

Unravelling the mechanism of complement activation via the lectin pathway

A thesis submitted to the Faculty of Medicine and Biological Sciences,
University of Leicester, in partial fulfilment of the requirements for the
degree of Doctor of Philosophy.

Christopher Michael Furze BSc MSc

Department of Infection, Immunity and Inflammation

University of Leicester

February, 2013

Unravelling the mechanism of complement activation via the lectin pathway.

Christopher Michael Furze

Department of Infection, Immunity and Inflammation

University of Leicester

Submitted for the degree of Doctor of Philosophy

February, 2013

Abstract

Activation of complement is involved in the clearance of foreign pathogens, altered-self and apoptotic cells. It proceeds through the Classical, Lectin or Alternative pathways and ultimately results in destruction of the cell through lysis by the membrane attack complex. Activation of the Lectin pathway is triggered by the recognition of carbohydrate targets by families of collagenous proteins, called mannose-binding lectins (MBLs) and ficolins. These recognition proteins activate associated proteases called MBL-associated serine proteases (MASPs). Once activated, the MASPs cleave downstream targets to initiate the complement cascade. MASP-2 circulates in a complex with MBL and activates when it recognises mannose-type sugars. The exact molecular mechanisms of the activation of MASP-2 by MBL remains unclear and this is the focus of the current work.

In this thesis I demonstrate that MASP binding can be introduced into pulmonary surfactant protein A (SP-A), a protein with a similar architecture to MBL, but which cannot bind MASPs to activate complement, through a series of substitutions to the collagenous domain. Surprisingly, introduction of the MASP-binding site results in constitutive activation of MASP-2, even in the absence of a carbohydrate target, thus lacking the control present in MBL. I then investigated the basis for this control by producing chimeric proteins of MBL and SP-A. The chimeras demonstrate that target-specific activation originates from a portion of MBL comprising the carbohydrate-recognition domains (CRDs) and neck region comprising an α -helical coiled coil. Additional work using a modified MBL that recognises galactose, subsequently demonstrates that the specific activity does not originate from the CRDs themselves, suggesting the neck domain plays an important role in the activation of MASPs.

Further work in this thesis revolves around a newly discovered collectin, CL-K1 which is probably involved in complement activation and also important developmental processes. Mutations within the CRD of CL-K1 result in developmental disorders known as 3MC syndrome. The work in this thesis shows that one of these mutants is destabilized and defective in sugar binding hence revealing the likely molecular basis of disease. In addition, the structure of the CRD has been solved by X-ray crystallography enabling the key residues to be visualised.

Acknowledgements

I would like to thank:

Dr. 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.

Dr. Umakhanth Girija for his help and support in the laboratory.

Dr. Daniel Mitchell and *Dr. Katrine Wallis* for providing the use of the surface plasmon resonance facilities at the University of Warwick.

Dr. Alexander Gingras for his knowledge and help with protein crystallisation.

Dr. Julia Toth for supplying the SP-A DNA.

Dr Catherine Williams and *Mrs Roshni Panchal* for the technician support.

The MRC and Department of Infection, Immunity and Inflammation for providing the funding to undertake this study.

Antonia, for her encouragement, patience and support.

My family for all they have done for me.

Table of Contents

Abstract.....	1
Acknowledgements.....	2
Table of Contents.....	3
Chapter 1 – General Introduction	7
1.1 - The Immune system.....	7
1.1.1 - The Complement system.....	8
1.1.2 - The Classical Pathway Activation	10
1.1.3 - The Lectin Pathway Activation.....	13
1.1.4 - The Alternative Pathway Activation	14
1.1.5 - The Terminal Pathway	15
1.1.6 - Other Roles of Complement Activation.....	17
1.1.7 - Regulation of Complement	18
1.2 - Lectins, C-type Lectins and Collectins	21
1.2.1 - Mannose-Binding Lectin.....	21
1.2.2 - Structural organisation of MBL	22
1.2.3 - Genetics of MBL.....	24
1.2.4 - Recognition of sugar ligands by MBL	27
1.2.5 - Binding to monosaccharides	27
1.2.6 - High avidity binding to sugar structures on foreign cells	29
1.2.7 - Microbial Targets of MBL	30
1.2.8 – MBL and disease association	30
1.2.9 – Pulmonary Surfactant Collectins	33
1.2.10 - Surfactant Protein-A.....	34
1.2.11 - Structural organisation of SP-A	34
1.2.12 - Genetics of SP-A.....	35
1.2.13 – MBL and SP-A receptors	35
1.3 - MBL-Associated Serine Proteases (MASPs)	38
1.3.1 - Structure and Function of MASPs	38
1.3.2 - MASP expression and gene organisation.....	40
1.3.3 - MASP regulators	42
1.4 – General aims of the thesis	43

Chapter 2 –Introducing MASP-binding and Complement Activation into Human Pulmonary Surfactant Protein-A.....	44
2.1 Introduction and Objectives	44
2.2 - Materials and Methods.....	48
2.2.1 – PCR Mutagenesis to introduce the MASP-binding motif into SP-A	48
2.2.2 – PCR Mutagenesis to modify the SP-A kink	48
2.2.2 - Bacterial Transformation	51
2.2.3 – Cell culture	52
2.2.4 – Mammalian Cell Transfection.....	52
2.2.5 – Protein Expression.....	54
2.2.6 – Sugar-Sepharose coupling	54
2.2.7 – MBL and SP-A Purification	55
2.2.8 – His-Tagged MASP-2 Purification	56
2.2.9 – Surface Plasmon Resonance.....	57
2.2.10 – MASP-2k Activation Assays.....	58
2.3 Results.....	59
2.3.1 - Engineering MASP binding into SP-A	59
2.3.1 – Cloning	60
2.3.2 – Expression and Purification.....	62
2.3.3 – Binding to MASP-2.....	64
2.3.4 – MASP-2k Activation Assays.....	67
2.4 Discussion	76
Chapter 3 – Control of MASP activation.....	80
3.1 Introduction and Objectives	80
3.2 Materials and Methods.....	83
3.2.1 – PCR construction of MBL/SP-A chimeras.....	83
3.2.2 – Cloning, Transfection and Expression	84
3.2.3 – MBL/SP-A and GBL Chimera Purification	85
3.2.4 – MASP-2k Activation Assays.....	85
3.3 - Results.....	86
3.3.1 – Cloning	86
3.3.2 Production of chimeras in CHO cells.....	87
3.3.3 – MASP-2K activation by MBL-SPA chimeras	88

3.3.4 - Switching sugar specificity in MBL from mannose to galactose: creation of a galactose-binding lectin	92
3.3.5 – Purification of GBL.....	92
3.3.5 – Activation of MASP-2K by the galactose-binding lectin.....	93
3.4 Discussion	97
Chapter 4 – Collectin 11	99
4.1 Introduction and Objectives	99
4.2 - Materials and Methods.....	102
4.2.1 – PCR amplification of the neck and CRD regions of CL-K1	102
4.2.2 – Cloning	103
4.2.3 – Competent Cell Production	103
4.2.4 – Expression	103
4.2.5 – Inclusion Body Preparation	104
4.2.6 – Refolding Protocol.....	105
4.2.7 – Ion Exchange Chromatography	105
4.2.8 – Gel Filtration	105
4.2.9 – Crystal Condition Screening.....	106
4.2.10 – X-Ray Diffraction.....	108
4.2.11 – Trypsin Digests.....	109
4.2.12 – Surface Plasmon Resonance.....	110
4.3 Results	111
4.3.1 – Purification of wild-type CL-K1 head and neck regions.....	111
4.3.2 - Binding of CL-K1 to HIV gp120.....	113
4.3.3 - Analysis of mutant CL-K1s associated with 3MC syndrome.....	115
4.3.4 – Wild-type and Ser ¹⁶⁹ Pro CRD Trypsin Digestion.....	118
4.3.5 – Purification of additional CL-K1 CRD mutants.....	120
4.3.6 – Crystallisation and X-ray diffraction.....	123
4.4 Discussion	124
4.4.1 - Mutant CL-K1s associated with 3MC syndrome.....	124
Chapter 5 – General Discussion.....	126
5.1 – Interactions between MBL-MASPs	126
5.2 – Collectin 11	130
5.3 – Future Work	137

5.3.1 – MBL/MASP Activation.....	137
5.3.2 – Galactose-binding lectin.....	137
5.3.3 – Collectin 11	138
Appendix 1 - DNA Sequences.....	139
Bibliography	150

Chapter 1 – General Introduction

1.1 - The Immune system

The immune system is the overarching name given to the collection of processes and molecules utilised by an organism to prevent and to clear an attack by invading viruses, bacteria and parasites. It can be divided between the innate immune response and the adaptive immune response.

The role of the innate immune response is to provide a first-line generic defence against invading micro-organisms by recognising self from non-self. This is achieved through a number of mechanisms including:

- The release of cytokines acting to control inflammation, induce an anti-viral state, recruit immune and regulate the adaptive response.
- Sentinel cells such as dendritic cells which bind cytokines and non-self targets using pattern recognition receptors on the cell surface. They can present foreign objects to the adaptive immune response via the MHC.
- The complement system can induce cytolysis, opsonisation, activation of inflammation and solubilisation of immune complexes.
- Natural Killer (NK) cells recognising and destroying infected cells by detecting ‘missing-self’ and ‘altered-self’ signals.

The innate immune response is permanently functioning against the constant barrage of infectious diseases encountered during life whilst the adaptive remains sequestered unless the innate response is unable to clear the infection faster than it occurs.

If the infection cannot be contained by the innate response, the adaptive response is recruited and acts in a target specific manner. This is achieved via two mechanisms:

- The humoral response uses antibodies to recognise the specific pathogen.
- The cell-mediated response recognising individual infected cells and destroying them without damaging neighbouring cells.

A key feature of the adaptive response is that of 'memory' where if the organism is exposed to a previously cleared infection, the action of the adaptive response is almost immediate.

1.1.1 - The Complement system

The complement system as previously mentioned belongs to the innate immune response, although it also serves as a bridge between the innate and adaptive immune systems. It provides a primary defence against both pathogens and altered-self cells (Muller-Eberhard, 1988, Reid, 1983). It consists of a number of proteins and enzymes circulating in their zymogenic or inactive state in the blood.

Once initiated it results in a series of specific downstream activations of other proteins in the pathway. This cascade leads to the release of cytokines, amplification of the response and the lysis of the target through assault by the membrane-attack complex (see below). The debris from cell lysis is subsequently cleared by host leukocytes that recognise complement proteins attached to the fragments.

There are three main modes of action by which complement activation can occur. These are the classical pathway, the lectin pathway and the alternative pathway. Complement can function in the absence of antibodies by recognising the cells directly or in the case of the classical pathway by recognising antibody-antigen complexes. In this way they play a secondary role in the clearance of targets of the adaptive immune response (Burton *et al.*, 1980, Duncan *et al.*, 1988, Perkins *et al.*, 1991, Sim *et al.*, 1991, Roos *et al.*, 2001).

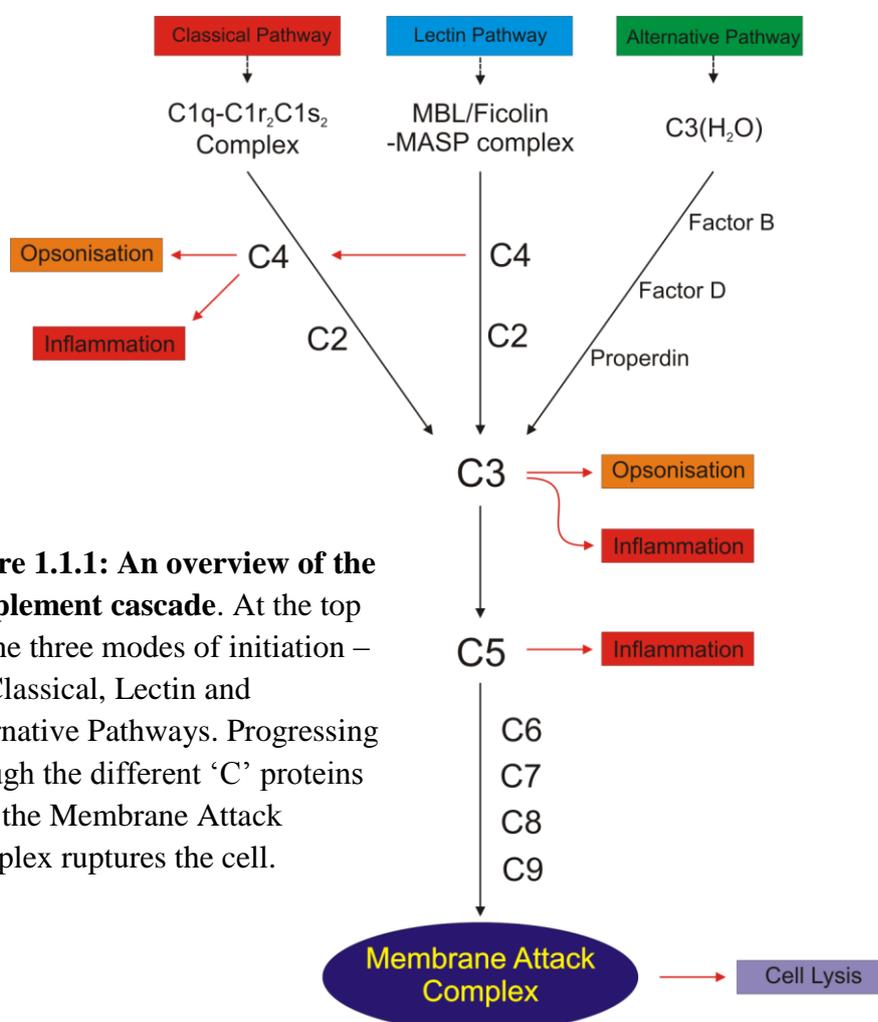


Figure 1.1.1: An overview of the complement cascade. At the top are the three modes of initiation – the Classical, Lectin and Alternative Pathways. Progressing through the different ‘C’ proteins until the Membrane Attack Complex ruptures the cell.

1.1.2 - The Classical Pathway Activation

The C1 complex is responsible for recognition and activation of the classical pathway.

The complex consists of C1q, C1r and C1s (Arlaud *et al.*, 1987). C1q is the recognition protein whilst C1r and C1s are serine proteases. C1q is a large multimeric protein which comprises three polypeptide chains (A, B and C) which are encoded by three homologous genes. Each chain consists of a short N-terminal domain, followed by a collagen-like domain and finally a globular domain at the C-terminus.

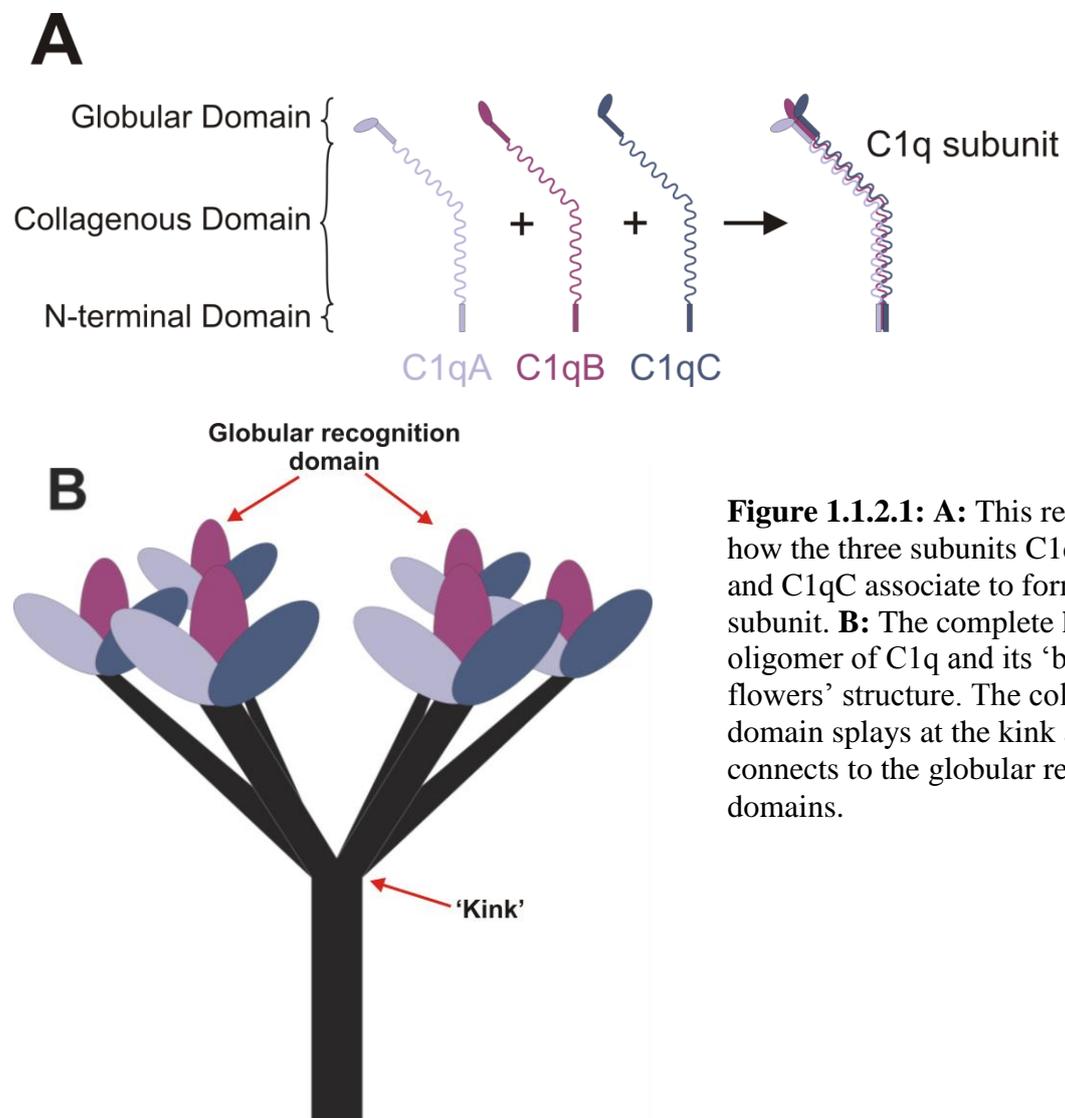


Figure 1.1.2.1: **A:** This represents how the three subunits C1qA, C1qB and C1qC associate to form the C1q subunit. **B:** The complete higher oligomer of C1q and its 'bouquet of flowers' structure. The collagenous domain splays at the kink and connects to the globular recognition domains.

The A, B and C chains associate to form the C1q subunit which clusters via the N-terminal domain and the first part of the collagenous domain into hexamers. The subunits splay apart at a kink that is present within the collagenous domain and form a bouquet like structure (Sim *et al.*, 1991). The globular recognition domains form the 'flowers' of the bouquet (Figure 1.1.2.1.B).

There are four protease components of the C1 complex, two C1r proteins and two C1s proteins forming heterotetramers. C1r and C1s are homologues and share the same domain organisation. Each consists of two CUB domains (CUB1 and CUB2) flanking a Ca^{2+} -binding EGF-like domain attached to this are two Complement Control Protein (CCP) modules before a final serine protease domain (Schwaeble *et al.*, 2002, Sim *et al.*, 2004). The composition of domains is also identical in MBL-Associated Serine Proteases described later in this chapter (figure 1.3.1.1).

The C1r proteases dimerise through binding interactions between the serine protease and CCP domains. In turn the two C1s proteases each bind to one of the C1r proteins in a Ca^{2+} -dependent manner through the CUB1-EGF domains to form the zymogenic $\text{C1r}_2\text{S}_2$. The zymogenic heterotetramer also binds in a Ca^{2+} -dependent manner to C1q producing the C1 complex (Gal *et al.*, 2007).

The Classical Pathway is initiated when the C1 complex binds to the surface of a pathogen or to immune complexes (such as IgM and IgG). When this occurs, a conformational change in the C1q protein leads to the activation of the two C1s serine proteases (Dodds *et al.*, 1978). These cleave complement component C4 to C4a and C4b. The C4a portion is released as an anaphylatoxin and the C4b binds rapidly to the

pathogen/cell surface via a reactive thioester bond (Dodds *et al.*, 1996). C2 binds to the C4b fragment and the C4bC2 complex is then cleaved by C1s to generating C4bC2a (C3 convertase) (Moller-Kristensen *et al.*, 2003, Wallis *et al.*, 2007). The C3 convertase protein is able to cleave numerous C3 molecules into C3b which is deposited onto the cell surface. C3 convertase can bind to the C3b to produce C4b2aC3b also known as C5 convertase which is responsible for cleaving C5 into C5a and C5b to initiate the terminal pathway (figure 1.1.2.2)

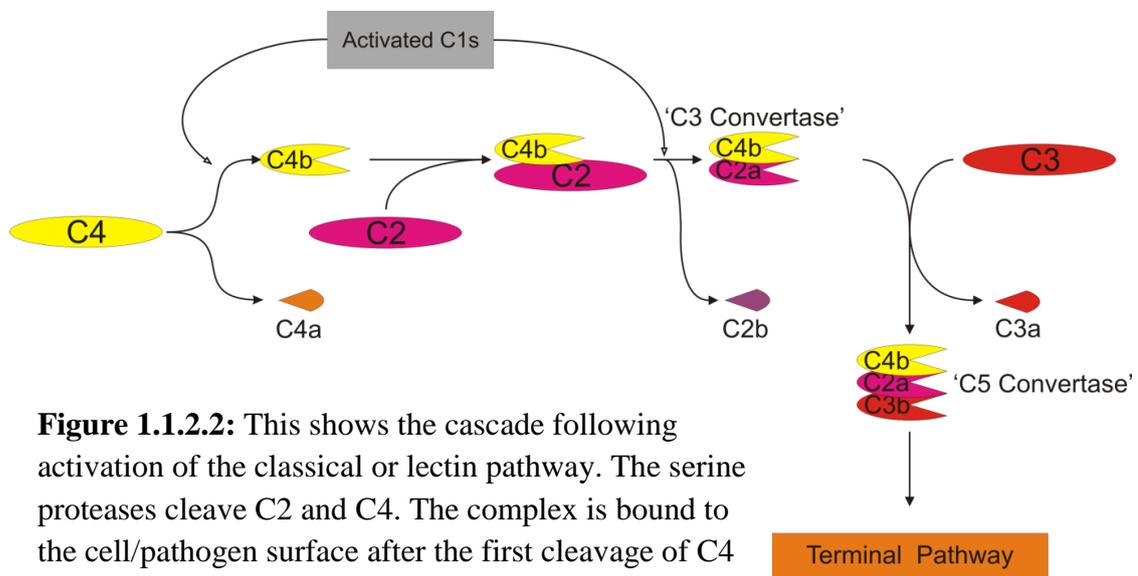


Figure 1.1.2.2: This shows the cascade following activation of the classical or lectin pathway. The serine proteases cleave C2 and C4. The complex is bound to the cell/pathogen surface after the first cleavage of C4 to C4b.

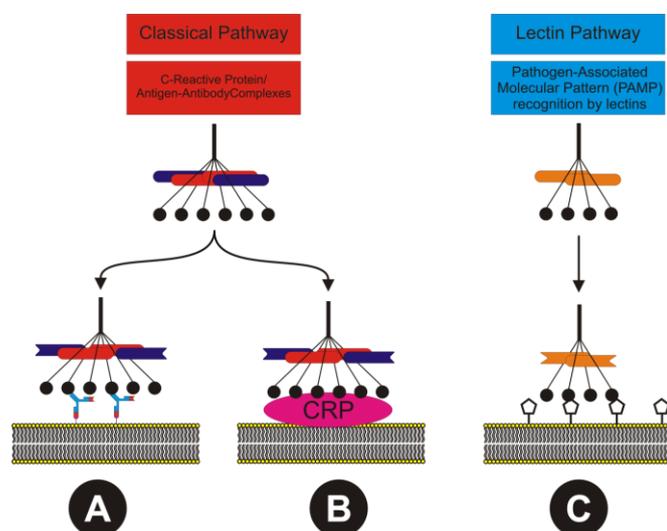
Although C1 recognises antibody-antigen complexes, it is also able to bind to a variety of bacterial cells wall components such as lipid A, nucleic acids, prions and ligand-bound C-reactive protein with no requirement for antibodies. This bridges the gap between the adaptive and innate systems (Blanquet-Grossard *et al.*, 2005, Mitchell *et al.*, 2007, Sim *et al.*, 1994).

1.1.3 - The Lectin Pathway Activation

Mannose-Binding Lectin (MBL), ficolins and collectin 11 circulate as heterogeneous complexes with different zymogen MBL-Associated Serine Proteases (MASPs) predominantly as 1:1 complexes (Chen *et al.*, 2001). Like C1q, MBLs and ficolins possess an N-terminal collagenous domain and a C-terminal pathogen recognition domain: a C-type carbohydrate-recognition domain in MBL and a fibrinogen-like domain in ficolins. There are three MASP proteins (called MASP-1, -2 and -3) and two small non-enzymatic proteins MAP-19 and MAP-44 that are the products of alternative splicing of the MASP-2 and MASP-1 genes, respectively (Schwaeble *et al.*, 2002, Wallis *et al.*, 2000). These will be discussed later in this chapter. For the lectin pathway, MASP-2 is the key enzyme that activates downstream complement components.

When MBL/MASP-2 or ficolin/MASP-2 complexes bind to the surface of a pathogen, MASP-2 activates through autolysis or via MASP-1 (Heja *et al.*, 2012). The activated MASP-2 cleaves C4 into C4a and C4b and the C4b binds to the pathogen surface (Rossi *et al.*, 2001). MASP-2 also cleaves C2 and the C2a fragment binds to C4b producing the C3 convertase (C4b2a). The cascade continues as previously described for the classical pathway through C5 convertase to the terminal pathway

Figure 1.1.5.1: Comparison between the Classical and Lectin Pathway recognition complexes, C1 and MBL/MASPs. C1q binds both the target specific IgG and IgM antibodies (A) as well the non-specific C-Reactive Protein (CRP) (B). In contrast MBL binds to mannose-like residues on the surfaces of pathogens (C).



1.1.4 - The Alternative Pathway Activation

Unlike the Classical and Lectin Pathways there is no target recognition molecule involved in the activation of complement via the Alternative Pathway (figure 1.1.4.1). Activation occurs by spontaneous hydrolysis of C3 in plasma to form C3(H₂O). The hydrolysed C3 resembles the structure of C3b and has a similar function. C3(H₂O) binds factor B in the alternative pathway and this is subsequently cleaved by factor D into two fragments. The Ba fragment diffuses away whilst the Bb fragment remains bound to C3(H₂O) resulting in a C3(H₂O)Bb complex. This functions as a mobile C3 convertase that cleaves C3 into the two fragments C3a and C3b (Fearon *et al.*, 1975a). This is able to cleave many C3 molecules depositing an increasing concentration of C3b fragments onto neighbouring structures. Recent studies suggest that MASP-1 and -3 cleave factor D and MASP-3 also cleaves factor B, suggesting that lectin pathway components also participate in activation of the alternative pathway (Iwaki *et al.*, 2011, Takahashi *et al.*, 2010).

Factor B attaches to membrane bound C3b in an Mg²⁺-dependent manner before being cleaved by factor D to produce C3bBb, the membrane-phase alternative pathway C3 convertase. The C3bBb is stabilised through association with properdin (P) (Farries *et al.*, 1988) (Fearon *et al.*, 1975c) and through further binding to additional C3b fragments forms the alternative pathway C5 convertase.

Although there is no differentiation directly by C3b between self and non-self cells, deposition of C3b onto self cells is prevented by regulatory and inhibitory mechanisms of the host. Factor H is the major plasma regulator of C3b and acts in two ways. Firstly it binds C3b blocking the binding of Factor B and displacing Bb. This prevents

formation of the C3bBb convertase protein, consequently no new C3b can be generated. Secondly Factor H acts as a cofactor of Factor I to promote the proteolytic cleavage of C3b to the inactive C3c and C3d fragments (Perkins *et al.*, 2012). Factor H is absent from most invading pathogens and so leaves them exposed to attack via the alternative pathway. Other regulators of complement are discussed later in the chapter in section 1.1.7.

1.1.5 - The Terminal Pathway

The classical, lectin and alternative pathway converge at the cleavage of the C5 protein by C5 convertase (either C4b2aC3b from the classical or lectin pathways or C3bBbC3b from the alternative pathway). This cleavage results in C5a being released into the plasma which in conjunction with C3a and C4a enable recruitment of immune cells to the site of activation (Frank *et al.*, 1991). C5b binds C6 and C7, the resulting C5b67 complex binds to plasma membranes and recruits C8 and numerous C9 molecules to form the membrane-attack complex. The C5b6789(n) complex forms a transmembrane channel (Muller-Eberhard, 1985) which permits diffusion of ions into and out of the cell, the loss of ATP from the cell and an influx of water due to the osmotic pressure (Bhakdi *et al.*, 1991, Muller-Eberhard, 1985). Once these pores have formed, lysis will occur resulting in the destruction of the cell (figure 1.1.5.1).

In addition to the complement-mediated killing of invading pathogens, it is utilised in the clearance of apoptotic and necrotic cells (Whaley *et al.*, 1989) and the regulation of the adaptive immune system (Dempsey *et al.*, 1996).

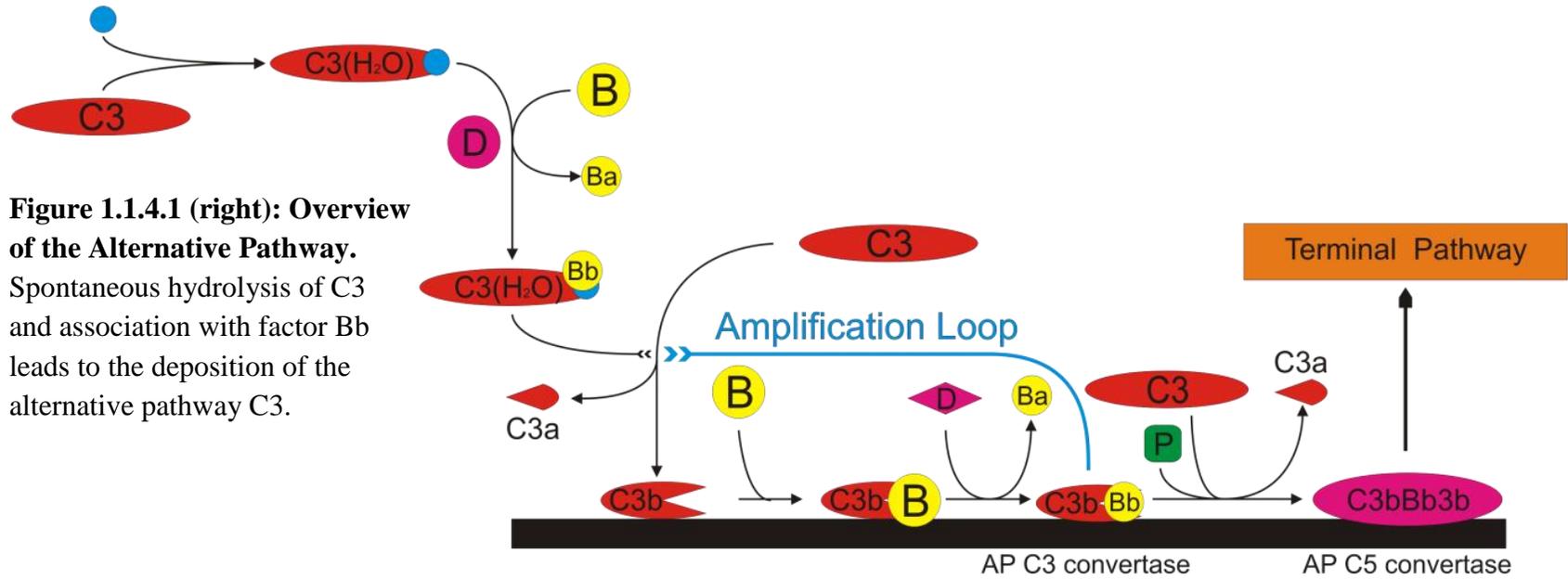
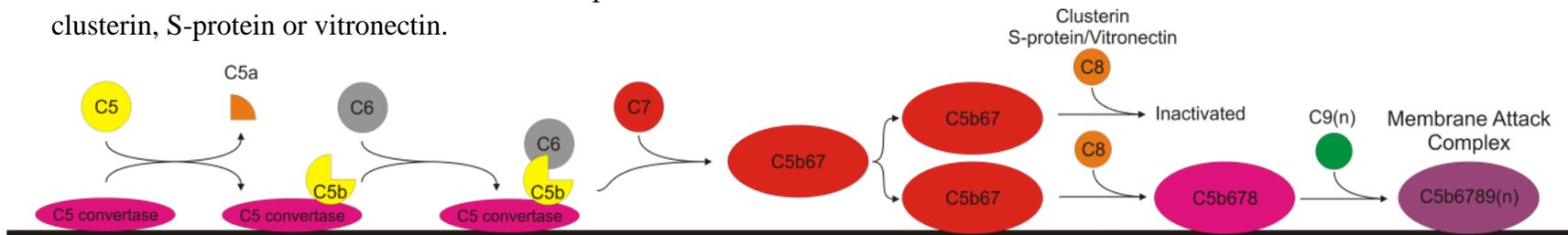


Figure 1.1.4.1 (right): Overview of the Alternative Pathway. Spontaneous hydrolysis of C3 and association with factor Bb leads to the deposition of the alternative pathway C3.

Figure 1.1.5.1 (below): Overview of the Terminal Pathway: The association of C5 convertase with downstream proteins can either lead to formation of the membrane attack complex or inactivation clusterin, S-protein or vitronectin.



1.1.6 - Other Roles of Complement Activation

The complement system aids in the neutralisation and clearance of pathogens through multiple mechanisms.

Complement activation stimulates phagocytosis through complement receptors (CR) present on the surface of macrophages and leukocytes. CR1 binds to C3b and C4b on the surface of pathogens and when this interaction is coupled with binding of IgG to the Fc γ receptor or C5a to the C5a receptor, it enhances phagocytosis (Campbell *et al.*, 1988). CR3 and CR4 function by stimulating phagocytosis against cells that have been opsonised by iC3b. This is a product of the degradation of C3b by factor I which releases the small C3f fragment.

B cell activation is another consequence of complement activation. Follicular dendritic cells (FDCs) which are located in secondary lymphoid organs such as the spleen and lymph nodes are believed to pass signals to B cells for their survival and continued recirculation. FDCs express CR1, CR2 and CR3 sequestering antigens in their organs of residence. Recruitment of the adaptive immune response can occur by lowering of the threshold of B lymphocyte activation. This occurs when iC3b or C3dg, another degradation fragment, binds to the CR2 of the B lymphocyte co-receptor complex (Campbell *et al.*, 1988).

Erythrocytes expressing CR1 bind C3b and C4b opsonins that have bound to cell fragments and immune complexes. The erythrocytes transport the cellular debris to the spleen and liver where they are degraded by macrophages (Campbell *et al.*, 1988).

The fragments C3a, C4a and C5a are anaphylatoxins. These act upon nearby blood vessels stimulating the local inflammatory responses. The recruitment of antibodies is achieved through an increase in vascular permeability. This also permits access by phagocytic cells and complement proteins to the site of infection. The anaphylotoxins induce histamine release via activation of mast cells and basophilic leukocytes. Additionally, they also cause the leukocytes to produce superoxide (Gerard *et al.*, 1994).

1.1.7 - Regulation of Complement

Uncontrolled activation of complement would have a massive destructive effect on the host organism. The alternative pathway initiates from the continued low-level activation of C3 which if it were to remain unregulated would drive the deposition of complement components onto any surface, whether it be either invading pathogen or host cell. Regulation is achieved through cell surface proteins that act as co-factors aiding in the degradation of host cell bound complement protein and also through circulating plasma proteins that inactivate unbound complement proteins.

Four of the main cell surface complement regulatory proteins are:

- Membrane co-factor protein (CD46)
- Decay-accelerating factor
- CR1
- CD59

Membrane co-factor protein is expressed fairly ubiquitously on most cells. It binds to C3b and C4b and aids degradation by the plasma protease factor I (Liszewski *et al.*, 1991). Decay-accelerating factor displaces C2a from the classical/lectin pathway C3 convertase and Bb from the C3 convertase of the alternative pathway (Burge *et al.*, 1981). Like decay-accelerating factor, CR1 disrupts C3 convertase and also functions as a co-factor for factor I mediated degradation of C3b and C4b. It is expressed on the surface of polymorphonuclear leukocytes, erythrocytes, B-lymphocytes and follicular dendritic cells (Sim, 1985). In the late stages of complement activation CD59 prevents C9 binding to C5b-8 with the result that the membrane-attack complex is unable to form, preserving the integrity of the host cell.

Plasma proteins that regulate complement activation include:

- C1 inhibitor
- C4-binding protein
- Factor H
- Factor I
- S-protein/vitronectin
- Clusterin
- Properdin

The target of the C1 inhibitor is the activated C1 complex, MBL/MASP complexes and ficolin/MASP complexes, preventing the cleavage of C2 and C4. It is a serpin that binds to C1r and C1s (and also MASPs) leading to irreversible inhibition and subsequent destruction (Reboul *et al.*, 1977, Sim *et al.*, 1979). C4-binding protein is a co-factor of

factor I and aids in degrading the C4b protein. Factor H competes with factor B for binding C3b and operates as another co-factor for factor I in the degradation of C3b. Factor H is localised to mammalian cell surfaces through binding to sialic acid. This helps protect the host cells from the alternative pathway whilst providing the opportunity for the C3b to bind to bacterial cells which lack this molecule and will lack localised levels of factor H as a result.

The target of clusterin and S-protein/vitronectin is the membrane-binding site of the C5b67 complex in the terminal pathway. By occupying this site, they prevent the binding to the cell surface and the completion of the membrane-attack complex (Preissner, 1991, Rosenberg *et al.*, 1995, Tschopp *et al.*, 1993).

Of all the plasma proteins that regulate complement, only properdin acts as a positive regulator of complement. This is achieved via properdin binding to and stabilising the C3 and C5 convertase complexes (Fearon *et al.*, 1975b, Muller-Eberhard *et al.*, 1980). Depletion of properdin from sera prevents activation of the alternative pathway which can be restored in a dose-dependent manner by the addition of purified properdin. Consequently, changes in the local concentrations of properdin could regulate alternative pathway activation (Schwaeble *et al.*, 1999).

1.2 - Lectins, C-type Lectins and Collectins

The term lectin is applied to sugar binding proteins. They are found ubiquitously throughout nature and can bind either soluble sugars or those that are part of a glycoprotein or glycolipid. Because of the broadness of the term, their method of binding and function can vary enormously and so they are divided into further groups, one of which is the C-type lectin superfamily.

C-type lectins bind to their specific carbohydrate targets in a calcium-dependent manner. The carbohydrate binding function exists within a compact globular structure of the protein, the C-type carbohydrate-recognition domain (CRD). The superfamily consists of 14 groups of which group III is of direct relevance to this thesis (Weis *et al.*, 1998).

The C-type lectins that belong to group III are the collectins, so called because they contain a collagen domain in addition to the Ca^{2+} -dependent CRD (Drickamer *et al.*, 1993). Two such proteins of the collectin family are MBL and Surfactant Protein-A (SP-A) both of which will be used in the thesis and so an introduction to each is essential.

1.2.1 - Mannose-Binding Lectin

MBL was first purified from rabbit liver nearly 35 years ago in the late 1970s (Kawasaki *et al.*, 1978) and subsequently was discovered in humans 5 years later (Kawasaki *et al.*, 1983). It is predominantly a serum protein, and member of the collectin family and shares the same domain organisation across the members. Other

proteins in the collectin family include the two surfactant proteins A and D, collectin liver 1 (CL-L1), collectin placenta 1 (CL-P1), conglutinin, a collectin of 43 kDa (CL-43) and one of 46 kDa (CL-46) (Holmskov, 2000) and more recently collectin 11 (CL-K1) (Rooryck *et al.*, 2011).

1.2.2 - Structural organisation of MBL

As previously discussed, the role of the collectin MBL is as a component of complement, responsible for recognition of pathogens and activation via the lectin pathway. It is assembled through the oligomerisation of the polypeptide initially into homotrimers and then into larger oligomers, consisting principally of two, three and four subunits resembling a bouquet-like structures (Lu *et al.*, 1990) in near planar fan-like conformation. The monomeric polypeptide consists of four domains as shown in figure 1.2.2.1.

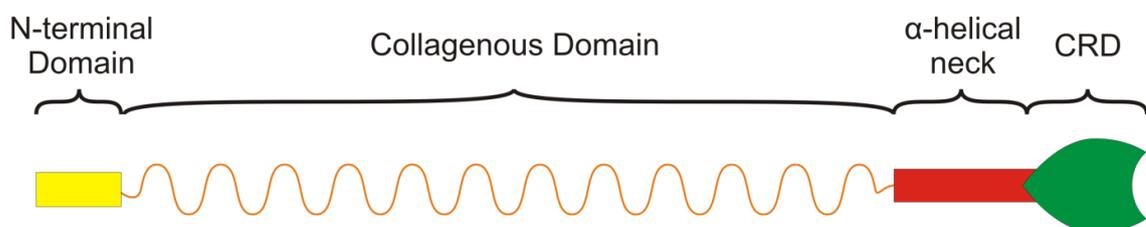


Figure 1.2.2.1. The domain organisation of MBL. The N-terminal domain forms disulphide bonds with other subunits and helps hold the oligomeric structure together. Within the collagenous domain is a kink (not shown) and the MASP binding site. The α -helical neck is where trimerisation of the polypeptide chain occurs and the CRD is the Ca^{2+} -dependent sugar binding globular head.

The N-terminal domain is cysteine-rich making it important for the stability of the oligomeric state by forming disulphide bonds between the subunits (Wallis *et al.*,

1997). The collagenous domain consists of the Gly-Xaa-Yaa repeating sequence characteristic of all vertebrate collagens (Drickamer *et al.*, 1986). There is an interruption of within the collagen domain of Gly-Gln-Gly creating a kink where the stalks splay apart. Following divergence of the stalks after the kink there is a MASP binding motif (Hyp-Gly-Lys-Xaa-Gly-Pro where Hyp is hydroxyproline and Xaa is generally an aliphatic residue); it is at this location that MASP homodimers are able to bind (Wallis *et al.*, 2000). After the collagen domain is the α -helical coiled coil neck domain from which trimerisation of the subunits initiates (Sheriff *et al.*, 1994). Finally, the CRD forms the trimeric sugar binding globular head (Weis *et al.*, 1994, Childs *et al.*, 1990).

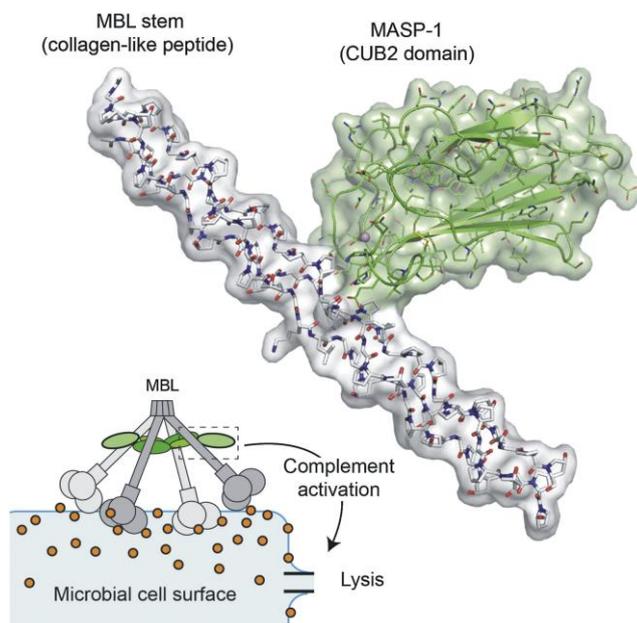
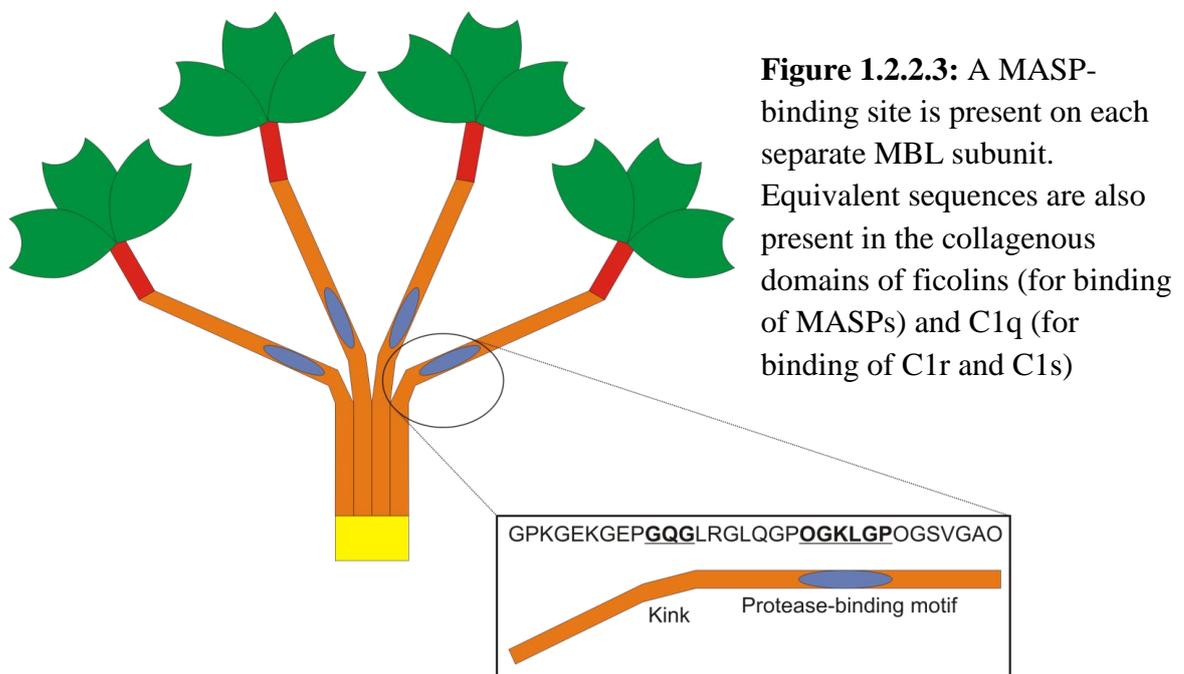


Figure 1.2.2.2: Ca^{2+} -dependent binding between the collagenous domain of MBL and the CUB2 domain of MASP-1. (Gingras *et al.*, 2011).

Oligomerisation occurs within the endoplasmic reticulum prior to secretion and any free cysteine residues are likely linked to small sulfhydryl groups such as cysteine or glutathione (Wallis *et al.*, 1999b). Following secretion, the oligomerisation state of the MBL is fixed and will not alter. MBL also undergoes post-translational modifications

of lysine and proline residues in the endoplasmic reticulum. Both residues are initially hydroxylated and the hydroxylysine residues can then become glycosylated.

Hydroxyprolines are important for the stability of collagen. The role of the lysine derivatisation is less clear although it might minimise non-specific interactions by coating the collagen triple helix with carbohydrate.



1.2.3 - Genetics of MBL

Mammals possess two MBL genes. In contrast, birds and fish only have one MBL gene.

It is highly probable that a gene duplication event occurred after the divergence of mammals and birds during their evolutionary history (White *et al.*, 1994).

In humans, chimpanzees and gorillas only one of the two genes yields a functional protein: the *MBL2* gene produces the 248-amino acid protein, MBL-C. It was initially

mapped to chromosome 10q11.2-q21 in 1989 and this was confirmed two years later. It contains of four exons and Sastry *et al.* concluded that it evolved through a recombination event between an ancestral nonfibrillar collagen gene and a gene encoding a carbohydrate-recognition domain (Sastry *et al.*, 1989). The homology between the human *MBL2* gene and the two rat MBL genes is approximately 50%.

The second MBL gene, the pseudogene *MBL1P1* in humans and primates was first found to be expressed in low levels in the liver and subsequently mapped near to the *MBL2* gene on chromosome 10q22.2-q22.3. It translates into a truncated 51-amino acid protein that shows homology to the MBL-A isoform of rodents (Guo *et al.*, 1998).

In contrast to humans, MBL-A is fully functional in rodents and other mammals. In fact of the two MBLs, MBL-A most closely resembles human MBL with respect to its ability to form large oligomers and to activate complement. By contrast, MBL-C comprises smaller oligomers and has low activity in terms of complement activation (Lipscombe *et al.*, 1995, Lu *et al.*, 1990, Yokota *et al.*, 1995).

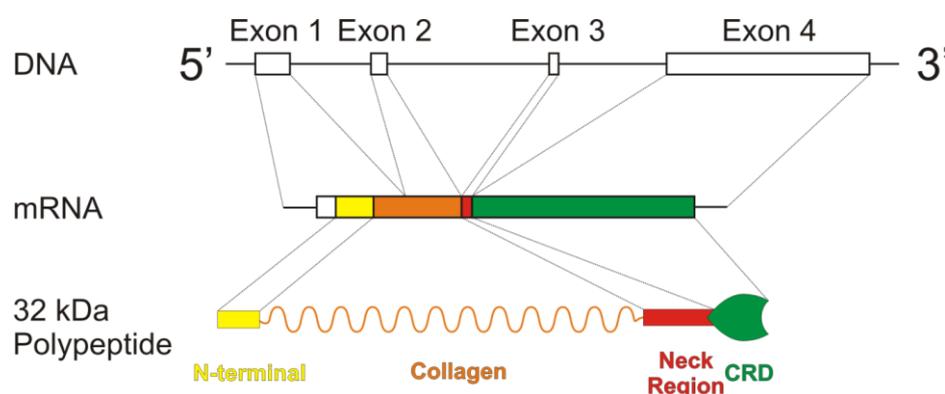


Figure 1.2.3.1: The MBL Gene. The picture shows how the *MBL2* gene is organised and which exons are responsible for encoding the multiple domains of the monomeric polypeptide.

