

The organisation and conservation of the  
mannan-binding lectin (MBL) associated serine  
protease-3 (MASP-3) in mammals, amphibia,  
reptiles and bony fish

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Steven Hanson BSc, MBChB

Department of Infection, Immunity & Inflammation

University of Leicester



September 2009

For Brian Hartley

# Abstract

Mannan binding lectin associated serine protease (MASP) are found associated with Mannan binding lectin (MBL) and ficolin, two alternative recognition molecules involved in activation of the lectin arm of the complement cascade. MBL and ficolin (H-ficolin, L-ficolin and M-ficolin) recognise carbohydrate moieties on the surface of pathogens; a resulting conformational change in the recognition molecule converts the associated MASP proenzyme to its active form.

MBL-MASP complexes play an essential role in the innate immune response in the period before an epitope specific adaptive antibody driven solution is found, to date there are three known MASP.

MASP and MASP-like proteins generate much interest due to their unique position in the activation apparatus of the complement pathway. Being the point of the first proteolytic cleavage for subsequent cleavage events, MASPs are an ideal point to apply a brake to any spurious activation leading to disease. The aim of this study is to further the knowledge of MASP-3 role in the lectin pathway and establish its connection with other members of the MASP family of proteins.

MASP-3 is the most recently characterised mannan binding lectin associated serine protease, here is presented the cloning, characterisation and recombinant expression of this novel complement serine protease in a number of mammalian and non-mammalian species. MASP-3 demonstrates a striking degree of sequence conservation between species.

MASP-3 is an alternatively spliced transcript of the MASP-1/3 locus; this study identifies in the current sequence database a truncated third alternatively splice transcript in human, rhesus monkey, rat, mouse and porcine species.

## Acknowledgements

Professor Wilhelm Schwaeble, for his continued and much appreciated support from the beginning.

Dr.Nick Lynch, Dr.Cordula Stover and all past members of Lab 231 for their help and advice.

To my Mum and family who enabled me to begin my studies due to their support, love and guidance.

To Sophie for the constant love and support.

## Standard abbreviations

A	adenosine
aa	amino acid
AP	alkaline phosphatase
ATP	adenosine triphosphate
bp	base pair
BMP	bone morphogenic protein
BSA	bovine serum albumin
CCP	complement control protein
cDNA	complementary DNA
cfu	colony forming units
cm	centimetre
CR	complement receptor
CRD	carbohydrate recognition domain
CRP	C-reactive protein
CUB	C1r/C1s/uEGF/BMP
Cx	complement factor number x
Da	Dalton
DAF	decay-accelerating factor
dATP	dideoxy adenosine triphosphate
dCTP	dideoxy cytidine triphosphate
DEPC	diethylpyrocarbonate
dGTP	dideoxy guanosine triphosphate
dH <sub>2</sub> O	distilled water
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dNTP	dideoxy nucleotide triphosphate
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
dTTP	dideoxy thymidine triphosphate
EBV	Epstein Barr virus
ECL	enhanced chemi-luminescence
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
ELISA	enzyme linked immunosorbent assay
EST	express sequence tags
FCN	ficolin
FPLC	fast performance liquid chromatography
g	gram
GalNAc	N-acetyl-D-galactosamine
GlcNAc	N-acetyl-D-glucosamine
HIV	human immunodeficiency virus
hr	hour
Ig	immunoglobulin
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
kb	kilo base pair
kDa	kilo Dalton
l	litre
LPS	lipopolysaccharide
M	molar
MAC	membrane attack complex
MAP	mannan-binding lectin association protein
MASP	mannan-binding lectin associated serine protease
MBL	mannan-binding lectin
MCP	membrane cofactor of proteolysis
MCS	multiple cloning site

MHC	major histocompatibility complex
mg	milligram
μl	micro litre
ml	millilitre
μm	micrometer
mm	millimetre
mM	millimolar
MOPS	3-[N-Morpholino] propanesulfonic acid
mRNA	messenger ribonucleic acid
NBT	nitroblue tetrazolium
NCBI	national centre for biotechnology information
ng	nanogram
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pfu	plaque forming units
RACE	rapid amplification of cDNA ends
rRNA	ribosomal RNA
RNA	ribonucleic acid
RNase	ribonuclease
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
SCR	short consensus repeat
SDS	sodium dodecyl sulfate
SLE	systemic lupus erythematosus
SSC	saline sodium citrate
TA	annealing temperature
TAE	tris-acetate-EDTA
TE	tris EDTA
TEMED	N,N,N.'N'-tetramethylethylenediamine
T <sub>m</sub>	melting temperature
UV	ultraviolet
V	volts
v/v	volume/volume
w/v	weight/volume

## List of figures

Fig.1 Schematic diagram of the Classical, Lectin and Alternative pathways of complement activation .....	6
Fig.2 A simplified representation of invertebrate and vertebrate immune systems.....	10
Fig.3 Specialisation of lymphoid tissues within the vertebrate phyla .....	12
Fig.4 Evolutionary processes of the complement system. Evolution and origins of the complement pathways.....	14
Fig.5 Schematic diagram of promotor and exon 1 mutations of the <i>mb12</i> locus.....	32
Fig.6 Protein domain structure of Mannan binding lectin Associated Serine protease (MASP) and Map19 .....	49
Fig.7 MASP-1 protein domain structure.....	61
Fig.8 MASP-2 protein domain structure.....	64
Fig.9 Map19 protein domain structure.....	66
Fig.10 MASP-2 locus.....	67
Fig.11 MASP-3 protein domain structure.....	68
Fig.12 MASP-1/3 locus exon structure.....	70
Fig.13 Human MASP-1/3 structural model of CUBI-EGF-CUBII domains .....	72
Fig.14 Human MASP-1/3 structural model of CUBI-EGF-CUBII domains .....	73
Fig.15 Model of rat MASP-3 CCPI-CCPII-Serine protease domain .....	74
Fig.16 pGEM-T Easy circle map and sequence reference points.....	89
Fig.17 Schematic overview of 5'RACE methodology.....	102
Fig.18 Schematic overview of 3'RACE methodology.....	106
Fig.19 Prokaryotic expression vector pRSET circle map.....	108
Fig.20 Human MASP-3 specific PCR primer binding positions with respect to MASP-3 protein domain structure.....	118
Fig.21 Gel electrophoresis showing product of human MASP-3 specific PCR utilising primer pairs.....	118
Fig.22 Restriction endonuclease (EcoRI/XhoI) digest of in vivo excised rat MASP-3 acute phase liver cDNA phage library clones (left panel). Southern blot hybridization (right panel) of restriction endonuclease (EcoRI/XhoI) digest of in vivo excised rat .....	119
Fig.23 Restriction endonuclease digestion (EcoRI) of in vivo excised rat MASP-3 acute phase liver cDNA phage library clone 1.1.1 (left panel). Southern blot hybridization (right panel) of restriction endonuclease (EcoRI) digest of in vivo excised rat MASP-3 acute phase liver cDNA phage library clone 1.1.1 .....	120
Fig.24 Restriction endonuclease characterisation of rat MASP-3 acute phase liver cDNA phage library clones (left panel) specifically hybridising with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) (right panel). .....	120
Fig.25 Restriction endonuclease analysis of rat MASP-3 acute phase liver cDNA phage library clones 7.1 and 13.1 specifically hybridising with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F).....	121
Fig.26 Schematic representation of rat MASP-3 cDNA clones characterised in Fig.24 and Fig.25, with respect to MASP-3 gene structure. ....	122

Fig.27 Schematic representation of rat MASP-3 5'RACE strategy and resulting cDNA clone. ....	123
Fig.28 Agarose gel electrophoresis of rat MASP-3 5' RACE PCR (left panel), Southern blot hybridisation (right panel) of rat MASP-3 5' RACE PCR agarose gel (left panel) utilising human MASP-3 specific cDNA probe .....	123
Fig.29 ClustalW 2.0.11 comparative alignment of translated rat MASP-3 (assembled contig, derived from this work) against human MASP-3 coding sequence (accession number AF284421) .....	125
Fig.30 Rat MASP-1/3 protein domain structure (lower panel), linked to respective exons of the rat MASP-1/3 locus (central panel). .....	129
Fig.31 Rat MASP-3 protein domain structure predicted by comparison to the Pfam database.....	131
Fig.32 Western blot analysis of rat MASP-3 recombinant and plasma purified protein .....	131
Fig.33 Rat chromosome 11q23, MASP-1/3 locus with respect to upstream and downstream gene loci .....	132
Fig.34 Agarose gel electrophoresis of mouse MASP-3 RT PCR utilising cDNA derived from mouse total liver RNA (left panel), Southern blot hybridisation (right panel) of mouse MASP-3 RT PCR agarose gel (left panel).....	133
Fig.35 Agarose gel electrophoresis of mouse MASP-3 5' RACE PCR utilising cDNA derived from mouse total liver RNA (left panel), Southern blot hybridisation (right panel) of mouse MASP-3 5' RACE PCR agarose gel (left panel) utilising rat MASP-3 specific cDNA probe (rat MASP-3 PCR p6F + p8F).....	135
Fig.36 Schematic presentation of mouse MASP-3 cDNA clones generated in this work .....	137
Fig.37 ClustalW 2.0.11 multiple sequence alignment of translated mouse MASP-3 (assembled contig, derived from this work) against mouse MASP-3 (accession number AB049755) and human MASP-3 coding sequence (accession number AF284421) .....	138
Fig.38 Mouse MASP-1/3 protein domain structure (lower panel), linked to respective exons of the mouse MASP-1/3 locus (central panel).....	141
Fig.39 Mouse chromosome 16B1 indicating mouse MASP-1/3 locus with respect to upstream and downstream gene loci.....	143
Fig.40 RT-PCR of rabbit cDNA utilising human MASP-3 specific primers.....	144
Fig.41 Schematic diagram of rabbit MASP-3 RT PCR products.....	149
Fig.42 ClustalW 2.0.11 multiple sequence alignment of translated rabbit MASP-3 (assembled contig, derived from this work) against human MASP-3 coding sequence (accession number AF284421).....	150
Fig.43 Rabbit MASP-3 protein domain structure predicted by comparison to the Pfam database.....	152
Fig.44 Guinea pig 5' RACE PCR utilising guinea pig MASP-3.....	154
Fig.45 Schematic diagram of guinea pig MASP-3 specific cDNA clones.....	157
Fig.46 ClustalW 2.0.11 multiple sequence alignment of translated guinea pig MASP-3 (contig derived from this work) against human MASP-3 coding sequence (accession number AF284421) .....	158
Fig.47 Guinea pig MASP-3 protein domain structure predicted by comparison to the Pfam database.....	161
Fig.48 DNA agarose gel electrophoresis of porcine EST clone (BE030550) restriction endonuclease (EcoRI/XbaI) digestion.....	162
Fig.48 Schematic diagram of porcine MASP-3 cDNA sequence.....	162
Fig.49 ClustalW 2.0.11 multiple sequence alignment of translated porcine MASP-3 (contig derived from this work) against human MASP-3 coding sequence (accession number AF284421) .....	163
Fig.50 Porcine chromosome 13 indicating the porcine MASP-1/3 locus with respect to upstream and downstream gene loci. ....	166

Fig.51. DNA agarose gel electrophoresis of bovine EST clone (AW558440) restriction endonuclease (EcoRI/XbaI) digestion. Lane 1 - 1kb DNA marker, lane 2 - bovine EST clone AW558440 EcoRI/XbaI restriction endonuclease digest. ....	167
Fig.52 Schematic diagram of partial bovine MASP-3 sequence.....	168
Fig.53 ClustalW 2.0.11 multiple sequence alignment of translated bovine MASP-3 (contig derived from this work) against human MASP-3 coding sequence (accession number AF284421).....	168
Fig.54 Bovine chromosome 1 indicating the porcine MASP-1/3 locus with respect to upstream and downstream gene loci.....	170
Fig.55 ClustalW 2.0.11 multiple sequence alignment of translated <i>Xenopus</i> MASP-3 (ENSXETT00000042781) against human MASP-3 coding sequence (accession number AF284421).....	172
Fig.56 <i>Danio rerio</i> MASP-3 protein domain structure predicted by comparison to the Pfam database.....	175
Fig.57 ClustalW 2.0.11 multiple sequence alignment of translated <i>Danio rerio</i> MASP-3 (LOC100002060) against human MASP-3 coding sequence (accession number AF284421).....	176
Fig.58 <i>Danio rerio</i> chromosome 15 indicating the <i>Danio rerio</i> MASP-1/3 locus with respect to upstream and downstream gene loci.....	177
Fig.59 ClustalW 2.0.11 multiple sequence alignment of translated <i>Pan troglodytes</i> MASP-3 (sequence derived from this work) against human MASP-3 coding sequence (accession number AF284421).....	180
Fig.60 ClustalW 2.0.11 multiple sequence alignment of translated <i>Pongo pygmaeus</i> MASP-3 (ensemble gene model ENSPPYT00000016741) against human MASP-3 coding sequence (accession number AF284421).....	183
Fig.61 ClustalW 2.0.11 multiple sequence alignment of translated Rhesus macaca MASP-3 (cDNA sequence derived from this work) against human MASP-3 coding sequence (accession number AF284421).....	185
Fig.62 ClustalW 2.0.11 multiple sequence alignment of translated Zebra finch MASP-3 (accession number XP_002191688) against human MASP-3 coding sequence (accession number AF284421).....	188
Fig.63 ClustalW 2.0.11 multiple sequence alignment of translated <i>Anolis carolinensis</i> MASP-3 (Ensembl gene ENSACAT00000006927) against human MASP-3 coding sequence (accession number AF284421).....	191
Fig.64 ClustalW 2.0.11 multiple sequence alignment of translated <i>Tetraodon nigroviridis</i> MASP-3 (Ensembl gene ENSTNIG00000011197 and ENSTNIG00000011199) against human MASP-3 coding sequence (accession number AF284421).....	194
Fig.65 ClustalW 2.0.11 multiple sequence alignment of translated <i>Gasterosteus aculeatus</i> MASP-3 (ENSGACG00000004811) against human MASP-3 coding sequence (accession number AF284421).....	196
Fig.66 Schematic diagram of truncated MASP-1/3 cDNA product predicted domain structure.....	200
Fig.67 ClustalW 2.0.11 multiple sequence alignment of twelve translated MASP-3 (contig derived from this work), human MASP-3 coding sequence (accession number AF284421).....	202
Fig.68 Evolutionary relationships of 35 MASP-3 serine protease domains.....	209
Fig.70 Phylogenetic tree analysis of MASP-like serine protease domain ClustalW alignment.....	211
Fig.71 Northern blot analysis of 10µg total liver RNA from rat and mouse.....	212
Fig.72 Quantitative RT PCR analysis of MASP-1 and MASP-3 mRNA expression in rat liver.....	213
Fig.73 Quantitative RT PCR analysis of MASP-1 and MASP-3 mRNA expression in a panel of human tissue total RNA preparations.....	214
Fig.74 Schematic diagram of the cloning strategy for prokaryotic recombinant expression of rat and human MASP-3.....	216
Fig.75 Agarose gel electrophoresis of human and rat MASP-3 expression cassette PCR (as presented in Fig.74) for restriction endonuclease directed cloning into pRSET prokaryote expression vector.....	217

Fig.76 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human and rat MASP-3 serine protease domain.....	218
Fig.77 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human MASP-3 serine protease domain and pRSET human MASP-3 TXP prokaryotic cell lysate .....	219
Fig.78 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human MASP-3 serine protease domain cell lysate at T1 post induction .....	220
Fig.79 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET rat MASP-3 serine protease domain and rat MASP-3 TXP cell lysate at T1 post induction. ....	221
Fig.80 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET rat MASP-3 serine protease domain T1 post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea).. ....	223
Fig.81 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET rat MASP-3 TXP post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea).. ....	224
Fig.82 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET human MASP-3 serine protease domain and human MASP-3 TXP post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea) .....	225

## List of tables

Table.1 Characterised complement component genes of metazoans.....	13
Table.2 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and rat MASP-1/3 cDNA.....	130
Table.3 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and mouse MASP-1/3 cDNA.....	142
Table.4 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and rabbit MASP-1/3 cDNA.....	151
Table.5 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and guinea pig MASP-1/3 cDNA.....	160
Table.6 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and porcine MASP-1/3 cDNA.....	165
Table.7 Identity between human and bovine MASP-1/3 protein domains as determined by ClustalW pairwise sequence alignment.....	170
Table.8 Identity between human and xenopus MASP-1/3 protein domains as determined by ClustalW pairwise sequence alignment.....	173

# Contents

Introduction.....	2
1.1 The immune system.....	3
1.2 Complement system.....	4
1.3 The immune systems of lower-order organisms.....	9
1.5 Alternative pathway amplification loop.....	22
1.6 Lectin pathway of complement activation.....	25
1.7 Terminal complement pathway.....	29
1.8 Mannan binding lectin.....	31
1.9 Ficolin.....	40
1.10 L-ficolin.....	43
1.11 H-ficolin.....	44
1.12 M-ficolin.....	45
1.13 MBL and ficolin receptors.....	46
1.14 Mannan binding lectin associated serine protease.....	48
1.15 MASP protein domain.....	53
1.16 MASP-1.....	59
1.17 MASP-2.....	63
1.18 MAP19/sMAP.....	66
1.19 MASP-3.....	68
1.20 MASP-3 3-dimensional protein structure.....	71
1.21 Thesis aims.....	75
Materials and Methods.....	76
2.1 Media, bacterial strains, chemicals and radioisotopes.....	77
2.2 DNA manipulation techniques.....	80
2.3 RNA techniques.....	95
2.4 Protein techniques.....	108
2.5 Bioinformatics.....	115
Results.....	117
3.1 Generation of MASP-3 specific cDNA probe.....	118
3.2 cDNA cloning of rat( <i>Rattus norvegicus</i> )MASP-3.....	119
3.3 Rat MASP-3 5' RACE-PCR.....	123
3.4 Rat MASP-3.....	125
Genomic organisation of rat MASP-1/3.....	129
3.5 cDNA cloning of mouse( <i>Mus musculus</i> )MASP-3.....	133
3.6 Mouse MASP-3 5' RACE-PCR.....	135
3.7 Mouse MASP-3.....	137
3.8 RT PCR cloning of rabbit ( <i>Oryctolagus cuniculus</i> ) MASP-3 cDNA.....	144
3.9 Rabbit MASP-3 5' RACE-PCR.....	146
3.10 Rabbit MASP-3 3' RACE-PCR.....	147
3.11 Rabbit MASP-3.....	148
3.12 RT PCR cloning of guinea pig ( <i>Cavia porcellus</i> ) MASP-3 cDNA.....	153
3.13 Guinea pig MASP-3 5' RACE-PCR.....	154
3.14 Guinea pig MASP-3 3' RACE-PCR.....	156
3.15 Guinea pig MASP-3.....	157
3.18 Xenopus ( <i>Xenopus tropicalis</i> ) MASP-3.....	171
3.19 cDNA cloning of zebra fish ( <i>Danio rerio</i> ) MASP-3.....	174
3.21 Truncated gene product of the MASP-1/3 locus.....	199
3.22 Evolutionary comparison of MASP-3 sequence.....	202
3.23 Phylogenetic analysis MASP-3 serine protease.....	206
3.24 Phylogenetic analysis of full length MASP-like serine protease.....	208
3.25 Phylogenetic analysis of MASP-like serine protease domain.....	210
3.26 Northern blot analysis of rat and mouse MASP-3 mRNA transcripts.....	212
3.27 Specific ratio of MASP-1: MASP-3 mRNA levels in rat liver.....	213
3.28 Tissue distribution of human MASP-3 as determined by quantitative PCR.....	214
3.29 Recombinant expression of human and rat MASP-3 serine protease domain.....	216
3.30 Expression time course of recombinant human MASP-3 serine protease domain and C-terminal region in pRSET vector.....	219
3.31 Solubility screening of human MASP-3 serine protease expression products.....	220

3.32 Purification of rat and human MASP-3 serine protease fragments under denaturing conditions	223
Discussion.....	227
Future work.....	245
Appendices .....	247
Bibliography .....	251
Publication .....	267

# Introduction

## Chapter 1

## 1.1 The immune system

The immune system is a collection of proteins, cells and tissues interacting to safeguard the integrity of the host organism from other invading organisms and also from deleterious effects arising from normal and abnormal cellular processes. This hugely complex system is traditionally divided into two major components, adaptive and innate immunity. The conventional delineation of the immune system into an adaptive and non-adaptive component is gradually being eroded, as evidence for an initiation and instructive role in the adaptive immune response is attributed to components of innate immunity.

The two arms of the immune response have a number of common features, both have specific recognition molecules to identify targets and effect mechanisms to neutralise potentially pathogenic organisms, cells or other factors such as prion protein or immune complexes. The major difference between the two systems lies in the mechanism for generating these recognition molecules, innate immune recognition proteins such as C1q, MBL and ficolins are encoded by inheritable genes that mutate and experience evolution, much the same as the vast majority of other genes.

In contrast to this is the adaptive immune system, where a series of recombination events and other genetic mechanisms acts in concert to rearrange a genomic region containing gene segments of the particular recognition molecule. Each rearrangement is essentially unique, producing a clone cell expressing a unique antibody or T-cell receptor. Such mechanisms are theoretically able to generate in excess of  $10^8$  unique recognition molecules; inevitably some are able to bind to host tissues, a situation that is minimised by tightly regulating further expansion of these particular clones. A delayed effective response and requirement of specific exposure to a particular antigen are characteristic features of the adaptive immune process. Therefore although the adaptive system greatly aids the successful survival of the host it is the innate immune system which provides the first line of defence against novel pathogens.

## 1.2 Complement system

The complement system has been a cause for study for more than a hundred years, emerging as an elusive heat labile factor identified by Bordet in 1899 able to complement the ability of antibody to lyse bacteria. 50 years elapsed before the first proteins of the system were characterised and the multicomponent nature of the system became evident, latterly burgeoning into a complex highly networked host defence system encompassing more than 35 serum proteins and a number of associated receptors. The drive to understand the mechanics of complement stems from the pathways pivotal role in innate immunity, its interactions with the adaptive and cellular immune response and its function in the aetiology of a number of disease states.

Traditionally the complement system is divided into three activation pathways, classical, alternative and lectin, each of which uses a different means to achieve non-self recognition. The function of each pathway once activated is to bring about the assembly of an enzyme complex leading to the activation of the terminal complement cascade. The basic mechanical model of the complement system is still broadly consistent, with the modification that the alternative pathway acts as an amplification loop in concert with the classical and lectin activation pathways rather than an independently functioning activation arm (Dodds 2002; Brouwer, Dolman et al. 2006; Takahashi, Ishida et al. 2010). The initial step in determining self from non-self resides in specific recognition molecules for each pathway, C1q of the classical pathway, MBL and ficolins of the Lectin pathway. Each recognition molecule had a specific affinity and avidity for its target molecule, broadly immune complexes for C1q and terminal sugar moieties for MBL and ficolins. The binding of complement recognition molecules can in isolation interact with specific cellular receptors as evidenced by increased phagocytosis of particles in conjunction with complement recognition proteins, forming a direct connection with the cellular immune response (Tenner 1989; Tenner 1998; Bajtay, Jozsi et al. 2000).

Initiation of an individual complement pathway is achieved by transduction of the recognition event to further downstream members of the cascade, key to this event are the MASP-like serine protease, C1r, C1s of the Classical pathway and MASP-1, MASP-

2, MASP-3 of the Lectin pathway. A series of highly regulated cleavage events drive the complement pathway forward, each cleavage event cascading into the next the substrate of the previous reaction becoming the catalyst of the next often in conjunction with other cleavage products (Fig.1). Two foci of cleavage activity are formed by the complement cascades, the C3 and C5 convertase; these are key control points of the complement pathways. The convertase generate complement fragments to further progress the complement cascade, also releasing important mediators of the cellular immune response, C3a and C5a. Individual complement cascades utilise a terminal pathway to generate the macromolecular Membrane Attack Complex (MAC), the catalyst for this is formation of the C5 convertase with subsequent cleavage of complement components C5-C9.

The distinct complement activation cascades provide the framework on which interactions with the cellular and adaptive immune response are held, the interface between these two systems, cellular receptors are described below. The complement system acts to guide opsonisation and initiate the assembly of the terminal complement pathway, each facet of the system is described in its relevant section.

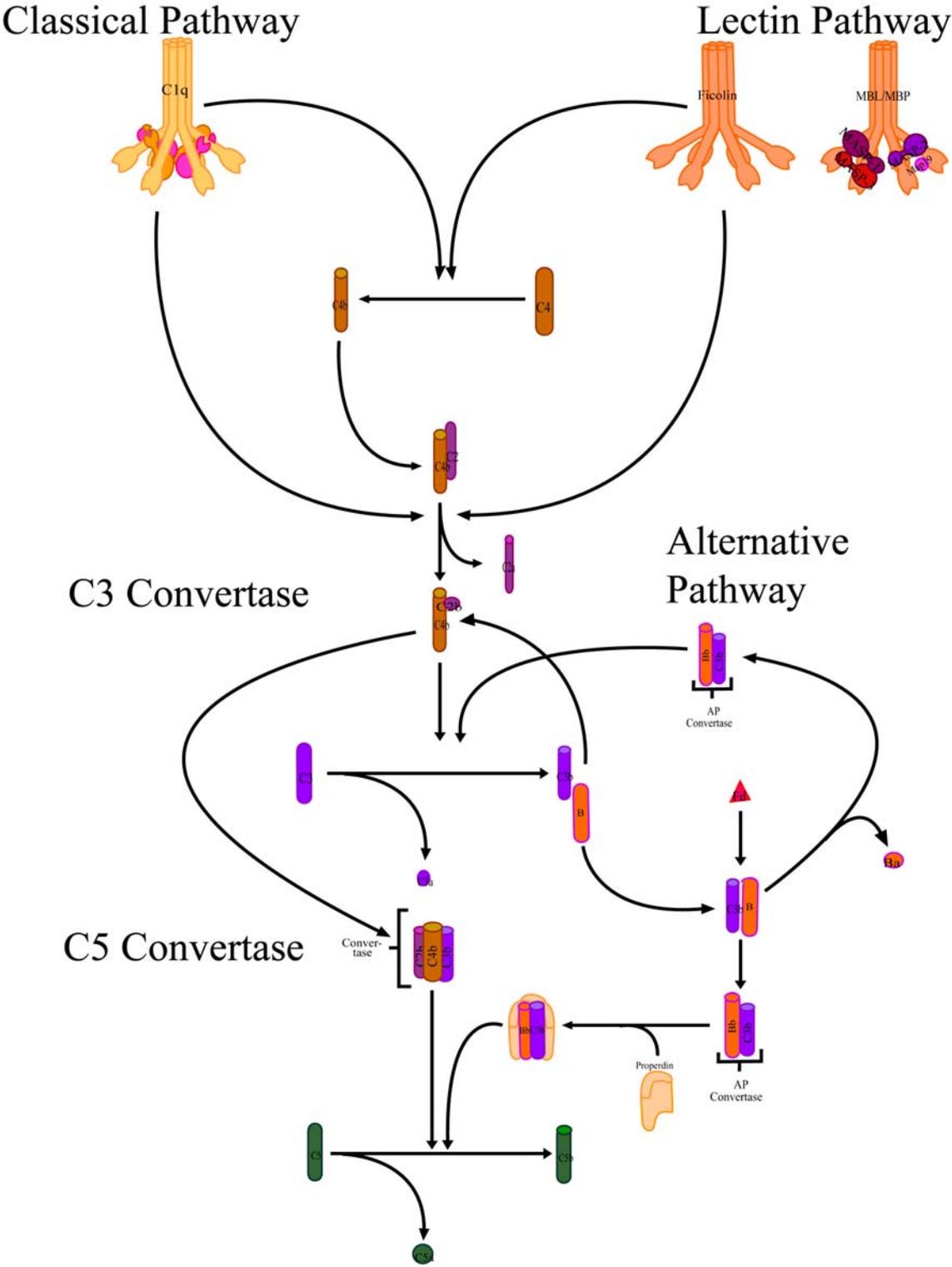


Fig.1 Schematic diagram of the Classical, Lectin and Alternative pathways of complement activation.

Recent research has shifted the emphasis away from the formation of a membrane attack complex as being the *raison d'être* of the complement system, the true power of complement activation resides in the production of C3a, C3b, iC3b, C3dg, C4a and C5a and subsequent involvement of the cellular immune response.

