UNIVERSITY OF LEICESTER

Characterisation of the classical and lectin pathways of complement activation and the roles of complement inhibitors LAIR-2 and BBK32

Thesis submitted for the degree of

Doctor of Philosophy

At the University of Leicester

By

Jamal Owayed Almitairi

Department of Infection, Immunity and Inflammation

University of Leicester

2019

Statement of originality

This accompanying thesis submitted for the degree of PhD entitled “Characterisation of the classical and lectin pathways of complement activation and the roles of complement inhibitors LAIR-2 and BBK32” is based on work conducted by the author at the University of Leicester mainly during the period between 27 September 2015 and 27 September 2019.

All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references.

None of the work has been submitted for another degree in this or any other university.

Signed Date:

Characterisation of the classical and lectin pathways of complement activation and the roles of complement inhibitors LAIR-2 and BBK32

Jamal Owayed Almitairi

Abstract

The complement system is part of innate immunity and fights invading pathogens via opsonisation and cell lysis. Three pathways activate complement: the classical (CP), the lectin (LP) and the alternative pathway (AP). The CP is initiated by a multimeric 790 kDa complex named C1. This complex is formed from a hexameric, bouquet-like protein called C1q assembled from three polypeptide chains A, B and C and four associated serine proteases; two C1r and two C1s, as a Ca^{2+}-dependant heterotetramer. C1r and C1s are homologous modular proteases composed of an N-terminal CUB1 domain, followed by an EGF-like domain, a CUB2 domain followed by two CCP modules, and a serine protease domain (SP). Several models have been proposed to explain how C1 is assembled. In this thesis I have determined the crystal structure of the C1r-C1s complex and propose a detailed model for assembly of C1. In this model, $C1r_2C1s_2$ complexes adopt an extended S-shaped structure that fold up to form a more compact structure when binding to C1q. Additional crystal structures of a fragment of C1s in complex with collagen-like peptides derived from the A, B and C chains of C1q suggest that C1qC is the leading chain and C1qA is the middle and C1qB is the lagging chain in each collagen-like domain. I also characterised the role of the immune modulator LAIR-2 in downregulating the activation of the CP and LP. Data reported here shows that LAIR-2 binds to the collagen like domain of MBL and thereby regulates activation of the LP. Finally, I assessed the interaction between the *Borrelia burgdorferi* surface protein BBK32 and C1r. Previous studies have suggested that BBK32 inhibits the CP by binding to C1r-CUB1-EGF-CUB2. However, the data reported here shows no interaction between BBK32 and C1r-CUB1-EGF-CUB2 domains.

Acknowledgement

I would like to express my gratitude to my supervisor, *Prof. Russell Wallis* for his support, guidance, time and encouragement during my PhD. I am grateful for the opportunity to undertake this study in his laboratory and the skills and knowledge that I take away from it.

I am also grateful to my friends Luay Alkanan and Abdulrahman Alzahrani for their help and support during this project.

Many thanks to my parents and my brothers Dr. Azzam, Eng. Malek and Dr. Abdulrahman and specially my sister Manal for her help and all of my family for their moral support.

Finally, I thank my wife Amwaj and my daughters Hoor and Leen for thier support and patience during the whole study period.

This project has succeeded by the grace of God, then, the role of the university officials and the department of Infection, Immunity and inflammation.

Publication arising from this thesis

ALMITAIRI, J. O., GIRIJA, U. V., FURZE, C. M., SIMPSON-GRAY, X., BADAKSHI, F., MARSHALL, J. E., SCHWAEBLE, W. J., MITCHELL, D. A., MOODY, P. C. & WALLIS, R. 2018. Structure of the C1r–C1s interaction of the C1 complex of complement activation. *Proceedings of the National Academy of Sciences,* 115**,** 768-773.

Table of content

Abstract…..i

Acknowledgement.....ii

Publication arising from this thesis..... iii

List of figures.....ix

List of tables.....xix

List of abbreviations.....xxi

Chapter 1 . General introduction.....1

1.1 The immune system:.....1

1.1.1 The Complement system:.....4

1.2 Lyme disease and evasion of complement:.....24

1.2.1 BBK32 surface protein:.....27

1.3 Leukocyte-associated immunoglobulin-like receptor (LAIR) proteins:.....28

1.3.1 Interaction between LAIR proteins and the recognition molecules of the CP and LP of complement:.....29

1.4 Hypothesis and general aim of the thesis:.....30

Chapter 2 . Structure of the C1r-C1s complex of the CP.....31

2.1 Introduction:.....31

2.2 Materials and methods:.....34

2.2.1 Materials:.....34

2.2.2 Production and purification of CUB1-EGF-CUB2 domains of C1r and C1s:.....34

2.2.3 Crystallisation of the C1r-CUB2 and 3D C1r-C1s complex:.....35

2.3 Results:.....42

2.3.1 Expression and purification of the 3D C1r and 3D C1s proteins: 42

2.3.2 Gel filtration and purification: 42

2.3.3 Generation of 3D C1r-C1s heterodimer complex by gel filtration:.....44

2.3.4 Interaction of the 3D C1r and C1s in solution:.....44

2.3.5 Structure of the CUB1-EGF-CUB2 domains of the heterodimer of C1r-C1s:.....46

2.4 Discussion:.....49

Chapter 3 . Structure of the C1qA, B and C chains in complex with the 3 domains CUB1-EGF-CUB2 of C1s.....51

3.1 Introduction:.....51

3.2 Materials and methods: 53

3.2.1 Materials: 53

3.2.2 Crystallisation of 3DC1s in complex with C1q chains A, B and C:.....53

3.2.3 Picking and storage of crystals: 55

3.2.4 Data collection and processing: 56

3.2.5 Data processing and structure refinement:.....56

3.3 Results:.....57

3.3.1 Structural analysis of the interaction between 3DC1s and C1q chains: 57

3.3.2 Structure of 3DC1s bound to C1qB:.....59

3.3.3 The interaction between the CUB1 domain of C1s and of C1qC:.....62

3.3.4 Structure of 3DC1s bound to C1q A and C chains:.....64

3.3.5 Structure of 3DC1s bound to C1q C and B: 66

3.4 Discussion: 68

Chapter 4 . Interaction of the immune modulator LAIR-2 with MBL of the lectin pathway for complement activation 71

4.1 Introduction and objectives: 71

4.2 Materials and methods: 73

4.2.1 Materials: 73

4.2.2 Cloning and amplification of FL-LAIR-2 gene: 73

4.2.3 FL-LAIR-2 expression and purification: 75

4.2.4 TEV protease digestion of the FL-LAIR-2 His_6 tag: 76

4.2.5 Preparation of Mannose Sepharose resin: 76

4.2.6 hMBL expression and purification: 77

4.2.7 FL-LAIR-2 crystallisation: 78

4.2.8 Binding of FL-LAIR-2 to hMBL: 80

4.3 Results: 82

4.3.1 Cloning of the FL-LAIR-2 cDNA: 82

4.3.2 Expression and purification of FL-LAIR-2: 83

4.3.3 TEV digestion and removal of the His_6 tag: 85

4.3.4 hMBL expression and purification: 87

4.3.5 FL-LAIR-2 crystallisation: 88

4.3.6 Binding of h-MBL to FL-LAIR-2: 89

4.4 Discussion: 90

Chapter 5 . Inhibition of the Classical pathway by *Borrelia burgdorferi* surface protein BBK32 92

5.1 Introduction: 92

5.2 Materials and methods: 94

5.2.1 Materials: 94

5.2.2 Cloning and production of the FL and C-terminal BBK32 portion: 94

5.2.3 Cloning and production of C1r FL and fragments CUB2, CUB2-CCP1, CUB2-CCP1-CCP2 and CUB2-CCP1CCP2-SP domains: 100

5.2.4 BBK32 crystallisation and optimisation: 109

5.2.5 C-BBK32/C1r-CUB2CCP1 binding assay using gel filtration: 111

5.2.6 BBK32/FL-C1r binding assay using BLI (Bio-Layer Interferometry): 111

5.3 Results: 112

5.3.1 Cloning, expression and purification of FL and C- terminal BBK32 protein: 112

5.3.2 Expression and purification of the FL-C1r protein: 120

5.3.3 Expression and purification of the C1r-CUB2 domain: 122

5.3.4 Cloning of the C1r-CUB2-CCP1, C1r-CUB2-CCP1-CCP2 and C1r-CUB2-CCP1-CCP2-SP constructs: 124

5.3.5 Crystallisation of C-BBK32: 129

5.3.6 Gel filtration binding assay for C-BBK32 to C1r-CUB2-CCP1: 129

5.3.7 BLI Binding assay of C-BBK32 to FL-C1r: 130

5.4 Discussion: 131

Chapter 6 . General discussion 134

6.1 Structure of C1r-C1s complex of the CP: 134

6.2 Structure of the C1qA, C1qB and C1qC chains in complex with the CUB1-EGF-CUB2 domains of C1s: 136

6.3 Interaction of the immune modulator LAIR-2 with MBL of the LP: 136

6.4 Evasion of the CP by the *Borrelia burgdorferi* surface protein BBK32: 137

6.5 Future work: 138

6.5.1 Crystallisation of LAIR-2 protein: 138

6.5.2 Binding of BBk32 to the C1r fragments: 138

APPENDICES 139

References. 145

List of figures

Figure 1-1: *This figure illustrates the difference between humoral and cellular immunity in the adaptive immune response (Abbas and Janeway, 2000). 4*

Figure 1-2: *This diagram illustrates the activation of the complement system via three pathways the CP, the LP and AP, which all results in the formation of the C3 convertase and the formation of MAC leading to target cell lysis.....6*

Figure 1-3: *A schematic illustration of the structure and function of the C1 complex showing how C1q(blue) binds to the immune complex (IgG+antigen) and activates the complement system via the classical pathway (Abbas and Janeway, 2000).....7*

Figure 1-4*: This diagram summarises the complement activation cascade and the biological role of each component. The C3 convertase produces from the CP, LP and alternative pathways cleave C3 protein and convert it into two products, C3a which act as a chemoattractant and recruit leukocytes to the site of infection. C3b on the other hand, remains bound to the pathogen surface leading to opsonaisation. cleavage product C5a also function in similar way (Ricklin and Lambris, 2007). 10*

Figure 1-5*: Schematic illustration of the C1q structure comprising three different subunits A, B and C chains as indicated by the colours, that associate to form the hexamer C1q.....12*

Figure 1-6: *Aligned proteases binding site on the recognition molecules subunits of the CP and LP. The arrows indicate the starting of the collagen domain of each subunit and the green highlight shows the specific protease binding motif within each subunit (Phillips et al., 2009). 13*

Figure 1-7: *A schematic of the MBL oligomer and the structure of each trimeric polypeptide chain subunits. Each subunit consists of a recognition domain as indicated in yellow, a neck region in green and a collagenous domain highlighted in red followed by a linking region. (Dommett, Klein and Turner, 2006). 14*

Figure 1-8: *Schematic illustration for the structure of ficolins. Similar to MBL, ficolins assembled of homotrimeric subunits. Each subunit consists of a fibrinogen recognition domain followed by a collagen-like domain and linking N-terminal region.* *15*

Figure 1-9: *Schematic illustration of CL-11 structure. As shown, CL-11 subunits are composed of three domains; the N-terminal region followed by a collagen-like domain, a neck region and a final C-terminal CRD domain. Three subunits assembled via oligomerisation to form a homotrimeric unit*..... *17*

Figure 1-10*: Structure representation of zymogenic and active C1r protease.*..... *18*

Figure 1-11: *Domain organisation and activation of the C1s protease.* *19*

Figure 1-12: *This diagram illustrates the formation of the MAC and the conformational changes leading to cell lysis (Abbas and Janeway, 2000).*..... *21*

Figure 1-13: *Schematic illustration of various complement regulatory proteins and their functions (Ehrnthaller et al., 2011)*..... *23*

Figure 1-14*: Structure and morphology of the B.burgdorferi bacteria. (A), an electron microscopy scan showing the overall helical shape (left) and a cross-section view (right), showing the periplasmic flagella. (B) is a diagram of the spirochaete showing the flagellar insertion points which are located at each end, six flagella bundle wind around the flexible protoplasmic cylinder. (C) showing each flagellum is inserted in the cytoplasmic membrane and extend to the periplasm. (Rosa, Tilly and Stewart, 2005).*..... *26*

Figure 1-15*: Schematic illustration of the BBK32 structure and the function of each domain.* *27*

Figure 1-16*: Schematic representation of the LAIR-1 and LAIR-2 structure.*..... *29*

Figure 2-1: *(A) Schematic illustration of the traditional C1r-C1s elongated S-shaped model.* *33*

Figure 2-2: *Phase diagram illustrating the process of crystal growth. At low protein/precipitant concentrations, unsaturation condition occurs, and the sample drop remains clear. At very high concentrations of protein/precipitant, precipitation occurs, which is a disordered arrangement of molecules. Between these two phases, in the middle of the diagram is the liable zone. This is where nucleation takes place, followed by crystal growth. Crystal growth can occur at lower concentrations than nucleation, at the right protein/precipitant concentrations (Russo Krauss et al., 2017).* 37

Figure 2-3: *Schematic representation of crystallisation techniques. (A) Vapor diffusion hanging drop, (B) Vapor diffusion sitting drop, (C) Micropatch and (D) Free interface diffusion (Russo Krauss et al., 2017).* 38

Figure 2-4: *Single crystal X-ray diffraction. When an X-ray beams hits a crystal, it is diffracted in different directions based on the arrangement of atoms in the structure. At specific angles, constructive wave interferences will produce spots on the detector*.....39

Figure 2-5: *Schematic representation of Bragg's law from two crystal planes. When an incident X-ray beam strikes one of the crystal planes at the angle (θ), the beam is reflected from an atom of both planes (upper and lower planes). The reflected beams from the upper and lower planes travel differently by a distance of d2sinθ. Both reflected beams combined to make a constructive interface when the path difference is equal to some integral multiple of the wavelength λ*.....40

Figure 2-6: *Non-reducing 15% SDS-PAGE analysis of the Ni^{2+} elution fractions of the 3D C1s. The gel indicates the presence of the protein in the eluted fractions 1, 2 and 3with a molecular mass of 32 kDa.*.....42

Figure 2-7: *Purification and elution profile of the 3DC1s protein. The fractions corresponding to the correct peak between the red dotted lines were collected and checked on a 15% non-reducing SDS-PAGE gel. A similar profile was observed for 3DC1r* 43

Figure 2-8:*15% non-reducing SDS-PAGE analysis showing the eluted fractions from the gel filtration column (fraction 1 to 8). Fractions containing the 3DC1s (32 kDa) protein were pooled together. 43*

Figure 2-9: *Gel filtration elution profile of the C1r-C1s complex. Fractions corresponding to the heterocomplex were collected and concentrated to 4mg/ml for crystallisation. 44*

Figure 2-10: *Analytical gel filtration of 3D C1r and 3D C1s. 3D C1r and 3D C1s are homodimers in the presence of Ca^{2+}. A peak of the heterodimer complex C1r-C1s was observed when mixing equimolar concentration of C1r and C1s which eluted slightly earlier than the homodimers. Dotted lines show elution in a buffer containing EDTA. 45*

Figure 2-11: *Overall structure of the 3DC1r-C1s complex. 3D fragment of C1r (blue) and C1s (green) are both L-shaped, and they interact in an antiparallel arrangement mediated via the CUB-EGF junction. Ca^{2+} is essential for the interaction between the C1r and C1s. Ca^{2+} are indicated in pink spheres. 46*

Figure 2-12: *C1r-C1s interface. (A) Residues buried at the interface of the C1r-C1s complex are indicated. (B) CUB1-EGF-CUB2 structures of C1r-C1s heterocomplex compared to C1s homodimer (PDB code: 4LMF), C1r homodimers and MASP-1/-3 homodimers (PDB code: 3DEM), the homologues serine proteases of the LP. (C) The C1r-C1s dimers overlay with CUB1-collagen (PDB: 4LOR, yellow) and CUB2-collagen structures (PDB: 3POB, wheat) showing the position of the collagen-binding sites at the CUB1 of C1s (blue) and CUB1, CUB2 of C1r (green). The collagen is in grey. 47*

Figure 2-13: *Structure of the C1r-C1s asymmetric unit. There are two dimers in the asymmetric unit which probably reflect the arrangement of the two C1r (green) and two C1s (cyan) molecules in the C1 complex. The binding sites for collagen are shown by red circles. The collagen-binding site in each CUB2 of C1r is blocked by the CUB1 domain of its partner making a total of four binding sites. Ca^{2+} molecules are shown as pink spheres. 50*

Figure 3-1: *Sequence of the peptides used for crystallisation with 3DC1s. Flanking GPO (where O is hydroxyproline) to stabilize the collagen allowing self-assembly of the collagen triple helix. Residues underlined shows the binding site in each chain of C1q.....57*

Figure 3-2: *This figure shows the overall structure of the 3D C1s in complex with C1qC peptide. The three domains of C1s are shown, which binds to the collagen peptide of C1q. Ca^{2+} is essential for the interaction and are shown as pink spheres.....59*

Figure 3-3*: The interaction between the CUB1 domain of C1s and of C1qB. The chains are colored as follow; black for the lagging strand, grey for the middle strand and white for the leading strand. 60*

Figure 3-4*: The interaction between the CUB1 domain of C1s and of C1qC. C1s residues Ser100, Asp98 and Glu45 interact with Lys15 of the lagging strand of C1qC as in the C1qB complex. The Ca^{2+} is shown as a pink sphere. 62*

Figure 3-5: *The interaction between C1qA/C with the CUB1 domain of C1s. Ser100, Asp98 and Glu45 of the C1s interact with Lys15 of the lagging strand of the collagen (chain C). Additionally, hydrogen bonds are present between the Tyr52 residue of CUB1 and the amide group of Val16 of the middle strand (chain A,) as well as the side chain of Lys15 of the middle strand (chain A). Interestingly, density for the Tyr18 of chain A was only observed in the middle strand of the three collagen chains (not the leading or lagging) the other two chains were both predominantly C1qC. A further interesting observation was that Tyr18 packs against the hydrophobic side chain of Met19 of the leading strand. These data are consistent with C1qA being the middle strand and C1qC being the lagging strand. 64*

Figure 3-6: *This figure presents the structure of the complex between C1s and C1qC/B. The ratio of C:B was 2:1 respectively. Density for C1qB was observed predominantly in the middle strands whereas density for C1qC was observed in the lagging strand. Residues Ser100, Asp98 and Glu45 of the CUB1 domain of C1s form hydrogen bonds with the Lys15 of the lagging*

strand of the collagen (chain C). Additionally, hydrogen bonds are present between the Tyr52 residue of the CUB1 and the amide group of Asn16 of the middle strand (chain B). 66

***Figure 4-1:** Schematic representation of the collagen-like domain of MBL and C1q. The recognition molecules of the CP and LP contain a collagen-like domain of 59 amino acids for the LP and 81 amino acids for the CP as indicated with light blue colour. The collagen domain is believed to be the binding site for LAIR-2.....72*

***Figure 4-2:** Vector map of the expression vector pLEICS-03**.** LAIR-2 gene was inserted in the EcoR1site. 75*

***Figure 4-3:** BLI principle showing the steps of sample loading and target detection (ForteBio manual). 80*

***Figure 4-4:** 1% agarose gel of the FL-LAIR-2 cDNA amplified via PCR. The expected size of the product was 459bp. 82*

***Figure 4-5:** 15% non-reducing SDS-PAGE analysis of six elution fractions (left to right) following affinity purification on a Ni-column. FL-LAIR-2 migrated at the expected size of 22 kDa. 83*

***Figure 4-6:** Gel filtration purification of FL-LAIR-2 on a Superdex 75 16/60 column. The protein eluted from the column at the expected molecular mass of a monomer 22kDa. 84*

***Figure 4-7:** SDS-PAGE analysis of gel filtration purification for the FL-LAIR-2 protein. No contaminants were detected after gel filtration. 84*

***Figure 4-8:** 15% non-reducing SDS-PAGE analysis of the TEV digestion of FL-LAIR-2. The digested protein eluted in the flow-through fractions (1 to 5) with a molecular mass of 18kDa. TEV enzyme itself possesses a his_6 tag and was separated successfully in the high imidazole elution fractions 6 to 10. .....85*

Figure 4-9: *Gel filtration purification for the digested FL-LAIR-2. The digested protein peak eluted slightly later from the column (78 ml) compared to un-digested FL-LAIR-2 (72 ml) confirming the successful removal of the His_6 tag. 86*

Figure 4-10: *Gel filtration fractions on a 15% non-reducing SDS-PAGE gel for the digested FL-LAIR-2. Fractions 3-10 (left to right) were pooled and concentrated. Fractions 1 and 2 were discarded to remove the minor contaminant present at ~30 kDa. 86*

Figure 4-11: *h-MBL purification analysis on non-reducing SDS-PAGE gel. Lanes 1 and 2 represent the flow through and wash fractions. Lanes 3 to 8 are the elution fractions and contain purified h-MBL.....87*

Figure 4-12: *Binding of FL-LAIR-2 to the immobilised h-MBL. LAIR-2 was injected at different concentration (800, 400, 200,100 and 50µM). The K_D calculation was performed by fitting to 1:1 model as indicated in red lines. 89*

Figure 5-1: *This figure shows the vector map of pLEIC-01. BBK32 protein Was inserted in the EcoR1 site. 96*

Figure 5-2: *pET-28a expression vector map. The C1r-CUB2 domain was inserted at the Sal1 site. 101*

Figure 5-3: *Vector map of pED4, the C1r inserts are cloned into the Sal1/EcoR1 polylinker (Kaufman et al., 1991). 106*

Figure 5-4: *1% agarose gel showing the 5' (N-BBK32) and 3' (C-BBK32) fragments. Bands of the expected sizes 500bp for C-BBK32 and 400bp for N-BBK32 were observed. 113*

Figure 5-5: *1% agarose gel showing the successful PCR amplification of the FL-BBK32 gene with an expected size of 1100bp.....113*

Figure 5-6: *15% non-reducing SDS-PAGE for testing the expression of C-BBK32. The protein migrated at the expected size 23kDa and was present in both total lysate (TL) on the left and*

in the soluble lysate (SL) on the right, 1,2 and 3 indicate different loading volumes (1 µl, 2 µl and 5 µl).....114

Figure 5-7: *Non-reducing SDS-PAGE of the Ni^{+2}- NTA Sepharose purification of C-BBK32. Lane 1 indicates the flow-through and lane 2 shows the non-specific wash with 10mM Imidazole. Lanes 3 to 12 shows the eluted fractions. The large band corresponds to C-BBK32 which migrated at the expected size of ~23kDa.....115*

Figure 5-8: *Gel filtration elution profile for C-BBK32.....115*

Figure 5-9: *SDS-PAGE analysis of the gel filtration elution fractions for C-BBK32. The protein eluted from the column with a molecular mass of 23kDa. 116*

Figure 5-10: *Ni^{2+}- NTA purification of the TEV-digested BBK32 protein analysed by SDS-PAGE. Most of the protein eluted in the flow-through (lane1) with a molecular mass of 18kDa, indicating successful digestion and removal of the his_6 tag. Lane 2 shows the low imidazole wash, and lanes 3 to 7 are the elution fractions. 117*

Figure 5-11:*15% SDS-PAGE analysis of digested C-BBK32 following gel filtration. C-BBK32 migrates with a molecular mass of 18kDa confirming successful digestion of the His_6 tag. 117*

Figure 5-12: *SDS-PAGE of the Ni column purification of the FL-BBK32. Lanes 1to 6 represents the elution fractions, and lanes 7 and 8 show the flow throw and low Imidazole wash respectively. 118*

Figure 5-13: *Gel filtration elution profile for the FL-BBK32. The elution peak indicated between the red arrows represent the FL-BBK32 protein which eluted at the expected position for a monomer.....119*

Figure 5-14: *SDS-PAGE analysis of the gel filtration elution fractions. The protein migrates on the gel as expected with a molecular weight of 46kDa of a monomer. 119*

Figure 5-15: *SDS-PAGE analysis following Ni^{2+}- NTA Sepharose purification of the harvested FL-C1r protein. Lane 1 shows the media flow through, and lane 2 represents the non-specific*

wash with low Imidazole. Lanes 3 to 5 shows the elution fractions. The eluted protein migrated on the gel with a molecular mass of ~80 kDa.....120

Figure 5-16: *Gel filtration elution of FL-C1r on a Superdex 200 16/60 column. 121*

Figure 5-17: *15% non-reducing SDS-PAGE analysis of the gel filtration eluted fractions. The FL-C1r protein migrated on the gel with the expected molecular mass of 150kDa. Fractions 1 to 10 were pooled and concentrated.....121*

Figure 5-18: *Reducing SDS-PAGE on a 15% gel of the inclusion bodies. The lanes 1 to 4 shows the wash steps and lanes 5,6 and 7 showed three different loading amounts of the inclusion bodies (1µl, 2 µl and 5 µl). The purified inclusion bodies migrated with a molecular mass of 13kDa, consistent with the mass of the CUB2 domain.....122*

Figure 5-19: *Gel filtration elution profile of C1r-CUB2. The first peak to elute from the column was the aggregated protein, and the peak indicated between the red arrows corresponds to the correctly folded C1r-CUB2.....123*

Figure 5-20: *15% non-reducing SDS-PAGE analysis of the gel filtration elution fractions of the C1r-CUB2 domain. The C1r-CUB2 protein eluted from the column with the expected molecular mass of 13kDa.....123*

Figure 5-21: *1% agarose gel showing amplification of the C1r fragments by PCR. Product 1 is the fragment encoding the C1r-CUB2-CCP1-CCP2 domains (~780bp), product 2 encodes the C1r-CUB2-CCP1-CCP2-SPdomains (~1500bp), and product 3 encodes the C1r-CUB2-CCP1(~550bp) domains. 124*

Figure 5-22: *EcoR1conformation digestion for each clone which shows two different sizes corresponding to the pED4 vector at ~5000bp and the size of the correct inserts. 125*

Figure 5-23: *SDS-PAGE analyses of the Ni^{+2}-NTA column elution for C1r-CUB2-CCP1-CCP2. Lanes 1 to 6 represent the elution fractions, 7 is the flow-through, 8 is the no imidazole wash and 9 the low imidazole wash. The protein migrated in the expected size of 31kDa. ..126*

Figure 5-24: *Gel filtration purification of C1r-CUB2-CCP1-CCP2 protein. 126*

Figure 5-25: *15% non-reducing SDS-PAGE gel for the gel filtration purification of C1r-CUB2-CCP1-CCP2 protein. The protein migrated on the gel as expected with a molecular mass of 31kDa.....127*

Figure 5-26: *SDS-PAGE analysis of the eluted fractions from Ni^{2+}- NTA column purification of the C1r-CUB2-CCP1. The fractions 1 to 4 are the high imidazole elutions and fractions 5 and 6 are the flow-through, and low imidazole washes respectively. Only one fraction contained the C1r-CUB2-CCP1. The molecular mass was ~25kDa.....127*

Figure 5-27: *Gel filtration purification of the Ni-column elutions for C1r-CUB2-CCP1 on a Superdex 200 column.....128*

Figure 5-28: *SDS-PAGE analysis for the gel filtration purification of C1r-CUB2-CCP1 protein. The gel filtration step was successful in eliminating the contaminants with a resulting protein purity of ~99%. The protein eluted from the column as expected with a molecular mass of 25kDa. Fractions were pooled together and concentrated. 128*

Figure 5-29: *Overlayed elution profiles of C-BBK32, C1r-CUB2-CCP1 and BBK-32/C1r complex. Individual elution profiles are shown in green and blue, and the mixture profile is shown in red.....129*

Figure 5-30: *Binding of FL-C1r to the immobilised C-BBK32 analysed by BLI. Flat lines shows no binding.....130*

Figure 5-31: *.....133*

List of tables

Table 1.1: *Table illustrating the main features of innate and adaptive immunity (Janeway, 2001)* *1*

Table 2.1: *Data collection and refinement statistics. The highest resolution shell is shown in parentheses.* *48*

Table 3.1: *This table summarises the different 3DC1s-C1q complexes used for crystallisation. Each 3DC1s protein was mixed with a deferent ratio of C1q chains A, B and C as presented in the table. 3DC1s was at 3.66mg/ml, and peptides were at C1qA- 1.2mg/ml, C1qB- 1.32mg/ml and C1qC- 1.39mg/ml.* *54*

Table 3.2: *Optimisation steps for 3DC1s protein in complex with C1qA, B and C crystallisation. Each buffer was used to set up a 48 well plate. The same optimization process was applied to each protein-peptide complex.* *55*

Table 3.3: *Summary of the data collection and refinement parameters of C1qB in complex with C1s-CUB1.* *61*

Table 3.4: *Summary of the data collection and refinement parameters of C1qC in complex with the C1s-CUB1domain.* *63*

Table 3.5: *Summary of the data collection and refinement parameters of the C1qA+C heterooligomer in complex with C1s-CUB1.* *65*

Table 3.6: *Summary of the data collection and refinement parameters of C1qC+B heterooligomer in complex with C1s-CUB.* *67*

Table 3.7*: This table shows the polar interactions between C1q chains in complex with the C1s-CUB1 domain. Hilighted resedues are conserved in all chains of C1q.* *68*

Table 4.1: *Oligonucleotides primers for the amplification of FL-LAIR-2 gene.* *74*

Table 4.2: *Summary of the PCR run and reaction mixture****.*** *74*

Table 4.3: *Optimisation conditions for FL-LAIR-2 crystallisation. Each buffer was used to set up a 48 well crystallisation plate.* *79*

Table 5.1: *Primers used to amplify the FL and C-terminal of BBK32.* *95*

Table 5.2*:This table summarises the PCR master mix composition and the conditions for PCR amplification.* *96*

Table 5.3: *Summary of the PCR mixture and conditions for C1r amplification.* *104*

Table 5.4: *Sequence list of the primers used to generate C1r fragments. Start and stop codon are highlighted in red, an optimised Kozak sequence was added to increase the level of expression and is underlined, the* His_6 *tag was added to enable purification with* Ni^{2+} *and is shown in bold.* *105*

Table 5.5: *Summary of the crystallisation optimisation conditions for BBK32 protein. Each buffer was used to set up a 48 well plate.* *110*

List of abbreviations

3MC	Malpuech, Michels and Mingarelli-Carnevale syndromes
aa	Amino acid
AMD	Age-related Macular Degeneration
AP	Alkaline Phosphate
Arg	Arginine
Asp	Aspartic acid
BBS	Barbital Buffer Saline
bp	base pair
BSA	Bovine Serum Albumin
C1-INH	C1-Inhibitor
C1r	Complement components 1r
C1s	Complement components 1s
C3aR	C3a Receptor
C4BP	C4-bBinding Protein
C5aR	C5a Receptor
CCP	Complement Control Protein
cDNA	Complementary DNA
CFU	Colony Forming Unit
CHO	Chinese Hamster Ovary
CL-11	Collectin-11
CL-K1	Collectin Kidney 1
CL-L1	Collectin Liver
CR	Complement Receptor
SRCR	Scavenger Receptor Cysteine-Rich
CRD	Carbohydrate Recognition Domain
CRP	C-reactive protein
CUB	Complement C1r/C1s, Uegf, Bone morphogenetic protein

DAF	Decay Accelerating Factor
dATP	Deoxyadenosine Triphosphate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
EGTA	Ethylene glycol-bis-N,N,Ń,Ń-tetraacetic acid
FB	Factor B
FD	Factor D
FH	Factor H
FI	Factor I
His	Histidine
Ig	Immunoglobulin
IgG	Immunoglobulin G
IgM	Immunoglobulin M
Ile	Isoleucine
IPTG	Isopropyl Thiogalactoside
kDa	Kilodalton
LB	Luria-Bertani
LTA	Lipoteichoic Acid
MAC	Membrane Attack Complex
MAp	MBL-Associated protein
MASP	MBL-Associated Serine Protease
MBL	Mannan Binding Lectin
MBP	Mannan Binding Protein
MEM	Minimum Essential Medium

ml	Milliliter
mM	Millimole
ng	Nanogram
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PAMP	Pathogen Associated Molecular Pattern
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RNAse	Ribonuclease
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
Ser	Serine
SLE	Systemic Lupus Erythematosus
SP	Serine Protease
TBE	Tris Buffered Saline with EDTA
TEMED	Tetramethylethylenediamine
v/v	volume/volume
Xaa	Any amino acid
µg	Microgram
µl	Microliter
BLI	Bio-Layer Interferometry

Chapter 1 . General introduction

1.1 The immune system:

The immune system is the main line of defence against pathogens (non-self) and self (damaged tissues). It is divided into two effector mechanisms; innate and adaptive responses (Table 1.1). Innate immunity includes a wide range of pattern-recognition receptors, it is rapid and does not possess memory. In contrast, the adaptive immune response is more specific and changes depending on the microbial threat. Memory cells trigger adaptive immune response rapidly and specifically upon a second infection (Janeway, 2001).

	Innate	Adaptive
Characteristics		
Specificity	For molecules shared by groups of related microbes and molecules produced by damaged host cells	For microbial and nonmicrobial antigens
Diversity	Limited; germline encoded	Very large; receptors are produced by somatic recombination of gene segments
Memory	None	Yes
Nonreactivity to self	Yes	Yes
Components		
Cellular and chemical barriers	Skin, mucosal epithelia; antimicrobial molecules	Lymphocytes in epithelia; antibodies secreted at epithelial surfaces
Blood proteins	Complement, others	Antibodies
Cells	Phagocytes (macrophages, neutrophils), natural killer cells	Lymphocytes

Table 1.1: *Table illustrating the main features of innate and adaptive immunity* (Janeway, 2001)

.

Through its physical and chemical barriers, innate immunity can prevent pathogens from penetrating and colonizing the host cells. These barriers include;

- Epithelial cells and anti-microbial agents on their surfaces and at mucous membranes.

- Phagocytic cells such as neutrophils, macrophages, dendritic cells (DCs) and natural killers (NKs), which can engulf the pathogens and destroy them.

- Inflammatory mediators and proteins of the complement system.

- Regulators of the immune response such as cytokines, that enable immune cells to coordinate and communicate with each other (Fallis, 2013).

The innate immune response is crucial during the early stages of an infection to educate and help trigger the adaptive response (Murphy, Travers and Walport, 2009). Innate or native immunity is commonly described as a non-memory response. Through its pattern recognition receptors (PRRs) present on the professional phagocytic cells, innate immunity can distinguish between self and non-self antigens (Kawai and Akira, 2010). These receptors recognise common features of pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharides (LPSs) or lipoteichoic acid (LTA)(Fujita, 2002). In contrast, the adaptive immune response is characterised by its high specificity for distinct molecules in pathogens, which can generate an immunological memory and thus responds more vigorously to the second exposure to the same pathogen (Denman, 1992). Several effector mechanisms carry out the functions of the adaptive immune response. These are listed as follows;

- **Antibody secretion** neutralises pathogens and blocks their ability to enter the host cell. Antibody binding promote the phagocytic activity.
- **Phagocytic cells** can coordinate with T-helper cells and ingest microbes to destroy them.

- **Cytotoxic T-lymphocytes (CTLs)** play a defensive role at sites that cannot be reached by antibodies or phagocytic cells.

The role of the adaptive immune response is to activate one or more of these mechanisms to eradicate pathogens at different anatomical positions in the body, either in circulation or inside the cells. There are two types of adaptive immune responses (Figure1.2); the first type is **humoral immunity** is mediated by the components in the circulation called antibodies, which are secreted by B-lymphocytes. Antibodies recognise microbial antigens in the three following ways (Janeway, 2001);

- **Neutralisation,** where antibodies can bind to bacterial toxins and viruses, preventing them from attaching to host cells.
- **Opsonisation,** where antibodies coat the pathogen surface, causing it to be ingested by phagocytic cells.
- **Complement activation,** where the antibodies bind to the bacterial cells and activates the first recognition molecule of the classical pathway of the complement system, leading to cell lysis via the activity of the membrane attack complex (MAC).

The second type of adaptive immune response is **cell-mediated immunity,** which is mediated by T-lymphocytes. This immune response is effective against intracellular pathogens to isolate them from circulation, via the mechanisms illustrated in (Figure 1.1).

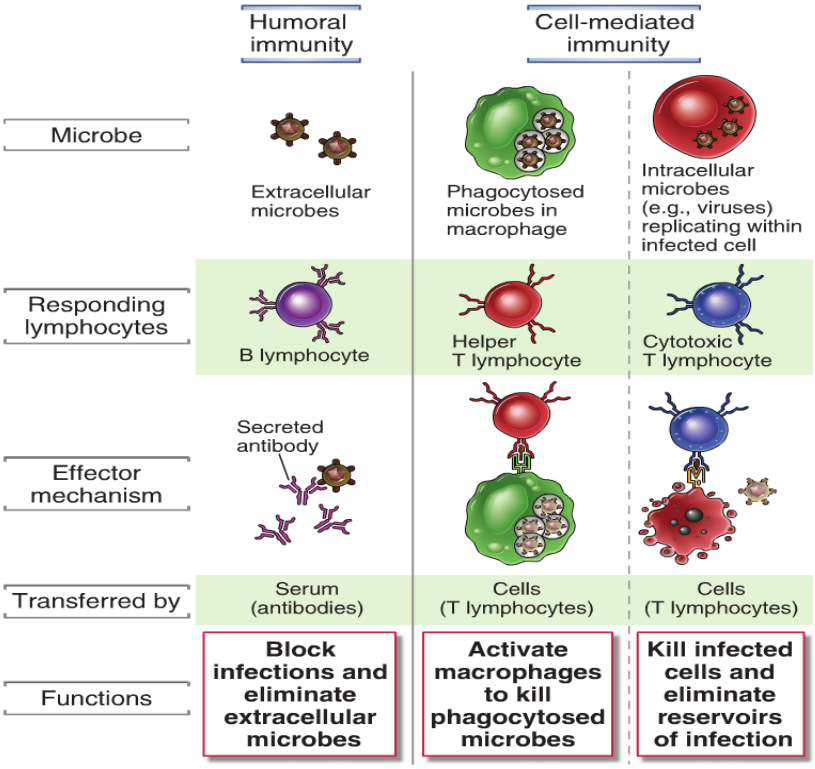


Figure 1-1: *This figure illustrates the difference between humoral and cellular immunity in the adaptive immune response* (Abbas and Janeway, 2000).

1.1.1 The Complement system:

The complement system is a major part of the innate immune response for host defence and is also important for the development of adaptive immunity. It is a cascade of more than 30 proteins produced mainly by the liver (Ricklin *et al.*, 2010), comprising soluble proteins and membrane-expressed receptors and regulators. Complement was discovered in the 1890s by Jules Bordet as a heat-labile bactericidal component of normal plasma/serum (Ehrnthaller *et al.*, 2011). It was named for its ability to complement the phagocytic cells function of clearing pathogens. Complement proteins present in the plasma in their inactive forms, known as proenzymes (zymogens). When they are activated by proteolytic cleavage, these zymogens are converted into their enzymatically active forms which then initiate a cascade of sequential protein cleavage (Noris and Remuzzi, 2013). This activation ultimately leads to the activation

of the terminal pathway and formation of the membrane attack complex (MAC), resulting in the lysis of the pathogen (Serna *et al.*, 2016). The complement system is responsible for many biological activities within the immune system. It generates opsonins that are covalently attached to pathogens to facilitate phagocytosis as well as stimulates the inflammatory response through the production of anaphylatoxins, it promotes the adaptive immune response by lowering the threshold for B cell activation, and lyses target cells via the MAC (Ward and Rosenthal, 2014). Complement also functions as a scavenger system for clearing immune complexes, cell debris and apoptotic cells (Fujita, 2002). The complement system is highly regulated to prevent unwanted activation and host-cell damage, through circulating and membrane-bound regulators that control activation.

Three distinct pathways activate the complement system; the classical (CP), lectin (LP) and alternative (AP) pathways (Figure 1.2). The CP and LP depend on pattern recognition molecules for activation (PRMs), C1q in CP, MBL, ficolins and collectins-K1 and L1 in LP, to recognise pathogen-associated molecular patterns (PAMPs) or danger associated molecular patterns (DAMPs) that are present on the surface of pathogens or self-stressed cells (Figure 1.2). Upon binding to a pathogen surface, the PRMs of the CP and LP activate their associated serine proteases C1r and C1s in the CP and MASP-1 -2 and -3 in the LP. All three pathways converge to generate a C3 convertase which cleaves C3 and activates the terminal pathway of complement by cleaving complement component C5. The product of C5 cleavage is C5b, which in turn associates with C6, C7 and C8 to form the C5bC6C7C8 terminal complement complex (TCC), which attaches to the surface of the target cell and initiates the formation of the MAC. The resulting pores formed from multiple copies of the transmembrane complement protein C9 lead to cell lysis via osmotic activity (Figure 1.2). Each pathway will be described in more detail below.

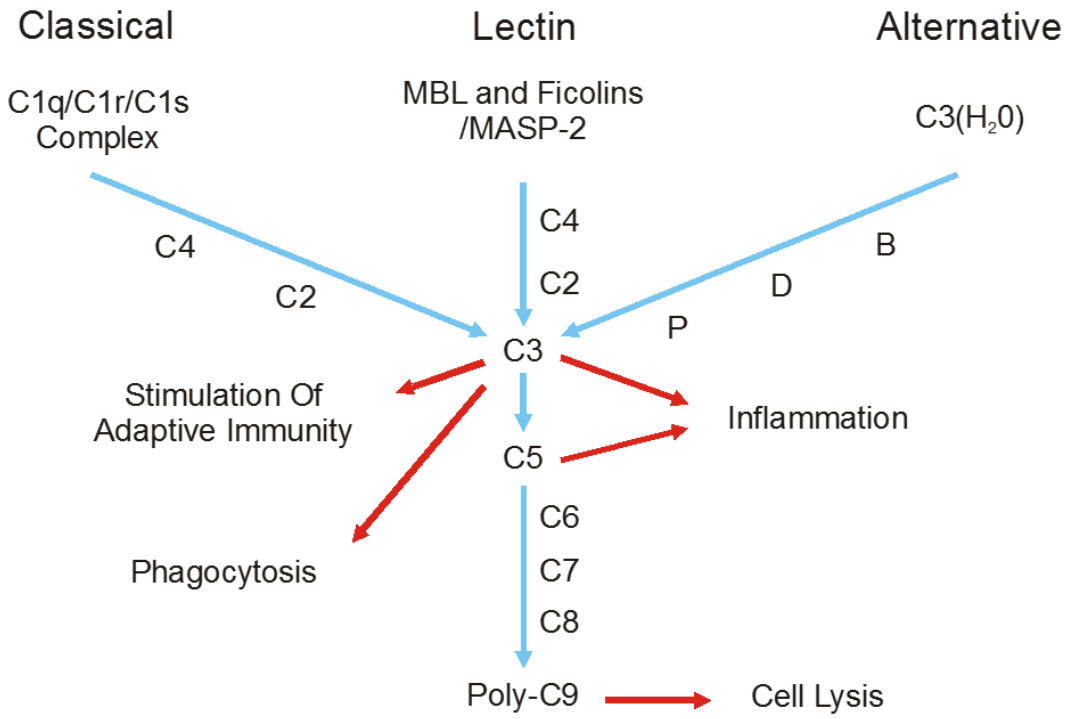


Figure 1-2: *Complement activation pathways.This diagram illustrates the activation of the complement system via three pathways the CP, the LP and AP, which all results in the formation of the C3 convertase and the formation of MAC leading to target cell lysis.*

1.1.1.1 Activation of the Classical pathway:

The CP is activated by the multi-component C1 complex, which is about 790 kDa (Figure 1.3). This complex consists of the PRM subcomponent C1q (460 kDa) which forms a functional complex with hetero-tetrameric serine proteases C1s-C1r-C1r-C1s. In the absence of C1q, the two C1r (90 kDa) and two C1s (80 kDa) form an elongated S-shaped structure. The proteases fold up within C1q in a Ca^{+2}-dependent manner. C1q recognises the Fc region of the immunoglobulin IgG classes IgG1, IgG2, IgG3 and IgM via its globular heads to activate C1r and C1s (Gál *et al.*, 2009). Furthermore, C1q can bind directly to polyanionic structures on the surface of pathogens (Nauta *et al.*, 2002). This binding results in the auto-activation of C1r, which in turn cleaves and activates C1s. Activated C1s cleaves C4 into C4b and C4a; C2 is

then cleaved and activated by C4b into C2a and C2b, after which the C2a fragment binds to C4b to cleave C3 and generate the C3 convertase (C4bC2a) (Bajic *et al.*, 2015).

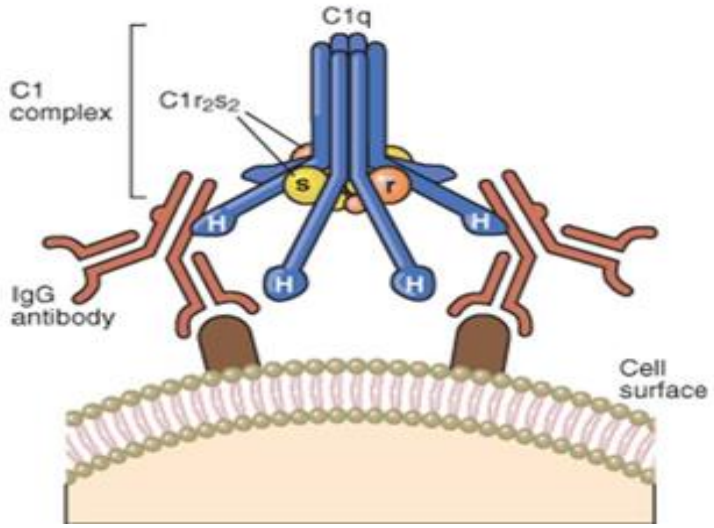


Figure 1-3: *A schematic illustration of the structure and function of the C1 complex. It show how C1q(blue) binds to the immune complex (IgG+antigen) and activates the complement system via the classical pathway* (Abbas and Janeway, 2000).

