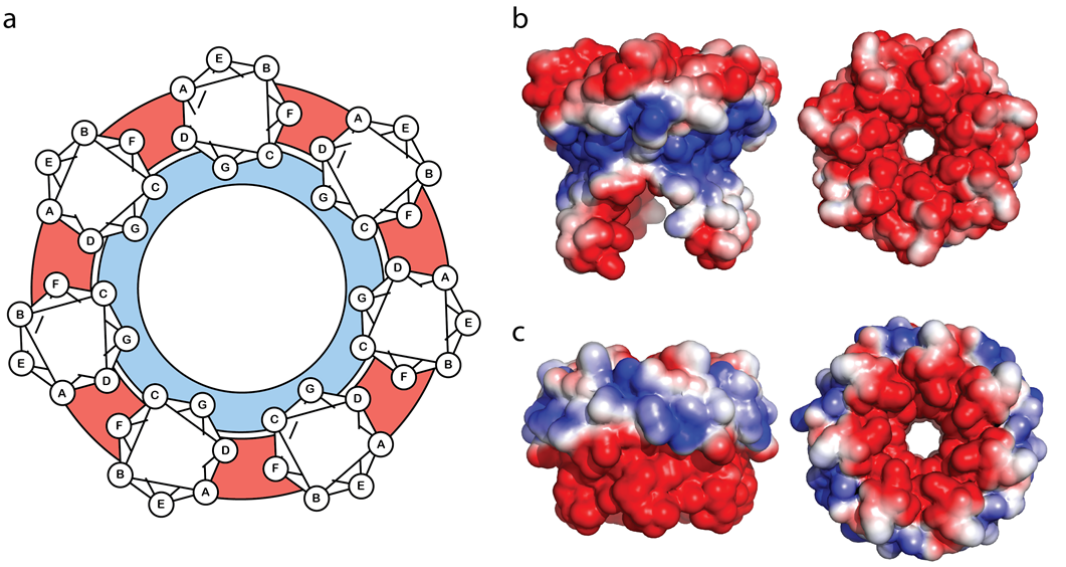


Figure 6.16 *Panel a*: Schematic describing the formation of the homoheptameric coiled-coils in CRES and C4bp. The position of each residue in the heptad repeat is displayed for each helix and the nature of their interactions described by the blue area (intermolecular hydrogen bonding network), the red area (intermolecular hydrophobic interactions) or the white area (solvent exposed). *Panel b*: Surface charge representation of the assembly of CRES oligomerisation domains calculated using APBS [Baker, Sept *et al.* 2001]. Charge scales from -50 mV (red) to +50 mV (blue). *Panel c*: Surface charge representation of the assembly of C4bp$_\alpha$ oligomerisation domains.

The protein fold in C4bp is not unique. Searching for structures with similar architecture using DALI [Holm and Rosenstrom 2010] reveals that the structure of Phospholamban (PDB entry 1FJK [Lamberth, Schmid *et al.* 2000], an integral

membrane protein that regulates the Ca^{2+} pump in cardiac and skeletal muscle cells [Rodriguez and Kranias 2005], is highly similar. Superposition of the monomer subunit of the $C4bp_{\alpha}$ oligomerisation domain with the NMR structure of Phospolamban shows 100% conservation of secondary structure elements over 52 residues with an RMSD of 2.41 Å and sequence identity of 19% (Figure 6.17). Phospholamban forms a pentameric species as described in PDB entry 2KYV [Verardi, Shi *et al.* 2011] mediated by formation of a parallel coiled-coil by the C-terminal helices (Figure 6.18). There are no cysteine residues in the N-terminal helix, and it is immediately obvious that these helices do not form a collar around the bundle in an analogous fashion to C4bp. Assignment of the heptad repeat using SOCKET reveals that the there is a stutter in the sequence and that although the general trend has hydrophobic residues in the **a**, **d** and **f** positions, residues **c** and **g** are also hydrophobic so a hydrophilic, hydrogen bonded core cannot be formed. The residues in positions **b** and **e** are a mixture of hydrophobic and hydrophilic residues which must aid stability within the membrane. These findings explain the specificity of Phospholamban for formation of a pentamer and not a heptamer. This adds further evidence that the formation of the hydrogen bonding network within the central channel of $C4bp_{\alpha}$ and CRES drives specificity for heptamerisation.

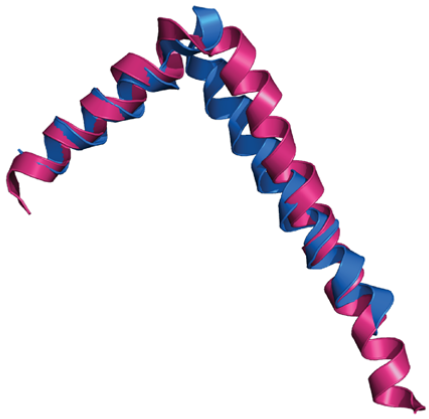


Figure 6.17 Superposition of a single copy of the $C4bp_{\alpha}$ oligomerisation domain (blue) with Phospholamban (pink. PDB entry 1FJK).

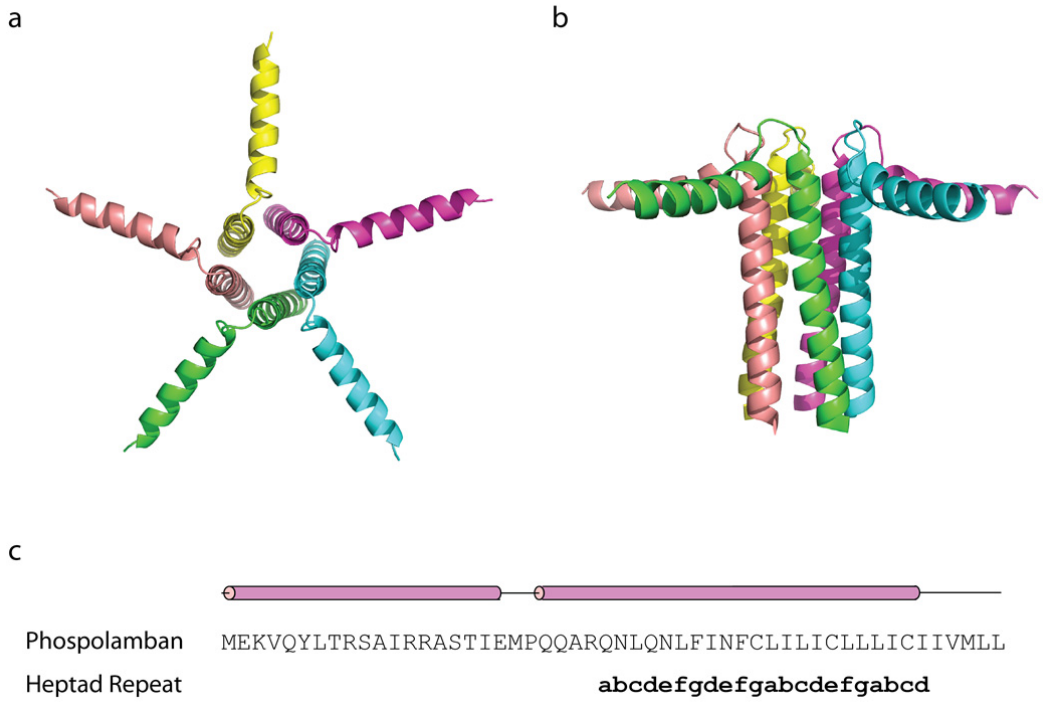


Figure 6.18 *Panel a*: Top view of the pentameric assembly of Phospholamban (PDB entry 2KYV) coloured by chain. *Panel b*: Side view of the pentameric assembly of Phospholamban. *Panel c*: Assignment of the heptad repeat within the coiled –coil.

Analysis of the sequence of the C4bp$_\beta$ chain explains why it has a specificity for a dimer and not a heptamer. Aligning the sequences of the alpha and beta chains and superposing the heptad repeat calculated previously shows that the beta chain does not maintain a conserved pattern of hydrophobic residues at the **a**, **d** and **f** positions (Figure 6.19). Although hydrophilic residues are conserved fairly highly at the other positions, the differences lead to an explanation for the altered specificity observed in assembly of the coiled-coil in the beta chain compared to the alpha.

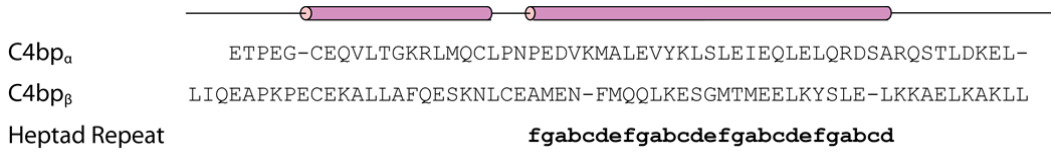
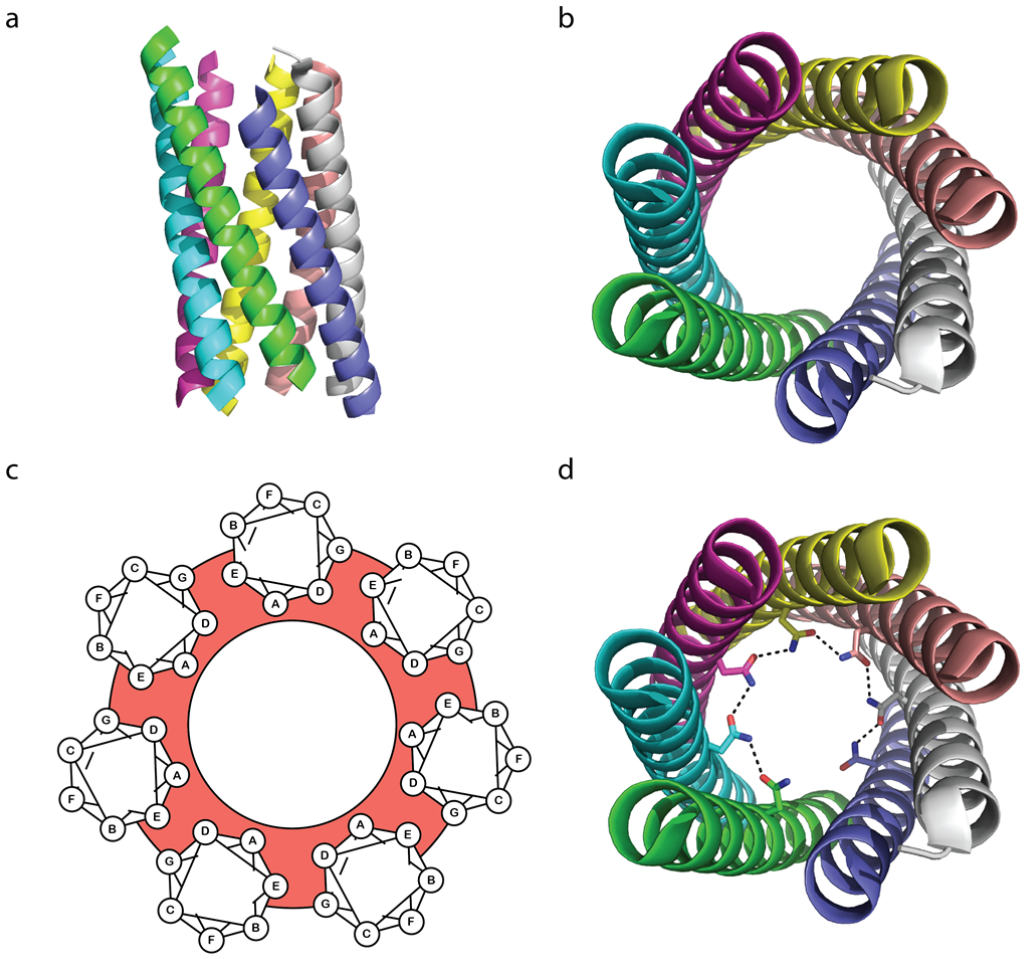


Figure 6.19 Alignment of the oligomerisation domains from the alpha and beta chains of C4bp shown with the heptad repeat assignment of the coiled-coil in the alpha chain.

6.5.2 Comparison of heptameric coiled-coils: C4bp, CRES and PDB entry 2HY6



(Figure 6.20 *Panel a*: Cartoon representation of PDB entry 2HY6. Each chain in the engineered homoheptameric coiled-coil is coloured separately. Note the phase shift in the heptad repeat between chain A (green) and chain G (dark blue). *Panel b*: Top-down view of the assembly showing formatin of a channel through the centre. *Panel c*: Schematic describing the formation of the homoheptameric coiled-coils in PDB entry 2HY6. The position of each residue in the heptad repeat is displayed for each helix and the nature of their interactions the red area (intermolecular hydrophobic interactions) or the white area (solvent exposed). *Panel d*: The single set of 6 hydrogen bonds between Asp17 in each chain that induces specificity for heptamer formation.

C4bp, CRES and PDB entry 2HY6 all display formation of a homo-heptameric parallel coiled-coil with a left handed supercoil axis and similar supercoil radii (Table 6.6). However, assembly of C4bp and CRES occurs differently to PDB entry 2HY6 which has the hydrophobic **a** and **d** residues lining the central channel with further inter-helix interactions between hydrophobic **e** and **g** residues [Liu, Zheng *et al.* 2006]. This results in a structure that has a heptad shift between the 1st and 7th chains (Figure 6.20),

lacking a “flat top” and pure rotational symmetry about the supercoil axis, characterised by a much greater radius of curvature compared to the natural proteins and a rise per superhelical turn that is not equal to 1 (Table 6.6). These differences are visible in the layers of interacting residues which form a screw in 2HY6 but distinct layers in C4bp and CRES (Figure 6.21).

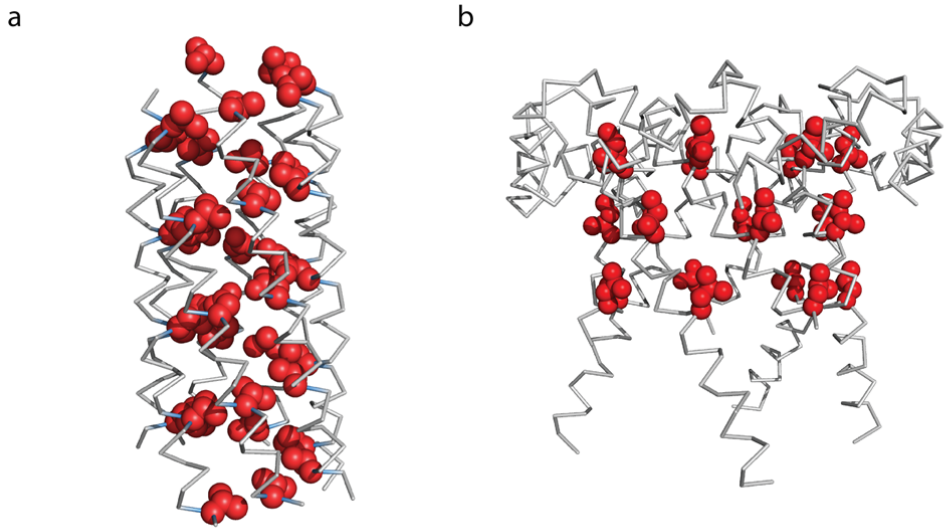


Figure 6.21 *Panel a*: Residues in heptad position **a** in PDB entry 2HY6 describe a left-handed screw moving from one chain to the next. *Panel a*: Residues at position **a** in the hepatemeric assembly of CRES oligomerisation domains form discrete layers.

2HY6 was generated from a tetrameric leucine-zipper protein by mutation of a single hydrophobic residue at an **a** position (residue 17) for asparagine which forms a hydrogen-bonding layer within the central channel (Figure 6.20). This mutation is critical for assembly of the heptameric species and likely causes the heptad phase shift [Liu, Zheng *et al.* 2006]. C4bp and CRES have evolved to have multiple layers of polar interactions down the central channel which provides both specificity for heptamer formation and distorts the hydrophobic packing enough to remove any phase shift, indicating that the polar interactions drive the formation of the coiled-coil by overcoming the energetic penalty imposed by disruption of the hydrophobic packing

and distortion of the helices. This is echoed in the increased distortion of the alpha helix parameters (Table 6.6) from their regular values in the natural proteins compared to 2HY6 and suggests that evolution of C4bp and CRES to be heptameric is critical to their function. It has also been previously noted that internal cavities in protein structures can lead to intrinsic instability [Akasako, Haruki *et al.* 1997], suggesting that by being hydrophilic down whole length instability can be outweighed by increased solvation and formation of hydrogen-bonds.

6.5.3 Disulphide-bond formation

Inter-molecular disulphide bonds in C4bp and CRES form between the first cysteine in the N-terminal helix and the second cysteine in the N-terminal helix of the clockwise monomer, locking the molecule together (Figure 6.22, panels a and b). As the N-terminal helix wraps around the C-terminal helix assembly, hydrophobic helix-helix interactions occur between the N-terminal helix of one monomer and the C-terminal helices in the clockwise and the clockwise plus 1 monomer accounting for why these residues in the **e** position of the heptad repeat are non-polar (Figure 6.22, panels a and b).

CRES has an extra residue between the two cysteine residues compared to C4bp, the result of which is a "tightening" of the disulphide-bonded collar without disruption of the coiled-coil which could account for differences observed in disulphide bond formation during purification. Disulphide bonds form much more readily in CRES whilst they form much more slowly in C4bp (author's observation), which could indicate that the tighter conformation is more strained and hence more difficult to assemble.

Another rationale for this observed difference in assembly rate could be due to formation of the coiled-coil itself. It has previously been reported that trigger sequences can promote coiled-coil formation [Kammerer, Schulthess *et al.* 1998]. The first residue in the heptameric coiled-coil of CRES is hydrophilic (threonine) not hydrophobic as predicted by its heptad assignment of **f**. It could be possible that the ability to form a hydrogen bond may aid initial assembly and may account for why CRES appears to assemble more readily. The equivalent residue in C4bp is proline which is not capable of behaving in this manner (Figure 6.22, panels c and d).

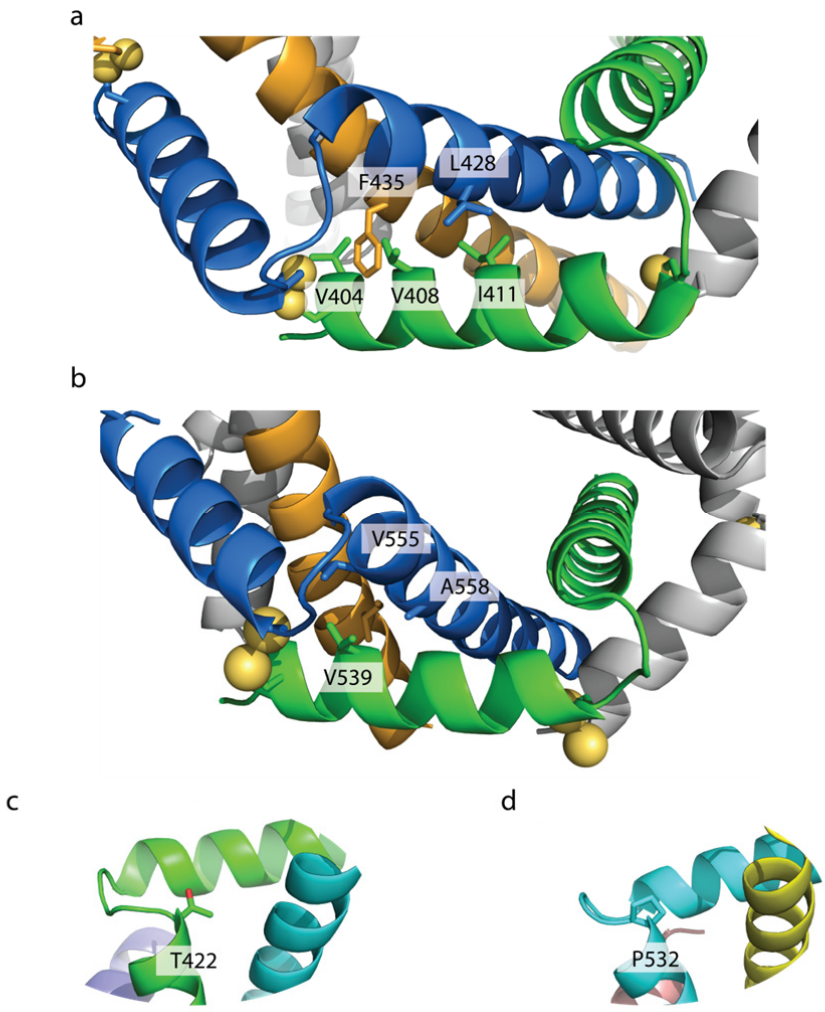


Figure 6.22 *Panel a*: Hydrophobic interactions between the N-terminal helix of one of the CRES oligomerisation domain monomer subunits (green) with the C-terminal helices in its clockwise neighbour (blue) and its clockwise + 1 neighbour (orange). *Panel b*: Hydrophobic interactions between the N-terminal and C-terminal helices in $C4bp_{\alpha}$. *Panel c*: The coiled-coil initiator residue Thr422 in CRES. *Panel d*: The coiled-coil initiator residue Pro532 in C4bp.

6.6 Investigation into the stoichiometry of C4bp

Previous studies have described the existence of multiple isoforms of C4bp based upon counting bands on gels during limited proteolysis, immunoblotting and their apparent mass on an SDS-PAGE gel [Hillarp, Hessing *et al.* 1989]. These experiments determined that there are three C4bp isoforms consisting of differing ratios of alpha and beta chains, 7α1β, 7α0β and 6α1β. Binding assays for protein S and immunoblotting have confirmed the presence and absence of the beta chain in C4bp species purified from serum [Hillarp and Dahlback 1988; Hillarp, Hessing *et al.* 1989]. Coupling these findings with electron microscopy images that allow counting of seven "legs" on each C4bp molecule [Dahlback, Smith *et al.* 1983] confirms the existence of the 7α0β and α7β1 isoforms. However, the presence of a species with six "legs" was not reported and other than counting of gel bands following chymotryptic digestion, the existence of the 6α1β isoform has never been robustly demonstrated.

The oligomerisation domains from both CRES and C4bp have been shown to form homoheptameric assemblies in solution (Figure 6.4 and Chapter 5.1.2) and in the crystal structures. However, when run on a non-reducing SDS-PAGE gel (Figure 6.4 and Chapter 5.1.2) the fully disulphide bonded structures run as a doublet. This doublet may be an artefact of how the coiled-coil formed between the oligomerisation domains interacts with the acrylamide gel and could possibly result from a fully and partially denatured forms of the assembly. Finding that the oligomerisation domain runs as a doublet could also account for the apparent lower molecular weight C4bp isoform (6α1β) if the fully assembled material behaves in a similar manner.

In order to confirm whether the 6α1β isoform of C4bp exists, renewed investigation was undertaken using a variety of characterisation methods including MALS and native mass spectrometry. These approaches may also help to demonstrate if the dimeric

C4bp$_\beta$ oligomerisation domain observed in the recombinant material is of any physiological relevance.

6.6.1 Analysis of C4bp using multiple angle light scattering

Unfortunately, due to the large molecular mass of intact C4bp, it was not possible to separate the sample scattering from the void scattering on an S200 column in order to analyse the sample using MALS. An attempt was made to resolve this issue using a Superose 6 column, but the reduced resolution of the column resulted in a larger error in the measurement, preventing any investigation of multiple isoforms using this technique.

6.6.2 Analysis of C4bp using native mass spectrometry

Native mass spectrometry is a technique that enables determination of the molecular weight of a protein without denaturing the sample. This also enables the mass of protein complexes to be determined and may enable accurate elucidation of the different C4bp isoforms found in serum.

The masses of the C4bp alpha and beta chains calculated from their sequences are presented in Table 6.7. However, both chains are known to be glycosylated with the beta chain being highly modified [Chung, Bentley *et al.* 1985; Hillarp and Dahlback 1990]. In order to assess the levels of glycosylation, C4bp containing the beta chain was purified from serum by barium citrate precipitation (Chapter 6.1.5) and deglycosylation attempted using PNGASE-F (New England Biosciences, USA) as per the manufacturers instructions for a native glycoprotein. SDS-PAGE analysis of the resulting material is shown in Figure 6.23, displaying a reduction in mass of intact (non-reduced) C4bp indicative of successful removal of some, or all, of the glycans. Examination of the masses of the individual chains (reduced) shows that the alpha chain has a mass of

approximately 70 kDa and shows a small change upon deglycosylation, indicative of a low level of modification. The beta chain has a mass of approximately 50 kDa which drops to approximately 30 kDa upon deglycosylation, corresponding well with the mass calculated from the protein sequence and demonstrates a very high level of modification.

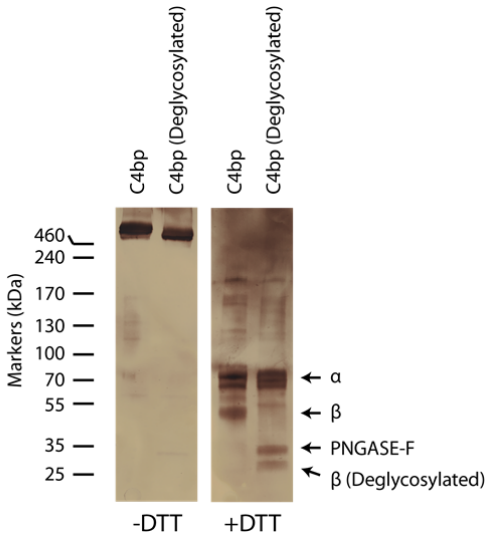


Figure 6.23 4 – 20% (w/v) SDS-PAGE analysis of C4bp pre and post deglycosylation in the presence and absence of reducing agent. The identities of the different species are marked.

Chain	Calculated Mass
α	61671 Da
β	26351 Da

Table 6.7 Masses of each C4bp chain calculated from its sequence using ProtParam.

Ionisation Conditions	Mass
150 V / 10 PSI Argon	517 ± 0.8 kDa
175 V / 10 PSI Argon	539 ± 0.6 kDa
200 V / 10 PSI Argon	506 ± 0.9 kDa
	516 ± 0.07 kDa
200 V *	51093.3 ± 0.93 Da
	488573 ± 344 Da

Table 6.8 Native mass spectrometry analysis of C4bp purified from serum under a variety of ionisation conditions. The condition marked with a star was performed in a different instrument with a higher energy collision cell.

C4bp was analysed using native mass spectrometry analysis by Dr. Julien Marcoux in the laboratory of Professor Carol Robinson (University of Oxford, UK) revealing a range of masses under different ionisation conditions (Table 6.8. Appendix 8). Under the strongest ionisation energies, two masses are observed of 51093 Da and 488573 Da. These masses suggest dissociation of C4bp with the smaller species corresponding to a single highly glycosylated beta chain (observed in Figure 6.23) whilst the larger fragment corresponds to seven alpha chains still in a heptameric assembly giving a monomer mass of 69786 Da. This latter finding is consistent with the mass of the alpha

chain calculated from the sequence with the small level of glycosylation suggested from Figure 6.23.

Stoichiometry	Calculated Mass
1α	70 kDa
1β	51 kDa
6α1β	470 kDa
7α0β	489 kDa
7α1β	540 kDa

Table 6.9 Approximate masses of the C4bp isoforms.

The masses of the 6α1β, 7α0β and 7α1β isoforms have been calculated using the measured masses of the alpha and beta chains (Table 6.9). The measured mass of the assembly of seven alpha chains gives a mass of 489 kDa to the 7α0β C4bp isoform whilst the 7α1β complex has a total mass of 540 Da which corresponds well to the 539 kDa species measured under weaker ionisation energies, suggesting that the complex does not dissociate under these conditions. The masses of the 7α0β and 7α1β isoforms also correspond to the findings of [Manabe and Jin 2008] who showed that the C4bp alpha chain could be detected by mass spectrometry in two SDS-PAGE bands with masses of 470 kDa and 540 kDa. The masses of the other species detected using native mass spectrometry, 506 kDa, 516 kDa and 517 kDa, do not correspond well to a 6α1β isoform, and are likely to be more 7α1β assemblies but with different glycosylation patterns that may account for why they have not ionised under these conditions and are further complicated by the fact that the C4bp was purified from mixed pool serum.

These findings suggest that 7α0β and 7α1β are correct stoichiometries of C4bp and the presence of other isoforms was not detected. They also suggest that dimerization of the oligomerisation domain from the beta chain is not relevant to its assembly with the alpha chain and is likely an artefact of expressing the domain alone. Another interesting observation is that the beta chain has been dissociated from the complex without the addition of reducing agent or disruption of the heptameric assembly of alpha chains, suggesting that the two are not joined by a disulphide linkage. This is in complete

contrast to the current opinion that the alpha and beta chains are linked [Hillarp and Dahlback 1988], but may also explain the original data if the beta chain is so tightly bound that it can only be released upon reduction or in the high energy collision cell.

6.7 A model for C4bp assembly

Having already discussed the complex assembly of the heptameric coiled-coil, it seems unlikely that the beta chain oligomerisation domain will insert into the arrangement to create an octameric coiled coil. This would also upset the tight geometry of the disulphide bonding network and possibly the hydrogen bonding network down the central core of the heptamer. One would also expect to have observed octamer formation of the alpha chain, but it displays a high specificity for assembly of heptameric species (Chapter 6.1.3). The findings in Chapter 6.6 also suggest that it is possible to remove the beta chain without disrupting the alpha assembly and that it is not disulphide bonded to the alpha. A proposed model could constitute the alpha chain oligomerisation domains assembled around a central beta chain, resulting in a heptamer of alpha chains with a single beta chain inserted down the centre (Figure 6.24). This model could also account for the tight binding of the latter, only being released upon disruption of the heptameric coiled-coil whilst maybe also explaining the low efficiency of beta chain incorporation demonstrated by Carlsson and Dahlback (2010) and the observation that no study has ever proposed C4bp isoforms with more than one beta chain.

Attempts have been made to verify this model by producing quantities of the fully formed oligomerisation assembly for characterisation. Both co-expression in *Escherichia coli* and refolding of the alpha and beta chains were attempted using multiple strategies and protein constructs. However, assembly of the two chains could not be achieved in this manner. Chymotryptic digestion was also employed in an

attempt obtain just the central body of oligomerisation domains from intact C4bp. However, samples could not be prepared of sufficient purity, homogeneity and volume for characterisation using this method.

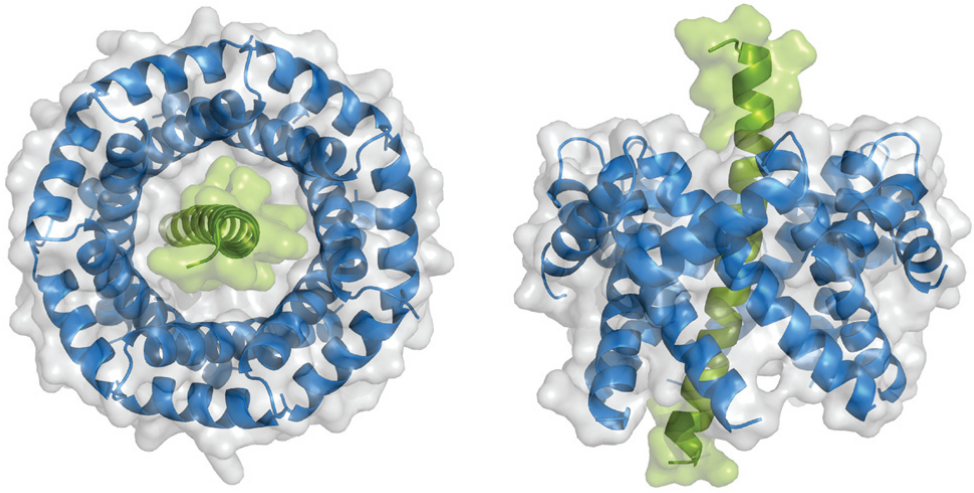


Figure 6.24 Structure of the proposed model for the combination of the oligomerisation domains from th alpha and beta chains showing that it is possible to fit an additional helix into the central channel formed by the heptameric assembly of the alpha chains. The beta chain is modeled by a canonical alpha helix taken from PDB entry 3HE4 [Reinke, Grant *et al.* 2010]).

6.8 Discussion

C4b-binding protein (C4bp) is a major negative regulator of the C3 and C5 convertases common to the classical and lectin activation pathways of the complement system. Possessing both decay-accelerating activity against the convertases and cofactor activity for factor I mediated inactivation of C4b, C4bp resembles a bunch of flowers with chains of SCR domains held together by a disulphide linked "central body" assembled from C-terminal oligomerisation domains on each chain. Two separate polypeptide chains exist in C4bp termed α (8 SCR domains) and β (3 SCR domains)chains with the most common isoform (approximately 70 - 75% of that found in plasma [Hillarp, Hessing *et al.* 1989]) regarded to be 7α1β. Preceding this investigation, the manner in which this "central body" is assembled was unknown and the pattern of disulphide bonding within this assembly had yet to be elucidated. Following solution of the atomic

resolution structure of the oligomerisation domain from the C4bp alpha chain, both of these unknowns can be resolved. The oligomerisation domains consist of two alpha helices in a helix-hinge-helix arrangement. Assembly is mediated via formation of a heptameric coiled coil between the C-terminal helices whilst the N-terminal helices form a collar around the coiled-coil. Cysteine residues are found at either end of the N-terminal helix and disulphide bonds form intermolecularly between a monomer and both of its neighbours, locking the N-terminal helices together in a continuous ring.

Coiled-coil formation is abundant in nature and they are found in a wide variety of proteins with diverse functions including formation of biological assemblies, molecular recognition and cell signalling (reviewed in Burkhard, Stetefeld *et al.* (2001)). Biologically significant proteins have been observed to contain coiled-coils with a wide variation in the number of constituent helices, but to date a homo-heptameric coiled-coil has yet to be found in nature. Presented in this chapter and Chapter 5 were the structures of the oligomerisation domains from C4bp and a functional homologue in *Gallus gallus*, CRES, showing formation of the first examples of this protein architecture in nature.

Analysis of the structures of the homo-heptameric oligomerisation domains provides a molecular rationale for the specificity of the molecules to assemble in this manner. Assignment of the heptad repeat shows that both C4bp and CRES have hydrophobic residues at the **a**, **d** and **f** positions and hydrophilic residues at **b**, **c**, **e** and **g**. This results in an assembly with an open ended channel through the centre which is lined with hydrophilic residues forming a hydrogen bonding network, generation of which is likely to provide the specificity for heptamerisation. This arrangement is in contrast to the accepted set of rules, known as the "Protein Velcro" theory suggesting that the rules could be modified to include these findings and may lead to the discovery of further homo-heptameric coiled-coils in nature.

These observations explain why the assemblies of oligomerisation domains from the C4bp alpha chain and CRES result in a molecule with a flat top, a surface that could be involved with binding to a ligand and requires further investigation. The complex set of interactions associated with heptamer formation also shows the importance of forming this precise stoichiometry of multimerisation upon the function of the molecule. It is highly likely that that C4bp and CRES are heptameric as duplication of the binding sites will result in higher apparent affinity for their ligands resulting from avidity. This suggests that this mechanism is crucial for efficient complement regulation by C4bp, which is highly evolved to providing the correct balance of complement activation in order to protect host cells, but enable rapid clearance of foreign entities, requiring C4bp to possess seven alpha chains.

Previous observations of the existence of different C4bp isoforms containing different numbers of alpha chains have been questioned in light of the structure of their assembly. As previously discussed, the interactions forming the homoheptameric coiled-coil formed by the alpha chain oligomerisation domains are highly specific. Addition or removal of a helix from this assembly would require distortion of the inter-helical interactions that are tuned towards heptamer formation and would also effect the geometry of disulphide bond formation. In order to form the 6α1β isoform, an alpha chain oligomerisation domain in the heptameric assembly may have to be replaced by one from the beta chain. However, the latter has been shown to lack the correct pattern of hydrophobic and hydrophilic residues in the heptad repeat to do this. The functional implications of decreased numbers of alpha chains would also result in decreased avidity for C4b and hence effect the homeostatic balance of complement activation. Thus it was hypothesised that only isoforms containing seven alpha chains exist. Native mass spectrometry was employed to investigate the existence of different C4bp

isoforms, but found no evidence of any isoforms other than 7α1β and 7α0β, agreeing with the hypothesis and suggesting that evidence of other species may be explained by differential glycosylation patterns and/or experimental artefacts.

The C4bp beta chain oligomerisation domain was observed to dimerise *in vitro* both within this thesis and in the study performed by [Carlsson and Dahlback 2010]. However, the native mass spectrometry investigation into the stoichiometry shows that dimerization is non-physiological as only one beta chain is incorporated with the alpha. Insertion of the C4bp beta chain into the alpha heptamer to form an octamer would also result in similar disruption of the coiled-coil packing and geometry of disulphide bonding previously discussed. It has been shown in this thesis that the beta chain can be dissociated from the alpha heptamer in the high-energy collision cell of a native mass spectrometer, in the absence of reducing agent, suggesting that the beta chain is not disulphide bonded to the alpha, but is instead attached in such a manner that it does not readily dissociate under non-reducing conditions. A model is proposed with the beta chain oligomerisation domain inserted into the central channel of the heptameric assembly of alpha chains and is not disulphide linked to them. This proposed new model means that the assembly of the alpha chains is identical in the 7α1β and 7α0β isoforms and could account for both the requirement for formation of the alpha structure with a channel through the centre and explaining why isoforms have not been observed with greater numbers of beta chains.

6.9 Summary

Presented in this chapter are significant findings that have led to the proposal of a new model for the assembly of the alpha and beta chains that comprise C4bp. The atomic resolution structure of the oligomerisation domain from the C4bp alpha chain has been solved using X-ray crystallography, showing formation of a species that is

architecturally homologous to the structure of the assembly of oligomerisation domains in CRES. The structure also describes the previously unknown pattern of disulphide bond formation, occurring between each chain. Heptamerisation in both assemblies is mediated by formation of homoheptameric coiled-coils which are the first examples of these structures in nature. Specificity for heptamerisation is driven by formation of a hydrogen bonding network within the central channel which results from non-classical population of the heptad repeat, possessing hydrophobic residues at the **a**, **d** and **f** positions, hydrophilic hydrogen bonded residues at the **c** and **g** positions and hydrophilic solvent-exposed residues at the **b** and **e** positions.

Native mass spectrometry has been employed to investigate reports of the existence of C4bp isoforms other than 7α1β and 7α0β, but no evidence was found suggesting that all molecules contain seven alpha chains. It was also possible to dissociate the 7α1β species within the native mass spectrometer into the beta chain and the assembly of seven alpha chains in the absence of reducing agent. Contrary to the currently accepted data, this suggests that the beta chain is not disulphide bonded to the alpha and a new model is proposed with the oligomerisation domain from the beta chain inserted into the channel in the centre of the assembly of oligomerisation domains from seven alpha chains.

Further work is required to verify the proposed model, but it is hoped that these findings will lead to improved understanding of the function of C4bp both as a regulator of complement and in any proposed mechanisms of involvement in the anti-coagulation pathway.

6.9.1 Future directions

Initial future experiments are to confirm whether the proposed model for assembly of the C4bp alpha and beta chains is correct. This will probably require the atomic

resolution structure of the central body of C4bp in order to completely rationalise the incorporation of the beta chain.

Continued searches will also be undertaken for other molecules that possess the proposed heptad repeat that confers specificity for formation of heptameric coiled-coils. This should confirm whether the hypothesis of the central hydrogen bonding network is true and may lead to the discovery of more homoheptameric coiled-coils in nature.

