The function and structure of the human complement protease, MASP-3

Tang Yongqing BBiomedSc (Hons)

Department of Biochemistry and Molecular Biology, School of Biomedical Science, Monash University

Submitted November, 2013

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other institution. To the best of my knowledge, this work contains no material previously published or written by another person, except where due reference is made in the text.

Signed:……. Date: 23th October, 2013

PUBLICATIONS ARISING FROM THIS WORK

Yongqing, T, Wilmann, PG, Reeve, SB, Coetzer, TH, Smith, AI, Whisstock, TC, Pike RN, and Wijeyewickrema LC. (2013) The X-ray crystal structure of Mannose-binding lectin associated serine proteinase-3 reveals a molecular basis for 3MC syndrome. *J Biol Chem*. **288**(31):22399-407.

Wijeyewickrema LC, **Yongqing T**, Tran TP, Thompson PE, Viljoen JE, Coetzer TH, Duncan RC, Kass I, Buckle AM, and Pike RN. (2013) Molecular Determinants of the Substrate Specificity of the Complement-initiating Protease, C1r. *J Biol Chem*. **288**(22):15571-80.

Yongqing T, Drentin N, Duncan RC, Wijeyewickrema LC, and Pike RN. (2012) Mannose-binding lectin serine proteases and associated proteins of the lectin pathway of complement: two genes, five proteins and many functions? *Biochim Biophys Acta*. **1824**(1):253-62.

ACKNOWLEDGEMENTS

I would like to acknowledge the assistance of the following people for their contribution to this thesis and my PhD study:

Prof. Rob Pike, whose expertise in the field of proteases has sparkled in every page of this thesis, and for his everlasting support, invaluable advice, enthusiasm and encouragement throughout the years when I worked in his laboratory. And Dr. Lakshmi Wijeyewickrema, for her hard work on supervising this project and nurturing my professional growth, and for her sizzling enthusiasm to put a smile on everyone's face.

Dr. Pascal Wilmann, for his invaluable advice, guidance and help on solving the structure of MASP-3, and Dr. Danuta Maksel, for her assistance and advice on protein crystallization. Dr. Shane Reeves, for her contribution on peptide sequencing and mass spectrometry. Dr. Oded Kleifeld and Dr. Jiang Ning Song, for their assistance and knowledge on proteomic analysis. Prof. James Whisstock, for his brilliant ideas on activation of MASP-3. A/Prof. Marcin Drag, for synthesizing the ACC-substrate library. Prof. Guiying Nie and Dr. Sarah Paule, for their contribution to study the PC enzymes.

Dr. Renee Duncan, Dr. Natasha Ng, Dr. Por Chee Ong, Ms. Nicole Dretin, Ms. Laura D'Andrea, Ms. Leigh Yang and Mrs. Usha Koul, for their friendship and keeping my emotional balance.

Very special thanks to my dearest friend, Assad Aboud. Without his encouragement and support, my study would have been clouded by shadows.

Finally, a great appreciation goes to all my families. Although you may never fully understand what I was doing, I did everything for you to be proud of me.

TABLE OF CONTENTS

TITLE PAGE……I

DECLARATION……...II

PUBLICATIONS ARISING FROM THIS WORK……...III

ACKNOWLEDGEMENTS……IV

TABLE OF CONTENTS……V

LIST OF TABLES……...X

LIST OF FIGURES……...XI

ABBREVIATIONS……XIV

ABSTRACT……XVI

CHAPTER 1: INTRODUCTION……..1

1.1 THE COMPLEMENT SYSTEM……2

1.2 THE ACTIVATION OF THE COMPLEMENT SYSTEM……..2

1.2.1 The classical pathway……...4

1.2.2 The alternative pathway……5

1.2.3 The lectin pathway……5

1.3 THE COMPONENTS OF THE LECTIN PATHWAY……6

1.3.1 The recognition molecules of the lectin pathway……6

1.3.2 The MASP components of the lectin pathway……9

1.4 MASP-2 AND MAP19……11

1.5 MASP-1……14

1.5.1 Gene structure and localization of expression of MASP-1……15

1.5.2 The structural chemistry of MASP-1……17

1.5.3 The activation and biological functions of MASP-1……..21

1.6 MASP-3……24
1.6.1 The gene and tissue distribution of MASP-3……24
1.6.2 The structure of MASP-3……25
1.6.3 The activation of MASP-3……27
1.6.4 The biological function of MASP-3……28
1.7 MAP44……30
1.8 PROJECT AIMS……31

CHAPTER 2: GENERAL MATERIALS AND METHODS……33
2.1 GENERAL MATERIALS……34
2.2 SDS-PAGE……35
2.3 WESTERN BLOT……36
2.4 CONSTRUCTION OF RECOMBINANT PLASMIDS FOR EXPRESSION OF RECOMBINANT C1r, MASP-1 AND MASP-2 PROTEINS……36
2.5 EXPRESSION, UNFOLDING AND REFOLDING, AND PURIFICATION OF THE RECOMBINANT C1r, MASP-1 AND MASP-2 PROTEINS……37
2.6 N-TERMINAL SEQUENCING PREPARATION……41

CHAPTER 3: THE ACTIVATION OF THE CATALYTIC PORTION OF MASP-3……42
3.1 INTRODUCTION……43
3.1.1 A new sight into the activity of C1r protease……43
3.1.2 The possible involvement of proprotein convertases in activation of MASP-3……45
3.1.3 The aims……47
3.2 MATERIALS AND METHODS……49
3.2.1 Materials……49
3.2.2 Construction of recombinant plasmids for expression of recombinant wtM3……49

3.2.3 Site directed mutagenesis.....50
3.2.4 Expression, unfolding, refolding and purification of recombinant MASP-3 Proteins……50
3.2.5 Preparation of a C1r column and activation of M3EEKQ/M3Q proteins…….52
3.2.6 Prediction of cleavage sites of wtM3 by PC enzymes……53
3.3 RESULTS……54
3.3.1 Production and characterization of recombinant MASP-3 proteins……..54
3.3.2 Cleavage of wtM3 and M3R454A by MASP-1 and MASP-2……...59
3.3.3 Activation of M3EEKQ/M3Q by C1r protease……..62
3.3.4 Activation of wtM3 by PC enzymes……...…66
3.4 DISCUSSION……..70

CHAPTER 4: THE SUBSTRATE SPECIFICITY OF MASP-3…….75
4.1 INTRODUCTION……76
4.2 MATERIALS AND METHODS……..78
4.2.1 Materials……78
4.2.2 Activity assay using synthetic AMC substrates……78
4.2.3 Characterization of substrate specificity of MASP-3 by using REPLi…..79
4.2.4 Activity assay using individual Abz substrates……..83
4.2.5 Determination of the P1-specificity of MASP-3……84
4.2.6 Determination of the P4-P2 specificity by using the ACC-substrate library…….....…..84
4.2.7 Aldehyde inhibitor library screening for M3Q……..85
4.2.8 Examination of the proteolytic activity of M3Q against protein Substrates……...…85
4.2.9 Silver staining……86
4.3 RESULTS……87
4.3.1 The activity of M3EEKQ and M3Q against AMC substrates indicated that

the two enzymes were equally active……87
4.3.2 Varying buffer conditions did not substantially improve the catalytic activity of MASP 3……87
4.3.3 The overall substrate specificity of MASP-3 determined by REPLI…….90
4.3.4 Identification of the optimal tripeptidyl substrates and the P1-specificity of MASP-3……..93
4.3.5 Determination of the P4-P2 specificity of MASP-3 by using the ACC-substrate library……..98
4.3.6 Aldehyde inhibitor library screening for M3Q……..98
4.3.7 Examination of the proteolytic activity of M3Q against protein Substrates……101
4.4 DISCUSSION……104

CHAPTER 5: THE CRYSTAL STRUCTURE OF A 3MC SYNDROME-ASSOCIATED MASP-3 MUTANT……..……110
5.1 INTRODUCTION……..……111
5.2 MATERIALS AND METHODS……114
5.2.1 Cloning……114
5.2.2 Expression, unfolding, refolding and purification of recombinant M3QG666E and M3QG687R proteins……115
5.2.3 Activation of M3QG666E and M3QG687R proteins……..115
5.2.4 Activity assays for the 3MC Syndrome-associated M3Q mutants……..116
5.2.5 Crystallization of the zymogenic M3QG666E protein……116
5.2.6 Data collection and processing……116
5.3 RESULTS……118
5.3.1 The purification and activation of recombinant M3QG666E and M3QG687R proteins……118
5.3.2 The catalytic activity of recombinant M3QG666E and M3QG687R…..118
5.3.3 The crystallization of M3QG666E……...122

5.3.4 The overall crystal structure of M3QG666E……122
5.3.5 The flexibility of the hinge regions within the catalytic portion of MASP-3……125
5.3.6 Comparisons of the loop structures around the active site of zymogenic MASP enzymes……129
5.3.7 The structural perturbation in the active site of M3QG666E……...131
5.4 DISCUSSION……134

CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS……140
6.1 THE ACTIVATION PARTNER OF MASP-3……..141
6.2 THE PHYSIOLOGICAL SUBSTRATE AND BIOLOGICAL FUNCTION OF MASP-3……143
6.3 THE STRUCTURAL BIOLOGY OF MASP-3……146

APPENDIX I (PUBLICATION): Mannose-binding lectin serine proteases and associated proteins of the lectin pathway of complement: two genes, five proteins and many functions?……150

APPENDIX II (PUBLICATION): Molecular Determinants of the Substrate Specificity of the Complement-initiating Protease, C1r ……161

APPENDIX III (PUBLICATION): The X-ray crystal structure of Mannose-binding lectin associated serine proteinase-3 reveals a molecular basis for 3MC syndrome ……172

REFERENCES……182

LIST OF TABLES

Table 1.1 The polypeptides of MASP members……10

Table 2.1 Preparation of a 12.5% running gel and a 4% stacking gel……35

Table 2.2 The calculated masses & absorption coefficients of the recombinant proteins……41

Table 3.1 Members of the human PC family enzymes……..……46

Table 3.2 Primers for MASP-3 CCP1-CCP2-SP mutants……..50

Table 3.3 The calculated masses & absorption coefficients of the recombinant proteins……52

Table 3.4 The predicted cleavage sites of wtM3 by at least one member of the proprotein convertase family of enzymes……...66

Table 5.1 Primers for 3MC syndrome-associated M3Q mutants……..114

Table 5.2 The calculated masses & absorption coefficients of the recombinant proteins……..115

Table 5.3 Data collection and refinement statistics for M3QG666E……124

LIST OF FIGURES

Figure 1.1 The Complement activation pathways ……3

Figure 1.2 A diagrammatic representation of the recognition molecules of the lectin pathway of complement……7

Figure 1.3 Genomic organization of the *masp-2* gene and resulting protein structures……12

Figure 1.4 Genomic organization of the *masp-1/3* gene and resulting protein structures……16

Figure 1.5 The crystal structure of a human MASP-1 CUB1-EGF-CUB2 homodimer……..18

Figure 1.6 The crystal structure of the active human MASP-1 CCP1-CCP2-SP segment……20

Figure 2.1 Purification of the recombinant C1r CCP1-CCP2-SP protein……38

Figure 2.2 Purification of the recombinant MASP-1 CCP1-CCP2-SP protein……39

Figure 2.3 Purification of the recombinant MASP-2 CCP1-CCP2-SP protein……40

Figure 3.1 The sequence alignment of protein members of the C1r/C1s/MASPs family at the activation site……43

Figure 3.2 Schematic representation of recombinant MASP-3……48

Figure 3.3 Purification of the CCP1-CCP2-SP fragment of wild type MASP-3 (wtM3)……55

Figure 3.4 Purification of the M3R454A CCP1-CCP2-SP fragment……..56

Figure 3.5 Purification of the M3EEKQ CCP1-CCP2-SP fragment……...57

Figure 3.6 Purification of the M3Q CCP1-CCP2-SP fragment……...58

Figure 3.7 SDS-PAGE analyses of cleavage of wtM3 by proteolytic enzymes of the complement and coagulation systems……60

Figure 3.8 The cleavage products of wtM3 after treatment with either MASP-1 or MASP-2……...……61

Figure 3.9 The cleavage products of M3R454A after treatment with either MASP-1 or MASP-2……..63

Figure 3.10 Cleavage of M3EEKQ and M3Q by recombinant C1r protease…….64

Figure 3.11 Activation of MASP-3 mutants by using a C1r column……...…..65

Figure 3.12 Cleavage of wtM3 by furin……68

Figure 3.13 Cleavage of wtM3 by PACE4……...69

Figure 4.1 A schematic display of the REPLi method……80

Figure 4.2 The allocation of tripeptide pools in REPLi……...82

Figure 4.3 The rate of cleavage of AMC substrates by MASP-3 mutants……..88

Figure 4.4 Kinetics of cleavage of VPR-AMC and LGR-AMC by M3EEKQ andM3Q……..89

Figure 4.5 The catalytic activity of M3Q in varying conditions……..91

Figure 4.6 The overall specificity profile for M3Q enzyme using REPLi……...92

Figure 4.7 The rates of cleavage of the individual Abz peptidyl substrates by M3Q……95

Figure 4.8 Kinetics of the cleavage of Abz substrates by M3Q……...96

Figure 4.9 Representative mass spectra identifying cleaved products of the Abz-peptides, RIY, KIY, RRL, and RKL after treatment with M3Q……...97

Figure 4.10 Substrate specificity of M3Q characterized by using the ACC-substrate library……..99

Figure 4.11 The sequences of aldehyde inhibitors that best attenuated the activity of M3Q (top twenty shown)……..100

Figure 4.12 Cleavage of protein substrates by M3Q……102

Figure 5.1 Conservation of residue C^{630}, G^{666} and G^{687} in MASP-3 among vertebrate species……..112

Figure 5.2 Purification of the M3QG666E (CCP1-CCP2-SP) fragment……...119

Figure 5.3 Purification of the M3QG687R (CCP1-CCP2-SP) fragment……...120

Figure 5.4 Activation of 3MC-associated MASP-3 mutants by using a C1r column……...……121

Figure 5.5 A crystal of zymogenic M3QG666E at day 10……122

Figure 5.6 Cartoon representations of the M3QG666E structure……...123

Figure 5.7 Domain flexibility of zymogenic M3QG666E…….……126, 127

Figure 5.8 Comparisons of loops surrounding the active site of MASP enzymes……..…130

Figure 5.9 The structural perturbation at the active site of M3QG666E……...133

Figure 5.10 The 3MC-associated mutation sites besides G^{666}……138

Figure 6.1 Structural alteration of the activation loop region of MASP-3 upon activation of enzyme……147

ABBREVIATIONS

a.a.	Amino acid
Abz	Ortho-aminobenzoic acid
Ac	Acetyl
ACC	7-amino-4-carbamoylmethylcoumarin
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AMC	7-amino-4-methylcoumarin
AU	Arbitrary unit
BMP	Bone morphogenetic protein
Boc	N-tert-butyloxycarbonyl
BSA	Bovine serum albumin
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid
CCP	Complement control protein
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECL	Electrochemiluminescence
EDTA	Ethylenediaminetetra-acetic acid
FAB	Fluorescent assay buffer
HCl	Hydrochloride acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Ig	Immunoglobulin
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
LB	Luria Bertani
MAC	Membrane attack complex
MASP	Mannan-binding lectin-associated serine protease

MBL	Mannan-binding lectin
MPD	2-methyl-2,4-pentanediol
NaCl	Sodium chloride
NaOH	Sodium hydroxide
Nle	Norleucine
OD	Optical density
PACE4	Paired basic amino acid-cleaving enzyme 4
PC	Prohormone convertase or proprotein convertase
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PVDF	Polyvinylidene difluoride
QFF	Q-Sepharose Fast Flow (column)
RNA	Ribonucleic acid
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrymide gel electrophoresis
SP	Serine protease
TEMED	Tetramethylethylenediamine
TGF	Tissue growth factor
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Z	α-benzyloxycarbonyl

ABSTRACT

Mannose-binding lectin (MBL)-associated serine proteases (MASPs) activate the lectin pathway of the complement system, an integral part of the human innate immune response. Thus far, three MASP enzymes (MASP-1, -2, and -3) and two non-catalytic derivatives, MAp19 and 44, have been identified. MASP-1, MASP-3 and MAp44 are products of alternative splicing of the same *masp-1/3* gene. MASP-1 and MASP-3 enzymes only differ in their C-terminal serine protease domains and thus are identical in their five N-terminal non-catalytic domains. However, the fact that MASP-3 is conserved among all vertebrate species examined, including those in which MASP-1 is not present, indicates that MASP-3 may play a role different to MASP-1 in vertebrate species. MASP-3 is ubiquitously distributed in the human body and exists in human plasma at a relatively high concentration of around 5 μg/ml, while its role in activation of complement is unlikely, since it does not cleave complement components C2, C3 and C4, and it is not inhibited by C1-inhibitor. Recent genetic studies have identified a critical correlation between genetic defects of MASP-3 and an autosomal recessive disorder, the Carnevale, Mingarelli, Malpuech and Michels (3MC) syndrome, indicating a possible role of MASP-3 in embryonic development. However, knowledge of the structure and function of MASP-3 was modest at the outset of this study.

MASP-3 is secreted as a single-chain zymogen and is activated by cleavage at the specific R^{449}-I^{450} activation bond. The lack of knowledge of the enzyme responsible for activation of MASP-3 has cast a great level of uncertainty on previous studies using active MASP-3. This study provides the first demonstration that MASP-2, not MASP-1, activates MASP-3. However, the presence of the R^{454}-N^{455} peptide bond near the activation bond limits the specificity of this activation. Further, data derived here also

first identified a specific *in vitro* activation of MASP-3 by the human proprotein convertase enzymes, furin and PACE4, which specifically cleave after dibasic residues, present in the activation site of MASP-3. However, the slow rate of cleavage indicates that the conditions of such activation need to be optimized.

In order to achieve efficient activation of MASP-3, two MASP-3 mutant proteins, M3EEKQ and M3Q, were constructed, both of which contain sequences N-terminal to the activation bond of the classical complement pathway serine protease homologue, C1s, which is specifically activated by C1r. Both M3EEKQ and M3Q were activated effectively and uniformly by recombinant C1r. The active forms of MASP-3 mutants consist of a serine protease domain identical to that of wild type MASP-3. Further, the substrate specificity of MASP-3 was characterized using combinational peptidyl substrate libraries, showing that MASP-3 most prefers the sequences GRIF/GRIY for cleavage, with specificity for R/K residues at the P1 position. The fluorescence-quenched tripeptide substrate, RIF, was found to be 20-fold more effectively cleaved by MASP-3 than the AMC substrate, VPR, which was previously claimed to be among the best substrates for MASP-3.

The present study of the 3MC syndrome-associated MASP-3 G666E and G687R mutants revealed that the active forms of these two mutants fail to cleave any of the substrates tested, including the optimal substrate, RIF. This data is therefore the first piece of evidence that conclusively shows that the 3MC syndrome is always related to a lack of MASP-3 activity. Solution of the three-dimensional structure of the zymogen form of the G666E mutant of MASP-3 to 2.6 Å shows that the mutation most likely results in an aberrant ionic bonding in the active site of the enzyme that is unlikely to be altered upon activation of MASP-3, thus providing the first insight into the mechanism whereby the mutation causes enzyme inactivation.

As a summary, by identifying the enzyme candidates responsible for activation of MASP-3, developing MASP-3 mutants that can be efficiently activated, identifying the optimal substrate for cleavage by MASP-3, solving the first structure of a 3MC syndrome-associated MASP-3 mutant and illustrating the molecular basis of the inactivation of this mutant, the present study has contributed strongly to knowledge about MASP-3, which had previously been very little characterized. These findings also provide a strong background for further study of the enzyme.

Chapter One

Introduction

1.1 THE COMPLEMENT SYSTEM

The complement system serves as one of the frontlines of defence against infection and cancer. It is an integral part of the innate immune response and networks more than 30 circulatory or cell surface-located proteins in a proteolytic cascade (Sim and Tsiftsoglou, 2004). Early studies of the complement system following its discovery in the late 19th century have highlighted its ability to 'complement' antibodies and cause cell lysis. It is now well-accepted that the complement system is not only capable of directly killing infectious organisms, but also promotes the elimination of these organisms by presenting them to phagocytic cells and enhancing inflammation (Frank and Fries, 1991). Furthermore, the complement system also plays a role in shaping B-cell and T-cell responses, and hence represents a bridge connecting the innate and adaptive immune systems (Dunkelberger and Song, 2010). The complement system also plays a role in eliminating dying cells, cell debris and immune complexes, which are potential sources of self-antigens (Markiewski and Lambris, 2007). Poor restraint of the complement system may result in inflammatory disorders, such as hereditary angioedema (Morgan and Walport, 1991). In contrast, lacking complement activity due to rare genetic impairment of complement components is often correlated to the development of autoimmune diseases and repetitious infection, particularly by the pyrogenic microorganisms (Markiewski and Lambris, 2007) and the Neisserial bacteria (Walport, 2001).

1.2 THE ACTIVATION OF THE COMPLEMENT SYSTEM

The complement system can be activated via three major routes (Figure 1.1): the classical, alternative, or lectin pathways. Each pathway utilizes unique recognition molecules to recognize a wide range of molecules presented on the surface of the pathogens and malignant cells. Binding of the recognition component triggers the

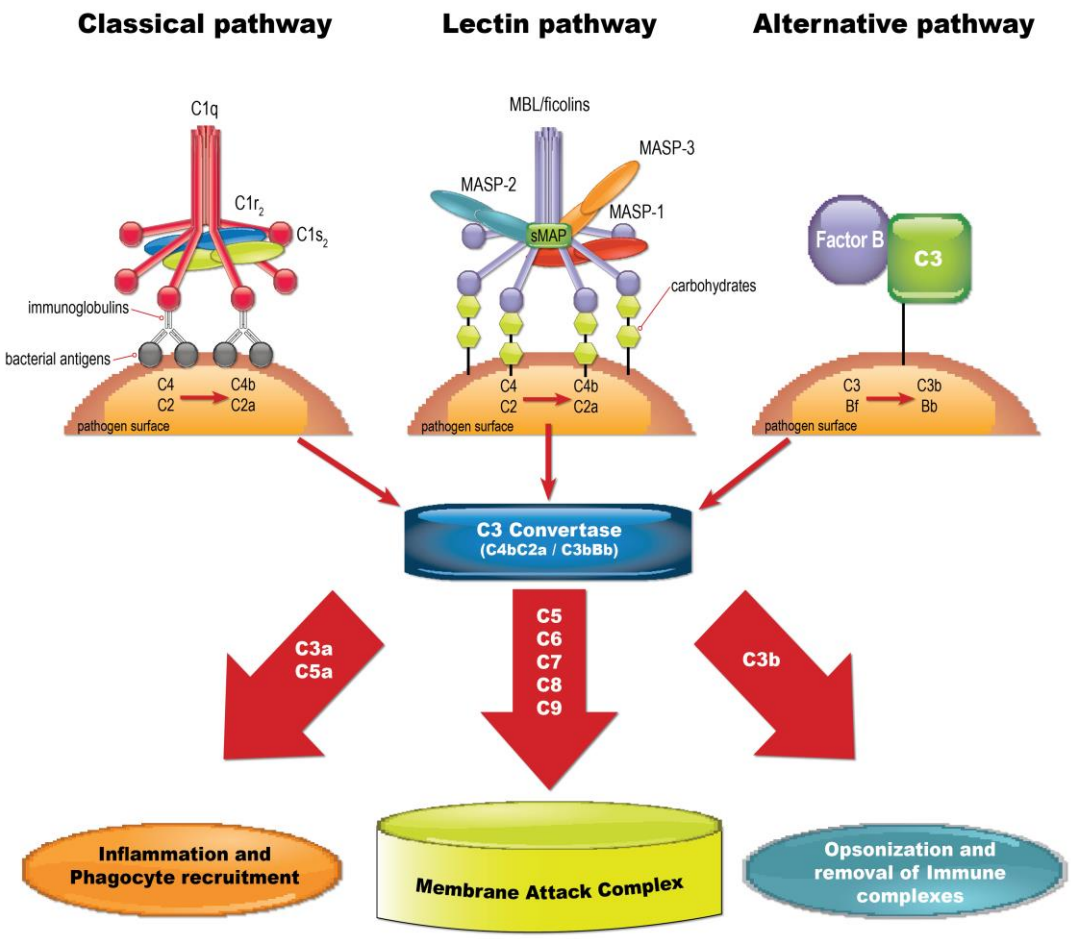


Figure 1.1 The Complement activation pathways (Duncan *et al*., 2008)
The complement system can be activated via three major routes: the classical, alternative, or lectin pathways. The activation of the complement system leads to the formation of a number of by-products, including C3a, C5a and C3b, as well as the membrane attack complex, which causes cell lysis.

activation of associated initiating enzymes, thus activating the proteolytic cascade of complement. Initiation of all three activation routes of the complement system leads to the formation of a C3 convertase complex (C4b2a or C3bBb). The complex cleaves the complement component C3 into a smaller peptide (C3a) and a larger fragment, C3b, which is homologous to C4b and possesses an exposed thioester group that can interact with hydroxyl or amino acid groups on the surface of target cells (Reid *et al*., 1986). On the target cell surface, C3b also binds the C4b or C3b component of C3 convertase to

form the C5 convertase complex (C4b2a3b or C3bBb3b), which cleaves C5 into C5a and C5b. Subsequently, the cell surface-localized C5b initiates the assembly of the membrane attack complex (MAC), which encompasses C5b, C6, C7, C8 and up to 13 molecules of C9 (Walport, 2001, Esser, 1994). Multiple units of MAC consequently cause a rapid imbalance of cellular Ca^{2+}, which leads to the breakdown of energy supply of the cells, and eventually causes lysis in the target cell (Papadimitriou *et al*., 1991).

1.2.1 The classical pathway

The classical pathway starts with the recognition of multiple surface molecules, including the Fc region of IgG and IgM in an antibody-antigen complex, the cell membrane-bound C-reactive protein, and a variety of charged or hydrophobic particles, by the plasma C1 complex (Sim and Tsiftsoglou, 2004). The C1 complex consists of a key recognition molecule, C1q, and a tetramer of zymogenic serine proteases, C1s-C1r-C1r-C1s (Gaboriaud *et al*., 2004). C1q is a 450 kDa protein consisting of six collagen-like N-terminal tri-helices, which associate with the C1r/C1s tetramer, and six C-termini that each contains triple globular domains that are responsible for binding to the pathogenic targets (Schumaker *et al*., 1987, Arlaud *et al*., 2001, Arlaud *et al*., 2002, Gaboriaud *et al*., 2003). Once the recognition binding occurs, the activation of the C1q-bound C1r molecules is triggered (Arlaud *et al*., 1986). The activated C1r strictly cleaves and activates C1s (Kardo *et al*., 2008). The resulting active C1s is highly effective in cleaving C4 and C2, forming the C3 convertase (C4b2a) on the surface of pathogens (Gaboriaud *et al*., 2000).

1.2.2 The alternative pathway

The alternative complement pathway is initiated when the surface carbohydrate of a foreign substance is recognized by the spontaneously produced C3b, most likely resulting from the slow hydrolysis of C3 in plasma (Pangburn and Muller-Eberhard, 1984). The circulatory C3b-like molecules, including hydrolysed C3 and amine-modified C3, are found to bind complement factor B, which is then processed by complement factor D to yield the fragments, Ba and Bb. The resulting C3b-like molecule in complex with Bb can function as a circulatory C3 convertase, which is believed to be another mechanism to produce the initial C3b for this pathway (Lachmann and Hughes-Jones, 1984). Subsequently, the pathogen surface-localized C3b binds Bb to form the C3 convertase (C3bBb). This C3 convertase can initiate the downstream events of the complement system, and can cleave further C3 molecules to produce more surface-localized C3b, further enhancing the complement response (Pangburn and Muller-Eberhard, 1984). Since the C3b resulting from activation of the classical and lectin pathways can also initiate the alternative complement pathway, the latter pathway can act as an amplification loop of complement following activation of the other two pathways (Lachmann, 2009).

1.2.3 The lectin pathway

Activation of the lectin pathway was observed in the mannose-binding lectin (MBL)-dependent activation of C4 in early research of the pathway (Ikeda *et al*., 1987). The initiating enzyme was initially identified as a serine protease in complex with MBL, which was thus termed an MBL-associated serine protease (MASP) (Matsushita and Fujita, 1992). It was believed that binding of MBL to the carbohydrate target triggers the activation of the MBL-bound zymogenic MASP enzyme. The active MASP enzyme in turn cleaves C4 and C2, forming the pivotal C3 convertase (C4b2a) (Matsushita and Fujita, 1992, Ji, *et al*., 1993, Kawasaki *et al*., 1989).

1.3 THE COMPONENTS OF THE LECTIN PATHWAY

Since its discovery, the lectin pathway has been gradually recognized as a much more complicated pathway than the other two activation routes for complement. Besides MBL, other groups of proteins have been found to recognize a range of surface structures to activate the lectin pathway. Adding more complexity, multiple MASP enzymes and derivatives were also found. However, a great deal of knowledge regarding these components remains incomplete.

1.3.1 The recognition molecules of the lectin pathway

The lectin pathway utilizes MBL, collectin 11 (CL-11, or collectin kidney 1, CL-K1) and ficolins as the recognition molecules (Figure 1.2) (Matsushita and Fujita, 1992, Hansen *et al*., 2010, Matsushita *et al*., 1996). These proteins and the classical pathway recognition molecule C1q are topological homologues, though they share low sequence similarity. Indeed, C1r and C1s do not associate with MBL (Thiel *et al*., 2000).

MBL is a plasma protein of the collectin family, which are a group of soluble collagenous pattern recognition receptors recognizing carbohydrate and lipid on the surface of pathogens (Holmskov *et al*., 2003). The structural subunit of MBL encompasses three identical polypeptides. Each polypeptide consists of a cysteine-rich N-terminus, a collagen-like region with sequence of G-Xaa-Yaa repeats, a 'neck' region with α-helical coiled coils, and a C-terminal Ca^{2+}-dependent carbohydrate recognition domain (CRD), which recognizes the 3' and 4' horizontal hydroxyl groups of the pyranose ring presented in many carbohydrates (Weis *et al*., 1992). Unlike the strict six oligomer form of a C1q molecule, which is similar in oligomeric structure to the

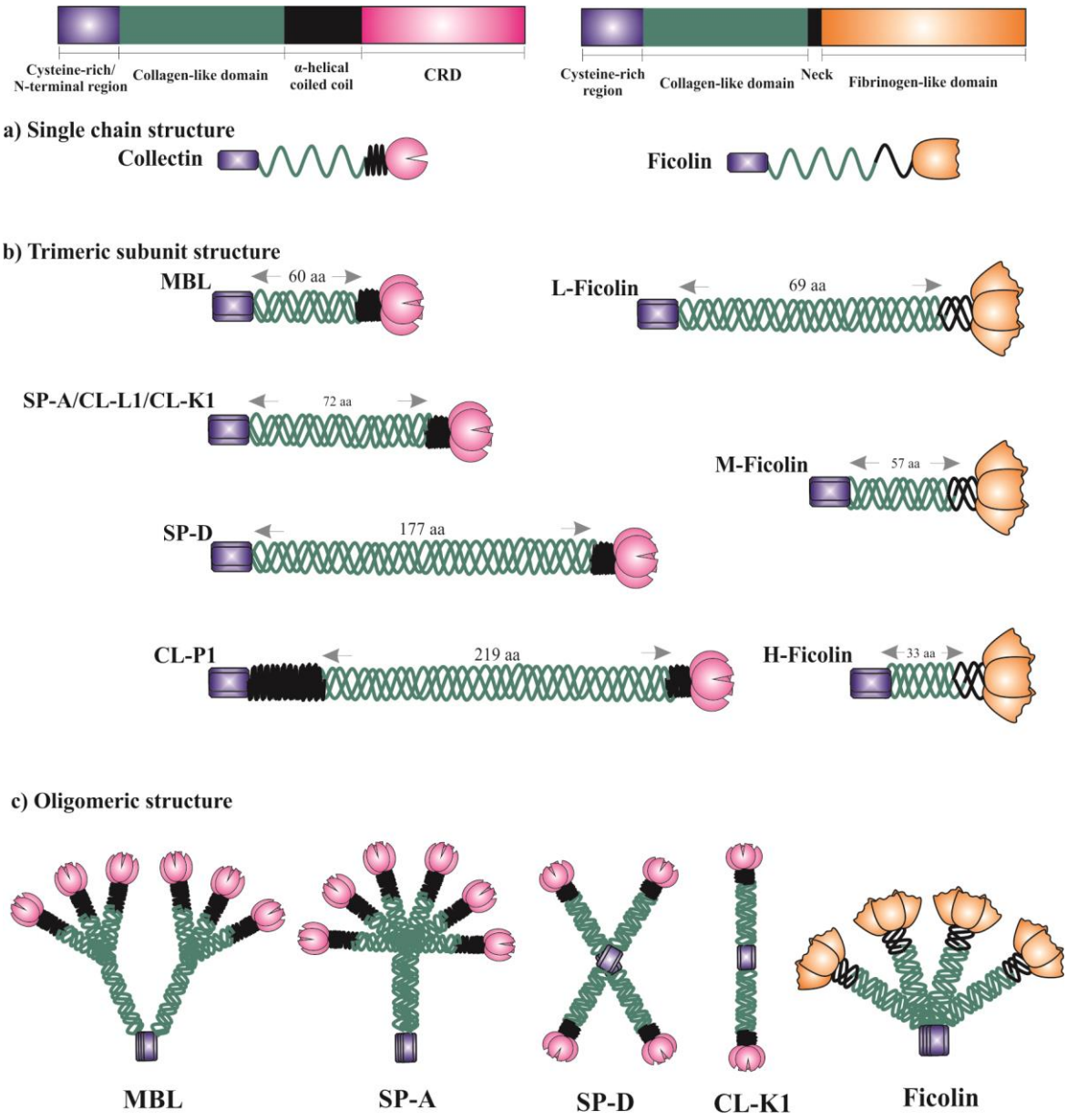


Figure 1.2 A diagrammatic representation of the recognition molecules of the lectin pathway of complement (Yongqing *et al.*, 2012)

Top, scheme for the domain structure of the collectins (left) and ficolins (right). For the collectins, the cysteine-rich domain is shown in purple, the collagen-like domain in green, the alpha-helical coiled-coil region in black and the carbohydrate recognition (CRD) domain in pink. The fibrinogen-like domain of the ficolin family members is shown in orange. a) The single chain organization of the collectin (left) and ficolin (right) family proteins is shown. b) The basic organization of the trimeric subunits of the collectin (left) and ficolin (right) family proteins. Colours correspond to the domains labelled above in the schematic domain representation. c) Potential oligomeric structures for family members of the collectin and ficolin family are shown. It should be noted that other oligomeric forms also exist, especially in the case of MBL.

collectin member, surfactant protein A (SP-A) (Figure 1.2), a few oligomeric MBL forms, termed MBL-I, -II, -III, and -IV, with molecular weights of 275, 345, 580, and 900 kDa, respectively, have been fractionated from human serum (Dahl *et al*., 2001). MBL-I and MBL-II were later determined to be the trimeric and tetrameric forms of MBL, respectively (Teillet *et al*., 2005). The serum MBL in healthy populations ranges widely from extremely low to higher than 5 μg/ml (Dommett *et al*., 2006). MBL deficiency is considered the most frequent congenital immunodeficiency state in humans (Turner and Hamvas, 2000). Low plasma concentrations of MBL have been found to correlate with repetitious infection and autoimmune diseases (Super *et al*., 1989, Boniotto *et al*., 2005, Øhlenschlæger *et al*., 2004).

CL-11, a new member of the collectin family, was recently found in complex with MASP enzymes in human plasma (Hansen *et al*., 2010). It shares the same domain organization with other collectin members like MBL and surfactant proteins, containing an N-terminal domain, a collagen-like region, a 'neck' region and a CRD. CL-11 consists of a C-Xaa-C motif in its 'neck' region, a characteristic motif of the CXC-chemokine family, though CL-11 is totally different to the chemokine family topologically. The presence of the CXC motif in the 'neck' region is also found in CL-L1 (collectin lung 1), but not CL-P1 (collectin placenta 1) or other collectins. CL-11 is localized in the liver, kidney and adrenal gland and primarily recognizes D-mannose and L-fucose on a wide range of organisms (Selman and Hansen, 2012).

Thus far, three forms of ficolins, termed L-(also named p35 or ficolin-2), M-(or ficolin-1), and H-ficolin (also called Hakata-antigen or ficolin-3), have been found in humans (Matsushita *et al*., 1996, Lu *et al*., 1996, Yae *et al*., 1991). L-ficolin and M-ficolin share about 80% similarity in primary structure, different from H-ficolin which shares only 45% sequence similarity with the other members. L-ficolin and H-ficolin are synthesized in liver and secreted into plasma. H-ficolin is also produced

by bile duct epithelial cells and hence exists in bile. The M-ficolin protein is found on the surface of monocytes and granulocytes, as reviewed by Matsushita *et al*. (2013). Among the three ficolin members, the collagen-like regions of the ficolin monomers are different in length. The ligand-binding regions of all ficolin members are fibrinogen-like recognition domains, which recognize the acetyl groups on a wide range of substances, including those on carbohydrates, natural and artificial acetylated compounds (Ichijo *et al*., 1993, Matsushita *et al*., 1996).

1.3.2 The MASP components of the lectin pathway

Five members of MASP proteins have thus far been identified, including three MASP enzymes termed MASP-1/-2/-3, a non-enzymatic protein, MAp19 (or sMAP), which shares an origin with MASP-2, and another non-catalytic protein, MAp44 (or MAP1), which shares an origin with MASP-3 (Matsushita and Fujita, 1992, Thiel *et al*., 1997, Dahl *et al*., 2001, Takahashi *et al*., 1999, Degn *et al*., 2009, Skjoedt *et al*., 2010). The results of sequence alignment of MASP enzymes and C1r/C1s revealed that these enzymes are members of the same family. All members possess a conserved H/S/D catalytic triad, and share the same domain organization: from the N-terminus, the first C1r/C1s/Uegf/bone morphogenetic protein 1 (CUB1) domain, an epidermal growth factor-like (EGF) domain, the CUB2 domain, followed by two complement control protein (CCP1 and CCP2) domains and the C-terminal serine protease (SP) domain. The SP domain is linked to the CCP2 domain by a small linker region.

Members of the C1r/C1s/MASPs family of enzymes are produced as single polypeptide zymogens (Table 1.1). A conserved R-I activation bond is located between the linker region and the SP domain of each member of the enzyme family. Activation of these enzymes, including MASP-3, requires cleavage at the activation bond (Endo *et al*., 1998). Resulting from the cleavage, the light chain contains a newly liberated NH^{3+} at

the N-terminal I residue preceding the SP domain, where the active site is located. Indeed, in order to convert zymogen molecules to their active forms, the requirement for the specific cleavage of an R-I peptide bond to generate an N-terminal I residue with a free NH^{3+} is common to most serine proteases, including chymotrypsin and trypsin, two of the most representative serine proteases. Studies on the activation of chymotrypsin and trypsin revealed that the newly liberated NH^{3+} at the N-terminal I residue forms a bond with the D residue preceding the active site S residue. The formation of this bond affects the structure of the surrounding molecules, which brings the active site S residue to its active state position and also forms a complete substrate-binding cleft. Therefore, the specific cleavage of the R-I activation site contributes to the maturation of the active site and the substrate-binding properties of the active serine proteases (Khan and James, 1998). The active MASP enzyme therefore consists of A (or heavy/H) and B (or light/L) chains, linked by a disulfide bond. The A chain is comprised of the five N-terminal domains and the linker region, and the B chain consists of the SP domain of the corresponding MASP enzyme (Endo *et al*., 1998).

Table 1.1 The polypeptides of MASP members

Material	Amino acids (signal peptide/polypeptide)	Molecular mass (kDa) (calculated/observed)	N-glycosylated sites
MASP-1	19/680	77/90	4
MASP-2	15/671	74/74	0
MASP-3	19/709	82/94	7
Map19	15/170	22/22	0
Map44	19/361	44/45	2

1.4 MASP-2 AND MAP19

MASP-2 was first described as a C1s-like serine protease that complexes with MBL in human plasma, whilst MAp19 was first identified as a MASP-1-binding protein, which consists of only the CUB1 and EGF domains identical to those of MASP-2 and a short unique C-terminal sequence (EQSL) (Thiel *et al*., 1997, Takahashi *et al*., 1999). MASP-2 and MAp19 are products of alternative splicing of the *masp-2* gene, located on human chromosome 1 p36.2-3 (Stover *et al*., 1999a, b). The genetic organization is shown in Figure 1.3. The production of the transcript of MAp19 is guided by an in-frame stop codon in exon 5 and a subsequent polyadenylation signal (Takahashi *et al*., 1999). The locations of the *masp-2* gene transcript in human tissues reveal that both proteins are predominantly expressed in the liver, whilst MASP-2 is also expressed in the small intestine and testes to a limited extent (Takahashi *et al*., 1999, Seyfarth *et al*., 2006).

Previous studies have suggested that the CUB1-EGF segment of MASP-2/MAp19 is responsible for the formation of a Ca^{2+}-dependent homodimer of these proteins (Wallis and Dodd, 2000, Chen and Wallis, 2001). Supporting this, the crystal structure of the CUB1-EGF-CUB2 segment of rat MASP-2 revealed a Ca^{2+} binding site on the EGF domain via which the two monomers interact (Feinberg *et al*., 2003). Further, the homodimer of MASP-2 or MAp19 was also thought to bind to MBL/ficolins via their CUB1-EGF region in a Ca^{2+}-dependent manner (Thielens *et al*., 2001, Cseh *et al*., 2002). Later, a Ca^{2+} binding site was found in the CUB1 domain in the crystal structure of human MAp19. It was found that this Ca^{2+} binding site participates in the interaction between MAp19 and MBL/ficolins, since mutating the residues of this site decreased the binding of MAp19 to the recognition molecules (Gregory *et al*., 2004). Another piece of evidence is that a natural MASP-2 mutant with a G^{105} residue, normally a D,

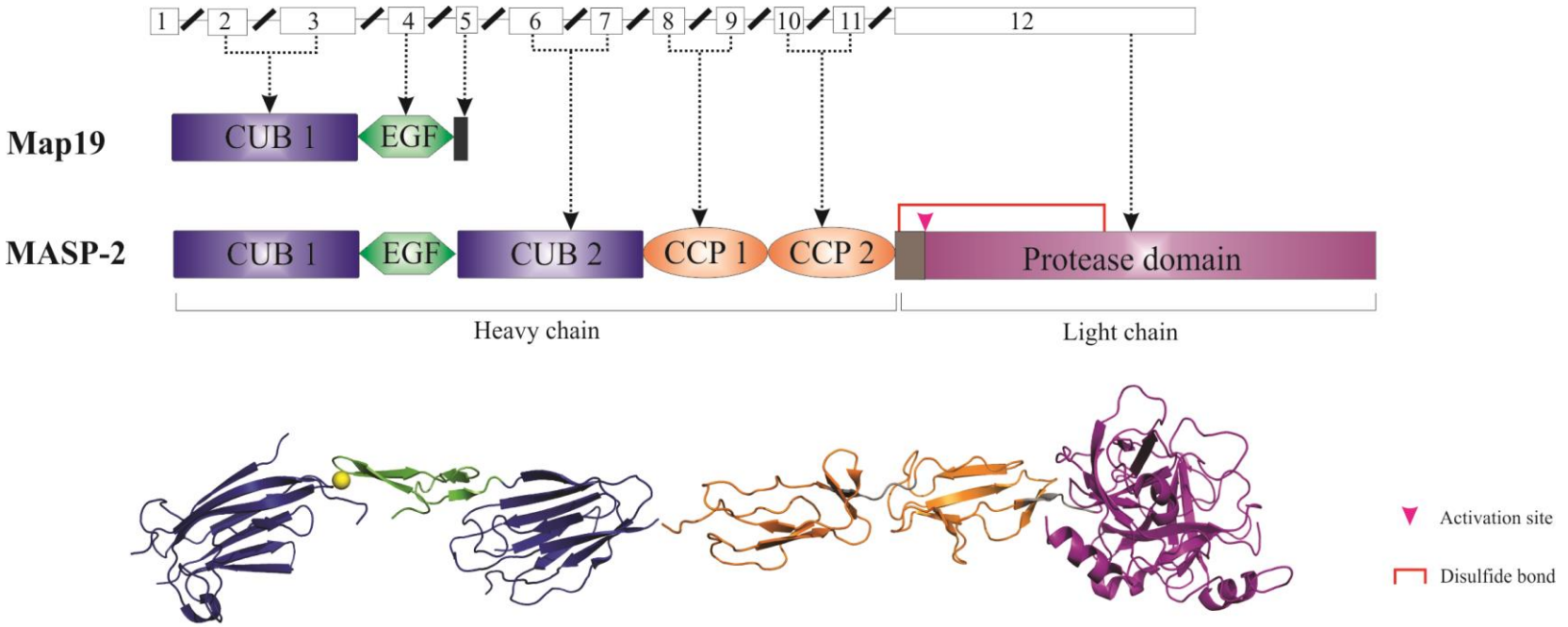


Figure 1.3 Genomic organization of the *masp*-2 gene and resulting protein structures (Yongqing *et al*., 2012)
Top, the exon–intron structure of the *masp-2* gene. Map19 contains the exons encoding the CUB1 domain and the EGF domain and is terminated by the stop codon in exon 5. An arrow head indicates the scissile Arg-Ile bond and a red line indicates the disulfide linkage holding the two chains together. Bottom, the crystal structures of human MASP-2 CUB1-EGF-CUB2 (1NT0.pdb) and the catalytic region of zymogenic MASP-2 (1ZJK.pdb) are shown as ribbon diagrams with the domains coloured in the same way as in the domain schemes above. The calcium ion at the interface of the CUB1 and EGF domains is shown as a yellow sphere.

which is a critical residue of the Ca^{2+} binding site in the CUB1 domain of MASP-2, also failed to bind to MBL/ficolins (Stengaard-Pedersen *et al*., 2003). Initially Dahl *et al.* (2001) proposed that MASP-2 and MAp19 preferentially bind to MBL-II and MBL-I, respectively. However, Thielens *et al*. (2001) studied the interaction between purified recombinant MBL and MASP-2/MAp19 by using surface plasmon resonance (SPR), and failed to observe such a binding preference. The crystal structures of the CCP2-SP fragment of active MASP-2 and the CCP1-CCP2-SP segments of an inactive MASP-2 mutant have also been respectively solved (Harmat *et al*., 2004, Gál *et al*., 2005). Both structures suggested that compared to C1s, MASP-2 possesses a more flexible interface between the CCP2 and SP domains due to the presence of a lower number of P residues at this interface. This was thought to possibly underlie the higher catalytic efficiency of MASP-2 over C1s. The SP domain of MASP-2 displays trypsin-like substrate specificity, cleaving after R/K residues.

The biological roles of MASP-2 have been well studied. The MASP-2 polypeptide undergoes auto-activation, which is more efficient when the MASP-2-MBL complex binds to the carbohydrate target (Gál *et al*., 2005). Active MASP-2 cleaves C4 highly effectively and C2, not C5 (Rossi *et al*., 2001, Matsushita *et al*., 2000a, b). The molecular basis for the high efficiency of MASP-2 in cleaving C4 was later revealed by Kidmose *et al*. (2012), who solved the structure of a MASP-2/C4 complex and found that the formation of the complex involves interaction of exosites at the CCP1/CCP2 junction and on the SP of MASP-2. They also noted an allosteric effect on the C4 molecule, induced by the binding of the complexed molecules. Beside C4 and C2, prothrombin was also observed to be cleaved by MASP-2 slowly (Krarup *et al*., 2005). Similar to thrombin and trypsin, MASP-2 is inhibited by C1-inhibitor and by antithrombin III in presence of heparin (Petersen *et al*., 2000, Presanis *et al*., 2004).

Inhibition of MASP-2 using a MASP-2-specific inhibitor completely blocks the activation of the lectin pathway, indicating the pivotal role of MASP-2 in this route of complement activation (Kocsis *et al*., 2010). Clinically, MASP-2 deficiency due to genetic mutation was observed to be racially associated. The previously mentioned D105G mutant of MASP-2 was not functional, resulting in patients with this mutation suffering from autoimmune diseases and repeat infections. The level of MASP-2 was found to be correlated with recurrence of colorectal cancer and with death (Ytting *et al*., 2005). Compared to that of MASP-2, the biological role of MAp19 is still debated. Iwaki *et al*. (2006) studied the activation of C4 on a mannan surface following the addition of different ratios of recombinant MASP-2 and MAp19 into the serum of a MAp19-null mice, and claimed that MAp19 down-regulates the activation of C4 mediated by MASP-2. For this to be effective, however, the concentrations of MASP-2 and MAp19 would have to be greater than the total number of binding sites on MBL and ficolin in the blood (Gál *et al*., 2009). Currently, this is believed to be unlikely, as it is thought that MBL and ficolins are present in excess amounts within the circulation (Mayilyan *et al*., 2006). And the affinity of interaction between MAp19 and MBL is found to be much lower than that between MASP-2 and MBL (Wallis and Dodd, 2000, Thielens *et al*., 2001).

1.5 MASP-1

MASP-1 was first isolated from an MBL preparation in human serum (Matsushita and Fujita, 1992). Later Takada *et al*. (1993) and Takahashi *et al*. (1993) published the amino acid sequence of the full-length MASP-1 in the sera of human and mouse, respectively. In these studies MASP-1 was referred to as P100, a 100 kDa protein that

formed a part of the Ra-reactive factor, which activates the complement system and hence acts as a bactericidal complex when it binds to polysaccharides on some bacteria. It was not until the discovery of MASP-2 (Thiel *et al*., 1997) that MASP-1 received its current name.

1.5.1 Gene structure and localization of expression of MASP-1

In humans, MASP-1 arises as one of the products of alternative splicing of the *masp-1/3* gene, located on chromosome 3q27-28 (Dahl *et al*., 2001). A corresponding gene in mouse was mapped to murine chromosome 16B2-B3 (Takada *et al*., 1995) and it was later found that the production of MASP-1 protein in mouse and rat also involves alternative splicing of this gene (Stover *et al*., 2003, Knittel *et al*., 1997). The *masp-1/3* gene spans about 50 kb which encodes for 17 exons (Figure 1.4) (Sørensen *et al*., 2005). The first exon encodes the signal peptide. Exons 2-11 encode the first five domains of MASP-1, with the last 6 exons encoding a small linker region plus the serine protease domain of MASP-1 (Endo *et al*., 1998). Homologues of the *masp-1/3* gene have been found in many species of vertebrates, including cartilaginous fish, bony fish, birds and mammals, as well as in the ascidian and the amphioxus invertebrate species (Fujita, 2002, Ji *et al*., 1997, Endo *et al*., 2003). However, MASP-1 is not expressed in lamprey, cartilaginous fish, bony fish and birds. The absence of MASP-1 protein in birds was found to be due to the absence of the 6 exons encoding the MASP-1 serine protease domain. This finding suggested that a loss of these exons may have occurred in birds during evolution of the *masp-1/3* gene (Lynch *et al*., 2005). The primary expression site of MASP-1 is the liver, from where MASP-1 is then secreted into plasma (Sato *et al*., 1994). Initially it was thought that the concentration of

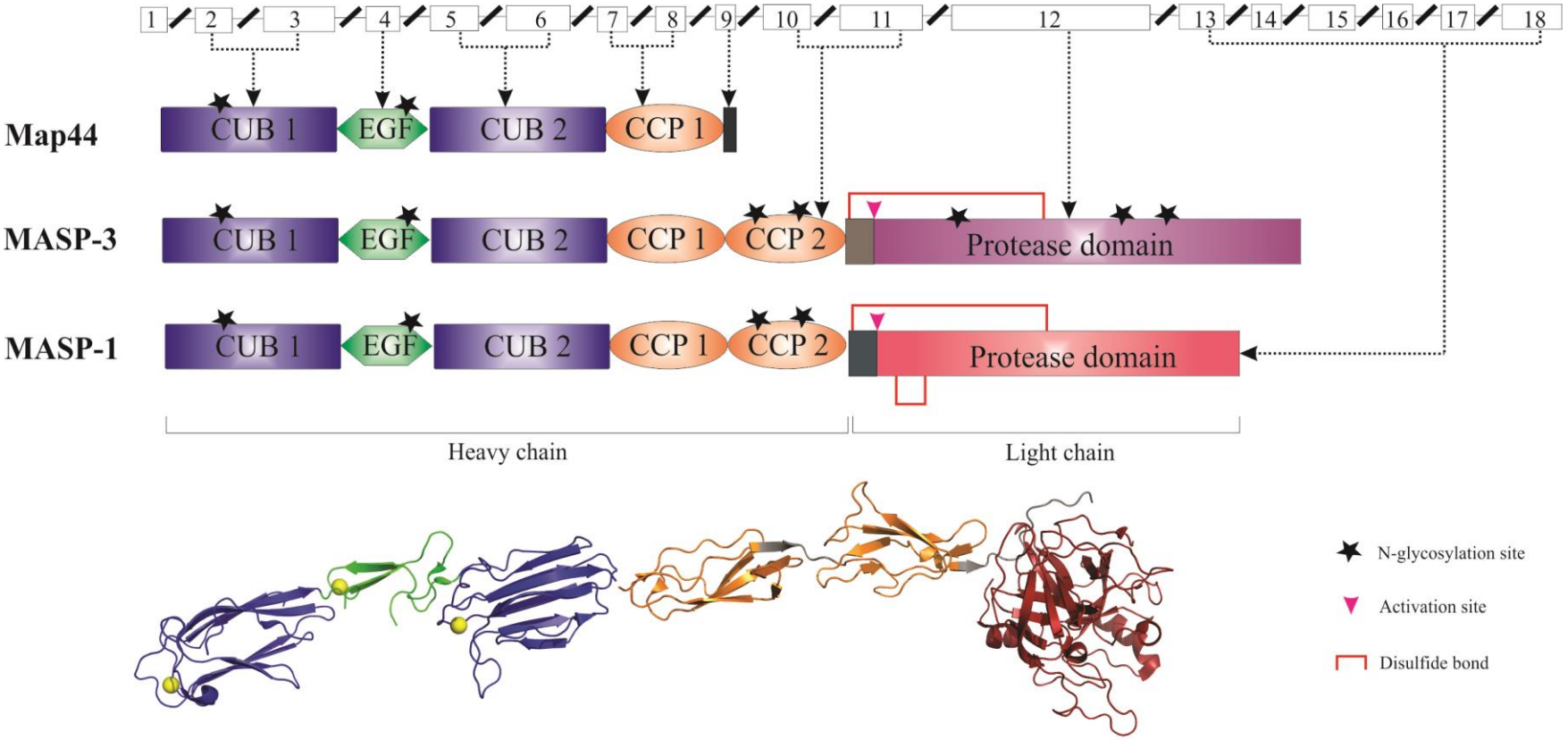


Figure 1.4 Genomic organization of the *masp-1/3* gene and resulting protein structures (Yongqing *et al*., 2012)
Top, the exon–intron structure of the *masp-1/3* gene. MAp44 contains the exons encoding the two CUB domains, the EGF domain and the first CCP domain together with the unique exon 9, which contains a stop codon (marked with a black box). MASP-3 is generated by splicing out exon 9 and terminated by the stop codon in exon 12. The protease domain of MASP-3 is encoded by a single exon, whereas this region in MASP-1 is encoded by six exons. The MASP-1 sequence contains a stop codon in exon 18. The potential glycosylation sites are shown using stars. An arrow head indicates the scissile Arg-Ile bond and a red line indicates the disulfide linkage holding the two chains together. Bottom, the crystal structures of human MASP-1 CUB1-EGF-CUB2 (3DEM.pdb) and the catalytic region of zymogenic MASP-1(4IGD.pdb) are shown as ribbon diagrams with the domains coloured in the same way as in the domain schemes above. The calcium ions in the CUB and EGF domains are shown as yellow spheres.

MASP-1 in human plasma was 6.27±1.85 μg/ml (Terai *et al*., 1997). Later, it was believed that the value includes other derivatives of the *masp-1/3* gene, and the plasma concentration of MASP-1 was deduced to be around 1 μg/ml (Degn *et al*., 2010). Recently Thiel *et al*. (2012) applied an antibody specific to MASP-1 and determined that the actual concentration of MASP-1 in human serum is about 11 μg/ml, the highest among all MASP proteins. MASP-1 is also expressed in a few extra-hepatic tissues, including the small intestines, kidneys and lungs, to a modest level (Seyfarth *et al*., 2006).

1.5.2 The structural chemistry of MASP-1

The MASP-1 polypeptide shares the features of an ancient serine protease, chymotrypsin. These features include six exons encoding for a serine protease domain, a TCN (N denotes the A, G, C or T nucleotide base) codon encoding the catalytic S residue, and the presence of a so-called 'histidine loop' (a.a.475-491) in the serine protease domain (Endo *et al*., 1998, Krem and Di Cerca, 2001). The crystal structure of the CUB1-EGF-CUB2 segment of human MASP-1 has been solved to a resolution of 2.3 Å (Teillet *et al*., 2008) (Figure 1.5). The crystal structure revealed that the CUB1 domain lacks two strands that were observed in the general structure of a CUB module, described as a sandwich of two five-stranded β-sheets (Romao *et al*., 1997), whilst the CUB2 domain lacks one of the two strands missing in CUB1. These structural features of the CUB domains in MASP-1 were also observed in human C1s, rat MASP-2 and human MAp19 (Gregory *et al*., 2003, Feinberg *et al*., 2003, Gregory *et al*., 2004). A Ca^{2+} is bound in each of the two CUB domains of MASP-1, contributing to the stability of the corresponding CUB domain. A homologous Ca^{2+} binding site was also found in

the CUB1 domain of human C1s and Map19 (Gregory *et al*., 2003, Gregory *et al*., 2004). The structure of the EGF domain of MASP-1 is similar to that of a typical EGF homologue, which is characterized by two double-stranded β-sheets linked by five loops and stabilized by three disulfide bonds (Rao *et al*., 1995). A Ca^{2+} binding site is also presented in the MASP-1 EGF domain. A few amino acids forming this binding site are strictly conserved in all other members of the MASP/C1r/C1s family, suggestive of the importance of this Ca^{2+} binding site (Teillet *et al*., 2008).

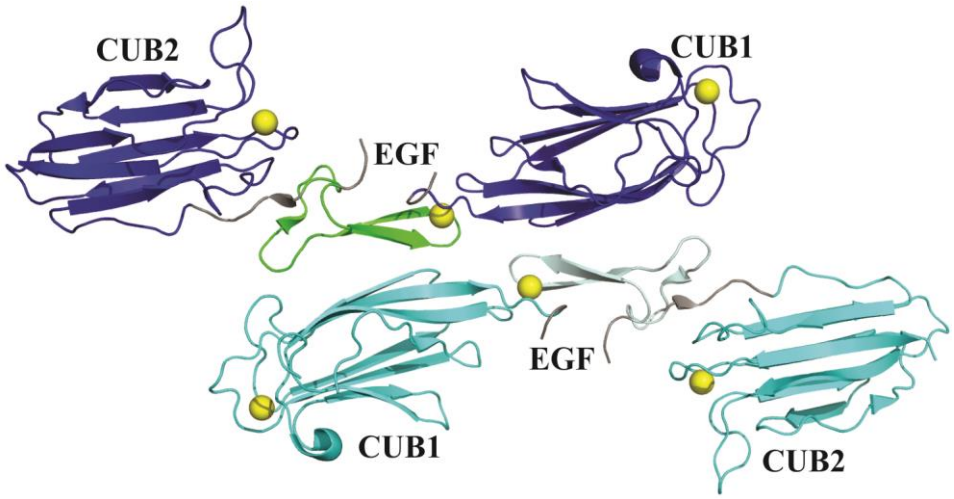


Figure 1.5 The crystal structure of a human MASP-1 CUB1-EGF-CUB2 homodimer
The crystal structure of a human MASP-1 CUB1-EGF-CUB2 homodimer (3DEM.pdb) is shown as ribbon diagrams. The two CUB domains of monomer A (top) are coloured blue, and those of monomer B (bottom) in cyan. The EGF domain of monomer A is in green and that of monomer B in light blue. The calcium ions in the CUB and EGF domains are shown as yellow spheres. The homodimer of the two CUB1-EGF-CUB2 monomers is formed by the CUB1 domain of one monomer interacting with the EGF domain of the other monomer in a 'head to tail' fashion.

Previous studies using truncated MASP-1 proteins from human and rat suggested that MASP-1 monomers form homodimers via a primary binding site on the CUB1 domain (Thielens *et al*., 2001, Chen and Wallis, 2001, Cseh *et al*., 2002). The crystal structure of the CUB1-EGF-CUB2 fragment of MASP-1 has provided evidence for this

hypothesis. The homodimer of the two CUB1-EGF-CUB2 monomers is formed by the CUB1 domain of one monomer interacting with the EGF domain of the other monomer in a 'head to tail' fashion, forming two interacting surfaces, each of which is stabilized by hydrophobic interactions and hydrogen bonds (Teillet *et al*., 2008) (Figure 1.5). Furthermore, it was also found that all three domains of the CUB1-EGF-CUB2 segment of MASP-1 were necessary for the MASP-1 homodimer to effectively bind to MBL and L-ficolin in a Ca^{2+}-dependent manner (Thielens *et al*., 2001, Chen and Wallis, 2001, Cseh *et al*., 2002). It was observed that MASP-1 preferentially binds to MBL-I (Dahl *et al*., 2001).

The structure of the zymogenic MASP-1 CCP1-CCP2-SP protein has also been most recently solved (Megyeri *et al*., 2013) (Figure 1.4). The crystal structure of the CCP1-CCP2-SP segments of active human MASP-1 has been solved with a resolution of 2.55 Å (Dobó *et al*. 2009) (Figure 1.6). In the structure of active MASP-1, each of the two CCP domains of MASP-1 resemble a common CCP homologue, featuring a rod-shaped module made of a number of β-sheets stabilized by two disulfide bonds (Barlow *et al*., 1991). However, an electrostatic interaction between a K and an N residue is presented uniquely in the CCP2-SP interface of MASP-1, accounting for the rigidity of this region of MASP-1 (Dobó *et al*., 2009). The crystal structure of the SP domain of MASP-1 revealed that the active site of MASP-1 locates to a long and open substrate-binding groove that topologically resembles the substrate-binding area of trypsin, formed by a number of loop structures. Therefore it is expected that the substrate specificity of MASP-1 is similar to that of trypsin, which exhibits a strict preference for R or K as the P1-residue [note that the nomenclature for serine proteases indicates that the cleavage occurs in substrates between the P1 and P1' residues, with substrate residues N-terminal to P1 labelled P2, P3 etc. and those C-terminal to P1' labelled P2',

P3' etc.; sub-sites binding the substrate residues in the enzymes are correspondingly labelled, e.g. The S1 sub-site binds the P1 residue, etc. (Schechter and Berger, 1967)]. A so-called '60-loop' (or loop B) with large aromatic residues shadows the area N-terminal to S1 sub-site, indicating a possible preference for small non-polar amino acids at the P2 position. Overall, the long and open substrate-binding area of MASP-1

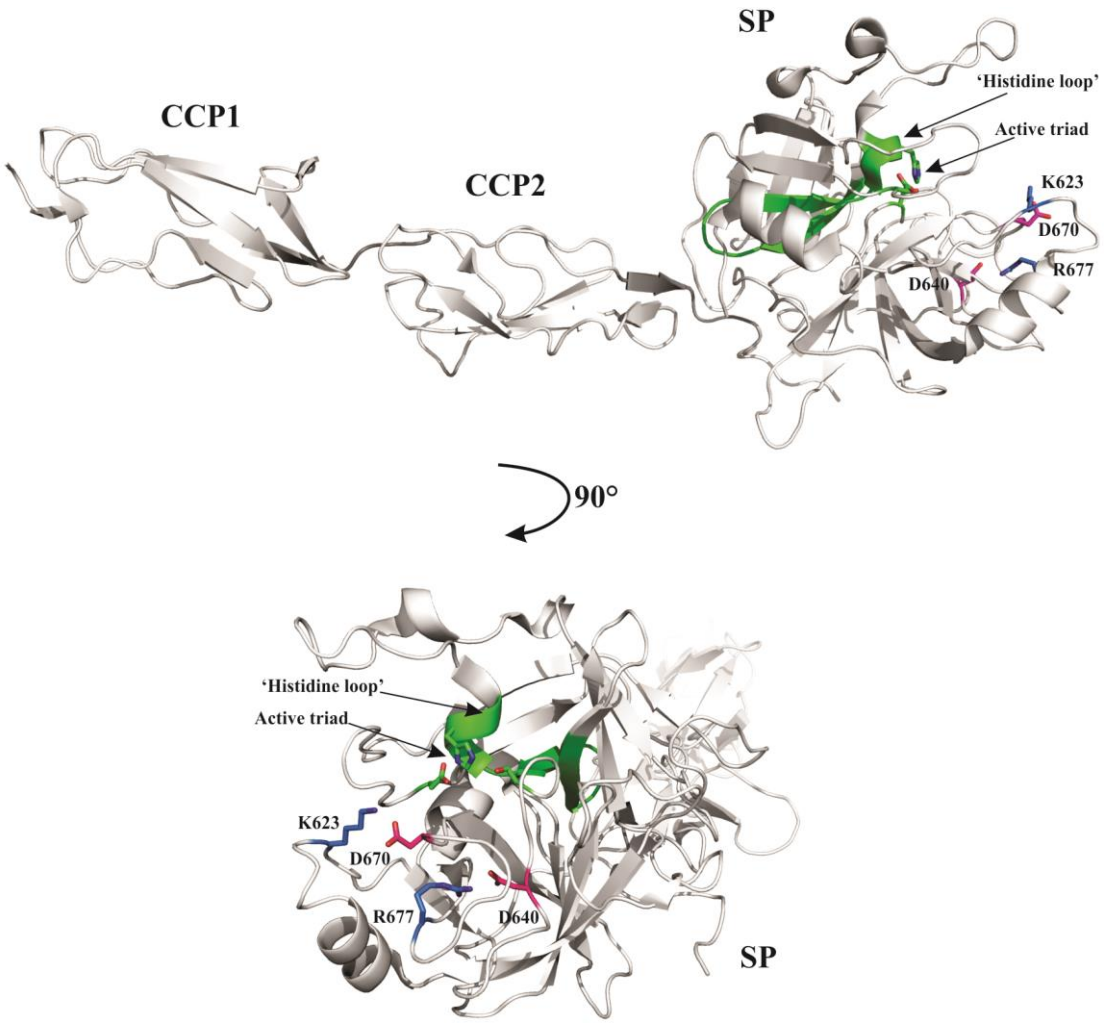


Figure 1.6 The crystal structure of the active human MASP-1 CCP1-CCP2-SP segment
Top, the crystal structure of the catalytic region of active human MASP-1 (3GOV.pdb) is shown as ribbon diagrams. The catalytic triad (H^{490}, D^{552} and S^{646}) are shown as sticks. The 'histidine loop' is in green. Two salt bridges, which are critical for the catalytic activity of MASP-1, are formed by residues R^{677} and D^{640}, and K^{623} and D^{670}, respectively. These residues are shown as sticks. R^{677} and K^{623} are coloured blue, and D^{640} and D^{670} are in magenta. Bottom, a view of the SP structure after a clockwise horizontal rotation of 90°.

suggests broad substrate specificity. Interestingly, a salt bridge is formed by the residues D^{640} and R^{677} in the MASP-1 active site, which would be predicted to restrict the ability of the D^{640} residue to bind substrate and thus reduce the cleavage efficiency of MASP-1 (Figure 1.6). Close to this region, another salt bridge is formed by the residues D^{670} and K^{623}. This second salt bridge not only hinders the substrate-binding ability of D^{670}, but also brings a patch of three K residues (a.a.622-624) into the substrate binding groove. These residues potentially further reduce the substrate-cleaving efficiency of MASP-1. It is proposed that these salt bridges are broken when the preferable substrate binds to the substrate-binding groove of MASP-1 and induces a conformational change in the binding site, restoring the substrate-binding capability of the restrained residues (Dobó *et al*., 2009).

1.5.3 The activation and biological functions of MASP-1

MASP-1 is believed to undergo auto-activation, since substituting the active site S^{627} residue with an A residue only resulted in the zymogenic form of MASP-1 (Zundel *et al*., 2004). Active MASP-1 exhibits a strong preference for synthetic peptidyl substrates with P1-R/K, in agreement with the predicted substrate specificity based on the crystal structure of the SP domain of MASP-1 (Zundel *et al*., 2004). MASP-1 was observed to cleave C3, and was thought to activate complement system directly via such cleavage (Matsushita and Fujita, 1995, Matsushita *et al*., 2000b, Dahl *et al*., 2001). Studies by Petersen *et al*. (2001) and Selander *et al*. (2006) confirmed the C3 cleavage by MASP-1, but showed that it occurred at relatively low efficiency. To the contrary, a study by Wong *et al.* (1999) was unable to observe C3 cleavage by human plasma-derived MASP-1, and subsequent studies using native MASP-1 and eukaryote-

expressed recombinant MASP-1 found that they cleaved only hydrolysed C3, instead of the native substrate. Rossi *et al*. (2001) and Ambrus *et al*. (2003) confirmed the C3 cleavage by MASP-1, but argued against the capability of MASP-1 to activate the complement system via such cleavage due to the low cleavage rate. Therefore, the involvement of MASP-1 in activating the complement system by cleaving C3 remains unlikely. MASP-1 was also found to cleave C2 (Matsushita and Fujita, 1995, Rossi *et al*., 2001, Ambrus *et al*., 2003), which in turn is able to strongly activate the entire complement system (Chen and Wallis, 2004, Møller-Kristensen *et al*., 2007, Rawal *et al*., 2008). The idea of MASP-1 activating or promoting the activation of MASP-2 was supported by a few studies (Takahashi *et al*., 2008, Kocsis *et al*., 2010). However, such contribution of MASP-1 is debatable since MASP-2 itself clearly undergoes auto-activation (Gál *et al*., 2005). Degn *et al*. (2012) found that recombinant MASP-1 in complex with MBL was able to directly activate MASP-2, and thus dramatically increase the MASP-2-dependent activation of lectin pathway. Such transactivation of MASP-2 by MASP-1 was also supported by Megyeri *et al*. (2013), who also claimed that even MASP-1 zymogen can activate MASP-2. Furthermore, MASP-1 exhibits thrombin-like activity, cleaving two substrates of thrombin, fibrinogen and factor XIII (Hajela *et al*., 2002). It was also found that, like thrombin, MASP-1 can be effectively inhibited by antithrombin in presence of heparin (Dobó *et al*., 2009). In addition, the activity of MASP-1 is tightly controlled in human plasma by C1-inhibitor and α-2-macroglobulin (Matsushita *et al*., 2000b, Petersen *et al*., 2000, Rossi *et al*., 2001, Terai *et al*., 1995).

A recent study utilized MASP-1-specific and MASP-2-specific inhibitors to examine the contribution of the two enzymes to activation of the lectin pathway. It was found that inhibiting either of the two enzymes prevented the activation of the lectin pathway,

and that inhibiting MASP-1 was more effective than inhibiting MASP-2. The authors therefore suggested that MASP-1 plays a determinant role in activating the lectin pathway (Kocsis *et al*., 2010). The suggestion is supported by Degn *et al*. (2012), who found that restoring MASP-1 into the serum of a *masp-1/3* gene-defected patient could reconstitute the lectin pathway of complement, apparently by activating MASP-2, and thus implicated MASP-1 as an important component of the lectin pathway.

Studies using transgenic mice have suggested a possible role of MASP-1 in the activation of alternative pathway (Takahashi *et al*., 2010, Banda *et al*., 2010). Takahashi *et al.* (2010) found that the masp1/3-null mouse failed to convert the zymogenic factor D (pro-Df) to active factor D (Df), and thus had an impaired alternative complement pathway. In their study, a recombinant MASP-1 was shown to activate the recombinant murine pro-Df in vitro. However, addition of this recombinant MASP-1 into the serum of the *masp1/3*-null mouse failed to detect the activation of pro-Df. Another study by Banda *et al*. (2010), using an alternative pathway-dependent arthritis murine model, revealed a direct correlation between MASP-1 and the activation of alternative pathway via activating pro-Df. In the latter study, mixing the serum from the *masp1/3*-null mouse and that from a Df-knock-out mouse resulted in activation of the pro-Df in the former serum. The cause of the difference between the results of these two studies remains to be solved. However, the involvement of MASP-1 in activating the alternative pathway of complement in humans is unlikely, since the patient with a nonsense mutation in the *masp 1/3* gene was found to have a functionally normal alternative pathway (Degn *et al*., 2012). Additionally, neither of the studies using *masp1/3*-null mice addressed the possible involvement of MASP-3, which is also knocked out with MASP-1 in the transgenic mouse since the two proteins are encoded by the same *masp1/3* gene.

1.6 MASP-3

MASP-3 was first separated from a preparation of MBL/MASP complexes in human plasma as a protein of 105 kDa which forms disulfide-linked dimers (Dahl *et al*., 2001). Compared to MASP-2 and MASP-1, MASP-3 has only featured in a limited number of studies.