In humans, three *MBL2* alleles have been identified that contain mutations in the coding region of the gene and result in a lack of functional MBL. These were first identified by an association between individuals susceptible to frequent and chronic infections, due to a defect in opsonisation (Miller *et al.*, 1986, Turner *et al.*, 1981), and low levels of MBL (Super *et al.*, 1989). The defect is caused by three single point mutations within exon 1 of the *MBL2* gene. As can be seen in figure 1.2.3.1, this exon codes for the signal peptide, the cysteine rich N-terminal domain and the first seven Gly-Xaa-Yaa repeats of the collagenous domain. The three mutations are all within the collagen domain result in the three amino acid substitutions R52C, G54D and G57E and are referred to as D, B and C respectively (Lipscombe *et al.*, 1995, Madsen *et al.*, 1994, Sumiya *et al.*, 1991). The variants with impaired function are referred to as O, whilst the wild-type MBL is termed A.

In addition to mutations within the coding sequence three promoter polymorphisms have been identified, termed H/L, X/Y and P/Q, which affect the levels of MBL circulating in the plasma (Madsen *et al.*, 1994, Madsen *et al.*, 1995). The H/L variant occurs at the sequence position -550, X/Y at position -221 and P/Q at position +4. Due to their proximity, the structural and promoter polymorphisms are in linkage disequilibrium and the seven haplotypes that frequently occur are HYP A, LY P A, LY Q A, LX P A, LY P B, LY Q C and HYP D. The H allele has a slightly higher level of MBL expression than the L allele and the Q/Q homozygotes have 2-fold greater expression than Q/P heterozygotes 2.5-fold greater than P/P homozygotes. Due to the similar levels of expression the H/L and P/Q alleles are usually ignored in disease association studies. The effect of the X/Y alleles is more substantial. In individuals who

are XA/XA the levels of MBL are comparable to those of individuals with an exon 1 mutation however those who are XA/YO show no detectable levels of MBL in serum.

Of the seven haplotype mentioned previously, these can be further grouped to give the three most common haplotypes – YA, XA and YO. This in turn produces six different genotypes that can be grouped into low, medium and high levels of functional MBL. The low expression group is effectively MBL deficient and comprises of the two genotypes XA/YO and YO/YO. In normal individuals the expression levels of functional MBL in Caucasians are 1.2-5 μ g/ml, this decreases to 0.05-1 μ g/ml in A/O heterozygotes to undetectable (<20ng/ml) in O/O homozygotes (Garred *et al.*, 2003).

1.2.4 - Recognition of sugar ligands by MBL

MBL is able to recognise a wide variety of sugar structures however, the binding between a sugar ligand and the CRD of a monomer is relatively weak with a dissociation constant (K_D) of ~1 mM (Iobst *et al.*, 1994b). More stable binding is achieved through multiple interactions involving more than one CRD, allowing complexes of greater stability to form. The sugar specificity in conjunction with the arrangement of CRDs in the full size MBL molecule permits the ability to discriminate between exogenous and endogenous sugar structures (Taylor *et al.*, 1993).

1.2.5 - Binding to monosaccharides

MBL is a collectin belonging to the C-type lectin superfamily. This means it has Ca^{2+} -dependent sugar binding. The CRD of a C-type lectin is generally about 120 amino acid residues long with two α -helices and two β -sheets forming the hydrophobic core. There

are four surface loops containing either one or two binding sites for the calcium ions (Drickamer, 1988, Weis *et al.*, 1992). The structure of the CRD is maintained by two disulphide bonds between conserved cysteine residues.

In MBL, there are three Ca^{2+} binding sites in each CRD and two of these are essential for carbohydrate binding. The first is present in loop 1 and acts to orientate the loops used in carbohydrate binding. The second is located between loops 3 and 4 and mediates the carbohydrate binding. Mutagenesis of the residues involved in Ca^{2+} binding leads to a total loss of sugar binding. The Ca^{2+} in the second binding site ligates to the conserved residues Asn²⁰⁵, Asp²⁰⁶ and Glu¹⁹³. The Asn and Glu residues form hydrogen bonds with the 3- and 4-OH groups of mannose whilst the adjacent residues (¹⁸⁵Gul-Pro-Asn¹⁸⁷) confer additional specificity for equatorial 3- and 4-OH groups present on mannose type sugars (Quesenberry *et al.*, 1992).

Crystal structures have been obtained of the MBL CRD monomer complexed with an asparaginyl-oligosaccharide (Man6GlcNAc2Asn) which demonstrates limited contact between the protein and sugar. MBL only binds to the terminal mannose moiety (Ng *et al.*, 2002). The limited nature of the contacts between protein and carbohydrate observed in the structure agrees with the observed low affinity of the interaction and explains the broad specificity of MBL which allows the recognition of a wide range of pathogens. Binding to the CRD can be achieved with any sugar that bears equivalent equatorial hydroxyl groups. Two such examples are N-acetyl glucosamine containing the 3- and 4-OH groups and fucose which has equivalent –OH groups at positions 2- and 3-. Both these sugars are commonly found on pathogen surfaces.

1.2.6 - High avidity binding to sugar structures on foreign cells

Although the binding affinity between a single MBL CRD and a monosaccharide ligand is in the mM range (Iobst *et al.*, 1994b), the overall affinity of the interaction increases as the number of binding sites present in the complex increase. This means an MBL tetramer (12 CRDs) can potentially form a complex with a much higher affinity than a trimer, which in turn will have a higher affinity than a dimer and greater still than the monomer with its three binding sites.

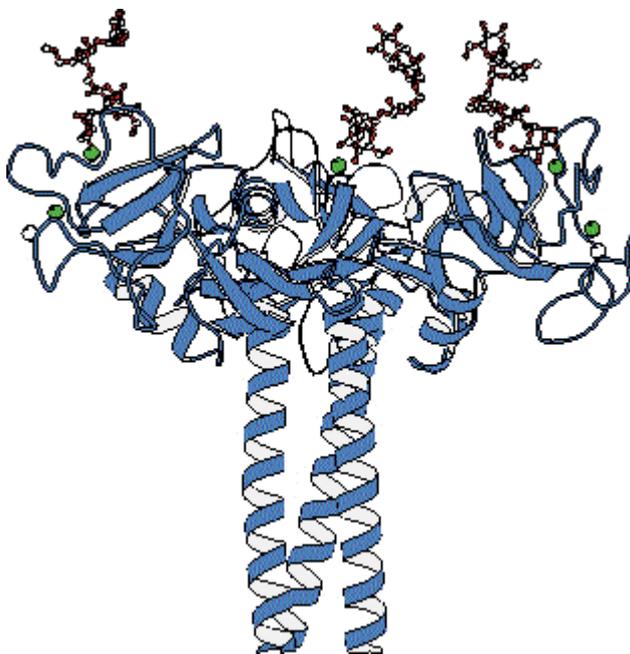


Figure 1.2.6.1: Structure of the neck and CRDs of rat MBL-A (blue) in complex with a carbohydrate ligand. In each CRD, the two Ca^{2+} essential for carbohydrate binding are in green (the third is in white). Adapted from (Weis *et al.*, 1994).

The binding sites in a single subunit of MBL are arranged in a single plane (figure 1.2.3.1). These are maintained in a fixed geometry by hydrophobic contacts between the base of a CRD and the upper region of the neck domain in the adjacent polypeptide (Sheriff *et al.*, 1994, Weis *et al.*, 1994). This flat orientation reveals that the protein is well adapted to the recognition of sugar residues present on the surface of invading pathogens. The spacing of the binding sites is too far apart to bind mammalian high-

mannose type oligosaccharides. Furthermore, most mammalian sugars terminate in sialic acid moieties, which are not recognised by MBL.

1.2.7 - Microbial Targets of MBL

As is consistent with the function of the innate immune response, MBL is able to bind to a wide variety of microorganisms. These include the bacteria *Staphylococcus aureus*, *Klebsiella aerogenes*, *Escherichia coli* (Neth *et al.*, 2000). MBL also binds gp120/gp41 of primary isolates of HIV (Saifuddin *et al.*, 2000) and also to influenza A viruses (White *et al.*, 2000). MBL can bind fungi and yeasts as well *C albicans*, *Cryptococcus* and *Aspergillus* (Levitz *et al.*, 1993, Neth *et al.*, 2000).

1.2.8 – MBL and disease association

Mutations in either the collagenous domain or the promoter region of the *MBL2* gene result in a deficiency of functional MBL. The consequence is that of a compromised innate immune response. Affected individuals are prone to bacterial, viral or parasitic infections in early life prior to the establishment of the adaptive immune response (Sumiya *et al.*, 1991). It also affects HIV infected patients, post-transplant patients on immunosuppressant drugs or following treatment with chemotherapy (Garred *et al.*, 1997, Neth *et al.*, 2001). It was first identified in children between the ages of 6 months to two years who experienced recurrent infections. The serum from the patients failed to opsonise *S. cerevisiae* with complement. Later an association between MBL deficiency and a defect of opsonisation that caused susceptibility to frequent and chronic infection was demonstrated (Super *et al.*, 1989).

MBL deficiency can refer to a number of disorders according to the genotypes. Patients who are XA/XA have low levels of functional MBL and those who are YO/YO have appreciable amounts of MBL that is functionally defective. Thus attempts to define MBL deficiency have revolved around the amount of functional MBL that is present in the serum. In most studies, amounts below 100 ng/ml are regarded as being deficient. Although this simplifies the studies, conflicting results arise. This could be due, in part, to differing genotypes belonging to the same 'deficiency' group. One MBL mutant (B, D or C) might be capable of opsonisation whilst failing to activate complement whilst another might not possess either function. Both of these genotypes come under the umbrella of 'O' mutants, but they might not have equal function. Disease association studies that examine the genotype of the patients in addition to the functional MBL concentrations should be more informative.

Variant MBL proteins give rise to reduced complement-fixing activities and low serum levels. These variants have been studied through creation of recombinant rat MBL-A and human MBL (Garred *et al.*, 2003, Wallis *et al.*, 1999a). In rat MBL the Arg²³Cys mutation forms disulphide bonds with other Cys²³ residues. This prevents MBL trimers and tetramers from forming during expression. This leaves the monomers and dimers which are less effective at fixing complement. The Gly²⁵Asp and Gly²⁸Glu mutations result in disruption of the triple-helical structure of the collagen domain due to the larger size of the side chains that have been incorporated. This disruption affects oligomerisation, MASP binding and MASP-2 activation (Wallis *et al.*, 1999b, Wallis *et al.*, 2005).

MBLs and ficolins are thought to act to help clear apoptotic cells (Kuraya *et al.*, 2005). Their serum levels will impact upon the pathology of autoimmune diseases, such as systemic lupus erythematosus (SLE) (Garred *et al.*, 2001) and rheumatoid arthritis (RA), in which inefficient removal of apoptotic cells is a key factor. With this in mind, variant MBLs that are functionally deficient might predispose patients to SLE. In studies on Spanish and Oriental populations a higher number of SLE patients are MBL deficient than is present in the control groups (Garred *et al.*, 2001, Huang *et al.*, 2003). The MBL defects found include the Gly⁵⁴Asp structural defect and the low plasma level LX promoter haplotype.

There have been conflicting studies on the effect of MBL on RA, some have shown a link between patients with structural variants or low serum levels of MBL with RA (Garred *et al.*, 2000, Graudal *et al.*, 2000) and some with no difference at all (Horiuchi *et al.*, 2000). Other studies have demonstrated that in late onset RA, increased inflammation was noted in patients with wild-type MBL whereas structural variants were observed with early onset RA and acceleration of the disease (Garred *et al.*, 2000). This suggests that MBL fights against the onset of RA however once the disease initiates the severity of the symptoms are increased.

Studies performed *in vitro* have demonstrated that MBL binds HIV derived from infected T cell lines as well as primary isolates (Ezekowitz *et al.*, 1989, Saifuddin *et al.*, 2000). Additionally, at physiological concentrations MBL can partially inhibit infection of CD4+ lymphocytes by T cell adapted HIV. MBL binds to HIV through high-mannose glycans present on the surface of gp120, although this interaction is regulated by the sialylation of the glycoprotein (Hart *et al.*, 2002). HIV can also infect T cells by

trans-infection through receptors present on dendritic cells known as dendritic cell-specific intracellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN) (Geijtenbeek *et al.*, 2000). MBL can inhibit this trans-infection of DC-SIGN expressing T cells (Spear *et al.*, 2003). Despite inhibition of trans-infection of T cells by MBL, it possibly affects virus trafficking since attachment to the virus is sufficient for opsonisation but not neutralisation (Ying *et al.*, 2004). The effect on viral transmission is uncertain since most studies indicate that MBL deficiency increases susceptibility to the virus (Garred *et al.*, 1997, Mombo *et al.*, 2003, Pastinen *et al.*, 1998). The same effect is observed in patients with low serum levels of MBL (Garred *et al.*, 1997).

Ischaemic reperfusion (IR) injury can occur as a result of complement-mediated cytotoxicity when oxygenated blood returns to the tissue following ischemia. This occurs via the lectin pathway of complement activation (Hart *et al.*, 2005, Moller-Kristensen *et al.*, 2005) It has been shown recently that targeting MASP-2 with an anti-MASP-2 antibody offers protection to tissues in myocardial and gastrointestinal IR injury. This result is of therapeutic value through the protection the inhibition of complement confers on the tissues during IR injury. The study also defined a MASP-2-dependent C4-independent pathway of C3 activation (Schwaeble *et al.*, 2011).

1.2.9 – Pulmonary Surfactant Collectins

MBL is produced in the liver and is predominantly found in serum, although it is also present at other sites such as gut and vaginal mucosa. Two other members of the collectin family, called pulmonary surfactant proteins-A and -D are associated with pulmonary surfactant, which lines the alveolar surface of the lungs. Pulmonary

surfactant consists of 90% lipids and 10% proteins and is produced by type II alveolar cells. Half of the protein components are plasma proteins whilst the remaining 5% is made up of the apolipoproteins SP-A, SP-B, SP-C and SP-D. Of these four proteins only SP-A and SP-D are involved with the innate immune response and both are collectins (Kishore *et al.*, 2006).

1.2.10 - Surfactant Protein-A

SP-A is one of the four constituent proteins of pulmonary surfactant. Like MBL, SP-A is a collectin and functions as part of the innate immune response recognising distinct surfactant phospholipids. However, unlike MBL, it does not activate complement but instead functions by binding a wide range of pathogens, acting as an opsonin and by promoting phagocytosis by alveolar macrophages (Crouch *et al.*, 2000, Crouch *et al.*, 2001, Pikaar *et al.*, 1995).

1.2.11 - Structural organisation of SP-A

The SP-A monomer is 28-36kDa and is 248 amino acid residues in length. The domain organisation is the same as that of MBL consisting of an N-terminal and collagenous domain before leading to the α -helical neck and CRD. Like MBL, SP-A oligomerises initially forming a trimer. Six of these subunits further oligomerise into octadecamers, stabilised through disulphide bond formation within the N-terminal domain. This produces a bouquet-like structure similar to that of MBL and C1q (McCormack, 1998).

1.2.12 - Genetics of SP-A

SP-A1 was first identified in 1985 (White *et al.*, 1985) and subsequently mapped to chromosome 10q21-q24 (Bruns *et al.*, 1987). A second isoform of SP-A (SP-A2) was later identified and mapped closely to the SP-A1 gene locus (Katyal *et al.*, 1992, Kolble *et al.*, 1993). The SP-A2 protein maintains over 98% amino acid identity with SP-A1 (Wang *et al.*, 2009). As a result of this homology, SP-A1 and SP-A2 can form both homotrimeric and heterotrimeric subunits (McCormack, 1998). Over 30 alleles have been characterised for SP-A and splicing variants add an extra dimension of complexity of protein (Floros, 2001). This complexity may arise from the variety of pathogens that humans are exposed to, leading to improved protection by the innate immune response on the alveolar surface. SP-A1 is comprised of 6 exons although exons 1, 2 and the 5' end of exon 3 are part of the untranslated region (figure 1.2.11.1).

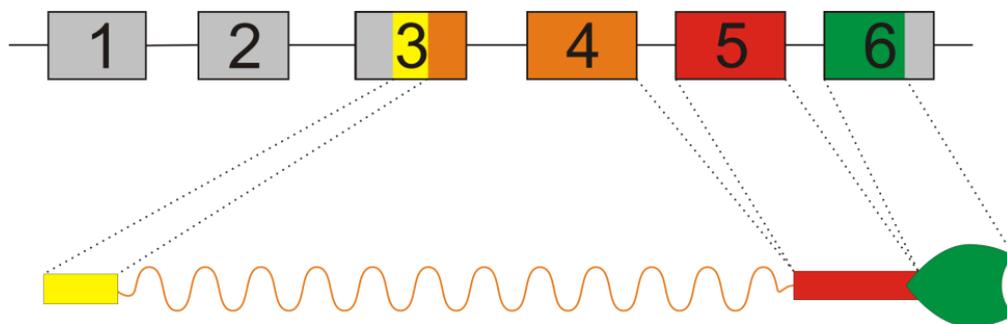


Figure 1.2.12.1: The organisation of the SP-A1 gene. The grey areas are the untranslated regions of the gene. The N-terminal domain is encoded by part of exon 3 (yellow). The remainder of exon 3 and the whole of exon 4 encode the collagenous domain (orange) which does not possess the MASP binding motif present in MBL or ficolins. Exons 5 (red) and 6 (green) encode the α -helical neck and Ca^{2+} -dependent CRD respectively.

1.2.13 – MBL and SP-A receptors

SP-A has no MASP-binding motif so unlike MBL does not activate complement.

Whilst it uses its multiple trimeric CRDs to aggregate pathogens, it requires various

cellular receptors in order to exhibit its other functions. A brief table detailing the receptors of both SP-A and MBL plus the functions they mediate can be found on the following page (table 1.2.12.1). A more detailed description of the functions of SP-A can be found in the reviews by Kishore *et al.* and Nayak *et al.* (Kishore *et al.*, 2006, Nayak *et al.*, 2012).

Receptor	MBL		SP-A	
	Binding	Mediates	Binding	Mediates
cC1qR (calreticulin)	Yes	Phagocytosis of apoptotic cells	Yes	Phagocytosis of microorganisms and apoptotic cells
C1qRp	?	Phagocytosis of microorganisms	?	Phagocytosis of microorganisms
CR1	Yes	Phagocytosis of microorganisms	?	?
CD14	Yes	?	Yes	Modulation of LPS-elicited cytokine release
SIRP α	?	?	Yes	Inhibition of LPS-elicited cytokine release
SP-R210	?	?	Yes	Phagocytosis of microorganisms. Inhibition phospholipid secretion by alveolar type-II cells. Enhancement of nitric oxide production. Enhancement of TNF- α production. Inhibition of T-lymphocyte proliferation.
gp-340	?	?	Yes	?
TLR2	?	?	Yes	Inhibition of peptidoglycan elicited cytokine release.
TLR4	?	?	?	Stimulation of cytokine synthesis

Table 1.2.13.1: A brief overview of MBL and SP-A receptors. For binding, ‘yes’ indicates direct binding studied and ‘?’ indicates no information available. Table has been adapted from a review of the collectins (van de Wetering *et al.*, 2004).

1.3 - MBL-Associated Serine Proteases (MASPs)

MBL-Associated Serine Protease-1 was the first of the MBL associated proteins to be discovered. It was purified from a human MBL preparation as a C1s-like protease. It was initially thought that it was the protein responsible for the cleavage of C4 and C2 in the initiation of complement (Matsushita *et al.*, 1992); however, it was subsequently shown that these activities were instead catalysed by another related protease, MASP-2 (Thiel *et al.*, 1997). Subsequently a third serine protease (MASP-3) has been identified, together with two additional proteins that bind MBL: MAp19 and MAp44 (Dahl *et al.*, 2001, Takahashi *et al.*, 1999, Degn *et al.*, 2009).

1.3.1 - Structure and Function of MASPs

The three MASPs comprise of two Ca²⁺-binding CUB domains (domain found in complement component C1r/C1s, Uegf, and bone morphogenic protein 1), separated by a Ca²⁺-binding epidermal growth factor-like domain (EGF). The second CUB domain is followed by two complement-control protein modules (CCP) and a terminal serine protease domain. It circulates as a zymogen in the form of a head-to-tail homodimer. MAp19 consists of the CUB1 and EGF domains of MASP-2, and MAp44 comprises the CUB1-EGF-CUB2-CCP1 domains of MASP-1 (Degn *et al.*, 2009, Schwaeble *et al.*, 2002).

Binding of MASP to MBL occurs at the distal end of each CUB domain where a Ca²⁺ is bound. A key lysine residue in the collagen (at the centre of the Hyp-Gly-Lys-Xaa-Gly-Pro motif) forms hydrogen bonds with three of the MASP residues, which coordinate the Ca²⁺ (Gingras *et al.*, 2011). Due to the antiparallel nature of the MASP dimer each

CUB domain is able to bind to a separate MBL subunit. This gives the MASP dimer the capacity to bind up to four MBL subunits, although CUB1 interactions are likely to be stronger than those of the CUB2 domain (Feinberg *et al.*, 2003, Gregory *et al.*, 2004).

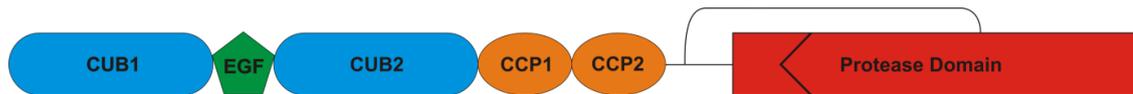


Figure 1.3.1.1: The domain organisation of MASPs. The CUB1 and EGF domains bind to a second MASP polypeptide in a head-to-tail arrangement to form the MASP dimer. Each CUB domain can potentially bind to a separate collagen-like domain, thus can contact four collagen-like domains of an MBL oligomer. The two CCP domains in MASP-2 are involved in substrate recognition (of C4) and the protease domain is responsible for MASP autoactivation as well as cleavage of downstream targets.

MASP circulate as zymogens bound to MBL. When MBL-MASP complexes recognise sugar residues on the surface of invading pathogens, MASP-1 and MASP-2 autoactivate. This is achieved by cleavage of a short linker at the N-terminus of the protease domain. The protease domain remains attached to the rest of the MASP through a sole disulphide bond (Chen *et al.*, 2004). Activation of MASP-3 differs from that of MASPs 1 and 2 in that it does not proceed through autocatalysis but instead most likely occurs through the action of MASP-1 or an as yet unknown protease.

MASP-2 is the key protein of the Lectin Pathway since it is the only one able to cleave both C4 and C2. As with C1s, the CCP modules participate with the SP domain in recognition of C4 (Rossi *et al.*, 2005). MASP-1 can cleave and activate MASP-2 and

C2, but not C4, so might participate in Lectin Pathway activation but cannot directly activate complement. In addition, MASP-1 can activate factor D, so probably facilitates activation of the Alternative Pathway. Likewise MASP-3 can cleave both factor D and factor B so may also participate in Alternative Pathway activation (Iwaki *et al.*, 2011, Matsushita *et al.*, 2000, Takahashi *et al.*, 2010).

The activation of the serine protease domain in MASP-2 occurs via autocatalysis when binding of the MBL/MASP-2 complex to a surface and this induces a conformational change within the structure. This allows the serine protease domain of one of the MASP proteins to access the cleavage site of its partner.

1.3.2 - MASP expression and gene organisation

Through alternative splicing events of two genes a total of five different proteins are produced. The *MASP1/3* gene is responsible for the production of MASP-1, MASP-3 and Map44 whereas the *MASP2* gene codes for the MASP-2 and Map19 proteins.

The *MASP1/3* gene is located on chromosome 3q27-28 and consists of 18 exons (Dahl *et al.*, 2001, Sato *et al.*, 1994). The MASP-1 and MASP-3 proteins share the same CUB, EGF and CCP domains but differ in their serine protease domain. The protease domain of MASP-1 is coded by exons 13-18 whereas MASP-3 is encoded by exon 12 alone. Map44 possesses neither the protease domain nor the CCP2 domain, instead exon 9 is encoded, truncating the protein at the end of the CCP1 domain (figure 1.2.15.1).

Although the same gene codes for both MASPs, MASP-1 is expressed in the liver whereas MASP-3 is expressed in the spleen, lung, small intestine, thymus and the brain.

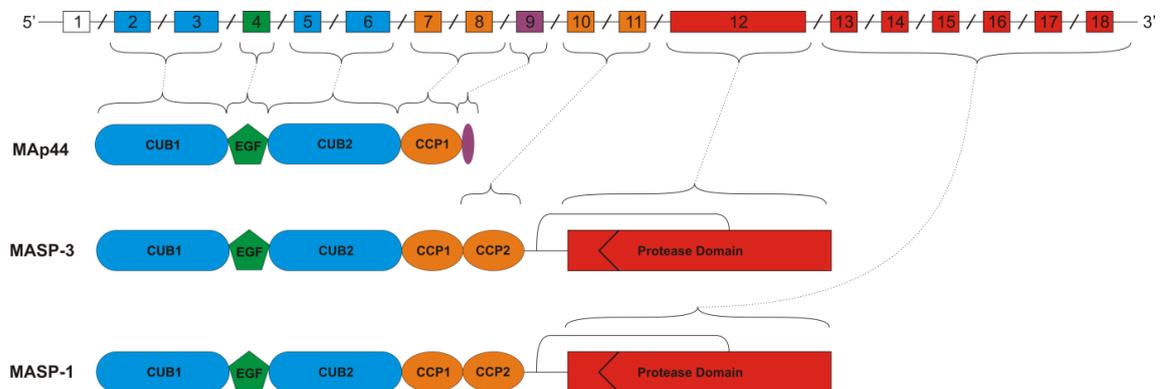


Figure 1.3.2.1: Alternative splicing of the *MASP1/3* gene results in three different products. Map44 incorporates exon 9 which truncates the protein after the CCP1 domain. MASP-1 and MASP-3 share the same CUB, EGF and CCP domains however the serine protease domain is encoded by different exons.

The *MASP2* gene is located on chromosome 1q36.2-3 (Stover *et al.*, 2001) and encodes two proteins, MASP-2 and Map19 which are primarily expressed in the liver. Map19 is produced by alternative splicing including exon 5 which results in the protein being truncated at the end of the EGF domain with the addition of four amino acid residues. (Stover *et al.*, 2004). Whilst it maintains the ability to bind to MBL through the same motif, the binding is around 5 times weaker than that of MASP-2 (Thielens *et al.*, 2001). This is because it lacks the MBL-binding site of CUB2.

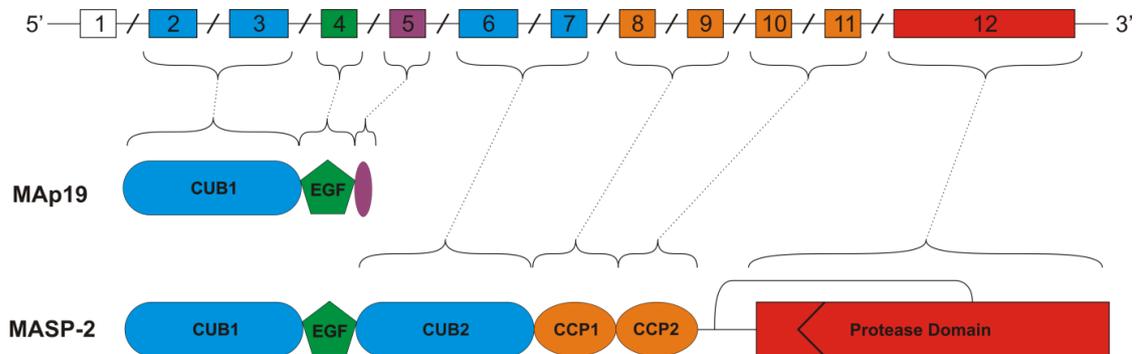


Figure 1.3.2.2: Alternative splicing of the *MASP2* gene produces two products. Inclusion of exon 5 produces the truncated protein MAp19 which consists of the CUB1 and EGF domains and an additional 4 residues. MASP-2 splices out exon 5 and the full-length protease is expressed.

1.3.3 - MASP regulators

As previously discussed, there needs to be control over the proteins of the complement system in order to prevent damage through uncontrolled escalation of the immune response. This is achieved through several inhibitors. For example, C1-inhibitor (C1-INH) which inhibits C1r and C1s, also inhibits MASP-1 and MASP-2. It forms covalent complexes with its target thereby permanently abolishing their protease function (Davis *et al.*, 2008, Sim *et al.*, 2004). MASP-3 is not inhibited by C1-INH (Zundel *et al.*, 2004). Additional inhibitors are anti-thrombin III which, in the presence of heparin, inhibits C4 cleavage by MBL-MASP complexes, aprotinin which inhibits MASP-2 *in vitro* (Petersen *et al.*, 2000) and α -2-macroglobulin which inhibits MASPs but especially MASP-1 (Ambrus *et al.*, 2003).

1.4 – General aims of the thesis

The lectin pathway of complement activation occupies a pivotal role in health and disease. This thesis describes a series of studies aimed at understand the mechanism of initiation of the lectin pathway. The main strategies I have used are:

- To introduce MASP-binding into a SP-A to determine if MASP binding and activation can be transferred in to a protein with a similar architecture but no intrinsic complement activity.
- To study the domains of MBL and determine the features that control MASP activation.
- To characterise a recently discovered member of the collectin family, collectin 11, and to investigate the properties of naturally occurring mutants associated with disease.

Chapter 2 –Introducing MASP-binding and Complement Activation into Human Pulmonary Surfactant Protein-A

2.1 Introduction and Objectives

Initiating complexes of the lectin and classical pathways of complement share a number of common features:

- They comprise two subcomponents: a recognition subcomponent (MBL, ficolins and CL-K1 in the lectin pathway and C1q of the classical pathway) which is able to distinguish between self and non-self structures; and a protease subcomponent (MASPs-1, -2 and -3 in the lectin pathway and C1r and C1s in the classical pathway), which are synthesised as zymogens and become activated when the recognition component binds to a target such as a microbial cell surface.
- The recognition subcomponent contains clusters of C-terminal globular recognition domains (C-type CRDs in MBL, fibrinogen-like domains in ficolins and C1q globular domains in C1q) enabling it to bind to the pathogen surface (Weis *et al.*, 1994). Binding can either be directly to cell surface structures e.g. to pathogen-specific carbohydrates in MBL or via modulator proteins e.g. antibodies or C-reactive protein in the case of C1q (Whaley *et al.*, 1989).
- All recognition subcomponents possess an N-terminal, collagen-like domain which transmits the activating signal from the recognition component to the protease (Sastry *et al.*, 1989). These domains are tethered at the N-termini of polypeptides and give the give recognition subcomponents their characteristic bouquet- or fan-like structures (Wallis *et al.*, 1997). In both pathways,

activation is believed to be driven by changes in the relative positions of the collagen-like domains, following docking to the microbial surface, which subsequently induce changes to the proteases subcomponents triggering activation (Matsushita *et al.*, 1992).

Previous work has identified how the recognition and protease components interact with one another: in each case, a protease bridges separate collagenous stalks of the recognition subcomponent via conserved binding sites on the CUB domains of the proteases (Wallis *et al.*, 2000). Nevertheless, the changes that trigger protease activation are poorly understood.

Somewhat surprisingly perhaps, complement activators are not the only secreted proteins to possess collagen-like domains or have bouquet-like structures. For example, certain forms of acetyl cholinesterase comprise clusters of catalytic domains linked by a central collagen stalk (Duval *et al.*, 1992). Moreover, most members of the collectin family of animal lectins, of which MBL is a member, and which share the same domain organizations do not activate complement. Instead they function as part of the innate immune system by aggregating pathogens, thereby facilitating their removal from the body. One such member of the collectin family is pulmonary surfactant protein-A (SP-A) a C-type lectin and member of the collecting family of lectins with a bouquet structure architecture similar to that of mannose-binding lectin (MBL) (McCormack, 1998). It also has a similar sugar specificity to MBL (although with a preference for L-fucose over D-mannose) (Childs *et al.*, 1992), but is unable to bind MASPs or to activate complement. These properties make it an ideal candidate to use as template to

investigate the features of MBL that are necessary to activate MASP and trigger the complement cascade.

The collagenous domain of MBL and ficolins contain the protease-binding motif common to all recognition-mediated complement activators (MBLs, ficolins, CL-K1 and C1q): Hyp-Gly-Lys-Xaa-Gly-Pro (Wallis *et al.*, 2004). What is not known is whether this sequence represents the only binding site for the MASP, or whether there are additional binding region on the collagenous domain of MBL. The binding-motif is absent in SP-A, and as a result it is unable to bind to MASPs and hence activate complement. By using SP-A as a template and engineering in the MASP-binding motif, I will determine both the extent of the binding site on the collagenous domain of MBL and also whether MASP binding is all that is required by an architecturally similar C-type lectin to activate complement.

Both MBL and SP-A have an interruption (sometime called a kink) in the collagen-like domain. In MBL the sequence Gly-Gln-Gly breaks the consensus Gly-Xaa-Yaa repeat characteristic of all collagens (Uemura *et al.*, 2006). The effect of this kink in the collagen structure is unknown but it might allow flexibility. In SP-A the interruption is more extensive: Pro-Cys-Pro-Pro. Furthermore the Cys forms interchain disulphide bonds that tether subunits together to form the bundle of stems at the base of the bouquet. Previous studies by electron microscopy have shown that changing the Cys to a Ser, opens the bouquet so that the stems diverge nearer to the base (figure 2.1.1) (Uemura *et al.*, 2006). Given the potential structural importance of the kink region, it was of interest to test the effect of its removal on MASP binding and activation.

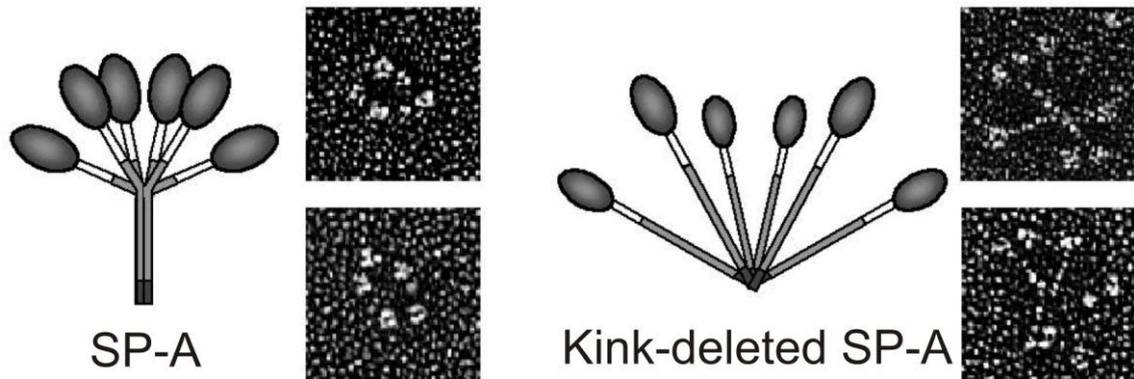


Figure 1.2.1: Electron micrographs by Uemura *et al.*: On the left the globular CRDs of SP-A are held closely together through the disulphide bond present at the kink. Removal of the kink opens the structure out providing improved access to the collagenous stalks. Figure taken from (Uemura *et al.*, 2006)

In this Chapter I will describe experiments where I have introduced the MASP-binding motif into SP-A, with the aim of addressing the following major questions:

1. Will the introduction of the MASP binding motif alone into SP-A be sufficient to permit high-affinity binding of SP-A to MASP?
2. What additional factors are required for SP-A to activate MASP upon binding to a carbohydrate surface?

Question 1 will further our understanding of MBL-MASP binding and question 2 will begin to address the important and as yet poorly understood process, of how MBL activates MASPs

I have also and modified/removed the kink region of the mutant SP-As to examine how large structural changes influence MASP binding and complement activation.

2.2 - Materials and Methods

2.2.1 – PCR Mutagenesis to introduce the MASP-binding motif into SP-A

Three SP-A mutants were created containing all or part of the MASP-binding motif.

Cloning of these mutants was performed by Dr Julia Toth:

Replacing Glu63 and Cys64 with Lys and Leu (in each case, changes to the SP-A cDNA sequence in the forward oligonucleotide primer are underlined)

Mutant	Description of mutation, Primer name and sequence (5'→3')	
SPA-KL	Glu63 and Cys64 to Lys and Leu	
	KLspaF	GT ATC CCT GGA <u>AAG TTA</u> GGA GAG AA
	KLspaRev	TT CTC TCC TAA CTT TCC AGG GAT AC
SPA-KLP	Glu63, Cys64 and Glu66 to Lys, Leu and Pro	
	KLPspaF	GT ATC CCT GGA <u>AAG TTA</u> GGA <u>CCG</u> AAA G
	KLPspaRev	C TTT CGG TCC TAA CTT TCC AGG GAT AC
SPA-KLPO	Glu63, Cys64, Glu66, Lys67 to Lys, Leu, Pro and Hyp	
	KLPOspaF	GT ATC CCT GGA <u>AAG TTA</u> GGA <u>CCG CCG</u> GGG GAG
	KLPOspaRev	CTC CCC CGG CGG TCC TAA CTT TCC AGG GAT AC

Table 2.2.1.1: Mutagenic primers used to produce the SPA-KL, SPA-KLP and SPA-KLPO mutants. Changes to the SP-A cDNA have been underlined in the forward primer sequence.

2.2.2 – PCR Mutagenesis to modify the SP-A kink

The collagen-like domain of SP-A contains a interruption or kink, which tethers the collagen stalks together. Two strategies were used to remove this kink:

Cys⁴⁷ was replaced by a serine residue to prevent the formation of interchain disulphide bonds that link the collagen stalks together

The entire kink region was removed (four residues between Pro⁴⁶ to Pro⁴⁹)

Mutation	Primer	Sequence (5' → 3')
Cys ⁴⁷ Ser	CSspaF	AG AAA <u>TGC</u> CAA GTC CTC CTG GA
	CSspaR	TC CAG GAG GAC TTG <u>GCA</u> TTT CT
'PCPP' Deletion	DeltaSpaF	TG GGT CCA CCT GGA GAA ATG <u>***</u> GGA AAT GAT GGG CTG CCT GG
	DeltaSpaR	CC AGG CAG CCC ATC ATT TCC <u>***</u> CAT TTC TCC AGG TGG A CCC A

Table 2.2.1.2: The mutagenic primers which remove the kink within the collagenous domain of SP-A. Changes to the SP-A cDNA have been underlined in the forward primer sequence.

The terminal primers used in all PCR reactions were:

Terminal Primer	Restriction Site	Sequence (5' → 3')
SP-Af	<i>XhoI</i>	ATACATGTCGACGCCACCATGTGGCTGTGCCCTCTG
SP-Arev	<i>EcoRI</i>	ATCGAATTCCTAATGCCTCTCAGAACTC

Table 2.2.1.3: Terminal Primers of SP-A. Highlighted in the primer sequences are the restriction sites (*pink*), Kozak sequence (*yellow*),

The proofreading polymerase *Pfu* from Promega was used in all PCR reactions. The final concentration of the components was: 1 μM forward primer, 1 μM reverse primer, 1.25 U/50 μl *Pfu* polymerase, 0.5 μg DNA template and 200 μM dNTP mix. The PCR heat cycle was 94 °C for 2 minutes followed by 25 cycles of 94 °C for 15 seconds, 45 °C for 30 seconds and 72 °C for 30 seconds. The 25 cycles were then followed by a final elongation step of 72 °C for 5 minutes to ensure that synthesis of the DNA strands was complete.

PCR fragments were run on a 1 % agarose gel made with TBE buffer containing ethidium bromide at 0.5 µg/ml. The fragments were visualised on a UV transilluminator and the DNA fragments excised with a sharp scalpel. The PCR products were purified from the gel using the QIAEX II gel extraction kit from QIAGEN.

Double digestions of the whole PCR fragments and expression vector were undertaken according to NEB protocols and the SP-A cDNA was ligated using NEB T4 DNA ligase enzyme. Once the gene had been cloned into the pED4 expression vector, the resulting clones were sequenced verified at the PNAACL facility and the University of Leicester.

Where cloning directly into pED4 did not work, the PCR product was cloned via pGEM-T a TA cloning vector from Promega. The PCR product was run on a 1 % agarose gel and gel purified as before. 7 µl of the product was then incubated at 70 °C for 30 minutes in the presence of 5 U *Taq* DNA polymerase, 0.2 mM dNTP and 1 µl 10x *Taq* polymerase buffer. This A-tailing to the 3' ends of the product allowed the DNA to be ligated into the pGEM-T linearised vector. The ligation reaction mixture was 1 µl of pGEM-T vector (50 ng), 5 µl 2x Rapid ligation buffer, 1 µl T4 DNA polymerase (3 weiss units) and 3 µl of the A-tailed PCR product were incubated at room temperature for 1 hour and the transformed into competent JM109 cells. The cells were plated onto IPTG and X-gal plates and white coloured colonies selected.

Expression vector pED4

The two key components of pED4 are the polylinker region into which the PCR product is cloned and the dihydrofolate reductase gene (DHFR) which is produced on a

bicistronic mRNA molecule along with the cloned DNA. Dihydrofolate reductase converts dihydrofolate into tetrahydrofolate, a methyl group shuttle required for the de novo synthesis of purines, thymidylic acid, and certain amino acids (Kaufman *et al.*). DXB11 is a cell line deficient in DHFR and as a result requires the growth media of naive cells to be supplemented with nucleotides. Following transfection into the DXB11 cells, only cells that have successfully obtained the plasmid will be able to grow in the absence of external nucleotides.

Through increasing concentrations of methotrexate (a competitive inhibitor of DHFR)

an increasing expression of the bicistronic DHFR mRNA occurs. The consequence of this is increased stable expression of the cloned cDNA.

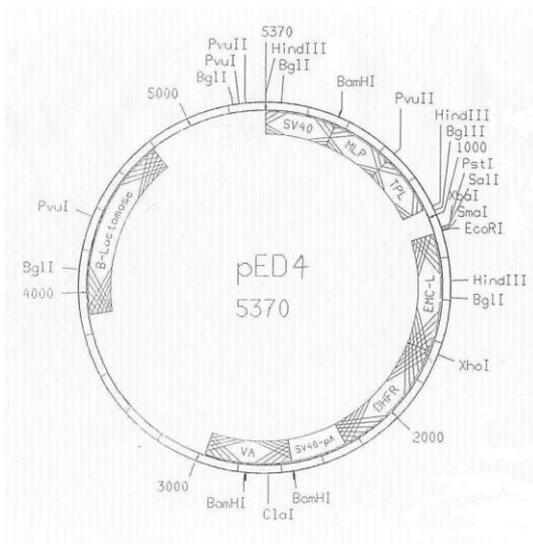


Figure 2.3.1.1: pED4 expression vector. The gene is cloned into the *SalI* and *EcoRI* sites within the polylinker. Modified from (Kaufman *et al.*).

2.2.2 - Bacterial Transformation

Competent JM109 cells were purchased from Promega. 4 μ l of ligation reaction was mixed with 50 μ l cells and incubated on ice for 20 minutes. The cells were then heated at 42 $^{\circ}$ C for 50 seconds and rested on ice for a further 5 minutes. 950 μ l of Luria broth (LB) was added and the cells grown at 37 $^{\circ}$ C for 1.5 hours. The cells were then spun at 1000 rpm for 10 minutes and plated on LB agar plates containing ampicillin at 50 μ g/ml and grown at 37 $^{\circ}$ C overnight.

2.2.3 – Cell culture

All proteins were expressed in a Chinese hamster ovary (CHO) cell line called DXB11 first described in 1980 (Urlaub *et al.*, 1980), which is deficient in the dihydrofolate reductase (DHFR) gene. Prior to transfection with DNA cells were grown in Minimal Essential Media α with nucleosides (MEM α^+) supplemented with 10 % dialysed, heat-treated foetal calf serum (DHFCs) and 50 units/ml penicillin and 50 μ g/ml streptomycin. Cells were grown in 25 cm² Nunc tissue culture flasks with filter caps and incubated at 37 °C and 5 % CO₂.

Once cells had grown to confluence the media was 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. The cells were incubated at room temperature for 5-10 minutes with 1 ml trypsin-EDTA solution to detach them from the flask. One drop of the resulting cell suspension was used to seed a fresh flask.

All cell culture reagents were purchased sterile from GIBCO®.

2.2.4 – Mammalian Cell Transfection

To produce sterile plasmid DNA for transfection, 5 μ g of DNA was diluted into 20 μ l sterile ddH₂O in a sterile Eppendorf tube. To this 2 μ l 3M NaOAc and 50 μ l ice cold ethanol were added and incubated on dry ice for 10 minutes. The Eppendorf tube was spun at 13000 rpm on a bench top centrifuge for 5 minutes. The pellet was washed with 500 μ l 70 % ice cold ethanol and spun for a further 2 minutes at 13000 rpm. In a tissue

culture hood the 70 % ethanol was removed and the pellet air dried. Finally the pellet was dissolved in 100 μ l sterile filtered ddH₂O.

The transfection was achieved through calcium phosphate precipitation first described in 1973 (Graham *et al.*, 1973). The contents of a 1 ml Eppendorf tube containing 100 μ l of 50 μ g/ml of plasmid DNA, 100 μ l of calf-thymus DNA 10 mg/ml, 800 μ l sterile filtered water and 120 μ l 2M CaCl₂ was added into a 15 ml falcon tube containing 1 ml 2x HEPES Buffered Saline (HBS) (1 g HEPES, 1.6 g NaCl / 100 μ l) and 40 μ l 100x Phosphate buffer (45 mM Na₂HPO₄, 45 mM NaH₂PO₄) and incubated at room temperature for 30 minutes. The solution was mixed constantly during addition.

After this time 1 ml of the mixture and precipitate was added evenly over a lawn of DXB11 cells that had been grown to 50 % confluence in a 25 cm² flask. The cells were returned to their incubator. The following day the media was replaced with fresh MEM α ⁺ to allow the cells to recover before selection

Once the cells had successfully taken up the expression vector DNA, they contain copies of the DHFR gene, so no longer require the addition of nucleotides to the culture medium. So, after another night in the incubator the cells were washed, trypsinised and plated into MEM α ⁻ media in a sterile culture dish. The cells were grown for 2-3 weeks to allow colonies to form on the plate which were individually picked and transferred to separate 25 cm² flasks containing MEM α ⁻.

2.2.5 – Protein Expression

In order to increase protein expression levels, increasing concentrations of methotrexate (MTX) were added to the MEM α media. MTX inhibits DHFR causing the cells to upregulate DHFR gene expression from the expression vector. Because the transcript is encoded on a single mRNA, increased expression of DHFR also results in increased expression of the protein of interest.

MTX was added initially at a concentration of 0.02 μ M and the cells grown over 2-3 passages until they were growing at a normal rate (typically needing one week to reach confluence). This was repeated at 0.1 μ M MTX and finally at 0.5 μ M MTX at which point the cells were tested for production.

For production, cells were passaged into 150 ml MEM α 0.5 μ M MTX in a triple layer flask and grown to confluence. The cells were washed twice with 50 ml sterile PBS and the media replaced with 100 ml (CHO-S-SFMII) containing pen-strep and 0.5 μ M MTX and 50 mM HEPES, pH 7.55 to help maintain the pH during cell growth. The SFM was replaced daily and the harvested media was centrifuged at 2000 rpm for 2 minutes to remove any cell debris. The media was then decanted and frozen at -20°C until it was required for purification.

2.2.6 – Sugar-Sepharose coupling

50 ml of Sepharose 6B from GE Healthcare was washed with 2 L ddH₂O and added to 50 ml 0.5 M NaHCO₃ pH 11. To activate the Sepharose beads, 5 ml divinyl sulfone was added to the suspension and the mixture was incubated for 70 minutes. The Sepharose

was then washed again with 2 L ddH₂O and added to 50 ml 0.5 M NaHCO₃ pH 10 containing 20 % sugar w/v and the mixture incubated overnight to couple the sugar to the Sepharose. The Sepharose was washed again with 2 L ddH₂O and finally incubated in 50 ml 0.5 M NaHCO₃ pH 8.5, 2ml β-mercaptoethanol to block further coupling to the Sepharose beads. Finally, the mixture was washed with 4 L ddH₂O and stored at 4°C. In this protocol all incubations were done at room temperature with gentle stirring in a fume hood.

2.2.7 – MBL and SP-A Purification

To purify these two proteins from the harvested SFM, the media was passed over a sugar-Sepharose column in the presence of Ca²⁺ ions.

To purify rMBL; 200 ml of harvested media was mixed in a 1:1 ratio with high-salt loading buffer (HSLB) 50 mM Tris-HCl pH7.5, 1.2 M NaCl and 10 mM CaCl₂. A 1 ml column of mannose-Sepharose was equilibrated with 10 ml HSLB and the media/buffer mix passed over the column overnight at 4 °C. The column was washed again with 10 ml HSLB and then 10 ml low salt loading buffer (LSLB) 50 mM Tris-HCl pH7.5, 150 mM NaCl and 10 mM CaCl₂. To release the protein from the column, elution buffer (EB) 50 mM Tris-HCl pH7.5, 150 mM NaCl and 2.5 mM EDTA was used. The protein was collected in 0.5 ml fractions.

The purification strategy for SP-A was modified compared to MBL, because SP-A precipitates in buffers containing high salt. Therefore, media (300 ml) was mixed in a 1:1 ratio with 50 mM Tris-HCl pH7.5 and 5 mM CaCl₂. A 1 ml mannose-Sepharose

column was equilibrated with 10 ml of the same buffer. The media was passed over the column overnight at 4 °C. Following washes with high salt buffer (10 mM Tris-HCl pH 7.4, containing 500 mM NaCl and 5 mM CaCl₂), triton buffer (10 mM Tris-HCl pH 7.5, containing 0.1% Triton X-100 and 5 mM CaCl₂) and low salt buffers (10 mM Tris-HCl pH 7.4, containing 5 mM CaCl₂), protein was eluted in 0.5 ml fractions of 10 mM Tris-HCl pH 7.4, containing 5 mM EDTA.

Fractions containing recombinant protein were identified by SDS-polyacrylamide gel electrophoresis and rMBL was dialysed into 50mM Tris-HCl pH7.5, 150mM NaCl and 5mM CaCl₂ and SP-A was dialysed into 50mM Tris-HCl pH7.5 and 2mM CaCl₂ using 10,000 MWCO dialysis tubing.