1.1.1.2 Activation of the Lectin pathway:

The lectin pathway of complement is antibody-independent pathway and is activated via three recognition components: mannan-binding lectin (MBL), three ficolins (L-H-M ficolins) and collectin-11 (CL-11), which bind to pathogen-associated molecular pattern (PAMPs) including polysaccharides, carbohydrates and acetylated sugars on the pathogen surface (Wallis *et al.*, 2007). The recognition proteins of the LP circulate as heterogeneous molecules and form complexes with three zymogen MBL-associated serine proteases (MASPs) at a ratio of 1:1 (one recognition molecule and one MASP dimer) (Chen *et al.*, 2001). Three MASPs have been identified; MASP-1, -2 and -3, as well as two non-enzymatically active proteins MAp-19 and MAp-44, which are produced by the alternative splicing of the MASP-2 and MASP-1 genes respectively (Sim, Schwaeble and Fujita, 2016). MASP-1 and -3 are alternatively splice

products that share the same N-terminal domains but differ in their serine protease domains. Activation of the LP takes place when the recognition molecules bind to the pathogen surface; following this, MASP-2 is activated either by autolysis or via MASP-1, and then catalyses the cleavage of C2 and C4, thereby generating the LP C3 convertase (C4b2a). Recent research has reported that human serum deficient in MASP-1 and MASP-3 does not activate the LP efficiently (Garred *et al.*, 2016). The LP convertase cleaves C3 to C3a and C3b, subsequently cleaving C5 and activating the terminal pathway of complement as described above.

1.1.1.3 Activation of the Alternative pathway:

In contrast to the CP and LP, the activation of the alternative pathway of complement does not require a recognition molecule. The AP is mediated by the spontaneous hydrolysis of the complement component C3 in plasma to generate C3(H_2O), which has a similar function to C3b. The hydrolysed C3 binds to factor B and is cleaved by factor D to generate two fragments; Ba which dissociates from the complex, and Bb, which remains attached to C3(H_2O) and results in the formation of the fluid phase AP C3 convertase C3(H_2O)Bb, which, cleaves C3 into C3b and C3a, creating an amplification loop (Harboe and Mollnes, 2008).

The membrane phase AP C3 convertase is generated by the binding of factor B to membrane-bound C3b in an Mg^{2+}-dependent manner. Subsequently, C3bB is cleaved by factor D to generate C3bBb (Farries *et al.*,1988). The binding of the protein properdin stabilises the alternative pathway C3 convertase (C3bBb) and extends its half-life by ~10-fold (Hourcade, 2006). The AP pathway serves as an amplification loop during the activation of the CP and LP (Thurman and Holers, 2006). Protection of the host cells from unwanted complement activation is mediated by several regulators, which will be discussed later in this chapter.

1.1.1.4 The biological roles of complement activation:

Complement activation leads directly to pathogen lysis via MAC formation. This function is active predominantly against Gram-negative bacteria (Hadders *et al.*, 2012). In addition, the small activation fragments C3a, C4a and C5a are chemoattractants (C5a in particular) which can guide leukocytes and macrophages to the site of activation. They also trigger an inflammatory reaction and increase the vascular permeability of endothelial cells (Iacovache, van der Goot and Pernot, 2008) as well as aid the production of chemokines such as tumour necrosis factor (TNF) and interleukins IL-1 and IL-8 (Markiewski and Lambris, 2007).

When cleavage products C3b, C4b and C5b are deposited on the pathogen surface, they act as opsonins to stimulate phagocytosis when they bind to complement receptor 1 (CR1) on the surface of macrophages and leukocytes. C3b and C4b attach covalently to cell surfaces via a reactive thioester bond that is exposed upon the cleavage of C3 and C4.

Another function of the complement system is clearing cell debris and immune complexes. When an immune complex is opsonised by C3b and C4b, they bind to erythrocytes expressing CR1 and are subsequently transported to the spleen and liver, where they are eliminated (Barnum, 2017).

Complement functions as a link between innate and adaptive immunity via the C3 cleavage product C3d binding the to complement receptor 2 (CR2) on B-cells, to lower the threshold for B-cell activation. CR1 and CR2 modulate the proliferation and activation of B-cells when the cleavage product iC3b (the product of C3b degradation by factor I) opsonises a pathogen and subsequently binds to the receptors on the surface of B-cells. This initiates the production of specific antibodies and the differentiation of B-memory cells (Carroll and Isenman, 2012). CR3 and CR4 also modulate phagocytosis for cells that have been opsonised by iC3b.

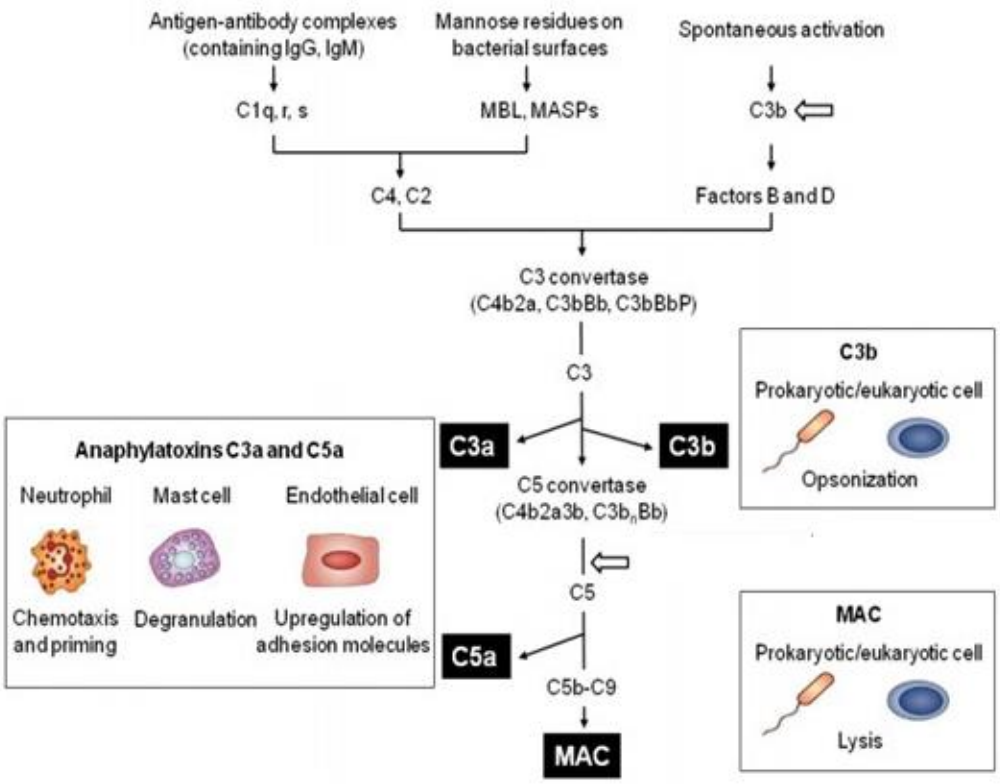


Figure 1-4*: This diagram summarises the complement activation cascade and the biological role of each component. The C3 convertase produces from the CP, LP and alternative pathways cleave C3 protein and convert it into two products, C3a which act as a chemoattractant and recruit leukocytes to the site of infection. C3b on the other hand, remains bound to the pathogen surface leading to opsonaisation. cleavage product C5a also function in similar way* (Ricklin and Lambris, 2007).

1.1.1.5 Complement pattern recognition molecules:

1.1.1.5.1 The structure of the classical pathway recognition molecule C1q:

The recognition subcomponent C1q is a 460 kDa hexameric glycoprotein composed of 18 polypeptide chains of three different types: A chain (223 residues), B chain (226 residues) and C chain (217 residues) (Gaboriaud *et al.*, 2004). Each chain consists of three regions; a short N-terminal region (3-9 residues), a collagen-like domain (~81 residues) which contains a proteases-binding motif and a globular domain (gC1q) at the C-terminus (~135 residues) which recognises immune complexes (Son *et al.*, 2012a). The C1q chains A, B and C are homologues and are encoded by the genes *C1qA*, *C1qB* and *C1qC* that are located on the human chromosome 1. Within C1q, the A and B chains are covalently linked together by a disulphide bond. Individual trimers are linked by disulphide bonds between two C chains, resulting in six heterotrimeric subunits ABC-CBA (Figure 1.5). The C1q subunits are assembled together through disulphide bonds to form a hexameric molecule. These chains spread apart at a kink that is present within the collagenous domain to form the bouquet-like structure (Agrawal *et al.*, 2001). C1q performs a wide range of complement and non-complement functions. It can recognise structures and molecules on pathogens surfaces, apoptotic cells and indirectly by recognising antibodies and C-reactive protein-bound ligands. Furthermore, C1q can bind directly to polyanionic structures on the surface of the pathogen (Bíró *et al.*, 2007). C1q via its globular heads can recognise the Fc portion of the immunoglobulin IgG class (IgG1, IgG2, IgG3) and IgM (Moreau *et al.*, 2016).

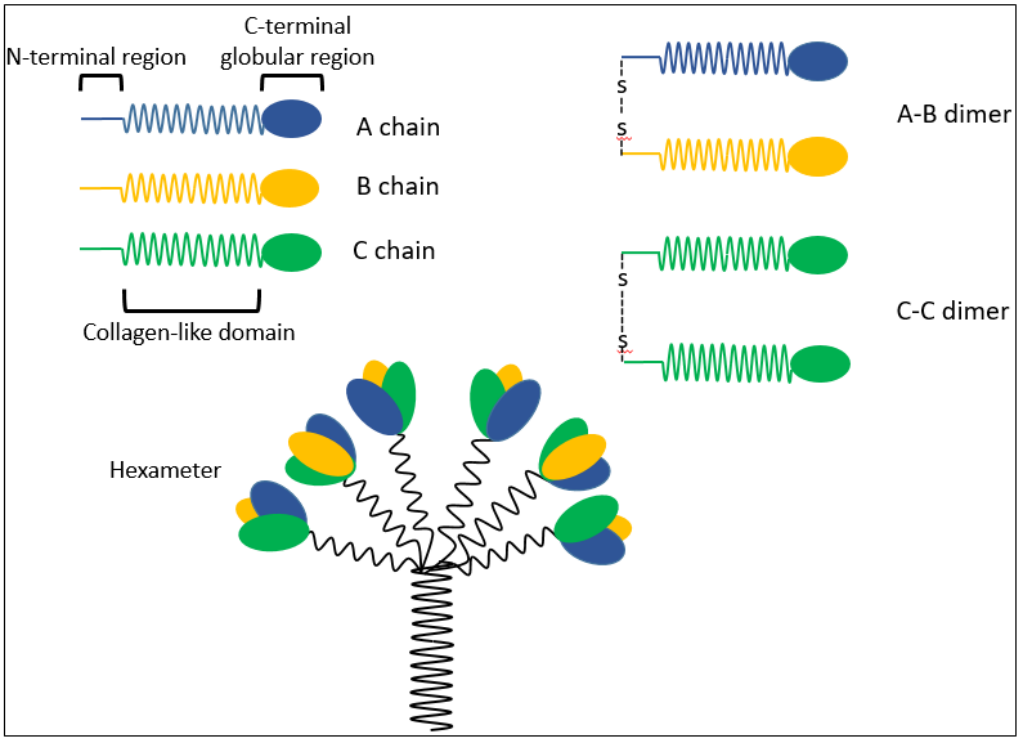


***Figure 1-5**: Schematic illustration of the C1q structure comprising three different subunits A, B and C chains as indicated by the colours, that associate to form the hexamer C1q*

The interaction between C1q and IgG requires multiple IgG interaction, so C1q will only bind to the surface of an immobilised IgG. Unlike most complement proteins, which are synthesised by the liver, C1q is produced by a wide range of cells, including monocytes, epithelial cells and dendritic cells (Gaboriaud *et al.*, 2012). Within the collagenous domain of C1q, there is a specific motif, Hyp-Gly-Lys-Xaa-Gly-Pro, by which the C1r and C1s attach to C1q. This motif is present in all of the recognition molecules of the CP and LP (Figure 1.6) (Phillips *et al.*, 2009). The unique structure of C1q by which having three different chains A, B and C accounts for its specificity towards its associated proteases $C1rs_2$. C1q forms a compact complex with $C1rs_2$ which, upon activation, undergoes a conformational change, allowing the stalks of C1q to splay and expose the SP domains of C1r dimers at the core of C1q enabling it to cleave its substrate C1s and subsequently activate the CP. A recent structure of the interaction between C1s CUB1-EGF-CUB2 domains and a collagen peptide derived from C1q containing the

proteases binding site has been solved (Venkatraman Girija *et al.*, 2013). From the solved structure, it was not possible to determine the order of the C1q chains A, B and C that interact with the CUB1 domain of C1s. My structural determination of the C1s-CUB1-EGF-CUB2 in complex with C1qA, C1qB and C1qC reveals the chain's order, as will be explained in the following chapter.

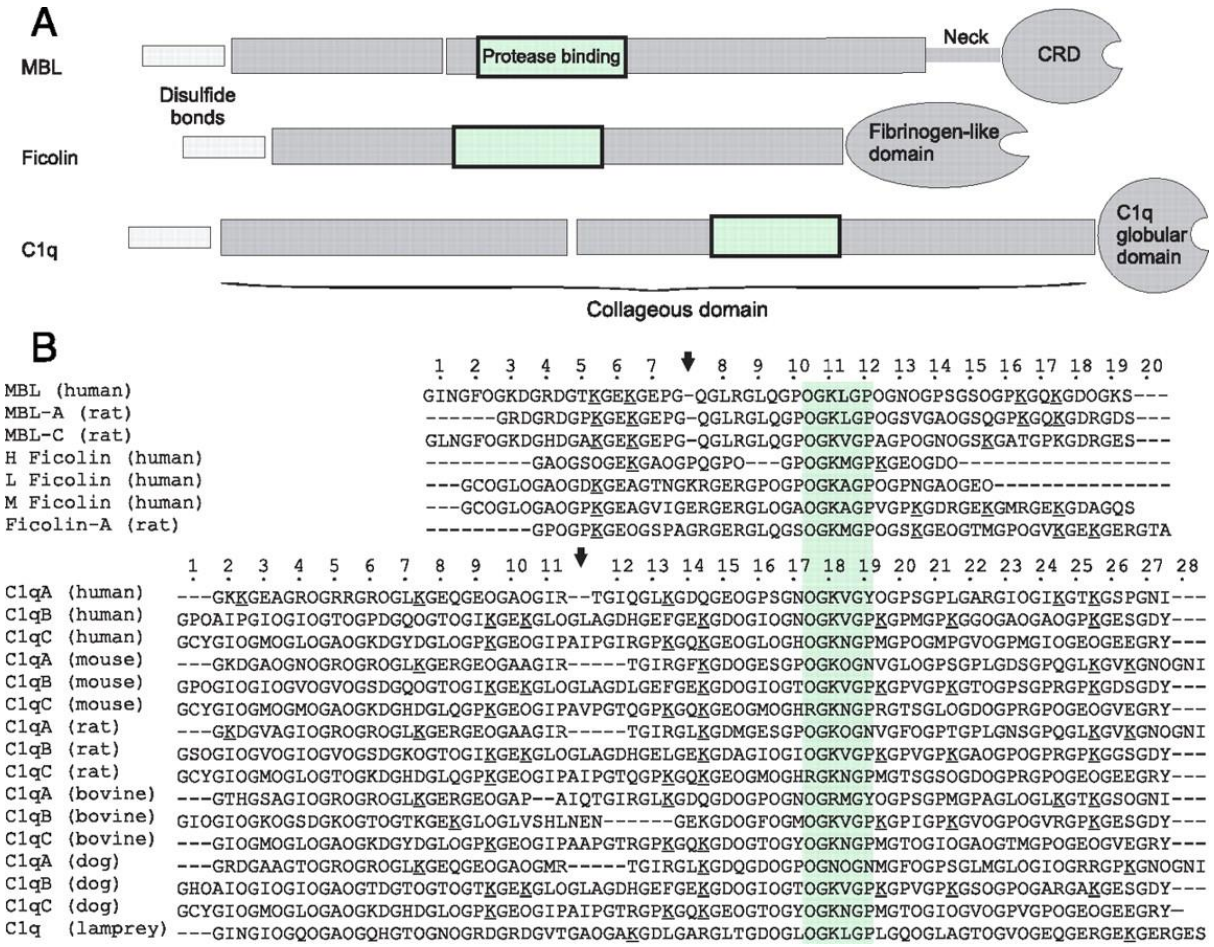


Figure 1-6: *Aligned proteases binding site on the recognition molecules subunits of the CP and LP. The arrows indicate the starting of the collagen domain of each subunit and the green highlight shows the specific protease binding motif within each subunit* (Phillips *et al.*, 2009).

1.1.1.5.2 Structure of the Lectin pathway recognition molecules:

1.1.1.5.2.1 Structure of MBL:

Mannan-binding lectin (MBL) is part of the collectin family with an N-terminal collagen-like domain and a C-terminal lectin domain. It is produced mainly by the liver. Unlike C1q of the CP, MBL is assembled from homotrimeric polypeptide subunits which can associate to form trimeric subunits. These subunits associate during biosynthesis in groups of two, three and four, forming a bouquet-like shape (Jensenius *et al.*, 2009). A single MBL molecule comprises a cysteine-rich N-terminal domain which stabilises the oligomer by forming disulphide bonds between the subunits (Wallis *et al.*, 1997); following this is a collagen-like domain, an α-helical coiled neck region and finally a carbohydrate-recognition domain (CRD) at the C-terminus (Dommett, Klein and Turner, 2006) see (Figure1.7).

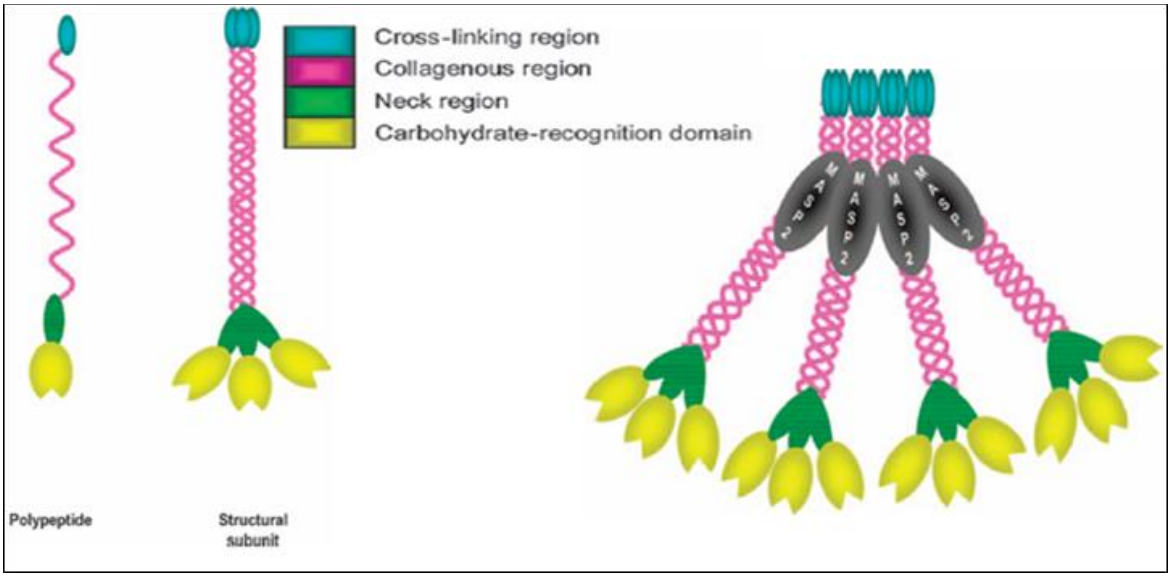


Figure 1-7: *A schematic of the MBL oligomer and the structure of each trimeric polypeptide chain subunits. Each subunit consists of a recognition domain as indicated in yellow, a neck region in green and a collagenous domain highlighted in red followed by a linking region.* (Dommett, Klein and Turner, 2006).

Through its CRD, MBL recognises patterns that are present on pathogens and exposed neoepitopes on apoptotic and injured cells and tissues. MBL forms a complex with MASPs in the ration of 1:1 (MBL oligomer: MASP dimer), which, when activated, leads to opsonisation and lysis of the pathogen via the LP of complement.

1.1.1.5.2.2 Structure of Ficolins:

Ficolins are similar in structure to MBL; they are composed of trimeric subunits in which each chain consists of a collagen-like domain at the N-terminus followed by a fibrinogen-like domain at the C-terminus (Figure 1.8). Three types of ficolins have been identified in humans: ficolin-1 (also called M-ficolin) which is present in the plasma and whose mRNA is expressed in monocytes, spleen and lung, ficolin-2 (also called L-ficolin), which is produced in the liver and circulates in the blood, and finally ficolin-3 (called H-ficolin or Hakata antigen), which is also produced in the liver as well as the lung (Zhang and Ali, 2008). Ficolins are all similar in structure but vary in size between 32 and 35 kDa.

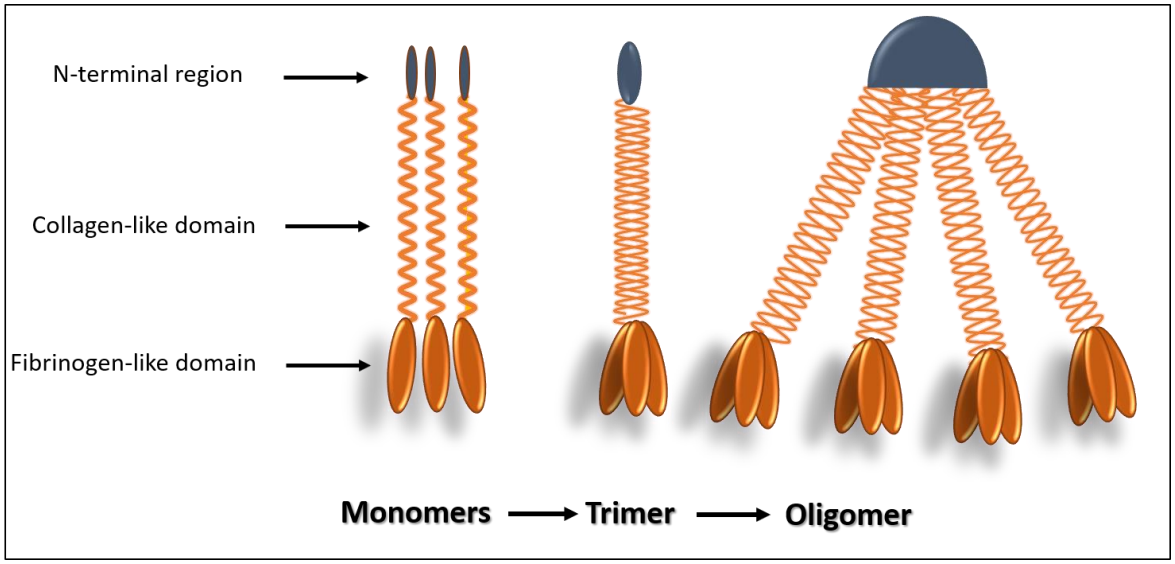


Figure 1-8: *Schematic illustration for the structure of ficolins. Similar to MBL, ficolins assembled of homotrimeric subunits. Each subunit consists of a fibrinogen recognition domain followed by a collagen-like domain and linking N-terminal region.*

Ficolins can form complexes with MASPs 1, 2 and 3, thereby initiating complement via the LP (Endo, Matsushita and Fujita, 2011). By recognising N-acetylated sugars via their fibrinogen-like domain, ficolins can bind to a variety of pathogens and activate their associated serine protease, leading to the lysis of the pathogens by forming C3 and C5 convertase and initiating the terminal pathway of complement (Matsushita, 2013).

1.1.1.5.2.3 Structure of collectin-11 and other collectins:

Like MBL and ficolins, collectin-11 is present in the serum and forms complexes with MASPs 1, 2 and 3 to activate the lectin pathway (Beltrame *et al.*, 2015). Collectin-11 is part of the C-type lectin superfamily; thus, it shares the same binding motifs as MBL within its CRD. In addition, like MBL, ficolins and C1q, it possesses a MASP-binding motif within its collagen-like domain (Hyp-Gly-Lys-Xaa-Gly-Pro) (Phillips *et al.*, 2009). The structure of collectin-11 is similar to MBL, with a collagen-like domain followed by a neck region and finally a C-type lectin recognition domain (Figure1.9). In addition to CL-11, other collectins have been identified, including collectin liver 1 (CL-L1 aka CL-10) and collectin placenta 1 (CL-P1). Recent studies have shown that CL-10 and CL-11 circulate in serum as heterooligomers (Selman and Hansen, 2012). However, not all tissue express CL-10 and CL-11, so homooligomers are also likely to be produced by some tissues e.g. CL-11 by the kidney.

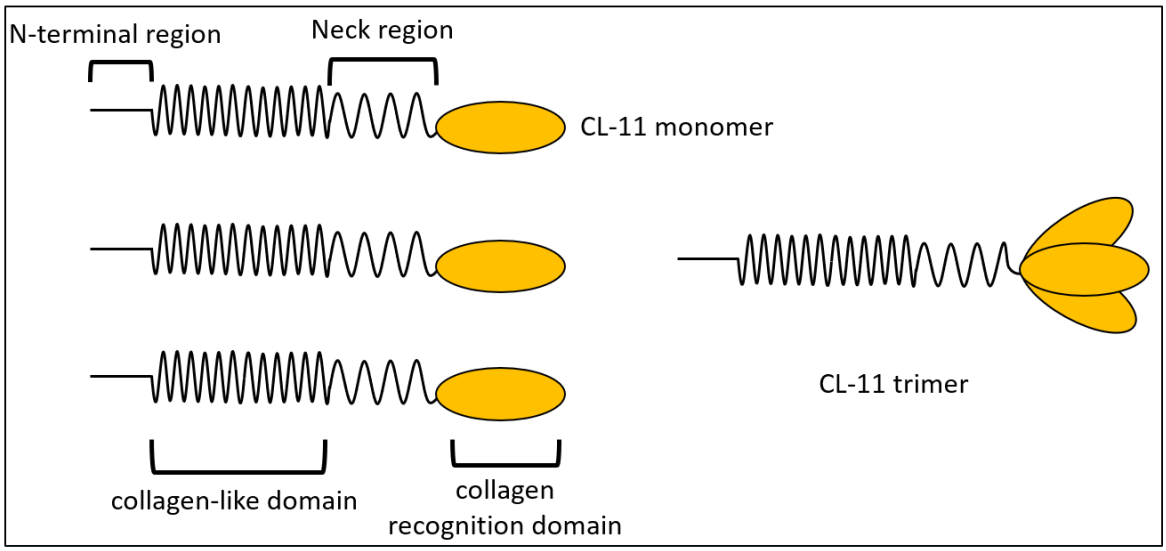


Figure 1-9: *Schematic illustration of CL-11 structure. As shown, CL-11 subunits are composed of three domains; the N-terminal region followed by a collagen-like domain, a neck region and a final C-terminal CRD domain. Three subunits assembled via oligomerisation to form a homotrimeric unit.*

1.1.1.6 Serine proteases of the complement system:

1.1.1.6.1 Proteases of the Classical pathway:

C1r and C1s are the initiating serine proteases of the CP. They form a complex with the CP recognition subcomponent C1q and circulate in their inactive form as zymogens (Almitairi *et al.*, 2018). C1r auto-activates and cleaves its substrate C1s, which subsequently cleaves C4 and C4b-bound C2 to initiate the downstream cascade of the CP. C1r and C1s are homologues and share the same domain structure, both consisting of six domains starting with two CUB domains at the N-terminus (Uegf for C1r and Bmp1 for C1s), separated by an epidermal growth factor (EGF)- like domain, and followed by two complement control modules, CCP1 and CCP2 and finally a C-terminal serine protease domain (SP) (Figure 1.10) (Venkatraman Girija *et al.*, 2013). In the absence of C1q, C1r/C1s forms an elongated S-shaped hetero-tetramer with two C1r proteases at the centre, C1s-C1r-C1r-C1s, which packs at the centre of the C1q molecule

with six Ca^{2+}-dependant binding sites between the CUB domains and the collagen stalks of C1q (4-binding sites for C1r and 2-for C1s) (Bally *et al.*, 2009). This binding is mediated via the Hyp-Gly-Lys-Xaa-Gly-Pro motif, located on the stalks of the recognition subcomponent (Bally *et al.*, 2009). The two C1r molecules dimerise though the interaction between the CCP1 domain of one molecule with the SP domain of its partner in a Ca^{2+}-independent anti-parallel manner (Almitairi *et al.*, 2018). Furthermore, C1r and C1s are also linked in an antiparallel arrangement with the CUB1 domain of C1r, interacting with the EGF domain of C1s and vice versa, forming the core of the C1 complex (C1s-C1r-C1r-C1s). The first enzymatic activity of the CP is initiated by the auto activation of the C1r protease through cleavage at the Arg446-Ile447 bond in the serine protease domain (Figure 1.10); the activated C1r then cleaves its substrate C1s (Figure 1.11). Following this, the activated C1s then initiates stepwise proteolytic activity via the CP pathway and cleavage of the C4/C2 proteases (Wijeyewickrema *et al.*, 2013).

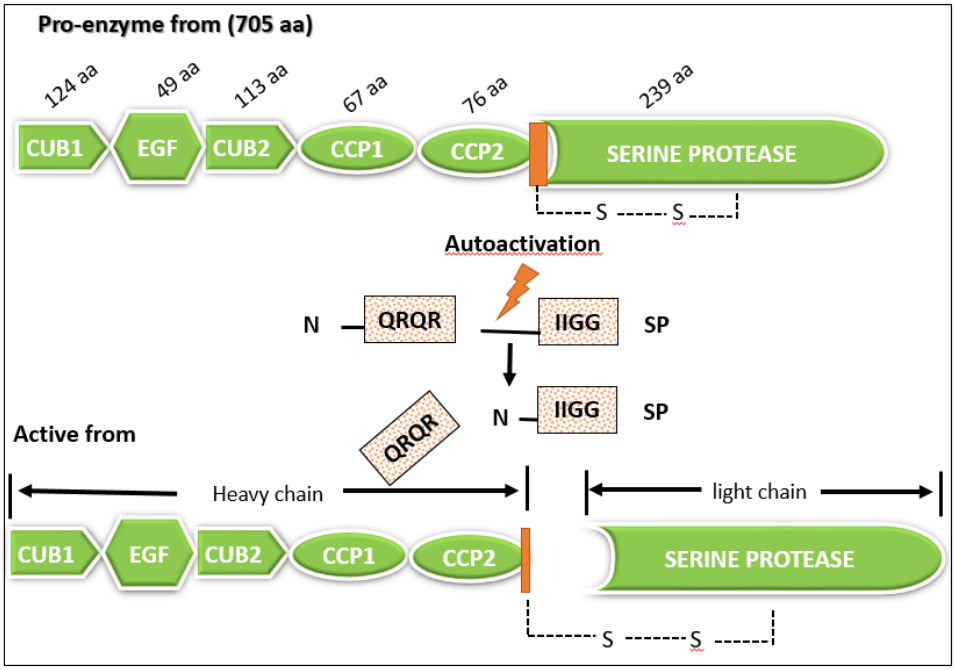


***Figure 1-10**: Structurale representation of zymogenic and active C1r protease.*

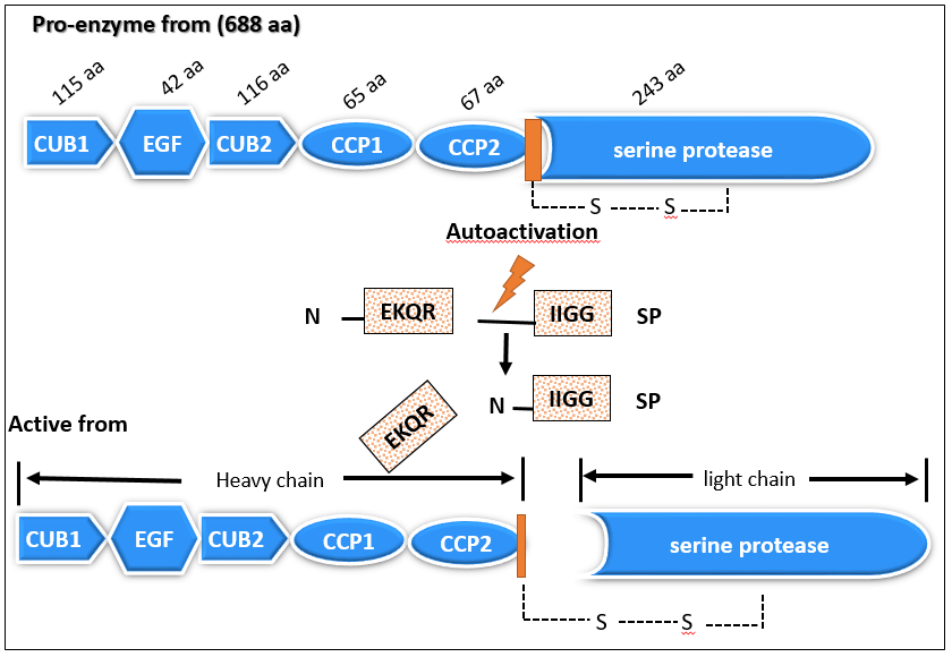


Figure 1-11: *Domain organisation and activation of the C1s protease.*

1.1.1.6.2 Proteases of the lectin pathway:

The LP is activated by three types of serine proteases called MASPs 1, 2 and 3. These proteases bind to the recognition molecules of the lectin pathway, MBL, CL-11 and ficolins at a ratio of one MASP dimer to one recognition oligomer. MASPs are homologues of each other and of C1r and C1s of the CP, with the same domain structure comprising two Ca^{2+}-CUB domains flanked by one Ca^{2+}-EGF domain and followed by two complement control domains CCP1 and CCP2, with a final catalytic serine protease domain SP. In addition, MAp-19 (19kDa), which is a product of the alternative splicing of the MASP-2 gene located on chromosome 1 of humans, consists of the CUB1-EGF domain of MASP-2. MAp-44 is a product of the alternative splicing the of MASPs 1/3 gene located on chromosome 3 of humans and is formed of the CUB1-EGF-CUB2-CCP1 domains of MASP-1. Similar to C1r/C1s of the CP, the Ca^{2+}-binding CUB domains of MASPs mediate the binding of the proteases to the recognition molecules. MASPs homodimerize through their CUB-EGF domains in an anti-parallel arrangement and have the ability to bind to the Hyp-Gly-Lys-Xaa-Gly-Pro motifs of up to 4 collagen-like stems

(Phillips *et al.*, 2009). The LP serine proteases circulate in their inactive zymogenic form alone or in complex with MBL, CL-11 or ficolins, and activate when the complexes bind to an activating surface. MASP-1 (81kDa) was the first protease to be discovered and is mainly expressed in the liver and other organs such as the placenta, heart, kidneys, lungs and small intestine. MASP-1 functions as an enhancer for the LP activation as its presence in the serum is 10 -fold higher than the concentration of MASP-2. Additionally, MASP-1 cannot activate LP alone as it cannot cleave C4. Instead, it cleaves MASP-2, thereby facilitating the LP activation via MASP-2. MASP-1 also cleaves C2, and therefore may contribute to the formation of the C3 convertase. MASP-2 is mainly produced by the liver and is essential for the lectin pathway activation. It cleaves C4 and C4b-bound C2 to form the LP C3 convertase (C4b2a). Recent studies suggest that MASP-2 can bypass the C3 convertase and activate C3 directly (Yaseen *et al.*, 2017). In this way, activation can proceed even in the absence of C4. MASP-3 is the most recently discovered protease of the LP. It shares the same heavy chain as MASP-1 but differs in the short linker region of 15 amino acid residues and the SP domain. MASP-3 is expressed in hepatocytes, the lungs, the small and large intestines, the heart, prostate, brain, ovary and spleen. Unlike MASP-1 and MASP-2, MASP-3 cannot auto-active to initiate the LP activation but requires activation by either MASP-1 or MASP-2. MASP-3 cleaves pro-factor D, the component of the AP providing a link between the LP and AP (Dobó *et al.*, 2016).

1.1.1.7 The Terminal pathway and the formation of the MAC complex:

The final stage of complement activation is the activation of the terminal pathway. The C3 convertase (C4b2a) generated from the CP and LP or (C3bBb) from the AP can cleave C3 to C3b (which remains bound to the surface of target cells via its thioester bond) and C3a (anaphylatoxin). The deposition of C3b in close proximity to C4b2a and C4bBb on the surface of the pathogen triggers an enzymatic step in which C5 is cleaved by either C4b2aC3b or C4bBbC3b to generate C5b, which binds to the target cell surface and the anaphylatoxin C5a, which in association with C3a and C4a, mediates the recruitment of immune cells to the site of infection. Following this, C5b binds to C6 and C7 to form the C5b67 complex, which can bind to the pathogen surface. Subsequent attachment of the pore forming C8 and the addition of multiple copies of the C9 proteins form a cylindrical pore structure called the MAC (Figure 1.12). Formation of the pore results in osmotic imbalance, leading to the lysis of the target cell (Lovelace *et al.*, 2011).

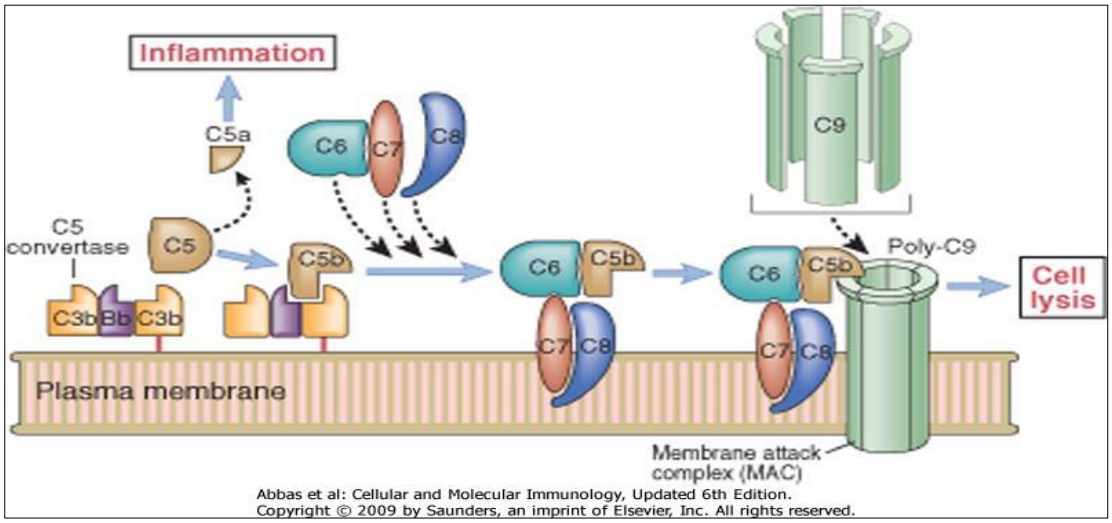


Figure 1-12: *This diagram illustrates the formation of the MAC and the conformational changes leading to cell lysis* (Abbas and Janeway, 2000).

1.1.1.8 Regulation of the Complement System:

The regulation of complement activation is crucial for the prevention of self-damage of the host tissues. Without tight regulation, constant activation of the AP forming low-level C3 would lead to the deposition of complement components onto the host cells, which resulting in a severely destructive effect (Zipfel and Skerka, 2009). There are two types of complement activation regulators; fluid-phase and membrane-bound regulators (Figure 1.13).

1.1.1.8.1 Fluid-Phase Regulators:

The first fluid-phase regulator is the C1 esterase inhibitor (C1-INH); this regulator is a glycoprotein of the serpin family that controls the activation of the CP and LP by the binding to their proteases C1r-C1s and MASPs to form covalent complexes. By doing this, it can prevent them from cleaving C2 and C4, thus stopping downstream activation.

The second regulator is the C4-binding protein (C4BP), which is a co-factor of the serine protease factor I; it regulates the CP and LP by blocking C2 from binding to C4b, and thus controls the formation of C3 convertase (Trouw, Blom and Gasque, 2008).

Furthermore, the AP is regulated by factor H. This regulator is another co-factor of factor I and regulates the AP C3 convertase either by removing Bb from the C3bBb complex (i.e. decay acceleration) or acting as a co-factor for factor I-mediated degradation of C3(Merle *et al.*, 2015). Factor H has lectin activity and associates to host tissues by binding to sialic acid residues on host glycoproteins. Importantly, it can also bind to heparin, thereby protecting the underlying extracellular matrix from complement attack. The lytic activity of the MAC complex is under the regulation of two plasma proteins called clusterin and S-protein/vitronectin. The target of these regulators is the binding site of the C5b67 complex on the cell surface, and therefore prevents the formation of MAC complex and C9 polymerisation (Frank, 2010).

Finally, properdin or factor P binds and stabilises C3 and C5 convertase complexes; it is the only positive fluid phase regulator (Miyaike *et al.*, 2002).

1.1.1.8.2 Cell Surface-Bound Regulators:

Complement receptor 1 (CR1) is a co-factor for factor I and is expressed on B-cells, red blood cells (RBCs) and leukocytes. It functions as a modulator of C3 and C5 convertase activity. The membrane co-factor protein (MCP), which is a co-factor for complement factor I, is a regulatory protease synthesised in the liver; it inactivates C3b via degradation to iC3b. Like CR1, the decay acceleration factor (DAF/CD55) regulates the formation of C3 convertase for the CP, AP and LP. Finally, protectin (CD59) regulates the formation of the MAC complex by binding to C8 and C9, preventing them from forming a complex with C5b67 (Ehrnthaller *et al.*, 2011).

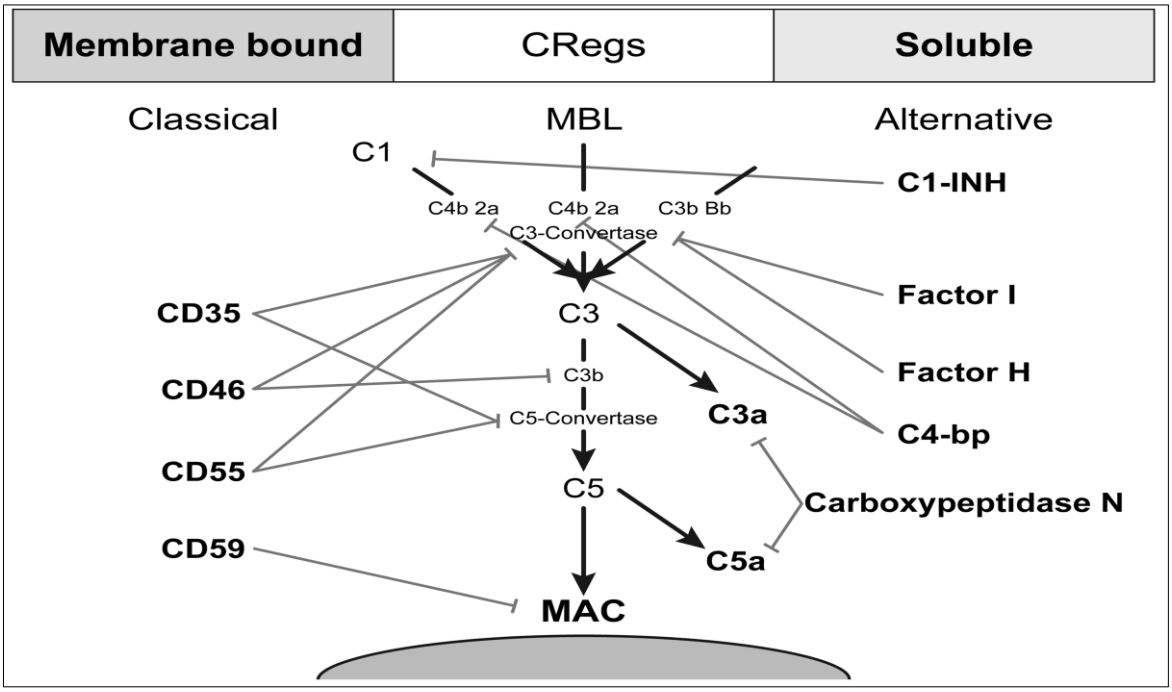


Figure 1-13: *Schematic illustration of various complement regulatory proteins and their functions* (Ehrnthaller *et al.*, 2011).

1.2 Lyme disease and evasion of complement:

Lyme disease, also known as Lyme borreliosis, is the most commonly reported tick-borne disease that is affecting 20,000 people around the world, with 1,136 cases reported in England and Wales in 2016 (Public Health England, 2017). It is a multi-organ infection caused by the spirochetes of *Borrelia burgdorferi.* This Bacterium is transmitted to humans via four species of *Ixodes* tick species, including *Ixodes ricinus,* which is the primary vector for spreading this disease in Europe. The tick can become infected by feeding on the natural reservoirs of *Borrelia spp.,* which includes birds, several strains of mice and pets such as dogs and cats. The disease can be subsequently transmitted to humans via a tick bite. The *Ixodes* ticks can have either 2 or 3 life cycle stages (larval, nymphal, and mature), and can become infected with *Borrelia* at any stage during feeding on an infected reservoir. The risk of disease transmission from an infected tick relates to the duration of the tick feeding. Feeding must exceed 24 hours before the risk becomes substantial (Huppertz and Girschick, 2016).

The causative agent for Lyme disease, *B.burgdorferi,* was first isolated in 1982 by Dr. Willy Burgdorfer from the typical skin lesions, blood and cerebrospinal fluid of infected patients (Garcia *et al.*, 2016). Three types of *Borrelia* species have been identified to cause with Lyme disease, *B.burgdorferi* sensu strictu (s.s), *B.garinii* and *B.afzelii,* which are collectively referred to as the *B.burgdorferi* sensu lato (s.l) complex.

The *B.burgdorferi* sensu lato group are Gram- negative bacteria, that are 10-30µm in length and 0.2-0.25µm in width. They are obligate bacterium which depends on the infected host for nutrition. Morphologically, *B.burgdorferi* form a rod-shaped helix with an inner and outer membrane surrounding 7 to 11 periplasmatic flagella that are located sub terminally to the protoplasmic cylinder, and overlapping at the middle of the cell (Rosa, Tilly and Stewart, 2005). The outer membrane of *B.burgdorferi* lacks lipopolysaccharides and is easily disrupted,

containing a peptidoglycan layer on the inner membrane of the bacterium (Figure1.14). Lyme disease is characterised by three stages based on clinical manifestation of symptoms; (early localised infection, early disseminated infection and persisting late infection). During the first stage which is also referred to as localised early disease and is recognised by the presence of an inflammatory skin rash, known as erythema migrans (EM), located at the site of the tick bite, indicating the localisation of the bacteria in the skin. Symptoms also include an influenza-like illness without specific symptoms (Huppertz and Girschick, 2016). The EM lesions usually appear within 4 to 30 days from the initial tick bite. The second stage occurs when the spirochetes spread and disseminated to multiple organs such as joints, causing arthritis, to multiple organs such as the heart, causing myocarditis, and to the nervous system, causing facial paralysis. The second stage of infection usually develops weeks or months after the first one. If left untreated, Lyme disease can progress to the third stage resulting in chronic arthritis and skin disorders. Because of the tick vector size, and lack of pain and itching at the bite site, nearly 40% of patients do not seek treatment before the EM appears. The spirochetes bacterium *B.burgdorferi* produces several outer surface membrane proteins that enables it to avoid clearance by the host's innate immune system, while spreading and colonising the host tissues (Bush and Vazquez-Pertejo, 2018). Many of the proteins produced by *B.burgdorferi* interact with complement regulatory proteins by recruiting them to its surface, in this way protecting itself from complement attack. These proteins include CspA and the OspE- related protein family (ErpA, ErpP and ErpC) which recruit factor H to the surface of the bacteria, thereby blocking the activity of the AP pathway. CspA also binds to C7 and C9 to inhibit the formation of the MAC. More recently, a surface protein called BBK32 was found to interact with the CP of complement, blocking its activity (Garcia *et al.*, 2016).