C3 cleavage products have long been shown to bind to circulating lymphocytes and follicular dendritic cells (FDC) within lymphoid tissue (Papamichail, Gutierrez et al. 1975), reduced levels of circulating C3 have a detrimental effect on the efficacy of the antibody response (Pepys 1976). Specific up regulation of B-cell responses occurs via CD21 (binding iC3b, C3dg and C3d) and CD35 receptors (binding iC3b, C3dg, C3d, C3b and C4b), both receptors work in conjunction with other cell surface molecules to generate a functional receptor complex. (CD, Cluster of Differentiation is an arbitrary nomenclature determined by the historical discovery of an immunologically determined surface epitope/marker)

CD21 forms a complex with CD19, a signalling component and CD81, a tetraspanin protein (Matsumoto, Kopicky-Burd et al. 1991). C3d-coated antigen taken up by B-cells utilising the CD21-CD19-CD81 co-receptor, lowering the threshold for B-cell activation and improving survival of the activated cell population (Carter and Fearon 1992). B-lymphocytes and follicular dendritic cells (FDC) cooperate via a shared affinity for C3d coated antigen, mediated by CD21 and CD35, FDC chemotaxis (Cyster, Ansel et al. 2000) attracts and concentrates mature B-lymphocytes within a germinal centre (Fang, Xu et al. 1998). C3d coated antigens are retained in lymphocyte germinal centres via follicular dendritic cells by virtue of their high expression of both CD21 and CD35. This brings in close contact B-cells and FDCs promoting retention and education of B-cells within a germinal centre, producing high affinity B-cells. Deficiency of C1q (Cutler, Botto et al. 1998), C3 or C4 (Fischer, Ma et al. 1996) demonstrate the involvement of complement components at a number of stages in the development of a B-cell response.

The stable complex between B-cell and FDC provides optimum conditions, and the correct context for association with T-cells. Priming of T-lymphocytes is complex and highly dependent on the antigen, associations with natural antibodies (Baumgarth, Herman et al. 1999) and influence of the prevailing cytokine milieu. Complement plays

a role in facilitating and directing T-lymphocytes to the site of antigen, via the chemotactic activity of complement components C3a and C5a. Interaction of T-cells with these small complement fragments increases the expression of their cell surface receptors, C3aR and C5aR, and in so doing altering the response of the T-lymphocyte as measured by cytokine profile.

The efficacy of the complement system is greatly enhanced via the interaction with the cellular immune response; this is mediated via complement fragments and complement receptors present on a variety of cell types. Complement cell surface receptors enhance the responsiveness of the innate immune system in destroying microorganisms, removing proteins, cellular debris and other foreign material from the circulation.

A group of seven receptors interacting with complement components have been identified; broadly these can be divided into those interacting with anaphylatoxin C3a and C5a via G protein coupled receptors and those able to bind C3 derivatives, C1q and C4b.

## 1.3 The immune systems of lower-order organisms

A system of immune protection exists in all living organisms ranging from single-celled to complex multinucleate species, providing an individual with the capacity to defend against threats both internal and external. Innate immune mechanisms provide an ever-present arm of defence throughout all species, exhibiting both complex and rudimentary approaches to the task of defending a host. Acquisition of an adaptive immune response may be thought of as delineating higher from lower order organisms with respect to the immune system, the ability to generate specific antigen-receptor molecules via somatic diversification and the exclusive expression of each unique receptor type by specialised clonal cell lines (vertebrate lymphocyte lineage).

Previously utilisation of immunoglobulin superfamily (Igsf) protein domain in the generation of antigen binding diversity was considered the hallmark of an adaptive immune response, adopted by the common ancestor of jawed vertebrates (Flajnik and Kasahara 2001). Recent research indicates the characteristics of an adaptive immune defence are present in varying forms across diverse species including invertebrates; there is no clear defining point to the beginning of this approach to immune protection, rather a series of solutions have been adopted to meet the requirements of specific species and their evolutionary decedents (Paul 2003). Ancestral phyla can be thought to possess an adaptive-like immune system; somatic diversity is achieved by a variety of mechanisms including multiple gene duplication, alternative splicing (Pancer 2000) and somatic hyper mutation but lacks the plasticity of the jawed vertebrate system (Du Pasquier 2005). Jawless vertebrates (agnathosomes) developed a radically different mechanism to generate an adaptive immune response (Alder, Rogozin et al. 2005), utilising assembly of multiple leucine rich repeat (LRR) cassettes to generate specific variable lymphocyte receptors (VLR).

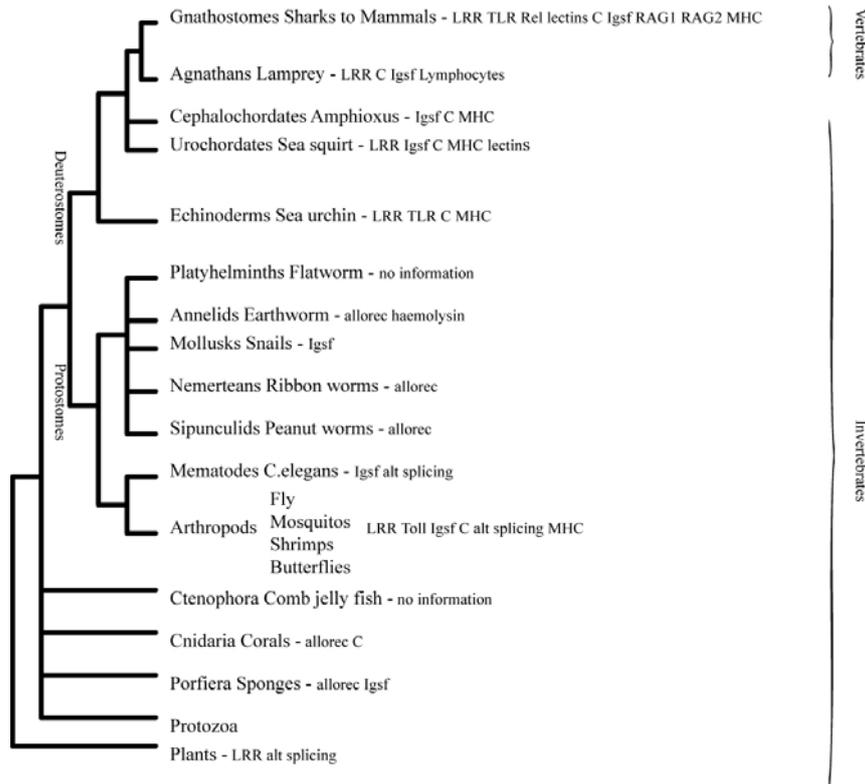


Fig.2 A simplified representation of invertebrate and vertebrate immune systems. Highlighting major components. Allorec (allorecognition), Alt splicing (alternative splicing), C (complement related component), haemolysin, Igsf (immunoglobulin superfamily), lectins, LRR (leucine-rich repeat), MHC (major histocompatibility complex), RAG1/RAG2 (genes mediating somatic recombination), Toll, TLR (toll like receptor).

Jawed vertebrate (gnathostome) adaptive immune systems are founded on the use of immunoglobulin (Ig), T-cell receptors (TCR) and major histocompatibility complex (MHC) molecules allowing clonal selection of immune cells (Danilova 2006). Generation of genetic diversity relies on a recombinatorial system, utilising RAG1/RAG2 gene products to mediate double-stranded DNA breaks and subsequent recruitment of a recipient gene segment (Oettinger, Schatz et al. 1990; Jung and Alt 2004). Novel combinations of genetic components can be constructed in this fashion controlled by specific recombination signal sequences, direct recombination of genomic DNA sequence significantly increasing variation of expressed protein product and enables propagation of those genetic events in subsequent clonal expansion of that unique cell line. The fundamental constituents of an advanced adaptive immune response are conserved amongst this evolutionary group despite a 450 million year timescale since a common ancestor and subsequent radical genome wide duplications experienced in the branch of teleost fish (Taylor, Van de Peer et al. 2001).

Important differences exist amongst jawed vertebrate species in the mechanisms of production, function and diversity of immunoglobulin repertoire, increasing complexity being introduced in to the system during evolution of more modern species (Paul 2003). Immunoglobulin genes are arranged in multiple clusters, cartilaginous fish (jawed fish) possessing 100-200 loci in which gene rearrangement is confined within a locus (Hsu, Pulham et al. 2006), this is in comparison to the 2-4 gene clusters exhibited by tetra pod species where rearrangement between clusters encourages variation. Diversity of immunoglobulin structure is determined by the ability to recombine genetic information from a variety of loci; in this respect jawed fish are limited to produce IgM/IgD immunoglobulin isotypes. The development of further immunoglobulin isotypes occurred in the amphibian species in conjunction with the ability to switch antibody class (Flajnik 2002), this established a modern complement of antibody isotypes IgM, IgG, IgE and IgA. Comparing the performance of immunoglobulin molecules in primitive jawed vertebrates and mammalian species we can see that antibody affinity is low, response time is slow and there is little affinity maturation in the jawed fish species. The limitations of primitive jawed vertebrate adaptive immune systems place a greater emphasis on the abilities of the innate immune system within such species.

In contrast to the comprehensive studies of immunoglobulin amongst jawed vertebrate species, the investigation of T-cell receptor (TCR) proteins is incomplete. T-cell receptor homologues have been identified in the most primitive jawed vertebrate fish, in all species examined TCR proteins are membrane bound and exhibit diversity comitant with the immunoglobulin molecules of that species (Paul 2003).

MHC molecules are vital in the process of determining self from non-self, in conjunction with the action of T-cells. The principals of T-cell and antigen processing cell (APC) collaboration leading to genetic restriction through the mechanisms of antigen processing and T-cell education are consistent amongst vertebrate classes (Kelley, Walter et al. 2005). MHC class I, II and III loci are identified in the most primitive cartilaginous fish, with the exception of the teleost lineage MHC loci exhibit genetic linkage conserved by modern mammalian species. Amongst the bony fish there is loss of the association between MHC class II loci and those of MHC class I and III precipitated by genomic duplication, this has enable a divergence in the rates of

sequence evolution between MHC class I and class II loci (Shum, Guethlein et al. 2001).

In contrast to the conservative development of immunoglobulin, TCR and MHC molecules amongst the jawed vertebrate lineage, lymphoid tissue exhibits significant variation (see Fig.3). Gnathostomes can be delineated from agnathostome species as the former possess both spleen and thymus structures, agnathostomes lymphoid tissue is limited to gut associated lymphoid tissue (GALT) present in all vertebrates (Paul 2003). The addition of specialised lymphoid tissues and increased organisation within such tissue is a reflection of the increased complexity of the interactions between T and B cells, enabling an efficient and controlled response to antigen to be mounted (Du Pasquier 2005).

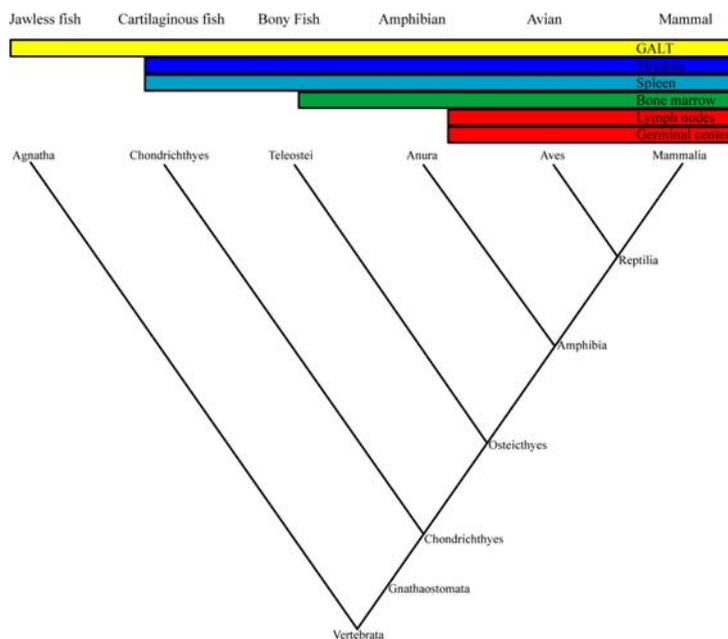


Fig.3 Specialisation of lymphoid tissues within the vertebrate phyla. GALT (gut associated lymphoid tissue).

Progressive development of an adaptive immune system during evolution places variable emphasis on innate immune system, leading to modification of constituents in a collaborative manner. Mechanisms of innate immunity are present throughout the vertebrate lineage, providing a first line of defence to maintain the integrity and

homeostasis of the host organism. Innate immune systems work in concert with the adaptive response within a particular species, the balance of this relationship shifting between species with the development of adaptive immunity throughout the jawed vertebrate class and selected members of the jawless vertebrate lineage.

Molecular phylogenetic analysis of the constituents of the complement system present in vertebrate and invertebrate species, has defined the extent of this important component of innate immune defence (Nonaka and Yoshizaki 2004). Until recently thought to be the preserve of deuterostomes phyla, complement proteins are now recognised in the evolutionarily distant Cnidaria phylum (Kimura, Sakaguchi et al. 2009). Complement components C3, factor B and MASP have been identified in the sea anemone, *Nematostella vectensis*, exhibiting conservation of protein domain architecture and crucial residues mediating catalytic and protein-protein interactions confirming the presence of a multicomponent complement system. The common ancestor of Cnidaria and Bilateria, a phylum lacking in complement components, diverged approximately 1,300 million years ago (MYA). The presence of individual complement components within respective classes of organisms is outlined in Table 1, due to a high degree of homology between many complement proteins significant gaps exist in the characterisation of complement proteins in a number of species.

	C3	C4	C5	Bf	C2	C1q	MBL	Ficolin	MASP-1	MASP-3	MASP-2	C1r	C1s	C6	C7	C8	C9	D	I	H
<b>Mammalia</b>																				
<b>Reptilia/Aves</b>					-				-								-			
<b>Amphibia</b>																			-	
<b>Teleostei</b>									-	-										
<b>Chondrichthyes</b>																				
<b>Agnatha</b>																				
<b>Cephalochordata</b>																				
<b>Urochordata</b>																				-
<b>Echinodermata</b>																				
<b>Cnidaria</b>						-	-	-			-	-	-	-	-	-	-			-

Table.1 Characterised complement component genes of metazoans. All complement component genes of humans are indicated as representative of mammals, with a confirmed orthologous genes for each animal class is indicated as a shaded box. Confirmed absence of an orthologous gene is indicated by -, no supporting evidence is indicated by a clear box.

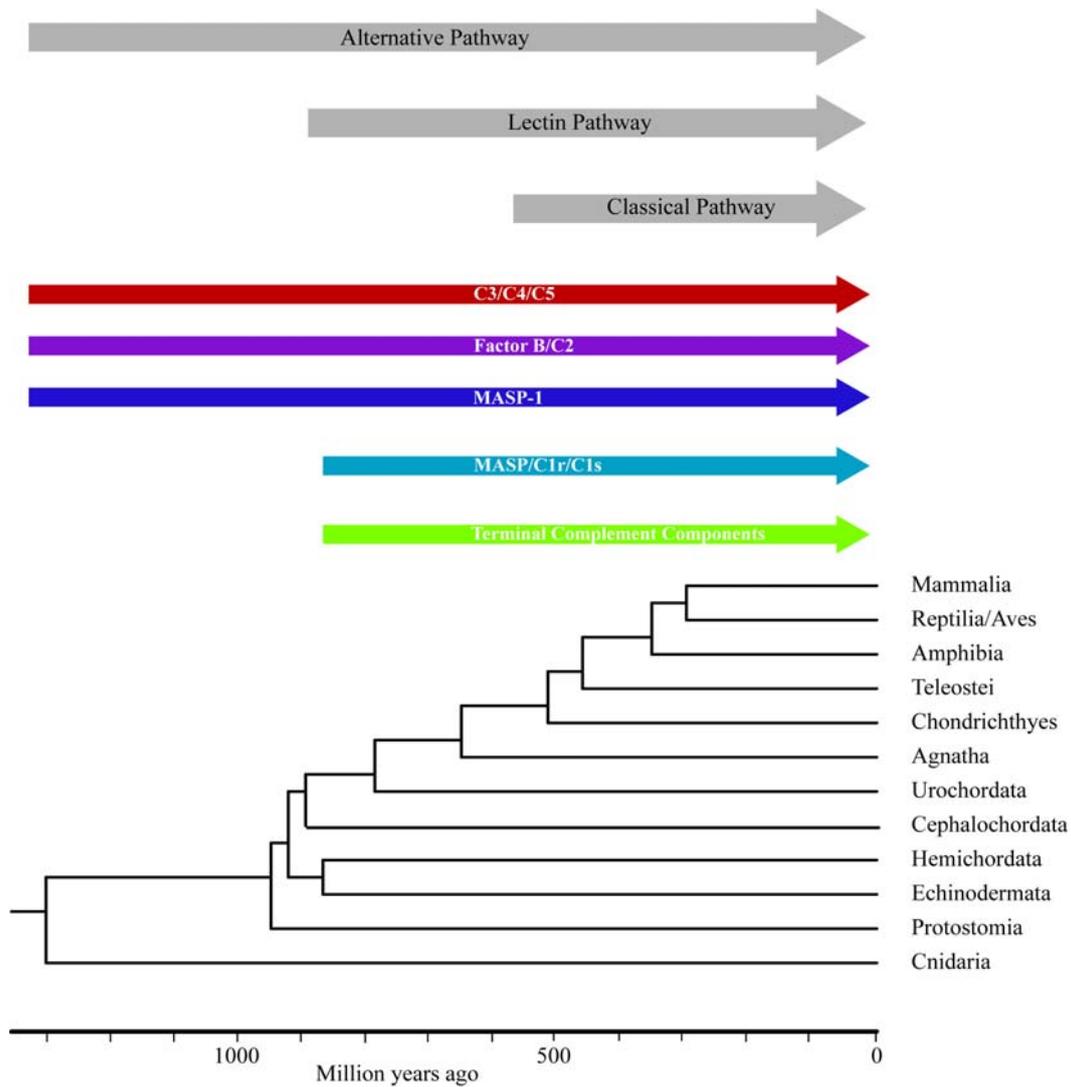


Fig.4 Evolutionary processes of the complement system. Evolution and origins of the 3 complement pathways. Origin and evolution of the major complement gene families, highlighted by arrows with respect to the most primordial class of organism currently characterised.

Progressive development of the complement system during evolution is detailed in figure 4 above. C3, factor B and MASP are present in the most primitive organisms. The alternative pathway is proposed to be the most ancient of the activation pathways, the novel function of MASP-1 in the activation of the alternative pathway component factor D (Takahashi, Ishida et al. 2010) further blurs the distinction between the lectin and alternative pathway activation pathways. The lectin pathway consolidates its independence during the chordate lineage (900 MYA), gene duplication of MASP and recruitment of MBL and ficolin recognition components in conjunction with the emergence of terminal complement components provides the foundations that are maintained in modern species (Nonaka and Kimura 2006).

Classical pathway complement cascade emerged within the vertebrate lineage prior to the emergence of the cartilaginous fish (600 MYA), arising as a result of gene duplication of the factor B/C2, MASP/C1r/C1s and C3/C4/C5 genes. The timing of the development of the classical complement cascade is approximate to the development of recombinatorial adaptive immune response.

As outlined the immune systems of lower-order organisms are complex, innate immune mechanisms provide consistent and evolving defence in conjunction with adaptive immunity. The development of both innate and adaptive immune systems occurs in concert, significant developments occurring synchronously within evolutionary time.

## 1.4 Classical pathway of complement activation

The classical activation pathway is commonly associated with the binding of antigen-antibody complexes, the recognition molecule C1q in association with dimers of the serine proteases C1r and C1s recognising specific conserved residues present in the Fc regions of both IgG (IgG1, IgG2 or IgG3) and IgM via its globular heads. Antibodies of the IgG and IgM class are only recognised when bound to an antigen, presenting the C1q molecule with a multivalent molecular pattern that is crucial to generate sufficient binding avidity. The requirement of epitopes to have not only specific binding characteristics, but also a particular spatial arrangement adds an extra dimension to the recognition properties of C1q. Pattern recognition is a central characteristic amongst complement recognition molecules including MBL, ficolins and other members of the collectin family of proteins. The structure of the C1q molecule is crucial to the ability to discern the spatial arrangement of a group of epitopes.

C1q is a calcium dependent glycoprotein composed of 18 polypeptide chains, three different forms associate, 6C1q-A, 6C1q-B and 6C1q-C, to produce six triple helical structures which again associate to form a single C1q molecule. Each triple helical structure consists of a linking region, a collagen-like stalk (Gly-Xaa-Yaa) and a C-terminal globular head. Further structuring of the C1q molecule is directed by the link region where six triple helical polypeptides are bound by inter chain A-B and C-C disulphide bridges (Reid and Porter 1976), this fixed point coupled with the effect of the neck region produce a structure that resembles a bunch of tulips the globular heads being precisely positioned to interact with suitable molecular patterns. There is often a bias towards solely classifying the classical pathway as an antigen-antibody complex recognition system, without regard to the numerous other proteins, ionic molecules and structures able to directly interact with C1q. The globular heads have been shown to bind certain retroviruses (Thielens, Tacnet-Delorme et al. 2002) and some microorganisms, while the collagen region of C1q has been shown to bind DNA, SAP (serum amyloid protein) and complexes containing CRP (C-reactive protein) (Jiang, Lint et al. 1991). Earlier studies have indicated C1q has the ability to bind the lipid A component of lipopolysaccharide (LPS) (Cooper and Morrison 1978), and mitochondrial membranes released from damaged cells (Giclas, Pinckard et al. 1979).

The relative importance of direct C1q complement activation leading to formation of the membrane attack complex, direct enhancement of phagocytosis and its role in immune complex clearance is difficult to quantify, certainly aberrant or absent function of the C1 complex results in the development in SLE (as does C4 deficiency). The emerging link between immune complex clearance and adaptive immunity emphasises the importance of antibody/immune complex recognition for the C1q molecule.

C1q interaction with the cellular immune system enhancing phagocytosis has been shown to be mediated via the collagen region of C1q (Bobak, Gaither et al. 1987), utilising a series of C1q proteolytic fragments. Identification of the sequence motifs responsible for this increased activity was achieved by a site directed mutagenesis approach; two important motifs have been identified. The first motif common to a number of collectins was identified via homologous sequence analysis based on experimental data generated on MBL, two triplets GEKGEP were shown to be key in the enhancement of phagocytosis (Arora, Munoz et al. 2001). A consensus sequence motif GE(K/Q/R)GEP is present in the A-chain of C1q. The second motif present in the C-chain of C1q has been shown to be crucial in the initiation of the neutrophil superoxide burst generated C1q (Ruiz, Henschen-Edman et al. 1999), proteolytic digestion experiments identified the discrete sequence Gly42-Lys58 as being responsible. C1q interaction with the cellular component of the immune system is limited by its interaction with the associated serine protease molecules (C1r, C1s); these interactions are only available once the C1 complex has dissolved into its component parts upon activation.

The search for C1q receptors or binding proteins began with the observation that C1q was able to bind to lymphocytes, particularly B-cell in a concentration dependent and saturable manner (Gabay, Perlmann et al. 1979).

A number of cells types are able to bind C1q, activating cells to increase phagocytosis, enhance microbial killing, induce chemotaxis and stimulate oxidative metabolism (Tenner 1998). Interactions with potential receptors and/or binding proteins have been mapped to both the globular head and tail region of C1q. The head region of C1q has been observed to bind to the globular head C1q binding protein (gC1qBP). The collagen-like tail regions of C1q have been proposed to bind to a number of proteins

including C1qRp, cC1qR (calreticulin also termed C1qR and collectin receptor) and CR1 (discussed above). Proposed C1q receptor molecules are discussed below.

gC1qBP was identified as a major 33 kDa component purified from Raji (B-cell line) lysate by purification on a sepharose C1q affinity column, found to bind to the globular head region of C1q in a saturable manner. Also termed p33, p32 and Tap (HIV Tat-associated protein), recombinant gC1qBP was shown to inhibit haemolysis of antibody-sensitized sheep erythrocytes by 50% when added to a  $1/10$  dilution of human serum. gC1qBP binds more avidly to intact C1q than does cC1qR (calreticulin) in both low and normal ionic strength, and the binding is not significantly affected by the presence of calcium and magnesium (Ghebrehiwet, Lim et al. 1994). The cellular localisation of gC1qBP remains controversial, being primarily an intracellular protein requiring pre-treatment of cultured cell to achieved detection with specific antibodies. The pre-processed protein contains an N-terminal mitochondria localization sequence, and has neither a transmembrane sequence nor a consensus sequence for the attachment of a GPI-anchor. Indirect functional evidence for binding of gC1qBP to C1q arises from inhibition studies utilising anti-gC1qbp antibodies, having the ability to ameliorate C1q mediated upregulation of P-selectin by platelets and mast cell chemotaxis. The entirety of gC1qBP function *in vivo* remains to be determined.

C1qRp or CD93 was identified by a series of monoclonal antibodies able to block binding to C4b/C3b opsonised targets in the presence of C1q (Guan, Robinson et al. 1994), the reconstituted protein able to enhance C1q-mediated removal of pathogens and immune complexes by phagocytosis. Collagenous tail preparations of C1q were used for affinity purification, under reducing conditions, of putative receptors from cell lysate of a monocyte cell line (U937) (Guan, Robinson et al. 1994). A 126 kDa protein was isolated by this method, preparation of a monoclonal antibody (mAb) (R139, IgG2b) against this protein was shown to inhibit the C1q-mediated enhancement of phagocytosis (Tenner 1998). Subsequent binding studies using a series of monoclonal antibodies noted that two, (IgG1 (U40.3) and IgG2b (R139)) failed to inhibit binding of C1q to a monocyte cell line. A third mAb (IgM (R3)) inhibited binding by only 30%, indicating an indirect interaction between C1qRp/C1q rather than a receptor ligand association. Primary structure analysis demonstrated that C1qRp was a type I transmembrane protein encoded by a 1959 bp cDNA. Sequence comparison revealed a

21 amino acid N-terminal signal peptide and a 156 amino acid C-type lectin-like domain also termed a carbohydrate recognition domain. The protein also contained five epidermal growth factor-like domains, three of which were predicted to have  $\text{Ca}^{2+}$  binding activity as well as a single transmembrane domain and a short cytoplasmic tail containing a tyrosine kinase recognition motif. C1qRp is the antigen recognized by several anti-CD93 antibodies and a pro-adhesive mAb mNI-11 (McGreal and Gasque 2002; Steinberger, Szekeres et al. 2002). CD93 is involved in C1q mediated enhancement of monocyte phagocytosis, with its apparent role limited to that of a co-receptor or co-signalling function.

A further molecule proposed to function as a C1q receptor is cC1qR, identified as binding to immobilised C1q with subsequent N-terminal sequencing identifying homology to calreticulin. cC1qR (CRT) is primarily an intracellular protein of the endoplasmic reticulum where it plays a role in glycoprotein quality control; it has also been detected in the extracellular environment, CRT which is tethered to CD91 at the monocyte cell surface (Ogden, deCathelineau et al. 2001) and with unfolded MHC class I on activated T cells (Arosa, de Jesus et al. 1999). cC1qR is able to bind the collagenous domains of C1q; it was also shown to bind the collagenous domains of two members of the collectin family, MBL and SPA, leading to the name “collectin receptor”. Calreticulin has a prominent role as part of autoantigen complexes found in SLE patients (Eggleton and Llewellyn 1999), forging a possible connection between C1q and calreticulin binding.

C1q is the detection component of the C1 complex able to interact with the cellular immune system via proposed receptor molecules (discussed above), but vital to the initiation of subsequent downstream events are the associated serine proteases, C1r and C1s. The stoichiometry of C1 is well characterised with a single C1q associating with dimers of both C1r and C1s. Dimers of C1r associate in antiparallel orientation, catalytic domains prevented from activating each partner C1r via their resting arrangement in the C1 complex (Budayova-Spano, Grabarse et al. 2002). C1r activation occurs via disruption of the resting conformation, binding of C1q to its target structure transmitting a mechanical stress to the region in which C1r binds allowing interaction between partner C1r protease (Budayova-Spano, Lacroix et al. 2002).