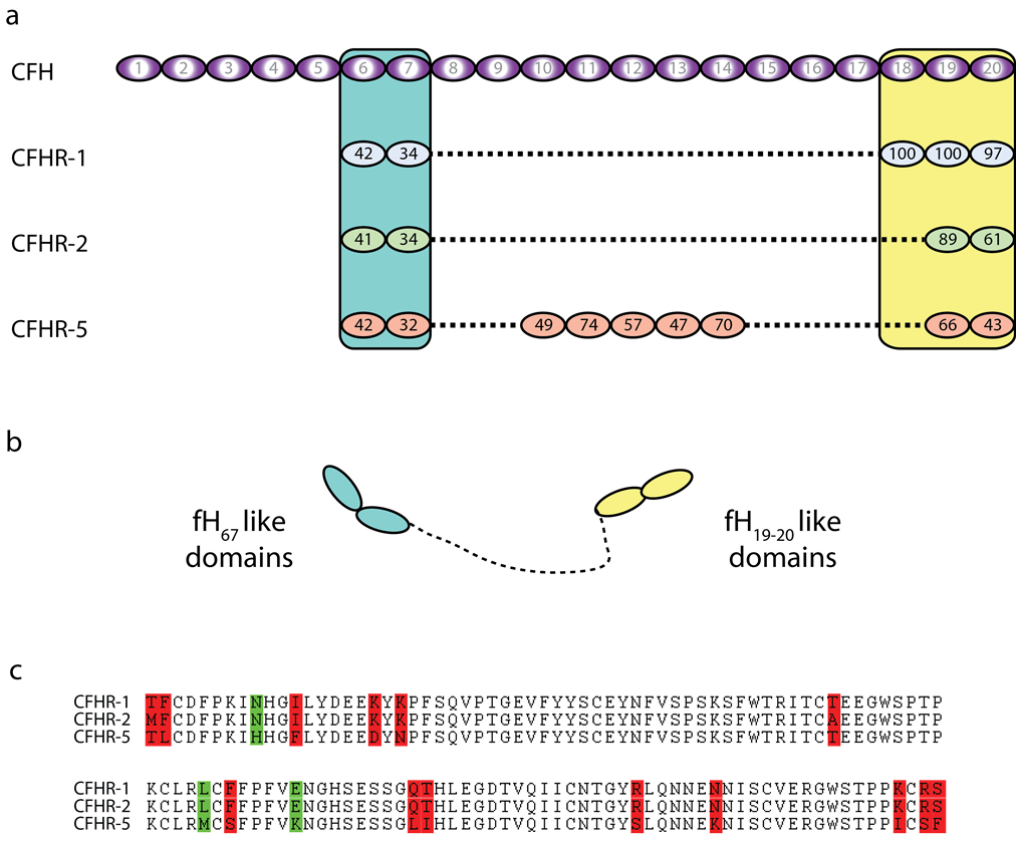


Figure 7.1 *Panel a*: Alignment of the individual domains of CFHR-1, CFHR-2 and CFHR-5 with their most similar in factor H. The percentage similarity between each pair of domains is displayed. The two N-terminal highly conserved, fH_{67} like, domains are highlighted in blue, whilst the C-terminal, fH_{19-20} like, domains are highlighted in yellow. *Panel b*: Common architecture of CFHR-1, CFHR-2 and CFHR-5 coloured as per panel a with a dotted line to represent the varying numberr of SCR domains between the pairs of domains under discussion. *Panel c*: Alignment of the two N-terminal domains of CFHR-1, CFHR-2 and CFHR-5. Major and minor differences are shown in red and green respectively. Percentage identity between these sets of domains in CFHR-1 and CFHR-2 or CFHR-5 are 98% and 88% respectively.

Previous *in vitro* studies have shown that CFHR-1 can inhibit the C5 convertase [Heinen, Hartmann *et al.* 2009], CFHR-3 can inhibit the C3 convertase whilst CFHR-5 inhibits the C3 convertase and possesses cofactor activity towards factor I mediated inactivation of C3b [McRae, Duthy *et al.* 2005]. However, many of these effects were observed at concentrations of proteins much higher than those found in serum. CFHR-4 and CFHR-5 have subsequently been shown not to exhibit cofactor activity at

physiological concentrations [Gale, de Jorge *et al.* 2010; Hebecker and Jozsi 2012]. These mixed findings have made it difficult to obtain convincing evidence to support a proposed mechanism for the CFHRs.

Analysis of the individual sequences suggests that a subfamily of CFHRs, consisting of CFHR-1, CFHR-2 and CFHR-5, may be identified by a highly conserved pair of N-terminal SCR domains and C-terminal domains that possess a high level of identity to the C-terminal domains of factor H (Figure 7.1). Domains one and two of CFHR-1 have been implicated in negative regulation of the C5 convertase belonging to the alternative pathway of complement activation [Heinen, Hartmann *et al.* 2009]. Due to the almost identical nature of these domains in CFHR-2 and CFHR-5, one might expect these domains to possess a similar function. However, no studies published to date have observed this property in either CFHR-2 or CFHR-5. The function of the C-terminal domains also remains unclear. Whilst the high levels of identity of these domains with the C-terminal domains of factor H suggests C3b and glycosaminoglycan (GAG) binding roles. Both CFHR-1 and CFHR-5 have been observed to bind C3b and heparin [McRae, Duthy *et al.* 2005; Heinen, Hartmann *et al.* 2009] but these functions have never been mapped to the C-terminal domains and it is unknown whether the equivalent domains in CFHR-2 behave similarly.

Knowledge of how they function may also explain two rare polymorphisms within the CFHRs that have been linked to increased susceptibility to C3 glomerulopathy. A familial set of individuals with Cypriot ancestry possess a heterozygous duplication of the N-terminal domains in CFHR-5 which has been associated with presentation of the disease [Gale, de Jorge *et al.* 2010; Athanasiou, Voskarides *et al.* 2011]. Another polymorphism consisting of a heterozygous hybrid of CFHR-3, plus the N-terminal domains of CFHR-1, present in a family of Irish origin [Malik, Lavin *et al.* 2012].

Another common polymorphism is a homozygous deletion of CFHR-1 and CFHR-3. This deletion has been associated with decreased susceptibility towards IgA nephropathy [Gharavi, Kiryluk *et al.* 2011] and age-related macular degeneration [Hughes, Orr *et al.* 2006], but increased susceptibility to systemic lupus erythematosus [Zhao, Wu *et al.* 2011]. Further investigation into how the domain duplications result in increased susceptibility to disease, and why the deletion can be both protective and linked to illness, is required in order to assess possible future treatments of all CFHR polymorphisms.

Presented here are data confirming that CFHR-1, CFHR-2 and CFHR-5 can be classified into a sub-family of CFHRs based upon their common N-terminal domains. These domains are shown to mediate dimerization of the molecules and the atomic resolution structure of CFHR-1 domains one and two has been solved by X-ray crystallography showing that the dimer is formed via a novel hydrophobic dimerization interface. Data are presented showing that due to the high levels of identity between the N-terminal domains of this CFHR subfamily both homo and heterodimers exist *in vivo* between CFHR-1, CFHR-2 and CFHR-5. The ability of the C-terminal domains of CFHR-1 and CFHR-2 to bind C3b and GAGs is shown alongside the atomic resolution structure of the two C-terminal domains of CFHR-2 enabling rationalisation of the differences and similarities between these CFHRs. Combination of all of these data allows a general model to be proposed of the architecture of CFHR-1, CFHR-2 and CFHR-5

Based on complement activity assays, the first universal mechanism of the function of the CFHRs is proposed: dimeric CFHR molecules compete with regulators for C3b binding, resulting in de-regulation of complement activation and an increased level of activity. This leads to full control over the levels of complement activation on the

surface of different serum exposed tissues which is important in maintaining homeostasis. The model also explains the susceptibility to disease of the polymorphisms, proposes a new therapy for individuals presenting with these polymorphisms, and suggests that pathogenic scavenging of the CFHRs needs closer examination.

The findings in this chapter were performed using CFHR-1, CFHR-2 and CFHR-5 which are referred to as CFHRs. Within this thesis this group does not include CFHR-3 and CFHR-4 unless specified.

7.1 Protein expression and purification

Due to the low serum concentrations of the CFHRs and mixed glycosylation states, purification of the proteins from serum is not a viable method to obtain homogenous samples in the quantity needed for crystallisation, requiring their production via recombinant methods. Initial attempts at refolding full length CFHR-1 from inclusion bodies, expressed in *Escherichia coli*, failed to generate soluble protein. In order to generate proteins for characterisation, a variety of expression systems and CFHR-1 constructs were explored.

7.1.1 Generation of plasmids pKLAC-CFHR1 and pKLAC-CFHR1$_{12}$

In order to facilitate eukaryotic expression of CFHR-1 in *Kluyveromyces lactis*, the gene encoding CFHR-1 was cloned into vector pKLAC-2 (New England Biosciences, USA) under the control of a $P_{LAC4\text{-}PBI}$ promoter. In order to achieve secretion of the mature protein, the *K. lactis* α-mating factor (α-MF) was appended to the N-terminus, coupled by a Kex protease site. Upon expression, a signal peptide in the α-MF domain directs the protein to the endoplasmic reticulum where the signal peptide is removed by signal peptidase. The immature protein is then transported to the Golgi apparatus where Kex

protease cleaves the remaining α-MF signalling secretion of the mature protein. The CFHR-1 sequence was amplified from a vector obtained from the mammalian gene collection (IMAGE ID: 4712577 [Strausberg, Feingold *et al.* 2002]) in a PCR reaction using KOD polymerase and primers CFHR1_1_For and CFHR1_5_Rev (detailed in Table 7.1) according to the manufacturers instructions. Amplified insert was digested with DpnI to remove template DNA as per standard protocols.Due to the presence of an NdeI restriction site within the CFHR-1 sequence, conventional cloning was rejected in favour of the InFusion dry-down system (Clontech, USA). The insert encoding CFHR-1 was ligated into the pKLAC-2 vector, pre-digested with NdeI and XhoI as per standard protocols, as per the manufacturers instructions. *Escherichia coli* strain Novablue (Novagen, USA) was transformed with the ligation reaction mixture and grown on LB/agar plates containing 100 μg/ml ampicillin. Positive colonies were selected using colony PCR (using GoTaq polymerase (Promega, USA)), as per standard protocols, and the sequence verified by Sanger sequencing using primers A_Mating_For and PKLAC2_Rev (Table 7.1). A map detailing the final vector, pKLAC-CFHR1, is shown in Figure 7.2.

CFHR1_1_For
5′-GCTGACAAGGATGATCTCGAGAAAAGAGAAGCAACATTTTGTGATTTTCC-3′
CFHR1_2_Rev
5′-GCCGCCCATGGACATCTAAGTGGACCTGCATTTGG-3′
CFHR1_5_Rev
5′-GCCGCCCATGGACATATGTCTTTTTGCACAAGTTGGATACTCCAG-3′
A_Mating_For
5′-GAAGAAGCCTTGATTGGA-3′
PKLAC2_Rev
5′-TTATCGCACAAGACAATC-3′

Table 7.1 Sequences of the primers used fo cloning CFHR-1 constructs into pKLAC2 and sequenciing the resulting vector.

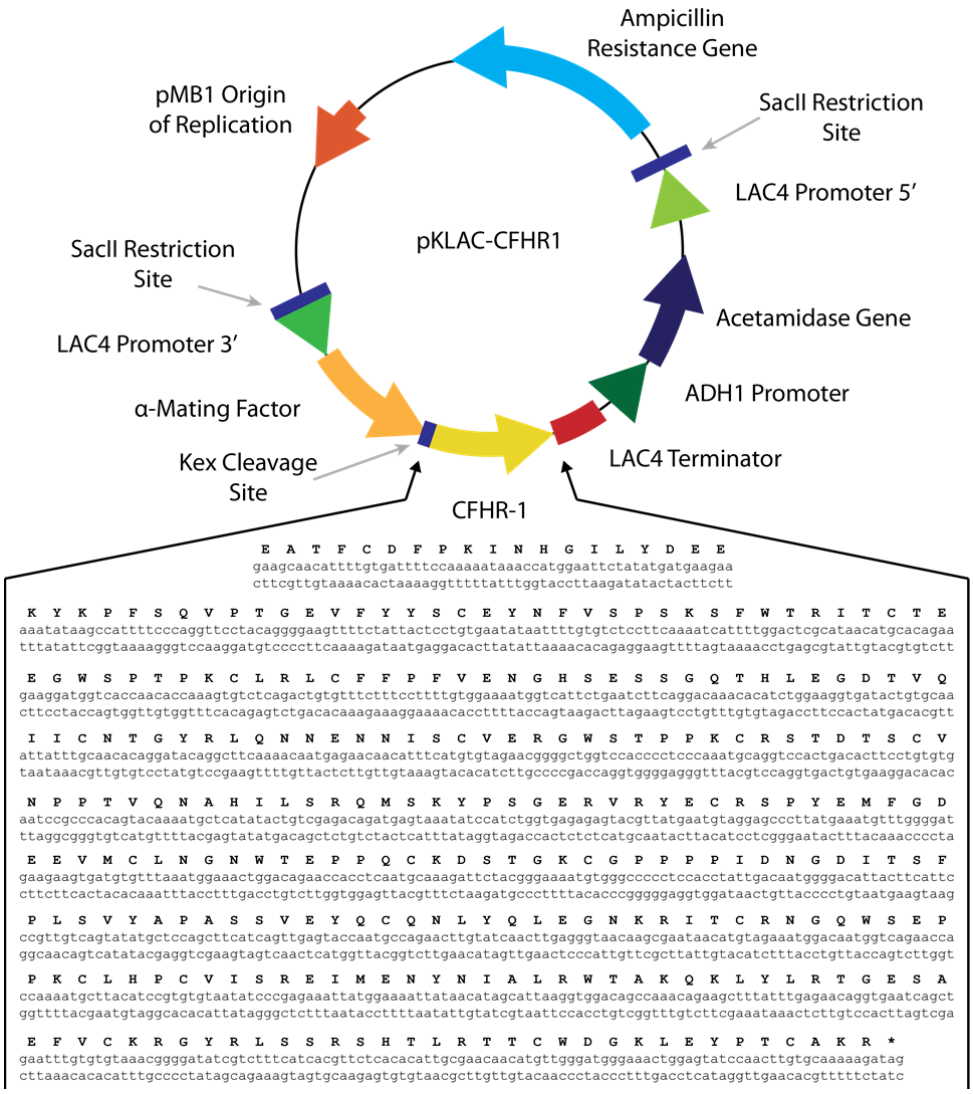


Figure 7.2 Schematic representation of pKLAC-CFHR1. Following chromosomal integration into the LAC4 locus of *K. lactis*, CFHR-1 is expressed without a purification tag, but is targetted for secretion by the α-mating factor peptide. The vector was generated by cloning the construct into pKLAC-2 (NEB, USA) between the NdeI and XhoI restriction sites.

Vector pKLAC-CFHR1$_{12}$, expressing CFHR-1 domains one and two only, was generated in an identical manner to pKLAC-CFHR1. An insert encoding these domains was amplified using primers CFHR1_1_For and CFHR1_2_Rev (detailed in Table 7.1) and ligated into the pKLAC-2 vector. The sequence was verified using Sanger sequencing and is described in Figure 7.3.

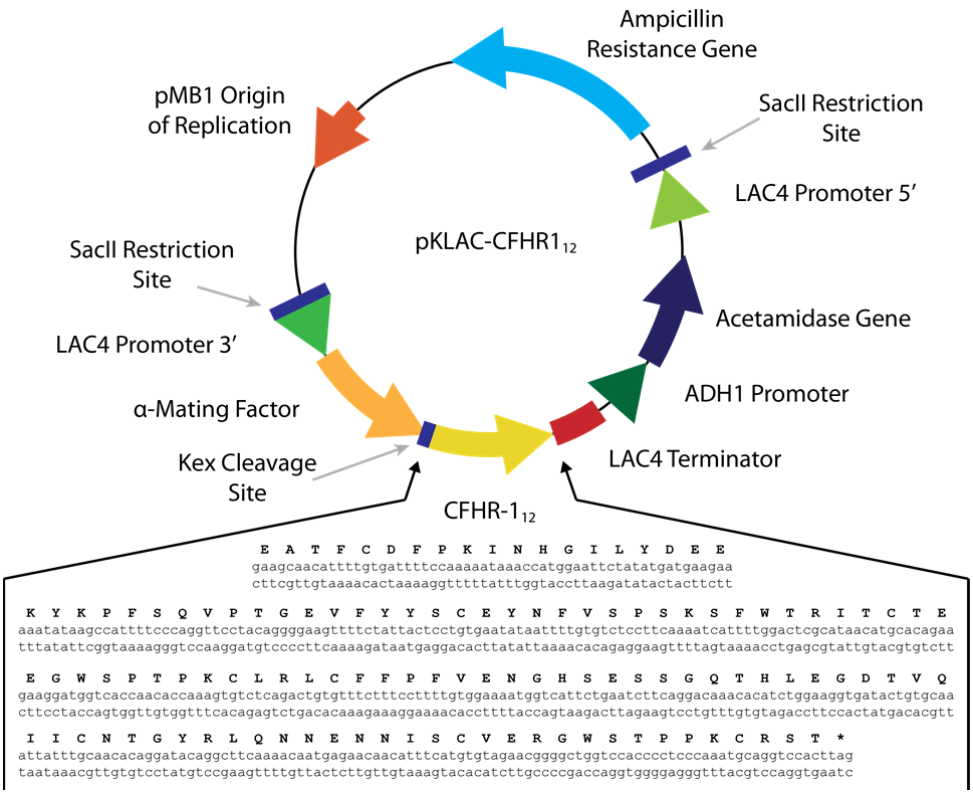


Figure 7.3 Schematic representation of pKLAC-CFHR1$_{12}$. Following chromosomal integration into the LAC4 locus of *K. lactis*, CFHR-1$_{12}$ is expressed without a purification tag, but is targetted for secretion by the α-mating factor peptide. The vector was generated by cloning the construct into pKLAC-2 (NEB, USA) between the NdeI and XhoI restriction sites.

7.1.2 Generation of plasmid pEF-CFHR1

In order to express CFHR-1 in mammalian cell culture using transient transfection, the gene encoding the target protein, including the native signal sequence, was cloned into plasmid pEF-BOS [Mizushima and Nagata 1990] with a C-terminal hexahistidine affinity tag. This should enable transfected cells to secrete the mature protein into the media from where it can be easily affinity purified using Ni-NTA.

An insert encoding CFHR-1 with its signal sequence was amplified from a vector obtained from the mammalian gene collection (IMAGE ID: 4712577) in a PCR reaction using KOD polymerase and primers CFHR1_PEF_For and CFHR1_PEF_Rev (detailed in Table 7.2) according to the manufacturers instructions. Amplified insert was digested

with DpnI to remove template DNA prior to digestion of the insert and vector, pEF-BOS, with restriction enzymes XbaI and SalI as per standard protocols. Following treatment of the doubly-cut vector with alkaline phosphatase (Fermentas, USA) the insert was ligated into the vector using a T4 DNA ligase ligation kit (Promega, USA) as per the manufacturers instructions.

CFHR1_PEF_For
5′-GGACTGTCTAGAGACTCAGATCCCCATCCGCTCAAGCAGGCCACCATGTGGCTCCTGGTCAGTGTAATTC-3′
CFHR1_PEF_Rev
5′-GTGTCGGGTCGACCTTTTTGCACAAGTTGGATACTCCAG-3′

Table 7.2 Sequences of the primers used for cloning CFHR-1 into pEF-BOS.

Escherichia coli strain DH5-α was transformed with the ligation reaction mixture and grown on LB/agar plates containing 100 µg/ml ampicillin. Positive colonies were selected using colony PCR (using GoTaq polymerase (Promega, USA) as per standard protocols, and the sequence verified by Sanger sequencing using a EF1a promoter primer. A map detailing the final vector, pEF-CFHR1, is shown in Figure 7.4.

7.1.3 Details of plasmids pET-CFHR1$_{345}$ and pET-CFHR2$_{34}$

Plasmids pET-CFHR1$_{345}$ and pET-CFHR2$_{34}$ were generated, under close supervision, by Mr. Matthew Colledge (Biochemistry Part II student. Lea Group. 2011-2012) and contain domains three through five of CFHR-1 and three through four of CFHR-2 respectively under the control of an inducible T7 promoter. Vector maps detailing each plasmid are presented in Figure 7.5.

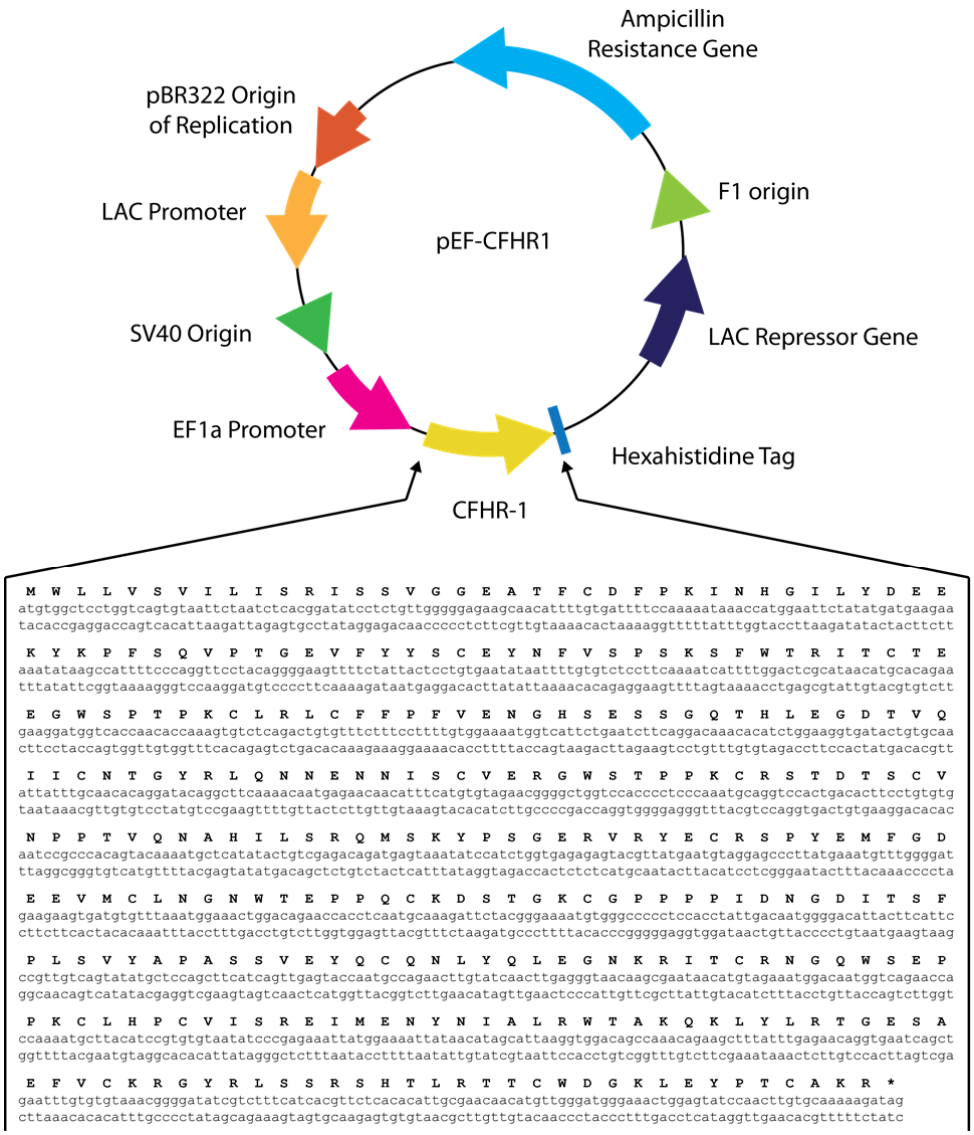


Figure 7.4 Schematic representation of pEF-CFHR. CFHR-1 is expressed with a C-terminal hexahistidine purification tag and is targetted for secretion during transient transfection by the native signal sequence. The vector was generated by cloning the whole CFHR-1gene into pEF-BOS between the XbaI and SalI restriction sites.

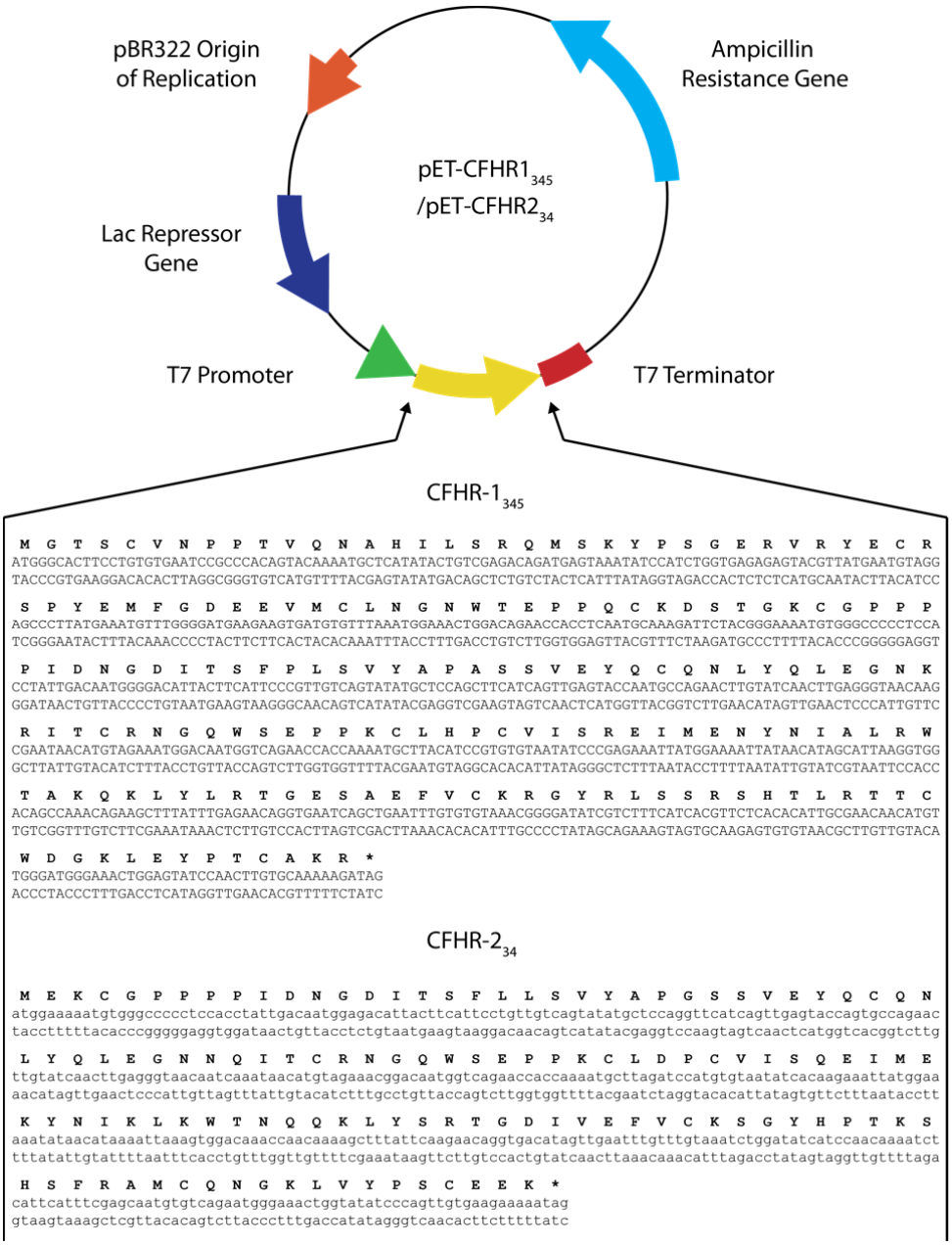


Figure 7.5 Schematic representation of pET-CFHR1$_{345}$ and pET-CFHR2$_{34}$. The proteins are expressed, without a purification tag, under the control of an inducible T7 promoter. The vector was produced by cloning of the fH construct into pET-15b (Novagen, USA) between the NcoI and XhoI restriction sites.

7.1.4 CFHR-1 expression and purification

pKLAC-2 can be linearised, generating an expression cassette with the ability to insert the gene of interest into the LAC4 locus of the *K. lactis* genome by homologous recombination. The cassette also contains a fungal acetamidase gene under a P_{ADH1} promoter that allows selection of colonies positive for recombination using acetamide as a nitrogen source. Constitutive expression of the target protein in the presence of galactose can then be performed in the absence of any selection.

Linearisation of pKLAC-CFHR1 was achieved by digestion with SacII. *K.lactis* (strain GG799) competent cells (New England Biosciences, USA) were transformed as per the manufacturer's instructions and plated on YCB/agar containing 5mM acetamide (Appendix 7). Colonies were selected after three days incubation at 30^{o}C and were re-plated for a further two days of selection.

Trial expressions in the manufacturer's recommended media, YPGal (Appendix 7), could not detect expression of CFHR-1 using SDS-PAGE. This was due to high background from the peptone in the media. Attempts to remove the peptone by dialysis, ammonium sulphate precipitation and ion exchange failed, as did purification of CFHR-1 from the peptone on a heparin column. In the absence of an antibody against CFHR-1, there was no evidence of expression. Coupled with the difficulties involved in removing the peptone, no further optimisation of expression in YPGal was performed.

In an effort to express CFHR-1 in the absence of peptone, growth of *Kluyveromyces lactis* was attempted in a minimal media. A modified minimal media for *K. lactis*, MMGal, was formulated, based on one used for selective NMR labelling [Sugiki, Shimada *et al.* 2008] with the addition of 100 mM phosphate buffer to keep the pH stable throughout expression (Appendix 7). Colonies containing the CFHR-1 gene were

grown overnight in 10 ml MMGal media at 30°C with shaking. Each overnight culture was then used to inoculate 1 L MMGal prior to incubation at 30°C, with 180 RPM shaking, for four days. The culture supernatant was harvested by centrifugation (4000 g, 4°C, 20 minutes), concentrated to 100 ml and dialysed against 50 mM Tris, 1 mM EDTA, pH 8.0. A 5 ml HiTrap Heparin column (GE Healthcare, USA) was loaded with the dialysed sample at 2ml/min before washing with 15 CV of 50 mM Tris, 1 mM EDTA, pH 8.0. The column was eluted with a 0 – 1.6 M gradient of NaCl over 40 CV (Figure 7.6, panel a). Fractions collected during the elution were analysed using SDS-PAGE in the presence and absence of reducing agent to check for disulphide bond formation (Figure 7.6, panel b).

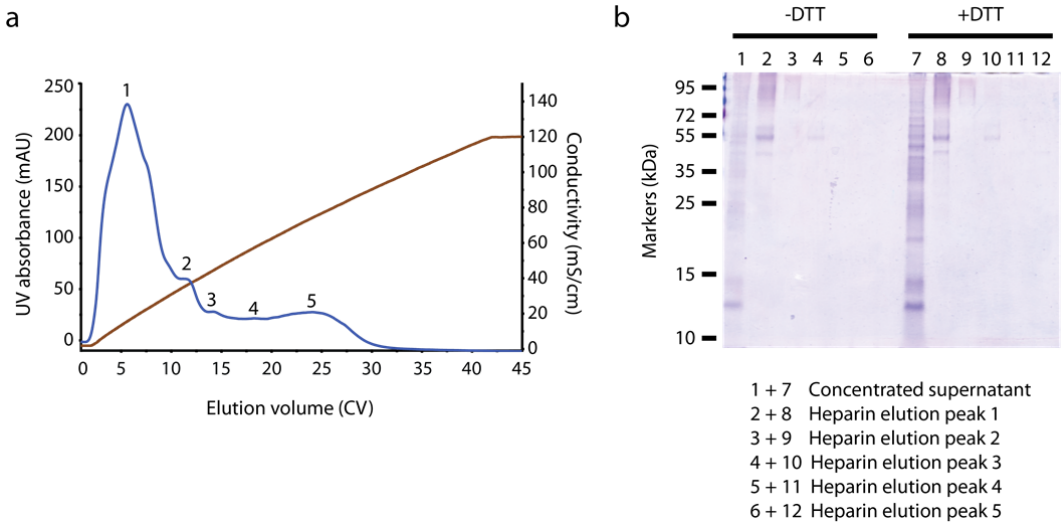


Figure 7.6 *Panel a*: Elution profile from the HiTrap Heparin column eluted with a salt gradient. The UV absorbance and conductivity traces are shown in blue and brown respectively. Each peak is labelled for analysis. *Panel b*: 15% SDS-PAGE gel of the elution peaks in the presence of absence of reducing agent.

Unfortunately, none of the elution peaks from the heparin column displays the gel-shift expected of a correctly folded species, and only one set of fractions has a mass close to that of the expected 36 kDa. Previously published studies have shown that CFHR-1 should bind heparin with a very high affinity under these conditions, supporting the

conclusion that full length CFHR-1 cannot be expressed in large quantities using *K. lactis*.

Following the failure to express CFHR-1 in *K. lactis*, an attempt to recombinantly produce the full length protein, in a correctly folded and soluble form, was performed using mammalian cell culture. In order to quickly assess whether expression of CFHR-1 is possible, transient expression was chosen as it does not require the extended period of selection needed to generate a stable cell line.

Cells belonging to human embryonic kidney (HEK) cell line 293T were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 5% foetal calf serum (FCS), non-essential amino acids, L-glutamine, penicillin and streptomycin as per standard protocols. In order to assess CFHR-1 expression in the media described above and an FCS free media, X-VIVO 10 (Lonza, Switzerland), the following steps were performed in duplicate using each media. Approximately 1×10^6 cells were transferred into a 75 cm^2 flask and cultured for 24 hours in 10 ml media. 10 µg of pEF-CFHR1 was diluted in 400 µl media, mixed with 100 µl of 1 mg/ml branched polyethyleneimine (PEI) and incubated at 298 K for 15 minutes. The media in the flask was replaced with 9 ml fresh media before the DNA:PEI mixture was added and incubated for 8 hours. Following incubation the transfection media was replaced again with 10 ml fresh media, and the flask incubated for four days.

In order to assess whether expression of CFHR-1 was successful, the culture supernatant was collected and concentrated down to 1ml prior to detection of the C-terminal hexahistidine tag via Western blotting (Figure 7.7). This shows that expression and secretion of CFHR-1 has been successful in HEK 293T cells in both types of media, resulting in a doublet in the blot which closely matches the doublet observed in material

purified from plasma, corresponding to two different glycosylation states. In order to generate larger quantities of protein for characterisation, an attempt was made to scale-up the transient transfection using roller-bottles containing 500 ml media. Unfortunately, no trace of CFHR-1 expression could be detected in these media which is possibly due to a low transfection efficiency in the larger culture volume. These trial experiments show that it is possible to produce CFHR-1 using transient expression in mammalian cell culture. However, in order to generate large quantities of protein it is likely that a stable cell line will have to be produced. Unfortunately, the timescale for producing and testing this cell line was outside that of this thesis.

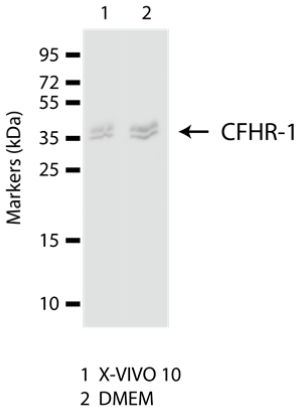


Figure 7.7 α-hexahistidine (Qiagen, USA) Western blot of the HEK 293T transient transfection culture supernatant from two different media.

7.1.5 CFHR-1_{12} expression and purification

pKLAC-CFHR1_{12} was linearised using SacII and transformed into *K. lactis* strain GG799 as per the manufacturer's instructions. After double selection on YCB/agar containing 5 mM acetamide, as described in Chapter 7.1.4, positive colonies were grown overnight in 10 ml MMGal at 30°C with shaking. Each overnight culture was used to inoculate 1 L MMGal prior to incubation at 30°C, with 180 RPM shaking. Secreted proteins in the culture supernatant were monitored daily

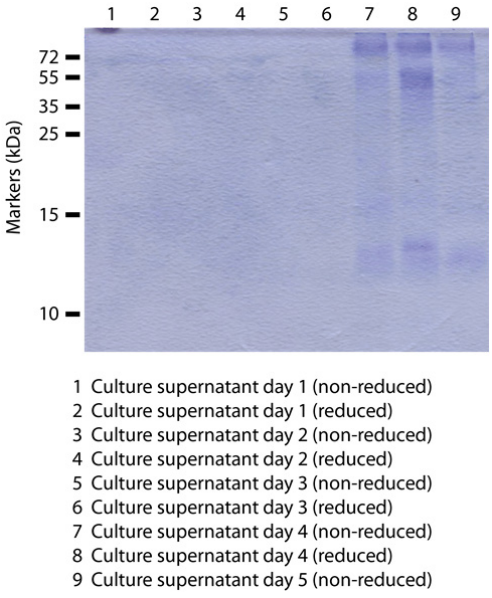


Figure 7.8 15% (w/v) SDS-PAGE gel (reducing and non-reducing) of culture supernant from *Kluyveromyces lactis* expressing CFHR-1_{12} at the given time points.

using SDS-PAGE (Figure 7.8) showing the existence of a band of the correct molecular weight (~14 kDa) that shifts to a higher molecular weight upon addition of reducing agent; indicative of the presence of disulphide bonds.

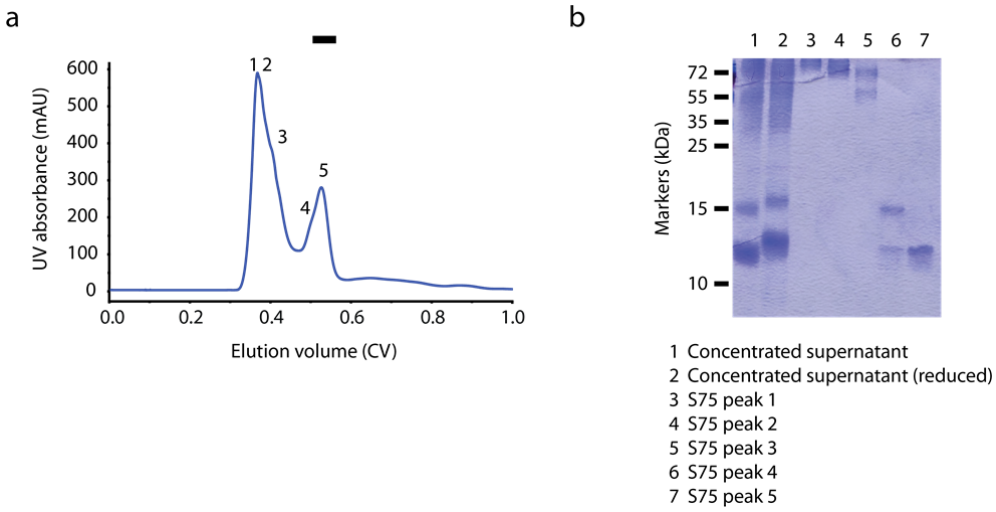


Figure 7.9 *Panel a*: Chromatogram detailing the elution profile of CFHR-1_{12} from an S75 16/60 column. The peak containing the target protein elutes at approximately 0.53 CV and the fractions collected are marked with a black bar. Peaks are also labelled for identification. *Panel b*: 15% (w/v) reducing and non-reducing SDS-PAGE gel of each peak eluted from the S75 column.