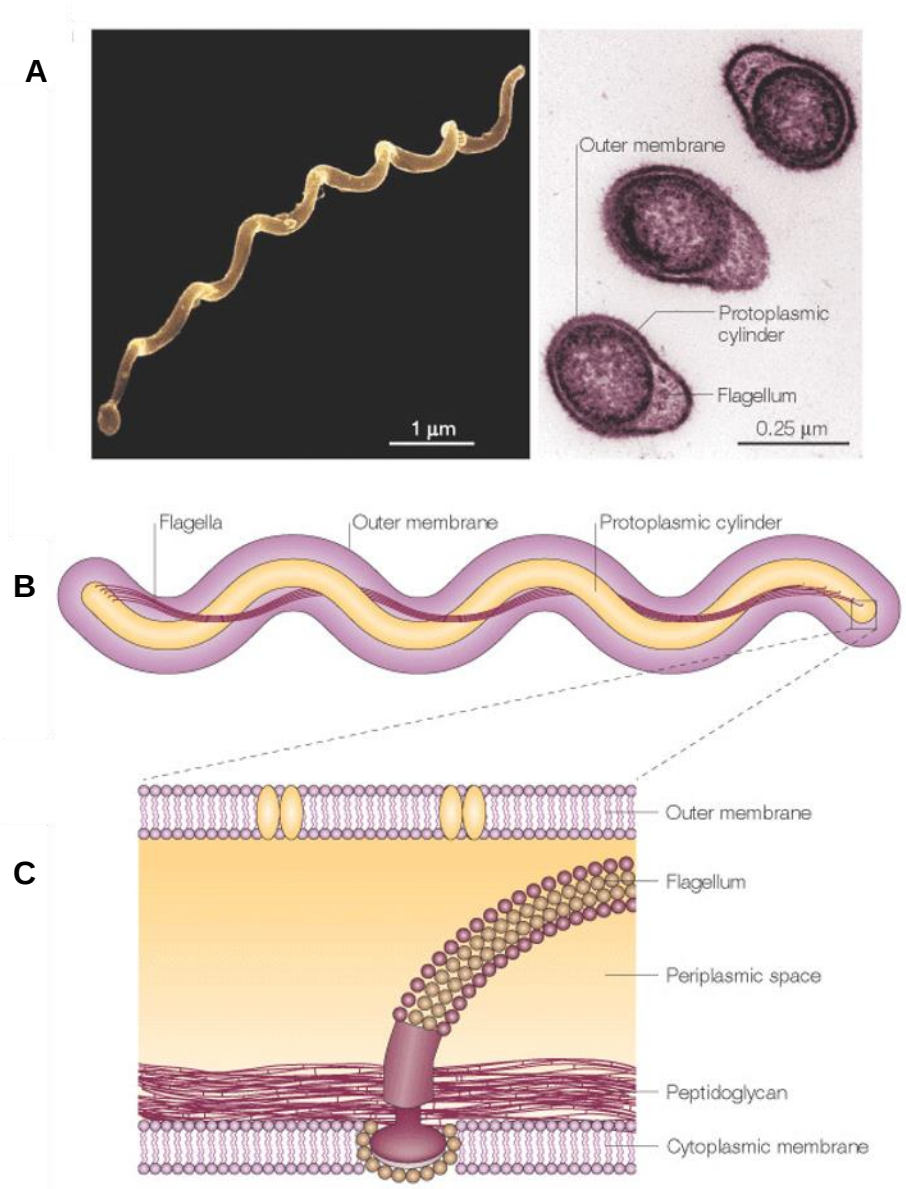


Figure 1-14*: Structure and morphology of the B.burgdorferi bacteria. (A), an electron microscopy scan showing the overall helical shape (left) and a cross-section view (right), showing the periplasmic flagella. (B) is a diagram of the spirochaete showing the flagellar insertion points which are located at each end, six flagella bundle wind around the flexible protoplasmic cylinder. (C) showing each flagellum is inserted in the cytoplasmic membrane and extend to the periplasm.* (Rosa, Tilly and Stewart, 2005).

1.2.1 BBK32 surface protein:

The *B.burgdorferi* surface protein BBK32 is a multifunctional lipoprotein that is expressed on the surface of the bacterium and aids the spread and colonisation to host cells (Lin *et al.*, 2015). BBK32 is a 47kDa adhesion protein that can bind to the extracellular matrix of the host organs via its ability to bind to the collagen-binding domain of fibronectin (Fn) (Probert and Johnson, 1998). Through its ability to interact with several macromolecules, (Fn) can bind to collagen, fibrin and heparin, which can collectively be hijacked by BBK32.

BBK32 is overexpressed during the first stage of infection, implying its important role in disseminating *B.burgdorferi* away from the bite site of the tick bite (Bush and Vazquez-Pertejo, 2018). Furthermore, BBK32 can also bind to glucosaminoglycans (GAGs), which are similar to fibronectins and present on the host cell surface, facilitating the attachment of BBK32 to host cells (Li *et al.*, 2006). Structurally, BBK32 comprises an N-terminal region that is responsible for the host cell adhesion and a C-terminal globular domain that functions as an immune evasion domain (Figure 1.15). The N-terminal domain of BBK32 and its interaction with GAGs and fibronectin is well characterised. More recently, BBK32 was reported to interact and block the activation of the CP (Xie *et al.*, 2019). In this study, the C-terminal of BBK32 was reported to block the activity of the zymogen C1r protease, thereby preventing activation of the CP. The nature of the interaction between BBK32 and C1r is still unknown.

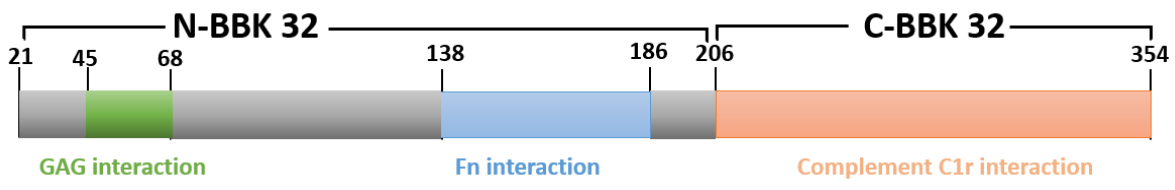


***Figure 1-15**: Schematic illustration of the BBK32 structure and the function of each domain.*

1.3 Leukocyte-associated immunoglobulin-like receptor (LAIR) proteins:

Regulation of the immune system is critical to prevent an excessive inflammatory response and autoimmunity. One form of regulation of immune cells is achieved by the inhibitory signal provided by receptors with cytoplasmic tails containing an immunoreceptor tyrosine-based inhibitory motif (ITIM) (Olde Nordkamp, Koeleman and Meyaard, 2014). Ligation of these receptors with their ligands results in the phosphorylation of the tyrosine residues within the ITIM cytoplasmic tail, leading to downregulation of immune cells activities. The availability of ligands during an immune response plays an important role in determining the type of immune regulatory cells that are recruited to the site of infection. LAIR proteins are immune regulation receptors and members of the Ig (immunoglobulin) superfamily. Expressed by monocytes and thymocytes, LAIR proteins are encoded by a gene located on human chromosome 19q13.4 of humans. Two LAIR proteins, LAIR-1 and LAIR-2 were discovered. LAIR-1 (CD305) is a transmembrane inhibitory receptor containing a C2-type Ig-like extracellular domain and two ITIM-baring domains in the cytoplasmic tail (Nordkamp *et al.*, 2014). LAIR-1 is a 32 kDa glycoprotein that is capable of inhibiting the activities of natural killer cells, resulting in the inhibition of target cell lysis. LAIR-2 (CD306), shares 84% sequence identity with LAIR-1. It is a 17 kDa secreted glycoprotein with a single Ig-like domain, similar to LAIR-1, but lacks both transmembrane and cytoplasmic domains (Figure 1.16). LAIR-2 functions as a competitor for LAIR-1 and plays an important role in regulating LAIR-1 induced downregulation of the immune response. The full function of the soluble LAIR-2 protein, however, is poorly understood.

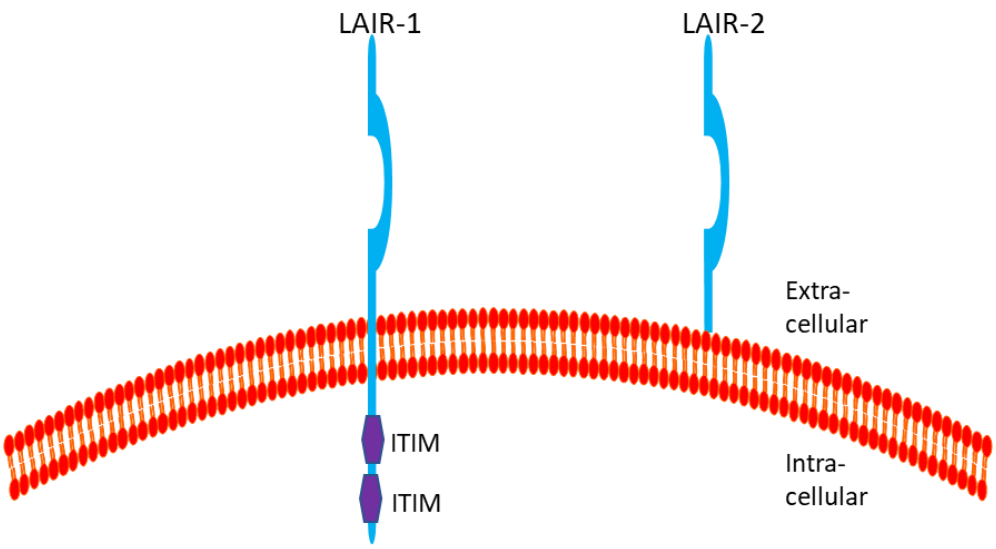


***Figure 1-16**: Schematic representation of the LAIR-1 and LAIR-2 structure.*

1.3.1 Interaction between LAIR proteins and the recognition molecules of the CP and LP of complement:

The recruitment of the inhibitory receptors to immune response sites is dependent on ligand availability. Specific ligands attract specific inhibitory receptors. It has been reported that LAIR proteins are receptors for extracellular matrix components and for transmembrane collagen. They bind specifically to collagen repeats (Gly-X-Y), where X is proline and Y is hydroxyproline (Lebbink *et al.*, 2006). Collagens are the most abundant proteins in the human body and play an important role in the growth and development of many tissues. Additionally, collagens function as a substrate for cell attachment, migration, coagulation and mediate several biological activities, such as binding to several cell surface receptors, including LAIR-1 and LAIR-2. Recognition molecules of the CP and LP that contain collagen-like domains are also possible targets for the inhibitory LAIR proteins, these interactions have been investigated in this thesis.

1.4 Hypothesis and general aim of the thesis:

The activation of the CP of complement is mediated by its C1 complex. This complex is formed of one recognition molecule C1q and four associated serine proteases assembled at core of C1q as heterotetramer. Several structural models have been proposed to explain the activation process based on the available structural data. One important part of the activation process of C1 is the flexibility of C1q during activation and the conformational changes to the C1r-C1s heterodimers. Therefore, the first aim of this thesis is to characterise the molecular organisation of the C1 complex of the CP of complement and assess how the subcomponents interact with each other. In addition, the recognition subcomponent of the CP and LP are susceptible for immune regulators such as LAIR-2 and BBK32. These regulators affect the pathways activation by interacting with the C1g-C1rs2 complex of the CP and MBL-MASPs complex of the LP. Therefore, the second aim of this thesis is to characterise the interaction between the CP and LP activator proteins and the immune inhibitors LAIR-1 and BBK32. Chapters 2 and 3 will focus on the structural and biochemical characterisation of the C1 complex, while chapter 4 will focus on the immune regulatory protein LAIR-2 and its putative role in regulating the CP and LP of complement. Finally, chapter 5 will focus on the interaction between the *B. burgdorferi* surface protein BBK32 and the CP serine protease C1r.

Chapter 2 . Structure of the C1r-C1s complex of the CP

2.1 Introduction:

The CP of complement is a major component of the innate immune system and is initiated by a multimeric 790kDa complex named C1. This complex is formed of a recognition hexameric-bouquet-like protein called C1q (Kishore and Reid, 2000) , and four associated serine proteases 2C1r and 2C1s assembled at the core of C1q as a Ca^{2+}- dependent heterotetramer (Gál *et al.*, 2009). C1r and C1s are homologous modular proteases and are composed of an N-terminal CUB domain (Bork and Beckmann, 1993) followed by an EGF-like domain (Hambleton *et al.*, 2004) and then another CUB domain, followed by two CCP modules (Norman *et al.*, 1991), and a final serine protease domain (SP). This modular structure is shared by the LP serine proteases (MASPs), which are associated with the LP recognition proteins that activate the LP of complement (Fujita, 2002). In the absence of C1q, the heterotetramer $C1r_2s_2$ forms an antiparallel elongated S-shaped structure that folds up to form a more compact structure when it associates with C1q (Venkatraman Girija *et al.*, 2013). Several models have been proposed to explain the C1 complex assembly, differing in the way component are associated together. The first model was proposed in 1985 (Villiers, Arlaud and Colomb, 1985), followed by several models based on the structural data availability. In the most well-established model, the two C1r proteases dimerizes at the centre of the C1 complex through the interaction between the CCP1 modules of one C1r and the SP domain of its partner. The CUB-EGF fragment of each C1r polypeptide forms a heterodimer interaction with the CUB-EGF junction of each C1s proteases, forming a heterotetramer with the two C1r polypeptides packed at the centre (Almitairi *et al.*, 2018). This assembly keeps the SP domains of each of the C1r proteases apart by a distance of 90 Å (Budayova-Spano *et al.*, 2002) thereby preventing unwanted activation. When C1q binds to an activator, it splays and disrupt the CCP1-SP interaction between the C1r dimers leading to autoactivation and cleavage of its substrate C1s (Wallis *et al.*, 2010).

This model was based on the crystal structure of the C1r-SP domain (PDB code: 1GPZ) (Budayova-Spano *et al.*, 2002), and the structure of the CUB-EGF-CUB domains of C1r and C1s (PDB code: 6F1C, 6F1H and 6F39) (Almitairi *et al.*, 2018). In the model, C1s-C1s packs inside C1q via interaction between the CUB1 and CUB2 of each C1r proteases and the CUB1 of C1s, providing a total of six binding sites for each C1q collagen stems (Bally *et al.*, 2009). Calcium plays a central role in mediating the assembly between the CUB domains with C1q and between the CUB-EGF junction of each proteases heterodimers (Venkatraman Girija *et al.*, 2013). More recently, an alternative model was proposed in which the CUB-EGF-CUB domains of C1r and C1s stack to form an anti-parallel, ring-shaped heterotetramer with the catalytic domains projecting from opposite sides (Mortensen *et al.*, 2017). However, this model does not explain the well characterised interactions between the CCP1-CCP2-SP domains of C1r that have been observed both in solution and in crystal structures (Villiers, Arlaud and Colomb, 1985; Budayova-Spano *et al.*, 2002) (Figure 2.1). In the absence of a high-resolution structural data for the C1r-C1s interaction, it is not possible to discriminate between these two models. Therefore, the aim of this chapter is to produce and crystallise the CUB-EGF-CUB domains of the C1r-C1s proteases (subsequently called 3DC1r and 3DC1s) and assess the complex formation both in crystal and in solution.

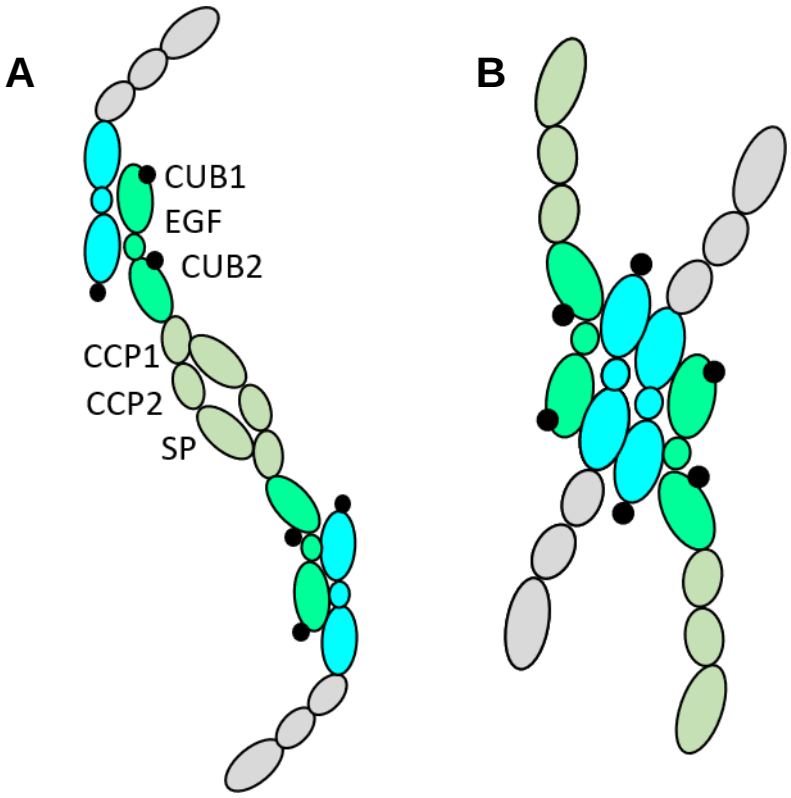


Figure 2-1: *(A) Schematic illustration of the traditional C1r-C1s elongated S-shaped model. (B) The recently proposed stacked model. C1r is presented in green, C1s is in blue and grey shows the CCP1-CCP2-SP domains of C1r. The black dots represent the C1q binding sites.*

2.2 Materials and methods:

2.2.1 Materials:

Cell lines expressing CUB1-EGF-CUB2 domains of both C1r and C1s were kindly provided by (Prof. Russell Wallis, University of Leicester). DNA and protein markers were supplied by New England Biolabs. Filters (0.2µm), Ni^{2+}- NTA affinity resin, Superdex 200 and 75 16/60 columns used for protein purification were supplied by GE Healthcare. All chemicals were purchased from Sigma and Thermo Fisher Scientific. Crystallisation screens JCSG, ProPlex, PACT, Morpheus and BCS were supplied by Molecular Dimensions. 96 well crystallisation screen plates were purchased from TTP Labtech, and the MRC Maxi 48 well plates and additives screen were purchased from Hampton Research. CryoCaps and Cryovials, as well as LithoLoops mounted on pins were all from Molecular Dimensions.

2.2.2 Production and purification of CUB1-EGF-CUB2 domains of C1r and C1s:

Chinse hamster ovary (CHO) cell lines expressing the CUB1-EGF-CUB2 domains of C1r and C1s later referred to as 3D C1r and C1s, was kindly provided by (Prof. Russell Wallis, University of Leicester). The cDNA of 3D C1r and C1s cDNA was cloned into expression vector pED4 for mammalian expression with the C-terminal His_6 tag and expressed as described previously (Phillips *et al.*, 2009). Cells were grown in Minimal Essential Media α without nucleotides (MEMα-) provided by molecular dimensions, supplemented with 10% dialysed, heat-treated foetal calf serum (DHFCS), 50 units/ml penicillin and 50 µg/ml streptomycin and 0.5 µM methotrexate (MTX) in a three-layered tissue culture flask. Once fully confluent, the media was removed, and cells were washed with 50 ml PBS, and the media was replaced with CHO-S-SFM II media without nucleotides, supplemented with 50 mM HEPES pH7.55, 0.5 µM MTX and 50 units/ml penicillin, and 50µ g/ml streptomycin. The media was collected every other day and replaced with 100 ml of fresh CHO-S-SFM II. 3D

C1r and 3D C1s were purified by affinity chromatography on a Ni^{+2} nitrilotriacetic acid (NTA) - Sepharose column. 300ml of the harvested media was mixed with an equal amount of (20 mM Tris-pH 7.5, 50 mM NaCl and 2 mM $CaCl_2$). The media was then loaded onto 2 ml of the pre-equilibrated Ni^{2+}-Sepharose column using the same dilution buffer and left at 4ºC overnight. After this, the column was then washed with a buffer containing 20 mM imidazole to remove any non-specific binding. The protein was eluted from the column in 1 ml fractions with a buffer containing 0.5 M imidazole. Eluted fractions were checked on a non-reducing 15% SDS-PAGE gel to assess the expression quality. The eluted fractions containing the protein were concentrated to 5 ml and further purified on a Superdex 200 16/60 gel filtration column using the same buffer. C1r-C1s heterocomplex was generated by mixing molar equivalent of both C1r and C1s followed by gel filtration on a Superdex 200 10/300 in the same buffer. The gel filtration fractions corresponding to the correct elution peak were collected and concentrated to ~4 mg/ml for crystallisation trials.

2.2.3 Crystallisation of the C1r-CUB2 and 3D C1r-C1s complex:

2.2.3.1 Crystallisation theory:

Crystallisation is a very powerful tool for understanding biological functions at a molecular level. A detailed atomic map of macromolecules can provide a deep insight into a protein's function. Protein crystallisation was first discovered in the 19th century as a means of purification (McPherson, 1991). The first step to obtaining protein crystals is to purify the target proteins as close as possible to homogeneity, which is a key factor for growing crystals that diffract to high resolution (Dessau and Modis, 2010). Protein crystallisation consists of two main steps. The first is to bring the sample to a supersaturation state generating the highest possible concentration without causing aggregation or precipitation. Supersaturation is a non-equilibrium condition by which protein concentration is increased above solubility limits

(McPherson, 2014). The second step is to introduce the concentrated sample to a suitable precipitating agent that can promote nucleation of protein crystals in solution and generate a three-dimensional crystal suitable for X-ray diffraction. The most commonly successful precipitants are polyethylene glycol (PEG) and ammonium sulphate, which together account for 60% of all recorded macromolecular precipitants used in crystallisation trials (Gilliland *et al.*, 1994; Gilliland, Tung and Ladner, 2002). Crystallisation is characterised by two main inseparable features, starting with nucleation and followed by crystal growth. Crystals are formed by a regular arrangement of building blocks (in this case of a protein) in three dimensions. Crystallisation is characterised by three main phases (Watts, 1993). The first phase is nucleation, which is the association of a certain number of molecules in an arranged order forming a three-dimensional shape. The second phase is crystal growth, where the nuclei formed in the first phase grows through the diffusion of molecules from the solution to its surface. The last phase is when crystal growth stops due to a decreased number of molecules (building blocks) in the solution (Russo Krauss *et al.*, 2017)(Figure 2.2).

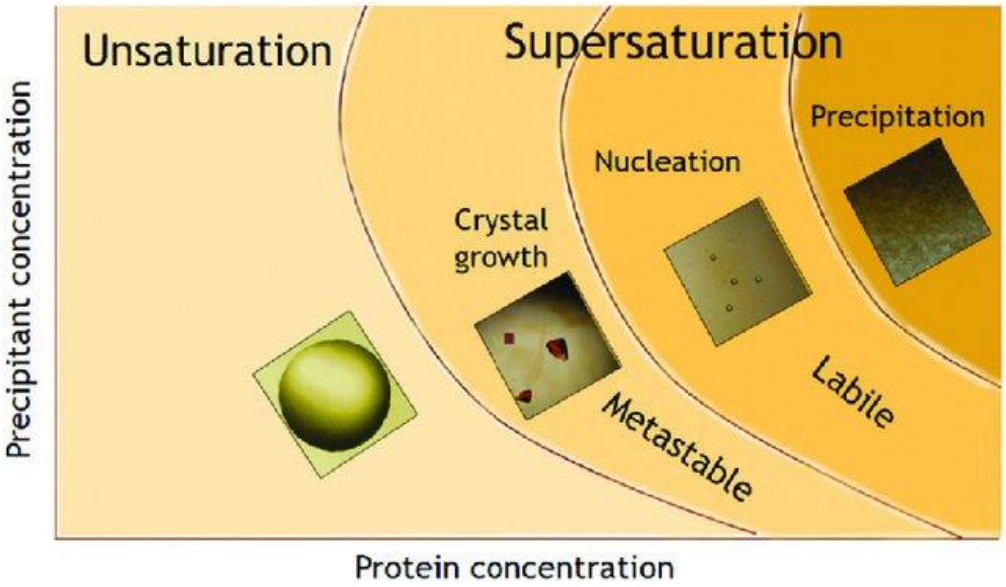


Figure 2-2: *Phase diagram illustrating the process of crystal growth. At low protein/precipitant concentrations, unsaturation condition occurs, and the sample drop remains clear. At very high concentrations of protein/precipitant, precipitation occurs, which is a disordered arrangement of molecules. Between these two phases, in the middle of the diagram is the liable zone. This is where nucleation takes place, followed by crystal growth. Crystal growth can occur at lower concentrations than nucleation, at the right protein/precipitant concentrations* (Russo Krauss *et al.*, 2017).

The search for the suitable crystallisation conditions of a protein is carried out in two steps, starting with the use of commercially available screens that can test a wide range of crystallisation conditions. The second step involves varying the precipitant concentration, pH and other components of the original hit condition. Several commonly used techniques were developed to set up crystallisation trials using 96 well crystallisation plates and changing the way the sample/precipitant were set (Figure 2.3).

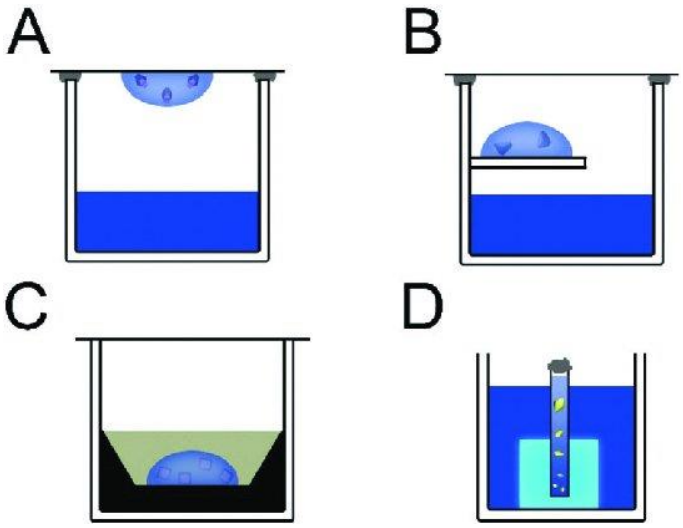


Figure 2-3: *Schematic representation of crystallisation techniques. (A) Vapor diffusion hanging drop, (B) Vapor diffusion sitting drop, (C) Micropatch and (D) Free interface diffusion* (Russo Krauss *et al.*, 2017).

2.2.3.2 X-ray diffraction:

X-ray imaging is a powerful tool for providing a detailed atomic map of proteins structures. Max von Laue *et al.* discovered in 1912 that crystalline substances act as three-dimensional diffraction gratings for X-ray wavelengths similar to the spacing of planes in a crystal lattice (Bunaciu, Udriştioiu and Aboul-Enein, 2015). When a single crystal is exposed to a beam of monochromatic X-rays, the electron density of any atom diffracts the X-ray radiation in all directions. Because of the lattice repetition, the radiation diffracted by one atom will interfere with that diffracted by all other atoms in the crystal, which resulting in a diffraction pattern (Guss, 2011) (Figure 2.4).

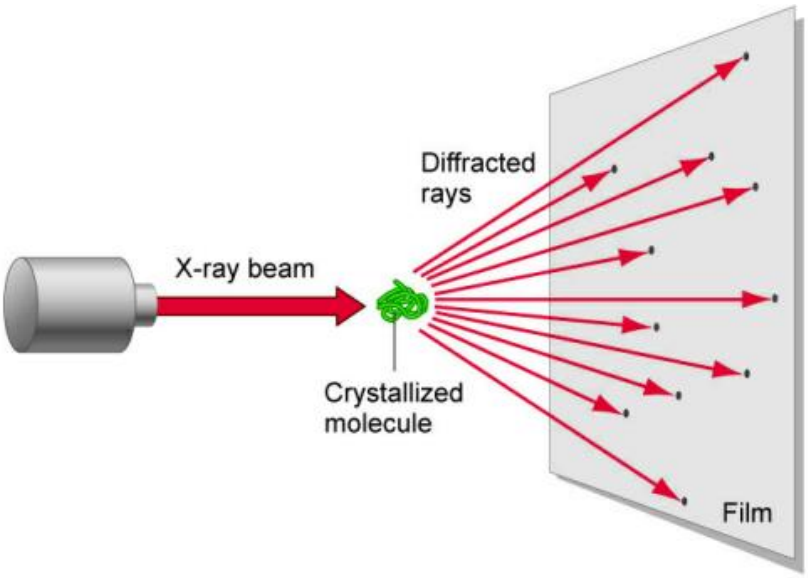


Figure 2-4: *Single crystal X-ray diffraction. When an X-ray beams hits a crystal, it is diffracted in different directions based on the arrangement of atoms in the structure. At specific angles, constructive wave interferences will produce spots on the detector.*

X-ray scattering of a single crystal can only be detected when a large number of molecules are arranged in a well-defined order within the crystal. X-rays are a form of electromagnetic radiation, with a wavelength (λ) ranging between 10 and 0.1 Å (1 Å = 0.1 nm) (Waseda, Matsubara and Shinoda, 2011). The diffraction of X-ray waves from crystal molecules can be calculated by Bragg's law; $\mathbf{n\lambda = \mathit{d}2\sin\theta}$, where (n) is an integer, (λ) is the wavelength of the X-ray, (*d*) is the spacing between the planes in the atomic lattice and **(θ)** represents the angle between the incident and scattering rays of the crystal planes (Figure 2.5).

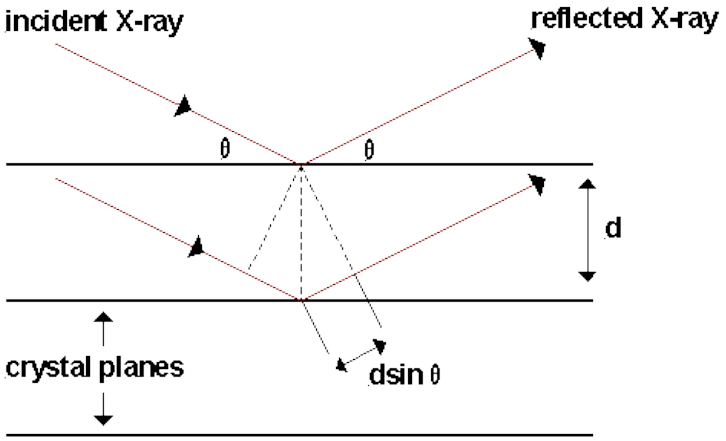


Figure 2-5: *Schematic representation of Bragg's law from two crystal planes. When an incident X-ray beam strikes one of the crystal planes at the angle (θ), the beam is reflected from an atom of both planes (upper and lower planes). The reflected beams from the upper and lower planes travel differently by a distance of d2sinθ. Both reflected beams combined to make a constructive interface when the path difference is equal to some integral multiple of the wavelength λ.*

2.2.3.3 Crystallisation of 3D-C1r-C1s complex:

All crystals were grown using the sitting-drop vapour diffusion method. Conditions were tested using a range of the precipitant PEG8K with buffers at 25 ºC and 4 ºC. Drops were set by mixing equal volumes (1.2 + 1.2 μl) of the protein and reservoir solution. Plates were checked regularly, and conditions were optimised by altering the precipitant and buffer concentrations. Crystals of the C1r-C1s complex (3-4 mg/ml) were grown in 12-18% PEG 8000, 100 mM imidazole at pH 8.0 and at 25 ºC. Similar conditions were used to crystallize C1r alone, except that the crystals were grown at 4 ºC. All crystals were transferred to a reservoir solution containing 20% glucose before being stored in liquid nitrogen and were maintained at 100 kelvin (K) during data collection.

Diffraction data were collected at Diamond Light Source in beam line I04 and were processed with the iMosflm software. Phases were determined by molecular replacement with Phaser using the C1s CUB-EGF-CUB2 (Venkatraman Girija *et al.*, 2013) structure as a search model. Models were optimised using cycles of manual refinement with Coot and refinement in Refmac5 (Emsley *et al.*, 2010), which are part of the CCP4 software suite, and in Phenix (Afonine *et al.*, 2012). A structure of the CUB2 domain of C1r (at 1.95 Å resolution) was determined independently and used as a reference for the C1r-C1s structures during refinement. This part of the project was carried out by Prof Wallis.

2.3 Results:

2.3.1 Expression and purification of the 3D C1r and 3D C1s proteins:

The pED4 expression plasmids containing the cDNAs encoding for 3D C1r and C1s were transfected into CHO cells, and the cells were grown and harvested as described in the methods section. The harvested media was then loaded onto a 2 ml Ni^{2+}- NTA Sepharose column as described in section 2.2.2. Following this step, the eluted fractions were loaded on to a non-reducing 15% SDS-PAGE (Figure 2.6). The gel image below shows the analysis for 3D C1s. The 3D C1r was purified in the same way.

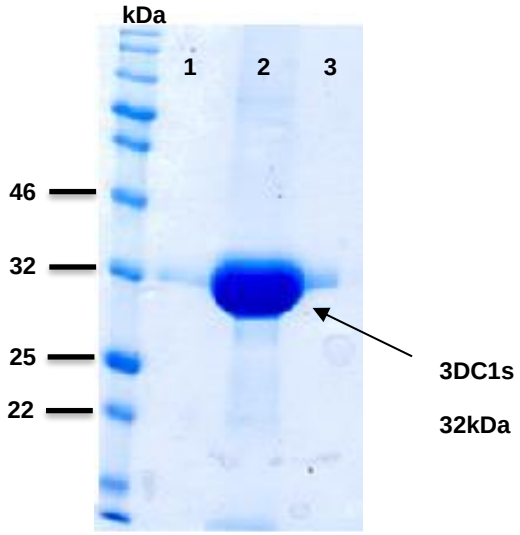


Figure 2-6: *Non-reducing 15% SDS-PAGE analysis of the Ni^{2+} elution fractions of the 3D C1s. The gel indicates the presence of the protein in the eluted fractions 1, 2 and 3with a molecular mass of 32 kDa.*

2.3.2 Gel filtration and purification:

The elution fractions from the nickel column were then ran on a Superdex 75 16/60 size exclusion chromatography column to further purify the proteins as shown in (Figure 2.7).

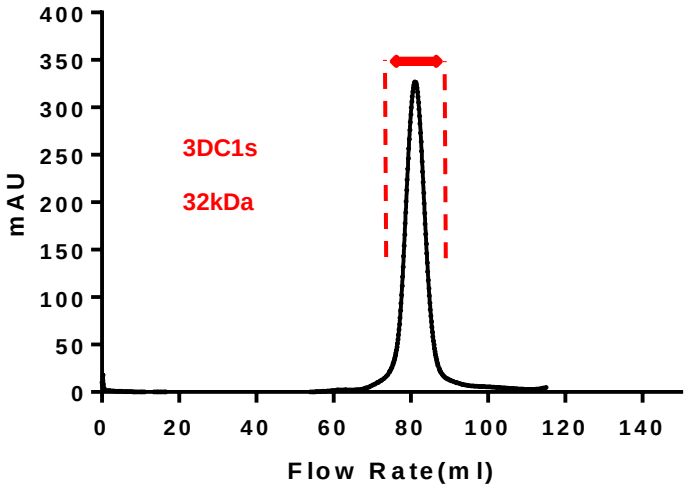


Figure 2-7: *Purification and elution profile of the 3DC1s protein. The fractions corresponding to the correct peak between the red dotted lines were collected and checked on a 15% non-reducing SDS-PAGE gel. A similar profile was observed for 3DC1r*

Fractions from the gel filtration step were analysed on a non-reducing 15% SDS-PAGE gel to check the purity of the protein (Figure 2.8).

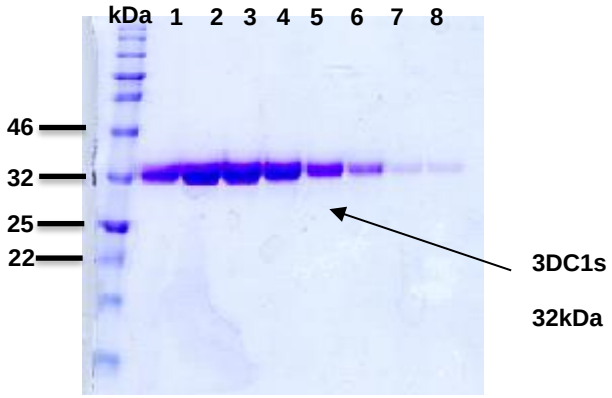


Figure 2-8: *15% non-reducing SDS-PAGE analysis showing the eluted fractions from the gel filtration column (fraction 1 to 8). Fractions containing the 3DC1s (32 kDa) protein were pooled together.*

2.3.3 Generation of 3D C1r-C1s heterodimer complex by gel filtration:

The 3D C1r-C1s heterocomplex was generated by mixing 3D C1r and 3D C1s at a 1:1 molar ratio. They were analyzed by gel filtration in 20 mM Tris pH7.5, 50 mM NaCl and 2 mM $CaCl_2$ on a Superdex 200 10/300 column (Figure 2.9).

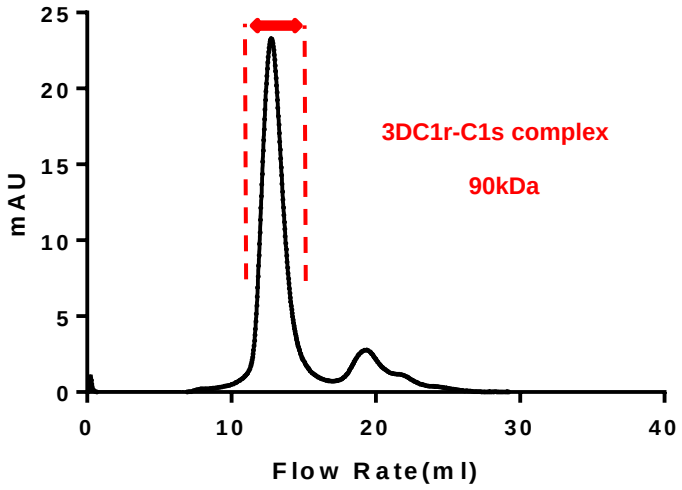


Figure 2-9: *Gel filtration elution profile of the C1r-C1s complex. Fractions corresponding to the heterocomplex were collected and concentrated to 4mg/ml for crystallisation.*

2.3.4 Interaction of the 3D C1r and C1s in solution:

Analytical gel filtration showed that 3D C1s is a Ca^{2+}-dependent dimer. 3D C1r was also dimeric in Ca^{2+} but aggregated in EDTA. The analysis of an equimolar mixture of C1r and C1s fragments revealed a new dimer peak that eluted slightly earlier than either the C1s or C1r homodimers, implying the formation of heterodimers (Figure 2.10). No peak corresponding to a tetrameric complex (expected to be ~150 kDa in size for the N-terminal domains) was observed when loading concentrations of up to 1 mg/mL (>10-fold the normal serum concentration of $C1r_2C1s_2$). The position of the heterodimer peak did not change over the concentration range examined, indicating a stable complex. Importantly, the gel filtration data

exclude the stacked tetramer model of C1 (Figure 2.10), as it shows that 3D C1r and 3D C1s are heterodimers, not heterotetramers as would be predicted.

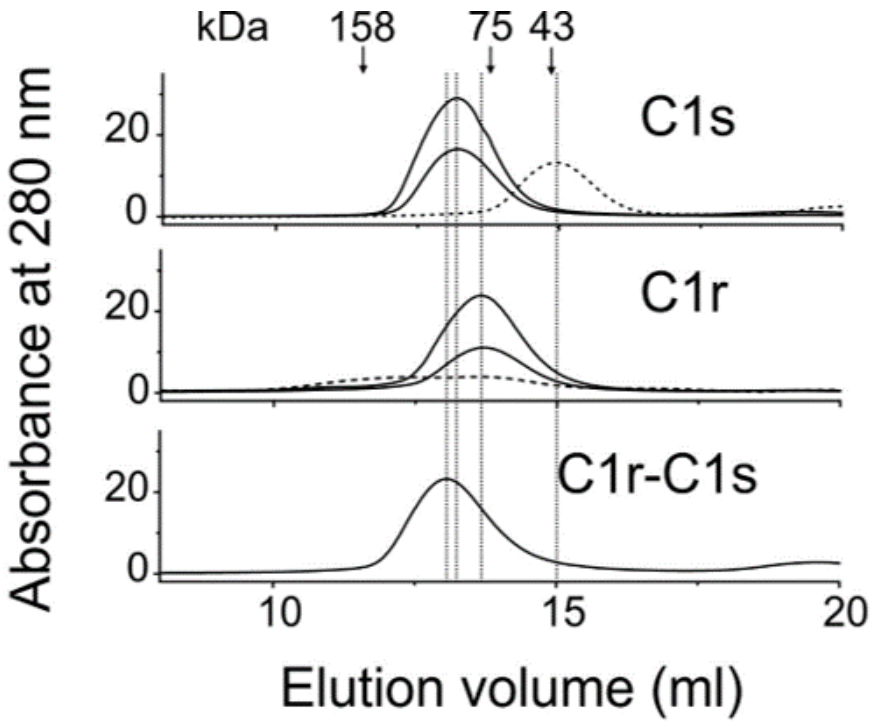


Figure 2-10: *Analytical gel filtration of 3D C1r and 3D C1s. 3D C1r and 3D C1s are homodimers in the presence of* Ca^{2+}*. A peak of the heterodimer complex C1r-C1s was observed when mixing equimolar concentration of C1r and C1s which eluted slightly earlier than the homodimers. Dotted lines show elution in a buffer containing EDTA.*

2.3.5 Structure of the CUB1-EGF-CUB2 domains of the heterodimer of C1r-C1s:

The crystal of the CUB1-EGF-CUB2 heterodimer of the C1r-C1s complex was obtained in an imidazole buffer at pH 8.0 with PEG 8000 as the precipitant at 25ºC. Data were collected at Diamond Light Source in beamline I04. The diffracted data was processed using the CCP4i software as described in section 2.2.3.3, and the structure was refined to 4.5Å (Figure 2.11).

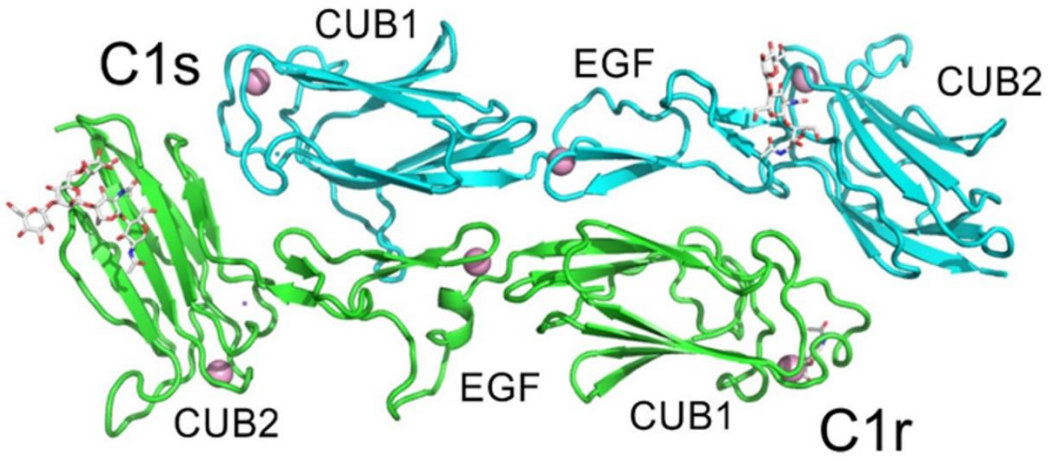


Figure 2-11: *Overall structure of the 3DC1r-C1s complex. 3D fragment of C1r (blue) and C1s (green) are both L-shaped, and they interact in an antiparallel arrangement mediated via the CUB-EGF junction.* Ca^{2+} *is essential for the interaction between the C1r and C1s.* Ca^{2+} *are indicated in pink spheres.*

Crystals were also obtained for 3D C1r homodimers under similar conditions but at 4 ºC. The structure was determined to 5.8 Å (Figure 2.12). C1r dimerises in an antiparallel arrangement with an L-shaped structure similar to the one found in the C1r-C1s heterodimers. The interface of the C1r-C1s heterocomplex is extensive and spans all three domains CUB1-EGF-CUB2 of each polypeptide (1,268 $Å^2$ for C1r and 1,304 $Å^2$ for C1s). The interface of the C1r-C1s heterodimer is more extensive than the interfaces observed between homodimers of C1s (Venkatraman Girija *et al.*, 2013), C1r (Figure 2.12) and the homologous serine proteases of the LP (Nan *et al.*, 2017). This explains why C1r and C1s preferentially form heterocomplexes rather than homocomplexes.