1.6.1 The gene and tissue distribution of MASP-3

Like MASP-1, human MASP-3 is also encoded by the Masp-1/3 gene and arises as one of the products of alternative splicing of the *masp-1/3* gene (Figure 1.4). The production of MASP-3 in mouse and rat also involves alternative splicing of the homologous gene in the species (Stover *et al*., 2003, Knittel *et al*., 1997). In human liver, the level of the MASP-3 mRNA was observed to be similar to that of MASP-1 (Seyfarth *et al*., 2006), but in rat liver, the level of the MASP-3 mRNA is about two thirds that of the MASP-1 mRNA, even though both transcripts arise from transcription using the same gene promoter (Stover *et al*., 2003). The cause of this difference in the levels of mRNA of MASP-3 and MASP-1 in rat liver remains unclear. Major suggestions include structural factors in the gene, such as the branch-point sequence and polypyrimidine tract in MASP-3 which might suppress transcription and/or post-transcriptional factors that may destabilize the MASP-3 transcript (Stover *et al*., 2003).

By measuring the levels of mRNA of MASP enzymes in human tissues, it was established that the liver was the primary expression site for all MASP enzymes, including MASP-3. However, compared to the limited numbers of expression sites for MASP-1 and MASP-2, tissues from a wide range of organs contain relatively high

levels of MASP-3 mRNA, including the pancreas, skeletal muscles, spleen, thymus, prostates and ovaries (Seyfarth *et al*., 2006). In addition, an astrocyte-derived human glioma cell line T98G also expresses MASP-3 and MASP-1, indicating these enzymes may play a role in the lectin pathway in the brain (Kuraya *et al*., 2003). By using a monoclonal MASP-3-specific antibody, Degn *et al*. (2010) determined the mean concentration of MASP-3 in human plasma as 5.2 μg/ml, which is close to the values of 6.4 μg/ml, previously measured by Skjoedt *et al*. (2009). The plasma level of MASP-3 in human appears to be stable 6 months after birth (Degn *et al*., 2010).

1.6.2 The structure of MASP-3

The primary structure of MASP-3 implies an evolutionary relationship that is further apart from the chymotrypsin origin than MASP-1, but closer to MASP-2, C1r and C1s. Similar to MASP-2, C1r and C1s, the MASP-3 polypeptide consists of a serine protease domain encoded by a single exon and a catalytic S residue encoded by an AGY (Y denotes the C or T nucleotide base) codon. The 'histidine loop' present in MASP-1 is absent in the serine protease domains of MASP-3, MASP-2, C1r and C1s (Endo *et al*., 1998, Krem and Di Cerca, 2001, Arlaud and Gagnon, 1981).

MASP-3 shares the sequence of the non-catalytic region (CUB1-EGF-CUB2-CCP1-CCP2) of MASP-1. The crystal structure of the CUB1-EGF-CUB2 segment of MASP-3 therefore is the same as that of MASP-1, discussed in the previous section (Figure 1.5). Consequently, the 'head to tail' dimerization model of two MASP-1 monomers by their CUB1-EGF-CUB2 segments is also adaptable in the case of MASP-3 (Teillet *et al*., 2008). In agreement with this, MASP-3 homodimers were observed in human plasma (Dahl *et al*., 2001). However, a heterodimer formed between a MASP-3 monomer and a

MASP-1 monomer via their respective, but identical CUB1-EGF-CUB2 segments, has thus far not been observed.

MASP-3, together with MASP-2, was initially found to be in complex with MBL-II (Dahl *et al*., 2001). Later, Zundel *et al*. (2004) reported that MASP-3 bound to immobilized MBL-I and MBL-II with similar equilibrium dissociation and rate constants. Recently it was found that in human plasma MASP-3 is predominantly in complex with H-ficolin, with only a small proportion associated with L-ficolin and MBL (Skjoedt *et al*., 2009, Skjoedt *et al*., 2012). The structural details underpinning these different preferences in forming MBL/ficolin-MASP complexes are unclear. Previous studies found that MASP-1 binds to MBL and L-ficolin via its CUB1-EGF-CUB2 segment in a Ca^{2+} dependent manner (Thielens *et al*., 2001, Chen and Wallis, 2001, Cseh *et al*., 2002). It is therefore expected that MASP-3 binds to its recognition molecule partner in a similar manner. A study using point mutations to residues that participate in binding of Ca^{2+} in the CUB1-EGF-CUB2 segment of MASP-3 revealed that E^{49} in CUB1 domain, D^{102} and F^{103} in EGF domain and Y^{225} and H^{218} in CUB2 domain all play a role in the interaction between MASP-3 and its binding partners. Of these residues, the acidic E^{49} and D^{102} residues are most likely to be involved in direct binding to MBL/ficolin, since a previous study revealed that the positively-charged residue K^{55} of MBL/ficolin is involved in their binding to all three human MASP enzymes and MAp19 (Teillet *et al*., 2007).

The crystal structure of the serine protease domain of MASP-3 was yet to be solved at the outset of the present study. Sequence alignment of the serine protease domain of MASP enzymes revealed five strictly conserved C residues, including two that are involved in forming the ‘methionine loop’ (or loop 3). This loop structure is near the catalytic residues and thus may affect the substrate binding and/or cleavage by the

enzyme (Stover *et al*., 2003). The methionine loop of MASP-3 (SRSGNYSVTEN) consists of neither the K patch presented in that of MASP-1 (KKKVTRD), which was believed to hinder the substrate-binding property and cleavage efficiency of MASP-1, nor the frequently presented P residue in that of MASP-2 (EKPPYPRGSVTAN). Instead, the 'methionine loop' of MASP-3 contains an N-glycosylation site (GNY) at its centre, which is not presented in those of MASP-1 and MASP-2. These features of the MASP-3 'methionine loop' indicate that it exerts a unique impact on the substrate specificity and/or substrate-cleavage efficiency of the enzyme.

1.6.3 The activation of MASP-3

Thus far, the precise activation mechanism of MASP-3 remains unclear. However, the majority of the human MASP-3 protein in complex with MBL or ficolins extracted by affinity chromatography appears to be in the active form (Dahl *et al*., 2001). Whether the activation occurred prior to or during the affinity chromatography process was unclear. Zundel *et al*. (2004) observed that the full length recombinant MASP-3 was activated during prolonged storage at 4 °C. The activation was diminished by inhibitors against cysteine proteases and metalloproteases, which were likely from the protein-expressing insect cell line in the study. By incubating active recombinant MASP-3 and the zymogen of a MASP-3 mutant with an A residue replacing the active site S residue, the authors also found that the MASP-3 mutant was not cleaved by active MASP-3, proving that MASP-3 does not undergo autoactivation. However, Iwaki *et al*. (2011) demonstrated that the recombinant murine MASP-3 in complex with mouse recombinant MBL-A formed a two-chained form when it was incubated with heat-killed *S. aureus*. Interestingly, the latter authors also found that human MASP-1

processes murine recombinant MASP-3 into a two-chain form, and thus proposed that MASP-1 activates MASP-3. The proposal was supported by Degn *et al*. (2012) and by Megyeri *et al*. (2013), since both studies observed that human MASP-1 cleaved MASP-3. However, whether the cleavage of MASP-3 by MASP-1 was at the correct activation site is unclear.

1.6.4 The biological function of MASP-3

Since MASP-3 is present in all of the vertebrate species analysed (Fujita *et al*., 2002), including lamprey, cartilaginous fish, bony fish and birds in which MASP-1 is absent (Matsushita *et al*., 2000), MASP-3 may have an important physiological role. However, due to the uncertainty of the activation method and the lack of a crystal structure, the substrate specificity and precise physiological function of MASP-3 are still unknown. Thus far only a few synthetic peptidyl substrates and extremely few protein candidates were suggested to be cleaved by MASP-3. Zundel *et al*. (2004) found that MASP-3 activated by an unknown enzyme in an insect cell line can weakly cleave a synthetic thioester, Z-Gly-Arg-S-Bzl. Cortesio and Jiang (2006) reported that MASP-3 cleaves a few synthetic substrates, including Pro-Phe-Arg-AMC, Boc-Val-Pro-Arg-AMC, Z-Gly-Gly-Arg-AMC, Z-Phe-Arg-AMC and Z-Leu-Arg-AMC, among which Pro-Phe-Arg-AMC and Boc-Val-Pro-Arg-AMC displayed the highest rate of cleavage by MASP-3. The authors also found that MASP-3 cleaves insulin-like growth factor-binding protein (IGFBP)-5. The three cleavage sites in IGFBP-5 all contain an R/K residue at P1-position and a P residue at P2-position. Therefore, the authors claimed that the substrate specificity of MASP-3 is P-R/K-X. However, the main drawback of this study is that the active MASP-3 enzyme used was a truncated MASP-3 SP domain that started

with the I^{450} residue of the activation bond at the N-terminus. Since previous studies have identified the involvement of the CCP domains of C1s and MASP-2 in shaping the proteolytic activity of the enzymes (Duncan *et al*., 2012), it is uncertain whether the kinetic behaviour of the truncated SP domain of MASP-3 resembles that of a full-length MASP-3 or the catalytic portion of MASP-3.

MASP-3 does not cleave complement components C2, C3 and C4, and it is not inhibited by C1-inhibitor as MASP-1 and MASP-2 are (Zundel *et al*., 2004). The plasma levels of MASP-3 in six patients receiving surgery for colorectal cancer displayed a slight decline post-operatively and recovered to baseline within another two days. Hence it is unlikely that MASP-3 acts as an acute phase protein (Degn, *et al*., 2010). Dahl *et al*. (2001) and Møller-Kristensen *et al*. (2007) found that incubating MASP-3 with an MBL-MASP-2 complex resulted in a lower level of cleavage of C4 mediated by MASP-2. Skjoedt *et al*. (2009) also claimed that MASP-3 acts at a negative regulator to the cleavage of C4 mediated by an H-ficolin-MASP complex in human plasma. It was thought that these down-regulating impacts of MASP-3 are via competing with MASP-2 for binding to the recognition molecules. However, whether this represents the main physiological function of MASP-3 is unclear.

The observation that the *masp1/3*-null mouse had an impaired alternative complement pathway gave rise to the hypothesis that MASP-3 may be involved in activation of the alternative pathway of complement (Takahashi *et al*., 2010). Iwaki *et al*. (2011) reported that recombinant murine MASP-3 can process pro-factor D and pro-factor B, and thus activate the alternative pathway in mice. More interestingly, the authors claimed that even proenzyme of recombinant murine MASP-3 was able to cleave pro-factor B. Such functions of MASP-3 in humans were denied by Degn *et al*. (2012), who

found that the alternative complement pathway of a patient genetically deficient in MASP-1/3 was functionally normal.

Recent genetic studies have identified a critical correlation between genetic defects of components of lectin complement pathway and an autosomal recessive disorder, the Carnevale, Mingarelli, Malpuech and Michels (3MC) syndrome. The affected genes include *COLEC11* and *masp-1/3*. Mutations in the *COLEC11* gene disrupt the expression of CL-11, one of the important binding ligands of MASP-3, while mutations in the *masp-1/3* gene include the introduction of an stop codon in the gene area encoding the CUB1-EGF-CUB2 fragment common to MASP-3 and MASP-1, and four point mutations affecting the residues C^{630}, G^{666}, G^{687} and the active site H^{497} of MASP-3 SP domain (Rooryck *et al*., 2011, Sirmaci *et al*., 2010). These findings shed light on the possible role of MASP-3 in embryonic development. However, due to the lack of structural information and a good substrate to characterize the catalytic activity of MASP-3, the molecular basis of the impact of these mutations on the function of MASP-3 is only postulated.

1.7 MAP44

Like MASP-1 and MASP-3, the recently discovered MAp44 is another product of alternative splicing of the *masp-1/3* gene. MAp44 mRNA consists of the first 8 exons of the *masp-1/3* gene and an extra exon encoding 17 amino acids unique to MAp44. Consequently, MAp44 protein is around 41 kDa and consists of 361 amino acids, encompassing the CUB1, EGF, CUB2 and CCP1 domains which are identical to those of MASP-1 and MASP-3, followed by the characteristic 17 amino acids (Degn *et al*., 2009).

By measuring the level of MAp44 mRNA, it was found that MAp44 is a highly specific protein which is predominantly expressed in the heart and skeletal muscle, followed by the liver. Further, by using MAp44-specific antibodies, MAp44 protein was strictly detected in the myocardial and skeletal muscles and liver tissue, but not in the smooth muscle tissue of aorta, giving rise to a possibility that MAp44 could be induced by contraction of striated muscles (Skjoedt *et al*., 2010). The plasma concentration of MAp44 is 1.4 μg/mL, determined by using an antibody recognizing the 17 characteristic C-terminal residues of MAp44. Further, similar to MASP-3, MAp44 is also most unlikely to be an acute phase protein, since it was found that the concentration of MAp44 in human serum following surgery only dropped slightly, and recovered in about a month. This finding reflects the highly tissue-specific distribution of MAp44 (Skjoedt *et al*., 2010).

The crystal structure of MAp44 was recently solved (Skjoedt *et al*., 2012). Monomers of MAp44 form head-to-tail dimers similar to the CUB1-EGF-CUB2 fragment of MASP-1/3. Increasing concentrations of recombinant MAp44 in complex with either recombinant MBL or recombinant H-ficolin resulted in decreasing levels of cleavage of C4 in sera deficient in MBL and H-ficolin, respectively. These results suggest that MAp44 may inhibit the complement activation via down-regulating cleavage of C4 (Skjoedt *et al*., 2010). It was also found that MAp44 competed with all MASP enzymes for the binding of H-ficolin, resulting in decreased activation of the lectin complement pathway (Skjoedt *et al*., 2012).

1.8 PROJECT AIMS

Mounting evidence has shown that the lectin pathway plays an important role in complement activation. Imbalanced activation of the lectin pathway due to gene mutations in the components of this pathway has been associated with autoimmune

diseases and repeated infections. Whilst the MASP-1 and MASP-2 enzymes of the lectin pathway have been intensively studied, very little is known about the structure and function of MASP-3. This project aims to gain an understanding of the structure of the enzyme and its possible biological function.

Aims:

(1) Revealing the activation mechanism of MASP-3.

(2) Characterization of the substrate specificity of MASP-3.

(3) Determination of the crystal structure of MASP-3.

Chapter Two

General Materials and Methods

2.1 GENERAL MATERIALS

The following materials were purchased from Sigma-Aldrich, Australia (Castle Hill, NSW): acrylamide, ampicillin, β-mercaptoethanol, bis-acrylamide, bis-tris propane, bromophenol blue, CAPS, chloramphenicol, Coomassie blue R-250, DMSO, ethanol, ethanolamine, ethidium bromide, ethylene glycol, glacial acetic acid, glycerol, HCl, heparin, iodoacetamide, L-cysteine, MPD, NaOH, PEG 400–20,000, rabbit anti-chicken antibody, silver nitrate, sodium azide, sodium HEPES and succinic acid. Sodium acetate and sodium carbonate were from Ajax Finechem (Taren Point, NSW). TEMED was from MP Biomedicals (Seven Hills, NSW). Agarose, ammonium persulfate, ammonium sulphate, DTT, EDTA, glycine, HEPES, IPTG, L-arginine, magnesium chloride, NaCl, oxidised/reduced glutathione, SDS, sodium carbonate, Tris, triton X-100, Tween 20 and urea were purchased from Amresco, US (Solon, OH). Acetylnitrile, agar, anhydrous citric acid, bacto tryptone, calcium chloride, formalderhyde, magnesium sulphate, sodium phosphate, potassium phosphate and yeast extract were purchased from Merck, Australia (Kilsyth, VIC). The 96-well Costar® white polystyrene assay plates were from Corning Life Science (Tewksbury, MA). The PCR reaction buffer, dNTPs, pfu polymerase, T4 ligase and 1kb DNA marker were from Promega (Madison, WI). *NheI*, *EcoRI* and *DpnI* restriction enzymes were from New England BioLabs (Ipswich, WA). All primers were produced by Geneworks (Hindmarsh, SA). The pET17b vector was from EMD Biosciences (Rockland, MA). The Precision plus protein™ standard and the Sequi-Blot™ PVDF membrane were from BioRad (Herculus, CA). HiTrap™ QFF, HiTrap™ NHS, and HiLoad™ superdex 75 16/60 columns and Amersham ECL reagent were from GE Healthcare, Australia (Sydney, NSW). All Vivaflow concentrator sets were purchased from Sartorius Stedim Biotech, US (Bohemia, NY). Except where specified, all chemicals and materials used in this work were of analytical grade of purity.

2.2 SDS-PAGE

A polyacrylamide SDS gel composed of a 12.5% running gel and a 4% stacking gel was prepared as per Table 2.1. The gel was fitted into a BioRad Minigel system (Hercules, CA). Samples were mixed with 25% (v/v) 4 X loading buffer (200 mM Tris [pH 6.8], 40% [v/v] glycerol, 0.8% [w/v] SDS, 0.02% [w/v] bromophenol blue, and for reduced condition, add 4% [v/v] β-mercaptoethanol), boiled for 5 min, loaded into the wells and electrophoresed at a constant voltage of 100 V in the SDS-PAGE running buffer (0.25 M Tris-HCl, 0.192 M glycine, 0.1% [w/v] SDS, pH 8.3). Once SDS-PAGE was completed, the proteins in the gel could be visualized by staining the gel in a Coomassie blue solution (40% [v/v] methanol, 7% [v/v] glacial acetic acid, 0.025% [w/v] Coomassie R-250 Brilliant Blue) for 30 min, followed by destaining in a solution containing 40% (v/v) methanol and 7% (v/v) glacial acetic acid to generate a clear background.

Table 2.1 Preparation of a 12.5% running gel and a 4% stacking gel

Material	Running gel	Stacking gel
Acrylamide/bis-acrylamide 30% (v/v) (ml)	4.2	1.0
1.5 M Tris, pH8.8 (ml)	2.5	
1.0 M Tris, pH6.8 (µl)		760
10% SDS (w/v) (µl)	100	60
Distilled water (ml)	3.2	8.4
10% Ammonium persulphate (w/v) (µl)	50	60
TEMED (µl)	3	3

2.3 WESTERN BLOT

Following SDS-PAGE, the proteins in the gel were transferred onto a BioTrap®NT nitrocellulose membrane from PALL Australia (Cheltenham, VIC) at a constant voltage of 100 V for 1 hr in the Western Blot transfer buffer (0.25 M Tris-HCl, 0.147 M glycine, 20% [v/v] methanol). The membrane was blocked with 3% (w/v) skim milk in TBS buffer (50 mM Tris, 150 mM NaCl, pH 7.5) for 1 hr, followed by probing with an appropriate primary antibody overnight at 4°C to recognize the protein of interest. After three 10 min washes in TBS buffer with 0.1% (v/v) Tween-20, the membrane was then probed with a horseradish peroxidase-conjugated secondary antibody which binds the primary antibody. After washing with TBS buffer, blots were developed with ECL reagent and exposed to a Fuji SuperRX medical X-ray film, which was finally developed in a Konica Minolta SRX-101A medical film processor (Tokyo, Japan).

2.4 CONSTRUCTION OF RECOMBINANT PLASMIDS FOR EXPRESSION OF RECOMBINANT C1r, MASP-1 AND MASP-2 PROTEINS

The cDNA for the catalytic portions of C1r (residues R^{296}-D^{705}), MASP-1 (residues A^{297}-N^{689}) and MASP-2 (residues T^{287}-F^{686}) were purchased (GeneScript, Piscataway, NJ). The gene was cloned into pET17b. The construct was then transformed into the DH5α *E. coli* cells, and the transformants were selected on LB plates containing 100 μg/ml ampicillin. Following DNA extraction by using the PureYield™ Plasmid Miniprep kit from Promega (Madison, WI), the recombinant constructs were sent to DNA sequencing to confirm the sequences.

2.5 EXPRESSION, UNFOLDING AND REFOLDING, AND PURIFICATION OF THE RECOMBINANT C1r, MASP-1 AND MASP-2 PROTEINS

The constructed plasmids were transformed into the One Shot® BL21(DE3)pLysS *E. coli* host strain, purchased from Invitrogen™ Life Technologies, Australia (Mulgrave, VIC). Single colonies of transformants were selected on LB plates containing 100 μg/mL ampicillin and 50 μg/mL chloramphenicol and were grown overnight at 37ºC. 25 mL overnight culture was then transferred into 1 L 2YT medium (1.6% [w/v] tryptone, 1.0% [w/v] yeast extract, 0.5% [w/v] NaCl) and further grown to an OD_{600} value of 0.6~0.8. IPTG was then added to the culture to yield a final concentration of 1 mM to induce the expression of proteins for 4 hr. The induced cells were collected by centrifugation and were subsequently resuspended in 25 mL Tris-EDTA buffer (50 mM Tris, 20 mM EDTA, pH 7.4), and were frozen at -80ºC. After thawing, the cells were then sonicated to release the inclusion bodies, which were collected by centrifugation at 27,000 X g for 20 min at 4ºC. After three sequential washes (first, Tris-EDTA buffer with 20 mM DTT and 1% [v/v] triton X-100; second, Tris-EDTA buffer with 1 M NaCl; third, Tris-EDTA buffer), the inclusion bodies were solubilised in 8 M urea, 0.1 M Tris-HCl with 100 mM DTT, pH 8.3, at RT for 3 hr. The solubilised proteins were refolded overnight in 5 L refold buffer (50 mM Tris-HCl, 5 mM EDTA, 500 mM L-arginine, 300 mM NaCl, 3 mM reduced glutathione and 1 mM oxidised glutathione, pH 9.0). The 5 L of refolded protein was then concentrated to 1/10 of the volume using Vivaflow 200 concentration cassettes. After dialysis against 20 mM Tris-HCl, pH 9.0 overnight at RT, the refolded proteins were purified by ion exchange chromatography followed by size-exclusion chromatography (Figures 2.1-2.3). First, the refolded proteins were separated on a QFF column, which was equilibrated with 20 mM Tris-HCl, pH 9.0 before they were loaded. The elution was conducted with a linear NaCl

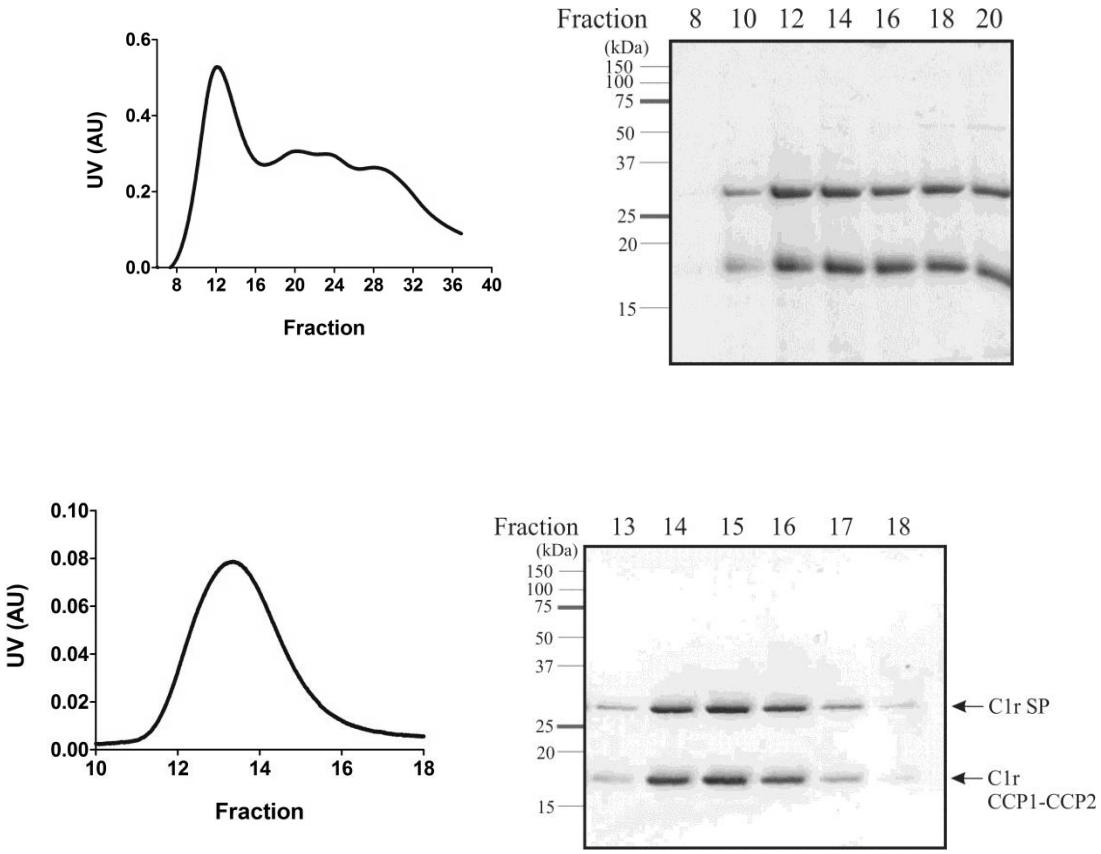


Figure 2.1 Purification of the recombinant C1r CCP1-CCP2-SP protein
Top panel, anion exchange chromatography of the recombinant C1r. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE analysis of the protein contents in the anion exchange fractions under reduced conditions. The reduced sample of auto-activated C1r displayed a band of 32 kDa, representing the SP domain, and another band of 17 kDa, which is presumed to represent the CCP1-CCP2 fragment of C1r. Each lane contained a 30 µl sample from the indicated fraction. Based on SDS-PAGE analysis, the C1r components in fractions 10-14 were relatively pure, and hence were pooled for further purification by size-exclusion chromatography.
Bottom panel, size-exclusion chromatography of the protein contents of previous anion exchange fractions 10-14. Left, chromatogram of gel filtration on a Superdex 75 16/60 column. Right, an SDS-PAGE analysis of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 30 µl sample from the indicated fraction. Based on SDS-PAGE analysis, fractions 13-17 contained C1r at high purity. The separated protein fragments are indicated by arrows.

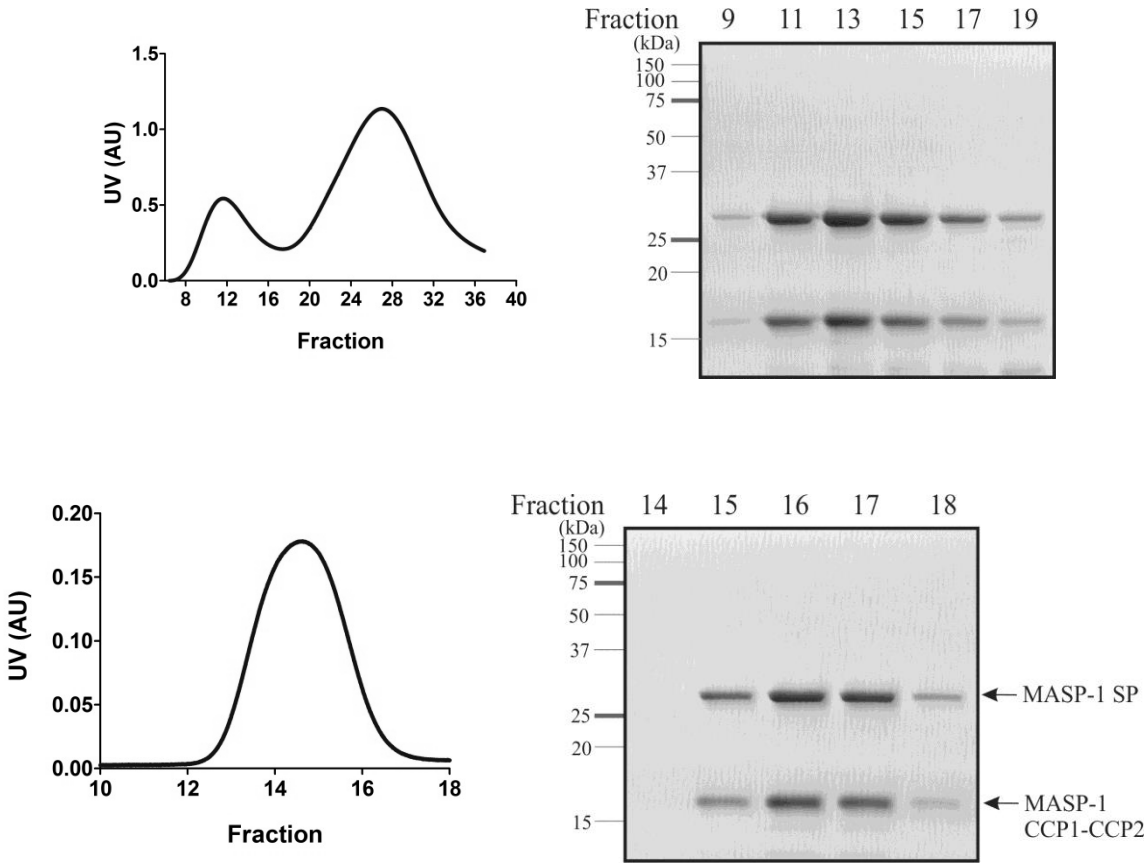


Figure 2.2 Purification of the recombinant MASP-1 CCP1-CCP2-SP protein
Top panel, anion exchange chromatography of the recombinant MASP-1. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE analysis of the protein contents in the anion exchange fractions under reduced conditions. The reduced sample of auto-activated MASP-1 displayed a band of 32 kDa, representing the SP domain, and another band of 17 kDa, which is presumed to represent the CCP1-CCP2 fragment of MASP-1. Each lane contained a 20 µl sample from the indicated fraction. Based on SDS-PAGE analysis, the MASP-1 components in fractions 9-13 were relatively pure, and hence were pooled for further purification by size-exclusion chromatography.
Bottom panel, size-exclusion chromatography of the protein contents of previous anion exchange fractions 9-13. Left, chromatogram of gel filtration on a Superdex 75 16/60 column. Right, an SDS-PAGE analysis of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 30 µl sample from the indicated fraction. Based on SDS-PAGE analysis, fractions 15-18 contained MASP-1 at high purity. The separated protein fragments are indicated by arrows.

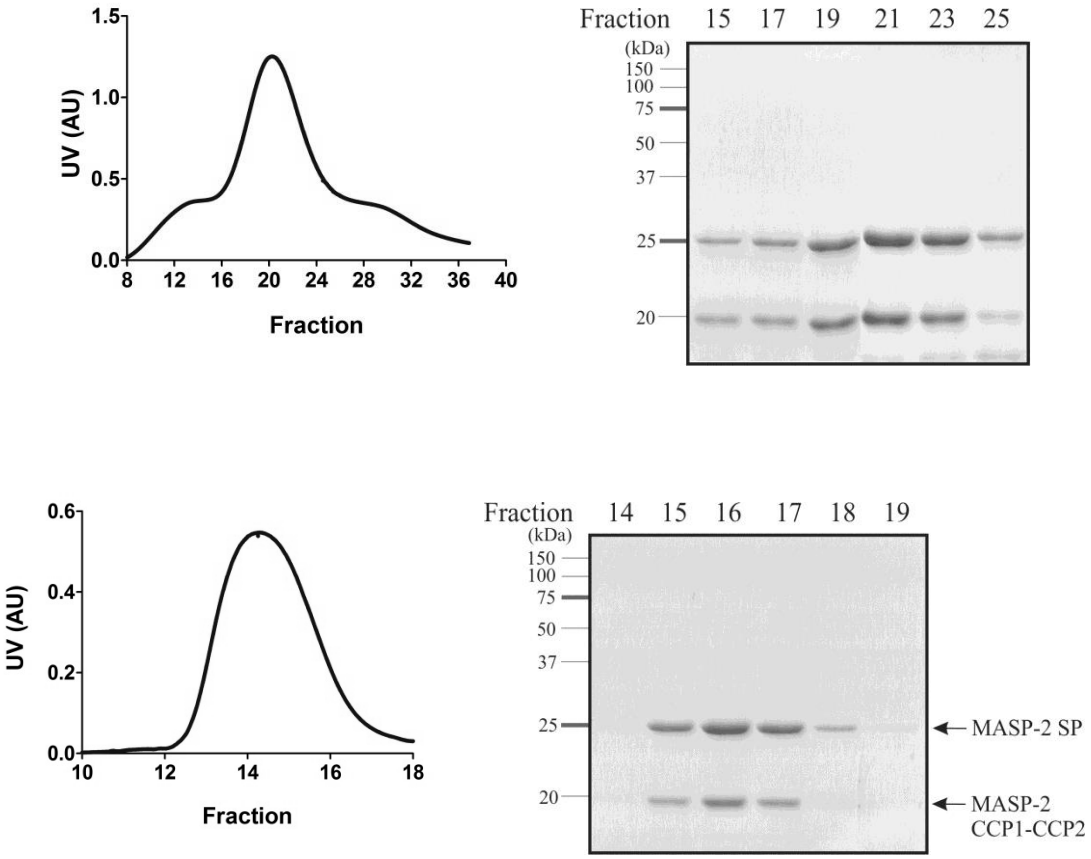


Figure 2.3 Purification of the recombinant MASP-2 CCP1-CCP2-SP protein
Top panel, anion exchange chromatography of the recombinant MASP-2. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE analysis of the protein contents in the anion exchange fractions under reduced conditions. The reduced sample of auto-activated MASP-2 displayed a band of 25 kDa, representing the SP domain, and another band of 19 kDa, which is presumed to represent the CCP1-CCP2 fragment of MASP-2. Each lane contained a 10 µl sample from the indicated fraction. Based on SDS-PAGE analysis, the MASP-2 components in fractions 19-23 were relatively pure, and hence were pooled for further purification by size-exclusion chromatography.
Bottom panel, size-exclusion chromatography of the protein contents of previous anion exchange fractions 19-23. Left, chromatogram of gel filtration on a Superdex 75 16/60 column. Right, an SDS-PAGE analysis of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 10 µl sample from the indicated fraction. Based on SDS-PAGE analysis, fractions 15-17 contained MASP-2 at high purity. The separated protein fragments are indicated by arrows.

gradient from 0 to 400 mM over 35 mL at 1 mL/min. SDS-PAGE was used to analyse the protein fractions. Fractions containing the majority of recombinant protein were further purified by a Superdex 75 16/60 column in the gel filtration buffer containing 50 mM Tris-HCl, 145 mM NaCl, pH 7.4. Based on analysis by SDS-PAGE, the gel filtration fractions containing the protein of interest were pooled, aliquoted, snap frozen and stored at -80°C. The concentrations of the recombinant C1r, MASP-1 and MASP-2 proteins were derived by dividing the OD_{280} value by the calculated absorption coefficient ($\varepsilon_{0.1\%,\ 1\ cm,}$ assuming all pairs of Cys residues form disulfide bonds). The $\varepsilon_{0.1\%,\ 1\ cm}$ value of each recombinant protein and its calculated molecular mass from its amino acid sequence are listed in Table 2.2.

Table 2.2 The calculated masses & absorption coefficients of the recombinant proteins

Protein fragments	Calculated mass (Da)	$\varepsilon_{0.1\%,\ 1\ cm}$
C1r CCP1-CCP2-SP	46,790	1.68
MASP-1 CCP1-CCP2-SP	45,144	1.55
MASP-2 CCP1-CCP2-SP	43,745	1.96

2.6 N-TERMINAL SEQUENCING PREPARATION

Following SDS-PAGE, the proteins was transferred onto a PVDF membrane at a constant voltage of 100 V for 1 hr in a transfer buffer containing 10 mM CAPS (pH 11) with 15% (v/v) methanol. The membrane was then stained in a stain solution of 0.025% (w/v) Coomassie R-250 Brilliant Blue with 40% (v/v) methanol, for 30 min, followed by destaining with 50% (v/v) methanol until the protein bands stood out from a clear background. The protein bands of interest were carefully cut out of the membrane and sent to the Monash Proteomics Facility (Monash University, Clayton) for N-terminal sequencing.

Chapter Three

The Activation of The Catalytic Portion of MASP-3

3.1 INTRODUCTION

3.1.1 A new sight into the activity of C1r protease

As introduced in Chapter 1, human MASP-3 does not undergo autoactivation and the enzyme that activates it remains unclear. In order to study the function of MASP-3, an alternative approach to produce a recombinant MASP-3 with a fully functional SP domain was necessary for this project. As mentioned previously, MASP enzymes and the classical complement pathway serine proteases, C1r and C1s, are members of the C1r/C1s/MASP family of enzymes. They share identical domain organization and a strictly conserved R-I activation bond in their proenzyme form, which needs to be cleaved to yield the active form of enzyme (Endo *et al*., 1998). Sequence alignment of members of the enzyme family reveals that C1r, C1s and MASP-3 share the identical P1'-P4' sequence (IIGG) at their activation sites (Figure 3.1). Of the three enzymes, C1r is the only serine protease that undergoes autoactivation, and the activated C1r protease is highly specific for cleavage and activation of the C1s proenzyme (Kardos *et al*., 2001).

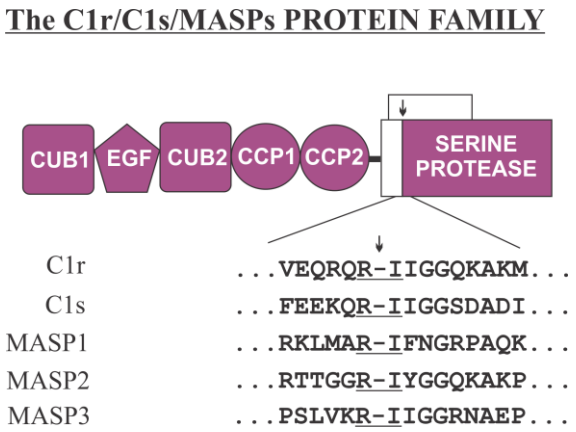


Figure 3.1 The sequence alignment of protein members of the C1r/C1s/MASPs family at the activation site
The domain organization of CUB1-EGF-CUB2-CCP1-CCP2-SP is shown schematically with a closed line representing the disulfide bond that connects the SP domain to the N-terminal fragment. Arrows represent the R-I activation bond, which is strictly conserved in all members of the family.

C1r is the initiating serine protease of the classical pathway of complement activation. Molecules of C1r proenzyme form a head-to-tail homodimer via the CCP1-CCP2-SP fragment, in which the SP domain of one C1r molecule interacts with the CCP1 domain of another (Budayova-Spano *et al*., 2002b). The homodimer of C1r proenzyme is proposed to provide the core to bind C1s proenzyme molecules to form the C1s-C1r-C1r-C1s heterotetramer, which is bound to the recognition molecule of the classical pathway, C1q, in a Ca^{2+}-dependent manner to form the C1 complex (Arlaud *et al*., 2001). Previous studies have identified that the N-terminal CUB1-EGF-CUB2 fragment of C1r provides Ca^{2+} binding sites for the interaction of C1r with C1s and C1q (Gaboriaud *et al*., 2007), and that the CCP1-CCP2-SP fragment of C1r is responsible for the catalytic activity of the C1r molecule (Arlaud *et al*., 1986). A further study by Kardos *et al*. (2001) showed that the CCP1 domain of C1r is not involved in the catalytic activity of C1r, whilst the CCP2 domain provides an exosite for substrate binding. When the C1 complex recognizes the targeting structures via the C-terminal globular domains of C1q, it is believed that the recognition binding induces a conformational change to the collagen-like N-terminal tri-helical stems of C1q. The molecular events of how this conformational change of C1q is converted to the intermolecular cleavage of C1r are still debated. In the latest model postulated by Kardos *et al*. (2008), the structural alteration of C1q causes a transient disassociation between the two C1r molecules in the C1r homodimer. The two C1r molecules mutually act as substrate and enzyme to each other, by adapting the exosite interaction between the CCP2 domain of the substrate C1r and the SP domain of the enzymatic C1r. Following autoactivation, the active C1r molecules reform the dimeric structure and subsequently activate C1s proenzymes in the same C1 complex. This model was supported by the demonstration of the Ca^{2+}-dependent flexibility of CUB2 domain of

C1r, which makes it possible for the CCP1-CCP2-SP fragments of C1r molecules to act on each other in the limited space within a C1q molecule (Major *et al*., 2010).

C1r has a clear specificity for the P1-R residue (Budayova-Spano *et al*., 2002a). The results of phage display to characterize the substrate specificity of C1r by Wijeyewickrema *et al*. (2013) indicated that C1r highly prefers substrates with P2-Q and P1'-I residues. This finding is evidenced by the presence of the motif -QRI- at the activation site of both C1r and C1s proenzymes, the two most important substrates of C1r. Further, substituting the P2-Q residue at the activation site of C1s zymogen with either an N or a G residue resulted in severely decreased cleavage by C1r, confirming that the P2-Q residue acts as a molecular determinant for the interaction between C1r and C1s. It can also be deduced that the P2-Q residue is also likely to be critical for the interaction between the two C1r molecules during autoactivation. The finding also explains why C1r does not activate wild type MASP-3, despite MASP-3 containing P1'-P4' residues (IIGG) at its activation site that are identical to those of C1r and C1s. It was therefore hypothesized that introducing the P4-P1 or simply the P2-Q residue found in C1s into the corresponding positions in MASP-3 would enable C1r to cleave and activate the MASP-3 mutants so constructed. Since the sequence C-terminal to the activation bond of such MASP-3 mutants represents that of wild type MASP-3, the C1r-activated MASP-3 mutants would also therefore contain an SP domain resembling that of a wild type enzyme.

3.1.2 The possible involvement of proprotein convertases in activation of MASP-3

The sequence of the activation site of MASP-3 contains a motif of dibasic residues (-KR-) adjacent to the activation bond at the N-terminal side, a common feature shared by many precursors of hormones and enzymes in humans. Protein precursors with such

characteristic activation bonds are specially activated by a group of subtilisin-like enzymes called proprotein or prohormone convertases (PCs). The human PC family of enzymes comprises nine members (Table 3.1, adapted from Seidah and Prat, 2012, review). Of these, PC1/3, PC2, furin, PC4, PC5/6, paired basic amino acid cleaving enzyme 4 (PACE4), and PC7 cleave after single or paired basic amino acids, whilst subtilisin kexin isozyme 1 (SKI-1/S1P) and proprotein convertase subtilisin kexin 9 (PCSK9) cleave at non-basic residues. Activities of PCs are involved in a diverse range of physiological events and biological pathways, both in healthy and disease states (Seidah and Prat, 2012, review). Since the expression of MASP-3 in humans is generally ubiquitous and the activation of MASP-3 is most likely not an intracellular event, the activation partner of MASP-3 may be a widely expressed secreted or cell membrane-bound protein. The PC members that fit these criteria include furin, PC5/6, and PACE4, although the other secreted PC members may also be involved in the activation of MASP-3 in a 'tissue-specific' manner: this topic may be of interest for future studies.

Table 3.1 Members of the human PC family enzymes

Protein name	Gene name	Tissue distribution	Secretion
PC1/3	*PCSK1*	Neuroendocrine	Secreted
PC2	*PCSK2*	Neuroendocrine	Secreted
furin	*PCSK3*	Ubiquitous	Membrane-bound
PC4	*PCSK4*	Germinal	Membrane-bound
PC5/6	*PCSK5*	Widespread: adrenal cortex, intestine, kidney, ovary	Secreted, bound
PACE4	*PCSK6*	Widespread: muscle, heart, pituitary, intestine, cerebellum, kidney	Secreted
PC7	*PCSK7*	Ubiquitous	Intracellular
SKI-1/S1P	*MBTPS1*	Ubiquitous	Intracellular
PCSK9	*PCSK9*	Liver, intestine, kidney	Secreted

Like MASP-3, furin, PACE4 and PC5/6 all appear to be vital factors for animal development. Murine embryos with a genetic knockout of furin died due to malformation of heart and gut tissues, resulting from a failure of ventral closure and

axial rotation (Roebroek *et al*., 1998). Genetic knockout of PCSK5 in mice, which encodes PC5/6, caused lethality at birth with severe malformations in the anteroposterior axis (Essalmani *et al*., 2008). Further, mutations in human PCSK5 have been correlated to the rare developmental disorder, VACTERL (vertebral, anorectal, cardiac, tracheoesophageal, renal and limb) syndrome (Szumska *et al*., 2008). PACE4-knockout mice displayed a lethality of 25% in murine embryos with defects in heart development, the left and right axis and bone morphogenetic features (Constam and Robertson, 2000).

3.1.3 The aims

The first aim of this part of project was to express and purify the catalytic portion of wild type MASP-3 (wtM3) in an *E. coli* cell line and test the cleavage of wtM3 by a series of enzymes from the complement and coagulation systems. Cleavage of MASP-3 by MASP-1 and MASP-2 was further studied by using a MASP-3 mutant with an A residue replacing the R^{454} residue (Figure 3.2). In order to develop a MASP-3 model which can be activated effectively and contains a fully functioning SP domain of wild type MASP-3, the catalytic portions of two MASP-3 mutants, M3EEKQ and M3Q (Figure 3.2), were engineered and purified. M3EEKQ contained the P5-P4' sequence (EEKQRIIGG) identical to that of C1s, and thus contained an activation site highly similar to that of C1s, whilst the M3Q mutant contained only a substituted P2-Q residue at its activation site. The activation of wtM3, M3EEKQ and M3Q by C1r was examined. In addition, the activation of the three recombinant proteins by members of the PC enzyme family was also investigated.

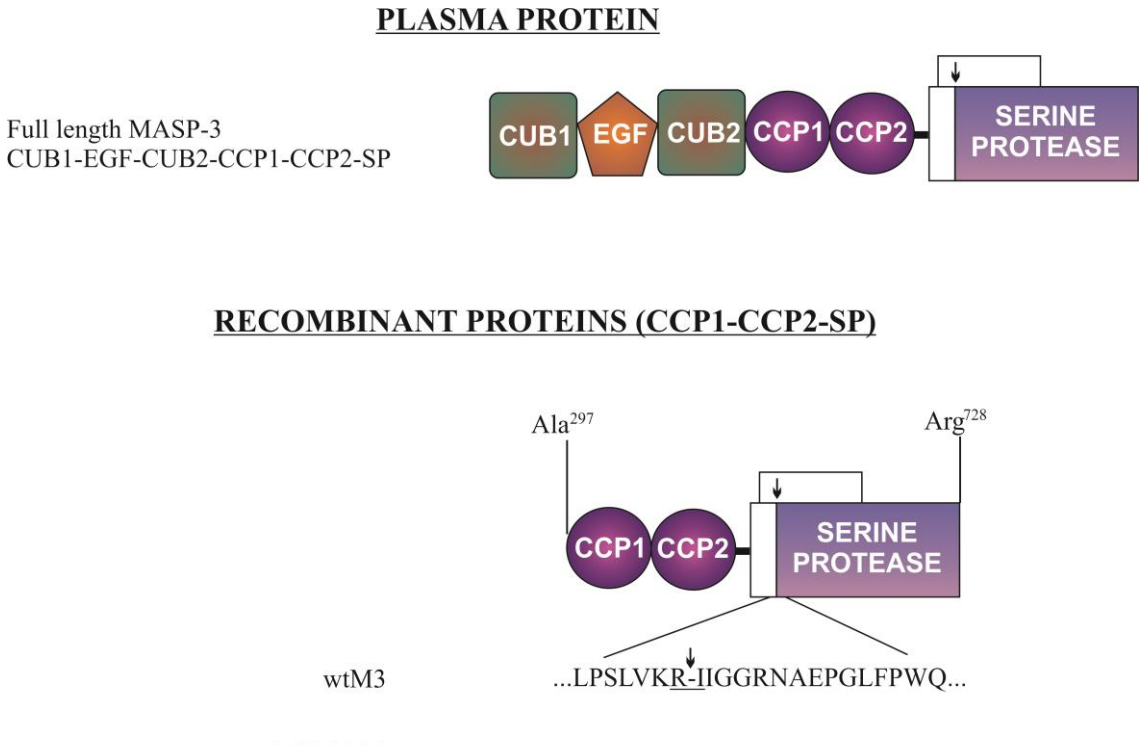


Figure 3.2 Schematic representation of recombinant MASP-3
The CCP1-CCP2-SP portion of MASP-3 represents the catalytic portion of the enzyme. The black arrow indicates the Arg-Ile activation bond for activation of MASP-3. The disulfide bond linking the SP domain to the CCP1-CCP2 fragment in active MASP-3 is indicated as the closed black line. The amino acid composition flanking the activation bond for each recombinant protein is also shown, with mutated residues highlighted in red.

3.2 MATERIALS AND METHODS

3.2.1 Materials

The catalytic portions of C1r, MASP-1 and MASP-2 enzymes were expressed and purified as described in Chapter 2. The catalytic portions of C1s protease and thrombin were purified and activated by members of Prof. Robert Pike's lab of Monash University (Clayton, VIC). A monoclonal antibody recognizing the unique peptide sequence, NPNVTDQIISSGTRT, of the MASP-3 SP domain was raised in chickens as previously described (Goldring *et al*., 2005) by Prof. Theresa Coetzer (University of Kwa-Zulu-Natal, South Africa), and was used as the primary antibody in western blotting to detect the wild type and mutants of MASP-3 protein. The PACE4 enzyme from PhenoSwitch Bioscience (Menlo Park, CA) and the DMEM/F-12 buffer were generously provided by Prof. Guiying Nie of Prince Henry's Institute for Medical Research (Clayton, VIC). Furin was purchased from New England Biolabs (Ipswich, MA). All enzymes tested were confirmed to be active by using specific synthetic peptidyl substrates.

3.2.2 Construction of recombinant plasmids for expression of recombinant wtM3

The DNA for the catalytic portions of wild type human MASP-3 (residue A^{296}-R^{728}) was synthesized by GeneScript (Piscataway, NJ). A short sequence of codons encoding three amino acids (A-S-M) of the T7-Tag was introduced prior to the sequences of the recombinant proteins to enhance protein expression. The gene was cloned into pET17b. The construct was then transformed into DH5α *E. coli* cells and the transformants were selected on LB plates containing 100 μg/ml ampicillin. Following DNA extraction by using the PureYield^{TM} Plasmid Miniprep kit from Promega (Madison, WI), the recombinant constructs were sent for DNA sequencing to confirm the sequences.

3.2.3 Site directed mutagenesis

The parental wtM3 DNA inserted into the pET-17b vector was used as a template. A set of forward and reverse primers encoding for the mutated sites was used for PCR to generate the constructs of the mutant. The nucleotide sequences of the primers for each MASP-3 protein mutant are listed in Table 3.2. The PCR products were processed with *DpnI* restriction enzyme to digest the parental DNA. The constructed plasmids were transformed into DH5α *E. coli* cells and the transformants were selected on LB plates containing 100 μg/ml ampicillin. Following DNA extraction by using the PureYieldTM Plasmid Miniprep kit from Promega (Madison, WI, USA), the recombinant constructs were sent to DNA sequencing for confirm the sequences.

Table 3.2 Primers for MASP-3 CCP1-CCP2-SP mutants

Mutant	Forward Primer	Reverse primer
M3R454A	5′GGTCAAGAGGATCATTGGGGG CGCCAATGC3′	5′GGCCAGGCTCAGCATTGGCGC CCCCAATGA3′
M3EEKQ	5′GGTCAGCCCTCCCGCTCCCTG CCAGAAGAAAAACAGAGGATCAT TGGGGGCCGAAATGCTGAGCC3′	5′GGCTCAGCATTTCGGCCCCCA ATGATCCTCTGTTTTTCTTCTGG CAGGGAGCGGGAGGGCTGACC3′
M3Q	5′CCCTGCCAAGCCTGGTCCAGA GGATCATTGGGGGC3′	5′GGCCCCCAATGATCCTCTGGA CCAGGCTTGGCAGGG3′

3.2.4 Expression, unfolding, refolding and purification of recombinant MASP-3 proteins

The constructed plasmids were transformed into the One Shot® BL21STAR(DE3) *E. coli* host strain, purchased from Invitrogen™ Life Technologies, Australia (Mulgrave,

VIC). Single colonies of transformants were selected on LB plates containing 100 μg/mL ampicillin and were grown overnight at 37ºC. A 25 mL overnight culture was then transferred into 1 L of 2YT medium (1.6 % [w/v] tryptone, 1.0 % [w/v] yeast extract, 0.5 % [w/v] NaCl) and further grown to an OD_{600} value of 0.6~0.8. IPTG was then added to the culture to yield a final concentration of 1 mM to induce the expression of proteins for 4 hr. Cells were collected by centrifugation and were subsequently resuspended in 25 mL Tris-EDTA buffer (50 mM Tris, 20 mM EDTA, pH 7.4), and were stored at -80ºC. After thawing, the cells were sonicated to release the inclusion bodies, which were collected by centrifugation at 27,000 X g for 20 min at 4ºC. After three sequential washes (first, Tris-EDTA buffer with 20 mM DTT and 1% [v/v] triton X-100; second, Tris-EDTA buffer with 1 M NaCl; third, Tris-EDTA buffer), the inclusion bodies were solubilised in 8 M urea, 0.1 M Tris-HCl with 100 mM DTT, pH 8.3, at RT for 3 hr. The solubilised proteins were refolded overnight in 5 L of refolding buffer (50 mM Tris-HCl, 5 mM EDTA, 500 mM L-arginine, 300 mM NaCl, 3 mM reduced glutathione and 1 mM oxidised glutathione, pH 9.0). The 5 L of refolded protein was then concentrated to 1/10 of the volume using Vivaflow 200 concentration cassettes. After dialysis against 20 mM Tris-HCl, pH 9.0 overnight at RT, the refolded proteins were purified by ion exchange chromatography followed by size-exclusion chromatography. First, the refolded proteins were separated on a QFF column, which was equilibrated with 20 mM Tris-HCl, pH 9.0 before they were loaded. The elution was conducted with a linear NaCl gradient from 0 to 400 mM over 35 mL at 1 mL/min. SDS-PAGE and western blotting were used to analyse the protein fractions. Fractions containing the majority of the recombinant protein were further purified by a Superdex 75 16/60 column in the gel filtration buffer containing 50 mM Tris-HCl, 145 mM NaCl, pH 7.4. Based on analyses by SDS-PAGE and western blotting, the gel filtration

fractions containing the protein of interest were pooled, aliquoted, snap frozen and stored at -80°C. The concentration of the recombinant MASP-3 proteins was derived by dividing the OD_{280} value by the calculated absorption coefficient ($\varepsilon_{0.1\%,\ 1\ cm,}$ assuming all pairs of Cys residues form disulfide bonds). The $\varepsilon_{0.1\%,\ 1\ cm}$ value of each recombinant protein and its calculated molecular mass from its amino acid sequence are listed in Table 3.3.

Table 3.3 The calculated masses & absorption coefficients of the recombinant proteins

Protein fragments	Calculated mass (Da)	$\varepsilon_{0.1\%,\ 1\ cm}$
wtM3 CCP1-CCP2-SP	48,047	1.69
M3EEKQ CCP1-CCP2-SP	48,134	1.69
M3Q CCP1-CCP2-SP	48,047	1.69
M3R454A CCP1-CCP2-SP	47,962	1.69

3.2.5 Preparation of a C1r column and activation of M3EEKQ/M3Q proteins

2.5 mg active recombinant C1r CCP1-CCP2-SP was coupled to a 5 ml HiTrap NHS-activated HP column from GE Healthcare (Uppsala, Sweden) by following the manufacture instruction to generate a C1r-column. To activate M3EEKQ and M3Q proteins, 2.5 mg of either MASP-3 mutant in 5 ml gel filtration buffer was loaded onto the C1r-column. The activation was carried out at 26°C for 18 hr. The specified temperature was selected as a compromise based on recommendations by the column manufacturer to ensure the best performance of the immobilised enzyme as possible whilst avoiding damage to the matrix due to prolonged exposure to higher temperatures, such as 37°C. The 18 hr incubation time was the minimum time required to activate the maximum amount of loaded MASP-3 mutant, determined by a preliminary study of the time-course cleavage of M3EEKQ/M3Q by the C1r column (data not shown). The activated MASP-3 mutant protein was eluted using 6.5 ml gel filtration buffer. SDS-PAGE, western blotting and N-terminal sequencing were carried out to confirm that the eluted protein was the active MASP-3 mutant.

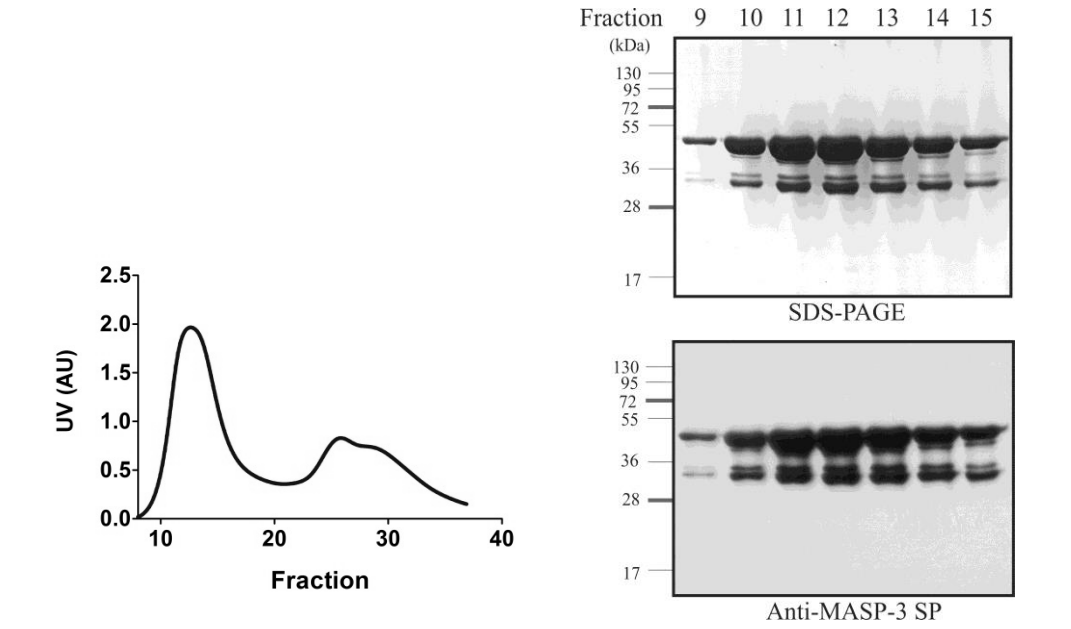
3.2.6 Prediction of cleavage sites of wtM3 by PC enzymes

The prediction of cleavage of wtM3 by members of the PC family was conducted by inputting the FASTA sequence of wtM3 into the ProP1.0 Server programme, developed by Technical University of Denmark. The programme is available on http://www.cbs.dtu.dk/services/ProP .

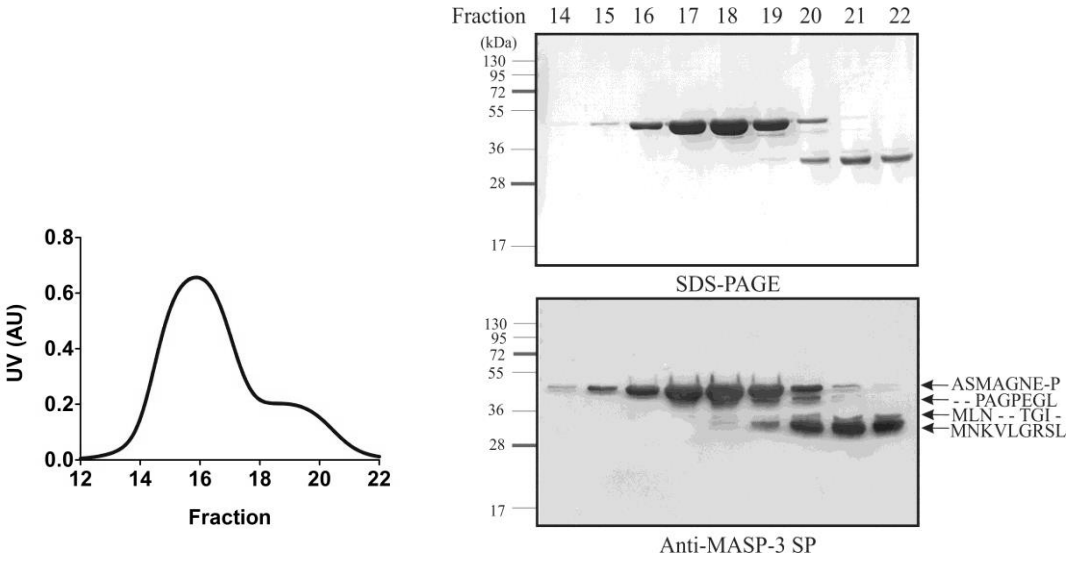
3.3 RESULTS

3.3.1 Production and characterization of recombinant MASP-3 proteins

The catalytic fragments of wtM3, M3EEKQ, M3Q and M3R454A were expressed and purified by the methods described in Section 3.2.4. SDS-PAGE analysis and western blotting were applied to the fractions of the two-step purification process to identify the fractions containing the proteins expected (Figures 3.3-3.6). SDS-PAGE analysis of the Q-Sepharose-separated fractions of all four recombinant MASP-3 proteins revealed that the newly refolded proteins contained four main components that were recognized by the anti-MASP-3 SP domain antibody in western blotting. A 48 kDa predominant protein in the Q-Sepharose-separated fractions of wtM3 was found to contain the N-terminal sequence of ASMAGNE-P (in which the blank residue was a Cys, which cannot be detected in N-terminal sequencing), indicating that it is the CCP1-CCP2-SP portion of MASP-3. The other three protein components were smaller, including a minor quality of a 44 kDa protein with the N-terminal sequence of - - PAGPEGL, which does not match the sequence of the wtM3 and is therefore most likely a contaminant, a fragment of 34 kDa and another of 32 kDa, respectively starting with M^{405} and M^{422} residues of the CCP2 domain, suggesting they are result from cleavage of wtM3 at the peptide bonds K^{404}-M^{405} and W^{421}-M^{422}. The 34 kDa and 32 kDa components most likely result from cleavage by intrinsic enzymes in *E. coli* cells. By using gel filtration, the majority of the 48 kDa wtM3 in the anion exchange chromatography fractions 10-14 was separated from the three smaller protein components (Figure 3.3). The gel filtration fractions 14-17 contained wtM3 at high homogeneity and were pooled as the wtM3 stock. The above methods of expression and purification were also applied to successfully produce M3R454A, M3EEKQ and M3Q proteins, with generally 1-3 mg of each recombinant protein obtained per litre of *E. coli*

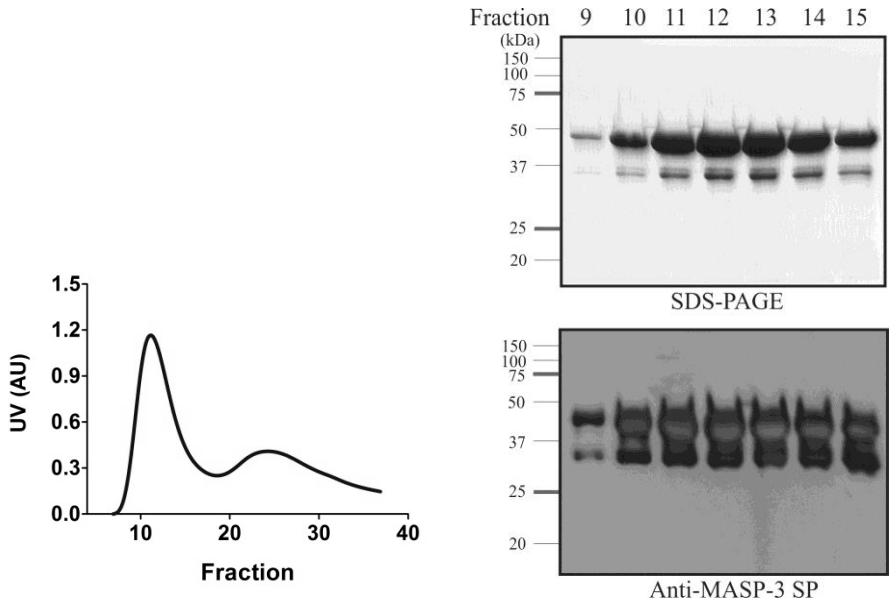


(A)

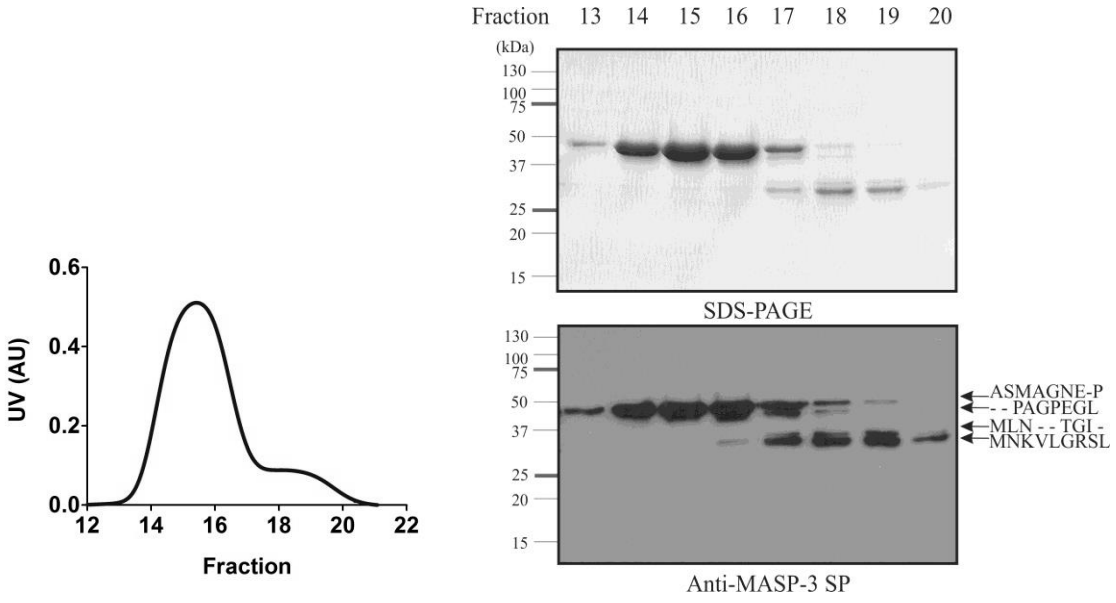


(B)

Figure 3.3 Purification of the CCP1-CCP2-SP fragment of wild type MASP-3 (wtM3)
Panel (A), anion exchange chromatography of wtM3. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. Fractions 10-14 were pooled for further purification by size-exclusion chromatography.
Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 10-14. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. The N-terminal sequences of the 48, 44, 34 and 32 kDa fragments are displayed with arrows indicating the bands corresponding to the sequences. This indicates that the 48 kDa protein was wtM3. Fractions 14-17 therefore contained wtM3 at high purity.



(A)



(B)

Figure 3.4 Purification of the M3R454A CCP1-CCP2-SP fragment

Panel (A), anion exchange chromatography of M3R454A. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. Fractions 9-13 were pooled for further purification by size-exclusion chromatography.

Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 9-13. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. The N-terminal sequences of the 48, 44, 34 and 32 kDa fragments are displayed with arrows indicating the bands corresponding to the sequences. This indicates that the 48 kDa protein was M3Q454A. Fractions 13-15 therefore contained M3R454A at high purity.

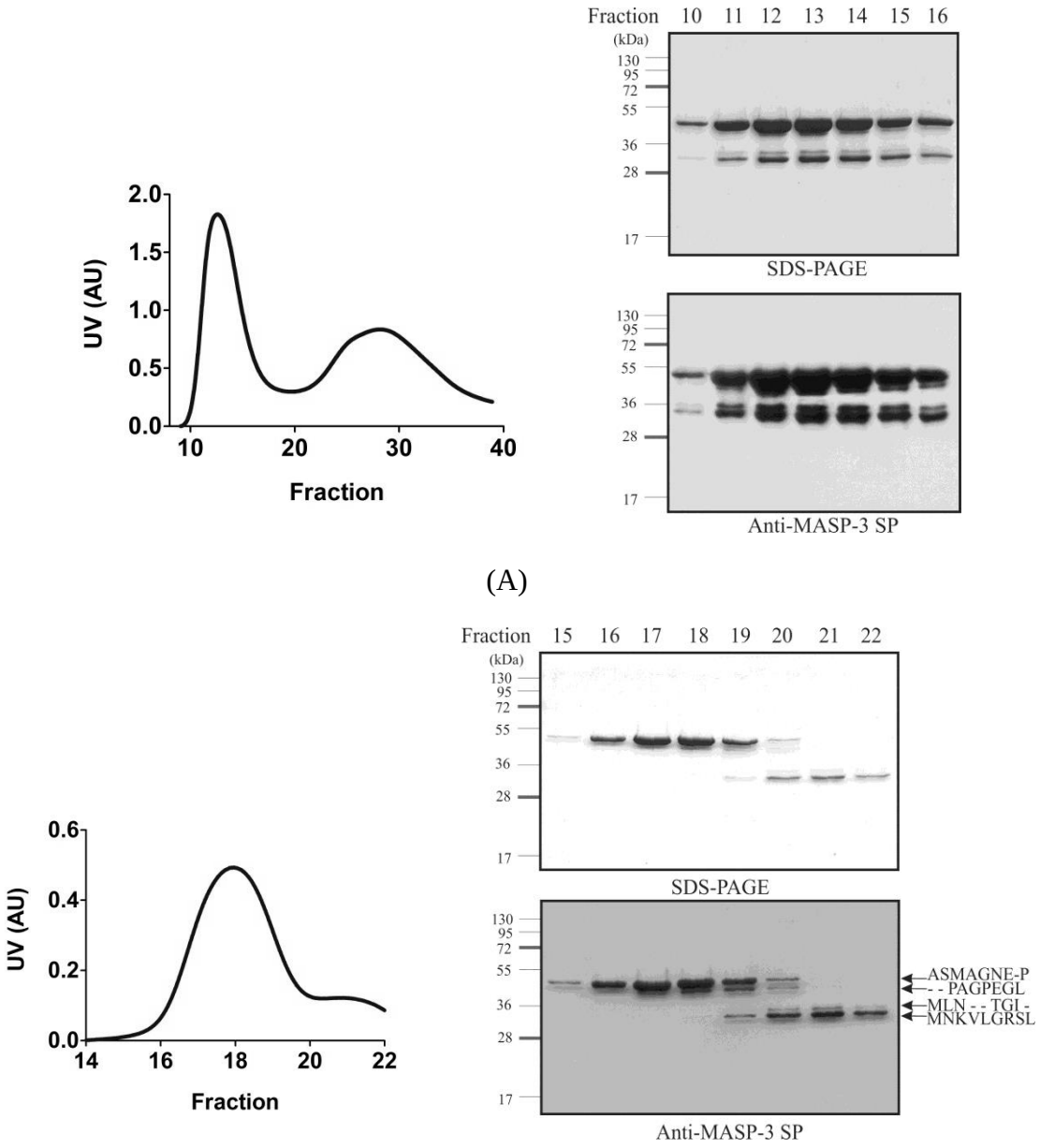


Figure 3.5 Purification of the M3EEKQ CCP1-CCP2-SP fragment

Panel (A), anion exchange chromatography of M3EEKQ. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 10 µl sample from the indicated fraction. Fractions 10-14 were pooled for further purification by size-exclusion chromatography.

Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 10-14. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 10 µl sample from the indicated fraction. The N-terminal sequences of the 48, 44, 34 and 32 kDa fragments are displayed with arrows indicating the bands corresponding to the sequences. This indicates that the 48 kDa protein was M3EEKQ. Fractions 15-17 therefore contained M3EEKQ at high purity.

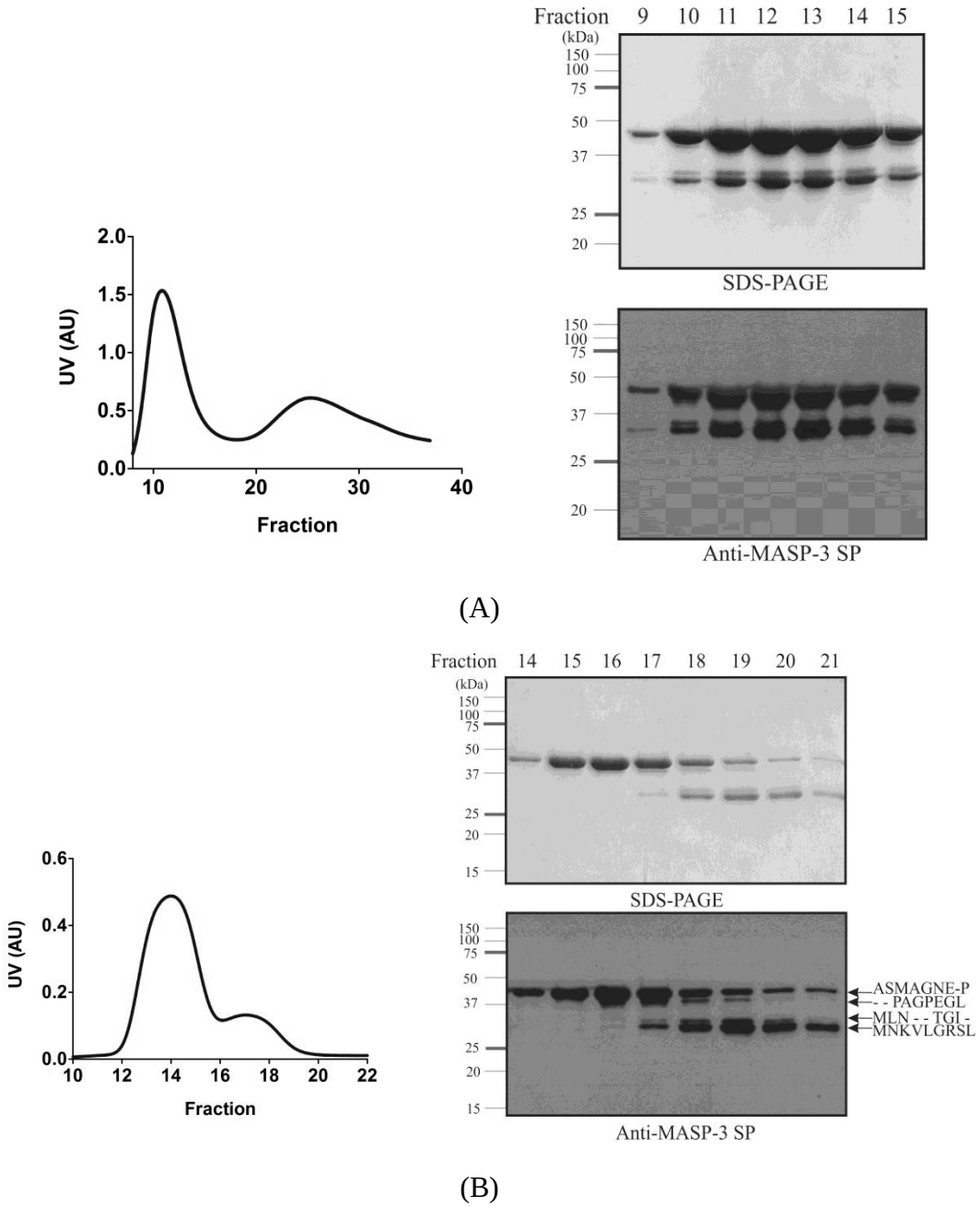


Figure 3.6 Purification of the M3Q CCP1-CCP2-SP fragment

Panel (A), anion exchange chromatography of M3Q. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. Fractions 9-13 were pooled for further purification by size-exclusion chromatography.

Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 9-13. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 20 µl sample from the indicated fraction. The N-terminal sequences of the 48, 44, 34 and 32 kDa fragments are displayed with arrows indicating the bands corresponding to the sequences. This indicates that the 48 kDa protein was M3Q. Fractions 14-16 therefore contained M3Q at high purity.

culture. All four recombinant MASP-3 proteins of 48 kDa were single polypeptide chains representing zymogenic forms of the enzymes.

3.3.2 Cleavage of wtM3 and M3R454A by MASP-1 and MASP-2

In order to investigate cleavage of MASP-3 by proteolytic enzymes of the complement and coagulation systems, wtM3 was incubated with varying concentrations of human recombinant enzymes, including the catalytic portions of C1r, C1s, MASP-1, MASP-2, thrombin and kallikrein, respectively, at 37°C for 2 hr in gel filtration buffer containing 50 mM Tris, 145 mM NaCl, pH 7.4. The reaction mixtures were then reduced, denatured and analysed by SDS-PAGE with Coomassie staining (Figure 3.7). The results showed that wtM3 was not cleaved by the classical pathway complements C1r and C1s, or the coagulation system proteases, thrombin and kallikrein. Recombinant MASP-1 appeared to cleave wtM3, indicated by the decrease of intensity of the wtM3 protein bands when increasing concentrations of MASP-1 were added. The reduced products of MASP-1-cleaved wtM3 appeared to overlap with the MASP-1 SP and MASP-1 CCP1-CCP2 bands in SDS gel due to their respectively similar molecular weights. MASP-2 at 0.25 μM cleaved all the wtM3 (0.35 μM) within the incubation time, resulting in two fragments of 32 kDa and 17kDa, of which, the 17 kDa band overlapped with MASP-2 CCP1-CCP2.

In order to further investigate cleavage of wtM3 by MASP-1 and MASP-2, 1 μM wtM3 was incubated with either 1 μM MASP-1 or 1 μM MASP-2 at 37°C in gel filtration buffer for 2 hr. The cleavage products were reduced, denatured and separated by SDS-PAGE, and then N-terminally sequenced (Figure 3.8). The results showed that the 32 kDa fragment resulting from wtM3 cleaved by MASP-1 was a mixture of a cleaved wtM3 with the N-terminal sequence of NAEPGLFP and a fragment with the N-terminal sequence of IFNGRPAQ, which represents MASP-1 SP. The 17 kDa fragment starting with the N-terminal sequence of ASMAGNE-P therefore represents a mixture of the

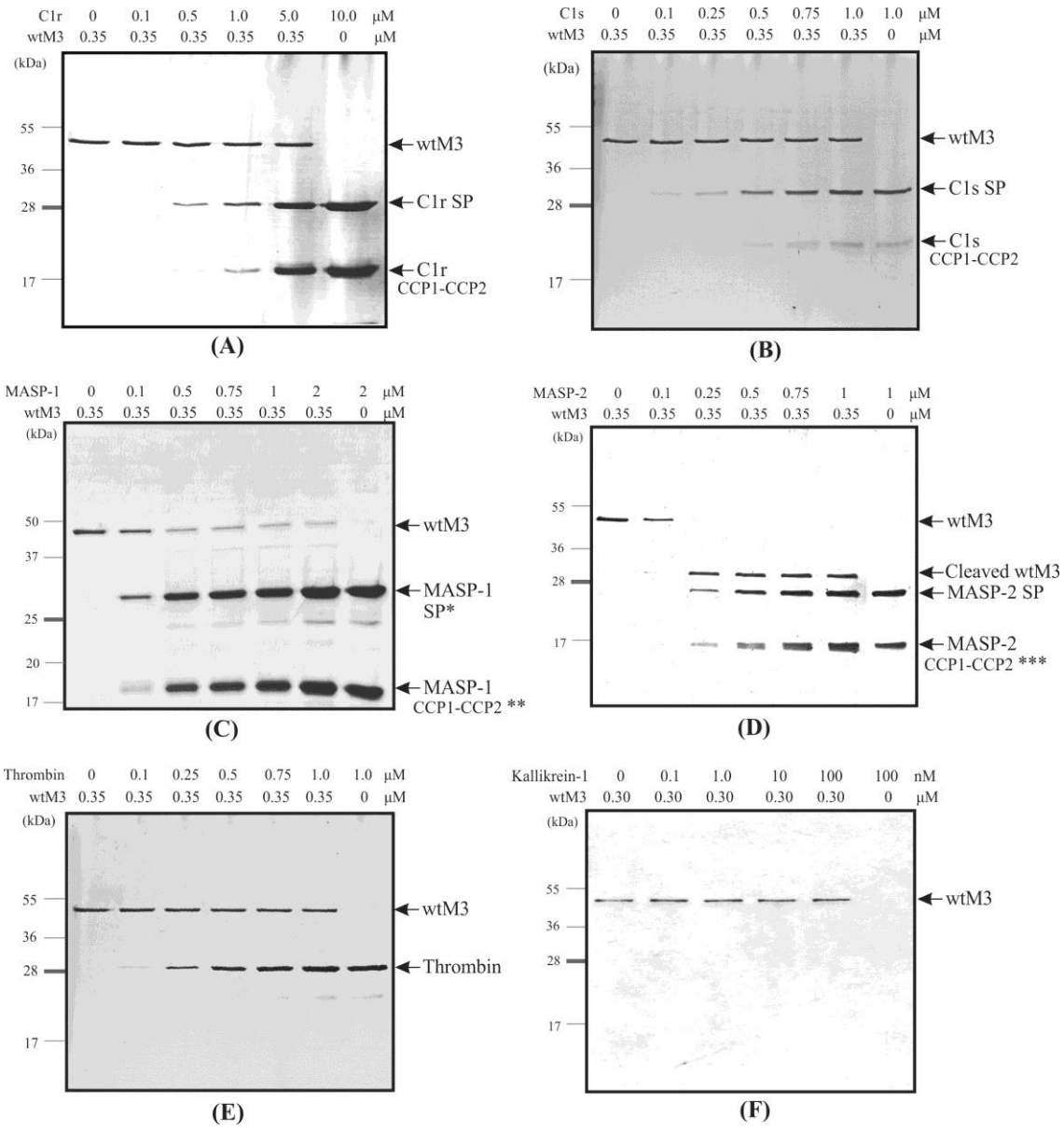


Figure 3.7 SDS-PAGE analyses of cleavage of wtM3 by proteolytic enzymes of the complement and coagulation systems

wtM3 was incubated with the indicated final concentrations of recombinant human enzymes, including the catalytic portion of C1r (A), C1s (B), MASP-1 (C), MASP-2 (D), thrombin (E) and kallikrein (F), respectively, at 37°C for 2 hr. The reaction mixtures were then reduced, denatured, and analysed by SDS-PAGE, followed by Coomassie staining. The identities of the protein bands are indicated using arrows. Notes: the MASP-1 SP* and the MASP-1 CCP1-CCP2** bands in the mixture of MASP-1 and wtM3 may also have contained cleaved wtM3 fragments, as the molecular weights of MASP-1 SP and MASP-1 CCP1-CCP2 are respectively similar to those of wtM3. A 24 kDa protein appeared to be a minor contaminant in the MASP-1 preparation. The MASP-2 CCP1-CCP2*** bands in the mixture of MASP-2 and wtM3 may also have contained cleaved wtM3 CCP1-CCP2 fragment due to its similar molecular weight to that of MASP-2 CCP1-CCP2.

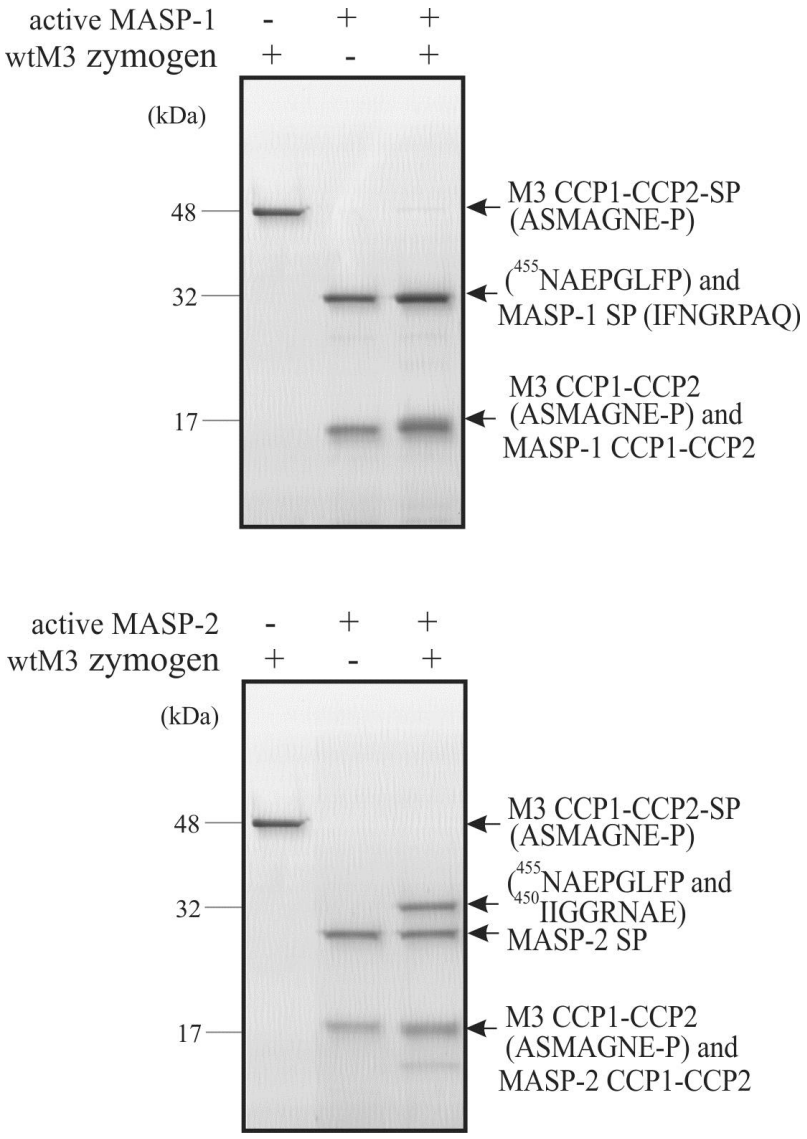


Figure 3.8 The cleavage products of wtM3 after treatment with either MASP-1 or MASP-2

Top panel, 1 μM wtM3 was incubated either with or without 1 μM MASP-1 at 37°C for 2 hr in a final volume of 20 μl. The mixture was then reduced and analysed by SDS-PAGE. Bottom panel, 1 μM wtM3 was incubated either with or without 1 μM MASP-2 at 37°C for 2 hr in a final volume of 20 μl. The mixture was then reduced and analysed by SDS-PAGE with Coomassie staining. The protein fragments in the cleavage products are indicated with arrows and the N-terminal sequences of these fragments are shown in brackets.

CCP1-CCP2 fragments of both MASP-1 and MASP-3. The heavier fragment indicated that the predominant cleavage of wtM3 by MASP-1 is at the peptide bond of R^{454}-N^{455}, four amino acids C-terminal to the R^{449}-I^{450} activation bond. Compared to this, MASP-2 also cleaves wtM3, resulting in two protein bands similar in molecular weights to those shown in the MASP-1-cleaved MASP-3 sample. However, the 32 kDa fragment in the MASP-2-cleaved wtM3 sample was found to be a mixture of two peptides with the N-terminal sequences of NAEPGLFP and IIGGRNAE, respectively. The presence of the two peptides indicates that MASP-2 cleaves wtM3 at two peptide bonds, R^{454}-N^{455}, similar to that cleaved by MASP-1, and the R^{449}-I^{450} activation bond.

The selectivity for the two closely located possible cleavage sites of MASP-3 by either MASP-1 or MASP-2 was further illustrated by testing cleavage of M3R454A by the two MASP enzymes (Figure 3.9). Both MASP-1 and MASP-2 cleave M3R454A only at the R^{449}-I^{450} activation bond, confirmed by N-terminal sequencing of the cleavage products.

3.3.3 Activation of M3EEKQ/M3Q by C1r protease

In order to demonstrate that recombinant C1r protease could activate the MASP-3 mutants, 1 μM recombinant C1r protease was mixed with wtM3, M3EEKQ or M3Q at 0.65 μM at 37°C for either 1 or 5 hr (Figure 3.10). The majority of either M3EEKQ or M3Q was processed by C1r protease within 1 hr, resulting in two fragments of 32 kDa and 17kDa in SDS gel, whilst wtM3 was not cleaved by C1r protease for up to 5 hr of incubation. The 32 kDa fragment was recognized by the MASP-3 SP-specific antibody and had the N-terminal sequence of IIGGRNAE, thereby representing the SP domain of

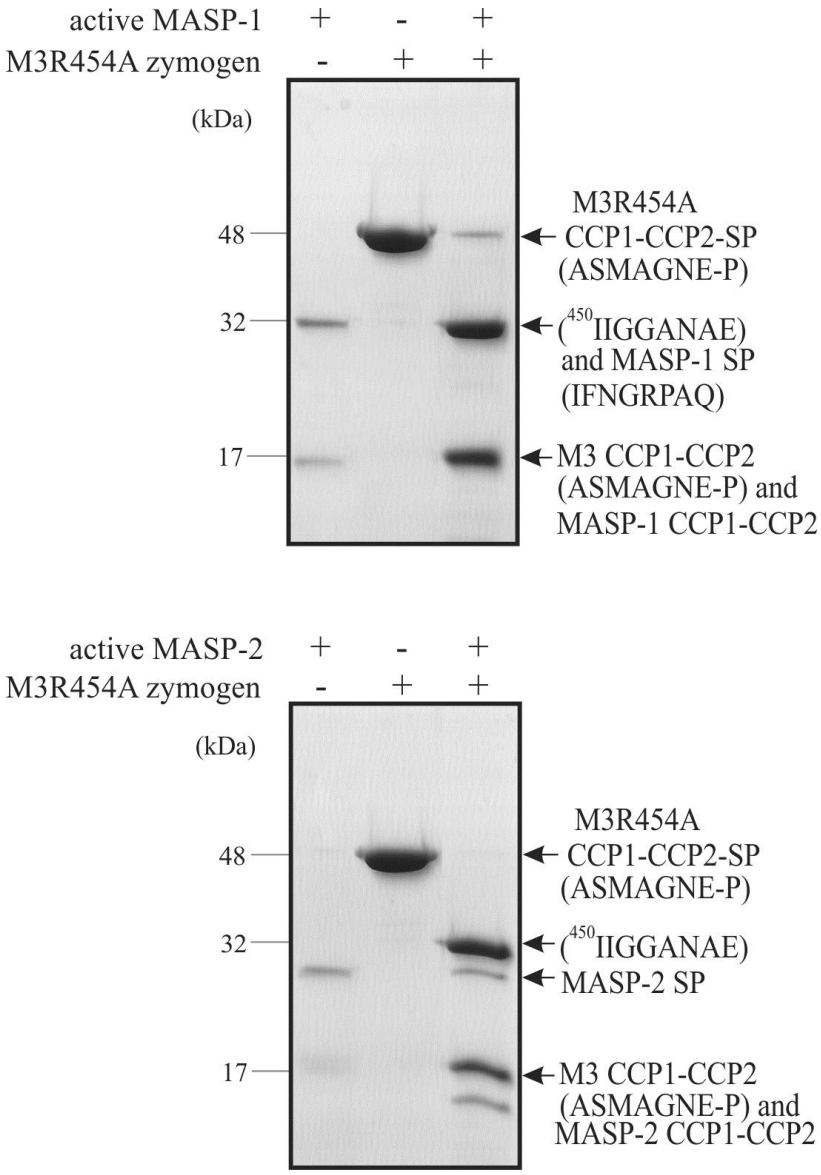


Figure 3.9 The cleavage products of M3R454A after treatment with either MASP-1 or MASP-2

Top panel, 10 μM wtM3 was incubated either with or without 1μM MASP-1 at 37°C for 2 hr in a final volume of 20 μl. The mixture was then reduced and analysed by SDS-PAGE. Bottom panel, 10 μM wtM3 was incubated either with or without 1 μM MASP-2 at 37°C for 2 hr in a final volume of 20 μl. The mixture was then reduced and analysed by SDS-PAGE with Coomassie staining. The protein fragments in the cleavage products are indicated with arrows and the N-terminal sequences of these fragments are shown in brackets.

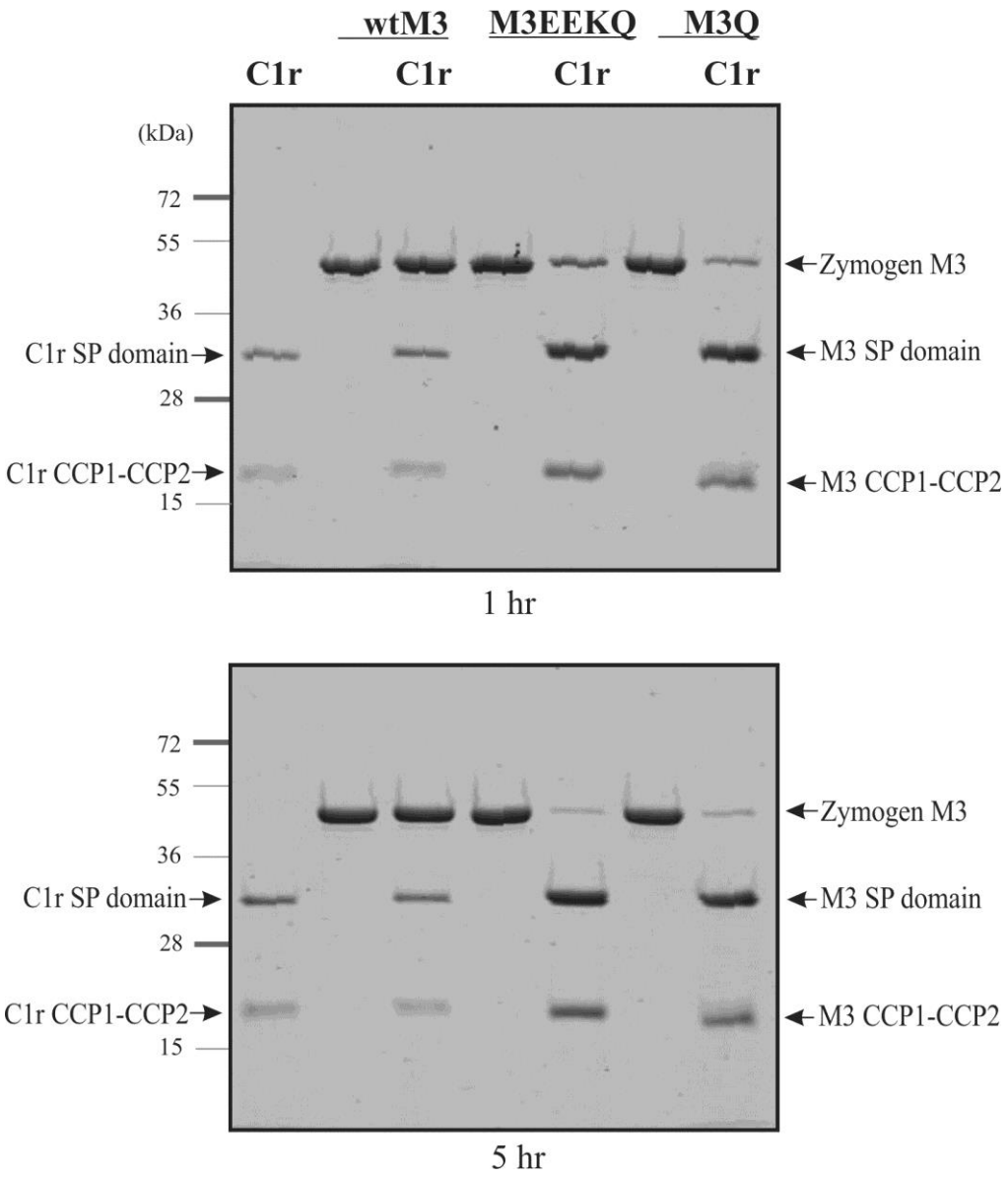


Figure 3.10 C**leavage of M3EEKQ and M3Q by recombinant C1r protease**
Zymogens of wtM3, M3EEKQ and M3Q were incubated either with or without recombinant C1r protease at 37°C for 1 hr (top panel) and 5 hr (bottom panel). The samples were reduced and analysed by SDS-PAGE with Coomassie staining.

activated MASP-3. The 17 kDa fragment with the N-terminal sequence of ASMAGNE-P thus is the CCP1-CCP2 fragment of the MASP-3 mutants. By using an NHS-activated column that was coupled with recombinant C1r protease, M3EEKQ and M3Q were effectively activated (Figure 3.11).

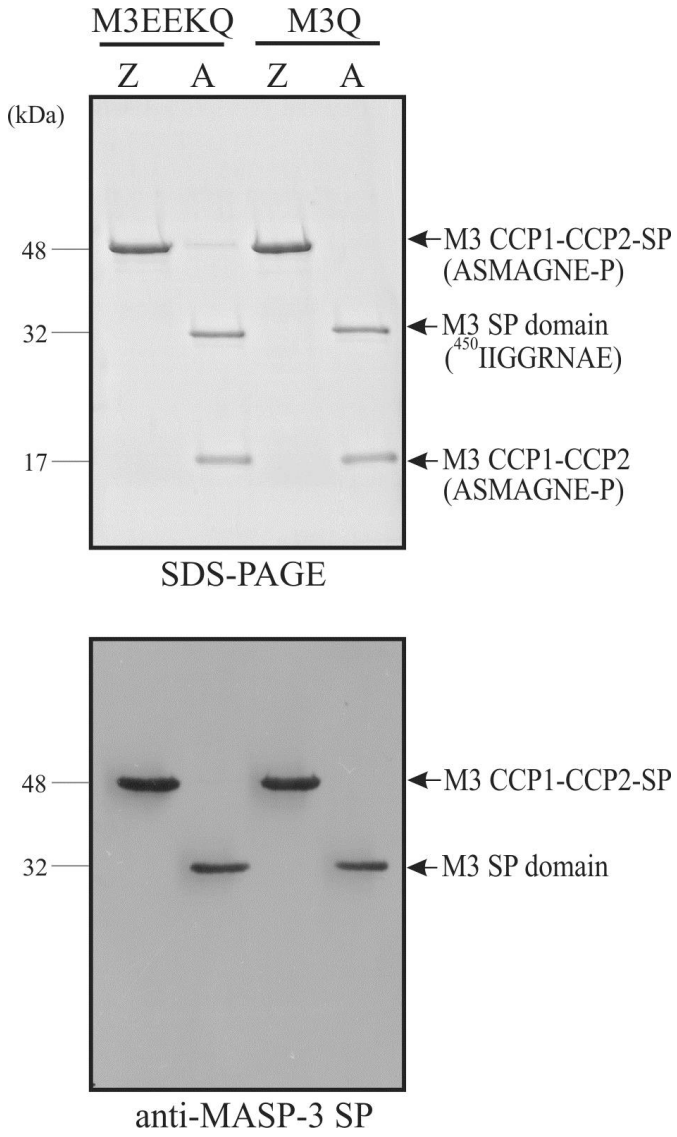


Figure 3.11 Activation of MASP-3 mutants by using a C1r column
The zymogen (Z) of either M3EEKQ or M3Q was loaded into a C1r-column and the active protein (A) was eluted after incubating the column at 26°C for 18 hr. The samples were reduced and analysed using SDS-PAGE (top panel) and by western blotting (bottom panel) using a monoclonal antibody recognizing the unique peptide sequence NPNVTDQIISSGTRT of MASP-3 SP domain. The represented protein fragments are indicated with arrows and the N-terminal sequences of these fragments are shown in brackets.

3.3.4 Activation of wtM3 by PC enzymes

The predicted probabilities of cleavage of wtM3 by members of the PC enzyme family suggested that there are two sites in wtM3 that are highly likely to be cleaved by at least one member of the PC enzyme family (Table 3.4). Interestingly, the top hit was at the activation bond R^{449}-I^{450}, with the highest probability score of 0.893. Another possible cleavage site was within the CCP-2 domain and located 22 amino acids N-terminal to the activation bond, with a score of 0.886.

Table 3.4 The predicted cleavage sites of wtM3 by at least one member of the proprotein convertase family of enzymes

Cleavage site		Score	Cleavage site		Score
LPSLVKR	II	0.893	LHDVRDK	SG	0.073
MNKVLGR	SL	0.886	SRVPNDK	WF	0.069
VLRSQRR	DT	0.377	KRIIGGR	NA	0.068
IISSGTR	TL	0.201	TVIPVSK	EH	0.067
VNSSAAR	VV	0.161	DVLQYVK	LP	0.067
FDDLSQR	WV	0.157	TGYKVLK	DN	0.066
TAAHVLR	SQ	0.152	NNLTTYK	SE	0.066
VVEDTSR	VP	0.121	FQIECLK	DG	0.063
VYGVYTK	VS	0.118	HVLRSQR	RD	0.062
TYKSEIK	YS	0.098	CQEPYYK	ML	0.061
SLPSLVK	RI	0.097	LGLHDVR	DK	0.060
IEPSQAK	YF	0.092	GYYEGGK	DT	0.056
QPPVHGK	IE	0.091	PEECGSK	QV	0.055
DGTWSNK	IP	0.088	KTSYESR	SG	0.053
LITFSTR	NN	0.087	CKIVDCR	AP	0.050
QGVWMNK	VL	0.083	MPVCLPR	LE	0.050
SCDTGYK	VL	0.081	NKIPTCK	IV	0.049
QAKYFFK	DQ	0.080	ECGQPSR	SL	0.046
VPHAECK	TS	0.079	VEPQVER	--	0.039

Cleavage of 2 μg wtM3 by 2 units of either furin (Figure 3.12A) or PACE4 (Figure 3.13A) was examined at 37°C for up to 24 hr. Incubating wtM3 alone in gel filtration buffer or the furin activation buffer or DMEM/F-12 buffer at 37°C for 24 hr did not yield any cleavage products of wtM3, whilst incubating wtM3 with either furin or PACE4 under the described experimental setting resulted in the gradual presence of two fragments of 32 kDa and 17 kDa. This indicated that both furin and PACE4 process wtM3 slowly. The 32 kDa fragment resulting from cleavage of wtM3 by either furin or PACE4 was recognized by the anti-MASP-3 SP antibody used in this project. N-terminal sequencing of this protein fragment revealed its N-terminal sequence to be IIGGRNAE, proving that both furin and PACE4 cleave wtM3 at the R^{449}-I^{450} activation bond. Cleavage of M3EEKQ and M3Q by either furin or PACE4 was also examined at the same conditions as those for cleavage of wtM3. It was found that neither furin (Figure 3.12B, C) nor PACE4 (Figure 3.13B, C) were able to cleave M3EEKQ or M3Q.

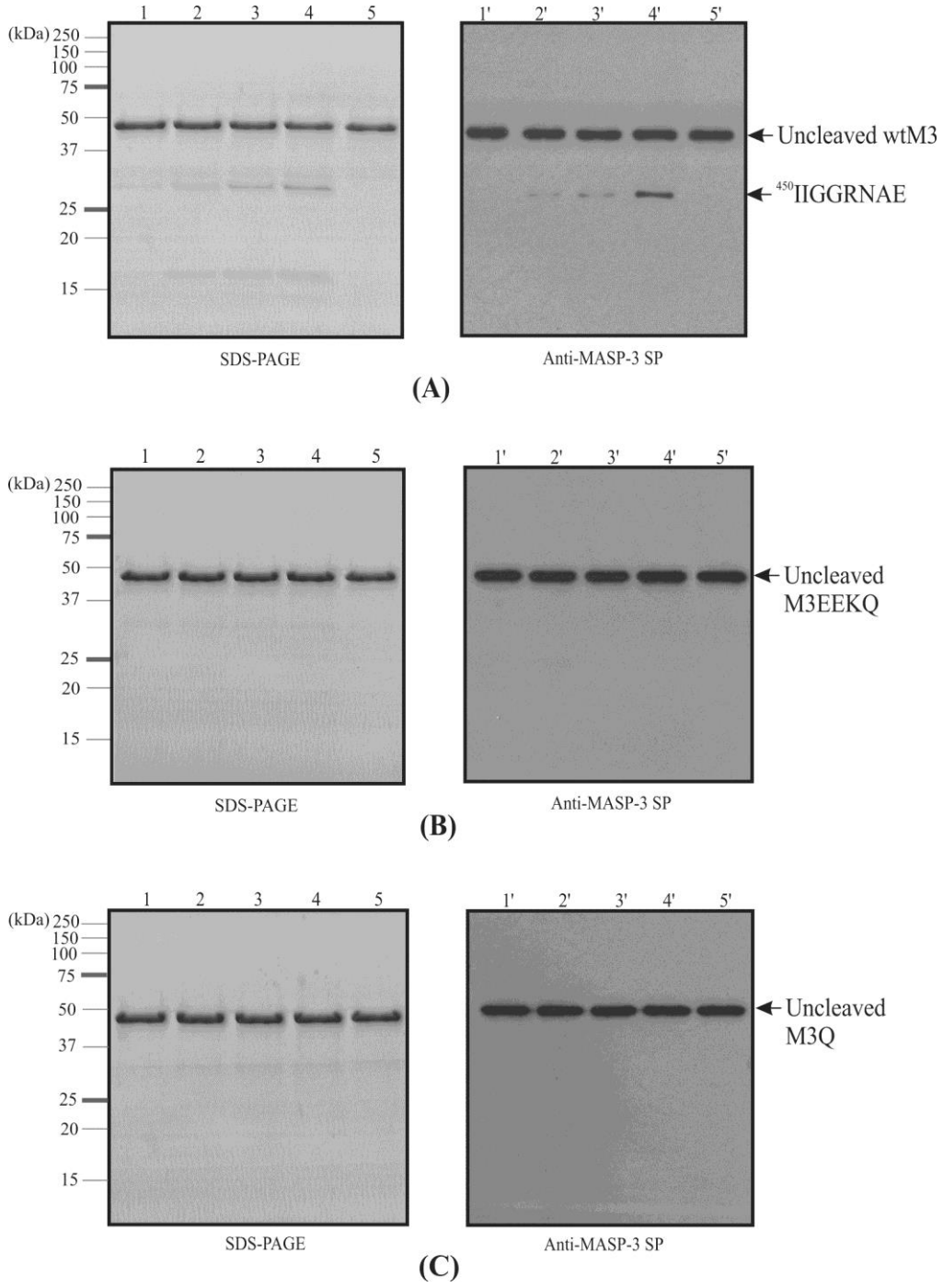


Figure 3.12 Cleavage of wtM3 by furin

2 μg wtM3 (A), M3EEKQ (B) or M3Q (C) were incubated with or without 2 units of furin at 37°C, for 1-24 hr. The samples were then reduced, denatured, followed by SDS-PAGE (left panels) and western blotting (right panels). The contents or N-terminal sequence of the indicated protein bands are shown with arrows.

Lanes 1/1'- 4/4' contained a mixture of 2 units of furin and 2 μg wtM3 (A), M3EEKQ (B), or M3Q (C), respectively, incubated for 1, 4, 8, and 24 hr. Lanes 5/5' contained 2 μg wtM3 (A), M3EEKQ (B) or M3Q (C), respectively, in the furin activation buffer incubated for 24 hr.

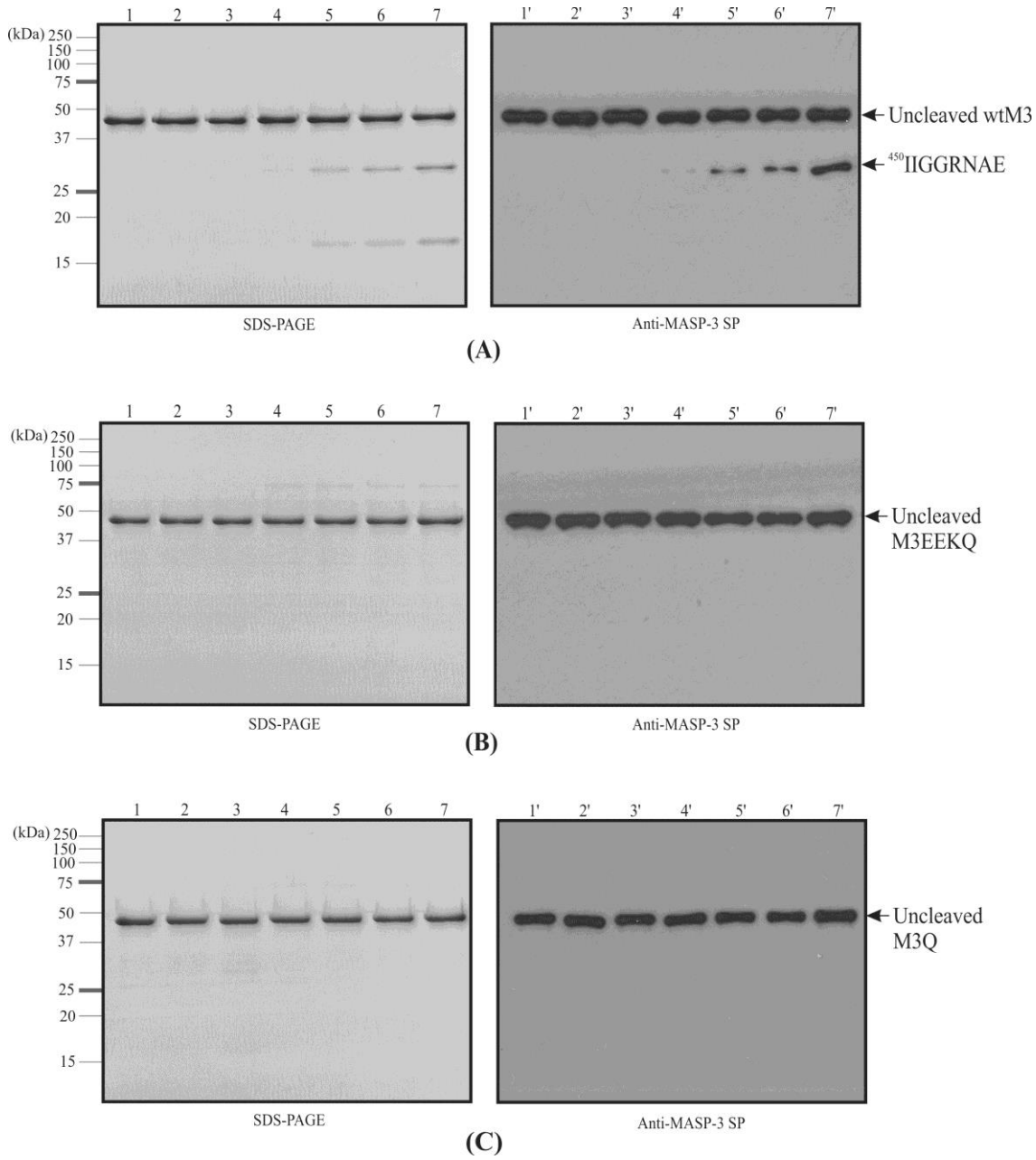


Figure 3.13 Cleavage of wtM3 by PACE4

2 μg wtM3 (A), M3EEKQ (B) or M3Q (C) were incubated with or without 2 units of PACE4 at 37°C, for 1-24 hr. The samples were then reduced, denatured, followed by SDS-PAGE (left panels) and western blotting (right panels). The contents or N-terminal sequence of the indicated protein bands are shown with arrows.

Lanes 1/1'- 3/3' contained 2 μg wtM3 (A), M3EEKQ (B) or M3Q (C), respectively. The proteins in lane 1/1' were not incubated, whereas proteins in lanes 2/2' were in the DMEM/F-12 buffer and proteins in lane 3/3' were in the gel filtration buffer, both incubated for 24 hr. Lanes 4/4'- 7/7' contained a mixture of 2 units of PACE4 and 2 μg wtM3 (A), M3EEKQ (B) or M3Q (C), respectively, incubated for 1, 4, 8, and 24 hr.

3.4 DISCUSSION

The prokaryotic BL21 *E. coli* cell line has been used to express the catalytic portions or smaller fragments of human complement proteins, including C1r protease (Wijeyewickrema *et al.*, 2013), C1s protease (Duncan *et al.*, 2012), MASP-1 and MASP-2 enzymes (Ambrus *et al.*, 2003), for studying the structure and function of these proteins. In this project, the BL21STAR(DE3) *E. coli* cell strain was used to express the catalytic portions of wild type human MASP-3 and MASP-3 protein mutants. Full length human wild type MASP-3 has been expressed in the form of a zymogen in the Chinese hamster ovary cell line, CHO-DG44 (Skjodt *et al*., 2010), a *Trichoplusia ni* insect cell line (Zundel *et al*., 2004) and the human embryonic kidney cell line, HEK293EBNA (Dahl *et al*., 2001). The full length wild type MASP-3 protein expressed by CHO-DG44 was found to be a protein mixture of varied glycosylated forms. Compared to the full length MASP-3 expressed in eukaryotic cells, MASP proteins expressed in *E. coli* cells were not glycosylated, and thus can avoid the possible variation in post-translational glycosylation, resulting in products of high homogeneity. It is possible that the function of unglycosylated MASP-3 is different to that of the wild type protein. Nevertheless, the abundant MASP-3 protein produced in *E. coli* cells provided a solid foundation for studying the enzyme in this project, and the functional alteration of MASP-3 by glycosylation may be addressed by further studies of eukaryotic cell- or plasma-derived MASP-3 in the future.

Under the described experimental conditions, wtM3 was not cleaved by C1r protease, C1s protease, thrombin and kallikrein, but was cleaved by the MASP-1 and MASP-2 enzymes. Previously, it was thought that MASP-1 activates MASP-3, based on the observations that human MASP-1 processes recombinant mouse and human MASP-3. The resulting reduced cleavage products include fragments that were recognized by antibodies specific for the SP domain of MASP-3 (Iwaki *et al*., 2011, Degn *et al*., 2012, Megyeri *et al*., 2013). However, our present data revealed that MASP-1 preferably cleaves MASP-3 zymogen at the peptide bond of R^{454}-N^{455}, four residues C-terminal to the activation bond. In contrast, our data showed that MASP-2 can cleave wtM3 at the

R^{449}-I^{450} activation bond, although the enzyme can also cleave at the R^{454}-N^{455} peptide bond. Our identification of MASP-2, not MASP-1, as an activator of MASP-3 is in agreement with previous findings that the MASP-1 and MASP-3 enzymes are found to be in complex with different oligomers of MBL in human plasma, and a heterodimer formed between MASP-1 and MASP-3 monomers via their respective, but identical CUB1-EGF-CUB2 segments, has never been observed. Instead, MASP-2 and MASP-3 in human plasma were co-localized in the same MBL-II oligomer, giving rise to the opportunity for interaction between these two enzymes (Dahl *et al*., 2001).

Further, our study revealed that replacing the R^{454} residue of MASP-3 with an A residue resulted in activation by MASP-1 and MASP-2, highlighting the importance of the sequence of the activation loop region of MASP-3 for interaction of the enzyme with its possible activation partner. The activation of serine proteases, including MASP-3, is often achieved by cleavage at the activation bond by another proteolytic enzyme with P1-R/K specificity. The presence of the R^{454}-N^{455} peptide bond in close proximity to the R^{449}-I^{450} activation bond therefore gives rise to a few possible events for cleavage of MASP-3 at this region by serine proteases like MASP-1 and MASP-2, highly effective in cleaving after R residues. Providing the primary cleavage site of MASP-3 is at the R^{454}-N^{455} peptide bond, as cleaved by MASP-1, MASP-3 is not activated since a newly liberated N-terminal I residue is not generated. However, if the two possible cleavage sites are equally preferable for cleavage by the protease partner, as shown in the case of cleavage by MASP-2, the proteolytic activity of the cleavage products of MASP-3 will need to be further studied. It is also possible that cleavage at the activation site and the R^{454}-N^{455} peptide bond are sequential events. In the last case, the proteolytic activity of the dually cleaved MASP-3 will need to be carefully studied. Previous studies of the activation of trypsinogen revealed that a small dipeptide (Nt)-IV, which resembles the sequence C-terminal to the K-I activation bond of trypsinogen, is able to induce

activation of trypsinogen (Menegatti *et al*., 1985), indicating that activation of this classic serine protease can be induced by insertion of the dipeptide into its active state position in the zymogen. It is therefore possible that when a subsequent cleavage at the R^{454}-N^{455} peptide bond occurs following activation of MASP-3, the dually cleaved MASP-3 containing the short fragment I^{450}IGGR between the two cleavage sites could still be active. In order to further study cleavage of the two closely located cleavage sites of MASP-3 by MASP-1 and MASP-2, a finer time-course analysis of the cleavage products is needed.

Although our data showed that MASP-2 is able to activate MASP-3, the cleavage products are a mixture of two fragments. By using the M3R454A mutant, both MASP-1 and MASP-2 enzymes can activate the MASP-3 mutant correctly. However, the activated M3R454A contains a serine protease with the mutation residue, and thus is different to that of wild type MASP-3. Therefore, an alternative approach to generate a wtM3 mutant with a homogeneous functional MASP-3 serine protease domain resembling that of the wild type MASP-3 is necessary. Here we have developed two MASP-3 mutant proteins, M3EEKQ, which contains the P5-P2 residues identical to those of the activation site of C1s, and M3Q, which contains the P2-Q residue at the activation site presented in both C1r and C1s. Both M3EEKQ and M3Q were activated by recombinant C1r. The active forms of MASP-3 mutants consist of a serine protease domain identical to that of wild type MASP-3, confirmed by western blotting and N-terminal sequencing.

The activation of both M3EEKQ and M3Q by recombinant C1r CCP1-CCP2-SP was equally effective. According to the kinetic data measured by Wijeyewickrema *et al*. (2013), the K_m and k_{cat} values for cleavage of M3EEKQ by recombinant C1r protease (64.0 μM and 1.75×10^{-2} s^{-1}) were extremely similar to those of M3Q (62.2 μM and

2.1 x 10^{-2} s^{-1}). Whist the K_m value for cleavage of C1s by C1r (22 μM) was found to be a third of that of either M3EEKQ or M3Q cleaved by C1r, the k_{cat} value was determined to be 6.45 x 10^{-2} s^{-1}, about three times greater than cleavage of MASP-3 mutants. Consequently, the k_{cat} /K_m value of the reaction of C1r activating C1s was determined to be 2.9 x 10^3 $M^{-1}s^{-1}$, which was similar to those of activating M3EEKQ (2.7 x 10^3 $M^{-1}s^{-1}$) and M3Q (3.4 x 10^3 $M^{-1}s^{-1}$) by C1r. The data indicates that the efficacy of activation of either MASP-3 mutant by recombinant C1r was comparable to that of zymogenic C1s CCP1-CCP2-SP. By applying a C1r-coupled NHS-activated column, M3EEKQ and M3Q proteins were successfully activated after incubation at 26°C for 18 hr. These findings not only established a MASP-3 module, which can be activated in a convincing scheme, but also contribute to the identification of the molecular determinants for interaction with substrates by the C1r protease.

One characteristic feature in the sequence of the activation loop of MASP-3 is the presence of a pair of basic residues (-KR-) N-terminal to the activation bond. Proproteins with this characteristic are often processed by members of the PCs, a family of enzymes that specifically cleave after dibasic residues (Seidah and Prat, 2012, review). Our results of the time-dependent activation of wtM3 by the PC enzymes, furin and PACE4, are the first showing specific activation of wtM3 by human enzymes. The results agreed with the prediction that the PC enzymes cleave wtM3 at the R^{449}-I^{450} activation bond with the highest probability (a score of 0.893). Although the predicted favourite cleavage sites of wtM3 by PC enzymes also include another site with similar probability (a score of 0.886), the cleavage site of wtM3 by furin and PACE4 was proved to be only at the R^{449}-I^{450} activation bond, emphasizing that of the two predicted sites, the activation site of wtM3 was the only site targeted by furin and PACE4. Furthermore, both furin and PACE4 failed to cleave M3EEKQ and M3Q, the two wtM3

mutants that contained activation sites without the dibasic residues. These results indicated that the characteristic -KR- motif preceding the activation bond of wild type MASP-3 was the driving force for its cleavage by furin and PACE4, in agreement with the substrate cleavage preference of the two PC enzymes (Seidah and Chretien, 1999, Turpeinen *et al.*, 2011).

Both furin and PACE4, and PC5/6, which was not featured in this project due to unavailability, are widely expressed in human tissues. Furin mainly processes precursors of constitutively secreted proteins, including the TGFβ-like growth factors Lefty and BMP10, which are vitally implicated in animal embryonic development. Similar to furin, PACE4 and PC5/6 also activate growth factors. The functions of these three PC enzymes are found to be strongly redundant among each other, both *in vitro* and *ex vivo* (Seidah and Prat, 2012, review). Therefore, MASP-3 may also be activated by PC5/6, as furin and PACE4 do. The biological implications of the activation of MASP-3 by the PC enzymes are not yet known. In addition, the activation of wtM3 by furin and PACE4 under the experimental conditions appeared to be slow, reflecting that the conditions for the activation might not be optimal, or such activation was affected by the lack of glycosylation of the MASP-3 protein in this project.

In conclusion, MASP-3 can be activated by MASP-2, not MASP-1. A slow activation of the catalytic portion of wild type MASP-3 by the PC enzymes, furin and PACE4, was also observed. Here we have developed two MASP-3 mutants, M3EEKQ and M3Q. Both mutants were efficiently activated by C1r protease, resulting in the active form of proteins with a serine protease domain resembling that of the wide type MASP-3. Therefore the specificity of C1r-activated M3EEKQ and M3Q can be predicted to be the same as that of the wild type MASP-3.

Chapter Four

The Substrate Specificity of MASP-3

4.1 INTRODUCTION

Currently, the substrate specificity of MASP-3 has not been fully profiled. By using the wild type MASP-3 activated by an unknown intrinsic enzyme in the High Five insect cells, Zundel *et al*. (2004) determined that MASP-3 weakly cleaves a synthetic thiobenzyl substrate, Z-Gly-Arg-S-Bzl. Later, Cortesio and Jiang (2006) found that a truncated MASP-3 SP domain cleaved a few synthetic substrates, including Pro-Phe-Arg-AMC, Boc-Val-Pro-Arg-AMC, Z-Gly-Gly-Arg-AMC, Z-Phe-Arg-AMC and Z-Leu-Arg-AMC, among which Pro-Phe-Arg-AMC displayed the highest rate of cleavage. The authors also found that the truncated protein cleaved insulin-like growth factor-binding protein (IGFBP)-5 at three sites containing an R/K residue at P1 position and a P residue at P2-position. Therefore, the authors claimed that the substrate specificity of MASP-3 is P-R/K-X. It is uncertain whether the kinetic behaviour of this truncated MASP-3 form resembles that of full-length MASP-3 or the catalytic portion of MASP-3.

In addition, the role of MASP-3 in activating the alternative pathway of complement is actively debated. Iwaki *et al*. (2011) stated that mouse recombinant MASP-3 after being incubated with mouse recombinant MBL-A and heat-killed *S. aureus,* cleaved proenzyme forms of factor B and factor D to activate the alternative pathway in MASP-1/-3 knock-out mice serum. The authors also claimed that even the proenzyme form of mouse recombinant MASP-3 was able to activate mouse factor D. In contrast, Degn *et al*. (2012) rejected these functions of MASP-3 in humans by reporting that a patient genetically deficient in both MASP-1 and MASP-3 had a normally functioning alternative pathway, which was not affected by the addition of MASP-3.

In this chapter, the C1r-activated M3Q was used to fully characterize the substrate specificity of MASP-3 by using a few synthetic peptide libraries. Initially the REPLi (Rapid Endopeptidases Profiling Library) library was used to determine the P1-residue and specificity of the flanking positions. Based on the REPLi results, an ACC-fluorescent substrate library with a fixed P1-residue and variable P4-P2 residues was applied to provide additional information on the preference for P4-P2 residues. An aldehyde inhibitor library with a fixed P1-residue adjacent to a C-terminal aldehyde group and variable P3-P2 residues was also used to obtain an effective inhibitor sequence, as well as providing extra information about preference of the enzyme for P3 and P2 residues. The physiological protein substrates of MASP-3 were also investigated using multiple approaches.

4.2 MATERIALS AND METHODS

4.2.1 Materials

The REPLi library and the PepSets aldehyde inhibitor library were manufactured by Mimotopes Inc. Australia (Clayton, VIC). The ACC-substrate library was synthesized by Associate Professor Marcin Drag of Wroclaw University of Technology, Poland. The AMC substrates were from Bachem (Bubendorf, Switzerland), and the Abz substrates were synthesized by GL Biochem (Shanghai, China). Vitronectin purified from human blood was purchased from Heamatologic Technologies Inc. US (Essex Junction, VT). Human C6 and factor H were purchased from Complement Technology Inc. US (Tyler, Texas). IGFBP-5 (Catalog No. GTX65125) was purchased from GeneTex, Inc. US (Irvine, CA).

4.2.2 Activity assay using synthetic AMC substrates

The synthetic peptide substrates Boc-Val-Pro-Arg-(AMC), Boc-Pro-Phe-Arg-(AMC), Boc-Gly-Gly-Arg-(AMC), Z-Phe-Arg-(AMC), Boc-Ala-Phe-Lys-(AMC) and Boc-Leu-Gly-Arg-(AMC) (denoted as VPR, PFR, GGR, ZFR, AFK, and LGR, respectively) were prepared as a 2 mM stock solution in 10% (v/v) dimethylsuloxide. Assays were carried out in fluorescence assay buffer (FAB) (50 mM Tris-HCl, 150 mM NaCl, 0.2% [w/v] PEG 8000, pH 7.4) at 37°C on a POLARstar fluorescent plate reader from BMG Labtech (Offenburg, Germany). The 96-well assay plates used were the Costar-3693 half area white polystyrene plates purchased from Corning Inc. US. For a typical activity assay, 100 nM M3Q was used to react with 100 µM substrate. The rate of increase of fluorescence was monitored by using an excitation wavelength of 360 nm and an emission wavelength of 460 nm for 30 min. The initial rate of cleavage of

substrate was measured as the slope of the linear portion of the progressive curve, and was then converted to the rate of increase of AMC molecules per unit enzyme ($M_{[AMC]}min^{-1}M_{[E]}^{-1}$) by using an AMC-fluorescence standard curve. To determine the values of the steady-state reaction constants, 100 nM M3Q was used to react with substrate at a range of concentrations from 0-0.5 M and the initial velocity of reaction was plotted against the concentrations of substrate. The Michaelis-Menten constant (K_m), the catalytic constant (k_{cat}) and the maximal velocity (V_{max}), were estimated by using the GraphPad Prism 5.0 Software (La Jolla, CA) via fitting a single site binding equation or, the Michaelis-Menten equation: $Y = V_{max}*X / (K_m + X)$, where Y is the initial rate of the proteolytic reaction and X is the concentration of substrate, and $k_{cat} = V_{max}/[E]$, where [E] is the enzyme concentration.

4.2.3 Characterization of substrate specificity of MASP-3 by using REPLi

The REPLi (Rapid Endopeptidases Profiling Library) method employing a library of tripeptidyl substrates was recently developed by Thomas *et al.* (2006) to study the amino acid specificity spanning the active site of a protease of interest. Based on the technology of fluorescence resonance energy transfer (FRET) (Figure 4.1), the tripeptidyl substrates are assembled in the sequence of donor fluorophore-(Nt)-G-G-G-tripeptide-G-G-quencher molecule-K-K-(Ct), in which the donor fluorophore (MeOC) and the quencher molecule (DPA) are a FRET pair, forming a donor-acceptor moiety. Therefore, when there is no cleavage of the core tripeptide by the protease, the whole substrate sequence is intact and the donor-acceptor moiety is in close proximity, consequently the quencher molecule quenches the fluorescent energy from the donor fluorophore, and thus no fluorescence change can be detected. In contrast, when the

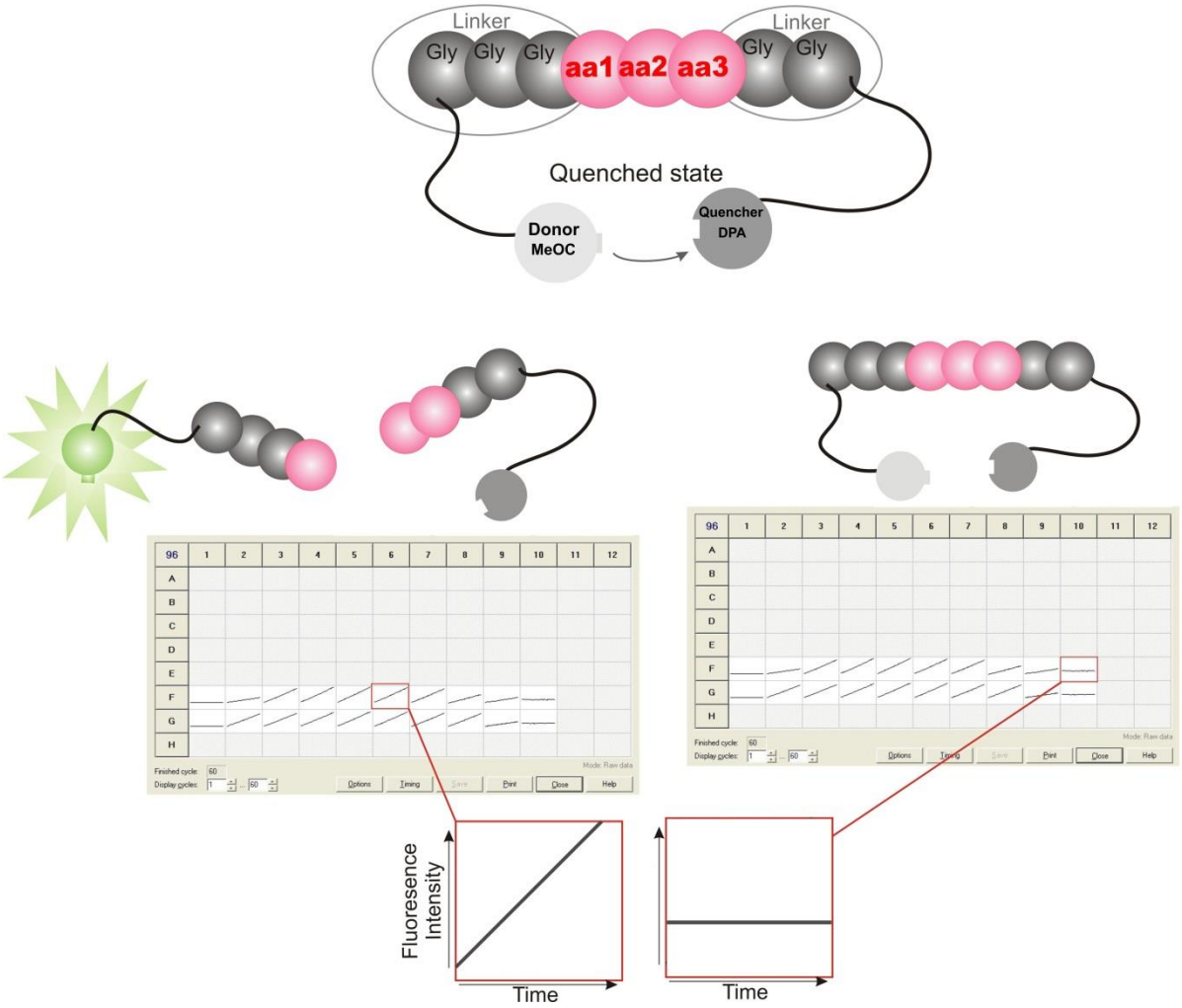
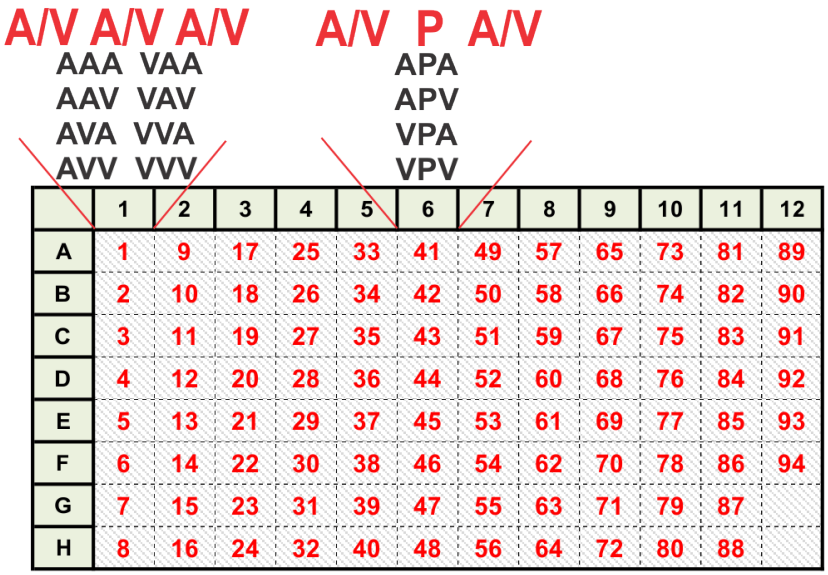


Figure 4.1 A schematic display of the REPLi method
Top, a tripeptidyl substrate in REPLi in which the donor fluorophore (MeOC) and the quencher molecule (DPA) are a FRET pair, forming a donor-acceptor moiety (arrow). The core tripeptide is illustrated in red, Gly in grey, donor fluorophore in pale grey, and quencher in light grey. The 15 amino acids individually making up the core tripeptide are grouped into 8 groups: Ala/Val, Arg/Lys, Asp/Glu, Asn/Gln, Leu/Ile. Ser/Thr, Phe/Tyr and Pro. Bottom left, when an enzyme cleaves the core tripeptide, the donor-acceptor moiety of the FRET pair moves apart, hence the fluorescence signal of MeOC (in Green) can be detected, which is displayed as increasing fluorescence intensity over time in a fluorescence graph. Bottom right, when there is no cleavage of tripeptide, the donor-acceptor moiety is intact. The fluorescent signal is therefore quenched by DPA and hence the fluorescence intensity does not increase over time.

protease cleaves the core tripeptide in a substrate, so the donor-acceptor moiety moves apart and the donor fluorophore is released from the quencher molecule; as a result, the fluorescence from the fragment containing the donor fluorophore can be detected.

The core tripeptides in the REPLi consist of combinations of all amino acids except 5 potentially problematic or unnecessary residues. These are: residue C, for its ability to form disulfide bonds; H residue, rarely observed as a residue close to the scissile bond; M residue, for its potential for being oxidised and for its similarity in size and hydrophobicity to I and L amino acids; W residue, since its emission wavelength (320 nM) overlaps the excitation wavelength of the donor fluorophore in this project; and G residue, since the information for cleavage between all other amino acids and G residue can be obtained from the G residues flanking the core tripeptide. The remaining 15 amino acids give rise to 3375 distinct tripeptides, which are allocated into 512 distinct pools (Figure 4.2). For more precise characterization of the tripeptides, the allocation of core tripeptides into pools is not random, but is based on the grouping of the 15 amino acids, that is, into 8 groups, which include 7 pairs of amino acids (A/V, R/K, D/E, N/Q, I/L, S/T and F/Y) according to their similarity in size and chemical properties, and the P residue, which provides additional structural information. Therefore, in each of the 512 pools, the variable tripeptide sequence is made of 3 of these 8 groups, giving rise to up to 8 tripeptides of distinct amino acid sequences (or 4, 2 or 1 tripeptide sequences when there are 1, 2 or 3 P residues, respectively).

To characterize the substrate specificity of MASP-3, a final concentration of 300 nM M3Q was mixed with 50 µM tripeptidyl substrates in FAB at 37˚C in each pool on the Costar-3693 white plates. Cleavage of the core tripeptides caused the MeOC-DPA donor-acceptor moiety to move apart, releasing the MeOC molecules from the quenched state conducted by the DPA quencher molecules. The fluorescence intensity



	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	
H	8	16	24	32	40	48	56	64	72	80	88	

PLATE A

Pool #	aa1	aa2	aa3
1	A/V	A/V	A/V
2	A/V	A/V	K/R
3	A/V	A/V	D/E
4	A/V	A/V	N/Q
5	A/V	A/V	I/L
6	A/V	A/V	P
7	A/V	A/V	S/T
8	A/V	A/V	F/V
9	A/V	K/R	A/V
10	A/V	K/R	K/R
11	A/V	K/R	D/E
12	A/V	K/R	N/Q
13	A/V	K/R	I/L
14	A/V	K/R	P
15	A/V	K/R	S/T
16	A/V	K/R	F/V
17	A/V	D/E	A/V
18	A/V	D/E	K/R
19	A/V	D/E	D/E
20	A/V	D/E	N/Q
21	A/V	D/E	I/L
22	A/V	D/E	P
23	A/V	D/E	S/T
24	A/V	D/E	F/V
25	A/V	N/Q	A/V
26	A/V	N/Q	K/R
27	A/V	N/Q	D/E
28	A/V	N/Q	N/Q
29	A/V	N/Q	I/L
30	A/V	N/Q	P
31	A/V	N/Q	S/T
32	A/V	N/Q	F/V
33	A/V	I/L	A/V
34	A/V	I/L	K/R
35	A/V	I/L	D/E
36	A/V	I/L	N/Q
37	A/V	I/L	I/L
38	A/V	I/L	P
39	A/V	I/L	S/T
40	A/V	I/L	F/V
41	A/V	P	A/V
42	A/V	P	K/R

Figure 4.2 The allocation of tripeptide pools in REPLi
The 3375 tripeptides with distinct sequences are allocated to 512 distinct pools, each of which contains 1~8 distinct tripeptides, depending on the number of Pro residues. For example, (right), the tripeptides in Pool 1 are made of A/V-A/V-A/V and in Pool 41 are of A/V-P-A/V. Consequently (left), the tripeptidyl substrates in Pool 1 include 8 peptides (AAA, VAA, AAV, VAV, AVA, VVA, AVV, VVV), whilst the tripeptides in Pool 41 include 4 (APA, APV, VPA, VPV).

from the MeOC molecules was detected at 55s per cycle for 30 cycles, with an excitation wavelength of 320 nm and an emission wavelength of 420 nm. The initial velocity of the cleavage was indicated by the slope per unit time of the linear region of the curves. These values were converted into molar MeOC molecules per unit time per molar enzyme ($M_{[MeOC]}min^{-1}M_{[E]}^{-1}$) by using an MeOC fluorescence standard curve, allowing comparison of the cleavage rate between pools and between enzymes.

4.2.4 Activity assay using individual Abz substrates

Based on the REPLi results, 17 characteristic individual peptides from the substrate pools containing the tripeptidyl sequences of R/K-I/L-F/Y, R/K-R/K-I/L and R/K-S/T-I/L, which displayed high or modest rates of cleavage by M3Q, were synthesized in the template layout of Abz-(Nt)-GGG-XXX-GG-(Ct)-Tyr(3NO2). The increase of Abz fluorescent intensity resulting from mixing 100 µM individual substrates with 500 nM M3Q was monitored by using an excitation wavelength of 320 nm and an emission wavelength of 420 nm for 30 min. The initial rate of cleavage of substrate was measured as the slope of the linear portion of the progressive curve, and was then converted to $M_{[Abz]}min^{-1}M_{[E]}^{-1}$ by using the standard curve. To determine the values of the steady-state reaction constants, 500 nM M3Q was mixed with substrate at a range of concentrations from 0-0.5 M and the initial velocity of reaction was plotted against the concentrations of the substrate. The Michaelis-Menten constant (K_m), the catalytic constant (k_{cat}) and the maximal velocity (V_{max}), were estimated by using GraphPad Prism 5.0 as previously described.

4.2.5 Determination of the P1-specificity of MASP-3

The Abz substrates RIY, KIY, RKL and RRL at the final concentration of 100 µM were incubated with 1 µM M3Q at 37°C overnight. Mass spectrometry analysis was applied for determination of the P1 specificity of M3Q. Samples containing the M3Q-cleaved products of RIY, KIY, RKL and RRL, respectively, were first separated by high performance liquid chromatography. Fractions containing peptide peaks were then co-spotted onto the MALDI target plate with a Matrix solution of 10 mg/mL α-cyano-4-hydroxycinnamic acid in 50% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid. The samples were analysed on an Applied Biosystems (Foster City, CA) 4700 Proteomics Analyzer MALDI TOF/TOF in reflectron mode with a mass range of 400-1500 Da, focus mass of 1000 Da at 1500 shots/spectrum using plate model calibration against 4700 tune mix. Peak detection on the spectra was performed using the software 4700 Series Explore 3.0.

4.2.6 Determination of the P4-P2 specificity by using the ACC-substrate library

The ACC-substrate library features substrates with a P1-Arg adjacent to a C-terminal ACC group and variable P4-P2 residues. Cleavage between the P1-R and ACC group by M3Q therefore releases the ACC group, the fluorescence of which can be detected. The library encompasses three sets, detecting the specificity at the P2-, P3- and P4-position, individually. Each set consists of 19 sub-libraries. Substrates in a single sub-library contain one of the 19 amino acids at the detecting position and variable residues at the other two positions. Therefore each sub-library contains 19 * 19 = 361 individual substrates. The 19 amino acids in this library include all natural amino acids, except amino acids C and M in order to avoid oxidation artefacts, and include residue

Nle, a synthetic isoform of amino acid L with a similar structure to M residue. In order to screen the ACC substrate library, a final concentration of 5 µM M3Q was incubated with 50 µM ACC substrates in each sub-library in FAB in a total final volume of 100 µl. The reaction was set at 37˚C on the Costar-3693 white plates. The rate of increase of fluorescence was monitored by using an excitation wavelength of 355 nm and an emission wavelength of 460 nm for 60 min. The initial rate of cleavage of substrates was measured as the slope of the linear portion of the progressive curve.

4.2.7 Aldehyde inhibitor library screening for M3Q

The PepSets aldehyde inhibitor library consists of synthetic peptidyl inhibitors with P1-R adjacent to a C-terminal aldehyde group and variable P3 and P2 residues. Therefore the library contains 20 * 20 = 400 individual inhibitors. In order to screen the library with M3Q, a final concentration of 600 nM M3Q was incubated with 100 µM of each inhibitor in FAB in a total volume of 50 µl. The inhibition reaction was set at 37˚C on the Costar-3693 white plates for 1 hr. The residual activity of M3Q in the reaction mixture was indicated by using 50 µM of an AMC substrate, either LGR or VPR. The rate of increase of fluorescence was monitored by using an excitation wavelength of 360 nm and an emission wavelength of 460 nm for 30 min. The initial rate of cleavage of substrate was measured as the slope of the linear portion of the progressive curve.

4.2.8 Examination of the proteolytic activity of M3Q against protein substrates

The proteolytic activity of M3Q was examined by incubating 1 µM protein substrate either with or without 1 µM M3Q in a total volume of 20 µl, at 37ºC for 4 hr. The samples were then reduced, denatured and analysed by SDS-PAGE or silver staining.

4.2.9 Silver staining

After SDS-PAGE, the proteins in the gel were fixed by submerging the gel in 40% (v/v) methanol with 7% (v/v) glacial acetic acid for 15 min. The gel was then soaked twice in 10% (v/v) ethanol for 7 min. After washing three times for 10 min with water, the gel was placed in a reduced solution with 60 ng/ml DTT for 20 min, following which it was stained with 0.15% (w/v) silver nitrate for 20 min in a light proof container and quickly rinsed with water. The proteins in the gel were visualized by gradually adding the developer (3% [w/v] sodium carbonate, 0.1% [v/v] formaldehyde, freshly made) into the gel container. Once the intensity of the protein bands reached the desired level, the development was terminated by immediately adding an excess amount of anhydrous citric acid to the developer mix.

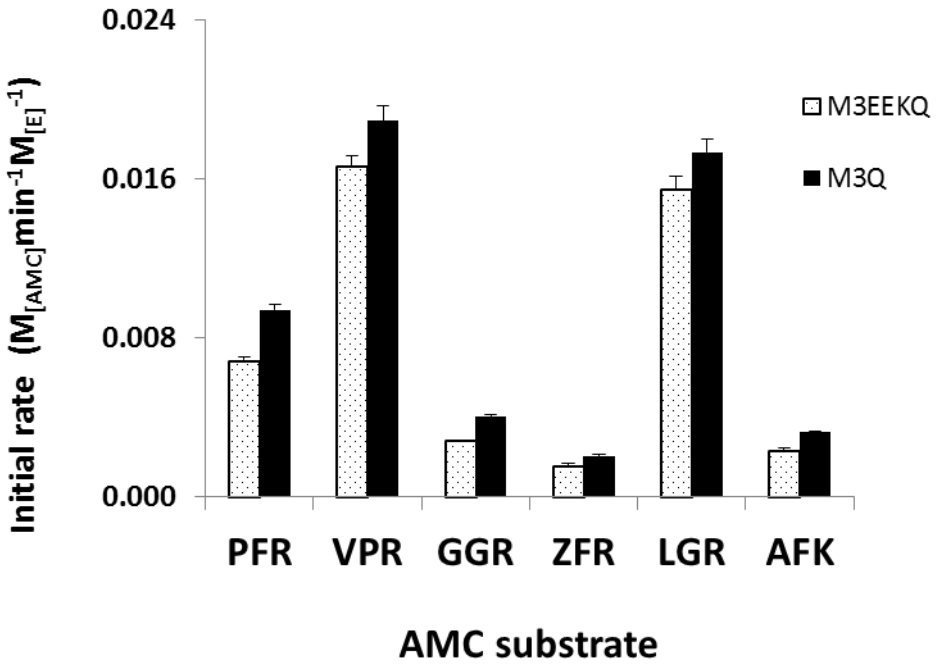
4.3 RESULTS

4.3.1 The activity of M3EEKQ and M3Q against AMC substrates indicated that the two enzymes were equally active

A number of synthetic peptidyl substrates containing a C-terminal fluorescent AMC group were used to investigate the peptidase activity of M3EEKQ and M3Q (Figure 4.3). The initial rate of proteolysis was indicated as the rate of release of AMC molecules per enzyme unit. M3EEKQ and M3Q cleaved all of AMC substrates tested, including AFK, although at a very low rate. Both enzymes exhibited a similar preference of cleavage of the six AMC substrates as VPR > LGR > PFR > GGR > AFK > ZFR, ranked by the initial rates of cleavage. Among the substrates tested, VPR and LGR yielded the greatest rates of proteolysis by both enzymes. The kinetics of catalysis of either M3EEKQ or M3Q against VPR and LGR was further characterized by determining the Michaelis-Menten curves (Figure 4.4). The k_{cat}/K_m value of the cleavage of VPR (5.71 $M^{-1}s^{-1}$ by M3EEKQ and 5.8 $M^{-1}s^{-1}$ by M3Q) was slightly higher than that of LGR (4.31 $M^{-1}s^{-1}$ by M3EEKQ and 4.84 $M^{-1}s^{-1}$ by M3Q), confirming that VPR was a more preferable substrate than LGR for cleavage by both MASP-3 mutants. Together, these results indicated that both M3EEKQ and M3Q were equally active and behaved very similarly in catalysis.