2.2.8 – His-Tagged MASP-2 Purification

Two different full-length MASP-2s were used called MASP-2A and MASP-2K (Chen *et al.*, 2001, Chen *et al.*, 2004), which are described in more detail in the Results Section. For production, existing cell lines were grown to confluence in triple flasks in MEM α - medium containing DHFCS, 0.5 μ M MTX and Pen/Strep as described for MBL in the previous section. The MASP-2 proteins were produced and harvested in serum-free media as described above expressed and harvested as previously described in '2.2.5 – Protein Expression'. Both MASPs contain an N-terminal hexahistidine tag enabling purification by affinity chromatography.

MASPs were purified on Ni-Sepharose FAST FLOW (GE Healthcare). Briefly media was diluted 1:1 in 25 mM Tris pH 7.5 containing 1 M NaCl. The column was washed

with 10 column volumes of loading buffer, followed by a further 10 column volumes of buffer containing 20 mM imidazole, to remove weakly bound contaminants and eluted using buffer containing 200 mM imidazole. Samples were monitored by SDS-PAGE and fractions were pooled and stored at 4° C prior to use.

2.2.9 – Surface Plasmon Resonance

Measurements were performed using a BIAcore 2000 instrument (GE Healthcare) or a Bio-Rad ProteOn XPR36 biosensor. Protein ligands were diluted into 10 mM sodium acetate (pH 4.5 for MBL or pH 5.0 for SP-A) and immobilized onto the carboxymethylated dextran surface of a CM5 sensor chip (GE Healthcare) or a GLM chip (Bio-Rad), using amine coupling chemistry. Binding was measured in 10 mM Tris-HCl (pH 7.4), containing 140 mM NaCl, 2 mM CaCl₂, and 0.005% surfactant SP40, at a flow rate of 25 µl/min and at 25 °C. After injection of ligand, the protein surface was regenerated by injection of 10 µl of 10 mM Tris-HCl buffer (pH 7.4), containing 1 M NaCl and 5 mM EDTA. Data were analyzed by fitting association and dissociation curves to Langmuir binding models for several protein concentrations simultaneously, using BIAevaluation 4.1 software (GE Healthcare). Increasingly complex models were tested until a satisfactory fit to the data was achieved. Apparent equilibrium dissociation constants (K_D) were calculated from the ratio of the dissociation and association rate constants (k_{off}/k_{on}). MBL was immobilized on the chip surface rather than used as a soluble ligand because it bound to the chip, thereby masking analysis of the protein-protein interactions. SP-A was also immobilized because it tended to precipitate on the chip surface.

2.2.10 – MASP-2k Activation Assays

Activation was measured by following MASP autolysis using a modified version of the protocol described previously (Chen *et al.*, 2004). Briefly, MASP-2K was mixed with wild-type or mutant MBL or SP-A and added to a suspension of fucose-Sepharose (5 μ l of a 1:1 v/v suspension in a total volume of 30 μ l), in 50 mM Tris-HCl (pH 7.5) containing 150 mM NaCl and 5 mM CaCl₂, at 37°C. A 1.2-fold molar excess of MBL or SP-A was used to ensure that all the MASP-2K was bound to the recognition molecule. The mixture was incubated at 37 °C with constant mixing and aliquots of the suspension were removed from the reaction mix at various times and immediately frozen on dry ice to quench the reaction. Proteins were separated by SDS-polyacrylamide gel electrophoresis under reducing conditions and the amount of MASP cleaved was quantified by densitometry.

2.3 Results

2.3.1 - Engineering MASP binding into SP-A

The first step of the project was try to engineer MASP binding into SP-A. Although both collectins, the collagen-like domains of MBL and SP-A share no apparent sequence identity apart from the Gly-Xaa-Yaa repeat, common to all collagens. Notably, the domain is significantly shorter in MBL with 17 Gly-Xaa-Yaa repeats compared to 23 in SP-A, so the first important consideration was where to position the MASP-binding site. Both domains contain an interruption to the consensus repeat, so this was used to align the domains. In MBL the MASP-binding site spans repeats 3, 4 and 5 beyond the interruption. Repeats 4, 5 and 6 were selected in SP-A to ensure sufficient room for the MASP to bind. Another advantage of this strategy was that the substitutions would remove a cysteine residue in SP-A (repeat 5), which might interfere with MASP binding through formation of disulfide bonds. Three different mutations were created by modifying the SP-A cDNA (figure 2.3.1.1). In SPA-KL, Glu63 and Cys64 were replaced by lysine and leucine residues, respectively. Lysine in the Xaa position is known to be essential for MASP binding and complement activation in both MBLs and ficolins. In SPA-KLP, a proline residue was introduced in place of the glutamate at position 66 to optimize the MASP-binding motif further: Hyp-Gly-Lys-Leu-Gly-Pro. One further change was made in SPA-KLPO, replacing Lys67 by a hydroxyproline residue. Lysine residues in the Yaa position of collagen are often hydroxylated and glycosylated during biosynthesis, so this change was designed to prevent any potential steric inhibition of MASP binding by the sugar residues.



Fig 2.3.1.1. Sequence alignment of the collagen-like domains of MBL and SP-A and the design of mutant SP-As. Sequences were aligned based on the positions of the kink of SP-A and the kink-like region of MBL (*red*). The MASP-binding motif is shaded *blue*. Changes to the sequence of SP-A are indicated.

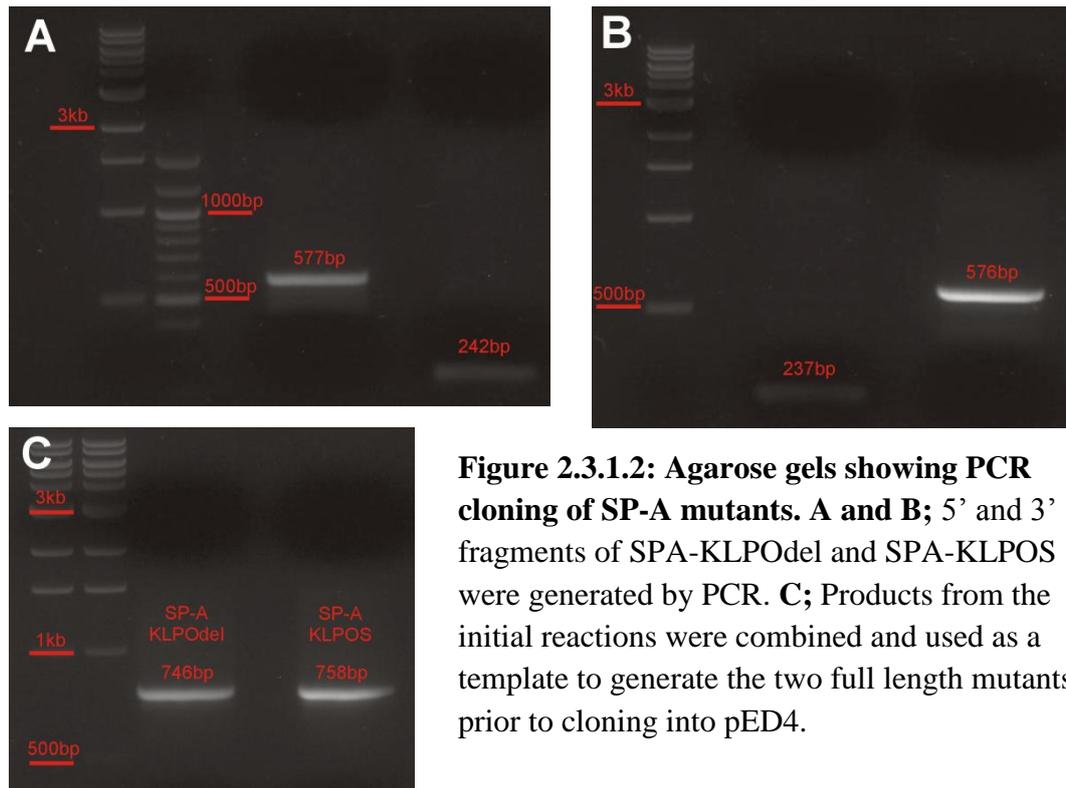
As well as introducing the MASP-binding site into SP-A, additional mutations were introduced to change the point at which the collagenous stalks splay apart by removing the kink. In one mutant, SPA-KLPOdel, the kink sequence (PCPP) was removed completely, and in the other, called SPA-KLPOS, the cysteine residue was replaced by a serine residue to remove the potential for disulfide bond formation, which tethers the separate stalks together in the native protein.

In each case mutations were introduced by PCR as before and the resulting cDNAs were cloned into the expression vector pED4.

2.3.1 – Cloning

All mutant clones were generated by PCR using a two step process in which the cDNA was initially amplified in two halves to introduce the mutation and these were then combined to form the full-length cDNA. For example, to clone the kink deletion mutations, a forward primer (spaf) was combined with the reverse mutagenic primer (either Deltasparev or CSsparev) to create a PCR fragment of approximately 240 bp and the reverse gene primer (sparev) was then combined with the forward mutagenic primer (either Deltaspaf or CSspaf) to produce a PCR fragment of approximately 577 bp. To

produce the full length construct, PCR products were purified and then mixed to become the template for the second PCR using the forward and reverse terminal primers. The resulting products were introduced into the polylinker of pED4 via *EcoRI* and *SalI* sites.



The cDNA fragments encoding SPA-KLPOdel and SPA-KLPOS generated by PCR were digested using *EcoRI* (G/AATTC) and *XhoI* (C/TCGAG) and cloned into the *EcoRI* and *SalI* (G/TCGAC) sites of pED4. *XhoI* and *SalI* have compatible cohesive ends, both enzymes producing a four base 5' overhang which can be subsequently ligated together. Both constructs were sequence verified by PNAACL at the University of Leicester.

2.3.2 – Expression and Purification

SP-A mutants were produced in a well characterised Chinese hamster ovary cells previously used for production of MBL , to ensure that collagenous domains were correctly modified. This is important because hydroxylation of proline residues within the collagenous domains is required for stability of the collagen and its assembly to form larger oligomers. In total, five recombinant SP-A constructs were produced. All were purified by affinity chromatography on mannose- or fucose-Sepharose columns. Typical yields were 0.2 – 0.5 mg/L of culture medium. Binding to sugar-Sepharose columns confirmed that all five proteins had folded correctly during biosynthesis with functional CRDs. These can be seen in figure 2.3.2.1.

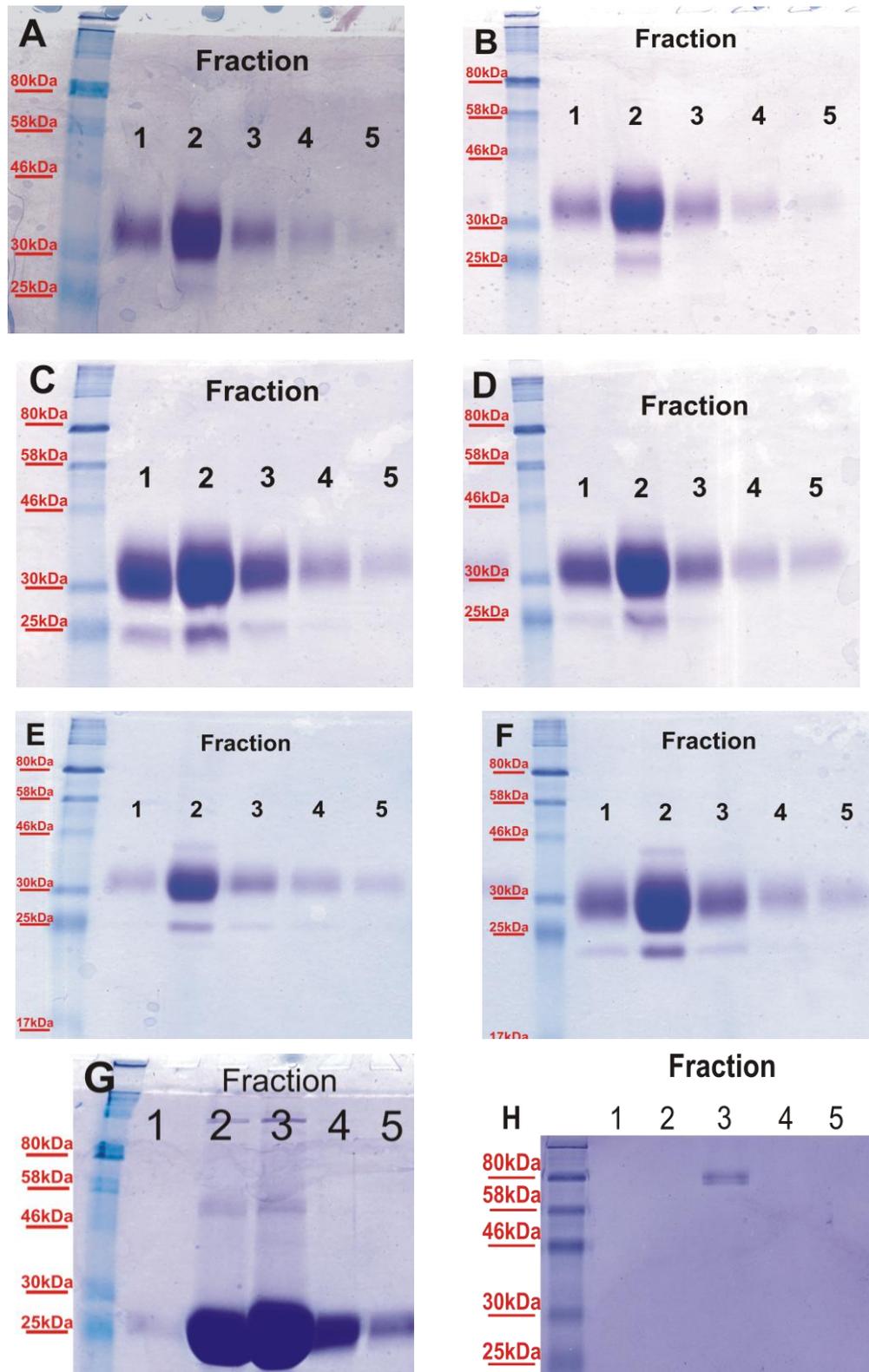


Figure 2.3.2.1: SDS-polyacrylamide gel electrophoresis showing purification of wild-type and mutant SP-As. Protein was purified by affinity chromatography on fucose-Sepharose and was eluted from the columns using EDTA, which chelates the Ca^{2+} required for binding. The gels (15%) were run under reducing conditions and show samples taken from each of 5, 0.5 ml elution fractions: **A: SPA wild-type. **B**: SPA-KL. **C**: SPA-KLP. **D**: SPA-KLPO. **E**: SPA-KLPOdel. **F**: SPA-KLPOS. **G**: rMBL **H**: MASP-2K. Proteins were stained with Coomassie blue.**

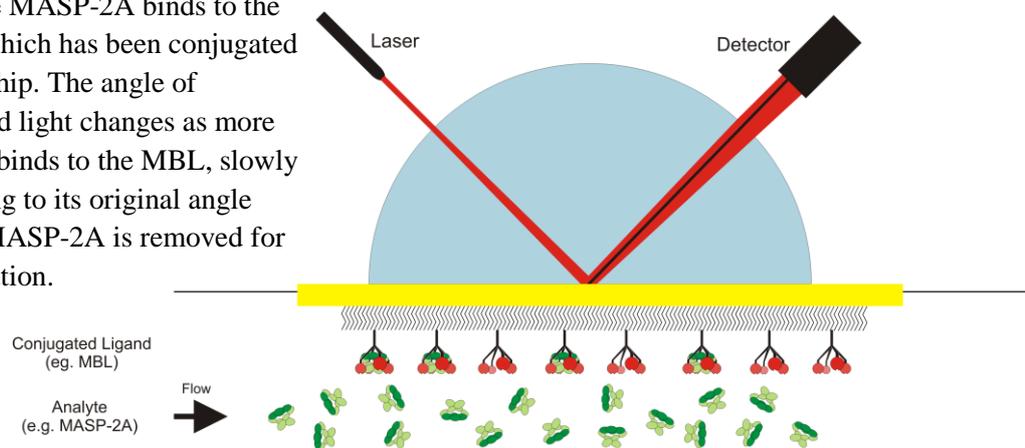
2.3.3 – Binding to MASP-2

The first property of the SP-A mutants tested was their ability to bind to MASP-2. Full length MASP-2 is difficult to purify in significant amounts, because it autoactivates during biosynthesis. This not only reduces the yields of recombinant protein greatly but also results in mixtures of zymogen (inactive) and activated forms. I therefore used a catalytically inactive form of MASP-2 to measure binding, called MASP-2A, in which the active site serine, Ser⁶¹³, is substituted by an alanine. This mutation prevents autoactivation, ensuring that all MASP is in the zymogen form (Wallis *et al.*, 2000).

To measure binding, SP-As were immobilized on to a CM5 chip (BIAcore) by amine coupling, and increasing concentrations of MASP-2 were flowed over the chip surface. Binding is detected by a change in the angle of reflected light, as more MASP becomes bound to the chip surface (through its association with MBL or SP-A) leading to an increase in signal expressed as Response Units. Dissociation is measured in the same experiment by stopping the flow of MASP-2 and just flowing buffer over the chip. Dissociation of bound MASP-2A leads to a reduction in the mass on the chip and a corresponding decrease in Response Units as the angle of reflected light returns to its original value.

Figure 2.3.3.1. Representation of Surface Plasmon Resonance.

In this example, the catalytically inactive MASP-2A binds to the MBL which has been conjugated to the chip. The angle of reflected light changes as more MASP binds to the MBL, slowly returning to its original angle when MASP-2A is removed for the solution.



An example of a typical experiment is shown in Figure 2.3.3.1, which shows binding between MBL and MASP-2A. As the MASP protein is passed over the chip the RU increase confirming that it binds to MBL. Removal of the MASP leads to a slow dissociation. Binding is reversible and concentration dependent. To determine the kinetic parameters data were fitted simultaneously to increasing complex binding models until a satisfactory fit was achieved. The data fitted poorly to a 1:1 model, but a satisfactory fit was achieved using a two-complex, parallel-reaction model in which there were two different binding sites for MASP (a high affinity site and a secondary weaker binding site). Association rate constants (k_{on}) were $40 \pm 0.1 \times 10^5$ and $2.9 \pm 0.1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ and dissociation rate constants (k_{off}) were $1.5 \pm 0.1 \times 10^{-3}$ and $4.75 \pm 0.1 \times 10^{-3} \text{ s}^{-1}$ giving K_D s of 3.7 ± 0.1 and $165 \pm 2 \text{ nM}$, respectively. These data are consistent with previous findings for MBL-MASP-2 interactions which also show complex binding.

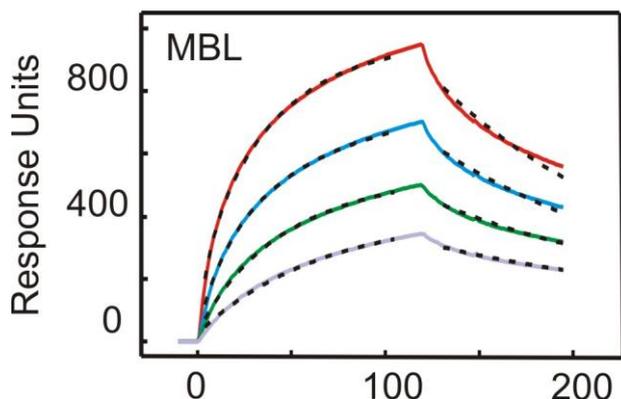


Fig 2.3.3.2. Binding of MASP-2A to MBL by surface plasmon resonance. MASP-2A was injected at 333 (*red*), 167 (*blue*), 83 (*green*), and 42 nM (*purple*) over immobilized MBL (~7000 response units). All data were fitted to a two-complex parallel binding model, and the fits are shown by the dotted lines.

As expected, wild-type SP-A did not bind to MASP-2A at any of the concentrations tested (up to 1 μ M). Surprisingly, however, SPA-KL containing just two substitutions bound with appreciable affinity. As with MBL, the data best fitted a two-complex, parallel-reaction model, with apparent K_{DS} 600 ± 110 and 2120 ± 100 nM, respectively. SPA-KLP bound with even higher affinity, with K_D values of 36 and 104 nM, only slightly weaker than MBL (3.7 and 165 nM). Additional changes had little effect on MASP binding. Replacing the lysine with a hydroxyproline did not improve binding significantly, so the lysine residue in SP-A (or its glycosylated derivative) does not impede MASP access. Similarly, removal of the kink region in SP-A also had little effect, indicating that the point at which the stalks splay apart and thus the angle between adjacent stalks is not limiting for MASP binding. Thus, just three amino acid changes to the collagenous domain of SP-A are sufficient to introduce MASP binding. Moreover the affinity of binding is comparable to MASP binding by MBL itself.

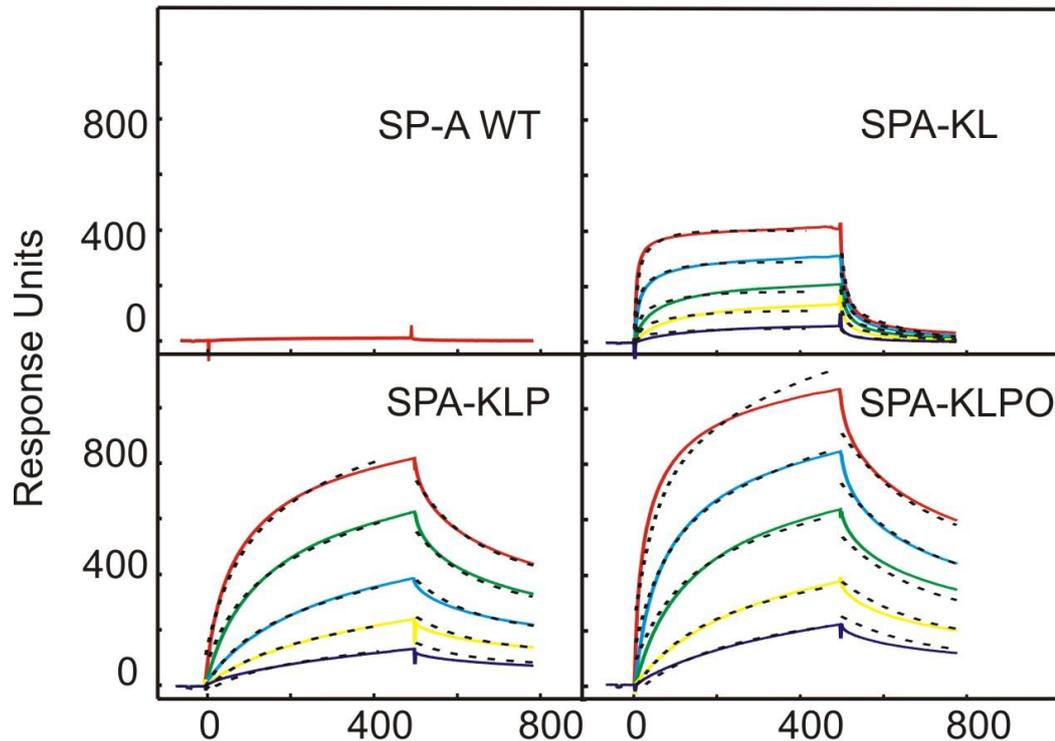


Fig 2.3.3.3. Binding of MASP-2A to SP-A mutants by surface plasmon resonance. MASP-2A was injected at 764 (red), 447 (green), 261 (blue), 152 (yellow), and 89 nM (dark blue) over each of the immobilized SP-As (~6000 response units). All data were fitted to a two-complex parallel binding model, and the fits are shown by the dotted lines.

2.3.4 – MASP-2k Activation Assays

MAASP-2A could not be used to measure MASP-2 activation, because it is catalytically inactive (it lacks the active site serine, Ser⁶¹³) so cannot autoactivate. Instead, I used a second modified MASP, called MASP-2K, which is a catalytically active form of MASP-2 (containing all of the normal catalytic machinery: catalytic triad; oxyanion hole and specificity pocket) but containing a point mutation in the linker region where MASP-2 autoactivation occurs. Replacement of Arg₄₂₄ by a lysine residue still permits autoactivation by MBL when it binds to a carbohydrate surface, but reduces the intrinsic activation rate, so that protein is synthesized, secreted and can be purified as a zymogen (Chen *et al.*, 2004).

I then tested whether the SP-A mutants could activate MASP-2K using fucose-Sepharose as the carbohydrate target (substrate). MBL was used as a positive control in these experiments. As shown in figure 2.3.4.1G, when MBL-MASP-2K complexes are mixed with fucose-Sepharose, initially MASP-2K is all in the zymogen form, which runs as broad band on SDS-PAGE gels with a molecular mass of ~80 kDa. However, upon incubation, the MASP-2 becomes progressively cleaved as it is converted to the active form. Activated MASP comprises two fragments which are linked by a disulphide bond. Under reducing conditions, the larger fragment migrates as two bands on SDS gels of molecular mass 50 and 54 kDa, due to differential glycosylation. The smaller fragment (~30 kDa) is masked by the MBL/SP-A. To determine the rate of activation, gels were scanned and the decrease in the amount of zymogen was monitored. The half-time for MASP-2K activation by MBL was approximately 50 minutes. By contrast, wild-type SP-A protein, which is unable to bind MASP, did not activate at all in the presence or absence of fucose-Sepharose even after prolonged incubation (16 hours). Similarly, SPA-KL, which bound MASP-2 only weakly was also unable to activate MASP-2K, showing no measureable increase above the basal rate of MASP-2K autoactivation in the absence of MBL ($t_{1/2} > 1000\text{min}$).

In order to establish how rapidly the MASP-2K was activated; densitometry was performed using Scion Image on the SDS-PAGE gels. Profiles of the density of the Coomassie-stained protein were obtained. The area under the trace corresponding to the zymogen MASP and activated MASP was calculated using Scion Image.

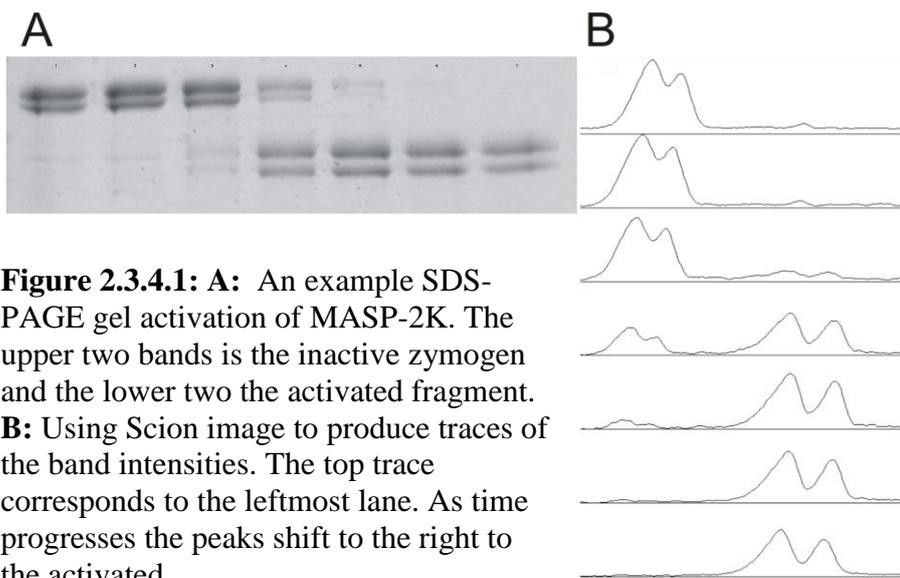


Figure 2.3.4.1: A: An example SDS-PAGE gel activation of MASP-2K. The upper two bands is the inactive zymogen and the lower two the activated fragment. **B:** Using Scion image to produce traces of the band intensities. The top trace corresponds to the leftmost lane. As time progresses the peaks shift to the right to the activated.

Once the areas have been calculated by the program the percentage of activated MASP can be calculated according to the following equation:

$$\left(\frac{1}{\frac{\text{Inactive} + 8/6 * \text{Activated}}{8/6 * \text{Activated}}} \right) \times 100$$

The activated MASP value is multiplied by 8/6 to correct for the loss of the C-terminal end of the serine protease domain. Once activated, the C-terminal end of the serine protease domain remains attached to the MASP-2K protein via the disulphide bond as described in Chapter 1. On a reducing gel this fragment detaches and runs separately to main body of the protein. This runs concurrently with the SP-A protein and so cannot be measured using densitometry.

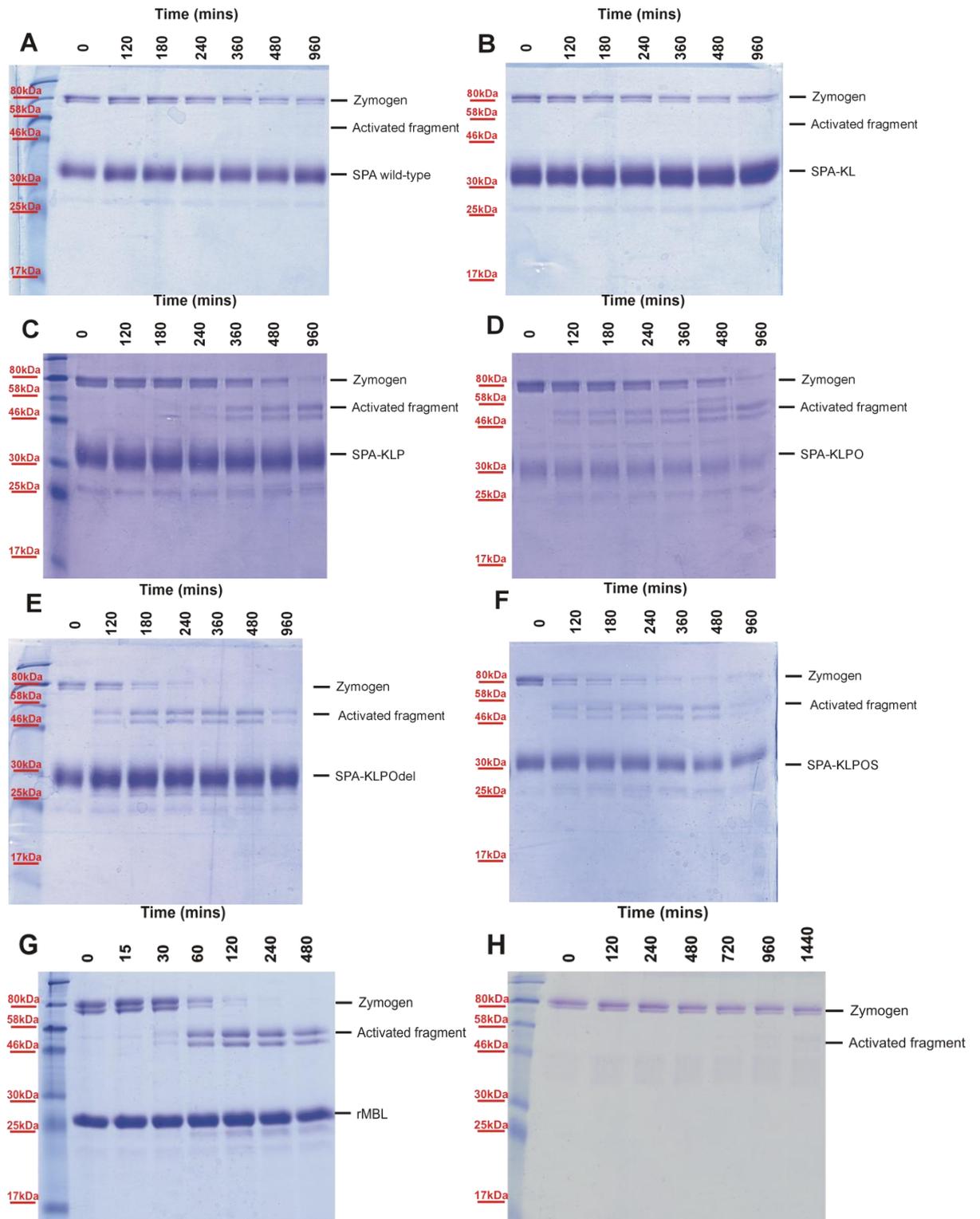


Figure 2.3.4.1: MASP-2K activation by SP-A mutants measured by SDS-PAGE in the presence of fucose-Sepharose. A and B: wild-type SP-A and SPA-KL are unable to activate MASP-2K. **C, D, E and F,** MASP-2K activation by SPA-KL, SPA-KLPO, SPA-KLPOdel and SPA-KLPOS. **G** a positive control wild-type rat MBL activation of MASP-2K. **H** The MASP-2K negative control does not activate in the absence of a suitable lectin.

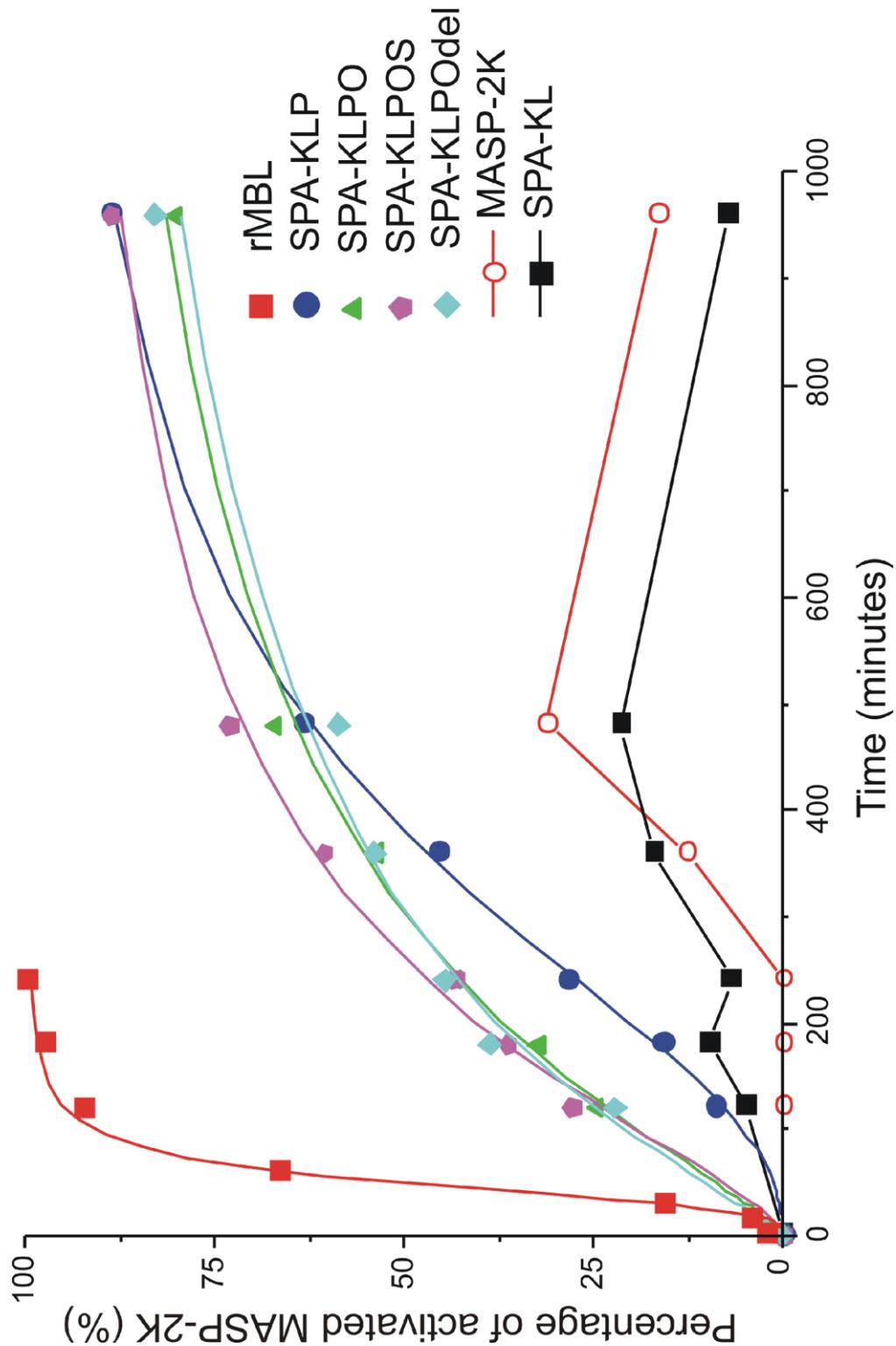


Figure 2.3.4.2: Graph of the different activation rates of the SP-A proteins against the MASP-2K and MBL controls. In the SP-A proteins previously shown to bind MASP activation occurs at comparable rates. SPA-KLP did not bind as tightly as the three different SPA-KLPO variants and there appears to be a slight lag in activation by SPA-KLP. SPA-KL showed no significant binding and the activation rate is comparable to the MASP control.

Protein	t_{1/2} (mins)	SE
MBL	50	5.2
SPA-KL	>1000	-
SPA-KLP	380	89.4
SPA-KLPO	310	79.4
SPA-KLPOS	260	8.3
SPA-KLPOdel	310	36.4
MASP-2	>1000	-

Table 2.3.4.1: Half-times of each of the SP-A proteins and MBL and MASP-2K controls. In the MASP-2K negative control and SPA-KL where we see no binding, activation rates show a t_{1/2} time of >1000 minutes. The other SP-A proteins all activate with t_{1/2} of approximately 300 minutes although this is still much slower than wild-type MBL which has a t_{1/2} of 50 minutes.

Remarkably, SPA-KLP and SPA-KLPO both activated MASP-2K significantly faster than the basal rate of autoactivation. Rates were comparable for both mutants with half-times of approximately 350 minutes compared to 50 min for MBL (~ 5-fold slower). Thus introducing the MASP-binding site into SP-A is sufficient to enable both binding and activation of MASP-2. SPA-KLPOdel and SPA-KLPOS also activated MASP-2K with similar rates to the KLP and KLPO mutants, indicating that the kink in SP-A does not prevent MASP activation, nor does it modulate the rate at which activation occurs.

MASP activation is normally carbohydrate-dependent so it was important to compare activation in the presence and absence of the carbohydrate target. Therefore, assays were repeated under both conditions using MBL, wild-type SP-A and the KLPO, KLPOdel and KLPOS mutants. As expected MASP-2 activation by MBL was dependent on the presence of fucose-Sepharose. In its absence there was no detectable activation over the time course of the experiment (figure 2.3.4.2.E). SP-A did not activate MASP-2 either in the presence or absence of fucose-Sepharose. This was an important control because it shows that preparations of SP-A are free from contaminating proteases, which could give a false-positive result by cleaving the MASP over the prolonged time course of the experiment. Unexpectedly, all three SPA-KLPO mutants activated MASP even in the absence of a carbohydrate target. Comparable rates were measured in the presence and absence of fucose-Sepharose. This constitutive activation could only be as a result of MASP binding to SP-A since wild-type SP-A, which cannot bind MASP, did not activate MASP.

Overall, these results indicate that following introduction of a function MASP-binding site, SP-A possesses the necessary architectural feature to activate MASP-2, and hence complement. However, it lacks the necessary control mechanisms that enable MBL and ficolins to activate MASP selectively, only in the presence of a carbohydrate surface.

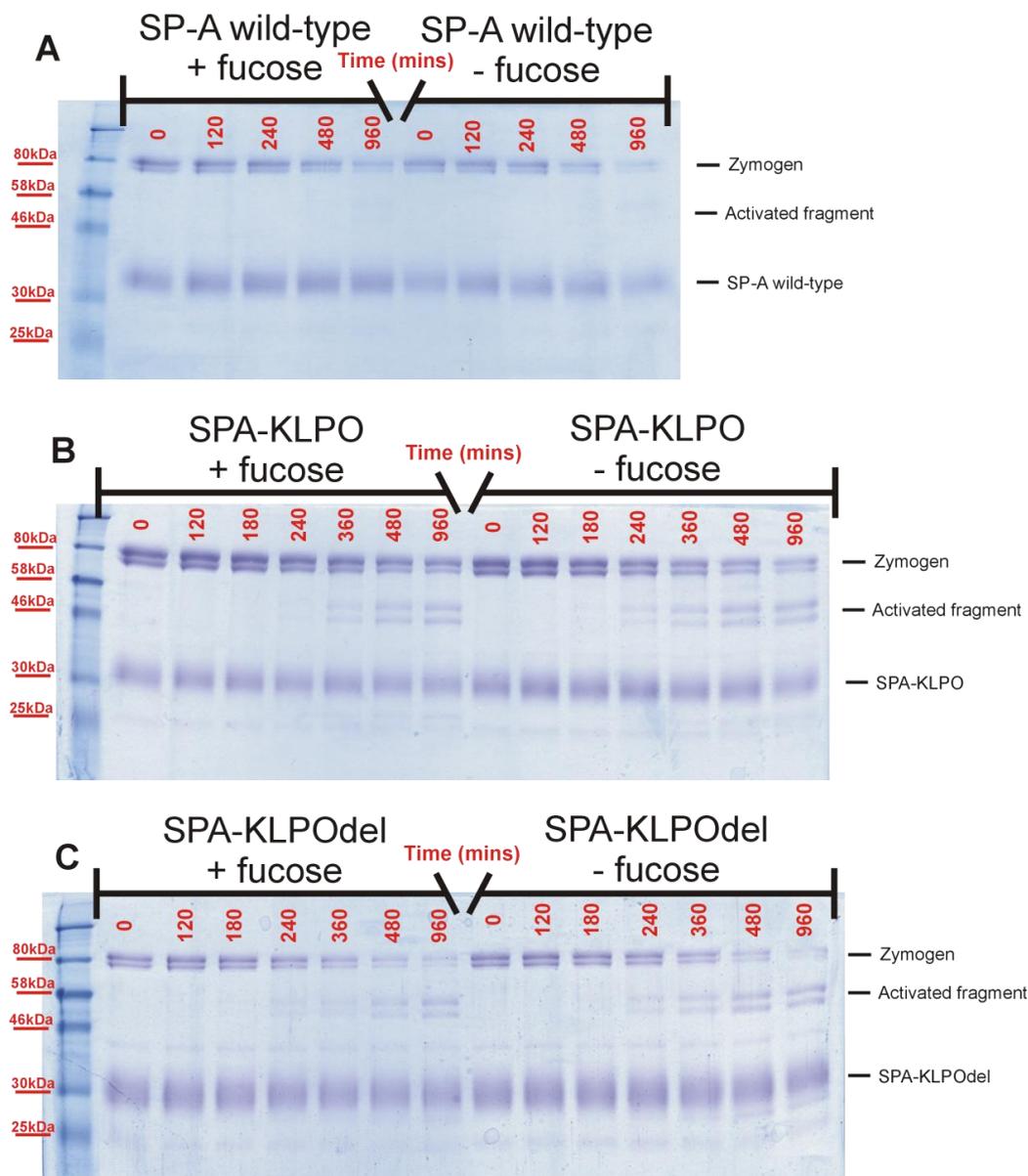
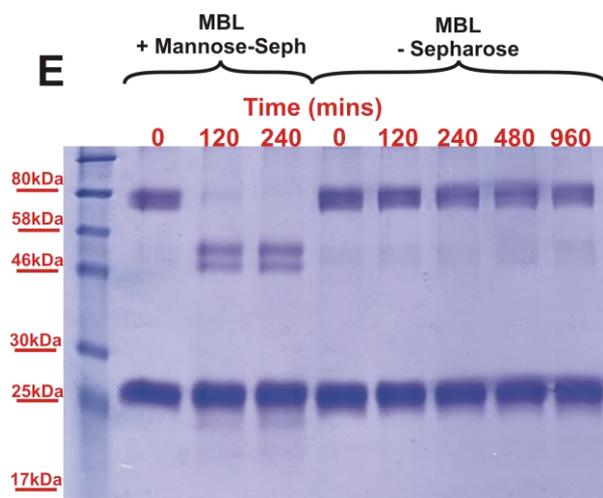
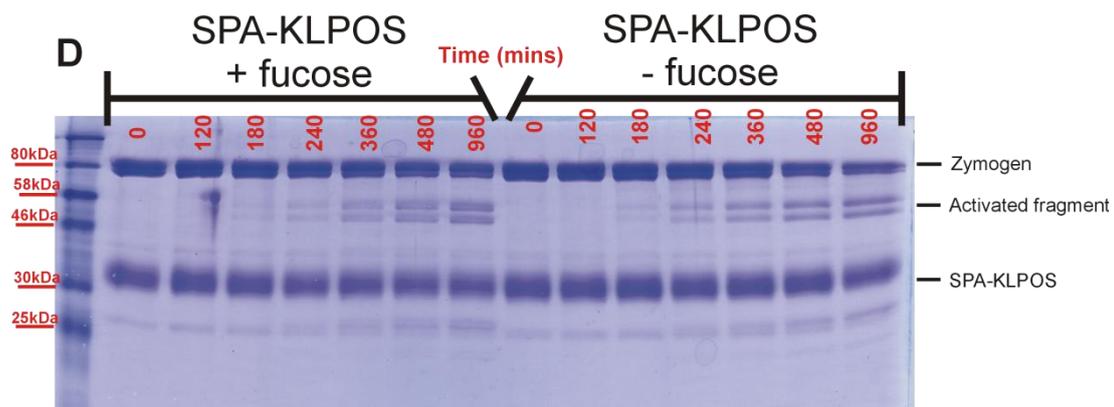


Figure 2.3.4.2: A: MASP-2K activation by SP-A mutants monitored by SDS-PAGE in the presence of fucose-Sepharose. A: wild-type SP-A cannot activate MASP-2K in the presence or absence of fucose-Sepharose. **B, C and D (see over):** By contrast, SPA-KLPO, SPA-KLPOdel and SPA-KLPOS all activate MASP-2K in the presence and absence of fucose-Sepharose. **E (see over):** rat MBL activates at the established rate in the presence of fucose-Sepharose. Activation does not take place in the absence of fucose-Sepharose.



2.4 Discussion

In this Chapter I have shown that MASP-2 binding can be introduced into the collagenous stalks of SP-A through just three amino acid substitutions, Glu⁶³Lys, Cys⁶⁴Leu and Glu⁶⁶Pro. Lysine has been shown previously to be the essential for binding of MBL to MASP and ficolin to MASP (Girija *et al.*, 2007, Wallis *et al.*, 2004) in both rats and humans. The adjacent residue is not strictly conserved in MBLs (leucine in rat MBL and the SP-A mutants) but is usually an aliphatic residue or a methionine. Mutagenesis of rat ficolins-A and -B have shown that small polar residues, such as serine, can also be accommodated in this position, but glutamate cannot (Girija *et al.*, 2007). It is likely that the substitution of the proline residue in place of Glu⁶⁶ in SPA-KLP removes unfavourable electrostatic interactions rather than providing direct contacts to the MASP. It has been shown that the proline can be replaced by an alanine residue in rat ficolin-A with only minimal decrease in affinity (Girija *et al.*, 2007) meaning that it is not essential for binding and may only help to stabilise the collagenous domain itself near the MASP-binding region.

The binding data described here demonstrate that inserting the MASP-binding motif alone is sufficient to introduce binding into SP-A almost equivalent to MBL, ruling out any interactions via additional sites on the MBL molecule. Given the lack of sequence conservation together with the high frequency of Pro and Hyp in the Xaa and Yaa positions of collagen, specificity appears to be provided almost entirely by the single lysine residue within the binding motif. Interestingly, all other lysine residues in the collagenous domain of MBL fall in the Yaa position of the Gly-Xaa-Yaa repeat and are likely to be post-translationally modified by hydroxylation and glycosylation. This

would block any MASP interactions with the collagenous domain, except at the binding site.

The other feature that dictates high-affinity binding of MBL to MASP is the architecture of the MBL (or ficolin) molecule. Tight binding requires multiple relatively weak interactions (at least two and up to four) between separate collagen-like stems and the binding sites on the MASP (Gingras *et al.*, 2011, Phillips *et al.*, 2009). As a consequence, the binding specificity is likely to be dependent on the architectures of the subcomponents and the multivalent nature of the interactions, which in this case is mimicked by the comparable structure of SP-A.

Complement activation is triggered by changes in the angle between collagenous stalks when complexes bind to an activating target (Wallis *et al.*, 2010, Wallis, 2007).

Significant conformational and rotational flexibility of the stalks is likely to be required, so MBL is probably quite flexible. The most surprising finding in the current work is that the modified SP-As can also activate complement, but that activation is uncontrolled. Two potential mechanisms can be postulated to explain activation. The first is that of a relaxed-to-strained system, in which, upon binding, MBL undergoes a change in structure that induces a high-energy strained state and the resulting changes are transmitted through the collagenous domain to the bound MASP zymogen. The second is a strained-to-relaxed mechanism in which MASP binding to MBL induces strain into the complex, which is released upon pathogen binding allowing MASP activation to occur freely.

A major difference between the two mechanisms is in the default state of the circulating complex. In the relaxed-to-strained mechanism the default complex is inactive, because it requires pathogen binding to convert to the high energy state and hence activate. In the strained-to-relaxed mechanism the default complex is active because pathogen binding releases strain thus permitting MASP auto-catalysis. The finding that SP-A activates MASPs constitutively, supports a strained-to-relaxed mechanism because the active state is default for the SP-A mutants. While, the possibility that modified SP-As are locked into a high-energy “activating” conformation cannot be completely excluded, it is clear that large structural changes (through removal of the kink) do not affect MASP activation by SP-A, which seems incompatible with such a mechanism.

Recent data has suggested that the C1 complex also activates via a strained-to-relaxed mechanism (Phillips *et al.*, 2009). In C1 however, the strain is probably imparted by interactions between the protease domains of C1r. This cannot be so for MASPs because the serine protease domains do not interact significantly until activation occurs. Interestingly, however, recent analysis suggests that the MBL stalks are not evenly distributed, but rather are highly asymmetrical and separated by $\sim 35\text{--}40^\circ$ between nearest neighbours (Jensenius *et al.*, 2009). In this case, significant strain would be induced when MBL binds to a MASP. Release of this strain upon target recognition might drive the changes that initiate complement activation.