Conversion from the proenzyme form allows activated C1r to cleave its substrate C1s again converting it to an active state. Activated C1s itself a serine protease, catalyses two downstream reactions, the proteolytic cleavage of C2 and C4. C4 is the first substrate of C1s, C4 is synthesised as a single polypeptide with posttranslational modification generating  $\alpha$ ,  $\beta$  and  $\gamma$  chains. C1s cleavage of C4 generating C4a, which is a weak anaphylatoxin (Gorski, Hugli et al. 1979), and C4b a larger highly unstable fragment with an exposed reactive electron-accepting thioester bond. In humans and other mammals two forms of C4 exist differing in their preference for electron donors to interact with the newly generated reactive C4b intermediate. The great majority of C4b is directly hydrolysed to iC4b, with water molecules donating electrons. Those in close proximity to a surface are able to form covalent bonds, two C4 isoforms exist C4A preferentially interacting with amine groups more often found in proteins and C4B with hydroxyl groups commonly found in carbohydrates. The genetics of C4 are complex with individuals possessing either a single C4 locus or due to unequal crossing over event duplicates or triplicate C4 genes (Teisberg, Jonassen et al. 1988; Chung, Yang et al. 2002). The highly localised attachment of C4b to a surface provides a nucleating point for further cascading complement events; C2 is now able to associate with surface bound C4b at which point the serine protease C2 becomes the second substrate for C1s. Using the revised nomenclature, C2 is cleaved to release a small fragment C2a and a larger serine protease domain, C2b that remains in association with C4b forming the C3 convertase, C4b2b. C3 in common with C4 has an unstable thioester bond, which upon cleavage by the C3 convertase, releases a smaller C3a fragment and generates a reactive C3b intermediately able to covalently bond with both amino and hydroxyl groups found in carbohydrates and protein residues. The generation of a covalent bond between proteins or sugars of an activating surface must compete with an inactivating hydrolysis reaction occurring with molecular water, this effectively limits the deposition of C3b to the immediate vicinity of the C3 convertase. Binding of C3b to the convertase itself forms the classical pathway C5 convertase, C4b2b3b, the C2b component again acts as the catalytic domain with C3b recognising C5. Cleavage of C5 leads the cascade towards the non-catalytic events of the terminal pathway, with the final stage of membrane attack complex formation.

Specific regulation of the classical pathway is controlled by C1 inhibitor, a member of the serine protease inhibitor family of proteins or serpins. C1 inhibitor interacts with activated C1r and C1s preventing further serine protease activity, C2 and C4 cleavage. The serious consequences of aberrant complement activation require that there are a number of control processes to curtail such events, a number of such processes are common to all activation pathways and are divided between those present on cell surfaces and those in the fluid phase.

## 1.5 Alternative pathway amplification loop

The alternative pathway of complement activation contributes to the efficacy of the complement system as a whole, not only providing a distinct and novel mechanism of activation but also a positive feedback amplification loop utilised by the classical and lectin activation pathways. In common with the classical and lectin activation cascades, the central aim is to generate the multimolecular C3 convertase enabling procession to the terminal complement pathway.

In a divergence from the lectin, classical pathway and indeed immunoglobulin of the adaptive immune response, activation of the alternative pathway is not governed by specific recognition molecules but rather the antagonistic effects of regulatory proteins. This finely poised system is highly sensitive to disturbance of this balance, interactions with surfaces that promote stability of the C3 convertase or reduce the effectiveness of regulatory proteins results in initiation of the amplification loop.

A vital step in the activation of the alternative complement pathway is exposure of a thioester group buried deep within the hydrophobic core of the native C3 molecule (Law 1983), a conformational change in the C3 protein is required. Enzymatic cleavage of C3 producing C3a and C3b via the action of a C3 convertase is a means to achieve this conformational change, generating a highly reactive C3b intermediate with an extremely short half life.

Exposure of the active thioester group is postulated to occur spontaneously without requiring an active external influence (Pangburn, Schreiber et al. 1981), but direct evidence for such spontaneous production of C3b remains scarce. The ability of MASP to cleave C3 (Rossi, Cseh et al. 2001), along with the more compelling data regarding the continual low level complement activation attributed to the classical pathway (Manderson A 2001), pose the question of whether the alternative pathway is indeed an independent means of complement activation or a vitally important amplifier and safety mechanism for both the classical and lectin activation routes. Cleavage of C3 is undoubtedly the first step in the initiation of the alternative pathway amplification loop, with the cleavage event occurring by a number of different mechanisms none of which

are exclusive. The fate of the reactive C3b intermediate generated by the complement system is determined by the molecules extremely short half-life, interacting with either water molecules to form C3i or covalently binding to carbohydrate or protein structures.

Spontaneous hydrolysis of C3 predominantly occurs in the fluid phase the resulting C3b being immediately hydrolysed to C3i, a further interaction with Factor B generates C3iB a substrate for the serine protease Factor D. Activation of Pro Factor D to the active form Factor D requires enzymatic cleavage by MASP-1 (Takahashi, Ishida et al. 2010), again highlighting the reliance of the alternative pathway on interaction with other complement pathways. The subsequent cleavage of C3iB by Factor D occurs within the Factor B subunit, releasing a small Ba fragment and activating the protease activity of Factor B, generating the fluid phase C3 convertase C3iBb. The C3 convertase can directly cleave C3a from C3 releasing the C3b reactive intermediate, supplying both the C3b required to form the C3 convertase itself and to opsonise surfaces to which it may contact.

Formation of a C3 convertase can more importantly occur at a surface; these can be either protected or non-protected by range of intrinsic and host regulatory mechanisms. Attachment of reactive C3b to a protected surface, such as bacterial cell wall or an immune complex, initiates C3 convertase assembly. Immobilised C3b forms a complex with Factor B generating C3bB, which is a substrate for Factor D the cleavage event releasing the Ba fragment of Factor B. The establishment of the C3 convertase C3bBb on a protected surface leads to the localised production of C3b, stability of the C3 convertase is enhanced by an interaction with properdin (Factor P) and the lack of regulatory proteins common to host cells. With the generation of the reactive C3b intermediate in close proximity to the activating surface, the majority now interacts directly with the surface producing more foci for convertase assembly. Production of large quantities of C3b is an important factor in the generation of the final stage of the activation event, the alternative pathway C5 convertase C3bBb(C3b)<sub>n</sub>.

Protected surfaces such as those found on normal host cells have a raft of regulatory mechanisms to prevent spurious complement activation, CR1, Factor H, MCP (membrane cofactor of proteolysis) and DAF (decay accelerating factor) act to bind C3b displacing Bb. In the case of CR1, Factor H and MCP the complex formed is a

substrate for Factor I, a cleavage event forming inactive iC3b preventing further C3 convertase activity.

Dependent on the nature of the surface to which reactive C3b interacts, generation of a C3 convertase can either be inhibited by regulatory mechanisms, maintained by the absence of regulatory intervention or the action of properdin, the only known positive regulator of complement (Hourcade 2006). Maintenance of the alternative pathway C5 convertase, C3bBb(C3b)<sub>n</sub> on protected surface moves the cascade forward into the terminal pathway and initiates generation of the membrane attack complex (MAC).

## 1.6 Lectin pathway of complement activation

The lectin activation pathway acts in an antibody independent manner to recognise non-self surface structures via specific recognition of carbohydrate arrays on target surfaces and signal transduction using a group of highly specialised serine protease molecules. The lectin activation pathway is the most recently described mechanism of activating the complement system, and has a gross architecture similar to that of the classical pathway. Characteristic recognition of surface carbohydrate patterns is balanced by a gross structural architecture familiar to the classical C1 complex, despite this similarity of form the detailed functioning of the MBL and ficolin activation complex is quite distinct.

Analogous to the recognition component C1q of the classical system are ficolins and MBL. Ficolins are collagen/fibrinogen domain containing proteins while MBL is a member of the collectin protein family. Collectins bind in a calcium dependent manner to carbohydrate arrays such as those found on the surface of microorganisms, these polypeptides are broadly composed of four regions: a cysteine rich N-terminal region important in oligomerisation, a collagen-like region, an  $\alpha$ -helical coiled-coil and the lectin domain. The functional collectin unit consists of a trimeric triple helix, utilising the N-terminal region to associate and stabilise the structure. Further oligomerisation of collectin units occurs to differing degrees, boosting the relatively low affinity binding by increasing the avidity of the collectin structure for its target. The multivalent nature of collectin structures function to agglutinate, enhance clearance and directly activate complement at non-host surfaces. There is a growing family of collectins but for the purposes of lectin pathway activation these can be divided into two groups, those that are associated with and induce the activation of MASP and others that are concerned solely with opsonisation and agglutination. Of the latter group SP-A identified in human, mouse, rat, rabbit, guinea pig, canine and porcine; SP-D identified in human, mouse, rat, bovine and porcine. Also related to the SP-D group are the bovine conglutinin and collectin-43; Cl-L1 found in humans (Hakansson and Reid 2000). Individuals deficient of either SP-A or SP-D have yet to be identified. Transgenic mouse models deficient of SP-A (*Spa*<sup>-/-</sup>) are susceptible to all forms of bacterial infection tested, whilst SP-D deficient mice (*Spd*<sup>-/-</sup>) develop emphysematous lung with

greatly expanded alveolar spaces (Holmskov, Thiel et al. 2003). Allelic variants of both SP-A and SP-D are thought to play a role in modulating infection and coexisting existing disease states.

MBL and ficolins also take part in opsonisation of activating particles, but in regard to the lectin activation pathway of complement they have a more fundamental role via their association with MASP.

Examining firstly MBL where the general picture is more clearly understood, to date MBL (mannan binding lectin) has been identified in human (chromosome 10), rhesus, mouse, rat, porcine and bovine. MBL in rhesus, mouse and rat is present in the genome as two copies being designated MBL-A and MBL-C. In bovine and porcine they appear to have a single MBL-C like gene although as whole genome sequence is not available it is not possible to confirm only one gene, in humans an expressed gene *mbi2* and a pseudogene *mbi1* have been characterised.

Structurally MBL is a calcium dependent or C-type lectin composed of 32 kDa subunits, three of which associate to form a functional unit. Each polypeptide contains an N-terminal region encompassing two or three cysteine residues, followed by 18-20 collagen like repeats of Gly-X-Y, a neck region responsible for the trimerisation process, and a C-terminal carbohydrate recognition domain (CRD) (Hoppe and Reid 1994). MBL circulates as oligomers ranging from dimers to hexamers, the hexameric structure resembling that of C1q, CRD's arrayed with a spacing of 45Å radiating from a central core (Sheriff, Chang et al. 1994). The multimeric nature of MBL is crucial in generating sufficient binding avidity, monomeric functional units utilising their three CRD's bind to carbohydrate with a dissociation constant of  $10^{-3}$  M (Lobst et al, 1994), oligomerisation greatly decreases this to  $10^{-9}$  M. The quaternary structure of MBL oligomers provides an extra dimension to the binding characteristics of the complex not only is a CRD able to recognise the particular structures presented by carbohydrates but the complex is able to recognise the way in which these carbohydrates are arrayed. Pattern recognition has identified a fundamental difference in the surface structures presented by microorganisms and invading parasites, in that they have a more rigid and repetitive nature than those of the host (which gives the relatively weak individual CRDs sufficient targets and reduced time to associate.)

Binding studies have been made with simple carbohydrates to define the binding specificity of the carbohydrate recognition domains. MBL preferentially recognises polycarbohydrate hexoses with characteristic 3- and 4- hydroxyl groups presented on terminal nonreducing units, such as mannose, glucose, fucose and N-acetylglucosamine groups present in glucans, lipophosphoglycans and glycoinositol-phospholipids. The crucial feature of the binding characteristics of MBL, as with C1q is the low affinity binding that is overcome by utilising the numerous recognition domains to build up a significant binding avidity.

Eukaryotic glycoconjugates predominantly terminate in sialic acid residues to which MBL shows limited binding. The low avidity of MBL for sialic acid, in conjunction with the rigidly presented binding sites of bacterial cell wall components, enables MBL to effectively determine self from non-self structures.

A Danish population study has shown the median level of circulating MBL to be 1.2 µg/ml, although individuals ranged from 0 – 5 µg/ml in part due to some individuals carrying one or more common mutations. Circulating levels of MBL are relatively constant for a particular individual although MBL has been shown to be a weak acute phase protein, with levels increasing between 1.5 and 3 fold. This is a relatively small increase compared to the variation seen between individuals carrying the same MBL genotype which can show a 6-fold difference in their MBL levels; this is mirrored in mice where individual inbred mice can exhibit up to 2.5 fold difference.

Ficolins are composed of a cysteine rich N-terminal region, a collagenous domain and a fibrinogen-like carboxy terminus (Ichijo, Hellman et al. 1993). In Humans three ficolin genes have been identified *FCN1*, *FCN2* and *FCN3*, encoding a surface bound M-ficolin expressed on monocytes (Lu, Tay et al. 1996), and two serum ficolins L-ficolin and H-ficolin (Matsushita and Fujita 2001) respectively. Circulating L-ficolin, H-ficolin and surface bound M-ficolin are recognition molecules of the lectin complement activation pathway, found to associate with MASP, Map19 and to activate complement (Matsushita, Endo et al. 2001; Matsushita, Kuraya et al. 2002; Liu, Endo et al. 2005).

Much progress has been made in elucidating the components and events surrounding activation of the lectin pathway, a combination of a relatively short history and its apparent complexity in comparison with the classical pathway leave us at this time with an incomplete picture and many questions.

## 1.7 Terminal complement pathway

The terminal pathway is the execution arm of the complement system, the dual function of which is to generate membrane-damaging pores in the target cell and to recruit cells of the immune system to the region. Terminal events in the complement cascade are common to all activation pathways and are initiated by the cleavage of C5, producing C5a and C5b. The remainder of the pathway is non-enzymatic, with proteins associating to form a macromolecular complex.

Two C5 convertase are generated by the three complement activation pathways, the Classical and Lectin pathways form C4b2b3b with the alternative pathway utilising the C3bBb(C3b)<sub>n</sub> enzyme complex. A common mechanism exists for capture of C5 at an acceptor site within the C3b molecule; the association of these components leaves C5 liable to cleavage by activated C2b or Bb serine proteases respectively. The reaction generates C5a and C5b in an extremely controlled fashion with a relatively limited number of molecules being generated, reflecting the seriousness of subsequent events. C5a is a potent anaphylatoxin linking the complement system to the cellular immune response and arguably has a more significant function than the assembly of the MAC for the outcome of an infection. C5b provides a point of nucleation for further events in the assembly of the MAC, an association with C6 and C7 leads to a conformational change within the complex unmasking a hydrophobic region in the C7 component leading to its insertion into the lipid bilayer. The C5, C6 and C7 further aggregate with C8 and C9<sub>n(10-16)</sub>, forming a macromolecular pore spanning the cell membrane leading to a general breakdown of cellular homeostasis.

Deficiencies of all terminal complement components have been identified; broadly they are associated with an increased incidence of *Neisseriae* infection, particularly meningococcal meningitis. Although such infections can be serious the remaining functional immune system appears to limit other potential infectious agents, despite the loss of the execution arm of the complement system.

In addition to the killing potential of the MAC, the C5b-9 complex can induce a variety of biological responses (Morgan 1989). At sublytic doses insertion of the MAC within the cellular membrane has been shown to induce the release of pro-inflammatory mediators (Kilgore, Schmid et al. 1997), induction of the cell cycle (Rus, Niculescu et al. 1996) and expression of adhesion molecules (Tedesco, Pausa et al. 1997).

The terminal complement system aside from generating the macromolecular membrane attack complex (MAC) also plays a role in shaping the adaptive immune response, recruiting components of the cellular immune system via C3 and C5 anaphylatoxin fragments.

## 1.8 Mannan binding lectin

The architecture of MBL lends itself to multiple binding events spatially arranged across a particular surface, these give the recognition event a number of important characteristics. Firstly recognition is not solely limited to particular carbohydrate residues but to the pattern in which they are presented and possibly their density and fluidity on a particular surface. Secondly the requirement to engage in multiple binding events by spatially arranged CRD's that are linked through a central core structure lend itself to the mechanical transmission of an activation signal to the associated serine protease molecules.

*mbl2* transcripts are found predominantly in liver with extrahepatic expression in the small intestine and testis identified (Seyfarth, Garred et al. 2006). MBL can be detected in all body fluids, including the normally immune privileged cerebrospinal fluid (Lanzrein, Jobst et al. 1998). Human MBL is a mild acute phase protein increasing approximately 3-fold on stimulation. A region of the *mbl2* promotor is homologous (90% identical) to that of the acute phase protein human serum amyloid-A (SAAg9)(Taylor, Brickell et al. 1989). However, the acute phase response is only moderate, with the increase being less than the distribution of protein levels within a population.

In the murine system MBL is primarily produced in the liver, (Wagner, Lynch et al. 2003) MBL-C and MBL-A are expressed with a relative ratio of 3:1. MBL-A and MBL-C are also expressed in the marginal macrophage zone of the spleen; the ratio is equal but a fraction of that seen in the liver. Extrahepatic MBL-A is seen in the glomerular region of the kidney, occurring at less than 30x the concentration seen in the liver while MBL-C is essentially absent. MBL-C is seen in comparable levels in the small intestine. MBL-C shows no acute-phase characteristics indicating a distinct regulatory apparatus.

A series of mutations have been identified which modulate serum MBL levels (Fig.5), three single nucleotide polymorphisms (SNPs) reside in the first exon at codons 52, 54 and 57; two are located in the promotor region (-550 H/L variant, -221 X/Y variant,

G→C) and a further resides in the 5' untranslated region (+4 P/Q variant, C→T) (Turner and Hamvas 2000). Utilising an alternative nomenclature the exon 1 polymorphisms are also referred to as wildtype A with B, C and D variants, corresponding to codons 54, 57 and 52 respectively, collectively these O alleles are present in 40% of certain populations (Minchinton, Dean et al. 2002). Those mutations associated with the promotor directly affect MBL expression levels, while amino acid substitutions at positions 52, 54 and 57 impair the function of MBL either by directly interfering with correct subunit assembly or disrupting the crucial MASP association. Arg<sup>23</sup>→Cys (codon 52, CGT→TGT) mutation interferes with formation of higher order MBL oligomers (Roos, Garred et al. 2004), while mutations Gly<sup>25</sup>→Asp (codon 54, GGC→GAC) and Gly<sup>28</sup>→Glu (codon 57, GGA→GAA) have an additional effect on MASP binding (Wallis and Dodd 2000) indicating the site of protein-protein interaction. Variant B is present in 22-28% of the Eurasian population, variant D in 14% of the European population while variant C is characteristic of sub-Saharan Africans reaching a prevalence of 50-60% (Turner 2003).

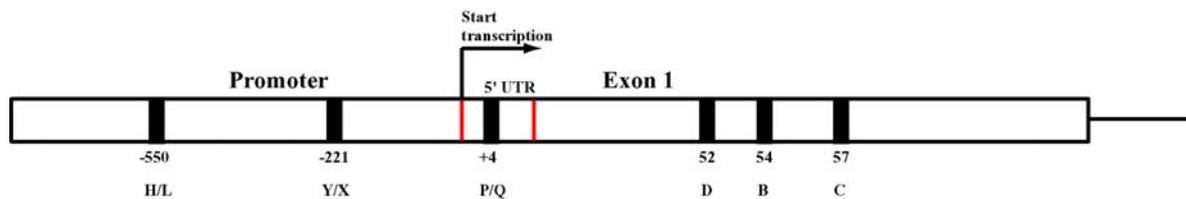


Fig.5 Schematic diagram of promoter and exon 1 mutations of the *mbi2* locus modulating Mannan Binding Lectin serum protein concentration.

The relationship between the genotype of an individual and the circulating level of MBL is complex, the interplay of 5 major polymorphic loci distributed across promoter, non-coding and coding regions of the MBL locus and the wide range of MBL protein levels observed for a particular genotype produces much overlap. A Danish population study has shown the median level of circulating MBL to be 1.2 µg/ml, although individuals ranged from 0 – 5 µg/ml largely due to some individuals carrying one or more of the mutations.

Examining firstly the coding mutations D, B and C in exon 1, homozygous or compound heterozygous individuals at these loci were deficient of circulating MBL (detection limit of 20µg/l). Heterozygotes for the common allele A and either mutation

B or C had significantly diminished MBL levels when compared with A/D individuals (Garred, Larsen et al. 2003). There is a great variety of MBL protein expression associated with different promotor haplotypes, to ease interpretation only the most significant in terms of MBL protein levels are considered. The promotor allele at position -221 (X/Y) is in linkage disequilibrium with the common structural allele A, exhibiting either a high (Y/A) or low (X/A) promotor activity and serum MBL level. Functional promotor studies using a  $\beta$ -galactosidase reporter system have stratified the alleles present at positions -550 and -221 into high, intermediate and low activities corresponding to the haplotypes HY, LY and LX respectively (Naito, Ikeda et al. 1999). Median serum levels for each of these haplotypes are 2.3 $\mu$ g/ml (HY), 1.9  $\mu$ g/ml (LY) and 0.35  $\mu$ g/ml (LX) (Madsen, Garred et al. 1995). A further polymorphic site within the 5' untranslated region of the *mb12* transcript at position +4 is associated with the intermediate promotor polymorphism LY. The allele has a distinct influence on serum MBL level; LYQA homozygotes exhibit a relatively high MBL level, LYQA/LYPA intermediate and LYPA homozygotes low (Madsen, Satz et al. 1998). The influence of the +4 (P/Q) allele on circulating MBL levels is minor and there is considerable overlap between homozygotes and heterozygotes, it should also be noted that a number of other polymorphisms identified within the promotor may account for the difference observed between P and Q alleles. Relegating the polymorphism at +4 to a marker, rather than the site of a functional polymorphism (Garred, Larsen et al. 2003).

The complex nature of the *mb12* gene polymorphisms and their interplay with detected circulating MBL levels makes interpretation of associations difficult; a more informative measure in terms of disease associations may be to look at the functional levels of circulating MBL.

Mutations of *mb12* are present in a surprisingly high proportion in populations in Europe and sub-Saharan Africa (Garred, Richter et al. 1997), studies of adult patients presenting with recurrent infections showed an increase in the frequency of individuals homozygous (8.3%) for *mb12* structural mutation but heterozygotes showed no increase in frequency. Two independent *mb12* mutations developed to reduce the effective concentration of circulating MBL, population studies in sub-Saharan Africa indicate increased *Mycobacterium tuberculosis* infection rates amongst those with normal or

high MBL levels (Garred, Richter et al. 1997). MBL deficiency may have a protective role in the development of atherosclerotic lesions; high MBL individuals with rheumatic heart disease exhibited significantly more damage when compared to deficient patients (Schafranski, Stier et al. 2004), MBL deficiency showed proportionally higher representation in the control group. Evolutionary pressure of this sort may contribute to the high MBL deficiency rates seen in a number of populations.

There appears not to be a direct correlation between MBL level and disease susceptibility as the concentration of MBL varies greatly between individuals, although levels for an individual are relatively constant. A decreased level or deficiency of MBL has been associated with an increased susceptibility to infectious disease, particularly with gram positive bacteria such as *Neisseria* (Jack 2001), *Salmonella spp* (van Emmerik, Kuijper et al. 1994) and *Streptococcus pneumoniae*. MBL levels play a role in the development of sepsis and modify the outcome of infection adults, MBL deficient individuals are overrepresented amongst septic hospitalised patients (Eisen, Dean et al. 2006) and experience increased level of organ damage as measured by SOFA (Sequential Organ Failure Assessment) score (Vincent, Moreno et al. 1996).

The structural mutations of the *mbi2* gene presented by individuals experiencing recurrent infection show both specific distribution and increased frequency when compared with the general population. In adult patients the frequency of homozygous individuals showed a marked increase as compared to the general population, but a similar increase in heterozygotes was not seen (Garred, Madsen et al. 1995). This contrasts with the situation seen in paediatric patients where the frequencies of both homozygotes and heterozygotes show a marked increase as compared to a control population (Summerfield and Taylor 1997). Respiratory infections in children aged between 1-16 years correlate well with the MBL status of the individual, those exhibiting the most severe recurrence rates were found to have both an MBL and humoral immune defect (Cedzynski, Szemraj et al. 2004). MBL plays a minor role in disease susceptibility but is predominantly a modifier of disease progression and outcome.

This difference in distribution of susceptible individuals indicates that MBL plays an important function in protecting children in the period between the disappearance of

maternally acquired immunity and the mature development of the individual's own immune defence. The disease modifying effect of MBL can also be seen in the adult population suffering from Chronic Obstructive Pulmonary Disease (COPD), *mbi2* polymorphic individuals were at an increased risk of exacerbation of their condition leading to hospital admission (Yang, Seeney et al. 2003). There was no association with the development of COPD and MBL deficiency.

Cystic fibrosis is a chronic disease caused by mutation (commonly  $\Delta 508$ ) of cystic fibrosis transmembrane conductance regulator (CFTR), absent or ineffective functioning of this chloride channel results in increased viscosity of secretions particularly in the lung and pancreas as a result of ineffective ciliary action. This combination predisposes individuals to chronic, recurrent respiratory infection. Bacteria such as *Pseudomonas aeruginosa*, *Burkholderia cepacia* and *Stenotrophomonas maltophilia* (Lambiase, Raia et al. 2006) increase secretion viscosity leading to air trapping and initiate cellular immune responses that damage the lung parenchyma. Lung function in carriers of MBL polymorphisms indicated a significant reduction when compared with normal homozygotes (Garred, Pressler et al. 1999), comparing structural *mbi2* gene mutations A/A individuals with a high MBL level (2472 $\mu$ g/ml) FVC of 90% and FEV<sub>1</sub> 69%, heterozygous individuals A/O intermediate MBL level (298 $\mu$ g/ml) FVC of 79% and FEV<sub>1</sub> 59% and homozygous individuals O/O undetectable MBL level (0 $\mu$ g/ml) FVC of 65% and FEV<sub>1</sub> 50%. An isolated case study demonstrated the therapeutic use of purified MBL, arresting the decline in lung function in an individual with low MBL levels and cystic fibrosis (Garred, Pressler et al. 2002). Individuals with a homozygous structural *mbi2* gene mutation are at 3 fold increased risk of developing end-stage Cystic fibrosis, while predicted survival was reduced by 8 years compared to normal homozygotes (Garred, Pressler et al. 1999). MBL levels appear to play a role in severity and progression of cystic fibrosis, in particular for those individuals developing chronic *Pseudomonas aeruginosa* infection; temporary reconstitution of normal MBL levels had a beneficial effect on lung function (Garred, Pressler et al. 2002).

Surveys of individuals undergoing chemotherapy for malignant disease, an effective model of an adaptive immunity deficit are currently in disagreement about the importance of MBL deficiency in relation to infection rates. Currently two studies

indicate that the incidence of febrile illness during the period of neutropenia following chemotherapy is significantly higher in those individuals with a defect in MBL (Neth, Hann et al. 2001; Peterslund, Koch et al. 2001), such findings indicate a period of prophylactic MBL infusion to be beneficial in reducing this incidence. More recent studies indicate that this correlation between febrile illness and periods of neutropenia associated with chemotherapy is not significant (Bergmann, Christiansen et al. 2003; Kilpatrick, McLintock et al. 2003), casting doubt on the value of MBL therapy in such individuals. Investigation of critically ill individuals serially admitted to a paediatric intensive care setting, both highlighted an over representation of variant *mb12* polymorphisms in this cohort and an increased propensity to progression from infection to septic shock. There is a significant association with the development of a systemic immune response and an MBL level <1000 ng/ml.

Major surgery and early post surgical complications such as pneumonia are associated with low preoperative MBL levels (Ytting, Christensen et al. 2005), there was no correlation with other forms of infection or long term postoperative outcome.

Mannan binding lectin has long been known to bind to HIV viral particles, interfering with viral entry into the susceptible cells (Ezekowitz, Kuhlman et al. 1989). MBL capture assay identified strong binding of HIV virions to MBL (Saifuddin, Hart et al. 2000), HIV strains not expressing the HIV surface protein gp120 were not retained. Microtiter plates coated with either recombinant gp120 or isolated gp120, showed MBL dependent complement activation (Haurum, Thiel et al. 1993).

HIV encoded proteins have a characteristic high-mannose/hybrid glycosylation that rarely occurs in mammalian-encoded glycoproteins but is conserved among different HIV strains (Geyer, Holschbach et al. 1988; Mizuochi, Matthews et al. 1990), increasing (Hart, Saifuddin et al. 2003) the mannan content of viral proteins or reducing the sialic acid content with neuraminidase increases the efficiency of MBL/HIV interaction.

A number of studies have investigated the effect of MBL level on the acquisition and progression of HIV, due to the differing methodologies and populations studied the results are conflicting and difficult to interpret. Initial studies indicated that low MBL levels were associated with an increased incidence of HIV, indicating a propensity for acquiring HIV when exposed (Garred, Madsen et al. 1997). More recent studies in

different population groups indicate that high MBL levels are associated with increased rates of infection (Pastinen, Liitsola et al. 1998), while other research indicates no association with MBL level (Malik, Arias et al. 2003). Similarly the effect of MBL level on the progression of HIV is unclear; studies indicate both a beneficial (Maas, de Roda Husman et al. 1998) and an adverse (Boniotto, Crovella et al. 2000) outcome of low MBL levels on the rate of HIV progression. Conflicting results may be due to the differing ethnicities of individuals recruited for the studies.