4.3.2 Varying buffer conditions did not substantially improve the catalytic activity of MASP-3

In comparison with FAB in which all activity assays were performed in this project, alteration of the buffer by adding up to 10 mM of Ca^{2+}, Mg^{2+} or K^{+} did not affect the

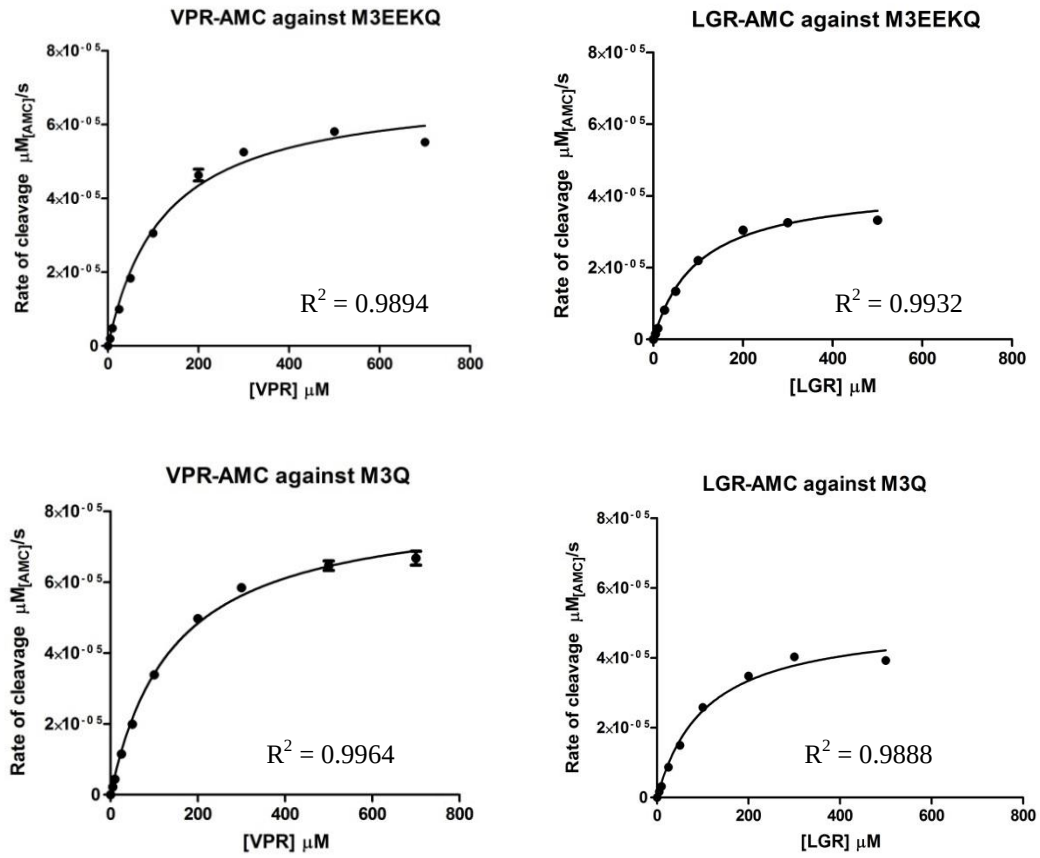


The rate of cleavage of AMC substrates by MASP-3 protein mutants ($M_{[AMC]}min^{-1}M_{[E]}^{-1}$)

Substrate	M3EEKQ	M3Q
VPR	0.018 ± 0.0008	0.019 ± 0.0008
LGR	0.016 ± 0.0007	0.017 ± 0.0007
PFR	0.007 ± 0.0003	0.009 ± 0.0003
GGR	0.003 ± 0.0001	0.004 ± 0.0001
ZFR	0.001 ± 0.0001	0.002 ± 0.0001
AFK	0.002 ± 0.0001	0.003 ± 0.0001

The Mean ± S.E.M. values of three independent tests were indicated.

Figure 4.3 The rate of cleavage of AMC substrates by MASP-3 mutants
Assays were carried out in fluorescence assay buffer (FAB) at 37°C on a POLARstar fluorescent plate reader. A final concentration of 100 µM substrate was incubated with 100 nM of either M3EEKQ or M3Q. The initial rate of cleavage of substrate was measured as the slope of the linear portion of the progressive curve using GraphPad Prism 5.0 software, and was then converted to the rate of increase of AMC molecules per unit enzyme ($M_{[AMC]}min^{-1}M_{[E]}^{-1}$) by using an AMC-fluorescence standard curve. For all measurements, the correlation co-efficient values (R^2 values) for the fitting of the lines were larger than 0.99.



Kinetics of the cleavage of *VPR* -AMC and *LGR* -AMC by M3EEKQ or M3Q

Substrate	k_{cat} (s^{-1})	K_m (µM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Cleaved by M3EEKQ			
VPR	0.0007 ± 0.00002	122.9 ± 12.0	5.71
LGR	0.0004 ± 0.00001	99.8 ± 8.6	4.31
Cleaved by M3Q			
VPR	0.0008 ± 0.00002	142.8 ± 8.2	5.8
LGR	0.0005 ± 0.00002	105.6 ± 11.3	4.84

Mean ± S.E.M. from three independent experiments are shown.

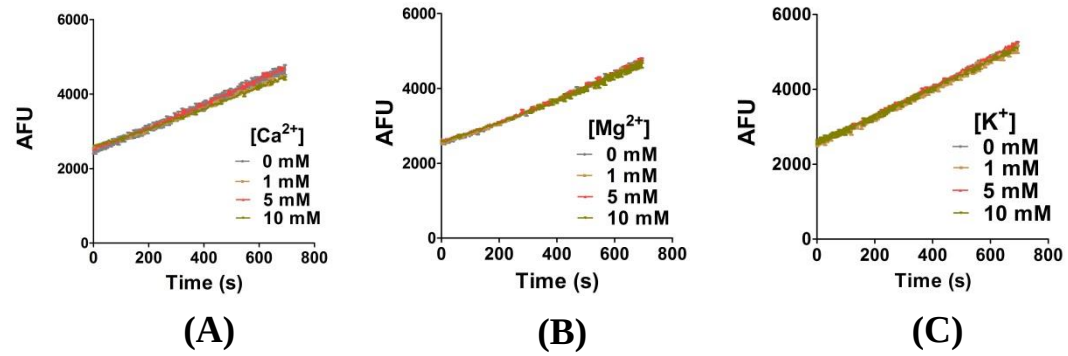
Figure 4.4 Kinetics of cleavage of *VPR*-AMC and *LGR*-AMC by M3EEKQ andM3Q
Either M3EEKQ or M3Q at a final concentration of 100 nM was incubated with substrate at a range of concentrations from 0-0.5 M. The initial velocity of reaction was plotted against the concentrations of substrate to yield the Michaelis-Menten curves using GraphPad Prism 5.0 software. From the curves, the K_m and k_{cat} values were obtained. The correlation co-efficient values (R^2 values) for the fit of each line are indicated.

catalytic activity of M3Q (Figure 4.5A-C). Various buffer conditions, including in presence of Tris, PEG8000, HEPES, EDTA, and Na_2HPO_4, did not significantly alter the rate of substrate cleavage by M3Q, either. Interestingly, a clear decrease of rate of cleavage of VPR was observed in a buffer containing a high concentration of 1 M NaCl, suggesting that high concentration of salt may be a negative cofactor for the catalysis by MASP-3 (Figure 4.5D). The pH value of the buffer did affect the catalytic activity of M3Q, and the optimal pH range was 7.0-8.5 (Figure 4.5E).

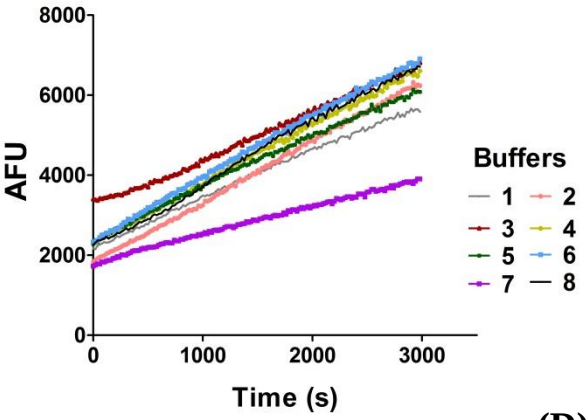
4.3.3 The overall substrate specificity of MASP-3 determined by REPLI

REPLi screening of the primary substrate specificity of M3Q (300 nM) was carried out and the initial rate of proteolysis of substrates in each pool was measured and converted into $M_{[MeOC]}min^{-1}M_{[E]}^{-1}$ for comparison. The overall REPLi profile for M3Q was summarized and shown in Figure 4.6. Out of the total 512 substrate pools in a REPLi set, 134 pools displayed proteolysis by M3Q, with rates of cleavage from 0 ~ 1.40 $M_{[MeOC]}min^{-1}M_{[E]}^{-1}$. The difference in the rate of proteolysis relative to other proteases screened against REPLi (data not shown) indicated that M3Q is a weak protease against most tripeptidyl substrates in REPLi. Out of the 134 pools which displayed M3Q proteolytic activity, 101 pools consisted of substrates with at least one K/R residue in the tripeptidyl sequences, and the remaining 33 pools exhibited only low rate of proteolysis by M3Q (less than 0.5 $M_{[MeOC]}min^{-1}M_{[E]}^{-1}$). Therefore, the proteolytic activity of M3Q favours the substrates with a K/R residue in their core tripeptides.

In order to obtain residue information from the characteristic pools in the REPLi, the residue compositions of the tripeptides in the top 20 pools best cleaved by M3Q were shown (Figure 4.6). Among the 20 pools, 17 pools consisted of tripeptides with at least three small hydrophobic residues, such as A/V, I/L, and the G residues in the substrate

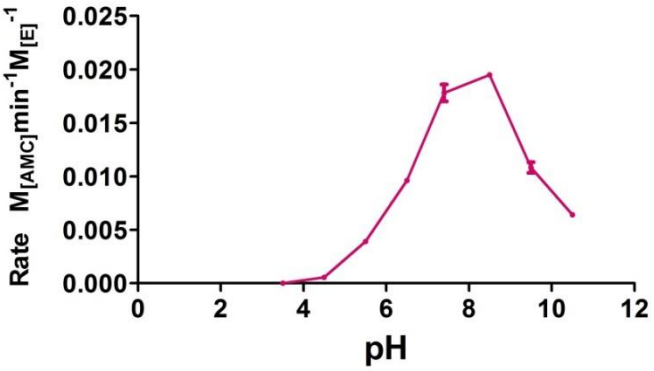


(A) **(B)** **(C)**



Buffer 1, FAB buffer.
Buffer 2, 50 mM Tris, 0.1M NaCl, 0.2% PEG8000.
Buffer 3, HEPES 100 mM.
Buffer 4, HEPES 50 mM, EDTA 1 mM.
Buffer 5, gel filtration buffer.
Buffer 6, Tris 50 mM.
Buffer 7, Tris 50 mM, 1M NaCl.
Buffer 8, 0.05 M Na_2HPO_4.

(D)



(E)

Figure 4.5 The catalytic activity of M3Q in varying conditions
The catalytic activity of 0.25 µM M3Q against 100 µM *VPR*-AMC was determined in varying conditions, including in the presence of (A), Ca^{2+}; (B), Mg^{2+}; and (C), K^{+}; (D) in various buffers and (E), in FAB at different pH values. For all measurements, the correlation co-efficient values (R^2 values) for the fitting of the lines were larger than 0.99. The data are representative of two independent experiments.

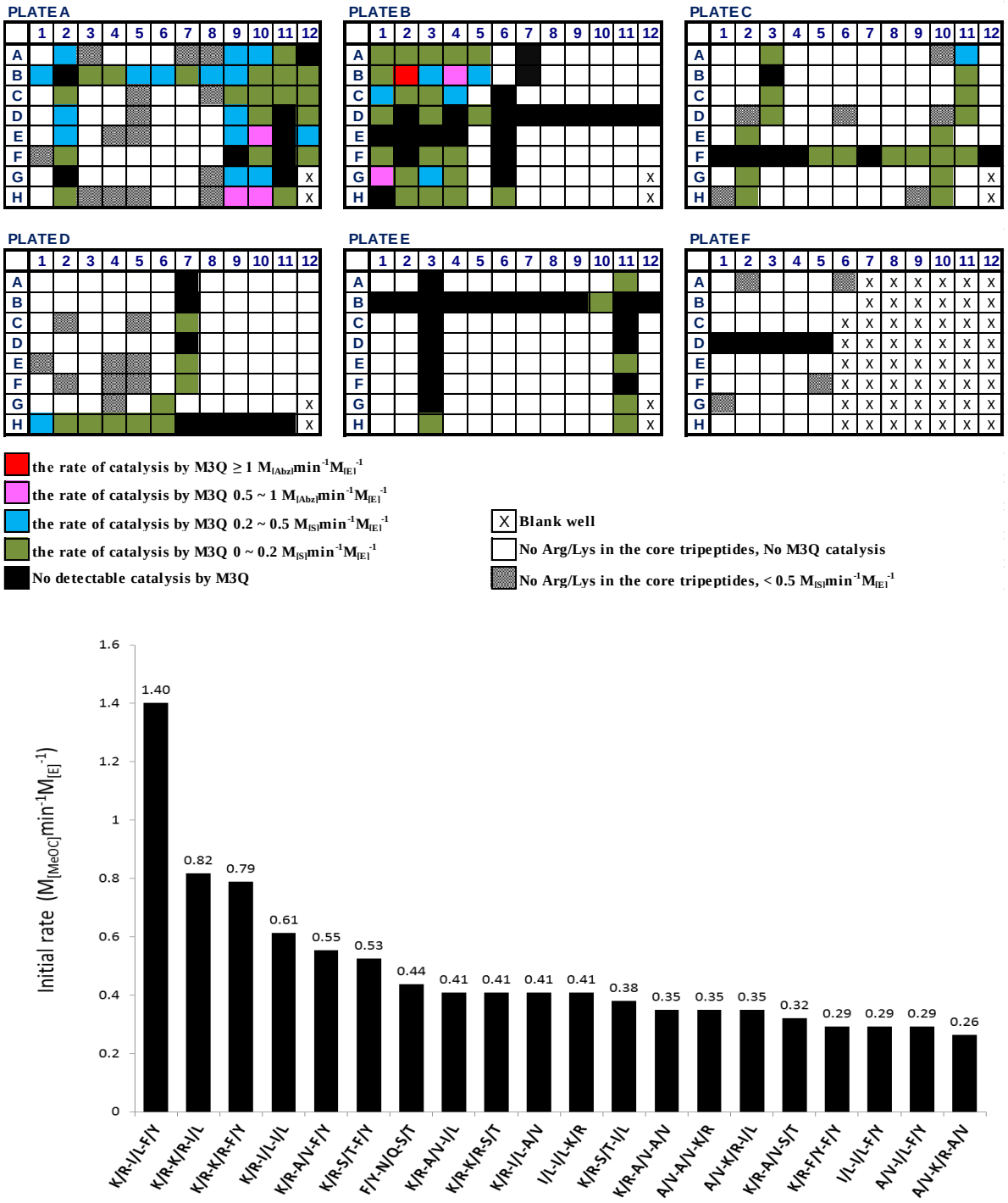


Figure 4.6 The overall specificity profile for M3Q enzyme using REPLi

Upper panel, the overall REPLi results by using coloured pools representing the rate of cleavage with M3Q. The peptide pools (50 μM) were incubated with 300 nM M3Q at 37°C. Fluorescence intensities were measured at 30 s intervals for 30 min. The coloured pools contain the substrates with at least one Arg/Lys residue in the core tripeptides and were coloured from red to black ranked by the rate of catalysis by 300 nM M3Q.

Lower panel, the rate of cleavage of twenty REPLi peptide pools best cleaved by M3Q. The initial velocities of cleavage for the pools are shown above each bar.

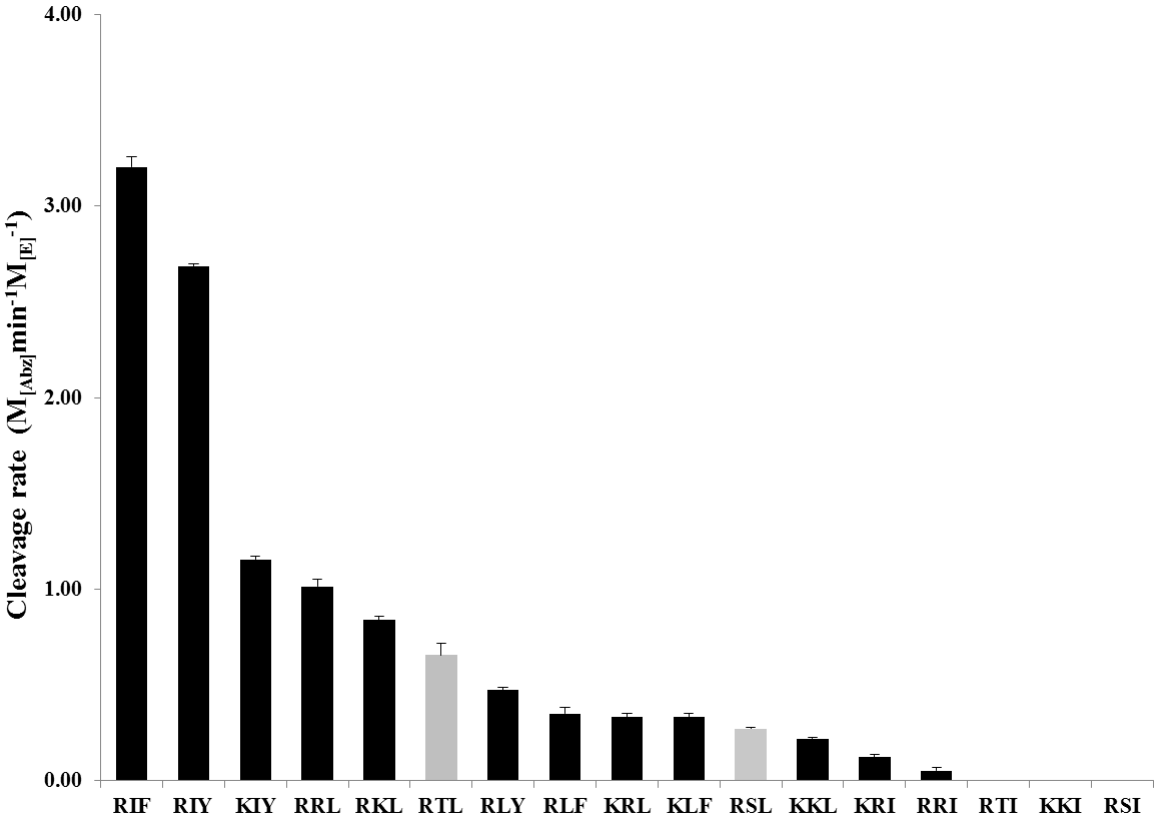
construct, before a K/R residue. This suggested that M3Q preferred substrates with the sequence of up to three small hydrophobic residues followed by a K/R residue. On the list, there were 7 substrate pools with tripeptidyl sequence of K/R-hydrophobic residue-hydrophobic residue. Among these sequences, M3Q appeared to have a narrow preference for K/R-I/L-F/Y (1.40 $M_{[MeOC]}min^{-1}M_{[E]}^{-1}$, the greatest rate of proteolysis in the catalogue), since substituting either the I/L residue or the F/Y residue of these tripeptides with other hydrophobic residues or hydrophilic residues hindered their rate of proteolysis by M3Q. M3Q also displayed a preference towards the tripeptidyl sequence of K/R-K/R-Xaa, of which a hydrophobic X residue was more preferable. Interestingly, 3 pools containing substrates with no K/R residue, but a common F/Y residue, in their tripeptidyl sequences showed low M3Q activity. Such low activity may be resulted from non-specific hydrolysis of the peptides themselves. The substrate pools containing tripeptides with at least one K/R residue, but not cleaved by M3Q were also reviewed to obtain information on the negative preference of residues. It was found that M3Q had preferences against a wide range of characteristic residues surrounding the K/R residue, including negatively charged residues, a P residue, small hydrophilic residues (S/T), and large hydrophobic residues (F/Y).

4.3.4 Identification of the optimal tripeptidyl substrates and the P1-specificity of MASP-3

Based on the REPLi results, the top pools for M3Q, including the pool K/R-I/L-F/Y and K/R-K/R-I/L, were most likely to contain the optimal tripeptidyl sequence for proteolysis by M3Q. In order to identify the optimal sequence, 13 characteristic tripeptidyl sequences (RIF, RIY, KIY, RRL, RRI, RKL, KRL, KRI, RLF, RLY, KLF, KKL and KKI) from these pools were synthesized as individual substrates. As controls,

4 characteristic tripeptidyl sequences (RTL, RSL, RTI and RSI) from the pool K/R-S/T-I/L were also selected, since M3Q exhibited a modest rate of proteolysis for this pool (0.38 $M_{[MeOC]}min^{-1}M_{[E]}^{-1}$). The 17 individual substrates were in the sequence layout of Abz-(Nt)-GGG-XXX-GG-(Ct)-Tyr($3NO_2$). The initial rates for the cleavage of these individual substrates by 500 nM M3Q were measured (Figure 4.7). RIF, RIY and KIY from the pool K/R-I/L-F/Y yielded the highest rates of cleavage with M3Q, followed by RRL and RKL from the pool K/R-K/R-I/L. These tripeptides contributed to the higher rate of total cleavage of the pool K/R-I/L-F/Y than the K/R-K/R-I/L pool. Among the 4 characteristic tripeptidyl sequences from the pool K/R-S/T-I/L, RTL and RSL displayed modest rates of cleavage with M3Q and the other two substrates were not cleaved by M3Q. Therefore collectedly the rate of cleavage of the pool K/R-S/T-I/L was modest, agreeing with the overall profile in REPLi. The Michaelis-Menten curves for the cleavage of the top five Abz substrates by 500 nM M3Q were also determined to identify the optimal substrate (Figure 4.8). The best individual substrate cleaved by M3Q was RIF, with a k_{cat}/K_m value of 98.55 $M^{-1}s^{-1}$, followed by RIY, with a k_{cat}/K_m value of 76.90 $M^{-1}s^{-1}$.

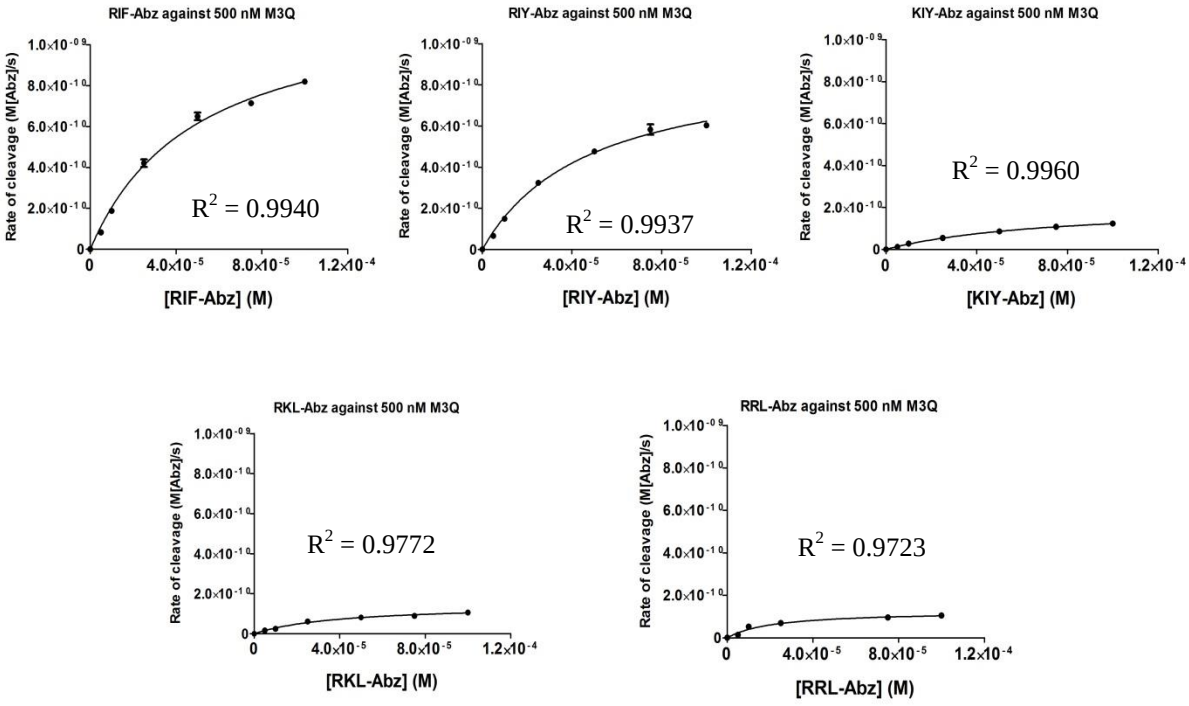
Mass spectrometry was applied to determine the cleavage sites in RIY, RRL, RKL, and KIY by M3Q (Figure 4.9). It was found that M3Q had a P1-R/K specificity and cleaved the peptide bonds between R-R in RRL, and between R-K in RKL. A P1-R residue was more preferable for M3Q proteolysis than a P1-K residue, since substituting the R residue in RIY with a K residue resulted in an 85.0% of decrease in the k_{cat}/K_m value.



Substrate	$M_{[Abz]}min^{-1}M_{[E]}^{-1}$	Substrate	$M_{[Abz]}min^{-1}M_{[E]}^{-1}$
RIF	3.20 ± 0.05	KRL	0.33 ± 0.02
RIY	2.69 ± 0.01	RSL	0.27 ± 0.01
KIY	1.16 ± 0.02	KKL	0.22 ± 0.01
RRL	1.02 ± 0.04	KRI	0.12 ± 0.02
RKL	0.84 ± 0.02	RRI	0.05 ± 0.02
RTL	0.66 ± 0.06	RTI	0.00
RLY	0.47 ± 0.02	RSI	0.00
RLF	0.35 ± 0.04	KKI	0.00
KLF	0.33 ± 0.02		

Mean ± S.E.M. of three independent experiments are shown.

Figure 4.7 The rates of cleavage of the individual Abz peptidyl substrates by M3Q
A final concentration of 500 nM M3Q was incubated with 100 µM individual Abz substrates in FAB at 37°C. The increase of Abz fluorescent intensity was monitored by using an excitation wavelength of 320 nm and an emission wavelength of 420 nm for 30 min. The initial rate of cleavage of substrate was converted to $M_{[Abz]}min^{-1}M_{[E]}^{-1}$ by using the standard curve. The rates of cleavage of RTL, RSL, RTI and RSI are shown in grey to visually discriminate these controls from the substrates derived from the peptide pools in REPLi best cleaved by M3Q. For all measurements, the correlation co-efficient values (R^2 values) for the fitting of the lines were larger than 0.99. The columns shown in the figure depict the Mean ± S.E.M. values derived from three independent tests.



Kinetics of the cleavage of Abz substrates by M3Q

Substrate	k_{cat} (s^{-1})	K_m (10^{-5}M)	k_{cat}/K_m ($M^{-1}s^{-1}$)
RIF	0.0049 ± 0.0002	4.97 ± 0.54	98.55
RIY	0.0037 ± 0.0002	4.80 ± 0.52	76.90
RRL	0.0005 ± 0.00004	2.17 ± 0.50	23.29
RKL	0.0006 ± 0.00005	4.00 ± 0.78	14.61
KIY	0.0009 ± 0.00006	7.46 ± 0.96	11.50

Mean ± S.E.M. of three independent experiments are shown.

Figure 4.8 Kinetics of the cleavage of Abz substrates by M3Q
M3Q at a final concentration of 500 nM was incubated with substrate at a range of concentrations from 0-0.5 M. The initial velocity of reaction was plotted against the concentrations of substrate to yield the Michaelis-Menten curves. The K_m and k_{cat} values were derived by non-linear regression fitting of the curves. The correlation co-efficient values (R^2 values) for the fit of each line are indicated.

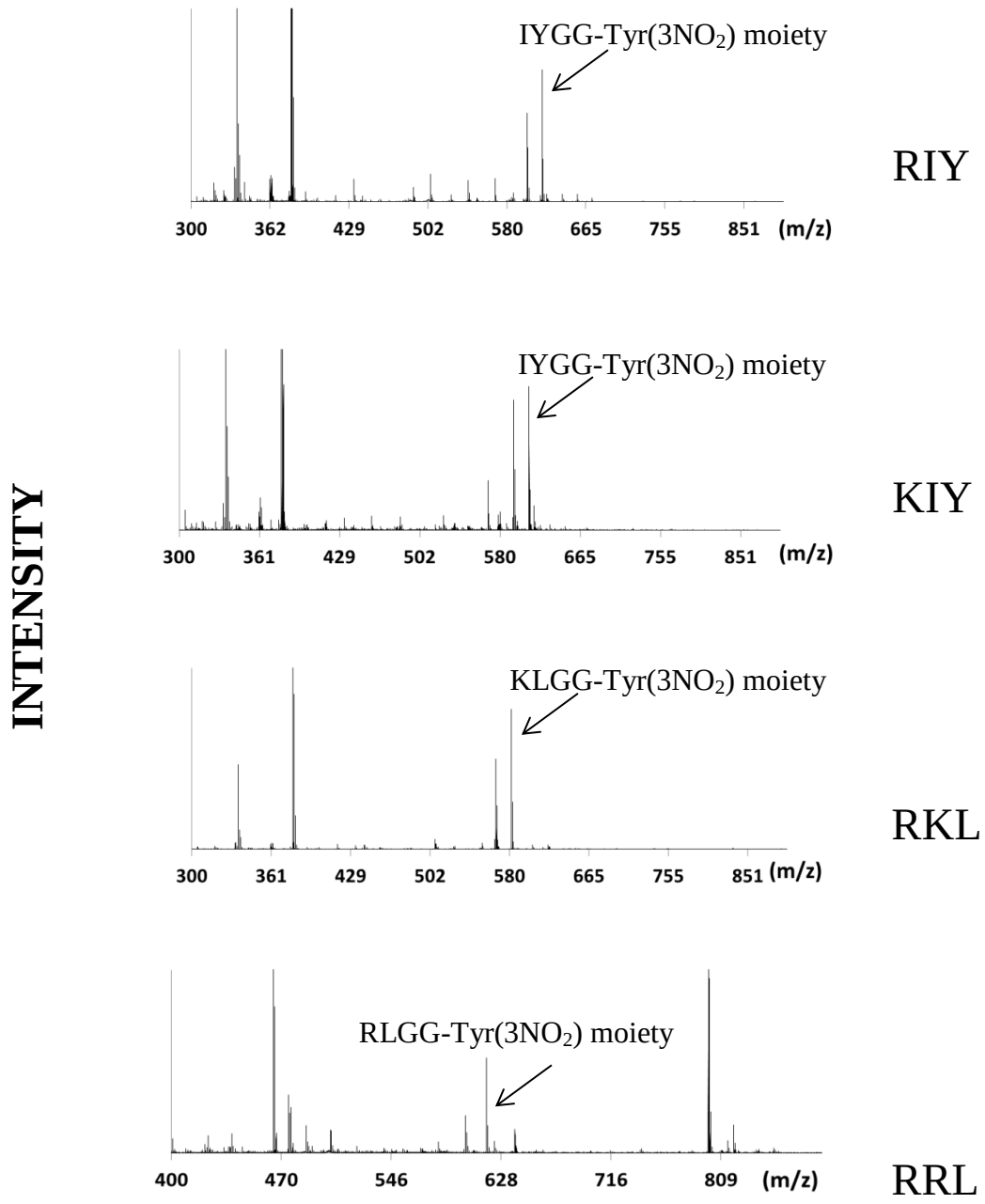


Figure 4.9 Representative mass spectra identifying cleaved products of the Abz-peptides, RIY, KIY, RRL, and RKL after treatment with M3Q.
The RIY, KIY, RKL and RRL peptides at a final concentration of 100 µM were incubated with 1 µM M3Q at 37°C overnight. Following high performance liquid chromatography, the samples were co-spotted onto the MALDI target plate and analysed on an Applied Biosystems (Foster City, CA) 4700 Proteomics Analyzer MALDI TOF/TOF. The fragments resulting from cleavage of Abz-peptides by M3Q were determined using the deduced m/z values, labelled with arrows.

4.3.5 Determination of the P4-P2 specificity of MASP-3 by using the ACC-substrate library

For the three sets of 19 sub-libraries characterizing the specificity at the P4, P3 and P2 positions, only 2, 3 and 5 sub-libraries, respectively, displayed detectable cleavage with 5 μM M3Q (Figure 4.10). Substrates with a P4-P residue were preferred for cleavage by M3Q. Interestingly, substrates with a P3-Nle residue, an unnatural amino acid with a side chain of four consecutive carbons, displayed a high preference for cleavage by M3Q, compared to those with a P3-L residue, the isoform of Nle with a shorter side chain of $-CH_2-CH-(CH_3)_2$. Substrates with a P2-G residue were preferred by M3Q. Substrates with a negative residue at the P2-position also exhibited cleavage by M3Q, indicating that M3Q was able to host a negatively charged residue at the P2. The overall rates of cleavage for the substrates showing detectable cleavage with M3Q were low at all three positions, indicating that the substrate specificity of MASP-3 may depend more on the sequence prime side to the scissile bond.

4.3.6 Aldehyde inhibitor library screening for M3Q

Out of the total 400 aldehyde inhibitors in the library, 31 attenuated the catalytic activity of M3Q by more than 50%. The 20 sequences of the inhibitors best attenuating M3Q activity are listed in Figure 4.11. Of these 20 sequences, 16 contain at least one W residue at the P2 and P3 positions, and the remaining 4 contain a C residue at P3. The inhibitor with the sequence WWR attenuated 94% of the catalytic activity of M3Q, and thus was the best inhibitor in the library against M3Q, followed by CWR (93.8%), IWR (92.6%) and YWR (92.3%).

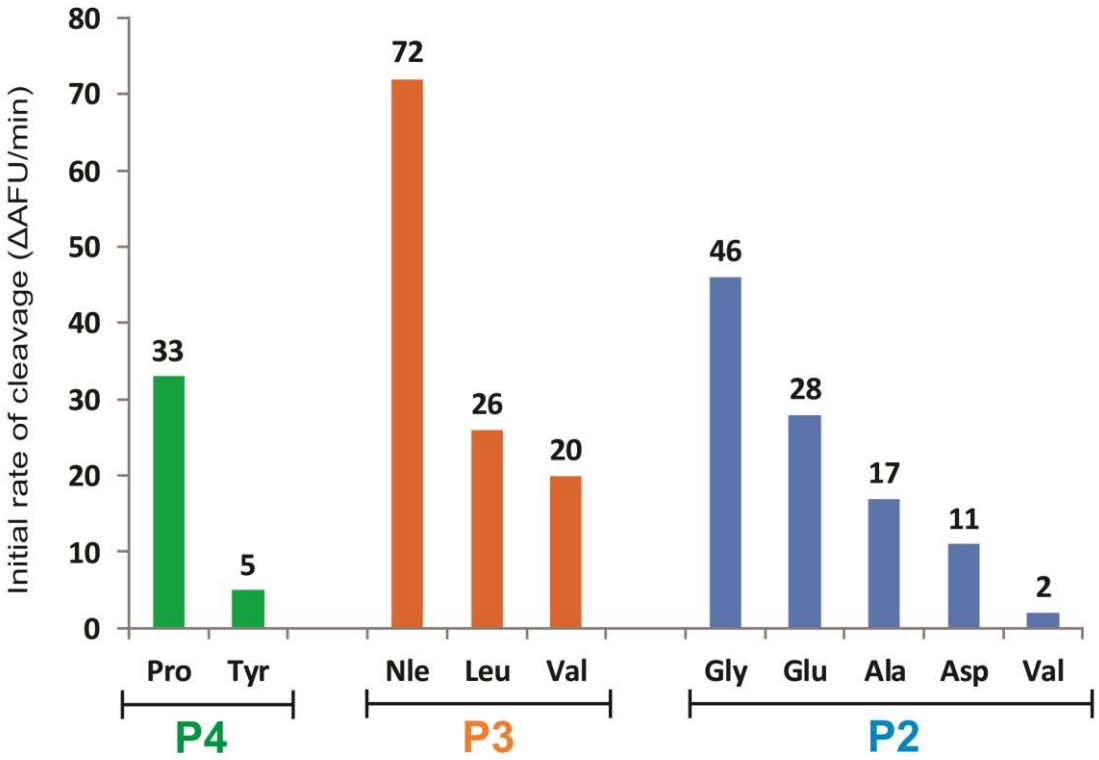


Figure 4.10 Substrate specificity of M3Q characterized by using the ACC-substrate library

The substrate pools or sub-libraries (50 μM) were incubated with 5 μM M3Q at 37°C. Fluorescence intensities were measured at 30s intervals for 30 min. Results of the substrate specificity at the P4, P3, P2 positions were illustrated in green, orange and blue, respectively. At each position, a marked amino acid represents a sub-library in which all of the substrates contain the marked amino acid at that position. The initial rates of cleavage for the sub-libraries that exhibited detectable cleavage are shown as the increase of arbitrary fluorescence units per min (ΔAFU/min). For all measurements, the correlation co-efficient values (R^2 values) for the fitting of the lines were larger than 0.99. The columns shown in the figure depict the results from a single measurement.

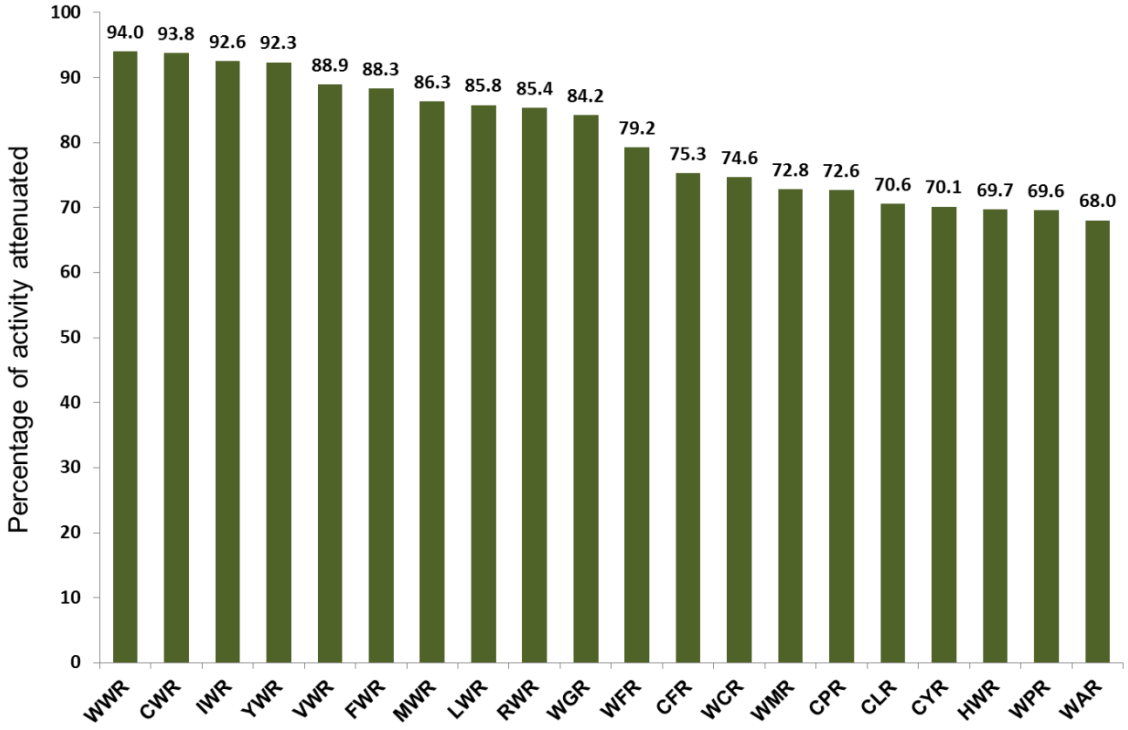


Figure 4.11 The sequences of aldehyde inhibitors that best attenuated the activity of M3Q (top twenty shown)

A final concentration of 600 nM M3Q was incubated either with or without 100 µM of inhibitor in the library in FAB in a total volume of 50 µl, at 37˚C for 60 min. The residual activity of M3Q was then measured by adding 50 µM of either LGR-AMC or VPR-AMC substrate and measured as the rate of increase of fluorescence intensity. The activity of M3Q attenuated by the inhibitors was compared with that of in absence of inhibitors to yield the percentage of activity attenuated. The values shown above each bar are the mean values resulting from two independent tests, of which, one used LGR-AMC, and another used VPR-AMC, as the indicated substrate. For all measurements, the correlation co-efficient values (R^2 values) for the fitting of the lines were larger than 0.99. The columns shown in the figure depict the results from a single measurement.

4.3.7 Examination of the proteolytic activity of M3Q against protein substrates

The proteolytic activity of M3Q was examined by incubating 1 μM protein substrate either with or without 1 μM M3Q, at 37ºC for 4 hr (Figure 4.12). Under the experimental conditions, M3Q did not cleave MASP-1-S195A, MASP-2-S195A, wild type MASP-3, C2, C4, C1r-S195A, C1s, Pro-thrombin, C6 or factor H. No clear formation of enzyme-inhibitor complex was observed when incubating M3Q with either C1-inhibitor or antithrombin. Cleavage of fibrinogen at the α-chain by M3Q was observed, indicating the M3Q was a functional serine protease.

Based on the results of the REPLi and the ACC-substrate library, the most preferred sequences for cleavage by MASP-3 include GRIF, GRIY, GKIF and GKIY. It is possible that some of the MASP-3 targets might contain one or more of these sequences to form cleavage sites for MASP-3. Therefore, the sequences GRIF, GRIY, GKIF and GKIY were used as probes to screen through the human proteome. The selection criteria of the resulting protein candidates included the locations of expression of the proteins and the location of these sequences in the 3D structure of the protein. Ideally, a good protein candidate for a MASP-3 target should be extracellular, either in plasma or outer-membrane-bound, and have an exposed target sequence on the surface of the protein structure. According to the criteria, vitronectin is a candidate substrate.

Vitronectin belongs to the pexin family. Full-length vitronectin is a 75 kD protein of 478 amino acids, contained within three domains: the N-terminal Somatomedin B domain with a 19-residue signal sequence, the central domain with homopexin sub-domains and a C-terminal domain which also contains a homopexin sub-domain (Xu *et al*., 2001). The sequence GRIY is located within the linker region between the central domain and the C-terminal domain, and thus may be exposed for cleavage by

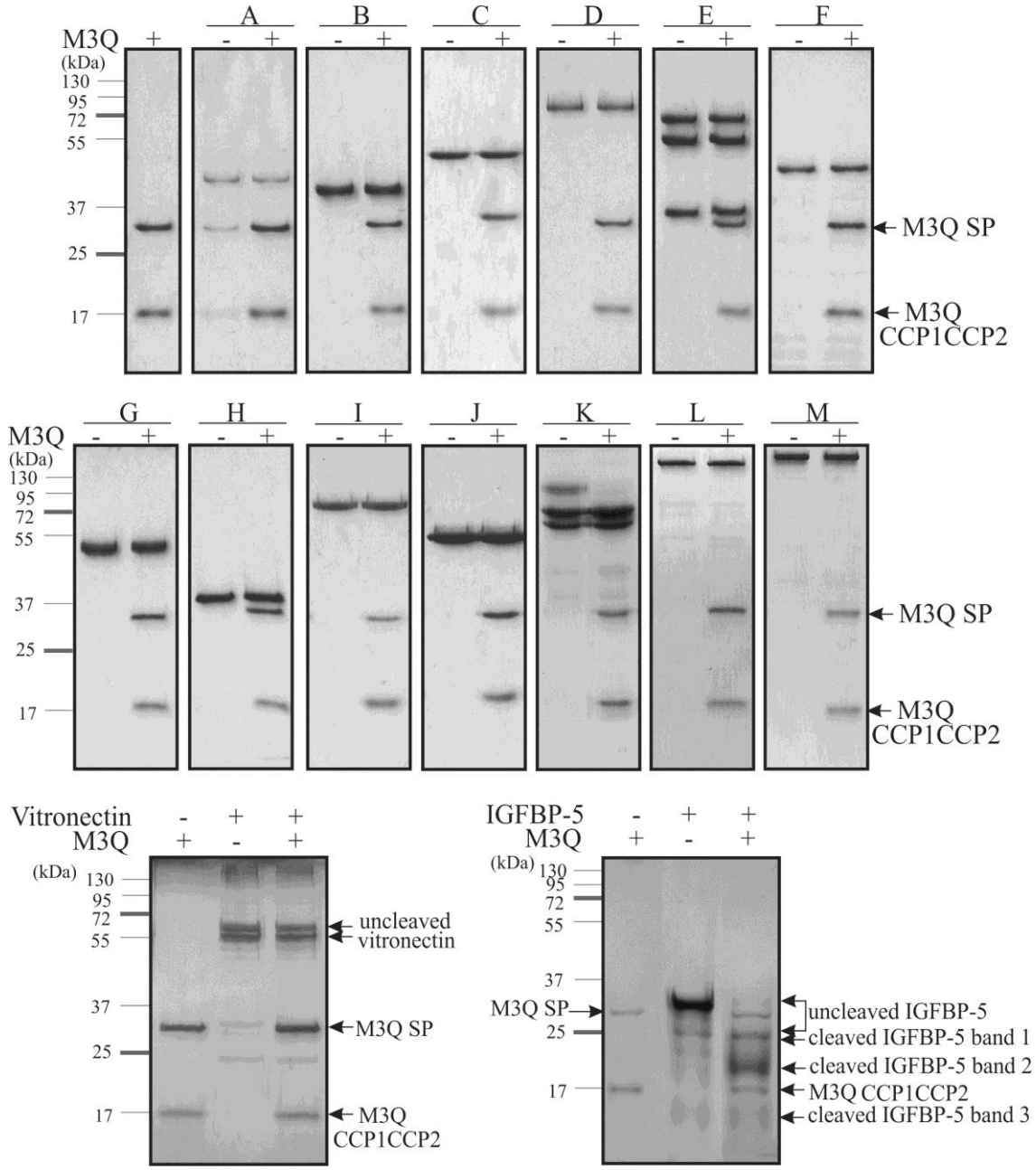


Figure 4.12 Cleavage of protein substrates by M3Q

Upper and middle panels, 1 µM of each tested protein, including the CCP1-CCP2-SP fragments of (A), MASP-1-S195A; (B), MASP-2-S195A; (C), MASP-3; (D), C2; (E), C4; the catalytic portions of (F), C1r-S195A and (G), C1s; (H), Pro-thrombin; (I), C1-inhibitor; (J), antithrombin; (K), fibrinogen; (L), C6; and (M), factor H, was incubated either with (+) or without (-) 1 µM active M3Q. The incubation was at 37°C for 4 hr. The samples were then reduced, denatured and analysed by SDS-PAGE with Coomassie staining. Bottom panel, silver stain analyses of cleavage of 1 µM vitronectin by by 1 µM M3Q (left) and 2.3 µM IGFBP-5 by 0.23 µM M3Q (right), both at 37°C for 4 hr. The cleavage of all tested proteins was conducted in 50 mM Tris, 145 mM NaCl, pH7.4, except for the cleavage of IGFBP-5, which was conducted in 50 mM Tris, pH8.5. The figures displayed here are representative of three independent experiments.

MASP-3. Vitronectin is responsible for cell-to-cell adhesion and communication (Zhou *et al*., 2003), and is also a regulatory component of the complement system (Sheehan *et al*., 1995). In order to examine whether MASP-3 cleaves vitronectin, human plasma vitronectin was incubated with M3Q at 37°C for 4 hr and analysed by SDS-PAGE followed by silver staining (Figure 4.12). However, M3Q did not cleave vitronectin.

Here, we also examined cleavage of IGFBP-5 by M3Q using similar buffer conditions and concentrations as those used in a previous study by Cortesio and Jiang (2006). The recombinant IGFBP-5 purchased for this project contained a main protein component of 34 kDa and a minor protein of 24 kDa. The tested IGFBP-5 was completely cleaved by M3Q after 4 hr, resulting in three fragments of 24, 22 and 15 kDa. We were unable to sequence the fragments due to their low protein concentration, resulting in poor visualization of protein bands after they were transferred onto a PVDF membrane and subsequently stained with Coomassie.

4.4 DISCUSSION

A few synthetic peptides were claimed to be cleaved by MASP-3, including AMC substrates cleaved by a truncated MASP-3 SP fragment (Cortesio and Jiang, 2006). Of these AMC substrates examined, the preference for catalysis by the truncated form of MASP-3 was PFR > VPR > ZGGR > ZFR > ZLR, and the truncated enzyme did not cleave GR, AFK and AAPF. In our study, both M3EEKQ and M3Q cleaved the AMC substrates in order of preferences: VPR > LGR > PFR > GGR > AFK > ZFR. The difference in the preferences for cleavage between the two studies indicates that there is a structural difference in the active site/substrate binding site of the proteins studied. Our data also showed that both M3EEKQ and M3Q proteins highly prefer the catalysis of VPR and LGR. These two synthetic substrates were therefore used to illustrate the activity of MASP-3. Further, the rates of cleavage of all six AMC substrates by M3Q were similar to those by M3EEKQ, indicating that both MASP-3 mutants are equally active, and that the configurations of the SP domains of the two active MASP-3 mutants are similar. Therefore, M3Q was used for the rest of the studies.

The focus of this chapter was to characterize the substrate specificity of MASP-3 enzyme. The REPLi library is a commercial substrate library encompassing a combination of 3365 unique tripeptides. It provides valuable specificity information flanking the scissile bond (P1-P1') of the substrate and has been validated for multiple enzymes, including trypsin, pepsin, matrix metalloprotease (MMP)-12, MMP-13 and calpains-1 and -2 (Thomas *et al*., 2006). In our study, the REPLi screening revealed an overall preference of substrates with at least one K/R residue in the tripeptide for cleavage by M3Q, suggesting a P1-K/R specificity, which was confirmed by subsequent mass spectrometry analysis. M3Q most preferred the sequences K/R-I/L-F/Y, among which the sequences RIF and RIY gave the highest rates of cleavage,

indicating a general preference for hydrophobic residues at the P1' and P2' positions. Under the experimental conditions, the sequence RIF was identified as the best substrate cleaved by M3Q, with a k_{cat}/K_m value of 98.55 $M^{-1}s^{-1}$, about 20-fold higher than that for VPR, which was claimed to be one of the best substrates for MASP-3, according to Cortesio and Jiang's study (2006) using the truncated MASP-3 SP fragment. Substituting the R residue of RIY with a K residue resulted in less than half the rate of cleavage of the substrate, indicating a P1-R preference over P1-K. Interestingly, M3Q also prefers cleavage of K/R-K/R-I/L, and cleaves between R-R and R-K dibasic bonds. The P1' residue of MASP-3 substrate therefore can also be a basic residue, not exclusively a hydrophobic residue. Compared to this, the P2' residue of MASP-3 displayed a clear preference of hydrophobic residue, since 8 of the 10 pools (K/R-I/L-F/Y, K/R-K/R-I/L, K/R-K/R-F/Y, K/R-I/L-I/L, K/R-A/V-F/Y, K/R-S/T-F/Y, K/R-A/V-I/L and K/R-I/L-A/V) best cleaved by M3Q in the REPLi contained substrates with a hydrophobic residue at the P2' position (Figure 4.6). Substrates in these 8 pools also contained a G residue at the P2 position, suggesting a high preference for a P2-G residue.

The P2-G specificity of MASP-3 was confirmed by the results of the ACC-substrate library screening. The results showed that the P2 position of MASP-3 substrates can be occupied by a small hydrophobic residue, such as residues G and A, or a negatively charged residue, but the enzyme most prefers a G residue. The ACC-substrate library screening also suggested a preference of hydrophobic residue as the P3 residue and a P4-P specificity. Overall, the specificity of P4-P2 residues appeared to be in favour of hydrophobic residues, suggesting a wide-spread hydrophobic region on the corresponding substrate-binding region of MASP-3. However, the overall rates of cleavage of substrates in the entire library appeared to be extremely low, indicating that

the determination of substrate specificity only by P4-P2 residues may not be efficient. The overall specificity for hydrophobic residues at the non-prime side to the scissile bond was also observed in the results of the aldehyde inhibitor library screening. The majority of inhibitors that best interacted with M3Q contained at least one large hydrophobic residue (W) at the P2 and P3 positions. The reason why the small hydrophobic G residue did not show up as a popular P2-residue as in the previous REPLi and ACC-substrate libraries is most likely due to a change of structure induced by the interaction between the aldehyde group and the active site of M3Q, widening the substrate-binding region of M3Q for the P3-P2 residues. Nevertheless, the preferences for hydrophobic residues at the P3 and P2 positions are confirmed.

Together, the results of screening with multiple libraries indicated that MASP-3 displayed a strong P2-P2' preference toward the sequences GRIF, GRIY and the structurally similar sequences GKIF and GKIY. It is possible that the physiological substrate of MASP-3 contains one or more of these sequences and is cleaved by MASP-3. These sequences therefore were used as probes to screen through the human proteome in search of the protein substrates for MASP-3. Although the results of the screening showed that human vitronectin is a good candidate to be cleaved by MASP-3, M3Q did not cleave vitronectin under the experimental conditions. It is common that the interaction of an enzyme to its substrate does not solely depend on individual residues 'best fitting' into the binding subsites on the enzyme, but also is strongly affected by allosteric effects induce by the interaction of exosite on the enzyme for the substrate. Thus structures outside the above P2-P2' sequences identified may also participate in and/or determine the interaction between MASP-3 and its protein substrates. Therefore, these 'best' sequences identified for cleavage by M3Q do not absolutely appear in the physiological substrates of MASP-3. Including more probes of

sequences with similar structures and/or with different patterns may increase the likelihood of identifying the real substrates of MASP-3, however, that may not be an effective and feasible way to identify the MASP-3 targets. In addition to such method of probing with actual sequences, another possible method to identify MASP-3 substrates is using common structures surrounding the scissile bond as probes to screen through the human proteome with known crystal structures. This latter method takes into account the dynamic aspects of an enzyme-substrate interaction, but it also requires specificity information at more positions and specially developed software to achieve.

Thus far, IGFBP-5 was claimed to be a MASP-3 substrate, cleaved by a truncated MASP-3 SP examined by Cortesio and Jiang (2006). We also observed cleavage of IGFBP-5 by M3Q. The recombinant IGFBP-5 used here appeared to be different to the IGFBP-5 protein used in the previous study, which displayed two components of 37 and 32 kDa after reduction and analysed using SDS-PAGE. Nevertheless, cleavage of IGFBP-5 by M3Q resulted in three fragments of 24, 22 and 15 kDa, which may reflect the same cleavage sites as those cleaved by the truncated MASP-3 SP in the previous study. In both studies, the tested IGFBP-5 was completely cleaved after 4 hr, indicating that the non-glycosylated nature of M3Q did not severely affect its catalytic activity toward IGFBP-5. According to Cortesio and Jiang's study, the three cleavage sites were found to be R^{176}-I^{171} of the sequence PRIISAP, K^{135}-I^{136} of PKIFRPK and K^{140}-H^{141} of PKHTRIS. All three cleavage sites contain a P residue on the N-terminal side. The author thereby proposed that the MASP-3 specificity is P-K/R-X, with a hypothetical P1-K/R specificity. Substrates with the sequences of P-K/R-X or X-P-K/R were not shown in the REPLi screening as one of the most preferable for cleavage by M3Q. Interestingly, the pattern of RII of the sequence PRIISAP and the KIF of PKIRPK stood

out in the REPLi results as being among the best substrates for cleavage by M3Q. In addition, although the H residue was not included in the REPLi library sequences, a cleavage in between dibasic residues with a P1-R/K specificity was found to be highly preferred by M3Q. Therefore, despite the different configurations of proteins used, the three cleavage sites in IGFBP-5 cleaved by the truncated MASP-3 SP appeared to fit with the REPLi data. Whether IGFBP-5 is a real substrate for MASP-3 and what biological outcomes are implicated by MASP-3 cleaving IGFBP-5 are yet to be determined. Besides IGFBP-5, mouse pro-factor B and pro-factor D were recently claimed to be substrates of mouse MASP-3, which was activated either by human MASP-1 or by sequentially incubating with mouse MBL-A and heat-killed *S. aureus* (Iwaki *et al*., 2011). Of the two components of murine alternative complement, mouse pro-factor D was also cleaved by the zymogen of mouse recombinant MASP-3. By using our catalytic portion of human MASP-3 and C1r-activated M3Q, we did not observe activation of MASP-3 by MASP-1, nor any catalytic activity of M3Q toward synthetic substrates or a series of proteins tested. Interestingly, despite all these puzzles, the cleavage site in pro-factor B to form Ba and Bb fragments is R^{259}-K^{260} of the sequence KRKIVL, and that in pro-factor D is R^{5}-I^{6} of PGRILGG. The patterns of RKI and GRIL were among the most popular sequences best cleaved by M3Q, identified by the REPLi screening. Although the involvement of MASP-3 in activating the alternative complement pathway via processing pro-factor B and pro-factor D was contested by Degn *et al*. (2012), who reported that a patient genetically deficient in both MASP-1 and MASP-3 had a normally functioning alternative pathway, which was not affected by the addition of MASP-3, it is still of interest to see that the processed sites on the two protein substrates actually contain sequences best cleaved by MASP-3.

By incubating active M3Q with a series of protein components of the immune system and the coagulation system, M3Q was found not able to activate C1r, C1s, MASP-1, MASP-2 or wide type MASP-3, all of which contain a conserved R-I activation site, within the sequences QRIIGG, QRIIGG, ARIFNG, GRIYGG, and KRIIGG, respectively. Among these sequences, the patterns RII, RIF, GRIY and KRI were all picked up in the REPLi analysis as among the best sequences cleaved by M3Q. The lack of cleavage of these proteins by M3Q highlighted that the interaction between protein molecules is determined mostly by structural determinants, rather than residues or sequences. M3Q was also shown not to cleave C2, C4, C6, pro-thrombin or factor H, and was not inhibited by C1-inhibitor and antithrombin, either. M3Q was able to cleave human fibrinogen at the α-chain. Fibrinogen can be cleaved by many human proteases with P1-K/R specificity because it contains multiple K/R residues in a long and exposed α-chain. However, among all the possible cleavage sites with a P1-K/R residue, none contains the sequences best cleaved by M3Q, based on the REPLi analysis. This once again emphasizes the importance of protein structure in determining the interaction between protein molecules.

Chapter Five

The Crystal Structure of a 3MC Syndrome-Associated MASP-3 Mutant

5.1 INTRODUCTION

The 3MC syndrome is a rare autosomal recessive disorder termed by combining four diseases, the Carnevale, Mingarelli, Malpuech and Michels syndromes. 3MC patients display a series of congenital characteristic anomalies, including abnormally increased distance between facial parts, bilateral ptosis with abnormally small and droopy eyelids, cleft lip and/or palate, abnormal pattern of skull, fusion of radius and ulna, as well as impairment in hearing and/or leaning, resulting from abnormal growth during embryogenesis (Leal *et al*., 2008).

The 3MC syndrome was recently found to be related to genetic defects of *COLEC11* and *masp-1/3*, both of which encode proteins of the lectin complement pathway (Sirmaci *et al*., 2010, Rooryck *et al*., 2011). A few missense mutations in the *COLEC11* gene were found in 3MC patients, resulting in disruption of the secretion of CL-11 (Rooryck *et al*., 2011), an important binding partner of MASP-3 (Hansen *et al*., 2010). The MASP-3-CL-11 protein complex thus would not be found in the serum of patients. Further, a nonsense mutation affects the residue W^{290} within the CUB1-EGF-CUB2 region, which is common to all proteins encoded by the *masp-1/3* gene, resulting in depletion of all *masp-1/3* protein products, including MASP-3 (Sirmaci *et al*., 2010). More importantly, four independent mutations in the *masp1-3* gene were identified to occur in exon 12, which is specific to the SP domain of MASP-3, affecting the residues C^{630}, G^{666}, G^{687}, and the active site H^{497}, which are substituted by an R, E, R, and Y residue, respectively (Sirmaci *et al*., 2010, Rooryck *et al*., 2011). Of these, besides the active site residue H^{497}, the C^{630}, G^{666} and G^{687} residues are conserved among MASP-3 of many vertebrate species (Figure 5.1), indicating the importance of these residues to the structure and function of MASP-3. Therefore, all genetic evidence points to a

correlation between the 3MC syndrome and a lack of function of MASP-3 or the mutated MASP-3 enzymes.

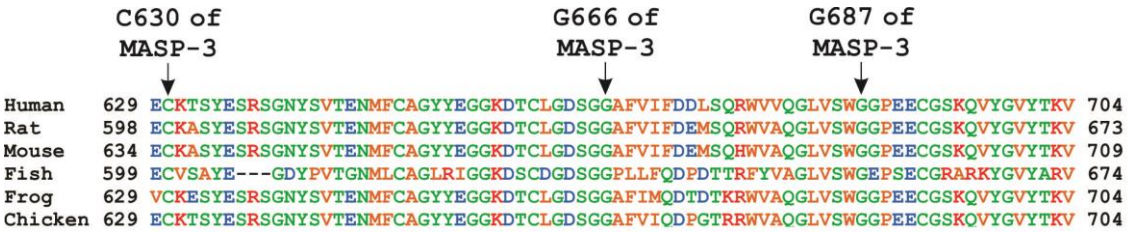


Figure 5.1 Conservation of residue C^{630}, G^{666} and G^{687} (arrows) in MASP-3 among vertebrate species.
A sequence alignment of aa629-704 of *Homo sapiens* MASP-3 (GenBank ID AAK84071.1), aa598-673 of *Rattus norvegicus* MASP-3 (GenBank ID CAD32171.1), aa634-709 of *Mus musculus* MASP-3 (GenBank ID AAN39849.1), aa599-674 of *Branchiostoma belcheri* MASP-3 (GenBank ID BAC75887.1), aa629-704 of *Xenopus laevis* MASP-3 (GenBank ID NP_001082342.1) and aa629-704 of *Gallus gallus* MASP-3 (GenBank ID AAS73180.1). The conserved residues C^{630}, G^{666} and G^{687} of human MASP-3 are marked with arrows.

Of the four 3MC syndrome-associated mutations specific to MASP-3 SP, mutation at H^{497}, a residue of the catalytic triad of MASP-3, is very likely to result in a dysfunctional MASP-3 mutant. Mutation of C630R may disrupt the structure and functions of MASP-3 enzyme since C^{630} is strictly conserved in all members of the C1r/C1s/MASP family of enzymes, and thus would most likely participate in forming a disulfide bond vital to the structure of MASP-3 enzyme. The other two MASP-3 mutants (G666E and G687R) were most likely to result in intact proteins. However, the effects of these changes in MASP-3 associated with the syndrome were less certain. Further, without a structure of the catalytic portion of MASP-3 at the time this study was carried out, the structural basis of the 3MC syndrome-related MASP-3 mutants and their catalytic activity is thus far only hypothesized. Based on a structural model of MASP-3 SP developed on the basis of human MASP-1, Sirmaci and her colleagues (2010) proposed that the G^{687} residue is located within the active site and that G^{687} is in close proximity to the residue D^{658}. The authors further suggested that the 3MC

syndrome-related G687R substitution results in a salt bridge formed by R^{687} and D^{658}, diminishing the catalytic activity of MASP-3. The D^{658} residue of MASP-3 corresponds to D^{640} of MASP-1 and D^{189} in trypsin. The D^{640} residue of MASP-1 was found to form an intermolecular salt bridge with R^{677}, restricting the catalytic activity of MASP-1 (Dobó *et al*., 2009) and substituting D^{189} in trypsin with a serine residue totally disrupts the catalytic activity of trypsin (Gráf *et al*., 1988). These findings support the unproven hypothesis by Sirmaci *et al*. (2010) that the 3MC syndrome-related MASP-3G687R mutant is a dysfunctional protein.

In this chapter, we applied the previously developed M3Q form of human MASP-3, which contains a single Q residue replacing the K residue prior to the R-I activation bond, to produce the 3MC syndrome-associated mutants, G666E and G687R. The resulting catalytic portions of M3QG666E and M3QG687R were therefore efficiently cleaved and activated by the C1r protease of the classical pathway of complement. Both mutants lacked detectable catalytic activity against all synthetic peptide substrates that we identified previously for MASP-3. We then determined the 2.6 Å crystal structure of the M3QG666E mutant protease in the zymogen form. The data strongly suggest substantial perturbation of the active site, consistent with a correlation between the lack of MASP-3 function and the 3MC syndrome.

5.2 MATERIALS AND METHODS

5.2.1 Cloning

The parental M3Q DNA incorporated in the pET-17b vector was produced as described in Chapter 3, Section 3.2.3, and was used as the template. A set of forward and reverse primers respectively encoding for the G666E and G687R mutations was used for PCR to generate the constructs of the mutant. The nucleotide sequences of the primers for each 3MC syndrome-associated M3Q mutant are listed in Table 5.1. The PCR products were processed with *DpnI* restriction enzyme to digest the parental DNA, following which the constructed plasmids were transformed into the DH5α *E. coli* cells, and the transformants were selected on LB plates containing 100 μg/ml ampicillin. Following DNA extraction by using the PureYieldTM Plasmid Miniprep kit from Promega, US (Madison, WI), the recombinant constructs were sent for DNA sequencing to confirm the sequences. The DNA for both M3Q mutants encodes the residues corresponding to the region A^{296}-R^{728} of wild type MASP-3. A short sequence of codons encoding three amino acids (Ala-Ser-Met) of the T7-Tag was introduced prior to the sequences of the recombinant proteins to enhance protein expression.

Table 5.1 Primers for 3MC syndrome-associated M3Q mutants

Mutant	Forward Primer	Reverse primer
G666E	5′GCCTTGGAGATAGCGGTGAGG CCTTTGTCATCTTTG3′	5′CAAAGATGACAAAGGCCTCAC CGCTATCTCCAAGGC3′
G687R	5′GGCCTGGTGTCCTGGAGGGGA CCTGAAGAATGCGGC3′	5′GCCGCATTCTTCAGGTCCCCT CCAGGACACCAGGCC3′

5.2.2 Expression, unfolding, refolding and purification of recombinant M3QG666E and M3QG687R proteins

The expression, unfolding, refolding and purification of recombinant M3QG666E and M3QG687R proteins were carried out using the methods described in Chapter 3. The concentrations of the recombinant proteins were derived by dividing the OD_{280} value by the calculated absorption coefficient ($\varepsilon_{0.1\%,\ 1\ cm,}$ assuming all pairs of Cys residues form disulfide bonds). The $\varepsilon_{0.1\%,\ 1\ cm}$ value of each recombinant protein and its calculated molecular mass from its amino acid sequence are listed in Table 5.2.

Table 5.2 The calculated masses & absorption coefficients of the recombinant proteins

Protein fragments	Calculated mass (Da)	$\varepsilon_{0.1\%,\ 1\ cm}$
M3QG666E CCP1-CCP2-SP	48,119	1.69
M3QG687R CCP1-CCP2-SP	48,146	1.69

5.2.3 Activation of M3QG666E and M3QG687R proteins

To activate the 3MC syndrome-associated M3Q mutant proteins, 2.5 mg of either M3QG666E or M3QG687R in 5 ml gel filtration buffer was loaded into the C1r-column. The activation was set at 26°C for 18 hr. The activated M3Q mutant protein was eluted by 6.5 ml gel filtration buffer. SDS-PAGE, western blotting and N-terminal sequencing were applied to confirm that the eluted proteins were activated M3Q mutant proteins.

5.2.4 Activity assays for the 3MC Syndrome-associated M3Q mutants

The catalytic activity of C1r-activated M3QG666E and M3QG687R was examined by using synthetic AMC substrates, including VPR, PFR, GGR, ZFR, AFK and LGR, and the 17 Abz substrates that were used for characterizing the substrate specificity of MASP-3 as described in Chapter 4.

5.2.5 Crystallization of the zymogenic M3QG666E protein

During the purification process of gel filtration, the M3QG666E protein was exchanged into a buffer containing 300 mM NaCl, 20 mM Tris-HCl, pH 7.4 and concentrated to 5 mg/ml. The crystallization experiment was carried out using the sitting-drop vapour-diffusion method at 4°C in a 96 well-Art Robbins LP Intelli-plate using a reservoir buffer comprised of 120 mM NaCl, 18% (w/v) polyethylene glycol 4000 and 100 mM imidazole, pH 7.5. 0.1 µl protein solution was mixed with 0.1 µl reservoir buffer and allowed to equilibrate over 50 µl reservoir solution. Crystals were observed after 1 day and grew to maximal size after 3 weeks. Crystals were cryogenically cooled in reservoir buffer supplemented with 30% (w/v) sucrose. A single crystal was then mounted in a nylon cryoloop prior to snap-frozen in liquid N_2.

5.2.6 Data collection and processing

Diffraction data was collected at cryogenic temperatures on the MX2 beamline at the Australian Synchrotron (Clayton, VIC) to a resolution of 2.6 Å. Data was indexed, integrated, scaled and merged using the Xia2 data reduction pipeline (Winter, 2009) which utilises XDS (Kabsch, 1988a, b, 1993), Pointless, Aimless (Evans, 2006) and the

CCP4 suite. A molecular replacement solution was found following utilisation of the EMBL-HH Automated Crystal Structure Determination Platform "Auto-Rickshaw" (Panjikar *et al*., 2005, 2009) which contained two molecules per asymmetric unit with clear unbiased fo-fc electron density observed in omitted loop regions. Manual iterative model building and refinement using Coot (Emsley *et al*., 2010) and phenix.refine (Adams *et al*., 2010) was performed by Dr. Pascal Wilmann, Monash University, to yield a final model of high quality (R-fac/R-free 19.5/26.0 – Molprobilty Score 3.29). All subsequent structural analyses were performed by using the PyMol system.

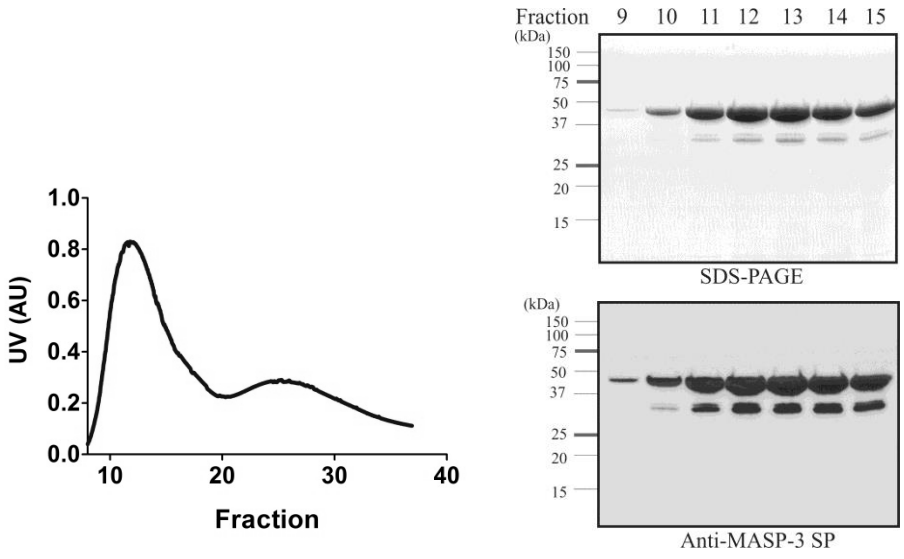
5.3 RESULTS

5.3.1 The purification and activation of recombinant M3QG666E and M3QG687R proteins

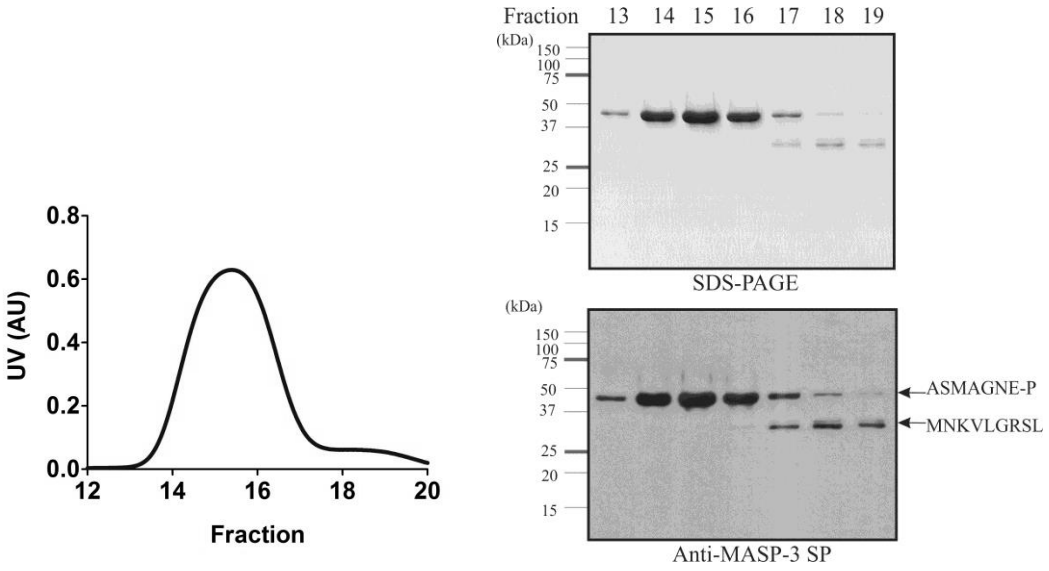
The catalytic fragments of M3QG666E and M3QG687 were expressed in the BL21STAR(DE3) *E. coli* hosts and purified by anion-exchange and size-exclusion chromatography (Figure 5.2-5.3). Both proteins at reduced conditions displayed a single band of 48 kDa in SDS-gel, with the N-terminal sequence of ASMAGNE-P, indicating that they are the single polypeptide chains of zymogenic MASP-3 CCP1-CCP2-SP (Figure 5.4). Generally 1-3 mg of each recombinant protein of high homogeneity could be yielded per litre *E. coli* culture. Both M3QG666E and M3QG687R zymogens were effectively processed by C1r protease, resulting in two bands of 32 kDa and 17 kDa in SDS-gel under reduced conditions, with the N-terminal sequence of IIGGRNAE and ASMAGNE-P, respectively (Figure 5.4). The two protein mutants therefore were activated correctly by C1r.