Overall, the data described in this Chapter suggest that MBL-MASPs are primed for activation as they circulate in serum. The major difference between these complexes and those formed with the modified SP-A is that activation is prevented until pathogen binding occurs. This conclusion highlights the need for further study to determine

where in the sequence of MBL this control originates from. This question will be the main focus of Chapter 3.

Chapter 3 – Control of MASP activation

3.1 Introduction and Objectives

The data in Chapter 2 shows that MASP-binding and complement activation can be introduced into SP-A by changing its collagenous domain to more resemble MBL. Although the resulting mutant SP-A initiated MASP-2 activation, activation was constitutive, and occurred even in the absence of a carbohydrate target. MBL must therefore possess additional structural features that control MASP-2 activation so that it occurs only in the correct place at the right time. The aim of the work described in this Chapter is to identify which regions of MBL are involved in this regulatory function.

Control of activation may be mediated through the N-terminal domain which links the trimeric subunits together, the collagenous domain or might originate in the α -helical neck and CRDs. To identify which regions are important I made pairs of reciprocal chimeras in which individual domains were swapped and tested their ability to activate complement.

In the second half of this Chapter I have investigated the importance of sugar specificity towards the ability of MBL to activate complement. C-type CRDs are ubiquitous domains and different family members display different sugar specificities. In general, recognition is achieved through interaction of a monosaccharide at the site equivalent to that in MBL, but specificity is influenced by surrounding residues. In addition, some CRDs, e.g. DC-SIGN and members of the selectin family possess secondary binding

sites involving additional contacts to a separate region on the CRD (Mitchell *et al.*, 2001, Weis *et al.*, 1998).

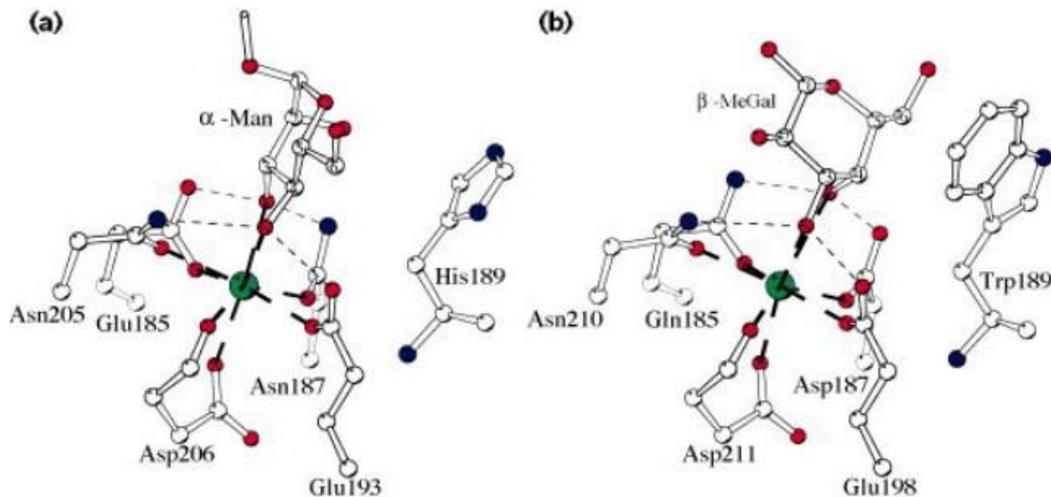


Figure 3.1.1: Comparison of sugar binding by a) MBL and b) MBL QPDWG (subsequently called galactose-binding lectin): Subtle changes at the carbohydrate binding site switch specificity from mannose to galactose. From Structure (1997) 5, 465–468. (Drickamer, 1997)

Previous studies have shown that galactose-binding activity can be introduced into the CRD of mannose-binding lectin by replacing residues at or near the sugar binding site in MBL with residues from the galactose-specific asialoglycoprotein receptor (Iobst *et al.*, 1994a). The asialoglycoprotein receptor is a scavenger receptor that functions to remove target glycoproteins from circulation. In particular, mutations Glu¹⁸⁵Gln and Asn¹⁸⁷Asp switch specificity from mannose to galactose, but with relatively low affinity for either sugar. Additional changes His¹⁸⁹Trp and insertion of a glycine-rich loop (replacement of Gly¹⁹⁰-Ser¹⁹¹ by Tyr-Gly-His-Gly-Leu-Gly-Gly to create the "QPDWG mutant CRD") results in a mutant with both high galactose selectivity and ~mM affinity. I will subsequently refer to full-length MBL containing the QPDWG mutations as Galactose-Binding Lectin (GBL).

GBL is of potential interest as a therapeutic. Previous research has shown that during oncogenesis the processes of glycosylation are affected often resulting in truncated structures on the cell surface (Hakomori, 1996), one result of which is to produce a terminal galactose. Two structures that have been identified are the cancer-associated T antigen (Gal β 1-3 GalNAc) and the Lewis^X trisaccharide (Gal β 1-4(Fuc α 1-3)GlcNAc) (Kim *et al.*, 1997).

Both monoclonal and polyclonal antibodies have been used for the identification of this differing glycosylation (Dube *et al.*, 2005). Published research also shows that the trimeric CRD fragment of this protein has the ability to bind a number of cell types presenting terminal galactose residues on their surface (Powlesland *et al.*, 2009). The full length GBL construct has the potential to both bind to and kill through complement mediated lysis cells expressing galactose residues on their surfaces. It is therefore important to establish whether this protein is able to activate complement in a target specific manner and if it does, it could be regarded as a potentially novel therapy for the killing of cancerous cells via the lectin pathway of complement.

This chapter aims to establish the domain from which control originates through producing chimeras containing domains from either MBL or SP-A. It will also investigate whether the carbohydrate specificity has any role in the activation process outside of target recognition.

3.2 Materials and Methods

3.2.1 – PCR construction of MBL/SP-A chimeras

To avoid potential problems arising from the cysteine residues in the collagen-like domain of SP-A forming incorrect disulphide bond during biosynthesis, SPA-KLPOdel in which the cysteine residue and the accompanying hinge region has been deleted, was used as a template for subsequent mutagenesis instead for of wild-type SP-A. As described in Chapter 2 above, these changes do not affect MASP binding or activation.

In each case, chimeras were constructed in two parts by separate PCR reactions and the products were then fused together in a third PCR reaction to generate the full length chimera cDNA. The primers used are detailed in table 3.2.1.1.

	Primer	Sequence (5'→3')
Terminal Primers	SP-A N-term F	ATACATGTCGACGCCACCATGTGGCTGTGCCCTCTG
	MBL N-term F	ATACATGTCGACGCCACCATGCTCCTGCTTCCACTG
	SP-A-CRD Rev	GATGAATTCTCAGAAGTCAACAGATGGT
	MBL-CRD Rev	GACGAATTCTCAGGCTGGGAAGTCA
SP-A Internal Primers	SP-A N-term Rev	ATCTCTCCCGTCTCTGCCAACACAAACGTCCTTAC
	SP-A-Col Rev	CTTCACCTCAATGGCTCTTGGGAAGCCCTGGAGGGCC
	SP-A-Col F	TGCTCTGTGATAGCCTGCGGAAGCCCTGGTATCCCC
	SPA-CRD F	GGGGATCGTGGAGACAGCGCTCATCTAGATGAGGAG
MBL Internal Primers	MBL N-term Rev	GGGGATACCAGGGCTTCCGAGGCTATCACAGAGCA
	MBL-Col Rev	CTCCTCATCTAGATGAGCGCTGTCTCCACGATCCCC
	MBL-Col F	GTGAAGGACGTTTGTGTTGGCAGAGACGGGAGAGAT
	MBL-CRD F	GGCCCTCCAGGGCTTCCAAGAGCCATTGAGGTGAAG

Table 3.2.1.1: Primers used to clone the MBL/SP-A Chimeras. The terminal primers are highlighted as follows; Methionine start codon (*green*), Stop codon (*red*), Kozak sequence (*yellow*), restriction sites (*purple*) (GTCGAC is *Sall* and GAATTC is *EcoRI*). The internal primers are highlighted for MBL sequence (*blue*) and SP-A sequence (*pink*).

The tables below detail the template and primers used in each of the two rounds of PCR along with the domains that the primers were designed to amplify. They have been grouped according to the chimera they form as shown in Fig 3.1.1.1. The conditions for the PCR reactions were the same as those detailed in section 2.2.1.

A				
First Round of PCR				
Chimera	Template	Domain	Forward Primer	Reverse Primer
1	MBL	N-terminal and collagen	MBL N-term F	MBL-Col Rev
	SPA-KLPOdel	CRD/neck	SP-A-CRD F	SPA-CRD Rev
2	SPA-KLPOdel	N-terminal and collagen	SPA1 N-term F	SPA-Col Rev
	MBL	CRD/neck	MBL-CRD F	MBL-CRD Rev
3	MBL	N-terminal	MBL N-term F	MBL N-term Rev
	SPA-KLPOdel	Collagen and CRD/neck	SP-A-Col F	SPA-CRD Rev
4	SPA-KLPOdel	N-terminal	SP-A1 N-term F	SPA1 N-term Rev
	MBL	Collagen and CRD/neck	MBL-Col F	MBL-CRD Rev

Table 3.2.1.2: PCR mutagenesis reactions. A (above): This shows the first round of PCR in which the domains are amplified. **B (below):** Shows the second round of PCR where the domains are combined to form the full length chimeric DNA sequence.

B				
Second Round of PCR				
Chimera	Template 1	Template 2	Forward Primer	Reverse Primer
1	MBL N-term & col	SP-A CRD/neck	SPA1 N-term F	MBL-CRD Rev
2	SP-A N-term & col	MBL CRD/neck	MBL N-term F	SPA-CRD Rev
3	MBL N-term	SP-A col and CRD/neck	SPA1 N-term F	MBL-CRD Rev
4	SP-A N-term	MBL col and CRD/neck	MBL N-term F	SPA-CRD Rev

The second round PCR products were A-tailed and ligated into pGEM-T according to NEB and Promega protocols respectively. Four white colonies of each construct from a blue/white screen were taken and the plasmid DNA prepared using a Qiagen miniprep kit. The plasmids were digested using *EcoRI* and *SalI* to test for the presence of the gene. Constructs were sequenced by PNAACL at the University of Leicester.

3.2.2 – Cloning, Transfection and Expression

The PCR fragments were digested using *EcoRI* and *SalI* and ligated into the *EcoRI* and *XhoI* sites of expression plasmid pED4 using T4 DNA ligase according to NEB

protocols. The calcium phosphate precipitation transfection method described in section 2.2.4 was used to introduce the plasmid into the DXB11 cells. Colony selection and protein expression was achieved by increasing the concentration of MTX in the growth media described in section 2.2.5.

3.2.3 – MBL/SP-A and GBL Chimera Purification

The chimaeric proteins were generated as described previously purified using 1ml mannose-sepharose columns as described in 2.2.7. They were eluted using 50 mM Tris-HCl pH7.4, 100 mM NaCl and 5 mM EDTA and subsequently dialysed into 50 mM Tris-HCl pH7.4, 100 mM NaCl and 5 mM CaCl₂. GBL was purified using the same protocol as MBL however galactose-Sepharose was used in place of mannose-Sepharose.

3.2.4 – MASP-2k Activation Assays

The assays were performed as described in section 2.2.10.

3.3 - Results

A total of four chimaeras were produced (see figure 3.1.1.1):

1. MBL with the neck and CRDs of SP-A
2. SP-A with the neck and CRDs of MBL
3. SP-A with an MBL N-terminal domain
4. MBL with an SP-A N-terminal domain

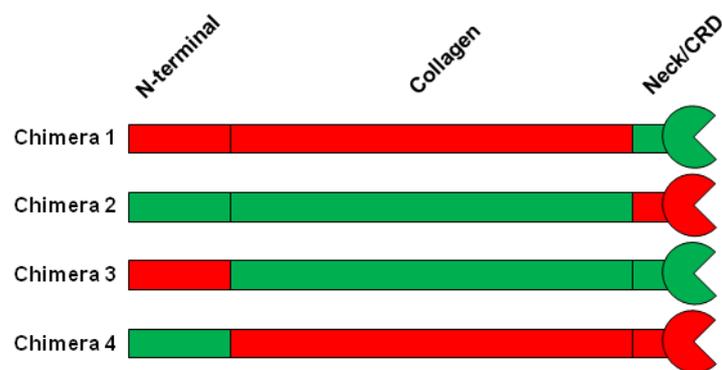


Fig 3.1.1.1. The domain organisation of the MBL/SP-A chimeras. SP-A domains are in green and MBL domains are in red.

In each case the junction was designed at the domain boundary (either at the N-terminal domain/collagen domain boundary or the collagen domain/neck boundary) to minimise disruption of folding during biosynthesis.

3.3.1 – Cloning

Chimeras were created using two rounds of PCR and clones were verified by restriction digest (Fig. 3.3.1.1) and by DNA sequencing.

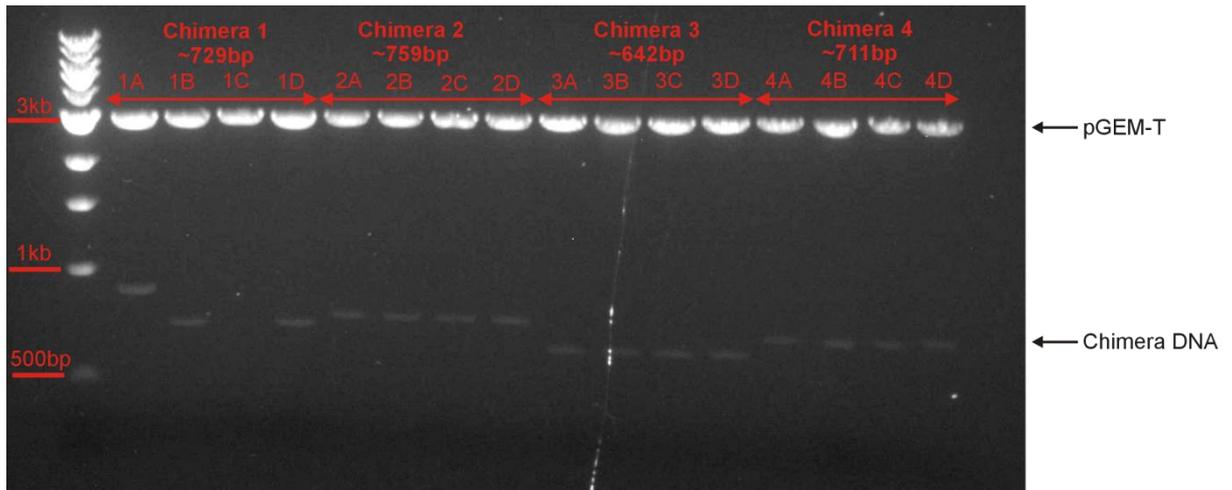


Figure 3.3.1.1: Agarose gel electrophoresis of four different MBL/SP-A chimeras. In each case, cDNAs in pGEM-T were digested with *EcoRI* and *SalI* to cut out the intact cDNA. Predicted sizes for the different chimeras were (1): 711 base pairs (bp) (2): 741 bp (3): 624 bp (4): 693 bp. Clones 1B and 1D, together with all four clones of chimera 2, 3 and 4 were of the expected size. The *EcoRI* – *SalI* fragments were cloned into mammalian expression vector pED4 for production.

3.3.2 Production of chimeras in CHO cells

Chimeras were produced by expression in mammalian cells and purified by affinity chromatography on mannose-Sepharose columns as for wild-type proteins. All eluted from the columns with similar yields to wild-type SP-A (~ 1-3 mg/L of culture) confirming that they had folded correctly during biosynthesis (Figure 3.3.2.1). All chimeras migrated with expected molecular masses on SDS polyacrylamide gels.

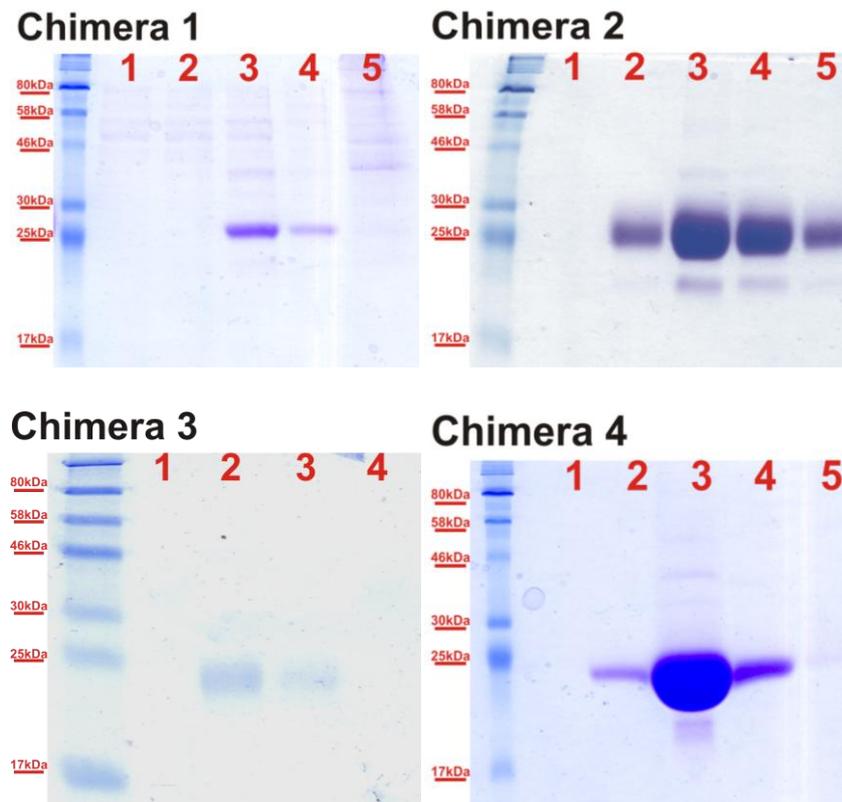


Figure 3.3.2.1: SDS-polyacrylamide gels of the elution fractions for the chimeric SP-A/MBL proteins by affinity chromatography. The expected molecular masses are Chimera 1: 25.27 kDa Chimera 2: 25.84 kDa Chimera 3: 22.40 kDa Chimera 4: 24.53 kDa.

3.3.3 – MASP-2K activation by MBL-SPA chimeras

The chimeric proteins were tested for their ability to activate MASP-2K. If control of activation originates from the CRD and neck region of MBL, I would expect to see target-specific activation occurring in those chimeras that contain the CRD and neck domains of MBL. If, however, control of activation is mediated via the N-terminal domain of MBL, those chimeras with the MBL N-terminal domain will show activation only on the carbohydrate target. The SDS-polyacrylamide gels and kinetics of activation for each chimera are shown in figures 3.3.3.1 and 3.3.3.2 respectively.

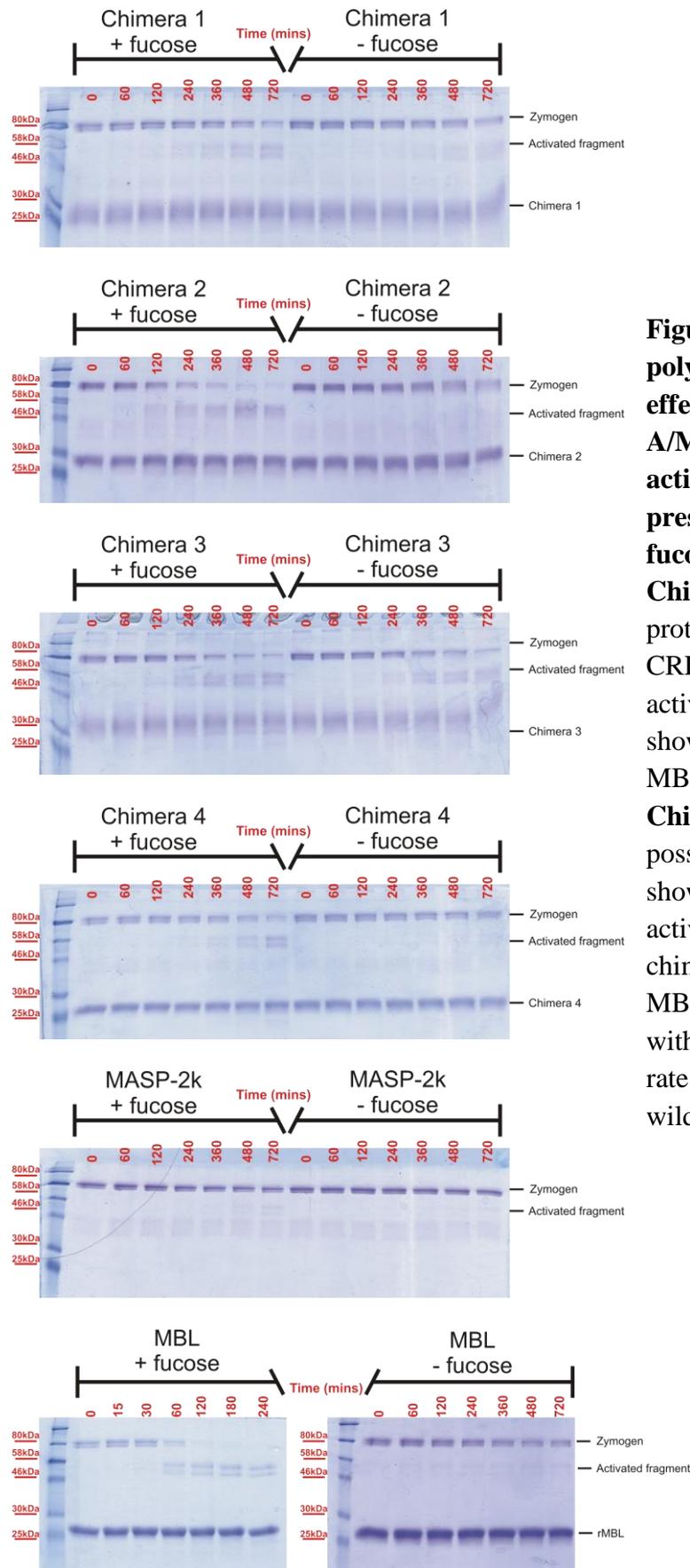


Figure 3.3.3.1: SDS-polyacrylamide gels of the effect of Chimeric SP-A/MBLs on MASP-2K activation. Both in the presence and absence of fucose-Sepharose. Chimeras 1 and 3: These proteins contain the SP-A CRD and are both seen to activate above the basal rate shown by the MASP-2K and MBL negative controls. **Chimeras 2 and 4:** These possess the MBL CRD and show target specific activation. In the case of chimera 4, replacing the MBL N-terminal domain with that of SP-A slows the rate considerably to that of wild-type MBL.

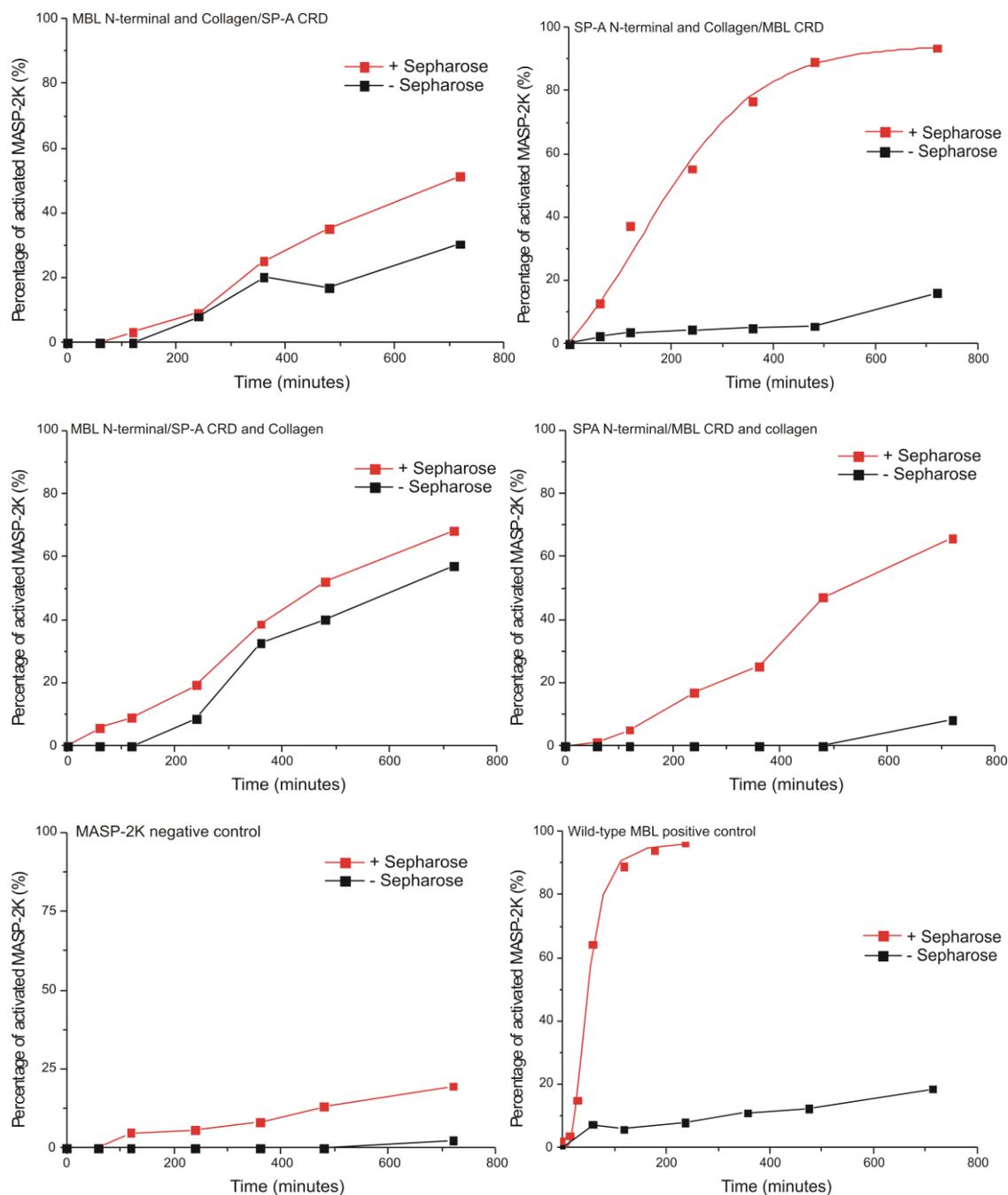


Figure 3.3.3.2: The activation behaviours of the four MBL/SP-A chimeric proteins and MASP-2K and MBL controls. Top row: Left; Chimera 1: Replacing the MBL CRD with that of SP-A leads to slow constitutive MASP-2K activation both in the presence or absence of fucose-Sepharose. **Right; Chimera 2:** The neck and CRD of MBL confers target specific activation to SP-A. Activation is much more efficient in the presence of fucose-Sepharose. **Middle row: Left; Chimera 3** has a phenotype similar to chimera 1, with enhanced MASP-2K activation both in the presence or absence of fucose-Sepharose. **Right; Chimeras 4** has a phenotype similar to chimera 2, although activation is slower in the presence of fucose-Sepharose. **Bottom row:** The basal level of MASP-2K activation is slow both in the presence and absence of fucose-Sepharose. MBL activates only in the presence of a carbohydrate target.

As shown in figure 3.3.3.2, in the absence of any chimera, activation of MASP-2K occurs at a basal rate which is relatively slow (<20% activation even after 700 min) both in the presence and absence of fucose-Sepharose. By comparison, chimers 2 and 4 show clear evidence of target-specific activation. In the absence of a fucose-Sepharose matrix there is relatively little activation of MASP-2K, no more than basal. However, in the presence of fucose-Sepharose activation is greatly enhanced. This is most evident in chimera 2, in which the halftime for activation ~ 200 min. Activation by chimera 4 is slower, with a half-time of ~500 min, but over the same period of time there is almost no activation in the absence of fucose-Sepharose demonstrating that it is still target specific. By contrast Chimeras 1 and 3 show evidence of enhanced activation both in the presence and absence of fucose-Sepharose. In each case similar rates are seen under indicating that they lack the regulatory mechanism of MBL.

Overall therefore, the neck and CRDs of MBL are important for ensuring activation occurs only in the presence of a carbohydrate surface. Additional regions of MBL do not further enhance activation. In addition, it is clear that replacing the N-terminal domain of MBL with the corresponding region of SP-A (chimera 4) severely compromises MASP-2 activation (half time increases from <60 min to ~500 minutes). Nevertheless, activation is still target specific, despite occurring more slowly.

3.3.4 - Switching sugar specificity in MBL from mannose to galactose: creation of a galactose-binding lectin

Given the clear importance of the neck and CRD of MBL not only for pathogen recognition but also in controlling activation, it was of interest to further investigate the role of the CRD itself. Each CRD may serve simply as a binding domain or alternatively might somehow transmit a binding signal via the collagen-like domain to initiate MASP activation. In the latter scenario, changing the specificity of the domain would alter its structure and thereby would likely disrupt its ability to activate complement. If on the other hand the CRD simply acts to tether MBL onto the surface of a pathogen, switching the sugar specificity would not be expected to affect MASP activation.

Previous studies have already shown that the specificity of the CRD towards sugars can be changed relatively easily. For example, selectin-like or galactose specificity can be introduced into MBL through relatively few changes to the collagen-like domain. The aim of this part of the project was to alter the specificity of MBL from mannose to galactose and test whether the resulting protein still activated complement.

3.3.5 – Purification of GBL

GBL was purified by affinity chromatography over a galactose-Sepharose matrix. This confirms both that it folds and possesses the ability to bind to galactose. The protein was run on an SDS-polyacrylamide gel and the protein band was visible at a size of approximately 27 kDa (figure 3.3.5.1).

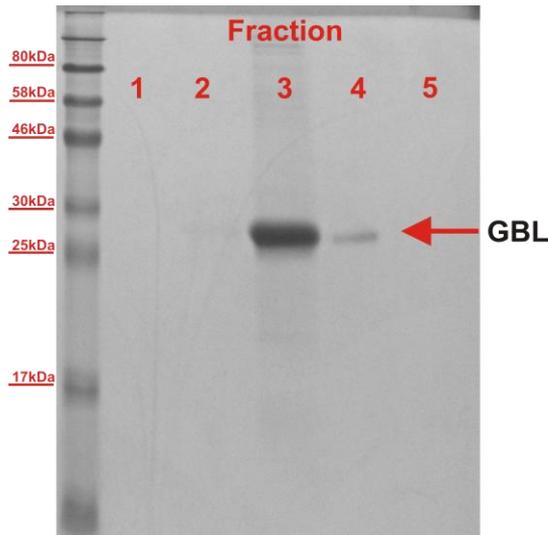


Figure 3.3.4.1: SDS-Polyacrylamide gel of the GBL elution fractions. GBL elutes from a galactose-sepharose column by the addition of buffer containing EDTA. This confirms correct folding and Ca^{2+} -dependent galactose-binding.

3.3.5 – Activation of MASP-2K by the galactose-binding lectin

To compare the differences in the activities of MBL and GBL, duplicate activation assays were undertaken using galactose-Sepharose, mannose-Sepharose or the absence of any Sepharose matrix. The latter condition was used to confirm that no MASP-2K autoactivation occurs in the absence of a carbohydrate target. The data are shown in figure 3.3.5.1. As before, MASP-2K band intensities were measured and the kinetics of activation were compared under the different conditions (figure 3.3.5.2).

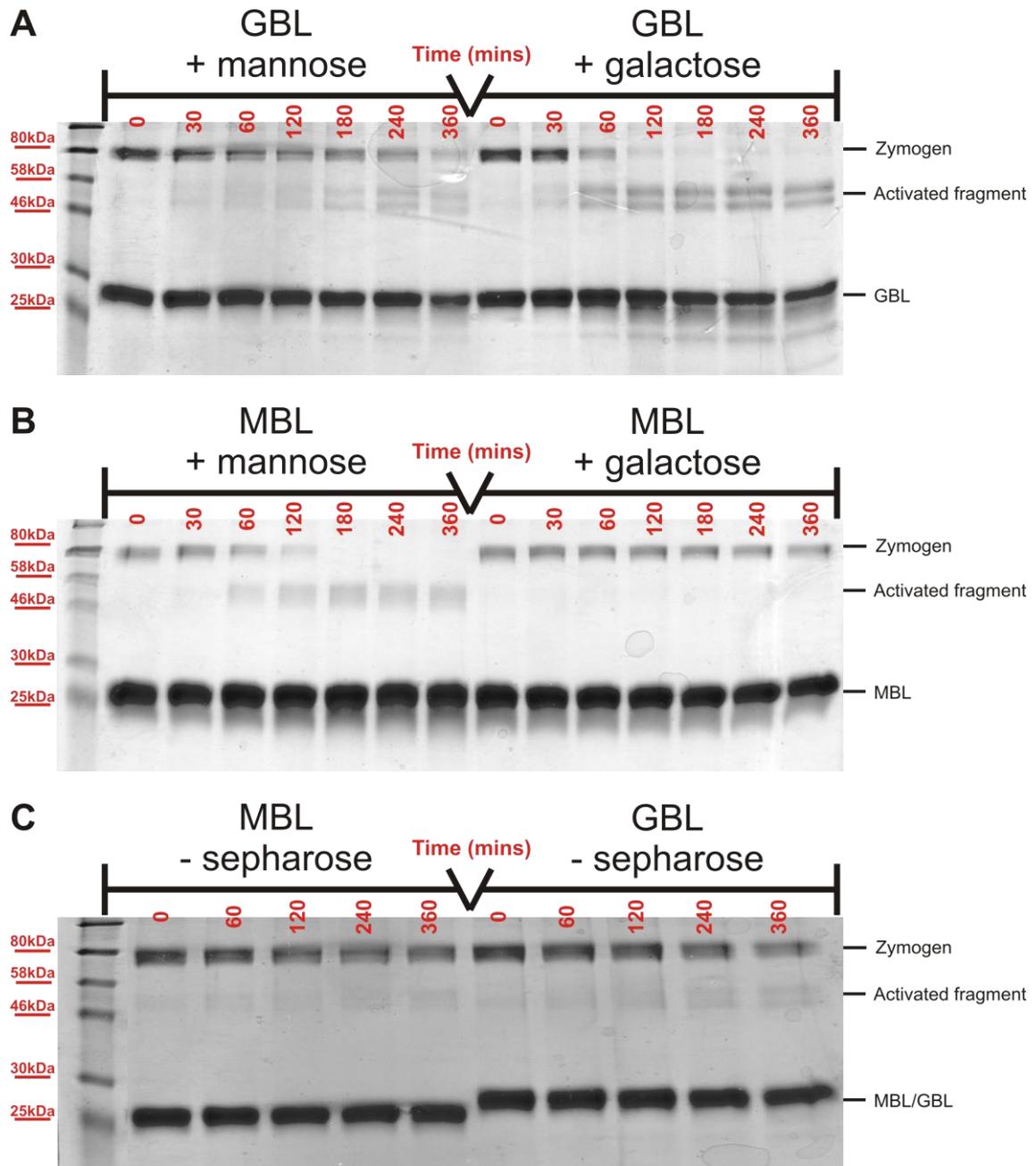


Figure 3.3.5.1: SDS-polyacrylamide gels of the MASP-2 activation by MBL and GBL. **A:** On galactose-Sepharose there is clear activation of MASP-2K by GBL with a half-time close to 60 minutes. Activation on mannose-Sepharose is much slower by some residual activation still occurs. **B:** MBL activates MASP-2 on a mannose target as previously demonstrated but is unable to activate on galactose-Sepharose. **C:** In the absence of any sugar-Sepharose, a basal level of activation occurs for both MBL and GBL.

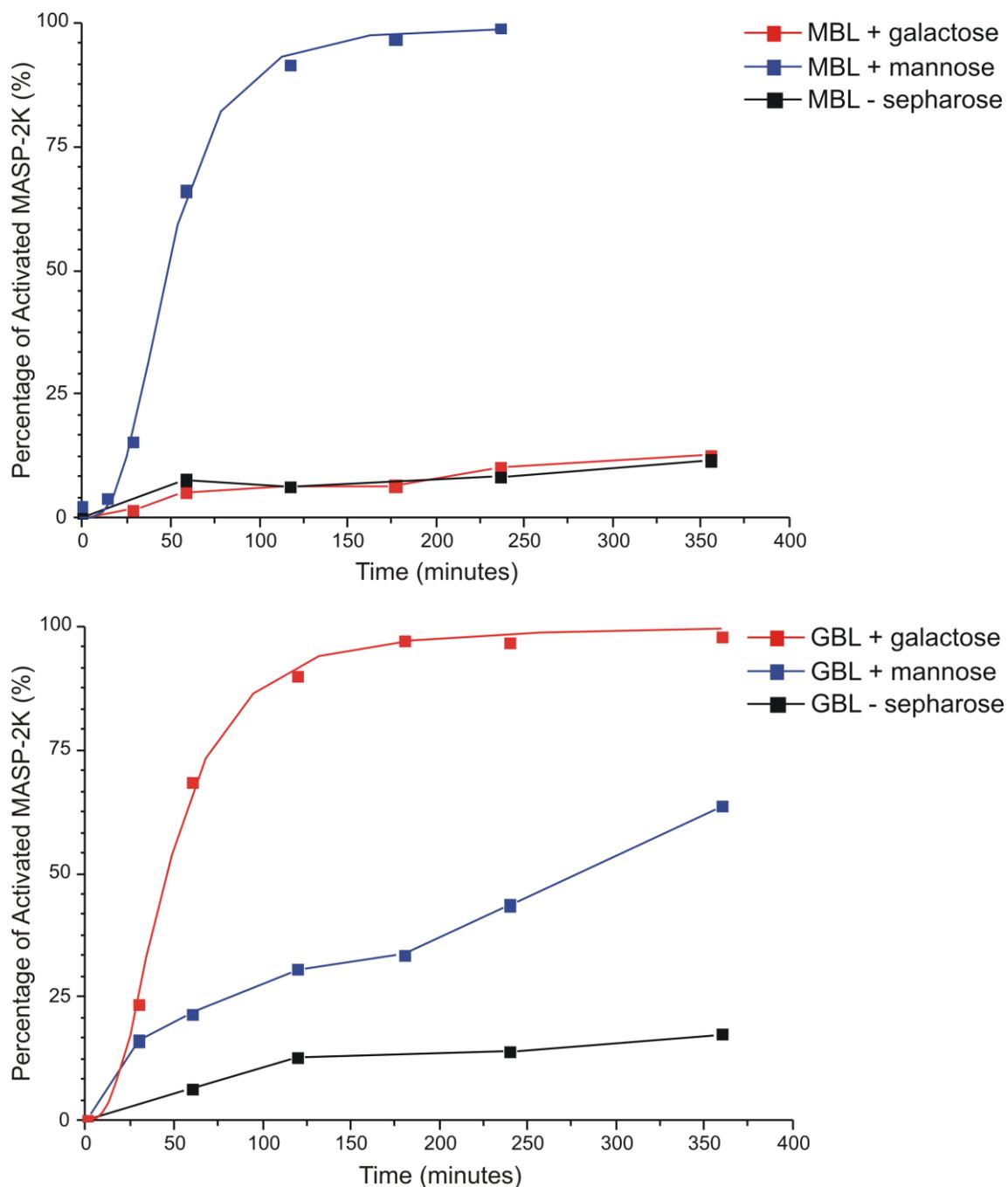


Figure 3.3.5.2: Graphs showing activation MASP-2K by MBL and GBL. Top: MBL only activates MASP-2K on mannose-Sepharose. Negligible activation is seen on galactose-Sepharose or in the absence of Sepharose. **Bottom:** On galactose-Sepharose, GBL activates MASP-2K at a comparable rate to that of MBL on mannose-Sepharose. On mannose-Sepharose, GBL still shows some activation but the rate is six times slower than on galactose-Sepharose. In the absence of Sepharose, MASP-2 autoactivation is slow in the presence of MBL and GBL, reflecting the basal autoactivation rate.

Protein	t_{1/2}	SE
MBL + Man-Seph	50	6.3
MBL - Seph	>400	-
MBL + Gal-Seph	>400	-
GBL + Man-Seph	260	63.7
GBL - Seph	>400	-
GBL + Gal-Seph	50	32.3

Table 3.3.5.2: Half-times of the MBL and GBL proteins in the presence of Mannose or Galactose Sepharose or no Sepharose controls. There is no discernible difference between the t_{1/2} values of MBL and GBL when incubated with their target sugar-Sepharose matrix. There is some residual mannose binding present in GBL. This results in a slower activation rate with half the MASP-2K cleaved after 260 minutes.

The data show that that GBL does indeed activate MASP-2K on a galactose-Sepharose. Moreover the rate of activation was comparable to MASP activation by MBL on mannose-Sepharose. Activation is target specific, occurring much faster on galactose-Sepharose than on mannose-Sepharose. Some residual activation on mannose-Sepharose was still observed, but this was much slower than on galactose-Sepharose and probably reflects low-level binding of GBL to mannose. As expected, activation of MASP-2K by MBL is specific to a mannose target with no activation seen in either the presence of galactose-Sepharose or absence of a Sepharose matrix.

3.4 Discussion

The data show that the CRD/neck region of MBL is important not only for recognising a carbohydrate target, but also for ensuring that MASP activation is target specific.

Replacement of this region with the corresponding neck and CRDs of SP-A still allows activation to occur but it does so constitutively in the presence or absence of a carbohydrate target.

The control of target specific complement activation is unlikely to involve changes in individual CRDs. As shown here, the specificity of MBL can be completely switched from mannose to galactose without impairing MASP activation. Crystal structures are available for the CRD/neck region of MBL and SP-A both in the presence and absence of carbohydrates. For each protein, binding sites are well separated from the neck (Shang *et al.*, 2011, Ng *et al.*, 2002). Furthermore, the presence of a carbohydrate in the binding pocket has very little effect on the CRD itself, so it is unlikely that a signal is transmitted from the sugar-binding site through the polypeptide chain to the neck region.

On the other hand, little is known about the junction between the neck and the collagen-like domain, so difference in this region could affect activation by MBL and modified SP-A. Such a mechanism would require the signal to be transmitted to the MASP via the collagen-like domain. The junction between the neck and collagen-like domain is likely to be flexible because of the change in register of the three polypeptide chains from a one-residue stagger in the collagen domain to perfect alignment in the α -helical coiled coil. Interestingly, flexibility is observed in macrophage scavenger receptors

which also possess linked α -helical coiled coils and collagen-like domains (although in this case the coiled coil is N-terminal to the collagen). Furthermore the angle between the domains is dependent on the pH (Resnick *et al.*, 1996). Thus, it is conceivable that specific changes in the CRD/neck region of MBL upon pathogen binding control activation.

A noticeable difference between the neck and CRDs of MBL and SP-A is the length of the neck itself (figure 3.4.1). In MBL this domain is 54 amino acid residues in length but it is much smaller in SP-A comprising only 31 residues. The length of the neck is likely to influence the stability of the CRD and neck trimers. This in turn could affect the adjacent collagen-like domain with its associated MASPs. Additional studies are required to elucidate how the activating signal may be transmitted down the MBL molecule, but investigating the relative stabilities of the domains may provide a good starting point.

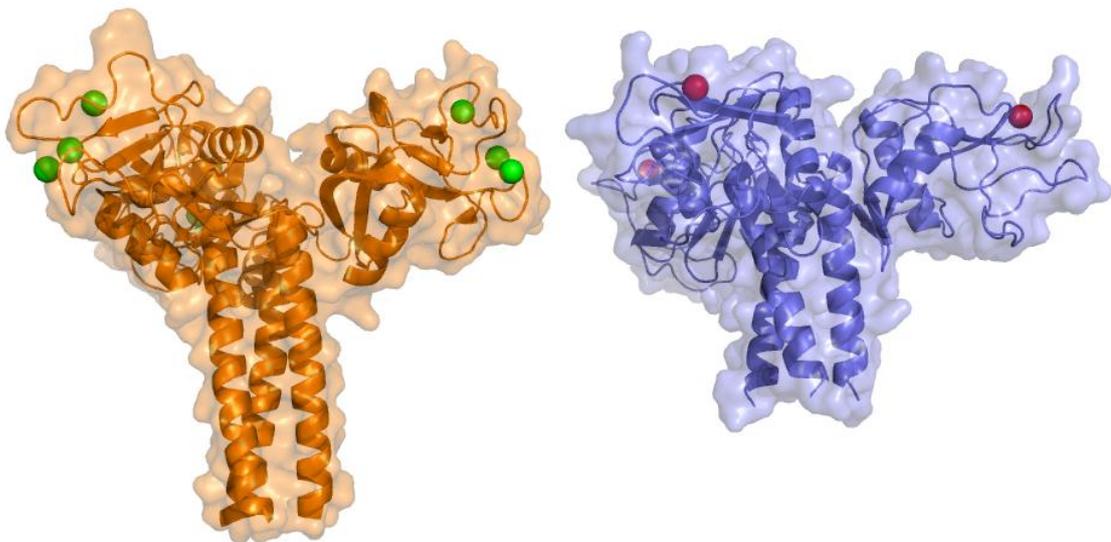


Figure 3.4.1: A comparison between the trimeric CRD and neck domains of MBL (left) and SP-A (right). The most striking difference from comparing MBL against SP-A is the length of the respective α -helical neck domains. In MBL this domain is significantly longer and this in turn may lead to increased stability of the trimeric unit.

Chapter 4 – Collectin 11

4.1 Introduction and Objectives

The aim of this Chapter is to investigate a recently discovered member of the collectin family called collectin 11 (CL-K1). Like MBL and SP-A, CL-K1 has a typical collectin domain organisation comprising an N-terminal domain followed by a collagenous domain with an α -helical neck and a C-type CRD at the C-terminal end. Polypeptides form disulfide-linked oligomers of 100 and 200 kDa indicative of dimers and trimers of the subunits (Selman *et al.*, 2012a). Preliminary studies have shown that CL-11 is present in serum (at $\sim 2 \mu\text{g/ml}$) and binds most avidly to L-fucose and D-mannose, but more weakly than MBL (20 mM compared to $\sim\text{mM}$) (Selman *et al.*, 2012b, Keshi *et al.*, 2006). It also binds to a variety of microorganisms and viruses including *Escherichia coli*, *Candida albicans* and Influenza A virus. Furthermore it possesses a MASP-binding motif within the collagen-like domain and has been shown to bind to MASP-1/-3, so is likely to activate complement (Hansen *et al.*, 2010).

Surprisingly, recent studies have shown that mutations within the *COLLEC11* gene, as well as mutations within the *MASPI* gene, result in a condition called 3MC syndrome (a term that encompasses the overlapping Carnevale, Mingarelli, Malpuech and Michels syndromes). This is the first time that constituents of the complement pathway have been shown to be involved in developmental processes. It also suggests that CL-K1 is likely to recognise both endogenous and exogenous ligands (Rooryck *et al.*, 2011).

The syndromes Carnevale, Mingarelli, Malpuech and Michels are all rare autosomal recessive diseases sharing many of the same features. In 2005 Titomanlio *et al.* proposed that the diseases be grouped together and termed 3MC syndrome (Titomanlio *et al.*, 2005). Following on from this work, Rooryck *et al.* identified a 2.2Mb region mapped to chromosome 2p25 that contained mutations within the *COLLEC11* gene (Rooryck *et al.*, 2011). 3MC syndrome has a negative effect on normal development. The more common features of this syndrome are facial dysmorphism, cleft lip and palate, postnatal growth deficiency, cognitive impairment and hearing loss as summarised by Rooryck (Rooryck *et al.*, 2011).

Within CL-K1, five separate mutations have been identified to cause 3MC syndrome: two substitutions and one deletion mutation within the CRD lead to changes: Ser¹⁶⁹Pro, Gly²⁰⁴Ser and deletion of Ser²¹⁷. Other mutations lead to frame shifts: in the collagenous domain there is a frameshift at the position of Gly¹⁰¹ to Val with a missense sequence until residue 113 and a second frameshift in the 5' leading sequence changing Phe¹⁶ to Ser with a missense sequence until residue 85. These are summarised in figure 4.1.1.1 (Rooryck *et al.*, 2011).

In addition to the mutations in *COLEC11*, it has been shown that mutations within the *MASP1* gene also results in 3MC syndrome for affected individuals. The mutations that have so far been identified are all nucleotide substitutions that are present within exon 12 which results in amino acid substitutions within the serine protease domain of MASP-3. The mutations in the DNA are c.1489C>T, c.1888T>C and 1997.G>A. These result in the following amino acid changes His⁴⁹⁷Tyr, Cys⁶³⁰Arg and Gly⁶⁶⁶Glu (Rooryck *et al.*, 2011).

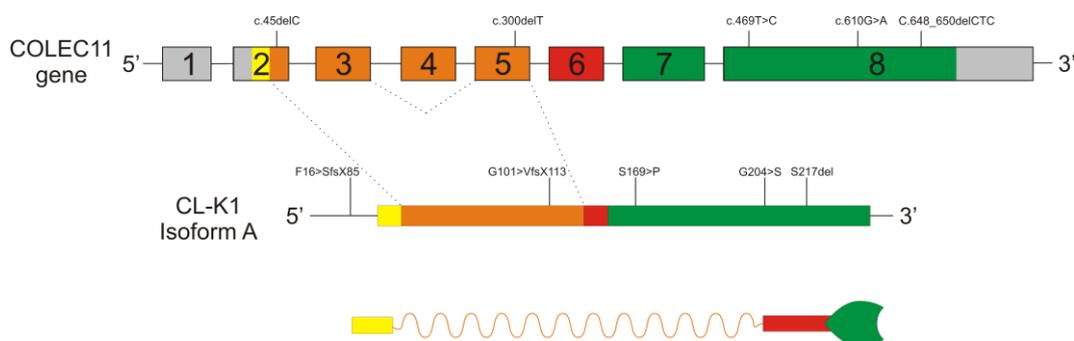


Figure 4.1.1.1. Overview of *COLEC11* mutations. The top row shows the exon organisation of the *COLEC11* gene and the DNA mutations of 3MC syndrome. The mutations in exons 2 and 5 are single nucleotide deletions resulting in missense amino acid sequence. The three mutations in exon 8 are a T>C and G>A substitutions and a CTC deletion, these translate to the two amino acid substitutions and the serine deletion shown underneath. Through alternative splicing exon 4 is omitted from isoform A. The stereotypical collectin ‘schematic’ (as previously described in Chapter 1) is shown at the bottom for comparison.