The culture supernatant was cleared by centrifugation (4000 g, 4^{o}C, 20 minutes) before being made 1 mM with respect to EDTA and PMSF (from a 1 M stock in DMSO) to inhibit proteases. Vivaflow blocks (10000 MWCO. Millipore, UK) were used to concentrate the sample to 100 ml and two EDTA-free protease inhibitor cocktail tablets (Roche, Germany) were added before dialysis against 50 mM Tris, 150 mM NaCl, pH 7.5 in Snakeskin tubing (3500 MWCO. Pierce. UK). The dialysed sample was concentrated to 10ml using an Ultracel 10,000 MWCO spin concentrator (Millipore, UK) before filtration and purification on an S75 26/60 size exclusion column pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 7.5 (Figure 7.9, panel a). The peak at 0.53 CV was collected and analysed using SDS-PAGE showing the formation of

species with the correct mass (~14.4 kDa) that shifts upon reducing suggesting the formation of disulphide bonds (Figure 7.9, panel b).

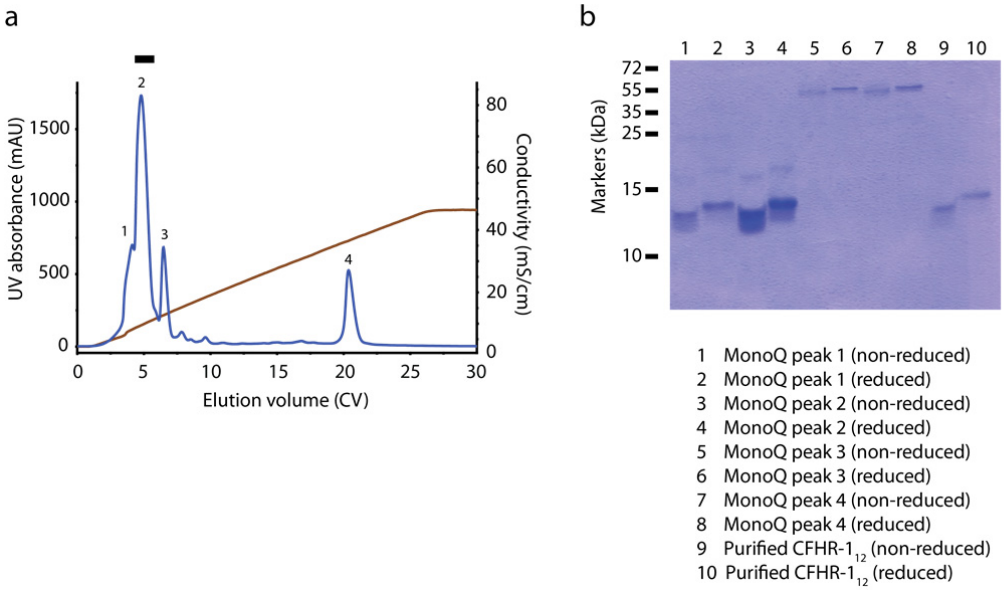


Figure 7.10 *Panel a*: Elution profile from the MonoQ 5/50 column eluted with a salt gradient. The UV absorbance and conductivity traces are shown in blue and brown respectively. Each peak is labelled for analysis and the collected fractions containing CFHR-1$_{12}$ are highlighted with a black bar. *Panel b*: 15% (w/v) SDS-PAGE gel of the elution peaks in the presence of absence of reducing agent.

In order to obtain a higher level of purity, the fractions containing CFHR-1$_{12}$ were diluted 1:10 into 25 mM Tris, pH 7.5 and loaded on a MonoQ 5/50 column (GE Healthcare, USA) pre-equilibrated in 25 mM Tris, pH 7.5. After washing with 15 CV 25 mM Tris, 10 mM NaCl, pH 7.5, the column was eluted with a 10 – 500 mM gradient NaCl over 25 CV. Fractions were collected and analysed on an SDS-PAGE gel (Figure 7.10, panels a and b) showing purification of CFHR-1$_{12}$ in peaks 1 and 2, eluting at 8.2 mS/cm and 9.5 mS/cm respectively. Analysis using ES-TOF mass spectrometry gave a mass of 14374 Da, which corresponds well to the expected mass of 14376 Da, and shows the absence of glycosylation. Typical yields were 1 mg/l culture (assuming ε_{280} = 25214 M^{-1} cm^{-1} calculated using ProtParam [Gasteiger E. 2005]).

7.1.6 CFHR-1_{345} and CFHR-2_{34} expression and purification

Refolding conditions for CFHR-1_{345} and CFHR-2_{34} were developed, under close supervision, by Mr. Matthew Colledge (Biochemistry Part II student. Lea Group. 2011-2012). A variety of conditions were trialled in order to obtain the highest yield of soluble, folded protein, arriving at the successful protocol described below.

BL21(DE3) *E. coli* cells transformed with pET-CFHR1_{345} or pET-CFHR2_{34} were grown overnight in 4 × 10 ml Luria-Bertani broth supplemented with 100 µg/ml ampicillin at 37^{o}C and 180 RPM shaking. Saturated cultures were used to inoculate four litres of LB broth containing 100 µg/ml ampicillin which were grown at 37^{o}C with 180 RPM shaking. The cells were induced with 1 mM IPTG when OD_{600} reached 0.6 and incubated for a further three hours. Cells were harvested by centrifugation (4000 g, 20 minutes, 4^{o}C) and the cell pellets stored at -20^{o}C until required. Proteins were refolded and purified in an identical manner to fH_{67} (Chapter 2.1.3) substituting the refold buffer for 1 mM cysteine, 2 mM cystine, 20 mM ethanolamine, 1 mM EDTA, pH 11.0. Example size exclusion elution profiles are shown in Figure 7.11.

Purity was assessed by SDS-PAGE (Figure 7.11) and ES-TOF mass spectrometry returned masses of 21435 Da and 14278 Da corresponding to the expected masses of 21432 Da and 14280 Da for CFHR-1_{345} and CFHR-2_{34} respectively. SDS-PAGE gels charting the purifications, alongside example traces from the final size exclusion chromatography steps, are displayed in Figure 7.11. Typical yields of correctly folded protein are approximately 2 and 3 mg/l of bacterial culture assuming A_{280} = 1.756 and 1.535 for CFHR-1_{345} and CFHR-2_{34} respectively.

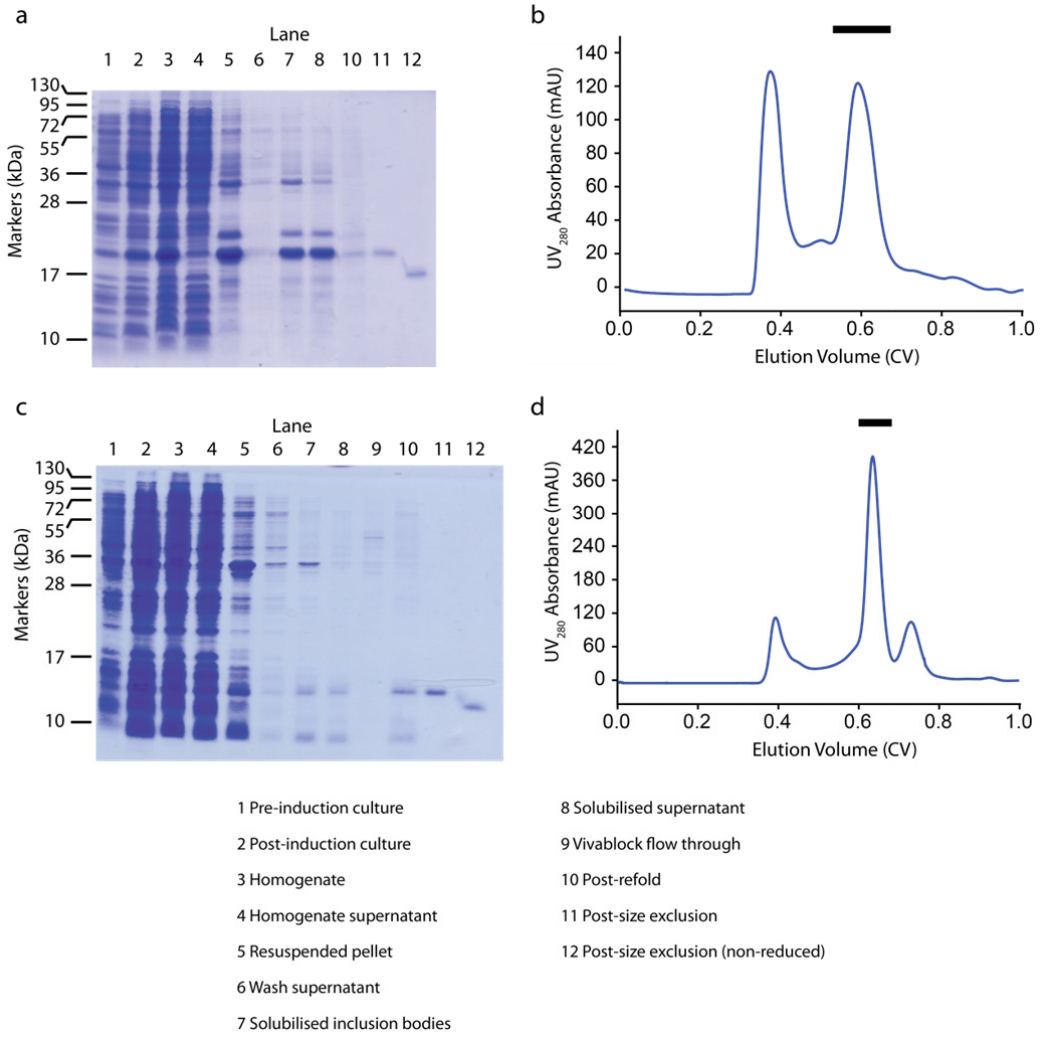


Figure 7.11 *Panel a*: 15% (w.v) reducing SDS-PAGE gel charting the preparation of CFHR-1_{345}. *Panel b*: Chromatogram detailing the elution profile of CFHR-1_{345} from an S75 16/60 column. The target protein elutes at approximately 0.58 CV in the peak highlighted with a black bar. *Panel c*: 15% (w/v) reducing SDS-PAGE gel charting the preparation of CFHR-2_{34}. *Panel d*: Chromatogram detailing the elution profile of CFHR-2_{34} from an S75 16/60 column. The target protein elutes at approximately 0.63 CV in the peak highlighted with a black bar.

7.1.7 Generation and testing of a monoclonal antibody against CFHR-2_{34}

Antibodies have been raised against the common N-terminus of CFHR-1, CFHR-2 and CFHR-5, the C-terminus of CFHR-1 and the C-terminus of CFHR-5. However, there are currently no specific antibodies against CFHR-2. Such a reagent would be highly useful as it would allow full quantification of the serum abundance of CFHR-2 and may be useful in determination of the mechanism of action of the CFHRs.

In order to address this need, monoclonal antibodies were raised against purified CFHR-2 by Dr. Paul Morgan and Dr. Claire Harris (University of Cardiff, UK). Following solid phase screening to discount any antibodies with cross reactivity towards factor H, tissue culture supernatant containing the most promising candidate was received and the antibody captured on a 1 ml NHS-sepharose column (GE Healthcare, USA) that had CFHR-2_{34} immobilised as per the manufacturers instructions. Following washing with 20 CV of 50 mM Tris, 150 mM NaCl, pH 7.5, the purified antibody was eluted in 4ml of 4 M $MgCl_2$.

The specificity of the purified α-CFHR-2_{34} antibody (determined to belong to the IgG1 isotype) was assessed in a dot blot (Figure 7.12), performed using standard protocols, to compare detection of CFHR-2_{34} and CFHR-1_{345}, the latter of which is almost identical to $fH_{18\text{-}20}$ (Figure 7.1, panel a). This clearly shows that the α-CFHR-2_{34} antibody is specific for CFHR-2_{34}, displaying no cross reactivity to CFHR-1_{345}, and likely none to factor H domains 18 through 20, at these primary antibody dilutions.

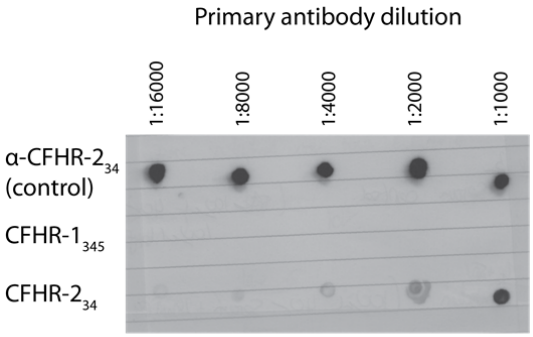


Figure 7.12 Dot blot with the α-CFHR-2_{34} antibody. The purified antibody, CFHR-1_{345} and CFHR-2_{34} were immobilised on nitrocellulose membranes prior to incubation in different dilutions of α-CFHR-2_{34} and secondary detection with a horseradish peroxidase conjugated goat anti-mouse IgG antibody.

7.2 Crystallisation, data collection and processing of CFHR-1_{12}

In order to investigate the role of the N-terminal domains of CFHR-1, CFHR-2 and CFHR-5, and compare them to factor H domains six and seven, attempts were made to obtain the atomic resolution structures of the N-terminal domains of CFHR-1 using X-ray crystallography.

7.2.1 Crystallisation of CFHR-1_{12}

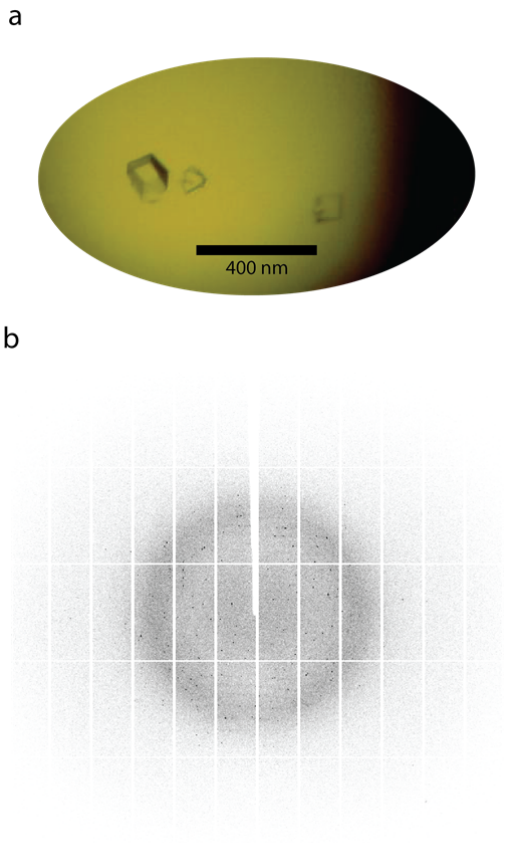


Figure 7.13 *Panel a*: Crystals of CFHR-1_{12}. *Panel b*: Example diffraction image.

CFHR-1_{12} was concentrated to A_{280} = 3.3 prior to crystallisation trials performed at 294K using vapour diffusion in 400 nL sitting drops produced by an Oryx Nano crystallisation robot (Douglas Instruments, UK) within MRC crystallisation plates (Molecular Dimensions, UK). Stura, Macrosol, Pact Premier, JCSG, Structure 1 and Structure 2 screens (Molecular Dimensions, UK) were used, with two drops set up for each condition containing protein to precipitant ratios of 1:1 and 1:3.

Crystals grew in 36% (w/v) PEG 2000 MME, 0.1 M MES pH 6.5 after approximately ten days (Figure 7.13, panel a), which were mounted in Cryoloops (Hampton research, USA) prior to cryoprotection with 20% ethylene glycol and flash cooling in liquid nitrogen.

7.2.2 Data collection of CFHR-1_{12}

Diffraction data were collected on beamline ID29 at the ESRF using the rotation method (example diffraction image in Figure 7.13, panel b). Crystals were indexed in a primitive orthorhombic cell (cell dimensions a = 45.3 Å, b = 46.9 Å, c = 111.7 Å, $\alpha = \beta = \gamma = 90^{o}$) with P222 symmetry prior to strategizing using EDNA [Incardona,

Bourenkov *et al.* 2009]. Two datasets were collected, a native and a S-SAD dataset at long wavelength (λ = 1.7 Å). Both were collected using the parameters that are presented in Table 7.3.

	Native	S-SAD
Diffraction source	ESRF ID29	ESRF ID29
Detector	PILATUS 6M	PILATUS 6M
Energy (eV)	7248	7248
Wavelength (Å)	1.7105	1.7105
Temperature (K)	120	120
Detector resolution (Å)	2.0	2.4
Exposure per image (s)	0.2	0.2
Φ_{start}	275	323
Φ_{end}	95	143
$\Delta\Phi$	0.2	0.2
Number Images	900	900
f′ (electrons)	-	0.34
f″ (electrons)	-	0.68

Table 7.3 Data collection parameters.

7.2.3 Data processing of CFHR-1_{12}

Data were processed using XIA2 [Winter 2010] invoking the 3dii option to enforce integration with XDS [Kabsch 2010] and scaling with XSCALE [Kabsch 2010]. Analysis of systematic absences using POINTLESS [Evans 2006] confirms the existence of two-fold screw axes along h, k and l suggesting a revision of the spacegroup to $P2_12_12_1$ (Figure 7.14). However, anomalous correlations within the S-SAD dataset are very weak and may only be present at low resolution. Data processing statistics are presented in Table 7.4.

	Native	S-SAD
Spacegroup	$P2_12_12_1$	$P2_12_12_1$
Unit-cell dimensions (Å)	a=45.3, b=46.9 c=111.7	a=45.1, b=46.8, c=111.8
Unit-cell angles (°)	α=90, β=90.0, γ=90.0	α=90, β=90.0, γ=90.0
Wilson B factor	32.6	58.6
Resolution (Å)	1.99-55.83 (1.99-2.04)	2.39-111.78 (2.39-2.46)
$\mathbf{R_{merge}}$	0.09 (0.54)	0.09 (0.58)
$\mathbf{R_{pim}}$	0.039 (0.224)	0.059 (0.383)*
<I/σ(I)>	11.2 (2.9)	10.0 (2.4)
Completeness (%)	96.6 (90.6)	96.8 (97.7)*
Multiplicity	6.2 (6.4)	3.2 (3.2)*
Anomalous correlations	-	0.05 (0.02)

Table 7.4 Data processing statistics. Values of R_{pim}, completeness and multiplicity marked with * represent the anomalous scattering statistics R_{pim}(I+/I-), anomalous completeness and anomalous multiplicity respectively.

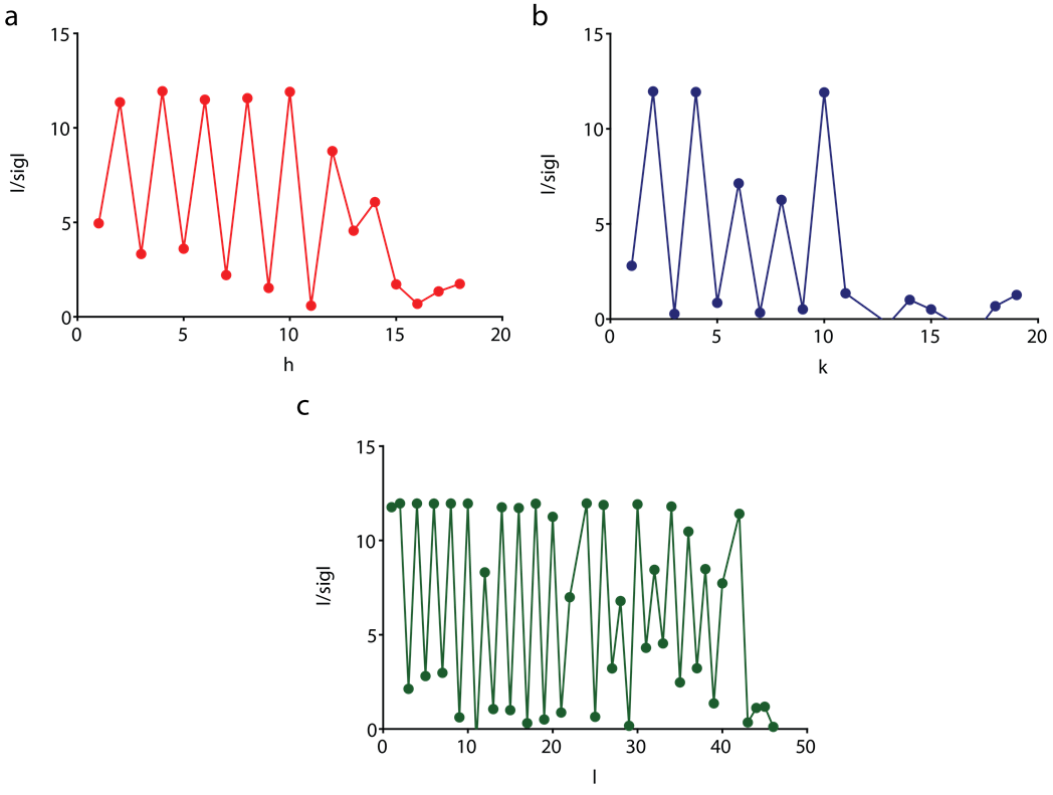


Figure 7.14 *Panel a*: Systematic absences amongst reflections with a given value of h. The lack of odd values of h suggest a screw axis parallel to a. *Panel b*: Screw axis parallel to b. *Panel c*: Screw axis parallel to c.

7.3 Structure determination and refinement of CFHR-1_{12}

Due to the high levels of sequence similarity between fH_{67} and CFHR-1_{12}, molecular replacement combined with molecular modelling was employed to solve the structure. Anomalous signal from the S-SAD experiment was used to validate the resulting model showing correct placement of the disulphide bonds.

7.3.1 Structure solution of CFHR-1_{12}

Analysis of the unit cell contents using MATTHEWS [Matthews 1968] suggested that there are two copies of CFHR-1_{12} in the asymmetric unit (solvent content = 40.6%). However, there were no significant non-origin peaks in the native Patterson or non-crystallographic rotation peaks in the self-rotation function (Figure 7.15). This could arise from only one copy in the unit cell, non-identical molecules in the asymmetric unit, or a two-fold symmetry axis close to one of the crystallographic screw axes.

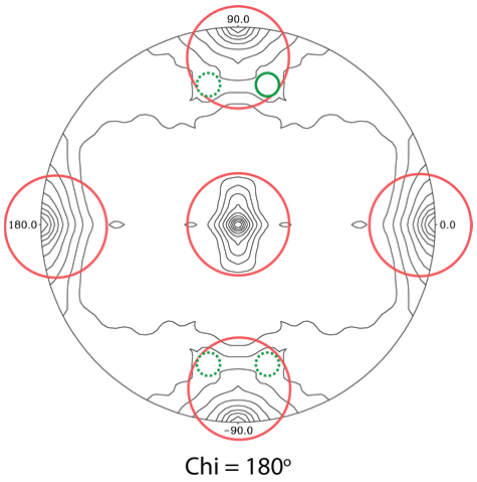


Figure 7.15 Self-rotation function calculated from the native dataset using MOLREP [Vagin and Teplyakov 1997]. Peaks resulting from crystallographic symmetry in spacegroup $P2_12_12_1$ are highlighted in red. The peak resulting from non-crystallographic symmetry and its combination with the crystallographic symmetry operations are shown in green.

MR_ROSETTA [Terwilliger, Dimaio *et al.* 2012] uses PHASER [McCoy, Grosse-Kunstleve *et al.* 2007] to place an initial molecular replacement model. Following placement, the model is rebuilt and improved using homology modelling by ROSETTA [DiMaio, Tyka *et al.* 2009] prior to final model building with PHENIX.AUTOBUILD [Terwilliger, Grosse-Kunstleve *et al.* 2008] and refinement with PHENIX.REFINE [Adams, Afonine *et al.* 2010].

	RFZ	TFZ	LLG
Copy 1	6.2	7.6	55
Copy 2	3.7	8.1	107

Table 7.5 Rotation function (RFZ) Z-scores, translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of two copies of the model in the data.

Using the structure of factor H domains six and seven (taken from the structure of fH_{678} in PDB entry: 2UWN [Prosser, Johnson *et al.* 2007b]), having 42% and 34% sequence identity to domains one and two of CFHR-1 respectively, as a model, the PHASER sub-process of MR_ROSETTA successfully placed two copies in the CFHR-1_{12} native dataset (solutions described in Table 7.5). Following rebuilding with the ROSETTA sub-process the model was improved, indicated by an increase in the log-likelihood gain to 204. The final model following autobuilding and refinement possessed two complete chains of CFHR-1_{12} with an R/R_{free} of 0.29/0.37.

7.3.2 S-SAD validation of the CFHR-1_{12} structure

The model of CFHR-1_{12} built by MR_ROSETTA was placed back into the S-SAD data using PHASER in order to generate phases described by Hendrickson-Lattmann coefficients. These phases were combined with the S-SAD anomalous data using CAD [Winn, Ballard *et al.* 2011] and used as external phases to generate an anomalous residual map in SHARP [Bricogne, Vonrhein *et al.* 2003]. Positive residuals form around the disulphide bonds in the model validating their placement and the correctness of the structure (Figure 7.16).

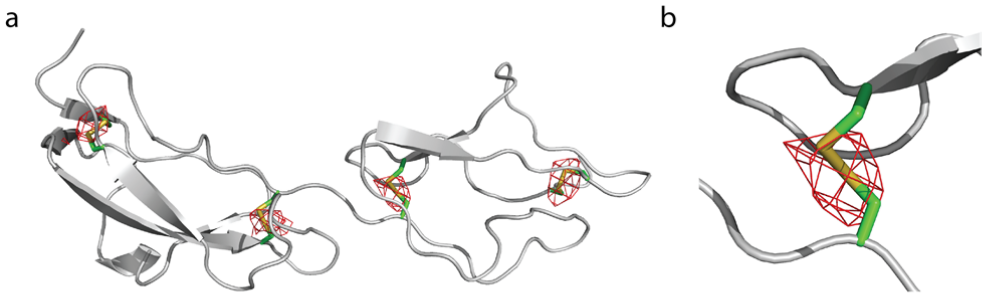


Figure 7.16 *Panel a*: Residuals around the sulphur sites are shown in red (contoured at 2 σ), displaying four sites which correspond to the disulphide bonds within the built structure. *Panel b*: Close-up of one of the residuals showing the location of the sulphur atoms within the mesh.

7.3.3 Refinement of CFHR-1_{12}

The validated model from MR_ROSETTA was refined iteratively between rounds of rebuilding in COOT [Emsley, Lohkamp *et al.* 2010] and refinement using AUTOBUSTER [Blanc, Roversi *et al.* 2004; Bricogne, Blanc *et al.* 2011] invoking usage of NCS restraints with the autoncs flag [Smart, Brandl *et al.* 2008]. During later refinement cycles the TLSbasic macro was used to carry out refinement defining each chain as a TLS group. Using this refinement technique models anisotropic displacement of rigid groups of residues in the structure. The final model, shown alongside representative electron density in Figure 7.17 had an R/R_{free} of 0.22/0.25 and was validated using the MOLPROBITY server [Chen.2010]. Refinement and validation statistics are shown in Table 7.6.

	CFHR-1_{12}
Resolution (A)	1.99-55.83 (1.99-2.13)
Number of reflections	16261 (2724)
Number of reflections for R_{free}	818 (126)
R_{work}/R_{free}	0.2222/0.2465 (0.2195/0.2552)
Number of atoms	
Protein	1973
Ligand/Ion	166 Ethylene glycol
Water	102
R.m.s deviations	
Bond lengths (A)	0.008
Bond angles (°)	0.98
Ramachandran plot analysis	
Residues in favoured regions (%)	93.4
Residues in disallowed regions (%)	0.4
Molprobity analysis	
Molprobity score	1.92
Molprobity score percentile	79th
Resolution range (A)	1.74-2.24

Table 7.6 Refinement and validation statistics for the final model of CFHR-1_{12}.

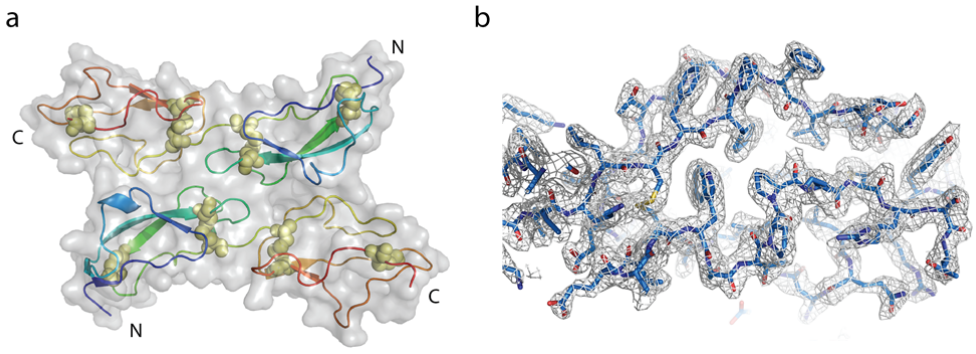


Figure 7.17 *Panel a*: Final model of CFHR-1_{12} showing both copies in the asymmetric unit as a cartoon of the secondary structure within a transparent surface rendering. The cysteine residues forming intra-molecular disulphide bonds are shown as gold spheres. *Panel b*: Representative electron density (2Fo-Fc map contoured at 1.5σ) with the refined structure shown in blue for clarity.

7.4 Structural analysis of CFHR-1_{12}

The structure of CFHR-1_{12} shows that the protein expressed in *K. lactis* has been secreted in a correctly folded form with the correct pattern of intramolecular disulphide bonds, the position of which were verified using S-SAD data. The asymmetric unit contains two identical copies of CFHR-1_{12} forming a dimer mediated by an interaction surface along the length of the molecule resulting in an antiparallel arrangement (Figure 7.18, panel a). Analysis of the dimerization interface using PISA [Krissinel and Henrick 2007] (Table 7.7) gives a complexation significance score (CSS) of 1.0, meaning that there is a very high likelihood that formation of the dimer is physiologically significant. By extension, it is highly likely that the dimerization interface observed in the crystal structure is responsible for formation of the dimer in solution observed using MALS.

	Dimerisation interface
CSS	1.000
Surface area (Å²)	749.0
Δ^{i}G (kcal/mol)	-13.6
Δ^{i}G P-value	0.05

Table 7.7 PISA analysis of the dimerisation interface.

Scrutiny of the interface shows that it is mostly hydrophobic in nature (Figure 7.18, panel b) resulting in a negative free energy of solvation, the high magnitude of which is indicative of a strong interaction. The hydrophobic interface is further strengthened by formation of four hydrogen bonds between the subunits. As the dimerization interface is two-fold symmetric, resulting from the antiparallel arrangement of the subunits, all of the interactions are duplicated resulting in a synergic relationship between the molecules. Mapping of the residues that differ between domains one and two of CFHR-1, CFHR-2 and CFHR-5 onto the structure of CFHR-1_{12} (Figure 7.19, panel a) shows that the binding surface is very highly conserved (99% and 88% identity to CFHR-1 for CFHR-2 and CFHR-5 respectively) suggesting that CFHR-2 and CFHR-5 can also dimerise in an identical manner. The high levels of

identity also suggest that hetero-dimers may be able to be formed between CFHR-1, CFHR-2 and CFHR-5.

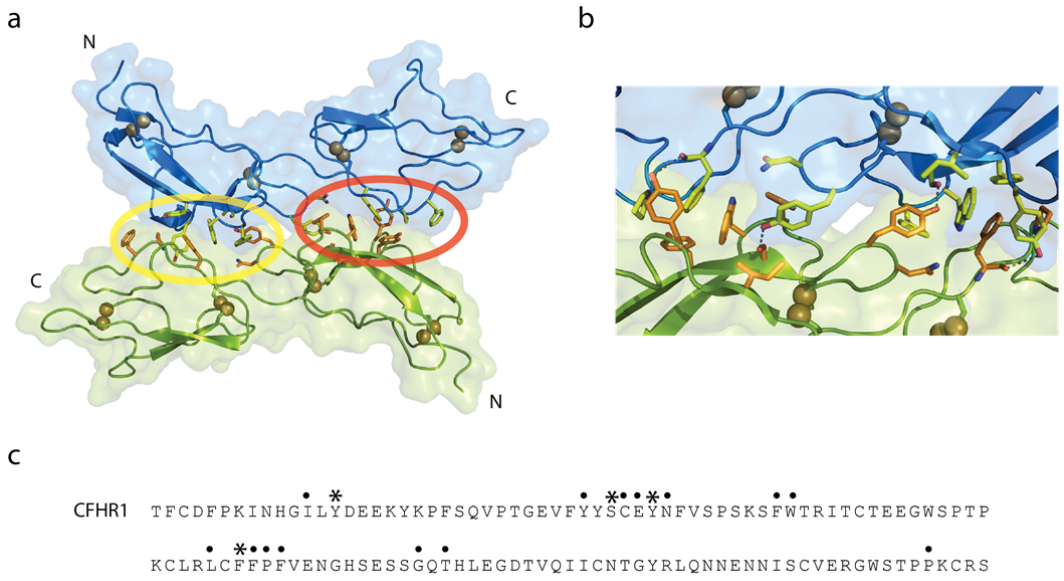


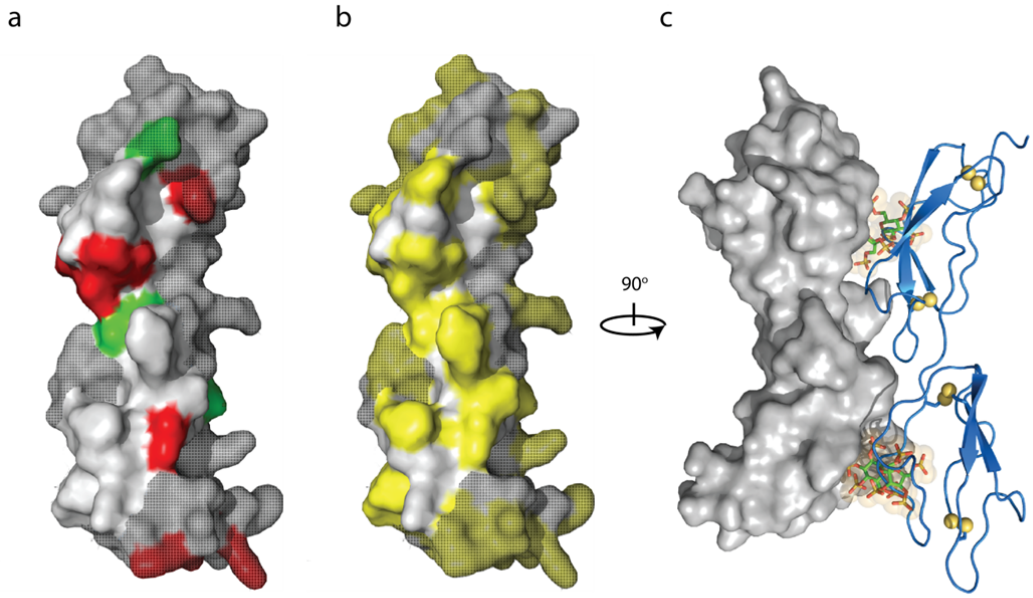
Figure 7.18 *Panel a*: Representation of the CFHR-1_{12} dimer with chain A in blue and chain B in green. Due to the anti-parallel arrangement of the homodimeric species, the binding site can be split into two identical regions, shown in yellow and orange, related by 2-fold symmetry. Thus each interaction is duplicated in a synergic receptor-donor donor-receptor type arrangement. *Panel b*: Close-up of the dimerisation interface showing intermolecular hydrogen bond formation and the highly hydrophobic interaction surface. *Panel c*: Sequence of CFHR-1_{12} highlighting the residues involved in the interface , as determined using PISA. Those residues involved in intermolecular hydrogen bonding are marked with a star.

Mapping the sequence conservation between CFHR-1_{12} and fH_{67} onto the structure of the former (Figure 7.19, panel b) shows that the dimerization interface is not conserved in factor H. Superposition of the structure of CFHR-1_{12} onto factor H domains six and seven in complex with the GAG analogue sucrose octasulphate (PDB entry: 2UWN [Prosser, Johnson *et al.* 2007a]) shows that the heparin binding site has evolved to form the dimerization interface (Figure 7.19, panel c) and accounts for the observations by [Heinen, Hartmann *et al.* 2009] that suggest that the two N-terminal domains of CFHR-1 do not possess heparin binding properties as the charged residues that promote these interactions are not conserved.

Calculation of the rotational transformation, in spherical polar coordinates, between both CFHR-1_{12} molecules (Table 7.8) allows assignment of the corresponding peak in the self-rotation function (shown in green in Figure 7.15) which is weak and overlaps with the crystallographic peak at $\omega = 90^o$, $\varphi = 90^o$, $\chi = 180^o$.

	ω	φ	χ
Chain A - Chain B	112.9	77.4	179.5

Table 7.8 Rotational transformation between chains A and B in the CFHR-1_{12} asymmetric unit given in spherical polar coordinates.



e

```
CFHR-1   TFCDFPKINHGILYDEEKYKPFSQVPTGEVFYYSCEYNFVSPSKSFWTRITCTEEGWSPTP
CFHR-2   MFCDFPKINHGILYDEEKYKPFSQVPTGEVFYYSCEYNFVSPSKSFWTRITCAEEGWSPTP
CFHR-5   TLCDFPKIHHGFLYDEEDYNPFSQVPTGEVFYYSCEYNFVSPSKSFWTRITCTEEGWSPTP
fH67     KPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYYCDEHFETPSGSYWDHIHCTQDGWSPAV

CFHR-1   KCLRLCFFPFVENGHSESSGQTHLEGDTVQIICNTGYRLQNNENNISCVERGWSTPPKCRS
CFHR-2   KCLRLCFFPFVENGHSESSGQTHLEGDTVQIICNTGYRLQNNENNISCVERGWSTPPKCRS
CFHR-5   KCLRMCSFPFVKNGHSESSGLIHLEGDTVQIICNTGYSLQNNEKNISCVERGWSTPPICSF
fH67     PCLRKCYFPYLENGYNQNHGRKFVQGKSIDVACHPGYALPKAQTTVTCMENGWSPTPRCIR
```

Figure 7.19 *Panel a*: Surface representation of chain A from the CFHR-1_{12} dimer with the interaction surface shown in light grey. Sequence differences between these domains of CFHR-1, CFHR-2 and CFHR-5 are coloured green (minor differences) and red (major differences). *Panel* b: The same representation is also shown with the differences between CFHR-1_{12} and fH_{67} (yellow). *Panel c*: CFHR-1_{12} dimer with the structure of fH_{67} in complex with the GAG analogue SOS superimposed over chain B. *Panel d*: Sequence alignment of the N-terminal domains of CFHR-1, CFHR-2, CFHR-5 and fH_{67} with differences coloured as per panel a.

7.5 Investigations into the role of dimerization by the CFHRs

In order to validate that the interface observed in the crystal structure is physiologically relevant, CFHR-1_{12} was analysed in solution to confirm that it forms a homodimer. Analysis of full length CFHR-1 and experiments probing the existence of heterodimeric species between CFHR-1 and CFHR-2 or CFHR-5 have been performed. Mutagenesis to ablate dimerization has been undertaken and the effects of monomerisation studied *in vivo*.

7.5.1 MALS analysis of CFHR-1_{12}

MALS was employed to show that CFHR-1_{12} exists as a dimer in solution. Analysis of the purified material gives an average mass of approximately 30 kDa which is consistent with dimer formation (2 × 14.4 kDa subunits) under these conditions (Figure 7.20).