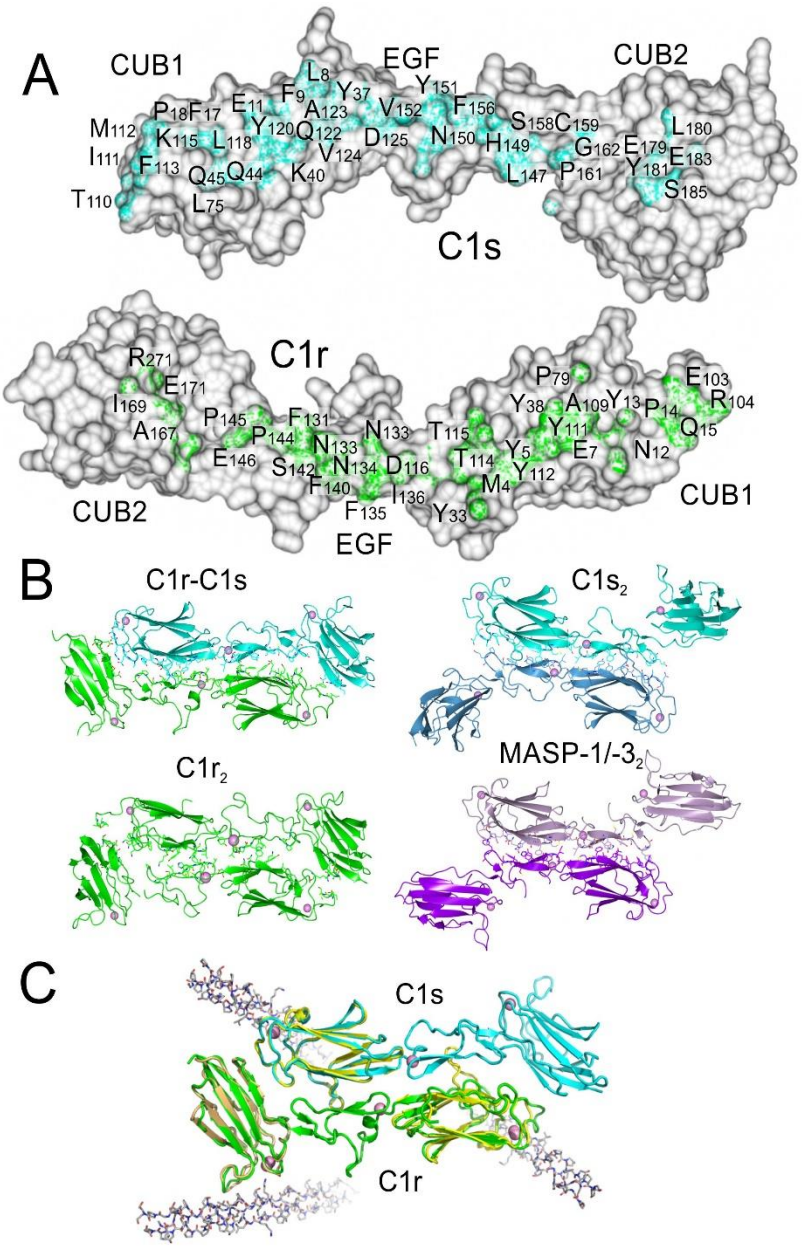


Figure 2-12: *C1r-C1s interface. (A) Residues buried at the interface of the C1r-C1s complex are indicated. (B) CUB1-EGF-CUB2 structures of C1r-C1s heterocomplex compared to C1s homodimer (PDB code: 4LMF), C1r homodimers and MASP-1/-3 homodimers (PDB code: 3DEM), the homologues serine proteases of the LP. (C) The C1r-C1s dimers overlay with CUB1-collagen (PDB: 4LOR, yellow) and CUB2-collagen structures (PDB: 3POB, wheat) showing the position of the collagen-binding sites at the CUB1 of C1s (blue) and CUB1, CUB2 of C1r (green). The collagen is in grey.*

	C1r-C1s 1 CUB1-EGF-CUB2	C1r-C1r CUB1-EGF-CUB2
PDB code	6F1C	6F39
Data collection		
Beam line	Diamond I04-1	Diamond I03
Space group	P 2_1 2_1 2_1	C 1 2 1
a, b, c, Å	96.8, 123.7, 195.5	140.5 54.6 137.6
α, β, γ, º	90, 90, 90	90 100.2 90
Resolution, Å	97.8 - 4.2 (4.70 - 4.20)	68.1 - 5.8 (6.49 - 5.80)
$R_{sym, \%}$	15.3 (93.3)	32.4 (195.2)
$R_{pim, \%}$	8.3 (53.3)	11.8 (69.2)
CC(1/2)	0.998 (0.814)	0.997 (0.610)
$I/\sigma I$	6.6 (1.9)	5.1 (1.8)
Completeness	99.8 (99.7)	99.8 (100)
Redundancy	7.6 (7.1)	15.2 (16.5)
Resolution, Å	86.73 - 4.20 (4.35 – 4.20)	68.06 - 5.80 (6.00 - 5.80)
No. reflections	17763 (1738)	2987 (293)
R_{work}/R_{free}	0.247/0.30.5	0.315/0.333
No. atoms	9260	4608
Protein	9003	4456
Ligand/ion	257	166
Water	-	-
B-factors, Å^2	226.8	354.8
Protein	225.0	353.3
Ligand/ion	286.4	395.3
Water	-	-
Bond lengths, Å	0.004	0.004
Bond angles, º	0.91	0.98

Table 2.1: *Data collection and refinement statistics. The highest resolution shell is shown in parentheses.*

2.4 Discussion:

The structure of the C1r–C1s interaction provides insight into the assembly of the C1 complex. C1r and C1s polypeptides bind via an extensive interface involving all three N-terminal domains (CUB1-EGF-CUB2). A Ca^{2+}-binding site in the EGF domain of each subcomponent forms part of the binding interface and explains the Ca^{2+}-dependence of the interaction. Additional Ca^{2+} sites are present in each CUB domain of C1r and in the CUB1 of C1s forming the binding sites for the collagen-like domains of C1q. The data reported here are incompatible with the recently proposed stacked-tetramer model for $C1r_2C1s_2$ via their (CUB1-EGF-CUB2) domains (Mortensen *et al.*, 2017), as the CUB1-EGF-CUB2 fragments of C1r and C1s do not form tetramers, even at higher concentrations than found in serum. Instead, each C1r-C1s heterodimer must be held together by contacts between the CCP-SP domains of C1r to form the elongated S-shaped structures first observed in EM images (Strang *et al.*, 1982). This arrangement would prevent the autoactivation of C1r, because of the SP domain of one C1r polypeptide packs against the CCP domains of its partner (Budayova-Spano *et al.*, 2002). Autoactivation of the complex would require disruption of the C1r–C1r interaction at the CCP-SP junction, followed by alignment of the catalytic site of one polypeptide with the cleavage site of the other. This intramolecular autoactivation mechanism is compatible with the first-order kinetics observed for spontaneous C1 activation (Ziccardi, 1982). The recent cryoEM structure of C1 bound to C4 (Sharp *et al.*, 2019), shows that our original model of the heterotetramer in C1 is incorrect. In the original model, C1s was placed at the centre of the tetramer with C1r on the outside creating six collagen binding sites (two on CUB1 and CUB2 of each C1r and one on CUB1 of C1s). However, recent data shows that C1r is at the centre with C1s is on the outside. Interestingly, there are two copies of the C1r-C1s dimer in our structure that probably mimic the $C1r_2C1s_2$ tetramer when bound to C1q. In this arrangement, there are only four binding sites for the collagen stems, as the binding site on each CUB2

domain of C1r is masked by the CUB1 domain of its partner, the C1s polypeptide. When bound to C4 on the cell surface, there is a change in the relative position of the C1r-C1s dimer so that the collagen bound to CUB1 of C1s also interacts with the CUB2 of the adjacent C1r polypeptide (Figure 2.13). This change may accompany C1 formation, or even the activation of C1 (Sharp *et al.*, 2019).

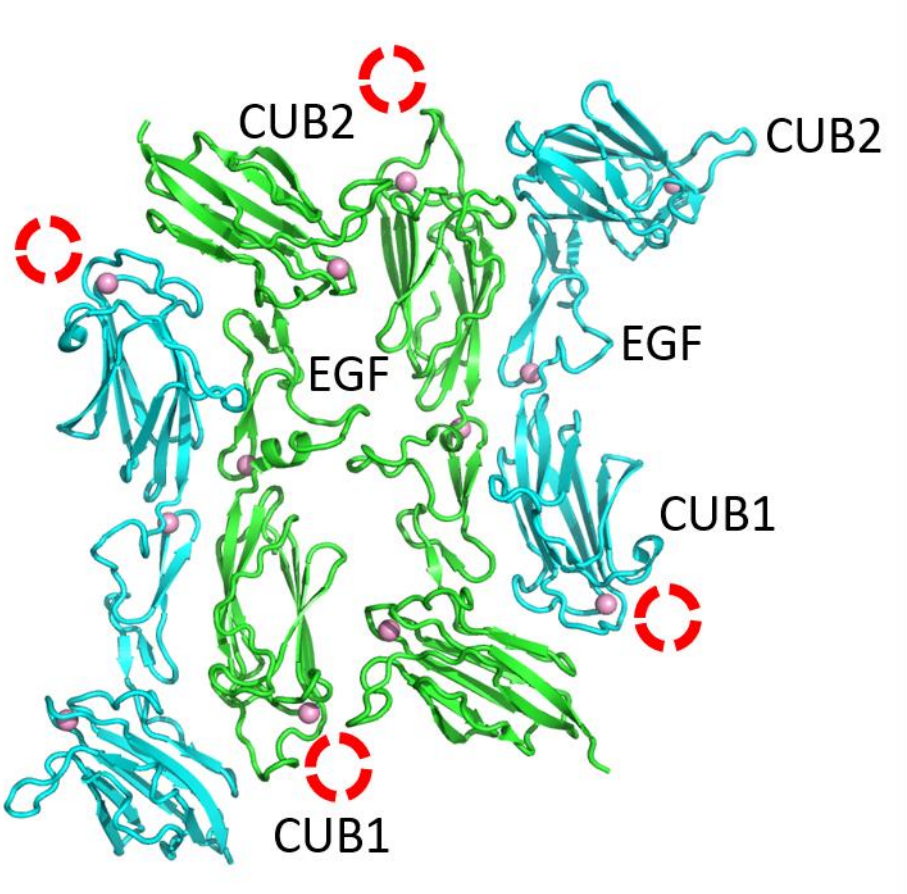


Figure 2-13: *Structure of the C1r-C1s asymmetric unit. There are two dimers in the asymmetric unit which probably reflect the arrangement of the two C1r (green) and two C1s (cyan) molecules in the C1 complex. The binding sites for collagen are shown by red circles. The collagen-binding site in each CUB2 of C1r is blocked by the CUB1 domain of its partner making a total of four binding sites. Ca^{2+} molecules are shown as pink spheres.*

Chapter 3 . Structure of the C1qA, B and C chains in complex with the 3 domains CUB1-EGF-CUB2 of C1s

3.1 Introduction:

The CP of the complement system is activated by a (460 kDa) recognition subcomponent named C1q, which is associated with four serine proteases, two C1r and two C1s proteins. These are arranged as an elongated S-shaped heterotetramer with the two C1r proteases at the center, stabilized through reciprocal contacts between the CCP1-CCP2-SP domains of each partner (Gál *et al.*, 2009), and a C1s protease at each end, mediated through contacts between the CUB1-EGF-CUB2 domains of C1r and C1s (Almitairi *et al.*, 2018). C1q has a bouquet-like structure assembled from six heterotrimeric collagenous stems. It is assembled from three different chains; the A chain (223 residues), B chain (226 residues) and C chain (217 residues). Each chain contains an (~81 residue) collagen-like region at the C-terminus, which contains the binding sites for C1r and C1s (Phillips *et al.*, 2009). In this region, all three collagen chains of C1q contain a sequence similar to the protease-consensus motif found in MBL, serum ficolins and CL-11: Hyp-Gly-Lys-Xaa-Gly-Pro (where Hyp is a hydroxyproline residue, and Xaa is typically a hydrophobic residue). This motif is also present in the lectin pathway recognition molecules and mediates the binding to their associated serine proteases MASPs (Phillips *et al.*, 2009). More recently, it has been shown that the same sequences form the binding site for C1r and C1s in the classical pathway activation complex (Bally *et al.*, 2013). Most notably, in all interactions, the lysine residue at the center of the motif is essential for binding. Unlike C1q, the LP recognition molecules are formed from homotrimeric subunits (Matsushita, Endo and Fujita, 2013). Presumably, the three different chains of C1q mediate its specificity towards its associated proteases $C1r_2s_2$ rather than MASPs (Venkatraman Girija *et al.*, 2013), at least in part, although components have been shown to cross-bind *invitro* (Phillips *et al.*, 2009). Recent work has revealed the binding sites for C1q in the CUB1 and CUB2

domains of C1r and CUB1 of C1s providing a total of six binding sites on the C1r-C1s tetramer one for each collagen stem of C1q (Bally *et al.*, 2009). Ca^{2+} plays a central role in the assembly and is necessary for the binding of the proteases to C1q (Colomb *et al.*, 1984).

In previous studies, the structure of the CUB1-EGF-CUB2 domains of C1s bound to the consensus motif: Hyp-Gly-Lys-Leu-Gly-Pro was determined (PDB code: 4LOR) (Venkatraman Girija *et al.*, 2013). This work derived from the observed cross-binding between C1s and MBL (Phillips *et al.*, 2009). However, from the structure, it is not possible to determine the order of the chains in C1q or why C1s binds preferentially to C1q rather than MBL. The three polypeptides of a collagen triple helix assemble with a ~1 residue staggered between them, giving rise to leading, middle and lagging strands. For this reason, there are six possible configurations of the 3 chains in C1q: ABC; ACB, BCA, BAC, CAB, CBA. Each arrangement would have a different structure due to the stagger in the collagen chains. Because C1s co-crystallises with collagen and the three chains have similar but distinct sequences within the binding region, there is the opportunity to further probe binding between C1q and C1s and to determine the likely order of the chains in C1q. The aim of this part of the project is to assess the interactions between the CUB1-EGF-CUB2 domains of C1s and collagen peptides containing the binding motif of the A, B and C chains thereby generating a better model for C1q-C1s interaction.

3.2 Materials and methods:

3.2.1 Materials:

Media from the cell line expressing the CUB1-EGF-CUB2 domains of C1s (3D-C1s) was harvested and purified as described previously in section 2.2.2. Peptides corresponding to the protease-binding site on C1q were synthesized for each of the three chains A, B and C, and were kindly provided by Prof. Wallis. Sequences are:

C1qA: Ac-GPO GPO GPO GNO GKV GYO GPO GPO GPO-NH2
C1qB: Ac-GPO GPO GPO GNO GKV GPO GPO GPO GPO-NH2
C1qC: Ac-GPO GPO GPO GHO GKN GPM GPO GPO GPO-NH2

Specific protease binding residues in each chain are underlined. O represents hydroxyproline. Three GPO repeats were added at the N- and C-termini to increase the stability of the collagen-like structures. Crystallisation screens were purchased from Molecular Dimensions. All chemicals were purchased from Sigma and Thermo Fisher Scientifics. Crystallisation plates and materials were from TTP Labtech and Hampton Research. Diffraction and data collection were performed at the Diamond Light Source synchrotron facility.

3.2.2 Crystallisation of 3DC1s in complex with C1q chains A, B and C:

Crystallization conditions were based around those previously shown to be successful for the crystallisation of C1s bound to a collagen-like peptide from MBL: 100 mM Trizma-HCl pH 8.5 or 8.0 containing PEG8K and 80 mM NaBr, as shown previously (Venkatraman Girija *et al.*, 2013). The C1s 3D protein was first concentrated to 3.6 mg/ml and then mixed with the C1q peptides A, B and C individually and with mixtures (A+B, A+C and C+B) prior to crystallisation trials using a molar excess of peptides to C1s, as detailed in (Table 3.1). The precipitant PEG8K concentration ranged from 28% to 18%. Drops were tested using sitting drop vapor diffusion method in 48 well MRC crystallisation plates. 200 µl of buffer was added

to each well, and then 1.2 µl of the protein-peptide mixture was mixed with 1.2 µl of buffer at 25 ºC (room temperature). After crystallisation was observed, the most promising condition was then optimized further by testing serial dilution of NaBr, precipitant and Trizma pH 8.5, and 8.0, as detailed in (Table3.2).

3DC1s purified protein	C1q chains A, B and C
12.5µl 3DC1s	2.5µl C1qA
12.5µl 3DC1s	2.5µl C1qB
12.5µl 3DC1s	2.5µl C1qC
12.5µl 3DC1s	1.75µl C1qA
	0.75µl C1qC
12.5µl 3DC1s	1.75µl C1qA
	0.75µl C1qB
12.5µl 3DC1s	1.75µl C1qB
	0.75µl C1qC

Table 3.1: *This table summarizes the different 3DC1s-C1q complexes used for crystallisation. Each 3DC1s protein was mixed with a deferent ratio of C1q chains A, B and C as presented in the table. 3DC1s was at 3.66mg/ml, and peptides were at C1qA- 1.2mg/ml, C1qB- 1.32mg/ml and C1qC- 1.39mg/ml.*

Buffer composition for each condition used in crystallisation	
1	0.1M Trizma-HCl pH 8.5, 23%, 22%, 21%, 21%, 19%, 18% w/v PEG8K, 80Mm NaBr.
2	0.1M Trizma-HCl pH 8.0, 23%, 22%, 21%, 20%, 19%, 18% w/v PEG8K, 80Mm NaBr.
3	0.1M Trizma-HCl pH 8.0, 19% w/v PEG8K, between 500 and 50 Mm NaBr
4	0.1M Trizma-HCl pH 8.0, 18% w/v PEG8K, range between 500 and 50 mM NaBr
5	Between 50 and 500 mM of Trizma-HCl pH 8.0, 19% w/v PEG8K, 200 mM NaBr.
6	Range between 50 and 500 mM of Trizma-HCl pH 8.0, 18% w/v PEG8K, 200Mm NaBr.

Table 3.2: *Optimisation steps for 3DC1s protein in complex with C1qA, B and C crystallisation. Each buffer was used to set up a 48 well plate. The same optimization process was applied to each protein-peptide complex.*

All crystals were obtained at 19% or 18% PEG8K, containing 200 mM NaBr, 100 mM Trizma-HCl at pH 8.0.

3.2.3 Picking and storage of crystals:

Crystals were picked from the sitting drops by scooping into appropriately sized LithoLoops after adding 30% v/v glycerol as a cryoprotectant. Crystals were immediately submerged into liquid nitrogen to fix and preserve the crystals and were then swiftly moved into a cryogenic storage unit.

3.2.4 Data collection and processing:

Complete data sets of various crystals of the C1q-C1s complex were collected in beamline I04 at Diamond Light Source synchrotron. Crystals were robotically mounted in a cryo stream to keep the immediate environment around the crystal at 100 K. Each crystal was then centred into the path of the x-ray beam, with manual adjustments made where necessary. Initially, diffraction images from a crystal were collected 45° apart and were loaded into the EDNA (Incardona et al, 2009) program, which indexes the data and returns a set of optimal parameters for collecting a full dataset. Datasets were then collected based on this information. Data collection were carried out with Prof. Wallis.

3.2.5 Data processing and structure refinement:

The collected diffraction data sets were analyzed with the CCP4i software (Potterton *et al.*, 2003). Images were indexed and integrated using iMosflm (Kabsch, 2010). The data were then scaled and merged using Aimless, followed by molecular replacement using Phaser MR all part of the software suite. The structure of C1q-C1s (PDB:4LOR) (Venkatraman Girija *et al.*, 2013) was used as a search model. Rounds of structural refinement were carried out manually in Coot (Emsley *et al.*, 2010), followed by multiple rounds of refinement in both Refmac5 (CCP4) and Phenix.refine. Part of the Phenix program suite (Afonine *et al.*, 2012). Structures were analysed with the help of Prof. Wallis.

3.3 Results:

3.3.1 Structural analysis of the interaction between 3DC1s and C1q chains:

C1q is assembled from six subunits of 3 chains called A, B and C. The binding site for C1r and C1s is located within the Hyp-Gly-Lys-Xaa-Gly-Pro motif within the collagen-like domain. In this region, all three collagen chains contain a sequence similar to the protease consensus motif but with minor differences (Figure 3.1).

Peptides corresponding to the C1qA, C1qB and C1qC chains, flanked by $(GPO)_3$ repeats (to stabilize the collagen) were synthesized. These were mixed with purified 3DC1s (~3.66 mg/ml) both individually and as mixtures at different molar ratios. Crystals were collected, and complete data sets were collected at Diamond Light Source facility in beamline I04. Phases were solved by molecular replacement using the structure of 3DC1s bound to a collagen-like peptide containing the binding motif of MBL: Hyp-Gly-Lys-Leu-Gly-Pro as a search model (PDB code:4LOR) (Venkatraman Girija *et al.*, 2013).

Protease-binding motif:

C1qA: Ac-GPO GPO GPO GNO GKV GYO GPO GPO GPO-NH2

C1qB: Ac-GPO GPO GPO GNO GKV GPO GPO GPO GPO-NH2

C1qC: Ac-GPO GPO GPO GHO GKN GPM GPO GPO GPO-NH2

Figure 3-1: *Sequence of the peptides used for crystallisation with 3DC1s. Flanking GPO (where O is hydroxyproline) to stabilize the collagen allowing self-assembly of the collagen triple helix. Residues underlined shows the binding site in each chain of C1q.*

By making mixtures of peptides, the hope was to crystallise C1s in complex with collagen heterooligomers. The stagger in the collagen chains of one residue gives rise to a leading, middle and lagging strand. Mixtures of collagen peptides will potentially form multiple collagen helices. For example, mixtures of two 2 different peptides can potentially from 2^3 different collagen helices (AAA, ABA, AAB, BAA, BBA, ABB, BAB and BBB), whereas 3 different peptides, can form 3^3 i.e. 27 different helices. Our hypothesis is that the correct arrangement of peptides (i.e. the arrangement found in C1q) is more likely to bind to C1s because it will form a more stable complex. The resulting crystal structure will reflect the average arrangement of collagen peptides. Where a particular chain comprises two or more collagen peptides in the crystal, density from both chains will be observed. To date and to my knowledge, no structures of proteins bound to collagen heterooligomers have been determined.

So far, I have determined the structures of C1s in complex with C1qB (data to ~2.0 Å resolution) and C1qC (to 2.8 Å resolution), but not with C1qA. Moreover, I have determined the structures of C1s bound to C1qA and C heterooligomer; (data to ~ 2.7Å resolution) and C1qC and B heterooligomer; (data to ~ 1.94Å resolution). Although the resolution is only medium in these structures, both chains can be observed in the electron density. This is the first structure of any protein bound to a collagen heterooligomer to be solved.

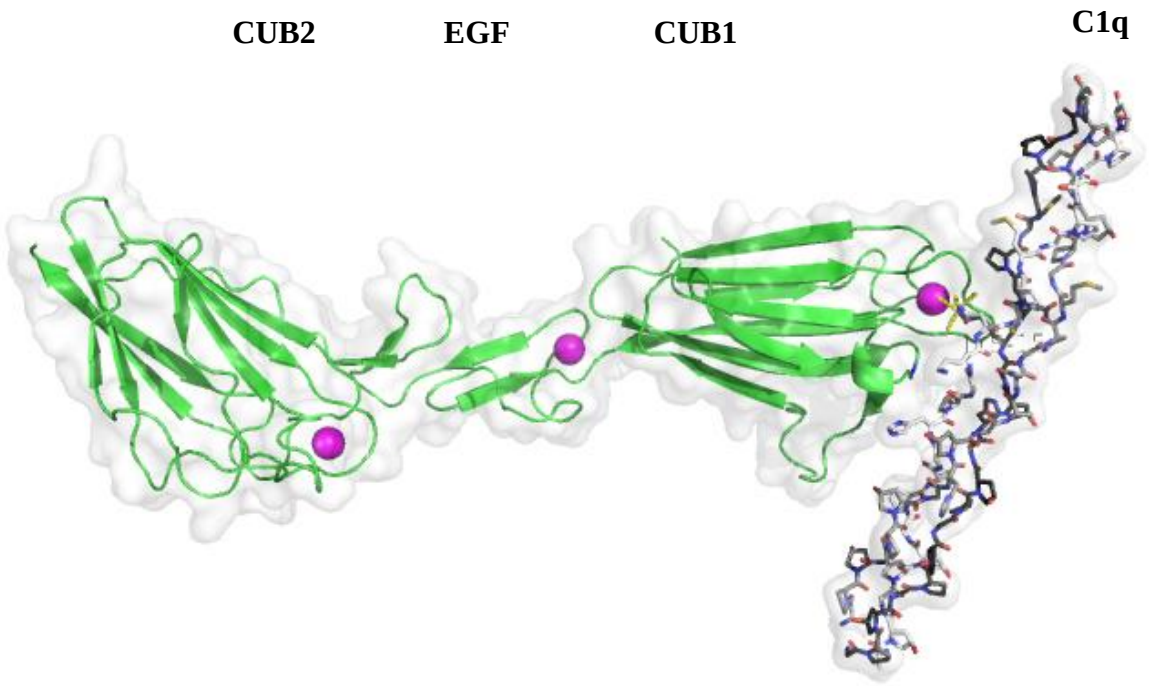


Figure 3-2: *This figure shows the overall structure of the 3D C1s in complex with C1qC peptide. The three domains of C1s are shown, which binds to the collagen peptide of C1q. Ca^{2+} is essential for the interaction and are shown as pink spheres.*

3.3.2 Structure of 3DC1s bound to C1qB:

Data were collected on beam line I04 at Diamond Light Source and diffracted to 2.0 Å resolution. The data were then processed by MOSFILM and Phenix to determine the structure as shown in (Figure 3.3). The asymmetric unit of the crystal contained a single complex, with the collagen peptide bound to the CUB1 of C1s. The CUB1-EGF-CUB2 domains of C1s are arranged linearly with three Ca^{2+} molecules; one present in each CUB domain and one within the EGF-like domain, near the CUB1-EGF junction. The side chain of Lys15 of the lagging strand of C1qB forms a hydrogen bonds with Ser100, Asp98 and Glu45 of the C1sCUB1 domain, which coordinates the Ca^{2+}. In addition, the amide nitrogen of Val16 of the middle

strand of C1qB forms a hydrogen bond with the side chain of Tyr52 of the CUB1 domain. The amide nitrogen of Val16 of the lagging strand of C1qB forms a hydrogen bond with Glu102 of C1s. The final interaction is between the side chains of Asn12 of the middle strand of C1qB and Glu48 of C1s.

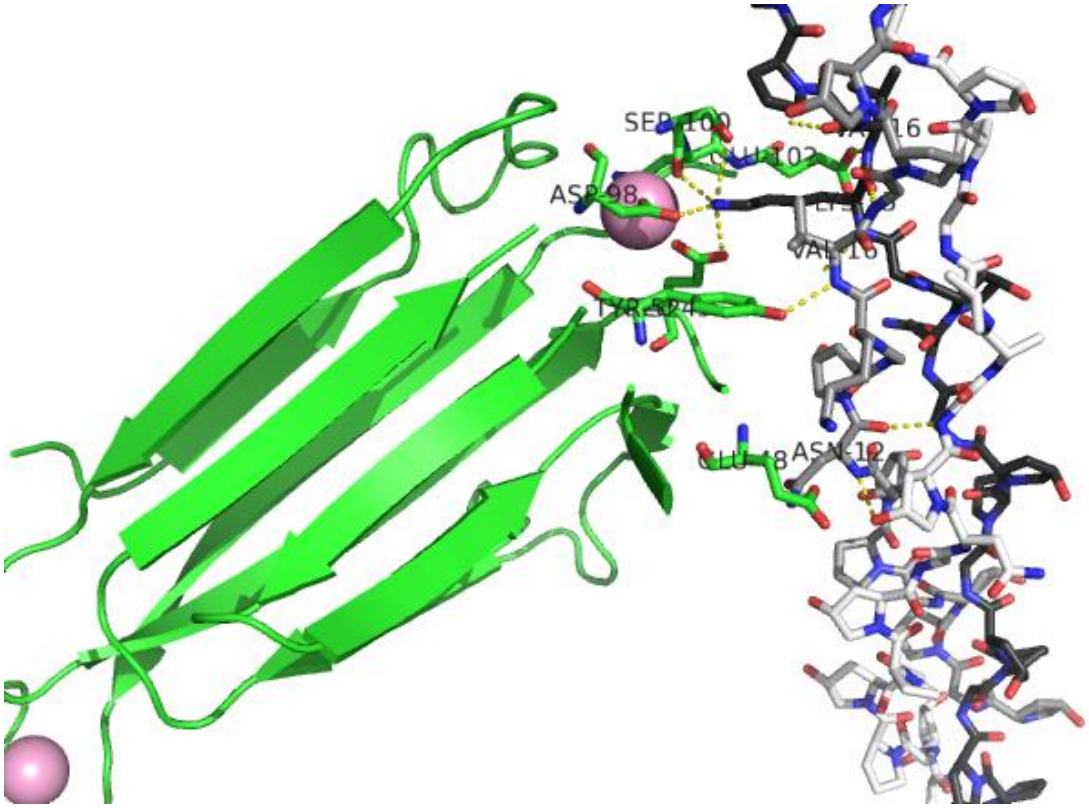


Figure 3-3*: The interaction between the CUB1 domain of C1s and of C1qB. The chains are colored as follow; black for the lagging strand, grey for the middle strand and white for the leading strand.*

Data collection and final refinement statistics summary for C1qB chain+3DC1s	
Resolution range (Å)	2.0
Space group	C1 2 1
a,b,c A	75.40 71.67 99.09
α,β,γ °	90.00 111.50 90.00
R_{sym}	10.12
I/σ(I)	5.8
Completeness	98.0
No, reflections	21857
R_{work}	0.21
R_{free}	0.23
RMSD bond lengths (Å)	0.038
RMSD bond angles (°)	3.0733
Number of protein atoms	30.12
Number of solvent atoms	8004
B-factors (Å^2)	22.6
water	-
Average multiplicity	3.4

Table 3.3: *Summary of the data collection and refinement parameters of C1qB in complex with C1s-CUB1.*

3.3.3 The interaction between the CUB1 domain of C1s and of C1qC:

The purified 3DC1s protein was mixed with the C1qC chain. Data were collected and processed as described in section 2.2.8. As before, residues Ser100, Asp98 and Glu45 of the CUB1 domain of C1s form hydrogen bonds with the Lys15 side chain of the lagging strand of C1qC chain, as shown in (Figure 3.4). However, there are no additional polar interactions.

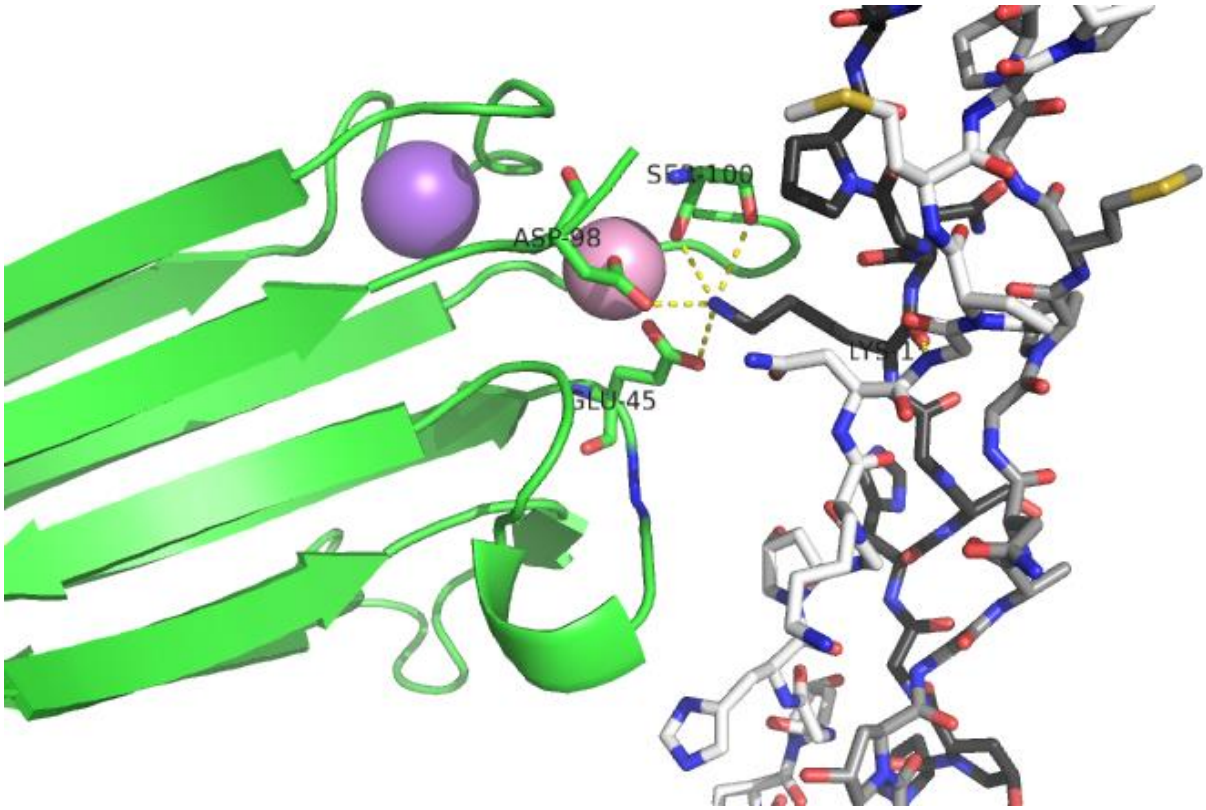


***Figure 3-4**: The interaction between the CUB1 domain of C1s and of C1qC. C1s residues Ser100, Asp98 and Glu45 interact with Lys15 of the lagging strand of C1qC as in the C1qB complex. The Ca^{2+} is shown as a pink sphere.*

Despite repeated attempts, no crystals were obtained for C1s bound to C1qA. The A chain contains a bulky tyrosine residue at position 18. It is likely that this residue cannot be accommodated in all three positions of the collagen chain without disrupting binding.

Data collection and final refinement statistics summary for C1qC+3DC1s	
Resolution range (Å)	2.8
Space group	C1 2 1
a,b,c A	75.40 71.67 99.09
α,β,γ °	90.00 111.50 90.00
R_{sym}	12.22
I/σ(I)	5.8
Completeness	98.0
No, reflections	21857
R_{work}	0.21
R_{free}	0.23
RMSD bond lengths (Å)	0.038
RMSD bond angles (°)	3.0733
Number of protein atoms	30.12
Number of solvent atoms	8004
B-factors (Å^2)	22.6
water	-
Multiplicity	3.4

Table 3.4: *Summary of the data collection and refinement parameters of C1qC in complex with the C1s-CUB1domain.*

3.3.4 Structure of 3DC1s bound to C1q A and C chains:

To assess the potential role of chain A of C1q, the C1qA and C1qC chains were mixed at a ratio of 3:1 ratio and cocrystallized with 3DC1s. An excess of chain A was used to ensure that it would be visible in the complex (Figure 3.5).

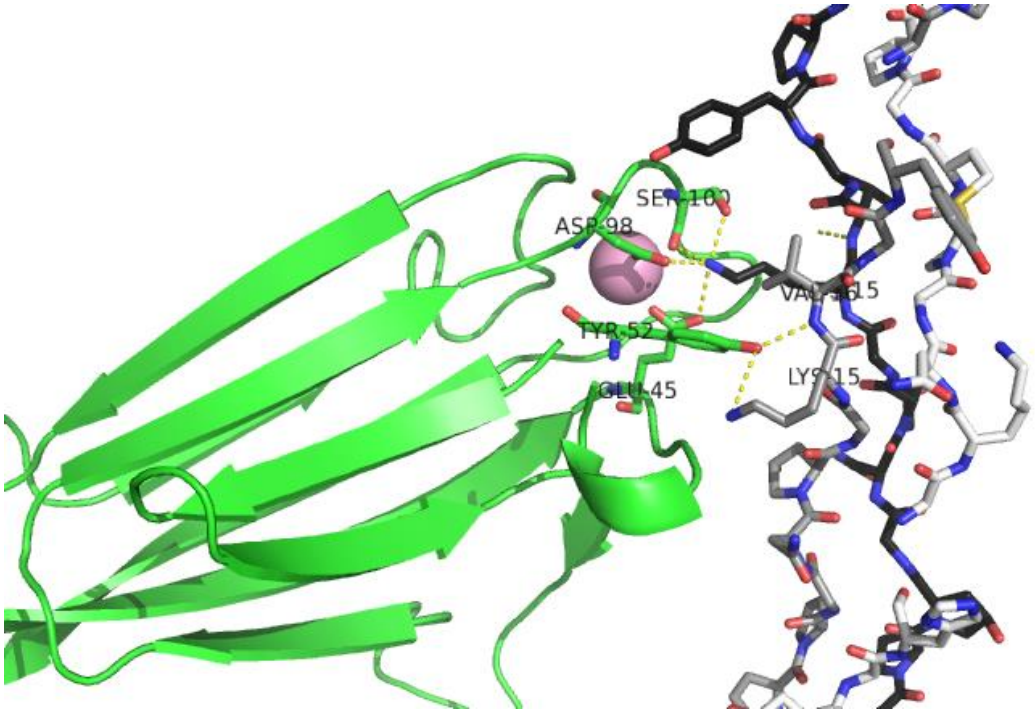


Figure 3-5: *The interaction between C1qA/C with the CUB1 domain of C1s. Ser100, Asp98 and Glu45 of the C1s interact with Lys15 of the lagging strand of the collagen (chain C). Additionally, hydrogen bonds are present between the Tyr52 residue of CUB1 and the amide group of Val16 of the middle strand (chain A,) as well as the side chain of Lys15 of the middle strand (chain A). Interestingly, density for the Tyr18 of chain A was only observed in the middle strand of the three collagen chains (not the leading or lagging) the other two chains were both predominantly C1qC. A further interesting observation was that Tyr18 packs against the hydrophobic side chain of Met19 of the leading strand. These data are consistent with C1qA being the middle strand and C1qC being the lagging strand.*

Data collection and final refinement statistics summary for C1qA/C +3DC1s	
Resolution range (Å)	2.7
Space group	C1 2 1
a,b,c A	73.74 71.67 99.09
α,β,γ °	90.00 111.50 90.00
R_{sym}	12.12
I/σ(I)	5.8
Completeness	82.0
No, reflections	21857
R_{work}	0.21
R_{free}	0.23
RMSD bond lengths (Å)	0.038
RMSD bond angles (°)	3.0733
Number of protein atoms	32.12
Number of solvent atoms	8022
B-factors (Å^2)	22.6
water	-
Multiplicity	3.4

Table 3.5: *Summary of the data collection and refinement parameters of the C1qA+C heterooligomer in complex with C1s-CUB1.*

3.3.5 Structure of 3DC1s bound to C1q C and B:

C1q peptides C and B were mixed at different ratios to assess the pattern of these strands and obtain full model of the C1q chains A, B and C.

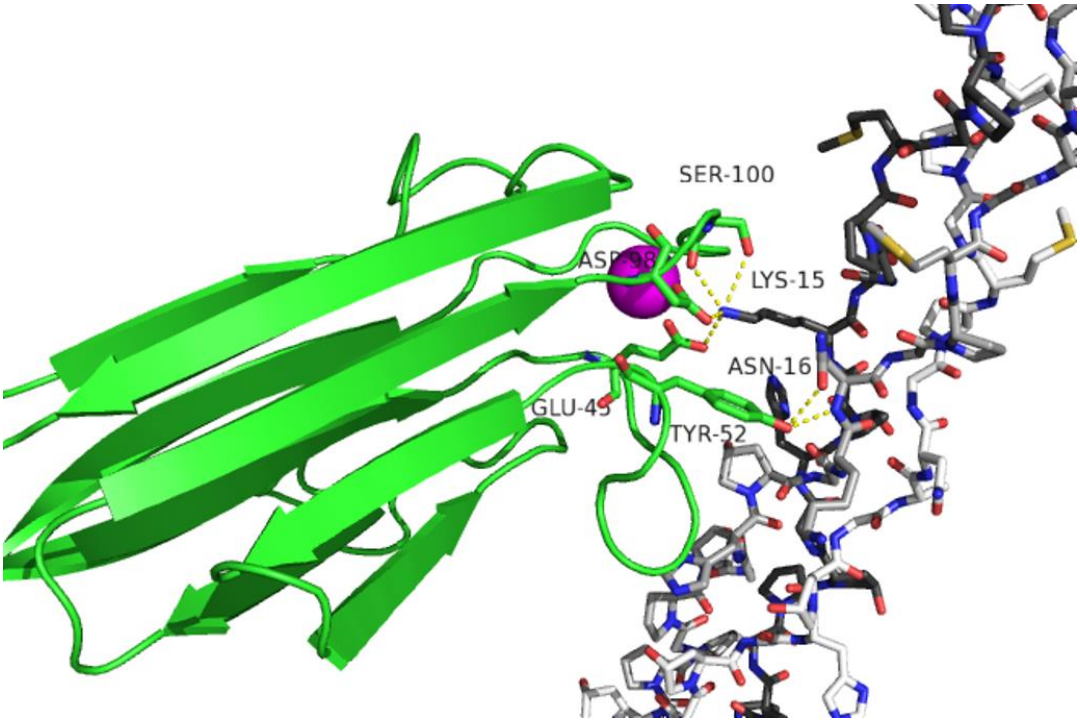


Figure 3-6: *This figure presents the structure of the complex between C1s and C1qC/B. The ratio of C:B was 2:1 respectively. Density for C1qB was observed predominantly in the middle strands whereas density for C1qC was observed in the lagging strand. Residues Ser100, Asp98 and Glu45 of the CUB1 domain of C1s form hydrogen bonds with the Lys15 of the lagging strand of the collagen (chain C). Additionally, hydrogen bonds are present between the Tyr52 residue of the CUB1 and the amide group of Asn16 of the middle strand (chain B).*

Data collection and final refinement statistics summary for C1qC/B +3DC1s	
Resolution range (Å)	1.94
Space group	C1 2 1
a,b,c A	73.74 71.67 99.09
α,β,γ °	90.00 111.50 90.00
R_{sym}	13.12
I/σ(I)	5.8
Completeness	82.0
No, reflections	21857
R_{work}	0.21
R_{free}	0.23
RMSD bond lengths (Å)	0.038
RMSD bond angles (°)	3.0733
Number of protein atoms	35.11
Number of solvent atoms	8002
B-factors (Å^2)	22.6
water	-
Multiplicity	3.4

Table 3.6: *Summary of the data collection and refinement parameters of C1qC+B heterooligomer in complex with C1s-CUB.*

3.4 Discussion:

In previous studies the structure of 3D C1s bound to the consensus motif: Hyp-Gly-Lys-Leu-Gly-Pro, was determined (Venkatraman Girija et al., 2013). However, the three chains of C1q are different (Gingras *et al.*, 2011). C1q is assembled from 3 chains called A, B and C. The binding site for C1r and C1s is located within the collagen-like domain. In this region, all three collagen chains contain a sequence similar to the protease consensus motif: Hyp-Gly-Lys-Xaa-Gly-Pro: chain A: GNOGKVGYO; chain B: GNOGKVGPO and chain C: GHOGKNGPM.

Structure	Collagen residues	C1s-CUB1 residues
C1qB	Lys15 (lagging strand)	Ser100 Asp98 Glu45
	Val16 (middle strand)	Tyr52
	Asn12 (middle strand)	Glu48
	Val16 (lagging strand)	Glu102
C1qC	Lys15 (lagging strand)	Ser100 Asp98 Glu45
C1qA+C	Lys15 (lagging strand)	Ser100 Asp98 Glu45
	Val16 (middle strand) Lys15 (middle strand)	Tyr52
C1qB+C	Lys15 (lagging strand)	Ser100 Asp98 Glu45
	Asn16 (middle strand)	Tyr52

Table 3.7: *This table shows the polar interactions between C1q chains in complex with the C1s-CUB1 domain. Hilighted resedues are conserved in all chains of C1q.*

The three collagen-like peptides form a right-handed helix with a characteristic one residue stagger between adjacent strands. Each chain binds to C1s, with the leading, middle, and trailing strands. Key to the interaction is the Lys15 of the lagging strand, which contacts the carboxylate groups of Glu45 and Asp98 and the hydroxyl and carbonyl groups of Ser100, three of the five residues that coordinate the Ca^{2+} of the CUB1 of C1s.

As presented in (Table 3.7), polar interactions highlighted in yellow are present in all of the structures indicating the key role of the lysine residue which is compatible with the previously solved structure of C1q/C1s (Venkatraman Girija *et al.*, 2013). Interestingly, density for chain A was only observed in the middle strand in the complex with chains A and C suggesting that it forms the middle strand in C1q. Chain B formed additional contacts with the CUB1 domain via the lagging strand of the collagen in C1s:C1qB compared to C1s:C1qC, implying that B is the lagging strand of the collagen in C1q, and that C is the leading strand. Thus, from the structures we can propose a model for the C1q:C1s interaction in which the leading strand of the collagen is C, the middle is A and the lagging is B. Notably, the C1qA/C:C1s and C1qB/C complexes are the first complexes between a protein and a collagen heterooligomer to be reported.

The structure of the collagen globular head has been determined previously by (Gaboriaud *et al.*, 2003). In this structure, the three chains are arranged clockwise in the order ACB≡CBA≡BAC (depending on which domain is the starting point clockwise), when observed from the base of the heterotrimer. In collagen the leading, middle and lagging strands are in the order of leading, lagging, middle from the N-terminus and in a clockwise direction i.e. C, B, A. Thus, the three collagen chains in my proposed arrangement are correctly ordered to connect to the globular domain of C1q without any requirements for strand crossing. Finally, mutagenesis studies using recombinant C1q (Kishore and Reid, 2000) have shown that the

lysine residues of the B and C chains of C1q are more important than lysine of C1qA with regard to $C1r_2s_2$ binding. This finding is also compatible with my model in which the lysine of C1qC interacts directly with C1s.

The method used here for identifying the leading lagging and middle strands of the collagen-like domain has great potential for use in other systems where protein-collagen complexes are assembled from collagen heterooligomers.

Chapter 4 . Interaction of the immune modulator LAIR-2 with MBL of the lectin pathway for complement activation

4.1 Introduction and objectives:

LAIR proteins are immune regulatory receptors that belong to the Ig superfamily. There are two members; a transmembrane protein LAIR-1 and a secreted soluble homologue named LAIR-2. Although LAIR-2 is similar to LAIR-1, it lacks the ITIM (Immunoreceptor tyrosine-based inhibitory motifs) baring transmembrane domain of the latter. LAIR-1 and LAIR-2 share 80% identity within the Ig like extracellular domain (Olde Nordkamp, Koeleman and Meyaard, 2014).

The LAIR-1 inhibitory function was characterised by the ability to inhibit the cytotoxicity in natural killer cells (NK) and T-cells (Meyaard *et al.*, 1997), thereby inhibiting the calcium mobilization and down regulation of cytokine production and maturation of dendritic cells (Meyaard *et al.*, 1999). The structure of LAIR-1 has been determined to 1.8 Aº (PDB entry code:3KGR) (Brondijk *et al.*, 2010). Previous work has shown that LAIR-1 binds specifically to collagen-like repeats (GPO)n and therefore, can directly down regulate immune responses (Lebbink *et al.*, 2006). LAIR-2 has been reported to compete with LAIR-1 for collagen binding and antagonises its role in regulating the immune response (Lebbink *et al.*, 2014). Recently, it has been reported that LAIR-1 can bind to the collagen-like domains of C1q and MBL to inhibit complement activation (Son *et al.*, 2012b). The subunits of MBL and C1q contain a collagen-like domain which is believed to be the binding site for LAIR proteins (Figure 4.1). Unlike LAIR-1, little is known about the function and structure of LAIR-2. The aim of this chapter is to investigate the potential interaction between LAIR-2 and the collagen-like domains of MBL.