The two missense mutations in *COLEC11* produce truncated proteins that lack the α -helical neck and CRD and would therefore be unable to trimerise like other collectins. However the deletion and point mutations could potentially yield functional protein and are the target of study in this Chapter.

To examine the binding properties of CL-KI, the strategy in this part of the project was to express and purify the wild-type neck and CRD region of CL-K1 and attempt to express the three mutations associated with 3MC syndrome. Characterisation of these proteins will further our understanding of CL-K1 function and explain how mutations lead to disease.

4.2 - Materials and Methods

4.2.1 – PCR amplification of the neck and CRD regions of CL-K1

The CL-K1 cDNA encodes a polypeptide of 271 amino acid residues. Two cysteine residues at the junction of the collagen-like domain and the neck probably link polypeptides together via disulphide bonds. To facilitate production in *E. coli* these residues were omitted and PCR primers were designed to amplify a region beginning at Ser¹¹⁶. The CL-K1 CRD DNA sequences and corresponding amino acid sequences can be found for the wild-type and 3 mutants in appendix 1. The amino acid sequence for the full length CL-K1 is shown below:

```

1  MRGNLALVGV LISLAFSLLL PSGHPQPAGD DACSVQILVP GLKGDAGEKG DKGAPGRPGR
61  VGPTGEKGDM GDKGQKGSVG RHGKIGPIGS KGEKGDSDGI GPPGPNGEPG LPCECSQLRK
121 AIGEMDNQVS QLTSELKFIK NAVAGVRETE SKIYLLVKEE KRYADAQLSC QGRGGTLSMP
181 KDEAANGLMA AYLAQAGLAR VFIGINDLEK EGAFVYSDHS PMRTFNKWSR GEPNNAYDEE
241 DCVEMVASGG WNDVACHTTM YFMCEFDKEN M

```

Signal sequence, N-terminal domain, collagen-like domain, neck and CRD

Terminal Primer	Restriction Site	Sequence (5' → 3')
CL11crdF	<i>Nco</i> I	ATAG CCATGG CTAGCCAGCTGCGCAAGGCCATCGG
CL11crdR	<i>Eco</i> RI	CGC GAATTC TCA CATGTTCTCCTTGTCAAACTCACAC

Table 4.2.1.1: Terminal Primers of CL-K1 CRD cloning. *Nco*I (containing the start codon ATG) and *Eco*RI sites are highlighted in purple. The stop codon is in red.

Mutant	Primer name and sequence (5' → 3')	
CL-K1 Ser ¹⁶⁹ Pro	S169PF	TACGCGGACGCCAGCTGC <u>CC</u> TTGCCAGGGCCGCGGGGGC
	S169PR	GCCCCGCGGCCCTGGCAAGGCAGCTGGGCGTCCGCGTA
CL-K1 Gly ²⁰⁴ Ser	G204SF	GCCCGTGTCTTCATCT <u>CC</u> ATCAACGACCTGGAGAAG
	G204SR	CTTCTCCAGGTCGTTGATGGAGATGAAGACACGGGC
CL-K1 ΔSer ²¹⁷	SdelF	GAGGGCGCCTTCGTGTAC***GACCACTCCCCATGCGG
	SdelR	CCGCATGGGGGAGTGGTC***GTACACGAAGGCGCCCTC

Table 4.2.1.2: Mutagenic primers for CL-K1 recognition domain. The mutations are underlined within the primer. Stars indicate nucleotides that have been deleted.

4.2.2 – Cloning

The PCR product was digested with the restriction enzymes *EcoRI* and *NcoI* according to NEB protocols and ligated, using a T4 DNA ligase protocol from NEB, into the bacterial expression vector pET28a which had been linearized with the same restriction enzymes.

4.2.3 – Competent Cell Production

A single colony of BL21 cells was used to inoculate 5 ml of LB in a 50 ml falcon tube and grown overnight. The following morning 25 ml of fresh LB was inoculated with the overnight culture to an OD₆₀₀ of 0.05. The cells were incubated and grown to an OD₆₀₀ of 0.6-0.8 and then decanted and placed on ice for 10 minutes to cool before being centrifuged at 3500 rpm, 4 °C for 10 minutes. The supernatant was then discarded and the pellet resuspended in 2.5ml cold 0.1 M CaCl₂. The cells were incubated on ice for 20 minutes and centrifuged under the same conditions. The pellet was gently resuspended in 5 ml cold 0.1 M CaCl₂, 15% glycerol and separated into 100 µl aliquots. The aliquots were frozen on dry-ice and stored at -80°C.

4.2.4 – Expression

The purified plasmid DNA was transformed into BL21 cells using the protocol described in Section 2.2.2. The cells were plated onto agar plates with kanamycin added to a concentration of 50 µg/ml and incubated at 37 °C overnight. The cells were suspended by adding 4 ml LB to each plate and agitating them with a sterile plastic spreader. The OD₆₀₀ measured and the cells diluted to a final OD₆₀₀ of 0.1 in 1 L of PowerPrime broth (AthenaES) containing 50 µg/ml kanamycin.

The cells were grown in baffled flasks in a shaking incubator at 37 °C until an OD₆₀₀ of 1.5-2 was achieved. IPTG was then added to a concentration of 1mM and the cells grown for 16 hours with shaking.

4.2.5 – Inclusion Body Preparation

The 1 L of overnight was centrifuged at 4000 g, 4 °C for 20 minutes. The supernatant was decanted and the cells resuspended in 40 ml of Bugbuster Protein extraction reagent (Merck) containing a protease inhibitor cocktail tablet (Roche). The suspension was mixed for 15 minutes at 4°C. The cells were lysed by sonication on ice with the large probe attached at amplitude 8 for 20 seconds and allowed to rest for 60 seconds. This was repeated 10 times to ensure complete lysis of cells.

The lysate was centrifuged and the pellet resuspended in 40 ml 1:10 dilution of Bugbuster. This step was repeated. The protein pellet was then washed in 40 ml 2 M urea, containing 0.5 M NaCl, 1 mM EDTA and 50 mM Tris-HCl pH 8.0 and centrifuged at 20,000 g and at 4 °C for 20 minutes. The inclusion bodies were washed again using 40 ml 50 mM Tris-HCl pH 8.0, containing 0.5 M NaCl, 1 mM EDTA and 0.25% sodium deoxycholate and centrifuged as before. After a final wash was with 40 ml ddH₂O the pellet was dissolved in 50 ml 8 M urea, 50 mM Tris-HCl pH 8.0, centrifuged to remove any insoluble debris and the supernatant collected. An OD₂₈₀ of 2-5 mg/ml was obtained and freshly made DTT added to a concentration of 5 mM to break any disulphide bonds.

4.2.6 – Refolding Protocol

50 ml of the solubilised protein was slowly dripped into 250 ml of 25 mM Tris-HCl pH 7.8, containing 1.25 M NaCl, and 25 mM CaCl₂ with gentle stirring at 4 °C. The resulting mixture was then dialysed extensively in 50 mM Tris-HCl pH7.5, containing 150mM NaCl and 5 mM CaCl₂ using 10,000 Da cut-off dialysis tubing

4.2.7 – Ion Exchange Chromatography

Refolded protein was initially purified by ion exchange chromatography (used primarily as a concentration step) and finally by gel filtration chromatography. A 20 ml column of Q Sepharose Fast Flow from GE Healthcare was equilibrated with filtered 10 mM Tris-HCl pH 7.5, containing 50 mM NaCl and 2 mM CaCl₂. Prior to loading, the dialysed protein was centrifuged at 3500 rpm for 20 minutes at 4 °C to remove any precipitate. The supernatant was then passed over the Q Sepharose column using a peristaltic pump. After loading, the column was attached to the AKTA purifier and washed with 10 mM Tris-HCl pH7.5 containing 50 mM NaCl, and 2 mM CaCl₂ to remove any unbound protein. Protein was eluted using an increasing NaCl gradient up to 1 M NaCl at a flow rate of 1 ml/min over 20 minutes and the protein fractions collected. At this stage all fractions containing protein were pooled giving a final volume of ~20 ml.

4.2.8 – Gel Filtration

The protein fractions from the ion exchange column were pooled and concentrated to 5 ml using a 10 kDa molecular weight cut off protein concentrator (Millipore). A Superdex 200 16/600 or Superdex 75 16/600 HiLoad column from GE Healthcare was

equilibrated with filtered 50 mM Tris-HCl pH7.5, containing, 150 mM NaCl, and 2 mM CaCl₂. Once equilibrated the 5 ml sample was injected into the loading loop of the AKTA purifier. The protein was run over the column at a flow rate of 1 ml/minute according to the manufacturers' specifications and the fractions collected. The fractions were run using SDS-PAGE to determine where the protein had eluted from the column.

4.2.9 – Crystal Condition Screening

To produce protein crystals, the protein solution is concentrated so that it can be considered 'supersaturated'. A precipitant is used that causes the protein to nucleate and grow. This is detailed in the figure below:

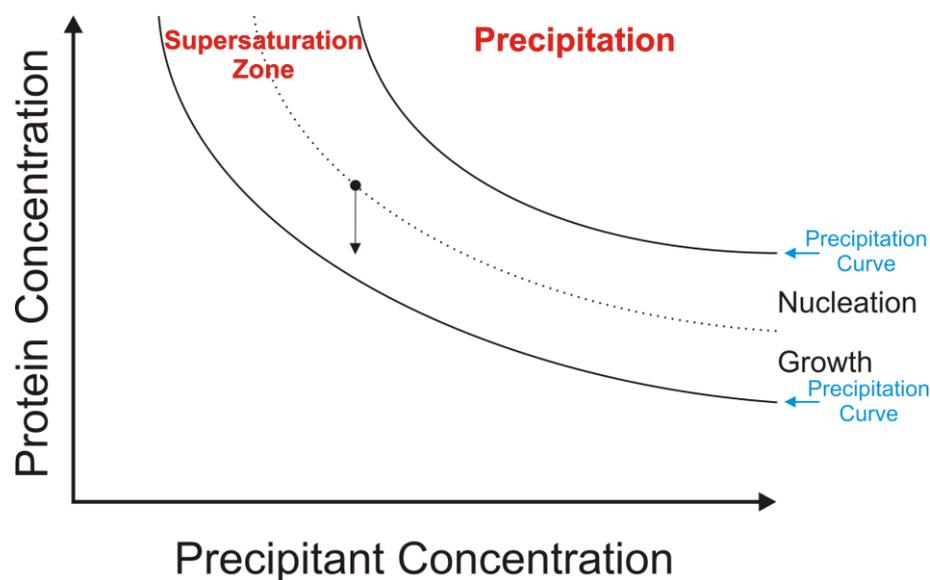


Figure 4.2.9.1: Protein crystallisation. The black dot represents an ideal condition for crystallization. The concentration of protein and precipitant must be high enough to allow nucleation. As the crystal grows the concentration of protein will decrease but growth will continue provided it is above the lower precipitation curve. Higher initial concentrations will lead to excessive nucleation or aggregation whereas a lower starting concentration will yield no crystals.

The prior diagram is previously described in (Asherie, 2004). As the protein concentration or precipitant concentration increases, the protein will move towards the supersaturation and precipitations zones. When the conditions are optimal (represented by black dot) a single nucleation centre will form around which the crystal will grow as more protein attaches on the surface; and will continue to grow at concentrations above the lower precipitation limit. At higher starting concentrations more crystals will grow, but they will be smaller as each 'competes' for the protein in solution. At concentrations below the nucleation curve, nucleation will not occur and no crystals will form.

The protein was concentrated to approximately 3 mg/ml in 20 mM Tris/HCl containing 50 mM NaCl and 2.5 mM CaCl₂ and a series of crystallization conditions were explored using readymade screens obtained from Emerald Biosystems (Wizard™ I, II, III and IV crystal screens). For each screen, crystallization experiments were set up by mixing 0.1 µL of protein and 0.1 µL buffer in the wells of an MRC 96-well sitting drop crystallization plate using an in-house crystallization robot. In the sitting drop method, the mixture of protein and crystallization buffer is incubated with a reservoir of crystallization buffer within a closed chamber. Over time the concentrations of protein and buffer in the drop increase and reach equilibrium with the reservoir buffer, such that the protein concentration approaches its initial value (in this case ~3 mg/ml).

Within minutes of setting up the initial crystallization screens, crystals were seen in a number of different conditions. The most promising condition (with regard to the size and number of crystals observed) was 100 mM HEPES, pH 7.0 containing 15 % ethanol. To generate larger crystals, drops (1 µL protein and 1 µL crystallisation

buffer) were set up again using the sitting drop method in Cloverleaf 96 well crystallization plates (Emerald Biosystems). The optimised condition for crystallization was 10 % ethanol, 100 mM HEPES pH 7.0 and 4 mM CaCl₂.

4.2.10 – X-Ray Diffraction

X-ray diffraction was to determine the structure of the neck and head regions of CL-K1. This part of the project was carried out by Alexander Gingras and Russell Wallis, so only brief details will be provided here. Crystals were transferred to reservoir solution containing 20 % glycerol before cryoprotection in liquid nitrogen, and were maintained at 100 K during data collection. Diffraction data were collected at Diamond Light Source and were processed with XDS (Kabsch, 1993). Phases were determined by molecular replacement with Phaser and models were optimized using cycles of manual refinement with Coot and maximum likelihood refinement in Refmac5, part of the CCP4 software suite (Collaborative, 1994).

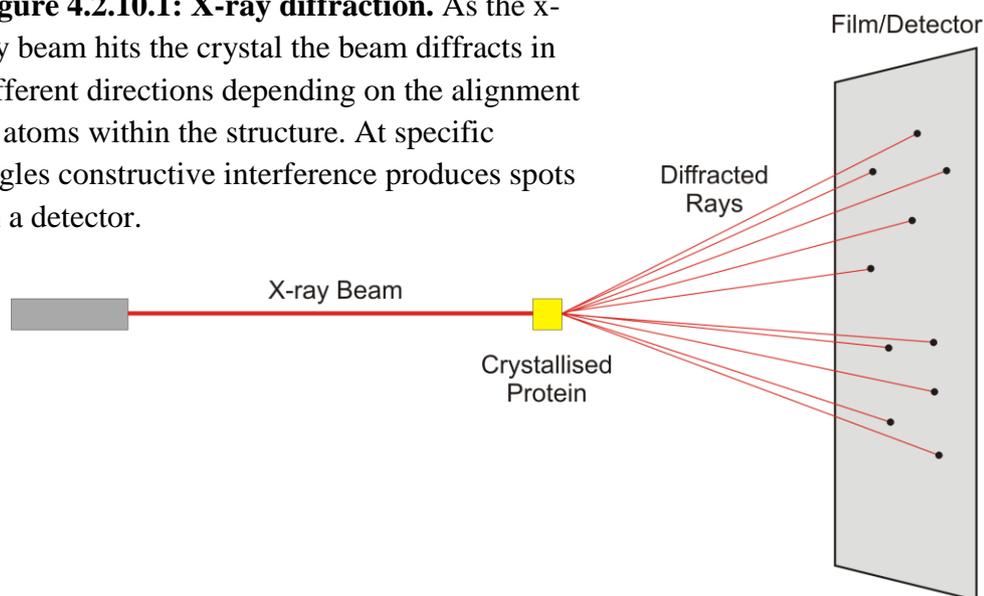
A protein crystal comprises ordered repeating units in the same orientation. When x-rays are fired at the crystal they diffract (Figure 5.1.1.2) and at specific angles the x-ray waves constructively interfere and will produce spots on the film or detector as governed by Braggs Law:

$$n\lambda = 2d\sin\theta$$

Where n is an integer, λ is the wavelength of the incident wave (X rays), d is the spacing between the planes in the atomic lattice, and θ is the angle between the incident ray and the scattering planes.

Rotating the crystal produces different diffraction patterns and using a range of rotations, the electron density of the atoms within the crystal can be deduced.

Figure 4.2.10.1: X-ray diffraction. As the x-ray beam hits the crystal the beam diffracts in different directions depending on the alignment of atoms within the structure. At specific angles constructive interference produces spots on a detector.



4.2.11 – Trypsin Digests

Limited digestion with trypsin was used to probe the stability of the neck and CRDs of CL-K1. The principle behind this experiment is that folded proteins are relatively resistant to proteolysis, whereas unfolded proteins are sensitive. Consequently, a protein containing a destabilising mutation is likely to be more sensitive to proteases than the wild type protein. Briefly, purified CL-K1 in 50 mM Tris-HCl, containing 150 mM NaCl and 2 mM CaCl₂ (10 µL of either wild type or mutant at 0.2 mg/ml) was incubated with serial dilutions of TCPK-treated trypsin (stock of 1 mg/ml prepared in water) for 1 hour at 37 °C. Samples were then separated by SDS-polyacrylamide gel electrophoresis.

4.2.12 – Surface Plasmon Resonance

Surface plasmon resonance was carried out essentially as described in section 2.2.9.

Wild-type and Ser¹⁶⁹Pro mutant CL-K1 proteins (150 µg/ml) were diluted in 10 mM sodium acetate pH 4.5 and immobilized on the surface of a GLM chip. Carbohydrate ligand was flowed over the chip to compare binding.

4.3 Results

PCR was used to amplify the portion of CL-K1 encoding the neck and CRDs and separate reactions were used to introduce the three mutations associated with 3MC syndrome (Ser¹⁶⁹Pro, Gly²⁰⁴Ser and Δ Ser²¹⁷). The resulting products were cloned into the expression vector pET28a. Clones were sequence verified to ensure that no errors had been introduced and the mutations were present.

4.3.1 – Purification of wild-type CL-K1 head and neck regions

The proteins were expressed in BL21 cells and the insoluble inclusion bodies washed and re-dissolved. The protein was refolded by drop dilution. Initial experiments showed that the CL-K1 fragment did not bind to mannose-Sepharose (or fucose-Sepharose), so instead an alternative purification strategy was used. Initially ion exchange chromatography was used to concentrate the sample. To separate folded protein from aggregates, CL-K1 was concentrated to 5 ml and loaded onto the Superdex 200 HiLoad 16/600 column and purified by gel filtration. A typical elution profile for wild-type protein is shown in figure 4.3.1.1. Molecular weight standards for the Superdex 200 16/600 column can be seen in figure 4.3.1.2.

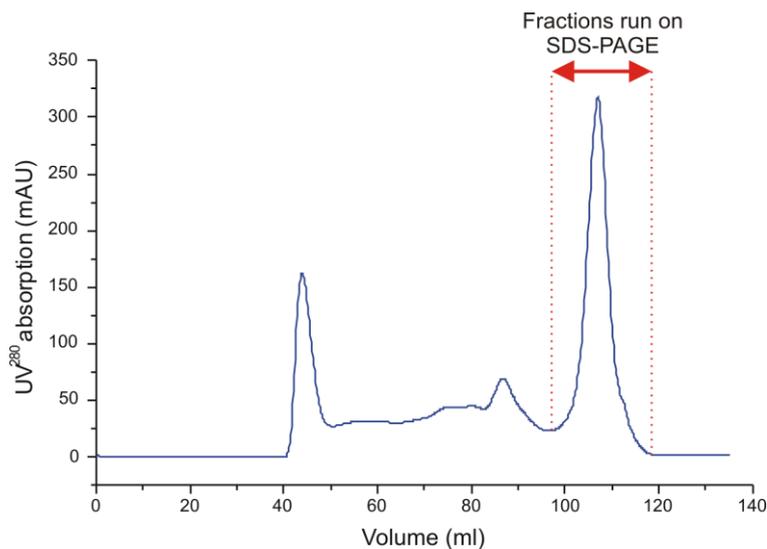


Figure 4.3.1.1.
Purification of CL-K1
on a Superdex 200
16/600 HiLoad
Column. Fractions were collected and those between the red lines were run on polyacrylamide gels to confirm the presence and purity of the sample.

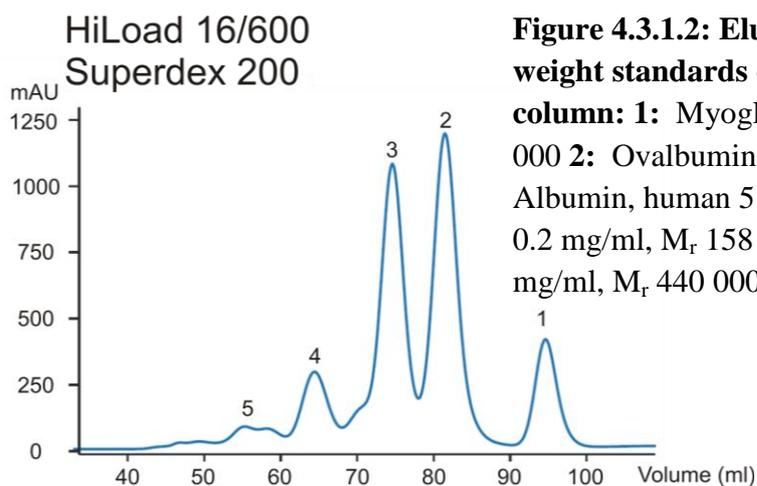


Figure 4.3.1.2: Elution profile of molecular weight standards on a Superdex 200

column: 1: Myoglobin, 1.5 mg/ml, M_r 17 000 **2:** Ovalbumin, 5 mg/ml, M_r 44 000 **3:** Albumin, human 5 mg/ml, M_r 66 000 **4:** IgG, 0.2 mg/ml, M_r 158 000 **5:** Ferritin, 0.24 mg/ml, M_r 440 000.

Protein eluted from the column in several peaks and these were checked by SDS-polyacrylamide gel electrophoresis under both reducing and non-reducing conditions. Folded CRDs contain 2 intrachain disulphide bonds but individual polypeptide are not linked by disulphide bonds. The early peaks that eluted from the column either comprised aggregates of CL-K1 (with interchain disulphide bonds) or contaminating proteins. The CL-K1 peak corresponding to folded protein eluted at 107 ml. Although,

the peak was symmetrical, compatible with a folded protein, it came off the column much later than expected from its size (17.7 kDa per polypeptide; 53 kDa for a trimer) with an estimated molecular mass of only <10 kDa. SDS-PAGE analysis showed that this was not due to proteolysis, but is more likely caused by CL-K1 binding weakly to the column. MBL also binds to Superdex columns (which comprise branched dextran chains) in the presence of Ca^{2+} .

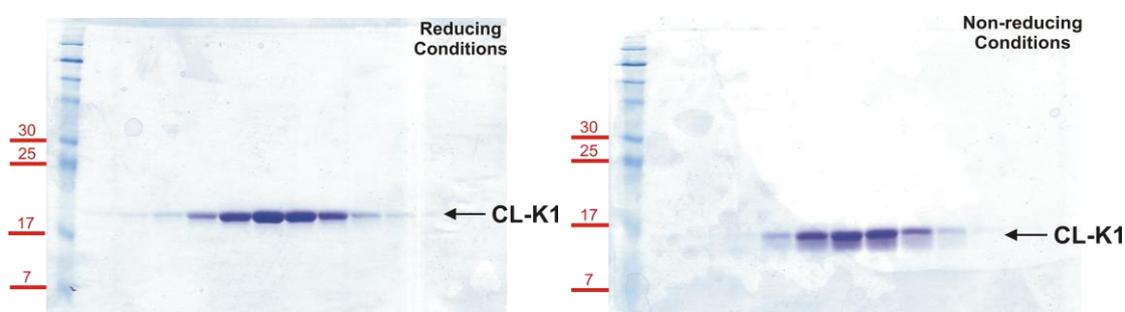


Figure 4.3.1.2. SDS-PAGE of CL-K1. The left gel shows the fractions run under reducing conditions which confirm the presence of CL-K1. The gel on the right is run in non-reducing conditions. In the absence of DTT, disulphide bonds remain unbroken and the polypeptide chain is more compact and migrates faster giving an apparent size slightly smaller than 17 kDa. Importantly, chains are not disulphide bonded together as is often seen in soluble aggregates. Samples were loaded in order of collection from left to right.

4.3.2 - Binding of CL-K1 to HIV gp120

To confirm that CL-K1 was folded, binding was measured to gp120, an envelope glycoprotein of the human immunodeficiency virus which contains high-mannose oligosaccharides. Previous studies using CL-K1 produced in CHO cells showed that this was a good ligand (Girija and Wallis, unpublished observations). When gp120 was

flowed over the chip in the presence of Ca^{2+} (Figure 4.3.2.1) there was a concentration dependent increase in the signal (response units), indicating that CL-K1 binds to gp120. No binding was detected in EDTA showing that binding was Ca^{2+} -dependent and implying that the CL-K1 binds to gp120 via its high-mannose sugars.

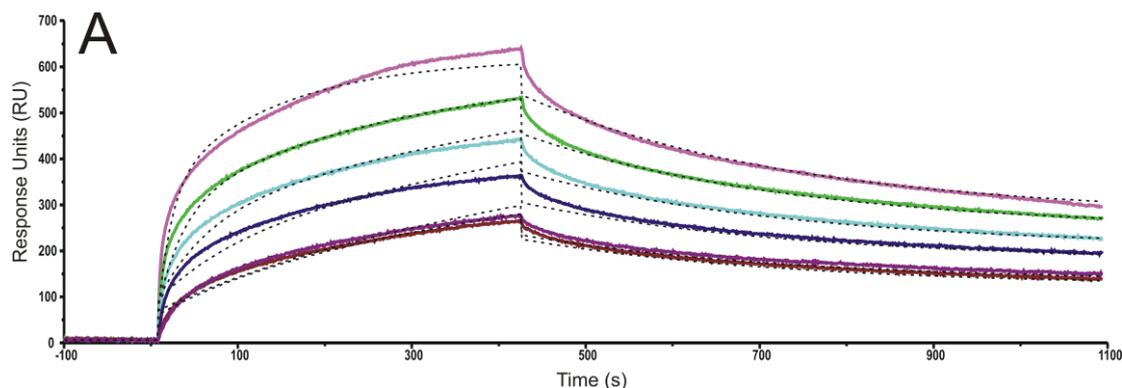


Figure 4.3.2.1: Surface plasmon resonance of the head and neck regions of CL-K1 binding to HIV gp120. CL-K1 was immobilized on the chip by amine coupling. Decreasing concentrations of soluble gp120: 200 nM (*pink*), 100 nM (*green*), 50 nM (*blue*), 25 nM (*purple*), 12.5 nM (*mauve*) and 6.25 nM (*maroon*) were flowed over the immobilized lectin.

The binding kinetics were analysed by fitting data simultaneously to increasing complex binding models until a satisfactory fit was achieved. The data fitted poorly to a 1:1 model, but a satisfactory fit was achieved using a two-complex, parallel-reaction model. Association rate constants were 3.05×10^5 and $3.77 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ and dissociation constants were 3.22×10^{-3} and $3.28 \times 10^{-6} \text{ s}^{-1}$ to yield K_D values of 10.55 nM and 0.087 nM. Thus, individual CRD heads of CL-K1 bind tightly to the sugars on gp120 with nM affinity. Although, complex kinetics such as seen here can result from many different reasons, in this case it probably reflects the nature of the target protein (gp120) which possesses multiple high-mannose oligosaccharides each of which can potentially function as a ligand.

4.3.3 - Analysis of mutant CL-K1s associated with 3MC syndrome.

To investigate the molecular basis of 3MC syndrome caused by mutations in CL-K1, equivalent constructs were expressed containing each of the mutations associated with disease. All three variants were expressed as inclusion bodies and the refolding and ion exchange steps were carried out as for the wild-type protein.

The Ser¹⁶⁹Pro variant showed a different elution profile than the wild-type by gel filtration. Two main peaks were observed, a larger peak which eluted in the excluded volume, probably comprising aggregated material and a slightly smaller peak, which eluted at ~65 ml. The calculated molecular mass, based on the elution positions of protein molecular mass standards was ~44 kDa, corresponding closely to the expected size of a trimer of polypeptides (53 kDa).

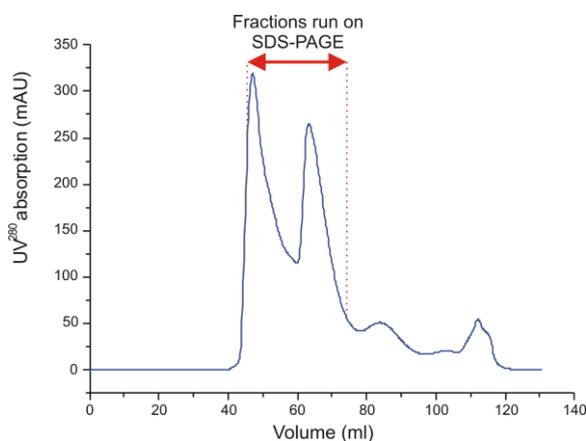


Figure 4.3.3.1: Purification of CL-K1S>P on a Superdex 75 16/600 HiLoad Column.

Fraction numbers 21 to 38 (spanning the red lines) were collected across the two peaks were run on polyacrylamide gels confirming that CL-K1 protein is present.

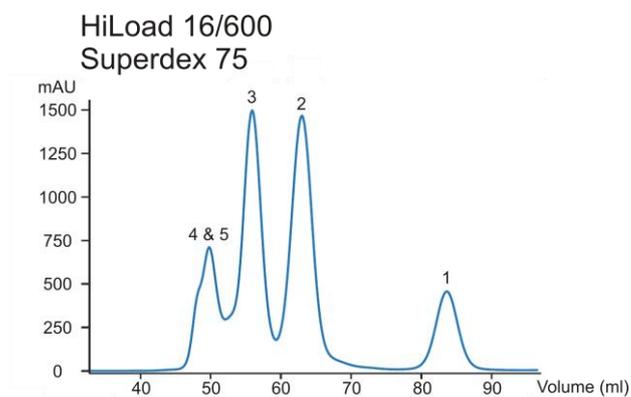


Figure 4.3.3.2: Elution profile of molecular weight standards on a Superdex 75 column: 1: Myoglobin, 1.5 mg/ml, M_r 17 000 2: Ovalbumin, 5 mg/ml, M_r 44 000 3: Albumin, human 5 mg/ml, M_r 66 000 4: IgG, 0.2 mg/ml, M_r 158 000 5: Ferritin, 0.24 mg/ml, M_r 440 000.

Analysis of fraction on polyacrylamide gels confirms that the second peak comprises pure CL-K1 with no evidence of disulphide linked aggregates. By contrast the first peak contains multiple high molecular weight impurities. Based on its high apparent molecular weight protein in this peak was not analysed further.

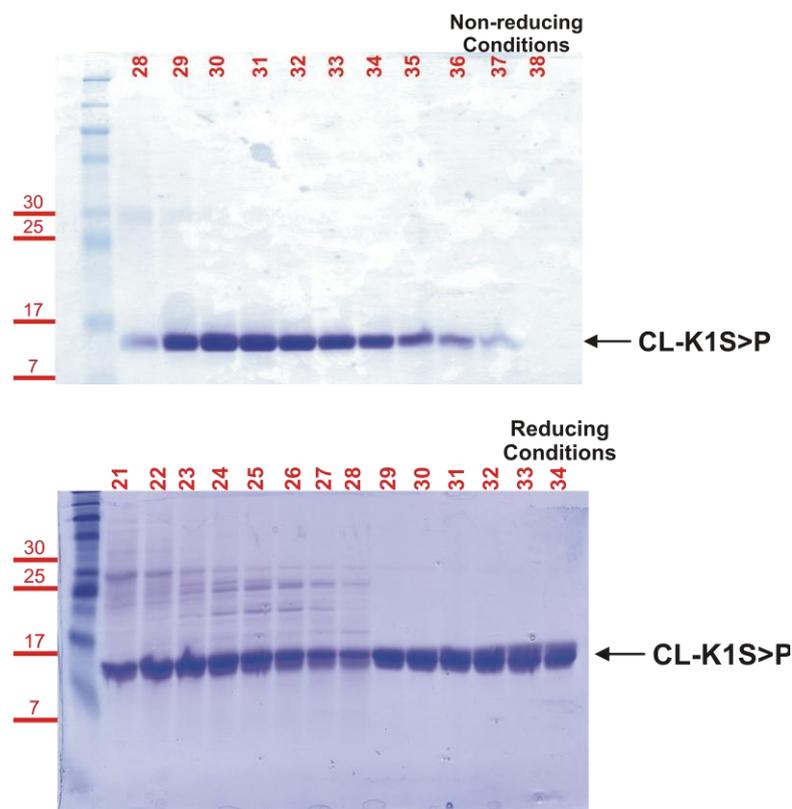


Figure 4.3.3.3. SDS-PAGE of gel filtration fractions of CL-K1 Ser¹⁶⁹Pro. Top: Under non-reducing conditions a single band is observed across the second gel filtration peak, with no evidence of disulphide linked aggregates. **Bottom:** The reducing gel shows high molecular weight impurities present in the fractions corresponding to the first gel filtration peak (fractions 21-28) but no impurities in the second peak (fractions 29-34).

Taken together the results can be explained if the Ser¹⁶⁹Pro variant folds correctly but no longer interacts with the dextran matrix of the Superdex column. To examine this possibility, fractions corresponding to the second gel filtration peak were pooled and tested for binding to gp120 by surface plasmon resonance, as before.

As shown in figure 4.3.3.4, the Ser¹⁶⁹Pro variant bound to gp120, but binding was less than for wild-type CL-K1. At comparable concentrations of gp120, the magnitude of

the signal was much lower despite equivalent amounts of CL-K1 having been immobilized. Furthermore, the affinity of binding was much lower than for wild type CL-K1. Data fitted well to a 1:1 complex model in which association and dissociation rate constants were $1.06 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ and $4.7 \times 10^{-3} \text{ s}^{-1}$ to give a K_D of 392 nM. Thus, gp120 binding by the Ser¹⁶⁹Pro mutant is >30-fold weaker than for wild-type CL-K1. Defective binding by the Ser¹⁶⁹Pro variant may explain, at least in part, the disease phenotype in 3MC syndrome. Thus, although the ligand for CL-K1 during development is not known, loss of binding might prevent normal developmental processes leading to the characteristic phenotype.

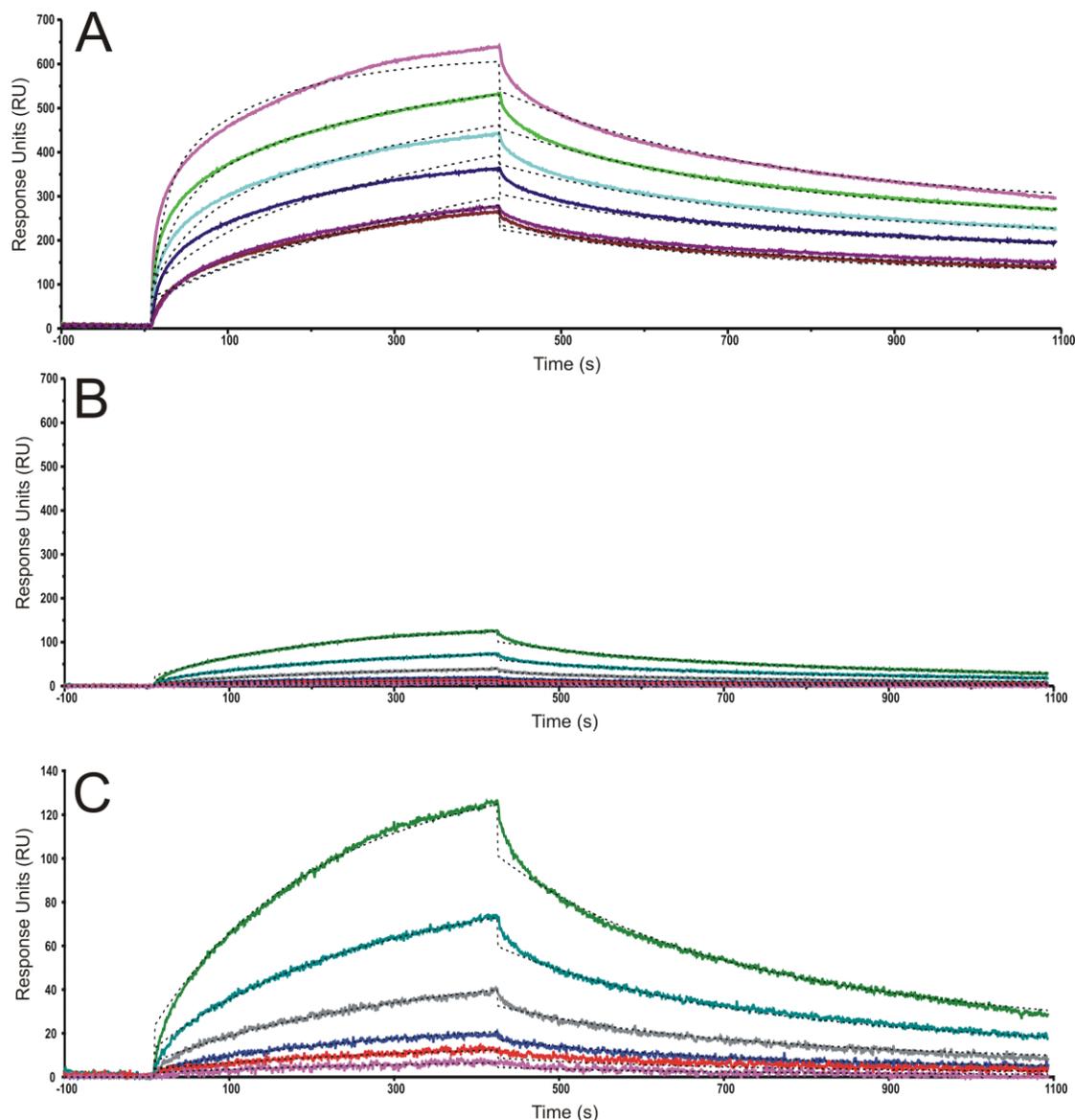


Figure 4.3.3.4: Surface Plasmon Resonance of CL-K1 binding to gp120. **A:** CL-K1 wild-type protein previously described in figure 4.3.2.1 **B:** Ser¹⁶⁹Pro mutant. In each case ~11000 RU were immobilized on the chip surface and gp120 was flowed over at concentrations of 200 nM (*green*), 100 nM (*sky blue*), 50 nM (*grey*), 25 nM (*mauve*), 12.5 nM (*red*) and 6.25 nM (*pink*). It can be seen that the binding response is considerably smaller than for wild-type. **C:** The Response Units axis has been adjusted to provide better resolution of the curves. Dotted lines show the fits to the data.

4.3.4 – Wild-type and Ser¹⁶⁹Pro CRD Trypsin Digestion

The serine to proline mutation may destabilise the CRD because of the limited flexibility of proline and/or its reduced capacity to form hydrogen bonds compared to

other residues. It was therefore important to examine the structural stability of the CRD. Therefore, wild type and Ser¹⁶⁹Pro mutant were treated with serial dilutions of trypsin to determine their relative sensitivities. Samples were run on an SDS polyacrylamide gel (figure 4.3.4.1).

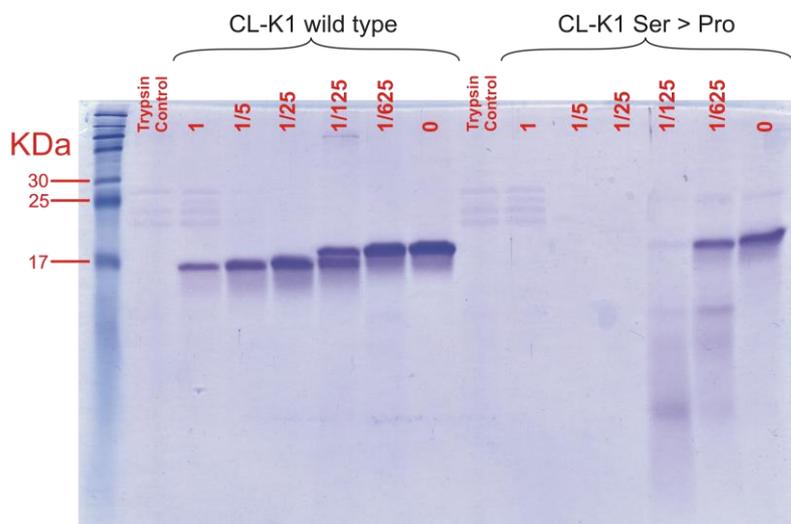


Figure 4.3.4.1: A reducing SDS-PAGE gel of the CL-K1 proteins treated with trypsin. The wild-type protein is resistant to trypsin showing minimal susceptibility. The ser > pro CL-K1 mutant is completely digested at the 1/25 dilution of the 1mg/ml trypsin stock and shows cleavage at the 1/625 dilution.

There is a marked difference between the effect of trypsin on the wild-type protein and the serine to proline mutant. In the wild-type sample, much of the protein still remains even at the highest concentration of trypsin (0.1 mg/ml), although it has been cleaved to a slightly smaller product than undigested control. The 1/125 dilution of trypsin provides an excellent mid-point between the cleaved and undigested forms. It is likely that the difference in size is caused by the digestion of the α -helical neck leaving the CRD intact.

The stability of the serine to proline mutant is much less than that of the wild-type. In the 1 mg/ml, 1/5 and 1/25 dilutions, the protein has been completely digested. Some bands are visible in the 1/125 and 1/625 dilutions, but these are smeared so are probably

relatively unstable. It is likely that the mutation reduces the stability of the CRD, exposing sites for proteolysis by trypsin.

4.3.5 – Purification of additional CL-K1 CRD mutants

Neither the Gly²⁰⁴ Ser variant nor the Ser²¹⁷ deletion mutant yielded enough folded protein for additional analysis. The Gly²⁰⁴ Ser eluted as a broad peak from the gel filtration column. No peak was observed at 107 ml (as for wild-type protein) or at 65 ml (for the Ser¹⁶⁹Pro mutant).

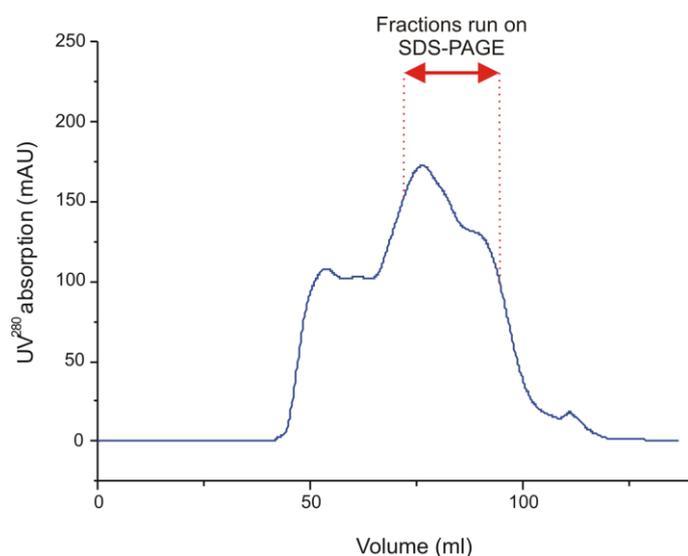


Figure 4.3.5.1: Purification of CL-K1 G²⁰⁴S on a Superdex 200 16/600 HiLoad Column. Fractions were collected and those between the red lines were run on polyacrylamide gels. The elution profile of the molecular weight standards can be seen in figure 4.3.1.2.

It is possible that polypeptides are in monomer-trimer equilibrium. In this case protein would be expected to elute between 90 ml (17.7 kDa) and ~75 ml (53.1 kDa). If association of the trimer is transient the protein could be spread across this range. To investigate this possibility, fractions were collected across the expected range and run on polyacrylamide gels under both reducing and non-reducing conditions (figure 4.3.5.2).

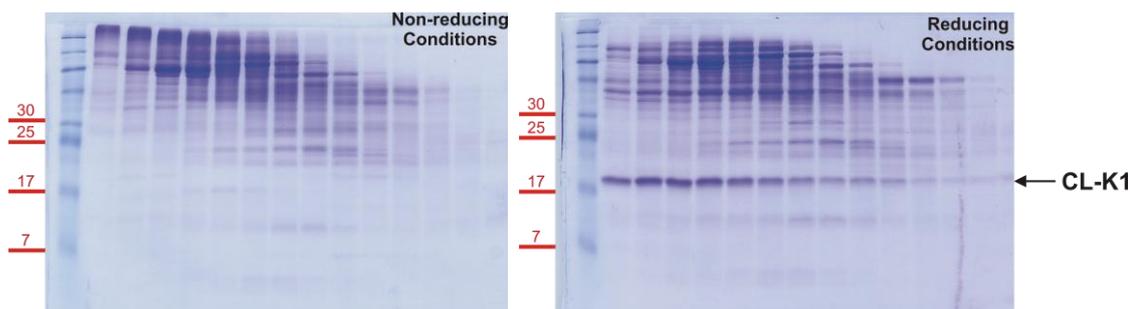


Figure 4.3.5.2: SDS-PAGE of CL-K1G^{204S}: Identical samples were run on each gel which are loaded from left to right. The polyacrylamide gel on the left was run under non-reducing conditions, there is very little protein present at around 17 kDa. In contrast, under reducing conditions the CL-K1 band is present at 17.7 kDa. Thus although protein is present, it forms disulphide-linked aggregates during refolding.

Under non-reducing conditions there is little or no protein present as monomers. This indicates that during refolding, the CL-K1 does not assemble correctly, but instead forms interchain disulphide bonds. Given that refolding did not occur *in vitro*, the G^{204S} mutant might be destabilized and not fold correctly *in vivo* leading to a loss of CL-K1 function. However this possibility needs to be tested by additional cell culture/*in vivo* studies.

The final mutant to be expressed was the serine deletion. After refolding, protein was run on the Superdex 200 as before. Although most protein eluted as aggregates in the void (~45 ml). Fractions were collected between 70 ml and 95 ml, which covers the expected elution position of trimeric and monomeric polypeptides, and were analysed on SDS gels.

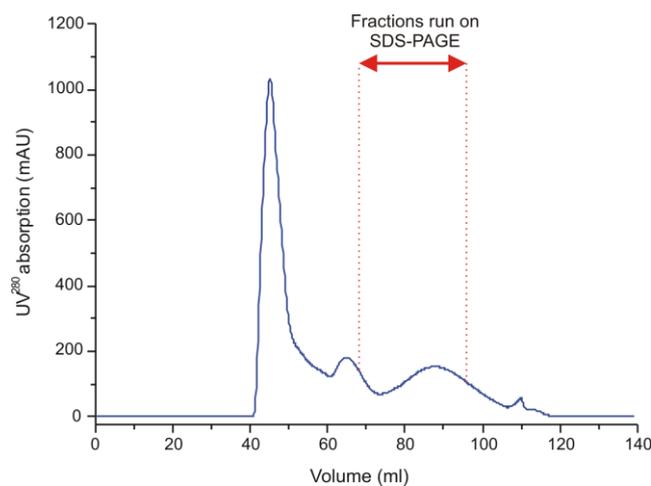


Figure 4.3.5.3: Purification of CL-K1 ΔS^{217} on a Superdex 200 16/600 HiLoad Column. The protein was collected and analysed on polyacrylamide gels. Most of the protein formed high molecular weight aggregates (eluting at ~ 45 ml) suggesting misfolding.

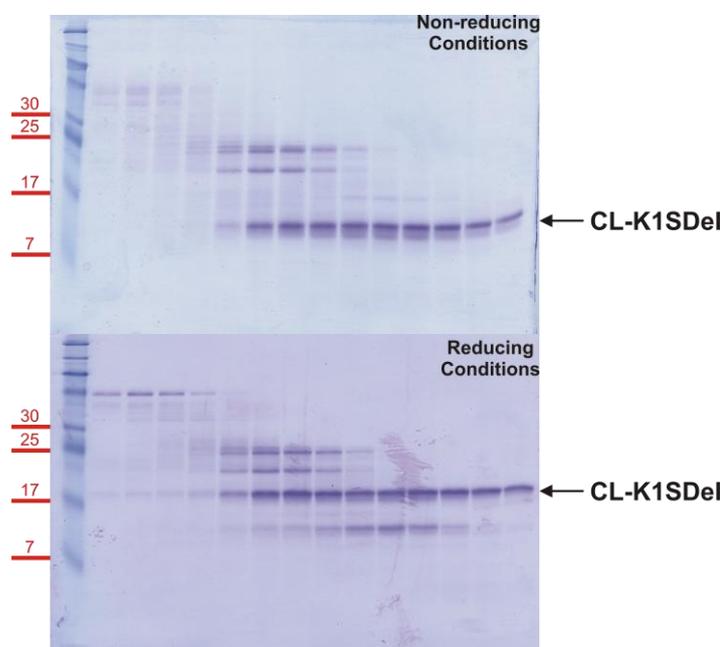


Figure 4.3.5.4: Non-reducing and reducing SDS-PAGE of CL-K1 ΔS^{217} gel filtration fractions. Samples were loaded from left to right in order of fraction collection between the two dotted red lines of figure 4.3.5.3.

As can be seen from figure 4.3.5.4, monomeric CL-K1 ΔS^{217} was present under non-reducing conditions, so some protein may have folded correctly. However, additional bands revealed the presence of multiple contaminants that eluted at the same place as the CL-K1 mutant. Consequently this mutant was not characterised further.

4.3.6 – Crystallisation and X-ray diffraction

The CL-K1 wild-type CRD was crystallised at 3 mg/ml concentration in 10% ethanol, 100 mM HEPES pH 7.0 and 4 mM CaCl₂. Although attempts were made, CL-K1 S¹⁶⁹P did not form crystals under similar conditions. For wild-type CL-K1, diffraction maps were collected at Diamond Light Source and the structure solved by Alexander Gingras and Russell Wallis by molecular replacement using the corresponding region of SP-D as a search model, as described in section 4.2.10. The structure of CL-K1 is discussed in more detail in the next Chapter.

4.4 Discussion

Relatively little is known about CL-K1. In this Chapter, I successfully produced a recombinant version of CL-K1 encompassing the neck and CRDs. Unlike the corresponding fragment of MBL, the neck and CRDs of CL-K1 did not bind significantly to mannose-Sepharose or fucose-Sepharose columns, however it bound to gp120 with high affinity. The differences between CL-K1 and MBL probably reflect differences in the way the proteins bind to ligands. MBL recognises sugars via binding of its CRDs to terminal mannose moieties. CL-K1 on the other hand does not bind to mannose-Sepharose, and recognises monosaccharides only weakly (20 mM compared to 1 mM) (Hansen *et al.*, 2010). One possibility is that it preferentially binds to larger carbohydrates (such as the high mannose-structure on gp120). This would be compatible with targeting both endogenous and exogenous sugars.