5.3.2 The catalytic activity of recombinant M3QG666E and M3QG687R

The catalytic activity of the C1r-activated M3QG666E and M3QG687R was examined by using up to 1 μM of each enzyme against the AMC substrates (VPR, PFR, GGR, ZFR, AFK and LGR) and the 17 Abz substrates (RIF, RIY, KIY, RRL, RRI, RKL, KRL, KRI, RLF, RLY, KLF, KKL and KKI), which were previously used for characterizing the specificity of M3Q. Neither protein mutant exhibited any detectable cleavage of any of the above substrates.



(A)



(B)

Figure 5.2 Purification of the M3QG666E (CCP1-CCP2-SP) fragment

Panel (A), anion exchange chromatography of M3QG666E. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 15 µl sample from the indicated fraction. Fractions 10-14 were pooled for further purification by size-exclusion chromatography.

Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 10-14. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 15 µl sample from the indicated fraction. The N-terminal sequences of the 48 and 32 kDa fragments are displayed and the corresponding bands are marked with arrows. This indicates that the 48 kDa protein was M3QG666E. Fractions 13-15 therefore contained M3QG666E at high purity.

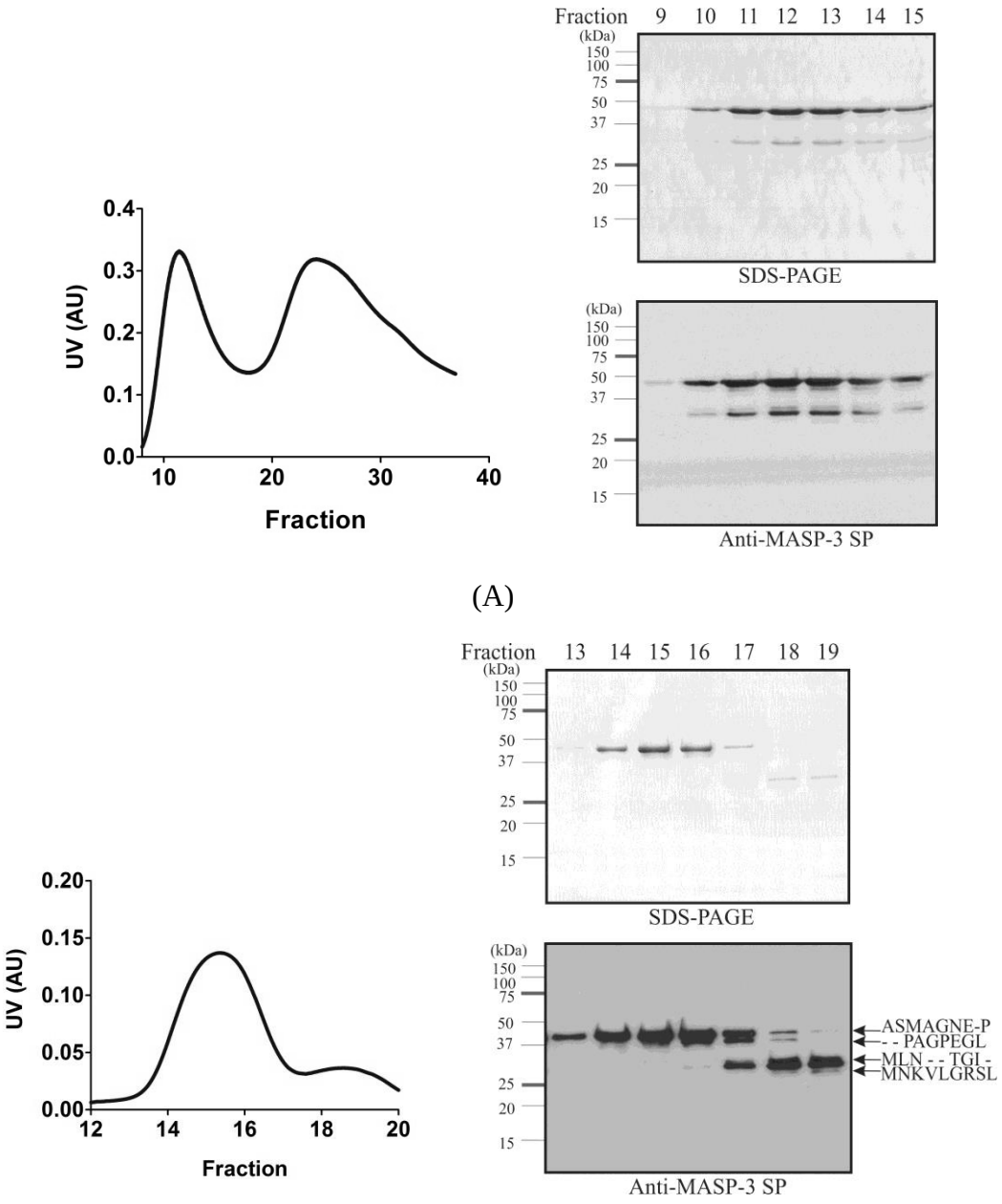


Figure 5.3 Purification of the M3QG687R (CCP1-CCP2-SP) fragment

Panel (A), anion exchange chromatography of M3QG687R. Left, chromatogram of anion exchange on a HiTrap Q-Sepharose column. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the anion exchange fractions under reduced conditions. Each lane contained a 15 µl sample from the indicated fraction. Fractions 10-14 were pooled for further purification by size-exclusion chromatography.

Panel (B), size-exclusion chromatography of the protein contents of previous anion exchange fractions 10-14. Left, chromatogram of gel filtration. Right, an SDS-PAGE (top) and a western blotting analysis (bottom) of the protein contents in the gel filtration fractions under reduced conditions. Each lane contained a 15 µl sample from the indicated fraction. The N-terminal sequences of the 48, 44, 34 and 32 kDa fragments are displayed and the corresponding bands are marked with arrows. This indicates that the 48 kDa protein was M3QG687R. Fractions 13-15 therefore contained M3QG687R at high purity.

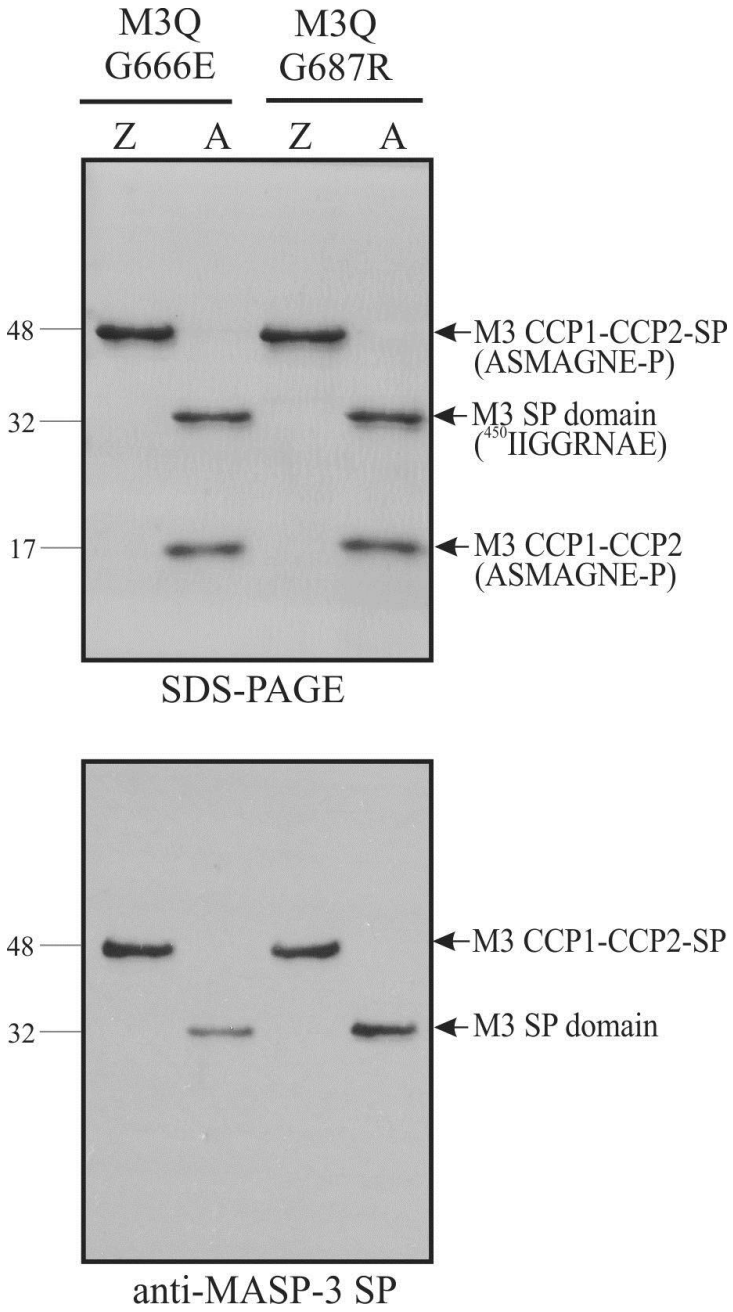


Figure 5.4 Activation of 3MC-associated MASP-3 mutants by using a C1r column
The zymogen form (Z) of M3QG666E or M3QG687R was loaded into a C1r column and the active protein (A) was eluted after incubating the column at 26°C for 18 hr. The reduced samples were analysed by SDS-PAGE (top panel) and western blotting (bottom panel) using an antibody recognizing the unique peptide sequence NPNVTDQIISSGTRT of the MASP-3 SP domain. The protein fragments are indicated with arrows and the N-terminal sequences of these fragments are shown in brackets.

5.3.3 The crystallization of M3QG666E

The crystallization experiment was carried out using the sitting-drop vapour-diffusion method at 4°C using equal volumes of zymogenic M3QG666E (5 mg/ml in 20 mM Tris, 300 mM NaCl, pH 7.4) over a reservoir buffer (120 mM NaCl, 18 % [w/v] PEG 4000, 100 mM imidazole, pH 7.5). Crystals were observed after 1 day and grew to maximal size after 3 weeks (Figure 5.5).

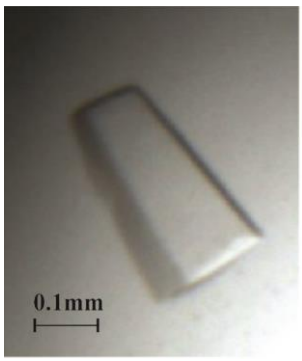


Figure 5.5 A crystal of zymogenic M3QG666E at day 10.

5.3.4 The overall crystal structure of M3QG666E

The structure of M3QG666E zymogen was solved by molecular replacement and refinement to a 2.6Å resolution (Figure 5.6). Final coordinates and structure factors have been deposited with the Protein Data Bank with the accession code 4KKD. The crystal belonged to the orthorhombic space group $P2_12_12$, with unit-cell parameters a = 72.84, b = 292.9, c = 43.59 Å. A table of full data collection and refinement statistics can be found in Table 5.3.

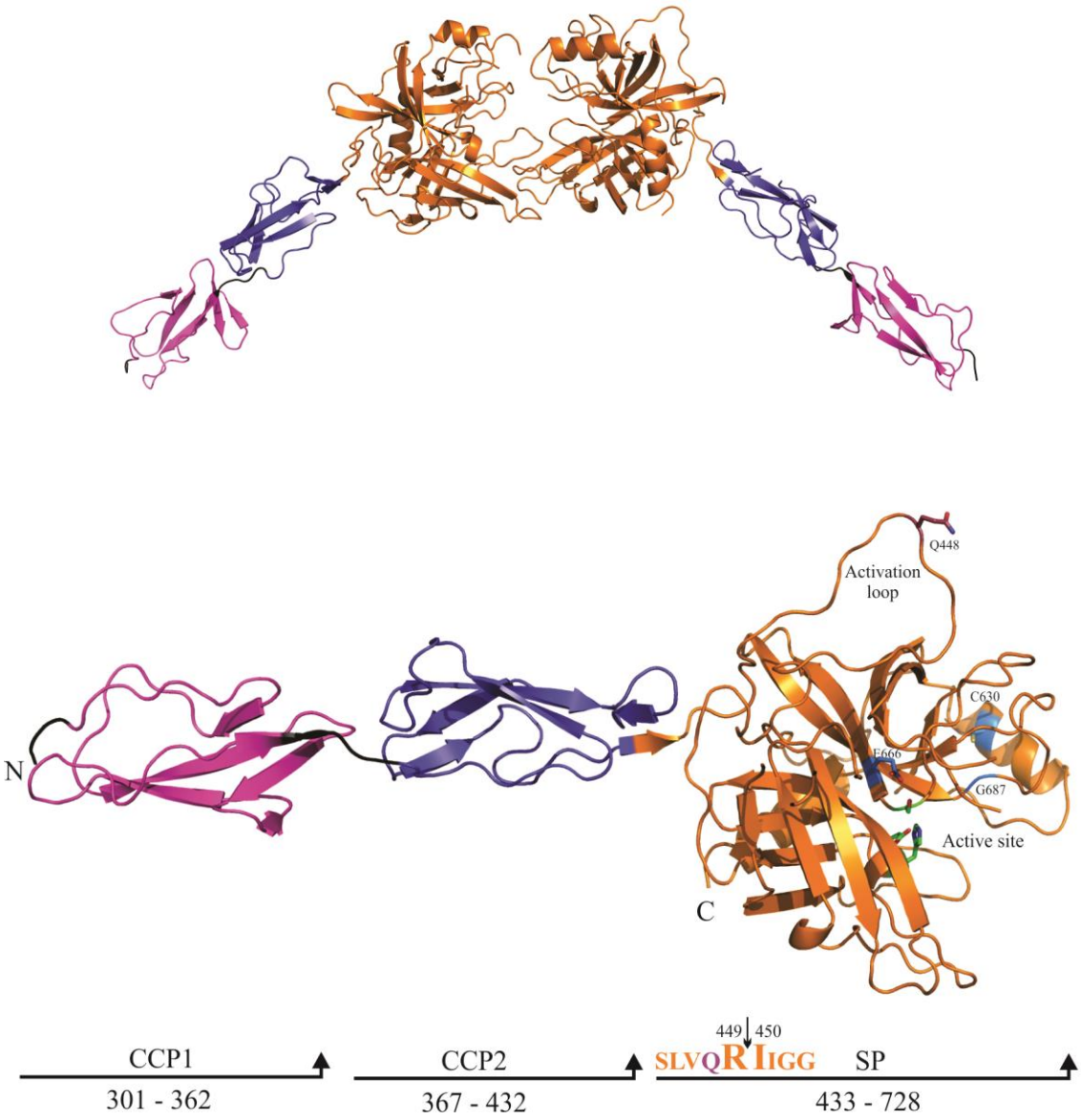


Figure 5.6 Cartoon representations of the M3QG666E structure
Upper panel, the space arrangement of one asymmetrical unit cell of the M3QG666E crystal containing two monomers; Lower panel, the zymogen M3QG666E monomer (PDB code 4KKD) with residues 301-362 in magenta representing the CCP1 domain, residues 367-432 in blue representing the CCP2 domain and residues 433-728 in orange, representing the SP domain. The linker regions are coloured black and the N and the C demote the N- and C-terminus, respectively. The activation loop substitution K448Q is shown in purple in stick format. The active site residues (H^{497}, D^{553}, S^{664}) are shown in green in stick format. The residues involved in point mutations associated with the 3MC syndrome, C^{630} and G^{687}, in addition to the incorporated G666E substitution are shown in blue in stick format (H^{497} in green stick is also the site of a point mutation associated with 3MC syndrome). The amino acid sequence of the activation loop is shown in orange font below the diagram, with the black arrow between R^{449} and I^{450} highlighting the activation point.

Table 5.3 Data collection and refinement statistics for M3QG666E

Data Collection Statistics	
Space group	$P2_12_12$
Cell parameters	
a, b, c (Å)	72.84, 292.9, 43.59
Resolution	48.82 - 2.60 (2.71 - 2.60)
Completeness	99.9% (99.1%)
I/σ (I)	9.6 (2.2)
Rmeas	16.4% (78.6%)
Rsym	15.2% (73.0%)
Rpim	6.1% (29.0%)
Multiplicity	7.2
Observations	214420 (25968)
Unique	29907 (3562)

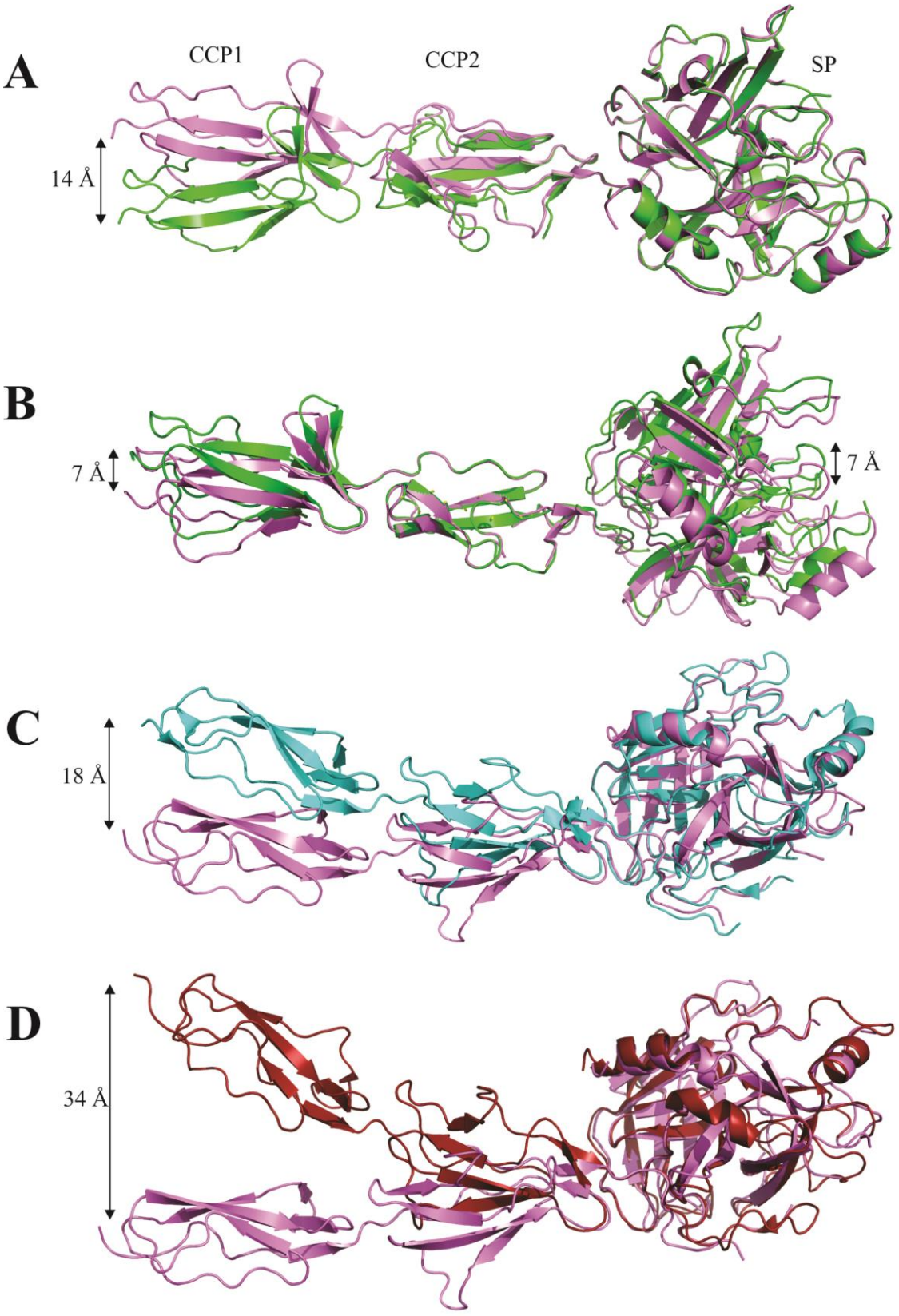
Refinement Statistics	
Resolution (Å)	48.82 – 2.60 (2.74 - 2.60)
Rwork / Rfree	19.40 / 25.98
Number of atoms	
Protein	6197
Imidazole Ion	10
Water	201
B-factors:	
Mainchain	17.5
Sidechain	19.7
Imidazole ion	18.2
Water	15.1
R.M.S deviations:	
Bond lengths (Å)	0.009
Bond angles (^{o})	1.2
Ramachandran favored (%)	92.55
Ramachandran outliers (%)	0.39
Molprobity score	3.29
Molprobity clashscore	36.94

Values in parentheses are for the highest resolution shell (2.74-2.60 Å)

Each asymmetrical unit consists of two monomers (A and B) of M3QG666E in a head to head configuration with interactions between the two SP domains (Figure 5.6). Molecule B is more complete than molecule A and includes the intact activation loop (a.a.440-453 {5-18}) (Throughout this work, the numbering of the SP domain of MASP-3 and other homologous proteins is given together with chymotrypsin numbering, marked in {}, for comparison). The activation loop with the K448Q mutation participates in crystallographic packing interactions, suggesting that this may have aided crystallisation. We were unable to crystallize wild type or activated enzyme. The interactions between the two SP domains in a unit-cell are mainly between loops A (see Figure 5.8 for labelling of the loop structures of all MASP enzymes with reference to chymotrypsin labelling and structural analysis) and the long extended loop E. The unusually long loop D in both chains was not visible in electron density. The buried surface area of the interaction between the SP domains of the asymmetric unit has been calculated to be 276 $Å^2$. The overall fold and topology of the recombinant MASP-3 CCP1-CCP2-SP fragment is the same as those of the other members of the C1r/C1s/MASP family of enzymes, exhibiting two rod-shaped CCP domains preceding the globular SP domain. The conformation of the SP domain of M3QG666E resembles that of the chymotrypsin-like serine proteases.

5.3.5 The flexibility of the hinge regions within the catalytic portion of MASP-3

Superimposition of the SP domain of molecule A onto molecule B of a single unit-cell reveals a significant difference in the positions of the CCP1 and CCP2 domains, with up to 14 Å of variance observed in Carbon α positions in CCP1 (Figure 5.7A). The interactions between the various domains remain the same, but differences in the


A
CCP1
CCP2
SP
14 Å
B
7 Å
7 Å
C
18 Å
D
34 Å

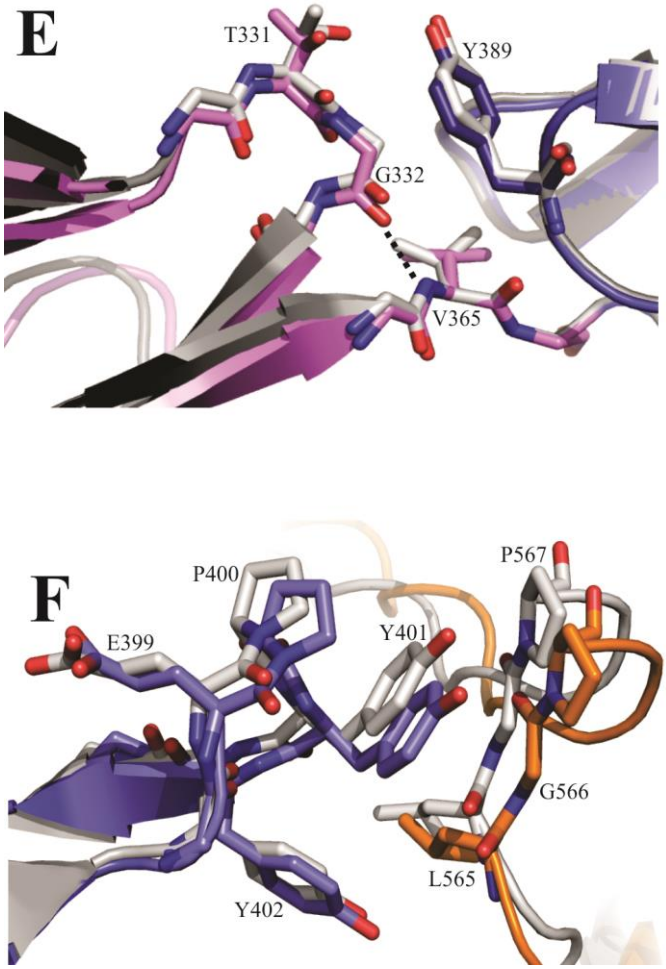


Figure 5.7 Domain flexibility of zymogenic M3QG666E
Cartoon representations of (A), SP superimposition of the M3QG666E chain A in green and chain B in pink; (B), CCP2 domain superimposition of M3QG666E; (C), SP domain superimposition of M3QG666E chain B in pink on zymogenic MASP-1, coloured cyan; and (D), SP superimposition of M3QG666E chain B on zymogenic MASP-2, coloured firebrick. The distances between the compared domains were marked with double arrows. (E), Stick representation of key residues involved in the CCP1-CCP2 domain movement seen between chains A and B of MASP-3. CCP1 of chain A was coloured magenta and CCP2 in blue, and the superimposed chain B was coloured grey. The dashed line indicates polar interactions. (F), Stick representation of key residues involved in the CCP2-SP domain movement seen between chains A and B of the MASP-3. CCP2 of chain A was coloured blue and SP in orange, and the superimposed chain B was coloured grey.

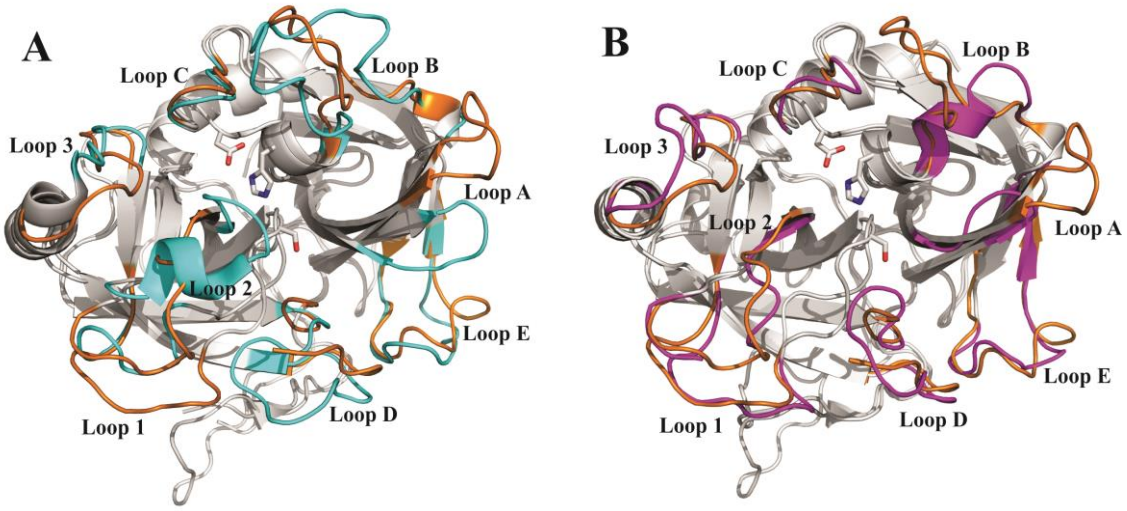
relative positions of residues in CCP domains account for the observed inter-domain flexion. The zMASP-3 SP chains of the two molecules superimpose well, with an RMSD value of 0.59 Å. Minor differences in side chain rotamers in addition to minor loop movements are observed, presumably due to differences in crystal packing interactions. Superimposition of the CCP2 domains of the two molecules reveals a disposition of the CCP1 and SP domains, with up to 7 Å of variance observed in Cα positions in CCP1 and SP domains, respectively (Figure 5.7B). Structural comparisons of zymogenic M3QG666E with zMASP-1 and zMASP2 further highlight the CCP2-SP inter-domain flexibility. Superimposition of the SP domain of zymogenic M3QG666E onto zMASP-1 (PDB ID: 4IGD) and zMASP-2 (PDB ID: 1ZJK) reveals a shift of up to 18 and 34 Å, respectively, in the relative position of the CCP1 domains between the enzymes (Figure 5.7C, D). The difference at the CCP2-SP junction of V^{433} and E^{435} in MASP-2 and MASP-3, respectively, sees the loss of a polar interaction in MASP-2, which contributes significantly to the different domain positioning in the two enzymes.

Molecular differences in the position of CCP1 relative to CCP2 are observed upon superimposition of CCP1 domains of the two M3QG666E molecules in a single unit-cell (Figure 5.7E). In CCP1, a backbone hydrogen bond between G^{332}-Oxygen atom and V^{365}-Nitrgen atom is observed in molecule A (2.78Å), but is lost in molecule B (3.34Å). Concurrently, the peptide backbone of T^{331} is flipped (chi angle A: 55.25 vs B: 175.89). Movement of G^{332} and T^{331} away from CCP2 serves to reduce steric hindrance to domain movement increasing the distance to Y^{389} in CCP2. A fine positional comparison of residues at the CCP2-SP hinge regions of the two monomer molecules reveals a different backbone position of residues 398-409 in CCP2 domain, causing the different domain juxtaposition. The greatest shift in this region is observed in the P^{400}-Carbon α, with a movement of 0.98 Å (Figure 5.7F). The positional variance

of the residues 398-409 in CCP2 domain gives rise to a greater displacement of residues in the SP domain at the CCP2-SP hinge region, for example, residues 565-567, with a movement of more than 2 Å generally.

5.3.6 Comparisons of the loop structures around the active site of zymogenic MASP enzymes

Comparisons of loops surrounding the active site of MASP enzymes are shown in Figure 5.8. Loop A of MASP-3 is similar in length with that of MASP-1 and is longer than that of MASP-2. It is more loosely packed and located farther away from the active site than in the cases of zMASP-1 and zMASP-2. Both ends of loop A of MASP-3 contain a negatively charged D residue, which is absent in MASP-1 and MASP-2. Loop B of MASP-3 is 8 amino acids shorter than that of MASP-1 and is 3 residues longer than that of MASP-2. The sequence of loop B of MASP-3 contains less acidic residues than that of MASP-1 and MASP-2. The conformation of loop B of zymogenic MASP-3 lacks the α-helical structure present in this region of MASP-2, and is similar to that of MASP-1. For both MASP-3 and MASP-1, the N-terminal side of loop B consists of a patch of four amino acids, similar between the two enzymes (LRSQ for MASP-3, and LHQS for MASP-1). These amino acids are close to the active site of their proteases. Both MASP-3 and MASP-1 also contain a patch of hydrophilic residues at the C-terminal side of loop B. In MASP-3, the sidechain of R^{504} of loop B is exposed to solvent and bends towards the active site, different to MASP-1, taken place by the sidechain of an acidic residue, D^{497}. Loop C is well preserved for both sequence and structure among all MASP enzymes, due to its importance in supporting the structure and function of the nearby active site $D^{\{102\}}$ residue. Comparison of the loop structures




```
                          Loop A                                  Loop B                                  Loop E
HUMAN MASP3   463 WQALIVVEDTSRVPNDKWFGSGA 485//497 HVLRSQRRDTTVI-----PVSKEHVTVYLGLH-DVRDKSGAVNSSAA 537
HUMAN MASP1   462 WIAMLSHL------NGQPFCGGS 478//490 HCLHQSLDPEDPTLRDSDLLSPSDFKIILGKHWRLRSDENEQHLGVK 536
HUMAN MASP2   458 WQVLILGG---------TTAAGA 471//483 HAVYEQKHDASALDIRMGTLK--RLSPHYTQAWS-----------E 515
CHYMOTRYPSIN 29  WQVSLQDK------TGFHFCGGS 45 // 57 HCGVTTSDVVVAGEFD----------------QGSSSEKIQKLKIA 86

                          Loop C                                  Loop D
HUMAN MASP3   538 RVVLHPDF-NI----Q--NYNHDIALV 557//588 GLVAGWGISNPNVTVDEIISSGTRTLSDVLQYVKLPVVPHAEC 630
HUMAN MASP1   537 HTTLHPQY-DP----N--TFENDVALV 556//584 VIVSGWGKQFLQ-----------RFPETLMEIEIPIVDHSTC 614
HUMAN MASP2   516 AVFIHEGYTHD----A--GFDNDIALI 536//568 GTASGWGLTQRG-----------FLARNLMYVDIPIVDHQKC 598
CHYMOTRYPSIN 87  KVFKNSKYNS-------LTINNDITLL 106//136 CVTTGWGLTRYT----------NANTPDRLQQASLPLLSNTNC 168

                          Loop 3          Loop 1                                  Loop 2
HUMAN MASP3   631 KTSYESRS----GNYSVTENMFCAGYYEGGKDTCLGDSGGAFVIFDDLS-QRWVVQGLVSWGGPEECGSKQVYGVYTKVS 705
HUMAN MASP1   615 QKAYAPLK------KKVTRDMICAGEKEGGKDACAGDSGGPMVTLNRER-GQWYLVGIVSWGD--DCGKKDRYGVYSYIH 685
HUMAN MASP2   599 TAAYEKPP---YPRGSVTANMLCAGLESGGKDSCRGDSGGALVFLDSET-ERWFVGGIVSWGSM-NCGEAGQYGVYTKVI 673
CHYMOTRYPSIN 169 KKYWGTK---------IKDAMICAGAS--GVSSCMGDSGGPLVCKKN---GAWTLVGIVSWGSS--TCSTSTPGVYARVT 232
```

Figure 5.8 Comparisons of loops surrounding the active site of MASP enzymes
Upper panel, (A), superimposition of the SP domains of zymogenic MASP-1 (1IGD.pdb) and MASP-3. The loops surrounding the active site of MASP-3 are coloured orange and those of MASP-1, coloured cyan. (B), superimposition of the SP domains of zymogenic MASP-2 (1ZJK.pdb) and MASP-3. The loop structures of MASP-2 are coloured magenta. The active site triad of MASP-3 (H^{497}, D^{553} and S^{664}) was presented as stick in both (A) and (B).
Bottom panel, a sequence alignment of MASP enzymes on the loops surrounding the active site. The loops are labelled according to both chymotrypsin-derived nomenclature and structural analysis.

of MASP enzymes reveals that MASP-3 possesses a long extended loop D with a couple of negatively charged residues. The definite structure of the centre of loop D in M3QG666E is unclear in the data, suggesting that this loop region of MASP-3 has strong flexibility and is prone to intermolecular interactions. Loop E is similar in length and conformation among all MASP enzymes. However, the sidechain of V^{525} in loop E of MASP-3 faces toward the active site and participates in forming the substrate binding groove of MASP-3 for residues prime side to the scissile bond. Compared to this, in the same position of MASP-1 and MASP-2, it is replaced by the sidechain of a basic residue, R^{524} in MASP-1 and R^{504} in MASP-2, respectively. Loop 1 of all MASP enzymes contains mainly polar residues. In MASP-3 zymogen, charged interactions between E^{654} and K^{695} and between K^{657} and E^{691} can be observed. These interactions appear to stabilize the loops 1 and 2 in MASP-3 zymogen. Similar interactions can be found in zMASP-1, not in zMASP-2. Loop 2 of both MASP-3 and MASP-1 consists of a motif of paired acidic residues. However, this motif in MASP-1 forms a small loop close to the active triad, whilst that in MASP-3 stays farther away from the triad. Further, loop 2 of the MASP-3 zymogen does not exhibit the α-helical structure of MASP-1. Loop 3 of zymogenic MASP-3 is brought close to loop 2 and the active site by a charged interaction between R^{637} and E^{690}. As introduced previously in Chapter 1, loop 3 of MASP-3 contains a glycosylation site unique among MASP enzymes.

5.3.7 The structural perturbation in the active site of M3QG666E

The catalytic motif GDSer$^{\{195\}}$GG is highly conserved across the chymotrypsin family. Our data reveal that mutation of the $G^{666\{197\}}$ in MASP-3 has marked structural consequences. Structural comparison of zymogenic M3QG666E with both zymogens

and active forms of homologous proteins reveals that the introduction of a negatively charged residue with a relatively large sidechain in place of the glycine residue has dramatic effects on the active site structure of MASP-3. The 3MC syndrome G666E mutation causes expulsion of the sidechain of residue $D^{663\{194\}}$, significantly impacting on the position of the adjacent active site $S^{664\{195\}}$, drawing the functional hydroxyl group away from the catalytic $H^{497\{57\}}$ to an inactive position and thereby eliminating any potential charge relay and subsequent proteolytic activity (Figure 5.9A). Specifically, $E^{666\{197\}}$ makes H-bonds with atoms $D^{663\{194\}}$-Nitrogen, $S^{664\{195\}}$-Nitrogen and $S^{664\{195\}}$-Oxygen γ, forcing the active site loop into an α-turn conformation. This structure is distinctly different to the β-turn conformation observed in the active MASP-3 SP domain (Figure 5.9B) (the structure of which was published during the time this thesis was being written) and in both zymogen and active forms of the other members of the C1r/C1s/MASP family of enzymes, displayed representatively by MASP-1 (Figure5.9C, D). A main chain H-bond between $D^{\{194\}}$-Oxygen atom and $G^{\{197\}}$-Nitrogen atom stabilizes the β-turn conformation. Further, rather than interacting with the continuation of the loop D, $D^{663\{194\}}$ residue is in a novel solvent-exposed conformation.

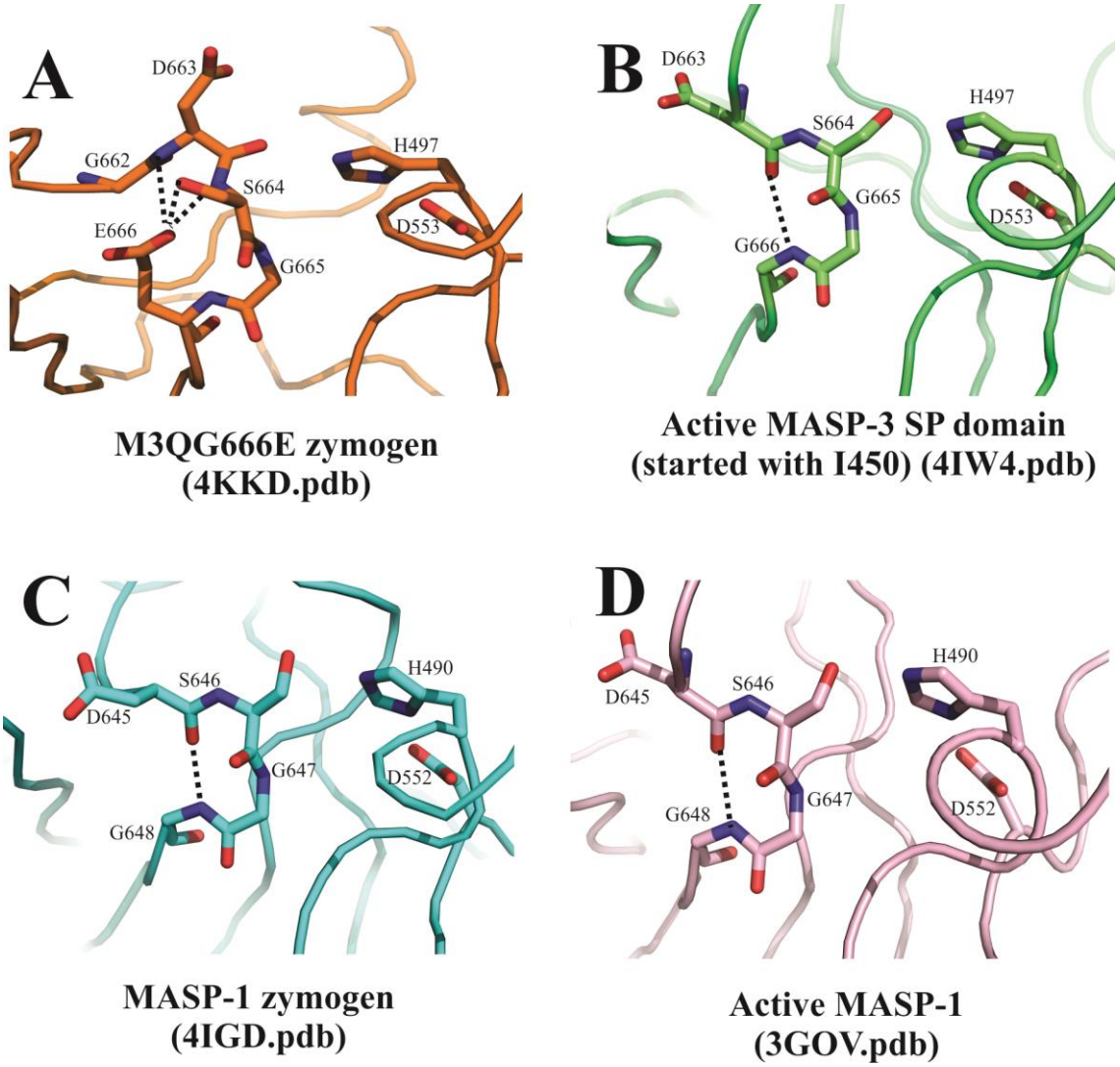


Figure 5.9 The structural perturbation at the active site of M3QG666E
(A), the α-turn structure of the GDS$^{664\{195\}}$GE motif in the zymogenic M3G666E active site, coloured orange and shown in stick representation. Dashed lines represent polar interactions. The α-turn structure is stabilized by interactions among the mutated E^{666} residue and residues D^{663} and S^{664}. The side chain of the active site residue S^{664} is withdrawn from its active state position. Instead of the α-turn structure, a β-turn structure is seen in (B), active MASP-3 SP domain (4IW4.pdb) in green, (C), zMASP-1 structure (4IGD.pdb) coloured cyan, and (D), active MASP-1 (3GOV.pdb), coloured pink.

5.4 DISCUSSION

The recent discovery of the association between 3MC syndrome and mutations of MASP-3 as well as one of its lectin ligand recognition partners, CL-11, has suggested a possible role of MASP-3 in development (Sirmaci *et al*., 2010, Rooryck *et al*., 2011). Four point mutations (C630R, G666E, G687R and H497Y) were found to be specific to the SP domain of MASP-3. Of these, the mutants G666E and G687R are most likely to result in intact proteins. However, the effects of these changes in MASP-3 associated with the syndrome were less certain. By using our M3Q form of human MASP-3, which contains a single Q residue replacing the K residue prior to the R-I activation bond, we similarly produced the two 3MC associated MASP-3 mutants (M3QG666E and M3QG687R), which were efficiently activated by the C1r protease. We found that the C1r-activated enzymes of both M3QG666E and M3QG687R failed to cleave the best substrate sequences RIF/RIY identified for cleavage by MASP-3. Indeed, both MASP-3 mutants did not cleave any of the AMC or Abz synthetic substrates available in this project, including those cleaved by M3Q. Our data was the first line of evidence that the G666E and G687R mutants of MASP-3 were indeed inactive enzymes, supporting the idea that the 3MC syndrome arises from a lack of MASP-3 activity.

We then solved the 2.6 Å X-ray crystal structure of the zymogenic G666E mutant enzyme. Each asymmetrical unit consists of two monomers of M3QG666E in a head to head configuration with interaction between the two SP domains. However, there is no evidence that MASP-3 forms a dimer *in vitro* or *in vivo* via interaction between SP domains. The activation loop (a.a.440-453) of monomer B in a single unit-cell was visible in the 3D structure. This is unusual since the flexible spatial movement of this loop is usually not beneficial to crystallization, and thus makes it difficult to be viewed in the structure. In this case, the activation loop of monomer B interacts with molecules

of other unit-cells, suggesting that this may have aided crystallization. The intermolecular interaction also restricts the movement of the loop, making it possible for its structure to be solved.

Minor differences in the orientations of the CCP1, CCP2 and SP domains were found between the two MASP-3 monomers in the crystallographic asymmetric unit, highlighting the limited flexibility of the CCP domain interactions with minimal hydrogen bonding observed between the domains. Superimposition of the SP domain of zymogenic M3QG666E onto zMASP-1 and zMASP-2 further underlines the CCP2-SP inter-domain flexibility of MASP-3, indicated by a displacement of Carbon α positions in CCP1 at up to 18 and 34 Å, respectively. In the CCP2-SP hinge region of zymogen MASP-1, an inter-domain salt bridge formed by N^{699} and K^{403} restricts the inter-domain flexibility of MASP-1. This salt bridge is not seen in MASP-3 and MASP-2, accounting for a higher flexibility in the CCP2-SP interface of the two MASP enzymes. It was postulated that the CCP2-SP interface is important for the MASP enzyme to adapt the fully mature form of active protease. The relatively flexible CCP2-SP hinge region of MASP-3 may reflect an enzymatic efficacy different to that of MASP-1 and MASP-2.

The conformation and length of the loops surrounding the active site are critical determinants for the substrate specificity and activity of serine proteases. Structural analysis of the zymogenic MASP enzymes shed light on some interesting features of zymogenic MASP-3. In the case of MASP-2, loops 1, 2 and D are believed to be involved in adaptation of structure of the fully functioning serine protease upon activation of the zymogen. As a result of the transition, the displacement of loop D toward loops 1 and 2 leads to the opening of the hidden substrate-binding groove, thus allowing the full proteolytic activity of MASP-2. Compared to loop D of MASP-2, the strikingly long extended loop D of MASP-3 may indicate that the substrate-binding

groove of MASP-3 is heavily buried by this loop, which may also affect the proteolytic efficacy of the active enzyme. Loops A, B, C and 3 are found to be involved in shaping the specificity of many serine proteases, including MASP-1, MASP-2, and thrombin. Loop B of zymogenic MASP-3 is structurally similar to that of MASP-1 zymogen, and also shares the feature of a hydrophilic patch at the C-terminal side of the loop similar to that of MASP-1. This suggests that active MASP-3, like MASP-1, may also prefer a small hydrophobic P2 residue, since loop B in MASP-1 is an important determinant for the P2 specificity of the enzyme. The sequence of loop 3 of MASP-3 is highly variable to that of the other MASP enzymes. Indeed, the presence of an N-glycosylation site in loop 3 of MASP-3 is one of the unique features among the loop structures of the protease. A charged interaction between R^{637} and E^{690} is found to bring the loop close to the active site, indicating that the loop may affect the proteolytic property of MASP-3. The exact impact of the glycosylation at this loop on the specificity and activity of MASP-3 is yet to be identified. In addition, loop E is located on the prime side of the enzymatic subsites, and very likely affects the P1'-P2' specificity of MASP enzymes. The sidechain of a hydrophobic residue V^{525} in loop E of MASP-3 is found to protrude towards the active site. This feature agrees with our previous data on the specificity of MASP-3, which highly prefers a hydrophobic residue at the P1' and P2' positions of substrates.

The structure of the M3QG666E protein reveals the mechanism of inactivation of this 3MC syndrome-associated mutant. The single amino acid mutation in MASP-3 causes a dramatic perturbation in the structure of the active site around the $GDS^{\{195\}}GG$ motif, forcing it to form an α-turn conformation, different to the critical β-turn structure observed in active MASP-3 SP domain and in both zymogen and active forms of C1r, C1s, MASP-1 and MASP-2. The mutated $E^{666\{197\}}$ residue not only makes the α-turn a

rigid conformation by forming multiple bonds with surrounding residues, but also forms a hydrogen bond with the side chain of the active site $S^{664\{195\}}$ residue, withdrawing it from the active triad configuration. Therefore, the zymogenic MASP-3G666E mutant is most likely to be unable to adopt the conformation required in order for the enzyme to be in an active state.

Our structure of the active site of MASP-3 also sheds light on the possible mechanism of inactivation of the other 3MC syndrome-associated MASP-3 mutants, G687R and C630R (Figure 5.10). It was proposed by Sirmaci and her colleagues (2010) that the G687R substitution results in a salt bridge formed by R^{687} and D^{658}, diminishing the catalytic activity of MASP-3. The hypothesis was based on a structural model of MASP-3 developed on the basis of the structure of MASP-1, according to which, G^{687} is expected to be in close proximity to the residue D^{658}. However, our data reveals that even though the sequence of loop 2 of MASP-3 shares some similarity with that of MASP-1, the structure of loop 2 of MASP-3 is strikingly different to that of MASP-1. In our structure of MASP-3, D^{658} is located on the top of the loosely packed loop 1, with the sidechain facing the solvent, and is 14.2 Å away from G^{687}. G^{687} stands at the start of loop 2 and is in close proximity to the active site and D^{663} (7.2 Å) of the GDSGG active site loop, the conformation of which is shown to be critical to the mechanism of inactivation of the G666E mutant. The G687R mutation therefore could be extremely close to D^{663}, since the sidechain of an R residue is 5.5-6.9 Å in dimension. It is therefore very likely that R^{687} of the G687R mutant forms a strong charge interaction with D^{663}, driving the adjacent active site S^{664} away from its active position. Our structure of MASP-3 also shows that C^{630} forms a disulfide bond with C^{649}, maintaining the structure of loop 3, which we believe to be important in determining the specificity and protease efficacy of MASP-3. Therefore the mutation at C^{630} may interfere with the

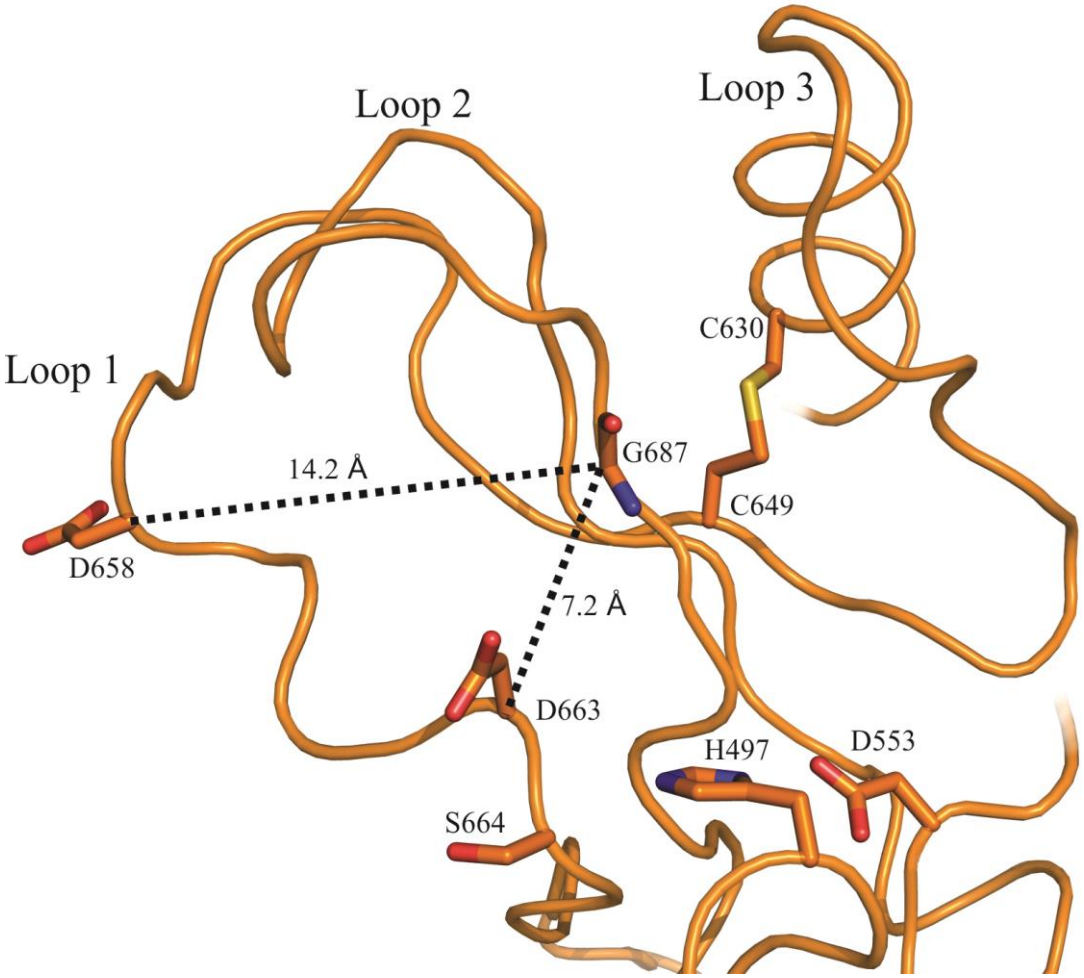


Figure 5.10 The 3MC-associated mutation sites besides G^{666}
The loop regions of the SP domain of MASP-3 are presented as a cartoon. The catalytic triad (H^{497}, D^{553} and S^{664}) and the 3MC-associated mutation sites G^{687} and C^{630} (and H^{497}) are displayed as sticks. C^{630} forms a disulfide bond with C^{649}, maintaining the structure of loop 3. G^{687} is in close proximity to D^{663} (7.2 Å), not D^{658} (14.2 Å), indicated by dash lines.

proteolytic properties of MASP-3. It is also highly possible that the C630R mutation of 3MC syndrome results in misfolded MASP-3 enzyme due to the breaking of the loop structure. In either case, the C630R mutation is also likely to be associated with a lack of MASP-3 activity.

Chapter Six

General Discussion and Future Directions

The complement system component, MASP-3, has remained an enigmatic enzyme since its isolation from human plasma in 2001. The isolated MASP-3 appeared to be the activated form of enzyme (Dahl *et al*., 2001). However, without information about a convincing scheme of activation of the enzyme and its substrate specificity, further studies of the function of MASP-3 were heavily hampered. Recent genetic evidence suggests a correlation between MASP-3 mutations and an embryonic developmental disorder, the 3MC syndrome, underlining the possible role of MASP-3 in development, outside the immune system (Sirmaci *et al*., 2010, Rooryck *et al*., 2011). However, further investigation of the molecular basis of the disease-associated MASP-3 mutants was also hindered by the lack of known structures of the enzyme. At the outset of this study, there were major questions about the structure and function of MASP-3 that were required to be answered, including:

(1) How is the enzyme activated?

(2) Is the activated enzyme active against peptide substrates and if so, what is its substrate specificity?

(3) What is the structure of the enzyme?

6.1 THE ACTIVATION PARTNER OF MASP-3

Our study provides the first demonstration that MASP-2, not MASP-1, activates MASP-3, illustrating the importance of the sequence of the activation loop of MASP-3 for the interaction between this enzyme and its possible activating partner protease. However, activation of the M3R454A mutant by both MASP-1 and MASP-2 indicates that both activating enzymes are able to cleave at both the activation bond and the R^{454}-N^{455} peptide bond in wild type MASP-3. Detailed information about the kinetics of these interactions needs to be obtained. Furthermore, our study of the activation of MASP-3 by MASP-2 was thus far conducted using recombinant proteins under *in vitro*

experimental conditions. It is therefore important to learn whether wild type proteins under physiological conditions can yield similar results.

Here, the specific *in vitro* activation of MASP-3 by the human PC enzymes, furin and PACE4, was shown for the first time. However, the slow rates of cleavage indicate that the conditions of activation may need to be optimized, or that such activation may not be physiologically relevant. Although our observation of cleavage of IGFBP-5 by M3Q suggests that the non-glycosylated nature of our recombinant MASP-3 did not affect the proteolytic activity of the enzyme, the lack of glycosylation may affect the interaction of MASP-3 with furin and PACE4. Therefore, a study of the activation of plasma-derived or eukaryotic cell-derived MASP-3 by PC enzymes and/or the activation of MASP-3 in PC enzyme-knock out cells may provide more information addressing the problem. Further, whether MASP-3 plays a role in the PC enzyme-associated developmental disorders is also currently unknown.

In order to further study MASP-3, an alternative approach to generate a MASP-3 mutant with a homogeneous functional serine protease domain resembling that of the wild type enzyme was necessary. Here, two MASP-3 mutant proteins, M3EEKQ and M3Q, were developed, both of which contain sequences N-terminal to the activation bond similar to that of C1s, which is specifically activated by C1r. Both M3EEKQ and M3Q were activated effectively and uniformly by recombinant C1r. The active forms of MASP-3 mutants consist of a serine protease domain essentially identical to that of wild type MASP-3, which was confirmed by western blotting and N-terminal sequencing. The production and activation of these MASP-3 mutants not only served as a convincing scheme for activation of MASP-3 for further study of this enzyme, but also contributed to the identification of the molecular determinants for interaction between C1r and its substrates.

6.2 THE PHYSIOLOGICAL SUBSTRATE AND BIOLOGICAL FUNCTION OF MASP-3

Besides the activation scheme of MASP-3, another barrier that hindered the study of MASP-3 was the lack of a good MASP-3 specific substrate for characterizing the activity of the enzyme. By using available commercial peptidyl substrate libraries, we determined that MASP-3 most prefers cleavage of the sequences GRIF/GRIY with a P1-R/K specificity. The fluorescence-quenched tripeptide substrate, RIF, was cleaved 20-fold more effectively by MASP-3 than the AMC substrate, VPR, which was previously claimed to be among the best substrates for MASP-3. Our finding generates a useful tool for further study of MASP-3 and the overall study provides useful insights into the specificity of the enzyme which may aid studies into its function. Nevertheless, our kinetic data showed that the M3Q mutant in this project cleaved synthetic AMC and Abz peptide substrates only at low rates, reflecting that the forms of AMC and Abz substrates may not be the most preferable for cleavage by M3Q, or that the proteolytic activity of M3Q was affected by its lack of glycosylation. To address the latter issue, further studies of fully glycosylated MASP-3, plasma-derived MASP-3 and/or MASP-3 in complex with physiological recognition molecules will be necessary.

The biological role of MASP-3 has long been one of the many enigmas of this enzyme. Genetic evidence speculates that MASP-3 arises from partial replication of the gene encoding the SP domain of MASP-1 (Dahl *et al*., 2001). However, the conservation of MASP-3 among all vertebrate species examined (Fujita *et al*., 2002), including those in which MASP-1 is not present (Matsushita *et al*., 2000), indicates that MASP-3 may play a role different to MASP-1 in vertebrate species. Unlike MASP-1 and MASP-2, MASP-3 does not cleave complement proteins C2, and it is not inhibited by C1-inhibitor (Zundel *et al*., 2004), indicating that MASP-3 may not act as an initiating enzyme of the complement activation as MASP-1 and MASP-2 do. Although MASP-1 and MASP-3 enzymes share identical domain structures responsible for binding to their recognition molecules and for dimerization and both enzymes thus can form complexes with MBL, ficolins and CL-11 *in vitro* (Degn, *et al*., 2012), the two enzymes are controlled by an unknown mechanism to form complexes with different oligomers of

MBL in plasma (Dahl *et al*., 2001). Thus far there is no evidence indicating that MASP-3 forms heterodimers with MASP-1.

It was previously observed that human plasma MASP-3 is co-localized with MASP-2 in complex with MBL-II and that recombinant human MASP-3 inhibits the cleavage of C2 and C4 by recombinant MASP-2 (Dahl *et al*., 2001, Møller-Kristensen *et al*., 2007, Degn, *et al*., 2010). Therefore, an early suggestion of the biological role of MASP-3 was that it competed with MASP-2 for the binding sites in MBL-II and hence down-regulated MASP-2-mediated complement activation (Dahl *et al*., 2001). Whether this represents the real biological function of MASP-3 is questionable, since there are many highly effective regulators, like C1-inhibitor, regulating complement activation. It is hardly convincing that a functional serine protease like MASP-3 only serves as an evolutionary remnant that does not exert any proteolytic activity. The role of MASP-3 in the immune response is therefore still unclear. Most recently, it was revealed that MASP-2 can form complexes with either MASP-1 or MASP-3 in complex with MBL or ficolins *in vivo* (Degn *et al*., 2013). Whilst the complexes of MASP-1, MASP-2 and MBL/ficolins were shown to play a role in activating complement (Degn *et al*., 2013), similar complexes with MASP-3, instead of MASP-1, are unlikely to be relevant for complement activation and might be involved in some other function.

It was not until recently that MASP-3-specific mutations were found to associate with the developmental disorder, the 3MC syndrome, pointing to a possible role of MASP-3 outside the immune system (Sirmaci *et al*., 2010, Rooryck *et al*., 2011). The four point mutations affect residues in the SP domain of MASP-3, including the active triad residue, H^{497}, and C^{630}, which forms a disulfide bond that is critical for the structure of MASP-3. Mutations at these two amino acids can be predicted to detrimental to the function of MASP-3 enzyme. Further, a nonsense mutation in the *COLEC11* gene

which encodes CL-11, a recognition molecule that forms complex with MASP-3, is also found to associate with the 3MC syndrome in humans and 3MC syndrome characteristics in zebrafish (Rooryck *et al*., 2011). All of these findings suggest that the disorder is associated with dysfunction of MASP-3. Our study of the 3MC syndrome-associated MASP-3 G666E and G687R mutants revealed that the active forms of these two mutants fail to cleave any of the substrates tested, including the optimal substrate RIF previously identified. Our data is therefore the first piece of evidence that conclusively shows that the 3MC syndrome is related to a lack of MASP-3 activity.

A congenital disorder, the 3MC syndrome affects embryonic development, leading to abnormalities in facial structures, brain functions, hearing and limb bones. Characteristics of the syndrome in humans include abnormal distances between facial organs, small or droopy eye lids, flat nasal bridge, and cleft lip and palate (Leal *et al*., 2008), indicating that the disorder most likely occurs during formation of the body axis in craniofacial development. This is a complicated developmental process which is finely controlled by quantitative migration of specific cell types to their correct location, the substances secreted by local and migrated cells and subsequent proliferation of the migrated cells. In zebrafish, craniofacial development is mainly controlled by the cranial neural crest cells (NCCs), a cell lineage that gives rise to brain cells, craniofacial cartilages and fibroblasts affecting the bone development of the mid-ear and jaw. The migration of NCCs in zebrafish during craniofacial development is found to be guided by the chemoattractant CL-11. It was observed that zebrafish defective in CL-11 displayed similar 3 MC syndrome-associated craniofacial abnormalities as those in defective forms of the CL-11-binding partner, MASP-1/3 (Rooryck *et al*., 2011). The correlation of the lectin pathway components, MASP-3 and CL-11, and the 3MC syndrome therefore points to a new direction for the identification

of the physiological substrates of MASP-3. It is possible that MASP-3 in complex with CL-11 may cleave surface structures of NCCs, affecting their migration and/or proliferation, and thus leads to the onset of 3 MC syndrome. Therefore, future approaches for the identification of the physiological substrate(s) of MASP-3 could involve isolation of NCCs or the NCC-derived cells, such as craniofacial fibroblasts from zebrafish during embryonic development, the establishment of the cell-surface membrane proteome of these cells and the use of subsequent methods, such as ProtoMap, for detecting changes to the proteome after treatment of MASP-3 in complex with CL-11.

6.3 THE STRUCTURAL BIOLOGY OF MASP-3

To further investigate the molecular basis of the inactivation of the 3MC syndrome-associated MASP-3 mutants, we solved the crystal structure of the zymogen of M3QG666E, which is the first published structure displaying the activation loop and the active site regions of MASP-3. Our structure of MASP-3 G666E mutant reveals a major perturbation at the active site region, resulting in a zymogen that is most unlikely to be able to adopt the active state conformation of MASP-3. Our data also provides a more precise hypothesis on the inactivation of the MASP-3 G687R mutant associated with the 3MC syndrome. For further validation of our results, crystals structures of active G666E mutant and of both zymogen and active forms of wild type MASP-3 and G687R mutant are needed.

During the time this thesis was written, the structure of an active form of a truncated MASP-3 SP domain that starts with I^{450}, in complex with a small protein inhibitor,

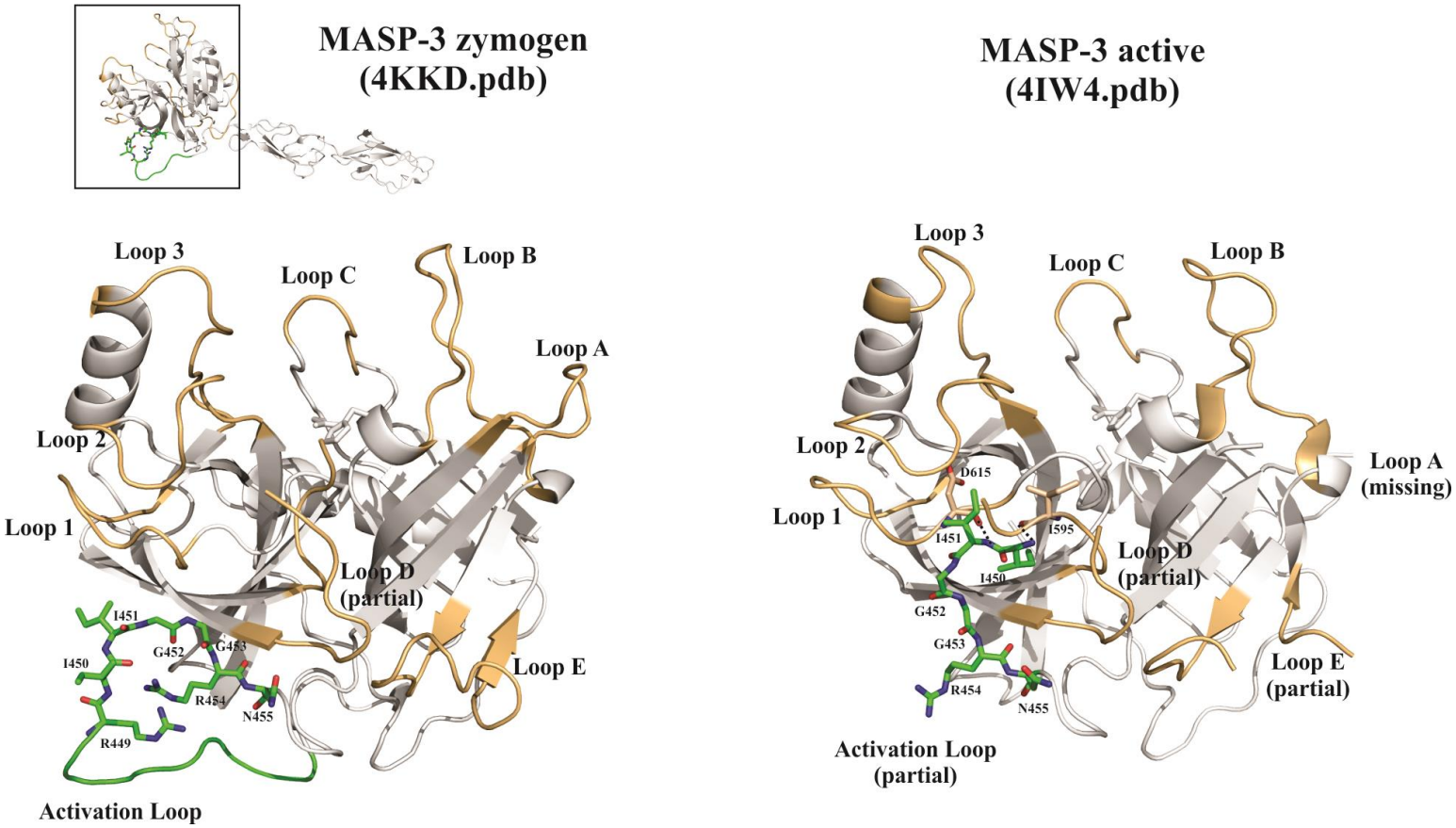


Figure 6.1 Structural alteration of the activation loop region of MASP-3 upon activation of enzyme

The structure of the activation loop is coloured green and the sequence following the activation R^{449}-I^{450} bond in the activation loop is shown as sticks. The loop structures surrounding the active site of MASP-3 are coloured light orange.

The structure of zymogenic MASP-3 (left) reveals an activation loop highly exposed to solvent. Upon activation (right), the cleaved activation loop starting with the I^{450} residue inserts into a pocket structure formed between loop D and loop 1. The insertion is facilitated by main chain interactions between I^{450} and I^{595} of loop D, and between I^{451} and D^{615} of loop 1, marked with dashed lines.

ecotin, was published (Gaboriaud *et al*., 2013). This structure, together with our structure of zymogenic MASP-3, provides precious information on the conformational transition for activation of MASP-3 (Figure 6.1). The activation loop of MASP-3 zymogen containing the R^{449}-I^{450} activation bond is highly exposed to the solvent. Upon activation by cleavage at the activation bond, the cleaved activation loop starting with the I^{450} residue undergoes structural alteration by inserting the I^{450} and I^{451} residues into a space between loop D and loop 1. The insertion is facilitated by main chain interactions between I^{450} residue and I^{595} residue of loop D, and between I^{451} residue and D^{615} residue of loop 1. A similar insertion of the newly liberated N-terminal isoleucine residue upon activating cleavage is critical for the maturation of active serine proteases, as previously discussed in Chapter 1.

Another puzzle of the structure and function of MASP-3 is that this enzyme appears to be weak in proteolytic activity towards synthetic peptide substrates. Such low activity of MASP-3 hinders further studies of the enzyme. A few modes of structural modulation of enzymatic activity have been observed in members of the C1r/C1s/MASP family of enzymes. The charge interactions between the residues D^{640} and R^{677} and between the residues D^{670} and K^{623} in the MASP-1 active site are predicted to restrict the proteolytic activity of MASP-1. Hence a conformational rearrangement probably occurs upon binding of preferable substrates to the substrate-binding groove of MASP-1, releasing the constraint mechanism of the enzyme (Dobó *et al*., 2009). Similar charge interactions are not observed in the structure of the active MASP-3 SP domain. However, it is still too early to draw a conclusion, since this structure was in complex with an inhibitor. Any structural alteration of MASP-3 induced by the binding

of a substrate or inhibitor is yet to be determined, preferably by the solution of an active form of the enzyme without an inhibitor bound. Moreover, the involvement of exosites in modulating the proteolytic activity of enzymes is observed in C1s and MASP-2 (Duncan *et al*., 2012, Perry *et al*., 2013). Such exosites are found on various domains in the catalytic portion of C1s and MASP-2, as well as on their substrates. Whether MASP-3 also adopts this mechanism for regulation of activity is thus another future direction for studying this enigmatic enzyme.

Thus the work in this study has answered many of our initial questions, with candidate activating enzymes known and a much greater understanding of the enzyme's structure and substrate specificity. The clues to MASP-3 function provided by the 3MC syndrome studies were unexpected and of good benefit. Much still needs to be learned, however, with the identity of the physiological substrate(s) of MASP-3 foremost of all questions about the biological role of this enigmatic enzyme.

APPENDIX I

(PUBLICATION)

Mannose-binding lectin serine proteases and associated proteins of the lectin pathway of complement: two genes, five proteins and many functions?

Yongqing T, Drentin N, Duncan RC, Wijeyewickrema LC, and Pike RN.

***Biochim Biophys Acta*. 1824(1):253-62 (2012).**

Biochimica et Biophysica Acta 1824 (2012) 253–262

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Mannose-binding lectin serine proteases and associated proteins of the lectin pathway of complement: Two genes, five proteins and many functions?☆

Tang Yongqing, Nicole Drentin, Renee C. Duncan, Lakshmi C. Wijeyewickrema, Robert N. Pike *

Department of Biochemistry & Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

a r t i c l e i n f o

Article history:
Received 10 March 2011
Received in revised form 27 May 2011
Accepted 27 May 2011
Available online 6 June 2011

Keywords:
Complement
Lectin pathway
Mannose-binding lectin associated serine protease

a b s t r a c t

The lectin pathway of the complement system is activated following the binding of carbohydrate-based ligands by recognition molecules such as mannose-binding lectin (MBL) or ficolins. Engagement of the recognition molecules causes activation of associated MBL-associated serine proteases or MASPs, which in turn activate downstream complement molecules to activate the system. Two MASP genes are alternatively spliced during expression to yield 5 proteins, including three proteases (MASP-1, -2 and -3) and two truncated proteins, MAp19 and MAp44. Here we discuss what is currently known about these proteins in terms of their structure and function. MASP-2 is autoactivated following the initial binding events of the pathway and is able to subsequently activate the C4 and C2 substrates required to activate the rest of the pathway. MASP-1 is able to augment MASP-2 activation, but also appears to play other roles, although the physiological significance of these is not yet clear. The roles of the truncated Map19 and Map44 proteins and the MASP-3 protease are currently unknown. The proteases form an interesting sub-family of proteins that clearly should be the focus of future research in order to establish their biological roles.
This article is part of a Special Issue entitled: Proteolysis 50 years after the discovery of lysosome.

1. Introduction

The complement system is an integral part of the innate immune system. It consists of more than 30 circulatory or cell surface-located proteins which work as a proteolytic cascade to fight against invading pathogens [1]. The complement system also plays a role in shaping B-cell and T-cell responses, and hence represents a bridge connecting the innate and the adaptive immune systems [2]. The complement system can be activated via three major routes: the classical, alternative, and lectin pathways. Each pathway utilizes specific recognition molecules to recognize certain structures on the surface of pathogens. The recognition in turn activates the initiating enzyme which is bound to the recognition molecules. Activation of complement pathways leads to the formation of the membrane attack complex (MAC) and eventually causes death in the target cell [3].