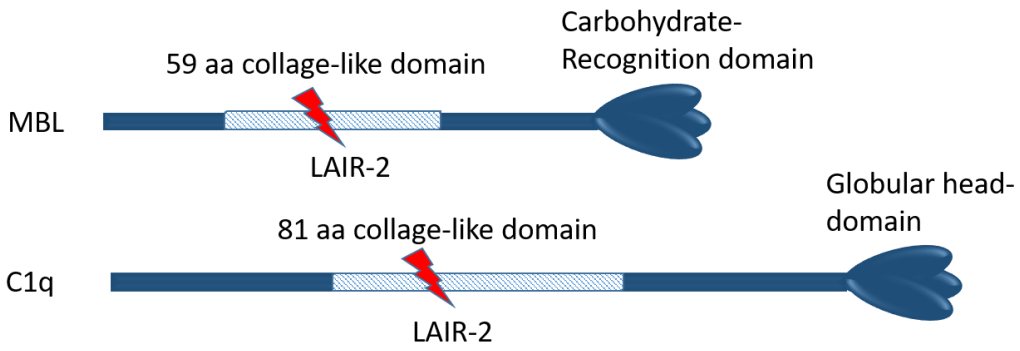


Figure 4-1: *Schematic representation of the collagen-like domain of MBL and C1q. The recognition molecules of the CP and LP contain a collagen-like domain of 59 amino acids for the LP and 81 amino acids for the CP as indicated with light blue colour. The collagen domain is believed to be the binding site for LAIR-2.*

Objectives:

- Production and purification of the full length (FL) LAIR-2 protein.
- Production and purification of human (hMBL).
- Attempt to crystallise FL-LAIR-2.
- Assess the binding between FL-LAIR-2 with hMBL using Bio-Layer Interferometry (BLI).

4.2 Materials and methods:

4.2.1 Materials:

Synthetic cDNA containing the FL-LAIR-2 protein was purchased from Thermofisher scientific. Cloning was carried out by the PROTEX (The University of Leicester cloning and protein production facility). Primers were designed following PROTEX guidelines. Platinum *Pfx* polymerase for PCR was purchased from Life Technologies. DNA and protein markers were supplied by New England Biolabs. Filters (0.2 µm), Ni^{+2}- NTA affinity resin, Superdex 200 and 75 16/60 columns used for protein purification were supplied by GE Healthcare. Mini and midi plasmid extraction and purification kits were supplied by Qiagen. SYBR safe DNA gel stain and loading dye were purchased from Invitrogen. The protein extraction and purification reagent BugBuster was supplied by Merck Millipore. Protease inhibitor tablets were purchased from Roche. All chemicals were purchased and supplied by Sigma and Thermo Fisher Scientific. Crystallisation screens JCSG, ProPlex, PACT, Morpheus and BCS were supplied by Molecular Dimensions. Mannose Sepharose resin was prepared and provided by Prof. Wallis (University of Leicester). The synthetic peptide comprising the sequence $(GPO)_6$ was obtained from Generon.

4.2.2 Cloning and amplification of FL-LAIR-2 gene:

FL-LAIR-2 cDNA was amplified by PCR from a synthetic cDNA. Primers were designed following the PROTEX guidelines to enable cloning into the expression plasmid pLEICS-03 by homologous recombination (Table 4.1). PCR amplification was carried out in 50 µl reactions containing 50 ng of DNA template in a master mix detailed in (Table 4.2). The PCR products were analysed on a 1% w/v agarose gel in TBE buffer. Bands were excised from the gel with a sterile scalpel and then purified from the agarose gel using the QIAgen quick gel extraction kit. Purified PCR products were cloned into the expression vector pLEICS-03 by

PROTEX to introduce an N-terminal His_6 tag and a Tobacco Etch Virus protease (TEV), cleavage site (Figure 4.2).

Primer	Sequence 5` ------- 3`
FL-LAIR-2 FP	TACTTCCAATCCATGTCTCCACACCTCACTGCTCTCCTG
FL-LAIR-2 RP	GACGGAGCTCGAATTTCATGGTGCATCAAATCCGGAGGCTTC

Table 4.1: *Oligonucleotides primers for the amplification of FL-LAIR-2 gene.*

PCR mixture and cycling conditions			
Component		**Volume**	**Final Concentration**
10x Pfx Amplification Buffer		5 µl	1 x
10 mM DNTp mixture		1.5 µl	0.3 mM each
50 mM $MgSO_4$		1 µl	1 mM
Forward primer (Table 4.1)		1 µl	2 pM
Reverse primer (Table 4.1)		1 µl	2 pM
FL-LAIR-2-template		1 µl	50 ng/µl
NanoPure water		39 µl	
Platinum Pfx polymerase		0.5 µl	1 unit
Cycling conditions			
94°C	35 Cycles	5 minutes	Pre-denaturation
94°C		15 seconds	Denaturation
63°C		30 seconds	Annealing
68°C		1.25 minutes	Extension
68°C		5 minutes	Final Extension

Table 4.2: *Summary of the PCR run and reaction mixture.*

After cloning, clones were verified by sequencing which was performed by the University of Leicester sequencing facility, PNACL.

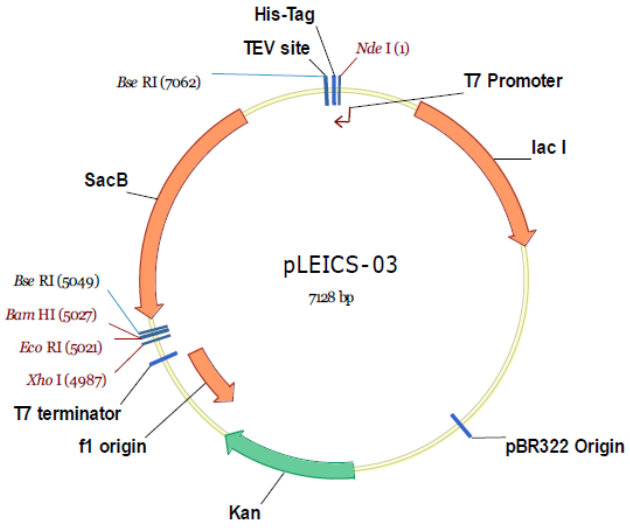


Figure 4-2: *Vector map of the expression vector pLEICS-03**.** LAIR-2 gene was inserted in the EcoR1site.*

4.2.3 FL-LAIR-2 expression and purification:

The pLECS-03 containing the FL-LAIR-2 cDNA was transformed into competent *E. coli BL-21 DE3* cells as follows; 2 µl of the recombinant DNA was added to 100 µl of competent cells and incubated on ice for 30 minutes, followed by 2 minutes heat shock treatment at 42ºC. The cells were then put back on the ice after which 300 ml on LB was added and incubated for 1 hour in a shaking incubator at 37ºC. Cells were then plated on a LB agar plate using a sterile L-shaped loop, containing 50 µg/ml Kanamycin and then incubated overnight at 37ºC. The freshly transformed cells were then used to inoculate 1 L of LB containing 50 µg/ml Kanamycin. After this, cells were then grown to an OD_{600} of 0.5 and induced with 1mM IPTG and incubated overnight in a shaking incubator at 16ºC. The cells were then harvested by centrifugation at 3396 g for 20 minutes at 4ºC and then resuspended in 30ml of lysis buffer containing (50 mM Tris-pH7.5, 150 mM $NaCl_2$, 1% v/v Tween 20) and one tablet of a protease inhibitor cocktail. The cell suspension was then sonicated on ice using large probe, with 8 pulse for 30 seconds at an amplitude of 8, with a one-minute break between each pulse to prevent

overheating. The cell lysate was then centrifuged at 20,000g for 20 minutes at 4ºC. The protein was expressed in the soluble fraction and purified by affinity chromatography on Ni^{2+}- NTA Sepharose column pre-equilibrated with 20 ml of 50 mM Tris-pH 7.5, containing 150 mM $NaCl_2$. The cell lysate was then diluted at a 1:1 ratio with the equilibration buffer and loaded at 4ºC overnight. After loading, the column was washed with 20ml of 50 mM Tris-pH 7.5, 150 mM $NaCl_2$ and 20 mM imidazole. Protein was then eluted from the column with the same buffer containing 500mM imidazole. Fractions were analysed on a non-reducing 15% SDS-PAGE gel after which they were pooled together and further purified by gel filtration on a Superdex 75 16/60 column equilibrated with 50 mM Tris pH7.5, containing 150 mM NaCl. Purified protein was then analysed on a non-reducing 15%SDS-PAGE gel, concentrated to 6mg/ml and stored at -80 ºC.

4.2.4 TEV protease digestion of the FL-LAIR-2 His_6 tag:

In preparation for crystallisation screens, the His_6 taq was removed by digestion with the TEV protease (PROTEX). Protein was first dialysed in TEV reaction buffer (50 mM Tris-pH7.5, 0.5 mM EDTA and 3 mM glutathione/0.3 mM oxidized glutathione). TEV protease was then added at the ratio of 1unit/500µl protein and then was left overnight at 4 ºC in a mixing roller. The mixture was separated by passing the mixture through a Ni^{+2}- NTA Sepharose column, and the efficiency of the digestion was analysed on a 15% non-reducing SDS-PAGE gel. Digested protein was then dialysed in 50 mM Tris-pH 7.5, containing 150 mM NaCl and was then purified further by gel filtration. Purified protein was then concentrated to 4mg/ml and stored at -80 ºC for further experiments.

4.2.5 Preparation of Mannose Sepharose resin:

To purify hMBL, Sepharose 6B beads were first coupled with mannose. Briefly, 40ml of the Sepharose 6B beads were washed with 2L of H_2O with gentle stirring and then were added to

40ml of 0.5 M $NaHCO_3$ pH 11 and 5ml of divinyl sulfone at room temperature for 1hour to activate the Sepharose beads. The beads were then washed with 2L of H_2O and then added to 40ml of 0.5 M $NaHCO_3$ pH 10 containing 20% mannose and incubated overnight to allow coupling of the sugar to the Sepharose beads. Following this, the Sepharose was washed with 2L of H_2O and blocked with 40ml of 0.5 M $NaHCO_3$ and 2ml β-mercaptoethanol for 30 minutes. Finally, the Sepharose was washed with 4L of H_2O and stored at 4ºC until further use.

4.2.6 hMBL expression and purification:

CHO cell line expressing hMBL was kindly provided by (Prof. Wallis, University of Leicester). hMBL was harvested from the cells following the same protocol described in section (2.2.2). The cells were grown in a triple-layer cell culture flask in a MEMα- media supplemented with (0.5µM MTX, 10% v/v DHFBS and 50units/ml P/S). Once confluence was reached, media was replaced with CHO-S-SFM II without nucleotides supplemented with 50 mM HEPES-pH 7.5 to help maintaining pH during cell growth, 0.5 µM MTX, and 50 units/ml of P/S. Media was collected every other day and spun at 671 g to remove cell debris and stored at -20 ºC. hMBL was then purified by affinity chromatography on a Mannose-Sepharose column as follows: 300 ml of the harvested hMBL media was mixed with an equal volume of high salt buffer (HSB) consisting of 50 mM Tris-pH 7.5, 1.2 M NaCl and 10 mM $CaCl_2$. 2ml of the mannose Sepharose column was equilibrated with 10 ml of HSB after which the media was passed through the column overnight at 4 ºC. The column was then washed with 10 ml of HSB followed by 10 ml of low salt buffer (LSB) (50 mM Tris-pH 7.5, 150 mM NaCl and 10 mM $CaCl_2$), to remove any non-specific binding to the column. The protein was then eluted from the column with the elution buffer (EB) (50 mM Tris-pH7.5, 150 mM $NaCl_2$ and 2.5 mM EDTA). Protein was collected in 1ml fractions. The fractions containing the eluted protein were identified by checking the elations on a non-reducing 15% SDS-PAGE gel. Fractions were then

pooled together and dialysed in 50 mM Tris-pH 7.5, 150 mM NaCl and 5 mM $CaCl_2$) using 10000 MWCO dialysis tubing and were concentrated to 2.2mg/ml.

4.2.7 FL-LAIR-2 crystallisation:

Crystallisation trials were carried out using commercial crystallisation screens provided by Molecular Dimensions. Using setting drop method as described in section 2.2.3, crystallisation screens PACT, Morpheus, JCSG, Proplex and BCS were set up by mixing 0.1µl of the screen reservoir buffer with 0.1µl of the digested FL-LAIR-2 (6mg/ml) using a Mosquito NanoDrop crystallisation robot. Screens were setup at room temperature and 4 ºC. Trials were set up of FL-LAIR-2 alone and in complex with a synthetic peptide containing 6 GPO repeats using an excess of peptide (Lebbink *et al.*, 2014). Plates were checked regularly, and observations were marked for each condition. After two weeks, small crystals were observed in the PACT screen: 20 % PEG 3350, 0.2 M Sodium citrate tribasic dehydrate and 0.1 M Bis-Tris propane pH7.5 and in the Proplex screen: 0.1 M Na-HEPES-pH7.0 and 1M Sodium citrate. Hits were then optimised further by setting up a rang of the condition component. Each condition was optimised in a 48 well Combi Clover crystallisation plates, using 0.7 to 1.2µl range of protein volume mixed with an equal volume of the condition buffer. Then the precipitant, pH and any other condition components were optimised individually (Table 4.3). Additionally, an additives screen was used for each condition to attempt to improve the crystal quality. All buffers were made by hand and plates were incubated at room temperature and at 4 ºC.

Buffer composition for each condition used in crystallisation	
1	0.2 M Sodium citrate tribasic dehydrate, 25%, 24%, 23%, 22%, 21%, 20% w/v PEG 3350, 0.1 M Bis-Tris propane pH7.5
2	0.2 M Sodium citrate tribasic dehydrate, 23% w/v PEG 3350, 0.6 to 0.1 M range of Bis-Tris propane pH7.5.
3	23% w/v PEG 3350, 0.2 M Bis-Tris propane pH7.5, 0.6 to 0.1 M range of Sodium citrate tribasic dehydrate.
4	0.1 M Na-HEPES-pH7.0, 1 to 0.5 M range of Sodium citrate.
5	0.8 M Sodium citrate, 0.6 to 0.1 M range of Na-HEPES-pH7.0.

Table 4.3: *Optimisation conditions for FL-LAIR-2 crystallisation. Each buffer was used to set up a 48 well crystallisation plate.*

4.2.8 Binding of FL-LAIR-2 to hMBL:

The interaction between FL-LAIR-2 and h-MBL was tested using Bio-Layer Interferometry (BLI), on an OCTETQK instrument. The principle of BLI works by detecting the binding between an immobilized protein, either by (amine coupling or Streptavidin/biotin interaction), on a sensor tip with a target sample (Figure 4.3).

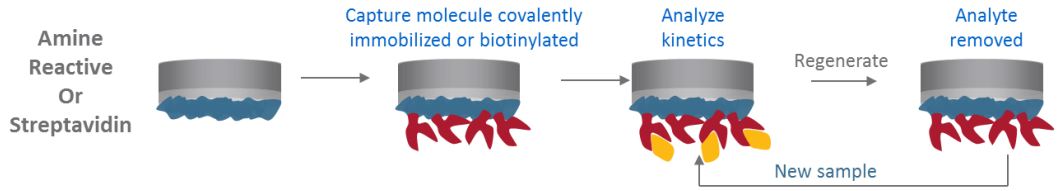


Figure 4-3: *BLI principle showing the steps of sample loading and target detection (ForteBio manual).*

hMBL was biotinylated to allow the protein to be immobilized onto a Streptavidin coated sensor tip using a 20-fold molar excess of biotin to insure complete modification. Following this, protein was dialysed in 50 mM HEPES-pH 7.5 and 150 mM NaCl to remove any excess biotin. Each experiment was carried out as follows: the biosensor tip was first hydrated with 200µl of running buffer (50 mM HEPES-pH 7.5 and 150 mM NaCl) for 10 minutes. Then, the biosensor was exposed to 250 µl of the same buffer to generate an initial baseline reading. 150µl of the biotinylated h-MBL protein (50µM) was then immobilized onto the biosensor tip in the same buffer for 120 seconds. Following this, the biosensor was then exposed to 200 µl of the running buffer for 30 seconds. Then FL-LAIR-2 was injected for 120 second at the following concentrations 50, 100, 200, 400 and 800 µM to measure association. Finally, the biosensor was then exposed to 250 µl of the running buffer to obtain a dissociation curve. Following the experiment run completion, the data were analysed using the ForteBio evaluation

software. The K_D affinity constant (equilibrium dissociation constant) was calculated by measuring the on (k_a) and off (k_d) rate using:

$$KD = \frac{[A] \,.\, [B]}{[AB]} = \frac{Kd}{Ka}$$

The K_D is in Molar units (M) and corresponds to the concertation at which 50% of the analyte is bound. The smaller the affinity constant, the tighter the interaction. In this experiment, the raw binding data were fitted to a 1:1 binding model using a global fit for all injected concentrations. The quality of the data fitted to the binding model was evaluated using steady-state analysis. An R^2 value higher than 0.9 indicates a good fit to the model.

4.3 Results:

Previous work has reported that the secreted LAIR-2 protein binds to the collagen repeats $(GPO)_n$ with higher affinity than its homologue LAIR-1 (Lebbink *et al.*, 2006). MBL, ficolins and C1q all contain a collagen-like domain which can potentially interact with LAIR-2 to prevent complement activation. To study this interaction, I produced recombinant FL-LAIR-2 and hMBL and tested the interaction using BLI.

4.3.1 Cloning of the FL-LAIR-2 cDNA:

The cDNA encoding the FL-LAIR-2 was amplified from synthetic cDNA (purchased from Invitrogen) using PCR. The product was then separated on a 1% agarose gel (Figure 4.4), purified and cloned into the expression vector pLEICES-03 by homologous recombination.

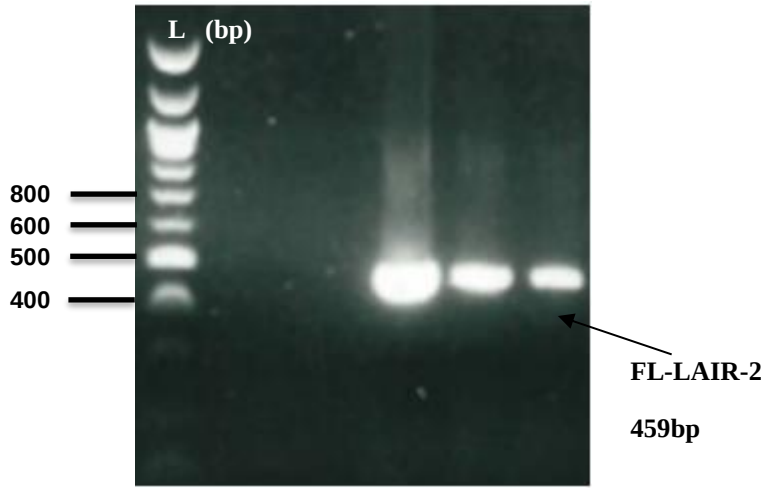


Figure 4-4: *1% agarose gel of the FL-LAIR-2 cDNA amplified via PCR. The expected size of the product was 459bp.*

The clone was then supplied to PNACL, the sequencing facility of the University of Leicester to confirm the sequence.

4.3.2 Expression and purification of FL-LAIR-2:

Following the successful cloning of the FL-LAIR-2 cDNA, the protein was expressed in *E coli BL-21* (DE3) cells. The protein was extracted from the soluble fraction following sonication of the cells and was purified by affinity chromatography on a Ni^{+2}- NTA Sepharose column. Elution fractions from the column were analysed on a non-reducing 15% SDS-PAGE gel (Figure 4.5).

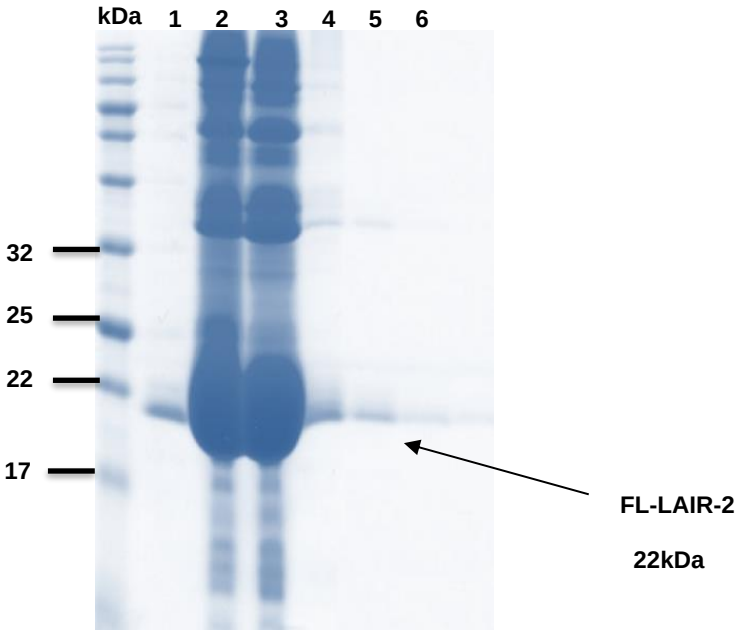


Figure 4-5: *15% non-reducing SDS-PAGE analysis of six elution fractions (left to right) following affinity purification on a Ni^{+2}- NTA column. FL-LAIR-2 migrated at the expected size of 22 kDa.*

Fractions containing the FL-LAIR-2 were then purified further by gel filtration on a Superdex 75 16/60 column (Figure 4.6) to remove contaminants. The protein eluted at an apparent molecular mass of ~22 kDa indicating that LAIR-2 is a monomer.

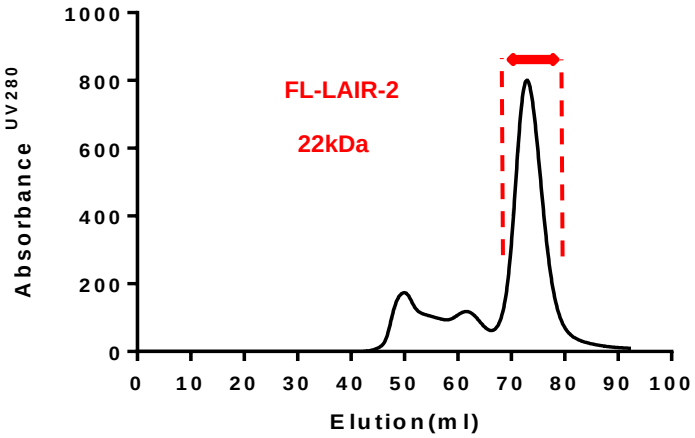


Figure 4-6: *Gel filtration purification of FL-LAIR-2 on a Superdex 75 16/60 column. The protein eluted from the column at the expected molecular mass of a monomer 22kDa.*

Following gel filtration, fractions were analysed on a 15% non-reducing SDS-PAGE gel to assess their purity (Figure 4.7).

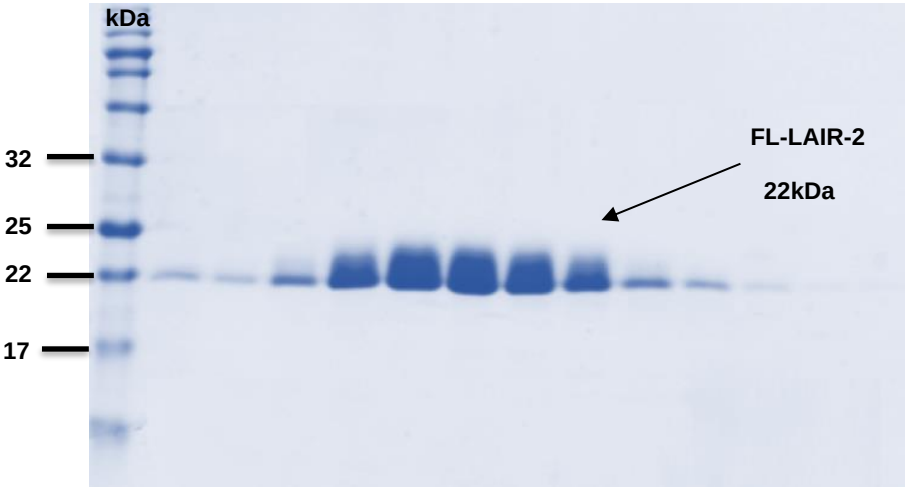


Figure 4-7: *SDS-PAGE analysis of gel filtration purification for the FL-LAIR-2 protein. No contaminants were detected after gel filtration.*

The additional gel filtration purification step was successful in removing impurities. Fractions were pooled and concentrated to 6 mg/ml using 10,000 MWCO concentrator tubes and stored at -80 ºC.

4.3.3 TEV digestion and removal of the His_6 tag:

For some of the LAIR-2 crystallisation experiments, the His-tag was removed prior to gel filtration. Following affinity chromatography, fractions were pooled and digested with the TEV as described in section 4.2.4. To remove both the tag and any TEV proteases (which itself contains a His-tag) the digested mix was passed through a second Ni^{+2}- NTA affinity column. The TEV protease and the tag remained bound to the column, while the digested LAIR-2 passes straight through. Both the flow through and elution fractions (500 µl) were collected and analysed by SDS-PAGE (Figure 4.8). After digestion, LAIR-2 migrated at ~18 kDa (compared to ~22 kDa for the undigested protein), confirming that the tag was successfully removed.

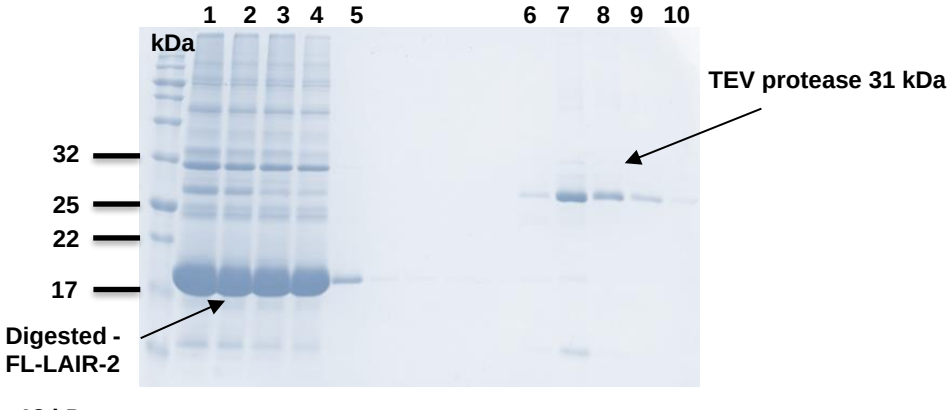


Figure 4-8: *15% non-reducing SDS-PAGE analysis of the TEV digestion of FL-LAIR-2. The digested protein eluted in the flow-through fractions (1 to 5) with a molecular mass of 18kDa. TEV enzyme itself possesses a his_6 tag and was separated successfully in the high imidazole elution fractions 6 to 10.*

Finally, impurities in the LAIR-2 fractions were removed by gel filtration chromatography as described previously (Figure 4.9).

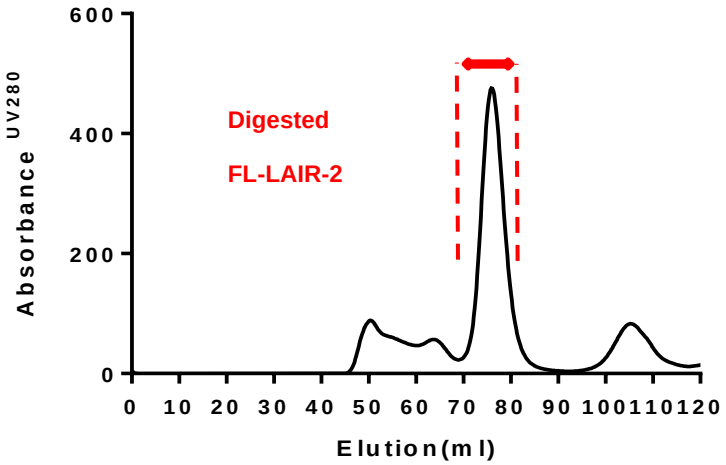


Figure 4-9: *Gel filtration purification for the digested FL-LAIR-2. The digested protein peak eluted slightly later from the column (78 ml) compared to un-digested FL-LAIR-2 (72 ml) confirming the successful removal of the* His_6 *tag.*

Fractions were checked on a non-reducing 15% SDS-PAGE gel to assess their purity (Figure 4.10), and then were pooled and concentrated to ~ 4 mg/ml and stored at -80ºC.

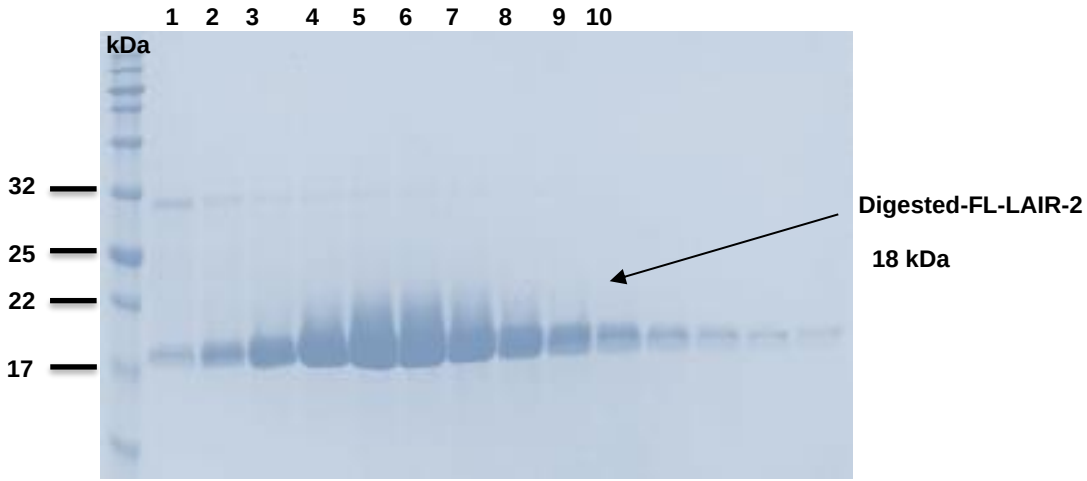


Figure 4-10: *Gel filtration fractions on a 15% non-reducing SDS-PAGE gel for the digested FL-LAIR-2. Fractions 3-10 (left to right) were pooled and concentrated. Fractions 1 and 2 were discarded to remove the minor contaminant present at ~30 kDa.*

4.3.4 hMBL expression and purification:

hMBL was expressed and purified from a stable CHO cell line by affinity chromatography on a Mannose-Sepharose column. Elution fractions were checked on a non-reducing 15% SDS-PAGE gel to assess the expression and purity of the fractions.

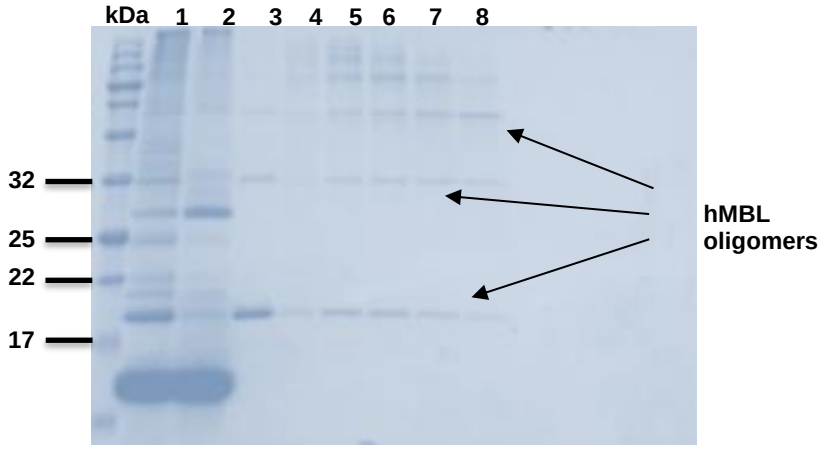


Figure 4-11: *h-MBL purification analysis on non-reducing SDS-PAGE gel. Lanes 1 and 2 represent the flow through and wash fractions. Lanes 3 to 8 are the elution fractions and contain purified h-MBL.*

hMBL migrates as multiple bands on a non-reducing SDS-PAGE gel due to the presence of disulphide linked oligomers of the hMBL polypeptide (Ahn *et al.*, 2013).

4.3.5 FL-LAIR-2 crystallisation:

Crystallisation trials were carried out initially using JCSG, PACT, Proplex, BCS and Morpheus screens. Plates were tested at room temperature and at 4ºC for both digested and un-digested FL-LAIR-2. The concentration of FL-LAIR-2 was ~6 mg/ml, whereas the cleaved FL-LAIR-2 protein was tested at ~4 mg/ml. Although no single crystals were obtained after several weeks, spherulites were observed in several conditions for the digested FL-LAIR-2. Spherulites are protein clusters but with lower order than crystals. The presence of spherulites is very encouraging and often leads to protein crystals following optimisation. Optimisation of the promising conditions was carried out in 48 well plates by altering each condition component as shown in (Table 4.3), using a systematic approach which involves changing one component at a time and assessing the effects on crystal quality. Further optimisation was done using a synthetic peptide containing six GPO repeats, which has been shown to be a ligand for LAIR-1 and LAIR-2 Additional optimisation was attempted using an additive screen from Hampton Research. Unfortunately, no crystals were obtained. Crystallisation process can take time considering the range of variables needed to be optimised.

4.3.6 Binding of h-MBL to FL-LAIR-2:

Previous work has reported that FL-LAIR-2 binds to collagen type I and III and competes with its partner, LAIR-1 in regulating immune responses (Olde Nordkamp, Koeleman and Meyaard, 2014). To investigate the potential interaction of LAIR-2 with hMBL, biotinylated hMBL was immobilised onto a Streptavidin sensor tip, and FL-LAIR-2 was injected over the tip at increasing concentrations (Figure 4.12). Binding was detected as evidenced by an increase in the sensor signal as the hMBL was immersed in the FL-LAIR-2 solution, compared to a tip immersed in a buffer-only control. The data were fitted globally to 1:1 model to calculate the K_D as shown in (Figure 4.12). The calculated affinity (K_D in M) was 5.6 ± 1.3 x 10^{-10} M (values are means ± SEM from two independent experiments). The association and dissociation constant were 3.7 x 10^{7} $M^{-1}s^{-1}$ and 2.09 x10^{-2} s^{-1}, respectively. The data achieved a good fit with R^2 value of 0.9314.

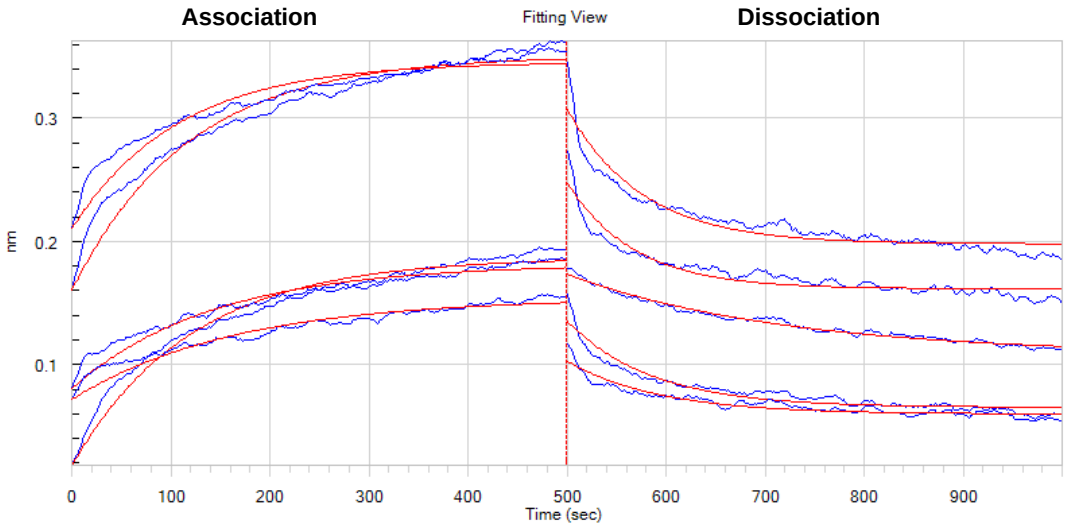


Figure 4-12: *Binding of FL-LAIR-2 to the immobilised h-MBL. LAIR-2 was injected at different concentration (800, 400, 200,100 and 50µM). The K_D calculation was performed by fitting to 1:1 model as indicated in red lines.*

4.4 Discussion:

The complement system is a host defence mechanism to aid in the fight against disease. It is activated via three pathways; the CP, the LP and the alternative. All pathways converge to generate C3 convertase and leading to the formation of MAC and target cells lysis. The recognition molecules of the CP and LP ; C1q and MBL respectively contain a collagen like domain by which their associated serine $C1rs_2$ and MASPs proteases are attached. This specific collagen-like domain is a target for several immune regulators by blocking the collagen-like domain and leaving the recognition molecules in their inactive state. Most vital organs such as kidneys, lungs and heart are susceptible to complement mediated injury. Therefore, any dysfunction and incorrect activation of this system is thought to be involved in the pathogenesis of disease states, particularly ischemia-reperfusion injury and autoimmune disease. The leukocyte associated immunoglobulin-like receptors (LAIRs), are a family of proteins involved in regulating host immune cell responses (Lebbink *et al.*, 2006). There are two members of the LAIRs family; the transmembrane inhibitory immune protein LAIR-1 and an 80% homolog soluble partner LAIR-2. It has been reported that collagen are high affinity ligands for LAIR-1 protein , interaction of LAIR-1 with collagens directly inhibits immunocyte activation in vitro and may represent a key mechanism of peripheral immune regulation through extracellular matrix. The identification of collagens as ligands for LAIR-1 revealed a novel function for extracellular matrix components as potential immune regulatory proteins. However, LAIR-2 is believed to play a regulatory role in the interaction between collagen and LAIR-1 and therefore compete with LAIR-1 for collagen interaction affecting its role on regulation of immune response. (Lebbink *et al.*, 2014).

To assess the interaction between LAIR-2 and MBL, I have successfully expressed and purified both FL-LAIR-2 and hMBL and investigated the interaction using BLI binding assay.

The binding data reported here shows that FL-LAIR-2 binds to the immobilised hMBL with a K_D of 5.6 x 10^{-10} M, indicating high affinity for the collagen domain in hMBL and supporting the recently reported study about MBL-LAIR-2 interaction in which the Fc domain of LAIR-2 bind to MBL collagen-like domain (Nordkamp *et al.*, 2014). The interaction between MBL and LAIR-2 is believed to affect the LP activation by blocking the associated serine proteases MASPs from attaching to the recognition molecule MBL and leaving it in an inactive state. However, the concertation of LAIR-2 in the circulation is less than 0.3 µg/ml (1.4 x 10^{-11} M), which is so low that it is unlikely to affect MBL and C1q function significantly (Olde Nordkamp, Koeleman and Meyaard, 2014). The normal serum levels of MBL is between 1 and 2 µg/ml (Bouwman, Roep and Roos, 2006). Attempt to crystallise FL-LAIR-2 for structure determination of the LAIR-2 protein and LAIR-2 in complex with MBL were unsuccessful with both digested and un-digested forms. Having said that, spherulites were observed in several crystallisation screens which is a promising indication for crystal formation. Despite multiple attempts, crystallisation trails can take time to optimise considering the range of variables needed in the crystallisation conditions and the time for crystal growth, this was not possible due to time constrains. Crystallisation trials will continue in the laboratory using the purified proteins produced in this part of the project to assess the specific residues responsible for the interaction.

Chapter 5 . Inhibition of the Classical pathway by *Borrelia burgdorferi* surface protein BBK32

5.1 Introduction:

Borrelia burgdorferi is the causative agent of Lyme disease. It produces multiple surface proteins that enable it to colonise host cells and evade the attack by the complement system. Many of the produced proteins function by recruiting complement regulators to the surface of *B.burgdorferi* and thereby protecting it from complement attack (Caine and Coburn, 2016). Recently a 47kDa secreted surface protein called BBK32 was found to interact with the C1r serine protease within the C1 complex of the CP and blocking the system activation (Garcia *et al.*, 2016). BBK32 comprises two distinct regions (Figure 5.1); the N-terminal region containing non-overlapping binding sites for glycosaminoglycan (GAG) residues 45-68 (Li *et al.*, 2006) and fibronectin (Fn) of the extracellular matrix (Probert and Johnson, 1998) and a globular C-terminal region of which residues 206-354 have been recently reported to interact with C1r (Xie *et al.*, 2019). The complement C1r protein is formed of six different domains as follow; CUB1-EGF-CUB2-CCP1-CCP2-SP. The most recent study at the time of preparing this thesis has shown that the C-terminal of BBK32 binds to the SP domain of C1r and inhibit its catalytic activity, therefore inhibiting the complement activation via CP (Xie *et al.*, 2019). To investigate the interaction between BBK32 and C1r, the aim of this chapter is to assess and characterise the interaction between the complement component C1r and the C-terminal of BBK32. The main objectives of this chapter are to:

- Produce and purify the full-length (FL) and C-terminal domains of BBK32.
- Produce and purify FL-C1r and C1r fragments CUB2, CUB2-CCP1, CUB2-CCP1CCP2 and CUB2-CCP1-CCP2-SP to assess the domain(s) responsible for the interaction.
- Assess the binding with Bio-Layer Interferometry (BLI) and in solution with gel filtration.
- Attempt to crystallise C-BBK32 on its own and in complex with C1r.

5.2 Materials and methods:

5.2.1 Materials:

A pET-28a expression vector containing C1r-CUB was kindly provided by Prof. Wallis. Cloning into the expression vector. pLEICS-01 was performed by the cloning facility of the University of Leicester PROTEX. Primers were designed following PROTEX guidelines. Platinum *Pfx* polymerase for PCR was purchased from Life Technologies. DNA and protein markers were supplied by New England Biolabs. All chemicals were purchased and supplied by Sigma and Thermo Fisher Scientific. Crystallisation screens JCSG, ProPlex, PACT, Morpheus and BCS were supplied by Molecular Dimensions.

5.2.2 Cloning and production of the FL and C-terminal BBK32 portion:

5.2.2.1 Preparation of chemically competent cells using $CaCl_2$ method:

To produce and purify the target proteins, *E. coli* competent cells BL-21 (*DE3*) and XL-10 were generated using the $CaCl_2$ protocol described in (Chang, Chau and Landas, 2017). The BL-21 (*DE3*) cells were used for protein expression, and the XL-10 cells were used for recombinant plasmid amplification and cloning. The *E. coli* strains were cultured onto LB agar plates and incubated overnight at 37ºC. The next day, one colony was inoculated into 10ml of LB broth. The culture was grown overnight in a shaking incubator at 37ºC. The following day the 1 ml of the overnight culture was used to inoculate 100 ml of LB broth. The culture was grown in a shaking incubator at 37ºC until an OD_{600} of 0.3-0.4 was reached. The cells were then harvested by centrifugation at 1509 g for 10 minutes at 4ºC. Following this, the cell pellet was resuspended with ice cold 50mM $CaCl_2$ and left on ice for 2-3 hours. The cell suspension was then centrifuged at 1509 g for 15 minutes at 4ºC. The harvested cell pellet was then resuspended gently in 50 mM $CaCl_2$ containing 25% glycerol. Cells were separated into 50µl aliquots and frozen immediately in dry ice and stored at -80ºC for further use in cloning and

protein expression. The $CaCl_2$ treatment method promotes the uptake of plasmid DNA in bacteria, allowing for the process of transformation to occur.

5.2.2.2 Cloning and amplification of FL and C-terminal BBK32:

The BBK32 gene constructs were amplified by PCR (polymerase chain reaction) from a synthetic gene purchased from Thermo Fisher Scientific. Primers were designed according to the PROTEX guidelines, which uses ligase-independent recombination (Table 5.1). PCR was carried out using 50 µl master mix detailed in (Table 5.2) in a Multi Gene II thermocycler instrument. The FL gene was amplified in two overlapping fragments. The final PCR products were analysed on a 1% agarose gel run in TBE buffer. DNA fragments were cut from the gel using a sterile scalpel and were purified using QIAgen quick gel extraction kit (Promega). Purified PCR products were supplied to PROTEX for cloning into the expression vector pLEICS-01 (Figure 5.1). Both the FL and C-BBK32 constructs contained an N-terminal 6-His tag with a tobacco etch virus (TEV) protease cleavage site, between the tag and the N-terminus of the inserted gene. After cloning, the clones were validated by sequencing carried out at the University of Leicester sequencing service (PNACL).