Studies in several population groups have indicated an association between MBL levels and systemic lupus erythematosus (SLE) an autoimmune disorder characterised by immune complex deposition and vasculitis. The frequency of *mbi2* structural polymorphisms are increased in individuals developing SLE, in a Danish population study homozygotes were significantly overrepresented in the SLE groups when compared to normal matched controls (Garred, Madsen et al. 1999). MBL B allele is found in higher frequency amongst SLE patients in both British (Davies, Snowden et al. 1995) and Hong Kong Chinese populations (Lau, Lau et al. 1996). These findings suggest that impairment of the MBL pathway due to production of dysfunctional MBL or due to a reduction in the amount of circulating MBL is associated with the development of SLE.

Serum MBL level and polymorphism of the *mbi2* locus are associated with the development of infectious and certain autoimmune disease states in susceptible individuals, recent studies have identified a number of additional pathological conditions in which MBL plays a role.

MBL level has an effect on reproduction and influences the period of gestation. Maternal MBL levels <100ng/ml are associated with an increased risk of recurrent spontaneous abortion (RSA) (Kruse, Rosgaard et al. 2002), being present in 18.9% of the RSA group as compared to 12.2% of matched controls. This confirmed previous work indicating an association between low MBL maternal level and complications of pregnancy, low or absent MBL was present in 16% of mothers experiencing miscarriage, while 5% of obstetrically normal individuals had low serum MBL (Kilpatrick, Bevan et al. 1995). MBL can be detected by immunohistochemistry in the first trimester fetoplacental unit; recent work has demonstrated that a degree of

inflammatory activity is required for the efficient implantation and normal pregnancy. Disturbance of this controlled inflammatory process appears to have a detrimental effect on the ability of the mother to support the developing foetus for a full gestational term, MBL playing a role in clearance of trophoblast debris and apoptotic cells (Christiansen, Nielsen et al. 2006). Premature birth was associated with an increased prevalence of the structural *mb12* B mutation (Annells, Hart et al. 2004), when compared with matched controls achieving gestation >37 weeks. The frequency of the *mb12* B polymorphism increases as the gestational age at birth decreases.

Cardiovascular disease and the outcome of myocardial infarction have serum MBL as a component in their pathogenesis. Atherosclerosis is a major contributory factor in the development of ischemic cardiovascular disease. Amongst Norwegian individuals who underwent coronary artery bypass grafting (CABG), valve replacement surgery or both, 13.2% were homozygous *mb12* defective as compared with 3% in matched blood donors (Madsen, Videm et al. 1998). In conjunction with the increased frequency of *mb12* mutations amongst individuals with cardiovascular disease, Hardy-Weinberg distribution of structural mutations D, B and C, differs greatly between the control and study group. Within the study group individuals homozygous for *mb12* mutation had a mean age of disease onset of 61 years, while normal or heterozygotes underwent their first operation at age 67. MBL-deficient individuals experience an earlier disease onset and/or a more progressive disease course than MBL-sufficient counterparts.

Mannan binding lectin plays a role in the damage inflicted on cells after a period of ischemia during subsequent reperfusion of the tissue; this is particularly relevant in both myocardial ischemia and cerebrovascular events. In vitro studies examining endothelial cells (HUVEC) exposed to oxidative stress clearly exhibited increased MBL surface binding, and was able to mediate iC3b deposition (Collard, Vakeva et al. 2000). Specific inhibition of MBL binding using an anti-MBL monoclonal antibody was able to significantly reduce iC3b deposition on the surface of endothelial cells exposed to oxidative stress.

The true weight of MBL in the immune response remains difficult to ascertain, functioning as it does in concert with other immune mechanisms. There certainly appears to be a reliance on MBL during periods of development or in the context of

primary immune deficiency, but the degree to which MBL is required to substitute for a secondary immune deficiency remains unclear.

## 1.9 Ficolin

Ficolins are the most recent family of recognition molecules to be conclusively associated with the lectin activation pathway, sharing the common theme of pattern recognition and MASP association. To date two serum ficolins have been characterised in humans, H-ficolin/Hakata antigen (Ficolin-3) and L-ficolin/P35 (EBP-37 or Hucolin) (Ficolin-2). A membrane bound ficolin-M/P35 (Ficolin-1), is predominantly expressed on peripheral blood leukocytes (Lu, Tay et al. 1996) of the monocyte lineage (Teh, Le et al. 2000). The ficolin family of protein continues to grow and shows a degree of heterogeneity that makes assigning homologues between species difficult, the question of synteny has been addressed between human and mouse species. Ficolin-3 homologue is not expressed in the mouse, but is present as a pseudogene located on chromosome 4 (Endo, Liu et al. 2004). Mouse ficolin A is located on chromosome 2 and is orthologous to human FCN2, FCN1 the human homologue of mouse ficolin-B, is also located on chromosome 2 separated by 2.3Mbp in the same orientation. Human FCN2 and FCN1 are located on chromosome 9q34, while FCN1 is located at 3p35. Ficolin homologues have been identified in human, rat, mouse and porcine species.

Ficolins share the common structural themes of MBL being composed of an N-terminal cysteine rich region, a collagen like domain, a neck region and finally a binding region that unlike MBL has a fibrinogen-like (FBG-like) domain. A structural unit is composed of triple helices of c. 35kDa, which form higher order functional units. Human ficolin-2 has been directly observed to form tetrameric functional units under the electron microscope, while Ficolin-3 (Hakata antigen) associates as a hexamer (Matsushita, Endo et al. 2001). Initial reports suggested ficolins played a purely opsonic role in immune defence (Matsushita M 1996), more recent studies have clearly shown a functional association of ficolin with MASP (Matsushita, Endo et al. 2000; Matsushita, Endo et al. 2001) and the activation of complement in an antibody independent manner (Tauber, Polley et al. 1976; Wilkinson, Kim et al. 1981; Loos, Clas et al. 1986). Fibrinogen domains determine binding specificity of ficolins in the same way that CRDs do in MBL. Both groups of proteins utilise collagen like regions for structural arrangement of their binding components enabling the recognition of micro patterns on the surface of foreign bodies. The range and scope of patterns recognised by

ficolins is only just emerging, essentially defining species of bacteria that possess the correct molecular lock to the ficolins key before more defined binding studies utilising purified cell wall components is undertaken.

A number of bacteria have been shown to both bind ficolin and activate complement, H-ficolin binds to *Salmonella typhimurium*, *S.minnesota* and *E. coli* (Sugimoto, Yae et al. 1998). L-ficolin was shown to bind whole fixed *Salmonella typhimurium* (Matsushita M 1996), and Lipoteichoic acid (LTA) preparations from *Staphalococcus aureus*, *Bacillus subtilis*, *Streptococcus pyrogenes*, *Streptococcus agalactiae* (Group B *Streptococci*) (Lynch, Roscher et al. 2003). A binding profile can be determined for each of the ficolins based on purified cell wall components that is distinct from that of MBL, the structural nature of these cell wall components often enables more than one of the ficolins and or MBL to bind to a particular bacterium preserving a redundancy in the system and potentially preventing bacteria from escaping the net by altering their surface expression profile.

Defined ficolin binding studies have identified acetyl groups as being crucial for Ficolin-2 binding; *N*-Acetylglucosamine exhibits greatest affinity with *N*-Acetylmannosamine, *N*-Acetylgalactosamine, *N*-Acetylcysteine, *N*-Acetylglycine and acetylcholine exhibiting decreasing binding profiles (Krarup, Thiel et al. 2004). *N*-Acetylglucosamine and *N*-Acetylmannosamine show binding to both MBL and ficolin-2, there is no shared binding with ficolin-3 for any of the above compounds. Ficolin-2 has been shown to have affinity for terminal *N*-acetyl-D-glucosamine, present on *Escherichia coli*, facilitating phagocytosis by the macrophage (Teh, Le et al. 2000).

It remains to be determined if the relative binding affinities of MBL and ficolins orchestrate subsequent events surround the destruction and disposal of invading microorganisms and foreign bodies, or play a role in the education of the adaptive immune system.

A number of polymorphisms in the human ficolin loci have been identified; functional consequences of such sequence variants have been assessed. 12 SNP's were characterised in the FCN1 loci, 5 in the promotor region, seven in the coding region, 4 at intron-exon boundaries. A single FNC3 polymorphism was identified, as a frame shift deletion in exon 5. FCN2 locus identified 5 polymorphisms in the promotor region

and a further 9 in the coding region, five of which reside in exons. Of the exonic polymorphisms two are silent mutation and three alter amino acid sequence, subsequently named FCN2-A (wild-type), FCN2-B (Thr<sup>236</sup>→Met) FCN2-C (Ala<sup>258</sup>→Ser) and FCN2-D (Ala<sup>264</sup> →Frame shift) observed in one individual only. Functional investigation of binding affinity towards GlcNAc revealed that variant FCN2-B has a reduced binding capacity, while FCN2-C showed greater binding capacity when compared with the FCN2-A genotype (Hummelshoj, Munthe-Fog et al. 2005).

Variation of serum FCN2 concentration is related to specific polymorphisms (position -986,-602 and -4) in the promotor region, no association was observed polymorphism within the coding region or promotor polymorphisms -557 or -64.

Ficolins have been measured in serum, bronchus/alveolus and bile (Akaiwa, Yae et al. 1999). Transcription of mRNA in the humans is primarily liver and spleen for ficolin-2 (Le, Lee et al. 1998), lung and liver tissue for ficolin-3 (Sugimoto, Yae et al. 1998; Akaiwa, Yae et al. 1999) and ficolin-1 expression in peripheral blood, lung, and spleen tissue (Endo, Sato et al. 1996; Teh, Le et al. 2000). In the murine system ficolin A mRNA can be detected in liver and spleen by RT-PCR, in situ hybridisation localised this expression to the macrophage cell type. Ficolin B transcript can be detected in bone marrow, spleen and embryonic liver, being produced by the myeloid lineage (Liu, Endo et al. 2005).

Ficolin act as part of the recognition arm of the lectin activation pathway, functioning to direct complement deposition or opsonise particles for phagocytosis. Ficolin-2 has a median serum concentration of 2.5-5.5µg/ml (Le, Lee et al. 1998; Atkinson, Cedzynski et al. 2004), individuals with circulating L-ficolin <1.8µg/ml show an increased propensity for atopy. Ficolin-3 has a serum concentration range of 7 to 23 µg/ml (Matsushita and Fujita 2001), recognized as an autoantigen in SLE patients.

## 1.10 L-ficolin

L-ficolin (Ficolin/P35 also called, ficolin L, elastin-binding protein (EBP)-37 or hucolin) is a polypeptide consisting of 35 kDa, forming a homo-trimeric subunit, associating into higher order tetramers to facilitate binding (Ohashi and Erickson 1998). The majority of L-ficolin expressed in the liver, the *FCN2* gene locus at 9q34. Circulating levels of L-ficolin range from 1.1-12.8 µg/ml with a median of 3.7µg/ml (Kilpatrick, Fujita et al. 1999).

L-ficolin differs in its binding capacity from that exhibited by MBL, exhibiting no affinity for mannose, while GlcNAc binding is achieved at a non-reducing terminal oligosaccharide in association with a galactose residue. Interaction with gram –ve *Salmonella typhimurium* and *Escherichia coli* is achieved via GlcNAc residues (Matsushita, Endo et al. 2000), in addition L-ficolin binds lipoteichoic acid, a gram positive bacterial cell wall component (Lynch, Roscher et al. 2004). In addition to a carbohydrate binding capacity L-ficolin also exhibits affinity for elastin (Harumiya, Omori et al. 1995) and corticosteroid (Edgar 1995), indicating a multifunctional role for this versatile protein.

## 1.11 H-ficolin

H-ficolin forms oligomers consisting of 34 kDa subunits linked by disulfide bonds, subunits associating to form a hexameric structure. H-ficolin has a broad range of tissue expression, being produced in pulmonary ciliated bronchial epithelial cells, type II alveolar epithelial cells and in the biliary system by hepatocytes and bile duct epithelia. *FCN3* locus at 1p35.5.

Circulating H-ficolin levels in serum from normal individuals ranges from 7 to 23 mg/ml (Matsushita and Fujita 2001).

H-ficolin (Hakata antigen) is a lectin that binds terminal GlcNAc and GalNAc but not mannose or lactose, binding activity towards PSA a polymer of repeating units of glucose, mannose, GlcNAc and xylose produced by *Aerococcus viridans*, (Tsujimura, Ishida et al. 2001) has been identified. H-ficolin is able to agglutinate human erythrocytes coated with lipopolysaccharides (LPS) derived from *S. typhimurium*, *S. minnesota* and *E. coli* (O111). The agglutination of erythrocytes coated with LPS from *S. typhimurium* is inhibited by GlcNAc, GalNAc and fucose, indicating that the binding of H-ficolin (Hakata antigen) to LPS is mediated by its lectin activity.

## 1.12 M-ficolin

M-ficolin, which can associate with MASP, activates the lectin pathway and specifically binds to *Staphylococcus aureus* (Liu, Endo et al. 2005). M-ficolin is synthesized by monocytes and is detected in both lung and spleen, located on the cell surface despite the absence of a membrane spanning anchor, a single polypeptide of 35kDa associates as a homo-trimer to generate a subunit with functional binding ability. The *FCN1* gene locus at 9q34, exhibiting a 79% homology with *FCN2* and thought to be a relatively recent gene duplication (Endo, Matsushita et al. 2007). M-ficolin exhibits binding to acetyl groups, maximal binding activity is associated with tetrameric subunits.

## 1.13 MBL and ficolin receptors

MBL has been shown to enhance phagocytosis independent of either complement activation or opsonisation with C3 and C4 fragments (Kuhlman, Joiner et al. 1989), conclusive identification of an MBL specific receptor remains to be gathered with evidence suggesting an overlap with previously characterised putative C1q receptors (Kuhlman, Joiner et al. 1989; Tenner, Robinson et al. 1995).

Interaction of polymorphonuclear cells with MBL bound surface ligands induce superoxide production (Uemura, Yamamoto et al. 2004), this activity can be completely abolished by treatment with pertussis toxin (PT) indicating a G-protein coupled receptor being involved in this interaction. CR1/CD35 present on leukocytes and erythrocytes has been shown to bind MBL, in addition to C3 fragments (Ghiran, Barbashov et al. 2000). C1q is also able to bind CR1 in a competitive manner with MBL, independent of divalent cations.

MBL has been shown to bind to monocytes, dendritic cells (Downing, Koch et al. 2003) and macrophages (Nadesalingam, Dodds et al. 2005). Binding can be inhibited by specific monosaccharide moieties and in contrast to previously characterised MBL/C1q associating receptors are not cation independent, binding being abolished by EDTA. Studies have indicated that this cation-dependent MBL receptor is markedly up-regulated on immature dendritic cells (Downing, MacDonald et al. 2005). Functional studies of MBL receptor binding in macrophages indicates that MBL/peptidoglycan complex binding is able to induce a cytokine response, inhibiting the inflammatory response to peptidoglycan via reduced TNF $\alpha$  expression and increased IL-8 and RANTES production (Nadesalingam, Dodds et al. 2005). Shifts in cytokine profile of this sort suggest MBL may down-regulate macrophage mediated inflammatory reactions, whilst enhancing recruitment of phagocytes.

Specific protein-protein interactions between MBL and cell surface receptors have been characterised, broadly these fall into two groups. CR1/CD35 that is able to bind MBL, C1q and C3 fragments, with significant overlap in the binding properties of collagenous

MBL and C1q. An uncharacterised, functionally defined and distinct receptor able to bind MBL, identified on the surface of monocytes and macrophages.

## 1.14 Mannan binding lectin associated serine protease

In common with the classical activation pathway recognition molecule C1q, MBL and ficolin are found to associate with a number of accessory molecules giving the complex complement activating ability. The associated molecules are a group of four related proteins termed MASP (mannan binding lectin associated serine protease) generated by alternative splicing events from two distinct genes (Takahashi, Endo et al. 1999; Dahl, Thiel et al. 2001), comprising three enzymatic and one non-enzyme protein.

The early characterisation of MASP is complex and often misleading due to their high degree of similarity. Mannan binding lectin associated serine protease (MASP) were identified as components of the Ra reactive factor, a 300kDa macromolecular complex found to bind Ra polysaccharide positive bacterial strains and cause lysis in a C4 dependent manner (Ihara, Harada et al. 1982). Protein separation techniques were used to isolate individual components and their fragments that were then subjected to N-terminal sequencing, this strategy successfully identified MBL as a constituent of the Ra reactive factor and also the first MASP. Protein sequencing data allowed the generation of specific degenerate primers to isolate a fragment of the protein coding gene, subsequent cDNA library screening or PCR techniques were used to complete the gene sequence. This strategy was successful in characterising three MASP and a related but non-catalytic molecule Map19 found to associate with MBL (Thiel, Vorup-Jensen et al. 1997; Stover, Thiel et al. 1999; Dahl, Thiel et al. 2001) and ficolins (Matsushita, Endo et al. 2000; Cseh, Vera et al. 2002; Schwaeble, Dahl et al. 2002), conducting the binding event to further downstream complement components.

MASP-1 (P-100, RaRf), MASP-2, MASP-3 and Map19 are the products of two distinct gene loci, generating distinct protein products via alternative splice mechanism (Stover, Thiel et al. 1999; Dahl, Thiel et al. 2001). Primary sequence information indicated MASP were modular serine protease of the chymotrypsin like family sharing a structural architecture that is also found in the classical pathway serine proteases C1r and C1s (Fig.6), each composed of six protein domains: An N-terminal CUB domain, an EGF-like domain, a second CUB domain, two CCP complement control protein

domains and a C-terminal serine protease domain. The related protein MAp19 consists of the N-terminal CUB and EGF domains, followed by four amino acids distinct from other MASP.

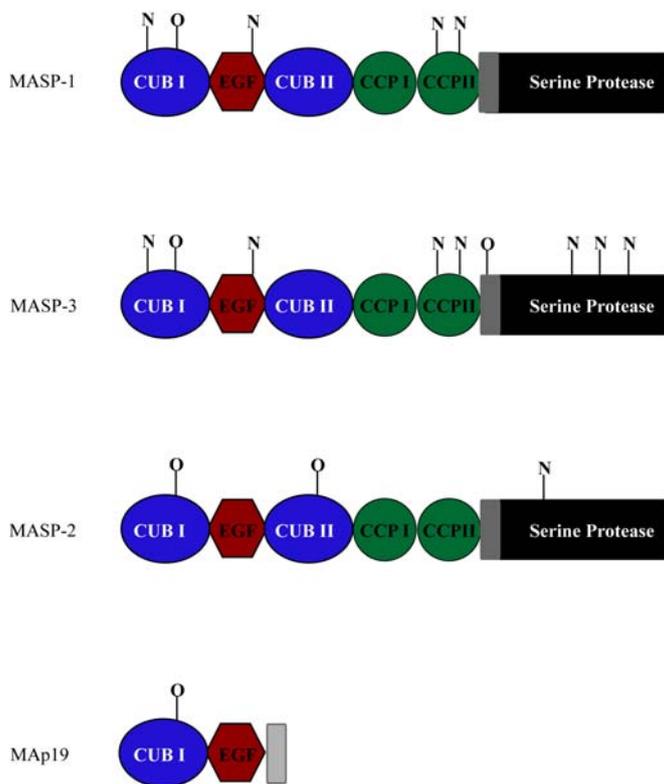


Fig.6 Protein domain structure of Mannan binding lectin Associated Serine protease (MASP) and Map19, predicted N and O-glycosylation sites are highlighted. CUB - Complement C1r/C1s, Uegf, Bmp1, EGF – Epidermal Growth Factor, CCP – Complement Control Protein.

MASP are secreted under the control of a signal peptide as a single chain zymogen, association with MBL or ficolin and subsequent activation induces an activation event producing two polypeptides linked by a disulphide bond. This active form of MASP consists of a heavy chain encompassing the 5 N-terminal domains and light chain consisting of the serine protease domain, the cleavage event occurring within a link region of 14-15aa separating CCP II from the serine protease domain at a conserved Arg↓Ile motif.

The multi domain nature of MASP and classical pathway complement serine proteases (MASP-like) set them apart from other members of the chymotrypsin-like family and

may account for their characteristically high specificity and low activity. MASP show a highly restricted substrate specificity this combined with a relatively low activity effectively constrains activation of the complement system, preventing unwanted complement deposition. MASP catalytic activity resides solely within the serine protease domain; N-terminal domains appear to be important in substrate selection and stabilisation greatly increasing the reaction rate.

In trying to order the events surrounding the initial activation steps of the lectin pathway, the enzymatic activities of the MASP enzymes have become a key focus. Much early work was based on effectively partially purified proteins that lead to some conflicting results (Ikeda, Sannoh et al. 1987; Ii, Wada et al. 1988) more recent studies utilising recombinant molecules and highly purified components has more clearly defined the properties of MASP (Rossi, Cseh et al. 2001).

MASP-2 functions as initial activator and signal transducer of lectin pathway activation (Vorup-Jensen, Petersen et al. 2000), combining the roles of classical pathway C1r and C1s. Autoactivation of MASP-2 occurs initially as a zymogen molecule cleaves another zymogen molecule (Gal, Harmat et al. 2005), association with MBL potentiates this activation event (Wallis, Lynch et al. 2005) via conformational changes in MBL elicited during multiple CRD binding. Conversion of proenzyme MASP-2 to activated MASP-2 confers further enzymatic activity against complement components C2 and C4 (Rossi, Cseh et al. 2001). Activated MASP-2 cleaves C4 to release C4a and C4b at an efficiency 23 fold greater than C1s, cleavage of C2 generating C2a and C2b occurs at a 3 fold greater rate (Rossi, Cseh et al. 2001). Cleavage of C4 exposes a highly reactive thioester group within the C4b fragment, this is rapidly hydrolysed by H<sub>2</sub>O or if sufficiently close to an activating surface or protein component is able to covalently bind to hydroxyl or amino groups. Assembly of the C3 convertase C2bC4b at or near the initiating surface is accomplished with subsequent association of C2b with surface bound C4b; cleavage of C3 marks the transition into the terminal pathway of complement activation.

MASP-2 appears to be sufficient to propagate lectin pathway activation in conjunction with recognition molecule binding; MASP-1, MASP-3 and Map19 are all found in

association with activating complexes of the lectin pathway (Dahl, Thiel et al. 2001) their role is less certain.

MASP-1 has evidenced minor activity towards C3 (Matsushita, Endo et al. 2000; Matsushita and Hamasaki 2000; Kenjo, Takahashi et al. 2001), recombinant studies have indicated that this activity is highly marginal and may not be biologically relevant. Cleavage of C3(H<sub>2</sub>O) has been observed at an appreciable rate at high enzyme concentrations (Hajela, Kojima et al. 2002; Kojima 2002), at a rate 100 fold that of C3. The function of this remains unclear. MASP-1 exhibits cleavage activity against components of the clotting cascade, fibrinogen and plasma transglutaminase (factor VIII) (Hajela, Kojima et al. 2002). Affinity purified MASP-1 has cleavage activity against fibrinogen and coagulation Factor VIII at 10-30% that of thrombin, this level of activity appears to be biologically relevant. Cleavage of fibrinogen by MASP-1 releases fibrinopeptide B, a potent chemotactic molecule for fibroblasts, macrophages and neutrophils (Senior, Skogen et al. 1986; Skogen, Senior et al. 1988). The most significant function of MASP-1 appears to lie outside of the lectin pathway, its role in the cleavage of Factor D (Takahashi, Ishida et al. 2010) of the alternative activation pathway.

Activated MASP-3 exhibited no enzymatic activity against complement proteins C2, C4 or C3 and does not react with C1 inhibitor. Synthetic substrate *N*-carboxybenzyloxyglycine-L-arginine thiobenzyl ester substrate elicited a significant cleavage activity.

There appears to be enormous complexity within the lectin pathway activation complexes, three fluid phase and one membrane coupled pattern recognition molecule orchestrating with three serine protease and a non-enzymatic binding protein has the flexibility of association to produce a complex route to achieve C3 cleavage. Elucidation of the enzymatic activities of the MASP has significantly simplified the view of these varied activation complexes, the lectin activation pathway predominantly generates a C3 convertase, C4bC2b, via a binding of a pattern recognition molecule and activation of MASP-2. Cleavage of C4 and C2 by MASP-2 is the crucial event enabling C3b to become covalently bound to an activating surface as the initiating step in assembly of a C5 convertase. A single enzyme MASP-2 both detects the activation

event and transduces the signal downstream, in contrast to the classical pathway, which separates these functions into two enzymes C1r and C1s.

## 1.15 MASP protein domains

Structurally MASP and the MASP-like serine protease of the classical activation pathway are composed of six domains, consisting of four distinct modular types. The interaction of these protein domains and the properties of each, combine to produce the unique enzymatic properties observed for each of the MASP-like serine protease.

### 1.15.1 CUB domain

The CUB domain was initially thought to be limited to components of developmental proteins (Bork 1991), the domain has subsequently been identified in a range of functionally diverse proteins (Bork and Beckmann 1993). The CUB domain is a feature of a number of complement proteins, deriving its name from its discovery in C1r/C1s, sea Urchin epidermal growth factor and human Bone morphogenic factor. The CUB domain in common with many other extracellular proteins appears to be limited to eukaryotic species, and is yet to be identified in prokaryotes, plants or fungi.

Classification of this structure is dependent on a number of characteristic features; each CUB domain spans approximately 110 amino acids containing four highly conserved cysteine residues. A highly conserved region of amino acids is present in the N-terminal region of the CUB domain, the consensus sequence is characteristic of the CUB domain (Bork 1991).

Serine – Proline – polar – Tyrosine – Proline – polar – x – Tyrosine

CUB domains share a conserved disulphide bond structure, 4 cysteine residues forming disulphide linkages in a C1-C2, C3-C4 pattern. Exceptions to this are those modules found in the complement components C1r and C1s, where only the second pair of cysteine residues is present. Predicted CUB domain structure shares a common beta-barrel similar to that of immunoglobulin. The CUB domains present in MASP and MASP-like serine protease of the complement system are characteristically paired and separated by an epidermal growth factor domain, in an N-terminal CUB1-EGF-CUB2 motif. This region facilitates interactions between the recognition and protease subunits.

### 1.15.2 Epidermal growth factor domain

The epidermal growth factor (EGF) motif is present in a wide range of extracellular proteins. Involved in protein-protein and protein-cellular interactions the vast array of EGF-like domains have arisen from a common ancestor by a process of gene duplication, point mutation and exon shuffling. EGF domains are single exons of approximately 50 amino acids, highly conserved cysteine residues (Appella, Weber et al. 1988) produce a disulphide bonding pattern characteristic of this protein domain. Primary structure and spatial arrangement of cysteine residues allow the structure to form 3 distinct loops, thought to enhance protein-protein interactions.

Ca<sup>2+</sup> binding within the EGF domain take place in conjunction the CUB module present in MASP and MASP-like serine protease (Thielens, Enrie et al. 1999), and is vital for efficient protein-protein interactions as seen in the model system of C1r-C1s association. Association of Ca<sup>2+</sup> allows the EGF-CUB module to form a more compact structure (Bersch, Hernandez et al. 1998), the intermolecular movement enabling the complex as a whole to undertake biological activity (Kelly, Dickinson et al. 1997).

### 1.15.3 Complement control protein domain

Complement control protein (CCP) domains as a paired units are found in a number of complement proteins, including MASP-like serine protease, Factor H, DAF, MCP, C3b-BP, CR1 and CR2. CCP domains mediate protein-protein interactions in the context of surrounding domains modulating considerably the binding properties of the protein as a whole, this has been clearly demonstrated in C1r (Budayova-Spano, Lacroix et al. 2002).

CCP modules are approximately 60 amino acid residues in length and characterised by a consensus sequence that includes four invariant cysteine residues, producing C1-C3 and C2-C4 disulphide bridges. Highly conserved tryptophan, proline, glycine and hydrophobic residues are located along 8  $\beta$ -stranded structures (Kirkitadze and Barlow 2001). CCP modules are separated by a linking region, which is commonly 4 amino acids of variable sequence, but can range from 2-7 residues. Structural data from C1s indicates the CCP module is highly flexible, allowing the serine protease domain considerable mobility within the C1 complex. This motility allows C1s to translate to a more external position in the complex and interact with C1r, inducing its activation (Gaboriaud, Rossi et al. 2000). This mobility is vital to enabling individual proteins to function in multi-protein complexes, opening more opportunities for interaction with other components (Gaboriaud, Thielens et al. 2004).