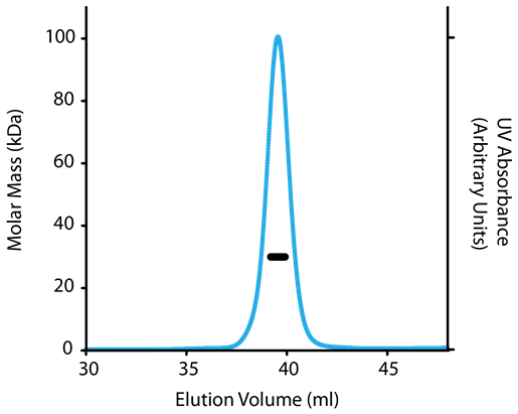


Figure 7.20 MALS analysis of CFHR-1_{12}.

7.5.2 MALS analysis of CFHR heterodimers

In order to investigate whether it is possible to form heterodimeric species between CFHR-1, CFHR-2 and CFHR-5, MALS was employed to analyse a sample of CFHR-1 that was purified from human serum using immunoprecipitation and kindly provided by Dr. Claire Harris (Cardiff University, UK). Following purification this material was found to be contaminated with CFHR-2 and CFHR-5 which could not be removed by affinity or ion exchange chromatography (Figure 7.21, panel a (inset)). A sample was subjected to analysis using MALS giving a single peak with a range of masses between 65 kDa and 80 kDa (Figure 7.21, panel a). This mass range can be explained by combination of a single CFHR-1 molecule with either another CFHR-1, CFHR-2 or

CFHR-5 (Figure 7.21, panel b). No masses were observed that are lighter than the smallest dimeric species (CFHR-1/CFHR-2) or heavier than the largest (CFHR-1/CFHR-5), showing that these CFHRs all exist as homo or heterodimers under these conditions.

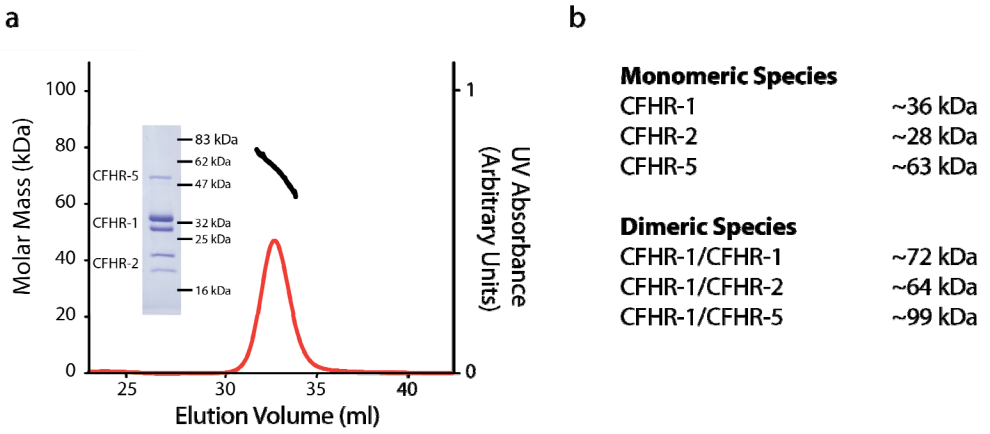


Figure 7.21 *Panel a*: MALS analysis of CFHR-1 purified from serum. Inset: 15% (w/v) SDS-PAGE analysis of the species present (labelled). *Panel b*: Masses of the possible species present.

7.5.3 Design of dimerization knockout mutants

It is expected that dimerization will increase the affinity of the CFHRs for their ligands via the avidity effect. In order to investigate the importance of avidity, mutations have been introduced into the dimerization interface that prevent assembly and their effects studied *in vivo*.

Analysis of the dimerization interface in the structure of CFHR-1_{12} shows the existence of four intermolecular hydrogen bonds, of which two are unique due to the two-fold symmetry of the binding site, highlighted in Figure 7.22. Comparison of the sequences of CFHR-1, CFHR-2 and CFHR-5 (Figure 7.22, panel d) reveals that the Ser36 and Tyr39 are conserved across all three proteins. A second hydrogen bond is formed between Tyr16 and the peptide carbonyl of Phe70. However, Tyr16 is conserved in fH_{67} and Phe70 is a serine residue in CFHR-5 suggesting that this hydrogen bond is not critical to the dimerization interaction. PISA [Krissinel and Henrick 2007] also

highlights another conserved residue, Tyr34, as having an important role in the interface with a large buried surface area (73 Å^2) and free energy of solvation (0.7 kcal/mol). Neither Ser36, Tyr39 nor Tyr34 are conserved in fH_{67}, which is known to be monomeric suggesting that mutating them to their equivalents in fH_{67} will ablate dimerization. Thus domains one and two of either CFHR-1, CFHR-2 or CFHR-5 containing the mutations Ser36Tyr, Tyr39Glu and Tyr34Ser are expected to be monomeric. These mutations should also have the added benefit of minimal perturbation to the folding and solubility of the protein due to the similarity between both stable molecules.

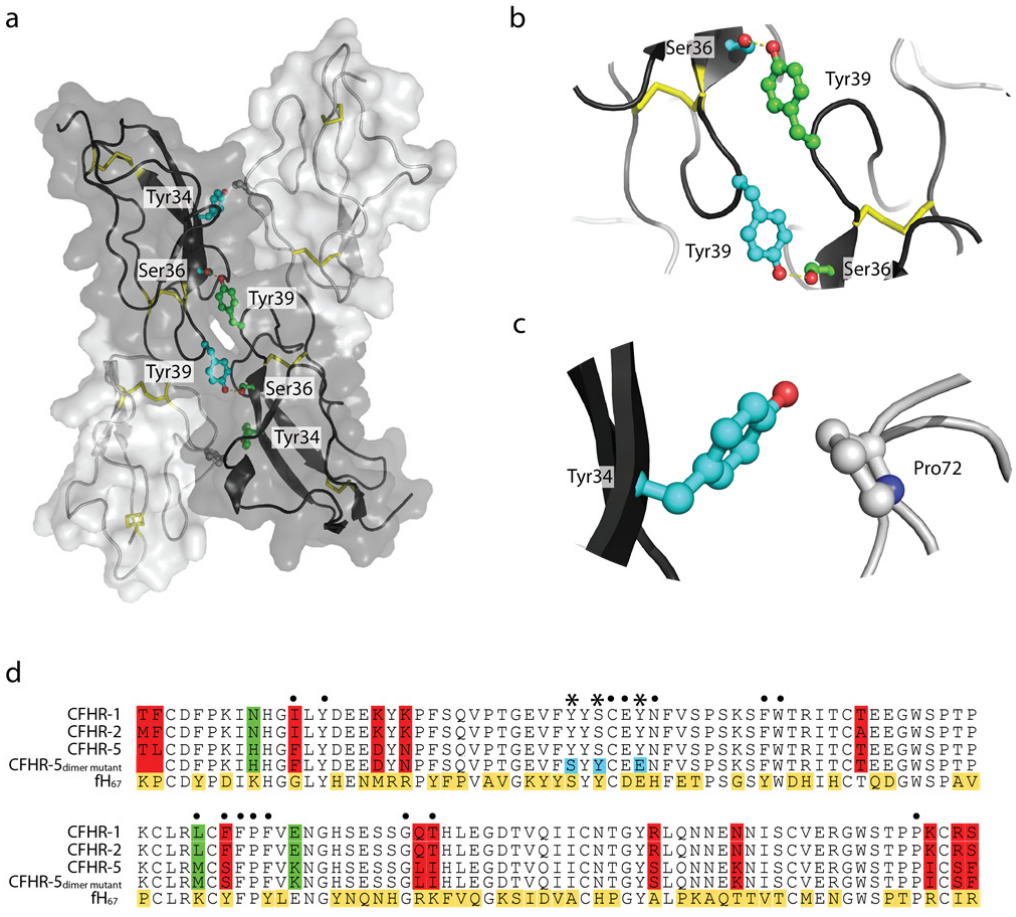


Figure 7.22 *Panel a*: Representation of the CFHR-1_{12} dimer showing the residues determined to b ecritical for dimerisation. Domain 1 is coloured black and domain 2 in grey. *Panel b*: Close-up of the intermolecular hydrogen bonds formed between Ser36 and Tyr39 in both molecules. *Panel c*: Close-up of the hydrophobic interaction between Tyr34 and Pro72 in the opposing molecule.

An attempt to test monomerisation of CFHR-1_{12} was performed by introduction of the mutations Ser36Tyr, Tyr39Glu and Tyr34Ser into pKLAC-CFHR1_{12} followed by expression in *Kluyveromyces lactis* as described for CFHR-1_{12} in Chapter 7.1.5. Unfortunately, the protein was degraded in the culture supernatant, possibly by proteases from the yeast. This suggests that the protein containing the mutations is not folding properly in this expression system allowing proteases to digest the peptide chain.

7.5.4 Effect of dimerization upon CFHR-5 binding to C3 *in vivo*

In order to fully assess the effects of introduction of the monomerisation mutants they were introduced into CFHR-5 and the mutant protein CFHR-$5_{\text{dimer mutant}}$ produced in mammalian cell culture (Figure 7.23, panel a) by Dr. Elena Goicoechea de Jorge (Imperial College, London) who then performed *in vivo* assays to probe the ability of CFHR-5 and CFHR-$5_{\text{dimer mutant}}$ to bind deposited C3. In order to confirm that the mutations confer monomerisation, the wild-type and mutant proteins were analysed by blue native PAGE and analytical gel filtration (Figure 7.23, panels b and c) showing monomer in the mutant sample and mainly dimer in the wild type.

Factor H deficient mice ($fH^{-/-}$) mice develop extensive deposition of activated C3 along the glomerular basement membrane in the kidney. 30µg of CFHR-5 or CFHR-$5_{\text{dimer mutant}}$ were injected into the mice and the levels bound to C3 in the kidney were observed after three hours by staining a histological section with an anti-CFHR-5 polyclonal antibody (Figure 7.23, panel d). The levels of deposited C3 were also assayed with an anti-C3 antibody. These findings clearly show that dimerization critically influences *in vivo* binding of CFHR-5 to deposited C3b, reducing the localised levels to barely detectable levels following monomerisation. Further experiments have also shown that

the interaction is both dose dependent and C3 dependent, confirming the result and proving that CFHR-5 is binding to the deposited C3b.

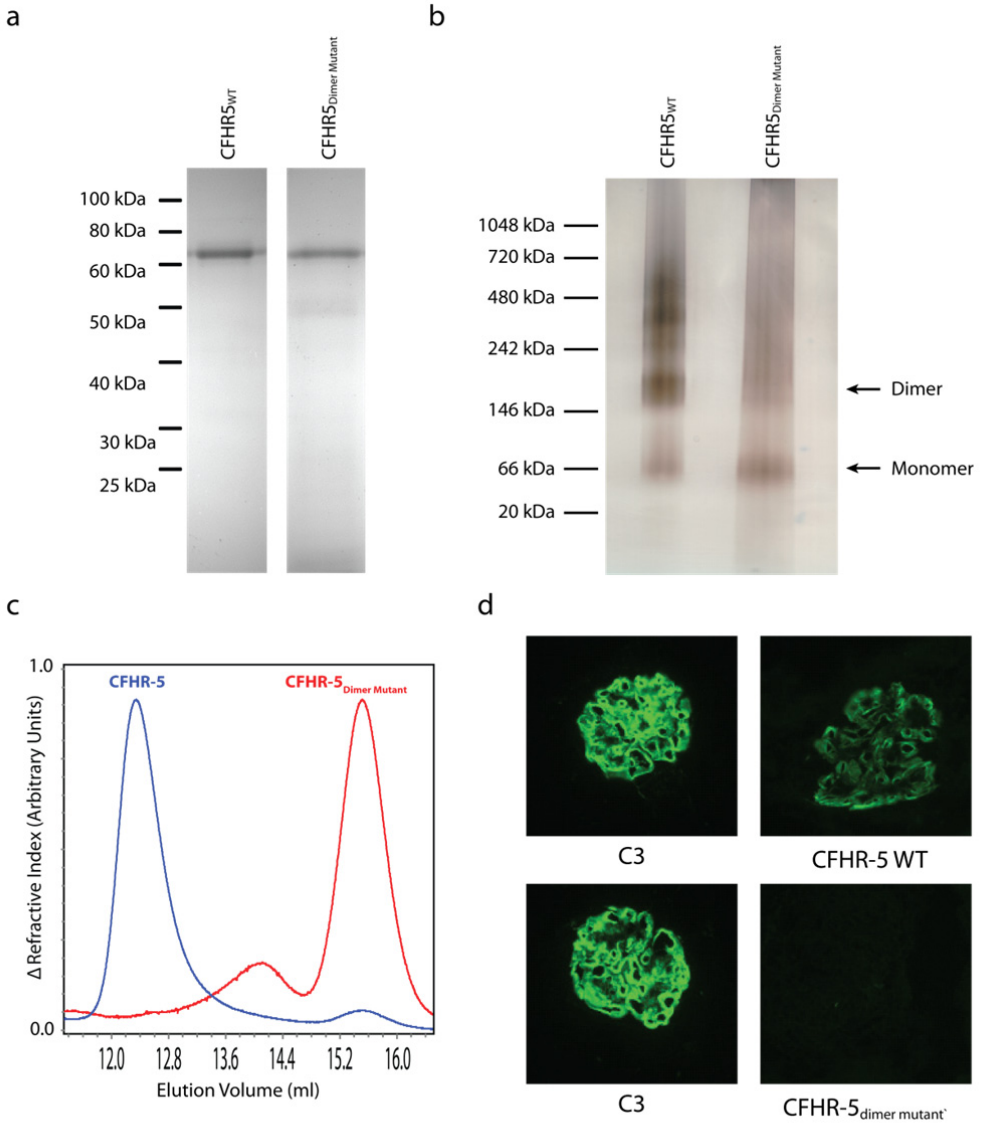


Figure 7.23 *Panel* a: 10% (w/v) SDS-PAGE analysis of CFHR-5 and the dimerisation knockout mutant. *Panel b*: 3 -12 % (w/v) blue native PAGE gel comparing the CFHR-5 wild type and dimerisation knockout mutant. *Panel* c: Analytical gel filtration analysis of CFHR-5 and the dimerisation knockout mutant. Performed using an S200 10/300 column (GE Healthcare, USA). Panel *d*: Staining of histological sections of the glomerular basement membrane with α-C3 and α-CFHR-5 polyclonal antibodies following injection of wild type CFHR-5 or the dimerization ablated mutant. These images were prepared by Dr. Elena Goicoechea de Jorge (Imperial College, London).

7.6 Investigation of the interactions of CFHR-1 and CFHR-2 with C3b or C5

[Heinen, Hartmann et al. 2009] reported that CFHR-1 inhibited the C5 convertase, resulting in negative regulation of complement, and bound C5 through its N-terminal domains. Following the revelation that these domains promote dimerization of the molecule, the C5 binding ability of the molecule must be reassessed alongside determination of the C3b binding site as it was not determined in the original paper.

7.6.1 Binding of C3b by CFHR-1 assessed using surface plasmon resonance

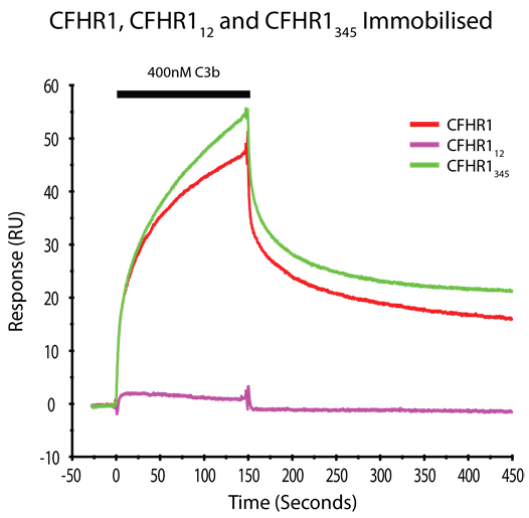


Figure 7.24 Reference subtracted curves from an injection of 400 nM C3b flowed over immobilised CFHR-1, CFHR-1_{12} and CFHR-1_{345}.

Binding of CFHR-1_{345} to immobilised C3b has already been presented (Chapter 7.2) showing that it binds with a very similar affinity to domains 19 and 20 of factor H. Surface plasmon resonance was employed to investigate whether these domains are the sole receptor for C3b, or whether the N-terminus of CFHR-1 also interacts with C3b. CFHR-1 (purified from serum by Dr. Claire Harris (Cardiff University, UK), CFHR-1_{12} and CFHR-1_{345} were immobilised on a CM5 chip (GE Healthcare, USA) using standard primary amine coupling protocols. Attention was paid to the immobilisation levels in order to have approximately equimolar ratios of the proteins on each channel. Final coupling levels were 2300 RU CFHR-1, 800 RU CFHR-1_{12} and 600 RU CFHR-1_{345} with a reference channel that was also mock activated and deactivated. A buffer comprising 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P-20, pH 7.5 (HBS-EP) was used throughout the experiment. 100 µl of 400nM C3b (Calbiochem,

Germany) was flowed over the chip surface at 40 µl/min using the KINJECT command at 25°C, matching the experiment reported by Heinen, Hartmann *et al.* (2009).

Following subtraction of the reference channel using BIAEVALUATION (GE Healthcare, USA) the processed data clearly show that the interaction between CFHR-1 and C3b is mediated solely by the three C-terminal domains (Figure 7.24) with the small difference between the CFHR-1 and CFHR-1_{345} curves likely to be a result of slightly different deposition levels on the chip surface.

7.6.2 Affinity measurement of CFHR-1 and CFHR-2 binding to C3b

Following the discovery that the C-terminal domains of CFHR-1 bind C3b, an attempt was made to measure the affinity of the interaction. The affinity of factor H domains 19 and 20 for C3b has previously been determined to have a K_D of 3.5 - 4.5 µM [Schmidt, Herbert *et al.* 2008b]. Due to the extremely high levels of sequence similarity between CFHR-1_{345} and $fH_{18\text{-}20}$, one would expect the former to have retained the C3b binding properties of the latter. Surface plasmon resonance was employed to investigate this supposition and also to extend the hypothesis to determine whether CFHR-2_{34} has also retained C3b binding properties.

All measurements were made on a Biacore 3000 instrument (GE Healthcare, USA) using CM5 chips to which proteins were immobilised via standard primary amine coupling protocols. A reference channel that was mock activated-deactivated was included on each chip and a buffer consisting of 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) surfactant P-20, pH 7.4 was used throughout. 800-1200 RU of C3b (Calbiochem, Germany), diluted into 10 mM sodium acetate, pH 4.0, was immobilised on the chip surface prior to dilution series of CFHR-1_{345} and CFHR-2_{34}, between 40 nM and 20 µM, being flowed over at 40 µl/min. 80 µl was injected using

KINJECT and a dissociation time of 500 seconds recorded. Curves were reference subtracted and analysed using BIAEVALUATION (Figure 7.25).

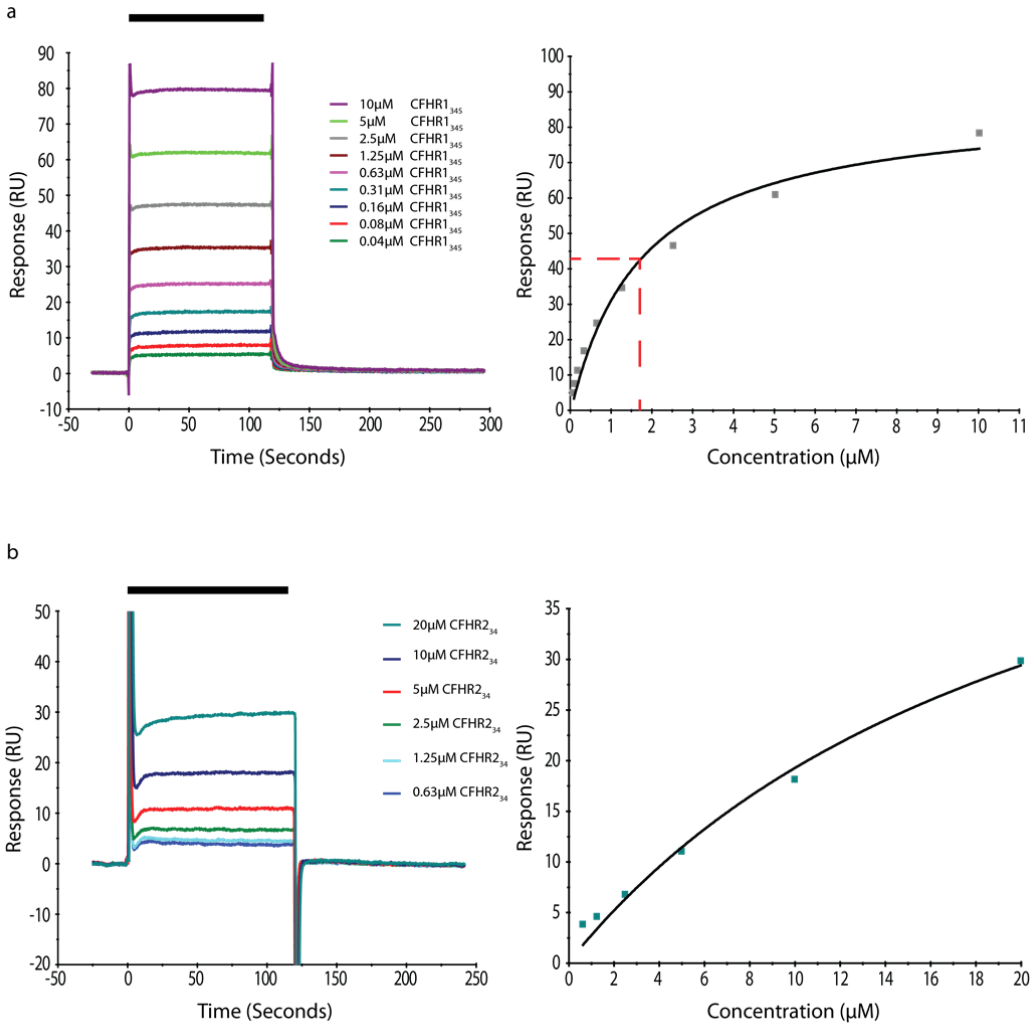


Figure 7.25 *Panel a*: Attempted measurement of the affinity of CFHR-1_{345} for C3b using a steady state fit. Example curves are inset showing equilibrium was reached at each concentration and the region over which as average response was measured. Also shown is a fit to the steady-state value but the quality of the fit is poor and questions the validity of the resulting K_D value. A rough estimate of the K_D can be interpolated from the concentration at which the response is half of R_{max} (shown in red) placing it around 2 μM. *Panel b*: Attempted measurement of the affinity of CFHR-2_{34} for C3b using a steady state fit. The resulting fit to the equilibrium responses is poor and due to the low curvature a reliable estimate of R_{max} is not attainable to perform the same treatment as in panel a.

Both CFHR-1_{345} and CFHR-2_{34} display concentration dependent binding to immobilised C3b. Unfortunately, the poor quality of fits to the data do not allow accurate calculation of the binding affinity between the interacting partners. However, a

rough estimate of the CFHR-1_{345} affinity for C3b may be obtained from the concentration at which the response is half that of the maximum, R_{max}. This treatment reveals an approximate K_D of 1.75 µM which is entirely in keeping with the previously published values of 3.5 - 4.5 µM for fH_{19-20} with C3b [Schmidt, Herbert *et al.* 2008b]. A similar treatment of the CFHR-2_{34} binding data was not possible as the low curvature of the fit does not allow precise determination of R_{max}. This could be due to the curves not quite reaching equilibrium and the small responses from the lower concentration samples. However, the responses over this concentration range display concentration dependent and suggest that the K_D belongs to the same order of magnitude as the interaction between CFHR-1_{345} and C3b. Binding of C3b by CFHR-1_{345} and CFHR-2_{34} also helps to confirm that the refolded proteins are correctly folded.

7.6.3 Binding of C5 by CFHR-1 assessed using surface plasmon resonance

C5 binding to the N-terminal domains of CFHR-1 was demonstrated by [Heinen, Hartmann *et al.* 2009] in an ELISA assay. Antibodies against the N or C-terminus of CFHR-1 were used to immobilise the protein prior to binding of C5 and detection with an anti-C5 monoclonal antibody. It is likely that the antibody against the C-terminus will result in immobilisation of molecules with the other C-terminus available for binding, whilst it is possible that the N-terminal antibody may hinder access to a C-terminal binding site. They also reported binding of C5 to immobilised CFHR-1 using surface plasmon resonance, but presented non reference subtracted curves which suggest a high level of non-specific binding of C5 to the chip surface.

In a repeat of the SPR experiment reported by [Heinen, Hartmann *et al.* 2009], a dilution series of C5 was flowed over the chip created in Chapter 8.2.1. This will also extend the assay to assess the binding of CFHR-1_{12}, to C5 in order to separately confirm if these domains are involved in an interaction. 100 µl of each dilution in a series of C5

concentrations ranging from 50 nM to 400 nM was flowed over the chip surface at 40 µl/min using the KINJECT command and the baseline was allowed to return to its initial value prior to each injection. The resulting reference subtracted curves (Figure 7.26) show that there is no concentration dependent binding between CFHR-1, or CFHR-1_{12}, and C5. This is in contrast to the previously reported results which may be explained by non-specific binding of C5 to the chip surface, a phenomenon observed in other experiments not reported in this thesis. Non-specific binding probably also accounts for the small responses observed in the sensorgrams which are negligible in comparison to those reported from a similar experiment with C3b.

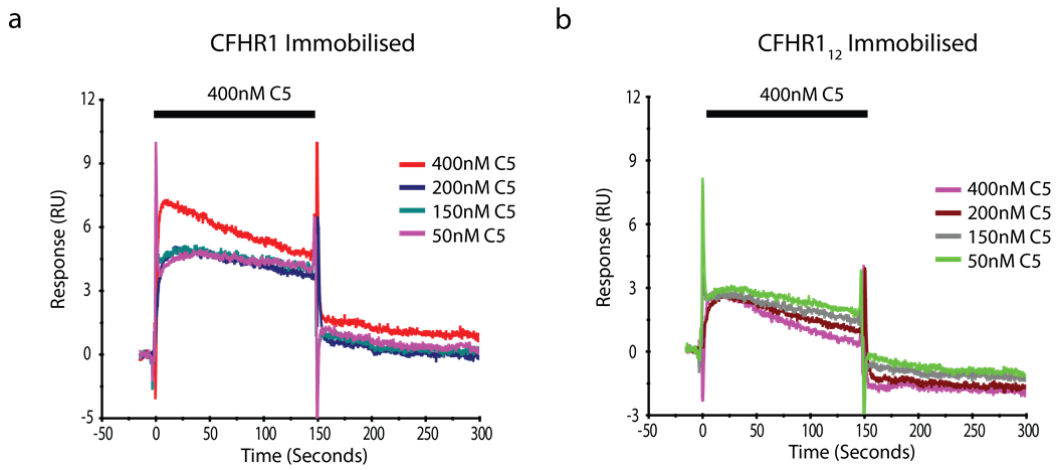


Figure 7.26 *Panel a*: Reference subtracted curves from a dilution series of C5 flowed over the CFHR-1 channel. *Panel b*: Reference subtracted curves from the CFHR-1_{12} channel.

7.7 Heparin binding properties of CFHR-1 and CFHR-2

In order to assess the glycosaminoglycan binding properties of CFHR-1_{345} and CFHR-2_{34}, their affinity for heparin, a commonly used GAG analogue, was investigated. Proteins were dialysed into 50 mM Tris, 10 mM NaCl, pH 7.5 before approximately 100 µg was loaded on a 1ml HiTrap Heparin column (GE Healthcare, USA) pre-equilibrated in 50 mM Tris, 10 mM NaCl, pH 7.5. Following washing with 5 CV of 50

mM Tris, 10 mM NaCl, pH 7.5 the column was eluted using a 10 – 500 mM gradient of NaCl over 15 CV (Figure 7.27). CFHR-2_{34} did not bind and was washed out during wash step where as CFHR-1_{345} exhibited strong binding and eluted at a conductivity of 29.6 mS/cm.

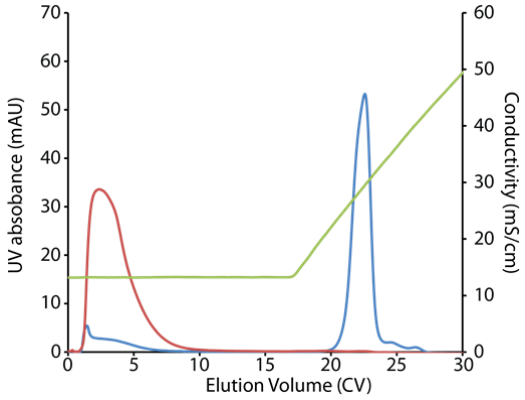


Figure 7.27 UV absorptoin profiles of the elution of CFHR-1_{345} (blue) and CFHR-2_{34} (red) from a HiTrap Heparin column with a salt gradient. Conductivity is shown in green.

7.8 Crystallisation, data collection and processing of CFHR-2_{34}

To further validate the correct folding of CFHR-2_{34}, rationalise its different GAG binding properties and possibly provide insight into its function, attempts were made to obtain the atomic resolution structure of these domains using X-ray crystallography.

7.8.1 Crystallisation of CFHR-2_{34}

Crystallisation trials of CFHR-2_{34} were carried out from a stock with A_{280} = 7.8 in an identical manner to CFHR-1_{12} (Chapter 7.3.1). Crystals grew after approximately 14 days in a condition containing 30% (w/v) PEG 8000, 0.2 M ammonium sulphate (Figure 7.28, panel a). However, these crystals consisted of multiple plates nucleating from a single point which had to be dissected prior to mounting in cryoloops (Hampton Research, USA), cryoprotection in 15% ethylene glycol prior and flash cooling in liquid nitrogen.

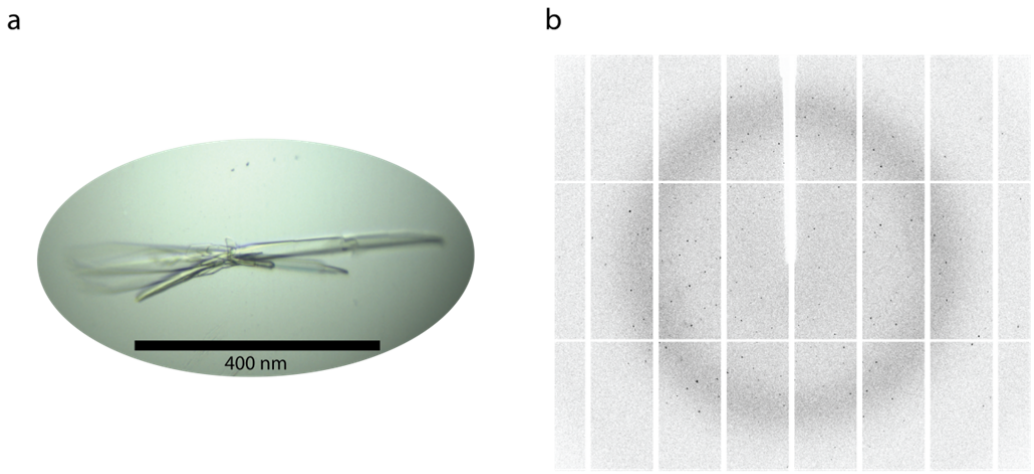


Figure 7.28 *Panel a*: Crystals of CFHR-2_{34}. *Panel b*: Example diffraction image.

	CFHR2$_{34}$
Diffraction source	DIAMOND I04-1
Detector	PILATUS 6M
Energy (eV)	13516
Wavelength (Å)	0.9173
Temperature (K)	120
Detector resolution (Å)	2.3
Exposure per image (s)	0.3
Φstart	346
Φend	166
ΔΦ	0.15
Number Images	1200

Table 7.9 Data collection parameters.

	CFHR-2_{34}
Spacegroup	P2
Unit-cell dimensions (Å)	a=53.0, b=25.2 c=95.7
Unit-cell angles (°)	α=90, β=93.8, γ=90.0
Wilson B factor	24.7
Resolution (Å)	2.00-95.47 (2.00-2.05)
R_{merge}	0.05 (0.259)
R_{pim}	0.033 (0.199)
<I/σ(I)>	15.2 (4.0)
Completeness (%)	96.7 (86.8)
Multiplicity	3.2 (2.6)

Table 7.10 Data processing statistics.

7.8.2 Data collection of CFHR-2_{34}

X-ray diffraction data were collected on beamline I03 at the Diamond light source using the rotation method. Crystals were indexed in a primitive monoclinic cell (cell dimensions a = 53.0, b = 25.2, c = 95.7, $\alpha = \gamma = 90^{o}$, $\beta = 93.8^{o}$) with P2 symmetry prior to calculation of the optimal data collection strategy using EDNA. A single high-resolution dataset was collected using the parameters detailed in Table 7.9.

7.8.3 Data processing of CFHR-2_{34}

Data were processed using XIA2 invoking the 3dii option to enforce indexing of every image prior to integration with XDS and scaling with XSCALE. Analysis using POINTLESS verified the likely spacegroup assignment of P2 and data processing statistics are presented in Table 7.10.

7.9 Structure determination and refinement of CFHR-2_{34}

Due to the similarity between CFHR-2_{34} and domains nineteen and twenty of factor H (89% and 61% sequence identity respectively), $fH_{19\text{-}20}$, these SCRs were used as a molecular replacement model to determine the structure.

7.9.1 Structure solution of CFHR-2_{34}

Analysis of the unit cell suggested that there are two copies of CFHR-2_{34} in the asymmetric unit. However, a peak in the native Patterson ~35% of the origin and lack of significant non-crystallographic peaks in the self-rotation function suggested the molecules are related by a non-crystallographic translational symmetry which must be taken into consideration during molecular replacement and refinement.

The structure of $fH_{19\text{-}20}$ (PDB accession code: 2G7I [Jokiranta, Jaakola *et al.* 2006]) was used as a model for molecular replacement. Two copies of the model, stripped of solvent and other ligands, were sequentially placed into the data using PHASER with the scores described in Table 7.11. The resulting model was subjected to a single round of rigid-body refinement in REFMAC [Murshudov, Skubak *et al.* 2011], declaring each SCR domain as a separate entity, and the resulting phases were input into BUCCANEER [Cowtan 2006] which

	RFZ	TFZ	LLG
Copy 1	12.2	6.4	172
Copy 2	7.6	30.0	707

Table 7.11 Rotation function (RFZ) Z-scores, translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of two copies of the model in the data.

rebuilt the starting model with the CFHR-2_{34} sequence. An initial model consisting of 247 residues in 7 chains, the longest of which contained 99 residues, was built with an R/R_{free} of 0.3692/0.4151. COOT was used to combine the chains and symmetry copies to generate a complete model.

7.9.2 Refinement of CFHR-2_{34}

The initial model of CFHR-2_{34} was refined iteratively between cycles of rebuilding in COOT and refinement with AUTOBUSTER. The final model has an R/R_{free} of 0.21/0.24 and was validated using the MOLPROBITY server. Refinement and validation statistics are displayed in Table 7.12. The fully refined structure is shown in Figure 7.29 alongside representative electron density.

	CFHR-2_{34}
Resolution (A)	2.0-19.09 (2.0-2.12)
Number of reflections	16963 (2567)
Number of reflections for R_{free}	859 (132)
R_{work}/R_{free}	0.2121/0.2400 (0.2122/0.2741)
Number of atoms	
Protein	1952
Ligand/Ion	117 Ethylene glycol
Water	77
R.m.s deviations	
Bond lengths (A)	0.01
Bond angles (°)	1.10
Ramachandran plot analysis	
Residues in favoured regions (%)	98.4
Residues in disallowed regions (%)	0.0
Molprobity analysis	
Molprobity score	1.35
Molprobity score percentile	99th
Resolution range (A)	1.75-2.25

Table 7.12 Refinement and validation statistics

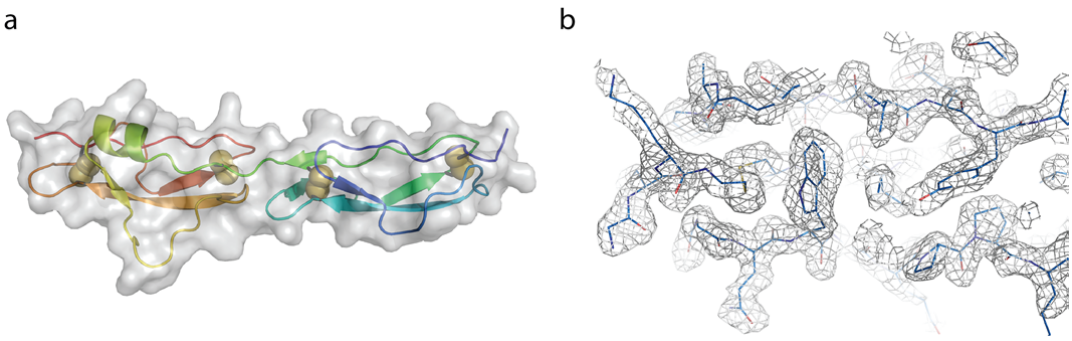


Figure 7.29 *Panel a*: Final model of CFHR-2_{34} shown as a cartoon of the secondary structure within a transparent surface rendering. The cysteine residues forming intra-molecular disulphide bonds are shown as gold spheres. *Panel b*: Representative electron density (2Fo-Fc map contoured at 1.5σ) with the refined structure shown in blue for clarity.

7.10 Structural analysis of CFHR-2_{34}

The atomic resolution structure of CFHR-2_{34} shows that the protein is correctly folded and has the correct number, and pattern, of disulphide bonds. There are two identical copies of CFHR-2_{34} in the asymmetric unit, but, as expected, analysis using PISA reveals no interfaces that are likely to be physiologically important. Comparison of the structures of CFHR-2_{34} and $fH_{19\text{-}20}$ (taken from PDB entry: 2XQW [Kajander, Lehtinen *et al.* 2011]. RMSD of superposition = 1.27 Å over 118 residues) shows that the charge distribution around the heparin binding region differs between the molecules. In $fH_{19\text{-}20}$ the heparin binding region is described by a positively charged groove running the length of both domains (Figure 7.30). However, in CFHR-2_{34} the positively charged region has disappeared and been replaced with a more neutral surface consisting of pockets of positive and negative charge explaining why CFHR-2_{34} is unable to bind the polyanionic GAG analogue heparin. Superposing the structure of $fH_{19\text{-}20}$ in complex with the C3d fragment of C3b (PDB entry: 2XQW [Kajander, Lehtinen *et al.* 2011]) onto CFHR-2_{34} shows that the sequence and charge distribution of the C3d binding site is very highly conserved, accounting for preservation of its ability to bind C3b (Figure 7.31).