4.4.1 - Mutant CL-K1s associated with 3MC syndrome

The most likely deleterious effect of any mutation on protein structure is the disruption of normal folding. Should misfolding occur the proteins are liable to be degraded within the cell and not secreted. In this Chapter I have attempted to produce CL-K1 containing the mutations associated with 3MC syndrome with varying success.

I successfully expressed and purified Ser¹⁶⁹Pro protein. The binding studies using gp120 showed that although Ser¹⁶⁹Pro retains some binding, it is reduced by at least 30-fold compared to wild-type protein, thus requiring much higher concentrations of gp120 to achieve the same binding response. This indicates a lack of functionality in the

protein resulting from the mutation. In addition, the CRD of CL-K1 S¹⁶⁹P was more susceptible to protease cleavage than its wild-type counterpart. This suggests that while it is able to refold in a manner similar to the wild-type, the substitution is destabilising. Should the protein be exported from the cell *in vivo* it is likely to be rapidly degraded by serum proteases. Furthermore, it will not recognise its physiological targets. Both reduced stability and loss of carbohydrate binding may contribute to loss of function *in vivo*.

The Gly²⁰⁴ to Ser and Ser²¹⁷ deletion mutants did not refold as well as wild type or S¹⁶⁹P mutant. Yields were significantly lower in each case. Destabilisation during biosynthesis and/or following secretion may contribute to the disease phenotype in 3MC syndrome but this awaits further analysis.

Chapter 5 – General Discussion

The work undertaken in this thesis has led to an increased understanding of the mechanism of complement activation by MBL/MASP-2 complexes. I have shown that introduction of the MASP binding motif is sufficient to introduce MASP-2 activation into a protein that possesses the collectin bouquet-like architecture. However, the neck and CRDs of MBL somehow ensure that activation occurs on a target. Mutagenesis of the carbohydrate binding site within the CRD to alter sugar specificity does not affect this control and crystal structures do not show conformational changes when the CRD binds to sugar ligands. This suggests that the control of activation originates in the physical properties of the neck. Whether this stems from the length, the angle between the α -helical neck and collagenous domain or the relative hydrophobicity between the α -helical neck chains is uncertain at present and the underlying mechanism that controls target specific activation remains to be elucidated.

5.1 – Interactions between MBL-MASPs

MASP zymogens bind to a conserved hexapeptide motif (Hyp-Gly-Lys-X-Gly-Pro) present within the collagenous domain of MBLs and ficolins. From the work contained within Chapter 2 we can see that lysine and leucine alone are sufficient to establish an interaction between MASP-2 and the collagenous domain of MBL. This agrees with previous studies that have shown the lysine residue to be a key player in binding in rat and human MBLs. It also is compatible with the recent structure of a CUB domain from MASP-1 in complex with a collagen-like peptide from MBL. In the structure, the lysine residue of the collagen forms hydrogen bonds with Ca^{2+} -coordinating residues in

the CUB domain both explaining the importance of the lysine residue and the Ca^{2+} -dependence of the interaction (Gingras *et al.*, 2011). Very few additional interactions are required to stabilize the complex explaining why so few amino acid substitutions are required to introduce MASP-binding into SP-A.

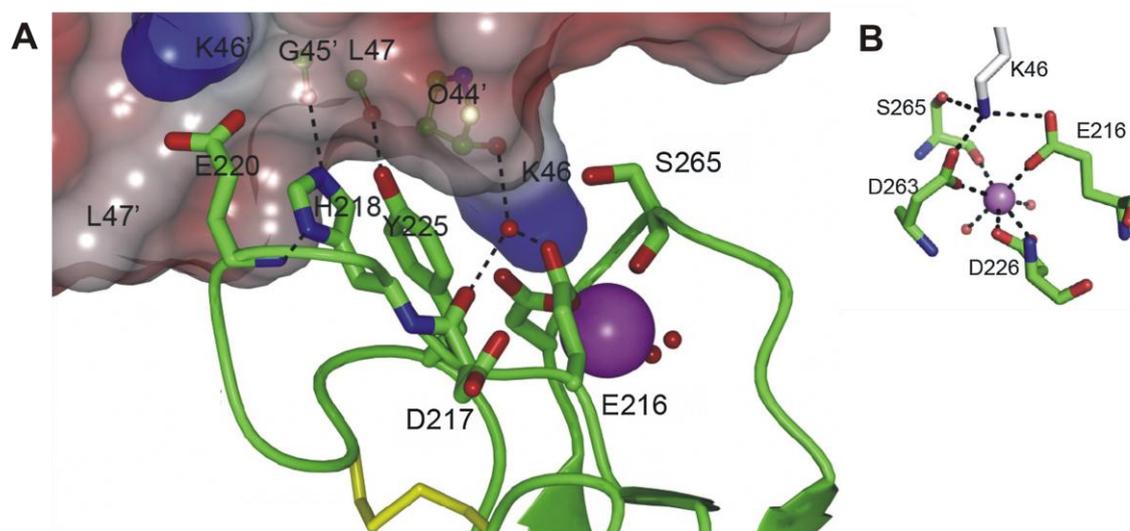


Figure 5.1.1: Interactions between MASP-1 CUB2 domain and MBL collagen domain (Adapted from Gingras *et al.*) **A:** This shows the overall interaction between the CUB2 domain and the collagen. The lysine residue at position 46 of MBL is the main point of interaction and the adjacent leucine at position 47 allows space in the structure for the tyrosine of MASP to fit. **B:** A closer view of the interactions between Lys⁴⁶ of the protease-binding motif of MBL and the calcium-binding residues of the CUB2 domain (Gingras *et al.*, 2011).

One of the most surprising features of the work describe here is that the modified SP-A proteins possessing a MASP-binding site could activate the MASP. Given that the collagenous domains of MBL and SP-A are very different, the size and length of these domains seems relatively unimportant. In addition, this rate of activation was also consistent in two additional SP-A mutants that had been modified to remove the kink within the collagenous domain. This indicates that the flexibility of the collagen stalks has little impact on the mechanism of activation, adding to the idea that MASP binding

is accepting of positional adjustments and changes to the orientation of the binding site in MBL.

One major result of this work was the lack of control that MASP-binding SP-A exerted over MASP-2K activation, suggesting a feature of MBL that SP-A lacks. This control has been located to the α -helical neck and recognition domain by work undertaken in Chapter 3.

By directly contrasting the CRD and neck domains of MBL and SP-A there does not appear to be any obvious disparity other than the length of the neck region itself (figure 5.1.2).

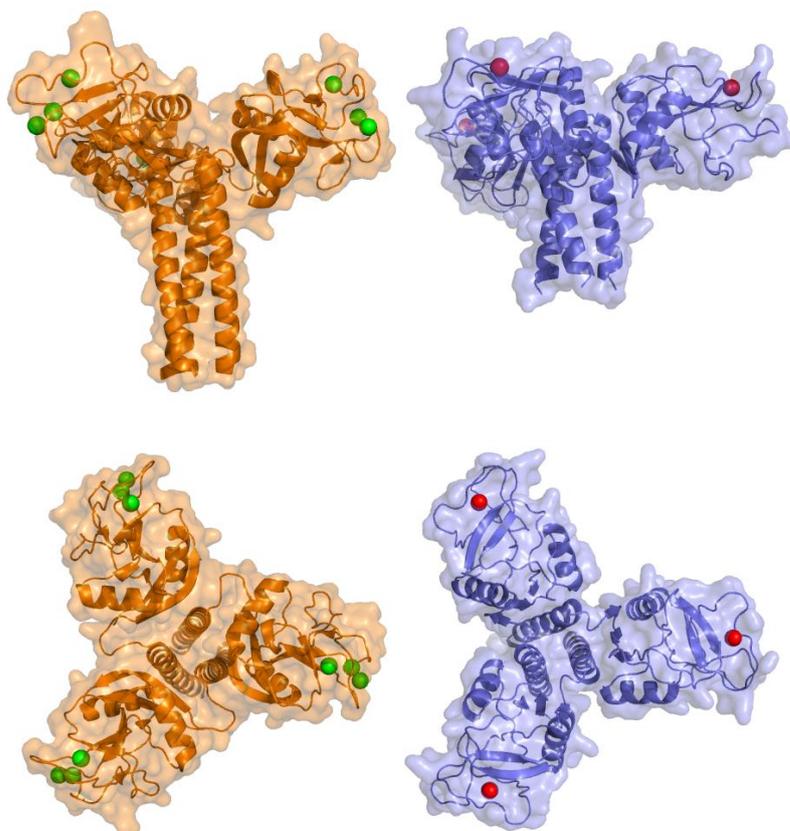


Figure 5.1.2: Comparison of MBL (orange) and SP-A (blue) trimeric heads. Top: Side view of the two collectins. The most striking feature is the length of the neck region. **Bottom:** Top-down view of the proteins. The calcium ions are located towards the tips of the globular heads. SP-A only possess one Ca^{2+} binding site whereas MBL possesses three.

Further comparison between the CRDs of the collectins reveals that whilst there is obvious similarity between the recognition domains the angle at which the MBL CRD and neck separate is distinctly different (figure 5.1.3).

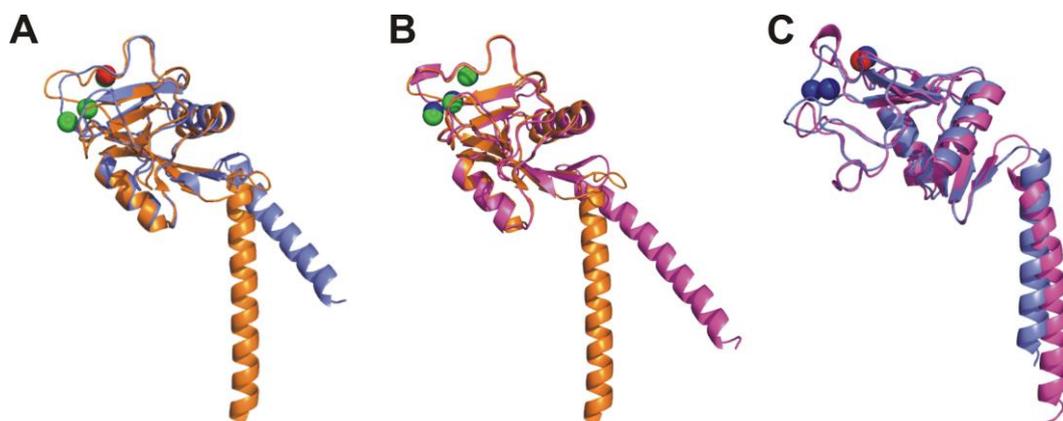


Figure 5.1.3: Comparison of MBL (orange), SP-A (blue) and SP-D (pink): A. MBL aligned against SP-A **B:** MBL aligned against SP-D **C:** SP-A aligned against SP-D. In A and B, the angle of the neck relative to the CRD of MBL is consistently different. In B the Ca²⁺ ions of MBL and SP-D are located in the same spatial positions indicating a similarity of binding. C demonstrates the similarity in CRD/neck angle shared between SP-A and SP-D.

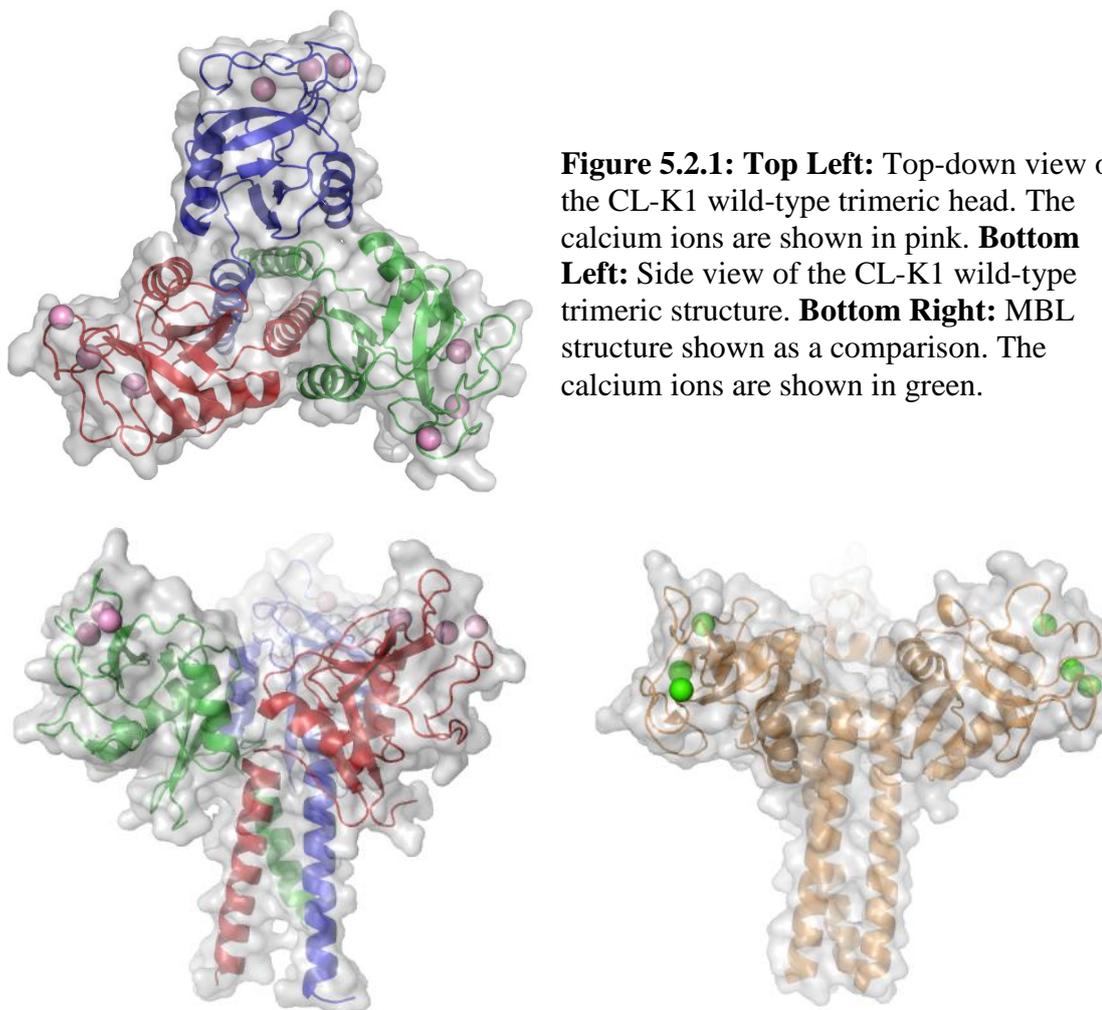
As discussed in Chapter 3 there were additional motivations for producing the MBL protein with the modified specificity for galactose. The Tn antigen and Lewis^x type structures are present on the surface of tumour cells making GBL an ideal target for further research as a potential therapeutic. This work has demonstrated that a galactose-binding lectin is able to selectively bind the sugar and activate MASP-2. There is still some latent mannose binding activity that results in activation demonstrating the loop introduced into the CRD to block mannose binding does not completely exclude the sugar (Iobst *et al.*, 1994a). External work has already demonstrated the ability of the

CRD/neck trimers to bind to cells of the human breast cancer cell line, MCF7, with a 200-fold greater affinity than wild-type MBL-A (Powlesland *et al.*, 2009). These cells express Lewis^x (Gal β 1-4(Fuc α 1-3)GlcNAc) structures and it is the terminal galactose residues to which GBP is able to bind. Combining the knowledge that GBP is able to bind to the Lewis^x structures along with its ability to activate MASP-2 provides credence to the idea of using GBL as a tool for complement mediated lysis of oncogenic cells displaying lewis-type structures on their surface.

5.2 – Collectin 11

The role of collectin 11 in development is important as it opens a new area of enquiry for the study of the lectin pathway of complement activation. Recent work has established the involvement of CL-K1 and MASP-3 as essential for normal foetal development. The crystallisation of the CL-K1 CRD and neck domains in this thesis has led to the structure being solved and other experiments undertaken in Chapter 4 have led to an understanding of the problems the CL-K1 mutants have in correct assembly, longevity and carbohydrate binding.

The structure of the wild-type CL-K1 trimeric CRD and neck domain can be seen in figure 5.2.1, alongside that of MBL for comparison.



One difference that can be noted between the two proteins is the orientation of the calcium ions which are, in CL-K1, rotated closer to the top of the CRD than those of MBL. This suggests a different target or mode of action from that of MBL. Further comparison between MBL and CL-K1 reveal a high level of conservation between the two recognition domains. In CL-K1 the residues involved in calcium binding are Glu232, Asn234, Glu240, Asn252 and Asp253 (figure 5.2.2.A). These residues are conserved from MBL (Glu185, Asn187, Glu193, Asn205 and Asp206) and both the spatial positioning of the calcium ion and the binding residues are the same (figure 5.2.2.B). It has been shown that CL-K1 binds both mannose and fucose in a calcium

dependent manner (Selman *et al.*, 2012a) which agrees with the homology observed between the binding sites of MBL and CL-K1.

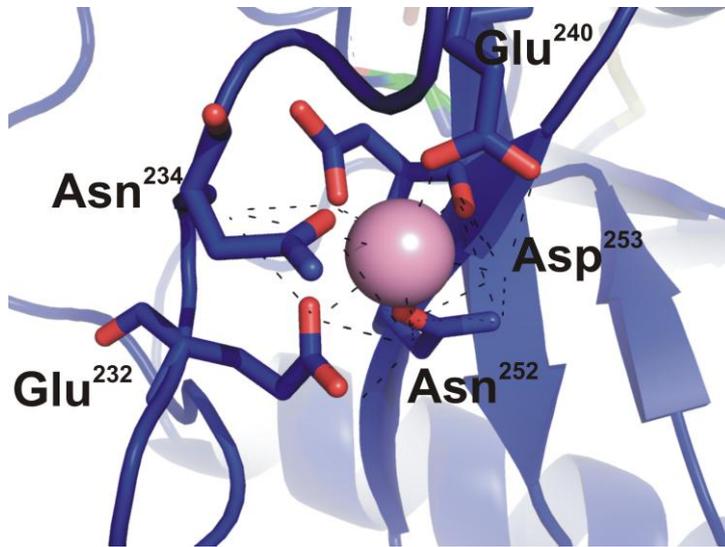


Figure 5.2.2: Calcium binding of CL-K1: The five conserved residues responsible for calcium binding are shown.

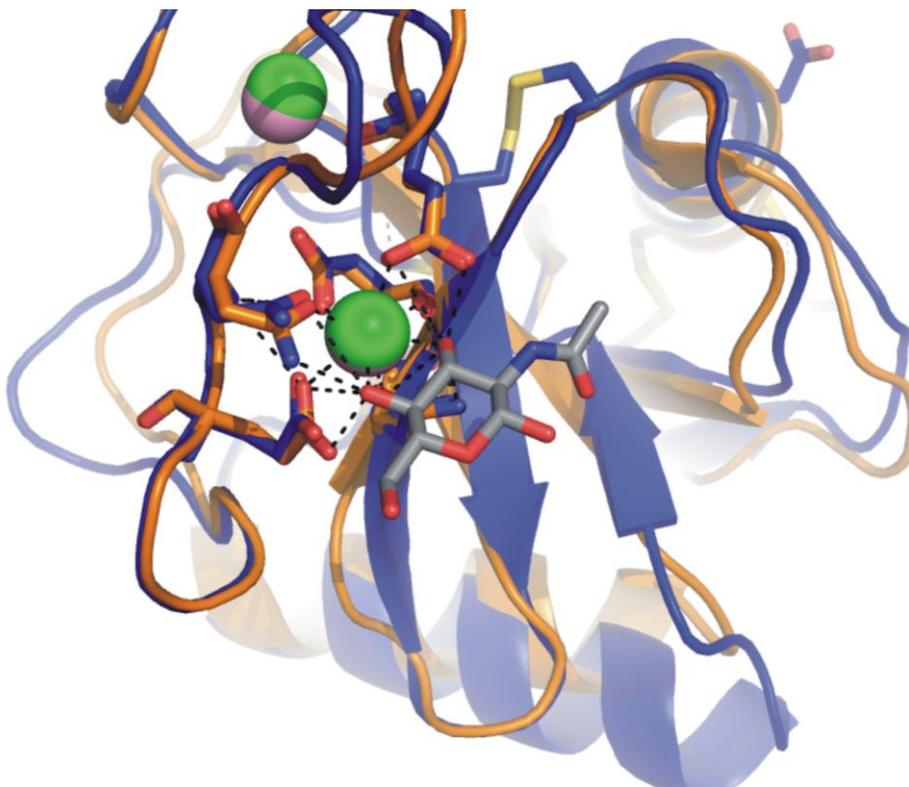


Figure 5.2.3: Alignment of the MBL (orange) and CL-K1 (blue) calcium binding sites. The high level of structural conservation results in the same spatial positions of calcium binding residues and calcium ions (green and pink). The sugar GlcNAc (grey) is part of the MBL structure but most likely will bind to CL-K1 in the same manner due to the high amount of homology.

Even though there is extensive structural homology between the CRDs, as noted above, there is a significant difference between the positioning of the calcium ions of the MBL and CL-K1 trimers. As with SP-A and SP-D, when the monomeric structures are aligned it is instantly obvious of the difference between the angles at which the α -helical neck of the each of the two proteins exit from the recognition domain (figure 5.2.3).

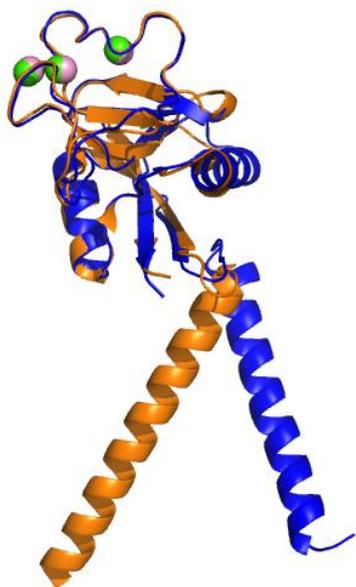


Figure 5.2.3: Alignment of the MBL (*orange*) and CL-K1 (*blue*) recognition domains. The polypeptide chains and calcium binding sites of the two proteins overlay with a highly conserved tertiary structure. At the junction between the CRD and α -helical neck there is an alteration that results in a significant change of direction for the two helices.

This suggests that CL-K1 may have different cellular targets and might not be involved in complement activation. Another indication of an alternative function is the presence of a CXC motif at the junction between the base of the α -helical neck and collagenous domain (Selman *et al.*, 2012a). The Cys⁸⁸ and Cys⁹⁰ residues form disulphide bonds between the three polypeptide chains of the trimer (Hansen *et al.*, 2010). These were not included in the bacterial expression of the CL-K1 CRD to prevent incorrect disulphide bonds forming during refolding. Within the collectin family, this motif is only found elsewhere in CL-L1 and is characteristic of the CXC chemokine family, however no structural resemblance between the two collectins and these chemokines is

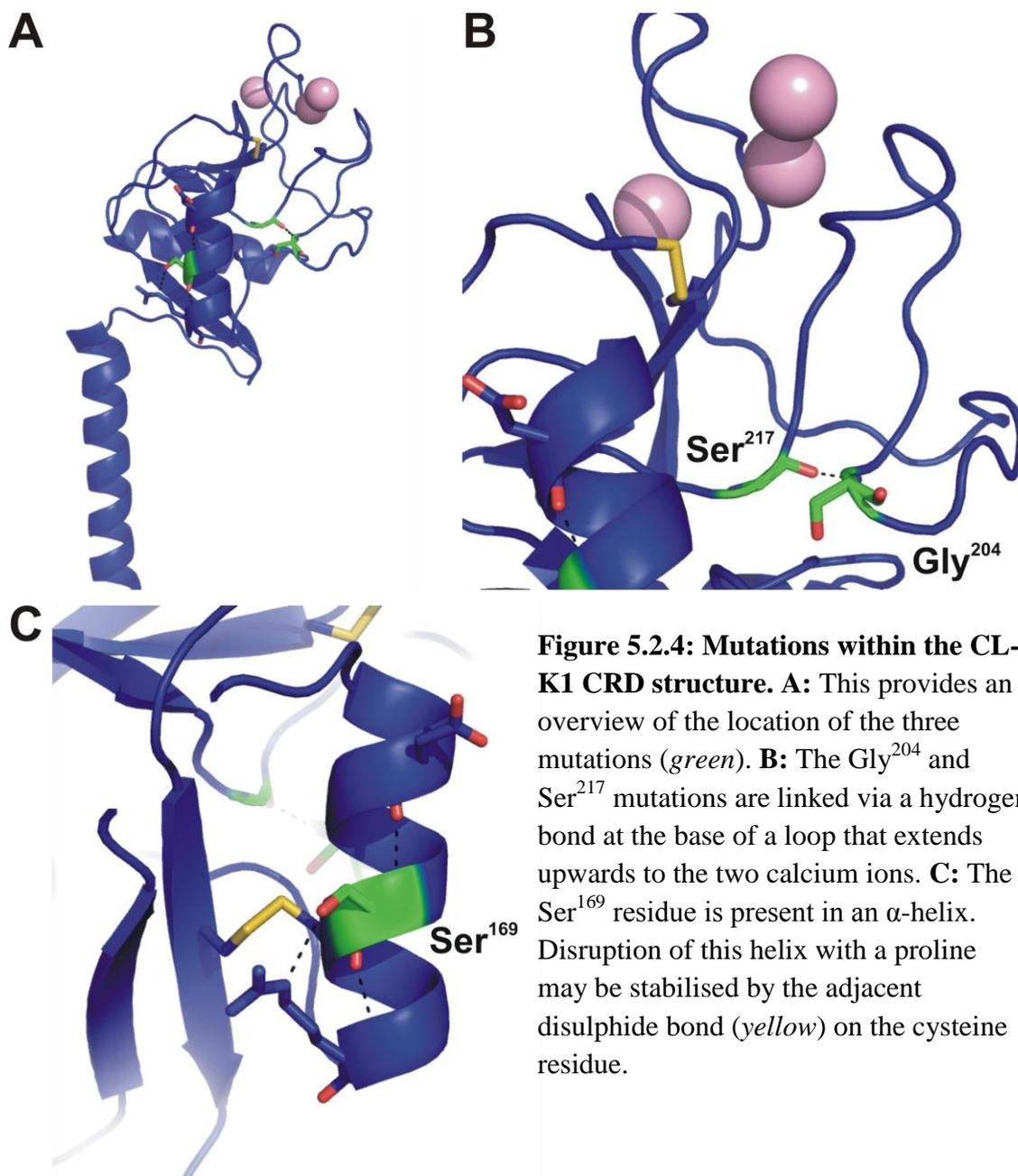
observed (Selman *et al.*, 2012a, Zlotnik *et al.*, 2000). The presence of disulphide bonds between the monomers is likely to interfere with allowing the passage of a signal to pass from the CRD to the collagenous domain, as is thought to be the case of MBL, again suggesting a different function of CL-K1.

Using the wild-type protein crystal structure it is possible to see the effect of the individual mutations that have been identified with 3MC syndrome (figure 5.2.4). The Gly²⁰⁴Ser and Δ Ser²¹⁷ mutations interact with each other through a hydrogen bond and form the base of a loop that extends out to two of the Ca²⁺ ions.

The location of the mutations relative to the overall structure can be seen in figure 5.2.4. In frame B, there is a hydrogen bond between the residues Gly²⁰⁴ to Ser²¹⁷. In the case the oxygen atom of the glycine peptide unit acts as a hydrogen acceptor to the –NH group of the serine residue. This hydrogen bond appears to tether the base of a loop that passes up to two calcium ions. Removal of this bond through mutagenesis of the amino acid residues is likely to disrupt the loop structure which may affect calcium binding and in turn the ability of the CRD to bind to sugars. Substitution of Gly²⁰⁴ with a serine adds a hydroxide side chain (hydrogen donor) into the space where the glycine residue was previously acting as acceptor. Deletion of the Ser²¹⁷ residue may ‘pull around’ the tyrosine residue at position 216 to interfere with the folding of the protein and preventing the hydrogen bond from forming.

Frame C shows the interactions made by Ser¹⁶⁹ in the α -helix. This residue forms three hydrogen bonds, two with Arg¹⁷³ and one with Asp¹⁶⁵. The side chain hydroxide of the

serine acts as a hydrogen donor to the distal guanidinium group of the arginine and the oxygen of the serine peptide unit acts as a hydrogen acceptor to the –NH group of arginine. The –NH group of the serine peptide unit forms a disulphide bond with the oxygen of Asp¹⁶⁵. Substitution of the serine for a proline disrupts the helix as it no longer acts as a hydrogen donor from the amide due to lacking the necessary hydrogen and because it introduces a bend of about 30° in the helix axis (Richardson, 1981). This may be stabilised by the presence of the adjacent Cys¹⁷⁰ forming a disulphide bond with Cys²⁶⁴ which allowed the CRD to fold but results in a protein that is highly susceptible to degradation. Exactly how this mutation affects the sugar binding is not initially obvious since the α -helix is present on the opposite side of the globular head to what is thought to be the sugar-binding calcium. It may reflect a general destabilization of the CRD accounting for the increased sensitivity to trypsin.



5.3 – Future Work

5.3.1 – MBL/MASP Activation

In order to understand the control over MASP-2 activation that MBL possesses, additional mutants could be produced in order to manipulate the properties of the α -helical neck. An initial experiment could be the effect of creating additional chimeras that incorporate the CRD and neck of SP-D into MBL. As can be seen in figure 5.1.3, there are three Ca^{2+} binding sites in both SP-D and MBL in the same positions thus the recognition of sugars is likely to be more similar than that of SP-A. Also, the lengths of the α -helical necks of MBL and SP-D are more comparable. This restricts the major structural differences to the angle present between the neck and CRD and will demonstrate if this has a role to play in activation of MASPs.

Additionally, MBL mutants could be produced that involve subtle changes such as altering the length of the neck or key amino acid substitutions into the SP-A. This could be either to alter the hydrophobicity between the subunit to attempt to adjust the ‘tightness’ of monomer association or to try and alter the angle between the CRD and neck. Since trimerisation of the subunit occurs from the neck, care needs to be taken when altering the properties of this domain. If the sequence is disrupted too much the protein will fail to assemble or even fold.

5.3.2 – Galactose-binding lectin

The next step with this modified lectin will be to determine whether it is possible that this protein is able to activate complement against cells displaying Lewis^x structures. Two cell lines that could be used are the MCF7 human breast cancer cells line and the T

cell leukemia cells, Jurkat cells. This could prove to be of therapeutic value should the modified lectin demonstrate the ability to specifically target and lyse oncogenic cells.

5.3.3 – Collectin 11

Although the structure of the wild-type CL-K1 protein has been demonstrated, this was achieved without any sugar bound to it. One route of experimentation will be to produce a crystal structure with a sugar molecules bound. This will confirm how CL-K1 binds mannose epitopes, but more importantly provide more detailed information about the orientation of the sugars. Additionally, further work is needed to attempt to produce the three mutant CRDs. There was some folding observed and this is likely to improve as the protocol is optimised. Purification of the full length CL-K1 wild-type protein has been achieved but the full length mutants have so far proved elusive.

If the mutants are unable to fold or be successfully transported out of the cell, then the focus of research needs to be the targets with which the wild-type CL-K1 interacts during foetal development. It has been shown that the protein can bind to linearised DNA in a calcium independent manner (Hansen *et al.*, 2010) but what function this serves is unclear.

Appendix 1 - DNA Sequences

Rat Mannose-Binding Lectin A

M L L L P L L V L L C V V S V S S S G S Q T C
 ATCCTCCTGCTTCCACTGCTCGTCCTTCTCTGTGTAGTGAGCGTGTCTCATCAGGGTCACAAACCTGTG
 10 20 30 40 50 60 70
 TACGAGGACGAAGGTGACGAGCAGGAAGAGACACATCACTCGCACAGGAGTAGTCCCAGTGTTTGGACAC

 E E T L K T C S V I A C G R D G R D G P K G E K
 AGGAAACCCTGAAGACTTGCTCTGTGATAGCCTGCGGCAGAGACGGGAGAGATGGGCCCAAAGGGGAGAA
 80 90 100 110 120 130 140
 TCCTTTGGGACTTCTGAACGAGACACTATCGGACGCCGTCTCTGCCCTCTCTACCCGGGTTTCCCCTCTT

 G E P G Q G L R G L Q G P P G K L G P P G S V
 GGGAGAACCAGGTCAAGGGCTCAGGGGCTTGCAGGGCCCTCCAGGAAACTGGGGCCTCCAGGAAGTGTA
 150 160 170 180 190 200 210
 CCCTCTTGGTCCAGTTCCTCGAGTCCCCGAACGTCCCGGGAGGTCCCTTTGACCCCGGAGGTCCCTTACAT

 G A P G S Q G P K G Q K G D R G D S R A I E V
 GGAGCCCTGGAAGTCAAGGACCAAAGGCCAAAAGGGGATCGTGGAGACAGCAGAGCCATTGAGGTGA
 220 230 240 250 260 270 280
 CCTCGGGGACCTTCAAGTTCCTGGTTTTCCGGTTTTTCCCCTAGCACCTCTGTCTCGGTAACCTCCACT

 K L A N M E A E I N T L K S K L E L T N K L H A
 AGCTGGCAAATATGGAGGCAGAGATAAACACCCTGAAGTCAAACCTGGAGCTAACCAACAAGTTGCATGC
 290 300 310 320 330 340 350
 TCGACCGTTTTATACCTCCGTCTCTATTTGTGGGACTTCAAGTTTTGACCTCGATTGGTTGTTCAACGTACG

 F S M G K K S G K K F F V T N H E R M P F S K
 CTTCTCCATGGGTAAAAAGTCTGGGAAGAAGTTCTTTGTGACCAACCATGAAAGGATGCCCTTTTCCAAA
 360 370 380 390 400 410 420
 GAAGAGGTACCCATTTTTTCAGACCCTTCTTCAAGAAACACTGGTTGGTACTTTTCTACGGGAAAAGGTTT

 V K A L C S E L R G T V A I P R N A E E N K A
 GTCAAGGCCCTGTGCTCAGAGCTCCGAGGCACTGTGGCTATCCCCAGGAATGCTGAGGAGAACAAGGCCA
 430 440 450 460 470 480 490
 CAGTTCGGGACACGAGTCTCGAGGCTCCGTGACACCGATAGGGTCCCTTACGACTCCTTGTTCGGT

 I Q E V A K T S A F L G I T D E V T E G Q F M Y
 TCCAAGAAGTGGCTAAAACCTCTGCCTTCTAGGCATCACGGACGAGGTGACTGAAGGCCAATTCATGTA
 500 510 520 530 540 550 560
 AGGTTCTTACCGATTTTTGGAGACGGAAGGATCCGTAGTGCCTGCTCCACTGACTTCCGGTTAAGTACAT

 V T G G R L T Y S N W K K D E P N D H G S G E
 TGTGACAGGGGGGAGGCTCACCTACAGCAACTGGAAAAGGATGAGCCCAATGACCATGGCTCTGGGGAA
 570 580 590 600 610 620 630
 AACTGTCCCCCTCCGAGTGGATGTCGTTGACCTTTTTTCTACTCGGGTTACTGGTACCGAGACCCCTT

 D C V T I V D N G L W N D I S C Q A S H T A V
 GACTGTGTCACTATAGTAGACAACGGTCTGTGGAATGACATCTCCTGCCAAGCTTCCCACACGGCTGTCT
 640 650 660 670 680 690 700
 CTGACACAGTGATATCATCTGTTGCCAGACACCTTACTGTAGAGGACGGTTTCAAGGGTGTGCCGACAGA

 C E F P A *
 GCGAGTTCAGGCTGA
 710
 CGCTCAAGGGTCCGACT

Human Surfactant Protein-A1

M W L C P L A L N L I L M A A S G A V C E V K
ATGTGGCTGTGCCCTCTGGCCCTCAACCTCATCTTGATGGCAGCCTCTGGTGCTGTGTGCGAAGTGAAGG
10 20 30 40 50 60 70
TACACCGACACGGGAGACCGGGAGTTGGAGTAGAACTACCGTCGGGAGACCACGACACACGCTTCACTTCC

D V C V G S P G I P G T P G S H G L P G R D G R
ACGTTTGTGTTGGAAGCCCTGGTATCCCCGGCACTCCTGGATCCCACGGCCTGCCAGGCAGGGACGGGAG
80 90 100 110 120 130 140
TGCAAACACAACCTTCGGGACCATAGGGGCCGTGAGGACCTAGGGTGCCGGACGGTCCGTCCCTGCCCTC

D G L K G D P G P P G P M G P P G E M P C P P
AGATGGTCTCAAAGGAGACCCTGGCCCTCCAGGCCCATGGGTCCACCTGGAGAAATGCCATGTCTCTCT
150 160 170 180 190 200 210
TCTACCAGAGTTTCTCTGGGACCGGGAGGTCCGGGGTACCCAGGTGGACCTCTTTACGGTACAGGAGGA

G N D G L P G A P G I P G E C G E K G E P G E
GGAAATGATGGGCTGCCTGGAGCCCCTGGTATCCCTGGAGAGTGTGGAGAGAAGGGGGAGCCTGGCGAGA
220 230 240 250 260 270 280
CCTTTACTACCCGACGGACCTCGGGGACCATAGGGACCTCTCACACCTCTCTTCCCCCTCGGACCGCTCT

R G P P G L P A H L D E E L Q A T L H D F R H Q
GGGGCCCTCCAGGGCTTCCAGCTCATCTAGATGAGGAGCTCCAAGCCACACTCCACGACTTTAGACATCA
290 300 310 320 330 340 350
CCCCGGGAGGTCCCGAAGGTCGAGTAGATCTACTCCTCGAGGTTCCGGTGTGAGGTGCTGAAATCTGTAGT

I L Q T R G A L S L Q G S I M T V G E K V F S
AATCCTGCAGACAAGGGGAGCCCTCAGTCTGCAGGGCTCCATAATGACAGTAGGAGAGAAGGTCTTCTCC
360 370 380 390 400 410 420
TTAGGACGTCTGTTCCCCTCGGGAGTCAGACGTCCCAGGATTACTGTTCATCTCTTCCAGAAGAGG

S N G Q S I T F D A I Q E A C A R A G G R I A
AGCAATGGGCAGTCCATCACTTTTGTATGCCATTTCAGGAGGCATGTGCCAGAGCAGGCGCCGATTGCTG
430 440 450 460 470 480 490
TCGTTACCCGTCAGGTAGTGAAAACCTACGGTAAGTCTCCGTACACGGTCTCGTCCGCCGGCGTAACGAC

V P R N P E E N E A I A S F V K K Y N T Y A Y V
TCCCAAGGAATCCAGAGGAAAATGAGGCCATTGCAAGCTTCGTGAAGAAGTACAACACATATGCCTATGT
500 510 520 530 540 550 560
AGGGTTCCTTAGGTCTCTCTTTTACTCCGGTAACGTTTCAAGCACTTCTTCATGTTGTGTATACGGATACA

G L T E G P S P G D F R Y S D G T P V N Y T N
AGGCCTGACTGAGGGTCCCAGCCCTGGAGACTTCCGCTACTCAGACGGGACCCCTGTAAACTACACCAAC
570 580 590 600 610 620 630
TCCGGACTGACTCCCAGGGTCCGGACCTCTGAAGGCGATGAGTCTGCCCTGGGGACATTTGATGTGGTTG

W Y R G E P A G R G K E Q C V E M Y T D G Q W
TGGTACCGAGGGGAGCCCGCAGGTCCGGGAAAAGAGCAGTGTGTGGAGATGTACACAGATGGGCAGTGGA
640 650 660 670 680 690 700
ACCATGGCTCCCCCTCGGGCGTCCAGCCCCTTTTCTCGTCACACACCTCTACATGTGTCTACCCGTCACCT

N D R N C L Y S R L T I C E F
ATGACAGGAACTGCCTGTACTCCCAGTACCATCTGTGAGTTCTGA
710 720 730 740
TACTGTCTTTGACGGACATGAGGGCTGACTGGTAGACACTCAAGACT

Chimera 1 – MBL N-terminal and Collagen domain/SP-A CRD

Salt
M L L L P L L V L L C V V S V S S S G
GTCGACGCCACCATGCTCCTGCTTCCACTGCTCGTCCTTCTCTGTGTAGTGAGCGTGTCTCATCAGGGT
10 20 30 40 50
CAGCTGCGGTGGTACGAGGACGAAGGTGACGAGCAGGAAGAGACACATCACTCGCACAGGAGTAGTCCCA
S Q T C E E T L K T C S V I A C G R D G R D G P
CACAAACCTGTGAGGAAACCCTGAAGACTTGCTCTGTGATAGCCTGCGGCAGAGACGGGAGAGATGGGCC
60 70 80 90 100 110 120
GTGTTTGGACACTCCTTTGGGACTTCTGAACGAGACACTATCGGACGCCGTCTCTGCCCTCTCTACCCGG
K G E K G E P G Q G L R G L Q G P P G K L G P
CAAAGGGGAGAAGGGAGAACCAGGTCAAGGGCTCAGGGGCTTGCGAGGGCCCTCCAGGGAAACTGGGGCCT
130 140 150 160 170 180 190
GTTTCCCTCTTCCCTCTTGGTCCAGTTCCCGAGTCCCCGAACGTCCCGGGAGGTCCCTTTGACCCCGGA
P G S V G A P G S Q G P K G Q K G D R G D S A
CCAGGAAGTGTAGGAGCCCCTGGAAGTCAAGGACCAAAGGCCAAAAGGGGATCGTGGAGACAGCGCTC
200 210 220 230 240 250 260
GGTCCTTACATCCTCGGGGACCTTCAGTTCCTGGTTTTCCGGTTTTTCCCTAGCACCTCTGTGCGGAG
H L D E E L Q A T L H D F R H Q I L Q T R G A L
ATCTAGATGAGGAGCTCCAAGCCACACTCCACGACTTTAGACATCAAATCCTGCAGACAAGGGGAGCCCT
270 280 290 300 310 320 330
TAGATCTACTCCTCGAGGTTCGGTGTGAGGTGCTGAAATCTGTAGTTTAGGACGTCTGTTCCCTCGGGA
S L Q G S I M T V G E K V F S S N G Q S I T F
CAGTCTGCAGGGCTCCATAATGACAGTAGGAGAGAAGGTCTTCTCCAGCAATGGGCAGTCCATCACTTTT
340 350 360 370 380 390 400
GTCAGACGTCCCAGGTATTACTGTCATCCTCTTCCAGAAGAGGTGCTTACCCGTCAGGTAGTGAAAA
D A I Q E A C A R A G G R I A V P R N P E E N
GATGCCATTGAGGAGCATGTGCCAGAGCAGGCGCCGATTGCTGTCCCAAGGAATCCAGAGGAAAATG
410 420 430 440 450 460 470
CTACGGTAAGTCCCTCCGTACACGGTCTCGTCCGCCGGCGTAACGACAGGGTTCCTTAGGTCTCCTTTTAC
E A I A S F V K K Y N T Y A Y V G L T E G P S P
AGGCCATTGCAAGCTTCGTGAAGAAGTACAACACATATGCCTATGTAGGCCTGACTGAGGGTCCCAGCCC
480 490 500 510 520 530 540
TCCGGTAACGTTGCAAGCACTTCTTCATGTTGTGTATACGGATACATCCGGACTGACTCCCAGGGTCCGGG
G D F R Y S D G T P V N Y T N W Y R G E P A G
TGGAGACTTCCGCTACTCAGACGGGACCCCTGTAAACTACACCAACTGGTACCGAGGGGAGCCCCGAGGT
550 560 570 580 590 600 610
ACCTCTGAAGGCGATGAGTCTGCCCTGGGGACATTTGATGTGGTTGACCATGGCTCCCCTCGGGCGTCCA
R G K E Q C V E M Y T D G Q W N D R N C L Y S
CGGGAAAAGAGCAGTGTGTGGAGATGTACACAGATGGGCAGTGAATGACAGGAACTGCCTGTACTCCC
620 630 640 650 660 670 680
GCCCCTTTTCTCGTCACACACCTCTACATGTGTCTACCCGTCACCTTACTGTCCTTGACGGACATGAGGG
EcoRI
R L T I C E F *
GACTGACCATCTGTGAGTTCTGAGAATTC
690 700 710
CTGACTGGTAGACACTCAAGACTCTTAAG

Chimera 2 – SPA-KLPOdel N-terminal and collagen domain/MBL CRD

Salt
M W L C P L A L N L I L M A A S G A V
GTTCGACGCCACCATGTGGCTGTGCCCTCTGGCCCTCAACCTCATCTTGATGGCAGCCTCTGGTGCTGTGT
10 20 30 40 50
CAGCTGCGGTGGTACACCGACACGGGAGACCGGGAGTTGGAGTAGAACTACCGTCCGGAGACCACGACACA
C E V K D V C V G S P G I P G T P G S H G L P G
GCGAAGTGAAGGACGTTTGTGTTGGAAGCCCTGGTATCCCCGGCACTCCTGGATCCCACGGCCTGCCAGG
60 70 80 90 100 110 120
CGTTCACTTCTTGCAAACACAACCTTCGGGACCATAGGGGCCGTGAGGACCTAGGGTGCCGGACGGTCC
R D G R D G L K G D P G P P G P M G P P G E M
CAGGGACGGGAGAGATGGTCTCAAAGGAGACCCTGGCCCTCCAGGCCCATGGGTCCACCTGGAGAAATG
130 140 150 160 170 180 190
GTCCCTGCCCTCTCTACCAGAGTTTCTCTGGGACCGGGAGGTCCGGGGTACCCAGGTGGACCTCTTTAC
G N D G L P G A P G I P G K L G P P G E P G E
GGAAATGATGGGCTGCCTGGAGCCCCTGGTATCCCTGGAAAGTTAGGACCGCCGGGGGAGCCTGGCGAGA
200 210 220 230 240 250 260
CCTTTACTACCCGACGGACCTCGGGGACCATAGGGACCTTTCAATCCTGGCGGCCCCCTCGGACCGCTCT
R G P P G L P R A I E V K L A N M E A E I N T L
GGGGCCCTCCAGGGCTTCCAAGAGCCATTGAGGTGAAGCTGGCAAATATGGAGGCAGAGATAAACACCCT
270 280 290 300 310 320 330
CCCCGGGAGGTCCCGAAGGTTCTCGGTAACCTCACTTCGACCGTTTATACTCCGTCTCTATTTGTGGGA
K S K L E L T N K L H A F S M G K K S G K K F
GAAGTCAAACTGGAGCTAACCAACAAGTTGCATGCCTTCTCCATGGGTAAAAAGTCTGGGAAGAAGTTC
340 350 360 370 380 390 400
CTTCAGTTTTGACCTCGATTGGTTGTTCAACGTACGGAAGAGGTACCCATTTTTTCAGACCCCTTCTTCAAG
F V T N H E R M P F S K V K A L C S E L R G T
TTTGTGACCAACCATGAAAGGATGCCCTTTTCCAAAGTCAAGGCCCTGTGCTCAGAGCTCCGAGGCACTG
410 420 430 440 450 460 470
AAACTGTTGGTACTTTTCTACGGGAAAAGTTTTAGTTCCGGGACACGAGTCTCGAGGCTCCGTGAC
V A I P R N A E E N K A I Q E V A K T S A F L G
TGGCTATCCCCAGGAATGCTGAGGAGAACAAGGCCATCCAAGAAGTGGCTAAAACCTCTGCCTTCTTAGG
480 490 500 510 520 530 540
ACCGATAGGGTCCCTTACGACTCCTCTTGTTCGGTAGGTTCTTACCAGATTTTGGAGACGGAAGGATCC
I T D E V T E G Q F M Y V T G G R L T Y S N W
CATCACGGACGAGGTGACTGAAGGCCAATTCATGTATGTGACAGGGGGGAGGCTCACCTACAGCAACTGG
550 560 570 580 590 600 610
GTAGTGCCTGCTCCACTGACTTCCGGTTAAGTACATACTGTCCCCCTCCGAGTGGATGTCTGTTGACC
K K D E P N D H G S G E D C V T I V D N G L W
AAAAAGGATGAGCCCAATGACCATGGCTCTGGGGAAGACTGTGTCACTATAGTAGACAACGGTCTGTGGA
620 630 640 650 660 670 680
TTTTTCTACTCGGGTTACTGGTACCGAGACCCCTTCTGACACAGTGATATCATCTGTTGCCAGACACCT
N D I S C Q A S H T A V C E F P A
ATGACATCTCCTGCCAAGCTTCCCACACGGCTGTCTGCGAGTTCCCAGCC*TGAGAATTC
690 700 710 720 730 740
TACTGTAGAGGACGGTTCGAAGGGTGTGCCGACAGACGCTCAAGGGTCCGACTCTTAAG

EcoRI

Chimera 3 – MBL N-terminal/SPA-KLP Odel collagen and CRD

Salt

M L L L P L L V L L C V V S V S S S G
 GTCGACGCCACCATGCTCCTGCTTCCACTGCTCGTCCTTCTCTGTGTAGTGAGCGTGTCTCATCAGGGT
 10 20 30 40 50
 CAGCTGCGGTGGTACGAGGACGAAGGTGACGAGCAGGAAGAGACACATCACTCGCACAGGAGTAGTCCCA

 S Q T C E E T L K T C S V I A C G S P G I P G K
 CACAAACCTGTGAGGAAACCCTGAAGACTTGCTCTGTGATAGCCTGCGGAAGCCCTGGTATCCCTGGAAA
 60 70 80 90 100 110 120
 GTGTTTGGACACTCCTTTGGGACTTCTGAACGAGACACTATCGGACGCCTTCGGGACCATAGGGACCTTT