CCP modules of the MASP and MASP-like serine protease are highly involved in substrate binding, playing a crucial role in determining enzyme activity (Rossi, Teillet et al. 2005). C4 association and substrate activity are highly dependent on the CCP domain; studies of chimeric C1s and MASP-2 constructs provide evidence that their respective serine protease have equal activity against C4. The utilization of distinct C4 binding sites within the respective CCP domains results in the differing activities observed in the native proteins.

#### 1.15.4 Serine protease domain

Serine proteases have a long evolutionary history, diversifying to provide a vast array of activities ranging from degradative processes to protein remodelling, fertilisation, embryogenesis, and blood clotting, cellular and innate immunity. The ability of a particular enzyme to carry out its specific function relies on precisely positioned and highly conserved amino acid residues in and around the active site, and it is through the study of these amino acids that enzymes can be classified into clans and families.

Serine proteases provide the backbone of the complement system, linking recognition events to the effector arm of the membrane attack complex. To date nine proteases have been identified together with a non-enzymatic but structurally related molecule, each function in concert with other complement proteins to transduce and regulate the activation signal. Complement serine proteases have a relatively low enzymatic activity and are highly restricted in both their substrate specificity and activation. Many proteases are synthesized as zymogens or require other enzymatic fragments as cofactors, a characteristic that gives the system as a whole a highly controlled and directional mechanism. At a structural level complement proteases can be divided into two groups, those consisting solely of a catalytic domain and multi domain enzymes. A group of serine protease working in conjunction with recognition molecules of the classical and lectin pathways act as the complement cascade trigger mechanism, detecting a binding event and initiating activation.

Associating with the recognition components of the complement system this group of related serine protease, due to their structural and evolutionary similarity, are known as the MASP-like serine protease.

Serine proteases utilise a triad of amino acids to drive the catalytic process, while other structural and amino acid charge features of the protease determine substrate specificity (Guy Dodson 1998). MASP-like serine proteases coordinate Histidine - Aspartic acid and Serine (H-D-S) at their active centre, the triad forms a highly polarised charge relay system capable of stabilising a substrate-enzyme transition state within the specificity (S1) pocket of the enzyme before proceeding to break the substrate peptide backbone (Rupert C. Wilmouth 2001).

The acid-base-ser/thr pattern is highly conserved amongst serine protease with codon usage being an evolutionary diagnostic feature (Krem and Di Cera 2001). A number of other highly conserved residues have been identified. The triad model of coordinated amino acids at the primary specificity site is sometimes expanded to include a fourth amino acid at position 225; although not directly in contact with the substrate this residue profoundly affects the architecture of the water channel of the primary recognition site. The water channel connects the active site to an exit point on the underside of the molecule, potentially providing an escape route for water molecules as substrate binding takes place. The correct and efficient displacement of water from the active site appears to be a component of efficient substrate recognition and specificity (Enriqueta R. Guinto and Cera 1999).

The C-terminal region of serine protease plays a vital function in determining substrate specificity; this region is particularly rich in conserved residues, the 220 loop ranges from residue 189 itself highly conserved and encompasses a number of crucial residues by virtue of its close association with the specificity pockets of the active site.

The link region provides a bridge between the second Complement Control Protein (CCP II) domain and the serine protease domain, consisting of 12-15 aa and containing a number of highly conserved amino acids particularly Arg↓Ile cleavage motif. The link region is encoded by the serine protease exon.

Trypsin and chymotrypsin are structurally similar, although they recognise different substrates. The proteins belong to MEROPS peptidase family S1 (chymotrypsin family, clan PA(S)), subfamily S1A. Trypsin initiates its cleavage event around lysine and arginine residues, whilst chymotrypsin has specificity for large hydrophobic residues such as tryptophan, tyrosine and phenylalanine, both with extraordinary catalytic efficiency. Both enzymes have a catalytic triad of serine, histidine and aspartate within the S1 binding pocket, although the hydrophobic nature of this pocket varies between the two, as do other structural interactions beyond the S1 pocket.

In seeking to determine the parent MASP gene or events leading to the specific combination of modular domains, it is important to order the events leading from that point by means of an evolutionary sequence or model of the molecular events that may have taken place. At this time this study is based primarily on the data obtained from cDNA sequence, due to the limited genomic sequence available from more exotic species.

## 1.16 MASP-1

MASP-1 was the first MBL associating serine protease to be identified, establishing the lectin pathway as independent of the classical pathway of complement activation. Identified in humans as an 83kDa protein that disassociated from the complex at low pH, and present as two disulphide linked polypeptides of 66kDa and 31kDa. N-terminal sequencing of the 31kDa light chain clearly indicated a serine protease similar to C1r and C1s (Matsushita and Fujita 1992), and to the previously described mouse RaRf protein P100 subsequently determined to be the mouse homologue of MASP-1.

Purified MASP-1 was ascribed the ability to cleave C2 and C3 (Matsushita, Thiel et al. 2000). Subsequent experiments using recombinant proteins show that C2 cleavage ability can be clearly associated with MASP-2 (Rossi, Cseh et al. 2001). The low C3 cleavage activity of MASP-1 appears to be a feature of serum purified protein as recombinant MASP-1 shows no such activity, the true nature of the C3 cleavage remains uncertain as the recombinant protein lacks the native posttranscriptional modifications of the purified protein while the purified protein may contain contaminating protease activity. Both C1-inhibitor (Petersen, Thiel et al. 2000) and  $\alpha$ 2-macroglobulin (Gulati, Sastry et al. 2002) have been shown to inhibit MASP-1 activity, although the mechanism is as yet unknown.

MASP-1 is found in association with MBL, H-ficolin and L-ficolin (Matsushita, Endo et al. 2000), as part of complexes exhibiting complement activating ability. The stoichiometry of the MBL MASP-1 functional unit has been studied using MBP-A and MBP-C in the rat (Chen and Wallis 2001) and the model has been extrapolated to the human system. MASP-1 dimers associate preferentially with MBP-A subunit trimers and tetramers creating a link between MBP stalks around the hinge region. Analytical ultra centrifugation techniques also indicate MASP-1 in human serum is preferentially associated with monomeric and dimeric MBL (Dahl, Thiel et al. 2001).

Unlike the C1 complex with its absolute requirement for divalent  $\text{Ca}^{2+}$  ions, the MASP-MBL complex has been shown to be resistant to dissociation in the presence of chelating agents (Chen and Wallis 2001). There is still a requirement for  $\text{Ca}^{2+}$  in the binding of carbohydrate recognition domains to their molecular pattern, but this is not matched by a similar need in the association of MASP with MBL.

MASP-1 occurs in serum at a mean concentration of  $6\mu\text{g/ml}$  (Terai, Kobayashi et al. 1997) ranging between  $1.5$  and  $13\mu\text{g/ml}$ , none could be detected in CSF or urine. In another departure from the model set by the C1 complex, MASP-1 levels appear to be far higher than those seen for MBL indicating that a considerable portion up to 95% (Thiel, Petersen et al. 2000) is not present in preformed MASP-MBL complexes. A proportion of this “free” MASP-1 may in fact be associated with ficolin pattern recognition molecules, although there is still a considerable excess of the serine protease. The differences in formation of the MBL-MASP complexes, coupled with the absence of extreme variation between individuals seen in MBL levels exhibits a very different profile compared with its functional partner the C1 complex. L-ficolins and H-ficolin associate with MASP-1, the complex including other MASP having C2, C3 and C4 proteolytic activity (Matsushita, Endo et al. 2000; Matsushita, Kuraya et al. 2002). MASP-1 is found in association with both surface bound and pelagic ficolin, as seen by solid matrix purification and immunoprecipitation respectively.

Isolation of MASP-1 gene sequence information was achieved using a probe generated using degenerate oligonucleotides based on protein sequencing data, the probe was successfully used to screen a human liver cDNA library (Sato, Endo et al. 1994) identifying a serine protease showing a 40% homology at the level of the DNA to C1r and C1s from the same species. The translated cDNA indicated that MASP-1 protein exhibits a modular domain structure familiar to the C1r and C1s serine protease of the classical pathway described at that time and subsequently other MASP, consisting of an N-terminal CUB domain, EGF-like domain, second CUB domain and two CCP domains followed by a short link region and a serine protease domain (Fig.7).

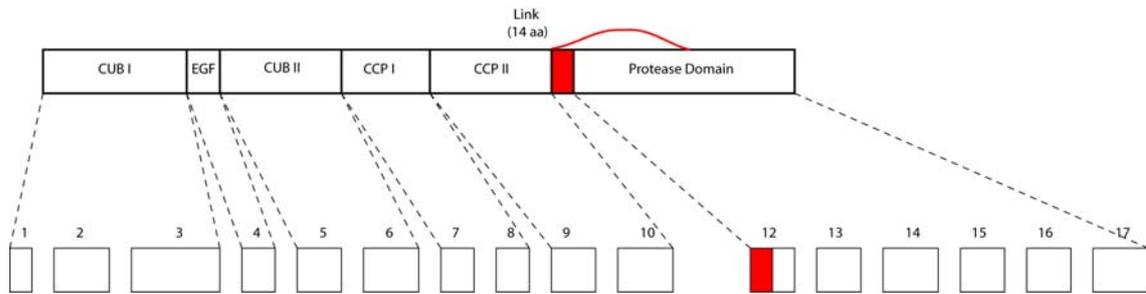


Fig.7 MASP-1 protein domain structure (upper panel) linked to each corresponding coding exon (lower panel). CUB - Complement C1r/C1s, Uegf, Bmp1, EGF - Epidermal Growth Factor, CCP - Complement Control Protein.

This similarity in overall structure is not matched at the level of the gene, where MASP-1 exhibits a number of unique features. MASP-1 sequences have been examined in human, rat, mouse, bovine and xenopus species all of which exhibit characteristics which differ from other characterised MASP-like serine protease. MASP-1 is absent from avian species as determined by genomic sequence in *Gallus gallus* and northern blot hybridisation for other avian species (Lynch, Khan et al. 2005). The human *MASP1* gene transcript derived from the human *MASP1/3* locus is composed of 16 exons; the first 10 exons encode protein domains characteristic of all MASP-like proteases, CUB domain, EGF-like domain, second CUB domain and two CCP domains. Unlike the serine protease domains of previously described classical pathway serine protease C1r and C1s, that of MASP-1 is the product of 6 spliced exons. MASP-1 active site histidine is stabilised by a histidine loop structure that is also a common feature amongst more primitive chymotrypsin proteases, along with a diagnostic active site serine codon TCN. Other members of the MASP-like serine protease have an AGY codon at the active site serine residue, analysis of this codon has been used to map evolutionarily associations between serine protease (Krem and Di Cera 2001).

Northern blot analysis using a MASP-1 cDNA probe indicated a transcript of 4.8kb expressed predominantly in liver and present in both human and mouse, a minor transcript of 3.4kb was also observed (Knittel, Fellmer et al. 1997). The availability of sequence data and cDNA probes allowed chromosomal localisation of the MASP-1 gene in humans to 3q27-28, 16B2-B3 in mouse (Takada, Seki et al. 1995), rat chromosome 11q23, *Bos taurus* chromosome 1, *Canis familiaris* chromosome 34.

MASP-1/3 promotor studies have localised significant expression elements from -1 to -283, within this region are a TATA-like sequence (TFIID) at -69, a hepatocyte nuclear factor (HNF-5) binding site at position -68 potentially conferring liver specific expression and Pit-1 (GHF-1) a pituitary specific element. Significant promotor activity is lost when the minimal region is deleted to -214bp, a 3' deletion of the -1 to -71bp promotor completely removes promotor activity. The promotor element was shown to be up regulated by IL-1 $\beta$  stimulation, an effect that can be abolished by IL-6 (Endo, Takahashi et al. 2002). INF- $\gamma$  stimulation had an inhibitory effect on the reporter construct, in contrast to the classical pathway C1s promotor activity that was enhanced by INF- $\gamma$ .

## 1.17 MASP-2

MASP-2 became the second lectin pathway serine protease to be identified (Thiel, Vorup-Jensen et al. 1997) and may in fact be the most crucial in terms of complement activation, the functional significance of other MASP as yet to be determined.

MASP-2 was identified as a copurifying component of  $\text{Ca}^{2+}$  dependent carbohydrate binding complexes obtained from human serum. Initial N-terminal protein sequencing of the 90kDa protein showed striking similarity to MASP-1, C1r and C1s (Vorup-Jensen, Jensenius et al. 1998). The purified protein was seen to be present as a 30kDa and 55kDa pair on reduction, indicating two disulphide linked domains, the larger of which, in keeping with complement nomenclature, is termed the A-chain and the smaller the B-chain.

MASP-2 in serum is predominantly found unbound to MBL; 95% circulates freely (Thiel, Petersen et al. 2000). Association of MASP-2 with MBL is mediated by homodimers of the first three modular domains (Wallis and Dodd 2000) the CUB domains making contact with MBL while calcium binding EGF domain maintains correct orientation, each dimer interacting with two MBL subunits (Chen and Wallis 2001) along the collagen axis. The crystal structure of MASP-2 CUB1-EGF-CUB2 demonstrates the potential arrangement of homodimers with MBL (Feinberg, Uitdehaag et al. 2003)

Using N-terminal protein sequence to design degenerate primers, a region of the gene was amplified from human liver RNA and subsequently used as a probe to obtain cDNA clones encompassing the entire coding sequence. A 2455bp clone (accession number Y09926) with an open reading frame of 686aa was isolated (gene structure figure), exhibiting an identical modular domain structure to MASP-1, C1r and C1s. Northern blots of human liver using a MASP-2 cDNA probe confirmed a transcript of 2.5kb; this was also demonstrated in mouse, rat and guinea pig liver RNA preparations. A number of other transcripts can also be detected in mouse and guinea pig RNA preparations at 2kb and 3kb, a further transcript of 1kb is present in abundance in all liver RNA preparations (Stover, Thiel et al. 1999).

Characterisation of the human gene was quickly followed by rat and mouse homologues (Stover, Thiel et al. 1999) and an alternatively spliced gene product MAP19. The presence of a MASP-2 transcript can be confirmed in human, rat, mouse and guinea pig. The genomic location of the MASP-2 locus has been determined in humans to be on chromosome 1p36.3-2 (Stover, Schwaeble et al. 1999), mouse on chromosome 4 (Stover, Lynch et al. 2004) and rat on 5q36 as determined by sequence database analysis.

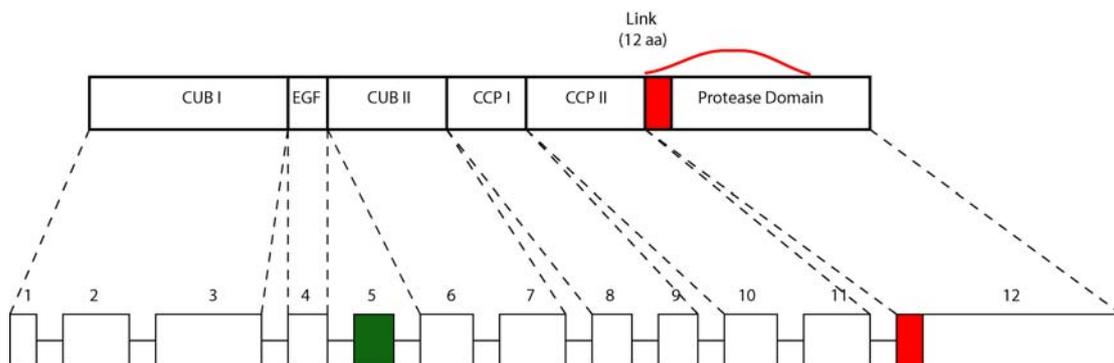


Fig.8 MASP-2 protein domain structure (upper panel) assigned to respective exons (lower panel). CUB - Complement C1r/C1s, Uegf, Bmp1, EGF – Epidermal Growth Factor, CCP – Complement Control Protein.

In common with the previously described MASP-1, C1r and C1s, MASP-2 cDNA derived amino acid sequences exhibit an identical domain architecture. In all species examined the 6 protein domains of MASP-2 (Fig.8) are encoded by 11 exons, the N-terminal 4 of which are shared by MAP19. The gene consists of 12 exons, generating two distinct mRNA products by alternative splice processes (Stover, Thiel et al. 1999). The proteolytic activity of activated MASP-2 has been studied using both purified (Matsushita, Thiel et al. 2000) and recombinant material (Rossi, Cseh et al. 2001), in both cases a C2 and C4 cleavage activity could be demonstrated at rates that are compatible with useful biological activity. MASP-2 shows substrate specificity similar to classical pathway C1s, in comparison MASP-2 cleaves C2 and C4 with 3- and 23-fold greater efficiency respectively. The previously described C3 cleavage by MASP-2 could only be reproduced at the extremes of enzyme concentrations and incubation times, rendering such activity biologically implausible (Rossi, Cseh et al. 2001).

Total MASP-2 protein levels in healthy individuals has been determined to have a range of 170ng/ml and 1196ng/ml (Moller-Kristensen, Jensenius et al. 2003), rather than a normal distribution about the mean, a logarithmic normal distribution (median 497ng/ml) is seen. An Asp – Gly mutation in the CUBI domain reduces circulating levels.

MASP-2 is synthesised primarily in the liver although mRNA can be detected in kidney and lung by both in situ hybridisation and RT-PCR (N Lynch, unpublished data). Expression and promoter studies have sought to determine controlling factors for MASP-2 regulation, in particular change in relation to the acute phase response and infection models.

The pivotal role of MASP-2 within the lectin pathway has led it to become a focus of interest regarding its association with recurrent infections and in particular the search for individuals who are deficient in MASP-2 expression or who exhibit point mutations associated with an altered immune response. To date there are three characterised individuals who are functionally deficient of MASP-2 (Stengaard-Pedersen, Thiel et al. 2003). Of most interest was the documented patient with a history of inflammatory disease and recurrent infection with *pneumococcus spp.* Heterozygous family members showed no such susceptibility to infection or inflammatory disease. Other individuals found to be functionally MASP-2 deficient are heterogeneous in their presenting clinical picture; indeed one individual is completely healthy and expressed no obvious phenotype. The results from a small number of patients should be taken advisedly; the identification of other homozygous individuals is required to fully explore such disease associations.

## 1.18 MAp19/sMAP

MAp19 is a plasma protein of 19kDa co-purifying with MBL and found to cross-react with antibodies directed against the N-terminus of MASP- 2 (Stover, Thiel et al. 1999; Takahashi, Endo et al. 1999). Mouse and rat sequence data for the MAp19 specific transcript were derived from EST database searches using MASP-2 N-terminal sequence information, complete sequencing revealed the alternative splice product (Stover, Thiel et al. 1999). Human MAp19 transcript was isolated along with MASP-2 cDNA clones while screening a human liver cDNA library using a 5' 450bp sub fragment of MASP-2. Several transcripts were identified, database searches yielding a 725bp cDNA termed sMAP while library screening characterised cDNA's of 736bp (phl-5) and 729bp (phl-6) designated MAp19. The independently isolated cDNA clones showed an identical 516bp open reading frame encoding an 171aa protein, exhibiting complete homology to MASP-2 5' untranslated region, CUB I and EGF domains with the addition of four amino acids (EQSL) (Fig.9), a stop codon and a 3' untranslated region. MAp19 was conclusively linked to the MASP-2 gene locus as a splice variant by genomic sequence analysis and confirmed as a single hybridising species by southern blot. MAp19 utilises the first 4 exons of the MASP-2 gene and a fifth MAp19 specific exon, absent from the mature MASP-2 transcript. Homologues of MAp19 have been characterised in human, rat, mouse (Stover, Thiel et al. 1999), chicken and a functional homologue (MRP) in the common carp has been described (Nagai, Mutsuro et al. 2000).

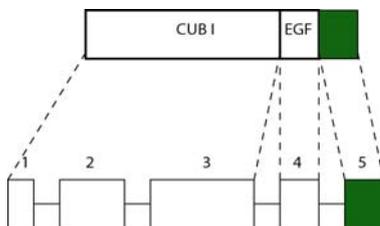


Fig.9 Map19 protein domain structure (upper panel) aligned with respective exons (lower panel). Highlighted green section (exon 5) indicating the Map19 specific amino acid residues. CUB - Complement C1r/C1s, Uegf, Bmp1, EGF – Epidermal Growth Factor.

MAp19 has no enzymatic activity, consisting of the N-terminal CUB domain, calcium binding EGF domain and four additional amino acids (Stover, Thiel et al. 1999). MAp19 is found in association with complexes of both MBL (Dahl, Thiel et al. 2001) and Ficolins (Cseh, Vera et al. 2002) with complement activating ability, although a clear role for MAp19 is yet to be delineated. MAp19 is conclusively linked to MASP-2 in all species examined to date, and so shares common chromosomal localisations (Fig.10).

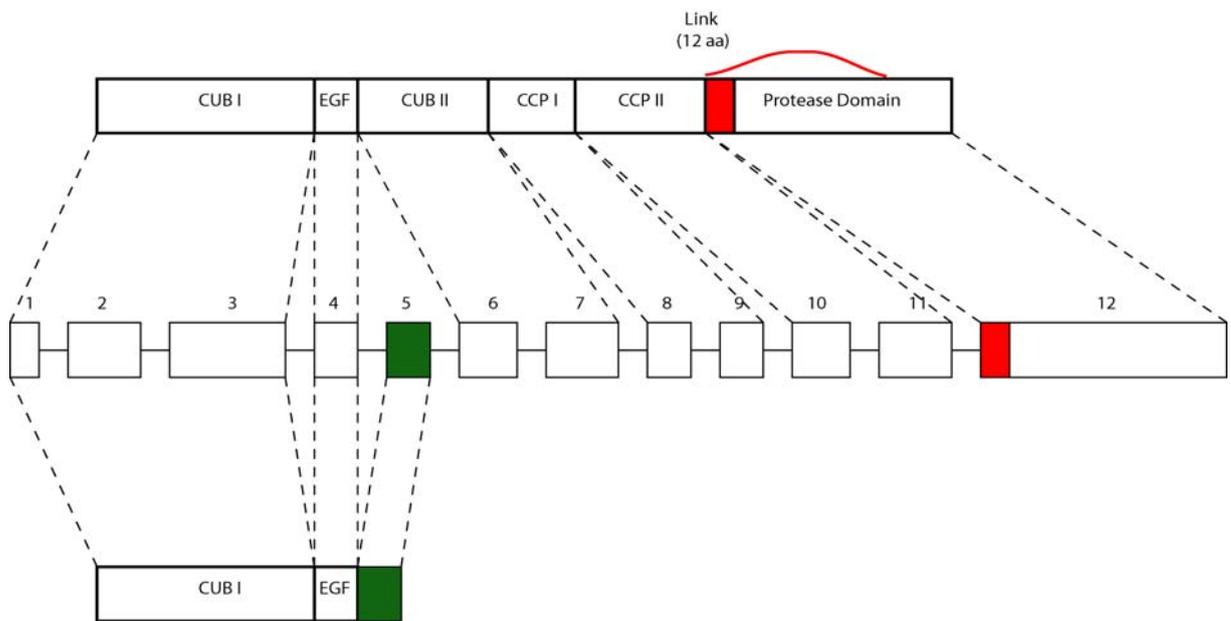


Fig.10 MASP-2 locus (centre panel) indicating MASP-2 (upper panel) and Map19 (lower panel) alternatively spliced transcripts, protein domains are indicated with respect to corresponding exons. Highlighted red section indicated the link region. Red line represents the Disulphide Bridge between A and B-chain of the activated protein. Highlighted green section (exon 5) indicating the Map19 specific amino acid residues. CUB - Complement C1r/C1s, Uegf, Bmp1, EGF - Epidermal Growth Factor, CCP - Complement Control Protein.

## 1.19 MASP-3

MASP-3 is the most recent MBL associated serine protease to be identified, using the strategy of protein purification and sequencing followed by a degenerate RT-PCR and library screening. In this case the library was an EST sequence database (Dahl, Thiel et al. 2001). Full sequence information was rapidly compared to the human genomic sequence database, clearly indicating that the newly characterised transcript was in fact an alternatively spliced variant of MASP-1 rather than a separate gene. A transcript of 2.4kb was characterised (accession number AF284421) encompassing an open reading frame of 725aa, exhibiting an identical domain structure to the previously described MASP and MASP-like serine protease (Fig.11). Predicted translation of the human cDNA indicated a high degree of homology with MASP-1, the first 5 domains showing complete identity while the serine protease domain exhibited 32% identity.

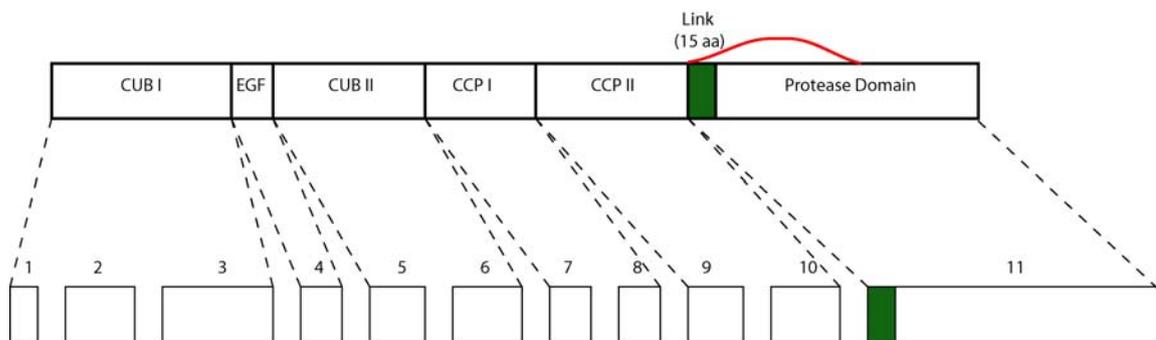


Fig.11 MASP-3 protein domain structure (upper panel) linked to respective exons (lower panel). Highlighted green section indicated the link region. Red line represents the Disulphide Bridge between A and B-chain of the activated protein. CUB - Complement C1r/C1s, Uegf, Bmp1, EGF – Epidermal Growth Factor, CCP – Complement Control Protein.

MASP-3 has been shown to complex with MBL exhibiting a regulatory function over the C2 and C4 cleavage activity of MASP-2, the mechanism of this is unknown (Dahl, Thiel et al. 2001). Ficolins also co purify with MASP-3, direct evidence for this is only available for Hakata antigen or H-ficolin (Matsushita, Kuraya et al. 2002).

MASP-3 shows no autoactivation activity, mutant constructs expressed in a baculovirus expression system exhibited no cleavage when incubated overnight at 37°C (Zundel, Cseh et al. 2004). Activated MASP-3 exhibits no detectable enzymatic activity towards complement components C2, C4, and C3; only the synthetic substrate *N*-

carboxybenzyloxyglycine-L-arginine thiobenzyl ester demonstrated significant cleavage activity. Recent studies have identified a further 3 synthetic substrates that are cleaved by MASP-3, insulin-like growth factor binding protein-5 is currently the only native protein identified as a substrate for MASP-3 (Cortesio and Jiang 2006). C1 inhibitor shows no association with activated MASP-3 (Zundel, Cseh et al. 2004). The *in vivo* MASP-3 enzymatic activity remains to be determined, although studies using synthetic substrates have yielded some avenues for further study.

Genomic sequence is available for human, mouse (Stover, Lynch et al. 2003) and rat showing an identical MASP1/3 locus structure (Fig.12), the entire locus consisting of 17 exons, the first 10 of which are shared by both MASP-1 and MASP-3 transcripts while exon 11 is MASP-3 specific and the remaining exons 12 to 17 are MASP-1 specific. The MASP-1/3 locus is present in all mammalian species where genomic sequence is available, notably avian species lack MASP-1. Genomic sequence data for *Gallus gallus* indicates that MASP-3, unpaired with MASP-1 is located on chromosome 6 (Lynch, Khan et al. 2005).

MASP-3 is the alternative splice product consisting of 11 exons in common with MASP-2 and the classical pathway MASP-like serine protease. MASP-3 serine protease domain is encoded within a single exon as is MASP-2, C1r and C1s. In common with the MASP-like serine protease, MASP-3 shares the AGY codon of the active site serine rather than TCN exhibited by MASP-1.