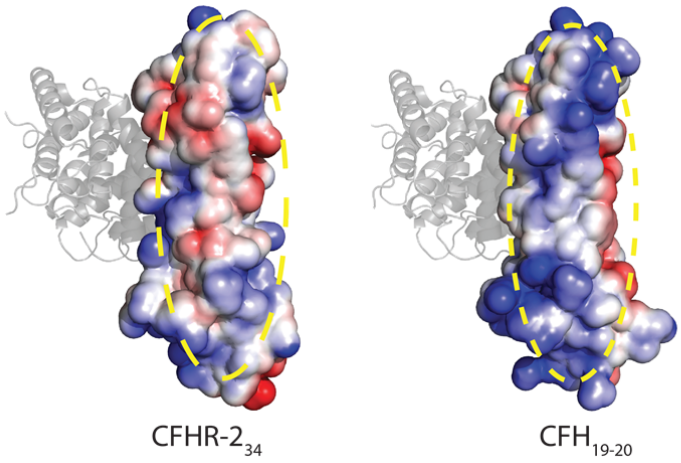


Figure 7.30 Comparison of the surface charge (calculated using APBS [Baker, Sept *et al.* 2001]) on both CFHR-2_{34} and fH_{19-20} shows oblation of the positively charged heparin binding patch in the former (highlighted in yellow).

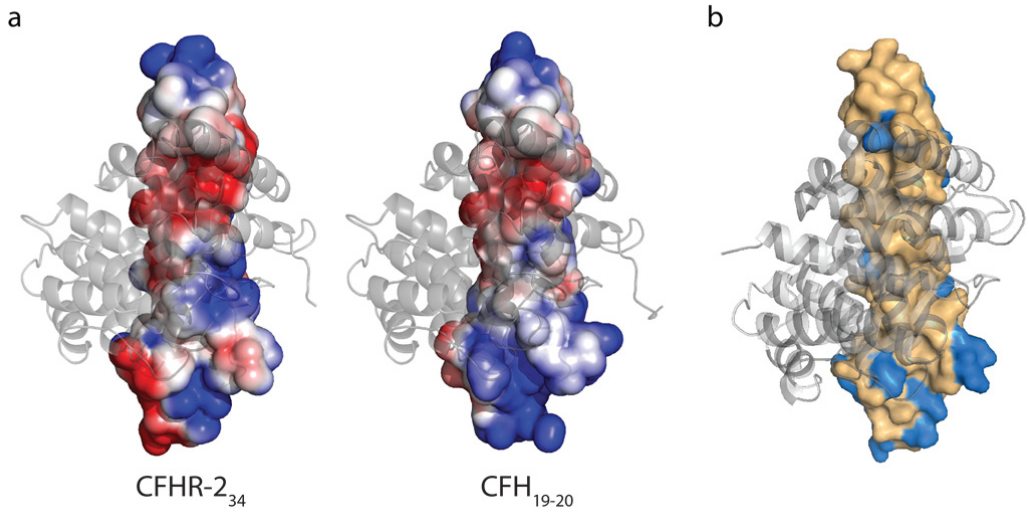


Figure 7.31 *Panel* a: Comparison of the surface charge (calculated using APBS[Baker, Sept *et al.* 2001] on both CFHR-2_{34} and fH_{19-20} shows conservation of the charge within the C3b binding interface. *Panel b*: Sequence conservation with fH_{19-20} mapped onto the structure of CFHR-2_{34} showing conservation of residues forming the interface with C3b. Conserved residues are coloured orange, differences in blue.

7.11 Crystallisation of CFHR-1_{345}

Crystallisation trials of CFHR-1_{345} were carried out from a stock with A_{280} = 10.3 in an identical manner to CFHR-1_{12} (Chapter 7.3.1). Large, single crystals grew after

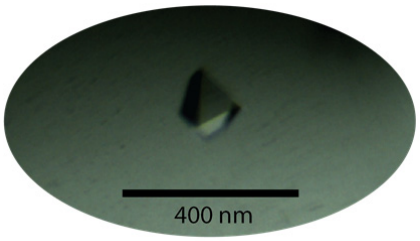


Figure 7.32 Potential crystals of CFHR-1_{345}.

approximately 8 weeks in a condition containing 2.0 M ammonium sulphate, 2% (v/v) PEG 400, 100 mM HEPES pH 7.5 (Figure 7.32). At the time of writing these crystals were still waiting to be sent to a synchrotron for data collection.

7.12 Regulatory activity assays

CFHR-1 was previously determined to be a negative regulator of the complement response by observing a decrease in complement mediated haemolysis of sheep erythrocytes [Heinen, Hartmann *et al.* 2009]. However, these experiments were performed in factor H deficient serum. In order to confirm these findings, haemolysis assays were repeated using factor H sufficient serum in a guinea pig erythrocyte haemolysis assay which is a standard assay for analysis of the alternative activation pathway.

Guinea pig erythrocytes (GPEs) possess the ability to activate complement in human serum via the alternative pathway. By performing the assay in the presence of ethyleneglycol tetraacetic acid (EGTA), Ca^{2+} is chelated, deactivating C1 complex and ensuring that activation only occurs via the alternative pathway. Mg^{2+} must also be supplemented back into the haemolysis reaction to ensure the activity of factor D. Following complement activation, GPEs are lysed by formation of membrane attack complex, releasing haemoglobin which can be measured spectrophotometrically. The levels of haemoglobin released are proportional to complement activity, allowing determination of whether molecules added to the reaction affect the activity levels.

7.12.1 Serum titration

Prior to commencement of haemolysis assays, the proportion of serum in each reaction must be titrated to determine the optimal serum concentration where there are high levels of haemolysis, but the serum is not in excess, hindering observation of any effects resulting from sample addition.

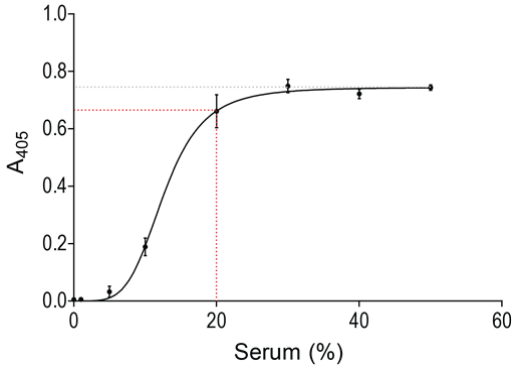


Figure 7.33 Dose-response curve showing how haemolysis of GPEs (measured by haemoglobin release at 405 nm) by complement activated via the alternative pathway varies with serum concentration in the reaction vessel. All data points are reported as the mean and standard error from three separate dilution series. The maximum level of lysis is marked with a grey line whilst the serum concentration used in subsequent experiments is shown in red.

Freshly collected human blood was allowed to clot at 298 K for 30 minutes before the serum was harvested by centrifugation (2200 g, 10 minutes, 4°C), flash frozen and stored at -80°C until required. 400 µl of GPEs (TCS Biosciences, UK) were washed seven times with 1 ml of 100 mM HEPES, 150 mM NaCl, 8 mM EGTA, 5 mM $MgCl_2$, 0.1% gelatin, pH 7.5 prior to dilution into 10 ml of the same buffer. Three dilution series containing 2 - 100 µl of serum made up to 150 µl with 100 mM HEPES, 150 mM NaCl, 8 mM EGTA, 5 mM $MgCl_2$, 0.1% gelatin, pH 7.5 were set up, with appropriate controls, in a 96 well conical bottomed plate. 50 µl of the diluted GPEs were added to the reactions prior to incubation at 37°C for 1 hour and clearance by centrifugation (4000 g, 5 minutes, 4°C). 150 µl of supernatant from each reaction was transferred to a 96 well plate and the absorbance at 405 nm measured for each well using a spectrophotometer.

A plot of the blank-subtracted absorbance against the percentage of serum (Figure 7.33) can be fitted with a four parameter dose-response curve that shows maximum lysis is

achieved at approximately 30% serum. Subsequent experiments are performed containing 20% serum as this concentration gives a high level of haemolysis (approximately 90%), but will also allow the effects of adding samples to be observed without adding them in great excess.

7.12.2 Factor H titration

In order to assess how the addition of factor H results in inhibition of complement-mediated lysis by regulation of the alternative activation pathway, haemolysis was measured in the presence of different concentrations of the complement regulator.

Reactions containing 20% serum and 0.001 - 10 µM factor H (kindly provided by Dr. Pietro Roversi, University of Oxford) in a total volume of 150 µl (made up with 100 mM HEPES, 150 mM NaCl, 8 mM EGTA, 5 mM $MgCl_2$, 0.1% gelatin, pH 7.5) were prepared in triplicate with appropriate controls. 50 µl of GPEs were added, the reactions incubated and the absorbance at 405 nm read as described in Chapter 8.3.1. All samples were blank-subtracted for absorption by the serum and then expressed as a percentage of the maximum haemolysis achieved in the absence of any factor H addition and graphed (Figure 7.34). Fitting the data points and interpolation of the concentration at which haemolysis has been inhibited by 50% gives an IC_{50} of 140 nM for factor H with these reaction conditions and reagents.

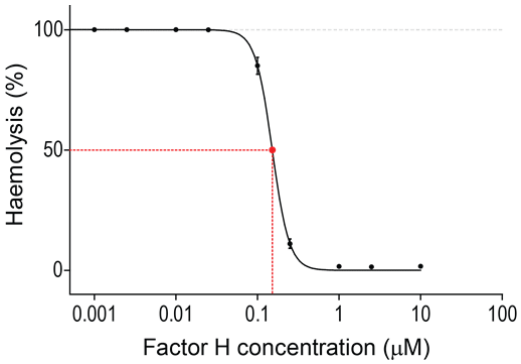


Figure 7.34 Dose-response curve for inhibition of GPE lysis by factor H in 20% serum. The IC_{50} (140 nM) is marked in red.

7.12.3 Effect of the CFHRs upon complement activity

In order to replicate the findings of [Heinen, Hartmann *et al.* 2009] in the guinea pig erythrocyte lysis assay, the effect of CFHR-1 on haemolysis was observed over an identical concentration range to that reported and extended to include the component constructs CFHR-1_{12} and CFHR-1_{345} alongside CFHR-2_{34} and CFHR-5. However, in order to observe either an increase or decrease in haemolysis, the experiment was modified by the addition of IC_{50} levels of factor H determined in Chapter 7.12.2.

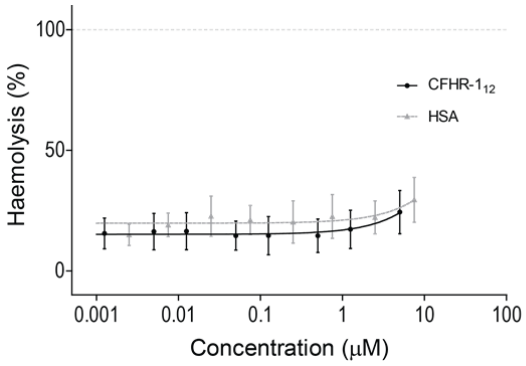


Figure 7.35 Linear fits to haemolysis in the presence of HSA and CFHR-1_{12}. Deviation from linearity is not significant (P (Runs test) = 0.43 and 0.11 respectively) and a non-zero gradient is also not significant (P = 0.21 and 0.18 respectively).

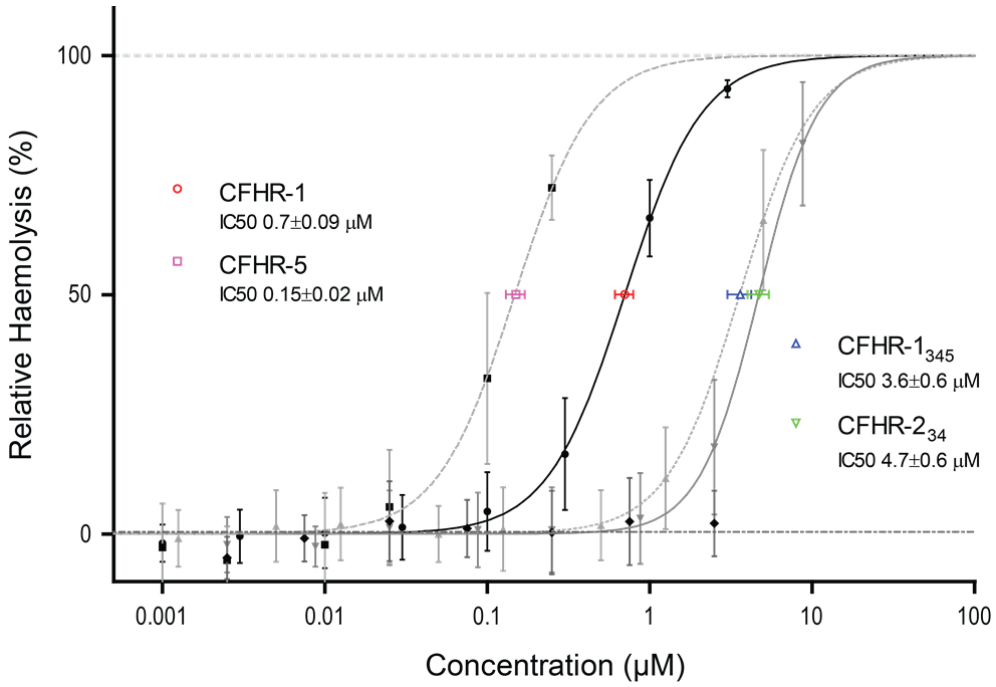


Figure 7.36 Dose-response curves describing the increase in haemolysis upon addition of CFHR-1, CFHR-1_{345}, CFHR-2_{34} and CFHR-5. IC_{50} concentrations have been interpolated and are displayed.

Haemolysis assays were performed in a total volume of 200 µl containing 20% serum, 140 nM factor H and approximately 1 x 10^6 guinea pig erythrocytes in 100 mM HEPES, 150 mM NaCl, 8 mM EGTA, 5 mM $MgCl_2$, 0.1% gelatin, pH 7.5. Haemolysis was measured by the absorbance at 405 nm after 60 minutes at 37°C and is reported as a percentage of lysis in the absence of factor H (100%) relative to the haemolysis levels in its presence (0%) following appropriate control subtraction. Dilution series of CFHR-1, CFHR-1_{12}, CFHR-1_{345}, CFHR-2_{34}, CFHR-5 and human serum albumen (HSA), ranging from 1 nM to 9 µM, were added to the reactions and the haemolysis levels measured. All measurements were recorded in triplicate and are presented in Figures 7.35 and 7.36.

No significant effects of haemolysis were detected in the reactions supplemented with CFHR-1_{12} and HSA (Figure 7.35). However, addition of CFHR-1, CFHR-1_{345}, CFHR-2_{34} and CFHR-5 results in increased levels of lysis and fitting of these curves allows interpolation of the IC_{50} for each sample (Figure 7.36).

7.12.4 Effect of CFHR-1 upon the activity of complement regulators

The findings in Chapter 7.12.3, showing that addition of CFHR-1 results in increased lysis, are in complete contrast to the findings of [Heinen, Hartmann *et al.* 2009] who reported decreased haemolysis. However, the inhibition effects in this study were only observed upon addition of 4 µM CFHR-1 which is higher than the serum concentration of the protein (1.8 - 2.5 µM) and much higher than the concentrations at which increased lysis was observed in Chapter 7.12.3 (IC_{50} = 700 nM), suggesting that the latter effect is more likely to be physiologically relevant. A likely explanation for the differences observed is that the experiment reporting negative regulation by CFHR-1 was performed in the absence of factor H using $fH^{-/-}$ serum, suggesting that the presence of the complement regulator is necessary for proper function. In order to assess whether

factor H is involved in the mechanism of CFHR-1, haemolysis was compared in the presence and absence of the complement regulator.

Haemolysis using 20% normal human serum (NHS) and fH deficient serum was measured in the presence and absence of 700 nM CFHR-1, replicating the IC_{50} concentration determined in Chapter 7.12.3. Reactions containing 20% serum (NHS or $fH^{-/-}$) and approximately 1 x 10^6 GPEs made up to 200 µl with 100 mM HEPES, 150 mM NaCl, 8 mM EGTA, 5 mM $MgCl_2$, 0.1% gelatin, pH 7.5, were set up with and without 700 nM CFHR-1. Haemolysis was measured by the absorbance at 405nm after 60 minutes at 37°C. All measurements were performed in triplicate with appropriate controls. Responses were subtracted for serum absorption and expressed as a percentage of 100% lysis achieved by osmolysis of the same quantity of GPEs with H_2O.

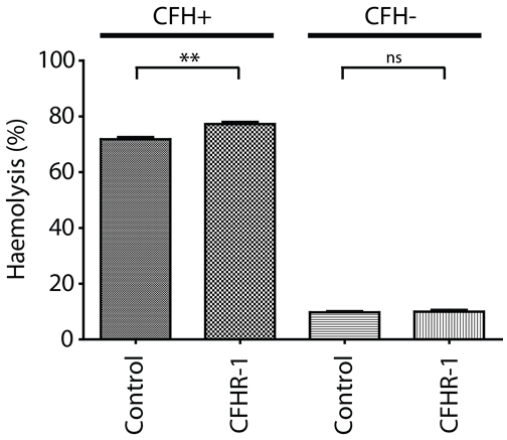


Figure 7.37 Levels of haemolysis in NHS and fH deficient serum in the presence and absence of 700 nM CFHR-1.

In the presence of factor H, the levels of haemolysis are significantly increased upon addition of CFHR-1 whilst in factor H deficient serum there was no significant difference observed upon its addition (Figure 7.37). This supports the hypothesis that factor H is involved in the mechanism of increased haemolysis by CFHR-1 and suggests that CFHR-1 is blocking the activity of fH by acting as a competitive antagonist, resulting in decreased regulation and associated increased haemolysis.

In order to assess whether the competitive behaviour of CFHR-1 is directed solely towards factor H, haemolysis was measured in the presence of inhibition by another

complement regulator, decay accelerating factor (DAF/CD55). Previous work in the laboratory of Professor Susan Lea has shown that addition of 100 - 200 nM DAF results in observable inhibition of lysis. Haemolysis of GPEs by NHS supplemented with 140 nM DAF (prepared by Miss. Imogen Breen), in the presence and absence of 700 nM CFHR-1 was performed in an identical manner to that described overleaf (Figure 7.38). The level of haemolysis is significantly increased in the presence of CFHR-1 showing that the antagonist is more general and is not solely directed towards competition against factor H.

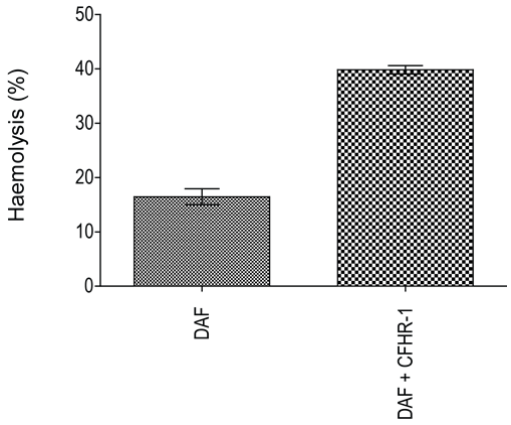


Figure 7.38 Haemolysis in the presence of 140 nM DAF with and without addition of 700 nM CFHR-1.

7.12.5 Effect of disease-associated polymorphisms within the CFHRs upon activity

It follows that CFHR-2 and CFHR-5 behave in a similar manner to CFHR-1, competing against regulators of complement for C3b binding, resulting in decreased regulation and a resulting increase in activation. As previously discussed, disease-related polymorphisms with duplication of the dimerization domains have the ability to assemble into higher order multimers than dimers. These multimers will have an even higher affinity for C3b than the dimeric forms due to avidity, resulting in increased competition for regulator binding and higher levels of complement activation that may be responsible for the manifestation of disease.

In order to test this hypothesis CFHR-1, CFHR-2 and CFHR-5 were purified from control subjects and patients with the polymorphisms using immunoprecipitation with an antibody against their common N-terminus by Dr. Elena Goicoechea de Jorge

(Imperial College, London). Haemolysis of GPEs was measured in the presence of 20% NHS and 140 nM factor H as described earlier but with the addition of the purified CFHRs reconstituted to serum concentrations. All measurements were performed in triplicate and responses were blank subtracted before being expressed as a percentage of complete lysis achieved using H_2O (Figure 7.39).

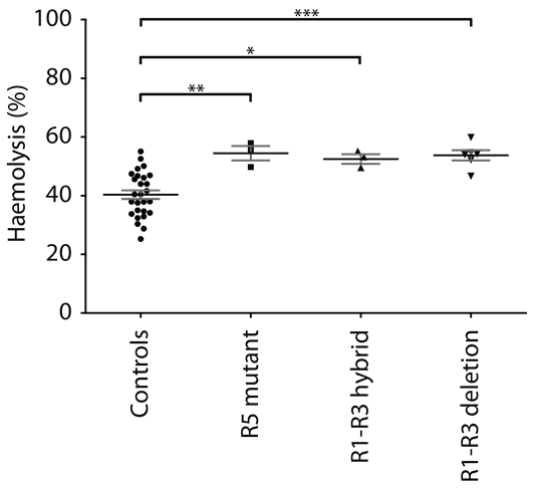


Figure 7.39 Effect of adding CFHRs from individuals with disease related polymorphisms upon haemolysis. All polymorphisms result in significantly increased levels of lysis with P values of 0.004, 0.011 and 0.0002 for the CFHR-5_{mutant}, CFHR-1/3 hybrid and heterozygous CFHR-1/CFHR-3 deletion.

As predicted, the CFHR-5_{mutant} and CFHR-1/3 hybrid samples result in significantly increased levels of haemolysis, supporting the hypothesis outlined above. CFHRs purified from patients presenting the CFHR-1/3 deletion also promote increased levels of haemolysis. The rationale behind this observation is not yet clear but could be due to increased levels of CFHR-5 homodimers in the absence of CFHR-1. CFHR-5 has a lower IC_{50} than CFHR-1 (Chapter 7.12.3) and will out compete factor H at lower C3b deposition levels, resulting in increased lysis. However, this does not account for the decreased susceptibility to disease and further investigations are required.

7.13 Discussion

Structural characterisation of CFHR-1 has revealed that it forms a homodimeric species mediated by a novel dimerization motif within the two N-terminal SCR domains. The atomic resolution structure of these domains has been solved showing formation of a dimer in which the monomeric subunits are arranged in an antiparallel arrangement. The

interface itself comprises a highly hydrophobic interaction surface, supplemented by the formation of four intermolecular hydrogen bonds, and is two-fold symmetric resulting in synergic binding and an extremely tight interaction. This is the first example of a high affinity interaction between SCR domains and the novel dimerization surface has evolved from the GAG binding interface accounting for the change in observed carbohydrate binding properties. The physiological implications of CFHR-1 dimerisation are likely to be profound as duplication of all the binding sites will result in increased affinity for its ligands driven by avidity. This could account for the low serum concentrations of these molecules and prompts re-evaluation of all previously published data in light of this discovery.

The two N-terminal domains of CFHR-1, CFHR-2 and CFHR-5 are almost identical suggesting both that CFHR-2 and CFHR-5 form homodimers, but also that heterodimers can form between combinations of all three proteins. CFHR-1 purified from serum is shown to form both homodimers and heterodimers with CFHR-2 and CFHR-5. The existence of heterodimers has also been confirmed by experiments by Dr. Elena Goicoechea de Jorge. Proteins were purified from plasma via immunoprecipitation with an anti-CFHR-5 antibody raised against the unique C-terminus. The presence of other CFHRs co-purified were detected via Western blot using an antibody against their common N-terminus, showing co-precipitation of CFHR-5, CFHR-2 and CFHR-1 and proving the existence of both homo and heterodimers. Work is currently being undertaken to quantify the relative abundance of each dimeric species using antibodies including the anti-CFHR-2_{34} monoclonal antibody produced and tested in Chapter 7.1.7.

Mutations have been designed that confer monomerisation upon CFHR-5. In an *in vivo* model, dimeric CFHR-5 accumulates to bound C3b in the glomerular basement

membrane. Addition of the monomeric form of CFHR-5 results in a dramatic decrease in accumulation suggesting that dimerisation, and the resulting increased affinity for C3b, plays an important role in the biological function of the CFHRs.

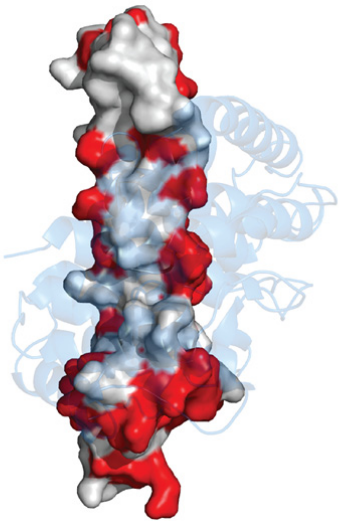


Figure 7.40 Mapping of the sequence differences with CFHR-5 (red) onto the structure of fH_{19-20} in complex with the C3d fragment of C3b.

The C-terminal domains of CFHR-1, CFHR-2 and CFHR-5 possess the highest identity to the C-terminal domains of factor H which have been shown to bind GAGs, characterised by a heparin binding groove, and C3b via the C3d fragment. CFHR-1 has been shown to also possess heparin and C3b binding properties, as expected due to the almost 100% identity. However, $CFHR\text{-}2_{34}$ still possesses C3b binding affinity of a similar magnitude to factor H but has altered its GAG binding properties. Solution of the atomic resolution structure of $CFHR\text{-}2_{34}$ shows conservation of the C3b binding site, but complete oblation of the heparin binding groove, suggesting that the former is critical to the biological function of the protein. One explanation could be that CFHR-2 is tailored to binding a different surface other than those presenting anionic GAGs, such as lipidated surfaces, and may explain why CFHR-2 has been observed to associate with serum lipoprotein particles [Park and Wright 1996].

Mapping of the sequence conservation between $CFHR\text{-}5_{8-9}$ and fH_{19-20} onto the structure of fH_{19-20} in complex with C3d suggests that the C3d binding site is conserved (Figure 7.40) and hints that binding to C3b is critical to the mechanism of these molecules. These findings allow a general model to be proposed for CFHR-1, CFHR-2 and CFHR-

5 where the two N-terminal domains promotes the formation of both hetero and homodimers whilst the C-terminal domains are involved in binding to C3b (Figure 7.41). As previously discussed, avidity resulting from dimerization will increase the effective affinity of all interactions with CFHR-1, CFHR-2 and CFHR-5.

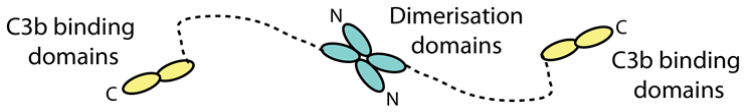


Figure 7.41 Cartoon demonstrating the common architecture of the CFHRs.

The function of the CFHRs within the complement system has never been properly elucidated due to the lack of reliable data showing regulatory function at physiological concentrations. Presented here are haemolysis assays showing that addition of CFHR-1, CFHR-2 or CFHR-5 increases the level of complement activation by competing for C3b binding with negative regulators. It is highly likely that this activity is directed mainly towards factor H, being the most highly abundant, serum bound regulator of the alternative activation pathway, suggesting that the purpose of the CFHRs is to compete with factor H for C3b binding.

Combination of all of these findings allows, for the first time, a mechanism of action of the CFHRs to be proposed. CFHR-1, CFHR-2 and CFHR-5 are all able to bind C3b and behave as competitive antagonists towards factor H, interfering with its ability to inhibit the C3 convertase. This process, termed de-regulation, gives rise to a feedback mechanism between regulation by factor H and de-regulation by the CFHRs enabling maintenance of the homeostatic levels of complement activation. The low serum concentrations of the CFHRs do not immediately suggest that they can significantly out-compete the highly abundant factor H. However, as CFHR-1, CFHR-2 and CFHR-5 exist as both homo and heterodimers their effective affinity for C3b, and hence their

ability to out compete factor H, is increased due to the influence of avidity. Under normal conditions, the nascent levels of C3b deposition on a host cell surface is low, maintaining the homeostatic balance between regulation by factor H and de-regulation by the CFHRs in favour of regulation, but retaining the correct level of activation to ensure a swift response if required. However, once the C3b deposition levels on a cell marked for clearance increase avidity will increase the effective affinity of the CFHRs, allowing them to out compete factor H driving the balance towards increased activation. A schematic diagram representing the proposed model is presented in Figure 7.42.

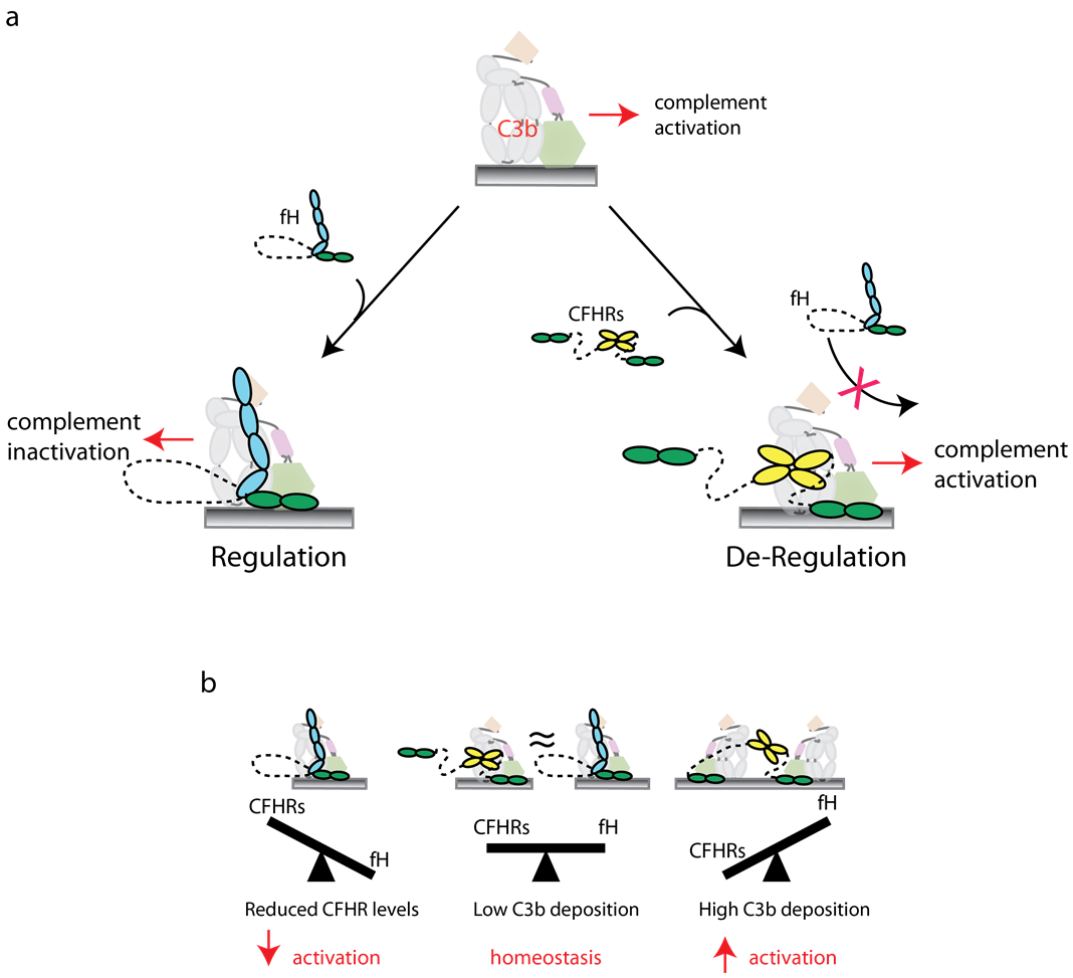


Figure 7.42 *Panel a*: Schematic desribing the proposed mechanism of action of the CFHRs. *Panel b*: Respresentation of the homeostatic balance of complement regulation and how it shifts upon alteration of the factors involves in its maintenance.

The ability to form homo and heterodimers between CFHR-1, CFHR-2 and CFHR-5 results in a set of molecules with similar de-regulatory properties, but with different degrees of separation between the C3b binding domains. Longer molecules will be more avid at lower C3b deposition levels leading to increased de-regulation. This is observed in the difference between CFHR-5 and CFHR-1 in the haemolysis assay, the IC_{50} of the longer CFHR-5 is approximately five times lower than that of CFHR-1. Each CFHR also has different GAG binding properties, suggesting that they are tailored to having a higher local concentration on the surface of serum exposed tissues presenting a specific pattern of carbohydrates. Coupling these two findings together suggests that each dimeric species of CFHRs plays a specific role in maintaining the appropriate level of complement activation on the surface of different serum exposed tissues and explains their different serum abundances to maintain this balance.

Previous studies have demonstrated that CFHR-1 is a negative regulator targeted towards the C5 convertase and showed a direct interaction with C5 via its N-terminal domains [Heinen, Hartmann *et al.* 2009]. However, the findings presented here show CFHR-1 de-regulates complement at physiological concentrations that are much lower than those needed to observe the regulatory effect. The absence of factor H in the published assays may also account for the different result. No interaction with C5 could also be detected with any portion of CFHR-1 which could be explained by a high propensity for non-specific binding to the chip surface by C5. In light of the findings presented here, further work is needed to clarify whether regulation of the C5 convertase is performed by CFHR-1.

Superposing the structure of factor H domains 19 and 20 in complex with C3d onto the structure of C3b in complex with factor H domains 1 through 4 gives a structural rationale for the action of the CFHRs (Figure 7.43). The C3d binding site in CFHR-1,

CFHR-2 and CFHR-5 is highly conserved, suggesting that these proteins bind in the same manner as fH_{19-20}. During regulation, both the N and C-termini of factor H bind to deposited C3b [Kajander, Lehtinen *et al.* 2011; Morgan, Schmidt *et al.* 2011] suggesting that binding of the CFHRs in the same binding site as fH_{19-20} would not block binding and of the regulatory domains of fH or DAF. However, increased steric occlusion of the binding site of the regulatory domains by the proximity of the dimerisation domains (shown in yellow in Figure 7.43) may prevent access to the site explaining why both regulators are de-regulated by the CFHRs. It is expected that the major function of the CFHRs is to compete with fH_{19-20} for C3b binding. This will reduce the overall affinity of factor H for C3b and result in decreased regulation. Thus the CFHRs act as competitive antagonists of factor H regulation resulting in de-regulation of complement activiation. [Ferreira, Herbert *et al.* 2006] showed that adding fH_{19-20} increased the levels of haemolysis without perturbing the solution activity of factor H. However, these effects were observed well above the physiological concentration. The CFHRs employ the same strategy, but employ dimerization to increase their effective concentration via avidity eliciting the same effect at low serum abundances.

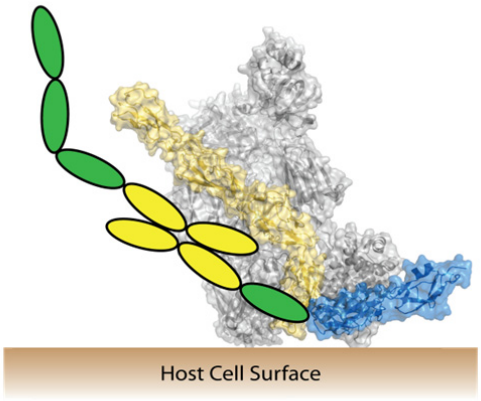


Figure 7.43 Superposition of the structures of C3b (grey) in complex with factor H domains 1 through 4 (gold) and factor H domains 19 and twenty (blue). A model of dimeric CFHR1 is shown with two of its N-terminal domains replaced with the structure of fH_{19-20}.

Two rare polymorphisms within the CFHRs have been linked to increased susceptibility to C3 glomerulopathy. Within a familial set of individuals of Cypriot ancestry, a duplication of the dimerization domains in CFHR-5 has been associated with

presentation of the disease as well as a hybrid of CFHR-3 plus the dimerization domains of CFHR-1 present in a family of Irish origin. Due to the multiple dimerization motifs one would expect these molecules to form higher order assemblies than dimers, most commonly trimers and tetramers with the most abundant CFHR-1. These assemblies will be even more avid for C3b resulting in increased competition for fH binding and have been shown to promote greater deregulation leading to amplified complement mediated damage and presentation of disease. The findings presented here also suggest a possible treatment for individuals possessing these polymorphisms. Depletion of the levels of CFHRs in their serum may be able to tip the activation balance back to reduced activation as factor H will be in excess, reducing damage from over activation of complement.

Another common polymorphism is a homozygous deletion of CFHR-1 and CFHR-3. This deletion has been associated with decreased susceptibility towards IgA nephropathy and age-related macular degeneration, but increased susceptibility to systemic lupus erythematosus. Haemolysis assays in the presence of CFHRs purified from individuals with the CFHR-1/3 deletion showed increased levels of lysis describing increased deregulation. This finding is at odds with what one might expect, that fewer CFHRs gives less deregulation and hence less lysis. However, it may be explained by increased levels of CFHR-5 homodimers which has a more potent deregulatory effect for reasons previously discussed. This could suggest that the protective effects are due to the absence of CFHR-1 and the increased susceptibilities to more CFHR-5 homodimers. This could also explain the manifestation, or protection, of disease in different tissues due to the differential GAG binding hypothesis described earlier. Further work is required to investigate the effects of this deletion and its relationship to disease

A variety of pathogens have been reported to bind the CFHRs including *Candida albicans* [Poltermann, Kunert et al. 2007], *Aspergillus fumigatus* [Behnsen, Hartmann et al. 2008], *Pseudomonas aeruginosa* [Kunert, Losse et al. 2007] and *Borrelia burgdorferi* [Siegel, Hallstrom et al. 2010]. In light of the findings that the CFHRs deregulate complement resulting in increased activation, the purpose of their being scavenged is not immediately clear. Presentation of the CFHRs on a pathogenic cell surface in a manner akin to factor H on the surface of *Neisseria meningitidis* (Chapter 2) will result in greater C3b deposition leading to increased clearance of the invading microbe. One hypothesis could be that CFHRs are bound by their C-termini, blocking the C3b interaction surface and rendering them unable to de-regulate. This would increase the regulatory effects of factor H, also presented on the surface, resulting in increased escape from complement mediated lysis.

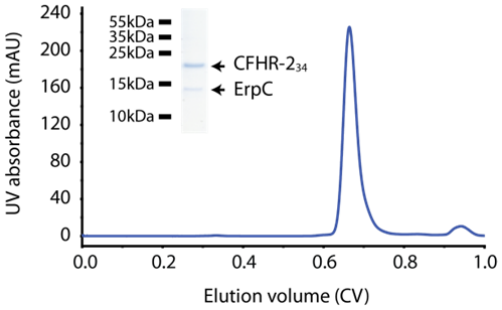


Figure 7.44 Elution profile of the complex between CFHR-2_{34} and ErpC from an S200 10/300 size exclusion column (GE Healthcare, USA). *Inset*: 15% (w/v) SDS-PAGE analysis of the elution peak showing the presence of both proteins.

The presence of CFHR heterodimers has profound implications upon study of thir scavenging by pathogens. Many of the CRASP family of proteins have been observed to bind multiple members of the CFHRs [Siegel, Hallstrom *et al.* 2010; Hammerschmidt, Hallstrom *et al.* 2012]. These results must now be re-evaluated as the presence of some of the CFHRs may be a secondary effect of binding another member of the family. Preliminary experiments have shown that only CFHR-2_{34} binds to ErpC with an affinity tight enough to allow purification of a complex between the two (Figure 7.44). No interaction was observed with the common N-terminal domains of CFHR-1, CFHR-2 and CFHR-5 or with

CFHR-1_{345} suggesting that the observation of CFHR-1 and CFHR-5 binding were secondary effects of binding CFHR-2. It is hoped that the atomic resolution structure of the complex between ErpC and CFHR-2_{34} will support the hypothesis that the CFHRs are bound in an inactive orientation.