Primer	Sequence 5` ------- 3`
C-BBK FP	TACTTCCAATCCATGGATGAGTATGATGAAGAGGATG
C-BBK RP	TATCCACCTTTACTGTCAGTAATACCACACGCCATTTTTATCGATG
FL-BBK FP	TACTTCCAATCCATGATGGACCTGTTTATCCGCTAT
FL-BBK RP	TATCCACCTTTACTGTCAGTAATACCACACGCCATTTTTATC

Table 5.1: *Primers used to amplify the FL and C-terminal of BBK32.*

PCR mixture and cycling conditions			
Component		**Volume**	**Final Concentration**
10x *Pfx* Amplification Buffer		5 µl	1 x
10 mM DNTp mixture		1.5 µl	0.3 mM each
50 mM $MgSO_4$		1 µl	1 mM
Forward primer (Table 5.1)		1 µl	2 pM
Reverse primer (Table 5.1)		1 µl	2 pM
BBK32-template for (BBK32-C)		1 µl	80 pg/µl
C-terminal PCR products N-terminal PCR products (for FL-BBK32)		1 µl 1 µl	
NanoPure water		39 µl	
Platinum *Pfx* polymerase		0.5 µl	1 unit
Cycling conditions			
94°C	35 Cycles	5 minutes	Pre-denaturation
94°C		15 seconds	Denaturation
63°C		30 seconds	Annealing
68°C		1.25 minutes	Extension
68°C		5 minutes	Final Extension

Table 5.2: *This table summarises the PCR master mix composition and the conditions for PCR amplification.*

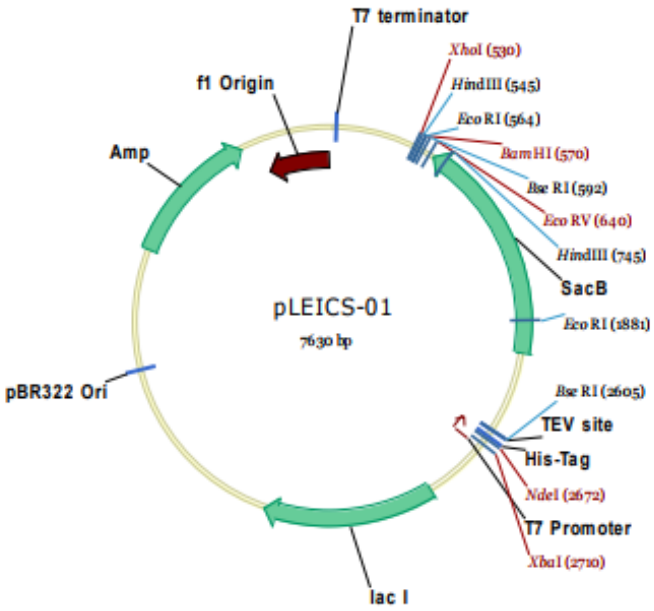


Figure 5-1: *This figure shows the vector map of pLEIC-01. BBK32 protein Was inserted in the EcoR1 site.*

5.2.2.3 Transformation and expression test of the FL and C-BBK32 constructs:

After cloning, the constructs were tested for protein expression. pLEICS-01 containing the BBK32 inserts were transformed into competent *E. coli* BL-21 (*DE3*) by transformation carried out as follows; 2µl of the recombinant DNA was added to 100µl of competent cells and incubated on ice for 30 minutes, followed by 2 minutes heat shock at 42ºC. The cells were then put back on the ice and then 300 ml of LB was added and incubated for 1 hour in a shaking incubator at 37ºC. Following this, the cells were plated on LB agar plates using a sterile L-shaped loop, containing 100 µg/ml of ampicillin and then incubated overnight at 37ºC. On the following day, one colony was used to inoculate 10ml of LB with 100µg/ml of ampicillin, and this was incubated until an OD_{600} of 0.4-0.6 was reached. Then, the culture was induced with IPTG at a final concentration of 1mM and incubated overnight at either 37ºC or 16ºC in a shaking incubator. The culture was harvested by centrifugation at 3396 g for 15 minutes. The cells were resuspended with 10µl BugBuster solution after being washed with PBS and incubated at room temperature for 15 minutes. After incubation, the cells were centrifuged, and the supernatant and pellet were analysed on a 15% SDS-PAGE gel. After investigating the expression test of BBK32 fragments, the optimal conditions for expressing were found to be incubated at 16ºC and induced with 1mM IPTG.

5.2.2.4 Large scale expression and purification of FL-BBK32 and C-BBK32:

The freshly transformed *BL-21*(DE3) cells were used to inoculate 500 ml flasks of LB broth medium containing 100µg/ml of ampicillin. The culture was grown in a shaking incubator at 37ºC until the OD_{600} reached 0.4-0.6, and then induced with IPTG (1mM). Then the cells were incubated overnight at 16ºC, in a shaking incubator. The cells were then harvested as described previously in section (4.2.3). The protein was expressed in the soluble fraction, therefore after centrifugation, the supernatant was collected and was kept on ice. The first step for purification

was carried out using Ni^{+2}- NTA Sepharose column (2ml). The column matrix was equilibrated with 20 ml of 50 mM Tris-pH7.5 containing 150 mM NaCl. The cell lysate was then diluted 1:1 with buffer and loaded onto the column at 4ºC overnight. The column was then washed with buffer containing 20 mM Imidazole to remove any non-specific binding to the column. Finally, the protein was eluted from the column in 1ml fractions with buffer containing 0.5 M imidazole. Fractions were then analysed on a non-reducing 15% SDS-PAGE gel to assess the purity and were further purified using gel filtration.

5.2.2.5 Purification by gel filtration chromatography:

A size exclusion Superdex 75/ 16/60 HiLoad column from GE Healthcare was used to purify BBK32 proteins. The column was equilibrated with filtered (50 mM Tris-pH 7.5and 150 mM $NaCl_2$). After the equilibration process was completed, 5 ml of the concentrated protein was injected into the loading loop. The sample was then run in a flow rate of 1ml/min. Fractions corresponding to the elution peaks were collected and analysed on a non-reducing SDS-PAGE gel. The protein elution was compared to an elution profile of standard molecular weights provided by the manufacturer. Collected fractions of the desired molecular weight were pooled together and concentrated to approximately 5mg/ml and stored at -80 ºC for further experiments.

5.2.2.6 SDS-polyacrylamide gel electrophoresis (SDS-PAGE):

Proteins were separated using BioRad mini protein II gel system. A 15% resolving gel was prepared as follows; 375mM Tris-pH8.8, 15% acrylamide, 0.1% SDS (sodium dodecyl sulphate), 0.1% ammonium persulfate and 0.0004% TEMED. The resolving gel was set in 1.5 mm gel cassettes and left to solidify under isopropanol. The stacking gel comprised; 125 mM Tris-pH6.8, 4% acrylamide, 0.1% SDS, 0.002% ammonium persulfate and 0.0004% TEMED. The stacking gel was cast on top of the resolving gel with a 1.5mm comb inserted and left to solidify. Samples were prepared by mixing 20µl of each sample fraction with 5µl of 6x loading dye comprising; 250 mM Tris-pH6.8, 50% glycerol, 10% SDS, 500 mM DTT and 0.25% bromophenol blue. Samples were then heated to 95ºC for 5 minutes to denature the proteins. After loading, the gel was ran at room temperature in an SDS running buffer containing; 25 mM Tris, 192 mM glycine and 1% SDS. Electrophoresis was carried out at 200v for ~50 minutes. The gel was the then stained for 30 minutes with Coomassie blue stain (0.4% Coomassie in 50% methanol and 10% acetic acid), washed with water and de-stained with 30% methanol containing 7% acetic acid.

5.2.3 Cloning and production of C1r FL and fragments CUB2, CUB2-CCP1, CUB2-CCP1-CCP2 and CUB2-CCP1CCP2-SP domains:

It has been previously shown that the BBK32 protein binds to the CP protease C1r, thereby blocking the activation of the system (Garcia *et al.*, 2016). However, the exact domain of C1r that binds to BBK32 has not been determined. To study this interaction, different C1r fragments were produced.

5.2.3.1 Expression the of C1r-CUB2 domain:

The purified plasmid pET-28a (Figure 5.2) containing the cDNA encoding the C1r CUB2 domain was kindly provided by (Professor Russell Wallis). The plasmid was transformed into competent *E.coli* cells BL-21 *DE3* as described in section 4.3.1.3. One colony of the transformed cells was used to inoculate 1L of Power Prime broth containing 1 ml of kanamycin (50µg/ml). The cells were then grown in baffled flasks in a shaking incubator at 37ºC. When an OD_{600} reading of 1.4-1.6 was achieved, the cells were then induced by adding 1 mM of IPTG and incubated overnight at 37ºC in a shaking incubator. Cells were harvested by centrifugation at 4000 g for 20 minutes at 4ºC. The supernatant was discarded, and the pellet was resuspended in 40 ml of phosphate-buffered saline (PBS). After verifying expression, the pellet was resuspended in 40 ml of lysis buffer (25 mM Tris-HCL pH 8.0, 150 mM NaCl, 0.5 mg/ml lysozyme, 0.5% (v/v) Triton X-100, 1 mM EDTA) containing one tablet of protease cocktail inhibitor (Roche). The suspension was then incubated for 30 minutes on a room temperature shaker until it became viscous due to release of genomic DNA. Then $MgCl_2$ (5 mM) and DNase (5 µg/ml) were added to digest the DNA. Then the mixture was incubated for further 15 minutes on a shaker at room temperature. Cells were then lysed by sonication on ice at an amplitude of 8 using a large probe; 8 to 9 pulses were applied for 30 seconds each, with a 1-minute break between each pulse to prevent overheating.

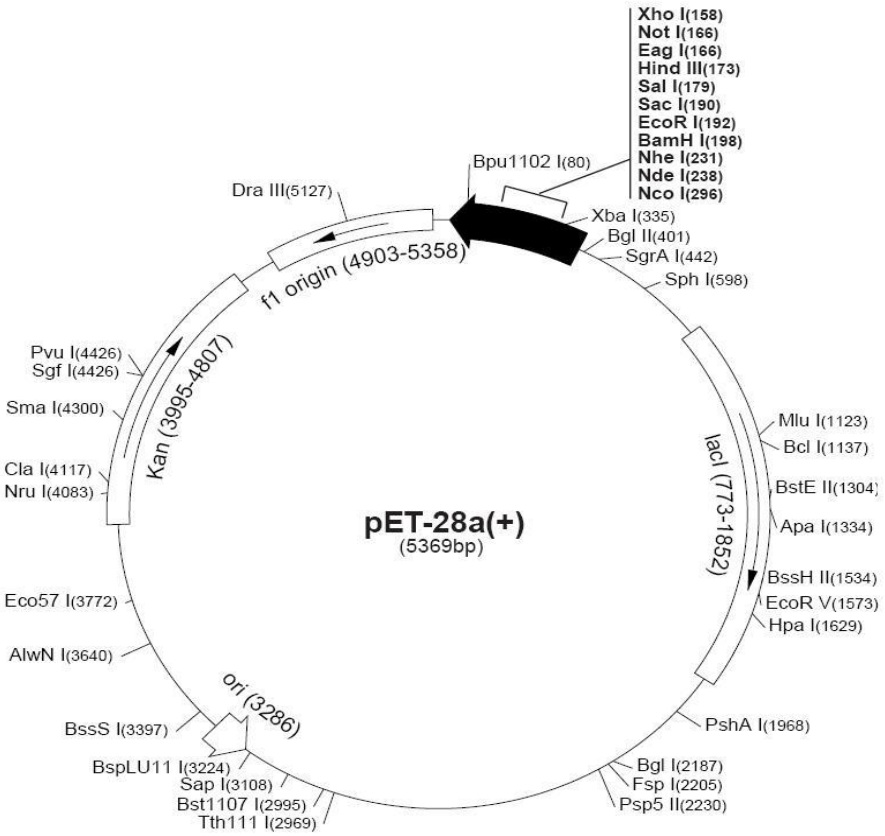


Figure 5-2: *pET-28a expression vector map. The C1r-CUB2 domain was inserted at the Sal1 site.*

Thereafter, the cell lysate was centrifuged at 20,000g for 20 minutes at 4ºC. The supernatant was discarded, and the pellet was washed in 40 ml of wash buffer (25 mM Tris-HCL pH 8.0, 0.5 M NaCl, 0.5% Triton X-100, and 1 mM EDTA). After resuspension, the mixture was sonicated for an additional 2-3 pulses to ensure full lysis of the cells. Then, the inclusion bodies were centrifuged at 20,000g for 20 minutes at 4ºC. After this, the resulting pellet was then resuspended with a 40 ml of (25 mM Tris-HCL pH 8.0, 0.5 M NaCl, 1 mg/ml Sodium Deoxy Cholate, 1 M Urea), and then centrifuged as before. The pellet was then washed with 40 ml of a 1:10 dilution of BugBuster and resuspended in 10 ml of 25 mM Tris-HCl pH 8.0 and snap frozen in liquid nitrogen after centrifugation and stored at -80ºC as a pellet aliquot. Aliquots

were collected throughout the purification procedure and checked on a reducing 15% SDS-PAGE gel to assess the purity of the inclusion bodies after each washing step.

5.2.3.2 Inclusion bodies solubilisation:

After the required purity was achieved, inclusion bodies were solubilised in (25 mM Tris-HCL pH 7.5 containing 8 M Urea and 5 mM DTT) which was added to disrupt any disulphide bonds. The mixture was then incubated in a water bath at 42ºC for 15 minutes and then centrifuged at 32869 g for 10 minutes to remove any insoluble cell debris. Finally, the supernatant was collected, and the protein concentration was measured at 280 nm using a Nanodrop 1000 spectrophotometer.

5.2.3.3 Refolding and protein recovery from solubilised inclusion bodies:

The solubilised inclusion bodies fractions were pooled, and the concentration was adjusted to 1mg/ml. Protein was refolded by drip dilution at 4ºC in refolding buffer containing (50 mM Tris-HCL pH 8.5, 240 mM NaCl, 10 mM KCL, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.4 M Sucrose, and 1 mM DTT) to reach a final concentration of 0.02 mg/ml. The refolding mixture was left overnight with gentle stirring. Following this, the refolded protein was dialysed in 20 mM Tris-HCL pH 8.5, containing 20 mM NaCl and 2 mM $CaCl_2$ using 10,000 cut-off dialysis tubing membrane, at 4ºC with gentle stirring. The dialysis buffer was changed three times every 4 hours and then left overnight. The dialysed protein was then concentrated using a Vivaflow 50 concentration system using a 10,000-cut-off membrane at 4ºC to a final volume of 5 ml. After concentration, the protein was further purified by gel filtration on a Superdex 75/ 16/60 HiLoad column in 50 mM Tris-pH 7.5 containing 150 mM $NaCl_2$ and 2 mM $CaCl_2$. Fractions were loaded on to a 15% non-reducing SDS-PAGE gel to assess purity. Fractions containing pure protein were pooled together, concentrated and stored at -80 ºC.

5.2.3.4 Expression and purification of FL-C1r protein:

A CHO cell line producing FL-C1r was kindly provided by Prof. Wallis. Cells contained the FL-C1r cDNA cloned into the expression vector pED4 with His_6 tag. Cells was grown in Minimal Essential Media α without nucleotides (MEMα -), containing 10 % dialysed, heat-treated foetal calf serum (DHFCS), 50 units/ml penicillin and 50 µg/ml streptomycin and 0.5µM methotrexate (MTX). Cells were grown in 25 cm^2 Nunc tissue culture flasks with filter caps and incubated at 37ºC and 5 % CO_2 until confluent. Following this the media was then removed, and the cells were washed with 2 ml phosphate buffered saline (PBS) pH 7.4, to remove any trypsin inhibitors present in the serum, then 1ml of trypsin-EDTA was added to the cells and then incubated at room temperature for 3 to 5 min, to detach the cells from the flasks. Then the resulting suspension was used to seed a fresh three-layers flasks used for protein production in 100 ml of media. Once fully confluent, the media was removed, and the cells were washed twice with 50 ml PBS, and then the media was replaced with 100 ml of (CHO-S-SFMII) without nucleosides containing 50 units/ml penicillin and 50 µg/ml streptomycin, 0.5µM MTX and 50 mM HEPES pH 7.55. The media was collected and replaced with a fresh 100ml of media every other day. The harvested media was collected and centrifuged at 671 g 4ºC for 2 min to remove any cell debris and stored at -20ºC until required for purification. Purification was carried out by affinity chromatography on a Ni^{2+}- NTA Sepharose column. 300 ml of the harvested media was mixed at a 1:1 ratio with 50 mM Tris-HCl pH 8.2 containing 150 mM $NaCl_2$ and 2 mM $CaCl_2$. The mixture was then loaded onto a 2ml Ni^{2+}- NTA Sepharose column and purified as described in section 2.2.2. Fractions were checked for purity on a non-reducing 15% SDS-PAGE gel, and then were pooled and run on a gel filtration column (Superdex 200 16/60) to remove minor contaminants. Protein was concentrated and stored at -80 ºC. The yield was typically 1.5mg from 1L of harvested media.

5.2.3.5 Cloning of C1r fragments CUB2-CCP1, CUB2-CCP1-CCP2 and CUB2-CCP1-CCP2-SP into pED4:

cDNA encoding C1r was kindly provided by (Prof. Wallis). Domains of interest were amplified using the PCR set up listed in (Table 5.3) with primers shown in (Table 5.4). The PCR products were separated by 1% agarose gel in TBE buffer and purified from an agarose gel using the gel extraction kit provided by QIAgen. The C1r fragments were introduced into pED4 expression vector through the *Sal1* site within the polylinker region of pED4. The pED4 vector was digested with *Sal1* by mixing 36µl of pED4 with 3µl *Sal1* and 10µl cut-smart buffer and 51µl H_2O. The mixture was then incubated overnight at 37ºC. The products were then separated on a 1% agarose gel and purified using the gel extraction kit (QIAgen).

PCR mixture and cycling conditions			
Component		**Volume**	**Final Concentration**
10x *Pfx* Amplification Buffer		5 µl	1 x
10 mM DNTp mixture		1.5 µl	0.3 mM each
50 mM $MgSO_4$		1 µl	1 mM
Forward primers 1and 2 (Table 5.4)		1 µl	2 pM
Reverse primer (Table 5.4)		1 µl	2 pM
C1r-template		1 µl	80 pg/µl
NanoPure water		39 µl	
Platinum *Pfx* polymerase		0.5 µl	1 unit
Cycling conditions			
94Cº	35 Cycles	5 minutes	Pre-denaturation
94Cº		15 seconds	Denaturation
64Cº		30 seconds	Annealing
72Cº		1 minutes	Extension
72Cº		5 minutes	Final Extension

Table 5.3: *Summary of the PCR mixture and conditions for C1r amplification.*

Primer	Sequence 5` ------- 3`
C1r-CUB-2 FP1	AGGTCCAACTGCAGGGCCACCATGTACAGGATGCAACTCCTGTCTTGCA TTGCACTAAGTCTTGCACTTGTCACGAATTCG
C1r-CUB-2 FP2	CTTGCACTTGTCACGAATTCGGCC**CACCACCATCACCATCAT**GCC
C1r-CUB2-CCP1 R	TCCCCGGGTCTAGAGTTCAGTCCTTGATCTTGCATCTGGGCATGGC
C1r-CUB2-CCP1-CCP2 R	TCCCCGGGTCTAGAGTTCACACTGGCAAGCACCGAGGAATCTTCTC
C1r-CUB2-CCP1-CCP2-SP R	TCCCCGGGTCTAGAGTTCAGTCCTCCTCCTCCATCTCTTTCTT

Table 5.4: *Sequence list of the primers used to generate C1r fragments. Start and stop codon are highlighted in red, an optimised Kozak sequence was added to increase the level of expression and is underlined, the His_6 tag was added to enable purification with Ni^{2+} and is shown in bold.*

The cloning was carried out by recombination using the Seamless cloning enzyme from Invitrogen using 50ng of cut vector + 200ng of each insert. The cloning reaction mixtures were incubated at 50ºC for 20 minutes and then transformed into *E. coli* XL-10 cells and then incubated overnight at 37ºC. A single colony was picked for each clone, and these were grown in 10ml LB containing 100µg/ml ampicillin. Plasmid was isolated after overnight growth using a mini-prep kit provided by (Promega), and the presence of the inserts was confirmed by restriction digestion with *EcoR1* enzyme. The constructs were then sequenced by the PNACL facility at University Leicester.

5.2.3.6 Expression vector pED4:

The mammalian expression vector pED4(Figure 5.3), contains two key components, the first one is the polylinker region in which the C1r gene fragments were inserted, and the second component is the marker dihydrofolate reductase (DHFR) gene which is expressed with the cloned gene as dicistronic m-RNA (Kaufman *et al.*, 1991). The DHFR gene can be selected for in a deficient CHO DXB11 cell line, which requires the growth media to be supplemented with nucleotides. Upon transfection, only cells that contain the plasmid would have the ability to grow in the absence of external nucleotides. The expression efficiency can be further increased by the growth in an increased concentration of a DHFR inhibitor methotrexate (MTX), up to 0.5µM.

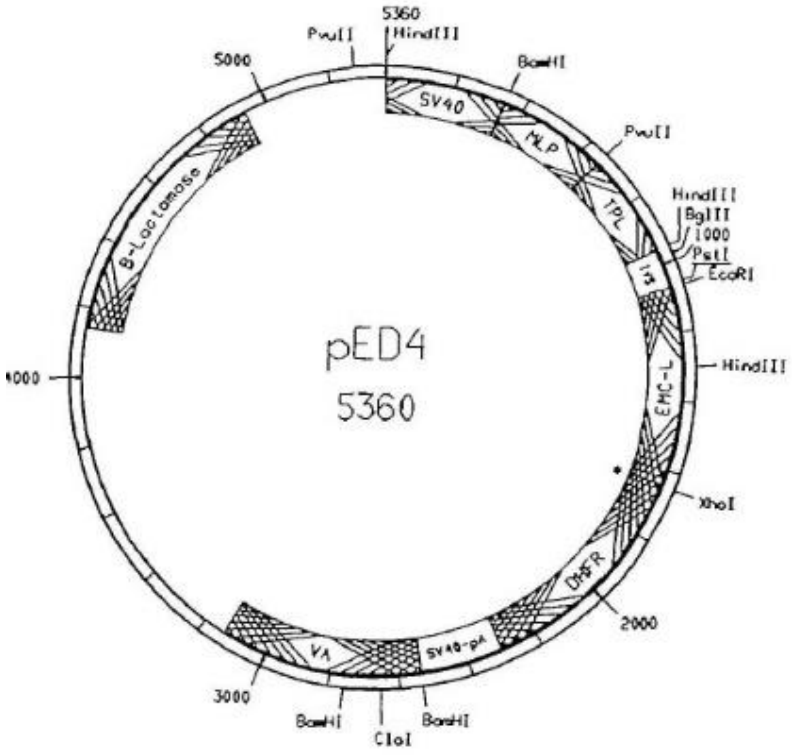


Figure 5-3: *Vector map of pED4, the C1r inserts are cloned into the Sal1/EcoR1 polylinker (Kaufman et al., 1991).*

5.2.3.7 Preparation of C1r DNA constructs and CHO DXB11 cells for transfection:

In preparation for transfection, the DNA was precipitated by mixing 1:10 of the total volume of DNA with 3M sodium acetate pH5.2, and two volume of 100% chilled ethanol. The mixture was then incubated on dry ice for 15 minutes and then was centrifuged for 15 minutes at 28341 g to pellet the DNA. The pellet was then washed with 70% ice cold ethanol and left to dry for 20-30 minutes under a sterile condition in tissue culture hood. CHO DXB11 cells were used for transfection and expression of C1r fragments. Cells were grown in Minimal Essential Medium containing nucleotides (MEMα+) supplemented with 10% DHFBS, Penicillin/Streptomycin at a final concentration of 500 units/ml and 500ug/ml respectively. The addition of nucleotides is vital for the survival of the cells prior to transfection. The cells were grown till confluent at 37ºC in a tissue culture incubator with 5% CO_2.

5.2.3.8 Transfection of CHO DXB11 cells:

Transfection was carried out using the calcium phosphate protocol, which was first described by (Graham and van der Eb, 1973). In this procedure, the recombinant plasmid is precipitated in a calcium phosphate buffer allowing it to be taken up by the cells via endocytosis. Briefly, 100 µl of 50 µg/ml of recombinant DNA and 100 µl of calf-thymus DNA at 10mg/ml were diluted in (1ml sterile ddH_2O with 120 µl of 2 M $CaCl_2$). The mixture was then added to a 15ml falcon tube containing 1ml of 2X HEPES buffered saline (150 mM NaCl and 20 mM HEPES-pH 7.4) (HBS) and 40µl of 100X phosphate buffer (70 mM Na_2HPO_4 and 70 mM NaH_2PO_4). The mixture was then incubated at room temperature for 30 minutes to allow precipitation. The precipitated mixture was then added to a 70% confluent CHO DXB11 cell line that was grown in MEMα+ medium containing 10% DHFBS and P/S in 25cm^2 tissue culture flasks. The transfected cells were then incubated overnight at 37ºC with 5% CO_2. The following day, media was replaced with fresh MEMα+ to allow cells to recover before selection, and then incubated

overnight. Once the cells had successfully uptake the recombinant plasmid, they would not require the added nucleotides and would be able to survive in MEMα- due to the function of DHFR gene. Cells were trypsinized and transferred to a sterile culture tissue culture dish with 3 ml of MEMα- supplemented with 10% DHFBS and P/S. Further selection and a high level of expression were achieved by adding an increased concentration of methotrexate (MTX) to up to 0.5µM. MTX inhibits DHFR which causes the cells to increase DHFR gene expression. Because pED4 expresses the gene of interest as dicistronic mRNA, increased expression of DHFR will increase the expression of the target gene.

5.2.3.9 Expression and purification of the transfected C1r fragments:

After transfection cells were grown to confluence and were passaged into a triple layer tissue culture flask for harvesting as described previously in section 2.2.2. The cells were grown till confluent in MEMα- media supplemented with 10% DHFBS, 0.5 µM MTX and P/S. The cells were then washed with 50ml PBS, and the media was replaced with 100ml of CHO-S-SFM media without nucleotides containing 50 mM HEPES pH7.55, 0.5 µM MTX and P/S. The media was harvested every two days and spun at 377 g to remove any cell debris and was then stored at -20ºC. C1r fragments were purified by affinity chromatography on a Ni^{2+}- NTA Sepharose column as described previously. Proteins were further purified by gel filtration in 50 mM Trizma-pH7.5, containing 150 mM NaCl and 2 mM $CaCl_2$ on a Superdex 200 16/60 column. The proteins were then concentrated to 2mg/ml and stored at -80ºC.

5.2.3.10 Transfected cell storage:

Transfected C1r cell line was stored by mixing 1ml of the trypsinized cell with 2ml of (DHFBS containing 10% DMSO). The cells were separated into two cryovials with 1ml each and were then snap frozen in dry ice and then were stored at -150ºC.

5.2.4 BBK32 crystallisation and optimisation:

For crystallisation, BBK32 was concentrated to 5mg/ml. Sitting drops were set up in 96 well MRC crystallisation plates using a Mosquito NanoDrop crystallisation robot. Commercial screens were purchased from Molecular Dimensions. Five screens were tested: JCSG, ProPlex, Morpheus, BCS and PACT. Briefly, 80µl of each screen was transferred into each plate reservoir, and 100nl of the reservoir buffer was mixed with 100nl of protein. Two plates were set up for each screen, one at room temperature and one at 4ºC. Plates were left for one week and then were checked using a microscope for a possible hit. After two weeks, BBK32 crystals were observed in the PACT and ProPlex screens in (0.2 M Potassium sodium tartrate tetrahydrate, 20 % w/v PEG 3350) and (0.2 M ammonium acetate, 0.1 M sodium acetate pH 4.0, 15 % w/v PEG 4000) respectively in room temperature plates. The conditions for each hit were further optimised by setting up large sitting drops in 48 well MRC plate, using 1.5µl of the screen condition and 1.5µl of the protein. A range of the precipitant and buffer concentrations were set up to improve crystal quality, as shown in (Table 5.5). All optimisation buffers were made by hand, and the 48 well plates were incubated at room temperature. After optimising precipitant and buffer concentration, an additives screen containing a library 96 different molecules were tested to further enhance crystal quality. Stock solutions of the optimised conditions were made to final concertation of 1.1M, and then 72µl of the stock solution was mixed with 8µl of each additive in the reservoir well. The protein was then mixed with the reservoir buffer using Mosquito NanoDrop crystallisation robot. BBK32 crystals grew in 0.2 M ammonium acetate, 0.2 M sodium acetate pH4.0, 20% PEG 4000, containing 3% ethylene glycol.

Conditions and buffers optimisation for BBK32 crystallisation screen.	
1	0.2 M Potassium sodium tartrate tetrahydrate, 25%, 24%, 23%, 22%, 21%, 20% w/v PEG 3350
2	0.2 M ammonium acetate, 0.1 M sodium acetate pH 4.0 20%, 19%, 18%, 17%, 16%, 15 % w/v PEG 4K
3	24% w/v PEG 3350, Potassium sodium tartrate tetrahydrate 0.6 M, 0.5 M, 0.4 M, 0.3 M, 0.2 M, 0.1 M
4	20% w/v PEG 4K, 0.2 M ammonium acetate, 0.6 to 0.1 M range of sodium acetate pH 4.0.
5	20% w/v PEG 4K, 0.2 M sodium acetate pH 4.0, 0.6 to 0.1 M range of ammonium acetate.
6	20% w/v PEG 4K, 0.2 M sodium acetate pH 4.0, 0.2 M ammonium acetate, 8 to 3% range of ethylene glycol.

Table 5.5: *Summary of the crystallisation optimisation conditions for BBK32 protein. Each buffer was used to set up a 48 well plate.*

5.2.4.1 X-ray diffraction for BBK32 crystals:

Crystals were picked by adding a cryoprotectant (crystallisation buffer containing 30% glycerol) scooping with appropriately sized LithoLoops and frozen in liquid nitrogen. Diffraction data were collected at 100K in beamline I04 at Diamond light source.

5.2.5 C-BBK32/C1r-CUB2CCP1 binding assay using gel filtration:

Equimolar concentrations of the C-BBK32 and C1r-CUB2-CCP1 were mixed and incubated for 20 min at room temperature and then loaded onto an analytical gel filtration column (Superdex 200 10/300). Separate samples of each protein were loaded onto the column at the same concentration to indicate the elution positions of the free components. The final concentration of all proteins on all occasions was 70 µM. Buffer containing 50 mM Tris-pH7.5, 150 mM NaCl and 2 mM $CaCl_2$ was used to dilute the proteins and equilibrate the gel filtration column. The flow rate used to elute the proteins from the analytical gel filtration column was 0.5 ml/min.

5.2.6 BBK32/FL-C1r binding assay using BLI (Bio-Layer Interferometry):

The binding was carried out using amine coupling and streptavidin/biotin sensor tips as described previously in section 4.2.8. C-BBK32 was immobilised in the sensor tip in 50 µM concentration in buffer containing 50 mM HEPES-pH 7.5, 150 mM NaCl and 2 mM $CaCl_2$. The FL-C1r was then injected over the immobilised C-BBK32 at the following concentrations (25, 50, 100, 200, 400, 800 µM). To insure the correct set up of the experiment, mannan binding lectin (MBL)/MASP-2 was used as a positive control (Thielens *et al.*, 2001).

5.3 Results:

Previous work has shown that BBK32 from *Borrelia burgdorferi* interacts with the serine protease C1r and inhibits its ability to cleave its substrate C1s, thus preventing activation of the CP of complement (Garcia *et al.*, 2016). It has been recently reported that the location of the binding site on C1r is within the SP domain (Xie *et al.*, 2019). To investigate the interaction between C1r protease and *Borrelia burgdorferi* surface protein BBK32, I produced full length (FL) and C-terminal BBK32 proteins and several fragments of the serine protease C1r encompassing the following domains; CUB2-CCP1, CUB2-CCP1-CCP2, CUB2-CCP1-CCP2-SP together with the full-length (FL) protease. I further tested the interaction between BBK32 and C1r fragments using BLI and gel filtration. BBK32 protein was also crystallised to determine its structure alone and in complex with C1r.

5.3.1 Cloning, expression and purification of FL and C- terminal BBK32 protein:

5.3.1.1 PCR amplification of FL and C-terminal BBK32:

The BBK32 constructs were amplified from two synthetic gene fragments encompassing the 5’ and 3’ regions of the BBK32 gene (purchased from Invitrogen). The FL-BBK32 was constructed in two overlapping segments using primers listed in (Table 5.3). The two PCR products were separated on a 1% agarose gel (Figure 5.4)

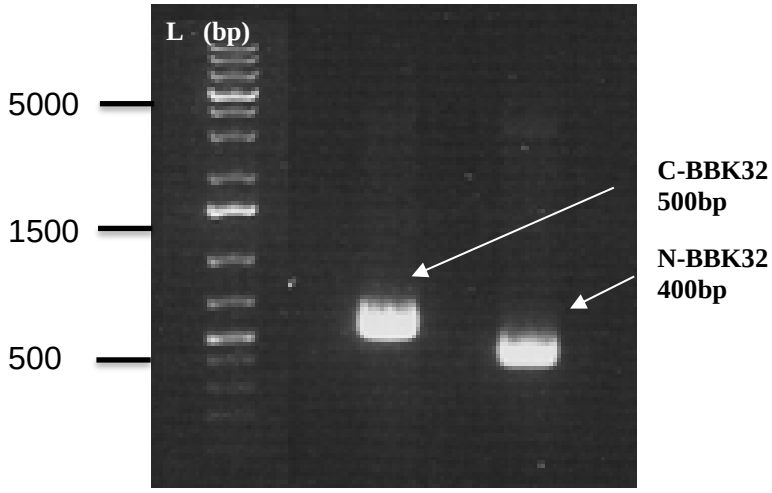


Figure 5-4: *1% agarose gel showing the 5' (N-BBK32) and 3' (C-BBK32) fragments. Bands of the expected sizes 500bp for C-BBK32 and 400bp for N-BBK32 were observed.*

After the successful amplification of the two gene fragments, the PCR products were combined and used as a template to generate the FL-BBK32 gene. Following the PCR, a single fragment was observed on a 1% agarose gel of the expected size (1100 bp) (Figure 5.6). The fragment was purified from the gel and supplied to the PROTEX cloning facility for cloning into expression vector pLEIC-01. The sequence was confirmed by DNA sequencing. The C-terminal fragment of BBK32 was cloned in the same way from a single PCR product.

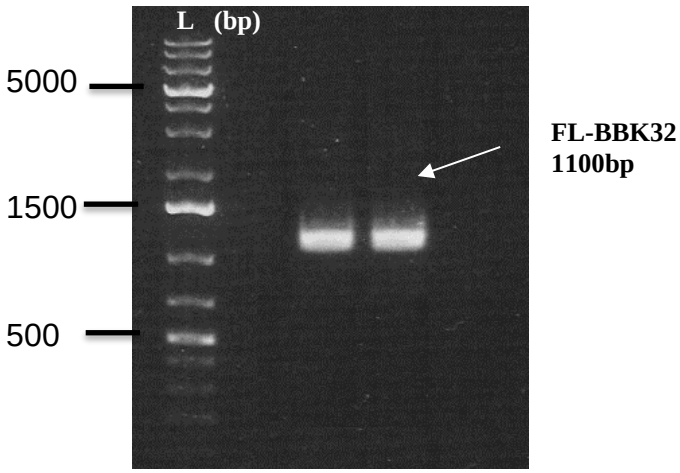


Figure 5-5: *1% agarose gel showing the successful PCR amplification of the FL-BBK32 gene with an expected size of 1100bp.*

5.3.1.2 Expression and purification of C-BBK32:

The FL and C-terminal BBK32 were expressed in *E. Coli* BL-21 (*DE3*). Expression was tested initially at 37 and 16ºC as described in section 5.3.1.3; the optimal expression condition for BBK32 was at16ºC (Figure 5.6).

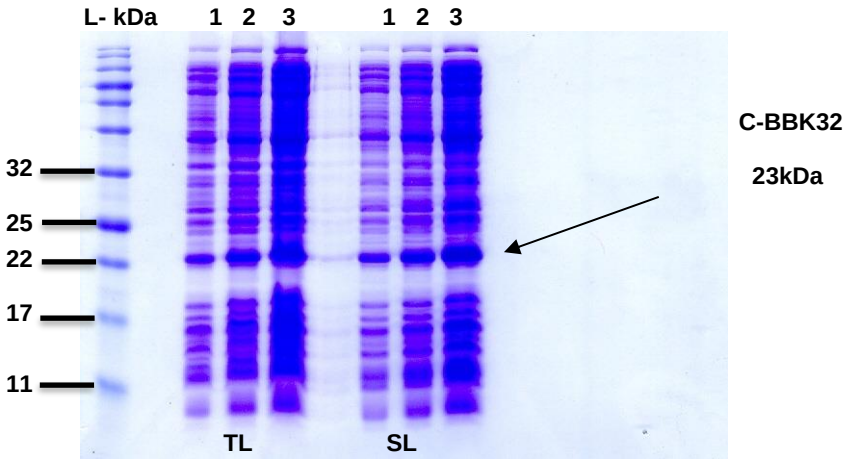


Figure 5-6: *15% non-reducing SDS-PAGE for testing the expression of C-BBK32. The protein migrated at the expected size 23kDa and was present in both total lysate (TL) on the left and in the soluble lysate (SL) on the right, 1,2 and 3 indicate different loading volumes (1 µl, 2 µl and 5 µl).*

Small scale expression tests indicated that C-BBK32 eluted in the soluble fractions, as seen in (Figure 5.7). Protein was then expressed from a larger scale (1L culture) and then purified by affinity chromatography on a Ni^{+2}- NTA Sepharose column. The elution fractions were loaded on a 15% non-reducing SDS-PAGE (Figure 5.7). The eluted C-BBK32 migrates at the expected size of ~23kDa.

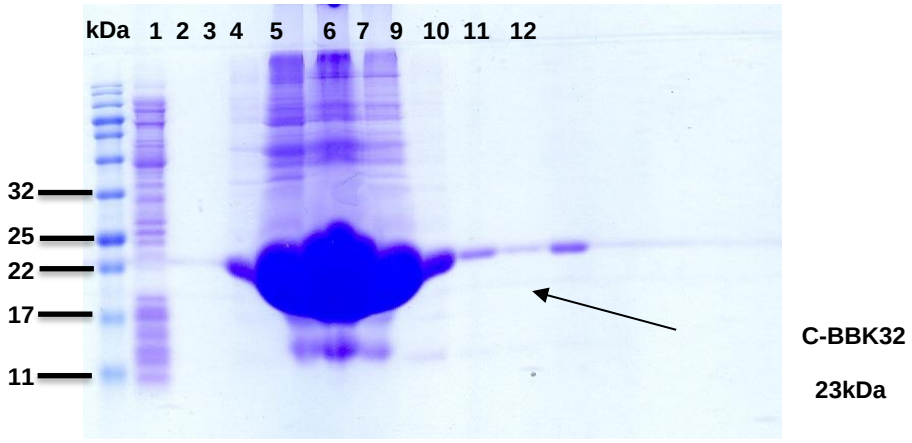


Figure 5-7: *Non-reducing SDS-PAGE of the Ni^{+2}- NTA Sepharose purification of C-BBK32. Lane 1 indicates the flow-through and lane 2 shows the non-specific wash with 10mM Imidazole. Lanes 3 to 12 shows the eluted fractions. The large band corresponds to C-BBK32 which migrated at the expected size of ~23kDa.*

The eluted fractions from the Ni^{2+}- NTA column were pooled together and then loaded on a Superdex 75 16/ 60 gel filtration column for further purification (Figure 5.8).

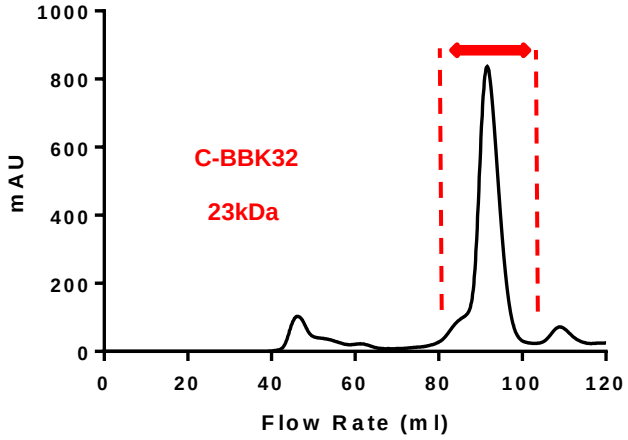


Figure 5-8: *Gel filtration elution profile for C-BBK32.*

The C-BBK32 protein eluted from the column at a molecular weight of 23kDa, indicating that it is a monomer in solution. Fractions across the peak were loaded onto a non-reducing 15% SDS PAGE gel (Figure 5.9). The final protein was of high purity with ~6 mg of protein obtained from 1L of culture.

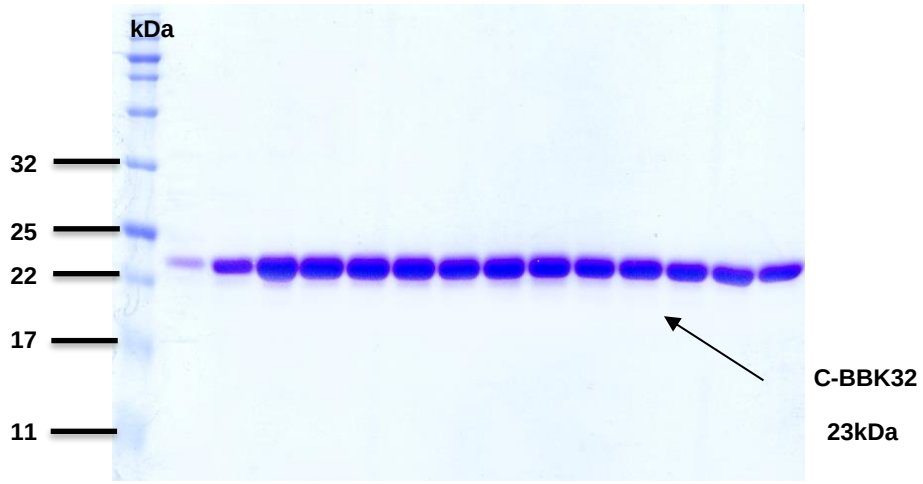


Figure 5-9: *SDS-PAGE analysis of the gel filtration elution fractions for C-BBK32. The protein eluted from the column with a molecular mass of 23kDa.*

The C-BBK32 protein was synthesised with an N-terminal His_6 tag to enable purification on a Ni^{2+}- NTA Sepharose column. Prior to crystallisation and binding trials, the His_6 tag was removed by digestion with TEV enzyme. Each 500ul of the eluted protein was treated with 1 unit of TEV enzyme and then re-loaded onto a Ni^{2+}- NTA Sepharose column to remove any un-cleaved protein (Figure 5.10). The digested protein was then loaded on a Superdex 75 16/60 gel filtration column to remove any impurities and was then analysed by a non-reducing 15% SDS-PAGE gel to assess the digestion quality (Figure 5.11). The digested protein was then stored at -80ºC for further analysis.

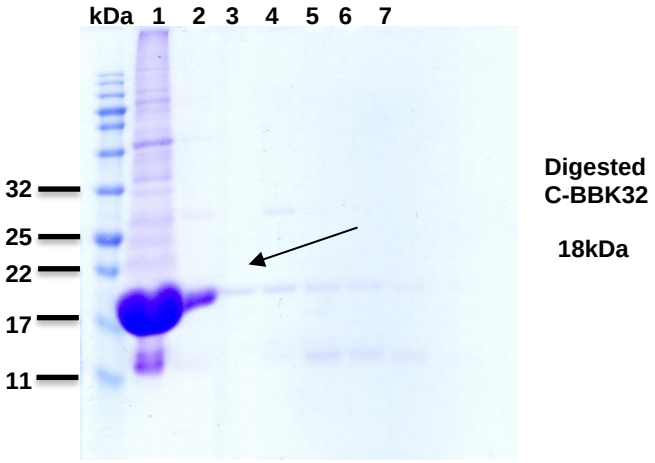


Figure 5-10: *Ni^{2+}- NTA purification of the TEV-digested BBK32 protein analysed by SDS-PAGE. Most of the protein eluted in the flow-through (lane1) with a molecular mass of 18kDa, indicating successful digestion and removal of the his_6 tag. Lane 2 shows the low imidazole wash, and lanes 3 to 7 are the elution fractions.*

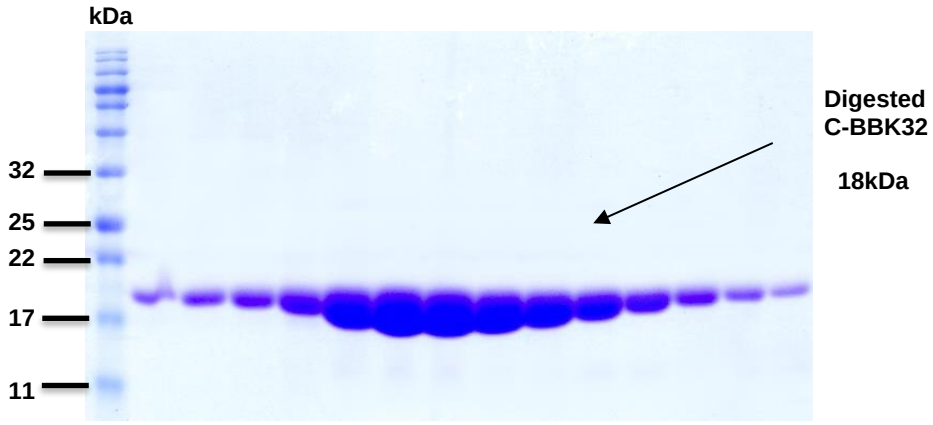


Figure 5-11:*15% SDS-PAGE analysis of digested C-BBK32 following gel filtration. C-BBK32 migrates with a molecular mass of 18kDa confirming successful digestion of the His_6 tag.*

5.3.1.3 Expression and purification of the FL-BBK32 protein:

The FL-BBK32 was expressed in *E. Coli* BL-21 (*DE3*) strain following the same expression conditions tested previously for the C-BBK32 protein. 1L of the harvested cells were lysed, and the supernatant was loaded on to a Ni^{2+}- NTA Sepharose column. Eluted fractions were analysed on a non-reducing 15% SDS-PAGE gel to check the purity of the harvested protein (Figure 5.12).

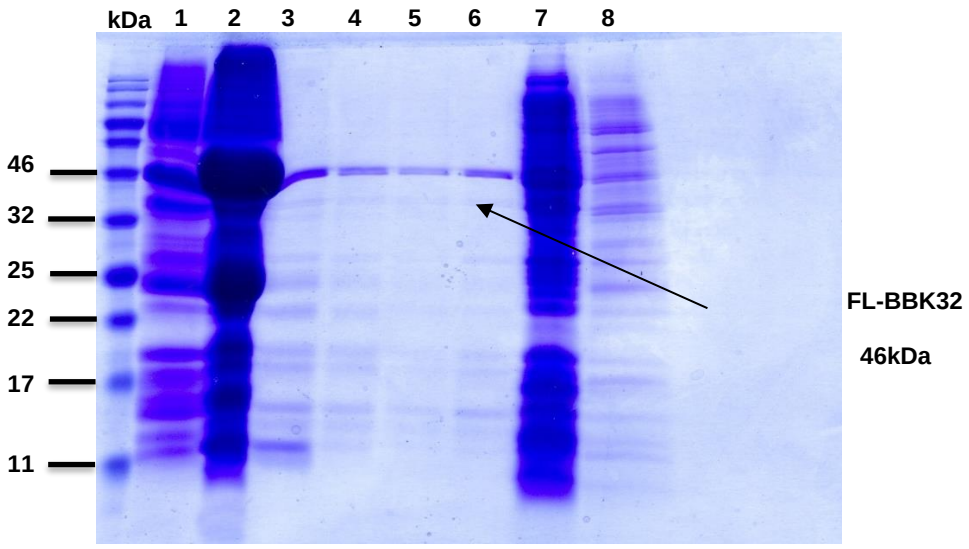


Figure 5-12: *SDS-PAGE of the Ni column purification of the FL-BBK32. Lanes 1to 6 represents the elution fractions, and lanes 7 and 8 show the flow throw and low Imidazole wash respectively.*

FL-BBK32 migrated with a molecular mass of 46kDa, as expected. The eluted fractions were pooled together and were subjected to additional purification via gel filtration on a Superdex 200 16/60 column (Figure 5.13) to remove any impurities. The harvested protein eluted at the expected position for a monomer ~46 kDa, based on the elution positions of molecular mass standards.