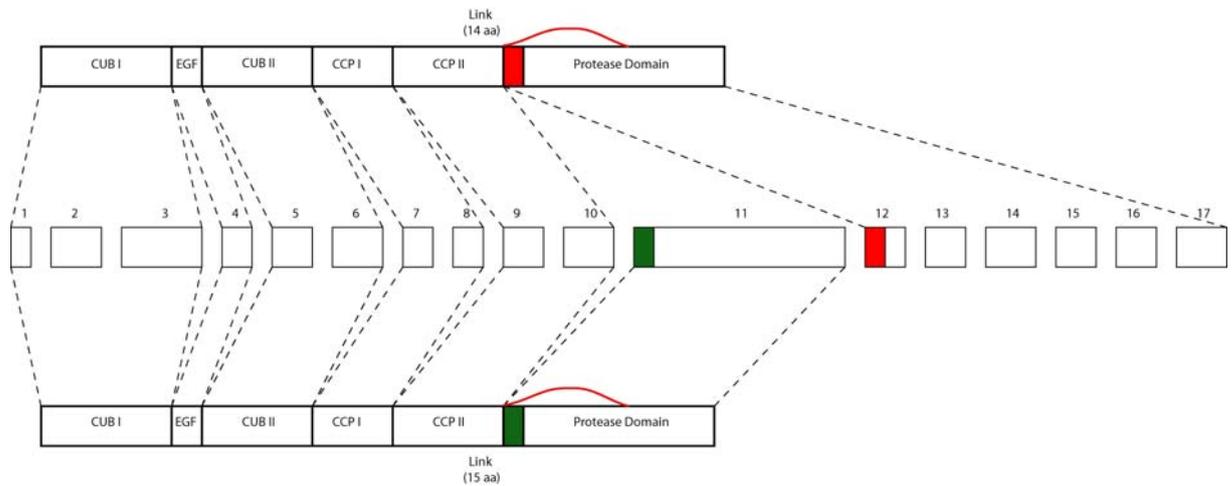


Fig.12 MASP-1/3 locus exon structure (centre panel), MASP-1 alternative splice transcript (upper panel) indicating protein domain structure. MASP-3 alternative splice transcript (lower panel) indicating protein domain structure. Highlighted red section indicated the MASP-1 link region, MASP-3 link region highlighted green. Red line represents the Disulphide Bridge between A and B-chain of the activated protein. CUB - Complement C1r/C1s, Uegf, Bmp1, EGF - Epidermal Growth Factor, CCP - Complement Control Protein.

MASP-3 has provided much scope for research, association with MBL, ficolin and the search for the enzymes substrate have formed the foundation for study. The cloning of homologues from diverse species and mapping of genomic loci have sought to provide insights into the evolution and potential function of MASP-3. The ancient origin of MASP-3 (Endo, Nonaka et al. 2003) suggests a conserved function for the protein, understanding of this role will require further investigation to identify a physiological substrate and the activating mechanism of MASP-3.

## 1.20 MASP-3 3-dimensional protein structure

Primary protein structure has been characterised for MASP-3 and for each of the MASP-like serine protease molecules, revealing a common domain configuration consisting of units present in a wide variety of protein molecules. Individual MASP-like protein domain characteristics are presented above; analysis of the interaction between domains and of tertiary protein structure enables new insights into molecular associations, conservation of amino acid residues and function.

Specific MASP-3 tertiary structural information is limited to the MASP-1/3 shared protein domains (Teillet, Gaboriaud et al. 2008); 3-dimensional structure regarding the MASP-3 serine protease domain is unavailable. The common modular domain architecture of MASP-1/3 gene products, sharing significant homology with other members of the MASP-like serine protease ensures there is a wealth of protein structural data on which to generate models of those yet experimentally undetermined structures.

Crystallographic data from the CUB<sub>1</sub>-EGF-CUB<sub>2</sub> domains of MASP-1/3 (Teillet, Gaboriaud et al. 2008) confirms structural studies of identical MASP-2 domains (Feinberg, Uitdehaag et al. 2003), defining associations between MASP-1/3 molecules, identifying calcium binding sites and residues important in MBL and ficolin interaction. MASP-1/3 molecules dimerise in a head-to-tail configuration, mediated by interaction between CUB<sub>1</sub> and EGF domains in respective partner molecules stabilized via Ca<sup>2+</sup> binding (Ca<sup>2+</sup> binding site I). The interaction between CUB<sub>1</sub> and EGF domain in association with Ca<sup>2+</sup> generates a rigid interface between domains. Dimer formation is identical to that exhibited by all MASP-like serine protease and Map19 (Thielens, Enrie et al. 1999; Feinberg, Uitdehaag et al. 2003; Gregory, Thielens et al. 2004), mediated by extended interactions at four distinct binding pockets and stabilized by hydrophobic interactions. The head-to-tail interaction of monomers produces a compact planar structure, exhibiting conservation of amino acids residues between molecules and species (Fig.13, Ca<sup>2+</sup> coordinating residues highlighted blue/**O**, dimer interaction residues highlighted in green and dark green).

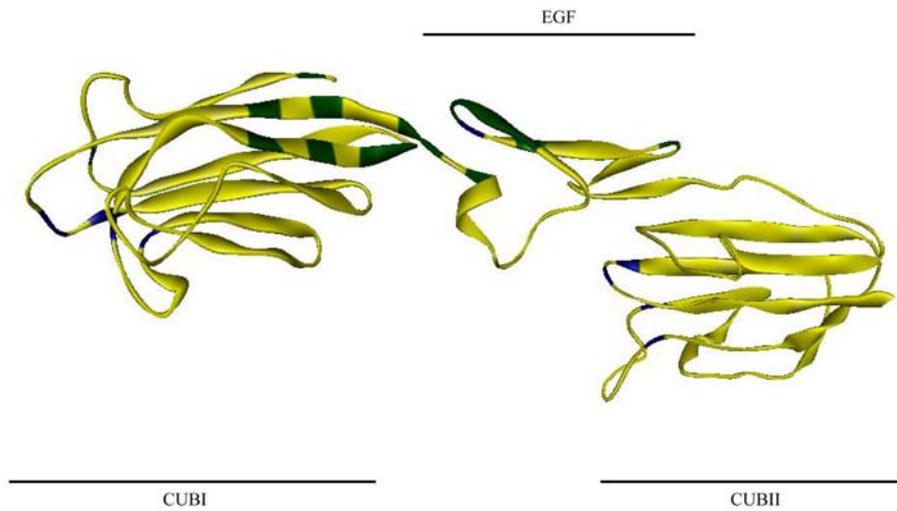


Fig.13 Crystal structure of human MASP-1/3 CUBI-EGF-CUBII domains. Highlighted Ca<sup>2+</sup> coordinating residues (blue), inter dimer interaction residues (green and dark green) representing residues of major and minor importance in binding. Respective domains are labelled.

Interaction between MASP-1/3, MBL and ficolin primarily occurs within the CUB<sub>1</sub> and CUB<sub>2</sub> domains, coordinated by Ca<sup>2+</sup> binding (Ca<sup>2+</sup> binding site II and III) to conserved amino acid side chains (Teillet, Gaboriaud et al. 2008). Mutational analysis efficiently mapped conserved amino acid residues whose side chains coordinate via carboxyl oxygen moieties to mediate Ca<sup>2+</sup> binding, facilitating interaction with a conserved lysine residue in the collagen region of MBL/ficolin (Teillet, Lacroix et al. 2007; Lacroix, Dumestre-Perard et al. 2009) (MBL binding residues highlighted in red and dark red, Ca<sup>2+</sup> coordinating residues highlighted in blue/○) (Fig.14).

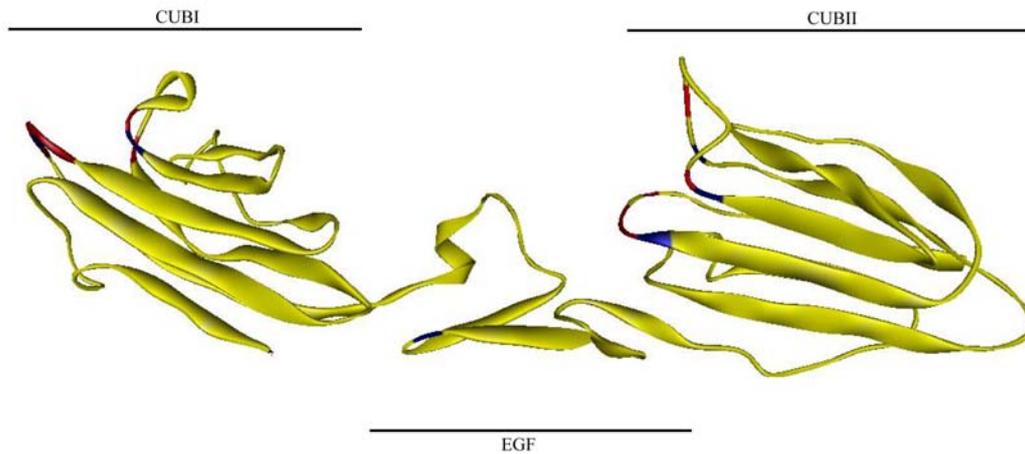


Fig.14 Structure of human MASP-1/3 CUBI-EGF-CUBII domains. Highlighted  $\text{Ca}^{2+}$  coordinating residues (blue), MBL binding residues highlighted in red and dark red, indicating major and minor residues involved in binding respectively. Respective domains are labelled.

Crystallographic determination of MASP-like serine protease N-terminal domain structures and their associations, between each other and to recognition molecules of the complement system, enabled stoichiometry of activation complexes to be modelled. As discussed above MASP-3 is able to form head-to-tail homodimers via interactions between the CUB<sub>1</sub>-EGF domains of each dimer pair, orientating catalytic serine protease domains at opposite ends of the complex.

Structural studies of MASP-1 including the CCP<sub>1</sub>-CCP<sub>2</sub>-SP domains (pdb 3gov) provides valuable information regarding the association between complement control protein domain and serine protease domains when considered in conjunction with previous studies of other MASP-like proteins (Dobo, Harmat et al. 2009). MASP-1 and MASP-like proteins conserve the same structural architecture, CCP domains extending in a rod like formation attached to a globular serine protease domain (Fig.15). In common with previously characterised CCP1-CCP2 structure there is a minor bend between the domains. There are conserved interactions between CCP2 and serine protease link region, pro-tyr-tyr and pro-vla-cyc (Dobo, Harmat et al. 2009)

respectively. These residues mediating the interaction exhibit conservation amongst species, as determined in this work (highlighted cyan, see Fig.15).

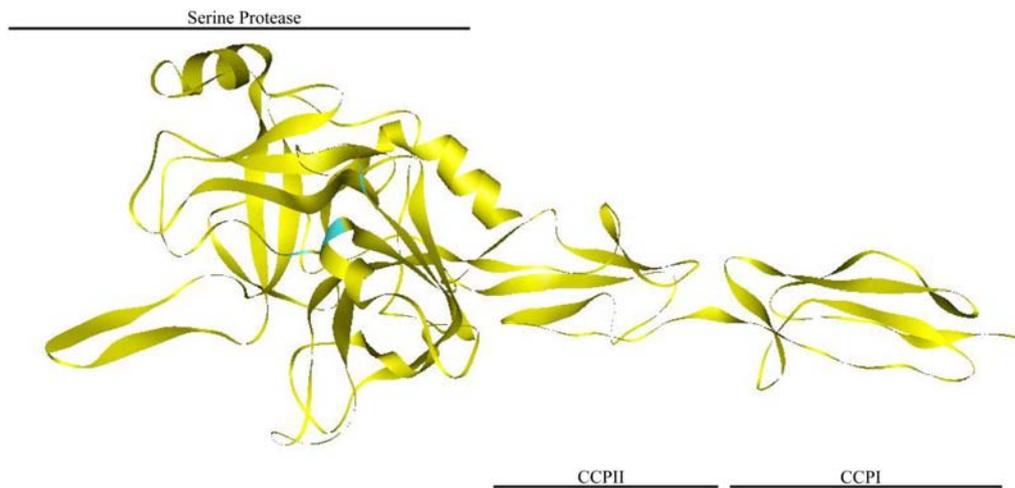


Fig.15 Model of rat MASP-3 CCPI-CCPII-Serine protease domain generated utilising homology to MASP-2 (pdb 1zjk). Highlighting in cyan the catalytic residues of the active site, respective domains are labelled.

Structural modelling of rat MASP-3 utilising MASP-2 (pdb 1zjk) as a template clearly highlights the coordination of catalytic site histidine, aspartic acid and serine residues (highlighted cyan).

Crystallographic studies in conjunction with determination of evolutionarily conserved residues, allows amino acid residues important in catalysis and protein-protein interactions to be viewed in context with respect to their position in the 3-dimensional protein structure. Utilisation of protein data in this way can provide useful information on the functional and binding properties of a protein, particularly important in the context of incomplete biochemical information regarding protein function.

## 1.21 Thesis aims

The primary aim of this thesis was to determine the existence in diverse species and extent of conservation of the MASP-3 transcript.

In doing this the following specific objectives were addressed:

- To examine MASP-3 sequence in diverse species and determine conserved amino acid residues vital to protein functional activity.
- To examine the structure of the MASP-1/3 locus in diverse species.
- To determine the evolutionary relationships amongst the MASP-like serine protease, with respect to MASP-3.
- To utilise sequence data derived from this work to determine differential expression of MASP-3 with respect to the alternatively spliced MASP-1.
- To determine the relative expression in tissue of alternatively spliced MASP-1 and MASP-3.
- To produce recombinant MASP-3 specific protein products to determine native protein expression and functional activity.

# Materials and Methods

## Chapter 2

## 2.1 Media, bacterial strains, chemicals and radioisotopes

### 2.1.1 Media

All media were obtained from Lab M, unless otherwise stated. Media were sterilised by autoclave at 121°C 15psi for 20 minutes.

#### Luria-Burtani (LB) broth

1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, were dissolved in 1 litre of dH<sub>2</sub>O.

#### Luria-Burtani (LB) agar

1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.5% (w/v) agar, were dissolved in 1 litre of dH<sub>2</sub>O.

#### SOB medium

2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>, were dissolved in 1 litre of dH<sub>2</sub>O.

#### SOB agar

2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>, and 1.5% (w/v) agar, were dissolved in 1 litre of dH<sub>2</sub>O.

#### SOC medium

2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>, 20mM glucose, were dissolved in 1 litre of dH<sub>2</sub>O.

#### NZY agar

1% (w/v) cassamino acids, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.2% (w/v) MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.5% (w/v) agar, were dissolved in 1 litre of dH<sub>2</sub>O.

Soft-top agarose

1% (w/v) cassamino acids, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.7% (w/v) agarose, were dissolved in 1 litre of dH<sub>2</sub>O.

Antibiotics were added to maintain plasmids where appropriate, at the following concentrations, Kanamycin at 25µg/ml, Ampicillin at 100µg/ml. Antibiotic stocks were dissolved in water, filter sterilised and stored at -20°C.

2.1.2 *Escherichia coli* strains

Top10F' (Invitrogen)

BL21 (DE3) pLYSs (Stratagene)

AD494 (DE3) pLYSs (Novagen)

XL-1 Blue MRF': Δ(mcrA) 183 Δ(mcrCB-hsdSMR-mrr) 173 endA1 supE44thi-1 recA1 gyrA96 relA1 lac[F' proAB lacIqZΔM15 Tn 10(Tetr)]

SOLR: e14-(McrA-) Δ(mcrCB-hsdSMR-mrr) 171 sbcC recB recJ uvrC umuC: Tn5 (Kanr) lac gyr A96 relA1 thi-1 endA1 λR [F' proAB lacIqZΔM15] Su- (nonsuppressing)

### 2.1.3 Chemicals and radioisotopes

All chemicals were obtained from the Sigma Chemical Company or from Fisher Chemicals. Radiochemicals, [ $\alpha^{32}\text{P}$ ] dCTP were purchased from Amersham Pharmacia biotech.

Chemicals (Fisher)

Fine chemicals (Sigma)

Radiochemicals (Amersham Pharmacia)

Restriction enzymes and buffers (Roche, Promega, Invitrogen, Gibco)

Culture media (Lab M)

pRSET prokaryotic expression system (Invitrogen)

Probond Affinity Resin (Invitrogen)

Mini Protean II (Bio-Rad)

Biologic HR FPLC system (Bio-Rad)

Biologic control software (Bio-Rad)

Sonicator (Hitachi)

## 2.2 DNA manipulation techniques

### 2.2.1 Polymerase chain reaction (PCR)

The polymerase chain reaction is used to amplify a segment of DNA bounded by two regions of known sequence (Mullis, Faloona et al. 1986), all enzymes and buffers were obtained from Promega unless otherwise stated. Oligonucleotide primers are designed to provide an initiation point for extension spanning the region to be amplified; primers were checked for self-complementarity and the potential to form hairpins and primer dimers. Two oligonucleotides; one is complementary to the antisense strand at the 5' end of the region to be amplified; one is complementary to the sense strand at the 3' end of the region to be amplified. Oligonucleotide primers provide a dNTP with a free –OH group as a substrate for the addition of subsequent nucleotides in a template dependent manner via the action of a DNA polymerase. A thermostable DNA polymerase (Taq DNA polymerase) is introduced to a  $Mg^{2+}$  containing buffered solution of template DNA, a molar excess of each primer, and dNTPs. The reaction mixture is initially denatured at 95°C, then cooled to a temperature that allows optimal primer annealing to their target sequences, a final temperature step allows efficient DNA polymerase activity. The three temperature transitions of denaturing, annealing and DNA synthesis are repeated many times under the control of a thermal cycler. Products of each cycle become templates in subsequent rounds of temperature cycling, theoretically doubling the amount of amplified DNA and leading to exponential growth of PCR product. Modifications of the oligonucleotide primers allow restriction sites, start codons, stop codons or point mutations to be engineered into the PCR product.

Primers are 18-25 nucleotides long (excluding modifications), have 45-60% G/C content, melting temperatures ( $T_m$ ) of 55-65°C and do not contain palindromic sequences. Primer pairs have approximately equal  $T_m$  and do not contain complementary sequences.  $T_m$  is calculated as 2°C per A or T base and 4°C per G or C base, excluding any mismatches or modifications. Computer programs Vector NTI and LaserGene were used to aid primer design, primers were checked for complementarity using a BLAST (<http://www.ncbi.nlm.nih.gov/PubMed>) search, identifying primers that may be expected to amplify homologous sequences.

A typical PCR reaction was assembled on ice using the following reagents

Forward primer (60pmoles/ $\mu$ l)	1 $\mu$ l
Reverse primer (60pmoles/ $\mu$ l)	1 $\mu$ l
10x PCR buffer (promega)	5 $\mu$ l
MgCl <sub>2</sub> (25mM)	2.5 $\mu$ l
dNTP's (10mM)	1.25 $\mu$ l

PCR amplification was done in a 50 $\mu$ l reaction volume containing: of each primer, 10nmoles of each dNTP, 1.5mM MgCl<sub>2</sub>, 20ng of DNA template and 2U of Taq DNA polymerase in the manufacturer's buffer diluted to 1x.

Thirty cycles of amplification were performed, each composed of 30 s at 95°C, 30 s at TA (temperature of annealing) and extension time at 72°C. The initial denaturation step was prolonged at to 2 min and the final extension step to 10 min. TA was typically 5°C below the calculated T<sub>m</sub>. PCR products were analysed by agarose gel electrophoresis.

The sequences of primers used in this work together with their calculated T<sub>m</sub> are listed in section appendices.

### 2.2.2 Agarose gel electrophoresis

DNA fragments were analysed by electrophoresis through horizontal agarose gels, which separated DNA molecules according to their size and/or topology (Fisher and Dingman 1971; Aaij and Borst 1972). Negatively charged at neutral pH, DNA is drawn towards the anode at a rate that decreases linearly with the logarithm of the molecules size. Ethidium bromide was included in the gel and the running buffer, the dye intercalates with dsDNA and can be visualised under UV illumination. This technique is sensitive enough to detect 10ng of DNA. Electrophoretic mobility decreases with increasing agarose concentration, so high concentrations (>2%) are appropriate for small DNA fragments (50-500 bp) and lower concentrations (0.6-1.5%) are used for large fragments (500-10,000 bp).

Agarose was mixed to the desired percentage in 1x TAE buffer (40mM Tris-acetate, 1mM EDTA, pH 8.0) before being melted and poured into the sealed casting chamber with a gel comb. Once set the gel comb was removed, the gel placed in the electrophoresis tank and covered with 1x TAE buffer before the samples were loaded. The gel was run at 2-5V/cm, monitored by migration of the loading dye and at a desired separation the gel run was stopped and the gel visualised under UV illumination and/or documented photographically.

### 2.2.3 Purification and recovery of DNA fragments from agarose gel

DNA fragments separated on Agarose gels were recovered using the Qiaex II Gel extraction kit (Qiagen), a uniform size silica matrix is used to bind DNA in high ionic strength buffer. DNA is then eluted from the silica matrix in dH<sub>2</sub>O or TE buffer (10mM Tris-HCl pH8.0, 1mM EDTA). A slice of agarose containing DNA is cut from the gel using a scalpel transferred to a 1.5ml microfuge tube and 3 volumes of buffer QX1 were added. 10µl of resuspended Qiaex II beads were added to the mixture which was heated to 50°C for 10 minutes, the tube was agitated throughout. The sample was centrifuged at 10000g for 1 minute, washed with 500µl buffer QX1 and twice with 500µl buffer PE which was then removed and the matrix pellet allowed to air dry. A buffer of low ionic strength (TE or dH<sub>2</sub>O) is used to elute the DNA fragment from the silica matrix;

centrifugation of the matrix allows the DNA to be recovered in the fluid phase. For most of the applications 2 $\mu$ g of DNA was purified and eluted in 20-30 $\mu$ l of TE.

#### 2.2.4 Ethanol precipitation of nucleic acid

Ethanol precipitation of nucleic acids allows the concentration and removal of contaminating protein components from a nucleic acid sample. To the sample is added  $\frac{1}{10}$  of a volume of 3M sodium acetate (pH 4.0) and 2.5 volumes of 100% ethanol. The sample was mixed prior to freezing in a dry ice/ethanol bath for 20 minutes, or a -80°C freezer for a minimum of 40 minutes. The precipitated nucleic acid sample was centrifuged at 15000g for 15 minutes, the supernatant was removed and the nucleic acid pellet retained. The pellet was washed with ice cold 70% ethanol, an allowed to air dry for 10 minutes. The recovered nucleic acid was resuspended in TE buffer or dH<sub>2</sub>O and stored at -20°C.

#### 2.2.5 Phenol:Chloroform extraction of nucleic acid

An equal volume of sample and phenol:chloroform:isoamylalcohol (25:24:1) were added, samples were vortexed briefly and centrifuged at 12,000g for 30 seconds. The upper aqueous phase was removed to a sterile microcentrifuge tube. The nucleic acid was recovered by ethanol precipitation.

### 2.2.6 Restriction endonuclease digest

Restriction digests of DNA molecules were used to confirm the presence of inserted DNA, prepare vectors and inserts for sub cloning and to generate restriction maps. All enzymes and buffers were obtained from Promega.

Restriction endonuclease digests were carried out in a volume of 20 $\mu$ l, each reaction containing 1 $\mu$ g DNA with 2 units of each restriction enzyme in the manufacturer's buffer diluted to a 1x concentration. The reaction was allowed to continue for 1-16 hours at the enzymes optimal reaction temperature according to the manufacturers instructions, before being analysed by agarose gel electrophoresis.

### 2.2.7 Ligation of DNA fragments

Ligation of foreign DNA molecules into plasmid vector DNA was accomplished using bacteriophage T4 ligase, catalysing the formation of the phosphodiester bonds between neighbouring 3' hydroxyl groups and 5' phosphate ends of double stranded DNA molecules. DNA fragment to be inserted and accepting plasmid vector DNA were restriction endonuclease digested with appropriate restriction enzymes, separated and analysed by agarose gel electrophoresis and purified using the Qiaex II Gel extraction kit (Qiagen). The concentrations of purified DNA solutions were estimated by ethidium bromide stained agarose gel electrophoresis.

Ligation reactions were performed in a 10 $\mu$ l reaction volume containing: 50-150ng of vector DNA, a three-fold molar excess of insert DNA and 4-6 Weiss units of T4 DNA ligase and 1 $\mu$ l of 10x ligation buffer (300mM Tris-HCl (pH7.8), 100mM MgCl<sub>2</sub>, 100mM DTT, 10mM ATP). The reaction volume was made to 10 $\mu$ l with sterile dH<sub>2</sub>O. Reactions were incubated overnight at 15°C, then transformed into *E. coli* or stored at -20°C.

### 2.2.8 Preparation of chemically competent *E.coli*

The following procedure, which was developed by (Hanahan 1983), can yield competent cultures of *E. coli* strains DH1, DH5 $\alpha$ , and TOP10F' that can be transformed at frequencies  $>5 \times 10^8$  transformed colonies per microgram of supercoiled plasmid DNA. The maximum frequency of transformation that can be obtained routinely with most other strains of *E. coli* is approximately five- to tenfold lower. Using a sterile platinum wire the required *E. coli* strain was taken directly from the frozen stock and streaked onto the surface of a SOB agar (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>, and 1.5% (w/v) agar) plate. The plate was incubated for 16 hours at 37°C to allow colonies to develop. One well-isolated colony was transferred into 1 ml of SOB containing 20mM MgSO<sub>4</sub> using a sterile platinum wire. The bacteria were dispersed by vortexing at moderate speed and then diluted in 30-100 ml of SOB (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>) containing 20mM MgSO<sub>4</sub> in a 1-litre flask. The bacterial culture was incubated for 2.5-3.0 hours at 37°C until the spectrographic absorption at OD<sub>550</sub> is 0.7 to 0.8, at which time the cells were harvested by centrifugation at 2000g for 10 min and the supernatant discarded. The pellet was resuspended in 30 ml sterile ice cold TfbI (30mM potassium acetate, 50mM MnCl<sub>2</sub>, 100mM KCl, 10mM CaCl<sub>2</sub>, 15% (v/v) glycerol) and incubated on ice for 5 - 30 min, depending on the strain chosen. 10 minute incubation time was used for XL1-blue and TOP10F'. The bacterial suspension was centrifuged at 2000g for 10 minutes at 4°C the supernatant discarded and the bacterial pellet carefully resuspended in 4ml TfbII (10mM Na-MOPS, 75mM CaCl<sub>2</sub>, 10mM KCl, 15% (v/v) glycerol). 0.2ml cell aliquots were prepared and either transformed immediately or rapidly frozen in a dry ice/ethanol bath and stored at -80°C.

### 2.2.9 Transformation of chemically competent *E.coli*

Competent cells were thawed slowly on ice. 2-10 $\mu$ l of a ligation reaction, or 0.1ng of plasmid DNA, and 2 $\mu$ l of 0.5 M  $\beta$ -mercaptoethanol (optional) were added to 100 $\mu$ l of competent cells and incubated on ice for 20 min. The cell and plasmid mix were heat shocked at 37°C for 5 min or at 42°C for 1 min, then return to ice for a further 2 min.

950µl of SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>, 20mM glucose) preheated to 37°C was added and the culture incubated at 37°C in a shaking incubator for 45-60 min to facilitate cell recovery. Aliquots (50µl - 150µl) of transformed cell culture were plated on to LB agar (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.5% Agar) plates with appropriate selective antibiotics and incubated overnight at 37°C.

#### 2.2.10 Small-scale plasmid DNA preparation

2 ml LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) supplemented with appropriate antibiotics was inoculated with a single plasmid-containing *E.coli* colony and grown overnight at 37°C in a shaking incubator. 1ml of the overnight culture was spun in a microcentrifuge at 14,000 x g: for 2 min, and the bacterial pellet resuspended in 100 µl Buffer P1-RNase (50mM glucose, 25mM Tris-Cl, 10mM EDTA, 10µg/ml RNase A, pH 8.0). 200 µl Buffer P2 (0.2M NaOH, 1% SDS) was added to lyse the cells. The mixture was neutralized by the addition of 150 µl Buffer P3 (3M KOH, 5M acetic acid, pH 5.2) mixed gently and incubated on ice for 10 min. The precipitate was centrifuged at 14,000 x g: for 10 min. 400 µl of the supernatant was added to 800 µl ethanol in a fresh tube and incubated at -20°C for 20 min. The precipitate of nucleic acid was centrifuged at 14,000 x g for 10 min, the supernatant discarded. The pellet was washed in ice cold 70% ethanol centrifuged as before and then air dried before being resuspended in a volume of dH<sub>2</sub>O or TE buffer (10mM Tris-Cl, 1mM EDTA, pH 7.5).

#### 2.2.11 High quality small-scale plasmid DNA purification

DNA for storage and sequencing was prepared using a high ionic strength DNA binding matrix in spin column format (Qiagen plasmid mini kit), yielding between 2-5µg of high quality plasmid DNA.