Activation of the lectin pathway was first observed in the mannose-binding lectin (MBL)-dependent activation of C4 [4]. Later, it was found that another group of proteins, called ficolins, was also used by the lectin pathway to recognize surface structures on bacteria, fungi and viruses [5]. The initiating enzyme of the lectin pathway was initially identified as a serine protease in complex with MBL, which was thus termed an MBL-associated serine protease (MASP) [6]. Thus initially the concept for the lectin pathway was that binding of MBL/ficolin to the target triggers the activation of the MBL-bound zymogenic MASP enzyme. The active MASP enzyme in turn cleaves C4 and C2, forming the pivotal C3 convertase (C4b2a), which carries forward the activation of the complement system [6–8].

2. The recognition molecules of the lectin pathway

The recognition molecules of the lectin pathway include MBL and ficolins (see Table 1 for nomenclature of proteins of the lectin pathway of complement) (Fig. 1), both of which are topological homologues of the classical pathway recognition molecule C1q, despite sharing low sequence similarity with the classical pathway molecule [9]. MBL belongs to the collectin family of proteins [10]. The monomer of MBL encompasses three identical polypeptides. Each polypeptide consists of an N-terminal collagen-like region and a C-terminal C-type carbohydrate recognition domain (CRD), which recognizes the 3′ and 4′ horizontal hydroxyl groups of the pyranose ring presented in many carbohydrates [11]. Unlike the strict six-oligomer form of a C1q molecule, a few oligomeric forms of MBL (termed MBL-I, -II, -III and -IV) with molecular weights of 275, 345, 580, and 900 kDa, respectively, have been fractionated from human serum [12]. MBL-I and MBL-II were later determined to be the trimeric and tetrameric forms of MBL, respectively [13]. MBL deficiency is considered the most frequent congenital immunodeficiency state in humans [14]. Low plasma concentrations of MBL have been found to correlate with repetitious

☆ This article is part of a Special Issue entitled: Proteolysis 50 years after the discovery of lysosome.

* Corresponding author. Tel.: +61 3 99029300; fax: +61 3 9902 9500.
E-mail addresses: tang.yongqing@monash.edu (T. Yongqing), nicole.drentin@monash.edu (N. Drentin), renee.duncan@monash.edu (R.C. Duncan), lakshmi.wijeyewickrema@monash.edu (L.C. Wijeyewickrema), rob.pike@monash.edu (R.N. Pike).

1570-9639/$ – see front matter © 2011 Elsevier B.V. All rights reserved. doi:10.1016/j.bbapap.2011.05.021

Table 1
Nomenclature of the protein components of the lectin pathway of complement.

Type of molecules abbreviation	Systematic name	Common	Alternative names	Nomenclature according to human genome organization	Gene origin in human
Recognition molecules	Mannose-binding lectin	MBL	Mannan-binding lectin	Mannose-binding lectin	MBL2
Collagen/fibrinogen domain containing lectin-1		Ficolin-1	M-ficolin	Ficolin-1	FCN1
Collagen/fibrinogen domain containing lectin-2		Ficolin-2	L-ficolin or p35	Ficolin-2 or hucolin	FCN2
Collagen/fibrinogen domain containing lectin-3		Ficolin-3 antigen	H-ficolin or Hakata-	Ficolin-3	FCN3
Collectin sub-family member 11		CL-11	Collectin kidney or CL-K1	Collectin sub-family member 11	COLEC11
Serine proteases and derivatives	Mannose-binding lectin associated serine protease-1	MASP-1		MASP1 isoform 1	MASP1
	Mannose-binding lectin associated serine protease-2	MASP-2		MASP2 isoform 1	MASP2
	Mannose-binding lectin associated serine protease-3	MASP-3		MASP1 isoform 2	MASP1
	Small mannose-binding lectin associated protein	sMAP	MAp19	MASP2 isoform 2	MASP2
	Mannose-binding lection associated protein 1	MAP1	MAp44	MASP1 isoform 3	MASP1

infection and autoimmune diseases [15–17]. Recently, a new member of the collectin family, designated CL-11, was found to be in complex with certain MASP enzymes in human plasma. CL-11, which is localized in the liver, kidney and adrenal gland, primarily recognizes D-mannose and L-fucose on a wide range of organisms [18].

Thus far, three forms of ficolins, termed L- (also named p35 or ficolin-2), M- (or ficolin-1), and H-ficolin (also called Hakata-antigen or ficolin-3) have been found in humans [5,19,20]. L-ficolin and M-ficolin share about 80% similarity in primary structure, different from H-ficolin which shares only 45% sequence similarity with the other members. L-ficolin and H-ficolin are synthesized in the liver and secreted into plasma, while M-ficolin is only expressed on the surface of monocytes. Among the three ficolin members, the collagen-like regions of the ficolin monomers are different in length, while the ligand-binding regions of all ficolin members are fibrinogen-like recognition domains, which recognize the acetyl groups on a wide range of substances, including those on carbohydrates, natural acetylated substances (e.g. N-acetyl-glycine and neo-glycoproteins) and artificial acetylated substances (e.g. acetylated LDL) [5,21].

3. MBL-associated serine proteases (MASPs) and their derivatives

In contrast to early assumptions that there was only one MASP protein, five MASP proteins (Table 1) have thus far been identified, including three MASP enzymes, termed MASP-1/-2/-3, a non-enzymatic protein, MAp19 (or sMAP), which shares its origin with MASP-2, and another non-catalytic protein, MAp44 (or MAP1), which shares its origin with MASP-1 and MASP-3 (Fig. 2) [6,12,22–25]. The results of sequence alignment of MASP enzymes and C1r/C1s revealed that these proteins are members of the same MASP/C1r/C1s family. These members possess a conserved His/Ser/Asp catalytic triad, and share the same domain organization: from the N-terminus, the first C1r/C1s/Uegf/bone morphogenetic protein 1 (CUB1) domain, an epidermal growth factor-like (EGF) domain, the CUB2 domain, followed by two complement control protein (CCP1 and CCP2) domains and the C-terminal serine protease (SP) domain. The SP domain is linked to the CCP2 domain by a small linker region [26].

3.1. Genes and sites of expression for MASP-1/-3

In humans, MASP-1 and MASP-3 arise as the products of alternative splicing of the *masp1/3* gene, located on chromosome 3q27-28 [12]. A corresponding gene in mouse was mapped to murine chromosome 16B2-B3 [27] and it was later found that the production of MASP-1 and MASP-3 in mouse and rat also involves alternative splicing of this corresponding gene [28,29]. The *masp1/3* gene spans about 50 kb which encodes for 17 exons. The first 10 of these 17 exons encode the first five domains shared by MASP-1 and -3, followed by a single exon that encodes a small MASP-3-specific linker region plus the serine protease domain of MASP-3, followed in turn by the last 6 exons encoding a small MASP-1 specific linker region plus the serine protease domain of MASP-1 [26]. Like MASP-1 and MASP-3, the recently discovered MAp44 is another product of alternative splicing of the *masp1/3* gene found in humans. MAp44 mRNA consists of the first 8 exons encoding the first four domains identical to those of MASP-1 and MASP-3 and an extra exon encoding 17 amino acids unique to MAp44 [24,25].

Homologues of the *masp1/3* gene have been found in many species of vertebrates, including cartilaginous fish, bony fish, birds and mammals, as well as in the ascidian and the amphioxus invertebrate species [30–32]. However, the MASP-1 transcript was not detected in lamprey, either fish species or birds [33]. The absence of MASP-1 expression in birds was found to be due to the absence of the 6 exons encoding the MASP-1 serine protease domain. This finding suggested that a loss of these exons may have occurred in birds during evolution of the *masp1/3* gene [34]. Compared to this, the MASP-3 transcript is present in all of the vertebrates analyzed, including those in which the MASP-1 transcript is absent, indicating that MASP-3 may have an important physiological role [33].

By measuring the levels of transcripts in different human tissues, it was established that the liver was the primary expression site for all MASP enzymes, while the heart and skeletal muscles are the main expression site for Map44. Compared to the limited numbers of expression sites for MASP-1 and MASP-2, tissues from a wide range of organs contain relatively high levels of the MASP-3 transcript, including the pancreas, skeletal muscle, spleen, thymus, prostate and ovary [35]. In addition, an astrocyte-derived human glioma cell line T98G also expresses MASP-1 and MASP-3, indicating that the MASP enzymes may play a role in the lectin pathway in the brain [36]. In human liver, the level of the MASP-3 transcript was observed to be similar to that of MASP-1 [35], but in rat liver, the level of the MASP-3 transcript is about two thirds of the MASP-1 transcript, even though both transcripts arise from transcription using the same gene promoter [28]. The cause of this difference in the levels of transcripts of MASP-3 and MASP-1 in rat liver remains speculative. Major suggestions include structural factors in the gene, such as the branchpoint sequence and poly-pyrimidine tract in MASP-3 which might suppress transcription

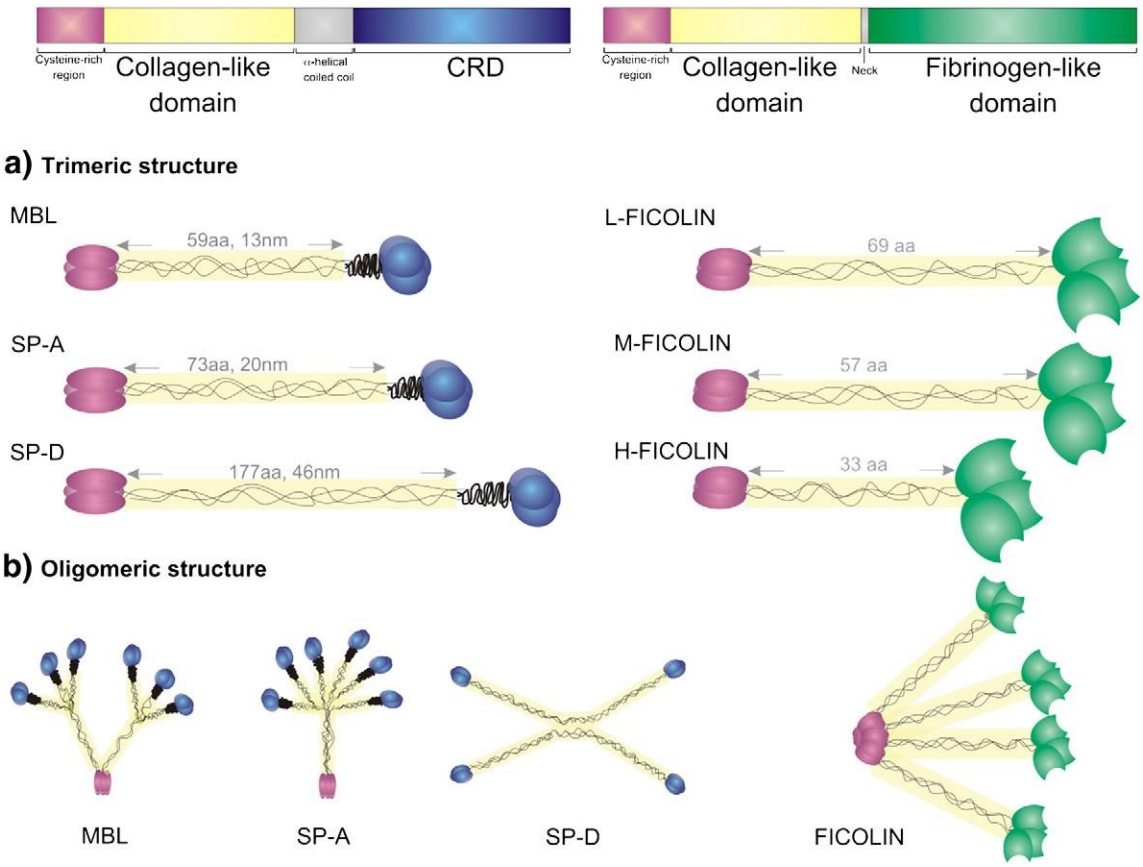


Fig. 1. Diagrammatic representation of the recognition molecules of the lectin pathway of complement. Top, scheme for the domain structure of the collectins (left) and ficolins (right). For the collectins, the cysteine-rich domain is shown in pink, the collagen-like domain in yellow, the alpha-helical coiled-coil region in gray and the carbohydrate recognition (CRD) domain in blue. The fibrinogen-like domain of the ficolin family members is shown in green. a) The basic organization of the trimeric units of the collectin (left) and ficolin (right) family proteins is shown. Colors correspond to the domains labeled above in the schematic domain representation. b) Potential oligomeric structures for each of the collectin and ficolin family members are shown. It should be noted that other oligomeric forms also exist, especially in the case of MBL.

and/or post-transcriptional factors that may destabilize the MASP-3 transcript [28].

3.2. Structural chemistry of MASP-1/-3

Both MASP-1 and MASP-3 are synthesized as a single polypeptide chain (Table 2) [37]. The MASP-1 polypeptide shares the features of an ancient serine protease, chymotrypsin. These features include six exons encoding one serine protease domain, a TCN (N denotes A, G, C or T) codon encoding the catalytic Ser residue, and the presence of a so-called 'histidine loop' in the serine protease domain formed bya disulfide bond [26,38]. In comparison, the MASP-3 polypeptide sequence implies an evolutionary relationship that is further apart from the chymotrypsin origin than MASP-1, but closer to MASP-2, C1r and C1s. Similar to MASP-2, C1r and C1s, the MASP-3 polypeptide consists of a serine protease domain encoded by a single exon and a catalytic Ser residue encoded by an AGY (Y denotes C or T) codon. The 'histidine loop' present in MASP-1 is absent in the serine protease domains of MASP-3, MASP-2, C1r and C1s [26,38,39].

3.2.1. The CUB1-EGF-CUB2 segment

The crystal structure of the CUB1-EGF-CUB2 segment shared by human MASP-1 and MASP-3 has been recently solved to a resolution of 2.3 Å [40]. The crystal structure revealed that the CUB1 domain lacks two strands that were observed in the general structure of a typical CUB module, described as a sandwich of two five-stranded β-sheets [41], while the CUB2 domain is only lacking of one of the two strands missing in CUB1. These structural features of the CUB domains in MASP-1/-3 were also observed in human C1s, rat MASP-2 and human MAp19 [42–44]. A Ca^{2+} is bound in each of the two CUB domains of MASP-1/-3, contributing to the stability of the corresponding CUB domain. A homologous Ca^{2+} binding site was also found in the CUB1 domain of human C1s and Map19 [42,44]. The structure of the EGF domain of MASP-1/-3 is similar to that of a typical EGF homologue, which is characterized by two double-stranded β-sheets linked by five loops and stabilized by three disulfide bonds [45]. A Ca^{2+} binding site is also present in the MASP-1/-3 EGF domain. A few amino acids forming this binding site are strictly conserved in all other members of the MASP/C1r/C1s family, suggestive of the importance of this Ca^{2+} binding site [40].

Previous studies using truncated MASP-1 proteins from humans and rat suggested that MASP-1 monomers form homodimers via a primary binding site at the CUB1 domain [46–48]. Homodimers of MASP-3 were also observed in the MBL-MASP-3 complex in human plasma [12]. Thus far, a heterodimer formed between MASP-1 and MASP-3 monomers via their respective, but identical CUB1-EGF-CUB2 segments, has not been observed. The crystal structure of the CUB1-EGF-CUB2 fragment of MASP-1/-3 revealed that the CUB1 domain of one monomer interacts with the EGF domain of the other monomer in a 'head to tail' fashion. Hence the two monomers form a homodimer with two interacting surfaces, each of which is stabilized by hydrophobic interactions and hydrogen bonds [40]. Furthermore, it was found that all three domains of the CUB1-EGF-CUB2 segment of MASP-1 were necessary for the MASP-1 homodimer to effectively bind to MBL and L-ficolin in a Ca^{2+}-dependent manner [46–48]. It is therefore expected that MASP- 3 binds to its recognition molecule partner in a similar manner. A study using point mutations to residues that participate in the binding of Ca^{2+}

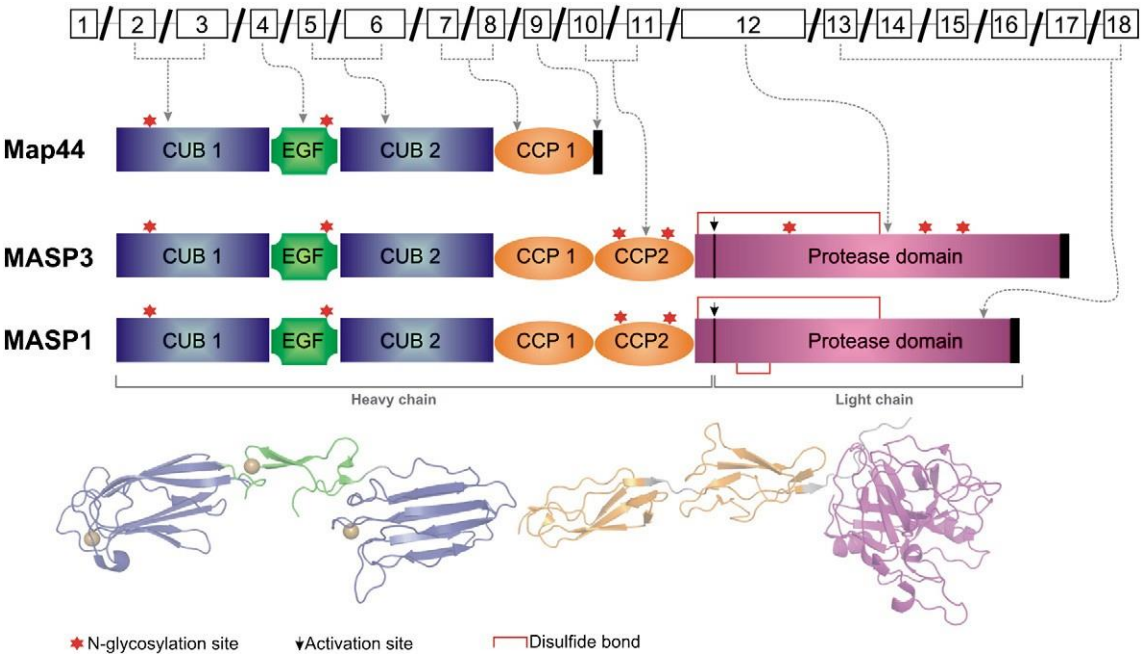


Fig. 2. Genomic organization of the MASP-1/3 gene and resulting protein structures. Top, the exon–intron structure of the MASP-1/3 gene. MAp44 contains the exons encoding the two CUB domains, the EGF domain and the first CCP domain together with the unique exon 9, which contains a stop codon (marked with black boxes). MASP-3 is generated by splicing out exon 9 and terminated by the stop codon in exon 12. The protease domain of MASP-3 is encoded by a single exon, whereas this region in MASP-1 is encoded by six exons. The MASP-1 sequence contains a stop codon in exon 18. The potential glycosylation sites are shown using red stars, an arrow indicates the scissile Arg\Ile bond and a red line indicates the disulfide linkage holding the two chains together. Bottom, the crystal structures of human MASP-1 CUB1-EGF-CUB2 (PDB ID: 3DEM) and the catalytic region of MASP-1 (PDB ID: 3GOV) are shown as ribbon diagrams with the domains colored in the same way as in the domain schemes above. The calcium ions in the CUB and EGF domains are shown as beige spheres.

in the CUB1-EGF-CUB2 segment of MASP-3 revealed that Glu 49 in the CUB1 domain, Asp 102 and Phe 103 in the EGF domain and Tyr 225 and His 218 in the CUB2 domain all played a role in the interaction between MASP-3 and its binding partners [40]. Of these residues, the acidic Glu 49 and Asp 102 residues are most likely to be involved in direct binding to MBL/ficolin, since a previous study revealed that the positively-charged residue Lys 55 of MBL/ficolin is involved in their binding to all three human MASP enzymes and MAp19 [49].

In human serum, MASP-1 and MAp19 were found to be in complex with MBL-I, and MASP-3, together with MASP-2, was initially found to be in complex with MBL-II [12]. Later, Zundel et al. reported that MASP-3 bound to immobilized MBL-I and MBL-II with similar equilibrium dissociation and rate constants [50]. Recently it was found that human MASP-3 is predominantly in complex with H-ficolin, with only a small proportion associated with L-ficolin and MBL [51]. The structural details underpinning these preferences in the formation of MBL/ficolin-MASP complexes are unclear.

3.2.2. The CCP1-CCP2-SP segment

The crystal structure of the CCP1-CCP2-SP segment of human MASP-1 has been solved recently by Dobó et al. at a resolution of 2.55 Å [52]. Each of the two CCP domains of MASP-1 resemble a common CCP homologue, featuring a rod-shaped module made of a number of β-sheets stabilized by two disulfide bonds [53]. However, a Lys-Asn electrostatic interaction is presented uniquely at the interface between the CCP2 domain and SP domain of MASP-1, accounting for the rigidity of this region of MASP-1 [52].

The active site in the SP domain of MASP-1 is located in a relatively exposed position and consists of a long substrate-binding groove that topologically resembles the substrate-binding area of trypsin, formed by a number of loop structures. Therefore it is expected that the substrate specificity of MASP-1 is similar to that of trypsin, which exhibits a strict preference for Arg or Lys as the P1-residue [note that the nomenclature for serine proteases indicates that the cleavage occurs in substrates between the P1 and P1′ residues, with substrate residues N-terminal to P1 labeled P2, P3 etc. and those C-terminal to P1′ labeled P2′, P3′ etc.; sub-sites binding the substrate residues in the enzymes are correspondingly labeled, e.g. the S1 sub-site binds the P1 residue, etc. [54]]. Indeed, a strong preference for P1-Arg/Lys residues by MASP-1 was confirmed by Rossi et al. using synthetic peptidyl substrates [55].

A so-called 60-loop with large aromatic residues shadows the area N-terminal to the S1 sub-site of MASP-1, indicating a possible preference for small non-polar amino acids at the P2 position. Overall, the long and open substrate-binding area of MASP-1 suggests broad substrate specificity. Interestingly, a salt bridge is formed by the residues Asp 640 and Arg 677 in the MASP-1 active site, which would be predicted to restrict the ability of the Asp 640 residue to bind substrates and thus reduce the cleavage efficiency of MASP-1. Close to this region, another salt bridge is formed by the residues Asp 670 and Lys 623. This second salt bridge not only hinders the ability of Asp 670 to bind substrates, but also brings a patch of three Lys residues (622–624) into the substrate binding groove, potentially further reducing the substrate-cleaving potency of MASP-1. It is proposed that these salt bridges are broken when the preferred substrate binds to the substrate-binding site of MASP-1 and induces a conformational change in the binding site, restoring the substrate-binding capability of restrained residues [52].

Table 2
Features of the proteins of the MASP sub-family.

MASP family member	Amino acids (signal peptide/polypeptide)	Molecular mass (kDa) (calculated/observed)	Glycosylation sites
MASP-1	19/680	77/90	4
MASP-2	15/671	74/74	0
MASP-3	19/709	82/94	7
Map19	15/170	22/22	0
Map44	19/361	44/45	2

The crystal structure of the SP domain of MASP-3 is yet to be solved. Sequence alignment of the SP domain of MASP enzymes reveals five strictly conserved Cys residues, including two that are involved in forming the 'methionine loop' (or loop 3). This loop structure is near the catalytic residues and thus may affect substrate binding and/or cleavage by the enzyme [28]. The methionine loop of MASP-3 (-KTSYESRSGNYSVTENMF-) consists of neither the Lys patch presented in MASP-1 (− QKAYAPLKKKVTRDMI-) as mentioned above, nor the frequently presented Pro residue in that of MASP-2 (− TAAYEKPPYPRGSVTANML-). Instead, the methionine loop of MASP-3 contains an N-glycosylation site at its center, which is not presented in those of MASP-1 and MASP-2. These features of the MASP-3 methionine loop indicate that it may exert a unique impact on the substrate specificity and/or substrate-cleavage efficiency of the enzyme.

3.3. The activation of MASP-1/-3

The activation of the MASP zymogens involves cleavage at a conserved Arg\Ile peptide bond, resulting in the active MASP enzyme consisting of A- (or H) and B- (or L) chains, linked by a disulfide bond [6]. The A chain is comprised of the five N-terminal domains and the linker region and the B chain consists of the SP domain. MASP-1 is believed to undergo auto-activation, since substituting the catalytic triad component, Ser 627, with an Ala residue resulted in only the zymogenic form of MASP-1 being found [50]. In contrast, the MASP-3 proenzyme does not appear to undergo auto-activation [50].

3.4. The biological functions of MASP-1/-3

Initially it was thought that the concentration of MASP-1 in human plasma was 6.27 ± 1.85 μg/mL [56]. It is now established that this value includes other derivatives of the *masp1/3* gene, and the actual plasma concentration of MASP-1 is around 1 μg/mL [57]. By using a monoclonal MASP3-specific antibody, Degn et al. determined the mean concentration of MASP-3 in human plasma to be 5.2 μg/mL, which is close to the value of 6.4 μg/mL, previously measured by Skjoedt et al. [51,57]. The plasma level of MASP3 in humans appears to be stable 6 months after birth. The plasma concentration of MAp44 is 1.4 μg/mL, determined by using an antibody that recognizes the 17 characteristic C-terminal residues of MAp44 [57].

MASP-1 was observed to cleave C3, and was thought to activate the complement system directly via such cleavage [12,58,59]. Studies by Petersen et al. and Selander et al. confirmed the C3 cleavage by MASP-1, but showed that it occurred at relatively low efficiency [60,61]. To the contrary, a study by Wong et al. was not able to observe C3 cleavage by human plasma-derived MASP-1, and subsequent studies using native MASP-1 and eukaryote-expressed recombinant MASP-1 found that they cleaved only hydrolysed C3, instead of the native substrate [62]. Rossi et al. and Ambrus et al. confirmed the C3 cleavage by MASP-1, but argued against the capability of MASP-1 to activate the complement system via such cleavage due to its low cleavage rate [55,63]. Therefore, the involvement of MASP-1 in activating the complement system by cleaving C3 remains unlikely. MASP-1 was also found to cleave C2 [55,59,63], which was believed to increase the formation of the C3 convertase complex mediated by MASP-2 [64–66]. It is also thought that MASP-1 might promote the activation of MASP-2, which in turn is able to strongly activate the entire complement system [67,68]. However, this contribution of MASP-1 is debatable since MASP-2 itself clearly undergoes auto-activation [69]. Furthermore, MASP-1 exhibits thrombin-like activity, cleaving two substrates of thrombin, fibrinogen and factor XIII [70]. It was also found that, like thrombin, MASP-1 can be effectively inhibited by antithrombin in presence of heparin [52]. In addition, the activity of MASP-1 is tightly controlled in human plasma by C1-inhibitor and α-2-macroglobulin [55,59,71,72].

A recent study utilized sunflower trypsin inhibitors engineered for selectivity for MASP-1 and MASP-2 to examine the contribution of the two enzymes to activation of the lectin pathway [68]. It was found that inhibiting either of the two enzymes prevented the activation of the lectin pathway, and that inhibiting MASP-1 was more effective than inhibiting MASP-2. The authors therefore suggested that MASP-1 plays a determinant role in activating the lectin pathway [68]. Studies using transgenic mice have suggested a possible role of MASP-1 in the activation of alternative pathway [73,74]. Takahashi et al. found that the *masp1/3*-null mouse failed to convert the zymogenic factor D (pro-Df) to active factor D (Df), and thus had an impaired alternative complement pathway. In their study, a recombinant MASP-1 was shown to activate the recombinant murine pro-Df in vitro. However, addition of this recombinant MASP-1 into the serum of the *masp1/3-null* mouse failed to detect the activation of pro-Df [73]. Another study by Banda et al. using an alternative pathway-dependent arthritis murine model revealed a direct correlation between MASP-1 and the activation of alternative pathway via activating pro-Df. In the latter study, mixing the serum from the *masp1/3-null* mouse and that from a Df-knock-out mouse resulted in activation of the pro-Df in the former serum [74]. The cause of the difference between the results of these two studies remains to be solved. However, neither of the two studies addressed the possible involvement of MASP-3, which is also knocked out with MASP-1 in the transgenic mouse since the two proteins are encoded by the same *masp1/3* gene.

Compared to MASP-1, studies of the substrate specificity of MASP-3 have been limited. Cortesio and Jiang reported that MASP-3 cleaves a few synthetic substrates, among which Pro-Phe-Arg-AMC and Boc-Val- Pro-Arg-AMC displayed the highest rate of cleavage by MASP-3. The authors also found that MASP-3 cleaves insulin-like growth factor-binding protein (IGFBP)-5, which is thus far the only physiological substrate of MASP-3 identified. The three sites in IGFBP-5 cleaved by MASP-3 all contain a Lys/Arg residue at P1-position and a Pro residue at P2-position. Therefore, the authors claimed that the substrate specificity of MASP-3 is Pro-Arg/Lys-X [75]. However, a main drawback in this study is that the active MASP-3 started with the Ile 450 at the N-terminus following the activation bond. It is uncertain whether this truncated serine protease of MASP-3 can fold correctly. Hence it is uncertain whether the kinetic behavior of this truncated mutant resembles that of a full-length MASP-3 or the catalytic portion (CCP1-CCP2-SP) of MASP-3.

Up to now, the biological functions of MASP-3 and the recently identified MAp44 remain uncertain. MASP-3 does not cleave any complement components and it is not inhibited by C1-inhibitor [50,76]. Dahl et al. and Møller-Kristensen et al. found that incubating MASP-3 with MBL-MASP-2 complexes on surface-coated MBL resulted in a lower level of cleavage of C4 mediated by MASP-2 [12,65]. Skjoedt et al. also claimed that MASP-3 acts as a negative regulator of the cleavage of C4 mediated by the H-ficolin-MASP complex in human plasma [51]. It was thought that these down-regulating impacts of MASP-3 occur via competition with MASP-2 for binding to the recognition molecules. However, whether this represents the main physiological function of MASP-3 is doubtful, since MASP-3 does not cleave MASP-2 and the physiological substrates of MASP-3 are yet to be identified. Similarly, increasing concentrations of recombinant MAp44 in complex with either recombinant MBL or recombinant H-ficolin caused a decreased level of cleavage of C4 in sera depleted of MBL and H-ficolin, respectively. This finding suggests that MAp44 exerts a negative impact on the activation of complement system [25]. Considering the striated muscle-localization of MAp44, this negative impact of MAp44 may represent a tissue-specific protective mechanism by which local inflammation is prevented. In addition, it was recently found that the plasma level of MASP-3 and MAp44 in humans does not severely drop following surgery, and recovers after about 4 days and 30 days, respectively. Hence it is unlikely that MASP-3 and MAp44 act as acute phase proteins [57].

3.5. MASP-2

Found a few years after MASP-1, MASP-2 was discovered while researchers were searching for conglutinin-like protein in human plasma [22]. For a short time it was thought that MASP-2 was the same as C1s in the classical pathway [77]. Indeed, MASP-2 was found to be homologous to the protease C1s, and to share the same substrates (C2 and C4) [76]. However, further analysis showed that despite the homology, MASP-2 was quite different in assembly and function to C1s [77]. Some examples of these differences include MASP-2 being able to autoactivate itself and the enzyme's ability to activate prothrombin [78]. The autoactivation of MASP-2 also represents the minimal requirement for lectin pathway activation [79].

MASP-2 is synthesized as a proenzyme. It binds to MBL, and when MBL recognizes pathogenic carbohydrates moieties, MASP-2 autoactivates itself by cleaving an arginine-isoleucine peptide bond at the N-terminal of the serine protease domain [76]. After autoactivation, MASP-2 goes on to cleave its substrates, C2 and C4, both of which contribute to the formation of the C3 convertase [80]. MASP-2 has only a few natural substrates/inhibitors: zymogenic MASP-2, C2, C4 and C1 inhibitor. It cleaves these substrates after arginine residues [81]. MASP-2 mRNA has been found to be expressed in the liver and thus far has not been found to be expressed in any other tissues in the body [82].

3.6. Structural chemistry of MASP-2

The MASP-2 protease is comprised of three N-terminal non-catalytic domains (CUB1-EGF-CUB2) and three catalytic domains (CCP1-CCP2-SP) (Fig. 3).

3.6.1. Non-catalytic domains

The non-catalytic domains are important in allowing MASP-2 to perform two functions. One is that these domains allow MASP-2 to form homodimers, and it appears that only the first two domains (CUB1 and EGF) are required for this to occur [46] and that the dimerization process is calcium-independent [83]. The second is that these domains allow the MASP-2 homodimer to bind to MBL in a calcium-dependant manner through its CUB1 and EGF domains [83]. A calcium binding site has been found at the N-terminus of the EGF domain [43]. Binding of calcium ions has been shown to protect against proteolysis and to stabilize the orientation of these domains [83]. The CUB regions may help to increase the EGF domain's affinity for calcium, as it has been shown that calcium binding EGF-like domains experience an increase in affinity for calcium ions by up to three orders of magnitude in the presence of an adjacent protein domain [84]. The structure of the non-catalytic region was elucidated in 2003 using rat CUB1-EGF-CUB2, and displayed an elongated 'letter C' shape [43].

3.6.2. Catalytic domains

The serine protease (SP) domain displays a typical chymotrypsin fold, with two six-stranded β barrel domains packed against one another [80]. At the junction of these barrels are the catalytic triad of Ser633, His483 and Asp532 (Ser195, His57, Asp102 using chymotryp- sin numbering) [80]. The specificity of the S1 pocket and other substrate binding sites are determined by eight surface loops [80]. Despite sharing characteristics with other complement proteases, such as the same substrates with C1s, many of these loops differ in length, amino acid composition and conformation [80]. In fact, some of the loops are quite similar to those of thrombin or trypsin [80]. It is likely that the conformation of these loops helps to ensure the relatively narrow selectivity for protein substrates by restricting access to the substrate-binding sites [77]. It is thought that loops 1, 2 and D, along with the activation peptide, show a high degree of flexibility and form the 'activation domain' [69,85].

Both an active form and a zymogen mutant (R444Q) form of MASP- 2's catalytic segment have been crystallized [69,80]. Crystal structures have shown that the β barrel sub-domains in the SP of both active and zymogen forms are very similar and that the catalytic triad is in an active conformation in both forms [69]. However, some major structural differences between active and zymogen form are seen in the loops [69]. A number of these loops (Loops A, B, C and 3) are normally unchanged in chymotrypsin-like enzymes upon activation, but in the active structure of MASP-2 structure they were significantly altered compared to the zymogen form, indicating their flexibility [69]. Another autoactivating enzyme, C1r, also undergoes major conformational changes in these loops upon activation [69,86]. Gal et al. hypothesize that this may be a feature of autoactivating enzymes, and that this flexibility permits the conformational changes required of the loops to form an active enzyme [69]. He suggests that these loops, along with the activation domain, form an 'autoactivation domain' [69].

Crystallization of the active form of MASP-2 has revealed two different orientations for the CCP2-SP domain interface [80]. Both conformations share new patterns of interactions in this CCP2-SP interface that differ from the usual hydrogen bonds and van der Waals contacts at the same interface seen in structures of C1r and C1s [80]. The interactions between the CCP-2 and SP domains are quite strong, and differential scanning calorimetry (DSC) performed by Harmat et al. have shown that CCP2 helps to increase the stability of the SP domain [80]. Despite this strong interaction, when compared to the rigid CCP2-SP interface of C1s, the interface is quite flexible. This may be due to a lower number of proline residues in the interface region (two for MASP-2, as opposed to four for C1s and five for C1r), as well as the reduced number of disulfide bonds (only two compared to the usual four to six bonds in human trypsin-like enzymes) [80,87]. It is thought that this flexibility allows MASP-2 to move its serine protease domains into the correct position during autoactivation, during C2 and C4 cleavage and may also contribute towards MASP-2's substrate specificity [69,80].

The CCP1 and CCP2 domains associate with the serine protease domain and help to stabilize it. As discussed above, it is known that CCP2 is necessary to stabilize the SP domain [69]. They may also provide an accessory binding site(s) for C4 [63]. The precise number of such sites and their exact location on the CCP domains is currently unknown. Studies have shown that for efficient C4 cleavage, the CCP domains are necessary [63]. This recognition appears to be a part of the MASP-2 CCP units, as when C1s had its CCP1 and CCP2 domains replaced with MASP-2 CCP1 and CCP2 domains, it experienced a strong increase in C4 cleavage [88]. Truncated forms of MASP-2 have demonstrated that the absence of one or both of these domains does not affect the rate of C2 cleavage [63] and so it appears that they do not assist cleavage of C2, and that the determinants for C2 cleavage thus reside in the serine protease domain [63].

3.7. MAp19/sMAP

MBL Associated Protein 19 (MAp19), also known as sMAP, is a truncated 19 kDa product of alternative splicing and polyadenylation of the primary RNA transcript of the MASP-2 gene [23,89]. It contains the same CUB1 and EGF domains as MASP-2, but has an extra four unique amino acids at the C-terminal end of the protein (Fig. 3) [44]. MAp19 mRNA has only been found in the liver [82]. Like MASP-2, it forms homodimers via the CUB1 and EGF domains and associates with MBL and L-ficolin in a calcium-dependent manner [44].

The structure of MAp19 was published in 2004, and in addition to the calcium-binding site on the EGF domain, it showed a second calcium-binding site on the CUB1 domain [44].

The function of MAp19 is not yet fully understood, but it is speculated that because of its ability to bind to MBL and ficolins, it competes with the MASPs to prevent the activation of the lectin pathway [90]. For this to be effective, however, the concentrations of the MASPs and MAp19 would have to be greater than the total number of binding sites on MBL and ficolin in the blood [77]. Currently, this is believed to be unlikely, as it is

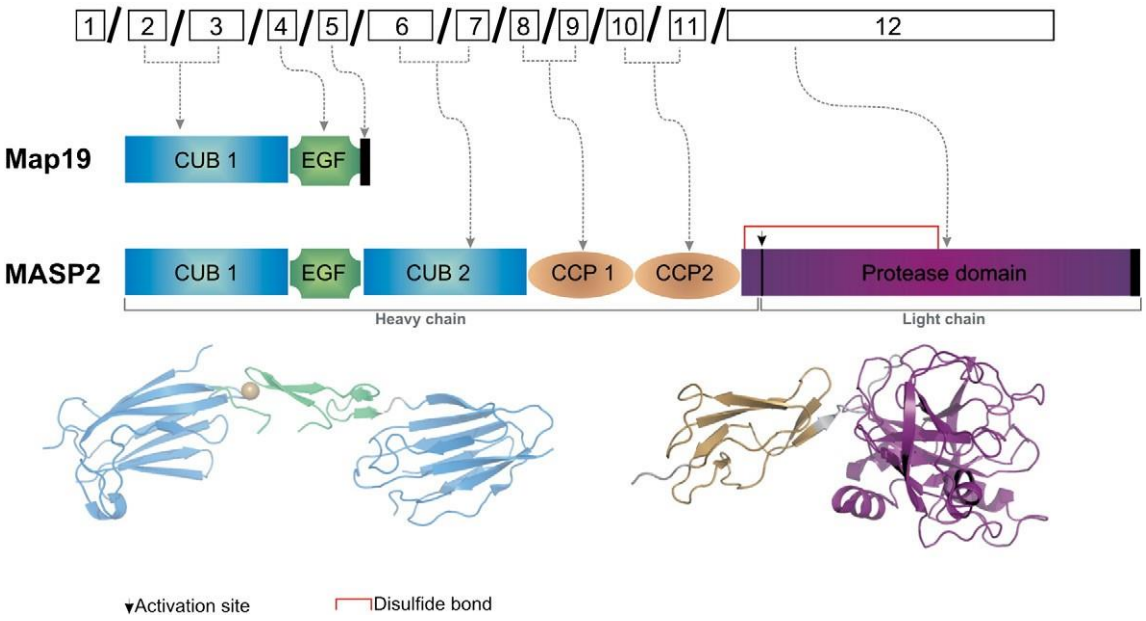


Fig. 3. Genomic organization of the MASP-2 gene and resulting protein structures. Top, the exon–intron structure of the MASP-2 gene. Map19 contains the exons encoding the CUB1 domain and the EGF domain. An arrow indicates the scissile Arg\Ile bond and a red line indicates the disulfide linkage holding the two chains together. Bottom, the crystal structures of human MASP-2 CUB1-EGF-CUB2 (PDB ID: 1NTO) and the catalytic region of MASP-2 (PDB ID: 1Q3X) are shown as ribbon diagrams with the domains colored in the same way as in the domain schemes above. The calcium ion at the interface of the CUB1 and EGF domains is shown as a beige sphere.

thought that MBL and ficolins are present in excess amounts within the circulation [91].

3.8. MASP-2 and disease

The lectin complement pathway plays an important role in the innate immune system, especially during early childhood, when the adaptive immune system is still developing, or when adaptive immunity is compromised in adults [92]. The clinical effects of MASP-2 deficiencies and mutations remain controversial. A number of studies have found no link between MASP-2 deficiency and increased infection rates [93,94]. However, there are two clinical examples of a person suffering serious effects as a result of MASP-2 deficiency. One patient was homozygous for a point mutation in exon 3 of the MASP-2 gene, which results in a D105G mutation (D120G if the signal peptide is included) [95,96]. This mutation occurs close to the C-terminus of the CUB1 domain [96]. This patient suffered recurrent infections and autoimmune disease manifestations [95]. The other example was another patient also homozygous for the D105G mutation. They were discovered after attending a lung clinic for allergic symptoms [96]. These cases are interesting because homozygosity for the D105G mutation has been found to occur at a relatively high frequency in certain populations (i.e. 1 in 1000 within the Danish population), yet cases like the ones above appear rare [96]. Some have suggested that while MASP-2 deficiency in itself does not normally cause severe disease, when combined with other diseases or immune suppressed states, such as cystic fibrosis or transplantation, it may have more severe consequences [97]. An example of this is the increased likelihood of Cytomegalovirus (CMV) infection in kidney transplant patients with low MASP-2 levels [98].

A growing area of study is the link between MASP-2 and various cancers. One of the cancers most prominently studied is colorectal cancer (CRC), the third most common cancer in the industrial world [99]. The blood serum levels of MASP-2 and activity levels of the enzyme were increased in CRC patients as compared to healthy blood donors [99]. It is believed that this may be due to chronic inflammation caused by the cancer, which may take several years to develop [99]. One group have found that increased post-operative MASP-2 serum levels in CRC are linked with a poor survival rate and high rates of recurrent cancer [100], although other findings have not supported these results [101]. It is hypothesized that this may be due to differences in the ethnic mix amongst the patients or the length of study in each case [101]. It is still yet to be established whether the increased MASP-2 levels influence the course of the disease (i.e. the increased MASP-2 levels cause a more aggressive disease course by increasing inflammation and facilitating proliferation, tumor invasion and metastasis) or the disease modifies the levels of MASP-2 itself [100]. Different MASP-2 genotypes have not been found to be linked to the increased MASP-2 levels in CRC patients, so the cause of the increased MASP-2 levels does not appear to be genetic [102].

Increased MASP-2 levels have also been detected in patients suffering esophageal squamous cell carcinoma (ESCC). Tumor cells in these patients have been shown to express MASP-2, where its expression is linked to advanced clinical cancer stage and nodular metastasis [103]. In an interesting variation, a study of pediatric patients suffering from hematological malignancies found a higher concentration of MASP-2 in sera was linked to greater event-free survival [104]. This was especially seen in patients suffering from lymphoma.

The study of the relationship of various cancers and MASP-2 is still in its infancy. Further studies are needed to strengthen and verify the link between MASP-2 levels and cancer, as well as to ascertain the details of this link, and how it differs depending on cancer type.

Surprisingly, there may also be a link between increased MASP-2 activity and schizophrenia. A few studies have shown that in the sera of schizophrenic patients there was an increase in MBL-bound MASP-2 activity [105,106]. However, how and why this occurs is not clear, especially as MASP-2 mRNA has not been found to be expressed in brain tissue [82]. Therefore, further research is required to fully establish the details of the link between schizophrenia and MASP-2.

4. Conclusions and future directions

The MASP enzymes of the complement system represent an interesting family of enzymes with much work to be carried out to understand their structure and function. It is clear that MASP-2 alone

is able to activate the lectin pathway of complement and yet it is also apparent that at least MASP-1 is able to augment this to some degree. The roles of MASP-2 in disease are starting to become more defined and some interesting and unexpected connections are being made in this regard. The precise mechanism/s whereby this enzyme is able to auto-activate upon recognition of ligands by its partner recognition molecule is/are also a potentially fruitful area for future study. The precise biological role/s for MASP-1 is/are not yet understood, despite a number of suggested substrates for the enzyme. It is clear that the lack of clear evidence for the role of the enzyme that can be generated by studies using mutant mice has hampered research in this regard. This also applies to MASP-3 and the other truncated gene product, MAP44. MASP-3 is arguably the most enigmatic member of the entire family. At present, its mechanism of activation, its structure and biological roles are almost entirely unknown. It is clearly an important enzyme given that it is highly conserved amongst species and has a more widespread tissue distribution than the other family members and yet its role is almost totally unknown at this time. Clearly, the biological roles of this enzyme and the truncated gene products for each enzyme should be a fruitful avenue for future research.

Acknowledgments

This work was supported by a Program Grant from the National Health & Medical Research Council of Australia.

References

[1] R.B. Sim, S.A. Tsiftsoglou, Protease of the complement system, Biochem. Soc. Trans. 32 (2004) 21–27.
[2] J.R. Dunkelberger, W.C. Song, Complement and its role in innate and adaptive immune responses, Cell Res. 20 (2010) 34–50.
[3] J.C. Papadimitriou, L.E. Ramm, C.B. Drachenberg, B.F. Trump, M.I. Shin, Quantitative analysis of adenine nucleotides during the prelytic phase of cell death mediated by C5-9, J. Immunol. 147 (1991) 212–217.
[4] K. Ikeda, T. Sannoh, N. Kawasaki, T. Kawasaki, I. Yamashina, Serum lectin with known structure activates complement through the classical pathway, J. Biol. Chem. 262 (1987) 7451–7454.
[5] M. Matsushita, Y. Endo, S. Taira, Y. Sato, T. Fujita, N. Ichikawa, M. Nakata, T. Mizuochi, A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin, J. Biol. Chem. 271 (1996) 2448–2454.
[6] M. Matsushita, T. Fujita, Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease, J. Exp. Med. 176 (1992) 1497–1502.
[7] Y.H. Ji, T. Fujita, H. Hatsuse, M. Kawakami, Activation of the C4 and C2 components of complement by a proteinase in serum bactericidal factor, Ra reactive factor, J. Immunol. 150 (1993) 571–578.
[8] N. Kawasaki, T. Kawasaki, I. Yamashina, A serum lectin (mannan-binding protein) has complement-dependent bactericidal activity, J. Biochem. Tokyo 106 (1989) 483–489.
[9] M. Ohta, M. Okada, I. Yamashina, T. Kawasaki, The mechanism of carbohydrate-mediated complement activation by the serum mannan-binding protein, J. Biol. Chem. 265 (1990) 1980–1984.
[10] R. Malhotra, J. Haurum, S. Thiel, R.B. Sim, Interaction of C1q receptor with lung surfactant protein A, Eur. J. Immunol. 22 (1992) 1437–1445.
[11] W.I. Weis, K. Drickamer, W.A. Hendrickson, Structure of a C-type mannose-binding protein complexed with an oligosacch, Nature 360 (1992) 127–134.
[12] M.R. Dahl, S. Thiel, M. Matsushita, T. Fujita, A.C. Willis, T. Christensen, T. Voruo-Jensen, J.C. Jensenius, MASP-3 and its association with distict complexes of the mannan-binding lectin complement activation pathway, Immunity 15 (2001) 127–135.
[13] F. Teillet, B. Dublet, J.P. Andrieu, C. Gaboriaud, G.J. Arlaud, N.M. Thielens, The two major oligomeric forms of human mannan-binding lectin: chemical characteriza- tion, carbohydrate-binding properties, and interaction with MBL-associated serine proteases, J. Immunol. 174 (2005) 2870–2877.
[14] M.W. Turner, R.M. Hamvas, Mannose-binding lectin:structure, function, genetics and disease association, Rev. Immunogenet. 2 (2000) 305–322.
[15] M. Super, S. Thiel, J. Lu, R.J. Levinsky, M.W. Turner, Association of low levels of mannan-binding protein with a common defect in opsonisation, Lancet 2 (1989) 1236–1239.
[16] M. Boniotto, L. Braid, V. Baldas, T. Not, A. Ventura, S. Vatta, O. Radillo, F. Tedesco, S. Percopo, M. Montico, A. Amoroso, S. Crovella, Evidence of a correlation between mannose binding lectin and celiac disease: a model for other autoimmune diseases, J. Mol. Med. 83 (2005) 308–315.
[17] T. Øhlenschlæger, P. Garred, H.O. Madsen, S. Jacobsen, Mannose-binding lectin variant alleles and the risk of arterial thrombosis in systemic lupus erythematosus, N. Engl. J. Med. 351 (2004) 260–267.
[18] S. Hansen, L. Selman, N. Palaniyar, K. Ziegler, J. Brandt, A. Kliem, M. Jonasson, M.O. Skjoedt, O. Nielsen, K. Hartshorn, T.J.D. Jørgensen, K. Skjødt, U. Holmskov, Collectin 11 (CL-11, CL-K1) is a MASP-1/-3-associated plasma collectin with microbial-binding activity, J. Immunol. 185 (2010) 6096–6104.
[19] J. Lu, P.N. Tay, O.L. Kon, K.B. Reid, Human ficolin, cDNA cloning, demonstration of peripheral blood leukocytes as the major site of synthesis and assignment of the gene to chromosome 9, Biochem. J. 313 (1996) 473–478.
[20] Y. Yae, S. Inaba, H. Sato, K. Okochi, F. Tokunaga, S. Iwanaga, Isolation and characterization of a thermolabile beta-2 macroglycoprotein ('thermolabile substance' or 'Hakata antigen') detected by precipitating (auto) antibody in sera of patients with systemic lupus erythematosus, Biochem. Biophys. Acta. 1078 (1991) 369–376.
[21] H. Ichijo, U. Hellman, C. Wernstedt, L.J. Gonez, L. Claesson-Welsh, C.H. Heldin, K.J. Miyazono, Molecular cloning and characterization of ficolin, a multimeric protein with fibrinogen- and collagen-like domains, J. Biol. Chem. 268 (1993) 14505–14513.
[22] S. Thiel, T. Vorup-Jensen, C.M. Stover, W. Schwaeble, S.B. Laursen, K. Poulsen, A.C. Willis, P. Eggleton, S. Hansen, U. Holmskov, K.B.M. Reid, J.C. Jensenius, A second serine protease associated with mannan-binding lectin that activates comple- ment, Nature 386 (1997) 506–510.
[23] M. Takahashi, Y. Endo, T. Fujita, M. Matsushita, 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 (1999) 859–863.
[24] S.E. Degn, A.G. Hansen, R. Steffensen, C. Jacobsen, J.C. Jensenius, S. Thiel, MAp44, a human protein associated with pattern recognition molecules of the complement system and regulating the lectin pathway of complement activation, J. Immunol. 183 (2009) 7371–7378.
[25] M.O. Skjoedt, T. Hummelshoj, Y. Palarasah, C. Honore, C. Koch, K. Skjodt, P. Garred, A novel mannose-binding lectin/ficolin-associated protein is highly expressed in heart and skeletal muscle tissues and inhibits complement activation, J. Biol. Chem. 285 (2010) 8234–8243.
[26] Y. Endo, M. Takahashi, M. Nakao, H. Saiga, H. Sekine, M. Matsushita, M. Nonaka, T. Fujita, Two lineages of Mannose-Binding Lectin-Associated Serine Protease (MASP) in vertebrates, J. Immunol. 161 (1998) 4924–4930.
[27] F. Takada, N. Seki, Y. Matsuda, Y. Takayama, M. Kawakami, Localization of the genes for the 100-kDa complement-activating component of Ra-reactive factor (CRARF and Craft) to human 3q27-28 and mouse 16B2-B3, Genomics 25 (1995) 757–759.
[28] C.M. Stover, N.J. Lynch, M.R. Dahl, S. Hanson, M. Takahashi, M. Frankenberger, L. Ziegler-Heitbrock, I. Eperon, S. Thiel, W.J. Schwaeble, Murine serine proteases MASP-1 and MASP-3, complements of the lectin pathway activation complex of complement, are encoded by a single structural gene, Gene Immun. 4 (2003) 374–384.
[29] T. Knittel, P. Fellmer, K. Neubauer, M. Kawakami, A. Grundmann, G. Ramadori, The complement activating protease P100 is expressed by hepatocytes and is induced by IL-6 in vitro and during the acute phase reaction in vivo, Lab. Invest. 77 (1997) 221–230.
[30] T. Fujita, Evolution of the lectin-complement pathway and its role in innate immunity, Nat. Rev. Immunol. 2 (2002) 346–353.
[31] X. Ji, K. Azumi, M. Sasaki, M. Nonaka, Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian *Halocynthia roretzi*, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 6340–6345.
[32] Y. Endo, M. Nonaka, H. Saiga, Y. Kakinuma, A. Matsushita, M. Takahashi, M. Matsushita, T. Fujita, Origin of mannose-binding lectin-associated serine protease (MASP)-1 and MASP-3 involved in the lectin complement pathway traced back to the invertebrate, amphioxus, J. Immunol. 170 (2003) 4701–4707.
[33] M. Matsushita, Y. Endo, T. Fujita, Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine-protease, J. Immunol. 164 (2000) 2281–2284.
[34] N.J. Lynch, S.H. Khan, C.M. Stover, S.M. Sandrini, J.S. Presanis, W.J. Schwaeble, Composition of the lectin pathway of complement in *Gallus gallus*: absence of mannan-binding lectin-associated serine protease-1 in birds, J. Immunol. 174 (2005) 4998–5006.
[35] J. Seyfarth, P. Garred, H.O. Madsen, Extra-hepatic transcription of the human mannose-binding lectin gene (*mbl2*) and the MBL-associated serine protease 1–3 genes, Mol. Immunol. 43 (2006) 962–971.
[36] M. Kuraya, M. Matsushita, Y. Endo, S. Thiel, T. Fujita, Expression of H-ficolin/Hakata antigen, mannose-binding lectin-associated serine protease (MASP)-1 and MASP-3 by human glioma cell line T98G, Int. Immunol. 15 (2003) 109–117.
[37] W. Schwaeble, M.R. Dahl, S. Thiel, C. Stover, J.C. Jensenius, The mannan-binding lectin-associated serine proteases (MASPs) and MAp19: four components of the lectin pathway activation complex encoded by two genes, Immunobiology 205 (2002) 455–466.
[38] M.M. Krem, E. Di Cerca, Molecular markers of serine protease evolution, EMBO J. 20 (2001) 3036–3045.
[39] G.J. Arlaud, J. Gagnon, C1r and C1s subcomponents of human complement: two serine proteases lacking the "histidine-loop" disulphide bridge, Biosci. Rep. 1 (1981) 779–784.
[40] F. Teillet, C. Gaboriaud, M. Lacroix, L. Martin, G.J. Arlaud, N.M. Thielens, Crystal structure of the CUB1-EGF-CUB2 domain of human MASP-1/3 and identification of its interaction sites with mannan-binding lectin and ficolins, J. Biol. Chem. 283 (2008) 25715–25724.
[41] M.J. Romao, I. Kolln, J.M. Dias, A.L. Carvalho, A. Romero, P.F. Varela, L. Sanz, E. Töpfer-Petersen, J.J. Calvete, Crystal structure of acid seminal fluid protein (aSFP)

at 1.9 Å resolution: a bovine polypeptide of the spermadhesin family, J. Mol. Biol. 274 (1997) 650–660.
[42] L.A. Gregory, N.M. Thielens, G.J. Arlaud, J.C. Fontecilla-Camps, C. Gaboriaud, X-ray structure of the Ca2 + − binding interaction domain of C1s. Insights into the assembly of the C1 complex of complement, J. Biol. Chem. 278 (2003) 32157–32164.
[43] H. Feinberg, J.C. Uitdehaag, J.M. Davies, R. Wallis, K. Drickamer, W.I. Weis, Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2, EMBO J. 22 (2003) 2348–2359.
[44] L.A. Gregory, N.M. Thielens, M. Matsushita, R. Sørensen, G.J. Arlaud, J.C. Fontecilla-Camps, C. Gaboriaud, 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 (2004) 29391–29397.
[45] Z. Rao, P. Handford, M. Mayhew, V. Knott, G.G. Brownlee, D. Stuart, The structure of a Ca^{2+}-binding epidermal growth factor-like domain: its role in protein-protein interactions, Cell 82 (1995) 131–141.
[46] N.M. Thielens, S. Cseh, S. Thiel, T. Vorup-Jensen, V. Rossi, J.C. Jensenius, G.J. Arlaud, Interaction properties of human mannan-binding lectin (MBL)-associ- ated serine protease-1 and -2, MBL-associated protein 19, and MBL, J. Immunol. 166 (2001) 5068–5077.
[47] C.B. Chen, R. Wallis, Stoichiometry of complexes between mannose-binding protein and its associated serine proteases: defining function units for complement activation, J. Biol. Chem. 276 (2001) 25894–25902.
[48] S. Cseh, L. Vera, M. Matsushita, T. Fujita, G.J. Arlaud, N.M. Thielens, Characterization of the interaction between L-ficolin/p35 and mannan-binding lectin-associated serine protease-1 and -2, J. Immunol. 169 (2002) 5735–5743.
[49] F. Teillet, M. Lacroix, S. Thiel, D. Weilguny, T. Agger, G.J. Arlaud, N.M. Thielens, Identification of the site of human mannan-binding lectin involved in the interaction with its partner serine proteases: the essential role of Lys55, J. Immunol. 178 (2007) 5710–5716.
[50] S. Zundel, S. Cseh, M. Lacroix, M.R. Dahl, M. Matsushita, J.P. Andrieu, W.L. Schwaeble, J.C. Jensenius, T. Fujita, G.J. Arlaud, N.M. Thielens, Characterization of recombinant mannan-binding lectin-associated serine protease (MASP)-3 suggests an activation mechanism difference from that of MASP-1 and MASP-2, J. Immunol. 172 (2004) 4342–4350.
[51] M.O. Skjoedt, Y. Palarasah, L. Munthe-Fog, Y.J. Ma, G. Weiss, K. Skjodt, C. Koch, P. Garred, MBL-associated serine protease-3 circulates in high serum concentrations predominantly in complex with ficolin-3 and regulates ficolin-3 mediated complement activation, Immunology 215 (2009) 921–931.
[52] J. Dobó, V. Harmat, L. Beinrohr, E. Sebestyén, P. Závodszky, P. Gál, MASP-1, a promiscuous complement protease: structure of its catalytic region reveals the basis of its broad specificity, J. Immunol. 183 (2009) 1207–1214.
[53] P.N. Barlow, M. Baron, D.G. Norman, A.J. Day, A.C. Willis, R.B. Sim, I.D. Campbell, Secondary structure of a complement control protein module by two-dimensional 'H NMR, Biochemistry 30 (1991) 997–1004.
[54] I. Schechter, A. Berger, On the size of the active site in protease. I. Papain, Biochem. Biophys. Res. Commun. 27 (1967) 157–162.
[55] V. Rossi, S. Cseh, I. Bally, N.M. Thielens, J.C. Jensenius, G.J. Arlaud, Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2, J. Biol. Chem. 276 (2001) 40880–40887.
[56] I. Terai, K. Kobayashi, M. Matsushita, T. Fujita, Human serum mannose-binding lectin (MBL)-associated serine protease-1 (MASP-1): determination of levels in body fluids and identification of two forms in serum, Clin. Exp. Immunol. 110 (1997) 317–323.
[57] S.E. Degn, L. Jensen, P. Gál, J. Dobó, S.H. Holmvad, J.C. Jensenius, S. Thiel, Biological variations of MASP-3 and MAp44, two splice products of the *MASP1* gene involved in regulation of the complement system, J. Immunol. Methods 361 (1–2) (2010) 37–50.
[58] M. Matsushita, T. Fujita, Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation, Immunology 194 (1995) 443–448.
[59] M. Matsushita, S. Thiel, J.C. Jensenius, I. Terai, T. Fujita, Proteolytic activities of two types of mannose-binding lectin-associated serine protease, J. Immunol. 165 (2000) 2637–2642.
[60] S.V. Petersen, S. Thiel, J.C. Jensenius, The mannan-binding lectin pathway of complement activation: biology and disease association, Mol. Immunol. 38 (2001) 133–149.
[61] B. Selander, U. Martensson, A. Weintraub, E. Holmstrom, M. Matsushita, S. Thiel, J.C. Jensenius, L. Truedsson, A.G. Sjoholm, Mannan-binding lectin activates C3 and the alternative complement pathway without involvement of C2, J. Clin. Invest. 116 (2006) 1425–1434.
[62] N.K. Wong, M. Kojima, J. Dobó, G. Ambrus, R.B. Sim, Activation of the MBL-associated serine proteases (MASPs) and their regulation by natural inhibitors, Mol. Immunol. 36 (1999) 853–861.
[63] G. Ambrus, P. Gál, M. Kojima, K. Szilágyi, J. Balczer, J. Antal, L. Gráf, A. Laich, B.E. Moffatt, W. Schwaeble, R.B. Sim, P. Závodszky, Natural substrates and inhibitors of mannan-binding lectin-associated serine protease-1 and -2: a study on recombinant catalytic fragments, J. Immunol. 170 (2003) 1374–1382.
[64] C.B. Chen, R. Wallis, Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases, J. Biol. Chem. 279 (2004) 26058–26065.
[65] M. Møller-Kristensen, S. Thiel, A. Sjöholm, M. Matsushita, J.C. Jensenius, Cooperatoin between MASP-1 and MASP-2 in the generation of C3 convertase through the MBL pathway, Int. Immunol. 19 (2007) 141–149.
[66] N. Rawal, R. Rajagopalan, V.P. Salvi, Activation of complement component C5: coparison of C5 convertases of the lectin pathway and classical pathway of complement, J. Biol. Chem. 283 (2008) 7853–7863.
[67] M. Takahashi, D. Iwaki, K. Kanno, Y. Ishida, J. Xiong, M. Matsushita, Y. Endo, S. Miura, N. Ishii, K. Sugamura, T. Fujita, Mannose-binding lectin (MBL)-associated serine protease (MASP)-1 contributes to activation of the lectin complement pathway, J. Immunol. 180 (2008) 6132–6138.
[68] A. Kocsis, K.A. Kékesi, R. Szász, B.M. Végh, J. Balczer, J. Dobó, P. Závodszky, P. Gál, G. Pál, Selective inhibition of the lectin pathway of complement with phage display selected peptides against mannose-binding lectin-associated serine protease (MASP)-1 and -2: significant contribution of MASP-1 to lectin pathway activation, J. Immunol. 185 (2010) 4169–4178.
[69] P. Gál, V. Harmet, A. Kocsis, T. Bián, L. Barna, G. Ambrus, B. Végh, J. Balczer, R.B. Sim, G. Náray-Szabó, P. Závodszky, A true autoactivating enzyme: structural insights into mannose-binding lectin-associated serine protease-2 activation, J. Biol. Chem. 280 (2005) 33435–33444.
[70] K. Hajela, M. Kojima, G. Ambrus, K.H. Wong, B.E. Moffatt, J. Ferluga, S. Hajela, P. Gál, R.B. Sim, The biological functions of MBL-associated serine proteases (MASPs), Immunobiology 205 (2002) 467–475.
[71] S.V. Petersen, S. Thiel, L. Jensen, T. Vorup-Jensen, C. Koch, J.C. Jensenius, Control of the classical and the MBL pathway of complement activation, Mol. Immunol. 37 (2000) 803–811.
[72] I. Terai, K. Kobayashi, M. Matsushita, T. Fujita, K. Matsuno, K. Okumura, α^2-Macroglobulin binds to and inhibits mannose-binding protein-associated serine protease, Int. Immunol. 7 (1995) 1579–1584.
[73] M. Takahashi, Y. Ishida, D. Iwaki, K. Kanno, T. Suzuki, Y. Endo, Y. Homma, T. Fujita, Essential role of mannose-binding lectin-associated serine protease-1 in activating of the complement factor D, J. Exp. Med. 207 (2010) 29–37.
[74] N.K. Banda, M. Takahashi, B. Levitt, M. Glogowska, J. Nicholas, K. Takahashi, G.L. Stahl, T. Fujita, W.P. Arend, V.M. Holers, Essential role of complement mannose-binding lectin-associated serine proteases-1/3 in the murine collagen antibody-induced model of inflammatory arthritis, J. Immunol. 185 (2010) 5598–5606.
[75] C.L. Cortesio, W. Jiang, Mannan-binding lectin-associated serine protease 3 cleaves synthetic peptide and insulin-like growth factor-binding protein 5, Arch. Biochem. Biophys. 449 (2006) 164–170.
[76] J.S. Presanis, K. Hajela, G. Ambrus, P. Gál, R.B. Sim, Differential substrate and inhibitor profiles for human MASP-1 and MASP-2, Mol. Immunol. 40 (2004) 921–929.
[77] P. Gál, J. Dobó, P. Závodszky, R.B.M. Sim, Early complement proteases: C1r, C1s and MASPs. A structural insight into activation and functions, Mol. Immunol. 46 (2009) 2745–2752.
[78] A. Krarup, R. Wallis, J.S. Presanis, P. Gál, R.B. Sim, Simultaneous activation of complement and coagulation by MBL-associated serine protease 2, PLoS One 2 (7) (2007) e623.
[79] T. Vorup-Jensen, S.V. Petersen, A.G. Hansen, K. Poulsen, W. Schwaeble, R.B. Sim, K.B.M. Reid, S.J. Davis, S. Thiel, J.C. Jensenius, Distinct pathways of Mannan- Binding Lectin (MBL) and C1-complex autoactivation revealed by reconstitution of MBL with recombinant MBL-associated serine protease-2, J. Immunol. 165 (2000) 2093–2100.
[80] V. Harmat, P. Gál, J. Kardos, K. Szilagyi, G. Ambrus, B. Végh, G. Naray-Szabo, P. Závodszky, The structure of MBL-associated serine protease-2 reveals that identical substrate specificities of C1s and MASP-2 are realized through different sets of enzyme–substrate interactions, J. Mol. Biol. 342 (2004) 1533–1546.
[81] P. Gál, G. Ambrus, Structure and function of complement activating enzyme complexes: C1 and MBL-MASPs, Curr. Protein Pept. Sci. 2 (1) (2001) 43–59.
[82] C.M. Stover, N.J. Lynch, S.J. Hanson, M. Windbichler, S.G. Gregory, W.J. Schwaeble, Organization of the MASP2 locus and its expression profile in mouse and rat, Mamm. Genome 15 (2004) 887–900.
[83] R. Wallis, R.B. Dodd, Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations, J. Biol. Chem. 275 (2000) 30962–30969.
[84] J. Stenflo, Y. Stenberg, A. Muryani, Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions, Biochim. Biophys. Acta 1477 (2000) 51–63.
[85] R. Huber, W. Bode, Crystal structure and analysis of two variants of trigonal trypsin: trigonal trypsin and PEG (Polyethylene glycol) trypsin and their comparison with orthorhombic and trigonal trypsinogen, FEBS Lett. 90 (1978) 265–269.
[86] M. Budayova-Spano, W. Grabarse, N.M. Thielens, H. Hillen, M. Lacroix, M. Schmidt, J.C. Fontecilla-Camps, G.J. Arlaud, C. Gaboriaud, Monomeric structures of the zymogen and active catalytic domain of complement protease C1r, Structure 10 (11) (2002) 1509–1519.
[87] E. Kenesi, G. Katona, L. Szilagyim, Structural and evolutionary consequences of unpaired cysteines in trypsinogen, Biochem. Biophys. Res. Commun. 309 (2003) 749–754.
[88] V. Rossi, F. Teillet, N.M. Thielens, I. Bally, G.J. Arlaud, 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 (5) (2005) 41811–41818.
[89] C.M. Stover, S. Thiel, M. Thelen, N.J. Lynch, T. Vorup-Jensen, J.C. Jensenius, W.J. Schwaeble, Two constituents of the initiation complex of the Mannan-Binding lectin activation pathway of complement are encoded by a single structural gene, J. Immunol. 162 (1999) 3481–3490.
[90] D. Iwaki, K. Kanno, M. Takahashi, Y. Endo, N.J. Lynch, W.J. Schwaeble, M. Matsushita, M. Okabe, T. Fujita, Small Mannose-Binding lectin-associated protein plays a regulatory role in the lectin complement pathway, J. Immunol. 177 (2006) 8626–8632.
[91] K.R. Mayilyan, J.S. Presanis, J.N. Arnold, R.B. Sim, Discrete MBL-MASP complexes show wide inter-individual variability in concentration: data from UK vs Armenian populations, Int. J. Immunopathol. Pharmacol. 19 (3) (2006) 567–580.

[92] R. Wallis, Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway, Immunobiology 212 (4–5) (2007) 289–299.
[93] A.St. Swierzko, M. Cedzynskia, I. Domzalska-Popadiuk, S.L. MacDonald, M. Borkowska-Klos, A.P.M. Atkinson, A. Szala, A. Jopek, J.C. Jensenius, M. Kawakami, J. Szczapa, M. Matsushita, J. Szemraj, M.L. Turner, D.C. Kilpatrick, Mannan-binding lectin-associated serine protease-2 (MASP-2) in a large cohort of neonates and its clinical associations, Mol. Immunol. 46 (2009) 1696–1701.
[94] M.I. Garcia-Laorden, J. Sole-Violan, F. Rodriguez de Castro, J. Aspa, M.L. Briones, A. Garcia-Saavedra, O. Rajas, J. Blanquer, A. Caballero-Hidalgo, J.A. Marcos-Ramos, J. Hernandez-Lopez, C. Rodriguez-Gallego, Mannose-binding lectin and mannose- binding lectin-associated serine protease 2 in susceptibility, severity, and outcome of pneumonia in adults, J. Allergy Clin. Immunol. 122 (2) (2008) 368–374.
[95] K. Stengaard-Pedersen, S. Thiel, M. Gadjeva, M. Møller-Kristensen, R. Sorensen, L.T. Jensen, A.G. Sjoholm, L. Fugger, J.C. Jensenius, Inherited deficiency of mannan-binding lectin-associated serine protease 2, N. Engl. J. Med. 349 (2003) 554–560.
[96] R. Sørensen, S. Thiel, J.C. Jensenius, Mannan-binding-lectin-associated serine proteases, characteristics and disease associations, Springer Semin. Immunopathol. 27 (2005) 299–319.
[97] H.V. Olesen, J.C. Jensenius, R. Steffensen, S. Thiel, P.O. Schiotz, The mannan-binding lectin pathway and lung disease in cystic fibrosis—dysfunction of mannan-binding lectin-associated serine protease 2 (MASP-2) may be a major modifier, Clin. Immunol. 121 (2006) 324–331.
[98] S. Sagedal, S. Thiel, T.K. Hansen, T.E. Mollnes, H. Rollag, A. Hartmann, Impact of the complement lectin pathway on cytomegalovirus disease early after kidney transplantation, Nephrol. Dial. Transplant. 23 (2008) 4054–4060.
[99] H. Ytting, J.C. Jensenius, I.J. Christensen, S. Thiel, H.J. Nielsen, Increased activity of the Mannan-Binding lectin complement activation pathway in patients with colorectal cancer, Scand. J. Gastroenterol. 7 (2004) 674–679.
[100] H. Ytting, I.J. Christensen, S. Thiel, J.C. Jensenius, H.J. Nielsen, Pre- and postoperative levels in serum of mannan-binding lectin associated serine protease-2—a prognostic marker in colorectal cancer, Hum. Immunol. 69 (2008) 414–420.
[101] J. Kocsis, T. Meszaros, B. Madaras, E.K. Toth, S. Kamondi, P. Gal, L. Varga, Z. Prohaszka, G. Fust, High levels of acute phase proteins and soluble 70 kDa heat shock proteins are independent and additive risk factors for mortality in colorectal cancer, Cell Stress Chaperones 16 (1) (2011) 49–55.
[102] H. Ytting, I.J. Christensen, R. Steffensen, J. Alsner, S. Thiel, J.C. Jensenius, U. Hansen, H.J. Nielsen, Mannan-Binding Lectin (MBL) and MBL-Associated Serine Protease 2 (MASP-2) genotypes in colorectal cancer, Scand. J. Immunol. 73 (2011) 122–127.
[103] A. Verma, A. Matta, N.K. Shukla, S.V.S. Deo, S.D. Gupta, R. Ralhan, Clinical significance of mannose-binding lectin-associated serine protease-2 expression in esophageal squamous cell carcinoma, Int. J. Cancer 118 (2006) 2930–2935.
[104] A. Zehnder, U. Fisch, A. Hirt, F.K. Niggli, A. Simon, H. Ozsahin, L.J. Schlapbach, R.A. Ammann, Prognosis in pediatric hematologic malignancies is associated with serum concentration of mannose-binding lectin-associated serine protease-2 (MASP-2), Pediatr. Blood Cancer 53 (1) (2009) 53–57.
[105] K.R. Mayilyan, J.N. Arnold, J.S. Presanis, A.F. Soghoyan, R.B. Sim, Increased complement classical and mannan-binding lectin pathway activities in schizophrenia, Neurosci. Lett. 404 (2006) 336–341.
[106] K.R. Mayilyan, D.R. Weinberger, R.B. Sim, The complement system in schizophrenia, Drug News Perspect. 21 (4) (2008) 200–210.