 L G P P G E P G E R G P P G L P A H L D E E L
 GTTAGGACCGCCGGGGAGCCTGGCGAGAGGGGCCCTCCAGGGCTTCCAGCTCATCTAGATGAGGAGCTC
 130 140 150 160 170 180 190
 CAATCCTGGCGGCCCTCGGACCGCTCTCCCCGGGAGGTCCCGAAGGTCGAGTAGATCTACTCCTCGAG

 Q A T L H D F R H Q I L Q T R G A L S L Q G S
 CAAGCCACACTCCACGACTTTAGACATCAAATCCTGCAGACAAGGGGAGCCCTCAGTCTGCAGGGCTCCA
 200 210 220 230 240 250 260
 GTTCGGTGTGAGGTGCTGAAATCTGTAGTTTAGGACGTCTGTTCCCCTCGGGAGTCAGACGTCCCGAGGT

 I M T V G E K V F S S N G Q S I T F D A I Q E A
 TAATGACAGTAGGAGAGAAGGTCTTCTCCAGCAATGGGCAGTCCATCACTTTTGTATGCCATTTCAGGAGGC
 270 280 290 300 310 320 330
 ATTACTGTCATCCTCTCTTCCAGAAGAGGTGCTTACCCGTCAGGTAGTGAAAACCTACGGTAAGTCTCCG

 C A R A G G R I A V P R N P E E N E A I A S F
 ATGTGCCAGAGCAGGCGGCCGATTGCTGTCCCAAGGAATCCAGAGGAAAATGAGGCCATTGCAAGCTTC
 340 350 360 370 380 390 400
 TACACGGTCTCGTCCGCGGCGTAACGACAGGGTTCCTTAGGTCTCCTTTTACTCCGGTAACGTTTCAAG

 V K K Y N T Y A Y V G L T E G P S P G D F R Y
 GTGAAGAAGTACAACACATATGCCTATGTAGGCCTGACTGAGGGTCCCAGCCCTGGGAGACTTCCGCTACT
 410 420 430 440 450 460 470
 CACTTCTTCATGTTGTGTATACGGATACATCCGGACTGACTCCCAGGGTCCGGGACCTCTGAAGGCGATGA

 S D G T P V N Y T N W Y R G E P A G R G K E Q C
 CAGACGGGACCCCTGTAAACTACACCAACTGGTACCGAGGGGAGCCCGCAGGTCCGGGAAAAGAGCAGTG
 480 490 500 510 520 530 540
 GTCTGCCCTGGGGACATTTGATGTGGTTGACCATGGCTCCCCTCGGGCGTCCAGCCCCTTTTCTCGTCAC

 V E M Y T D G Q W N D R N C L Y S R L T I C E
 TGTGGAGATGTACACAGATGGGCAGTGAATGACAGGAAGTGCCTGTACTCCCGACTGACCATCTGTGAG
 550 560 570 580 590 600 610
 ACACCTCTACATGTGTCTACCCGTCACCTTACTGTCCTTGACGGACATGAGGGGCTGACTGGTAGACTC

EcoRI

F *
 TTCGAGAATTC
 620
 AAGACTCTTAAG

Chimera 4 – SP-A N-terminal/MBL collagen and CRD

Sall
 M W L C P L A L N L I L M A A S G A V
GTTCGACGCCACCATGTGGCTGTGCCCTCTGGCCCTCAACCTCATCTTGATGGCAGCCTCTGGTGCTGTGT
 10 20 30 40 50 60
CAGCTGCGGTGGTACACCGACACGGGAGACCGGGAGTTGGAGTAGAACTACCGTCCGGAGACCACGACACA
 C E V K D V C V G R D G R D G P K G E K G E P G
 GCGAAGTGAAGGACGTTTTGTGTTGGCAGAGACGGGAGAGATGGGCCCAAAGGGGAGAAGGGAGAACCAGG
 70 80 90 100 110 120 130
 CGCTTCACTTCTTGCAAACACAACCGTCTCTGCCCTCTCTACCCGGGTTTTCCCTCTTCCCTCTTGGTCC
 Q G L R G L Q G P P G K L G P P G S V G A P G
 TCAAGGGCTCAGGGGCTTGCAGGGCCCTCCAGGGAACTGGGGCCTCCAGGAAGTGTAGGAGCCCCCTGGA
 140 150 160 170 180 190 200
 AGTTCCTCCGAGTCCCCGAACGTCCCGGGAGGTCCCTTTGACCCCGGAGGTCCTTACATCTCGGGGACCT
 S Q G P K G Q K G D R G D S R A I E V K L A N
 AGTCAAGGACCAAAGGCCAAAAGGGGATCGTGGAGACAGCAGAGCCATTGAGGTGAAGCTGGCAAATA
 210 220 230 240 250 260 270
 TCAGTTCCTGGTTTTCCGGTTTTTCCCTTAGCACCTCTGTCTCGTCTCGGTAACCTCCACTTCGACCGTTTTAT
 M E A E I N T L K S K L E L T N K L H A F S M G
 TGGAGGCAGAGATAAACACCCTGAAGTCAAACCTGGAGCTAACCAACAAGTTGCATGCCTTCTCCATGGG
 280 290 300 310 320 330 340
 ACCTCCGTCTCTATTTGTGGGACTTCAGTTTTGACCTCGATTGGTTGTTCAACGTACGGAAGAGGTACCC
 K K S G K K F F V T N H E R M P F S K V K A L
 TAAAAAGTCTGGGAAGAAGTTCTTTGTGACCAACCATGAAAGGATGCCCTTTTCCAAAGTCAAGGCCCTG
 350 360 370 380 390 400 410
 ATTTTTTCAGACCCTTCTTCAAGAAACTGGTTGGTACTTTTCTACGGGAAAAGGTTTTTCAGTTCCGGGAC
 C S E L R G T V A I P R N A E E N K A I Q E V
 TGCTCAGAGCTCCGAGGCACTGTGGCTATCCCCAGGAATGCTGAGGAGAACAAGGCCATCCAAGAAGTGG
 420 430 440 450 460 470 480
 ACGAGTCTCGAGGCTCCGTGACACCGATAGGGGTCTTACGACTCCTCTTGTTCGGTAGGTTCTTACC
 A K T S A F L G I T D E V T E G Q F M Y V T G G
 CTA AACCTCTGCCTTCTTAGGCATCACGGACGAGGTGACTGAAGGCCAATTCATGTATGTGACAGGGGG
 490 500 510 520 530 540 550
 GATTTTGGAGACGGAAGGATCCGTAGTGCCTGCTCCACTGACTTCCGGTTAAGTACATACACTGTCCCCC
 R L T Y S N W K K D E P N D H G S G E D C V T
 GAGGCTCACCTACAGCAACTGGAAAAGGATGAGCCCAATGACCATGGCTCTGGGGAAGACTGTGTCACT
 560 570 580 590 600 610 620
 CTCCGAGTGGATGTCGTTGACCTTTTTTCTACTCGGGTTACTGGTACCGAGACCCCTTCTGACACAGTGA
 I V D N G L W N D I S C Q A S H T A V C E F P
 ATAGTAGACAACGGTCTGTGGAATGACATCTCCTGCCAAGCTTCCACACGGCTGTCTGCGAGTTCCCAG
 630 640 650 660 670 680 690
 TATCATCTGTTGCCAGACACCTTACTGTAGAGGACGGTTTCAAGGGTGTGCCGACAGACGCTCAAGGGTC

EcoRI
 A *
 CCTGAGAATTC
 700
 GGACTCTTAAG

Galactose-Binding Lectin

M L L L P L L V L L C V V S V S S S G S Q T C
ATGCTCCTGCTTCCACTGCTCGTCCTTCTCTGTGTAGTGAGCGTGTCTCATCAGGGTCACAAACCTGTG
 10 20 30 40 50 60 70
 TACGAGGACGAAGGTGACGAGCAGGAAGAGACACATCACTCGCACAGGAGTAGTCCCAGTGTGGACAC

 E E T L K T C S V I A C G R D G R D G P K G E K
 AGGAAACCCCTGAAGACTTGCTCTGTGATAGCCTGCGGCAGAGACGGGAGAGATGGGCCCAAAGGGGAGAA
 80 90 100 110 120 130 140
 TCCTTTGGGACTTCTGAACGAGACACTATCGGACGCCGTCTCTGCCCTCTCTACCCGGGTTTTCCCTCTT

 G E P G Q G L R G L Q G P P G K L G P P G S V
 GGGAGAACCAGGTCAAGGGCTCAGGGGCTTGCAGGGCCCTCCAGGGAAACTGGGGCCTCCAGGAAGTGTA
 150 160 170 180 190 200 210
 CCCTCTTGGTCCAGTTCCCGAGTCCCCGAACGTCCCGGGAGGTCCCTTTGACCCCGGAGGTCCCTTCACAT

 G A P G S Q G P K G Q K G D R G D S R A I E V
 GGAGCCCTGGAAGTCAAGGACCAAAAGGCCAAAAGGGGATCGTGAGACAGCAGAGCCATTGAGGTGA
 220 230 240 250 260 270 280
 CCTCGGGGACCTTCAGTTCCTGGTTTTCCGGTTTTTCCCTAGCACCTCTGTCTCGTCTCGGTAACCTCACT

 K L A N M E A E I N T L K S K L E L T N K L H A
 AGCTGGCAAATATGGAGGCAGAGATAAACACCCTGAAGTCAAAACTGGAGCTAACCAACAAGTTGCATGC
 290 300 310 320 330 340 350
 TCGACCGTTTTATACCTCCGTCTCTATTTGTGGGACTTCAGTTTTGACCTCGATTGGTTGTTCAACGTACG

 F S M G K K S G K K F F V T N H E R M P F S K
 CTTCTCCATGGGTAAAAAGTCTGGGAAGAAGTTCTTTGTGACCAACCATGAAAGGATGCCCTTTTCCAAA
 360 370 380 390 400 410 420
 GAAGAGGTACCCATTTTTTCAGACCCTTCTTCAAGAAACTGGTTGGTACTTTTCTACGGGAAAAGGTTT

 V K A L C S E L R G T V A I P R N A E E N K A
 GTCAAGGCCCTGTGCTCAGAGCTCCGAGGCACTGTGGCTATCCCCAGGAATGCTGAGGAGAACAAGGCCA
 430 440 450 460 470 480 490
 CAGTTCGGGACACGAGTCTCGAGGCTCCGTGACACCGATAGGGGTCTTACGACTCCTCTTGTTCGGT

 I Q E V A K T S A F L G I T D E V T E G Q F M Y
 TCCAAGAAGTGGCTAAAACCTCTGCCTTCTAGGCATCACGGACGAGGTGACTGAAGGCCAATTCATGTA
 500 510 520 530 540 550 560
 AGGTTCTTACCGATTTTTGGAGACGGAAGGATCCGTAGTGCCTGCTCCACTGACTTCCGGTTAAGTACAT

 V T G G R L T Y S N W K K D Q P D D W Y G H G
 TGTGACAGGGGGGAGGCTCACCTACAGCAACTGAAAAAGGATCAGCCCGATGACTGGTACGGCCATGGC
 570 580 590 600 610 620 630
 AACTGTCCCCCTCCGAGTGGATGTCGTTGACCTTTTTTCTAGTCCGGCTACTGACCATGCCGGTACCG

 L G G G E D C V T I V D N G L W N D I S C Q A
 CTAGGCGGTGGTGAAGACTGTGCTCACTATAGTAGACAACGGTCTGTGGAATGACATCTCCTGCCAAGCTT
 640 650 660 670 680 690 700
 GATCCGCCACCACTTCTGACACAGTGATATCATCTGTTGCCAGACACCTTACTGTAGAGGACGGTTCGAA

 S H T A V C E F P A Y P Y D V P D Y A *****
 CCCACACGGCTGTCTGCGAGTTCCAGCCTACCCCTACGACGTACCCGACTACGCC**TGA**
 710 720 730 740 750 760
 GGGTGTGCCGACAGACGCTCAAGGGTCGGATGGGGATGCTGCATGGGCTGATGCGGACT

CL-K1 wild-type - α -helical neck and CRD**NcoI**

M A S Q L R K A I G E M D N Q V S Q L T S E
 TA**CCATG**CTAGCCAGCTGCGCAAGGCCATCGGGGAGATGGACAACCAGGTCTCTCAGCTGACCAGCGAG
 500 510 520 530 540 550

AT**GGTACC**GATCGGTTCGACGCGTTCGGTAGCCCCTCTACCTGTTGGTCCAGAGAGTTCGACTGGTTCGCTC

L K F I K N A V A G V R E T E S K I Y L L V K
 CTCAAGTTCATCAAGAATGCTGTGCGCGGTGTGCGCGAGACGGAGAGCAAGATCTACCTGCTGGTGAAGG
 560 570 580 590 600 610 620
 GAGTTC AAGTAGTTCTTACGACAGCGGCCACACGCGCTCTGCCTCTCGTTCTAGATGGACGACCACTTCC

E E K R Y A D A Q L S C Q G R G G T L S M P K D
 AGGAGAAGCGCTACGCGGACGCCAGCTGTCTGCCAGGGCCGCGGGGGCAGCTGAGCATGCCCAAGGA
 630 640 650 660 670 680 690
 TCCTCTTCGCGATGCGCCTGCGGGTTCGACAGGACGGTCCCGGCGCCCCCGTTCGACTCGTACGGGTTCTC

E A A N G L M A A Y L A Q A G L A R V F I G I
 CGAGGCTGCCAATGGCCTGATGGCCGCATACCTGGCGCAAGCCGGCCTGGCCCGTGTCTTTCATCGGCATC
 700 710 720 730 740 750 760
 GCTCCGACGGTTACCGGACTACCGGCGTATGGACCGGTTTCGGCCGGACCGGGCACAGAAGTAGCCGTAG

N D L E K E G A F V Y S D H S P M R T F N K W
 AACGACCTGGAGAAGGAGGGCGCCTTCGTGTACTCTGACCACTCCCCATGCGGACCTTCAACAAGTGGC
 770 780 790 800 810 820 830
 TTGCTGGACCTCTTCTCCCGCGGAAGCACATGAGACTGGTGAGGGGTACGCTGGAAGTTGTTACCCG

R S G E P N N A Y D E E D C V E M V A S G G W N
 GCAGCGGTGAGCCCAACAATGCCTACGACGAGGAGGACTGCGTGGAGATGGTGGCCTCGGGCGGCTGGAA
 840 850 860 870 880 890 900
 CGTCCCACTCGGGTTGTTACGGATGCTGCTCCTCCTGACGCACCTCTACCACCGGAGCCCCGCCGACCTT

D V A C H T T M Y F M C E F D K E N M * **EcoRI**
 CGACGTGGCCTGCCACACCACCATGTACTTTCATGTGTGAGTTTGACAAGGAGAACATG**TGAGAATTC**
 910 920 930 940 950 960 970
 GCTGCACCGGACGGTGTGGTGGTACATGAAGTACACACTCAAACCTGTTCTCTTGTACT**CTTAAG**

CL-K1 S²¹⁷ deletion - α -helical neck and CRD

NcoI
M A S Q L R K A I G E M D N Q V S Q L T S E
TACCATGGCTAGCCAGCTGCGCAAGGCCATCGGGGAGATGGACAACCAGGTCTCTCAGCTGACCAGCGAG
500 510 520 530 540 550
ATGGTACC GATCGGTTCGACGCGTTCCGGTAGCCCCTCTACCTGTTGGTCCAGAGAGTTCGACTGGTTCGCTC
L K F I K N A V A G V R E T E S K I Y L L V K
CTCAAGTTCATCAAGAATGCTGTGCGCGGTGTGCGCGAGACGGAGAGCAAGATCTACCTGCTGGTGAAGG
560 570 580 590 600 610 620
GAGTTCAAGTAGTTCTTACGACAGCGGCCACACGCGCTCTGCCTCTCGTTCTAGATGGACGACCACTTCC
E E K R Y A D A Q L S C Q G R G G T L S M P K D
AGGAGAAGCGCTACGCGGACGCCAGCTGTCTGCCAGGGCCGCGGGGGCACGCTGAGCATGCCCAAGGA
630 640 650 660 670 680 690
TCCTCTTCGCGATGCGCCTGCGGGTTCGACAGGACGGTCCC GGCGCCCCCGTTCGCGACTCGTACGGGTTCTT
E A A N G L M A A Y L A Q A G L A R V F I G I
CGAGGCTGCCAATGGCCTGATGGCCGCATACCTGGCGCAAGCCGGCCTGGCCCCGTGTCTTCATCGGCATC
700 710 720 730 740 750 760
GCTCCGACGGTTACCGGACTACCGGCGTATGGACCGGTTTCGCGCGGACCGGGCACAGAAGTAGCCGTAG
N D L E K E G A F V Y - D H S P M R T F N K W
AACGACCTGGAGAAGGAGGGCGCCTTCGTGTACXXXGACCACTCCCCATGCGGACCTTCAACAAGTGGC
770 780 790 800 810 820 830
TTGCTGGACCTCTTCTCCCGCGGAAGCACATGXXXCTGGTGAGGGGTACGCTGGAAGTTGTTACCG
R S G E P N N A Y D E E D C V E M V A S G G W N
GCAGCGGTGAGCCCAACAATGCCTACGACGAGGAGGACTGCGTGAGATGGTGGCCTCGGGCGGCTGGAA
840 850 860 870 880 890 900
CGTCGCCACTCGGGTTGTTACGGATGCTGCTCCTCCTGACGCACCTCTACCACCGGAGCCCGCCGACCTT
D V A C H T T M Y F M C E F D K E N M * EcoRI
CGACGTGGCCTGCCACACCACCATGTACTTCATGTGTGAGTTTGACAAGGAGAACATGTGAGAATTC
910 920 930 940 950 960
GCTGCACCGGACGGTGTGGTGGTACATGAAGTACAACTCAAAGTTCCTTCTTGTACTCTTAAG

Bibliography

- Ambrus, G., Gal, P., Kojima, M., Szilagyi, K., Balczer, J., Antal, J., Graf, L., Laich, A., Moffatt, B. E., Schwaeble, W., Sim, R. B., Zavodszky, P., 2003. Natural substrates and inhibitors of mannan-binding lectin-associated serine protease-1 and -2: a study on recombinant catalytic fragments. *J Immunol*, **170**, 1374-1382.
- Arlaud, G. J. & Colomb, M. G., 1987. Modelling of C1, the first component of human complement: towards a consensus? *Molecular Immunology*, **24**, 317.
- Asherie, N., 2004. Protein crystallization and phase diagrams. *Methods*, **34**, 266-272.
- Bhakdi, S. & Trantum-Jensen, J., 1991. Complement lysis: a hole is a hole. *Immunol Today*, **12**, 318-320.
- Blanquet-Grossard, F., Thielens, N. M., Vendrely, C., Jamin, M., Arlaud, G. J., 2005. Complement protein C1q recognizes a conformationally modified form of the prion protein. *Biochemistry*, **44**, 4349-4356.
- Bruns, G., Stroh, H., Veldman, G. M., Latt, S. A., Floros, J., 1987. The 35 kd pulmonary surfactant-associated protein is encoded on chromosome 10. *Hum Genet*, **76**, 58-62.
- Burge, J., Nicholson-Weller, A., Austen, K. F., 1981. Isolation of C4-binding protein from guinea pig plasma and demonstration of its function as a control protein of the classical complement pathway C3 convertase. *J Immunol*, **126**, 232-235.
- Burton, D. R., Boyd, J., Brampton, A. D., Easterbrook-Smith, S. B., Emanuel, E. J., Novotny, J., Rademacher, T. W., van Schravendijk, M. R., Sternberg, M. J., Dwek, R. A., 1980. The C1q receptor site on immunoglobulin G. *Nature*, **288**, 338-344.
- Campbell, R. D., Law, S. K., Reid, K. B., Sim, R. B., 1988. Structure, organization, and regulation of the complement genes. *Annu Rev Immunol*, **6**, 161-195.
- Chen, C.-B. & Wallis, R., 2004. Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases. *Journal of Biological Chemistry*, **279**, 26058-26065.

- Chen, C. B. & Wallis, R., 2001. Stoichiometry of complexes between mannose-binding protein and its associated serine proteases. Defining functional units for complement activation. *J Biol Chem*, **276**, 25894-25902.
- Childs, R. A., Feizi, T., Yuen, C. T., Drickamer, K., Quesenberry, M. S., 1990. Differential recognition of core and terminal portions of oligosaccharide ligands by carbohydrate-recognition domains of two mannose-binding proteins. *J Biol Chem*, **265**, 20770-20777.
- Childs, R. A., Wright, J. R., Ross, G. F., Yuen, C. T., Lawson, A. M., Chai, W., Drickamer, K., Feizi, T., 1992. Specificity of lung surfactant protein SP-A for both the carbohydrate and the lipid moieties of certain neutral glycolipids. *Journal of Biological Chemistry*, **267**, 9972-9979.
- Collaborative, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D*, **50**, 760-763.
- Crouch, E., Hartshorn, K., Ofek, I., 2000. Collectins and pulmonary innate immunity. *Immunol Rev*, **173**, 52-65.
- Crouch, E. & Wright, J. R., 2001. Surfactant proteins a and d and pulmonary host defense. *Annu Rev Physiol*, **63**, 521-554.
- Dahl, M. R., Thiel, S., Matsushita, M., Fujita, T., Willis, A. C., Christensen, T., Vorup-Jensen, T., Jensenius, J. C., 2001. MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway. *Immunity*, **15**, 127-135.
- Davis, A. E., 3rd, Mejia, P., Lu, F., 2008. Biological activities of C1 inhibitor. *Mol Immunol*, **45**, 4057-4063.
- Degn, S. E., Hansen, A. G., Steffensen, R., Jacobsen, C., Jensenius, J. C., Thiel, S., 2009. MAp44, a human protein associated with pattern recognition molecules of the complement system and regulating the lectin pathway of complement activation. *J Immunol*, **183**, 7371-7378.
- Dempsey, P. W., Allison, M. E., Akkaraju, S., Goodnow, C. C., Fearon, D. T., 1996. C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science*, **271**, 348-350.

- Dodds, A. W., Ren, X. D., Willis, A. C., Law, S. K., 1996. The reaction mechanism of the internal thioester in the human complement component C4. *Nature*, **379**, 177-179.
- Dodds, A. W., Sim, R. B., Porter, R. R., Kerr, M. A., 1978. Activation of the first component of human complement (C1) by antibody-antigen aggregates. *Biochem J*, **175**, 383-390.
- Drickamer, K., 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J Biol Chem*, **263**, 9557-9560.
- Drickamer, K., 1997. Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure*, **5**, 465-468.
- Drickamer, K., Dordal, M. S., Reynolds, L., 1986. Mannose-binding proteins isolated from rat liver contain carbohydrate-recognition domains linked to collagenous tails. Complete primary structures and homology with pulmonary surfactant apoprotein. *J Biol Chem*, **261**, 6878-6887.
- Drickamer, K. & Taylor, M. E., 1993. Biology of animal lectins. *Annu Rev Cell Biol*, **9**, 237-264.
- Dube, D. H. & Bertozzi, C. R., 2005. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov*, **4**, 477-488.
- Duncan, A. R. & Winter, G., 1988. The binding site for C1q on IgG. *Nature*, **332**, 738-740.
- Duval, N., Krejci, E., Grassi, J., Coussen, F., Massoulie, J., Bon, S., 1992. Molecular architecture of acetylcholinesterase collagen-tailed forms; construction of a glycolipid-tailed tetramer. *Embo J*, **11**, 3255-3261.
- Ezekowitz, R. A., Kuhlman, M., Groopman, J. E., Byrn, R. A., 1989. A human serum mannose-binding protein inhibits in vitro infection by the human immunodeficiency virus. *J Exp Med*, **169**, 185-196.
- Farries, T. C., Lachmann, P. J., Harrison, R. A., 1988. Analysis of the interaction between properdin and factor B, components of the alternative-pathway C3 convertase of complement. *Biochem J*, **253**, 667-675.

- Fearon, D. T. & Austen, K. F., 1975a. Initiation of C3 cleavage in the alternative complement pathway. *J Immunol*, **115**, 1357-1361.
- Fearon, D. T. & Austen, K. F., 1975b. Properdin: binding to C3b and stabilization of the C3b-dependent C3 convertase. *Journal of Experimental Medicine*, **142**, 856-863.
- Fearon, D. T. & Austen, K. F., 1975c. Properdin: initiation of alternative complement pathway. *Proc Natl Acad Sci U S A*, **72**, 3220-3224.
- Feinberg, H., Uitdehaag, J. C., Davies, J. M., Wallis, R., Drickamer, K., Weis, W. I., 2003. Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2. *Embo J*, **22**, 2348-2359.
- Floros, J., 2001. Human surfactant protein A (SP-A) variants: why so many, why such a complexity? *Swiss Med Wkly*, **131**, 87-90.
- Frank, M. M. & Fries, L. F., 1991. The role of complement in inflammation and phagocytosis. *Immunol Today*, **12**, 322-326.
- Gal, P., Barna, L., Kocsis, A., Zavodszky, P., 2007. Serine proteases of the classical and lectin pathways: similarities and differences. *Immunobiology*, **212**, 267-277.
- Garred, P., Larsen, F., Madsen, H. O., Koch, C., 2003. Mannose-binding lectin deficiency--revisited. *Mol Immunol*, **40**, 73-84.
- Garred, P., Madsen, H. O., Balslev, U., Hofmann, B., Pedersen, C., Gerstoft, J., Svejgaard, A., 1997. Susceptibility to HIV infection and progression of AIDS in relation to variant alleles of mannose-binding lectin. *Lancet*, **349**, 236-240.
- Garred, P., Madsen, H. O., Marquart, H., Hansen, T. M., Sorensen, S. F., Petersen, J., Volck, B., Svejgaard, A., Graudal, N. A., Rudd, P. M., Dwek, R. A., Sim, R. B., Andersen, V., 2000. Two edged role of mannose binding lectin in rheumatoid arthritis: a cross sectional study. *J Rheumatol*, **27**, 26-34.
- Garred, P., Voss, A., Madsen, H. O., Junker, P., 2001. Association of mannose-binding lectin gene variation with disease severity and infections in a population-based cohort of systemic lupus erythematosus patients. *Genes Immun*, **2**, 442-450.
- Geijtenbeek, T. B., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Middel, J., Cornelissen, I. L., Nottet, H. S., KewalRamani, V. N., Littman, D.

- R., Figdor, C. G., van Kooyk, Y., 2000. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*, **100**, 587-597.
- Gerard, C. & Gerard, N. P., 1994. C5A anaphylatoxin and its seven transmembrane-segment receptor. *Annu Rev Immunol*, **12**, 775-808.
- 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., Dodds, A. W., Roscher, S., Reid, K. B. M., Wallis, R., 2007. Localization and characterization of the mannanose-binding lectin (MBL)-associated-serine protease-2 binding site in rat ficolin-A: equivalent binding sites within the collagenous domains of MBLs and ficolins. *Journal of Immunology*, **179**, 455-462.
- 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.
- Graudal, N. A., Madsen, H. O., Tarp, U., Svejgaard, A., Jurik, G., Graudal, H. K., Garred, P., 2000. The association of variant mannanose-binding lectin genotypes with radiographic outcome in rheumatoid arthritis. *Arthritis Rheum*, **43**, 515-521.
- Gregory, L. A., Thielens, N. M., Matsushita, M., Sorensen, R., Arlaud, G. J., Fontecilla-Camps, J. C., Gaboriaud, C., 2004. The X-ray structure of human mannan-binding lectin-associated protein 19 (MAp19) and its interaction site with mannan-binding lectin and L-ficolin. *J Biol Chem*, **279**, 29391-29397.
- Guo, N., Mogue, T., Weremowicz, S., Morton, C. C., Sastry, K. N., 1998. The human ortholog of rhesus mannanose-binding protein-A gene is an expressed pseudogene that localizes to chromosome 10. *Mamm Genome*, **9**, 246-249.
- Hakomori, S., 1996. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res*, **56**, 5309-5318.
- Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M. O., Nielsen, O., Hartshorn, K., Jorgensen, T. J., Skjodt, K., Holmskov, U., 2010. Collectin 11 (CL-11, CL-K1) is a MASP-1/3-associated plasma collectin with microbial-binding activity. *J Immunol*, **185**, 6096-6104.

- Hart, M. L., Ceonzo, K. A., Shaffer, L. A., Takahashi, K., Rother, R. P., Reenstra, W. R., Buras, J. A., Stahl, G. L., 2005. Gastrointestinal ischemia-reperfusion injury is lectin complement pathway dependent without involving C1q. *J Immunol*, **174**, 6373-6380.
- Hart, M. L., Saifuddin, M., Uemura, K., Bremer, E. G., Hooker, B., Kawasaki, T., Spear, G. T., 2002. High mannose glycans and sialic acid on gp120 regulate binding of mannose-binding lectin (MBL) to HIV type 1. *AIDS Res Hum Retroviruses*, **18**, 1311-1317.
- Heja, D., Harmat, V., Fodor, K., Wilmanns, M., Dobo, J., Kekesi, K. A., Zavodszky, P., Gal, P., Pal, G., 2012. Monospecific inhibitors show that both mannan-binding lectin-associated serine protease-1 (MASP-1) and -2 Are essential for lectin pathway activation and reveal structural plasticity of MASP-2. *J Biol Chem*, **287**, 20290-20300.
- Holmskov, U. L., 2000. Collectins and collectin receptors in innate immunity. *APMIS Suppl*, **100**, 1-59.
- Horiuchi, T., Tsukamoto, H., Morita, C., Sawabe, T., Harashima, S., Nakashima, H., Miyahara, H., Hashimura, C., Kondo, M., 2000. Mannose binding lectin (MBL) gene mutation is not a risk factor for systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) in Japanese. *Genes Immun*, **1**, 464-466.
- Huang, Y. F., Wang, W., Han, J. Y., Wu, X. W., Zhang, S. T., Liu, C. J., Hu, Q. G., Xiong, P., Hamvas, R. M., Wood, N., Gong, F. L., Bittles, A. H., 2003. Increased frequency of the mannose-binding lectin LX haplotype in Chinese systemic lupus erythematosus patients. *Eur J Immunogenet*, **30**, 121-124.
- Iobst, S. T. & Drickamer, K., 1994a. Binding of sugar ligands to Ca(2+)-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis. *Journal of Biological Chemistry*, **269**, 15512-15519.
- Iobst, S. T., Wormald, M. R., Weis, W. I., Dwek, R. A., Drickamer, K., 1994b. Binding of sugar ligands to Ca(2+)-dependent animal lectins. I. Analysis of mannose binding by site-directed mutagenesis and NMR. *Journal of Biological Chemistry*, **269**, 15505-15511.
- Iwaki, D., Kanno, K., Takahashi, M., Endo, Y., Matsushita, M., Fujita, T., 2011. The role of mannose-binding lectin-associated serine protease-3 in activation of the alternative complement pathway. *J Immunol*, **187**, 3751-3758.

- 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. *J Mol Biol*, **391**, 246-259.
- Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography*, **26**, 795-800.
- Katyal, S. L., Singh, G., Locker, J., 1992. Characterization of a second human pulmonary surfactant-associated protein SP-A gene. *Am J Respir Cell Mol Biol*, **6**, 446-452.
- Kaufman, R. J., Davies, M. V., Wasley, L. C., Michnick, D., 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.
- Kawasaki, N., Kawasaki, T., Yamashina, I., 1983. Isolation and characterization of a mannan-binding protein from human serum. *J Biochem*, **94**, 937-947.
- Kawasaki, T., Etoh, R., Yamashina, I., 1978. Isolation and characterization of a mannan-binding protein from rabbit liver. *Biochem Biophys Res Commun*, **81**, 1018-1024.
- Keshi, H., Sakamoto, T., Kawai, T., Ohtani, K., Katoh, T., Jang, S. J., Motomura, W., Yoshizaki, T., Fukuda, M., Koyama, S., Fukuzawa, J., Fukuoh, A., Yoshida, I., Suzuki, Y., Wakamiya, N., 2006. Identification and characterization of a novel human collectin CL-K1. *Microbiol Immunol*, **50**, 1001-1013.
- Kim, Y. J. & Varki, A., 1997. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj J*, **14**, 569-576.
- Kishore, U., Greenhough, T. J., Waters, P., Shrive, A. K., Ghai, R., Kamran, M. F., Bernal, A. L., Reid, K. B., Madan, T., Chakraborty, T., 2006. Surfactant proteins SP-A and SP-D: structure, function and receptors. *Mol Immunol*, **43**, 1293-1315.
- Kolble, K., Lu, J., Mole, S. E., Kaluz, S., Reid, K. B., 1993. Assignment of the human pulmonary surfactant protein D gene (SFTP4) to 10q22-q23 close to the surfactant protein A gene cluster. *Genomics*, **17**, 294-298.

- Kuraya, M., Ming, Z., Liu, X., Matsushita, M., Fujita, T., 2005. Specific binding of L-ficolin and H-ficolin to apoptotic cells leads to complement activation. *Immunobiology*, **209**, 689-697.
- Levitz, S. M., Tabuni, A., Treseler, C., 1993. Effect of mannose-binding protein on binding of *Cryptococcus neoformans* to human phagocytes. *Infect Immun*, **61**, 4891-4893.
- Lipscombe, R. J., Sumiya, M., Summerfield, J. A., Turner, M. W., 1995. Distinct physicochemical characteristics of human mannose binding protein expressed by individuals of differing genotype. *Immunology*, **85**, 660-667.
- Liszewski, M. K., Post, T. W., Atkinson, J. P., 1991. Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu Rev Immunol*, **9**, 431-455.
- Lu, J. H., Thiel, S., Wiedemann, H., Timpl, R., Reid, K. B., 1990. Binding of the pentamer/hexamer forms of mannan-binding protein to zymosan activates the proenzyme C1r2C1s2 complex, of the classical pathway of complement, without involvement of C1q. *J Immunol*, **144**, 2287-2294.
- Madsen, H. O., Garred, P., Kurtzhals, J. A., Lamm, L. U., Ryder, L. P., Thiel, S., Svejgaard, A., 1994. A new frequent allele is the missing link in the structural polymorphism of the human mannan-binding protein. *Immunogenetics*, **40**, 37-44.
- Madsen, H. O., Garred, P., Thiel, S., Kurtzhals, J. A., Lamm, L. U., Ryder, L. P., Svejgaard, A., 1995. Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J Immunol*, **155**, 3013-3020.
- Matsushita, M. & Fujita, T., 1992. Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. *J Exp Med*, **176**, 1497-1502.
- Matsushita, M., Thiel, S., Jensenius, J. C., Terai, I., Fujita, T., 2000. Proteolytic Activities of Two Types of Mannose-Binding Lectin-Associated Serine Protease. *The Journal of Immunology*, **165**, 2637-2642.
- McCormack, F. X., 1998. Structure, processing and properties of surfactant protein A. *Biochim Biophys Acta*, **19**, 2-3.

- Miller, J. & Germain, R. N., 1986. Efficient cell surface expression of class II MHC molecules in the absence of associated invariant chain. *J Exp Med*, **164**, 1478-1489.
- Mitchell, D. A., Fadden, A. J., Drickamer, K., 2001. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J Biol Chem*, **276**, 28939-28945.
- Mitchell, D. A., Kirby, L., Paulin, S. M., Villiers, C. L., Sim, R. B., 2007. Prion protein activates and fixes complement directly via the classical pathway: implications for the mechanism of scrapie agent propagation in lymphoid tissue. *Mol Immunol*, **44**, 2997-3004.
- Moller-Kristensen, M., Thiel, S., Hansen, A. G., Jensenius, J. C., 2003. On the site of C4 deposition upon complement activation via the mannan-binding lectin pathway or the classical pathway. *Scand J Immunol*, **57**, 556-561.
- Moller-Kristensen, M., Wang, W., Ruseva, M., Thiel, S., Nielsen, S., Takahashi, K., Shi, L., Ezekowitz, A., Jensenius, J. C., Gadjeva, M., 2005. Mannan-binding lectin recognizes structures on ischaemic reperfused mouse kidneys and is implicated in tissue injury. *Scand J Immunol*, **61**, 426-434.
- Mombo, L. E., Lu, C. Y., Ossari, S., Bedjabaga, I., Sica, L., Krishnamoorthy, R., Lapoumeroulie, C., 2003. Mannose-binding lectin alleles in sub-Saharan Africans and relation with susceptibility to infections. *Genes Immun*, **4**, 362-367.
- Muller-Eberhard, H. J., 1985. Transmembrane channel-formation by five complement proteins. *Biochem Soc Symp*, **50**, 235-246.
- Muller-Eberhard, H. J., 1988. Molecular organization and function of the complement system. *Annual Review of Biochemistry*, **57**, 321-347.
- Muller-Eberhard, H. J. & Schreiber, R. D., 1980. Molecular biology and chemistry of the alternative pathway of complement. *Advances in Immunology*, **29**, 1-53.
- Nayak, A., Dodagatta-Marri, E., Tsolaki, A. G., Kishore, U., 2012. An Insight into the Diverse Roles of Surfactant Proteins, SP-A and SP-D in Innate and Adaptive Immunity. *Front Immunol*, **3**, 7.

- Neth, O., Hann, I., Turner, M. W., Klein, N. J., 2001. Deficiency of mannose-binding lectin and burden of infection in children with malignancy: a prospective study. *Lancet*, **358**, 614-618.
- Neth, O., Jack, D. L., Dodds, A. W., Holzel, H., Klein, N. J., Turner, M. W., 2000. Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infect Immun*, **68**, 688-693.
- Ng, K. K., Kolatkar, A. R., Park-Snyder, S., Feinberg, H., Clark, D. A., Drickamer, K., Weis, W. I., 2002. Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J Biol Chem*, **277**, 16088-16095.
- Pastinen, T., Liitsola, K., Niini, P., Salminen, M., Syvanen, A. C., 1998. Contribution of the CCR5 and MBL genes to susceptibility to HIV type 1 infection in the Finnish population. *AIDS Res Hum Retroviruses*, **14**, 695-698.
- Perkins, S. J., Nan, R., Li, K., Khan, S., Miller, A., 2012. Complement factor H-ligand interactions: self-association, multivalency and dissociation constants. *Immunobiology*, **217**, 281-297.
- Perkins, S. J., Nealis, A. S., Sim, R. B., 1991. Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry*, **30**, 2847-2857.
- Petersen, S. V., Thiel, S., Jensen, L., Vorup-Jensen, T., Koch, C., Jensenius, J. C., 2000. Control of the classical and the MBL pathway of complement activation. *Mol Immunol*, **37**, 803-811.
- 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., Keeble, A. H., Wallis, R., 2009. Analogous interactions in initiating complexes of the classical and lectin pathways of complement. *J Immunol*, **182**, 7708-7717.
- Pikaar, J. C., Voorhout, W. F., van Golde, L. M., Verhoef, J., Van Strijp, J. A., van Iwaarden, J. F., 1995. Opsonic activities of surfactant proteins A and D in phagocytosis of gram-negative bacteria by alveolar macrophages. *J Infect Dis*, **172**, 481-489.
- Powlesland, A. S., Hitchen, P. G., Parry, S., Graham, S. A., Barrio, M. M., Elola, M. T., Mordoh, J., Dell, A., Drickamer, K., Taylor, M. E., 2009. Targeted glycoproteomic identification of cancer cell glycosylation. *Glycobiology*, **19**, 899-909.

- Preissner, K. T., 1991. Structure and biological role of vitronectin. *Annu Rev Cell Biol*, **7**, 275-310.
- Quesenberry, M. S. & Drickamer, K., 1992. Role of conserved and nonconserved residues in the Ca(2+)-dependent carbohydrate-recognition domain of a rat mannose-binding protein. Analysis by random cassette mutagenesis. *J Biol Chem*, **267**, 10831-10841.
- Reboul, A., Arlaud, G. J., Sim, R. B., Colomb, M. G., 1977. A simplified procedure for the purification of C1-inactivator from human plasma. Interaction with complement subcomponents C1r and C1s. *FEBS Lett*, **79**, 45-50.
- Reid, K. B., 1983. Proteins involved in the activation and control of the two pathways of human complement. *Biochemical Society Transactions*, **11**, 1-12.
- Resnick, D., Chatterton, J. E., Schwartz, K., Slayter, H., Krieger, M., 1996. Structures of class A macrophage scavenger receptors. Electron microscopic study of flexible, multidomain, fibrous proteins and determination of the disulfide bond pattern of the scavenger receptor cysteine-rich domain. *J Biol Chem*, **271**, 26924-26930.
- Richardson, J. S., 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem*, **34**, 167-339.
- Rooryck, C., Diaz-Font, A., Osborn, D. P. S., Chabchoub, E., Hernandez-Hernandez, V., Shamseldin, H., Kenny, J., Waters, A., Jenkins, D., Kaissi, A. A., Leal, G. F., Dallapiccola, B., Carnevale, F., Bitner-Glindzicz, M., Lees, M., Hennekam, R., Stanier, P., Burns, A. J., Peeters, H., Alkuraya, F. S., Beales, P. L., 2011. Mutations in lectin complement pathway genes COLEC11 and MASP1 cause 3MC syndrome. *Nature Genetics*, **43**, 197-203.
- Roos, A., Bouwman, L. H., van Gijlswijk-Janssen, D. J., Faber-Krol, M. C., Stahl, G. L., Daha, M. R., 2001. Human IgA activates the complement system via the mannan-binding lectin pathway. *Journal of Immunology*, **167**, 2861-2868.
- Rosenberg, M. E. & Silkens, J., 1995. Clusterin: physiologic and pathophysiologic considerations. *Int J Biochem Cell Biol*, **27**, 633-645.
- Rossi, V., Cseh, S., Bally, I., Thielens, N. M., Jensenius, J. C., Arlaud, G. J., 2001. Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. *J Biol Chem*, **276**, 40880-40887.

- Rossi, V., Teillet, F., Thielens, N. M., Bally, I., Arlaud, G. J., 2005. Functional characterization of complement proteases C1s/mannan-binding lectin-associated serine protease-2 (MASP-2) chimeras reveals the higher C4 recognition efficacy of the MASP-2 complement control protein modules. *J Biol Chem*, **280**, 41811-41818.
- Saifuddin, M., Hart, M. L., Gewurz, H., Zhang, Y., Spear, G. T., 2000. Interaction of mannose-binding lectin with primary isolates of human immunodeficiency virus type 1. *J Gen Virol*, **81**, 949-955.
- Sastry, K., Herman, G. A., Day, L., Deignan, E., Bruns, G., Morton, C. C., Ezekowitz, R. A., 1989. The human mannose-binding protein gene. Exon structure reveals its evolutionary relationship to a human pulmonary surfactant gene and localization to chromosome 10. *J Exp Med*, **170**, 1175-1189.
- Sato, T., Endo, Y., Matsushita, M., Fujita, T., 1994. Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Int Immunol*, **6**, 665-669.
- Schwaeble, W., Dahl, M. R., Thiel, S., Stover, C., Jensenius, J. C., 2002. The mannan-binding lectin-associated serine proteases (MASPs) and MAp19: four components of the lectin pathway activation complex encoded by two genes. *Immunobiology*, **205**, 455-466.
- Schwaeble, W. J., Lynch, N. J., Clark, J. E., Marber, M., Samani, N. J., Ali, Y. M., Dudler, T., Parent, B., Lhotta, K., Wallis, R., Farrar, C. A., Sacks, S., Lee, H., Zhang, M., Iwaki, D., Takahashi, M., Fujita, T., Tedford, C. E., Stover, C. M., 2011. Targeting of mannan-binding lectin-associated serine protease-2 confers protection from myocardial and gastrointestinal ischemia/reperfusion injury. *Proc Natl Acad Sci U S A*, **108**, 7523-7528.
- Schwaeble, W. J. & Reid, K. B., 1999. Does properdin crosslink the cellular and the humoral immune response? *Immunol Today*, **20**, 17-21.
- Selman, L. & Hansen, S., 2012a. Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). *Immunobiology*, **217**, 851-863.
- Selman, L., Henriksen, M. L., Brandt, J., Palarasah, Y., Waters, A., Beales, P. L., Holmskov, U., Jorgensen, T. J., Nielsen, C., Skjodt, K., Hansen, S., 2012b. An enzyme-linked immunosorbent assay (ELISA) for quantification of human collectin 11 (CL-11, CL-K1). *J Immunol Methods*, **375**, 182-188.

- Shang, F., Rynkiewicz, M. J., McCormack, F. X., Wu, H., Cafarella, T. M., Head, J. F., Seaton, B. A., 2011. Crystallographic complexes of surfactant protein A and carbohydrates reveal ligand-induced conformational change. *J Biol Chem*, **286**, 757-765.
- Sheriff, S., Chang, C. Y., Ezekowitz, R. A., 1994. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat Struct Biol*, **1**, 789-794.
- Sim, R. B., 1985. Large-scale isolation of complement receptor type 1 (CR1) from human erythrocytes. Proteolytic fragmentation studies. *Biochem J*, **232**, 883-889.
- Sim, R. B., Arlaud, G. J., Colomb, M. G., 1979. C1 inhibitor-dependent dissociation of human complement component C1 bound to immune complexes. *Biochem J*, **179**, 449-457.
- Sim, R. B. & Malhotra, R., 1994. Interactions of carbohydrates and lectins with complement. *Biochem Soc Trans*, **22**, 106-111.
- Sim, R. B. & Reid, K. B., 1991. C1: molecular interactions with activating systems. *Immunology Today*, **12**, 307-311.
- Sim, R. B. & Tsiftoglou, S. A., 2004. Proteases of the complement system. *Biochem Soc Trans*, **32**, 21-27.
- Spear, G. T., Zariffard, M. R., Xin, J., Saifuddin, M., 2003. Inhibition of DC-SIGN-mediated trans infection of T cells by mannose-binding lectin. *Immunology*, **110**, 80-85.
- Stover, C., Endo, Y., Takahashi, M., Lynch, N. J., Constantinescu, C., Vorup-Jensen, T., Thiel, S., Friedl, H., Hankeln, T., Hall, R., Gregory, S., Fujita, T., Schwaeble, W., 2001. The human gene for mannan-binding lectin-associated serine protease-2 (MASP-2), the effector component of the lectin route of complement activation, is part of a tightly linked gene cluster on chromosome 1p36.2-3. *Genes Immun*, **2**, 119-127.
- Stover, C. M., Lynch, N. J., Hanson, S. J., Windbichler, M., Gregory, S. G., Schwaeble, W. J., 2004. Organization of the MASP2 locus and its expression profile in mouse and rat. *Mamm Genome*, **15**, 887-900.

- Sumiya, M., Super, M., Tabona, P., Levinsky, R. J., Arai, T., Turner, M. W., Summerfield, J. A., 1991. Molecular basis of opsonic defect in immunodeficient children. *Lancet*, **337**, 1569-1570.
- Super, M., Thiel, S., Lu, J., Levinsky, R. J., Turner, M. W., 1989. Association of low levels of mannan-binding protein with a common defect of opsonisation. *Lancet*, **2**, 1236-1239.
- Takahashi, M., Endo, Y., Fujita, T., Matsushita, M., 1999. A truncated form of mannose-binding lectin-associated serine protease (MASP)-2 expressed by alternative polyadenylation is a component of the lectin complement pathway. *Int Immunol*, **11**, 859-863.
- Takahashi, M., Ishida, Y., Iwaki, D., Kanno, K., Suzuki, T., Endo, Y., Homma, Y., Fujita, T., 2010. Essential role of mannose-binding lectin-associated serine protease-1 in activation of the complement factor D. *J Exp Med*, **207**, 29-37.
- Taylor, M. E. & Drickamer, K., 1993. Structural requirements for high affinity binding of complex ligands by the macrophage mannose receptor. *J Biol Chem*, **268**, 399-404.
- Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B., Poulsen, K., Willis, A. C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K. B., Jensenius, J. C., 1997. A second serine protease associated with mannan-binding lectin that activates complement. *Nature*, **386**, 506-510.
- 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. *J Immunol*, **166**, 5068-5077.
- Titomanlio, L., Bennaceur, S., Bremond-Gignac, D., Baumann, C., Dupuy, O., Verloes, A., 2005. Michels syndrome, Carnevale syndrome, OSA syndrome, and Malpuech syndrome: variable expression of a single disorder (3MC syndrome)? *American Journal of Medical Genetics, Part A*. **137A**, 332-335.
- Tschopp, J., Chonn, A., Hertig, S., French, L. E., 1993. Clusterin, the human apolipoprotein and complement inhibitor, binds to complement C7, C8 beta, and the b domain of C9. *J Immunol*, **151**, 2159-2165.
- Turner, M. W., Mowbray, J. F., Robertson, D. R., 1981. A study of C3b deposition on yeast surfaces by sera of known opsonic potential. *Clin Exp Immunol*, **46**, 412-419.

- Uemura, T., Sano, H., Katoh, T., Nishitani, C., Mitsuzawa, H., Shimizu, T., Kuroki, Y., 2006. Surfactant protein A without the interruption of Gly-X-Y repeats loses a kink of oligomeric structure and exhibits impaired phospholipid liposome aggregation ability. *Biochemistry*, **45**, 14543-14551.
- Urlaub, G. & Chasin, L. A., 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 4216-4220.
- van de Wetering, J. K., van Golde, L. M., Batenburg, J. J., 2004. Collectins: players of the innate immune system. *Eur J Biochem*, **271**, 1229-1249.
- Wallis, R., 2007. Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway. *Immunobiology*, **212**, 289-299.
- Wallis, R. & Cheng, J. Y., 1999a. Molecular defects in variant forms of mannose-binding protein associated with immunodeficiency. *J Immunol*, **163**, 4953-4959.
- Wallis, R. & Dodd, R. B., 2000. Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations. *J Biol Chem*, **275**, 30962-30969.
- Wallis, R., Dodds, A. W., Mitchell, D. A., Sim, R. B., Reid, K. B. M., 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. & Drickamer, K., 1997. Asymmetry adjacent to the collagen-like domain in rat liver mannose-binding protein. *Biochem J*, **325**, 391-400.
- Wallis, R. & Drickamer, K., 1999b. Molecular determinants of oligomer formation and complement fixation in mannose-binding proteins. *J Biol Chem*, **274**, 3580-3589.
- Wallis, R., Lynch, N. J., Roscher, S., Reid, K. B., Schwaeble, W. J., 2005. Decoupling of carbohydrate binding and MASP-2 autoactivation in variant mannose-binding lectins associated with immunodeficiency. *J Immunol*, **175**, 6846-6851.
- 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.