1.5ml (high-copy-number plasmid) of bacterial culture was harvested by centrifugation at 10,000 x g for 5 minutes, the supernatant was discarded. 250µl of Buffer P1-RNase (50mM glucose, 25mM Tris-Cl, 10mM EDTA, 10µg/ml RNase A, pH 8.0) was added

and the bacterial pellet completely resuspended. 250µl of Buffer P2 (0.2M NaOH, 1% SDS) was added and mixed by inverting the tube 4 times; incubate for 1-5 minutes until the cell suspension clears. Add 350µl of Buffer P3 (3M KOH, 5M acetic acid, pH 5.2) and immediately mix by inverting the tube 4 times, the bacterial lysate was centrifuged at  $14,000 \times g$  for 10 minutes to pellet bacterial debris. The Qiagen tip column was prepared by equilibration with 1ml of buffer QBT (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol (v/v), 0.15% Triton® X-100 (v/v)). The cleared lysate (c.850µl) was recovered and applied to a prepared Qiagen tip column, after centrifugation at  $10,000 \times g$  for 1 minute the flow through was discarded. 750µl of Column Wash Solution was added to the column, centrifuged at  $10,000 \times g$  for 1 minute. Washing of the plasmid DNA bound to the column was repeated a further 3 times with 250µl buffer QC (1.0 M NaCl, 50mM MOPS, pH 7.0, 15% isopropanol (v/v)), a final centrifugation at  $10,000 \times g$  for 2 minutes removes traces of wash solution. Bound plasmid DNA was eluted from the spin column with 100µl dH<sub>2</sub>O, centrifuged at  $10,000 \times g$  for 1 minute captured in a sterile 1.5ml microcentrifuge tube.

### 2.2.12 Large scale plasmid preparation

A single isolated colony was used to inoculate 25-50ml of LB medium with appropriate antibiotic selection, incubate for 12-16 hours at 37°C with shaking. Bacterial cells were harvested by centrifugation at  $6000 \times g$  for 15 minutes at 4°C, the supernatant was discarded. The bacterial pellet was completely resuspended in 4ml of buffer P1-RNase (50mM glucose, 25mM Tris-Cl, 10mM EDTA, 10µg/ml RNase A, pH 8.0), 4ml of buffer P2 (0.2M NaOH, 1% SDS) was added and mixed by inverting the tube 4-6 times. Incubate at room temperature for 5 minutes, immediately add 4ml of buffer P3 (3M KOH, 5M acetic acid, pH 5.2) and mix by inversion 4-6 times. Incubate on ice for 15 minutes. Centrifuge at  $20,000 \times g$  for 30 minutes at 4°C, recover the plasmid DNA containing supernatant. Apply the supernatant to an anion exchange column (Qiagen-tip 100) equilibrated with 10ml QBT buffer (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol (v/v), 0.15% Triton® X-100 (v/v)), the supernatant was allowed to flow through the column. The column retaining plasmid DNA was washed twice with 10ml of buffer QC (1.0 M NaCl, 50mM MOPS, pH 7.0, 15% isopropanol (v/v)), before being

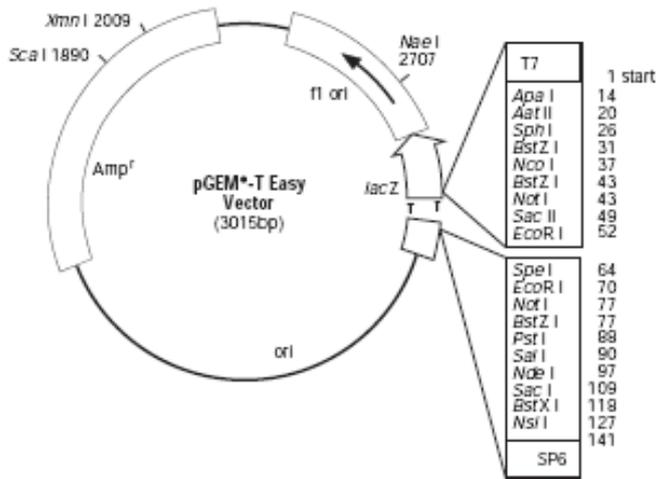
eluted into a sterile tube with 5ml of buffer QF (1.25 M NaCl, 50mM Tris-Cl, pH 8.5, 15% isopropanol (v/v)). Plasmid DNA was precipitated by addition of 3.5ml isopropanol, mix and centrifuge at 15,000 x g for 30 minute. The supernatant was discarded carefully not to disturb the plasmid DNA pellet, 2ml of 70% ethanol was added and centrifuged at 15,000 x g for 10 minutes to wash the pellet. The supernatant was carefully aspirated and the plasmid DNA pellet was allowed to air dry for 5-10 minutes. The Qiagen 100 kit was used to prepare up to 100µg of ultrapure plasmid DNA, the final DNA pellet being resuspended in 100µl of dH<sub>2</sub>O for storage at -20°C.

### 2.2.13 DNA sequencing

Sequencing of plasmid inserts was carried out by PNACL (Leicester) using universal (T7 or sp6) or specifically designed primers. Isolated plasmids were prepared using the Qiagen plasmid mini kit and supplied to PNACL (Leicester) to be sequenced. Sequence data was returned in electronic format as two files, a text file containing sequence data as determined by ABI prism sequence analysis software and a chromatogram of the original un-interpreted sequence data.

### 2.2.14 Cloning of PCR amplification product

Cloning of PCR products generated by Taq thermostable DNA polymerase can be efficiently achieved, utilising the enzymatic property to add single deoxyadenosine residues to the 3'-ends of PCR products (Clark 1988; Hu 1993). T-overhang vectors such as Promega's pGEM<sup>®</sup>-T Vectors are linearised vectors with 3'-T overhangs complementary to the A-overhangs on the PCR product generated by the thermostable Taq polymerase.



T7 RNA polymerase transcription initiation site	1
multiple cloning region	10-128
SP6 RNA polymerase promoter (-17 to +3)	139-158
SP6 RNA polymerase transcription initiation site	141
pUC/M13 Reverse Sequencing Primer binding site	176-197
<i>lacZ</i> start codon	180
<i>lac</i> operator	200-216
$\beta$ -lactamase coding region	1337-2197
phage <i>f1</i> region	2380-2835
<i>lac</i> operon sequences	2836-2996, 166-395
pUC/M13 Forward Sequencing Primer binding site	2949-2972
T7 RNA polymerase promoter (-17 to +3)	2999-3

Fig.16 pGEM-T Easy circle map and sequence reference points.

The high copy number pGEM<sup>®</sup>-T Easy Vector contain T7 and SP6 RNA Polymerase promoters flanking a multiple cloning site (MCS) within the alpha-peptide coding region of the enzyme beta-galactosidase. Insertional inactivation of the alpha peptide allows recombinant clones to be directly identified by blue/white colour screening on indicator plates, containing IPTG (isopropyl-beta-D-thiogalactopyranoside) as a promoter inducer and X-Gal (5-brom-4-chloro-3-indolyl-beta-D-galactopyranoside) indicator.

PCR products were either directly ligated or processed to remove excess nucleotides prior to being added to a ligation reaction. Low temperature ligation favours the desired annealing of the A-overhang of the insert with the T-overhang of the vector; reactions were allowed to proceed overnight at 4°C. An aliquot of the ligation reaction was used to transform chemically competent *E.coli*, after appropriate selection and plasmid preparation; restriction endonuclease digestion was used to confirm a cloned insert of expected size. The pGEM<sup>®</sup>-T Easy vector multiple cloning site (MCS) is flanked by the restriction enzymes *NotI* or *EcoRI*, either of these may be used in a single digest to release a cloned insert.

### 2.2.15 Random primed DNA labelling

Labelling of DNA fragments for use as a specific probe was carried out using the random primer method, utilising the Random primers DNA labelling system (Invitrogen). The DNA fragment to be labelled was heat denatured at 100°C for 10min, before being rapidly cooled on ice to prevent reannealing of complementary DNA strands. A random decamer primer was added to 100ng of denatured template providing a point of synthesis for the Klenow fragment of DNA polymerase I, which incorporates nucleotides in a 3'-5' direction. The labelling reaction was allowed to take place in the presence of dATP, dGTP, dTTP and radiolabeled P<sup>32</sup> dCTP, extending the random decamer primer in a template dependent fashion under the action of the Klenow fragment of DNA polymerase I. The reaction was allowed to proceed for 12 hours at room temperature or 2 hours at 37°C, unincorporated nucleotides were removed using a Microspin G-50 size exclusion column (Amersham biosciences).

### 2.2.16 Southern blot

Agarose gel electrophoresis of DNA fragments was carried out as described, after documentation of the gel including a ruler as a size marker the gel was treated to produce single stranded DNA within the gel before being transferred to a nitrocellulose membrane. The gel was first placed in denaturing solution (1.5M NaCl, 0.5M NaOH) for 2 x 15 minutes, this solution decanted and replaced with neutralising solution (1.5M NaCl, 0.5M Tris-Cl, pH 7.0) again for 2 x 15 minutes. Capillary action was used to transfer the now single stranded DNA to a nitrocellulose support, 20x SSC (3M NaCl, 300mM Sodium citrate pH 7.0) being used to carry the DNA from the gel to the membrane as it flowed from a reservoir beneath the gel through the membrane towards dry filter paper on top.

Transfer of denatured DNA from the gel to the nitrocellulose is allowed to proceed for a minimum of 3 hours to a maximum of 12 hours. The nitrocellulose membrane was marked on the non-DNA side to orientate sample lanes; the DNA was fixed to the nitrocellulose membrane by exposure to UV illumination for 1 minute.

Prehybridisation of the membrane was carried out at 42°C in Southern blot hybridisation buffer (0.9M NaCl, 90mM Sodium citrate pH 7.0, 5x Denhardt's solution (0.1% (w/v) Polyvinylpyrrolidone, 0.1% (w/v) Ficoll (400 kDa), 0.1% (w/v) BSA (fraction 5) in dH<sub>2</sub>O), heat denatured salmon sperm DNA to 100-200µg/ml, and 0.5% SDS) for 3-4 hours. The membrane was hybridised overnight with radioactively labelled probe at 42°C in fresh Southern blot hybridisation buffer with 7.5% (w/v) dextran sulphate, in a rotating cylinder air incubator. The hybridised membranes were washed twice each for 10 minutes in 2x SSC (0.3M NaCl, 30mM Sodium citrate pH 7.0) at room temperature, then for 30 minutes in 1x SSC (0.15M NaCl, 15mM Sodium citrate pH 7.0) at 65°C. Washed membranes were dried and exposed to autoradiography film in a film cassette with intensifying screens for 1-12 hours at -80°C. The photographic film was processed using an automatic developer (Gevamatic 60, AGFA-Gevaert Ltd).

### 2.2.17 Autoradiography

Detection of radioactively labelled probes the membrane was wrapped in Clingfilm/Saran Wrap and exposed to autoradiography film (Hyperfilm MP, Amersham Pharmacia biotech) in an autoradiography cassette overnight at -80°C. Photographic film was developed using an automatic developer (Gevamatic 60, AGFA-Gevaert Ltd) according to manufacturer's instructions.

### 2.2.18 Colony blots

Colony blotting allows the screening of transformed colonies for a specific DNA sequence. Transformed colonies on an agar plate are directly transferred/replicated on to a nitrocellulose membrane; DNA is liberated from the colonies, denatured and then fixed to the membrane. Remaining DNA binding sites are blocked with BSA blocking buffer (10mM Tris-Cl pH7.5, 2mM EDTA, 150mM NaCl, 5% (w/v) BSA, 0.05% (v/v) Tween 20) prior to hybridising a specific <sup>32</sup>P radiolabeled probe. The blots were

developed and realigned to the original agar plates, enabling colonies containing the desired DNA sequence to be identified and propagated.

Agar plates for colony blots were pre-cooled for 30 minutes at 4°C; a disc of nitrocellulose membrane was placed on top of the bacterial colonies. The position of the disc was marked by poster paint pushed through the membrane and into the agar using a needle; this enables reorientation of the membrane after hybridisation and developing. The membrane was allowed to adhere to the colonies for 1-2 minutes, before being removed in a continuous movement and placed colony side up on a sheet of 3MM filter paper. The membrane was then transferred on to a series of 3MM paper filters saturated with solution; 1 minute 10% (w/v) SDS, to lyse cells, 5 minutes on denaturing solution (1.5M NaCl, 0.5M NaOH), 5 minutes neutralising solution (1.5M NaCl, 0.5M Tris-Cl, pH 7.0), finally the membrane was washed in 2xSSC (0.3M NaCl, 30mM Sodium citrate pH 7.0). The membrane disc was allowed to air dry, before the DNA was fixed to the membrane either by baking at 80°C for 2 hours or by UV crosslinking for 1 minute. Membranes were probed as described for Southern blot procedures.

### 2.2.19 cDNA phage library screening

The phage library to be screened was diluted in  $\lambda$  dilution buffer/SM buffer (100mM NaCl, 8mM MgSO<sup>4</sup>, 0.01% gelatine, 50mM Tris-Cl, pH 7.5), test plates using a range of phage library dilutions were made to determine pfu/ $\mu$ l. A representative phage library screen aimed to examine 5-10x10<sup>5</sup> pfu.

1-5x10<sup>4</sup> pfu of the phage library was adsorbed with 0.6ml XL-1 Blue (OD<sub>600</sub> 0.5) and 0.3ml  $\lambda$  dilution buffer/SM buffer for 15 minutes at 37°C. 7ml of soft-top agarose (1% (w/v) cassamino acids, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.7% (w/v) agarose) was then added, mixed and poured onto a pre warmed 150mm NZY agar (1% (w/v) cassamino acids, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.2% (w/v) MgSO<sup>4</sup>.7H<sub>2</sub>O, 1.5% (w/v) agar) plate. Once set the plates were incubated at 42°C for a period of 5 hours, allowing phage plaques to form. 132 mm nitrocellulose filters were placed onto

the cooled plates surface avoiding the creation of air bubbles; the membrane was allowed to adsorb some of the phage material for 2-5 min before each filter was marked to identify and orientate it to its respective plate and then removed. Filters were sequentially soaked in denaturing solution (1.5M NaCl, 0.5M NaOH), neutralising solution (1.5M NaCl, 0.5M Tris-Cl, pH 7.0) and 20x SSC (3M NaCl, 300mM Sodium citrate pH 7.0), for 1 min, 8 min and 15 min respectively. The filters were allowed to air dry and then baked at 80°C for 2 hours, fixing the DNA to the filter. The remaining DNA binding capacity of the filter was occupied using a prehybridisation solution (7% (w/v) SDS, 1% (w/v) BSA, 1mM EDTA, 250mM sodium phosphate, pH 7.5) incubated overnight with shaking at 65°C.

The fixed and blocked filters were exposed to the desired denatured, <sup>32</sup>P labelled DNA probe in hybridisation buffer (denatured probe, 7% (w/v) SDS, 1% (w/v) BSA, 1mM EDTA, 250mM sodium phosphate, pH7.5) at 65°C overnight with shaking. The probe was removed and saved for rescreening, the filters were washed (4x15min) in washing buffer (300mM NaCl, 0.03mM sodium citrate, 0.1% (w/v) SDS) at 65°C, air dried and fixed to Whatman paper (3mm chromatography paper). Filters were protected with plastic wrap before being exposed to x-ray film overnight at -80°C, in a cassette with intensifying screens. X-ray film was processed in using an automated film processor.

Positive filters were realigned to their respective plates and the region of a positive phage plaque was excised and allowed to elute from its soft-top agarose support overnight in 1ml of  $\lambda$  dilution buffer/SM buffer at 4°C. The enriched phage stock was then replated to 0.5-1x10<sup>3</sup> pfu per plate and the procedure repeated, a further enrichment was used to isolate single positive phage plaque.

### 2.2.20 In vivo excision of single positive phage

The in vivo excision process removes a large proportion of the phage genome from the cloned cDNA, leaving a circular pBluescript-cDNA plasmid molecule.

Two bacterial strains XL-1 Blue and SOLR were grown overnight at 30°C in a shaking incubator. XL-1 Blue were pelleted and resuspended to an OD<sub>600</sub> of 1.0 in sterile

10mM MgSO<sub>4</sub>, before 200µl were combined with 100µl of single positive phage stock and 1µl ExAssist helper phage. After incubation at 37°C for 15 minutes, 3ml of LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) was added before the culture was allowed to grow at 37°C for 2.5 hours. This procedure generated a stock of pBluescript-cDNA phagemids that are unable to infect XL-1 Blue. pBluescript-cDNA molecules were introduced into SOLR *E.coli* by mixing 5µl of phagemid stock with 200µl SOLR cells prior to 15 minutes incubation at 37°C, aliquots were spread onto LB agar (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.5% (w/v) agar) plates containing 100µg/ml ampicillin and incubated overnight at 37°C. Plasmids DNA was prepared and analysed as previously described.

### 2.2.21 PCR recovery of positive cDNA from phage vector

Recovery of positive cDNA phage integrates was achieved using T3 and T7 primers flanking the cloning site, standard PCR techniques were applied. Single positive phage clones were isolated and excised from the soft-top agarose, phage was allowed to elute in a 1.5ml microcentrifuge tube into sterile dH<sub>2</sub>O. PCR was carried out directly on samples of eluted phage solution, reactions were analysed by agarose gel electrophoresis.

## 2.3 RNA techniques

### 2.3.1 RNA manipulation and analysis

To maximise the integrity and minimise RNase contamination of RNA material several general precautions were taken, RNA work space was physically separated from other laboratory work, dedicated reagents, filter tips and DEPC treated disposable plasticware were used throughout.

### 2.3.2 Diethyl pyrocarbonate (DEPC) treatment of solutions and plastic ware

Diethyl pyrocarbonate forms histidine residue derivatives and is therefore an effective method to inactivate nucleases including RNase. DEPC was prepared as a 1:100 dilution in absolute ethanol, diluted to a working concentration of 0.1% (v/v) prior to use. DEPC solution (0.1% v/v) was heated for 1 hour at 65°C prior to autoclave sterilisation to prepare it for use. Material to be treated was submerged in a working solution (0.1% v/v) of DEPC and incubated at 65°C for 1 hour, after treatment materials were autoclaved prior to use.

### 2.3.3 Total RNA preparation

Total RNA isolation from cells and tissue was achieved using TRIzol reagent (Invitrogen), based on the Chomczynski method (Chomczynski and Sacchi 1987). TRIzol maintain the integrity of RNA whilst disrupting and denaturing protein and cell components. 1ml of TRIzol reagent was added for each 50-100mg homogenised tissue or  $5 \times 10^6$  cells in an RNase free microfuge tube, and incubated at room temperature for 5 minutes. 0.2ml chloroform was added per 1ml TRIzol reagent, agitated and incubated at room temperature for 5 minutes. The microfuge tube was centrifuged at 12,000g for 2-3 minutes at 2°C-8°C; the upper aqueous phase containing RNA was removed to a fresh micro centrifuge tube. RNA was precipitated by adding 0.5ml isopropyl alcohol

per 1ml TRIzol, the mixture was incubated at room temperature for 15 minutes and centrifuged at 12,000g for 10 minutes at 2°C-8°C. The supernatant was discarded. The RNA pellet was wash with 70% ethanol, centrifuge at 12,000g for 2-3 minutes at 2°C-8°C. The precipitated RNA pellet was allowed to air dry prior to resuspension in RNase free dH<sub>2</sub>O.

### 2.3.4 DNase treatment of RNA

High quality total RNA was prepared as previously described using the TRIzol reagent; RQ1 RNase-free DNase (promega) was used to degrade contaminating DNA whilst maintaining the integrity of the RNA. The following reagents were added to a microfuge tube on ice

RNA in water or TE buffer	1–8µl
RQ1 RNase-Free DNase 10X Reaction Buffer	1µl
RQ1 RNase-Free DNase	1u/µg RNA
Nuclease-free water	final reaction volume of 10µl

Samples were incubated at 37°C for 30 minutes; the reaction was terminated by incubation at 65°C for 10 minutes to inactivate the DNase enzyme. Phenol:Chloroform extraction was used to remove inactivated enzyme.

### 2.3.5 Northern blot

Northern blot analysis is the transfer of total RNA from an electrophoresis gel, to a supported membrane where it is immobilised. RNA transcripts of interest can be detected on the membrane by hybridisation with a specific RNA or DNA probe.

Total RNA was isolated from tissue or cells using the TRIzol method, RNA handling precautions were used throughout minimise potential degradation or contamination.

### 2.3.6 Northern blot denaturing gel electrophoresis RNA sample preparation

RNA samples were prepared for electrophoresis by the addition of 4 $\mu$ l 1x MOPS (20mM MOPS, 1mM EDTA, 5mM sodium acetate, pH7.0), 3.5 $\mu$ l (12.3M) formaldehyde, and 8 $\mu$ l formamide to a total of 5 $\mu$ l RNA. Samples were denatured at 56°C for 15 minutes, placed on ice and 2 $\mu$ l of loading dye added (10mM EDTA, pH 8.0, 97.5% deionised formamide, 0.3% bromophenol blue).

### 2.3.7 Northern blot denaturing gel electrophoresis

A 1% agarose gel was prepared with DEPC treated water, 0.6% (v/v) formaldehyde and 10x MOPS buffer (200mM MOPS, 10mM EDTA, 50mM sodium acetate, pH7.0) to a final 1x concentration. The gel was cast in a fume hood and allowed to set for 1 hour. The gel was submerged in 1x MOPS buffer in an RNase free electrophoresis tank and pre run for 10 minutes at 60V. Prepared RNA samples and RNA molecular weight markers (RNA markers included ethidium bromide at 40 $\mu$ g/ml for visualisation) (Promega) were loaded into individual sample wells and electrophoresis continued at 60V for 2-3 hours until the bromophenol blue dye front migrated  $\frac{2}{3}$  -  $\frac{3}{4}$  of the way along the gel. The gel was removed and photographed alongside a ruler under UV transillumination to facilitate RNA transcript size determination.

### 2.3.8 Transfer of RNA to membrane

The northern blot gel was soaked in 200 ml RNase-free water for 10-15 min and then for 15 min in 50mM NaOH, and neutralised by soaking in 10x SSPE buffer (1.8M NaCl, 0.1M NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O, 0.001M EDTA, pH 7.4) for 30 min. RNA was transferred overnight to a nylon membrane, using capillary action in 10x SSPE buffer as described by Sambrook (Sambrook, Fritsch et al. 1989). The membrane was labelled and the position of the wells marked, rinsed in 4x SSPE (0.72M NaCl, 0.04M NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O, 0.4mM EDTA, pH 7.4) and the RNA immobilised to the membrane by UV crosslinking.

### 2.3.9 Hybridisation and Washing

Prehybridisation of the membrane was carried out at 42°C in Northern blot hybridisation buffer (50% deionised formamide, 5x SSPE (0.9M NaCl, 0.05M NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O, 0.5mM EDTA, pH 7.4), 50mM sodium phosphate buffer, pH 6.8, 1x Denhardt's solution (0.02% (w/v) Polyvinylpyrrolidone, 0.02% (w/v) Ficoll (400 kDa), 0.02% (w/v) BSA (fraction 5) in dH<sub>2</sub>O), heat denatured salmon sperm DNA to 100-200µg/ml, and 0.5% SDS) for 3-4 hours. The membrane was hybridised overnight with radioactively labelled probe at 42°C in fresh Northern blot hybridisation buffer with 7.5% (w/v) dextran sulphate, in a rotating cylinder air incubator. The hybridised membranes were washed twice each for 10 minutes in 2x SSPE (0.36M NaCl, 0.02M NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O, 0.2mM EDTA, pH 7.4, 0.1% (w/v) SDS) at room temperature, then for 30 minutes in 0.5x SSPE (0.09M NaCl, 5mM NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O, 0.05mM EDTA, pH 7.4, 0.1% (w/v) SDS) at room temperature. Washed membranes were dried and exposed to autoradiography film in a film cassette with intensifying screens for 1 week at -80C. The photographic film was processed using an automatic developer (Gevamatic 60, AGFA-Gevaert Ltd) according to the manufacturer's instructions.

### 2.3.10 Reverse transcriptase PCR (RT-PCR)

The reverse transcriptase enzyme is used to copy an RNA transcript into a complementary DNA product, allowing it to be used as a template in a standard PCR reaction. First strand cDNA synthesis was performed using the Superscript II first strand synthesis system from Invitrogen. Amplification of specific regions of cDNA can be used to identify, quantify and characterise transcribed mRNA in a particular sample. Specific primer pairs are utilised to amplify cDNA sequence via PCR, using a first strand cDNA template. Primers were designed utilising either Vector NTI or Lasergene computer software, where possible the region to be amplified spanned an exon/intron boundary, identifying and preventing aberrant amplification of contaminating genomic DNA. Primers were checked for complementarity to currently characterised sequence using a BLAST (<http://www.ncbi.nlm.nih.gov/PubMed>) search, identifying primers that may be expected to amplify homologous sequences.

RNA (0.1µg-5µg of total RNA or 50ng-500ng poly-A RNA) prepared by the TRIzol method was used as a template in the first strand cDNA synthesis reaction. The RNA template was mixed with an appropriate oligonucleotide primer (a random hexamer primer, an oligo-dT oligonucleotide primer or a gene specific primer), and dNTP's in a PCR tube. The following reagents were assembled on ice

RNA	0.1µg-5µg
Oligo-dT primer (0.5/µl)/hexamer/GSP	1µl
dNTP's (10mM)	1µl
Nuclease free dH <sub>2</sub> O	up to 10µl

Samples were incubated at 65°C for 5 minutes to denature RNA secondary structure, and then rapidly cooled on ice to anneal the chosen primer to the RNA template. Extension of the annealed primer was mediated by Superscript II reverse transcriptase enzyme (Invitrogen) in the presence of 1x RT buffer, DTT, Mg<sup>2+</sup> ions and an RNase inhibitor. The following components were added to a final volume of 20µl:

10x RT buffer	2µl
MgCl <sub>2</sub> (25mM)	4µl
DTT (0.1M)	2µl
RNase inhibitor	1µl

Reaction components were then mixed and incubated at 42°C for 2 minutes, prior to the addition of 1µl (50 units) superscript II RT enzyme, the reaction was continued for 50min at 42°C before being terminated by incubation at 70°C for 15min. A no RT control, with the addition of dH<sub>2</sub>O in place of reverse transcriptase enzyme, and a no RNA control, with the addition dH<sub>2</sub>O in place of RNA, was carried out simultaneously for all samples analysed.

Following first strand cDNA synthesis, 1µl of cDNA was used in a standard PCR reaction. RT-PCR products were analysed by agarose gel electrophoresis.

### 2.3.11 Quantitative RT-PCR

The Roche Lightcycler (LC) instrument and the Roche fast start DNA master SYBR Green I kit were used for all quantitative RT-PCR experiments. The Roche Lightcycler instrument allows rapid PCR cycling, continuous monitoring and analysis during amplification via fluorescence detection of SYBR Green dye DNA complexes. Post-amplification melting curve analysis allows product characterisation and mutation detection.

PCR primers were designed to specifically amplify the desired sequence, producing a product of between 200-350bp in length. Amplification primers were designed where possible to span an intron/exon boundary, eliminating the detection of contaminating genomic DNA remaining from cDNA synthesis. Standard curves were prepared for each gene product analysed, dilutions of cloned plasmids containing target cDNA's at a defined concentration. Lightcycler quantification software referenced results from the standard curve samples to calculate the numbers of target molecules present in the samples under investigation.

Template cDNA for was prepared from 2µg total RNA isolated by the TRIzol method, using SuperScript first strand synthesis system for RT-PCR. Template cDNA was used directly in the real-time PCR reaction; fluorescent monitoring of amplification and melting curve analysis was controlled using the Lightcycler control software.

The following components were assembled in a Lightcycler capillary tube on ice

cDNA template	0.2µl
SYBR Green Mix (2x)	12.5µl
Forward primer (5pmol/ml)	0.5µl
Reverse primer (5pmol/ml)	0.5µl
H <sub>2</sub> O	11.3µl
Total 25µl reaction volume	

The following PCR conditions were used

Enzyme activation step	95°C	10 min,	1 cycle
Cycling	95°C	15s	40 cycles
	60°C - 70°C	30s	
	72°C	30s	
Melting curve analysis	60°C - 98°C		1 cycle

The initial denaturing step allows activation of the DNA polymerase for a hot start PCR, reducing the generation of non-specific amplification products. The Lightcycler instrument detects the fluorescent signal from each sample at the end of each extension step in the PCR cycle, SYBR green dye binds to double stranded DNA to generate a fluorescent signal, therefore as the amount of double stranded DNA product increases in each reaction so does the fluorescent signal from the respective PCR reaction. During the melting curve analysis SYBR green fluorescence is monitored continuously as samples are heated from the annealing temperature to 98°C, generating a specific profile dependent on the nucleotide content of the PCR product or products generated. Lightcycler analysis of the standard curve samples run in parallel to the samples of interest was carried out using the best fit point's method, reference to the standard curve allowed quantification of target sequence in the samples of interest.