All the work carried out so far has been using CFHR-1, CFHR-2 and CFHR-5. Further work is required to extend the research to see whether CFHR-3 and CFHR-4 behave similarly. Analysis of their sequences does not immediately suggest that heterodimers can form between CFHR-3 and CFHR4, defining another sub-family of CFHRs, but previous studies have shown them to bind C3b and heparin [Hellwage, Jokiranta *et al.* 1999], suggesting that they might participate in an identical mechanism to CFHR-1, CFHR-2 and CFHR-5. Further work should be carried out to characterise CFHR-3 and CFHR-4 in order to provide a full description of the function of this family of proteins.

7.14 Summary

Presented in this chapter were data that describe the previously unknown function of the CFHR family of proteins. A sub-family of CFHRs consisting of CFHR-1, CFHR-2 and CFHR-5 has been proposed based upon a common set of two N-terminal domains. The atomic resolution of these domains from CFHR-1 structure has been solved using X-ray crystallography showing formation of a dimeric species mediated by a novel hydrophobic dimerization motif. Subsequent investigation of CFHR-1 confirmed that the full length protein is dimeric in solution alongside showing that heterodimers can form between CFHR-1, CFHR-2 and CFHR-5. Mutations that successfully ablate dimerization were introduced into CFHR-5 and *in vivo* studies showed that this resulted in dramatically reduced accumulation in the glomerular basement membrane, supporting the suggestion that CFHR dimers are critical to their function.

The three C-terminal domains of CFHR-1 and the two C-terminal domains of CFHR-2 are capable of binding C3b with an affinity similar to the C-terminal domains of factor H. However, only CFHR-1 has the ability to bind heparin, implying both proteins have different glycosaminoglycan binding capabilities. The atomic resolution structure of the two C-terminal domains of CFHR-2 has been solved showing conservation of the C3b binding interface but ablation of the heparin binding groove described in the published structures of factor H domains 19 and 20 in complex with C3d rationalising these differences. Combination of these findings have described a novel common architecture between CFHR-1, CFHR-2 and CFHR-5 which describes formation of both homo and heterodimers between these proteins mediated by their two N-terminal domains whilst their C-terminal domains all possess C3b binding activity and differential glycosaminoglycan binding properties.

A mechanism for the previously unknown function of the CFHR family of proteins has been proposed. The CFHRs have been shown to compete for factor H binding to C3b resulting in increased complement activity leading to a proposed mechanism of maintenance of complement activation homeostasis driven by regulation by factor H and de-regulation by the CFHRs. Under normal circumstances the regime results in protection of host cells due to more regulation by factor H than de-regulation. However, if the level of deposited C3b on a cell surface increases following activation, the effective affinity of the CFHRs for C3b is amplified due to avidity arising from their dimeric nature. This results in increased competition for factor H and tips the homeostatic balance even further into activation.

Due to the lack of convincing functional data for the CFHRs, it was previously though that they were unimportant. However, discovery of this mechanism suggests that they play a vital role in maintaining homeostasis. Polymorphisms in the CFHRs resulting in

duplication of the dimerization domains has been associated with development of diseases resulting from improper regulation of complement. It has been hypothesised that these molecules are able to form higher order oligomers than dimers, resulting in increased de-regulation of complement due to increased avidity for C3b. Data have been presented that support this hypothesis, showing that CFHRs with the dimerization duplication polymorphisms result in increased complement activation at physiological concentrations. This has led to a proposed therapy for individuals possessing these polymorphisms by depletion of CFHRs in their serum.

7.14.1 Future directions

Potential crystals of the three C-terminal domains of CFHR-1 have been grown and it is hoped that these will yield the atomic resolution structure of these domains in order to confirm that the minor sequence differences to factor H domains 19 and 20 do not cause major structural modifications. Full length CFHR-1 was successfully produced in small quantities using mammalian cell culture and further efforts to use this expression system should be pursued in an attempt to obtain the structure and characterise the intact molecule. This may yield information regarding the flexibility and structural architecture of the molecule and a similar approach should also be employed to further investigate CFHR-2 and CFHR-5.

An antibody has been raised against the C-terminal domains of CFHR-2 and has been shown to have no cross-reactivity to factor H or CFHR-1. The serum abundance of CFHR-2 has never been reliably reported. This antibody now provides a way to accurately measure the abundance of CFHR-2 as well as allowing analysis and quantification of the heterodimeric species described earlier.

The proposed mechanism accounts for CFHR-1, CFHR-2 and CFHR-5. Further investigation is required to investigate whether CFHR-3 and CFHR-4 also form dimers and de-regulate factor H in a similar manner. The mechanism also questions the methodology behind bacterial scavenging of the CFHRs and suggests that they bind them in a manner that renders them unable to de-regulate in order to prevent increased complement activity on their surfaces. The discovery of CFHR heterodimers also questions which CFHRs are being actively bound on the bacterial cell surfaces requiring further investigations into these mechanisms and rationalisation of which proteins are actually bound. This investigation has been started by the purification of a complex between the C-terminal domains of CFHR-2 and ErpC from *Borrelia burgdorferi* and attempts will be made to obtain the atomic resolution structure of this molecule.

A homozygous deletion of CFHR-1 and CFHR-3 has been linked with both protective activity and increased susceptibility towards certain complement related diseases [Hughes, Orr et al. 2006; Gharavi, Kiryluk et al. 2011; Zhao, Wu et al. 2011]. CFHRs purified from these individuals have been shown to result in increased haemolysis, likely due to a higher concentration of CFHR-5 homodimers. This helps to explain the increased levels of susceptibility to some diseases, but required further investigation to understand how this deletion can also be protective towards others.

Bibliography

(2007). Centres for Disease Control and Prevention. MMWR Morb. Mortal. Wkly Rep. **56:** 573-576.

Abrahams, J. P. and A. G. W. Leslie (1996). "Methods used in the structure determination of bovine mitochondrial F-1 ATPase." Acta Crystallographica Section D-Biological Crystallography **52**: 30-42.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Akasako, A., M. Haruki, M. Oobatake and S. Kanaya (1997). "Conformational stabilities of Escherichia coli RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core." Journal of Biological Chemistry **272**(30): 18686-18693.

Anderson, D. H., R. F. Mullins, G. S. Hageman and L. V. Johnson (2002). "A role for local inflammation in the formation of drusen in the aging eye." Am J Ophthalmol **134**(3): 411-431.

Appel, G. B., H. T. Cook, G. Hageman, J. C. Jennette, M. Kashgarian, M. Kirschfink, J. D. Lambris, L. Lanning, H. U. Lutz, S. Meri, N. R. Rose, D. J. Salant, S. Sethi, R. J. Smith, W. Smoyer, H. F. Tully, S. P. Tully, P. Walker, M. Welsh, R. Wurzner and P. F. Zipfel (2005). "Membranoproliferative glomerulonephritis type II (dense deposit disease): an update." J Am Soc Nephrol **16**(5): 1392-1403.

Arndt, K. M., J. N. Pelletier, K. M. Muller, A. Pluckthun and T. Alber (2002). "Comparison of in vivo selection and rational design of heterodimeric coiled coils." Structure **10**(9): 1235-1248.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods **34**(3): 266-272.

Aslam, M. and S. J. Perkins (2001). "Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling." J Mol Biol **309**(5): 1117-1138.

Athanasiou, Y., K. Voskarides, D. P. Gale, L. Damianou, C. Patsias, M. Zavros, P. H. Maxwell, H. T. Cook, P. Demosthenous, A. Hadjisavvas, K. Kyriacou, I. Zouvani, A. Pierides and C. Deltas (2011). "Familial C3 glomerulopathy associated with CFHR5 mutations: clinical characteristics of 91 patients in 16 pedigrees." Clin J Am Soc Nephrol **6**(6): 1436-1446.

Bailey, S. (1994). "The Ccp4 Suite - Programs for Protein Crystallography." Acta Crystallographica Section D-Biological Crystallography **50**: 760-763.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-10041.

Barbour, A. G. and S. F. Hayes (1986). "Biology of Borrelia Species." Microbiological Reviews **50**(4): 381-400.

Barnum, S. R. and B. Dahlback (1990). "C4b-binding protein, a regulatory component of the classical pathway of complement, is an acute-phase protein and is elevated in systemic lupus erythematosus." Complement Inflamm **7**(2): 71-77.

Battye, T. G. G., L. Kontogiannis, O. Johnson, H. R. Powell and A. G. W. Leslie (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallographica Section D-Biological Crystallography **67**: 271-281.

Beernink, P. T. and D. M. Granoff (2008). "Bactericidal antibody responses induced by meningococcal recombinant chimeric factor H-binding protein vaccines." Infection and Immunity **76**(6): 2568-2575.

Behnsen, J., A. Hartmann, J. Schmaler, A. Gehrke, A. A. Brakhage and P. F. Zipfel (2008). "The opportunistic human pathogenic fungus Aspergillus fumigatus evades the host complement system." Infection and Immunity **76**(2): 820-827.

Benvenuti, M. and S. Mangani (2007). "Crystallization of soluble proteins in vapor diffusion for x-ray crystallography." Nature Protocols **2**(7): 1633-1651.

Blackmore, T. K., J. Hellwage, T. A. Sadlon, N. Higgs, P. F. Zipfel, H. M. Ward and D. L. Gordon (1998). "Identification of the second heparin-binding domain in human complement factor H." Journal of Immunology **160**(7): 3342-3348.

Blackmore, T. K., T. A. Sadlon, H. M. Ward, D. M. Lublin and D. L. Gordon (1996). "Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H." Journal of Immunology **157**(12): 5422-5427.

Blanc, E., P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea and G. Bricogne (2004). "Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT." Acta Crystallographica Section D-Biological Crystallography **60**: 2210-2221.

Blaum, B. S., J. A. Deakin, C. M. Johansson, A. P. Herbert, P. N. Barlow, M. Lyon and D. Uhrin (2010). "Lysine and arginine side chains in glycosaminoglycan-protein

complexes investigated by NMR, cross-linking, and mass spectrometry: a case study of the factor H-heparin interaction." J Am Chem Soc **132**(18): 6374-6381.

Blom, A. M., J. Webb, B. O. Villoutreix and B. Dahlback (1999). "A cluster of positively charged amino acids in the C4BP alpha-chain is crucial for C4b binding and factor I cofactor function." Journal of Biological Chemistry **274**(27): 19237-19245.

Blom, A. M., A. F. Zadura, B. O. Villoutreix and B. Dahlback (2000). "Positively charged amino acids at the interface between alpha-chain CCP1 and CCP2 of C4BP are required for regulation of the classical C3-convertase." Molecular Immunology **37**(8): 445-453.

Bricogne, G., E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O. S. Smart, C. Vonrhein and T. O. Womack (2011). "BUSTER version 2.11.2." Cambridge, United Kingdom: Global Phasing Ltd.

Bricogne, G., C. Vonrhein, C. Flensburg, M. Schiltz and W. Paciorek (2003). "Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0." Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 2023-2030.

Brodbeck, W. G., C. Mold, J. P. Atkinson and M. E. Medof (2000). "Cooperation between decay-accelerating factor and membrane cofactor protein in protecting cells from autologous complement attack." J Immunol **165**(7): 3999-4006.

Buchan, D. W., S. M. Ward, A. E. Lobley, T. C. Nugent, K. Bryson and D. T. Jones (2010). "Protein annotation and modelling servers at University College London." Nucleic Acids Res **38**(Web Server issue): W563-568.

Burkhard, P., J. Stetefeld and S. V. Strelkov (2001). "Coiled coils: a highly versatile protein folding motif." Trends Cell Biol **11**(2): 82-88.

Bykowski, T., M. E. Woodman, A. E. Cooley, C. A. Brissette, R. Wallich, V. Brade, P. Kraiczy and B. Stevenson (2008). "Borrelia burgdorferi complement regulator-acquiring surface proteins (BbCRASPs): Expression patterns during the mammal-tick infection cycle." International Journal of Medical Microbiology **298 Suppl 1**: 249-256.

Cantini, F., S. Savino, M. Scarselli, V. Masignani, M. Pizza, G. Romagnoli, E. Swennen, D. Veggi, L. Banci and R. Rappuoli (2006). "Solution structure of the immunodominant domain of protective antigen GNA1870 of Neisseria meningitidis." J Biol Chem **281**(11): 7220-7227.

Carlsson, S. and B. Dahlback (2010). "Dependence on vitamin K-dependent protein S for eukaryotic cell secretion of the beta-chain of C4b-binding protein." J Biol Chem **285**(42): 32038-32046.

Carroll, M. C. (2004). "The complement system in B cell regulation." Molecular Immunology **41**(2-3): 141-146.

Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D-Biological Crystallography **66**: 12-21.

Chung, L. P., D. R. Bentley and K. B. Reid (1985). "Molecular cloning and characterization of the cDNA coding for C4b-binding protein, a regulatory protein of the classical pathway of the human complement system." Biochem J **230**(1): 133-141.

Clark, S. J., V. A. Higman, B. Mulloy, S. J. Perkins, S. M. Lea, R. B. Sim and A. J. Day (2006a). "His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-

disease-associated form." Journal of Biological Chemistry **281**(34): 24713-24720.

Clark, S. J., V. A. Higman, B. Mulloy, S. J. Perkins, S. M. Lea, R. B. Sim and A. J. Day (2006b). "His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form." Journal of Biological Chemistry **281**(34): 24713-24720.

Comp, P. C., R. R. Nixon, M. R. Cooper and C. T. Esmon (1984). "Familial protein S deficiency is associated with recurrent thrombosis." J Clin Invest **74**(6): 2082-2088.

Cordes, F. S., P. Kraiczy, P. Roversi, M. M. Simon, V. Brade, O. Jahraus, R. Wallis, L. Goodstadt, C. P. Ponting, C. Skerka, P. F. Zipfel, R. Wallich and S. M. Lea (2006). "Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of Borrelia burgdorferi." International Journal of Medical Microbiology **296**: 177-184.

Cordes, F. S., P. Roversi, P. Kraiczy, M. M. Simon, V. Brade, O. Jahraus, R. Wallis, C. Skerka, P. F. Zipfel, R. Wallich and S. M. Lea (2005). "A novel fold for the factor H-binding protein BbCRASP-1 of Borrelia burgdorferi." Nature Structural & Molecular Biology **12**(3): 276-277.

Cowtan, K. (1998). "Modified phased translation functions and their application to molecular-fragment location." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 750-756.

Cowtan, K. (1999). "Error estimation and bias correction in phase-improvement calculations." Acta Crystallographica Section D-Biological Crystallography **55**: 1555-1567.

Cowtan, K. (2001). "Fast Fourier feature recognition." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1435-1444.

Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." Acta Crystallogr D Biol Crystallogr **62**(Pt 9): 1002-1011.

Cowtan, K. (2010). "Recent developments in classical density modification." Acta Crystallographica Section D-Biological Crystallography **66**: 470-478.

Criado Garcia, O., P. Sanchez-Corral and S. Rodriguez de Cordoba (1995). "Isoforms of human C4b-binding protein. II. Differential modulation of the C4BPA and C4BPB genes by acute phase cytokines." J Immunol **155**(8): 4037-4043.

Crick, F. H. C. (1953). "The Packing of Alpha-Helices - Simple Coiled-Coils." Acta Crystallographica **6**(8-9): 689-697.

Cromer, D. T. (1983). "Calculation of Anomalous Scattering Factors at Arbitrary Wavelengths." Journal of Applied Crystallography **16**(Aug): 437-437.

Dahlback, B. (1983). "Purification of human C4b-binding protein and formation of its complex with vitamin K-dependent protein S." Biochem J **209**(3): 847-856.

Dahlback, B. and H. J. Muller-Eberhard (1984). "Ultrastructure of C4b-Binding Protein-Fragments Formed by Limited Proteolysis Using Chymotrypsin." Journal of Biological Chemistry **259**(19): 1631-1634.

Dahlback, B., C. A. Smith and H. J. Muller-Eberhard (1983). "Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b." Proc Natl Acad Sci U S A **80**(11): 3461-3465.

Dahlback, B. and J. Stenflo (1981). "High molecular weight complex in human plasma between vitamin K-dependent protein S and complement component C4b-binding protein." Proc Natl Acad Sci U S A **78**(4): 2512-2516.

Dhillon, B., A. F. Wright, A. Tufail, I. Pappworth, C. Hayward, I. Moore, L. Strain, D. Kavanagh, P. N. Barlow, A. P. Herbert, C. Q. Schmidt, A. M. Armbrecht, A. Laude, I. J. Deary, S. J. Staniforth, L. V. Holmes, T. H. Goodship and K. J. Marchbank (2010). "Complement factor h autoantibodies and age-related macular degeneration." Invest Ophthalmol Vis Sci **51**(11): 5858-5863.

DiMaio, F., M. D. Tyka, M. L. Baker, W. Chiu and D. Baker (2009). "Refinement of protein structures into low-resolution density maps using rosetta." Journal of Molecular Biology **392**(1): 181-190.

Doolittle, R. F. (1989). Redundancies in protein sequences. New York, Plenum press.

Dragon-Durey, M. A., V. Fremeaux-Bacchi, C. Loirat, J. Blouin, P. Niaudet, G. Deschenes, P. Coppo, W. Herman Fridman and L. Weiss (2004). "Heterozygous and homozygous factor h deficiencies associated with hemolytic uremic syndrome or membranoproliferative glomerulonephritis: report and genetic analysis of 16 cases." J Am Soc Nephrol **15**(3): 787-795.

Draper, S. J., A. C. Moore, A. L. Goodman, C. A. Long, A. A. Holder, S. C. Gilbert, F. Hill and A. V. Hill (2008). "Effective induction of high-titer antibodies by viral vector vaccines." Nat Med **14**(8): 819-821.

Embers, M. E., R. Ramamoorthy and M. T. Philipp (2004). "Survival strategies of Borrelia burgdorferi, the etiologic agent of Lyme disease." Microbes and Infection **6**(3): 312-318.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallographica Section D-Biological Crystallography **66**: 486-501.

Evans, G. and R. F. Pettifer (2001). "CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra." Journal of Applied Crystallography **34**: 82-86.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallographica Section D-Biological Crystallography **62**: 72-82.

Evenas, P., P. Garcia De Frutos, S. Linse and B. Dahlback (1999). "Both G-type domains of protein S are required for the high-affinity interaction with C4b-binding protein." Eur J Biochem **266**(3): 935-942.

Feavers, I. M. (2000). "Bacterial genomics - ABC of meningococcal diversity." Nature **404**(6777): 451-452.

Ferreira, V. P., A. P. Herbert, H. G. Hocking, P. N. Barlow and M. K. Pangburn (2006). "Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces." J Immunol **177**(9): 6308-6316.

Fletcher, L. D., L. Bernfield, V. Barniak, J. E. Farley, A. Howell, M. Knauf, P. Ooi, R. P. Smith, P. Weise, M. Wetherell, X. L. Xie, R. Zagursky, Y. Zhang and G. W. Zlotnick (2004). "Vaccine potential of the Neisseria meningitidis 2086 lipoprotein." Infection and Immunity **72**(4): 2088-2100.

Foltyn Zadura, A., P. F. Zipfel, M. Bokareva, G. Sturfelt, A. Jonsen, S. Nilsson, A. Hillarp, T. Saxne, L. A. Trouw and A. M. Blom (2012). "Factor H autoantibodies and deletion of complement factor H-related protein-1 in rheumatic diseases in comparison to atypical hemolytic uremic syndrome." Arthritis Res Ther **14**(4): R185.

Fontaine, M., M. J. Demares, V. Koistinen, A. J. Day, C. Davrinche, R. B. Sim and J. Ripoche (1989). "Truncated forms of human complement factor H." Biochemical Journal **258**(3): 927-930.

Forneris, F., D. Ricklin, J. Wu, A. Tzekou, R. S. Wallace, J. D. Lambris and P. Gros (2010). "Structures of C3b in Complex with Factors B and D Give Insight into Complement Convertase Formation." Science **330**(6012): 1816-1820.

Fremeaux-Bacchi, V., M. A. Dragon-Durey, J. Blouin, C. Vigneau, D. Kuypers, B. Boudailliez, C. Loirat, E. Rondeau and W. H. Fridman (2004). "Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome." J Med Genet **41**(6): e84.

Fremeaux-Bacchi, V., E. A. Moulton, D. Kavanagh, M. A. Dragon-Durey, J. Blouin, A. Caudy, N. Arzouk, R. Cleper, M. Francois, G. Guest, J. Pourrat, R. Seligman, W. H. Fridman, C. Loirat and J. P. Atkinson (2006). "Genetic and functional analyses of membrane cofactor protein (CD46) mutations in atypical hemolytic uremic syndrome." J Am Soc Nephrol **17**(7): 2017-2025.

Friedman, D. S., B. J. O'Colmain, B. Munoz, S. C. Tomany, C. McCarty, P. T. de Jong, B. Nemesure, P. Mitchell and J. Kempen (2004). "Prevalence of age-related macular degeneration in the United States." Arch Ophthalmol **122**(4): 564-572.

Gaboriaud, C., N. M. Thielens, L. A. Gregory, V. Rossi, J. C. Fontecilla-Camps and G. J. Arlaud (2004). "Structure and activation of the C1 complex of complement: unraveling the puzzle." Trends Immunol **25**(7): 368-373.

Gale, D. P., E. G. de Jorge, H. T. Cook, R. Martinez-Barricarte, A. Hadjisavvas, A. G. McLean, C. D. Pusey, A. Pierides, K. Kyriacou, Y. Athanasiou, K. Voskarides, C. Deltas, A. Palmer, V. Fremeaux-Bacchi, S. R. de Cordoba, P. H. Maxwell and M. C. Pickering (2010). "Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis." Lancet **376**(9743): 794-801.

Garcia de Frutos, P., R. I. Alim, Y. Hardig, B. Zoller and B. Dahlback (1994). "Differential regulation of alpha and beta chains of C4b-binding protein during acute-phase response resulting in stable plasma levels of free anticoagulant protein S." Blood **84**(3): 815-822.

Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). "Protein Identification and Analysis Tools on the ExPASy Server." The Proteomics Protocols Handbook, Humana Press.

Gerhard, D. S., L. Wagner, E. A. Feingold, C. M. Shenmen, L. H. Grouse, G. Schuler, S. L. Klein, S. Old, R. Rasooly, P. Good, M. Guyer, A. M. Peck, J. G. Derge, D. Lipman, F. S. Collins, W. Jang, S. Sherry, M. Feolo, L. Misquitta, E. Lee, K. Rotmistrovsky, S. F. Greenhut, C. F. Schaefer, K. Buetow, T. I. Bonner, D. Haussler, J. Kent, M. Kiekhaus, T. Furey, M. Brent, C. Prange, K. Schreiber, N. Shapiro, N. K. Bhat, R. F. Hopkins, F. Hsie, T. Driscoll, M. B. Soares, T. L. Casavant, T. E. Scheetz, M. J. Brown-stein, T. B. Usdin, S. Toshiyuki, P. Carninci, Y. Piao, D. B. Dudekula, M. S. Ko, K. Kawakami, Y. Suzuki, S. Sugano, C. E. Gruber, M. R. Smith, B. Simmons, T. Moore, R. Waterman, S. L. Johnson, Y. Ruan, C. L. Wei, S. Mathavan, P. H. Gunaratne, J. Wu, A. M. Garcia, S. W. Hulyk, E. Fuh, Y. Yuan, A. Sneed, C. Kowis, A. Hodgson, D. M. Muzny, J. McPherson, R. A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madari, A. C. Young, K. D. Wetherby, S. J. Granite, P. N. Kwong, C. P. Brinkley, R. L. Pearson, G. G. Bouffard, R. W. Blakesly, E. D. Green, M. C. Dickson, A. C. Rodriguez, J. Grimwood, J. Schmutz, R. M. Myers, Y. S. Butterfield, M. Griffith, O. L. Griffith, M. I. Krzywinski, N. Liao, R. Morin, D. Palmquist, A. S. Petrescu, U. Skalska, D. E. Smailus, J. M. Stott, A. Schnerch, J. E. Schein, S. J. Jones, R. A.

Holt, A. Baross, M. A. Marra, S. Clifton, K. A. Makowski, S. Bosak and J. Malek (2004). "The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)." Genome Res **14**(10B): 2121-2127.

Gern, L., A. Estrada-Pena, F. Frandsen, J. S. Gray, T. G. Jaenson, F. Jongejan, O. Kahl, E. Korenberg, R. Mehl and P. A. Nuttall (1998). "European reservoir hosts of Borrelia burgdorferi sensu lato." Zentralbl Bakteriol **287**(3): 196-204.

Gharavi, A. G., K. Kiryluk, M. Choi, Y. Li, P. Hou, J. Xie, S. Sanna-Cherchi, C. J. Men, B. A. Julian, R. J. Wyatt, J. Novak, J. C. He, H. Wang, J. Lv, L. Zhu, W. Wang, Z. Wang, K. Yasuno, M. Gunel, S. Mane, S. Umlauf, I. Tikhonova, I. Beerman, S. Savoldi, R. Magistroni, G. M. Ghiggeri, M. Bodria, F. Lugani, P. Ravani, C. Ponticelli, L. Allegri, G. Boscutti, G. Frasca, A. Amore, L. Peruzzi, R. Coppo, C. Izzi, B. F. Viola, E. Prati, M. Salvadori, R. Mignani, L. Gesualdo, F. Bertinetto, P. Mesiano, A. Amoroso, F. Scolari, N. Chen, H. Zhang and R. P. Lifton (2011). "Genome-wide association study identifies susceptibility loci for IgA nephropathy." Nat Genet **43**(4): 321-327.

Gigli, I., T. Fujita and V. Nussenzweig (1979). "Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator." Proc Natl Acad Sci U S A **76**(12): 6596-6600.

Giuliani, M. M., L. Santini, B. Brunelli, A. Biolchi, B. Arico, F. Di Marcello, E. Cartocci, M. Comanducci, V. Masignani, L. Lozzi, S. Savino, M. Scarselli, R. Rappuoli and M. Pizza (2005). "The region comprising amino acids 100 to 255 of Neisseria meningitidis lipoprotein GNA 1870 elicits bactericidal antibodies." Infection and Immunity **73**(2): 1151-1160.

Goicoechea de Jorge, E., C. L. Harris, J. Esparza-Gordillo, L. Carreras, E. A. Arranz, C. A. Garrido, M. Lopez-Trascasa, P. Sanchez-Corral, B. P. Morgan and S. Rodriguez de Cordoba (2007). "Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome." Proc Natl Acad Sci U S A **104**(1): 240-245.

Gray, S. J., C. L. Trotter, M. E. Ramsay, M. Guiver, A. J. Fox, R. Borrow, R. H. Mallard and E. B. Kaczmarski (2006). "Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit." Journal of Medical Microbiology **55**(7): 887-896.

Griffin, J. H., A. Gruber and J. A. Fernandez (1992). "Reevaluation of total, free, and bound protein S and C4b-binding protein levels in plasma anticoagulated with citrate or hirudin." Blood **79**(12): 3203-3211.

Grosse-Kunstleve, R. W. and P. D. Adams (2003). "Substructure search procedures for macromolecular structures." Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1966-1973.

Haines, J. L., M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins, K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E. A. Postel and M. A. Pericak-Vance (2005). "Complement factor H variant increases the risk of age-related macular degeneration." Science **308**(5720): 419-421.

Hammerschmidt, C., T. Hallstrom, C. Skerka, R. Wallich, B. Stevenson, P. F. Zipfel and P. Kraiczy (2012). "Contribution of the infection-associated complement regulator-acquiring surface protein 4 (ErpC) to complement resistance of Borrelia burgdorferi." Clin Dev Immunol **2012**: 349657.

Harrison, L. H. (2006). "Prospects for vaccine prevention of meningococcal infection." Clinical Microbiology Reviews **19**(1): 142-+.

Haupt, K., P. Kraiczy, R. Wallich, V. Brade, C. Skerka and P. F. Zipfel (2007). "Binding of human factor H-related protein 1 to serum-resistant Borrelia burgdorferi is mediated by borrelial complement regulator-acquiring surface proteins." J Infect Dis **196**(1): 124-133.

Hayward, S. and H. J. C. Berendsen (1998). "Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme." Proteins-Structure Function and Genetics **30**(2): 144-154.

Hebecker, M. and M. Jozsi (2012). "Factor H-related protein 4 activates complement by serving as a platform for the assembly of alternative pathway C3 convertase via its interaction with C3b protein." J Biol Chem **287**(23): 19528-19536.

Heinen, S., A. Hartmann, N. Lauer, U. Wiehl, H. M. Dahse, S. Schirmer, K. Gropp, T. Enghardt, R. Wallich, S. Halbich, M. Mihlan, U. Schlotzer-Schrehardt, P. F. Zipfel and C. Skerka (2009). "Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation." Blood **114**(12): 2439-2447.

Hellwage, J., T. S. Jokiranta, V. Koistinen, O. Vaarala, S. Meri and P. F. Zipfel (1999). "Functional properties of complement factor H-related proteins FHR-3 and FHR-4: binding to the C3d region of C3b and differential regulation by heparin." FEBS Lett **462**(3): 345-352.

Hellwage, J., S. Kuhn and P. F. Zipfel (1997). "The human complement regulatory factor-H-like protein 1, which represents a truncated form of factor H, displays cell-attachment activity." Biochemical Journal **326 (Pt 2)**: 321-327.

Herbert, A. P., J. A. Deakin, C. Q. Schmidt, B. S. Blaum, C. Egan, V. P. Ferreira, M. K. Pangburn, M. Lyon, D. Uhrin and P. N. Barlow (2007). "Structure shows that a glycosaminoglycan and protein recognition site in factor H is perturbed by age-related macular degeneration-linked single nucleotide polymorphism." Journal of Biological Chemistry **282**(26): 18960-18968.

Herbert, A. P., D. Uhrin, M. Lyon, M. K. Pangburn and P. N. Barlow (2006). "Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure." Journal of Biological Chemistry **281**(24): 16512-16520.

Hessing, M., R. A. Vlooswijk, T. M. Hackeng, D. Kanters and B. N. Bouma (1990). "The localization of heparin-binding fragments on human C4b-binding protein." Journal of Immunology **144**(1): 204-208.

Hillarp, A. and B. Dahlback (1988). "Novel subunit in C4b-binding protein required for protein S binding." The Journal of biological chemistry **263**(25): 12759-12764.

Hillarp, A. and B. Dahlback (1990). "Cloning of cDNA coding for the beta chain of human complement component C4b-binding protein: sequence homology with the alpha chain." Proc Natl Acad Sci U S A **87**(3): 1183-1187.

Hillarp, A., M. Hessing and B. Dahlback (1989). "Protein S binding in relation to the subunit composition of human C4b-binding protein." FEBS Lett **259**(1): 53-56.

Hillarp, A., F. Pardo-Manuel, R. R. Ruiz, S. Rodriguez de Cordoba and B. Dahlback (1993). "The human C4b-binding protein beta-chain gene." J Biol Chem **268**(20): 15017-15023.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Res **38**(Web Server issue): W545-549.

Hostetter, M. K., R. A. Krueger and D. J. Schmeling (1984). "The biochemistry of opsonization: central role of the reactive thiolester of the third component of complement." J Infect Dis **150**(5): 653-661.

Hourcade, D., V. M. Holers and J. P. Atkinson (1989). "The regulators of complement activation (RCA) gene cluster." Advances in Immunology, Vol 104 **45**: 381-416.

Howell, P. L. and G. D. Smith (1992). "Identification of Heavy-Atom Derivatives by Normal Probability Methods." Journal of Applied Crystallography **25**: 81-86.

Hughes, A. E., N. Orr, H. Esfandiary, M. Diaz-Torres, T. Goodship and U. Chakravarthy (2006). "A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration." Nat Genet **38**(10): 1173-1177.

Ikeda, K., T. Sannoh, N. Kawasaki, T. Kawasaki and I. Yamashina (1987). "Serum lectin with known structure activates complement through the classical pathway." J Biol Chem **262**(16): 7451-7454.

Incardona, M. F., G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov and O. Svensson (2009). "EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis." J Synchrotron Radiat **16**(Pt 6): 872-879.

Isenman, D. E. (1983). "The role of the thioester bond in C3 and C4 in the determination of the conformational and functional states of the molecule." Ann N Y Acad Sci **421**: 277-290.

Janssen, B. J., L. Gomes, R. I. Koning, D. I. Svergun, A. J. Koster, D. C. Fritzinger, C. W. Vogel and P. Gros (2009). "Insights into complement convertase formation based on the structure of the factor B-cobra venom factor complex." EMBO J **28**(16): 2469-2478.

Jenkins, H. T., L. Mark, G. Ball, J. Persson, G. Lindahl, D. Uhrin, A. M. Blom and P. N. Barlow (2006). "Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor." Journal of Biological Chemistry **281**(6): 3690-3697.

Jokiranta, T. S., V. P. Jaakola, M. J. Lehtinen, M. Parepalo, S. Meri and A. Goldman (2006). "Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome." EMBO J **25**(8): 1784-1794.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." Journal of Molecular Biology **292**(2): 195-202.

Kabsch, W. (1976). "Solution for Best Rotation to Relate 2 Sets of Vectors." Acta Crystallographica Section A **32**(Sep1): 922-923.

Kabsch, W. (2010). "Xds." Acta Crystallographica Section D-Biological Crystallography **66**: 125-132.

Kajander, T., M. J. Lehtinen, S. Hyvarinen, A. Bhattacharjee, E. Leung, D. E. Isenman, S. Meri, A. Goldman and T. S. Jokiranta (2011). "Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement." Proc Natl Acad Sci U S A **108**(7): 2897-2902.

Kammerer, R. A., T. Schulthess, R. Landwehr, A. Lustig, J. Engel, U. Aebi and M. O. Steinmetz (1998). "An autonomous folding unit mediates the assembly of two-stranded coiled coils." Proc Natl Acad Sci U S A **95**(23): 13419-13424.

Kazatchkine, M. D., D. T. Fearon and K. F. Austen (1979). "Human Alternative Complement Pathway - Membrane-Associated Sialic-Acid Regulates the Competition between B and Beta-1h for Cell-Bound C3b." Journal of Immunology **122**(1): 75-81.

Kemper, C. and D. E. Hourcade (2008). "Properdin: New roles in pattern recognition and target clearance." Molecular Immunology **45**(16): 4048-4056.

Kerr, M. A. (1980). "The human complement system: assembly of the classical pathway C3 convertase." Biochemical Journal **189**(1): 173-181.

Kidmose, R. T., N. S. Laursen, J. Dobo, T. R. Kjaer, S. Sirotkina, L. Yatime, L. Sottrup-Jensen, S. Thiel, P. Gal and G. R. Andersen (2012). "Structural basis for activation of the complement system by component C4 cleavage." Proc Natl Acad Sci U S A **109**(38): 15425-15430.

Kimberley, F. C., B. Sivasankar and B. P. Morgan (2007). "Alternative roles for CD59." Molecular Immunology **44**(1-3): 73-81.

Kohler, P. F. and H. J. Muller-Eberhard (1967). "Immunochemical Quantitation of Third Fourth and Fifth Components of Human Complement - Concentrations in Serum of Healthy Adults." Journal of Immunology **99**(6): 1211-&.

Kraiczy, P., C. Hanssen-Hubner, V. Kitiratschky, C. Brenner, S. Besier, V. Brade, M. M. Simon, C. Skerka, P. Roversi, S. M. Lea, B. Stevenson, R. Wallich and P. F. Zipfel (2009). "Mutational analyses of the BbCRASP-1 protein of Borrelia burgdorferi identify residues relevant for the architecture and binding of host complement regulators FHL-1 and factor H." International Journal of Medical Microbiology **299**(4): 255-268.

Kraiczy, P., J. Hellwage, C. Skerka, H. Becker, M. Kirschfink, M. M. Simon, V. Brade, P. F. Zipfel and R. Wallich (2004). "Complement resistance of Borrelia burgdorferi correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins." Journal of Biological Chemistry **279**(4): 2421-2429.

Kraiczy, P., C. Skerka, V. Brade and P. F. Zipfel (2001a). "Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi." Infection and Immunity **69**(12): 7800-7809.

Kraiczy, P., C. Skerka, V. Brade and P. F. Zipfel (2001b). "Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi." Infection and Immunity **69**(12): 7800-7809.

Kraiczy, P., C. Skerka, M. Kirschfink, V. Brade and P. F. Zipfel (2001c). "Immune evasion of Borrelia burgdorferi by acquisition of human complement regulators FHL-1/reconectin and Factor H." European Journal of Immunology **31**(6): 1674-1684.

Kraiczy, P., C. Skerka, M. Kirschfink, P. F. Zipfel and V. Brade (2001d). "Mechanism of complement resistance of pathogenic Borrelia burgdorferi isolates." International Immunopharmacology **1**(3): 393-401.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallographica Section D-Biological Crystallography **60**: 2256-2268.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of Molecular Biology **372**(3): 774-797.

Kristensen, T., R. A. Wetsel and B. F. Tack (1986). "Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2." Journal of Immunology **136**(9): 3407-3411.

Kuhn, S., C. Skerka and P. F. Zipfel (1995). "Mapping of the Complement Regulatory Domains in the Human Factor-H-Like Protein-1 and in Factor-H." Journal of Immunology **155**(12): 5663-5670.

Kuhn, S. and P. F. Zipfel (1996). "Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H." Eur J Immunol **26**(10): 2383-2387.

Kunert, A., J. Losse, C. Gruszin, M. Huhn, K. Kaendler, S. Mikkat, D. Volke, R. Hoffmann, T. S. Jokiranta, H. Seeberger, U. Moellmann, J. Hellwage and P. F. Zipfel (2007). "Immune evasion of the human pathogen Pseudomonas aeruginosa: elongation factor Tuf is a factor H and plasminogen binding protein." J Immunol **179**(5): 2979-2988.

Kvalsvig, A. J. and D. J. Unsworth (2003). "The immunopathogenesis of meningococcal disease." Journal of Clinical Pathology **56**(6): 417-422.

Lachmann, P. J. (2009). "The Amplification Loop of the Complement Pathways." Advances in Immunology, Vol 104 **104**: 115-149.

Lamberth, S., H. Schmid, M. Muenchbach, T. Vorherr, J. Krebs, E. Carafoli and C. Griesinger (2000). "NMR solution structure of phospholamban." Helvetica Chimica Acta **83**(9): 2141-2152.

Lambris, J. D., Z. Lao, T. J. Oglesby, J. P. Atkinson, C. E. Hack and J. D. Becherer (1996). "Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component." J Immunol **156**(12): 4821-4832.

Langer, G., S. X. Cohen, V. S. Lamzin and A. Perrakis (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." Nature Protocols **3**(7): 1171-1179.

Law, S. K. and A. W. Dodds (1990a). "C3, C4 and C5: the thioester site." Biochem Soc Trans **18**(6): 1155-1159.

Law, S. K. A. and A. W. Dodds (1990b). "C3, C4 and C5 - the Thioester Site." Biochemical Society Transactions **18**(6): 1155-1159.

Law, S. K. A. and A. W. Dodds (1997). "The internal thioester and the covalent binding properties of the complement proteins C3 and C4." Protein Science **6**(2): 263-274.

Leslie, A. G. W. and H. R. Powell (2007). "Processing diffraction data with MOSFLM." Evolving Methods for Macromolecular Crystallography **245**: 41-51.

Liu, J., Q. Zheng, Y. Deng, C. S. Cheng, N. R. Kallenbach and M. Lu (2006). "A seven-helix coiled coil." Proc Natl Acad Sci U S A **103**(42): 15457-15462.