- Wallis, R., Shaw, J. M., Uitdehaag, J., Chen, C.-B., Torgersen, D., Drickamer, K., 2004. Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein: indirect effects of naturally occurring mutations on protease binding and activation. *Journal of Biological Chemistry*, **279**, 14065-14073.
- Wang, Y., Kuan, P. J., Xing, C., Cronkhite, J. T., Torres, F., Rosenblatt, R. L., DiMaio, J. M., Kinch, L. N., Grishin, N. V., Garcia, C. K., 2009. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am J Hum Genet*, **84**, 52-59.
- Weis, W. I. & Drickamer, K., 1994. Trimeric structure of a C-type mannose-binding protein. *Structure*, **2**, 1227-1240.
- Weis, W. I., Drickamer, K., Hendrickson, W. A., 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127-134.
- Weis, W. I., Taylor, M. E., Drickamer, K., 1998. The C-type lectin superfamily in the immune system. *Immunol Rev*, **163**, 19-34.
- Whaley, K. & Ahmed, A. E., 1989. Control of immune complexes by the classical pathway. *Behring Inst Mitt*, **84**, 111-120.
- White, M. R., Crouch, E., Chang, D., Sastry, K., Guo, N., Engelich, G., Takahashi, K., Ezekowitz, R. A., Hartshorn, K. L., 2000. Enhanced antiviral and opsonic activity of a human mannose-binding lectin and surfactant protein D chimera. *J Immunol*, **165**, 2108-2115.
- White, R. A., Dowler, L. L., Adkison, L. R., Ezekowitz, R. A., Sastry, K. N., 1994. The murine mannose-binding protein genes (Mbl 1 and Mbl 2) localize to chromosomes 14 and 19. *Mamm Genome*, **5**, 807-809.
- White, R. T., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B., Cordell, B., 1985. Isolation and characterization of the human pulmonary surfactant apoprotein gene. *Nature*, **317**, 361-363.
- Ying, H., Ji, X., Hart, M. L., Gupta, K., Saifuddin, M., Zariffard, M. R., Spear, G. T., 2004. Interaction of mannose-binding lectin with HIV type 1 is sufficient for virus opsonization but not neutralization. *AIDS Res Hum Retroviruses*, **20**, 327-335.

- Yokota, Y., Arai, T., Kawasaki, T., 1995. Oligomeric structures required for complement activation of serum mannan-binding proteins. *J Biochem*, **117**, 414-419.
- Zlotnik, A. & Yoshie, O., 2000. Chemokines: a new classification system and their role in immunity. *Immunity*, **12**, 121-127.
- Zundel, S., Cseh, S., Lacroix, M., Dahl, M. R., Matsushita, M., Andrieu, J. P., Schwaeble, W. J., Jensenius, J. C., Fujita, T., Arlaud, G. J., Thielens, N. M., 2004. Characterization of recombinant mannan-binding lectin-associated serine protease (MASP)-3 suggests an activation mechanism different from that of MASP-1 and MASP-2. *J Immunol*, **172**, 4342-4350.

Please find overleaf a reprint of a peer reviewed publication that reports parts of my work described in this PhD thesis.

Engineering Novel Complement Activity into a Pulmonary Surfactant Protein*

Received for publication, December 21, 2009, and in revised form, January 21, 2010. Published, JBC Papers in Press, January 29, 2010, DOI 10.1074/jbc.M109.097493

Umakhanth Venkatraman Girija^{†1}, Christopher Furze^{†1}, Julia Toth[‡], Wilhelm J. Schwaeble[‡], Daniel A. Mitchell^{§2}, Anthony H. Keeble^{‡3}, and Russell Wallis^{†¶4}

From the Departments of [†]Infection, Immunity, and Inflammation and [¶]Biochemistry, University of Leicester, Leicester LE1 9HN and the [§]Clinical Sciences Research Institute, University of Warwick, Coventry CV4 7AL, United Kingdom

Complement neutralizes invading pathogens, stimulates inflammatory and adaptive immune responses, and targets non- or altered-self structures for clearance. In the classical and lectin activation pathways, it is initiated when complexes composed of separate recognition and activation subcomponents bind to a pathogen surface. Despite its apparent complexity, recognition-mediated activation has evolved independently in three separate protein families, C1q, mannose-binding lectins (MBLs), and serum ficolins. Although unrelated, all have bouquet-like architectures and associate with complement-specific serine proteases: MBLs and ficolins with MBL-associated serine protease-2 (MASP-2) and C1q with C1r and C1s. To examine the structural requirements for complement activation, we have created a number of novel recombinant rat MBLs in which the position and orientation of the MASP-binding sites have been changed. We have also engineered MASP binding into a pulmonary surfactant protein (SP-A), which has the same domain structure and architecture as MBL but lacks any intrinsic complement activity. The data reveal that complement activity is remarkably tolerant to changes in the size and orientation of the collagenous stalks of MBL, implying considerable rotational and conformational flexibility in unbound MBL. Furthermore, novel complement activity is introduced concurrently with MASP binding in SP-A but is uncontrolled and occurs even in the absence of a carbohydrate target. Thus, the active rather than the zymogen state is default in lectin-MASP complexes and must be inhibited through additional regions in circulating MBLs until triggered by pathogen recognition.

Complement is a central part of the immune system that neutralizes pathogens via antibody-dependent and -independent mechanisms and stimulates a variety of protective responses including phagocytosis, inflammation, and adaptive immunity (1). It is triggered via three routes called the classical, lectin, and alternative pathways. The classical and lectin path-

ways both selectively target pathogen-associated molecular patterns via circulating complexes composed of recognition and zymogen protease subcomponents (2). Upon binding to a target, conformational changes in the recognition subcomponent trigger activation of the protease subcomponent, which in turn activates the downstream complement cascade. In the classical pathway, C1q binds to microorganisms, immune complexes, apoptotic and necrotic cells as well as amyloids to initiate the stepwise activation of C1r and C1s. In the lectin pathway, mannose-binding lectin (MBL)⁵ and serum ficolins bind to terminal mannose-like epitopes or *N*-acetyl groups on pathogens to activate MBL-associated serine protease-2 (MASP-2).

Initiating complexes of the classical and lectin pathways have a number of properties in common (2). Although unrelated, recognition subcomponents have similar domain organizations consisting of a short cysteine-containing domain followed by a collagen-like domain linked to a C-terminal recognition domain (Fig. 1) (3). During biosynthesis, polypeptides associate into trimeric subunits (4), which in turn assemble to form larger oligomers resembling bouquets (5, 6). In C1q, the collagenous stalks associate with each other through their N-terminal portions and splay apart at a short interruption within the repeating Gly-Xaa-Yaa sequence, called the kink. Although MBL and some ficolins also possess a kink-like region, recent measurements have revealed that the stalks probably diverge nearer the N terminus, at the junction with the cysteine-containing domain, and thus form spider-like structures rather than the classical bouquets of C1q (7). Recognition subcomponents possess different numbers of subunits. For example, human C1q is a hexamer assembled from three different polypeptide chains (8), and rat MBLs are homo-oligomers comprising dimers, trimers, and tetramers of subunits (9).

The protease subcomponents, C1r and C1s of the classical pathway and MASP-2 of the lectin pathway, are homologous and bind to the collagenous domains of their respective recognition subcomponent through Ca²⁺-dependent interactions (10–13). They each comprise two Ca²⁺-binding CUB domains (domain found in complement component C1r/C1s, Uegf, and bone morphogenic protein 1), separated by a Ca²⁺-binding epidermal growth factor-like domain (EGF), followed by two complement-control protein modules and a serine protease domain

* This work was supported by Grant G0501425 from the Medical Research Council and Grant 077400 from The Wellcome Trust.

✂ Author's Choice—Final version full access.

¹ Both authors contributed equally to this work.

² Research Council United Kingdom Academic Fellow.

³ Wellcome Trust Value in People Award Fellow.

⁴ Research Council United Kingdom Academic Fellow. To whom correspondence should be addressed: Dept. of Infection, Immunity, and Inflammation, Maurice Shock Bldg., University of Leicester. P. O. Box 138, Leicester LE1 9HN, United Kingdom. Tel.: 44 (0)116 252 5089; Fax: 44 (0)116 252 5030; E-mail: rw73@le.ac.uk.

⁵ The abbreviations used are: MBL, mannose- or mannan-binding lectin; MASP, MBL-associated serine protease; CUB, domain found in complement component C1r/C1s, Uegf, and bone morphogenic protein 1; EGF, epidermal growth factor; SP-A, pulmonary surfactant protein-A; Hyp, hydroxyproline.

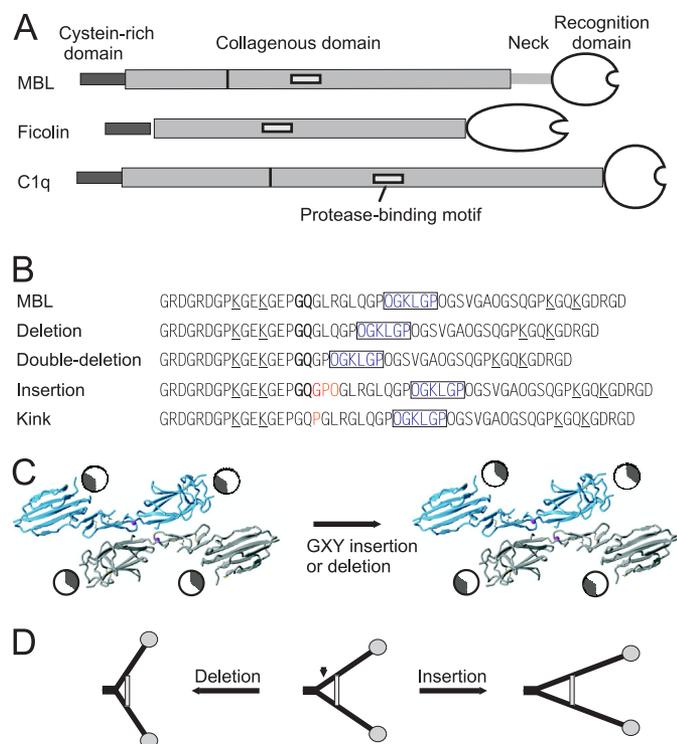


FIGURE 1. Changes to the collagenous domains of MBLs. *A*, domain organizations of MBL, ficolins, and C1q. Interruptions within the collagenous domains of C1q and MBL are marked by a **bold line**. The binding sites for MASPs and C1r/C1s within the collagenous domains of MBL, ficolins, and C1q are indicated by **light gray boxes**. The recognition domains are: carbohydrate-recognition domain in MBL, fibrinogen-like domain in ficolins, and C1q-globular domain in C1q. *B*, aligned sequences of the collagenous domains of wild-type and modified MBLs. Residues forming the protease-binding motif are in **blue** and are **boxed**. The GQ residues that form the kink-like region are in **bold**. Residues inserted into the collagenous domain are shown in **red**. Glycosylated hydroxylysine residues are **underlined**. *C*, predicted changes to the alignment of the MASP-binding sites on MBL subunits as a result of a single insertion or deletion of a Gly-Xaa-Yaa into the collagenous domain based upon a pitch of 3.5 for a collagen helix. The figure shows a cross-section through the four collagenous stalks, which are represented by **circles**. Each binding site is shaded in **gray**. The structure of the CUB1-EGF-CUB2 domains of MASP-2 is from Ref. 14. *D*, schematic representation of the effects of changes to the position of the MASP-binding sites on MBL. Only two MBL stalks are shown for clarity. The MASP is shown as a **box** and is shaded in **light gray**. Carbohydrate recognition domains are represented as **gray circles**. The **arrow** shows the position of the insertion/deletion. To accommodate the MASP, the collagenous stalks would have to move closer together for an insertion or splay further apart for a deletion.

(14, 15). C1r and C1s assemble to form heterotetramers (C1s-C1r-C1r-C1s), and MASPs are homodimers. Although the stoichiometries of the resulting C1 and MBL·MASP complexes differ, interactions between subcomponents are analogous, with equivalent contacts between the CUB domains of the proteases and the collagenous domains of the recognition subcomponents (2). Overall, each C1r tetramer presents a total of six binding sites, one for each of the collagenous domains of C1q (one site on each CUB1 of C1r and C1s and each CUB2 of C1r). Each MASP-2 dimer presents four binding sites to MBL and ficolins (one site on each CUB domain).

A number of other proteins have bouquet-like architectures, such as adiponectin (16), emilins (17), and some collectins (18), including pulmonary surfactant protein-A (SP-A). Nevertheless, only MBLs, ficolins, and C1q molecules are able to activate complement. As yet, the only established difference between

these proteins and their structural analogues is their ability to bind to MASPs or C1r tetramers. The MASP-binding site in MBLs and ficolins is characterized by a distinct motif within the collagenous domain: Hyp-Gly-Lys-Xaa-Gly-Pro, where Hyp is hydroxyproline and Xaa is generally an aliphatic residue. Point mutations in this sequence disrupt MASP binding (10, 13). C1q also possesses a similar motif, and we have proposed that this region forms the binding sites for C1r and C1s (2).

In this paper, we have further probed the functions of the recognition subcomponents by examining the structural requirements for complement activation. Initially, we modified the collagenous domain of MBL to change the position and relative orientations of the MASP-binding site on the collagenous stalks. The data reveal that MASP binding and activation are surprisingly tolerant to such changes, implying considerable rotational and conformational flexibility. To take these studies further, we have engineered the MASP-binding motif into SP-A, which also has a C-type lectin domain and an architecture similar to that of MBL but cannot bind MASPs or activate complement. We find that MASP binding, comparable with that by MBL, can be engineered through just three amino acid substitutions to the collagenous domain of SP-A. The resulting proteins also activate the MASP but lack the control mechanisms necessary to target carbohydrate surfaces selectively.

EXPERIMENTAL PROCEDURES

Protein Components—Recombinant rat MBL and modified forms of rat MASP-2 were produced in a Chinese hamster ovary cell expression system and purified as described previously by affinity chromatography on mannose-Sepharose and nickel-Sepharose columns, respectively (9, 19). Recombinant wild-type rat MASP-2 is toxic to producing Chinese hamster ovary cells and autoactivates during biosynthesis. We therefore used two modified forms of MASP-2, both of which are secreted as zymogens and have been characterized extensively with regard to their structures, activation, and catalytic properties (19, 20). MASP-2A is a catalytically inactive form in which the active site serine at position 613 has been replaced by an alanine. MASP-2K is a catalytically active form, in which the arginine residue at the cleavage site for zymogen activation (Arg⁴²⁴) has been changed to a lysine residue to slow down the rate of spontaneous autoactivation during biosynthesis. Binding and activation assays were carried out using MASP-2A and MASP-2K, respectively. Mutant forms of MBL were created by PCR and expressed in the same way as the wild-type protein. All mutations were verified by DNA sequencing of the entire cDNA within the expression vector, prior to transfection of the Chinese hamster ovary cells.

Cloning and Production of Recombinant Wild-type and Mutant SP-As—The cDNA of SP-A isoform A1 was cloned into the polylinker of the eukaryotic expression vector pED-4 (21), which contains the dihydrofolate reductase cDNA as a selectable marker. Proteins were produced by expression in the Chinese hamster ovary cell line DXB11 following amplification using the dihydrofolate reductase inhibitor, methotrexate, to a final concentration of 25 μ M. Culture medium was harvested using a protocol that has been described previously for produc-

Engineering Complement Activity into SP-A

tion of recombinant MBL (9, 22). Protein was purified by affinity chromatography on mannose-Sepharose columns (1-ml column for 250 ml of medium). Following washes with high salt (10 mM Tris-HCl (pH 7.4), containing 500 mM NaCl and 5 mM CaCl_2), Triton (10 mM Tris-HCl (pH 7.5), containing 0.1% Triton X-100 and 5 mM CaCl_2), and low salt buffers (10 mM Tris-HCl (pH 7.4), containing 5 mM CaCl_2), protein was eluted in 0.5-ml fractions of 10 mM Tris-HCl (pH 7.4), containing 5 mM EDTA. Wild-type SP-A and all mutants bound to the affinity matrix, confirming that they were folded correctly. Fractions containing recombinant protein were identified by SDS-PAGE. Mutations were introduced into the cDNA of SP-A by PCR, and the encoded proteins were produced in the same way as the wild-type protein. Wild-type and mutant SP-As were stored in low ionic strength buffer (10 mM Tris (pH 7.4), containing 2 mM CaCl_2) because they tended to aggregate at high protein concentrations (>0.1 mg/ml) in salt concentrations >10 mM.

Surface Plasmon Resonance—Measurements were performed using a BIAcore 2000 instrument (GE Healthcare) or a Bio-Rad ProteOn XPR36 biosensor. Protein ligands were diluted into 10 mM sodium acetate (pH 4.5 for MBL or pH 5.0 for SP-A) and immobilized onto the carboxymethylated dextran surface of a CM5 sensor chip (GE Healthcare) or a GLM chip (Bio-Rad), using amine coupling chemistry. Binding was measured in 10 mM Tris-HCl (pH 7.4), containing 140 mM NaCl, 2 mM CaCl_2 , and 0.005% surfactant SP40, at a flow rate of 25 $\mu\text{l}/\text{min}$ and at 25 °C. After injection of ligand, the protein surface was regenerated by injection of 10 μl of 10 mM Tris-HCl buffer (pH 7.4), containing 1 M NaCl and 5 mM EDTA. Data were analyzed by fitting association and dissociation curves to Langmuir binding models for several protein concentrations simultaneously, using BIAevaluation 4.1 software (GE Healthcare). Increasingly complex models were tested until a satisfactory fit to the data was achieved. Apparent equilibrium dissociation constants (K_D) were calculated from the ratio of the dissociation and association rate constants ($k_{\text{off}}/k_{\text{on}}$). MBL was immobilized on the chip surface rather than used as a soluble ligand because it bound to the chip, thereby masking analysis of the protein-protein interactions. SP-A was also immobilized because it tended to precipitate on the chip surface.

MASP-2 Activation Assays—Activation was measured by following MASP autolysis using a modified version of the protocol described previously (20). Briefly, MASP-2K was mixed with wild-type or mutant MBL or SP-A and added to a suspension of mannose- or fucose-Sepharose (5 μl of a 1:1 v/v suspension in a total volume of 30 μl) in 50 mM Tris-HCl (pH 7.5), containing 150 mM NaCl and 5 mM CaCl_2 , at 37 °C with mixing. A 1.2-fold molar excess of MBL or SP-A was used to ensure that all of the MASP-2K was bound to the recognition molecule. The mixture was incubated at 37 °C with shaking, and aliquots of the suspension were removed from the reaction mix at various times and immediately frozen on dry ice to quench the reaction. Proteins were separated by SDS-PAGE under reducing conditions, and the amount of MASP cleaved was quantified by densitometry. Data are the mean \pm S.E. from at least two separate experiments using different protein preparations, unless otherwise stated.

RESULTS

MASP Binding and Activation Are Tolerant of Changes in the Position and Orientation of Binding Sites on MBL—MASPs bind inside the cone created by the collagenous stalks of MBL and bridge up to four stalks simultaneously (Fig. 1) (3). Consequently, during complex formation, the binding sites on the MBL stalks must either face inward or reorient so that the correct collagen-CUB domain interactions are formed. If the stalks are relatively rigid or are unable to reorient independently, any change in their position or alignment would lead to loss of function. To examine the tolerance of MBL to such changes, we made a series of modified MBLs in which the collagenous domain was modified. Because the pitch of collagen is typically ~ 3.5 residues (24), deletion or insertion of a Gly-Xaa-Yaa triplet would change both the position of each binding site and its relative orientation. For example, deletion of a triplet would effectively rotate the orientation of the binding site on the collagen by $\sim 180^\circ$ as well as shift the position of further toward the N terminus (Fig. 1). Deletion of a second triplet would restore the original alignment but move the binding sites even further toward the N terminus. To retain the ability to bind MASP, the collagenous stalks would not only have to splay further apart (for a deletion) or move closer together (for an insertion), but also rotate independently of each other. Thus, analysis of MASP binding and activation by the modified MBLs provides a useful probe of the conformational and rotational flexibility of MBL.

Two deletion mutants were created in which either one or two Gly-Xaa-Yaa triplets were removed from the collagenous domain, thus moving the MASP-binding motif closer to the N terminus of MBL (Fig. 1). In addition, a single Gly-Pro-Hyp was inserted near the kink-like region to move the MASP-binding motif further toward the C terminus. We also made an MBL in which the kink-like region itself was removed by insertion of a proline residue to restore the Gly-Xaa-Yaa tandem repeats. A previous study has demonstrated that the kink-like region is not essential for complement activation by MBL (25). However, the kinetics of activation were not investigated in this work, so more subtle changes might have been missed. Analysis of the purified, recombinant MBLs by SDS-PAGE under nonreducing conditions, shown previously to be a sensitive indicator of MBL assembly (26), demonstrated that all assembled correctly during biosynthesis (Fig. 2).

Initially, MBLs were tested for their abilities to bind to MASP-2A, using surface plasmon resonance. As expected, wild-type bound with complex kinetics compatible with the formation of 1:1 complexes, together with binding of a second MASP at relatively high protein concentrations (19). The K_D values were 3.7 and 164.5 nM, consistent with previous measurements (2). Surprisingly, the entire set of mutant MBLs bound to MASP-2A with similar affinities and kinetics (Table 1), indicating that none of the changes perturbed MASP binding (Fig. 2). We next examined their complement activities by monitoring MBL-dependent MASP-2K autoactivation with mannose-Sepharose as an activating target (Fig. 3). In the presence of the target, all mutant MBLs activated the MASP with rates comparable to that of wild-type ($t_{1/2} \sim 20$ min). The greatest difference observed was for the double-deletion mutant, which activated

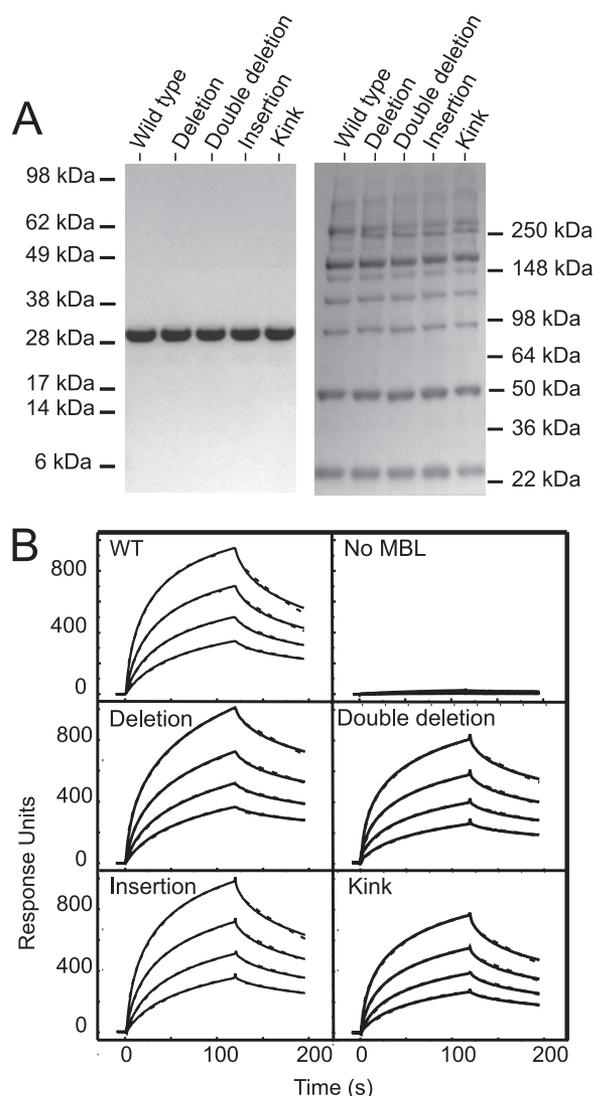


FIGURE 2. MASP-binding by modified MBLs. *A*, SDS-PAGE of wild-type and mutant MBLs under reducing (*left*) and nonreducing (*right*) conditions. Mutant MBLs are assembled from single polypeptide chains, which migrate as a ladder of covalently linked polypeptides under nonreducing conditions, characteristic of the heterogeneous nature of oligomers of wild-type MBL (9). *B*, binding of MASP-2A to immobilized MBLs by surface plasmon resonance. MASP-2A was injected at 333, 167, 83, and 42 nm over immobilized MBL (~7000 response units). A sensor chip blocked by treatment with ethanolamine was used as a negative control (*No MBL*). All data were fitted to a two-complex parallel binding model, and the fits are shown by the *dotted lines*. *WT*, wild-type.

TABLE 1
Kinetic properties of modified MBLs

MBL modification	k_{on} ($\times 10^{-4}$)	k_{off} ($\times 10^3$)	K_D^a	Half-time for
				MBL activation
	$M^{-1} s^{-1}$	s^{-1}	<i>nm</i>	<i>Min</i>
Wild-type	40 ± 1	1.5 ± 0.1	3.7 ± 0.1	20 ± 1
	2.9 ± 0.1	4.75 ± 0.1	165 ± 2	
Deletion	38 ± 4	3.8 ± 1.7	11 ± 6	22 ± 1
	3.1 ± 0.7	5.4 ± 1.3	175 ± 2	
Double deletion	39 ± 1	3.1 ± 0.1	8.2 ± 0.3	39 ± 3
	3.3 ± 0	6.4 ± 0.1	195 ± 3	
Insertion	44 ± 3	2.2 ± 0.6	4.9 ± 1.2	21 ± 1
	4.6 ± 0	9.1 ± 0.2	199 ± 5	
Kink	57 ± 1	6.2 ± 2.7	10 ± 2	21 ± 1
	5.2 ± 0.2	6.3 ± 0.2	136 ± 9	

^a K_D values were calculated from k_{off}/k_{on} for each experiment and were averaged from two separate experiments.

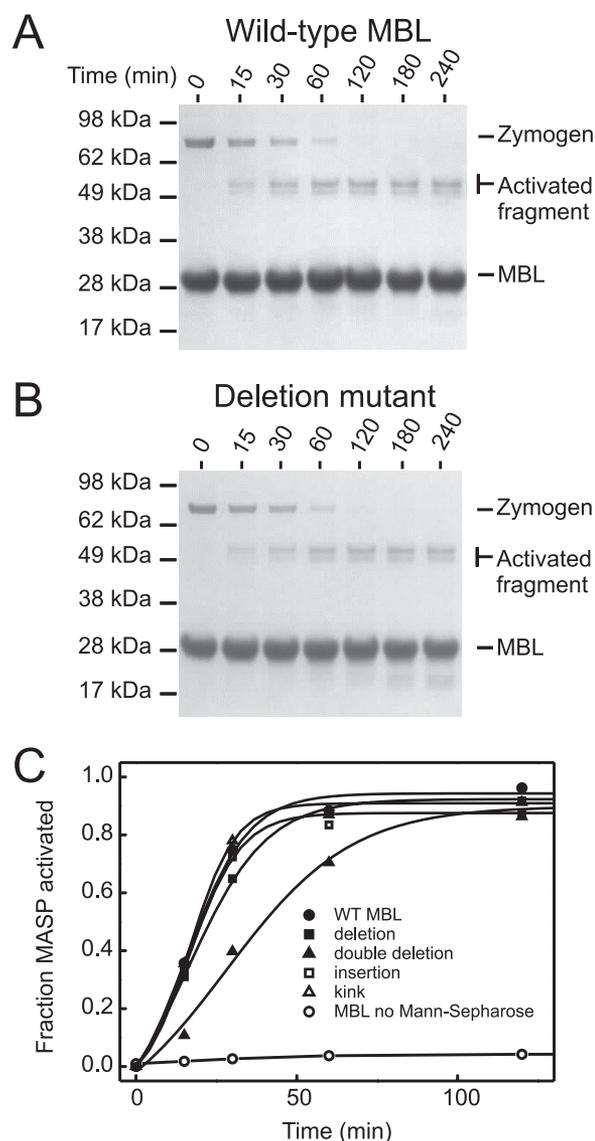


FIGURE 3. MASP activation by modified MBLs. Kinetics of MBL-MASP-2 activation analyzed by SDS-PAGE. *A* and *B*, wild-type MBL and the deletion mutant were incubated with MASP-2K together with mannose-Sepharose as an activating target. Proteins were separated on a 4–12% linear gradient gel under reducing conditions and were stained with Coomassie Blue. The N-terminal fragment of MASP-2K runs as a double band due to differential glycosylation. MASP activation was measured by quantifying cleavage of the MASP polypeptide. The other mutants were tested in the same way (data not shown). *C*, comparison of MASP-2K activation by wild-type (*WT*) and mutant MBLs. Data for all mutants are shown.

the MASP just 2-fold more slowly ($t_{1/2} \sim 40$ min), despite removal of six amino acid residues from the collagenous domain. In the absence of the target, activation was relatively slow and similar to autoactivation by MASP-2 alone (>1000 min), indicating that all MBLs activate only on a carbohydrate target. Thus, neither insertions, deletions, nor removal of the kink-like region disrupts MASP-2 binding or activation, despite large changes to the position and relative orientations of the MASP-binding sites. The tolerance of MBL to these changes must reflect significant adjustment and realignment of the collagenous stalks. Given that collagenous domains themselves are relatively rigid and inflexible (27), changes must occur at the N terminus, where the stalks converge, probably at

Engineering Complement Activity into SP-A

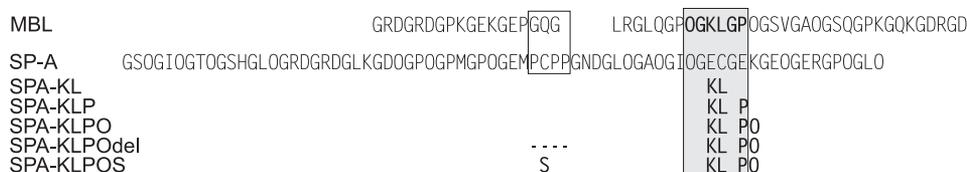


FIGURE 4. **Design of modified SP-As.** Sequences of the collagenous domains of MBL (top) and SP-A (bottom) Sequences were crudely aligned based on the positions of the kink of SP-A and the kink-like region of MBL (boxed). The MASP-binding motif is shaded gray. Changes to the sequence of SP-A are indicated.

chromatography and assessment of complement activity using similar activation assays. Although structurally analogous to MBL, the collagenous domains of SP-A and MBL are of different lengths and share no apparent sequence identity (other than the repeating Gly-Xaa-Yaa motif characteristic of all collagens)

(Fig. 4). SP-A possesses a natural kink near the middle of the collagenous domain where the collagenous stalks splay apart, similar to C1q (30). Previous studies have shown that upon removal of the kink, the stalks diverge nearer the N terminus of polypeptides, at the junction between the N-terminal domain and the collagenous domain, thus becoming more like MBL and ficolins (30).

Three new recombinant constructs were created in which all or part of the MASP-binding motif was introduced into the collagenous domain of SP-A (Fig. 4). To allow enough space for the MASP to bind within the SP-A bouquet, changes were made five Gly-Xaa-Yaa triplets toward the C-terminal side of the kink. In SPA-KL, Glu⁶³ and Cys⁶⁴ were replaced by lysine and leucine residues, respectively. Lysine in the Xaa position is known to be essential for MASP binding and complement activation in both MBLs and ficolins. The adjacent cysteine residue was also replaced to avoid any potential disulfide bond formation that might interfere with MASP binding. In SPA-KLP, a proline residue was introduced in place of the glutamate at position 66 to optimize the MASP-binding motif further: Hyp-Gly-Lys-Leu-Gly-Pro. One further change was made in SPA-KLPO, replacing Lys⁶⁷ by a hydroxyproline residue. Lysine residues in the Yaa position of collagen are often hydroxylated and glycosylated during biosynthesis, so this change was designed to prevent any potential steric inhibition of MASP binding by the sugar residues (24). We created two further variants of SPA-KLPO, in which the kink was removed or modified to change the point at which the stalks splay apart (30) and thus more resemble MBL and ficolins. In one mutant, SPA-KLPOdel the kink sequence (PCPP) was removed completely, and in the other, called SPA-KLPOS, the cysteine residue was replaced by a serine residue to remove the potential for disulfide bond formation, which is believed to tether the separate stalks together in native SP-A (30).

Following production in Chinese hamster ovary cells, all proteins bound to mannose- or fucose-Sepharose columns, demonstrating that they were folded correctly. As expected, wild-type SP-A did not bind to MASP-2A at any of the concentrations tested (up to 1 μ M). Surprisingly, however, SPA-KL bound with appreciable affinity (Fig. 5). The data best fitted a two-complex, parallel-reaction model, with apparent dissociation constants K_{D1} and K_{D2} of 600 and 2120 nM, respectively, compared with 3.8 and 166 nM for MBL·MASP (Table 2). SPA-KLP bound with even higher affinity, with K_D values of 36 and 104 nM, only slightly weaker than MBL. Additional changes had little effect on MASP binding, so the naturally occurring lysine residue in SP-A (or its potential glycosylated derivative) does not prevent MASP access. Removal of the kink region also had little effect, indicating that the point at which the stalks splay

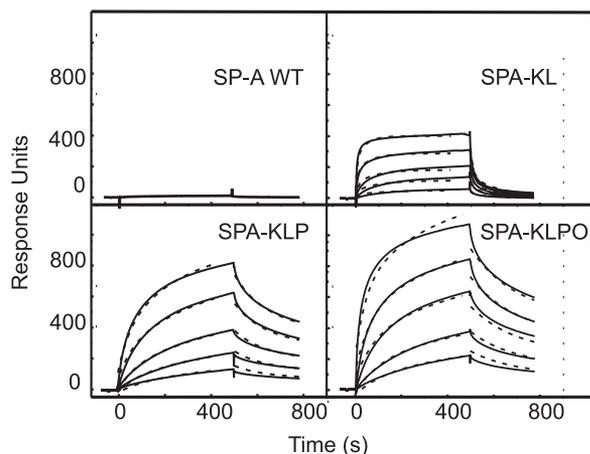


FIGURE 5. **MASP-binding by modified SP-As.** Binding of MASP-2A to immobilized SP-As by surface plasmon resonance. MASP-2A was injected at 764, 447, 261, 152, and 89 nM over each of the immobilized SP-As (~6000 response units). All data were fitted to a two-complex parallel binding model, and the fits are shown by the dotted lines.

TABLE 2
Kinetic properties of engineered SP-As

Engineered SP-A	k_{on} ($\times 10^{-4}$)	k_{off} ($\times 10^3$)	K_D^a	Half-time for MASP-2K activation ^b
	$M^{-1} s^{-1}$	s^{-1}	nM	Min
Wild-type	N.B. ^c	N.B. ^c	N.B. ^c	>1000 ^d
SPA-KL	1.2 \pm 0.2	7.0 \pm 0.1	600 \pm 110	>1000 ^d
	4.3 \pm 0.1	91 \pm 2	2120 \pm 100	
SPA-KLP	5.4 \pm 2.9	1.4 \pm 0.1	36 \pm 20	270 \pm 56
	36 \pm 24	17.6 \pm 1.0	104 \pm 77	
SPA-KLPO	2.6 \pm 0.8	2.5 \pm 0.6	97 \pm 7	256 \pm 96
	9.6 \pm 1.0	22 \pm 3.1	232 \pm 9	
SPA-KLPOdel	5.3 \pm 2.3	1.7 \pm 0.1	29 \pm 17	316 \pm 171
	18 \pm 0.0	21 \pm 1	114 \pm 2	
SPA-KLPOS	4.5 \pm 2.6	4.6 \pm 1.2	99 \pm 18	166 \pm 43
	21 \pm 2	37 \pm 15	203 \pm 31	

^a K_D values were calculated from k_{off}/k_{on} for each experiment and were averaged from two separate experiments.

^b Averaged data from three separate experiments.

^c No binding detected.

^d Activation occurred at a rate similar to autoactivation of MASP-2 alone.

the junction between the cysteine-rich domain and the collagenous domain.

Engineering Complement Activity into a Pulmonary Surfactant Protein—Given its surprising tolerance to even large changes in MBL structure, we sought to characterize MASP activation further by engineering novel activity into a bouquet-like template. We chose SP-A for these studies because of its similar structure but lack of innate complement activity. SP-A, like MBL, is a member of the collectin family of animal lectins with an N-terminal collagen-like domain and a C-terminal carbohydrate-binding domain (28). It has sugar specificity similar to that of MBL (but with a preference for L-fucose over mannose) (29), allowing purification by affinity

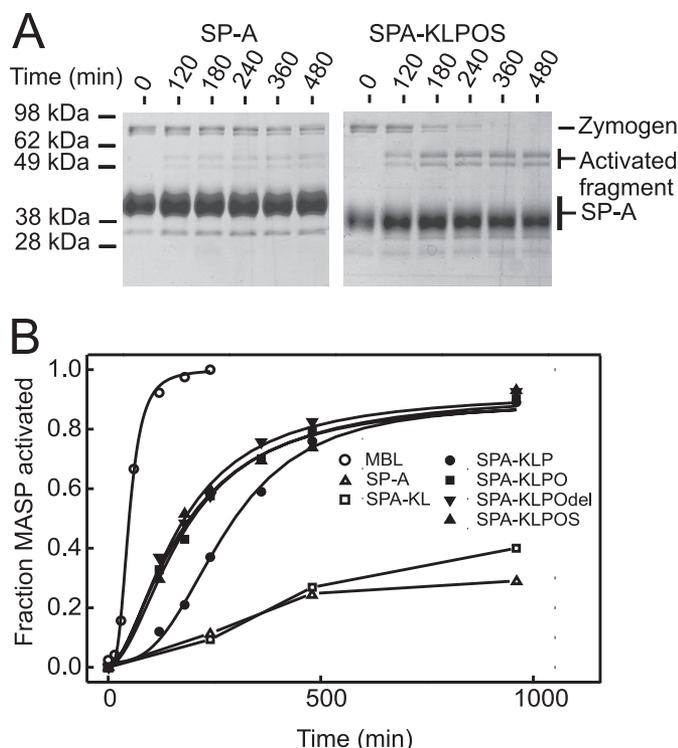


FIGURE 6. **MASP-2 activation by modified SP-As.** Fucose-Sepharose was used as the activating ligand. Proteins were separated on a 15% SDS-polyacrylamide gel and were stained with Coomassie Blue. *B*, MASP-2K activation by wild-type and mutant SP-As. Averaged data from three separate experiments are shown.

apart and thus the angle between adjacent stalks are not limiting for MASP binding. Thus, just three amino acid changes to the collagenous domain are sufficient to introduce MASP binding to SP-A almost comparable with MBL itself.

We next examined whether the modified SP-As could activate MASP-2K. The half-time for activation by MBL was ~50 min, using fucose-Sepharose as a target. As expected, wild-type SP-A, which does not bind to the MASP at all, and SPA-KL, which binds only weakly, did not activate the MASP (Fig. 6). In each case, the measured rate was similar to the intrinsic auto-activation rate of MASP-2K alone ($t_{1/2} > 1000$ min). Surprisingly, however, SPA-KLP and SPA-KLPO activated the MASP significantly faster than the basal rate ($t_{1/2} \sim 260$ min). Similar rates were also measured for SPA-KLPOdel and SPA-KLPOS, demonstrating that the kink is neither necessary for MASP activation nor modulates the rate of activation significantly. To determine whether activation was target-dependent, assays were repeated in the absence of fucose-Sepharose. Surprisingly, comparable rates were observed (Fig. 7). Activation was still much faster than the basal rate, so it must be driven by SP-A binding, but was constitutive, occurring even in the absence of a carbohydrate target. Thus, modified SP-As lack the control mechanisms required to activate MASP selectively.

DISCUSSION

The data presented here reveal that novel MASP binding and complement activation can be introduced into a bouquet-like template through relatively minor changes to the chemical makeup of the collagenous stalks. Just three substitutions:

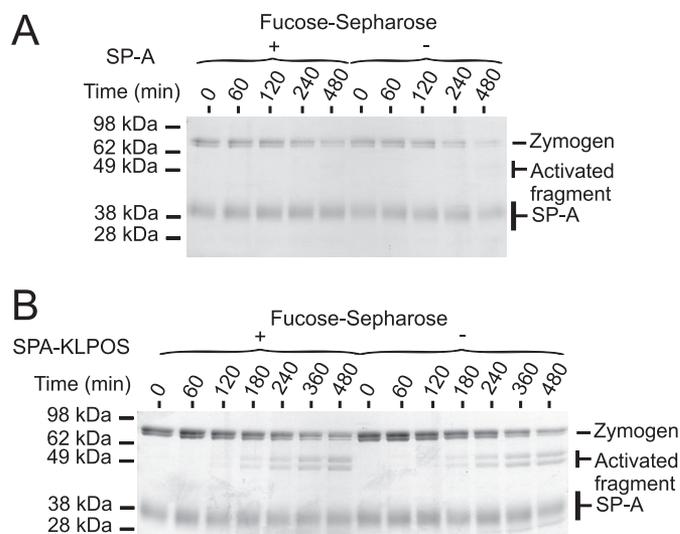


FIGURE 7. **MASP-2 activation in the presence and absence of a carbohydrate target.** Proteins were separated on a 15% SDS-polyacrylamide gel and were stained with Coomassie Blue. MASP-2K activation by wild-type SP-A (*A*) and SPA-KLPOS (*B*) in the presence and absence of a carbohydrate target.

Glu⁶³ to Lys, Cys⁶⁴ to Leu, and Glu⁶⁶ to Pro, establish MASP binding in SP-A, which is almost equivalent to that of MBL. Of these residues, only the lysine is essential for binding (10, 13). Generally, the adjacent residue in the Yaa position (leucine in rat MBL and the SP-A mutants) has an aliphatic side chain, although a small polar residue such as serine can also be accommodated, but glutamate cannot (10). The proline residue in the Xaa position of the next Gly-Xaa-Yaa triplet is also not essential for MASP binding and can be replaced by an alanine with only a small decrease in affinity. The main function of the proline is probably to help stabilize the binding region (10). It is notable that a significant component of the difference in MASP binding by SPA-KLP relative to SPA-KL is due to faster association, so it probably reflects removal of the unfavorable electrostatic interactions mediated by the glutamate, rather than the introduction of the proline residue itself. The presence of any additional binding sites for MASP on MBL can be completely ruled out from the data presented here. Overall, therefore, specificity of the MBL-MASP interaction appears to be driven mainly by a single lysine residue in the collagen-like domain.

Several other binding motifs have been identified in collagens and typically extend across two or more Gly-Xaa-Yaa triplets. For example, the I domain of $\alpha 2$ integrin recognizes the sequence GFOGER (31), and the extracellular matrix protein SPARC binds to the sequence GVMGFO (32). Thus, the less stringent binding requirements of MASP-2 are relatively unusual. Nevertheless, it is notable that almost all of the other lysine residues in the collagenous domains of MBL and ficolins are in the Yaa position of the Gly-Xaa-Yaa repeat, so they are likely to be post-translationally modified by hydroxylation and glycosylation, which probably blocks any potentially incorrect MASP interactions. Perhaps even more importantly, MASP binds to MBL and ficolins through multiple relatively weak interactions involving up to four separate CUB-collagen contacts. Thus, much of the binding specificity is probably medi-

Engineering Complement Activity into SP-A

ated through the architectures of the subcomponents and the multivalent nature of the interactions.

Complement activation is thought to be triggered by changes in the angle between collagenous stalks when complexes bind to an activating target (3, 23). The data presented here indicate that significant conformational and rotational flexibility of the stalks must be permitted, implying that unbound MBL is probably quite flexible. Both relaxed-to-strained and strained-to-relaxed mechanisms could explain the activation process. The traditional (relaxed-to-strained) view is that activation is transmitted to the MASP through a change in the structure of MBL that induces a high energy strained state. However, in an alternative but equally valid possibility, binding of the zymogen MASP to MBL (or ficolin) induces strain into circulating complexes, which releases upon pathogen binding to trigger activation. A major difference between these mechanisms is that the relaxed or default states of the complexes would differ: inactive in a relaxed-to-strained mechanism as activation is reliant on surface binding, but active in a strained-to-relaxed mechanism. In the work described here, MASP binding leads to constitutive activation by SP-A, even in the absence of a target. Thus, activation is the default state, as would be expected in a strained-to-relaxed mechanism. The possibility that modified SP-As are locked into a high energy conformation that activates MASP constitutively cannot be completely excluded. However, SP-A must also be flexible to bind to the MASP multivalently, which would seem to be at odds with this possibility. Furthermore, large structural changes (through removal of the kink) do not affect MASP activation by SP-A, which also seems incompatible with such a mechanism.

Based on these data, we propose that MBL·MASPs also activate via a strained-to-relaxed mechanism. We have recently suggested that activation of the C1 complex also operates in this way (2) and have proposed that strain is induced into the zymogen complex through the interactions between the protease domains of C1r. In this respect, the mechanism must differ in MBL·MASP complexes because the SP domains of a MASP dimer do not interact significantly until the moment of autocatalysis. Interestingly, however, recent analysis suggests that the MBL stalks are not evenly distributed, but rather are highly asymmetrical and separated by $\sim 35\text{--}40^\circ$ between nearest neighbors (7). In this case, significant strain would be induced when MBL binds to a MASP. Release of this strain upon target recognition might drive the changes that initiate complement activation.

From an evolutionary perspective, these data suggest that recognition-mediated complement activation could have arisen relatively rapidly in a collectin-like molecule such as SP-A as result of a small number of changes to the collagen-like domain. It would also explain how similar activation mechanisms could have arisen multiple times in structurally analogous but unrelated recognition components. Clearly however, regulation of activation necessitates additional sequences, and we are currently undertaking studies to identify these regions in MBL and ficolins.

Acknowledgments—We thank Robert Freedman and Katrine Wallis, of the Department of Biological Sciences, University of Warwick, for use of the departmental BIAcore facility.

REFERENCES

- Porter, R. R., and Reid, K. B. M. (1978) *Nature* **275**, 699–704
- 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., Keeble, A. H., and Wallis, R. (2009) *J. Immunol.* **182**, 7708–7717
- Wallis, R., Mitchell, D. A., Schmid, R., Schwaeble, W. J., and Keeble, A. H. (2010) *Immunobiology* **215**, 1–11
- Heise, C. T., Nicholls, J. R., Leamy, C. E., and Wallis, R. (2000) *J. Immunol.* **165**, 1403–1409
- Brodsky-Doyle, B., Leonard, K. R., and Reid, K. B. M. (1976) *Biochem. J.* **159**, 279–286
- Strang, C. J., Siegel, R. C., Phillips, M. L., Poon, P. H., and Schumaker, V. N. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 586–590
- Jensenius, H., Klein, D. C., van Hecke, M., Oosterkamp, T. H., Schmidt, T., and Jensenius, J. C. (2009) *J. Mol. Biol.* **391**, 246–259
- Reid, K. B., and Porter, R. R. (1976) *Biochem. J.* **155**, 19–23
- Wallis, R., and Drickamer, K. (1999) *J. Biol. Chem.* **274**, 3580–3589
- Girija, U. V., Dodds, A. W., Roscher, S., Reid, K. B., and Wallis, R. (2007) *J. Immunol.* **179**, 455–462
- Reid, K. B., Sim, R. B., and Faiers, A. P. (1977) *Biochem. J.* **161**, 239–245
- Teillet, F., Lacroix, M., Thiel, S., Weilguny, D., Agger, T., Arlaud, G. J., and Thielens, N. M. (2007) *J. Immunol.* **178**, 5710–5716
- Wallis, R., Shaw, J. M., Uitdehaag, J., Chen, C. B., Torgersen, D., and Drickamer, K. (2004) *J. Biol. Chem.* **279**, 14065–14073
- Feinberg, H., Uitdehaag, J. C., Davies, J. M., Wallis, R., Drickamer, K., and Weis, W. I. (2003) *EMBO J.* **22**, 2348–2359
- Sim, R. B., and Tsiftoglou, S. A. (2004) *Biochem. Soc. Trans.* **32**, 21–27
- Scherer, P. E., Williams, S., Fogliano, M., Baldini, G., and Lodish, H. F. (1995) *J. Biol. Chem.* **270**, 26746–26749
- Doliana, R., Mongiat, M., Bucciotti, F., Giacomello, E., Deutzmann, R., Volpin, D., Bressan, G. M., and Colombatti, A. (1999) *J. Biol. Chem.* **274**, 16773–16781
- Reid, K. B. M., Colomb, M. G., and Loos, M. (1998) *Immunol. Today* **12**, 56–59
- Chen, C. B., and Wallis, R. (2001) *J. Biol. Chem.* **276**, 25894–25902
- Chen, C. B., and Wallis, R. (2004) *J. Biol. Chem.* **279**, 26058–26065
- Kaufman, R. J., Davies, M. V., Wasley, L. C., and Michnick, D. (1991) *Nucleic Acids Res.* **19**, 4485–4490
- Wallis, R., and Drickamer, K. (1997) *Biochem. J.* **325**, 391–400
- Wallis, R. (2007) *Immunobiology* **212**, 289–299
- Shoulders, M. D., and Raines, R. T. (2009) *Annu. Rev. Biochem.* **78**, 929–958
- Kurata, H., Cheng, H. M., Kozutsumi, Y., Yokota, Y., and Kawasaki, T. (1993) *Biochem. Biophys. Res. Commun.* **191**, 1204–1210
- Wallis, R., and Cheng, J. Y. (1999) *J. Immunol.* **163**, 4953–4959
- Fan, P., Li, M. H., Brodsky, B., and Baum, J. (1993) *Biochemistry* **32**, 13299–13309
- Hoppe, H.-J., and Reid, K. B. M. (1994) *Protein Sci.* **3**, 1143–1158
- Haurum, J. S., Thiel, S., Haagsman, H. P., Laursen, S. B., Larsen, B., and Jensenius, J. C. (1993) *Biochem. J.* **293**, 873–878
- Uemura, T., Sano, H., Katoh, T., Nishitani, C., Mitsuzawa, H., Shimizu, T., and Kuroki, Y. (2006) *Biochemistry* **45**, 14543–14551
- Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) *Cell* **101**, 47–56
- Hohenester, E., Sasaki, T., Giudici, C., Farndale, R. W., and Bächinger, H. P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18273–18277