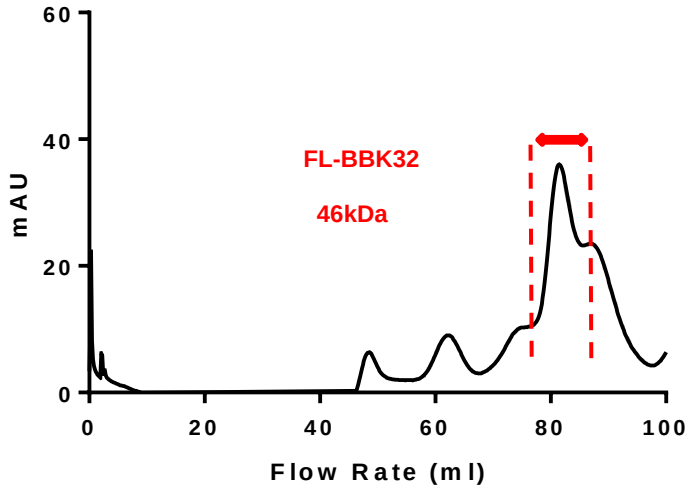


Figure 5-13: *Gel filtration elution profile for the FL-BBK32. The elution peak indicated between the red arrows represent the FL-BBK32 protein which eluted at the expected position for a monomer.*

Fractions from the target peak were analysed by non-reducing SDS-PAGE on a 15% gel to assess their purity (Figure 5.14).

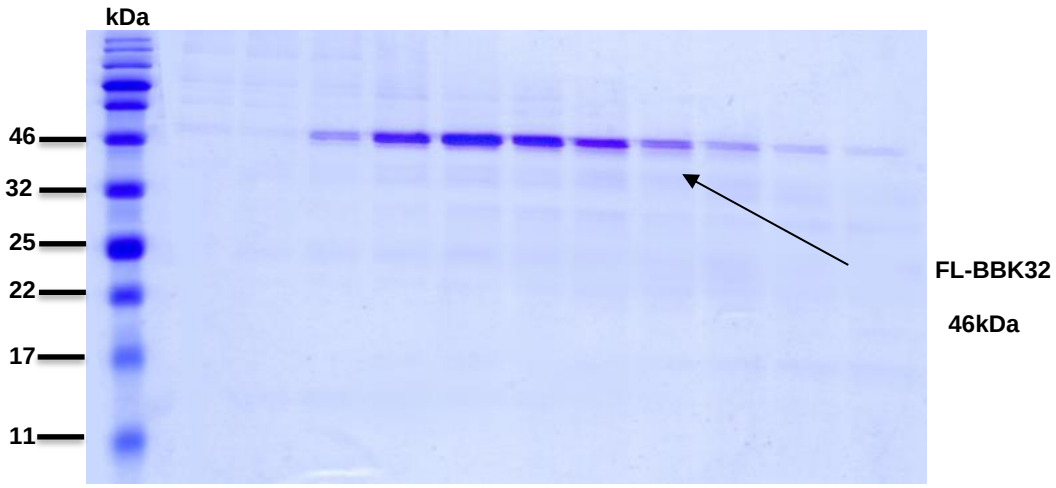


Figure 5-14: *SDS-PAGE analysis of the gel filtration elution fractions. The protein migrates on the gel as expected with a molecular weight of 46kDa of a monomer.*

The final preparation of FL-BBK32 was of high purity with ~1.7mg of protein from 1L of culture.

5.3.2 Expression and purification of the FL-C1r protein:

The FL-C1r protein was expressed in a CHO mammalian cell line. Serum free medium was harvested from producing cells and protein was purified by affinity chromatography on a Ni^{2+}-NTA Sepharose column. The eluted fractions were then analysed by 15% non-reducing SDS-PAGE gel to assess the cell expression and protein purity (Figure 5.15).

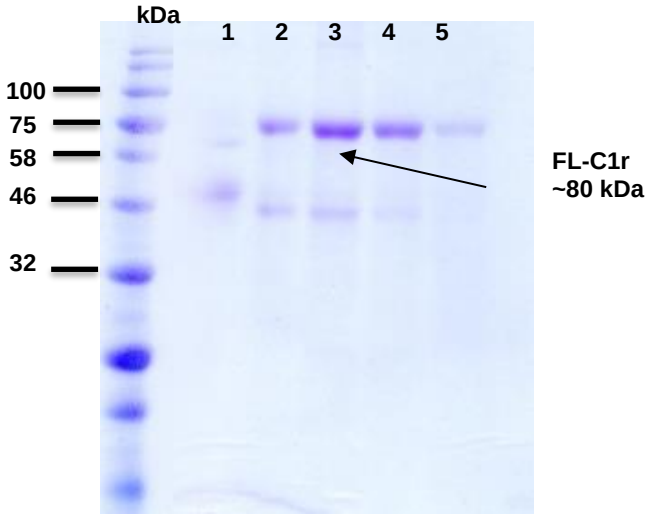


Figure 5-15: *SDS-PAGE analysis following Ni^{2+}- NTA Sepharose purification of the harvested FL-C1r protein. Lane 1 shows the media flow through, and lane 2 represents the non-specific wash with low Imidazole. Lanes 3 to 5 shows the elution fractions. The eluted protein migrated on the gel with a molecular mass of ~80 kDa.*

Fractions from the Ni^{2+}- NTA column were pooled and loaded on a Superdex 200 16/60 gel filtration column (Figure 5.16). FL-C1r migrated as a dimer on gel filtration with an apparent molecular mass of ~150 kDa (Figure 5.17). Previous studies have shown that dimerization of C1r is mediated through interactions between the CCP domains of one C1r monomer and the SP domain of its partner (Sharp *et al.*, 2019).

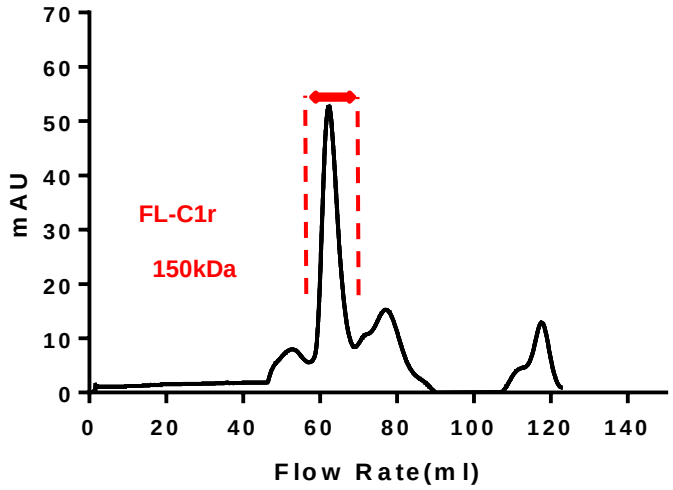


Figure 5-16: *Gel filtration elution of FL-C1r on a Superdex 200 16/60 column.*

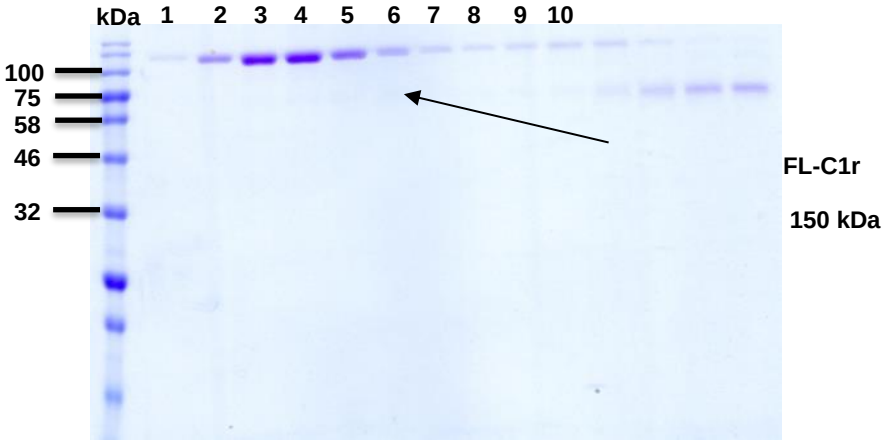


Figure 5-17: *15% non-reducing SDS-PAGE analysis of the gel filtration eluted fractions. The FL-C1r protein migrated on the gel with the expected molecular mass of 150kDa. Fractions 1 to 10 were pooled and concentrated.*

SDS-PAGE analysis revealed that the additional gel filtration purification step removed all minor contaminants. The protein yield was 1.2 mg from 1L of harvested media.

5.3.3 Expression and purification of the C1r-CUB2 domain:

The cDNA of interest was cloned into pET-28 expression vector and CUB2 domain of C1r was purified from inclusion bodies. After each wash step, 20 µl aliquot was collected and checked on a reducing 15% SDS-PAGE gel to determine the purity of the inclusion bodies before solubilisation and refolding (Figure 5.18).

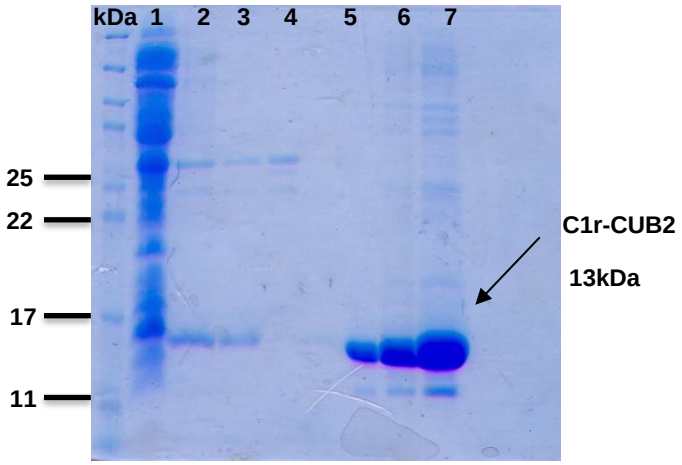


Figure 5-18: *Reducing SDS-PAGE on a 15% gel of the inclusion bodies. The lanes 1 to 4 shows the wash steps and lanes 5,6 and 7 showed three different loading amounts of the inclusion bodies (1µl, 2 µl and 5 µl). The purified inclusion bodies migrated with a molecular mass of 13kDa, consistent with the mass of the CUB2 domain.*

The CUB2 domain was refolded by drop dilution at 4ºC. Although yields were relatively low (0.6 mg from 2L of culture) refolded protein was isolated by gel filtration on a Superdex 75 16/60 gel filtration column (Figure 5.19). The CUB2 domain eluted from the column at the expected position of a monomer.

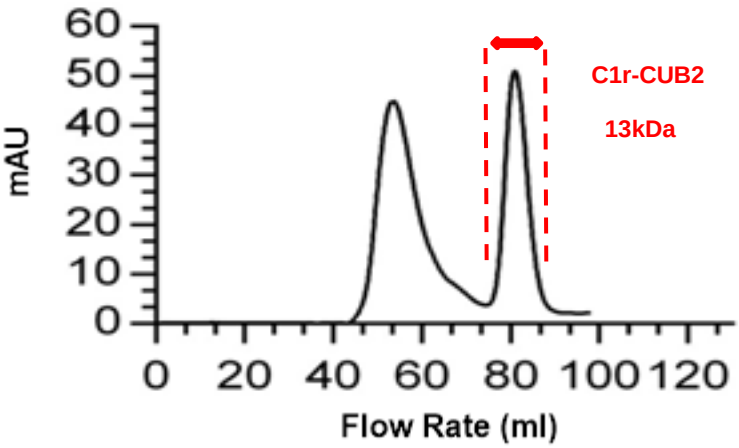


Figure 5-19: *Gel filtration elution profile of C1r-CUB2. The first peak to elute from the column was the aggregated protein, and the peak indicated between the red arrows corresponds to the correctly folded C1r-CUB2.*

The eluted fractions from the column were then analysed by SDS-PAGE gel to assess the purity (Figure 5.20).

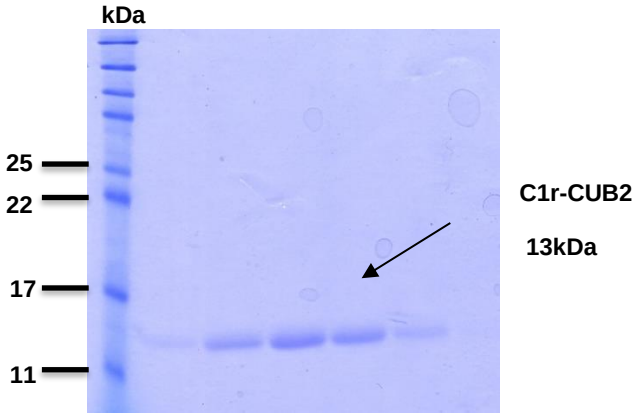


Figure 5-20: *15% non-reducing SDS-PAGE analysis of the gel filtration elution fractions of the C1r-CUB2 domain. The C1r-CUB2 protein eluted from the column with the expected molecular mass of 13kDa with reletivley high yeild.*

5.3.4 Cloning of the C1r-CUB2-CCP1, C1r-CUB2-CCP1-CCP2 and C1r-CUB2-CCP1-CCP2-SP constructs:

5.3.4.1 Cloning and amplification of the C1r constructs:

The additional C1r fragments were amplified by PCR (Figure 5.21) and cloned into the mammalian expression vector pED4 by recombination.

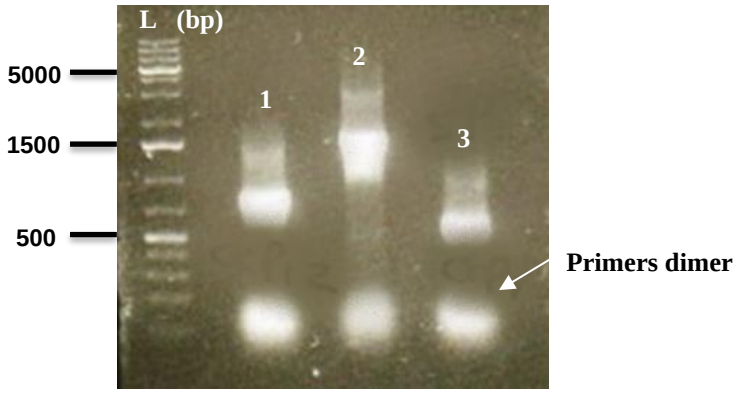


Figure 5-21: *1% agarose gel showing amplification of the C1r fragments by PCR. Product 1 is the fragment encoding the C1r-CUB2-CCP1-CCP2 domains (~780bp), product 2 encodes the C1r-CUB2-CCP1-CCP2-SPdomains (~1500bp), and product 3 encodes the C1r-CUB2-CCP1(~550bp) domains.*

Clones for the three different constructs were identified by digestion with *EcoR*1, which removes the insert (Figure 5.22).

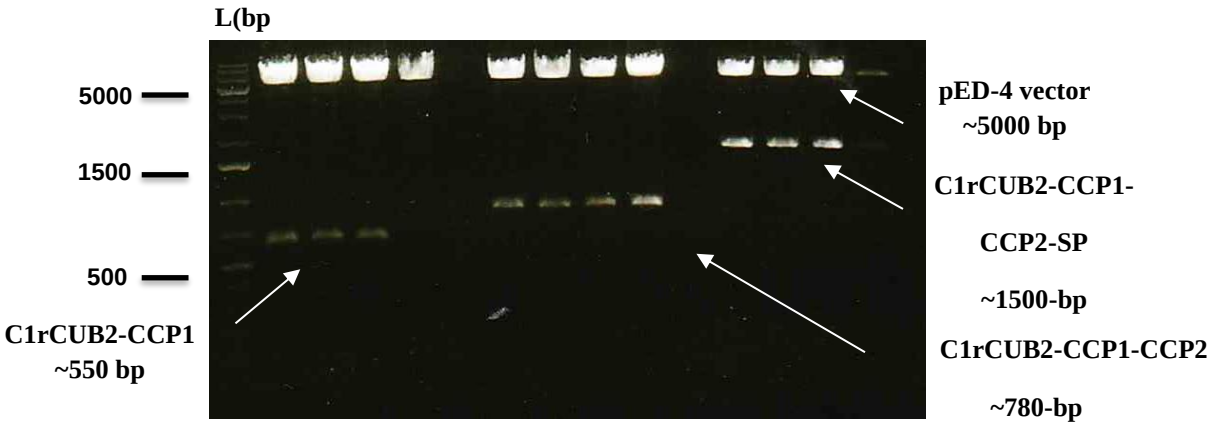


Figure 5-22: *EcoR1conformation digestion for each clone which shows two different sizes corresponding to the pED4 vector at ~5000bp and the size of the correct inserts.*

All clones were sequenced to confirm the correct insert.

5.3.4.2 Expression and purification of C1r constructs: CUB2-CCP1, CUB2-CCP1-CCP2 and CUB2-CCP1-CCP2-SP:

After successful transfection of the recombinant DNA into the CHO cell line, the cells were harvested in SFM and then purified by affinity chromatography in a Ni^{2+}- NTA Sepharose column. The eluted fractions were analysed on a 15% non-reducing SDS-PAGE gel to check for expression and assess purity (Figure 5.23).

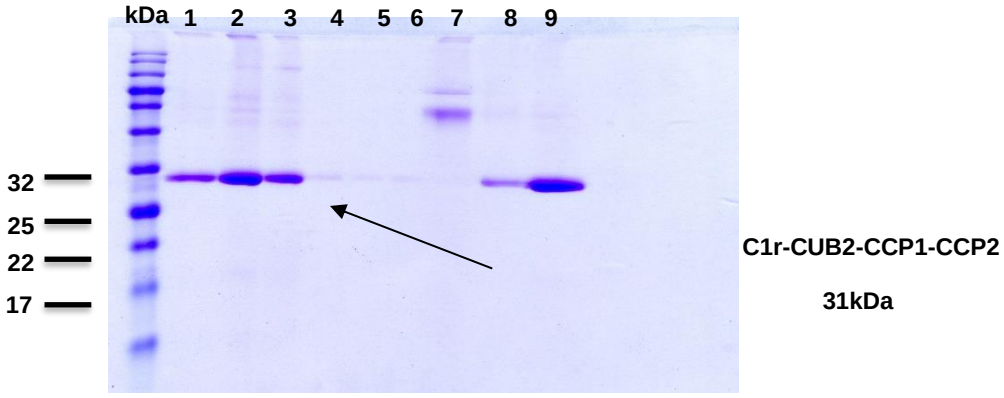


Figure 5-23: *SDS-PAGE analyses of the Ni^{+2}-NTA column elution for C1r-CUB2-CCP1-CCP2. Lanes 1 to 6 represent the elution fractions, 7 is the flow-through, 8 is the no imidazole wash and 9 the low imidazole wash. The protein migrated in the expected size of 31kDa.*

Protein was purified further by loading the elution fractions on gel filtration on a Superdex 200 16/60 column (Figure 5.24).

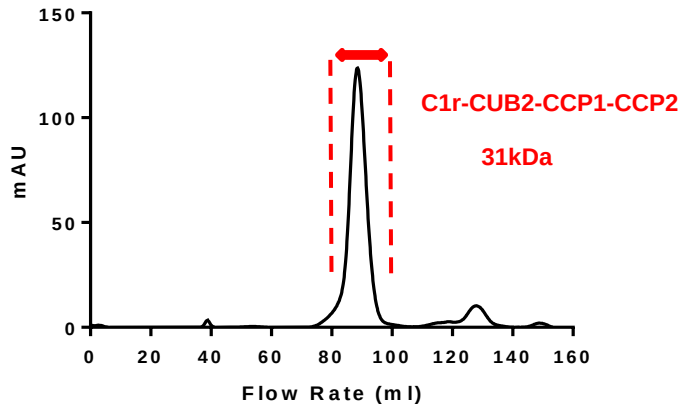


Figure 5-24: *Gel filtration purification of C1r-CUB2-CCP1-CCP2 protein.*

Protein eluted from the gel filtration as a monomer with a molecular mass of ~31kDa and was pure based on SDS-PAGE gel (Figure 5.25). Yields were ~1.5mg from 1L of culture medium.

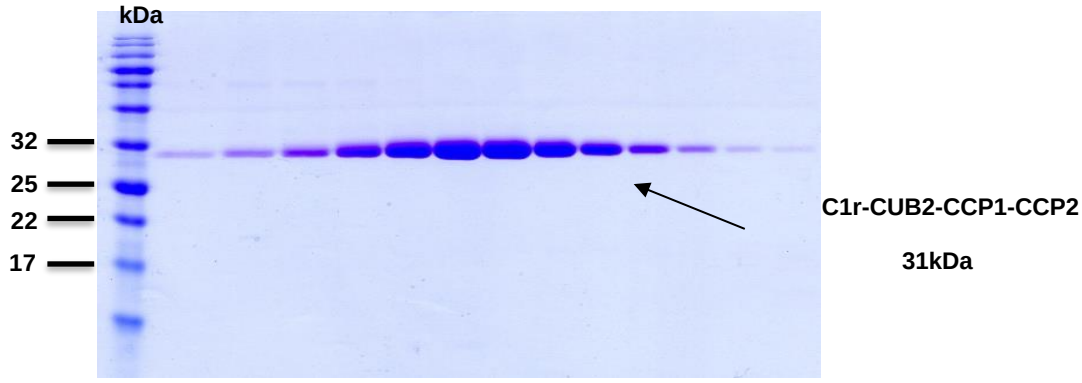


Figure 5-25: *15% non-reducing SDS-PAGE gel for the gel filtration purification of C1r-CUB2-CCP1-CCP2 protein. The protein migrated on the gel as expected with a molecular mass of 31kDa.*

C1r-CUB2-CCP1 purification was carried out following the same protocol used in the purification of C1r-CUB2-CCP1-CCP2. The protein was analysed on 15% non-reducing SDS-PAGE gel to assess the purity (Figure 5.26).

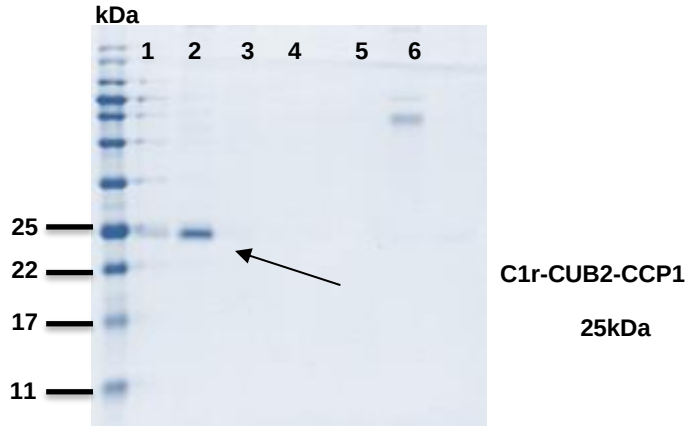


Figure 5-26: *SDS-PAGE analysis of the eluted fractions from Ni^{2+}- NTA column purification of the C1r-CUB2-CCP1. The fractions 1 to 4 are the high imidazole elutions and fractions 5 and 6 are the flow-through, and low imidazole washes respectively. Only one fraction contained the C1r-CUB2-CCP1. The molecular mass was ~25kDa.*

Eluted fractions of C1r-CUB2-CCP1were purified further using gel filtration (Figure 5.27).

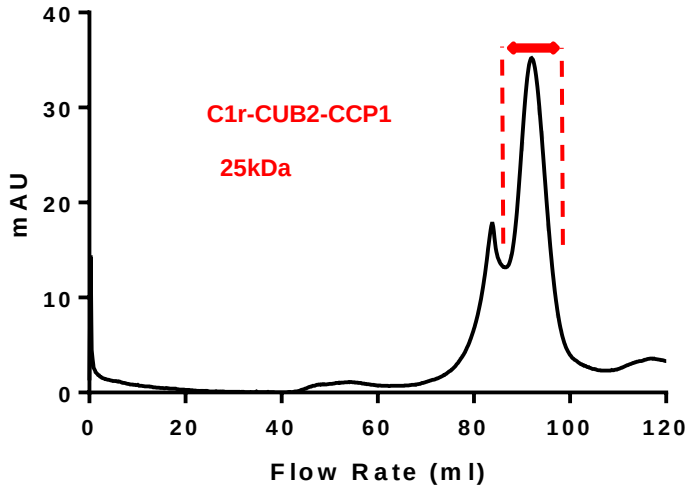


Figure 5-27: *Gel filtration purification of the Ni-column elutions for C1r-CUB2-CCP1 on a Superdex 200 column.*

The C1r-CUB2-CCP1 protein eluted from the gel filtration at the expected position for a monomer ~25 kDa (Figure 5.28).

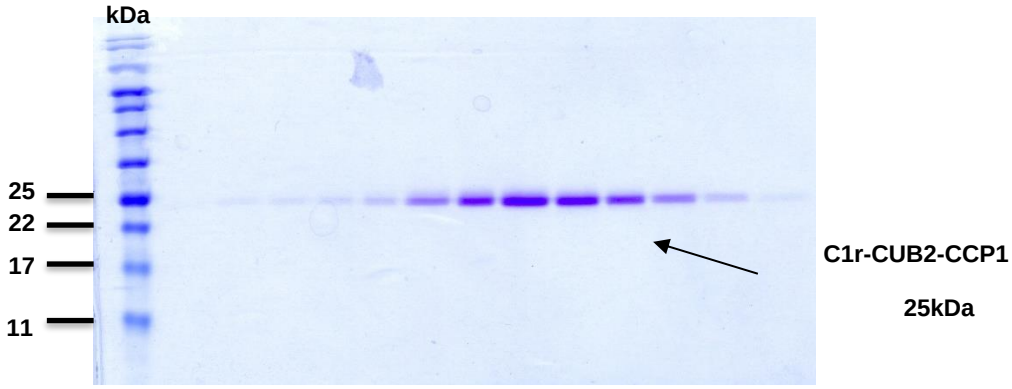


Figure 5-28: *SDS-PAGE analysis for the gel filtration purification of C1r-CUB2-CCP1 protein. The gel filtration step was successful in eliminating the contaminants with a resulting protein purity of ~99%. The protein eluted from the column as expected with a molecular mass of 25kDa. Fractions were pooled together and concentrated.*

Unfortunately, no protein was obtained from cells expressing the C1r-CUB2-CCP1-CCP2-SP fragment. This might be due to toxicity as a result of protease activity in the cell.

5.3.5 Crystallisation of C-BBK32:

Crystallisation trials were performed using crystallisation screens. Although crystals were obtained, the diffraction was relatively weak (10Å).

5.3.6 Gel filtration binding assay for C-BBK32 to C1r-CUB2-CCP1:

The binding of C-BBK32 and C1r-CUB2-CCP1 was investigated using an analytical gel filtration column. C-BBK32 was mixed with C1r and incubated to allow binding. The mixture was then loaded onto gel filtration column, C-BBK32 and C1r alone were also loaded separately onto the same column to give standard curves to which the mixture's elution profile could be compared to. The gel filtration elution profiles of the individually eluted C-BBK32 and C1r was overlaid with the elution profile of the mixture of the two (Figure 5.29). From the gel filtration elution, the complex peak eluted at the same position as the free component indicating that there was no binding.

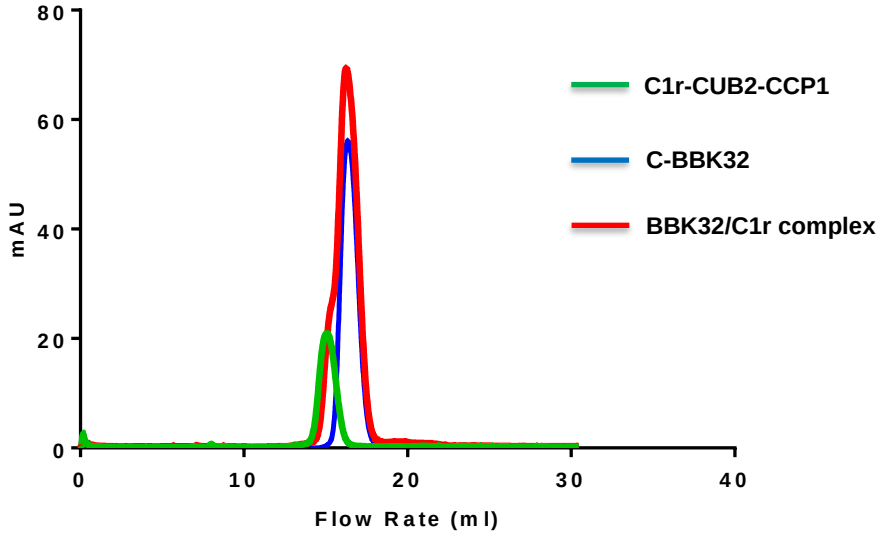


Figure 5-29: *Overlayed elution profiles of C-BBK32, C1r-CUB2-CCP1 and BBK-32/C1r complex. Individual elution profiles are shown in green and blue, and the mixture profile is shown in red.*

5.3.7 BLI Binding assay of C-BBK32 to FL-C1r:

It has been reported previously that *Borrelia burgdorferi* surface lipoprotein BBK32 can interact directly with the C1r serine protease via its C-terminal fragment (Xie *et al.*, 2019). I, therefore, used the panel of protein fragments to identify the binding site on C1r. To analyse this interaction, C-BBK32 was immobilized on a sensor tip, and the different soluble C1r fragments were injected at increasing concentrations. The C-terminal fragment was used because it was produced at a higher yield compared to the FL-BBK32 yet is reported to retain full C1r-binding activity. Surprisingly, no interaction was detected with any of the C1r fragments at any concentration tested (up to 800 µM). To ensure that the immobilisation of BBK32 itself did not prevent binding, two different immobilisation methods were used: amine coupling using the amine-reactive biosensor tips and biotinylation using the streptavidin biosensor tips. In each case, no binding was detected. The previous report claimed that the interaction was Ca^{2+}-dependent (Garcia *et al.*, 2016). However, the addition of Ca^{2+} at 2 and 5 mM $CaCl_2$ did not affect the outcome (Figure 5.30). To ensure that the machine was set up correctly, positive control using immobilised mannan-binding lectin with soluble MASP-2 was used, and binding was detected as expected (Thielens *et al.*, 2001).

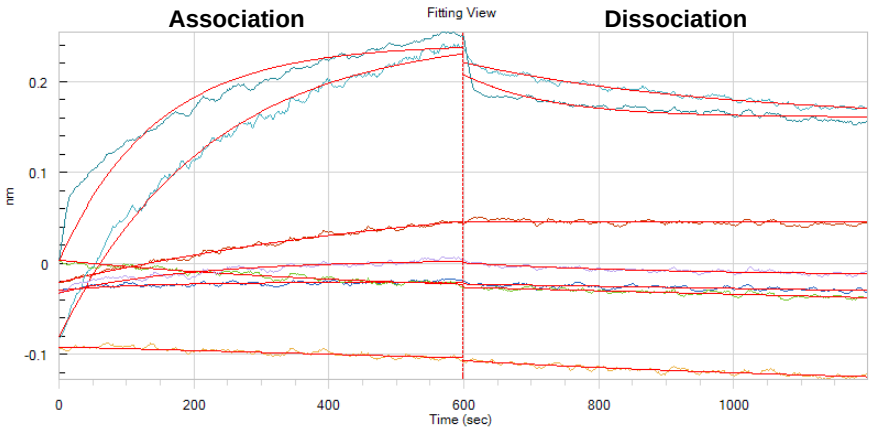


Figure 5-30: *Binding of FL-C1r to the immobilised C-BBK32 analysed by BLI. Flat lines shows no binding.*

5.4 Discussion:

The CP of complement is activated by immune complexes via its C1 component. In addition to immune complexes, it has been reported that CP can be activated and kill *B. burgdorferi* in the absence of specific antibodies (Van Dam *et al.*, 1997). The causative agent of Lyme disease *B. burgdorferi* produces several virulence factors including surface protein BBK32 which can protect the bacterium from lyses by specifically targeting and inactivating complement proteins. *B. burgdorferi* surface protein BBK32 is a multifunctional lipoprotein that is expressed on the surface of the bacteria and aids its spread and colonisation on host cells (Lin *et al.*, 2015). More recently, BBK32 was reported to interact with the CP serine protease C1r in a non-covalent manner and blocking the ability of the proenzyme to auto-activate and cleave it substrate C1s leaving the C1q in a non-active state and therefore affecting the activation of the CP of complement (Garcia *et al.*, 2016). In this study, the C-terminal domain of BBK32 was shown to bind to the FL-C1r, however, C1r is formed of 6 different functional domains. The exact domain of C1r that interact with BBK32 is still uncharacterised. To understand the nature of this interaction I successfully produced FL and C-terminal BBK32 proteins and several fragments of C1r protease to test the possible domain of C1r that is being specifically blocked by BBk32. Each domain of C1r serine protease has a specific function toward the normal activity of the C1 complex of the CP as previously described in chapter 2.

The binding data collected using Bio-layer interferometry reported here shows no interaction between the C-terminal fragment of BBK32 and FL-C1r. BBK32 protein was immobilised using two different methods (amine coupling and biotin/streptavidin) biosensor tips, to reduce the possibility that immobilisation itself prevents binding by blocking the binding sites in BBK32 protein. Interestingly, in the previous report, BBK32 was immobilised using amine coupling chemistry, so it is unlikely that coupling prevents binding (Garcia *et al.*, 2016).

However, it would have been useful to immobilise C1r fragments and test with soluble BBK32 to examine all possibilities. Unfortunately, this was not possible because of time constraints. Additionally, I tested the binding between C1r and BBK32 in solution using size exclusion chromatography column. The binding data from the gel filtration (Figure 5.30), for C-BBK32 in complex with C1r-CUB2-CCP1 also show no binding in solution and therefore excluding the CUB2-CCP1 domains of C1r as a possible binding site for BBK32.

In a more recent study the interaction between BBK32 and C1r (Xie *et al.*, 2019) was characterised and the protease fragment of C1r (SP) was identified as containing the BBK32 binding site. The structure of the C-BBK32 was also determined by (Xie *et al.*, 2019) at 1.7Aº which reveals a five-helix bundle (Figure 5.31). No Ca^{2+} was observed in the structure, suggesting that the interaction with C1r-SP is not Ca^{2+}- dependent and therefore, excluding the CUB1-EGF-CUB2 domains of C1r as they are Ca^{2+} dependant domains as previously described in chapter 2 (Almitairi *et al.*, 2018). BBK32 most likely binds to the SP domain of C1r and therefore disrupt the formation of the C1r-C1s heterodimers. The CP C1r and C1s proteases dimerised at the core of C1q by the binding between the CCP2-SP domains of each C1r protease in the heterotetramer complex. Further work will be carried out in the lab to attempt to resolve the discrepancy between our data and that in the literature. Further crystallisation attempt using complexes of C1r-BBK32 will explain the nature of this interaction. The C1r fragments generated in this work were purified with relatively high yields to give pure protein. These fragments are useful candidates for X-ray crystallography to examine the structure and potential conformational flexibility of C1r within the C1 complex of the CP during activation.

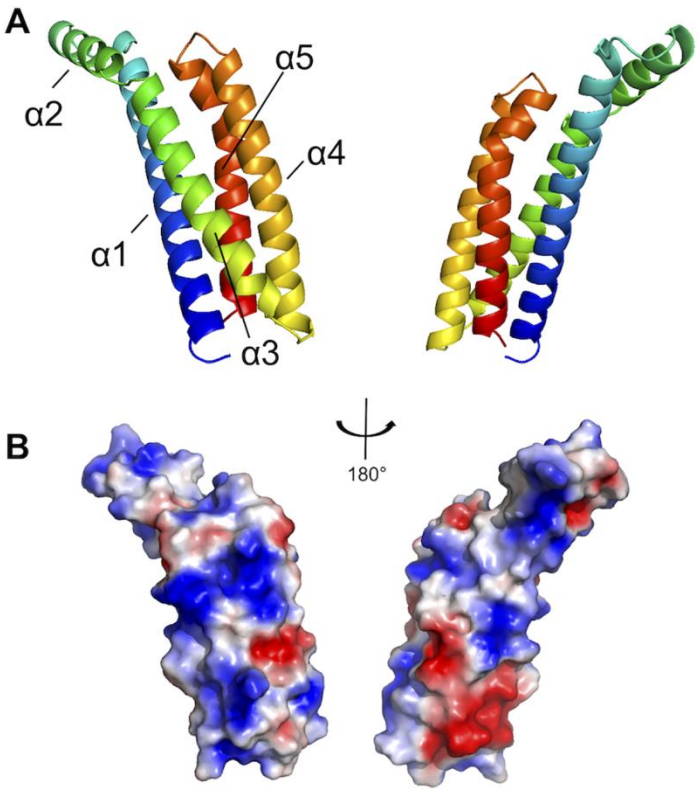


Figure 5-31: *The Crystall structure of BBK32 protien. From the structure, five helix was observes as indiacted in α1, α2, α3 and α4, no* Ca^{+2} *presence in the structure.*

Chapter 6 . General discussion

The complement system is a humoral recognition system and is a vital effector arm of the innate immunity. It plays an important role in the detection and lysis of pathogens and clearing of immune complexes. Three Pathways activates the complement system; the CP, LP and AP, all three pathways converge via the terminal pathway (TP) to lyse the target cells via the formation of the MAC. CP and LP activation are mediated via their collagen-containing pattern recognition molecules C1q and MBL respectively which are associated with specific serine protease C1r/C1s in the CP and MASPs1/2 and 3 in the LP (Fujita, 2002). Understanding the molecular interactions affecting the CP and LP of complement activation is of great interest from biochemical, immunological and medical perspectives and will form a basis for rational drug design. The data presented in this thesis reveals how the subcomponent of the CP associate and assembled to form the C1 complex, as presented in chapter 2 and 3.

6.1 Structure of C1r-C1s complex of the CP:

The crystal structure of the C1r-C1s heterodimer presented in chapter 2, provides a better insight into the C1 complex assembly. The C1r-C1s heterodimers via an extensive interface involving all three domains CUB1-EGF-CUB2. This interaction is calcium dependant indicated by the Ca^{2+} binding site in each EGF domain of the subcomponent. Additional Ca^{2+} binding sites at the CUB domains of C1r and CUB1 of C1s mediates the binding to C1q collagen stems. The structural data reported here are incompatible with the recently presented stacked tetramer model for $C1r_2C1s_2$ via CUB1-EGF-CUB2 domains (Mortensen *et al.*, 2017). The analytical gel filtration results (Figure 2.10), shows that the 3D C1r and 3D C1s do not form a tetrameric complex even at 10-fold higher the normal serum concertation of $C1r_2C1s_s$ which exclude the stacked tetramer model of C1. The presented data support the traditional model of C1 first observed in EM images in which the C1r-C1s heterodimers are formed by

the interaction between the CCP-SP domains of C1r forming an anti-parallel elongated S-shaped heterotetramer (Ziccardi, 1982). Autoactivation of C1 will require disruption of the $C1r_2$-$C1s_2$ heterodimers complex at the CCP-SP junction of C1r followed by the alignment of the catalytic site of one polypeptide with the cleavage site of the other. A more recent model for C1 activation proposed by (Sharp *et al.*, 2019), shows that the original model presented here is incorrect. In the original model, the C1r was positioned in the outside of the tetramer with C1s at the centre creating six binding sites for each C1q stems at the CUB domains of C1r and the CUB1 domain of C1s (Almitairi *et al.*, 2018). However, the recently proposed model shows that C1r is positioned at the centre of the tetramer instead of being on the outside. In the new model, there are two copies of the C1r-C1s heterodimers that associate with three adjacent C1q-collagen helices with the outer two C1q-collagen helices bound to the CUB1 domains of C1r and C1s, while the third collagen helix in the middle is not bound. Interestingly, there are two copies of C1r-C1s heterodimers in the obtained crystal structure that most likely mimic the $C1r_2C1s_2$ when bound to C1q. In this arrangement four binding sites for the collagen stems in the CUB1 domain of each C1r-C1s heterodimers. The binding sites on the CUB2 domains of C1r is blocked by the CUB1 domain of its anti-parallel partner C1s. The absence of collagen binding site for the C1s-CUB2 domain is likely provides an increased flexibility for the C1s CUB2-CCP1-CCP2-SP arms outside the complex, accounting for diverse arrangements needed for C1s to be cleaved by C1r and to reach and subsequently cleave its substrates C4. Upon binding to C4 on cell surface, the relative position of the C1r-C1s heterodimers shifts allowing the collagen bound CUB1 of C1s to interact with adjacent CUB2 of C1r which may accompany the formation of C1.

6.2 Structure of the C1qA, C1qB and C1qC chains in complex with the CUB1-EGF-CUB2 domains of C1s:

The assembly of the C1r-C1s heterotetramer inside the recognition molecule C1q is mediated via six Ca^{2+} dependant binding sites in each collagen stem at the CUB domains of C1r and the CUB1 domain of C1s (Bally *et al.*, 2013). The proteases are attached to C1q via a specific motif (Hyp-Gly-Lys-Xaa-Gly-Pro), present in the all C1q chains A, B and C. The C1q chains form a right-handed helix with the characteristic of one residue stagger between adjacent strands giving rise to the leading, middle and lagging strands. From the polar interactions highlighted in (Table 3.7), the Lys15 residue of the lagging strand is present in all structures highlighting the key role of the Lys 15 for the interaction with the CUB1 domain of C1s and is compatible with the previously solved structure of C1q/C1s (Venkatraman Girija *et al.*, 2013). From the solved structures, it is possible to propose a model for the C1g/C1s interaction in which the leading strand of C1q collagen is C, the middle is A, and the lagging is B. The C1qA/C:C1s and C1qB/C:C1s complexes are the first complexes between protein and collaging heterooligomer to be reported. The method used to identify the order of the chains in C1q has a great potential for other systems where protein-collagen complexes are assembled from collagen heterooligomers.

6.3 Interaction of the immune modulator LAIR-2 with MBL of the LP:

Analysis of the interaction between the immune modulator LAIR-2 and the hMBL shows that LAIR-2 can bind to the collaging like domain of MBL. LAIR-2 was reported to compete with its partner LAIR-1 and therefore affecting its role in regulating immune response (Lebbink *et al.*, 2014). The data presented here support the recently reported study about LAIR-2 binding to MBL (Nordkamp *et al.*, 2014), and show that the binding occurs with high affinity. However, concertation of LAIR-2 in the circulation is less than 0.3 µg/ml, which is unlikely to be

sufficient to block the complement activation. The normal concertation of MBL the serum is between 1 and 2mg/ml (Bouwman, Roep and Roos, 2006).

6.4 Evasion of the CP by the *Borrelia burgdorferi* surface protein BBK32:

The BLI binding data reported in this chapter shows no interaction between FL-C1r and the C-terminal BBK32. Two methods were used for BBK32 immobilisation (amine coupling using the amine-reactive biosensor tips and biotinylation using the streptavidin biosensor tips), to ensure immobilisation itself did not affect the binding. Immobilisation of the C1r would have been a useful complementary approach, however, this was not possible due to time constrains. In a more recent study of the interaction between BBK32 and C1r, the SP domain of C1r was identified as containing the BBK32 binding site, and the structure of the C-BBK32 was determined (Xie *et al.*, 2019). Our results are in direct conflict with the exciting literature, and it will be interesting to use additional methods (e.g. surface plasmon resonance and/or isothermal titration calorimetry and/or gel filtration) to probe the putative interaction. The structure reveals a five-helix bundle. No Ca^{2+} was observed in the structure suggesting that the interaction with C1r is not Ca^{2+}-dependent and therefore, excluding the CUB1-EGF-CUB2 domains of C1r as they are Ca^{2+} dependant domains. This is in line with the gel filtration binding experiment reporter in (Figure 5.30) in which no binding was observed between the C1r-CUB2-CCP1 domains and C-BBK32 suggesting that BBK32 binds to the SP domain of C1r as the SP domain do not bind calcium (Gál *et al.*, 2009). Several fragments of C1r were expressed in good yield and can be used for crystallisation to test and examine the potential flexibility of C1r within C1 complex.

6.5 Future work:

6.5.1 Crystallisation of LAIR-2 protein:

Attempt to crystallise FL-LAIR-2 for structure determination were unsuccessful with both digested and undigested forms (ie where the His-tag was removed). Spherulites were observed which is a promising indication. Spherulites are crystal that lack order, hence are spherical. Despite multiple attempts, crystallisation trails can take time to optimise considering the range of variables in the crystallisation conditions and the time for crystal growth, this was not possible due to time constrain. However, the structure of LAIR-2 will be informative to understand its role in the immune regulation and how LAIR-2 interacts with collagen domains and this project is worth pursuing in the future.

6.5.2 Binding of BBk32 to the C1r fragments:

Studies in this thesis revealed no interaction between BBK32 and the C1r-SP domain as has been reported previously. The BLI binding data for the FL-C1r and C-BBK32 shows no binding with the BBK32 immobilised. It will be useful to test the binding with the C1r fragments immobilised to confirm that the immobilisation its self-did not prevents binding. Other techniques can also be used to test for an interaction e.g. surface plasmon resonance, isothermal titration calorimetry and/or gel filtration chromatography.