APPENDIX II

(PUBLICATION)

Molecular Determinants of the Substrate Specificity of the Complement-initiating Protease, C1r.

Wijeyewickrema LC, Yongqing T, Tran TP, Thompson PE, Viljoen JE, Coetzer TH, Duncan RC, Kass I, Buckle AM, and Pike RN.

***J Biol Chem*. 288(22):15571-80 (2013).**

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 22, pp. 15571–15580, May 31, 2013

Molecular Determinants of the Substrate Specificity of the Complement-initiating Protease, C1r*[S]

Received for publication, January 9, 2013, and in revised form, March 21, 2013 Published, JBC Papers in Press, April 15, 2013, DOI 10.1074/jbc.M113.451757

Lakshmi C. Wijeyewickrema[‡], Tang Yongqing[‡], Thuy P. Tran[‡], Phillip E. Thompson[§], Jacqueline E. Viljoen[¶], Theresa H. Coetzer[¶], Renee C. Duncan[‡1], Itamar Kass[‡], Ashley M. Buckle[‡], and Robert N. Pike[‡2]

From the [‡]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia, the [§]Monash Institute of Pharmaceutical Sciences, Parkville, Victoria 3052, Australia, and the [¶]Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg Campus), Private Bag X01, Scottsville 3209, South Africa

Background: Classical complement pathway activation depends on cleavage of inactive C1s by C1r.
Results: P2 Gln and P1' Ile residues in activation loop of C1s are crucial for activation by C1r.
Conclusion: Residues at P2 and P1' in cleavage position of C1s make important interactions with C1r active site.
Significance: Critical determinants identified for activation of the classical complement pathway.

The serine protease, C1r, initiates activation of the classical pathway of complement, which is a crucial innate defense mechanism against pathogens and altered-self cells. C1r both autoactivates and subsequently cleaves and activates C1s. Because complement is implicated in many inflammatory diseases, an understanding of the interaction between C1r and its target substrates is required for the design of effective inhibitors of complement activation. Examination of the active site specificity of C1r using phage library technology revealed clear specificity for Gln at P2 and Ile at P1', which are found in these positions in physiological substrates of C1r. Removal of one or both of the Gln at P2 and Ile at P1' in the C1s substrate reduced the rate of C1r activation. Substituting a Gln residue into the P2 of the activation site of MASP-3, a protein with similar domain structure to C1s that is not normally cleaved by C1r, enabled efficient activation of this enzyme. Molecular dynamics simulations and structural modeling of the interaction of the C1s activation peptide with the active site of C1r revealed the molecular mechanisms that particularly underpin the specificity of the enzyme for the P2 Gln residue. The complement control protein domains of C1r also made important contributions to efficient activation of C1s by this enzyme, indicating that exosite interactions were also important. These data show that C1r specificity is well suited to its cleavage targets and that efficient cleavage of C1s is achieved through both active site and exosite contributions.

Complement activation represents a crucial innate defense mechanism against invading microorganisms, providing an immediate response against microbial invasion (1). It also plays a vital role in the maintenance of immune tolerance and, because it can also target altered-self structures, is a key player in tissue homeostasis through clearance of apoptotic and necrotic cells. The complement system can be activated by the classical, lectin, or alternative pathway. The C1r[3] protease is responsible for the first enzymatic events in the classical pathway of complement activation, through autoactivation and subsequent initiation of the cascade by cleaving and activating pro-enzyme C1s (2). The lectin pathway is activated by the MASP-1 and MASP-2 enzymes, whereas MASP-3, a splice variant of MASP-1, plays a presently less well characterized role in the system. The complement system is strongly implicated in many inflammatory disease states, and therefore inhibitors of the initiating proteases could be powerful anti-inflammatory agents (3). Understanding how C1r interacts with its target substrates is the key to the knowledge required to design effective inhibi- tors of complement activation by targeting this initiating enzyme.

It has previously been shown that C1r and C1s pro-enzymes form a heterotetrameric structure that associates with the rec- ognition molecule, C1q, in the C1 complex (4). Binding of C1q to target ligands, such as antigen-bound antibodies, causes autoactivation of C1r by an unknown mechanism (5). It has been postulated that binding of the multiple C1q recognition sites to their ligands essentially transmits a mechanical signal to the heterotetrameric protease structure that loosens the con- straints on C1r that prevent its autoactivation in the C1 com- plex. The activated C1r molecule is then able to cleave and activate the associated C1s enzyme, allowing it to in turn cleave its C4 and C2 physiological substrates in sequence to thus activate the complement system (6).

The C1r, C1s enzymes, and MASP enzymes are composed of six domains (Fig. 1) (7): the N-terminal CUB1-EGF-CUB2 seg- ment of the structure contains calcium binding sites and is most likely required to bind to recognition molecules such as C1q and for homo- and heterodimerization of the enzymes. The

* This work was supported by a program grant from the National Health and Medical Research Council of Australia.

[S] This article contains supplemental Movies S1 and S2 and an additional reference.

[1] Present address: Center for Virology, Burnet Institute, Melbourne, VIC, Australia.

[2] To whom correspondence should be addressed. Tel.: 61-3-99029300; Fax: 61-3-99029500; E-mail: rob.pike@monash.edu.

[3] The abbreviations used are: C1r, complement subcomponent 1r; AMC, 7-amino-4-methyl coumarin; C1, complement component 1; C1q, complement subcomponent 1q; C1s, complement subcomponent 1s; CCP, complement control protein; CUB, module originally found in complement C1r/C1s, Uegf, and bone morphogenetic protein; MASP, mannose-binding lectin-associated serine protease; SP, serine protease.

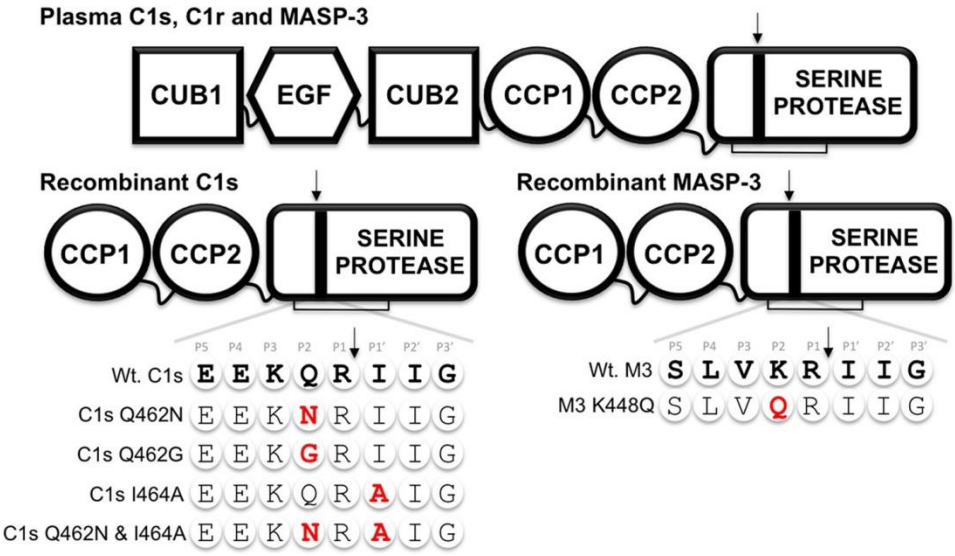


FIGURE 1. **C1s and MASP-3 constructs used in this study.** Schematic represents the domain structures of plasma and recombinant C1r, C1s, and MASP-3. The two CUB domains, an EGF-like domain, two CCP domains, and the SP domain are indicated. The *black arrow* indicates the Arg-Ile cleavage site, which is required for activation of the enzyme, held together by a disulfide bond. The amino acid sequence surrounding the activation point of wild type MASP-3 and C1s are in *bold type*. The amino acids altered by site-directed mutagenesis are shown in *red*.

C-terminal CCP1-CCP2-SP segment is responsible for catalysis of both a neighboring C1r molecule (autoactivation) and C1s. The CCP1 domain has been shown to play a major role in dimerization of C1r but no role in catalytic processes, whereas the CCP2 domain apparently provides an additional binding site for substrate C1r that facilitates catalysis of this substrate by the active site of the SP domain (8, 9). The precise active site specificity of C1r has never been mapped, and the relative roles played by the active site and proposed exosites in the catalysis of substrates have not previously been determined. Here we have used phage display technology to map the substrate specificity of the enzyme and then validated the data obtained using acti- vation of the C1s protein substrate as a molecular "readout." The data obtained and the analysis performed, including molecular dynamics simulations, indicate that the P2 Gln resi- due in substrates plays a vital role in catalysis by C1r, in addition to important contributions by exosite(s) most likely contained on the C1r CCP2 domain.

EXPERIMENTAL PROCEDURES

Construction of Recombinant Plasmids for Expression of the C1r, C1s, and MASP-3 Fragments—Recombinant C1r CCP12SP (residues Arg^{296}–Asp^{705}), recombinant C1r SP (res- idues Pro^{449}–Asp^{705}), recombinant C1s CCP12SP (residues Lys^{281}–Asp^{688}), recombinant C1s SP (residues Pro^{423}–Asp^{688}), and recombinant MASP-3 CCP12SP (residues Lys^{298}–Arg^{728}) were expressed and refolded with some modifications to previ- ously described methods (10, 11). Briefly, genes for all recom- binant proteins were synthesized (GenScript), and the DNA was cloned into the pET17b vector (EMD Biosciences). After transformation of the vector into *Escherichia coli* strain BL21(DE3)pLysS, cells were cultured at 37 °C in 2XTY (tryp- tone/yeast extract) broth with 50 μg/ml ampicillin and 34 μg/ml chloramphenicol to an A_{595} of 0.6, followed by induction with 1 mM isopropyl ⍺-D-thiogalactopyranoside for 4 h. Following induction, the culture was centrifuged (27,000 × *g*, 20 min, 4 °C), and the cells were collected in 30 ml of 50 mM Tris-HCl, 20 mM EDTA, pH 7.4, and then frozen at -80 °C. The cells were thawed and sonicated on ice for 6 × 30 s. After centrifugation at 27,000 × *g* for 20 min, inclusion body pellets were sequentially washed and centrifuged with 10 ml of 50 mM Tris-HCl, 20 mM EDTA, pH 7.4. The washed pellet was resuspended in 10 ml of 8 M urea, 0.1 M Tris-HCl, 100 mM DTT, pH 8.3, at room temperature for 3 h. Refolding was initiated by rapid dilution dropwise into 50 mM Tris-HCl, 3 mM reduced glutathione, 1 mM oxidized glutathione, 5 mM EDTA, and 0.5 M arginine, pH 9.0. The renatured protein solutions were concentrated and dialyzed against 50 mM Tris-HCl, pH 9.0, and renatured proteins were purified on a 5-ml Q-Sepharose Fast Flow column (GE Healthcare). The bound protein was eluted with a linear NaCl gradient from 0 to 400 mM over 35 ml at 1 ml/min. The recombinant proteins were further purified using a Superdex 75 16/60 column (GE Healthcare) in a buffer of 50 mM Tris, 145 mM NaCl, pH 7.4, aliquoted, snap frozen, and maintained at -80 °C. The purity of the protein was confirmed by SDS-PAGE followed by Western blotting and N-terminal sequencing. Typically protein yields were between 2 and 4 mg/liter.

Western Blotting and Antibodies—Proteins were resolved by SDS-PAGE, transferred, and immunoblotted with various antibodies. The antibodies used were polyclonal C1r (Abcam), a C1s antibody directed against the unique peptide sequence CSTSVQTSRLAKSKM, and a MASP-3 antibody directed against the unique peptide sequence NPNVTDQIISSGTRT. The latter antibodies were raised in chickens as described pre- viously (12).

Phage Display—The Novagen T7Select1-1b Phage Display system was used to generate a randomized substrate peptide library as described previously (13, 14), following the approach of Cwirla *et al.* (15). Amino acid peptides were displayed in low copy number (0.1–1/phage) from the T7Select1-1b vector used, making them suitable for the selection of displayed pep-

tides that were highly susceptible to protease cleavage. As described previously (14), the substrate library was constructed by synthesizing a degenerate oligonucleotide, annealing it to complementary half-site oligonucleotides, ligating the result- ing heteroduplex to vector arms and adding to a T7 phage pack- aging extract. The half-site oligonucleotides were 5'-GCCGC-CTGGAGTGAGAG-3' and 5'-AGCTTAGTGATGGT-GATGGTGATG-3'. This library was made by using the degenerate oligonucleotide 5'-AATTCTCTCACTCCAGG-CGGC-(NNK)9CATCACCATCACCATCACA-3' (where N represents any nucleotide and K is either T or C). This added a randomized unconstrained nonameric peptide (apart from a fixed arginine residue at the fifth position) and a His_6 tag to the C terminus of the 10B coat protein. The complexity of this randomized library was 7×10^6 plaque-forming units (pfu).

Approximately 10^9 pfu of amplified phage in phage extraction buffer were bound to nickel-Sepharose beads at 4 °C. Unbound phage were removed by washing the beads with phage wash buffer (850 mM NaCl, 0.1% (w/v) Tween 20 in PBS), followed by 1 mM $MgSO_4$ in PBS. Selection commenced by the addition of 500 nM human C1r (C1r purified from human plasma (EMD Biosciences) to the treatment tubes for rounds 1– 6 of selection. Equal volumes of 1 mM $MgSO_4$ in PBS were added to the control tube instead of protease. Both the treatment and control tubes were incubated overnight at 37 °C. Cleaved phage were recovered from the supernatant and subsequently titrated and amplified to form the sublibrary for the next round of selection. Phage that remained bound to the beads were eluted with 0.5 M imidazole and titrated to assess protease cleavage efficiency. Randomly selected individual phage plaques from round 4 were chosen for DNA sequencing. Phage DNA was amplified by PCR using dedicated primers (T7Select cloning kit; Novagen). Sequencing of PCR products using the same primers was performed using the Big-Dye 3.1 kit (GE Healthcare).

The sequencing results were analyzed to determine the statistical distribution of each amino acid at each position of the nonamer (16). This analysis allowed for codon redundancy, as well as the fact that only 32 of a possible 64 codons were represented by NNK. In the following equation, Licr indicates the difference of the observed frequency from the expected frequency in terms of standard deviations

$$\mathrm{Licr} = \mathrm{Obs}(X) - nP(X)/(nP(X)[1 - P(X)])1/2 \quad \text{(Eq. 1)}$$

where Obs(X) is the number of times amino acid X occurs in the selected sequences, P(X) is the theoretical probability of amino acid X occurring, and n is the total number of sequences analyzed.

Measurement of the Kinetics of Activation of Zymogen Proteases by C1r—Recombinant C1r at 100 nM was added to varying concentrations of zymogenic C1s or MASP-3 in 20 mM Tris-HCl, 100 mM NaCl, pH 7.4, and 50 μM Z-Leu-Gly-Arg-AMC (for C1s measurements (LGR-AMC)) or Z-Val-Pro-Arg-AMC (for MASP-3 measurements (VPR-AMC)) previously preincubated at 37 °C for 10 min. The appearance of fluorescence was measured using excitation and emission wavelengths of 355 and 460 nm, respectively. C1r was not active against the peptide substrates at the concentrations used, and therefore the increase in fluorescence seen was due entirely to the activity of C1s or MASP-3 activated by C1r.

The observed increase in fluorescence over time was fitted to an equation for exponential increase by nonlinear regression in GraphPad Prism: $Y = Y0*\exp(k_{obs}*X)$. This gave a k_{obs} value that equated to the observed rate of increase in fluorescence. The k_{obs} values obtained at different concentrations of substrate (zymogen C1s or MASP-3) were plotted against the substrate concentrations to yield a Michaelis-Menten plot that could be fitted by nonlinear regression in GraphPad Prism to the equation: $Y = V_{max}*[S]/(K_m + [S])$. The V_{max} values obtained from this analysis were converted into k_{cat} values by taking into account the k_{cat} of activated C1s or MASP-3 for the cognate reporter substrate used for each enzyme to yield estimates of product formation at the V_{max}values obtained The Michaelis-Menten plots could thus be used to derive K_m, k_{cat}, and k_{cat}/K_m values for the reaction of C1r with the zymogen forms of C1s and MASP-3.

Molecular Modeling and Dynamics—Missing residues (493–497) in the C1r x-ray crystal (Protein Data Bank ID code 1MD8 (7) were modeled using Modeller version 9.8 (17) The model was then superimposed onto the kallikrein chain of the kallikrein-hirustasin complex (18) using PyMOL version 1.3r2 (19). Peptides EEKQRIILG, EEKNRIILG, EEKQRAILG, EEKGRIILG, and EEKNRAILG were threaded onto the hirustasin coordinates, resulting in two C1r-peptide complexes. Each complex was then placed in a cubic unit cell with a minimum distance of 1.4 nm to the box edge and solvated in explicit SPC water (20). To neutralize the system at a physiological salt concentration of 0.1 M, Cl^- and Na^+ ions were randomly replaced with water molecules.

All models were then subjected to energy minimization with the conjugate gradient algorithm and a tolerance of 100 kJ $mol^{-1}{\cdot}nm^{-1}$. Following the EM stage, systems were subjected to a positional restraints procedure in which a harmonic restraint was applied to all heavy atoms in C1r and bound peptides. In the procedure, the restraint was gradually decreased from 1000 to 0 kJ $mol^{-1}{\cdot}nm^{-1}$ during 0.5-ns simulations. All models were then subjected to a 100-ns-long molecular dynam- ics simulation, each repeated three times with different random initial velocities. All simulations and trajectories analysis were conducted using the GOMACS package version 4.0.7 in conjunction with the GROMOS 53A6 united atom force field (21). During the energy minimization, the lengths of all bonds within the system were constrained using the LINCS algorithm (22). Nonbonded interactions were evaluated using a twin range cut- off scheme: interactions falling within the 0.8-nm short range cutoff were calculated every step, whereas interactions within the 1.4-nm-long cutoff were updated every three steps, together with the pair list. A reaction-field correction was applied to the electrostatic interactions beyond the long range cutoff (23), using a relative dielectric permittivity constant of $\varepsilon_{RF} = 62$ as appropriate for SPC water (24). Temperature and pressure were kept constant during simulations using the Berendsen coupling algorithm (25). Temperature was maintained at 300 K by independently coupling both protein and solvent to external temperature baths with a coupling constant of $\tau = 0.1$ ps. The pressure was maintained at 1 bar by weakly coupling the system

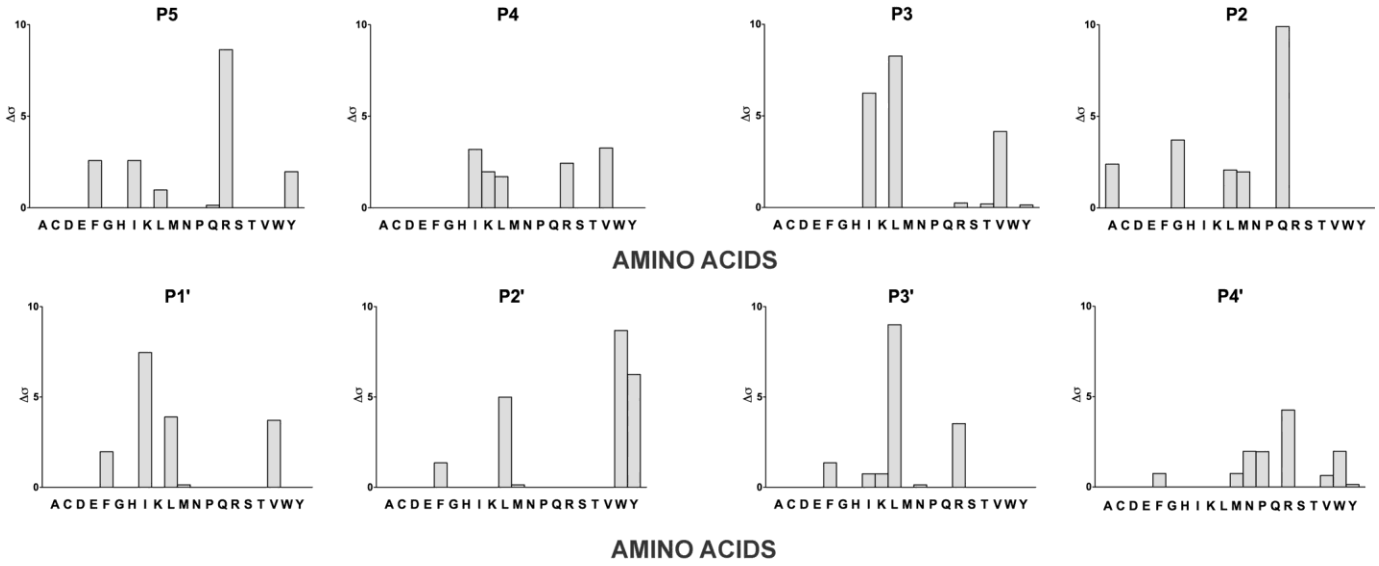


FIGURE 2. **Subsite profiling of human C1r using a phage display library with a fixed P1 arginine.** A library of peptides exploring the P5–P4' positions was exposed to 500 nM human C1r over six rounds of panning. Sequences of phage cleaved by C1r were analyzed, yielding Licr values for each subsite. The Licr values represent the number of S.D. away from an expected "normal" to identify overrepresentation of particular amino acids at any given substrate position.

to an isotropic pressure bath, using an isothermal compressibility of $4.6 \times 10{-}5$ bar^{-1} and a coupling constant of $\tau_P = 1$ ps. The electrostatic potentials were calculated using APBS version 1.3 (26). Atomic parameters for the calculation were taken from the GROMOS 53A6 force field (21). Electrostatic potential was visualized using PyMOL version 1.3r2 (19) with positive potential in blue and negative potential in red in a range between −1 and +1 k_bT/e_c, where k_b is the Boltzmann constant, T is the temperature (set to 300 K), and e_c is electron charge.

RESULTS

Analysis of the Active Site Specificity of C1r Using Phage Display Technology—The specificity of C1r for positively charged residues at the P1 position of physiological substrates was known (7), and therefore the phage library was constructed with an Arg residue fixed at the fifth position in the randomized sequence. A large concentration of the enzyme had to be used to obtain selection with the library. Six rounds of panning using cleavage by C1r were conducted, with the titer of the protease- selected sublibrary increasing at each round until round five. 94 samples were selected for sequencing, with 29 viable sequences obtained. Once adjusted statistically, the results (Fig. 2) clearly revealed that the enzyme displays considerable specificity at every position apart from P4, P3', and P4'. The most significant results (Licr $>$ 5) were: Gln at P2 (Licr = 6.4), Leu at P3 (Licr = 5.8) (Ile was nearly as high as Leu with Licr = 4.3), Ile at P1' (Licr = 5.3) (Val is quite significant with Licr = 3.9), and Tyr at P2' (Licr = 6.4), with Trp also significant at this position (Licr = 4.3). The presence of Arg residues at P5 was significant (Licr = 6.4), although it must be noted that this position is close to the phage capsid. Of the residues identified to be important at each position, it was notable that the preference for Gln residues at P2 and Ile residues at P1' matched that of the physiological sub- strates for C1r, particularly those found in zymogen C1s.

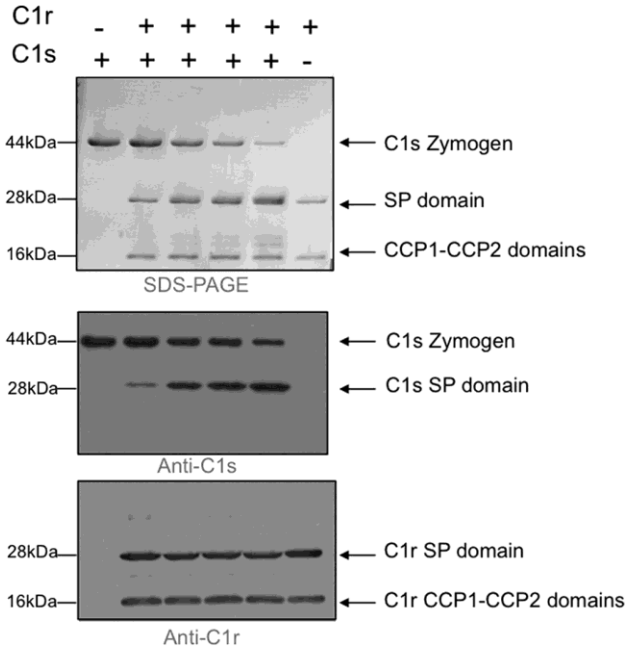


FIGURE 3. **C1r cleavage of zymogen C1s.** Recombinant wild type zymogen C1s (1 μM) was incubated with recombinant wild type C1r (10 nM) and sub- jected to SDS-PAGE, with subsequent immunoblot analysis using antibodies raised against a SP domain peptide of C1s (*middle panel*) and a polyclonal C1r antibody (*bottom panel*). Each *panel* of the experiment shows a time course (0, 1, 5, 10, and 30 min) for incubation at 37 °C (*second* through *fifth lanes*).

Analysis of the Importance of Cleavage Site Residues to Catal- ysis by C1r—Having noted that Gln and Ile residues at P2 and P1', respectively, were important for cleavage of phage dis- played substrates, we set out to confirm the importance of these residues for cleavage by C1r. A series of fluorescence-quenched peptide substrates, comprising residues from the cleavage site in C1s and some peptides found among the phage-displayed peptides, were synthesized. Unfortunately, very high concen- trations of C1r were required to cleave such peptides, and

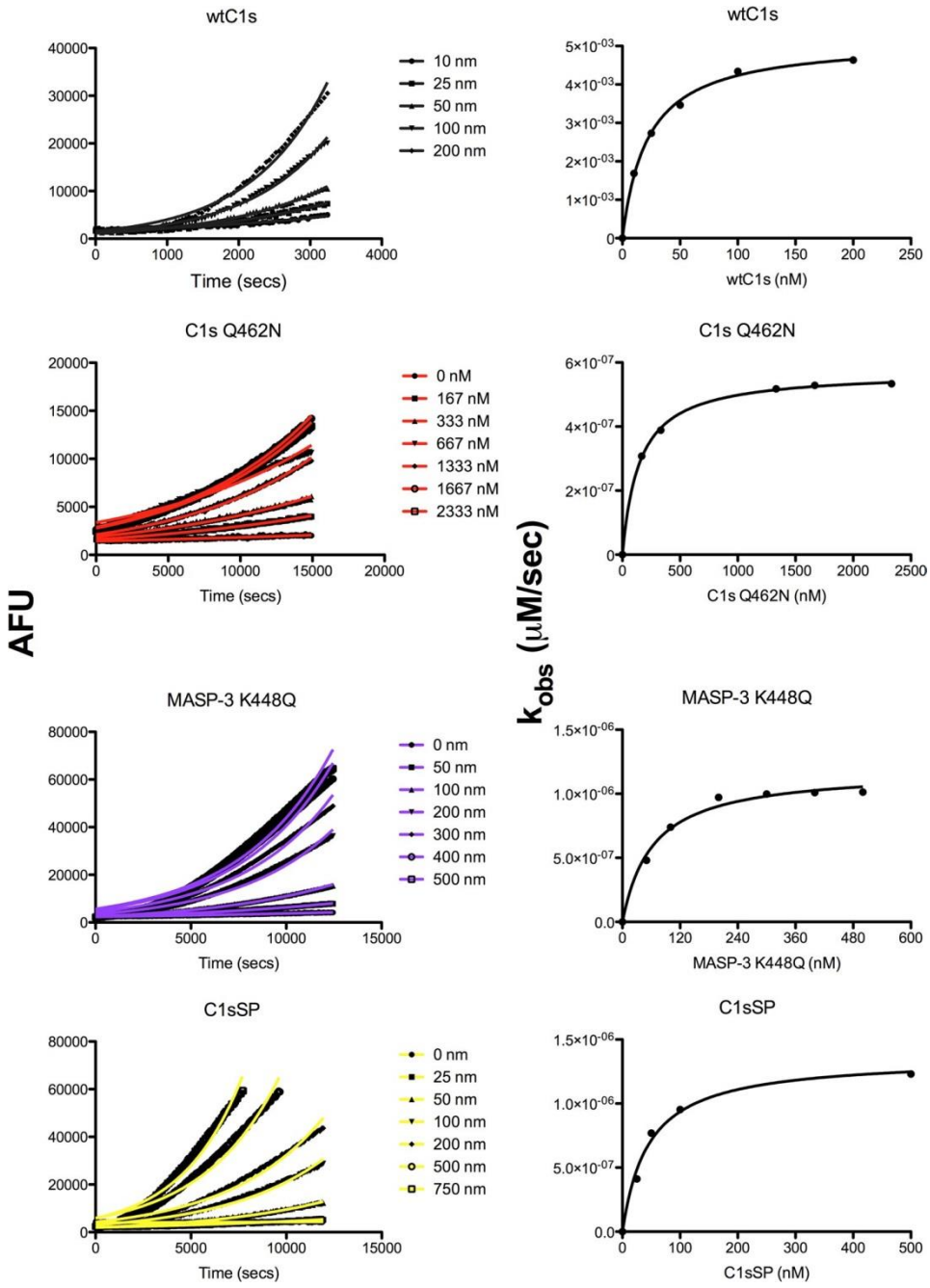


FIGURE 4. **Kinetics of activation of C1s and MASP-3 mutants by C1r.** *Left panels*, progress curves of C1r (100 nM) activation of indicated concentrations of WT C1s, Q462N C1s CCP12SP, MASP-3 K448Q, and C1s SP in the presence of either 50 μM LGR-AMC (for C1s) or VPR-AMC (for MASP-3). *AFU*, arbitrary fluorescence units. *Right panels*, plots of the observed rate of activation (k_{obs}) as a function of the concentration of substrate.

erratic results were recorded for kinetic analyses. This indicated that the enzyme displayed poor activity in general against peptide substrates, and therefore another means had to be found to investigate cleavage by C1r.

We therefore decided to use a CCP1-CCP2-SP form of zymogen C1s to test for cleavage by recombinant C1r and C1r purified from human plasma. We could show that the protein substrate was efficiently cleaved by both forms of C1r, making this a better means of testing the specificity determinants for C1r cleavage (Fig. 3). We therefore constructed mutants of zymogenic C1s in which the P2 and P1' positions were altered (Fig. 1) and tested the kinetics of cleavage of the substrates using a coupled assay in which activation of the C1s was revealed by measuring its activity against the peptide substrate, LGR-AMC (Fig. 4). The C1r had no activity against the peptide substrate at the concentrations of enzyme used. Varying the concentration of the C1s substrate allowed the effect of substrate concentration on the observed rate of activation of C1s (k_{obs}) to be measured. The curves of fluorescence obtained could be fitted to an exponential function (Fig. 4). Plots of the k_{obs} values obtained *versus* the substrate concentration could be fitted to the Michaelis-Menten equation (Fig. 4), thus allowing K_m and k_{cat} values to be estimated once the kinetics of cleavage of the peptide substrate by C1s was taken into account (Table 1).

TABLE 1

Kinetic parameters for cleavage of wild type and mutant forms of C1s and MASP-3 by C1r CCP12SP and C1r SP enzymes

Cleavage of the wild type and mutant forms of C1s and MASP-3 was followed by monitoring the appearance of their activity using fluorescent substrates, and data were fitted to allow the determination of the kinetic parameters shown below.

Substrate	K_m	k_{cat}	k_{cat}/K_m
	nM	s^{-1}	$M^{-1}s^{-1}$
Substrate with C1r CCP1-CCP2SP			
Wild type C1s CCP1-CCP2SP	22.0	6.45×10^{-2}	2.9×10^{6}
C1s CCP1-CCP2SP Q462N	148	6.3×10^{-3}	4.2×10^{4}
C1s CCP1-CCP2SP Q462G	42.6	3.9×10^{-2}	9.1×10^{5}
C1s CCP1-CCP2SP I464A	38.8	4.75×10^{-3}	1.2×10^{5}
C1s CCP1-CCP2SP Q462N and I464A	445	ND[a]	ND
C1s SP	45.4	3.25×10^{-2}	7.2×10^{5}
Wild type MASP-3 CCP1-CCP2SP	237	ND	ND
MASP-3 K448Q	62.2	2.1×10^{-2}	3.4×10^{5}
Substrate with C1r SP			
Wild type C1s CCP1-CCP2SP	191	7.4×10^{-3}	3.9×10^{4}
C1s SP	1730	3.3×10^{-2}	1.9×10^{4}

[a] ND, not determined because rates of cleavage were too low to allow determination of these parameters.

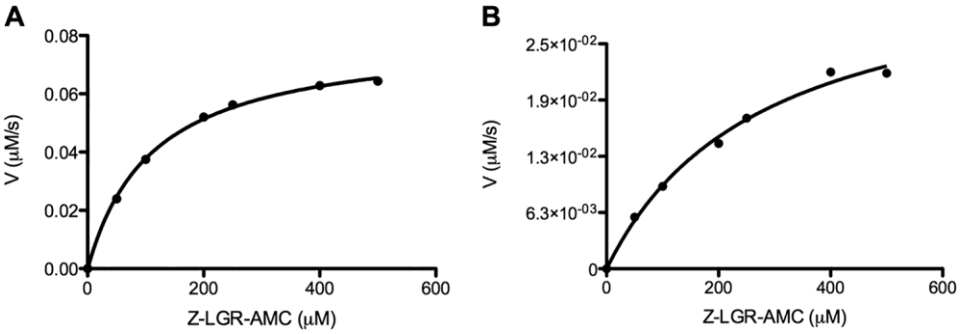


FIGURE 5. **Dependence of initial velocity on substrate concentration for wild type C1s and C1s I464A.** Initial velocity values for cleavage of Z-LGR-AMC (0–500 μM) by 100 nM wild type C1s (*A*) and C1s I464A (*B*), to the substrate are shown. Data were fitted by nonlinear regression to the Michaelis-Menten equation.

The data obtained showed that wild type zymogen C1s was very efficiently cleaved by C1r, with a very low K_m value of 22 nM and an overall k_{cat}/K_m value of 2.9×10^{6} $M^{-1}\cdot s^{-1}$. Substitution of the P2 Gln residue of zymogen C1s by a chemically similar Asn residue (Q462N) resulted in a 69-fold decrease in the k_{cat}/K_m value for the reaction, strongly influenced by a 7-fold increase in the K_m value. Interestingly, substitution of the P2 Gln by a Gly residue (Q462G) brought about a much smaller 3-fold decrease in the k_{cat}/K_m value, indicating that the substi- tution by the Asn residue was especially detrimental at this position.

Altering the P1' Ile residue to an Ala residue (I464A) also had a strong effect on the k_{cat}/K_m value for the reaction, decreasing it 24-fold, mainly due to a decrease in the k_{cat} value for the reaction (14-fold). Interestingly, the activated C1s with an Ala residue at the new N terminus was still active against the peptide substrate, with the k_{cat}/K_m value of 1.2×10^{3} $M^{-1}\cdot s^{-1}$ for the mutant only 6-fold lower than that for wild type C1s (7.1×10^{3} $M^{-1}\cdot s^{-1}$) (Fig. 5), indicating that the Ala residue was able to substitute for Ile at this crucial position. It is worth noting that substituting the Ile at the new N terminus of activated trypsin with an Ala residue resulted in a similar reduction in activity against most peptide substrates tested (27). Alteration of both the P2 and P1' residues simultaneously resulted in a form of C1s which C1r was only able to cleave very weakly, such that k_{cat} residues could not be estimated and the K_m value was increased >20-fold.

These data indicated that the Gln residue found at the P2 position of the physiological substrates was of high importance for efficient cleavage by C1r. To further verify this, we substituted the Lys residue found at the P2 position of zymogen MASP-3 with a Gln residue (K448Q) (Fig. 1) and investigated whether C1r, which was essentially unable to cleave wild type MASP-3 (Fig. 5), could efficiently activate this protease with a similar domain structure to C1s. Interestingly, the mutated MASP-3 was efficiently activated by C1r (Fig. 6), with a k_{cat}/K_m value only 8.5-fold lower than that for wild type C1s zymogen (Fig. 4 and Table 1). The k_{cat} value was comparable with that found for C1s, whereas the K_m value was nearly 3-fold higher.

Investigation of the Relative Effect of CCP Domains on Cleav- age by C1r—These data indicate that the residues found in the activation loop of the zymogens capable of being activated by C1r play a major role in recognition of the active site of C1r and in turn validate the results demonstrated using phage display technology. It appears that there was still some activation of C1s occurring even when both the important P2 and P1' residues were altered, however, indicating that other parts of the enzyme might be playing a role in recognizing cognate physiological substrates of the enzyme. It has previously been demonstrated that the CCP domains of C1r play an important role in the recognition of substrate C1r molecules in the autoactiva- tion reaction (8). We therefore set out to determine the importance of such exosites on the body of the protease by examining the effect of eliminating the CCP domains from the C1s substrate, reasoning that these domains may be playing an important role in the recognition of the substrate protein. We found that the C1s SP domain alone was activated with a 4-fold lower k_{cat}/K_m value than that found for the CCP1-CCP2-SP

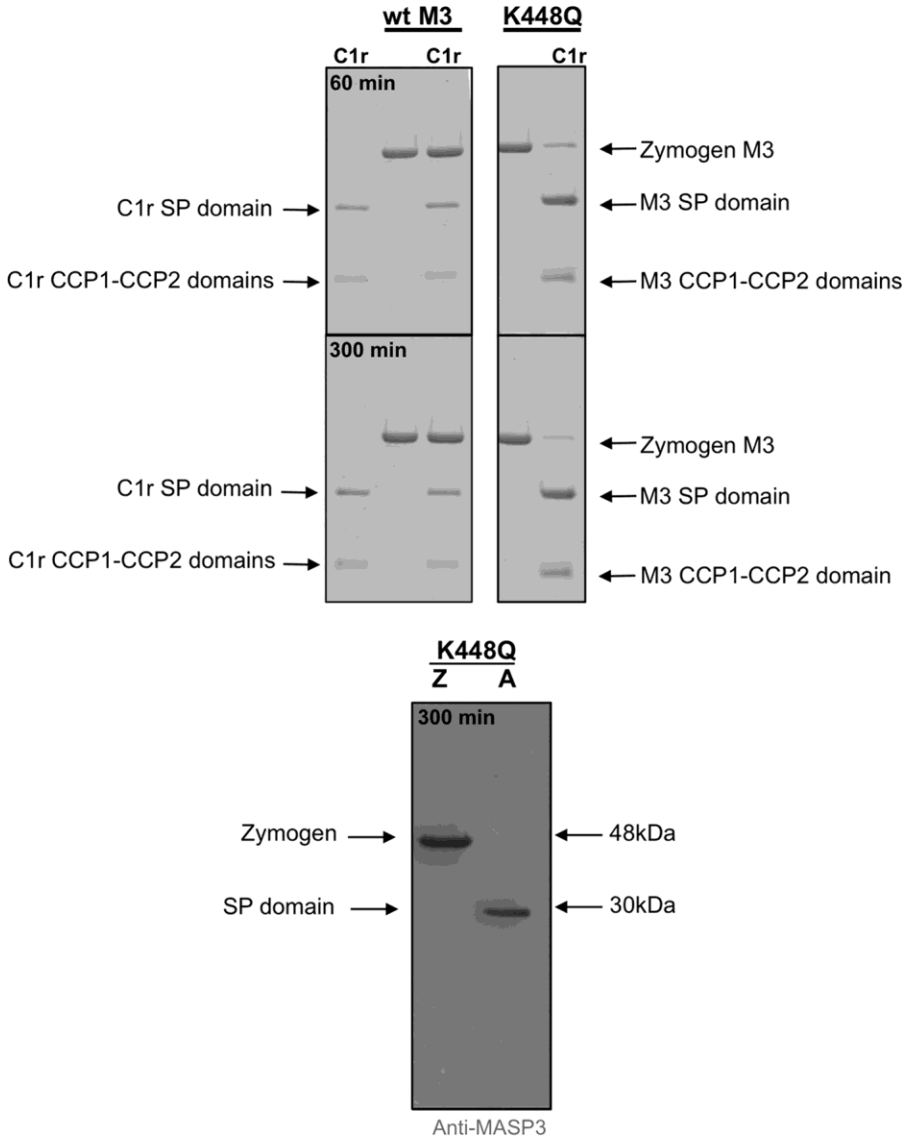


FIGURE 6. **C1r cleavage of wild type and mutant MASP-3.** *Top panel*, zymogen forms of wild type MASP-3 and K448Q MASP-3 were incubated with or without C1r at 37 °C for 60 or 300 min, as indicated, and samples were analyzed by SDS-PAGE. *Bottom panel*, cleavage of the K448Q mutant after 300 min was further analyzed by immunoblotting using antibodies raised against a MASP-3 SP domain peptide.

form of the substrate, due to changes in both the K_m and k_{cat} values for the reaction (Fig. 4 and Table 1). Removal of the CCP domains from C1r had a much stronger effect on the interaction, decreasing the k_{cat}/K_m value 74-fold. Removal of the CCP domains from the C1s substrate resulted in only a further 2-fold decrease in the k_{cat}/K_m value, indicating that the CCP domains of C1r played the dominant role in recognition of the substrate protein.

Molecular Dynamics Simulations of the Interaction between C1r and Zymogen C1s—Both phage display libraries and kinetic studies have shown the importance of Gln and Ile residues at position P2 and P1', respectively, for substrate cleavage by C1r. To gain a structural and dynamic insight into C1r-substrate interactions, we modeled a series of five nona-peptides bound to C1r and subjected them to molecular dynamics simulations. Snapshots after 100 ns of simulations indicate that whereas the WT C1s peptide maintains the modeled canonical and stable conformation in the active site (Fig. 7*A*), other peptides, representing mutations at P2 and P1', undergo rapid conformational fluctuations, resulting in the loss of most protein-peptide interactions (Fig. 7, *B–F*, and supplemental Movies S1 and S2). The calculated root mean square deviation for the last 90 ns of simulation for the peptide similar to WT C1s was 0.22 ± 0.05 nm (EEKQRIILG), significantly lower than the root mean square deviation calculated for the other peptides: 0.62 ± 0.09 nm (EEK**N**RIILG), 0.52 ± 0.07 nm (EEKQR**A**ILG), 0.56 ± 0.09 nm (EEK**G**RIILG), and 0.53 ± 0.1 nm (EEK**N**R**A**ILG). The simulations provide a simple and physically straightforward molecular insight into the differences in measured K_m and k_{cat}. The muta- tion Ile3 Ala at P1' causes a weakening of its interactions at S1'. This is also the case for the mutation Gln3 Gly at position P2, in which an amide group is removed, resulting in a loss of interactions between P2 and S2. In contrast, the mutation Gln3 Asn at P2 results in conformational change of the side chain and subsequent loss of interactions with S2 (Fig. 8). Modeling and molecular dynamics simulations suggest that the

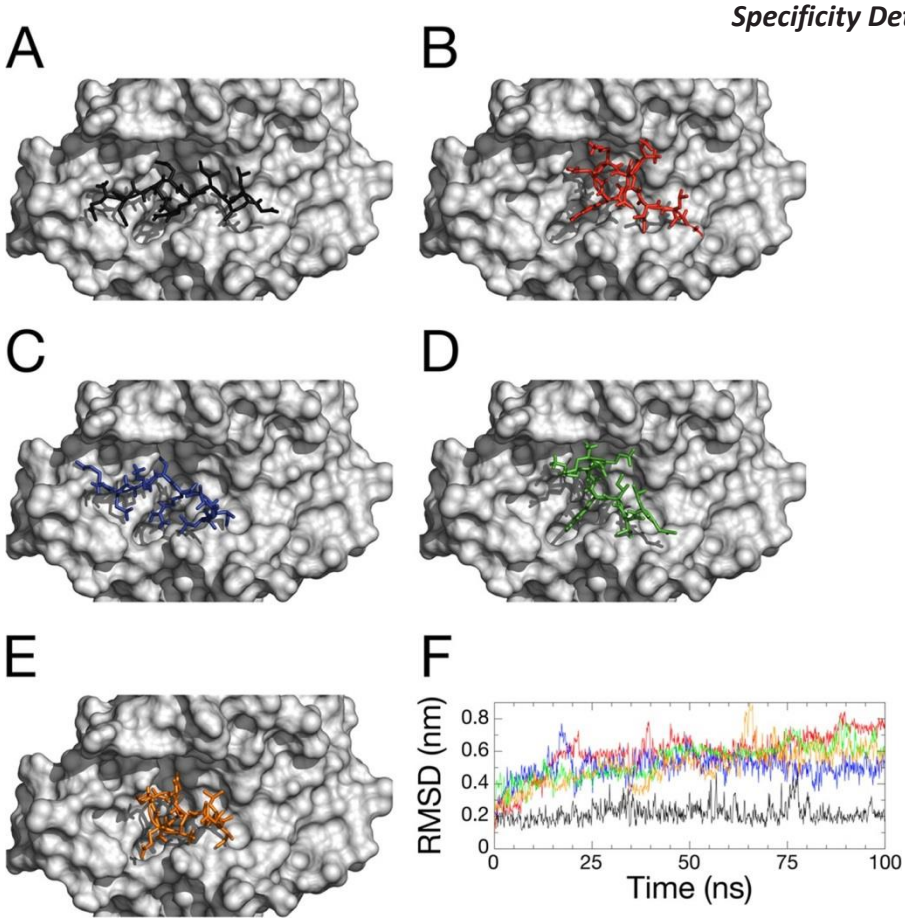


FIGURE 7. **Structural models of complexes between C1r protease and bound peptides.** *A–E*, snapshot after 100 ns of simulation of C1r protease, repre- sented as a *gray scheme*, with bound EEKQRIILG (*A*), EEKNRIILG (*B*), EEKQRAILG (*C*), EEKGRIILG (*D*), and EEKNRAILG (*E*) represented as *sticks*. *F*, root mean square deviation (*RMSD*) of backbone atoms of the EEKQRIILG (*black*), EEKNRIILG (*red*), EEKQRAILG (*blue*), EEKGRIILG (*green*), and EEKNRAILG (*orange*) peptides calculated after structural fitting of C1r backbone atoms.

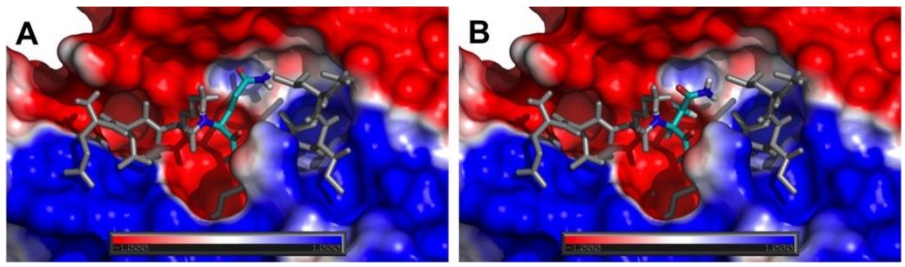


FIGURE 8. **Electrostatic potential of C1r protease with bound EEKQRIILG (*A*) and EEKNRIILG (*B*) represented as *sticks* mapped on its solvent accessible surface.** *Color coding* is according to the electrostatic potential gradient, where positively and negatively charged areas are represented in *blue* and *red* (iso-values from +1 k_bT/e_c to -1 k_bT/e_c), respectively.

shortening of the side chain and different side chain rotamer preferences of Asn compared with Gln may result in the introduction of its polar amide group into the nonpolar portion of the S2 binding pocket, destabilizing the C1r-peptide complex.

DISCUSSION

The complement system is vital for the proper function of the immune system, but also contributes to inflammatory diseases, therefore understanding initiating events in the pathways controlling activation is crucial to the design of inhibitors that can precisely target them to alleviate diseases in which complement is involved (3). Here we have provided evidence that indicates that the recognition of residues in the cleavage site by the active site of C1r, the initiating protease of the classical pathway, is equally important to recognition of substrate C1s via exosites found on the CCP domains of the activating C1r enzyme.

Analysis of the specificity of C1r using phage display technol- ogy revealed that the enzyme displayed strong specificity for residues at the P2, P1', and P2' of substrates. The strong preference displayed for Gln residues at P2 and Ile residues at P1' matched those found in physiological substrates of C1r, *i.e.* itself and zymogen C1s. The lack of convincing activity of C1r for peptide substrates meant that entire protein substrates had to be used to map the importance of the cleavage site residues for recognition by C1r. These analyses confirmed the impor- tance of the P2 Gln residue in particular and introduction of this residue alone into MASP-3, a lectin pathway protease with

similar domain structure (6), but different cleavage site residues at the nonprime side in particular, were sufficient to render this protease efficiently activated by C1r.

Molecular dynamics simulations confirmed that the Gln res- idue found at P2 was indeed highly important for recognition by the active site of C1r. These studies particularly explained why substituting an Asn residue at the P2 site was so much more detrimental than altering this residue to a Gly. Reduction of the side chain group of the P2 residue by one carbon most likely brings the polar head groups of the Asn residue into contact with hydrophobic residues surrounding the S2 pocket of C1r, which is clearly highly detrimental for binding, as demon- strated in the molecular dynamics simulations. Alteration of the P2 Gln to a Gly residue eliminates the binding interaction predicted to occur at S2 for the polar head group of the Gln residue, but the clash between the polar group on Asn and the hydrophobic surrounds of the S2 pocket is not found for the Gly residue, thus explaining its lower effect on the k_{cat}/K_m value for C1r activation of C1s.

Our results indicate that the CCP domains of C1r also play an important role in the activation of C1s by C1r, whereas the CCP domains of C1s play a lesser role in recognition of the activating protease. The crystal structure of C1r (7) shows the protease in a head to tail dimer, with the CCP domains of the protease mediating strong contacts with the dimer partner. C1r and C1s apparently form a tetramer in the C1 complex, and it has been postulated that the head to tail dimer of C1r forms the core of the tetrameric structure. Such an arrangement would indeed facilitate contact between the active site of one C1r molecule with the cleavage loop of the other partner, albeit that a rear- rangement would still be required over that found in the crystal structure to allow these regions to form the intimate contacts required for activation (28). Our results indicate that a substan- tial change must occur in the C1 complex following autoactiva- tion of C1r to allow the C1r CCP domains to mediate contacts with the C1s molecule. Such exosite interactions would then be in addition to the contacts formed between active site of C1r and the cleavage loop of C1s. Our results have therefore pro- vided important insights into the contacts required between the active site of C1r and the cleavage loop of C1s, with the P2 Gln residue being of particular importance in this regard. These data will facilitate the development of inhibitors of C1r for the treatment of inflammatory diseases.

Acknowledgments—We thank Antony Matthews for supplying the phage display library used in this study, Mrs. Usha Koul for excellent technical assistance, and Prof. J. A. Huntington (University of Cambridge) for useful discussions.

REFERENCES

1. Frank, M. M., and Fries, L. F. (1991) The role of complement in inflammation and phagocytosis. *Immunol. Today* **12,** 322–326
2. Duncan, R. C., Wijeyewickrema, L. C., and Pike, R. N. (2008) The initiating proteases of the complement system: controlling the cleavage. *Biochimie* **90,** 387–395
3. de Cordoba, S. R., Tortajada, A., Harris, C. L., and Morgan, B. P. (2012) Complement dysregulation and disease: from genes and proteins to diag- nostics and drugs. *Immunobiology* **217,** 1034–1046
4. Gaboriaud, C., Teillet, F., Gregory, L. A., Thielens, N. M., and Arlaud, G. J. (2007) Assembly of C1 and the MBL- and ficolin-MASP complexes: structural insights. *Immunobiology* **212,** 279–288
5. Arlaud, G. J., Gaboriaud, C., Thielens, N. M., Budayova-Spano, M., Rossi, V., and Fontecilla-Camps, J. C. (2002) Structural biology of the C1 complex of complement unveils the mechanisms of its activation and proteo- lytic activity. *Mol. Immunol.* **39,** 383–394
6. Gál, P., Dobó, J., Závodszky, P., and Sim, R. B. (2009) Early complement proteases: C1r, C1s and MASPs. A structural insight into activation and functions. *Mol. Immunol.* **46,** 2745–2752
7. Budayova-Spano, M., Grabarse, W., Thielens, N. M., Hillen, H., Lacroix, M., Schmidt, M., Fontecilla-Camps, J. C., Arlaud, G. J., and Gaboriaud, C. (2002) Monomeric structures of the zymogen and active catalytic domain of complement protease C1r: further insights into the C1 activation mech- anism. *Structure* **10,** 1509–1519
8. Kardos, J., Gál, P., Szilágyi, L., Thielens, N. M., Szilágyi, K., Lõrincz, Z., Kulcsár, P., Gráf, L., Arlaud, G. J., and Závodszky, P. (2001) The role of the individual domains in the structure and function of the catalytic region of a modular serine protease, C1r. *J. Immunol.* **167,** 5202–5208
9. Lacroix, M., Ebel, C., Kardos, J., Dobó, J., Gál, P., Závodszky, P., Arlaud, G. J., and Thielens, N. M. (2001) Assembly and enzymatic properties of the catalytic domain of human complement protease C1r. *J. Biol. Chem.* **276,** 36233–36240
10. Duncan, R. C., Bergström, F., Coetzer, T. H., Blom, A. M., Wijeyewick- rema, L. C., and Pike, R. N. (2012) Multiple domains of MASP-2, an initi- ating complement protease, are required for interaction with its substrate C4. *Mol. Immunol.* **49,** 593–600
11. Duncan, R. C., Mohlin, F., Taleski, D., Coetzer, T. H., Huntington, J. A., Payne, R. J., Blom, A. M., Pike, R. N., and Wijeyewickrema, L. C. (2012) Identification of a catalytic exosite for complement component C4 on the serine protease domain of C1s. *J. Immunol.* **189,** 2365–2373
12. Goldring, J. P., Thobakgale, C., Hiltunen, T., and Coetzer, T. H. (2005) Raising antibodies in chickens against primaquine, pyrimethamine, dap- sone, tetracycline, and doxycycline. *Immunol. Invest.* **34,** 101–114
13. Kerr, F. K., Thomas, A. R., Wijeyewickrema, L. C., Whisstock, J. C., Boyd, S. E., Kaiserman, D., Matthews, A. Y., Bird, P. I., Thielens, N. M., Rossi, V., and Pike, R. N. (2008) Elucidation of the substrate specificity of the MASP-2 protease of the lectin complement pathway and identification of the enzyme as a major physiological target of the serpin, C1-inhibitor. *Mol. Immunol.* **45,** 670–677
14. Kaiserman, D., Bird, C. H., Sun, J., Matthews, A., Ung, K., Whisstock, J. C., Thompson, P. E., Trapani, J. A., and Bird, P. I. (2006) The major human and mouse granzymes are structurally and functionally divergent. *J. Cell Biol.* **175,** 619–630
15. Cwirla, S. E., Peters, E. A., Barrett, R. W., and Dower, W. J. (1990) Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 6378–6382
16. Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) A survey of furin substrate specificity using substrate phage display. *Protein Sci.* **3,** 1197–1205
17. Eswar, N, Marti-Renom, M. A, Webb, B, Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative Protein Struc- ture Modeling Using Modeller. *Curr. Protoc. Bioinform.* **5,** 5.6.1–5.6.30
18. Mittl, P. R., Di Marco, S., Fendrich, G., Pohlig, G., Heim, J., Sommerhoff, C., Fritz, H., Priestle, J. P., and Grütter, M. G. (1997) A new structural class of serine protease inhibitors revealed by the structure of the hirustasin- kallikrein complex. *Structure* **5,** 253–264
19. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, version 1.3r1, Schrödinger, LLC, New York
20. Berendsen, H. J., Postma, J. P., and van Gunsteren, W. F. (1981) in *Inter- molecular Forces* (Pullman, B., ed) pp. 331–342, Reidel, Dordrecht
21. Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and sol- vation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Com- put. Chem.* **25,** 1656–1676
22. Hess, B., Berendsen, H. J., and Fraaije, J. G. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472
23. Tironi, S. R., Smith, P. E., and van Gunsteren, W. F. (1995) A generalized

reaction field method for molecular-dynamics simulations. J. Chem. Phys. 102, 5451–5459

24. Hu¨nenberger, H. T., and van Gunsteren, W. F. (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. J. Chem. Phys. 115, 1125–1136

25. Berendsen, H. J., van Gunsteren, J. P., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684–3690

26. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribo- some. Proc. Natl. Acad. Sci. U.S.A. 98, 10037–10041

27. Hedstrom, L., Lin, T. Y., and Fast, W. (1996) Hydrophobic interactions control zymogen activation in the trypsin family of serine proteases. Bio- chemistry 35, 4515– 4523

28. Wallis, R., Mitchell, D. A., Schmid, R., Schwaeble, W. J., and Keeble, A. H. (2010) Paths reunited: initiation of the classical and lectin pathways of complement activation. Immunobiology 215, 1–11

APPENDIX III

(PUBLICATION)

The X-ray crystal structure of Mannose-binding lectin associated serine proteinase-3 reveals a molecular basis for 3MC syndrome.

Yongqing, T, Wilmann, PG, Reeve, SB, Coetzer, TH, Smith, AI, Whisstock, TC, Pike RN, and Wijeyewickrema LC.

***J Biol Chem*. 288(31):22399-407 (2013).**

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 31, pp. 22399–22407, August 2, 2013

The X-ray Crystal Structure of Mannose-binding Lectin-associated Serine Proteinase-3 Reveals the Structural Basis for Enzyme Inactivity Associated with the Carnevale, Mingarelli, Malpuech, and Michels (3MC) Syndrome*

Received for publication, May 7, 2013, and in revised form, June 18, 2013 Published, JBC Papers in Press, June 21, 2013, DOI 10.1074/jbc.M113.483875

Tang Yongqing[‡1], Pascal G. Wilmann[‡1], Shane B. Reeve[‡], Theresa H. Coetzer[§], A. Ian Smith[‡], James C. Whisstock[‡¶], Robert N. Pike[‡2,3], and Lakshmi C. Wijeyewickrema[‡2,4]

From the [‡]Department of Biochemistry & Molecular Biology and the [¶]Australian Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia and the [§]Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg Campus), Scottsville 3209, South Africa

Background: Mutants of the serine protease, mannose-binding lectin associated-protease-3 (MASP-3), are associated with Carnevale, Mingarelli, Malpuech and Michels (3MC) syndrome.

Results: The lack of activity and the structure of the G666E mutant of MASP-3 reveals a molecular basis for 3MC syndrome.

Conclusion: Mutants of MASP-3 associated with 3MC syndrome are inactive.

Significance: The catalytic activity of MASP-3 is required for normal developmental processes.

The mannose-binding lectin associated-protease-3 (MASP-3) is a member of the lectin pathway of the complement system, a key component of human innate and active immunity. Mutations in MASP-3 have recently been found to be associated with Carnevale, Mingarelli, Malpuech, and Michels (3MC) syndrome, a severe developmental disorder manifested by cleft palate, intellectual disability, and skeletal abnormalities. However, the molecular basis for MASP-3 function remains to be understood. Here we characterize the substrate specificity of MASP-3 by screening against a combinatorial peptide substrate library. Through this approach, we successfully identified a peptide substrate that was 20-fold more efficiently cleaved than any other identified to date. Furthermore, we demonstrated that mutant forms of the enzyme associated with 3MC syndrome were completely inactive against this substrate. To address the structural basis for this defect, we determined the 2.6-Å structure of the zymogen form of the G666E mutant of MASP-3. These data reveal that the mutation disrupts the active site and perturbs the position of the catalytic serine residue. Together, these insights into the function of MASP-3 reveal how a mutation in this enzyme causes it to be inactive and thus contribute to the 3MC syndrome.

The lectin pathway of complement plays a crucial role in host defense against invading pathogens and malfunctioning host cells (1). The pathway is initiated by binding of specific recognition molecules, including multiple oligomeric forms of mannose-binding lectin (2), ficolins (-1, -2, and -3) (3) and the recently identified collagenous lectin CL-11 (4), to carbohy- drate and/or acetylated components on the surface of target cells. This binding triggers the activation of mannose-binding lectin-associated serine proteases (MASPs),[5] named MASP-1/- 2/-3, which are in complex with the recognition molecules. The MASP enzymes are all produced as single chain pro-enzymes and share the same domain organization: from the N terminus, the first C1r/C1s/Uegf/bone morphogenetic protein 1 (CUB1) domain, an epidermal growth factor-like (EGF) domain, the CUB2 domain, followed by two complement control protein (CCP1 and CCP2) domains, and the C-terminal serine protease (SP) domain, which is linked to the CCP2 domain by a small linker region (5). Although the CUB1-EGF-CUB2 portion of MASP enzymes is responsible for their dimerization (6, 7) and binding to mannose-binding lectin, ficolins, or CL-11, the CCP1-CCP2-SP portion mediates the catalytic properties of the enzymes (8, 9).

MASP-3 is ubiquitously distributed in the human body and exists in human plasma at a median concentration of 5.0 μg/ml (≈50 nM) (range: 1.8 –10.6 μg/ml) (10). Thus far, the only substrates of human MASP-3 reported are a few synthetic peptide substrates and the insulin-like growth factor-binding protein-5 (11). The physiological role of MASP-3 is currently unknown. Recently, two independent genetic studies have found a strong correlation between MASP-3-specific mutations and a

* This work was supported by a program grant from the National Health and Medical Research Council of Australia.

The atomic coordinates and structure factors (code 4KKD) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] Both authors contributed equally to this work.

[2] Joint senior and corresponding authors.

[3] To whom correspondence may be addressed. E-mail: rob.pike@monash.edu.

[4] To whom correspondence may be addressed. Tel.: 61-3-99029300; Fax: 61-3-99029500; E-mail: lakshmi.wijeyewickrema@monash.edu.

[5] The abbreviations used are: MASP, mannose-binding lectin-associated serine protease; Boc, tert-butyloxycarbonyl; C1, complement component 1; C1r, complement subcomponent 1r; C1s, complement subcomponent 1s; C1q, complement subcomponent 1q; CCP, complement control protein- like domain; SP, serine protease domain; zMASP, zymogen form of MASP; M3Q, K448Q mutant of MASP-3; CCP, complement control protein; CUB, module originally found in complement C1r/C1s, Uegf, and bone morpho- genetic protein; AMC, 7-amino-4-methyl coumarin group; PDB, Protein Data Bank; Z, benzyloxycarbonyl.

developmental disorder, the 3MC syndrome, which suggests a possible role for MASP-3 outside of the immune system (12, 13). The 3MC syndrome is the collective term used to identify four allelic variants of the same disease category, the Carnevale, Mingarelli, Malpuech, and Michels syndromes (14, 15). Patients with 3MC display a series of congenital characteristic anomalies, including abnormally increased distance between facial parts, bilateral ptosis with abnormally small and droopy eyelids, cleft lip and/or palate, abnormal growth pattern of skull, fusion of radius and ulna, as well as impairment in hearing and/or learning (16 –19). Four mutations (C630{168}R, G666{197}E, G687{216}R, and H497{57}Y (12, 13)) (numbers in {brackets} indicate chymotrypsin numbering) in the *MASP1* gene were identified to be located in exon 12, which encodes the SP domain of MASP-3. All these mutations map to the active site region of the enzyme, including the catalytic $His^{497\{57\}}$ res- idue. Both studies make elegant predictions of the effects of these mutations on the MASP-3 enzyme, however, the struc- ture of MASP-3 and the structural basis for MASP-3 deficiency remain to be characterized.

In a previous study (20), we produced a form of human MASP-3 with a single Gln residue replacing the Lys residue N-terminal to the R-I activation bond (M3Q). This enzyme could be efficiently cleaved and activated by the C1r protease of the classical pathway complement (20). Because the amino acid sequence C-terminal to the activation bond represents that of the wild type MASP-3, the C1r-activated recombinant M3Q protein contains a fully functional SP domain of the wild type MASP-3. We were similarly able to produce cleaved, activated (M3Q cleaved) forms of 3MC syndrome-related mutants G687{216}R and G666{197}E mutant. Both mutants lacked detectable protease activity. Finally, we determined the 2.6-Å structure of the G666{197}E mutant protease in the zymogen form. These data reveal substantial perturbation of the active site, consistent with a correlation between the lack of MASP-3 function and the 3MC syndrome.

EXPERIMENTAL PROCEDURES

Expression and Purification of MASP-3—Recombinant MASP-3 CCP12SP (residues Lys^{298}-Arg^{728} and mutants, M3Q, M3QG666{197}E, and M3QG687{216}R) were expressed and refolded with some modifications to previously described methods (21). Briefly, genes for all recombinant proteins were synthesized (GenScript, Piscataway, NJ) and the DNA was cloned into pET17b (EMD Biosciences, Rockland, MA). After transformation of the vector into *Escherichia coli* strain BL21(DE3)pLysS, cells were cultured at 37 °C in 2X TY (tryptone/yeast extract) broth with 50 μg/ml of ampicillin and 34 μg/ml of chloramphenicol to an A_{595} of 0.6, followed by induction with 1 mM isopropyl 13-D-thiogalactoside for 4 h. Following induction, the culture was centrifuged (27,000 X *g*, 20 min, 4 °C), the cells were collected in 30 ml of 50 mM Tris-HCl, 20 mM EDTA, pH 7.4, and frozen at -80 °C. The cells were thawed and sonicated on ice 6 times for 30 s. After centrifugation at 27,000 X *g* for 20 min, inclusion body pellets were sequentially washed and centrifuged with 10 ml of 50 mM Tris-HCl, 20 mM EDTA, pH 7.4. The washed pellet was resuspended in 10 ml of 8 M urea, 0.1 M Tris-HCl, 100 mM DTT, pH 8.3, at room temperature (RT) for 3 h. Refolding was initiated by rapid dilution dropwise into 50 mM Tris-HCl, 3 mM reduced glutathione, 1 mM oxidized glutathione, 5 mM EDTA, and 0.5 M arginine, pH 9.0. The renatured protein solutions were concentrated and dialyzed against 50 mM Tris-HCl, pH 9.0, and renatured proteins were purified on a 5-ml Q-Sepharose-Fast Flow column (GE Healthcare). The bound protein was eluted with a linear NaCl gradient from 0 to 400 mM over 35 ml at 1 ml/min. The recombinant proteins were further purified using a Superdex 75 16/60 column (GE Healthcare) in a buffer of 50 mM Tris, 145 mM NaCl, pH 7.4, aliquoted, snap frozen, and maintained at -80 °C. The purity of the protein was confirmed by SDS-PAGE followed by Western blotting and N-terminal sequencing. Typically protein yields were between 1 and 2 mg/liters.

Western Blotting and Antibodies—Proteins were resolved by SDS-PAGE, transferred, and immunoblotted using an anti-MASP-3 antibody directed against the unique peptide sequence, NPNVTDQIISSGTRT, which was raised in chickens as previously described (22).

Activation of Zymogen MASP-3—To activate the recombi- nant M3Q and 3MC syndrome-related MASP-3 mutants, 0.5 mg of active human C1r enzyme (Complement Technology, TX) was coupled to a 1-ml HiTrap™ NHS column (GE Health- care), according to the manufacturer's instructions. 0.5 mg of pure recombinant protein was loaded into the column for acti- vation at 26 °C for 16 h. The C1r-activated MASP-3 was then eluted, in a buffer containing 50 mM Tris, 145 mM NaCl, pH 7.4. Wild type MASP-3 and all recombinant MASP-3 mutant pro- teins were subjected to SDS-PAGE under reduced conditions and then transferred onto a PVDF membrane, followed by N-terminal sequencing to identify the cleavage site by C1r.

N-terminal Sequencing—Protein samples were reduced and denatured. The protein fragments in the samples were separated by SDS-PAGE and transferred onto a polyvinylidene difluoride membrane in a transfer buffer containing 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11, with 15% (v/v) methanol. The protein band on the membrane was visualized by Coomassie R-250 Brilliant Blue staining and subsequently excised. After three alternating washes of water and 50% (v/v) methanol, the band was cut into small pieces and loaded onto a Procise Protein Sequencer 492/492C with a Microgradient Delivery System 140C and a UV Detector 785A (Applied Biosystems) for sequencing. The sequencing process applied automated Edman degradation and phenylthiohydan- toin amino acid analysis. The data were analyzed by applying the Sequence Pro software (Applied Biosystems).

Enzyme Kinetic Analysis—Kinetic analysis of MASP-3 was performed in fluorescence assay buffer (FAB) (50 mM Tris-HCl, 150 mM NaCl, 0.2% (w/v) PEG 8000, pH 7.4), using the synthetic peptide substrates Boc-VPR-AMC, Boc-PFR-AMC, Boc-GGR-AMC, *Z*-FR-AMC, Boc-AFK-AMC, and Boc-LGR-AMC (denoted as VPR, PFR, GGR, ZFR, AFK, and LGR, respectively) at 37 °C on a POLARstar fluorescent plate reader (BMG Labtech, Offenburg, Germany). For a typical activity assay, 100 nM M3Q (or up to 1 μM of either M3QG666{197}E or M3QG687{216}R) was incubated with 100 μM substrate. The rate of increase of fluorescence was monitored for 30 min using excitation and emission wavelengths of 360 and 460 nm,

respectively. The initial rate of proteolysis of substrate was measured as the slope of the linear portion of the progressive curve, and was then converted to the rate of increase of AMC molecules per unit of enzyme ($M_{[AMC]}min^{-1}M_{[E]}^{-1}$) using a standard curve for arbitrary fluorescence units *versus* AMC concentration and taking into account the enzyme concentra- tion used. To determine the values of the steady-state reaction constants, K_m and V_{max}, 100 nM M3Q was used to react with substrate at a range of concentrations from 0 to 500 μM and the initial velocity of the reactions was plotted against the substrate concentration. The K_m and V_{max} values were estimated by using GraphPad Prism 5.0 (GraphPad Software, San Diego, CA) to fit the data to the Michaelis-Menten equation: $Y = V_{max} \times X/(K_m + X)$, where Y is the initial rate of the proteolytic reaction and X is the concentration of substrate. The $k_{cat} = V_{max}/[E]$, where E is the enzyme concentration.

TABLE 1

Data collection and refinement statistics for structure of MASP-3- CCP1-CCP2-SP (G666E 3MC mutant, K448Q activation loop mutant)

Values in parentheses are for the highest resolution shell (2.74–2.60 Å).

Data collection statistics	
Space group Cell parameters	$P2_12_12$
a, *b*, *c* (Å)	72.84, 292.9, 43.59
Resolution	48.82-2.60 (2.71-2.60)
Completeness	99.9% (99.1%)
I/u (*I*)	9.6 (2.2)
R_{meas}	16.4% (78.6%)
R_{sym}	15.2% (73.0%)
R_{pim}	6.1% (29.0%)
Multiplicity	7.2
Observations	214420 (25968)
Unique	29907 (3562)
Refinement statistics	
Resolution (Å)	48.82-2.60 (2.74-2.60)
R_{work}/R_{free}	19.40/25.98
Number of atoms	
Protein	6197
Imidazole ion	10
Water	201
B factors	
Main chain	17.5
Side chain	19.7
Imidazole ion	18.2
Water	15.1
R.M.S deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.2
Ramachandran favored (%)	92.55
Ramachandran outliers (%)	0.39
Molprobity score	3.29
Molprobity clashscore	36.94

Determination of the Substrate Specificity of M3Q Using the Rapid Endopeptidase Profiling Library (REPLi)—The REPLi library is a combinatorial peptide library that contains 512 pools of peptides with each pool containing up to eight different variable tripeptides with the template layout of MeOC-GG*XXX*GG-dipicolinic acid-KK, where each *X* represents a variable alternative amino acid based on similar physiochemi- cal properties, *i.e.* A/V, F/Y, I/L, D/E, R/K, D/E, S/T, Q/N, and P. There are no Gly, His, Trp, Cys, or Met residues in the variable tripeptide region. The resulting combinations of variable tripep- tides give rise to 3375 different peptides in the library in total. Methoxycarbonyl (MeOC) is the fluorophore, and dipicolinic acid is the fluorophore quencher. The soluble peptide library pools (*i.e.* 512 wells) in six 96-well plates were diluted using FAB to a final concentration of 50 μM. A final concentration of 300 nM M3Q was incubated with the substrate pools in FAB at 37 °C. Cleavage of the substrates was monitored by measuring the increase in fluores- cence intensity from the MeOC fluorophores using 55-s cycles for 30 cycles, with an excitation wavelength of 320 nm and an emis- sion wavelength of 420 nm. The initial velocity of the cleavage was indicated by the slope per unit time of the linear region of the curves. These values were converted into molar MeOC molecules per unit time per molar enzyme ($M_{[MeOC]}min^{-1}M_{[E]}^{-1}$) by using an MeOC-fluorescence standard curve, allowing comparison of the cleavage rate between pools and between enzymes.