Lo, H., C. M. Tang and R. M. Exley (2009). "Mechanisms of avoidance of host immunity by Neisseria meningitidis and its effect on vaccine development." Lancet Infect Dis **9**(7): 418-427.

Lupas, A., M. Van Dyke and J. Stock (1991). "Predicting coiled coils from protein sequences." Science **252**(5009): 1162-1164.

Madico, G., J. A. Welsch, L. A. Lewis, A. McNaughton, D. H. Perlman, C. E. Costello, J. Ngampasutadol, U. Vogel, D. M. Granoff and S. Ram (2006). "The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance." Journal of Immunology **177**(1): 501-510.

Malik, T. H., P. J. Lavin, E. Goicoechea de Jorge, K. A. Vernon, K. L. Rose, M. P. Patel, M. de Leeuw, J. J. Neary, P. J. Conlon, M. P. Winn and M. C. Pickering (2012). "A hybrid CFHR3-1 gene causes familial C3 glomerulopathy." J Am Soc Nephrol **23**(7): 1155-1160.

Manabe, T. and Y. Jin (2008). "Noncovalent interactions in human plasma proteins analyzed by the comparison of non-denaturing and denaturing micro-2-D gel

electrophoresis patterns after polypeptide assignment using matrix-assisted laser desorption/ionization-mass spectrometry and peptide mass fingerprinting." Electrophoresis **29**(12): 2672-2688.

Marshall, J. S. (2004). "Mast-cell responses to pathogens." Nat Rev Immunol **4**(10): 787-799.

Masignani, V., M. Comanducci, M. M. Giuliani, S. Bambini, J. Adu-Bobie, B. Arico, B. Brunelli, A. Pieri, L. Santini, S. Savino, D. Serruto, D. Litt, S. Kroll, J. A. Welsch, D. M. Granoff, R. Rappuoli and M. Pizza (2003). "Vaccination against Neisseria meningitidis using three variants of the lipoprotein GNA1870." Journal of Experimental Medicine **197**(6): 789-799.

Mason, J. M. and K. M. Arndt (2004). "Coiled coil domains: Stability, specificity, and biological implications." Chembiochem **5**(2): 170-176.

Matsushita, M., Y. Endo, N. Hamasaki and T. Fujita (2001). "Activation of the lectin complement pathway by ficolins." International Immunopharmacology **1**(3): 359-363.

Matthews, B. W. (1968). "Solvent Content of Protein Crystals." Journal of Molecular Biology **33**(2): 491-&.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**: 658-674.

McDowell, J. V., M. E. Harlin, E. A. Rogers and R. T. Marconi (2005). "Putative coiled-coil structural elements of the BBA68 protein of Lyme disease spirochetes are required for formation of its factor H binding site." J Bacteriol **187**(4): 1317-1323.

McRae, J. L., P. J. Cowan, D. A. Power, K. I. Mitchelhill, B. E. Kemp, B. P. Morgan and B. F. Murphy (2001). "Human factor H-related protein 5 (FHR-5). A new complement-associated protein." J Biol Chem **276**(9): 6747-6754.

McRae, J. L., T. G. Duthy, K. M. Griggs, R. J. Ormsby, P. J. Cowan, B. A. Cromer, W. J. McKinstry, M. W. Parker, B. F. Murphy and D. L. Gordon (2005). "Human factor H-related protein 5 has cofactor activity, inhibits C3 convertase activity, binds heparin and C-reactive protein, and associates with lipoprotein." J Immunol **174**(10): 6250-6256.

Meri, S. and M. K. Pangburn (1990). "Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H." Proc Natl Acad Sci U S A **87**(10): 3982-3986.

Milder, F. J., H. C. Raaijmakers, M. D. Vandeputte, A. Schouten, E. G. Huizinga, R. A. Romijn, W. Hemrika, A. Roos, M. R. Daha and P. Gros (2006). "Structure of complement component C2A: implications for convertase formation and substrate binding." Structure **14**(10): 1587-1597.

Mizushima, S. and S. Nagata (1990). "pEF-BOS, a powerful mammalian expression vector." Nucleic Acids Res **18**(17): 5322.

Morgan, B. P. (1995). "Physiology and pathophysiology of complement: progress and trends." Crit Rev Clin Lab Sci **32**(3): 265-298.

Morgan, H. P., C. Q. Schmidt, M. Guariento, B. S. Blaum, D. Gillespie, A. P. Herbert, D. Kavanagh, H. D. Mertens, D. I. Svergun, C. M. Johansson, D. Uhrin, P. N. Barlow and J. P. Hannan (2011). "Structural basis for engagement by complement factor H of C3b on a self surface." Nat Struct Mol Biol **18**(4): 463-470.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Musingarimi, P., M. E. Plumb and J. M. Sodetz (2002). "Interaction between the C8 alpha-gamma and C8 beta subunits of human complement C8: role of the C8 beta N-terminal thrombospondin type 1 module and membrane attack complex/perforin domain." Biochemistry **41**(37): 11255-11260.

Nagasawa, S., K. Mizuguchi, C. Ichihara and J. Koyama (1982). "Limited chymotryptic cleavage of human C4-binding protein: isolation of a carbohydrate-containing core domain and an active fragment." J Biochem **92**(4): 1329-1332.

Nagasawa, S. and R. M. Stroud (1980). "Purification and characterization of a macromolecular weight cofactor for C3b-inactivator, C4bC3bINA-cofactor, of human plasma." Mol Immunol **17**(11): 1365-1372.

Neumann, H. P., M. Salzmann, B. Bohnert-Iwan, T. Mannuelian, C. Skerka, D. Lenk, B. U. Bender, M. Cybulla, P. Riegler, A. Konigsrainer, U. Neyer, A. Bock, U. Widmer, D. A. Male, G. Franke and P. F. Zipfel (2003). "Haemolytic uraemic syndrome and mutations of the factor H gene: a registry-based study of German speaking countries." Journal of Medical Genetics **40**(9): 676-681.

Noris, M. and G. Remuzzi (2005). "Hemolytic uremic syndrome." Journal of the American Society of Nephrology **16**(4): 1035-1050.

Nukatsuka, M. and S. Nagasawa (1987). "Characterization of the interaction between human protein S and C4b-binding protein (C4bp)." J Biochem **102**(3): 599-605.

Ogun, S. A., L. Dumon-Seignovert, J. B. Marchand, A. A. Holder and F. Hill (2008). "The oligomerization domain of C4-binding protein (C4bp) acts as an adjuvant,

and the fusion protein comprised of the 19-kilodalton merozoite surface protein 1 fused with the murine C4bp domain protects mice against malaria." Infection and Immunity **76**(8): 3817-3823.

Okemefuna, A. I., R. Nan, J. Gor and S. J. Perkins (2009). "Electrostatic interactions contribute to the folded-back conformation of wild type human factor H." J Mol Biol **391**(1): 98-118.

Oshiumi, H., K. Shida, R. Goitsuka, Y. Kimura, J. Katoh, S. Ohba, Y. Tamaki, T. Hattori, N. Yamada, N. Inoue, M. Matsumoto, S. Mizuno and T. Seya (2005). "Regulator of complement activation (RCA) locus in chicken: identification of chicken RCA gene cluster and functional RCA proteins." Journal of Immunology **175**(3): 1724-1734.

Pangburn, M. K. and H. J. Muller-Eberhard (1983). "Initiation of the Alternative Complement Pathway Due to Spontaneous Hydrolysis of the Thioester of C-3." Ann N Y Acad Sci **421**(Dec): 291-298.

Pangburn, M. K. and N. Rawal (2002). "Structure and function of complement C5 convertase enzymes." Biochemical Society Transactions **30**: 1006-1010.

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

Park, C. T. and S. D. Wright (1996). "Plasma lipopolysaccharide-binding protein is found associated with a particle containing apolipoprotein A-I, phospholipid, and factor H-related proteins." J Biol Chem **271**(30): 18054-18060.

Peitsch, M. C. and J. Tschopp (1991). "Assembly of macromolecular pores by immune defense systems." Curr Opin Cell Biol **3**(4): 710-716.

Perkins, S. J., A. S. Nealis and R. B. Sim (1991). "Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering." Biochemistry **30**(11): 2847-2857.

Phillips, A. E., J. Toth, A. W. Dodds, U. V. Girija, C. M. Furze, E. Pala, R. B. Sim, K. B. M. Reid, W. J. Schwaeble, R. Schmid, A. H. Keeble and R. Wallis (2009). "Analogous Interactions in Initiating Complexes of the Classical and Lectin Pathways of Complement." Journal of Immunology **182**(12): 7708-7717.

Poltermann, S., A. Kunert, M. von der Heide, R. Eck, A. Hartmann and P. F. Zipfel (2007). "Gpm1p is a factor H-, FHL-1-, and plasminogen-binding surface protein of Candida albicans." Journal of Biological Chemistry **282**(52): 37537-37544.

Porter, R. R. and K. B. Reid (1979). "Activation of the complement system by antibody-antigen complexes: the classical pathway." Adv Protein Chem **33**: 1-71.

Prosser, B. E., S. Johnson, P. Roversi, S. J. Clark, E. Tarelli, R. B. Sim, A. J. Day and S. M. Lea (2007a). "Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulfate/human complement regulator factor HSCRs 6-8." Acta Crystallographica Section F-Structural Biology and Crystallization Communications **63**: 480-483.

Prosser, B. E., S. Johnson, P. Roversi, A. P. Herbert, B. S. Blaum, J. Tyrrell, T. A. Jowitt, S. J. Clark, E. Tarelli, D. Uhrin, P. N. Barlow, R. B. Sim, A. J. Day and S. M. Lea (2007b). "Structural basis for complement factor H-linked age-related macular degeneration." Journal of Experimental Medicine **204**(10): 2277-2283.

Rawal, N., R. Rajagopalan and V. P. Salvi (2009). "Stringent regulation of complement lectin pathway C3/C5 convertase by C4b-binding protein (C4BP)." Molecular Immunology **46**(15): 2902-2910.

Read, R. J. and A. J. Schierbeek (1988). "A Phased Translation Function." Journal of Applied Crystallography **21**: 490-495.

Reinke, A. W., R. A. Grant and A. E. Keating (2010). "A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering." Journal of the American Chemical Society **132**(17): 6025-6031.

Riedo, F. X., B. D. Plikaytis and C. V. Broome (1995). "Epidemiology and prevention of meningococcal disease." Pediatr Infect Dis J **14**(8): 643-657.

Ripoche, J., A. J. Day, T. J. R. Harris and R. B. Sim (1988). "The Complete Amino-Acid Sequence of Human-Complement Factor-H." Biochemical Journal **249**(2): 593-602.

Rodriguez de Cordoba, S., P. Sanchez-Corral and J. Rey-Campos (1991). "Structure of the gene coding for the alpha polypeptide chain of the human complement component C4b-binding protein." J Exp Med **173**(5): 1073-1082.

Rodriguez, P. and E. G. Kranias (2005). "Phospholamban: a key determinant of cardiac function and dysfunction." Arch Mal Coeur Vaiss **98**(12): 1239-1243.

Rooijakkers, S. H. M., J. Wu, M. Ruyken, R. van Domselaar, K. L. Planken, A. Tzekou, D. Ricklin, J. D. Lambris, B. J. C. Janssen, J. A. G. van Strijp and P. Gros (2009). "Structural and functional implications of the alternative complement pathway C3 convertase stabilized by a staphylococcal inhibitor." Nature Immunology **10**(7): 721-U772.

Roversi, P., S. Johnson, J. J. Caesar, F. McLean, K. J. Leath, S. A. Tsiftsoglou, B. P. Morgan, C. L. Harris, R. B. Sim and S. M. Lea (2011a). "Structural basis for complement factor I control and its disease-associated sequence polymorphisms." Proc Natl Acad Sci U S A **108**(31): 12839-12844.

Roversi, P., S. Johnson, J. J. Caesar, F. McLean, K. J. Leath, S. A. Tsiftsoglou, B. P. Morgan, C. L. Harris, R. B. Sim and S. M. Lea (2011b). "Structures of the rat complement regulator CrrY." Acta Crystallogr Sect F Struct Biol Cryst Commun **67**(Pt 7): 739-743.

Saeki, T., S. Hirose, M. Nukatsuka, Y. Kusunoki and S. Nagasawa (1989). "Evidence that C4b-binding protein is an acute phase protein." Biochem Biophys Res Commun **164**(3): 1446-1451.

Sanchez-Corral, P., O. Criado Garcia and S. Rodriguez de Cordoba (1995). "Isoforms of human C4b-binding protein. I. Molecular basis for the C4BP isoform pattern and its variations in human plasma." J Immunol **155**(8): 4030-4036.

Scharfstein, J., A. Ferreira, I. Gigli and V. Nussenzweig (1978). "Human C4-binding protein. I. Isolation and characterization." J Exp Med **148**(1): 207-222.

Schmidt, C. Q., A. P. Herbert, H. G. Hocking, D. Uhrin and P. N. Barlow (2008a). "Translational mini-review series on complement factor H: structural and functional correlations for factor H." Clin Exp Immunol **151**(1): 14-24.

Schmidt, C. Q., A. P. Herbert, D. Kavanagh, C. Gandy, C. J. Fenton, B. S. Blaum, M. Lyon, D. Uhrin and P. N. Barlow (2008b). "A new map of glycosaminoglycan and C3b binding sites on factor H." Journal of Immunology **181**(4): 2610-2619.

Schneider, M. C., R. M. Exley, H. Chan, I. Feavers, Y. H. Kang, R. B. Sim and C. M. Tang (2006). "Functional significance of factor H binding to Neisseria meningitidis." Journal of Immunology **176**(12): 7566-7575.

Schneider, M. C., B. E. Prosser, J. J. E. Caesar, E. Kugelberg, S. Li, Q. Zhang, S. Quoraishi, J. E. Lovett, J. E. Deane, R. B. Sim, P. Roversi, S. Johnson, C. M. Tang and S. M. Lea (2009). "Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates." Nature **458**(7240): 890-U899.

Schwaeble, W., J. Zwirner, T. F. Schulz, R. P. Linke, M. P. Dierich and E. H. Weiss (1987). "Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver." European Journal of Immunology **17**(10): 1485-1489.

Seya, T. and J. P. Atkinson (1989). "Functional-Properties of Membrane Cofactor Protein of Complement." Biochemical Journal **264**(2): 581-588.

Sharma, A. K. and M. K. Pangburn (1996). "Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis." Proc Natl Acad Sci U S A **93**(20): 10996-11001.

Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallographica Section A **64**: 112-122.

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

Siegel, C., T. Hallstrom, C. Skerka, H. Eberhardt, B. Uzonyi, T. Beckhaus, M. Karas, R. Wallich, B. Stevenson, P. F. Zipfel and P. Kraiczy (2010). "Complement factor H-related proteins CFHR2 and CFHR5 represent novel ligands for the infection-associated CRASP proteins of Borrelia burgdorferi." PLoS One **5**(10): e13519.

Sim, R. B., K. Kolble, M. A. McAleer, O. Dominguez and V. M. Dee (1993). "Genetics and deficiencies of the soluble regulatory proteins of the complement system." Int Rev Immunol **10**(1): 65-86.

Sinilnikova, O. M., J. D. McKay, S. V. Tavtigian, F. Canzian, D. DeSilva, C. Biessy, S. Monnier, L. Dossus, C. Boillot, L. Gioia, D. J. Hughes, M. K. Jensen, K. Overvad, A. Tjonneland, A. Olsen, F. Clavel-Chapelon, V. Chajes, V. Joulin, J. Linseisen, J. Chang-Claude, H. Boeing, S. Dahm, A. Trichopoulou, D.

Trichopoulos, M. Koliva, K. T. Khaw, S. Bingham, N. E. Allen, T. Key, D. Palli, S. Panico, F. Berrino, R. Tumino, P. Vineis, H. B. Bueno-de-Mesquita, P. H. Peeters, C. H. van Gils, E. Lund, G. Pera, J. R. Quiros, M. Dorronsoro, C. Martinez Garcia, M. J. Tormo, E. Ardanaz, G. Hallmans, P. Lenner, G. Berglund, J. Manjer, E. Riboli, G. M. Lenoir and R. Kaaks (2007). "Haplotype-based analysis of common variation in the acetyl-coA carboxylase alpha gene and breast cancer risk: a case-control study nested within the European Prospective Investigation into Cancer and Nutrition." Cancer Epidemiol Biomarkers Prev **16**(3): 409-415.

Sjoberg, A. P., L. A. Trouw and A. M. Blom (2009). "Complement activation and inhibition: a delicate balance." Trends Immunol **30**(2): 83-90.

Skerka, C., J. Hellwage, W. Weber, A. Tilkorn, F. Buck, T. Marti, E. Kampen, U. Beisiegel and P. F. Zipfel (1997). "The human factor H-related protein 4 (FHR-4). A novel short consensus repeat-containing protein is associated with human triglyceride-rich lipoproteins." J Biol Chem **272**(9): 5627-5634.

Skerka, C., N. Lauer, A. A. Weinberger, C. N. Keilhauer, J. Suhnel, R. Smith, U. Schlotzer-Schrehardt, L. Fritsche, S. Heinen, A. Hartmann, B. H. Weber and P. F. Zipfel (2007). "Defective complement control of factor H (Y402H) and FHL-1 in age-related macular degeneration." Mol Immunol **44**(13): 3398-3406.

Smart, O. S., M. Brandl, C. Flensburg, P. Keller, W. Paciorek, C. Vonrhein, T. O. Womack and G. Bricogne (2008). "Refinement with Local Structure Similarity Restraints (LSSR) Enables Exploitation of Information from Related Structures and Facilitates use of NCS." Abstr. Annu. Meet. Am. Crystallogr. Assoc., Abstract TP139, p. 117.

Sontheimer, R. D., E. Racila and D. M. Racila (2005). "C1q: its functions within the innate and adaptive immune responses and its role in lupus autoimmunity." J Invest Dermatol **125**(1): 14-23.

Sprengart, M. L., E. Fuchs and A. G. Porter (1996). "The downstream box: an efficient and independent translation initiation signal in Escherichia coli." EMBO J **15**(3): 665-674.

Stanek, G. and F. Strle (2003). "Lyme borreliosis." Lancet **362**(9396): 1639-1647.

Steere, A. C. (1989). "Lyme-Disease." New England Journal of Medicine **321**(9): 586-596.

Stein, N. (2008). "CHAINSAW: a program for mutating pdb files used as templates in molecular replacement." Journal of Applied Crystallography **41**: 641-643.

Stephens, D. S., B. Greenwood and P. Brandtzaeg (2007). "Epidemic meningitis, meningococcaemia, and Neisseria meningitidis." Lancet **369**(9580): 2196-2210.

Stevenson, B., N. El-Hage, M. A. Hines, J. C. Miller and K. Babb (2002). "Differential binding of host complement inhibitor factor h by Borrelia burgdorferi erp surface proteins: a possible mechanism underlying the expansive host range of Lyme disease Spirochetes." Infection and Immunity **70**(2): 491-497.

Strausberg, R. L., E. A. Feingold, L. H. Grouse, J. G. Derge, R. D. Klausner, F. S. Collins, L. Wagner, C. M. Shenmen, G. D. Schuler, S. F. Altschul, B. Zeeberg, K. H. Buetow, C. F. Schaefer, N. K. Bhat, R. F. Hopkins, H. Jordan, T. Moore, S. I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A. A. Farmer, G. M. Rubin, L. Hong, M. Stapleton, M. B. Soares, M. F. Bonaldo, T. L. Casavant, T. E. Scheetz, M. J. Brownstein, T. B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S. S. Raha, N. A. Loquellano, G. J. Peters, R. D. Abramson, S. J. Mullahy, S. A. Bosak, P. J. McEwan, K. J. McKernan, J. A. Malek, P. H. Gunaratne, S.

Richards, K. C. Worley, S. Hale, A. M. Garcia, L. J. Gay, S. W. Hulyk, D. K. Villalon, D. M. Muzny, E. J. Sodergren, X. Lu, R. A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. C. Young, Y. Shevchenko, G. G. Bouffard, R. W. Blakesley, J. W. Touchman, E. D. Green, M. C. Dickson, A. C. Rodriguez, J. Grimwood, J. Schmutz, R. M. Myers, Y. S. Butterfield, M. I. Krzywinski, U. Skalska, D. E. Smailus, A. Schnerch, J. E. Schein, S. J. Jones and M. A. Marra (2002). "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." Proc Natl Acad Sci U S A **99**(26): 16899-16903.

Strelkov, S. V. and P. Burkhard (2002). "Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation." J Struct Biol **137**(1-2): 54-64.

Strobel, S., P. F. Hoyer, C. J. Mache, E. Sulyok, W. S. Liu, H. Richter, M. Oppermann, P. F. Zipfel and M. Jozsi (2010). "Functional analyses indicate a pathogenic role of factor H autoantibodies in atypical haemolytic uraemic syndrome." Nephrol Dial Transplant **25**(1): 136-144.

Sugiki, T., I. Shimada and H. Takahashi (2008). "Stable isotope labeling of protein by Kluyveromyces lactis for NMR study." J Biomol NMR **42**(3): 159-162.

Terwilliger, T. C., F. Dimaio, R. J. Read, D. Baker, G. Bunkoczi, P. D. Adams, R. W. Grosse-Kunstleve, P. V. Afonine and N. Echols (2012). "phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta." J Struct Funct Genomics **13**(2): 81-90.

Terwilliger, T. C., R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read and P. D. Adams (2008). "Iterative model

building, structure refinement and density modification with the PHENIX AutoBuild wizard." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 61-69.

Thiel, S. (2007). "Complement activating soluble pattern recognition molecules with collagen-like regions, mannan-binding lectin, ficolins and associated proteins." Molecular Immunology **44**(16): 3875-3888.

Thiel, S., T. Vorup-Jensen, C. M. Stover, W. Schwaeble, S. B. Laursen, K. Poulsen, A. C. Willis, P. Eggleton, S. Hansen, U. Holmskov, K. B. Reid and J. C. Jensenius (1997). "A second serine protease associated with mannan-binding lectin that activates complement." Nature **386**(6624): 506-510.

Thompson, D., M. B. Pepys and S. P. Wood (1999). "The physiological structure of human C-reactive protein and its complex with phosphocholine." Structure **7**(2): 169-177.

Tofte, R. W., P. K. Peterson, Y. Kim and P. G. Quie (1980). "Influence of serum concentration on opsonization by the classical and alternative complement pathways." Infect Immun **27**(2): 693-696.

Vagin, A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." Journal of Applied Crystallography **30**: 1022-1025.

Vallota, E. H. and H. J. Muller-Eberhard (1973). "Formation of C3a and C5a anaphylatoxins in whole human serum after inhibition of the anaphylatoxin inactivator." J Exp Med **137**(5): 1109-1123.

van de Poel, R. H., J. C. Meijers and B. N. Bouma (1999). "Interaction between protein S and complement C4b-binding protein (C4BP). Affinity studies using chimeras containing c4bp beta-chain short consensus repeats." J Biol Chem **274**(21): 15144-15150.

Verardi, R., L. Shi, N. J. Traaseth, N. Walsh and G. Veglia (2011). "Structural topology of phospholamban pentamer in lipid bilayers by a hybrid solution and solid-state NMR method." Proc Natl Acad Sci U S A **108**(22): 9101-9106.

Vik, D. P., P. Munozcanoves, H. Kozono, L. G. Martin, B. F. Tack and D. D. Chaplin (1990). "Identification and Sequence-Analysis of 4 Complement Factor H-Related Transcripts in Mouse-Liver." Journal of Biological Chemistry **265**(6): 3193-3201.

Vonrhein, C., E. Blanc, P. Roversi and G. Bricogne (2007). "Automated structure solution with autoSHARP." Methods Mol Biol **364**: 215-230.

Vonrhein, C., C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack and G. Bricogne (2011). "Data processing and analysis with the autoPROC toolbox." Acta Crystallographica Section D-Biological Crystallography **67**: 293-302.

Wallich, R., J. Pattathu, V. Kitiratschky, C. Brenner, P. F. Zipfel, V. Brade, M. M. Simon and P. Kraiczy (2005). "Identification and functional characterization of complement regulator-acquiring surface protein 1 of the Lyme disease spirochetes Borrelia afzelii and Borrelia garinii." Infection and Immunity **73**(4): 2351-2359.

Wallis, J. W., J. Aerts, M. A. Groenen, R. P. Crooijmans, D. Layman, T. A. Graves, D. E. Scheer, C. Kremitzki, M. J. Fedele, N. K. Mudd, M. Cardenas, J. Higginbotham, J. Carter, R. McGrane, T. Gaige, K. Mead, J. Walker, D. Albracht, J. Davito, S. P. Yang, S. Leong, A. Chinwalla, M. Sekhon, K. Wylie, J. Dodgson, M. N. Romanov, H. Cheng, P. J. de Jong, K. Osoegawa, M. Nefedov, H. Zhang, J. D. McPherson, M. Krzywinski, J. Schein, L. Hillier, E. R.

Mardis, R. K. Wilson and W. C. Warren (2004). "A physical map of the chicken genome." Nature **432**(7018): 761-764.

Walport, M. J. (2001). "Complement. First of two parts." N Engl J Med **344**(14): 1058-1066.

Walshaw, J. and D. N. Woolfson (2001). "SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures." J Mol Biol **307**(5): 1427-1450.

Ware, C. F. and W. P. Kolb (1981). "Assembly of the functional membrane attack complex of human complement: formation of disulfide-linked C9 dimers." Proc Natl Acad Sci U S A **78**(10): 6426-6430.

Waters, A. M. and C. Licht (2011). "aHUS caused by complement dysregulation: new therapies on the horizon." Pediatr Nephrol **26**(1): 41-57.

Weiler, J. M., M. R. Daha, K. F. Austen and D. T. Fearon (1976). "Control of Amplification Convertase of Complement by Plasma-Protein Beta-1h." Proceedings of the National Academy of Sciences of the United States of America **73**(9): 3268-3272.

Whaley, K. and S. Ruddy (1976). "Modulation of Alternative Complement Pathway by Beta-1h Globulin." Journal of Experimental Medicine **144**(5): 1147-1163.

White, J., P. Lukacik, D. Esser, M. Steward, N. Giddings, J. R. Bright, S. J. Fritchley, B. P. Morgan, S. M. Lea, G. P. Smith and R. A. G. Smith (2004). "Biological activity, membrane-targeting modification, and crystallization of soluble human decay accelerating factor expressed in E-coli." Protein Science **13**(9): 2406-2415.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N.

Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallographica Section D-Biological Crystallography **67**: 235-242.

Winter, G. (2010). "xia2: an expert system for macromolecular crystallography data reduction." Journal of Applied Crystallography **43**: 186-190.

Woodman, M. E., A. E. Cooley, J. C. Miller, J. J. Lazarus, K. Tucker, T. Bykowski, M. Botto, J. Hellwage, R. M. Wooten and B. Stevenson (2007). "Borrelia burgdorferi binding of host complement regulator factor H is not required for efficient mammalian infection." Infection and Immunity **75**(6): 3131-3139.

Wormser, G. P. (2006). "Hematogenous dissemination in early Lyme disease." Wien Klin Wochenschr **118**(21-22): 634-637.

Wu, J., Y. Q. Wu, D. Ricklin, B. J. Janssen, J. D. Lambris and P. Gros (2009). "Structure of complement fragment C3b-factor H and implications for host protection by complement regulators." Nature Immunology **10**(7): 728-733.

Xu, Q., B. A. Traag, J. Willemse, D. McMullan, M. D. Miller, M. A. Elsliger, P. Abdubek, T. Astakhova, H. L. Axelrod, C. Bakolitsa, D. Carlton, C. Chen, H. J. Chiu, M. Chruszcz, T. Clayton, D. Das, M. C. Deller, L. Duan, K. Ellrott, D. Ernst, C. L. Farr, J. Feuerhelm, J. C. Grant, A. Grzechnik, S. K. Grzechnik, G. W. Han, L. Jaroszewski, K. K. Jin, H. E. Klock, M. W. Knuth, P. Kozbial, S. S. Krishna, A. Kumar, D. Marciano, W. Minor, A. M. Mommaas, A. T. Morse, E. Nigoghossian, A. Nopakun, L. Okach, S. Oommachen, J. Paulsen, C. Puckett, R. Reyes, C. L. Rife, N. Sefcovic, H. J. Tien, C. B. Trame, H. van den Bedem, S. Wang, D. Weekes, K. O. Hodgson, J. Wooley, A. M. Deacon, A. Godzik, S. A. Lesley, I. A. Wilson and G. P. van Wezel (2009). "Structural and functional

characterizations of SsgB, a conserved activator of developmental cell division in morphologically complex actinomycetes." Journal of Biological Chemistry **284**(37): 25268-25279.

Zhao, J., H. Wu, M. Khosravi, H. Cui, X. Qian, J. A. Kelly, K. M. Kaufman, C. D. Langefeld, A. H. Williams, M. E. Comeau, J. T. Ziegler, M. C. Marion, A. Adler, S. B. Glenn, M. E. Alarcon-Riquelme, B. A. Pons-Estel, J. B. Harley, S. C. Bae, S. Y. Bang, S. K. Cho, C. O. Jacob, T. J. Vyse, T. B. Niewold, P. M. Gaffney, K. L. Moser, R. P. Kimberly, J. C. Edberg, E. E. Brown, G. S. Alarcon, M. A. Petri, R. Ramsey-Goldman, L. M. Vila, J. D. Reveille, J. A. James, G. S. Gilkeson, D. L. Kamen, B. I. Freedman, J. M. Anaya, J. T. Merrill, L. A. Criswell, R. H. Scofield, A. M. Stevens, J. M. Guthridge, D. M. Chang, Y. W. Song, J. A. Park, E. Y. Lee, S. A. Boackle, J. M. Grossman, B. H. Hahn, T. H. Goodship, R. M. Cantor, C. Y. Yu, N. Shen and B. P. Tsao (2011). "Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility." PLoS Genet **7**(5): e1002079.

Zipfel, P. F. and C. Skerka (2009). "Complement regulators and inhibitory proteins." Nat Rev Immunol **9**(10): 729-740.

Zipfel, P. F., R. Wurzner and C. Skerka (2007). "Complement evasion of pathogens: Common strategies are shared by diverse organisms." Molecular Immunology **44**(16): 3850-3857.

Appendix One: X-ray crystallography theory

A1.1 Bragg's law

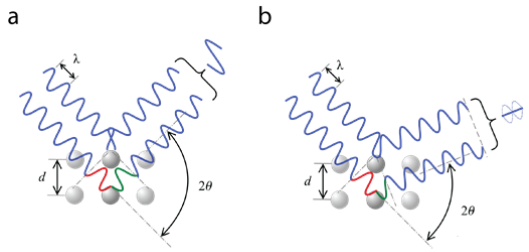


Figure A1.1 *Panel a*: Diagram describing Braggs law. The scattered rays are coherent resulting in an observable diffraction maxima. *Panel b*: scattered rays are non-coherent resulting in destructive interference and a diffraction minima. Reproduced under the creative commons attribution-share alike 3.0 unported license.

When an electromagnetic wave is incident on an atom, energy is absorbed by the electron density cloud and re-emitted in all directions as waves with the same wavelength in a process known as Raleigh scattering. Scattered waves from each atom in a lattice interfere. At most angles this is destructive. However, at certain angles the waves interfere constructively to produce the phenomenon known as diffraction. At these angles, the path length difference is equal to the wavelength; producing waves that are in phase (Figure A1.1). The equation that describes the angles at which diffraction occurs is known as Bragg's law (Equation A1.1).

$$n\lambda = 2d \sin\theta$$

Equation A1.1 Bragg's law. N is an integer, λ the wavelength, d the separation between diffracting planes and θ the angle between the incident ray and the diffracting plane.

In order to observe diffraction, the wavelength of the electromagnetic radiation must be similar to the interatomic separation which has magnitude of approximately 10^{-10} m. These separation distances correspond to an optimal wavelength in the X-ray band of the electromagnetic spectrum.

A1.2 Lattice planes

A protein crystal may be thought of as an infinite three dimensional lattice of identical points, each residing in an identical environment with respect to its neighbours. Associated with each lattice point is a unit cell which contains the protein molecule(s). The overall diffraction from the crystal is a convolution of the diffraction from the lattice with the diffraction from the protein, resulting in a three dimensional diffraction pattern representing the reciprocal lattice points with each maxima describing the protein diffraction pattern at that point. Diffraction from the lattice occurs when the three dimensional Bragg condition (Equation A1.2) is fulfilled by sets of planes running through the crystal, termed lattice planes, described by Miller indices (Figure A1.2). Thus each diffracted ray can be described using its corresponding set of indices, h, k and l.

$$\lambda = 2d_{hkl} \sin\theta$$

Equation A1.2 Bragg's law in three dimensions describing diffraction from a set of planes described by Miller indices h, k and l. n from Equation 2.3 has been dropped as these indices can be chosen to represent any value of n.

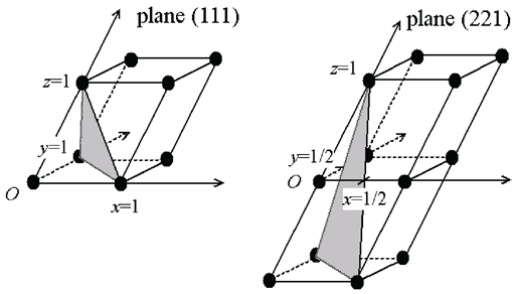


Figure A1.2 Description of Miller notation of lattice planes. Reproduced under the creative commons attribution-share alike 3.0 unported license

A1.3 Structure factor equation

$$F_{hkl} = |F_{hkl}| \exp[i\phi_{hkl}]$$

Equation A1.3 The structure factor of a reflection described in terms of its magnitude and phase.

Each diffracted maxima is described by its structure factor, a complex number that describes the amplitude and phase of the scattered wave (Equation A1.3). The structure factor can also be described by the scattering object, in this case the protein electron density (Equation A1.4). This equation may be simplified by considering that diffraction from every atom in the unit cell has a contribution in each reflection. By summing the individual components, the structure factor can be described

using Equation A1.5, which introduces both the atomic scattering factor, *f*, a measure of an atoms ability to scatter X-rays, and the thermal factor, B, a measure of atomic vibrations and disorder within the crystal. The last term, encompassing both of these measures, is known as the Debye-Waller factor and explains why the intensity of reflections decreases with increased scattering angle, and why they are not infinitely sharp.

$$F_{hkl} = V \iiint_{0}^{1} \rho(x,y,z) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz$$

Equation A1.4 The structure factor described in terms of the electron density at a point x, y, z in the unit cell.

$$F_{hkl} = \sum_{N} f_i \exp[2\pi i(hx_i + ky_i + lz_i)] \exp\left[-B_i \sin^2\theta/\lambda^2\right]$$

Equation A1.5 The structure factor expressed as a sum of the individual components from each atom.

As described in Appendix 1.2, each maximum describes a different part of the diffraction pattern of the protein, which in turn is the Fourier transform of the diffracting structure. Applying a reverse Fourier transform to all of the structure factors from the diffracted rays (Equation A1.6) allows calculation of the electron density at each point in the unit cell, into which an atomic model of the protein may be built. This introduces an important factor concerning resolution; the more structure factors that are known gives a more accurate, hence higher resolution, calculation of the electron density. Thus the resolution is determined by total the number of reflections observed which increases with scattering angle, but is limited by the Debye-Waller component of Equation A1.5.

$$\sigma(x,y,z) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp[-2\pi i(hx + ky + lz)]$$

Equation A1.6 The inverse Fourier transform describing how the electron density at a point x, y, z can be described by the sum of individual components from each structure factor.

A1.4 The phase problem

In order to solve the structure of the protein, both the amplitude and phase of each diffracted wave are required to calculate the structure factors (Equation A1.7). Amplitudes can be experimentally measured as they are related to the intensity of the scattered waves by Equation A1.7. However, experimental measurement of the phases of the diffracted rays is not possible. This is known as the phase problem and presents the largest hurdle in solving a crystal structure. Different ways of solving it are discussed in Appendix 2.

$$I \propto |F_{hkl}|^2$$

Equation A1.7

A1.5 Symmetry

Diffraction by a crystal is self-amplifying as diffraction occurs throughout the crystal, resulting in strong maxima that can be accurately measured. This is aided by Friedel's law (the intensity of a reflection, I_{hkl}, must be equal to the intensity of I_{-h-k-l}) as the number of repeated measurements is increased. Further accuracy is often achieved due to symmetry within the unit cell. There are seven different crystal lattices that proteins may adopt. The centering within each unit cell then gives rise to fourteen Bravais lattices. Internal symmetry elements within each unit cell then determines the spacegroup, of which there are 230, which fully describes the symmetry of the entire crystal. Invoking the symmetry operators adds extra constraints meaning that reflections are measured multiple times resulting in more accurate intensity measurements and also reduces the overall number of reflections that need to be measured.

Symmetry also places constraints upon which reflections are allowed, which manifest themselves as systematic absences in the diffraction pattern. It is from these absences that the spacegroup can be determined.

Appendix Two: X-ray crystallography practice

A2.1 Crystallisation using the sitting drop vapour diffusion method

Nucleation of crystal growth occurs in the upper region of the supersaturation zone of the protein crystallisation phase diagram (Figure A2.1) [Asherie 2004]. In order for it to occur, the ratio of protein to precipitant must be brought into the correct regime. In the sitting drop vapour diffusion method [Benvenuti and Mangani 2007], protein and precipitant are mixed at concentrations that fall below the solubility curve. A drop is then placed in a sealed well with a reservoir of precipitant. As the concentration of water is higher in the drop versus the reservoir, due to the protein buffer, water molecules diffuse from the former to the latter. This has the desired effect of increasing the protein and precipitant concentration into the nucleation zone. As a crystal forms, the concentration of protein in solution decreases as the concentration of precipitant continues to increase, maintaining the system in the crystal growth zone.

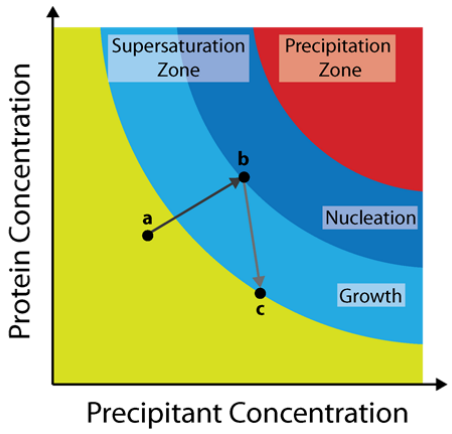


Figure A2.1 Protein crystallisation phase diagram

A2.2 Data collection

Within this thesis, all data were collected at a synchrotron radiation source using the rotation method. As previously discussed, monochromatic X-ray radiation with a wavelength of approximately 1 x 10^{-10} m was used to match the average interatomic distance and data were collected at 120 K in order to reduce the effects of radiation damage. A crystal is mounted on a goniometer and aligned so that the collimated X-ray

beam hits the sample. Diffraction spots are then measured on a charge coupled device (CCD) detector as the crystal is rotated in small incremented steps, recording a new image for each step.