APPENDICES

Appendix 1: Full-length C1r amino acids sequence

```
        10         20         30         40         50
MWLLYLLVPA LFCRAGGSIP IPQKLFGEVT SPLFPKPYPN NFETTTVITV
        60         70         80         90        100
PTGYRVKLVF QQFDLEPSEG CFYDYVKISA DKKSLGRFCG QLGSPLGNPP
       110        120        130        140        150
GKKEFMSQGN KMLLTFHTDF SNEENGTIMF YKGFLAYYQA VDLDECASRS
       160        170        180        190        200
KSGEEDPQPQ CQHLCHNYVG GYFCSCRPGY ELQEDTHSCQ AECSSELYTE
       210        220        230        240        250
ASGYISSLEY PRSYPPDLRC NYSIRVERGL TLHLKFLEPF DIDDHQQVHC
       260        270        280        290        300
PYDQLQIYAN GKNIGEFCGK QRPPDLDTSS NAVDLLFFTD ESGDSRGWKL
       310        320        330        340        350
RYTTEIIKCP QPKTLDEFTI IQNLQPQYQF RDYFIATCKQ GYQLIEGNQV
       360        370        380        390        400
LHSFTAVCQD DGTWHRAMPR CKIKDCGQPR NLPNGDFRYT TTMGVNTYKA
       410        420        430        440        450
RIQYYCHEPY YKMQTRAGSR ESEQGVYTCT AQGIWKNEQK GEKIPRCLPV
       460        470        480        490        500
CGKPVNPVEQ RQRIIGGQKA KMGNFPWQVF TNIHGRGGGA LLGDRWILTA
       510        520        530        540        550
AHTLYPKEHE AQSNASLDVF LGHTNVEELM KLGNHPIRRV SVHPDYRQDE
       560        570        580        590        600
SYNFEGDIAL LELENSVTLG PNLLPICLPD NDTFYDLGLM GYVSGFGVME
       610        620        630        640        650
EKIAHDLRFV RLPVANPQAC ENWLRGKNRM DVFSQNMFCA GHPSLKQDAC
       660        670        680        690        700
QGDSGGVFAV RDPNTDRWVA TGIVSWGIGC SRGYGFYTKV LNYVDWIKKE

MEEED
```

N.B: C1r domains colour: CUB1-EGF-CUB2-CCP1-CCP2-SP

Appendix 2: C1r-CUB2-CCP1 cloned into pED4 sequence blast

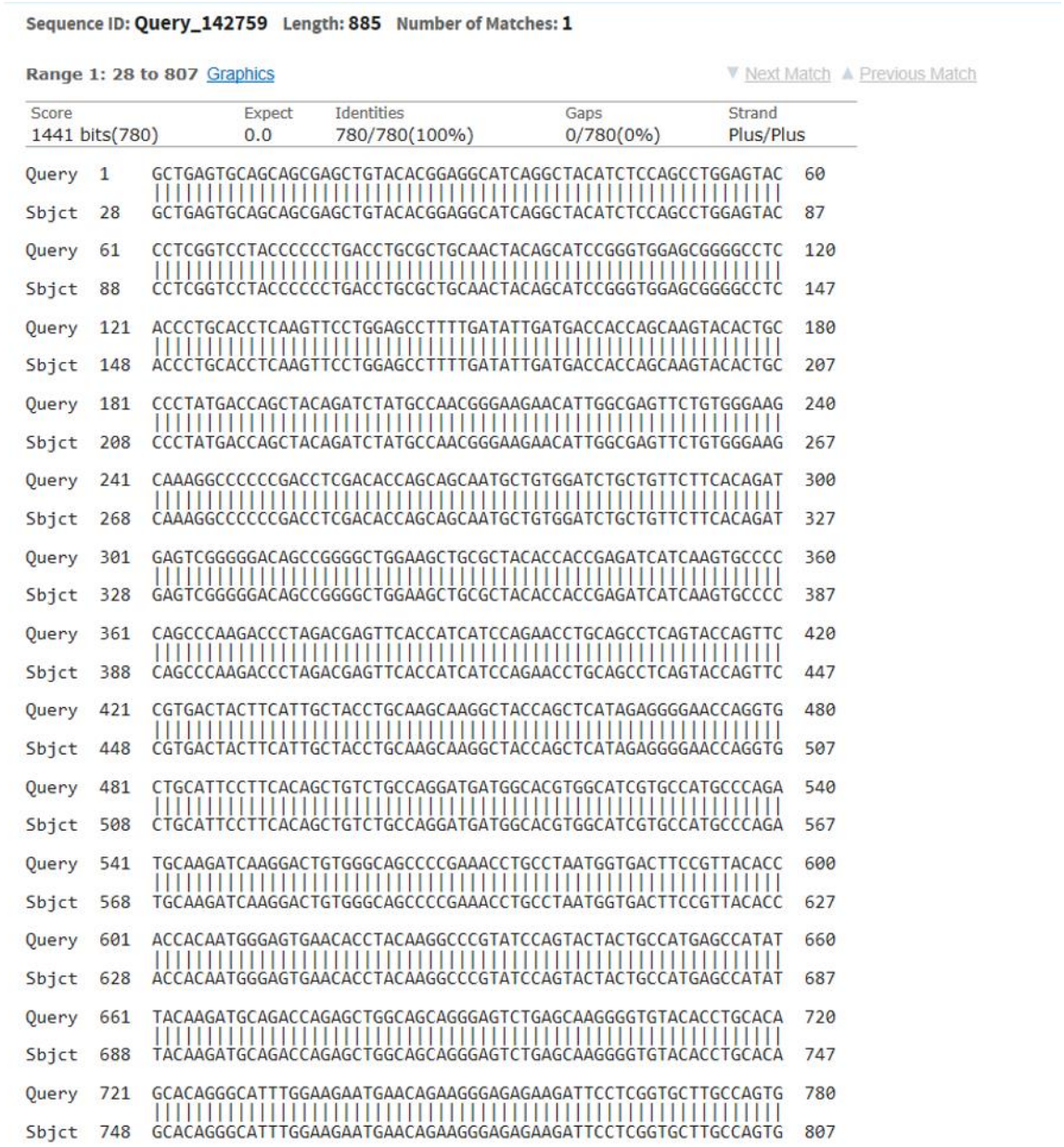
Sequence ID: Query_86611 **Length:** 647 **Number of Matches:** 1

Range 1: 71 to 625 Graphics ▼ Next Match ▲ Previou

Score	Expect	Identities	Gaps	Strand
1026 bits(555)	0.0	555/555(100%)	0/555(0%)	Plus/Plus

```
Query  1    GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  71   GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  130

Query  61   CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  131  CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  190

Query  121  ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  191  ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  250

Query  181  CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  251  CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  310

Query  241  CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  311  CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  370

Query  301  GAGTCGGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  371  GAGTCGGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  430

Query  361  CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  431  CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  490

Query  421  CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  491  CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  550

Query  481  CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  551  CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  610

Query  541  TGCAAGATCAAGGAC  555
            |||||||||||||||
Sbjct  611  TGCAAGATCAAGGAC  625
```

Appendix 3: C1r-CUB2-CCP1-CCP2 cloned into pED4 sequence blast



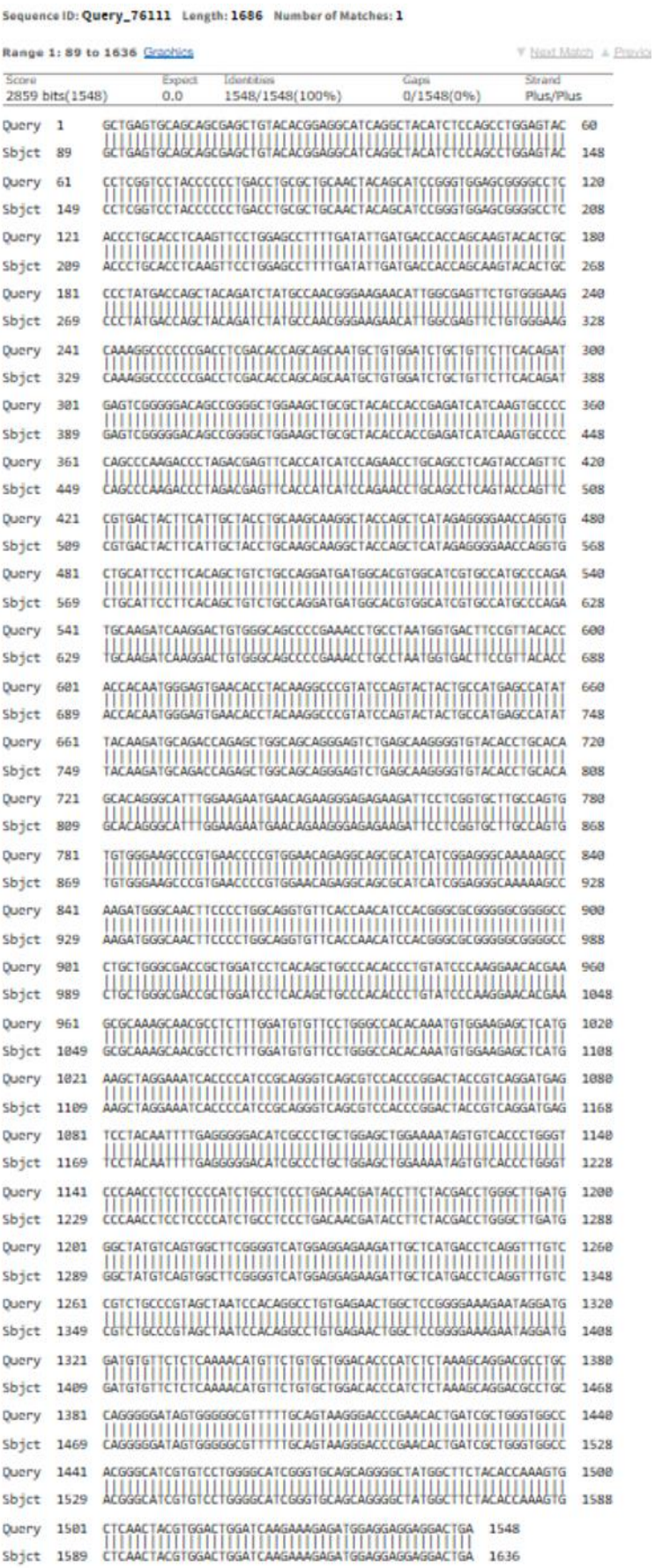
Sequence ID: Query_142759 **Length:** 885 **Number of Matches:** 1

Range 1: 28 to 807 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1441 bits(780)	0.0	780/780(100%)	0/780(0%)	Plus/Plus

```
Query  1    GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  28   GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  87

Query  61   CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  88   CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  147

Query  121  ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  148  ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  207

Query  181  CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  208  CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  267

Query  241  CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  268  CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  327

Query  301  GAGTCGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  328  GAGTCGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  387

Query  361  CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  388  CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  447

Query  421  CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  448  CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  507

Query  481  CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  508  CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  567

Query  541  TGCAAGATCAAGGACTGTGGGCAGCCCCGAAACCTGCCTAATGGTGACTTCCGTTACACC  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  568  TGCAAGATCAAGGACTGTGGGCAGCCCCGAAACCTGCCTAATGGTGACTTCCGTTACACC  627

Query  601  ACCACAATGGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATAT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  628  ACCACAATGGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATAT  687

Query  661  TACAAGATGCAGACCAGAGCTGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACA  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  688  TACAAGATGCAGACCAGAGCTGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACA  747

Query  721  GCACAGGGCATTTGGAAGAATGAACAGAAGGGAGAGAAGATTCCTCGGTGCTTGCCAGTG  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  748  GCACAGGGCATTTGGAAGAATGAACAGAAGGGAGAGAAGATTCCTCGGTGCTTGCCAGTG  807
```

Appendix 4: C1r-CUB2-CCP1-CCP2-SP cloned into pED4 sequence blast



Sequence ID: **Query_76111** Length: **1686** Number of Matches: **1**

Range 1: 89 to 1636 Graphics

Score	Expect	Identities	Gaps	Strand
2859 bits(1548)	0.0	1548/1548(100%)	0/1548(0%)	Plus/Plus

```
Query  1     GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  60
Sbjct  89    GCTGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAGTAC  148

Query  61    CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  120
Sbjct  149   CCTCGGTCCTACCCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTC  208

Query  121   ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  180
Sbjct  209   ACCCTGCACCTCAAGTTCCTGGAGCCTTTTGATATTGATGACCACCAGCAAGTACACTGC  268

Query  181   CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  240
Sbjct  269   CCCTATGACCAGCTACAGATCTATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAG  328

Query  241   CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  300
Sbjct  329   CAAAGGCCCCCCGACCTCGACACCAGCAGCAATGCTGTGGATCTGCTGTTCTTCACAGAT  388

Query  301   GAGTCGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  360
Sbjct  389   GAGTCGGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCACCGAGATCATCAAGTGCCCC  448

Query  361   CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  420
Sbjct  449   CAGCCCAAGACCCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTCAGTACCAGTTC  508

Query  421   CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  480
Sbjct  509   CGTGACTACTTCATTGCTACCTGCAAGCAAGGCTACCAGCTCATAGAGGGGAACCAGGTG  568

Query  481   CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  540
Sbjct  569   CTGCATTCCTTCACAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCATGCCCAGA  628

Query  541   TGCAAGATCAAGGACTGTGGGCAGCCCCGAAACCTGCCTAATGGTGACTTCCGTTACACC  600
Sbjct  629   TGCAAGATCAAGGACTGTGGGCAGCCCCGAAACCTGCCTAATGGTGACTTCCGTTACACC  688

Query  601   ACCACAATGGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATAT  660
Sbjct  689   ACCACAATGGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATAT  748

Query  661   TACAAGATGCAGACCAGAGCTGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACA  720
Sbjct  749   TACAAGATGCAGACCAGAGCTGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACA  808

Query  721   GCACAGGGCATTTGGAAGAATGAACAGAAGGGAGAGAAGATTCCTCGGTGCTTGCCAGTG  780
Sbjct  809   GCACAGGGCATTTGGAAGAATGAACAGAAGGGAGAGAAGATTCCTCGGTGCTTGCCAGTG  868

Query  781   TGTGGGAAGCCCGTGAACCCCGTGGAACAGAGGCAGCGCATCATCGGAGGGCAAAAAGCC  840
Sbjct  869   TGTGGGAAGCCCGTGAACCCCGTGGAACAGAGGCAGCGCATCATCGGAGGGCAAAAAGCC  928

Query  841   AAGATGGGCAACTTCCCCTGGCAGGTGTTCACCAACATCCACGGGCGCGGGGGCGGGGCC  900
Sbjct  929   AAGATGGGCAACTTCCCCTGGCAGGTGTTCACCAACATCCACGGGCGCGGGGGCGGGGCC  988

Query  901   CTGCTGGGCGACCGCTGGATCCTCACAGCTGCCCACACCCTGTATCCCAAGGAACACGAA  960
Sbjct  989   CTGCTGGGCGACCGCTGGATCCTCACAGCTGCCCACACCCTGTATCCCAAGGAACACGAA  1048

Query  961   GCGCAAAGCAACGCCTCTTTGGATGTGTTCCTGGGCCACACAAATGTGGAAGAGCTCATG  1020
Sbjct  1049  GCGCAAAGCAACGCCTCTTTGGATGTGTTCCTGGGCCACACAAATGTGGAAGAGCTCATG  1108

Query  1021  AAGCTAGGAAATCACCCCATCCGCAGGGTCAGCGTCCACCCGGACTACCGTCAGGATGAG  1080
Sbjct  1109  AAGCTAGGAAATCACCCCATCCGCAGGGTCAGCGTCCACCCGGACTACCGTCAGGATGAG  1168

Query  1081  TCCTACAATTTTGAGGGGGACATCGCCCTGCTGGAGCTGGAAAATAGTGTCACCCTGGGT  1140
Sbjct  1169  TCCTACAATTTTGAGGGGGACATCGCCCTGCTGGAGCTGGAAAATAGTGTCACCCTGGGT  1228

Query  1141  CCCAACCTCCTCCCCATCTGCCTCCCTGACAACGATACCTTCTACGACCTGGGCTTGATG  1200
Sbjct  1229  CCCAACCTCCTCCCCATCTGCCTCCCTGACAACGATACCTTCTACGACCTGGGCTTGATG  1288

Query  1201  GGCTATGTCAGTGGCTTCGGGGTCATGGAGGAGAAGATTGCTCATGACCTCAGGTTTGTC  1260
Sbjct  1289  GGCTATGTCAGTGGCTTCGGGGTCATGGAGGAGAAGATTGCTCATGACCTCAGGTTTGTC  1348

Query  1261  CGTCTGCCCGTAGCTAATCCACAGGCCTGTGAGAACTGGCTCCGGGGAAAGAATAGGATG  1320
Sbjct  1349  CGTCTGCCCGTAGCTAATCCACAGGCCTGTGAGAACTGGCTCCGGGGAAAGAATAGGATG  1408

Query  1321  GATGTGTTCTCTCAAAACATGTTCTGTGCTGGACACCCATCTCTAAAGCAGGACGCCTGC  1380
Sbjct  1409  GATGTGTTCTCTCAAAACATGTTCTGTGCTGGACACCCATCTCTAAAGCAGGACGCCTGC  1468

Query  1381  CAGGGGGATAGTGGGGGCGTTTTTGCAGTAAGGGACCCGAACACTGATCGCTGGGTGGCC  1440
Sbjct  1469  CAGGGGGATAGTGGGGGCGTTTTTGCAGTAAGGGACCCGAACACTGATCGCTGGGTGGCC  1528

Query  1441  ACGGGCATCGTGTCCTGGGGCATCGGGTGCAGCAGGGGCTATGGCTTCTACACCAAAGTG  1500
Sbjct  1529  ACGGGCATCGTGTCCTGGGGCATCGGGTGCAGCAGGGGCTATGGCTTCTACACCAAAGTG  1588

Query  1501  CTCAACTACGTGGACTGGATCAAGAAAGAGATGGAGGAGGAGGACTGA  1548
Sbjct  1589  CTCAACTACGTGGACTGGATCAAGAAAGAGATGGAGGAGGAGGACTGA  1636
```

Appendix 5: Full length LAIR-2 cloned into pLEIC-03 sequence blast

Sequence ID: Query_115959 Length: 541 Number of Matches: 1

Range 1: 42 to 500 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
848 bits(459)	0.0	459/459(100%)	0/459(0%)	Plus/Plus

```
Query  1    ATGTCTCCACACCTCACTGCTCTCCTGGGCCTAGTGCTCTGCCTGGCCCAGACCATCCAC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  42   ATGTCTCCACACCTCACTGCTCTCCTGGGCCTAGTGCTCTGCCTGGCCCAGACCATCCAC  101

Query  61   ACGCAGGAGGGGGCCCTTCCCAGACCCTCCATCTCGGCTGAGCCAGGCACTGTGATCTCC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  102  ACGCAGGAGGGGGCCCTTCCCAGACCCTCCATCTCGGCTGAGCCAGGCACTGTGATCTCC  161

Query  121  CCGGGGAGCCATGTGACTTTCATGTGCCGGGGCCCGGTTGGGGTTCAAACATTCCGCCTG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162  CCGGGGAGCCATGTGACTTTCATGTGCCGGGGCCCGGTTGGGGTTCAAACATTCCGCCTG  221

Query  181  GAGAGGGAGGATAGAGCCAAGTACAAAGATAGTTATAATGTGTTTCGACTTGGTCCATCT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  222  GAGAGGGAGGATAGAGCCAAGTACAAAGATAGTTATAATGTGTTTCGACTTGGTCCATCT  281

Query  241  GAGTCAGAGGCCAGATTCCACATTGACTCAGTAAGTGAAGGAAATGCCGGGCTTTATCGC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  282  GAGTCAGAGGCCAGATTCCACATTGACTCAGTAAGTGAAGGAAATGCCGGGCTTTATCGC  341

Query  301  TGCCTCTATTATAAGCCCCCTGGATGGTCTGAGCACAGTGACTTCCTGGAGCTGCTGGTG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  342  TGCCTCTATTATAAGCCCCCTGGATGGTCTGAGCACAGTGACTTCCTGGAGCTGCTGGTG  401

Query  361  AAAGAAAGCTCTGGAGGCCCGGACTCCCCGGACACAGAGCCCGGCTCCTCAGCTGGGACT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  402  AAAGAAAGCTCTGGAGGCCCGGACTCCCCGGACACAGAGCCCGGCTCCTCAGCTGGGACT  461

Query  421  GTGCCAGGCACTGAAGCCTCCGGATTTGATGCACCATGA  459
            |||||||||||||||||||||||||||||||||||||||
Sbjct  462  GTGCCAGGCACTGAAGCCTCCGGATTTGATGCACCATGA  500
```

Appendix 6: BBK32 cloned into pLEIC-01 sequence blast

Sequence ID: **Query_40665** **Length:** **525** **Number of Matches:** **1**

Range 1: 25 to 516 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
909 bits(492)	0.0	492/492(100%)	0/492(0%)	Plus/Plus

```
Query  1    ATGGATGAATATGATGAAGAGGATGAAGAGGAAATAAGATTAAGCAATCGATATCAATCT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  25   ATGGATGAATATGATGAAGAGGATGAAGAGGAAATAAGATTAAGCAATCGATATCAATCT  84

Query  61   TATCTAGAAGGTGTTAAATATAATGTAGATTCAGCAATTCAAACAATTACTAAGATATAT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  85   TATCTAGAAGGTGTTAAATATAATGTAGATTCAGCAATTCAAACAATTACTAAGATATAT  144

Query  121  AATACTTATACATTATTTTCAACAAAGCTAACCCAAATGTATTCTACACGCCTTGACAAC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  145  AATACTTATACATTATTTTCAACAAAGCTAACCCAAATGTATTCTACACGCCTTGACAAC  204

Query  181  TTTGCTAAAGCCAAAGCTAAAGAAGAAGCTGCAAAGTTTACAAAAGAAGACCTTGaaaaa  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  205  TTTGCTAAAGCCAAAGCTAAAGAAGAAGCTGCAAAGTTTACAAAAGAAGACCTTGAAAAA  264

Query  241  aaTTTCAAGACCTTATTAAATTATATTCAAGTAAGTGTAAAGACTGCAGCAAATTTTGTA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  265  AATTTCAAGACCTTATTAAATTATATTCAAGTAAGTGTAAAGACTGCAGCAAATTTTGTA  324

Query  301  TACATAAATGACACGCATGCAAAAAGGAAATTAGAGAACATTGAAGCAGAAATAAAAACT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  325  TACATAAATGACACGCATGCAAAAAGGAAATTAGAGAACATTGAAGCAGAAATAAAAACT  384

Query  361  TTAATTGCAAAGATCAAAGAACAATCTAATTTATACGAAGCATATAAAGCAATAGTAACG  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  385  TTAATTGCAAAGATCAAAGAACAATCTAATTTATACGAAGCATATAAAGCAATAGTAACG  444

Query  421  TCAATCTTATTAATGAGGGATTCTCTTAAAGAAGTGCAAGGTATCATTGACAAGAATGGC  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  445  TCAATCTTATTAATGAGGGATTCTCTTAAAGAAGTGCAAGGTATCATTGACAAGAATGGC  504

Query  481  GTTTGGTACTAA  492
            ||||||||||||
Sbjct  505  GTTTGGTACTAA  516
```

References

ABBAS, A. K. & JANEWAY, C. A. 2000. Immunology: improving on nature in the twenty-first century. *Cell,* 100**,** 129-138.

AFONINE, P. V., GROSSE-KUNSTLEVE, R. W., ECHOLS, N., HEADD, J. J., MORIARTY, N. W., MUSTYAKIMOV, M., TERWILLIGER, T. C., URZHUMTSEV, A., ZWART, P. H. & ADAMS, P. D. 2012. Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography,* 68**,** 352-367.

AGRAWAL, A., SHRIVE, A. K., GREENHOUGH, T. J. & VOLANAKIS, J. E. 2001. Topology and structure of the C1q-binding site on C-reactive protein. *The Journal of Immunology,* 166**,** 3998-4004.

AHN, B. C., PARK, J. S., KIM, D., PARK, J., PI, J., YUM, J. S., JEONG, Y., BAEK, K., MOON, H. M. & YOON, J. 2013. Overproduction of recombinant human mannose-binding lectin (MBL) in Chinese hamster ovary cells. *Protein Expression and Purification,* 88**,** 1-6.

ALMITAIRI, J. O., GIRIJA, U. V., FURZE, C. M., SIMPSON-GRAY, X., BADAKSHI, F., MARSHALL, J. E., SCHWAEBLE, W. J., MITCHELL, D. A., MOODY, P. C. & WALLIS, R. 2018. Structure of the C1r–C1s interaction of the C1 complex of complement activation. *Proceedings of the National Academy of Sciences,* 115**,** 768-773.

BAJIC, G., DEGN, S. E., THIEL, S. & ANDERSEN, G. R. 2015. Complement activation, regulation, and molecular basis for complement-related diseases. *The EMBO journal,* 34**,** 2735-2757.

BALLY, I., ANCELET, S., MORISCOT, C., GONNET, F., MANTOVANI, A., DANIEL, R., SCHOEHN, G., ARLAUD, G. J. & THIELENS, N. M. 2013. Expression of recombinant human complement C1q allows identification of the C1r/C1s-binding sites. *Proceedings of the National Academy of Sciences,* 110**,** 8650-8655.

BALLY, I., ROSSI, V., LUNARDI, T., THIELENS, N. M., GABORIAUD, C. & ARLAUD, G. J. 2009. Identification of the C1q-binding Sites of Human C1r and C1s A refined three-dimensional model of the C1 complex of complement. *Journal of Biological Chemistry,* 284**,** 19340-19348.

BARNUM, S. R. 2017. Complement: a primer for the coming therapeutic revolution. *Pharmacology & therapeutics,* 172**,** 63-72.

BELTRAME, M. H., BOLDT, A. B., CATARINO, S. J., MENDES, H. C., BOSCHMANN, S. E., GOELDNER, I. & MESSIAS-REASON, I. 2015. MBL-associated serine proteases (MASPs) and infectious diseases. *Molecular immunology,* 67**,** 85-100.

BÍRÓ, A., ROVÓ, Z., PAPP, D., CERVENAK, L., VARGA, L., FÜST, G., THIELENS, N. M., ARLAUD, G. J. & PROHÁSZKA, Z. 2007. Studies on the interactions between C-reactive protein and complement proteins. *Immunology,* 121**,** 40-50.

BORK, P. & BECKMANN, G. 1993. The CUB domain: a widespread module in developmentally regulated proteins. *Journal of molecular biology,* 231**,** 539-545.

BOUWMAN, L. H., ROEP, B. O. & ROOS, A. 2006. Mannose-binding lectin: clinical implications for infection, transplantation, and autoimmunity. *Human immunology,* 67**,** 247-256.

BRONDIJK, T. H. C., DE RUITER, T., BALLERING, J., WIENK, H., LEBBINK, R. J., VAN INGEN, H., BOELENS, R., FARNDALE, R. W., MEYAARD, L. & HUIZINGA, E. G. 2010. Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood, The Journal of the American Society of Hematology,* 115**,** 1364-1373.

BUDAYOVA-SPANO, M., LACROIX, M., THIELENS, N. M., ARLAUD, G. J., FONTECILLA-CAMPS, J. C. & GABORIAUD, C. 2002. The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *The EMBO Journal,* 21**,** 231-239.

BUNACIU, A. A., UDRIŞTIOIU, E. G. & ABOUL-ENEIN, H. Y. 2015. X-ray diffraction: instrumentation and applications. *Critical reviews in analytical chemistry,* 45**,** 289-299.

BUSH, L. M. & VAZQUEZ-PERTEJO, M. T. 2018. Tick borne illness—Lyme disease. *Disease-A-Month,* 64**,** 195-212.

CAINE, J. A. & COBURN, J. 2016. Multifunctional and redundant roles of Borrelia burgdorferi outer surface proteins in tissue adhesion, colonization, and complement evasion. *Frontiers in immunology,* 7**,** 442.

CARROLL, M. C. & ISENMAN, D. E. 2012. Regulation of humoral immunity by complement. *Immunity,* 37**,** 199-207.

CHANG, A. Y., CHAU, V., LANDAS, J. A. & PANG, Y. 2017. Preparation of calcium competent Escherichia coli and heat-shock transformation. *JEMI methods,* 1**,** 22-25.

COLOMB, M., ARLAUD, G. & VILLIERS, C. 1984. Activation of C1. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* 306**,** 283-292.

DENMAN, A. 1992. Cellular and Molecular Immunology. *Postgraduate Medical Journal,* 68**,** 305.

DESSAU, M. A. & MODIS, Y. 2011. Protein crystallization for X-ray crystallography. *JoVE (Journal of Visualized Experiments)***,** e2285.

DOBÓ, J., SZAKÁCS, D., OROSZLÁN, G., KORTVELY, E., KISS, B., BOROS, E., SZÁSZ, R., ZÁVODSZKY, P., GÁL, P. & PÁL, G. 2016. MASP-3 is the exclusive pro-factor D activator in resting blood: the lectin and the alternative complement pathways are fundamentally linked. *Scientific reports,* 6**,** 1-12.

DOMMETT, R., KLEIN, N. & TURNER, M. 2006. Mannose-binding lectin in innate immunity: past, present and future. *Tissue antigens,* 68**,** 193-209.

EHRNTHALLER, C., IGNATIUS, A., GEBHARD, F. & HUBER-LANG, M. 2011. New insights of an old defense system: structure, function, and clinical relevance of the complement system. *Molecular medicine,* 17**,** 317-329.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 486-501.

ENDO, Y., MATSUSHITA, M. & FUJITA, T. 2011. The role of ficolins in the lectin pathway of innate immunity. *The international journal of biochemistry & cell biology,* 43**,** 705-712.

FRANK, M. M. 2010. Complement disorders and hereditary angioedema. *Journal of Allergy and Clinical Immunology,* 125**,** S262-S271.

FUJITA, T. 2002. Evolution of the lectin–complement pathway and its role in innate immunity. *Nature Reviews Immunology,* 2**,** 346-353.

GABORIAUD, C., FRACHET, P., THIELENS, N. & ARLAUD, G. 2012. The human c1q globular domain: structure and recognition of non-immune self ligands. *Frontiers in immunology,* 2**,** 92.

GABORIAUD, C., JUANHUIX, J., GRUEZ, A., LACROIX, M., DARNAULT, C., PIGNOL, D., VERGER, D., FONTECILLA-CAMPS, J. C. & ARLAUD, G. J. 2003. The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *Journal of Biological Chemistry,* 278**,** 46974-46982.

GABORIAUD, C., THIELENS, N. M., GREGORY, L. A., ROSSI, V., FONTECILLA-CAMPS, J. C. & ARLAUD, G. J. 2004. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends in immunology,* 25**,** 368-373.

GÁL, P., DOBÓ, J., ZÁVODSZKY, P. & SIM, R. B. 2009. Early complement proteases: C1r, C1s and MASPs. A structural insight into activation and functions. *Molecular immunology,* 46**,** 2745-2752.

GARCIA, B. L., ZHI, H., WAGER, B., HÖÖK, M. & SKARE, J. T. 2016. Borrelia burgdorferi BBK32 inhibits the classical pathway by blocking activation of the C1 complement complex. *PLoS pathogens,* 12**,** e1005404.

GARRED, P., GENSTER, N., PILELY, K., BAYARRI-OLMOS, R., ROSBJERG, A., MA, Y. J. & SKJOEDT, M. O. 2016. A journey through the lectin pathway of complement—MBL and beyond. *Immunological reviews,* 274**,** 74-97.

GILLILAND, G. L., TUNG, M., BLAKESLEE, D. & LADNER, J. 1994. Biological Macromolecule Crystallization Database, Version 3.0: new features, data and the

NASA archive for protein crystal growth data. *Acta Crystallographica Section D: Biological Crystallography,* 50**,** 408-413.

GILLILAND, G. L., TUNG, M. & LADNER, J. E. 2002. The biological macromolecule crystallization database: crystallization procedures and strategies. *Acta Crystallographica Section D: Biological Crystallography,* 58**,** 916-920.

GINGRAS, A. R., GIRIJA, U. V., KEEBLE, A. H., PANCHAL, R., MITCHELL, D. A., MOODY, P. C. & WALLIS, R. 2011. Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: implications for complement activation. *Structure,* 19**,** 1635-1643.

GIRIJA, U. V., GINGRAS, A. R., MARSHALL, J. E., PANCHAL, R., SHEIKH, M. A., HARPER, J. A., GÁL, P., SCHWAEBLE, W. J., MITCHELL, D. A. & MOODY, P. C. 2013. Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation. *Proceedings of the National Academy of Sciences,* 110**,** 13916-13920.

GRAHAM, F. L. & VAN DER EB, A. J. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *virology,* 52**,** 456-467.

GUSS, J. M. 2011. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology, by Bernard Rupp. Taylor & Francis.

HADDERS, M. A., BUBECK, D., ROVERSI, P., HAKOBYAN, S., FORNERIS, F., MORGAN, B. P., PANGBURN, M. K., LLORCA, O., LEA, S. M. & GROS, P. 2012. Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell reports,* 1**,** 200-207.

HAMBLETON, S., VALEYEV, N. V., MURANYI, A., KNOTT, V., WERNER, J. M., MCMICHAEL, A. J., HANDFORD, P. A. & DOWNING, A. K. 2004. Structural and functional properties of the human notch-1 ligand binding region. *Structure,* 12**,** 2173-2183.

HARBOE, M. & MOLLNES, T. E. 2008. The alternative complement pathway revisited. *Journal of cellular and molecular medicine,* 12**,** 1074-1084.

HOURCADE, D. E. 2006. The role of properdin in the assembly of the alternative pathway C3 convertases of complement. *Journal of Biological Chemistry,* 281**,** 2128-2132.

HUPPERTZ, H.-I. & GIRSCHICK, H. 2016. Lyme Borreliosis. *Handbook of Systemic Autoimmune Diseases.* Elsevier.

IACOVACHE, I., VAN DER GOOT, F. G. & PERNOT, L. 2008. Pore formation: an ancient yet complex form of attack. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778**,** 1611-1623.

JANEWAY, C. A., MURPHY, K., TRAVERS, P. & WALPORT, M. 2009. *Immunobiologie*, De Boeck Supérieur.

JANEWAY JR, C. A. 2001. How the immune system protects the host from infection. *Microbes and infection,* 3**,** 1167-1171.

JENSENIUS, H., KLEIN, D. C., VAN HECKE, M., OOSTERKAMP, T. H., SCHMIDT, T. & JENSENIUS, J. C. 2009. Mannan-binding lectin: structure, oligomerization, and flexibility studied by atomic force microscopy. *Journal of molecular biology,* 391**,** 246-259.

KABSCH, W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 133-144.

KAUFMAN, R. J., DAVIES, M. V., WASLEY, L. C. & MICHNICK, D. 1991. Improved vectors for stable expression of foreign genes in mammalian cells by use of the untranslated leader sequence from EMC virus. *Nucleic acids research,* 19**,** 4485-4490.

KAWAI, T. & AKIRA, S. 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature immunology,* 11**,** 373.

KISHORE, U. & REID, K. B. 2000. C1q: structure, function, and receptors. *Immunopharmacology,* 49**,** 159-170.

KRAUSS, I. R., FERRARO, G., PICA, A., MÁRQUEZ, J. A., HELLIWELL, J. R. & MERLINO, A. 2017. Principles and methods used to grow and optimize crystals of protein–metallodrug adducts, to determine metal binding sites and to assign metal ligands. *Metallomics,* 9**,** 1534-1547.

LEBBINK, R. J., DE RUITER, T., ADELMEIJER, J., BRENKMAN, A. B., VAN HELVOORT, J. M., KOCH, M., FARNDALE, R. W., LISMAN, T., SONNENBERG, A. & LENTING, P. J. 2006. Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *The Journal of experimental medicine,* 203**,** 1419-1425.

LEBBINK, R. J., VAN DEN BERG, M. C., DE RUITER, T., RAYNAL, N., VAN ROON, J. A., LENTING, P. J., JIN, B. & MEYAARD, L. 2008. The soluble leukocyte-associated Ig-like receptor (LAIR)-2 antagonizes the collagen/LAIR-1 inhibitory immune interaction. *The Journal of Immunology,* 180**,** 1662-1669.

LI, X., LIU, X., BECK, D. S., KANTOR, F. S. & FIKRIG, E. 2006. Borrelia burgdorferi lacking BBK32, a fibronectin-binding protein, retains full pathogenicity. *Infection and immunity,* 74**,** 3305-3313.

LIN, Y. P., CHEN, Q., RITCHIE, J. A., DUFOUR, N. P., FISCHER, J. R., COBURN, J. & LEONG, J. M. 2015. Glycosaminoglycan binding by B orrelia burgdorferi adhesin BBK 32 specifically and uniquely promotes joint colonization. *Cellular microbiology,* 17**,** 860-875.

LOVELACE, L. L., COOPER, C. L., SODETZ, J. M. & LEBIODA, L. 2011. Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. *Journal of Biological Chemistry,* 286**,** 17585-17592.

MARKIEWSKI, M. M. & LAMBRIS, J. D. 2007. The role of complement in inflammatory diseases from behind the scenes into the spotlight. *The American journal of pathology,* 171**,** 715-727.

MATSUSHITA, M. 2013. Ficolins in complement activation. *Molecular immunology,* 55**,** 22-26.

MATSUSHITA, M., ENDO, Y. & FUJITA, T. 2013. Structural and functional overview of the lectin complement pathway: its molecular basis and physiological implication. *Archivum immunologiae et therapiae experimentalis,* 61**,** 273-283.

MCPHERSON, A. 1991. A brief history of protein crystal growth. *Journal of crystal growth,* 110**,** 1-10.

MCPHERSON, A. & GAVIRA, J. A. 2014. Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications,* 70**,** 2-20.

MERLE, N. S., CHURCH, S. E., FREMEAUX-BACCHI, V. & ROUMENINA, L. T. 2015. Complement system part I–molecular mechanisms of activation and regulation. *Frontiers in immunology,* 6**,** 262.

MEYAARD, L., ADEMA, G. J., CHANG, C., WOOLLATT, E., SUTHERLAND, G. R., LANIER, L. L. & PHILLIPS, J. H. 1997. LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes. *Immunity,* 7**,** 283-290.

MEYAARD, L., HURENKAMP, J., CLEVERS, H., LANIER, L. L. & PHILLIPS, J. H. 1999. Leukocyte-associated Ig-like receptor-1 functions as an inhibitory receptor on cytotoxic T cells. *The Journal of Immunology,* 162**,** 5800-5804.

MIYAIKE, J., IWASAKI, Y., TAKAHASHI, A., SHIMOMURA, H., TANIGUCHI, H., KOIDE, N., MATSUURA, K., OGURA, T., TOBE, K. & TSUJI, T. 2002. Regulation of circulating immune complexes by complement receptor type 1 on erythrocytes in chronic viral liver diseases. *Gut,* 51**,** 591-596.

MOREAU, C., BALLY, I., CHOUQUET, A., BOTTAZZI, B., GHEBREHIWET, B., GABORIAUD, C. & THIELENS, N. 2016. Structural and functional characterization of a single-chain form of the recognition domain of complement protein C1q. *Frontiers in immunology,* 7**,** 79.

MORTENSEN, S. A., SANDER, B., JENSEN, R. K., PEDERSEN, J. S., GOLAS, M. M., JENSENIUS, J. C., HANSEN, A. G., THIEL, S. & ANDERSEN, G. R. 2017. Structure and activation of C1, the complex initiating the classical pathway of the complement cascade. *Proceedings of the National Academy of Sciences,* 114**,** 986-991.

NAN, R., FURZE, C. M., WRIGHT, D. W., GOR, J., WALLIS, R. & PERKINS, S. J. 2017. Flexibility in mannan-binding lectin-associated serine proteases-1 and-2 provides insight on lectin pathway activation. *Structure,* 25**,** 364-375.

NAUTA, A. J., TROUW, L. A., DAHA, M. R., TIJSMA, O., NIEUWLAND, R., SCHWAEBLE, W. J., GINGRAS, A. R., MANTOVANI, A., HACK, E. C. & ROOS, A. 2002. Direct binding of C1q to apoptotic cells and cell blebs induces complement activation. *European journal of immunology,* 32**,** 1726-1736.

NORDKAMP, M. J. O., BOROSS, P., YILDIZ, C., JANSEN, J. M., LEUSEN, J. H., WOUTERS, D., URBANUS, R. T., HACK, C. E. & MEYAARD, L. 2014a. Inhibition of the classical and lectin pathway of the complement system by recombinant LAIR-2. *Journal of innate immunity,* 6**,** 284-292.

NORDKAMP, M. J. O., KOELEMAN, B. P. & MEYAARD, L. 2014b. Do inhibitory immune receptors play a role in the etiology of autoimmune disease? *Clinical Immunology,* 150**,** 31-42.

NORIS, M. & REMUZZI, G. Overview of complement activation and regulation. Seminars in nephrology, 2013. Elsevier, 479-492.

NORMAN, D., BARLOW, P., BARON, M., DAY, A., SIM, R. & CAMPBELL, I. 1991. Three-dimensional structure of a complement control protein module in solution. *Journal of molecular biology,* 219**,** 717-725.

PHILLIPS, A. E., TOTH, J., DODDS, A. W., GIRIJA, U. V., FURZE, C. M., PALA, E., SIM, R. B., REID, K. B., SCHWAEBLE, W. J. & SCHMID, R. 2009. Analogous interactions in initiating complexes of the classical and lectin pathways of complement. *The Journal of Immunology,* 182**,** 7708-7717.

POTTERTON, E., BRIGGS, P., TURKENBURG, M. & DODSON, E. 2003. A graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Biological Crystallography,* 59**,** 1131-1137.

PROBERT, W. S. & JOHNSON, B. J. 1998. Identification of a 47 kDa fibronectin-binding protein expressed by Borrelia burgdorferi isolate B31. *Molecular microbiology,* 30**,** 1003-1015.

RICKLIN, D., HAJISHENGALLIS, G., YANG, K. & LAMBRIS, J. D. 2010. Complement: a key system for immune surveillance and homeostasis. *Nature immunology,* 11**,** 785-797.

RICKLIN, D. & LAMBRIS, J. D. 2007. Complement-targeted therapeutics. *Nature biotechnology,* 25**,** 1265-1275.

ROSA, P. A., TILLY, K. & STEWART, P. E. 2005. The burgeoning molecular genetics of the Lyme disease spirochaete. *Nature Reviews Microbiology,* 3**,** 129-143.

SELMAN, L. & HANSEN, S. 2012. Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). *Immunobiology,* 217**,** 851-863.

SERNA, M., GILES, J. L., MORGAN, B. P. & BUBECK, D. 2016. Structural basis of complement membrane attack complex formation. *Nature communications,* 7**,** 1-7.

SHARP, T. H., BOYLE, A. L., DIEBOLDER, C. A., KROS, A., KOSTER, A. J. & GROS, P. 2019. Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proceedings of the National Academy of Sciences,* 116**,** 11900-11905.

SIM, R., SCHWAEBLE, W. & FUJITA, T. 2016. Complement research in the 18th–21st centuries: progress comes with new technology. *Immunobiology,* 221**,** 1037-1045.

SON, M., SANTIAGO-SCHWARZ, F., AL-ABED, Y. & DIAMOND, B. 2012. C1q limits dendritic cell differentiation and activation by engaging LAIR-1. *Proceedings of the National Academy of Sciences,* 109**,** E3160-E3167.

STRANG, C. J., SIEGEL, R. C., PHILLIPS, M. L., POON, P. H. & SCHUMAKER, V. N. 1982. Ultrastructure of the first component of human complement: electron microscopy of the crosslinked complex. *Proceedings of the National Academy of Sciences,* 79**,** 586-590.

THIELENS, N. M., CSEH, S., THIEL, S., VORUP-JENSEN, T., ROSSI, V., JENSENIUS, J. C. & ARLAUD, G. J. 2001. Interaction properties of human mannan-binding lectin (MBL)-associated serine proteases-1 and-2, MBL-associated protein 19, and MBL. *The Journal of Immunology,* 166**,** 5068-5077.

THURMAN, J. M. & HOLERS, V. M. 2006. The central role of the alternative complement pathway in human disease. *The Journal of Immunology,* 176**,** 1305-1310.

TROUW, L., BLOM, A. & GASQUE, P. 2008. Role of complement and complement regulators in the removal of apoptotic cells. *Molecular immunology,* 45**,** 1199-1207.

VILLIERS, C., ARLAUD, G. & COLOMB, M. 1985. Domain structure and associated functions of subcomponents C1r and C1s of the first component of human complement. *Proceedings of the National Academy of Sciences,* 82**,** 4477-4481.

WALLIS, R., DODDS, A. W., MITCHELL, D. A., SIM, R. B., REID, K. B. & SCHWAEBLE, W. J. 2007. Molecular interactions between MASP-2, C4, and C2 and their activation fragments leading to complement activation via the lectin pathway. *Journal of Biological Chemistry,* 282**,** 7844-7851.

WALLIS, R., MITCHELL, D. A., SCHMID, R., SCHWAEBLE, W. J. & KEEBLE, A. H. 2010. Paths reunited: Initiation of the classical and lectin pathways of complement activation. *Immunobiology,* 215**,** 1-11.

WARD, A. E. & ROSENTHAL, B. M. 2014. Evolutionary responses of innate immunity to adaptive immunity. *Infection, Genetics and Evolution,* 21**,** 492-496.

WASEDA, Y., MATSUBARA, E. & SHINODA, K. 2011. *X-ray diffraction crystallography: introduction, examples and solved problems,* Springer Science & Business Media.

WATTS, A. 1993. Crystallization of nucleic acids and proteins. A practical approach: edited by A. Ducruix and R. Giegé, Oxford University Press; Oxford, 1992; xxiv+ 331 pages.£ 25.00. ISBN 019-963246-4. *FEBS Letters,* 319**,** 283-284.

WIJEYEWICKREMA, L. C., YONGQING, T., TRAN, T. P., THOMPSON, P. E., VILJOEN, J. E., COETZER, T. H., DUNCAN, R. C., KASS, I., BUCKLE, A. M. & PIKE, R. N. 2013. Molecular determinants of the substrate specificity of the complement-initiating protease, C1r. *Journal of Biological Chemistry,* 288**,** 15571-15580.

XIE, J., ZHI, H., GARRIGUES, R. J., KEIGHTLEY, A., GARCIA, B. L. & SKARE, J. T. 2019. Structural determination of the complement inhibitory domain of Borrelia burgdorferi BBK32 provides insight into classical pathway complement evasion by Lyme disease spirochetes. *PLoS pathogens,* 15**,** e1007659.

YASEEN, S., DEMOPULOS, G., DUDLER, T., YABUKI, M., WOOD, C. L., CUMMINGS, W. J., TJOELKER, L. W., FUJITA, T., SACKS, S. & GARRED, P. 2017. Lectin pathway effector enzyme mannan-binding lectin-associated serine protease-2 can activate native complement C3 in absence of C4 and/or C2. *The FASEB Journal,* 31**,** 2210-2219.

ZHANG, X.-L. & ALI, M. A. 2008. Ficolins: structure, function and associated diseases. *Current topics in complement II.* Springer.

ZICCARDI, R. J. 1982. Spontaneous activation of the first component of human complement (C1) by an intramolecular autocatalytic mechanism. *The Journal of Immunology,* 128**,** 2500-2504.

ZIPFEL, P. F. & SKERKA, C. 2009. Complement regulators and inhibitory proteins. *Nature Reviews Immunology,* 9**,** 729-740.