Based on the REPLi results, 17 individual peptides from the substrate pools containing tripeptidyl sequences of Arg/Lys- Ile/Leu-Phe/Tyr, Arg/Lys-Arg/Lys-Ile/Leu, and Arg/Lys-Ser/ Thr-Ile/Leu, which displayed high or modest rates of cleavage by M3Q, were synthesized (GL Biochem, Shanghai, China). Each assay used 100 μM of the fluorescent quenched substrate with 500 nM M3Q (or up to 1 μM of either M3QG666{197}E or M3QG687{216}R) and cleavage was monitored for 30 min as described above. The initial rate of the proteolysis of substrate was measured as the slope of the linear portion of the progressive curve, and was then converted to $M_{[Abz]}min^{-1}m_{[E]}^{-1}$ using the standard curve. To determine the values of the steady state reaction constants, 500 nM M3Q was mixed with substrate at a range of concentrations from 0 to 500 μM and the initial velocity of reaction was plotted against the substrate concentration allow- ing the determination of the K_m, V_{max}, and k_{cat} values as described above. Selected wells that showed the highest increase in fluores-

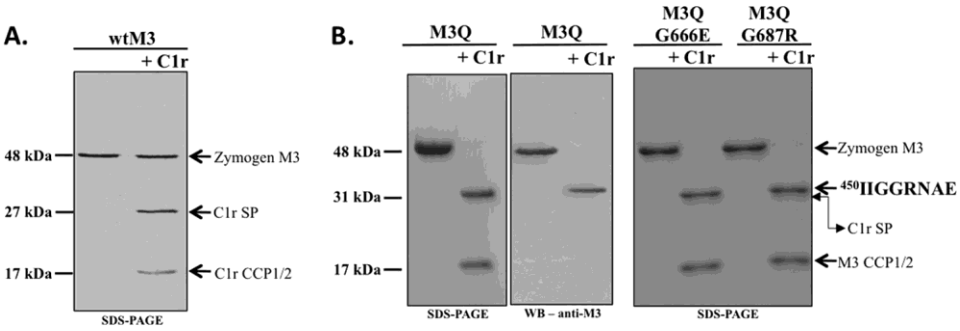


FIGURE 1. **The activation of MASP-3.** *A,* complement C1r does not cleave wild type MASP-3 (*wtM3*). Samples were analyzed using SDS-PAGE (reducing conditions). *Arrows* indicate the location and size of the domains of each protein. *B*, complement C1r cleaves and activates M3Q and its mutants. The N-terminal sequence indicated was identified in all three proteins. The location and size of the domains of each protein are indicated. Samples were analyzed using SDS-PAGE (reducing conditions) and immunoblotted with chicken anti-MASP-3 antibodies visualized using horseradish peroxidase-conjugated secondary antibody and ECL. *WB*, Western blot.

cence were subjected to mass spectrometry analysis for determination of the P1 specificity of MASP-3. The protease-treated peptide samples were co-spotted onto the MALDI target plate with Matrix solution of 10 mg/ml of cx-cyano-4-hydroxycinnamic acid (Laser BioLabs, Sophia-Antipolis, France) in 50% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid. The samples were analyzed on an Applied Biosystems (Foster City, CA) 4700 Proteomics Analyzer MALDI TOF/TOF in reflectron mode with a mass range of 400 – 1500 Da, focus mass of 1000 Da at 1500 shots/spectrum using plate model calibration against the 4700 tune mixture. Peak detection on the spectra was performed using the software 4700 Series Explorer version 3.0.

Protein Crystallization and X-ray Structure Determination— Purified zMASP-3 CCP12SP_QG666{197}E was buffer exchanged into 100 mM NaCl, 10 mM Tris-HCl, pH 7.4, and concentrated to 5 mg/ml for crystallization trials.

Crystals of zMASP-3 were obtained utilizing the sitting drop vapor diffusion method at 4 °C with a reservoir solution comprised of 0.12 M NaCl, 18% (w/v) polyethylene glycol 4000, and 0.1 M imidazole, pH 7.5. With an equal reservoir to protein ratio, crystal were observed after a day and grew to maximal size after 3 weeks. Crystals were cryogenically cooled in sucrose supplemented mother liquor prior to data collection.

TABLE 2
The rate of cleavage of synthetic peptidyl-AMC substrates by MASP-3 protein mutants ($M_{[AMC]}min^{-1}M_{[E]}^{-1}$)
Mean ± S.E. from triplicate experiments are shown.

Substrate	M3Q	M3QG666E	M3QG687R
Boc-VPR-AMC	0.019±0.0008	ND[a]	ND
Boc-LGR-AMC	0.017±0.0007	ND	ND
Boc-PFR-AMC Boc-GGR-AMC	0.009±0.0003	ND	ND
	0.004±0.0001	ND	ND
Z-FR-AMC	0.002 ± 0.0001	ND	ND
Boc-AFK-AMC	0.003 ± 0.0001	ND	ND

[a] ND, none detected.

TABLE 3
The activity of M3Q against selected synthetic AMC-peptidyl substrates, VPR and LGR
Mean ± S.E. from triplicate experiments are shown.

Substrate	k_{cat} s^{-1}	K_m μM	k_{cat}/K_m $M^{-1}s^{-1}$
Boc-VPR-AMC	0.0008 ± 0.00002	142.8±8.2	5.80
Boc-LGR-AMC	0.0005 ± 0.00002	105.6 ± 11.3	4.84

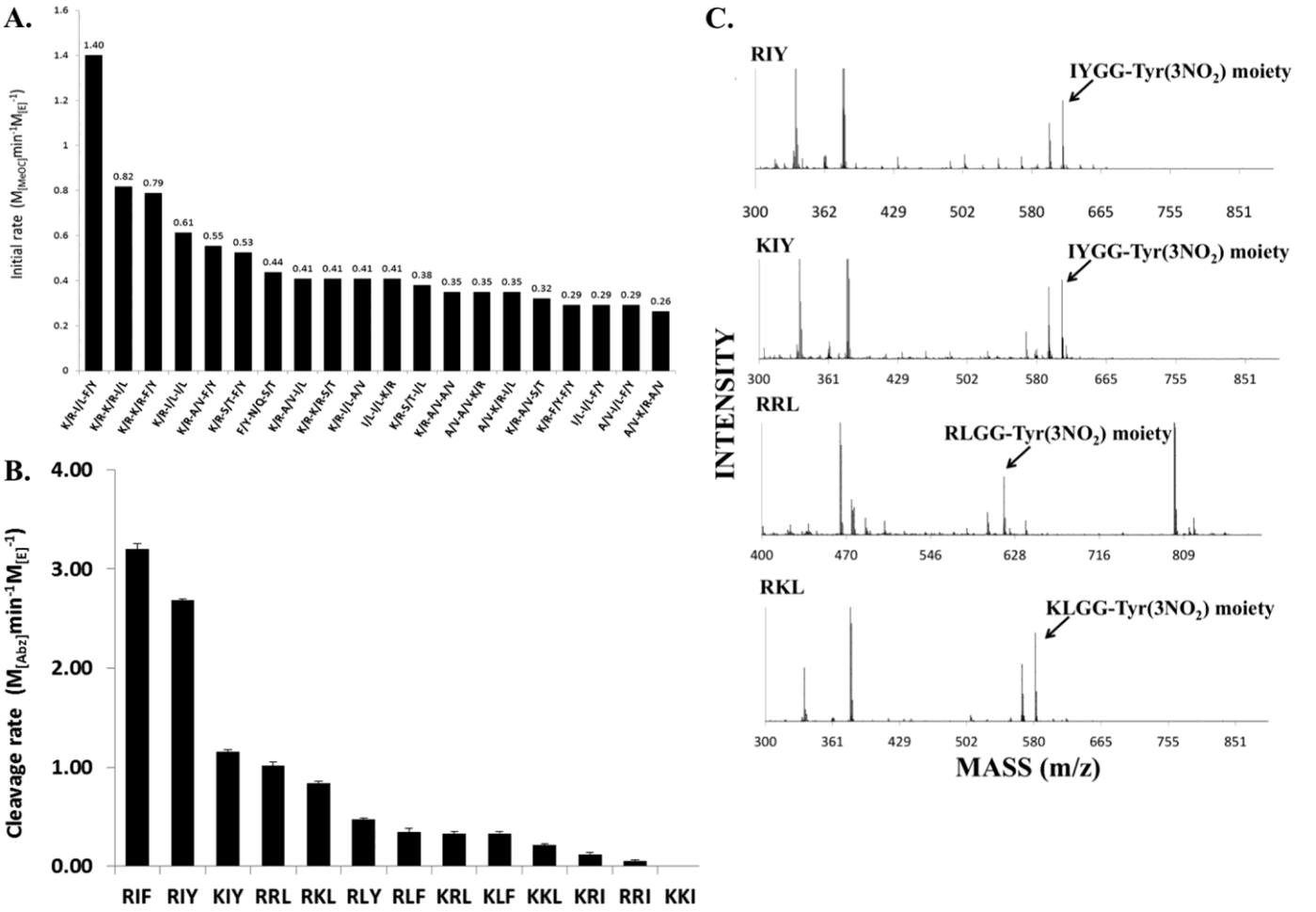


FIGURE 2. **The substrate specificity of MASP-3.** *A,* the rate of cleavage of 20 REPLi peptide pools best cleaved by M3Q. The peptide pools (50 μM) were incubated with 300 nM M3Q at 37 °C. Fluorescence intensities were measured at 30 s intervals for 30 min. The initial velocities of cleavage for the pools are shown. *B*, the rate of cleavage of thirteen synthesized peptides based on the (K/R)-(I/L)-(F/Y) and (K/R)-(K/R)-(I/L) pools of the REPLi library by M3Q. The protease (300 nM) was incubated with each peptide substrate (50 μM) at 37 °C, and fluorescence intensities were measured at 30 s intervals for 30 min. The initial velocities of cleavage of each peptide are shown. Mean ± S.E. (*error bars*) from a triplicate experiment are shown. *C*, representative mass spectra identifying cleaved products of the peptides, RIY, KIY, RRL, and RKL after cleavage with M3Q.

Data were collected at cryogenic temperatures on the MX2 beamline at the Australian Synchrotron to 2.6-Å resolution. Data were indexed, integrated, scaled, and merged using the Xia2 data reduction pipeline (23), which utilizes XDS (24–26), Pointless, Aimless (27), and the CCP4 suite (28).

A molecular replacement solution was found following utilization of the EMBL-HH Automated Crystal Structure Determination Platform "Auto-Rickshaw" (29, 30), which contained 2 molecules per asymmetric unit with clear unbiased $F_o - F_c$ electron density observed in omitted loop regions. Manual iterative model building and refinement using Coot (31) and phenix.refine (32) was performed yielding a high quality final model (R_{factor}/R_{free} 19.5/26.0, Molprobilty Score 3.29).

Final coordinates and structure factors have been deposited with the Protein Data Bank with the accession code 4KKD. A table of full data collection and refinement statistics can be found in Table 1.

RESULTS

The Activity and Specificity of Activated MASP-3—We used a mutant of MASP-3 in which a Gln residue had been introduced at position 448 in place of the Lys residue at that position to facilitate its efficient cleavage and activation by C1r (20) (Fig. 1). In general, active MASP-3 had very low activity (Table 2), with VPR being the best substrate, (k_{cat}/K_m value of 5.8 $M^{-1} s^{-1}$ (Table 3)). The similarly activated (Fig. 1) mutant forms of MASP-3 associated with the 3MC syndrome were completely inactive against any substrate tested (Table 2).

TABLE 4

The activity of M3Q against individual substrates derived from REPLi library analysis ($M_{[Abz]}min^{-1} M^{-1}$) (substrates were of the form MeOC-GG*XXX*GG-DPA-KK, where the XXX sequence represents the tri- peptide motif shown in the table)

Substrate	k_{cat}	K_m	k_{cat}/K_m
		μM	$M^{-1} s^{-1}$
RIF	0.0049 ± 0.0002	4.97 ± 0.54	98.55
RIY	0.0037 ± 0.0002	4.80 ± 0.52	76.90
RRL	0.0005 ± 0.00004	2.17 ± 0.50	23.29
RKL	0.0006 ± 0.00005	4.00 ± 0.78	14.61
KIY	0.0009 ± 0.00006	7.46 ± 0.96	11.50

We comprehensively mapped the specificity of activated wild type MASP-3 using the REPLi combinatorial peptide substrate library. The majority of the substrate pools in the library were not cleaved by MASP-3 (Fig. 2*A*). However, we were able to identify optimal substrates for the enzyme, with the top ranked pool con- taining substrates of the sequence Lys/Arg-Ile/Leu-Phe/Tyr (Fig. 2*A*).

The individual substrates contained in the top two pools from the REPLi analysis were individually synthesized and tested for their rate of cleavage by wild type MASP-3 (Fig. 2*B*). These data collectively reveal that the enzyme had a strong preference for an Arg residue at P1, an Ile residue at P1', and an amino acid with an aromatic side chain (Phe or Tyr) at P2'. Our studies also revealed that the enzyme was able to cleave substrates containing the diba- sic RR or RK motif. Analysis using LC-MS revealed that the enzyme cleaved after the first basic amino acid (Fig. 2*C*), thus it is likely that the enzyme additionally can accommodate a positively charged residue at the P1' position of substrates.

Kinetic analysis of the cleavage of the top 5 substrates selected from the REPLi library revealed that the RIF and RIY substrates were optimal for MASP-3, with k_{cat}/K_m values 3-fold better than any of the other substrates tested (Table 4). The top substrate, RIF, had a k_{cat}/K_m value of 98.5 $M^{-1} s^{-1}$, which is ~20-fold better than the coumaryl peptide substrates used thus far to characterize MASP-3 *in vitro*.

We tested the 3MC syndrome-associated mutants G666E and G687R against this substrate. These data revealed no detectable activity, suggesting that the mutant enzymes are indeed inactive.

Structure Determination of Zymogen MASP-3—To address the structural basis of 3MC syndrome, we determined the 2.6-Å resolution structure of a MASP-3 CCP12SP zymogen

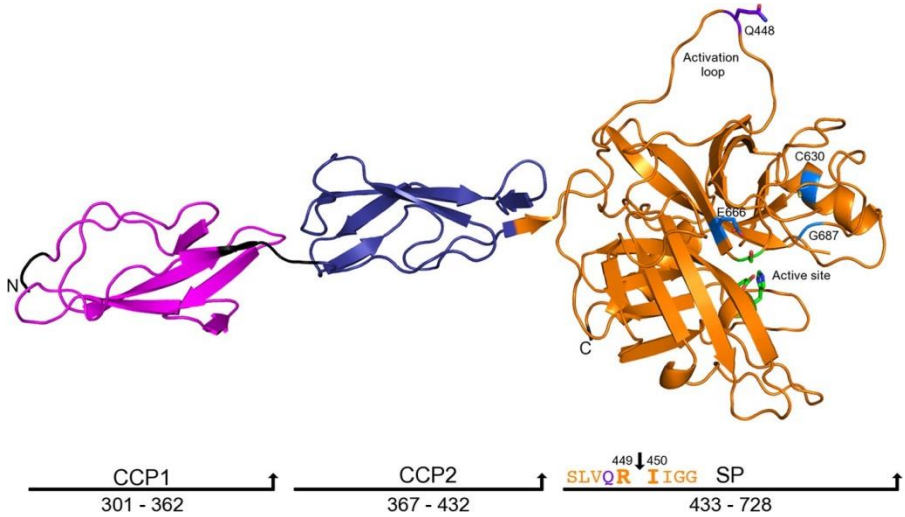


FIGURE 3. **Schematic representations of the MASP-3_G666E structure.** *A,* the zymogen MASP-3 G666E monomer (PDB code 4KKD) with residues 301–362 in *magenta* representing the CCP1 domain, residues 367– 432 in *dark blue* representing the CCP2 domain, and residues 433–728 in *gold* representing the SP domain. The linker regions are colored *black* and the *N* and *C* denote the N- and C-terminal ends, respectively. The activation loop substitution K448{14}Q is shown in *purple stick* format. The active site residues (His$^{497\{57\}}$, Asp$^{553\{102\}}$, and Ser$^{664\{195\}}$) are shown in *green stick* format. The residues involved in point mutations associated with the 3MC syndrome, Cys$^{630\{168\}}$ and Gly$^{687\{216\}}$, in addition to the incorporated G666{197}E substitution are shown in *blue stick* format (His$^{497\{57\}}$ in *green stick* is also the site of a point mutation associated with 3MC syndrome). The amino acid sequence of the activation loop is shown in *gold* font below, with the *black arrow* between Arg$^{449\{15\}}$ and Ile$^{450\{16\}}$ highlighting the activation point.

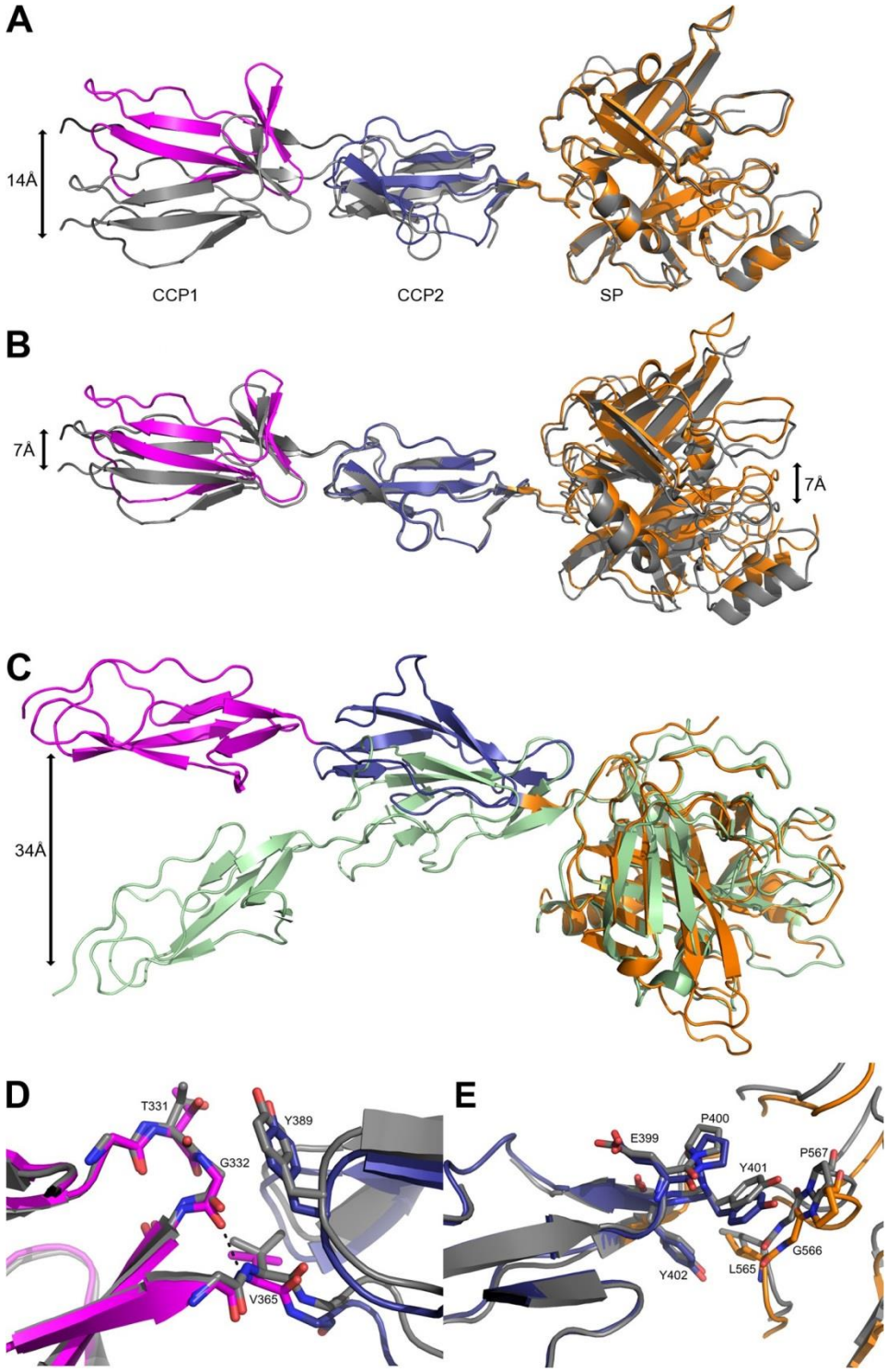


FIGURE 4. **Domain flexibility of zMASP-3.** *A,* schematic representation of SP domain superimposition of zMASP-3 chain A colored *magenta*, *violet*, and *orange* and chain B colored *gray*. *B,* schematic representation of CCP2 domain superimposition of zMASP-3, showing significant inter-domain movement. *C,* schematic representation of SP domain superimposition of zMASP-3 chain A, colored as in *panel A*, on zMASP-2, colored *green*. *D,* stick representation of key residues involved in the CCP1-CCP2 domain movement seen between chains A and B of the MASP-3. The *dashed line* indicates polar interactions. *E,* stick representation of key residues involved in the CCP2-SP domain movement seen between chains A and B of the MASP-3.

(zMASP-3) containing the mutation G666{197}E (Fig. 3). Two molecules are present in the asymmetric unit. Molecule B is more complete than molecule A and includes the intact activation loop (amino acids 440 – 453{5–19}). The activation loop mutation (K448{14}Q) participates in crystallographic packing interactions, suggesting that this may have aided crystallization. We were unable to crystallize wild type or activated enzyme.

The SP domain of MASP-3 is in the zymogen conformer of the chymotrypsin-like serine proteases. Superimposition of the SP domain of chain A onto chain B reveals a significant differ-

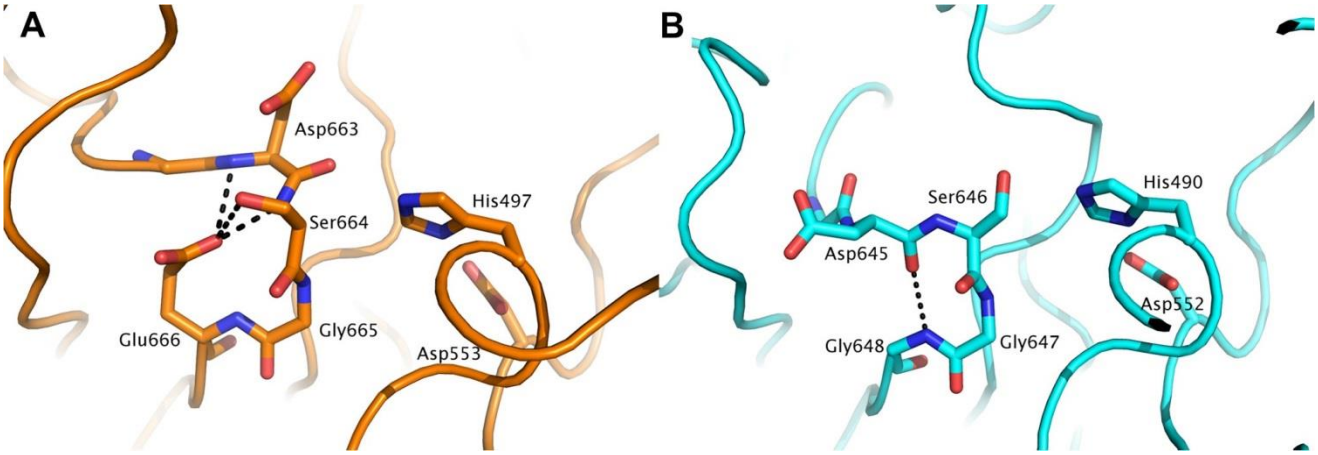


FIGURE 5. **Active site comparison of zymogen MASP-3 G666E and zymogen MASP-1.** *A,* zMASP-3 G666{197}E mutant active site structure (PDB code 4KKD) cx-turn colored *orange* and shown in stick representation. *Dashed lines* represent polar interactions. *B,* zMASP-1 structure (PDB code 4IGD) showing the 13-turn colored in *cyan*.

ence in the positions of the CCP1 and CCP2 domains, with up to 14 Å of variance observed in Ccx positions in CCP1. The interactions between the various domains remain the same, but differences in the relative positions of CCP residues account for inter-domain flexion (Fig. 4). The zMASP-3 SP chains A and B superimpose well (root mean square deviation = 0.588 Å). Minor differences in side chain rotamers in addition to minor loop movements are observed, presumably due to differences in crystal packing interactions. Superimposition of the CCP2 domains reveals a different backbone position of residues 398 – 409 causing the different domain juxtaposition. The greatest shift is observed in the Pro^{400} Ccx with a movement of 0.98 Å (Fig. 4, *B* and *E*).

Differences in the positions of CCP1 relative to CCP2 are observed upon superimposition of CCP1 domains (Fig. 4*D*). In CCP1, a backbone hydrogen bond between Gly^{332}-O and Val^{365}-N is observed in chain A (2.78 Å), but is lost in chain B (3.34 Å). Concurrently, the peptide backbone of Thr^{331} is flipped (x angle A: 55.25 *versus* B: 175.89). Movement of Gly^{332} and Thr^{331} away from CCP2 serves to reduce steric hindrance to domain movement increasing the distance to Tyr^{389} in CCP2. The different orientations of the CCP1, CCP2, and SP domains between the two chains in the crystallographic asym- metric unit highlight the flexibility of the CCP domain interac- tions with minimal hydrogen bonding observed between the domains.

Comparison of zMASP-3 with zMASP2 further highlights the CCP2-SP inter-domain flexibility. Superimposition of the SP domains of zMASP-3 and zMASP-2 (PDB code 1ZJK) reveal a shift of up to 34 Å in the relative position of the CCP1 domain between the enzymes (Fig. 4*C*). The difference at the CCP2-SP junction of Val^{433} and Glu^{435} in MASP-2 and MASP-3, respec- tively, sees the loss of a polar interaction in MASP-2, which contributes significantly to the different domain positioning in the two enzymes.

In the zMASP-3 unit cell, the 2 molecules are found in a head to head configuration, with several interactions between the long extended A and B loops observed. The unusually long D loop (amino acids 595– 612{143–150}) in both chains was not visible in electron density. The buried surface area of the inter- action between the SP domains of the asymmetric unit has been calculated to be 276 $Å^2$. However, there is no evidence that MASP-3 forms a dimer *in vitro*.

The catalytic motif (GDSer$^{\{195\}}$GG) is highly conserved across the chymotrypsin family. Our data reveal that mutation of the $Gly^{666\{197\}}$ in MASP-3 has marked structural conse- quences. Comparison of zMASP-3_G666{197}E with zymogen forms of MASP-1, MASP-2, C1r, and C1s reveals that the intro- duction of a relatively large charged side chain in place of the glycine residue has dramatic effects. The 3MC syndrome G666{197}E mutation significantly impacts on the position of the active site $Ser^{664\{195\}}$, drawing the functional hydroxyl group away from the catalytic $His^{497\{57\}}$ to an inactive position and thereby eliminating any potential charge relay and subse- quent proteolytic activity (Fig. 5*A*). Specifically, $Glu^{666\{197\}}$ makes H-bonds with $Asp^{663\{194\}}$-N, $Ser^{664\{195\}}$-N, and $Ser^{664\{195\}}$-Oγ, forcing the active site loop into a cx-turn confor- mation. This structure is distinctly different to the 13-turn con- formation observed in other zymogens (MASP-1, MASP-2, C1s, C1r (where the 13-turn is stabilized by a main chain H-bond between $Asp^{\{194\}}$-O and $Gly^{\{197\}}$-N)) (Fig. 5*B*). Furthermore, rather than interacting with the continuation of the D loop, $Asp^{663\{194\}}$ is in a novel solvent-exposed conformation.

DISCUSSION

The lectin pathway of complement is an important compo- nent of the innate immune system of the body (1). The MASP-1 and MASP-2 enzymes of the pathway play clear roles in activa- tion of the complement system via this pathway, with MASP-1 activating MASP-2 and playing a role in subsequent C2 cleav- age, whereas MASP-2 is crucial for C4 cleavage and plays a major role in C2 cleavage as well (33). The role of MASP-3 in this system is presently contentious.

Iwaki *et al.* (34) presented findings to suggest that murine MASP-3 was able to cleave Factors B and D of the alternative pathway of complement and thus play a role in activation of this system. Recently, Degn *et al.* (35) showed that there was no alteration in the activation of the alternative pathway in serum from a patient deficient in both MASP-1 and MASP-3. It is therefore uncertain what relevance the cleavage of Factors B

and D will have physiologically in humans, but it is worth noting that activating cleavage of these enzymes would be predicted to occur after the Arg residues in the P2-P2' sequence KR-KI for Factor B and GR-IL for Factor D. This matches the cleavage between basic residues noted for the enzyme here and also the preference for a hydrophobic amino acid at P2'. In addition to these cleavages, the reported cleavage of insulin growth factor-binding protein-5 by MASP-3 (11) has been shown to occur at the P2-P2' sites PR-II (major cleavage site), PK-IF, and PK-HT. The first cleavage matches the R-I P1-P1' specificity noted here and the general preference for a hydrophobic amino acid at P2', the second cleavage matches the Ile specificity at P1' and the preference for an aromatic amino acid at P2'. In general, the cleavage specificity noted here is consistent with cleavages in proteins by MASP-3 reported in the literature.

Recently, some light has possibly been shed on the function of MASP-3 due to findings that mutations to the enzyme or one of its lectin ligand recognition partners, CL-11, are associated with the 3MC syndrome (12, 13). Work in this regard suggests that the enzyme and lectin molecule are involved in neural crest migration during early developmental processes (13). The molecular targets of the lectin and enzyme are currently unknown. Many of the mutations to the enzyme clearly will yield an inactive enzyme because they result in predicted trun- cation of the protein (12, 13). Two of the other point mutations that are likely to result in full-length protein are also most likely to be inactive because one (H497{57}Y) targets the His residue involved in the catalytic triad of the enzyme and the other (C630{168}R) would most likely disrupt a disulfide bond vital to the structure of the enzyme. The effects of the other two point mutations in MASP-3 (G666{197}E and G687{216}R) associ- ated with the syndrome were less certain, although molecular modeling has suggested that these might be inactivating due to their active site-associated location and the major change in amino acid side chain involved. Here we show that activated forms of these two mutants of MASP-3 associated with the 3MC syndrome were indeed unable to cleave a substrate that was optimal for wild type MASP-3, supporting the idea that the 3MC syndrome arises from a lack of MASP-3 activity. Further- more, the x-ray crystal structure of the G666{197}E mutant enzyme provides an understanding of the structural basis for the inactivating effect of the 3MC-associated point mutation.

The physiological roles of MASP-3 remain to be found. It is clear that the enzyme plays a role in an important developmen- tal process as evidenced by the severe effect that an inactivating mutation has in determining the phenotype observed. Elucidat- ing the role of the enzyme in this system and in the complement system at maturity remain crucial objectives in understanding the biology of this enzyme.

Acknowledgments—We thank Usha Koul for excellent technical assistance. We also thank Professor James Huntington,Cambridge Institute for Medical Research, Cambridge, UK, for insightful discussions.

REFERENCES

1. Sim, R. B., and Tsiftsoglou, S. A. (2004) Proteases of the complement system. *Biochem. Soc. Trans.* **32,** 21–27
2. Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., and Yamashina, I. (1987) Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.* **262,** 7451–7454
3. Matsushita, M., Endo, Y., Taira, S., Sato, Y., Fujita, T., Ichikawa, N., Na- kata, M., and Mizuochi, T. (1996) A novel human serum lectin with col- lagen- and fibrinogen-like domains that functions as an opsonin. *J. Biol. Chem.* **271,** 2448–2454
4. Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M. O., Nielsen, O., Hartshorn, K., Jørgensen, T. J., Skjødt, K., and 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
5. Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H., Matsushita, M., Nonaka, M., and Fujita, T. (1998) Two lineages of mannose-binding lec- tin-associated serine protease (MASP) in vertebrates. *J. Immunol.* **161,** 4924–4930
6. Wallis, R., and Dodd, R. B. (2000) Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations. *J. Biol. Chem.* **275,** 30962–30969
7. Thielens, N. M., Cseh, S., Thiel, S., Vorup-Jensen, T., Rossi, V., Jensenius, J. C., and 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
8. Ambrus, G., Gál, P., Kojima, M., Szilágyi, K., Balczer, J., Antal, J., Gráf, L., Laich, A., Moffatt, B. E., Schwaeble, W., Sim, R. B., and Závodszky, P. (2003) Natural substrates and inhibitors of mannan-binding lectin-asso- ciated serine protease-1 and -2. A study on recombinant catalytic frag- ments. *J. Immunol.* **170,** 1374–1382
9. Rossi, V., Cseh, S., Bally, I., Thielens, N. M., Jensenius, J. C., and Arlaud, G. J. (2001) Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. *J. Biol. Chem.* **276,** 40880–40887
10. Degn, S. E., Jensen, L., Gál, P., Dobó, J., Holmvad, S. H., Jensenius, J. C., and Thiel, S. (2010) Biological variations of MASP-3 and MAp44, two splice products of the *MASP1* gene involved in regulation of the complement system. *J. Immunol. Methods* **361,** 37–50
11. Cortesio, C. L., and Jiang, W. (2006) Mannan-binding lectin-associated serine protease 3 cleaves synthetic peptides and insulin-like growth factor-binding protein 5. *Arch. Biochem. Biophys.* **449,** 164–170
12. Sirmaci, A., Walsh, T., Akay, H., Spiliopoulos, M., Sakalar, Y. B., Hasanefendioğlu-Bayrak, A., Duman, D., Farooq, A., King, M. C., and Tekin, M. (2010) MASP1 mutations in patients with facial, umbilical, coc- cygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am. J. Hum. Genet.* **87,** 679–686
13. Rooryck, C., Diaz-Font, A., Osborn, D. P., 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., and Beales, P. L. (2011) Mutations in lectin complement pathway genes *COLEC11* and *MASP1* cause 3MC syndrome. *Nat. Genet.* **43,** 197–203
14. Leal, G. F., Silva, E. O., Duarte, A. R., and Campos, J. F. (2008) Blepharo- phimosis, blepharoptosis, defects of the anterior chamber of the eye, cau- dal appendage, radioulnar synostosis, hearing loss and umbilical anoma- lies in sibs. 3MC syndrome? *Am. J. Med. Genet. A* **146,** 1059–1062
15. Al Kaissi, A., Klaushofer, K., Safi, H., Chehida, F. B., Ghachem, M. B., Chaabounni, M., and Hennekam, R. C. (2007) Asymmetrical skull, ptosis, hypertelorism, high nasal bridge, clefting, umbilical anomalies, and skele- tal anomalies in sibs. Is Carnevale syndrome a separate entity? *Am. J. Med. Genet. A* **143,** 349–354
16. Carnevale, F., Krajewska, G., Fischetto, R., Greco, M. G., and Bonvino, A. (1989) Ptosis of eyelids, strabismus, diastasis recti, hip defect, cryptorchid- ism, and developmental delay in two sibs. *Am. J. Med. Genet.* **33,** 186–189
17. Michels, V. V., Hittner, H. M., and Beaudet, A. L. (1978) A clefting syn- drome with ocular anterior chamber defect and lid anomalies. *J. Pediatr.* **93,** 444–446
18. Mingarelli, R., Castriota Scanderbeg, A., and Dallapiccola, B. (1996) Two sisters with a syndrome of ocular, skeletal, and abdominal abnormalities (OSA syndrome). *J. Med. Genet.* **33,** 884–886
19. Malpuech, G., Demeocq, F., Palcoux, J. B., and Vanlieferinghen, P. (1983)

A previously undescribed autosomal recessive multiple congenital anomalies/mental retardation (MCA/MR) syndrome with growth failure, lip/palate cleft(s), and urogenital anomalies. *Am. J. Med. Genet.* **16,** 475–480

20. Wijeyewickrema, L. C., Yongqing, T., Tran, T. P., Thompson, P. E., Vil- joen, J. E., Coetzer, T. H., Duncan, R. C., Kass, I., Buckle, A. M., and Pike, R. N. (2013) Molecular determinants of the substrate specificity of the complement initiating protease, C1r. *J. Biol. Chem.* **288,** 15571–15580
21. Duncan, R. C., Mohlin, F., Taleski, D., Coetzer, T. H., Huntington, J. A., Payne, R. J., Blom, A. M., Pike, R. N., and Wijeyewickrema, L. C. (2012) Identification of a catalytic exosite for complement component C4 on the serine protease domain of C1s. *J. Immunol.* **189,** 2365–2373
22. Goldring, J. P., Thobakgale, C., Hiltunen, T., and Coetzer, T. H. (2005) Raising antibodies in chickens against primaquine, pyrimethamine, dapsone, tetracycline, and doxycycline. *Immunol. Invest.* **34,** 101–114
23. Winter, G. (2009) Xia2. An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190
24. Kabsch, W. (1988) Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* **21,** 67–72
25. Kabsch, W. (1988) Evaluation of single-crystal x-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21,** 916–924
26. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800
27. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82
28. Collaborative Computational Project, Number 4 (1994) The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
29. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2005) Auto-rickshaw. An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 449–457
30. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **65,** 1089–1097
31. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
32. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., Mc- Coy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solu- tion. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
33. Héja, D., Kocsis, A., Dobó, J., Szilágyi, K., Szász, R., Závodszky, P., Pál, G., and Gál, P. (2012) Revised mechanism of complement lectin-pathway activation revealing the role of serine protease MASP-1 as the exclusive activator of MASP-2. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 10498–10503
34. Iwaki, D., Kanno, K., Takahashi, M., Endo, Y., Matsushita, M., and 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
35. Degn, S. E., Jensen, L., Hansen, A. G., Duman, D., Tekin, M., Jensenius, J. C., Thiel, S. (2012) Mannan-binding lectin-associated serine protease (MASP)-1 is crucial for lectin pathway activation in human serum, whereas neither MASP-1 nor MASP-3 is required for alternative pathway function. *J. Immunol.* **189,** 3957–3969

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. **66**:213-221.

Ambrus, G., Gál, P., Kojima, M., Szilágyi, K., Balczer, J., Antal, J., Gráf, L., Laich, A., Moffatt, B.E., Schwaeble, W., Sim, R.B., and Závodszky, 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**(3):1374-82.

Arlaud, G.J., Gaboriaud, C., Thielens, N.M., Budayova-Spano, M., Rossi, V., and Fontecilla-Camps, J.C. (2002) Structural biology of the C1 complex of complement unveils the mechanisms of its activation and proteolytic activity. *Mol Immunol*. **39**:383-394.

Arlaud, G.J., Gaboriaud, C., Thielens, N.M., Rossi, V., Bersch, B., Hernandez, J.F., and Fontecilla-Camps, J.C. (2001) Structural biology of C1: dissection of a complex molecular machinery. *Immunol Rev*. **180**:136-145.

Arlaud, G.J., and Gagnon, J. (1981) C1r and C1s subcomponents of human complement: two serine proteases lacking the “histidine-loop” disulphide bridge. *Biosci Rep*. **1**:779-784.

Arlaud, G.J., Gagnon, J., Villiers, C.L., and Colomb, M.G. (1986) Molecular characterization of the catalytic domains of human complement serine protease C1r. *Biochem*. **25**:5177-5182.

Banda, N.K., Takahashi, M., Levitt, B., Glogowska, M., Nicholas, J., Takahashi, K., Stahl, G.L., Fujita, T., Arend, W.P., and Holers, V.M. (2010) Essential role of complement mannose-binding lectin-associated serine proteases-1/3 in the murine collagen antibody-induced model of inflammatory arthritis. *J Immunol*. **185**:5598-5606.

Barlow, P.N., Baron, M., Norman, D.G., Day, A.J., Willis, A.C, Sim, R.B. and Campbell, I.D. (1991) Secondary structure of a complement control protein module by two-dimensional 'H NMR. *Biochem*. **30**:997-1004.

Boniotto, M., Braid, L., Baldas, V., Not, T., Ventura, A., Vatta, S., Radillo, O., Tedesco, F., Percopo, S., Montico, M., Amoroso, A., and Crovella, S. (2005) Evidence of a correlation between mannose binding lectin and celiac disease: a model for other autoimmune diseases. *J Mol Med*. **83**: 308-315.

Budayova-Spano, M., Grabarse, W., Thielens, N.M., Hillen, H., Lacroix, M., Schmidt, M., Fontecilla-Camps, J.C., Arlaud, G.J., and Gaboriaud, C. (2002a) Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism. *Structure.* **10**(11):1509-1519.

Budayova-Spano, M., Lacroix, M., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., and Gaboriaud, C. (2002b) The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *EMBO J*. **21**(3):231-239.

Chen, C.B., and Wallis, R. (2001) Stoichiometry of complexes between mannose-binding protein and its associated serine proteases: defining function units for complement activation. *J Biol Chem*. **276**:25894-25902.

Chen, C.B., and Wallis, R. (2004) Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases. *J Biol Chem*. **279**:26058-26065.

Constam, D.B., and Robertson, E.J. (2000) SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes Dev*. **14**(9):1146-1155.

Cortesio, C.L., and Jiang, W. (2006) Mannan-binding lectin-associated serine protease 3 cleaves synthetic peptide and insulin-like growth factor-binding protein 5. *Arch Biochem Biophy*. **449**:164-170.

Cseh, S.L., Vera, L., Matsushita, M., Fujita, T., Arlaud, G.J., and Thielens, N.M. (2002) Characterization of the interaction between L-ficolin/p35 and mannan-binding lectin-associated serine protease-1 and -2. *J Immunol*. **169**:5735-5743.

Dahl, M.R., Thiel, S., Matsushita, M., Fujita, T., Willis, A. C., Christensen, T., Voruo-Jensen, T., and Jensenius, J.C. (2001) MASP-3 and its association with distict complexes of the mannan-binding lectin complement activation pathway. *Immunity*. **15**:127-135.

Degn, S.E., Hansen, A.G., Steffensen, R., Jacobsen, C., Jensenius, J.C., and Thiel, S. (2009) MAp44, a human protein associated with pattern recognitionmolecules of the complement system and regulating the lectin pathway of complement activation. *J Immunol*. **183**:7371-7378.

Degn, S.E., Jensen, L., Gál, P., Dobó, J., Holmvad, S.H., Jensenius, J.C., and Thiel, S. (2010) Biological variations of MASP-3 and MAp44, two splice products of the MASP-1 gene involved in regulation of the complement system. *J Immunol Methods*. **361**(1-2):37-50.

Degn, S.E., Jensen, L., Hansen, A.G., Duman, D., Tekin, M., Jensenius, J.C., and Thiel, S. (2012) Mannan-binding lectin-associated serine protease (MASP)-1 is crucial for lectin pathway activation in human serum, whereas neither MASP-1 nor MASP-3 is required for alternative pathway function. *J Immunol*. **189**(8):3957-3969.

Degn, S.E., Jensen, L., Olszowski, T., Jensenius, J.C., and Thiel, S. (2013) Co-complexes of MASP-1 and MASP-2 associated with the soluble pattern-recognition molecules drive lectin pathway activation in a manner inhibitable by MAp44. *J Immunol*. **191**(3):1334-1345.

Dobó, J., Harmat, V., Beinrohr, L., Sebestyén, E., Závodszky, P., and Gál, P. (2009) MASP-1, a promiscuous complement protease: structure of its catalytic region reveals the basis of its broad specificity. *J Immunol*. **183**:1207-1214.

Dommett, R.M., Klein, N., and Turner, M.W. (2006) Mannose-binding lectin in innate immunity:past, presentand future. *Tissue Antigens*. **68**(3):193-209.

Duncan, R. C., Wijeyewickrema, L. C., and Pike, R.N. (2008) The initiating proteases of the complement system: controlling the cleavage. *Biochimie*. **90**(2):387-395.

Duncan, R.C., Bergström, F., Coetzer, T.H., Blom, A.M., Wijeyewickrema, L.C., and Pike, R.N. (2012) Multiple domains of MASP-2, an initiating complement protease, are required for interaction with its substrate C4. *Mol Immunol*. **49**(4):593-600.

Dunkelberger, J.R., and Song, W.C. (2010) Complement and its role in innate and adaptive immune responses. *Cell Res*. **20**:34-50.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. **66**:486-501.

Endo, Y., Nonaka, M., Saiga, H., Kakinuma, Y., Matsushita, A., Takahashi, M., Matsushita, M., and Fujita, T. (2003) Origin of Mannose-Binding Lectin-Associated Serine Protease (MASP)-1 and MASP-3 Involved in the Lectin Complement Pathway Traced Back to the Invertebrate, Amphioxus. *J Immunol*. **170**:4701-4707.

Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H., Matsushita, M., Nonaka, M., and Fujita, T. (1998) Two lineages of Mannose-Binding Lectin-Associated Serine Protease (MASP) in vertebrates. *J Immunol*. **161**:4924-4930.

Essalmani, R., Zaid, A., Marcinkiewicz, J., Chamberland, A., Pasquato, A., Seidah, N.G., and Prat, A. (2008) In vivo functions of the proprotein convertase PC5/6 during mouse development: Gdf11 is a likely substrate. *Proc Natl Acad Sci U S A*. **105**(15):5750-5755.

Esser, A.F. (1994) The membrane attack complex of complement. Assembly, structure and cytotoxic activity. *Toxicol*. **87**:229-247.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*. **62**:72-82.

Feinberg, H., Uitdehaag, J.C., Davies, J.M., Wallis, R., Drickamer, K., and Weis, W.I. (2003) Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2. *EMBO J*. **22**(10):2348-2359.

Frank, M.M., and Fries, L.F. (1991) The role of complement in inflammation and phagocytosis. *Immunol Today*. **12**:322-326.

Fujita, T. (2002) Evolution of the lectin-complement pathway and its role in innate immunity. *Nat Rev Immunol*. **2**:346-353.

Gaboriaud, C., Gupta, R.K., Martin, L., Lacroix, M., Serre, L., Teillet, F., Arlaud, G.J., Rossi, V., and Thielens, N.M. (2013) The serine protease domain of MASP-3:

enzymatic properties and crystal structure in complex with ecotin. *PLoS One*. **8**(7):e67962.

Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J.C., and Arlaud, G.J. (2003) The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem*. **278**:46974-46982.

Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G.J. and Fontecilla-Camps, J.C. (2000) Crystal structure of the catalytic domain of human complement C1s: a serine protease with a handle. *EMBO J*. **19**(8):1755-1765.

Gaboriaud, C., Teillet, F., Gregory, L.A., Thielens, N.M., and Arlaud, G.J. (2007) Assembly of C1 and the MBL- and ficolin-MASP complexes: structural insights. *Immunobiolo*. **212**(4-5):279-288.

Gaboriaud, C., Thielens, N.M., Gregory, L.A., Rossi, V., Fontecilla-Camps, J.C., and Arlaud, G.J. (2004) Structure and activation of the C1 complex of complement: unravelling the puzzle. *Trends Immunol*. **25**:368-373.

Gál, P., Dobó, J., Závodszky, P., and Sim, R.B.M. (2009) Early complement proteases: C1r, C1s and MASPs. A structural insight into activation and functions, *Mol Immunol*. **46**:2745–2752.

Gál, P., Harmat, V., Kocsis, A., Bián, T., Barna, L., Ambrus, G., Végh, B., Balczer, J., Sim, R.B., Náray-Szabó, G., and Závodszky, P. (2005) A true autoactivating enzyme. Structural insight into mannose-binding lectin-associated serine protease-2 activations. *J Biol Chem*. **280**(39):33435-33444.

Goldring, J.P., Thobakgale, C., Hiltunen, T., and Coetzer, T.H. (2005) Raising antibodies in chickens against primaquine, pyrimethamine, dapsone, tetracycline, and doxycycline. *Immunol Invest.* **34**:101-114.

Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, G., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradszky, K., and Rutter, W.J. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc Natl Acad Sci U S A*. **85**(14):4961-4965.

Gregory, L.A., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., and Gaboriaud, C. (2003) X-ray structure of the Ca2+-binding interaction domain of C1s. Insights into the assembly of the C1 complex of complement. *J Biol Chem*. **278**(34):32157-32164.

Gregory, L.A., Thielens, N.M., Matsushita, M., Sorensen, R., Arlaud, G.J., Fontecilla-Camps, J.C., and 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**(28):29391-29397.

Hajela, K., Kojima, M., Ambrus, G., Wong, K.H., Moffatt, B.E., Ferluga, J., Hajela, S., Gál, P., and Sim, R.B. (2002) The biological functions of MBL-associated serine proteases (MASPs). *Immunobiol*. **205**:467-475.

Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M.O., Nielsen, O., Hartshorn, K., Jørgensen, T.J.D., Skjødt, K., and Holmskov, U. (2010) Collectin 11 (CL-11, CL-K1) is a MASP-1/-3-associated plasma collectin with microbial-binding activity. *J Immuol*. **185**:6096-6104.

Harmat, V., Gál, P., Kardos, J., Szilágyi, K., Ambrus, G., Végh, B., Náray-Szabó, G., and Závodszky, P. (2004) The structure of MBL-associated serine protease-2 reveals that identical substrate specificities of C1s and MASP-2 are realized through different sets of enzyme-substrate interactions. *J Mol Biol*. **342**(5):1533-1546.

Holmskov, U., Thiel, S., and Jensenius, J.C. (2003) Collections and ficolins: humoral lectins of the innate immune defense. *Annu Rev Immunol*. **21**:547-578.

Ichijo, H., Hellman, U., Wernstedt, C., Gonez, L.J., Claesson-Welsh, L., Heldin, C.H., and Miyazono, K.J. (1993) Molecular cloning and characterization of ficolin, a multimeric protein with fibrinogen- and collagen-like domains. *J Biol Chem*. **268**(19):14505-14513.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., and Yamashina, I. (1987) Serum lectin with known structure activates complement through the classical pathway. *J Biol Chem*. **262**:7451-7454.

Iwaki, D., Kanno, K., Takahashi, M., Endo, Y., Lynch, N.J., Schwaeble, W.J., Matsushita, M., Okabe, M., and Fujita, T. (2006) Small mannose-binding lectin-

associated protein plays a regulatory role in the lectin complement pathway. *J Immunol*. **177**:8626-8632.

Iwaki, D., Kanno, K., Takahashi, M., Endo, Y., Matsushita, M., and Fujita, T. (2011) The role of mannose-binding lectin-associated serine protease-3 in activation of the alternative complement pathway. *J Immunol*. **187**(7):3751-2758.

Ji, X., Azumi, K., Sasaki, M., and Nonaka, M. (1997) Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian Halocynthia roretzi. *Proc Natl Acad Sci U S A*. **94**:6340-6345.

Ji, Y.H., Fujita, T., Hatsuse, H., and Kawakami, M. (1993) Activation of the C4 and C2 components of complement by a proteinase in serum bactericidal factor, Ra reactive factor. *J Immunol*. **150**:571-578.

Kabsch, W. (1988a) Automatic indexing of rotation diffraction patterns. *J Appl Cryst*. **21**:67-72.

Kabsch, W. (1988b) Evaluation of single-crystal X-ray diffraction data from a positionsensitive detector. *J Appl Cryst*. **21**:916-924.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst*. **26**:795-800.

Kardos, J., Harmat, V., Palló, A., Barabás, O., Szilágyi, K., Gráf, L., Náray-Szabó, G., Goto, Y., Závodszky, P., and Gál, P. (2008) Revisiting the mechanism of the autoactivation of the complement protease C1r in the C1 complex: structure of the active catalytic region of C1r. *Mol Immunol*. **45**:1752-1760.

Kardos, J., Gál, P., Szilágyi, L., Thielens, N.M., Szilágyi, K., Lõrincz, Z., Kulcsár, P., Gráf, L., Arlaud, G.J., and Závodszky, P. (2001) The role of the individual domains in the structure and function of the catalytic region of a modular serine protease, C1r. *J Immunol*. **167**(9):5202-5208.

Kawasaki, N., Kawasaki, T., and Yamashina, I. (1989) As erum lectin (mannan-binding protein) has complement-dependent bactericidal activity. *J Biochem Tokyo*. **106**:483-489.

Khan, A.R., and James, M.N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci*. **7**(4):815-836.

Kidmose, R.T., Laursen, N.S., Dobó, J., Kjaer, T.R., Sirotkina, S., Yatime, L., Sottrup-Jensen, L., Thiel, S., Gál, P., and Andersen, G.R. (2012) Structural basis for activation of the complement system by component C4 cleavage. *Proc Natl Acad Sci U S A*. **109**(38):15425-15430.

Knittel, T., Fellmer, P., Neubauer, K., Kawakami, M., Grundmann, A., and Ramadori, G. (1997) The complement activating protease P100 is expressed by hepatocytes and is induced by IL-6 in vitro and during the acute phase reaction in vivo. *Lab Invest*. **77**:221–230.

Kocsis, A., Kékesi, K.A., Szász, R., Végh, B.M., Balczer, J., Dobó, J., Závodszky, P., Gál, P., and Pál, G. (2010) Selective inhibition of the lectin pathway of complement with phage display selected peptides against mannose-binding lectin-associated serine protease (MASP)-1 and -2: significant contribution of MASP-1 to lectin pathway activation. *J Immunol*. **185**(7):4169-4178.

Krarup, A., Presanis, J.S., Gál, P., and Sim, R.B. (2005) Exploring the substrate activity of MBL-associated serine protease 1 (MASP-2). *Immunol*. **116**(supplement 1), 96.

Krem, M.M., and Di Cerca, E. (2001) Molecular markers of serine protease evolution. *EMBO J*. **20**:3036-3045.

Kuraya, M., Matsushita, M., Endo, Y., Thiel, S., and Fujita, T. (2003) Expression of H-ficolin/Hakata antigen, mannose-binding lectin-associated serine protease (MASP)-1 and MASP-3 by human glioma cell line T98G. *Int Immunol*. **15**(1):109-117.

Lachmann, P.J. (2009) The amplification loop of the complement pathways. *Adv Immunol*. **104**:115-149

Lachmann, P. J., and Hughes-Jones, N.C. (1984) Initiation of complement activation. *Spring Semin Immunopathol*. **7**:143-162.

Leal, G.F., Silva, E.O., Duarte, A.R., and Campos, J.F. (2008) Blepharophimosis, blepharoptosis, defects of the anterior chamber of the eyes, caudal appendage, radioulnar synostosis, hearing loss and umbilical anomalies in sibs: 3MC syndrome? *Am J Med Genet A*. **146A**:1059-1062.

Lu, J., Tay, P.N., Kon, O.L., and Reid, K.B. (1996) Human ficolin, cDNA cloning, demonstration of peripheral blood leukocytes as the major site of synthesis and assignment of the gene to chromosome 9. *Biochem J*. **313**:473-478.

Lynch, N.J., Khan, S-H., Stover, C.M., Sandrini, S.M., Presanis, J.S., and Schwaeble, W.J. (2005) Composition of the Lectin Pathway of Complement in Gallus gallus: Absence of Mannan-Binding Lectin-Associated Serine Protease-1 in Birds. *J Immunol*. **174**:4998-5006.

Major, B., Kardos, J., Kékesi, K.A., Lorincz, Z., Závodszky, P., and Gál, P. (2010) Calcium-dependent conformational flexibility of a CUB domain controls activation of the complement serine protease C1r. *J Biol Chem*. **285**(16):11863-11869.

Markiewski M.M., and Lambris, J.D. (2007) The role of complement in inflammatory diseases from behind the scenes into the spotlight. *Am J Pathol*. **171**:715-727.

Matsushita, M., Endo, Y., and Fujita, T. (2013) Structural and functional overview of the lectin complement pathway: its molecular basis and physiological implication. *Arch Immunol Ther Exp*. **61**:273–283.

Matsushita M., Endo, Y., Taira. S., Sato, Y., Fujita, T., Ichikawa, N., Nakata, M., and Mizuochi, T. (1996) A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin. *J Biol Chem*. **271**:2448-2454.

Matsushita, M., and 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., and Fujita, T. (1995) Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. *Immunol*. **194**:443-448.

Matsushita, M., Endo, Y., and Fujita, T. (2000b) Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine-protease. *J Immunol*. **164**:2281-2284.

Matsushita, M., Thiel, S., Jensenius, J.C., Terai, I., and Fujita, T. (2000a) Proteolytic activities of two types of mannose-binding lectin-associated serine protease. *J Immunol*. **165**:2637-2642.

Mayilyan, K.R., Presanis, J.S., Arnold, J.N., and Sim, R.B. (2006) Discrete MBL-MASP complexes show wide inter-individual variability in concentration: data from UK vs Armenian populations. *Int J Immunopathol Pharmacol*. **19**(3):567–580.

Megyeri, M., Harmat, V., Major, B., Végh, Á., Balczer, J., Héja, D., Szilágyi, K., Datz, D., Pál, G., Závodszky, P., Gál, P., and Dobó. J. (2013) Quantitative characterization of the activation steps of mannan-binding lectin (MBL)-associated serine proteases (MASPs) points to the central role of MASP-1 in the initiation of the complement lectin pathway. *J Biol Chem*. **288**(13):8922-8934.

Menegatti, E., Guarneri, M., Bolognesi, M., Ascenzi, P., and Amiconi, G. (1985) Activating effect of the Ile-Val dipeptide on the catalytic properties of bovine trypsinogen. *Biochim Biophys Acta*. **832**(1):1-6.

Møller-Kristensen, M., Thiel, S., Sjöholm, A., Matsushita, M., and Jensenius, J.C. (2007) Cooperatoin between MASP-1 and MASP-2 in the generation of C3 convertase through the MBL pathway. *Int Immunol*. **19**:141-149.

Morgan, B.P., and Walport, M.J. (1991) Compplement deficiency and disease. *Immunol Today*. **12**(9):301-306.

Øhlenschlæger, T., Garred, P., Madsen, H.O., and Jacobsen, S. (2004) Mannose-binding lectin variant alleles and the risk of arterial thrombosis in systemic lupus erythematosus. *N Engl J Med*. **351**:260-267.

Pangburn, M.K., and Muller-Eberhard, H.J. (1984) The alternative pathway of complement. *Spring Semin Immunopathol*. **7**:163-192.

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., and Tucker, P.A. (2005) Autorickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr*. **61**:449-457.

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., and Tucker, P.A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr*. **65**:1089-1097.

Papadimitriou, J.C., Ramm, L.E., Drachenberg, C.B., Trump, B.F., and Shin, M.I. (1991) Quantitative analysis of adenine nucleotides during the prelytic phase of cell death mediated by C5-9. *J Immunol*. **147**:212-217.

Perry, A.J., Wijeyewickrema, L.C., Wilmann, P.G., Gunzburg, M.J., D'Andrea, L., Irving, J.A., Pang, S.S., Duncan, R.C., Wilce, J.A., Whisstock, J.C., and Pike, R.N. (2013) A molecular switch governs the interaction between the human complement protease C1s and its substrate, complement C4. *J Biol Chem*. **288**(22):15821-9.

Petersen, S.V., Thiel, S., Jensen, L., Vorup-Jensen, T., Koch, C. and Jensenius, J.C. (2000) Control of the classical and the MBL pathway of complement activation. *Mol Immunol*. **37**:803-811.

Petersen, S.V., Thiel, S., and Jensenius, J.C. (2001) The mannan-binding lectin pathway of complement activation: biology and disease association. *Mol Immunol*. **38**:133-149.

Presanis, J.S., Hajela, K., Ambrus, G., Gál, P and Sim, R.B. (2004) Differential substrate and inhibitor profiles for human MASP-1 and MASP-2. *Mol Immunol*. **40**(3):921-929.

Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G.G., and Stuart, D. (1995) The structure of a Ca2+-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell*. **82**:131−141.

Rawal, N., Rajagopalan, R., and Salvi, V.P. (2008) Activation of complement component C5: coparison of C5 convertases of the lectin pathway and classical pathway of complement. *J Biol Chem*. **283**:7853-7863.

Reid, K.B., Bentley, D.R., Campbell, R.D., Chung, L.P., Sim, R.B., Kristensen, T., and Tack, B.F. (1986) Complement system proteins which interact with C3b or C4b. *Immunol Today*. **7**:230-234.

Roebroek, A.J., Umans, L., Pauli, I.G., Robertson, E.J., van Leuven, F., Van de Ven, W.J., and Constam, D.B. (1998) Failure of ventral closure and axial rotation in embryos lacking the proprotein convertase Furin. *Development*. **125**(24):4863-4876.

Romao, M.J., Kolln, I., Dias, J.M., Carvalho, A.L., Romero, A., Varela, P.F., Sanz, L., Töpfer-Petersen, E., and Calvete, J.J. (1997) Crystal structure of acid seminal fluid protein (aSFP) at 1.9 Å resolution: a bovine polypeptide of the spermadhesin family. *J Mol Biol*. **274**:650-660.

Rossi, V., Cseh, S., Bally, I., Thielens, N.M., Jensenius, J.C., and Arlaud, G.J. (2001) Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. *J Biol Chem*. **276**:40880-40887.

Rooryck, C., Diaz-Font, A., Osborn, D. P., 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., and Beales, P. L. (2011) Mutations in lectin complement pathway genes COLEC11 and MASP1 cause 3MC syndrome. *Nat Genet*. **43**:197-203.

Sato T., Endo, Y., Matsushita M., and Fujita, T. (1994) Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Int Immunol*. **6**(4):665-669.

Schechter, I., and Berger, A. (1967) On the size of the active site in protease. I. Papain. *Biochem Biophys Res Commun*. **27**:157-162.

Schumaker, V.N., Zavodszky, P., and Poon, P.H. (1987) Activation of the first component of complement. *Annu Rev Immunol*. **5**:21-42.

Seidah, N.G., and Chrétien, M. (1999) Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res*. **848**(1-2):45-62.

Seidah, N.G., and Prat, A. (2012) The biology and therapeutic targeting of the proprotein convertases. *Nat Rev Drug Discov*. **11**(5):367-383.

Selander, B., Martensson, U., Weintraub, A., Holmstrom, E., Matsushita, M., Thiel, S., Jensenius, J.C., Truedsson, L., and Sjoholm, A.G. (2006) Mannan-binding lectin activates C3 and the alternative complement pathway without involvement of C2. *J Clin Invest*. **116**(5):1425-1434.

Selman, L., and Hansen, S. (2012) Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). *Immunobiolo*. **217**(9):851-863.

Seyfarth, J., Garred, P., and Madsen, H.O. (2006) Extra-hepatic transcription of the human mannose-binding lectin gene (mbl2) and the MBL-associated serine protease 1-3 genes. *Mol Immunol*. **43**:962-971.

Sheehan, M., Morris, C.A., Pussell, B.A., and Charlesworth, J.A. (1995) Complement inhibition by human vitronectin involves non-heparin binding domains. *Clin Exp Immunol*. **101**(1):136-141.

Sim, R.B., and Tsiftsoglou, S.A. (2004) Protease of the complement system. *Biochem Soc Tran*. **32**:21-27.

Sirmaci, A., Walsh, T., Akay, H., Spiliopoulos, M., Sakalar, Y. B., Hasanefendioglu-Bayrak, A., Duman, D., Farooq, A., King, M. C., and Tekin, M. (2010) MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am J Hum Genet*. **87**:679-686.

Skjoedt, M.O., Hummelshoj, T., Palarasah, Y., Honore, C., Koch, C., Skjodt, K., and Garred, P. (2010) A novel mannose-binding lectin/ficolin-associated protein is highly expressed in heart and skeletal muscle tissues and inhibits complement activation. *J Biol Chem*. **285**(11):8234-8243.

Skjoedt, M.O., Palarasah, Y., Munthe-Fog, L., Ma, Y.J., Weiss, G., Skjodt, K., Koch, C., and Garred, P. (2009) MBL-associated serine protease-3 circulates in high serum concentrations predominantly in complex with ficolin-3 and regulates ficolin-3 mediated complement activation. *Immunol*. **215**(11):921-931.

Skjoedt, M.O., Roversi, P., Hummelshøj, T., Palarasah, Y., Rosbjerg, A., Johnson, S., Lea, S.M., and Garred, P. (2012) Crystal structure and functional characterization of the complement regulator mannose-binding lectin (MBL)/ficolin-associated protein-1 (MAP-1). *J Biol Chem*. **287**(39):32913-32921.

Sørensen, R., Thiel, S., and Jensenius, J.C. (2005) Mannose-binding lectin-associated serine proteases, characteristic and disease association. *Springer Semin Immune*. **27**:299-319.

Stengaard-Pedersen, K., Thiel, S., Gadjeva, M., Møller-Kristensen, M., Sørensen, R., Jensen, L.T., Sjøholm, A.G., Fugger, L., and Jensenius, J.C. (2003) Inherited deficiency of mannan-binding lectin-associated serine protease 2. *N Engl J Med*. **349**(6):554-560.

Stover, C.M., Lynch, N.J., Dahl, M.R., Hanson, S., Takahashi, M., Frankenberger, M., Ziegler-Heitbrock, L., Eperon, I., Thiel, S., and Schwaeble, W.J. (2003) Murine serine proteases MASP-1 and MASP-3, complements of the lectin pathway activation complex of complement, are encoded by a single structural gene. *Gene and Immun*. **4**:374-384.

Stover, C.M., Schwaeble, W. J., Lynch, N. J., Thiel, S., and Speicher, M. R. (1999a) Assignment of the gene encoding mannanbinding-lectin-associated serine protease 2 (MASP-2) to human chromosome 1p36.3→p36.2 by in situ hybridization and somatic-cell hybrid analysis. *Cytogenet Cell Genet*. **84**:148–149.

Stover, C.M., Thiel, S., Lynch, N.J., and Schwaeble, W.J. (1999b) The rat and mouse homologues of MASP-2 and MAp19, components of the lectin activation pathway of complement. *J Immunol*. **163**:6848-6859.

Super, M., Thiel, S., Lu, J., Levinsky, R.J., and Turner, M.W. (1989) Association of low levels of mannan-binding protein with a common defect in opsonisation. *Lancet*. **2**:1236-1239.

Szumska, D., Pieles, G., Essalmani, R., Bilski, M., Mesnard, D., Kaur, K., Franklyn, A., El Omari, K., Jefferis, J., Bentham, J., Taylor, J.M., Schneider, J.E., Arnold, S.J., Johnson, P., Tymowska-Lalanne, Z., Stammers, D., Clarke, K., Neubauer, S., Morris, A., Brown, S.D., Shaw-Smith, C., Cama, A., Capra, V., Ragoussis, J., Constam, D., Seidah, N.G., Prat, A., and Bhattacharya, S. (2008) VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5. *Genes Dev*. **22**(11):1465-1477.

Takada, F., Seki, N., Matsuda, Y., Takayama, Y., and Kawakami, M. (1995) Localization of the genes for the 100-kDa complement-activating component of Ra-reactive factor (CRARF and Craft) to human 3q27-28 and mouse 16B2-B3. *Genomics*. **25**:757-759.

Takada, F., Takayama, Y., Hatsuse, H., and Kawakami, M. (1993) A new member of the C1s family of complement proteins found in a bactericidal factor, Ra-active factor, in human serum. *Biochem Biophys Res Commun*. **196**:1003-1009.

Takahashi, A., Takayama, Y., Hatsuse, H., and Kawakami, M. (1993) Presence of a serine protease in the complement-activating component of the complement-dependent bactericidal factor, RaRF, in mouse serum. *Biochem Biophys Res Commun*. **196**:1003-1009.

Takahashi, M., Endo, Y., Fujita, T., and 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., and Fujita, T. (2010) Essential role of mannose-binding lectin-associated serine protease-1 in activating of the complement factor D. *J Exp Med*. **207**:29-37.

Takahashi, M., Iwaki, D., Kanno, K., Ishida, Y., Xiong, J., Matsushita, M., Endo, Y., Miura, S., Ishii, N., Sugamura, K., and Fujita, T. (2008) Mannose-binding lectin (MBL)-associated serine protease (MASP)-1 contributes to activation of the lectin complement pathway. *J Immunol*. **180**:6132-6138.

Teillet, F., Dublet, B., Andrieu, J.P., Gaboriaud, C., Arlaud, G.J., and Thielens, N.M. (2005) The two major oligomeric forms of human mannan-binding lectin: chemical characterization, carbohydrate-binding properties, and interaction with MBL-associated serine proteases. *J Immunol*. **174**(5):2870-2877.

Teillet, F., Gaboriaud, C., Lacroix, M., Martin, L., Arlaud, G.J., and Thielens, N.M. (2008) Crystal structure of the CUB1-EGF-CUB2 domain of human MASP-1/3 and identification of its interaction sites with mannan-binding lectin and ficolins. *J Biol Chem*. **283**(37):25715-25724.

Teillet, F., Lacroix, M., Thiel, S., Weilguny, D., Agger, T., Arlaud, G.J., and Thielens, N.M. (2007) Identification of the site of human mannan-binding lectin involved in the interaction with its partner serine proteases: the essential role of Lys55. *J Immunol*. **178**:5710-5716.

Terai, I., Kobayashi, K., Matsushita, M., and Fujita, T. (1997) Human serum mannose-binding lectin (MBL)-associated serine protease-1 (MASP-1): determination of levels in body fluids and identification of two forms in serum. *Clin Exp Immunol*. **110**:317-323.

Terai, I., Kobayashi, K., Matsushita, M., Fujita, T., Matsuno, K., and Okumura, K. (1995) α2-Macroglobulin binds to and inhibits mannose-binding protein-associated serine protease. *Int Immunol*. **7**(10):1579–1584.

Thiel, S., Peteesen, S.V., Vorup-Jensen, T., Matsushita, M., Fujita, T., Stover, C.M., Schwaeble, W.J., and Jensenius, J.C. (2000) Interaction of C1q and mannan-binding lectin (MBL) with C1r, C1s, MBL-associated serine proteases 1 and 2, and the MBL-associated protein Map19. *J Immunol*. **165**:878-887.

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.M., and Jensenius, J.C. (1997) A second serine protease associated with mannan-binding lectin that activates complement, *Nature*. **386**:506-510.

Thiel, S., Jensen, L., Degn, S.E., Nielsen, H.J., Gál, P., Dobó, J., and Jensenius, J.C. (2012) Mannan-binding lectin (MBL)-associated serine protease-1 (MASP-1), a serine protease associated with humoral pattern-recognition molecules: normal and acute-phase levels in serum and stoichiometry of lectin pathway components. *Clin Exp Immunol*. **169**(1):38-48.

Thielens, N.M., Cseh, S., Thiel, S., Vorup-Jensen, T., Rossi, V., Jensenius, J.C., and Arlaud, G.J. (2001) Interaction properties of human mannan-binding lectin (MBL)-associated serine protease-1 and -2, MBL-associated protein 19, and MBL. *J Immunol*. **166**:5068-5077.

Thomas, D.A., Francis, P., Smith, C., Ratcliffe, S., Ede, N.J., Kay, C., Wayne, G., Martin, S.L., Moore, K., Amour, A., and Hooper, N.M. (2006) A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates. *Proteomics*. **6**(7):2112-2120.

Turner, M.W., and Hamvas, R.M. (2000) Mannose-binding lectin:structure, function, genetics and disease association. *Rev Immunogenet*. **2**:305-322.

Turpeinen, H., Kukkurainen, S., Pulkkinen, K., Kauppila, T., Ojala, K., Hytönen, V.P., and Pesu, M. (2011) Identification of proprotein convertase substrates using genome-wide expression correlation analysis. *BMC Genomics*. **12**:618-627.

Wallis, R., and Dodd, R.B. (2000) Interaction of mannose binding protein with associated serine proteases. Effects of naturally occurring mutations. *J Biol Chem*. **275**(40):30962-30969.

Walport, M.J. (2001) Complement, First of two parts. *N Engl J Med*. **344**(14):1058-1066.

Weis, W.I., Drickamer, K., and Hendrickson, W.A. (1992) Structure of a C-type mannose-binding protein complexed with an oligosacch. *Nature*. **360**:127-134.

Wijeyewickrema, L.C., Yongqing, T., Tran, T.P., Thompson, P.E., Viljoen, J.E., Coetzer, T.H., Duncan, R.C., Kass, I., Buckle, A.M., and Pike, R.N. (2013) Molecular determinants of the substrate specificity of the complement-initiating protease, C1r. *J Biol Chem*. **288**(22):15571-15580.

Winter, G. (2009) Xia2: an expert system for macromolecular crystallography data reduction. *J Appl Cryst*. **43**:186-190.

Wong, N.K., Kojima, M., Dobó, J., Ambrus, G., and Sim, R.B. (1999) Activation of the MBL-associated serine proteases (MASPs) and their regulation by natural inhibitors. *Mol Immunol*. **36**:853-861.

Xu, D., Baburaj, K., Peterson, C.B., and Xu, Y. (2001) Model for the three-dimensional structure of vitronectin: predictions for the multi-domain protein from threading and docking. *Proteins*. **44**(3):312-320.

Yae, Y., Inaba, S., Sato, H., Okochi, K., Tokunaga, F., and Iwanaga, S. (1991) Isolation and characterization of a thermolabile beta-2 macroglycoprotein ('thermolabile substance' or 'Hakata antigen') detected by precipitating (auto) antibody in sera of patients with systemic lupus erythematosus. *Biochem Biophys Acta*. **1078**(3):369-376.

Ytting, H., Christensen, I.J., Thiel, S., Jensenius, J.C., and Nielsen, H.J. (2005) Serum mannan-binding lectin-associated serine protease 2 levels in colorectal cancer: relation to recurrence and mortality. *Clin Cancer Res*. **11**(4):1441-1446.

Yongqing, T., Drentin, N., Duncan, R.C., Wijeyewickrema, L.C., and Pike, R.N. (2012) Mannose-binding lectin serine proteases and associated proteins of the lectin pathway of complement: two genes, five proteins and many functions? *Biochim Biophys Acta*. **1824**(1):253-262.

Zhou, A., Huntington, J.A., Pannu, N.S., Carrell, R.W., and Read, R.J. (2003) How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration. *Nat Struct Biol*. **10**(7):541-544.

Zundel, S., Cseh, S., Lacroix, M., Dahl, M.R., Matsushita, M., Andrieu, J.P., Schwaeble, W.L., Jensenius, J.C., Fujita, T., Arlaud, G.J., and Thielens, N.M. (2004) Characterization of recombinant mannan-binding lectin-associated serine protease (MASP)-3 suggests an activation mechanism difference from that of MASP-1 and MASP-2. *J Immunol*. **172**:4342-4350.