Prior to collection of a dataset, the crystal is indexed from an initial set of diffraction images in order to find the unit cell parameters and the Bravais lattice. Once the lattice is known, it is possible to calculate which reflections are equivalent, due to symmetry or Friedel's law, resulting in a set of unique reflections that need to be measured in order to obtain a complete dataset. A data collection strategy can then be devised to collect as many of these unique reflections in the minimum rotation range possible; reducing the overall X-ray exposure and associated radiation damage. The number of measured unique reflections expressed as the percentage of possible unique reflections is known as the completeness (Equation A2.1; a measure of how whole the dataset is. Another measure, known as the multiplicity (Equation A2.2), describes the average number of times that the same reflection was observed resulting in a more accurate determination of the intensity.

$$\text{Completeness} = \frac{\text{Number unique reflections measured}}{\text{Total number unique reflections}}$$

Equation A2.1

$$\text{Multiplicity} = \frac{\text{Total number reflections measured}}{\text{Total number unique reflections measured}}$$

Equation A2.2

Another data collection factor that is calculated is the optimal oscillation range. Use of an oscillation range that is too wide can result in diffraction spots overlapping which hinders measurement of the intensity of each individual reflection.

A2.3 Data Processing

After collection of a dataset, all of the diffraction spots are indexed to determine the Miller indices of each reflection. The intensity of each reflection is then measured by integration of the diffraction image over the spot. Due to the size of the crystal, crystal mosaicity, non-perfect monochromatic x-ray beams and other experimental factors, spots are not infinitely sharp and are blurred into a distribution which is often fit with a Gaussian distribution during processing. This manifests itself by blurred spots that are usually spread over several diffraction images. Within this thesis, all diffraction spots have been integrated in three dimensions; meaning Gaussian distributions have been fit to the intensity distribution in both the plane of the diffraction image and across diffraction images in the orthogonal direction. This results in a more accurate integration and determination of the overall reflection intensity.

$$R_{merge} = \frac{\sum_{hkl}\sum_{i}|I_i - \langle I_{hkl}\rangle|}{\sum_{hkl}\sum_{i}\langle I_{hkl}\rangle}$$

Equation A2.3

$$R_{pim} = \frac{\sum_{hkl}\left[\frac{1}{N_{hkl}-1}\right]^{\frac{1}{2}}\sum_{i}|I_i - \langle I_{hkl}\rangle|}{\sum_{hkl}\sum_{i}\langle I_{hkl}\rangle}$$

Equation A2.4

In order to merge all of the repeated measurements into a single dataset containing average intensities for each reflection, scaling must first be applied. The purpose of scaling is to account for variance between reflections that should have identical intensity; often occurring due to fluctuations in the X-ray beam intensity during data collection, crystal orientation and detector settings. In order to assess the quality of a merged dataset, two statistics are often used, R_{merge} (Equation A2.3) and R_{pim} (Equation A2.4). R_{merge} is the mean difference between the measured intensities and the mean, expressed as a fraction of the measurements, reflecting the mean error in the measured

intensity. Unfortunately, R_{merge} increases with the number of measurements and does not fully reflect the increased accuracy achieved from more observations. R_{pim} addresses this issue with a weighting dependent on the number of observations, giving a better reflection of the accuracy of the mean intensity.

In order to determine the resolution of a dataset, the signal to noise ratio of the intensities is often used to define a cut off. Scrutiny of the mean I/σI in each resolution bin often suggests a resolution limit when I/σI = 2. However, the quality of the data at this resolution is also dependent upon acceptable $R_{merge,}$ R_{pim} and completeness values in the outer shell of reflections. Analysis of the intensities also allows determination of the spacegroup. Observation of specific systematic absences in the h, k and l Miller indices describes the lattice centering and the presence of screw axes.

$$|F_{hkl}| = \left[\frac{I}{LP}\right]^{\frac{1}{2}}$$

Equation A2.5

As described by Equation A1.7 the measured intensity of a reflection is proportional to the square of the magnitude of the structure factor. This relationship is given in full in Equation A2.5, introducing Lorenz (L) and polarization (P) factors which arise from experimental setup. The former is due to higher resolution spots passing through the detector quicker than low resolution, giving lower than expected intensities, where as the latter is a correction for effects of using a polarised X-ray beam. Applying these corrections enables calculation of the structure factor amplitudes.

A2.4 Solving the phase problem I: Phasing by molecular replacement

The Patterson function is generated by convolution of the protein diffraction pattern with its inverse and is described by Equation A2.6. Calculation of the Patterson function throughout the whole asymmetric unit results in a three dimensional map with peaks

describing the vectors between each atom. These Patterson maps can be calculated directly from experimentally measured intensities as the function requires only the amplitude of the structure factors and not the phases. Patterson maps also present a way of solving the phase problem using a model that is similar to the structure of the diffracting protein in a technique known as molecular replacement.

$$\mathrm{P(x,y,z)} = \sum_{hkl} |F_{hkl}|^2 exp[-2\pi i(hx + ky + lz]$$

Equation A2.6

Molecular replacement uses a Patterson map calculated from the model and attempts to fit it into the experimentally derived Patterson map in order to find the position of the protein and generate initial phases. Firstly a rotation function search is performed about the origin in order to fit the calculated map to the experimental map and give the rotational orientation of the model in the unit cell. This is performed within a sphere about the origin with a radius equal to that of the protein in order to discount any Patterson peaks that occur from vectors between multiple identical molecules in the unit cell. Once the orientation of the model is known, a translation function search is performed to find its absolute position. The model is placed at each position in the unit cell and the Patterson map is calculated. A solution is found when the calculated map best fits the experimental map. Once the correct model placement has been found, structure factors can be calculated from it and used to generate initial phases allowing elucidation of the electron density into which a refined model can be built.

Patterson maps can also yield important information about the arrangement of molecules within the asymmetric unit. Calculation of a self-rotation function by looking for rotations within the map at different orientations, described in spherical polar

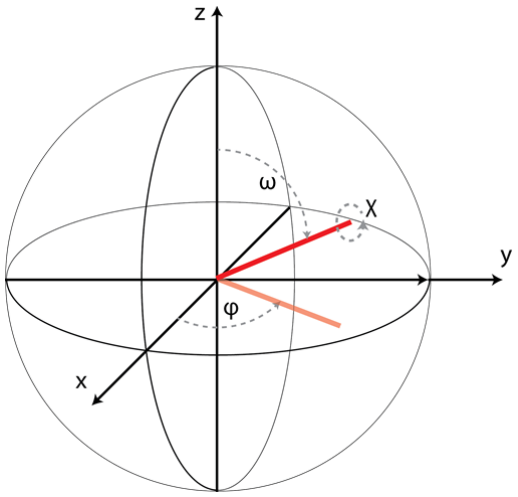


Figure A2.2 Diagram explaining the conventions in spherical polar coordinates

coordinates (Figure A2.2), suggests the presence of identical molecules related by a rotational symmetry. This symmetry is only within the asymmetric unit, not part of the crystal symmetry, and is termed non-crystallographic symmetry (NCS). The presence of a large, strong peak in the Patterson map can also be indicative of a translational relationship between molecules in the asymmetric unit. This can be a problem as it may lead to pseudo-translational symmetry which can cause crystal twinning.

A2.5 Solving the phase problem II: Phasing using anomalous dispersion

As described in Appendix 1.1, an incident X-ray is absorbed by an atom and re-emitted with the same wavelength and a constant phase shift of π radians. When the energy of the X-ray is close to the transition energy of moving an electron to an excited state, some of the energy is absorbed to stimulate the change, resulting not only in diffracted rays with a lower intensity, but also with an altered phase shift. This effect is known as anomalous dispersion and can be described by splitting the atomic scattering factor into two components (Equation A2.7); *f'* describes the normal coherent scattering and accounts for the altered intensity of the scattering at the given wavelength, where as the i*f''* term describes the change in the phase and is known as the anomalous scattering.

$$f = f' + if''$$

Equation A2.7

Considering Equation A2.7 allows the total diffraction of a crystal containing anomalous scatterers to be split into normal and anomalous components (Equation A2.8). Whilst $F_{hkl}(norm)$ and $F_{-h-k-l}(norm)$ are complex conjugates of each other giving rise to Friedel's law, $F_{hkl}(anom)$ is the complex conjugate of $-F_{-f-k-l}(anom)$. This means that $F_{hkl}(total)$ is not the complex conjugate of $F_{-h-k-l}(total)$ resulting in the breakdown of Friedel's law (Figure A2.3) as the intensity of the related reflections is not equal. These pairs of reflections differing in intensity are known as Bijvoet pairs.

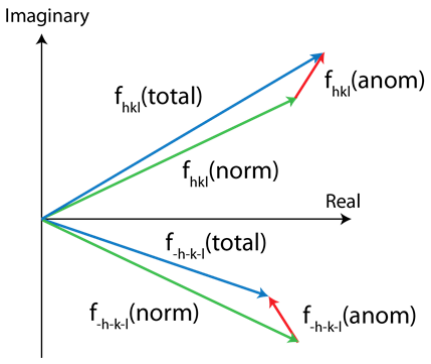


Figure A2.3 Argand diagram depicting the breakdown of Friedel's law in the presence of anomalous scattering.

$$f_{hkl}(total) = f_{hkl}(norm) + f_{hkl}(anom)$$

Equation A2.8

Most light atoms do not have any electronic transitions around the energy of the X-rays used for crystallography. Instead, heavy atoms that are electron dense and do have transitions in the right energy region are either co-crystallised with the protein, or soaked into crystals, where it is hoped that they bind in specific locations which can be determined from the vectors between them calculated from the Patterson map. By experimentally recording the differences in intensity of the Bijvoet pairs and the values of *f'* and *f''* it is possible to both calculate $F_{hkl}(anom)$ and the magnitudes of $F_{hkl}(total)$ and $F_{-h-k-l}(total)$, the latter of which is identical to the magnitude of its complex conjugate $F_{-h-k-l}(total)^*$. Equation A2.9 can now be used to generate initial phases for $F_{hkl}(total)$.

$$F_{hkl}(total) - F_{-h-k-l}(total)^* = 2F_{hkl}(anom)$$

Equation A2.9

Single-wavelength anomalous dispersion (SAD) is a technique that uses a dataset collected at a wavelength corresponding to the energy of the maximum *f''*, called the peak wavelength, to generate phases. Another technique known as multi-wavelength anomalous dispersion (MAD) uses data collected at the peak wavelength, but also at the *f'* minimum known as the inflection wavelength and at a remote wavelength where *f'* is normal. Combination of phases from all the wavelengths can result in a map that is less noisy than that calculated from the peak wavelength alone.

A2.6 Automated model building

Automated model building in BUCCANEER works using statistical main-chain tracing directed using a likelihood-based density target [Cowtan 2001]. Possible C_α sites are found by searching for electron density spheres with a 4Å radius in the map [Cowtan 1998]. These initial sites are scored by their mean electron density compared to that around each C_α in a reference structure (by default PDB accession 1TQW). Scores are then weighted by the conservation of density at any one point around the potential C_α compared to the reference structure. Chains are then built and sequenced from these initial sites according to the scheme described in Figure A2.4.

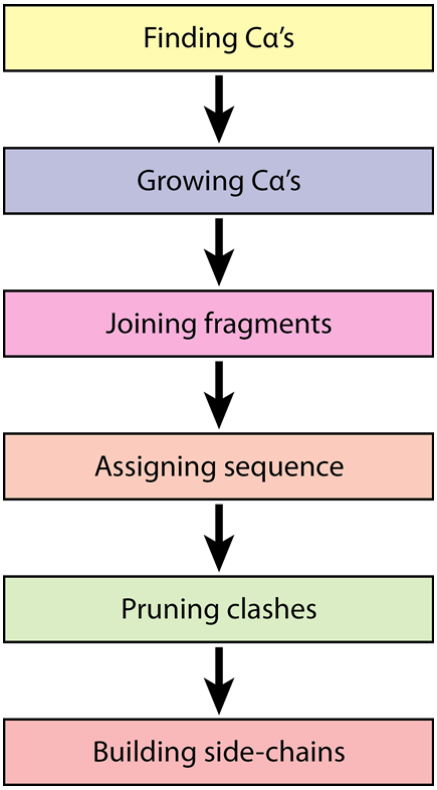


Figure A2.4 Flow-scheme of the procedures carried during autobuilding using BUCCANEER.

The initial model is then refined using REFMAC5 to generate a better model and phases which are used in subsequent rounds of BUCCANEER building and refinement. This approach allows building from experimental, or molecular replacement, phases, but also

enables extension and rebuilding of an incomplete or partial model. Other programs that can perform similar functions and run in a similar manner include ARP/wARP [Langer, Cohen *et al.* 2008] and PHENIX_AUTOBUILD [Terwilliger, Grosse-Kunstleve *et al.* 2008].

A2.7 Refinement

In order to obtain the best possible model of the protein, refinement of the individual atom positions and temperature factors is performed to obtain the best match between the magnitude of the structure factors calculated from the model and those observed. By refining in this manner the phases also improve, showing more detail in the resulting electron density map, making refinement an iterative procedure. Refinement is also aided by using restraints, for example on bond lengths, dihedral and torsion angles, penalising the model if they differ too much from mean values. The presence of non-crystallographic symmetry within the unit cell can also aid refinement by assuming that all copies of the same molecule must be identical.

$$R = \frac{\sum_{hkl}\left||F_{obs}| - |F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

Equation A2.10

The progress of refinement is measured by the R-value (Equation A2.10), comparing the observed amplitudes with those calculated from the model. R drops as the quality of the model increases. However, there is nothing to signify creation of a model that is biased towards a reduced R-value but isn't actually improved. In order to measure this, a set of reflections are not refined against. Calculation of an R-value, called R_{free}, against these amplitudes gives a reference against which to measure bias.

Appendix Three: Surface plasmon resonance

Surface Plasmon Resonance (SPR) is a technique that may be used to investigate intermolecular interactions and is particularly useful for measurement of protein-protein, and protein-ligand, interactions.

A3.1 Theory

A surface plasmon is a quantum mechanical phenomenon whereby a propagating electron density wave occurs at a metal surface (an example being gold) that travels parallel to the interface between metal and dielectric. Associated with the electron density wave is an electric field, known as the evanescent field, which is perpendicular to the surface and decays exponentially with distance. This field makes the frequency of the surface plasmon sensitive to changes in the dielectric constant near the surface and provides the mechanism by which interactions can be probed.

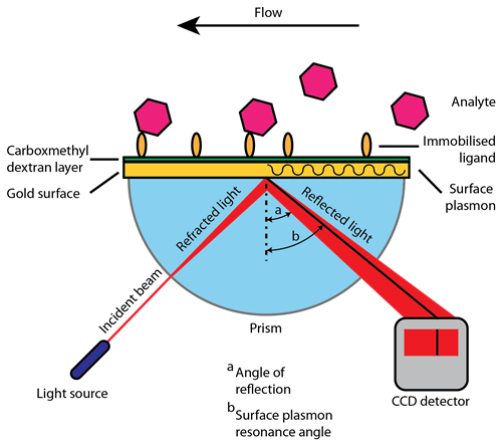


Figure A3.1 Diagram showing the simplified experimental setup for SPR. A change in the dielectric constant on the upper surface upon analyte binding to the immobilized protein species is observed by a change in the angle of incidence of the plane polarized light required to observe a minimum in the intensity of the reflected light.

In order to measure changes in the resonant frequency of the surface plasmon, a thin gold surface is mounted onto a prism which is then subjected to a beam of plane-polarised monochromatic light above the critical angle for total internal reflection. As the beam enters the prism it is refracted resulting in dispersion of incident angle. Each incident photon has a frequency component parallel to the prism-metal boundary that is dependent on its incident angle. When this frequency component matches that of the

surface plasmon the resonance condition is fulfilled and photon energy absorbed. This manifests itself as a reduction in intensity of totally internally reflected light (when measured on a CCD detector) at an angle dependent on the frequency of the surface plasmon. The dielectric constant on the prism side of the gold surface is constant, so any changes in the surface plasmon frequency are due to perturbations of the dielectric constant on the other side. Figure A3.1 details this experimental setup.

The upper surface of the gold surface is coated with carboxymethyl dextran to which proteins can be immobilised via primary amine coupling to activated carboxyl groups. An interaction between the coupled protein and another protein, or ligand, will alter the dielectric constant which will be observed as a change in the angle of minimum reflected intensity. This technique allows the study of such interactions in real time and a plot of the change in SPR angle versus time, known as a sensorgram, allows analysis of interaction rates from which affinity can be determined.

A3.2 Analysis of sensorgrams

It is usual to divide a surface plasmon resonance experiment into two phases: association and dissociation. During the association stage, protein molecules in solution are binding to the immobilised protein on the chip, resulting in a gradual increase in the number of bound species. In the dissociation stage, buffer is flowed over the chip to observe how the concentration of bound protein species decreases. These curves can be fitted using a Langmuir binding model which assumes a simple bimolecular interaction between the bound and analyte proteins, that all the bound protein interaction sites are the same, and that when one immobilised protein is bound to an analyte molecule it does not impair the binding of its neighbours. Assuming this model is correct, the association and dissociation kinetics are given by Equation A3.1 and A3.2. Where [A]

is the analyte protein concentration, [B] is the immobilised species concentration and [AB] is the concentration of the complex between the two proteins.

$$Association: \frac{d[AB]}{dt} = k_{on}[A][B] - k_{off}[AB]$$

Equation A3.1

$$Dissociation: \frac{-d[AB]}{dt} = k_{off}[AB]$$

Equation A3.2

In order to obtain the most accurate measurement of k_{on} and $k_{off,\ a}$ range of concentrations of the analyte are flowed over the surface which allow global fits to be made to the data. Affinities can also be determined from steady state data collected by flowing a dilution series of analyte over a chip, allowing each response to reach equilibrium. Fitting a curve to these responses plotted against analyte concentration allows determination of the maximum possible response, R_{max}. The K_A, and hence K_D, can then be estimated as the concentration at which the response is half of R_{max}.

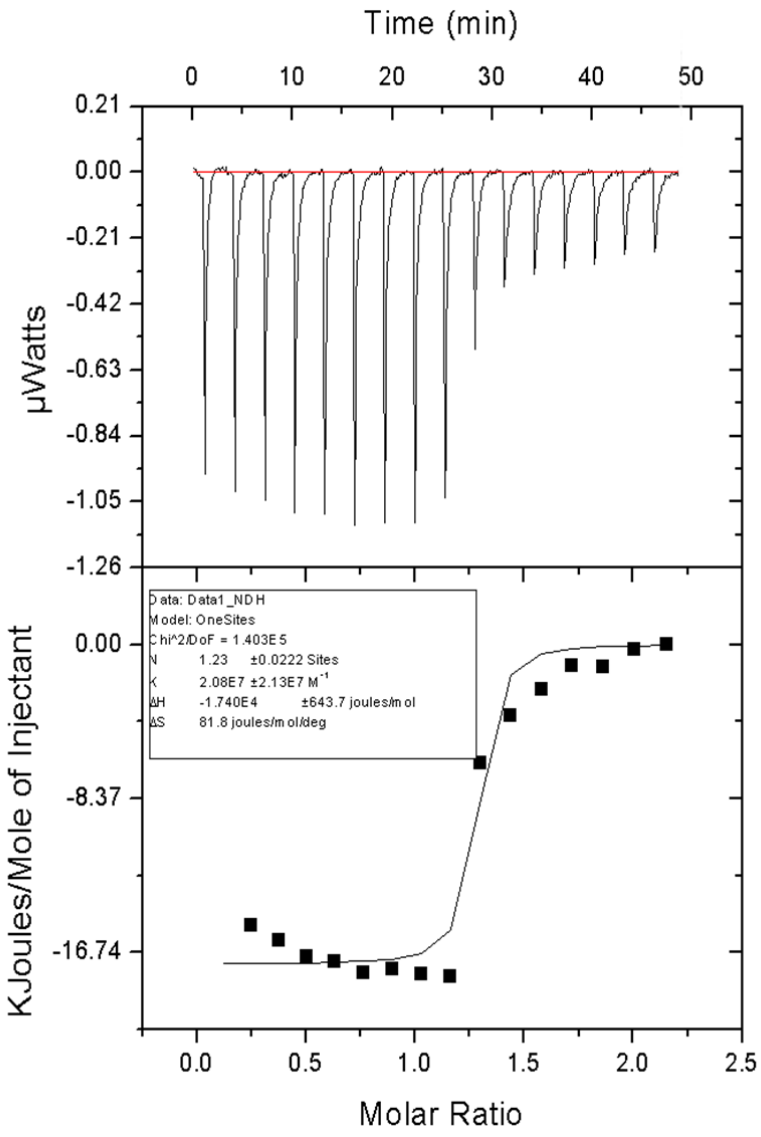
Appendix Four: Multiple angle light scattering

Multiple angle light scattering (MALS) is a technique that can be used to determine the average molecular weight of a protein. The intensity of monochromatic laser light scattered by a sample is measured at multiple angles, allowing calculation of the root mean square hydrodynamic radius. This radius, coupled with an accurately measured refractive index, can then be used to obtain the molecular weight in an orientation independent manner.

MALS can also be used with a constant flow-cell, allowing measurement of the mass distribution across the elution profile of an analytical size exclusion chromatography column. This is the setup used throughout this thesis. 100µl sample (1mg/ml) was injected onto an S200 16/60 column (GE Healthcare, USA) and the elution monitored using a Dawn Helios II (Wyatt Technology, USA) and an Optilab T-rEX (Wyatt Technology, USA) to measure the scattering and refractive index respectively. All data and were analysed using ASTRA (Wyatt Technology, USA). Within this thesis, all samples were kindly run and analysed by Dr Steven Johnson.

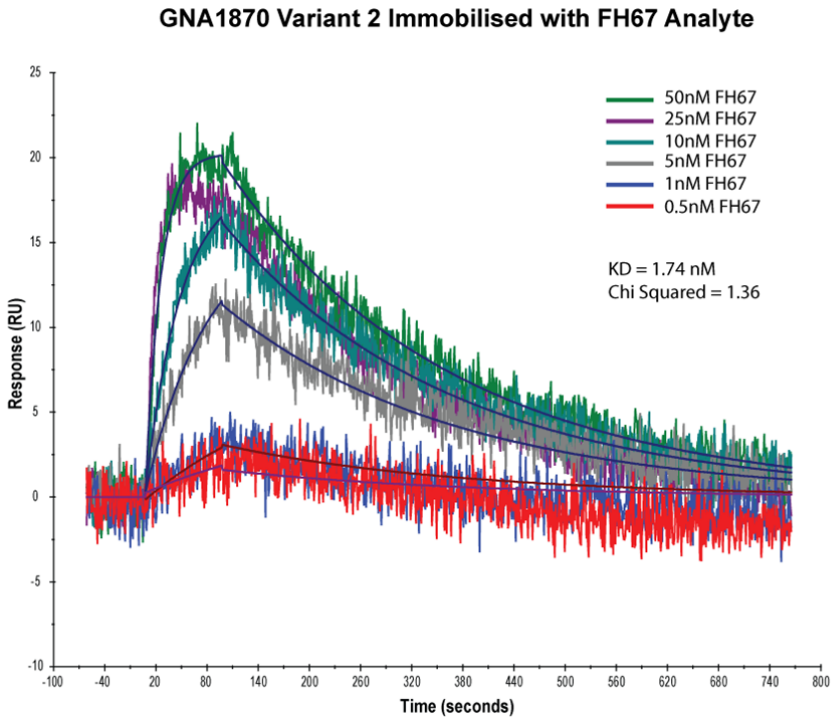
Appendix Five: fH binding by all three fHbp variants

A5.1 Affinity of fH_{67} for fHbp variant 1 measured using isothermal titration calorimetry



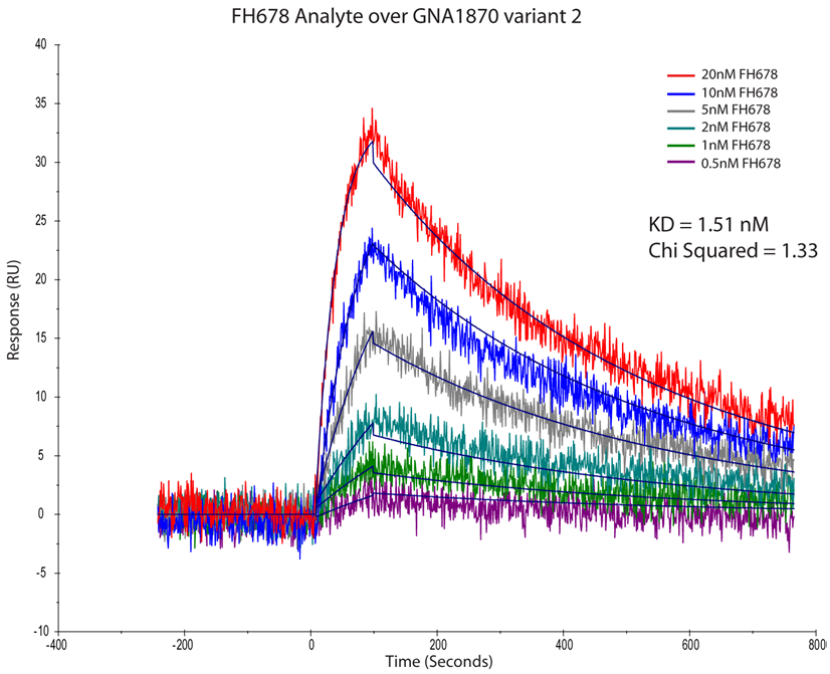
Isothermal titration calorimetry was attempted by titrating 300 µM fHbp Variant 1 into 30 µM fH_{67}. Due to the tightness of binding, a poor fit was made to the data. However, the interaction is observable in solution and the calculated K_d falls in the low nanomolar range as expected.

A5.2 fH_{67} binding by fHbp variant 2



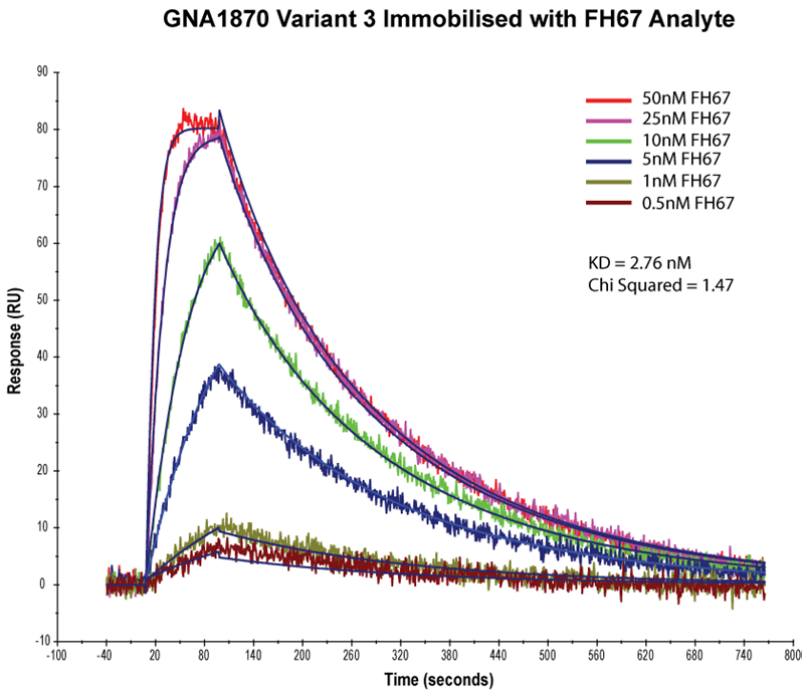
SPR curves and 1:1 Langmuir fits for fH_{67} binding to fHbp variant 2 (approximately 1000 RU immobilised). These data were recorded on a Proteon XPR36 machine.

A5.3 fH_{678} binding by fHbp variant 2



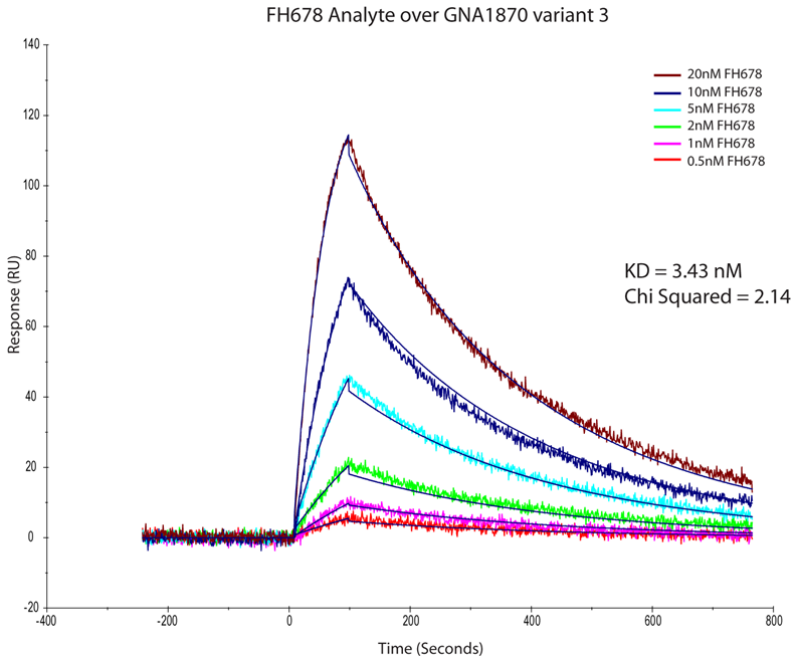
SPR curves and 1:1 Langmuir fits for fH_{687} binding to fHbp variant 2 (approximately 1000 RU immobilised). These data were recorded on a Proteon XPR36 machine.

A5.4 fH_{67} binding by fHbp variant 3



SPR curves and 1:1 Langmuir fits for fH_{67} binding to fHbp variant 3 (approximately 1000 RU immobilised). These data were recorded on a Proteon XPR36 machine.

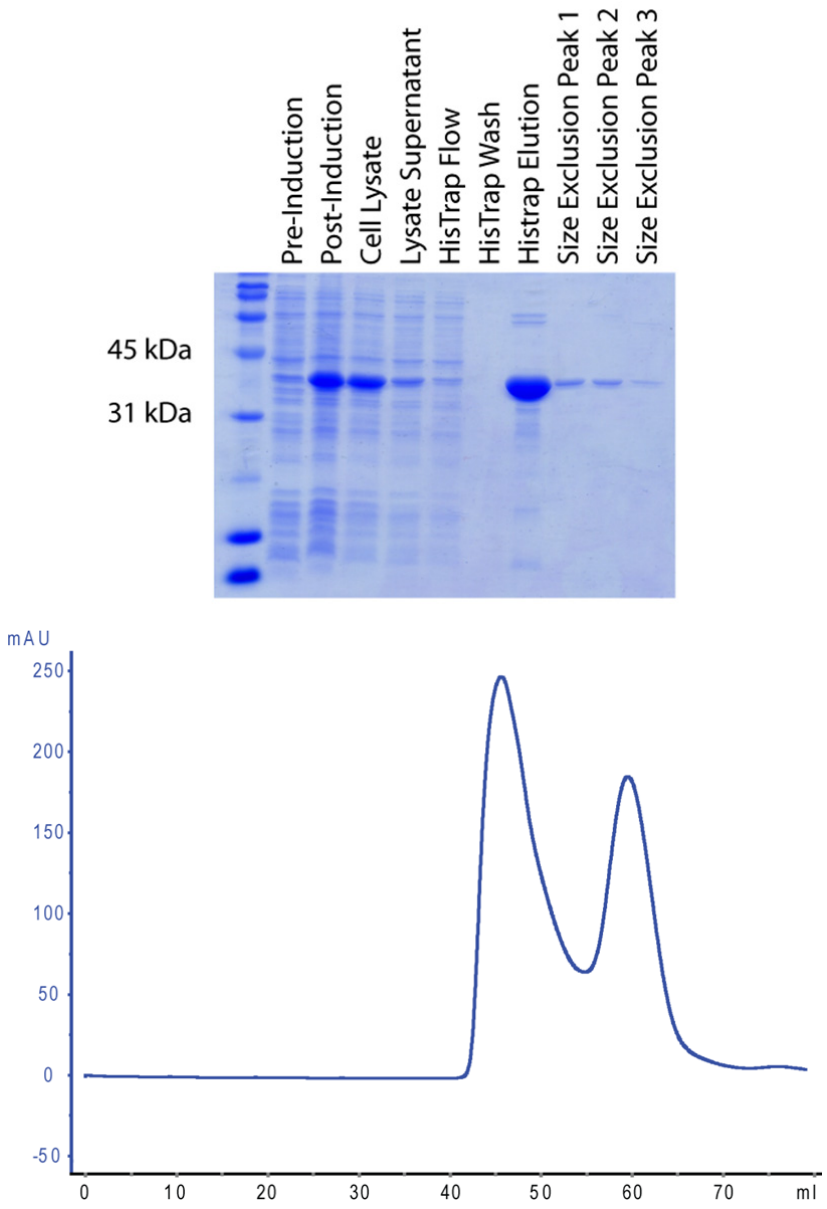
A5.5 fH_{678} binding by fHbp variant 3



SPR curves and 1:1 Langmuir fits for fH_{678} binding to fHbp variant 3 (approximately 1000 RU immobilised). These data were recorded on a Proteon XPR36 machine.

Appendix Six: Investigation into fH binding by NGO0033

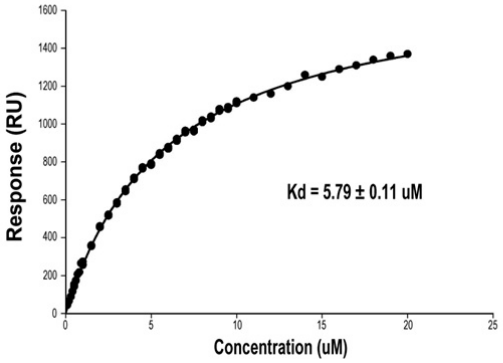
A6.1 Preparation of NGO0033

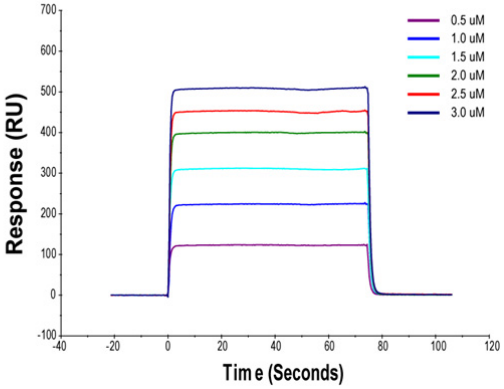


Top- 15% SDS-PAGE gel showing the expression and purification of NGO0033 using the same protocol as fHbp (Chapter 2.1)

Bottom- Example UV_{280} trace from NGO0033 run on an S75 16/60 size exclusion column. Samples from each peak were run on the gel above identifying peak 1 as the correct species.

A6.2 fH_{67} binding by NGO0033





	Fc1	Fc2	Fc3
Parameter Values			
Bmax	1765	1517	1492
Kd	5.9439E-06	5.7012E-06	5.7311E-06
Std. Errors			
Bmax	10.4699	8.6499	8.43
Kd	8.26E-06	7.75E-06	7.70E-06
95% Confidence Intervals			
Bmax	1744. to 1786.	1500. to 1534.	1476. to 1510.
Kd	5.779e-6 to 6.109e-6	5.546e-6 to 5.856e-6	5.577e-6 to 5.885e-6
Goodness of Fit			
Degrees of Freedom	58	58	58
R2	0.9991	0.9991	0.9992
Residual Sum of Square	8425	6183	5819
Sy.x	12.0524	10.3252	10.0172
Fit Status	Converged	Converged	Converged
Data			
Number of X Values	60	60	60
Number of Y Replicates	1	1	1
Total Number of Y Values	60	60	60
Number of Missing Values	0	0	0

Top left- 1:1 Langmuir fit to equilibrium responses for a range of FH67 concentrations flowed over an SPR chip with approximately 4500 RU undigested NGO0033 immobilised. The mean K_D from all three flowchannels is reported.

Top right- Example SPR curves from which equilibrium responses were measured.

Bottom- Fit statistics from all three flowchannels.

Appendix Seven: *Kluyveromyces lactis* media recipes

A7.1 YPGal

10 g Yeast extract

20 g Peptone

Made up to 950 ml prior to autoclaving. Once cooled at 50 ml of filter sterilised 40% (w/v) D-galactose.

A7.2 MMGal (10x stock)

100 g D-galactose

8.5 g Yeast nitrogen base (no amino acids or ammonium sulphate)

25 g Ammonium chloride

60 g Monosodium phosphate (anhydrous)

Dissolve in 500 ml of warm sterile H_2O and adjust pH to 7.0. Filter sterilise prior to use.

A7.3 YCB/agar with 5 mM acetamide

15 ml 1M Tris-HCl pH 7.0

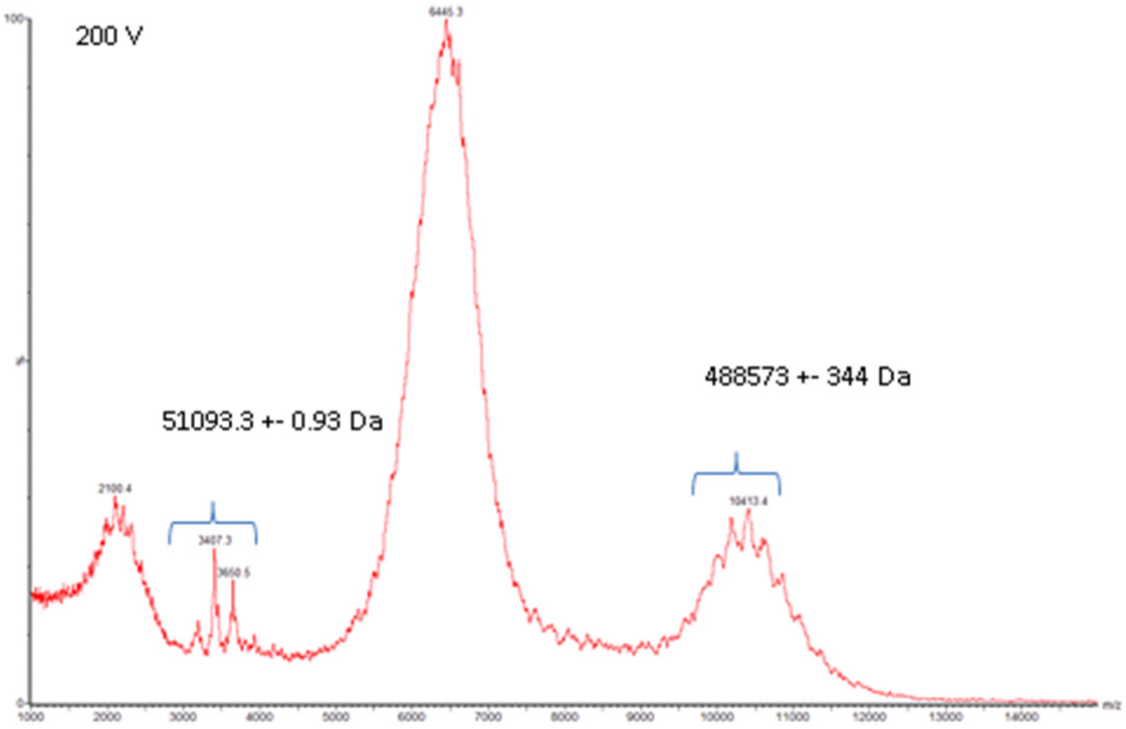
6.85 g Yeast carbon base

10g Agar

Dissolve in 495 ml sterile H_2O and autoclave. Once cooled to <60^{o}C add 5 ml of filter sterilised 500 mM acetamide.

Appendix Eight: Native mass spectrometry spectra

A8.1 Masses of the dissociated C4bp assembly



A8.2 Masses of the intact C4bp complex