## 2.3.12 Rapid amplification of cDNA ends PCR (RACE-PCR)

## 2.3.12.1 5'RACE-PCR

Rapid amplification of cDNA ends allows the extension of a messenger RNA sequence from a known internal site towards undefined sequence at the 5' end (Fig.12). The Gibco 5'RACE kit (5'RACE System for Rapid Amplification of cDNA Ends, Version 2 Gibco Life Technologies) was used for all RACE reactions.

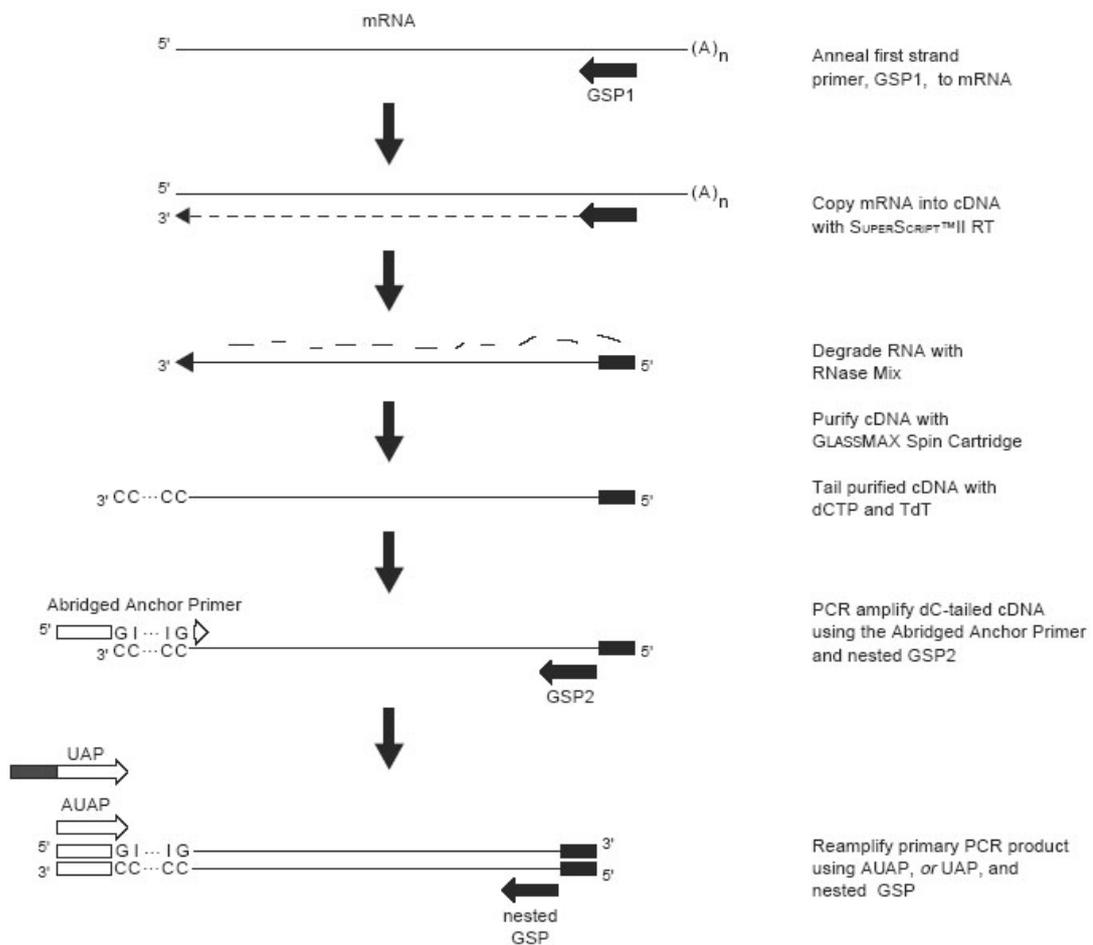


Fig.17 Schematic overview of 5'RACE methodology.

First strand cDNA synthesis is carried out using a gene specific internal antisense primer (GSPI); reverse transcription converts a specific subset of mRNAs or related mRNAs into a DNA template for further specific amplification.

The following reagents were added to a PCR tube on ice for each sample

RNA template	1µg-5µg
GSPI	2.5pmoles
dH <sub>2</sub> O	final volume 15.5µl

The reaction was incubated at 70°C for 10 minutes to denature RNA secondary structure and then cooled to 50°C. The following were added to a separate PCR tube and equilibrated to 50°C prior to mixing with the RNA template reaction above.

10x PCR buffer	2.5µl
MgCl <sub>2</sub> (25mM)	2.5µl
dNTP's (10mM)	1µl
DTT (0.1M)	2.5µl

1µl of Superscript II reverse transcriptase enzyme (50 units) was added and mixed. The reaction was incubated at 50°C for 50 minutes; the reaction was terminated by incubation at 70°C for 10 minutes. Residual RNA was degraded from the synthesised cDNA with the addition of 1µl RNase enzyme and incubation at 30°C for 30 minutes. Purification of the first strand cDNA synthesis products was carried out using the GlassMax DNA isolation spin cartridge according the manufacturers instructions.

Purified first strand cDNA is extended at the 3' end, terminal deoxynucleotidyl transferase (TdT) is used to add a homopolymeric cytosine tail to the end of the cDNA.

The following reagents were mixed in a PCR tube on ice.

Purified cDNA sample	10µl
5x tailing buffer	5µl
dCTP (2mM)	2.5µl
dH <sub>2</sub> O	6.5µl

The dCTP tailing mix was incubated for 2 minutes at 95°C and cooled on ice for 1 minute. 1µl of TdT enzyme was added and the reaction incubated at 37°C for 10 minutes, prior to deactivation of the TdT enzyme with incubation at 65°C for 10 minutes.

Utilizing the homopolymeric cytosine tail as a primer (Abridged Anchor Primer - AAP) binding site and a nested primer (GSPII) downstream of GSPI, efficient amplification of this region can be achieved using standard PCR.

The following were added to a PCR tube on ice.

dC-tailed cDNA	5µl
10x PCR buffer	5µl
MgCl <sub>2</sub> (25mM)	3µl
dNTP's (10mM)	1µl
GSPII primer (10µM)	2µl
AAP primer (10µM)	2µl
dH <sub>2</sub> O	31.5µl

0.5µl of Taq polymerase (5units/µl) was added before mixing and the PCR tubes transferred to a thermal cycler, the following PCR cycling protocol was used.

Initial denaturing	95°C	2 min	1 cycle
Cycling	95°C	60s	30 cycles
	55°C - 65°C	40s	
	72°C	90s	
Terminal elongation	72°C	5 min	1 cycle
Cycle hold	4°C		

Amplification products are analysed by agarose gel electrophoresis, extension products of an expected size were available to be subcloned for further analysis.

### 2.3.12.2 Nested RACE-PCR amplification

To achieve further specificity of amplification, primary PCR product were diluted 5 $\mu$ l in to 495 $\mu$ l TE buffer and 5  $\mu$ l of this dilution was used in the PCR reaction mix. The reaction mix was set up as above except that 1.0 $\mu$ l of nested GSP (10 $\mu$ M) and 1.0 $\mu$ l of abridged universal amplification primer (AUAP) replaced the original primers used. Subsequent rounds of PCR are analysed by agarose gel electrophoresis, products are subcloned for further analysis.

## 2.3.12.3 3'RACE-PCR

3' RACE takes advantage of the poly-A tail found in eukaryotic mRNA as a generic priming site for PCR (Fig.13), converting mRNA into cDNA using reverse transcriptase and an oligo-dT adapter primer (AP). The Gibco 3'RACE kit (3'RACE system for rapid amplification of cDNA ends, Gibco Life technologies) was used for all RACE reactions.

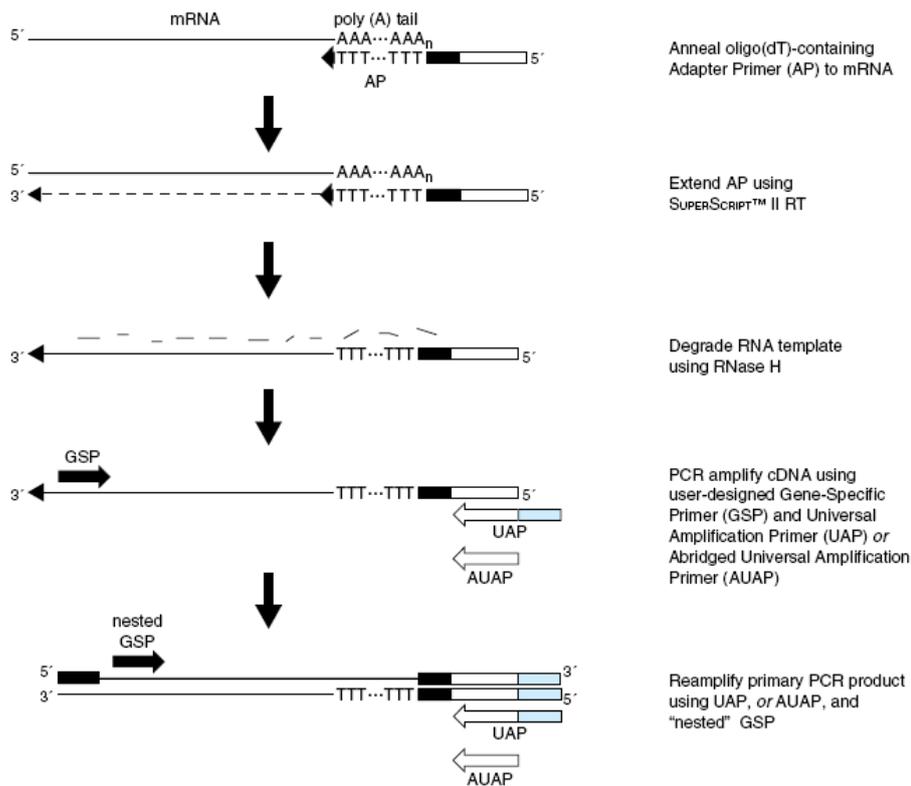


Fig.18 Schematic overview of 3'RACE methodology.

First strand cDNA synthesis utilising the Adapter Primer (AP) was carried out using total RNA isolated by the TRIzol method.

The following reagents were assembled in a PCR tube on ice.

RNA (1µg-5µg)	
dH <sub>2</sub> O	11µl final volume
AP (10µM)	1µl

The reaction was incubated at 70C° for 10 minutes before rapid cooling on ice.

Extension of the AP primer was achieved by the Superscript II reverse transcriptase enzyme; the reaction was assembled in a PCR tube and added to the AP mix described above.

10x PCR buffer	2 $\mu$ l
MgCl <sub>2</sub> (25mM)	2 $\mu$ l
dNTP's (10mM)	1 $\mu$ l
DTT (0.1M)	2 $\mu$ l

The combined AP and reaction mixture was incubated at 42°C for 2 minutes, 1 $\mu$ l of Superscript II reverse transcriptase enzyme (50 units) was added and the reaction allowed to proceed for 50 minutes. The reaction was terminated by heating to 70°C for 10 minutes. Template RNA was degraded by RNase to improve the sensitivity of subsequent amplification reactions.

Specific amplification is then achieved using a gene-specific sense primer (GSP) annealing in a known region of exonic sequence and a universal amplification primer (UAP) binding to the adapter primer of the first strand synthesis. The protocol permits the capture of unknown 3'-mRNA sequences that lie between the exon and the poly-A tail.

The following components were assembled in a PCR tube on ice.

AP primed cDNA	2 $\mu$ l
10x PCR buffer	5 $\mu$ l
MgCl <sub>2</sub> (25mM)	3 $\mu$ l
dNTP's (10mM)	1 $\mu$ l
GSP primer (10 $\mu$ M)	2 $\mu$ l
AAP/AUAP primer (10 $\mu$ M)	2 $\mu$ l
dH <sub>2</sub> O	31.5 $\mu$ l

0.5 $\mu$ l of Taq polymerase (5units/ $\mu$ l) was added before mixing and the PCR tubes transferred to a thermal cycler, a standard PCR cycling protocol was used. Amplification products were analysed by agarose gel electrophoresis, extension products of an expected size were subcloned for further analysis.

## 2.4 Protein techniques

### 2.4.1 pRSET Expression System (Invitrogen)

The pRSET vectors are pUC-derived expression vectors designed for high-level protein expression and purification from cloned genes in *E. coli*. High levels of recombinant protein expression was achieved from cDNA sequences cloned in frame with an N-terminal fusion protein, driven by a T7 promoter under the control of elements of the lactose operon.

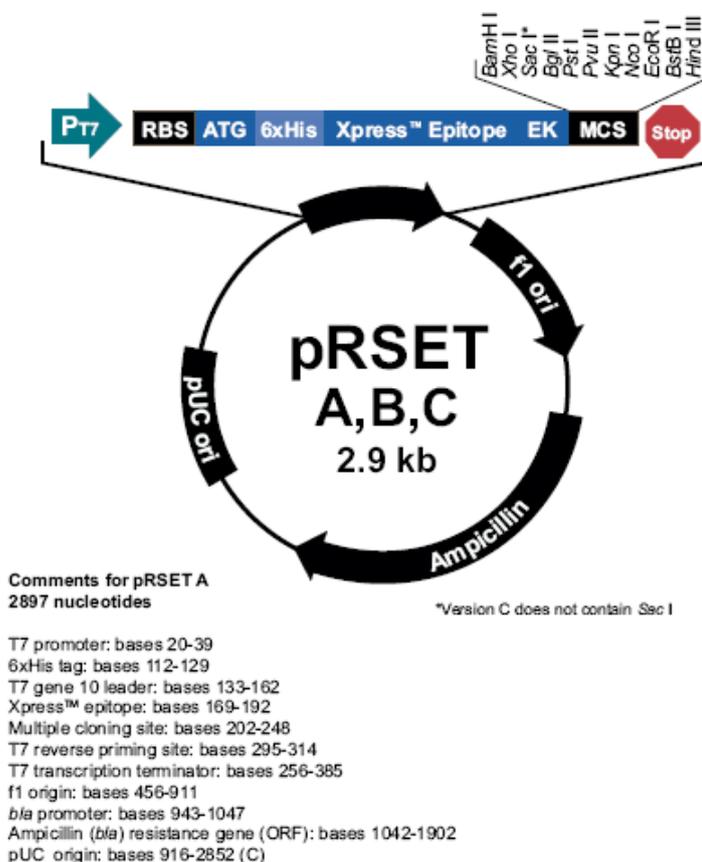


Fig.19 Prokaryotic expression vector pRSET circle map. Indicating elements of the multiple cloning site.

Gene specific primers were designed including restriction endonuclease sequences to generate an expression cassette, using standard PCR techniques. PCR products were first cloned into the pGEM T-Easy vector before being transferred inframe to the pRSET vector using standard molecular biology techniques. The correct incorporation of each expression cassette as inframe with the N-terminal fusion protein of the pRSET vector was confirmed by DNA sequencing using the T7 primer.

Confirmed pRSET and expression cassette plasmid constructs were transformed into the expression strain of *E.coli*, BL21 (DE3) pLYSs using the standard heat shock technique. Transformed BL21 (DE3) pLYSs *E.coli* were selected on LB Agar plates (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.5% Agar) containing ampicillin (50µg/ml), incubation overnight at 37°C produced bacterial colony clones available for protein expression. A single recombinant colony was inoculated into 10ml of SOB medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5mM KCl, 0.01M MgCl<sub>2</sub>) containing ampicillin (50µg/ml), and grown overnight at 37°C with shaking. 5ml of the starter culture was then added to 500ml SOB medium in a 2 litre baffled flask, and grown until the OD<sub>600</sub> was 0.4-0.6 at which point protein expression was induced with the addition of IPTG to a final concentration of 1mM. Induction was continued for up to 5 hours after which cells were harvested.

#### 2.4.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by denaturing discontinuous SDS-PAGE, under reducing conditions, according to the method of Laemmli (Laemmli 1970). Gels were cast and run using the Mini-protean II electrophoresis cell (Biorad). A 12.5% resolving gel (3.12ml 40% 29:1 acrylamide:bis-acrylamide (Protogel), 2.5ml 4x resolving gel buffer (0.5M Tris-Cl pH8.0, 0.4% SDS), 50µl 10% APS, 10µl Temed, 3.73ml dH<sub>2</sub>O) was cast between glass plates separated by 1.5mm spacers. 100µl of dH<sub>2</sub>O-saturated butanol was layered on top of the gel which was left to polymerise for 1 hour, after which the dH<sub>2</sub>O-saturated butanol was removed. A 4.5% stacking gel (0.56ml 40% 29:1 acrylamide:bis-acrylamide (Protogel), 1.25ml 4x stacking gel buffer (0.5M Tris-Cl pH6.8, 0.4% SDS), 30µl 10% APS, 10µl Temed, 2.96ml dH<sub>2</sub>O) was poured on to the surface of the polymerised resolving gel, a Teflon comb was inserted and the gel allowed to polymerise. The comb was removed to form sample wells and the gels placed in the electrophoresis tank filled with SDS running buffer (0.19M Glycine, 25mM Tris-Cl, 0.1% (w/v) SDS). Protein samples were placed in a boiling water bath for 10 minutes then cooled directly prior to loading 15µl-25µl into the preformed sample well, 10µl of protein molecular weight marker (Biorad) were run alongside protein samples. Gels were run at 100V until the dye front reached the resolving gel, at which time the voltage

can be increased to a maximum of 200V. After the completion of the run the apparatus was disassembled and the gel recovered for further analysis.

### 2.4.3 Coomassie blue staining of SDS-polyacrylamide gels

Gels were fixed and stained with coomassie blue staining solution (0.05% (w/v) Coomassie blue R-250, 50% (v/v) methanol, 10% (v/v) acetic acid) for a period of 1 hour, removal of non-specific background coomassie blue was achieved by immersing the gels in Coomassie blue destain solution (10% (v/v) methanol, 5% (v/v) acetic acid). Polyacrylamide gels were shaken in Coomassie blue destain solution for 5 - 20 hr, with regular change of the Coomassie blue destain, until protein bands could be clearly visualised. PAGE gels were archived by drying between sheets of clear acetate at 80°C for 2 hours using a Bio-Rad gel dryer (Bio-Rad).

### 2.4.4 Transfer of SDS-PAGE separated proteins to nitrocellulose membrane

A Bio-Rad Mini Protean Trans Blot system™ was used to transfer proteins separated by SDS-PAGE onto a supported nitrocellulose membrane. After running, SDS-PAGE gels were equilibrated in pre-chilled western blot transfer buffer (25mM Trizma base, 192mM Glycine, 20% (v/v) methanol) at 4°C for 15 min. The SDS-PAGE gel was then placed on top of a piece of Whatman 3MM chromatography paper. Supported nitrocellulose membrane was wetted in western blot transfer buffer and placed on top of the gel, taking care to remove bubbles, and the sandwich completed with wetted Whatman 3MM chromatography paper and pads supplied with the system. The sandwich was closed into the blotting cassette and placed in the transfer tank according to the manufacturers instructions with the supported nitrocellulose membrane towards the anode. An ice block was placed in the tank, which was then topped up with cold transfer buffer. Transfer of SDS-PAGE separated proteins to the supported nitrocellulose membrane was achieved in 1 hour at 100V or overnight at 30V.

### 2.4.5 Immunodetection of nitrocellulose membrane immobilised protein

After protein transfer the nitrocellulose filters were placed in 10ml blocking buffer (10mM Tris-Cl pH7.5, 2mM EDTA, 150mM NaCl, 5% (w/v) non fat milk powder, 0.05% (v/v) Tween 20) at room temperature for a minimum of 1 hour, to occupy all available protein binding sites remaining on the nitrocellulose membrane with a non-reactive protein. The nitrocellulose membrane was then exposed to the primary antibody diluted in blocking buffer to the appropriate concentration, and allowed to incubate for 1 hour with constant agitation. The primary antibody solution was removed and the nitrocellulose membrane was washed to remove non-specifically binding antibody, 3 x 10 minutes in western blot washing buffer (10mM Tris-Cl pH7.5, 2mM EDTA, 150mM NaCl, 0.1% (v/v) Tween 20). The nitrocellulose membrane with specifically bound primary antibody was exposed to the appropriate horseradish peroxidase-conjugated secondary antibody for 20 - 45 min in 10ml blocking buffer (10mM Tris-Cl pH7.5, 2mM EDTA, 150mM NaCl, 5% (w/v) non fat milk powder, 0.05% (v/v) Tween 20). Nitrocellulose membrane was washed to remove non-specifically binding antibody, 3 x 10 minutes in western blot washing buffer (10mM Tris-Cl pH7.5, 2mM EDTA, 150mM NaCl, 0.1% (v/v) Tween 20). The position of the protein of interest with specifically bound primary and labelled secondary antibody on the nitrocellulose membrane was visualised using the enhanced chemi-luminescence reagents (ECL; Amersham), according to the manufacturers instructions. Briefly excess liquid was removed from the nitrocellulose membrane before being incubated for 60s in the combined reagents of the ECL kit. The nitrocellulose membrane was covered in Saran wrap and exposed to blue sensitive X-ray film; the photographic film was processed using an automatic developer (Gevamatic 60, AGFA-Gevaert Ltd) according to the manufacturer's instructions.

### 2.4.6 Protein purification (FPLC)

All protein purification procedures were carried out using the Bio-Rad HR Biologic FPLC apparatus (Bio-Rad) under the control of the Biologic software. The purification system controlled the flow of two independent pumps, precisely delivering chromatography buffer mixtures to the chromatography column. The Bio-Rad HR Biologic FPLC apparatus detected and recorded the UV absorbance of the column eluent (450nm) and controlled fraction collection, enabling reproducible protein separation.

### 2.4.7 Buffer preparation

Chromatography buffers were prepared from analytical grade chemicals at the appropriate concentration, deionised water and equilibrated to the appropriate pH. Buffers were filtered through a 22µm membrane and degassed under negative pressure using a vacuum pump to remove dissolved gases.

### 2.4.8 Immobilised metal affinity chromatography (IMAC)

Immobilised metal affinity chromatography utilises the reversible interaction between the side chains of certain amino acid residues and immobilised metal ions (Porath, Carlsson et al. 1975). The Probond resin (Invitrogen) uses nitrilotriacetic acid (NTA) crosslinked to a 6% agarose matrix to immobilise nickel ( $\text{Ni}^{2+}$ ) ions, using its tetradentate chelating properties. Bound  $\text{Ni}^{2+}$  ions are able to form reversible coordinated interactions with histidine residues under both native and denaturing conditions, this binding interaction can be disrupted by the addition of excess or an increasing concentration gradient of histidine, glycine, glutamate, sulphhydryl reagents and imidazole. Recombinant expression systems utilise the affinity of histidine residues to immobilised metal ions by the inclusion of a 6xHis residue tag to be included in the recombinant protein, facilitating affinity protein purification. The Probond resin was prepared according to the manufacturer's instructions.

### 2.4.9 Native immobilised metal affinity chromatography

The chromatography step was carried out under native conditions, the harvested cells from recombinant protein expression induction being resuspended directly in the binding buffer (formulation referred to in appropriate results section) and ultrasonicated to disrupt the bacterial cell wall. Bacterial cell lysate was prepared as described in the chromatography sample preparation section, prior to application to 1ml column of Probond resin equilibrated with the binding buffer and allowed to flow through via gravity. The column was then washed with several column volumes of wash buffer (binding buffer + 500mM NaCl), and elution achieved by the addition of imidazole to a final concentration of 300mM in a step gradient controlled by the Bio-Rad HR Biologic FPLC apparatus. 1ml fractions of each elution were collected.

### 2.4.10 Denaturing immobilised metal affinity chromatography

The chromatography step was carried out under denaturing conditions, the harvested cells from the induction being solubilised directly in the binding buffer (6M Guanidine-HCL, pH7.8, 20mM NaPO<sub>4</sub>, 500mM NaCl) then cleared by centrifugation at maximum speed. The soluble lysate was then added to a 1ml column of Probond resin equilibrated with the binding buffer and allowed to flow through via gravity. The column was then washed with several column volumes of wash buffer (8M urea, pH7.8, 20mM NaPO<sub>4</sub> 500mM NaCl), and elution achieved by the addition of imidazole to a final concentration of 300mM in a step gradient controlled by the Bio-Rad HR Biologic FPLC apparatus. 1ml fractions of each elution were collected.

### 2.4.11 Ultrasonication of bacterial cells

Expressing cell cultures were harvested by centrifugation at 10,000g for 10 min at 4°C. The culture media supernatant was discarded. The cell pellet was resuspended in an appropriate buffer solution, +/- protease inhibitor solution. Disruption of the bacterial cell wall was achieved using ultrasonication on ice for 5 x 30 seconds. Insoluble cellular debris was removed by centrifugation at 10,000g for 10min at 4°C; the supernatant was removed for further analysis.

#### 2.4.12 Chromatography sample preparation

Samples for protein purification were prepared for chromatography by removal of insoluble debris. Samples were centrifuged at 10,000g for 10min, the supernatant was retained and filtered through a 22µm membrane. The soluble filtrate was suitable for application to chromatography columns.

#### 2.4.13 Protein sample dialysis

Dialysis membranes were prepared as per the manufacturer's instructions, before being clamped at one end. The sample was introduced into the membrane before the open end was itself clamped; the membrane was checked for leaks before being immersed in the desired buffer. Dialysis was allowed to proceed with stirring for a minimum of 4 hours with several changes of buffer.

## 2.5 Bioinformatics

### 2.5.1 DNA and amino acid sequence alignment

Alignment of DNA and amino acid sequences was carried out using the Basic Local Alignment Search Tool (BLAST) (Altschul, Gish et al. 1990) hosted by the National Centre for Biotechnology Information (NCBI) at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. All sequences were entered into the search tool in FASTA format via the web based interface, sequence information was compared against the current sequence held in the NCBI database, and the nature of the search is modifiable dependent on the variant of the BLAST algorithm chosen from the search tool menu.

Nucleotide BLAST – Direct comparison of DNA nucleotide sequence against the NCBI nucleotide database.

Protein BLAST – Direct comparison of amino acid (single letter code) sequence against the NCBI nucleotide database.

#### Translated BLAST

Blastx – Converts a nucleotide query sequence into protein sequences in all 6 reading frames, translated protein products are then compared against the NCBI protein databases.

TBlastx – Converts a nucleotide query sequence into protein sequences in all 6 reading frames and then compares this to an NCBI nucleotide database which has been translated in all six reading frames.

DNA and protein alignments were presented graphically, with similar regions of sequence aligned.

Similarity of the query and database sequence is assessed by statistical significance, dependent on the query sequence and the database being interrogated. Presented as an E-value, this is the measure of the odds that the similarity identified occurred due to chance.

Pairwise sequence analysis was carried out using the BLAST and ClustalW (Thompson, Gibson et al. 2002) (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) internet based algorithm tools and the AlignX module of the Vector NTI 11 software (Invitrogen).

### 2.5.2 Genomic sequence analysis

Analysis of genomic sequence data was facilitated by the online search and analysis tools Ensemble (<http://www.ensembl.org/index.html>) and UCSC genome browser (<http://genome.ucsc.edu/>). The Ensemble project provides automatic genomic sequence annotation utilising publicly available sequence database from all available species. Automatically generated Ensemble genebuild genes are assigned a HGNC (HUGO gene nomenclature committee) symbol, Ensemble ID begin ENSG.

Predicted gene sequences based on genomic sequence data were pairwise aligned using the BLAST sequence analysis algorithm, against human and rat reference sequence to confirm the validity of the gene model.

### 2.5.3 Phylogenetic analysis

Phylogenetic trees were constructed based on the ClustalW alignment of amino acid sequences, using the MEGA4 software suite (Tamura, Dudley et al. 2007). The Neighbour Joining method was used to determine the relationships between aligned sequence, excluding gaps by pairwise deletion. Bootstrap analyses of the resulting phylogenetic tree were displayed as percentages.

# Results

## Chapter 3

### 3.1 Generation of a MASP-3 specific cDNA probe

A MASP-3 specific cDNA probe was generated from a full-length human MASP-3 cDNA clone (accession number AF284421) by PCR. Primers (hu MASP-3 p5 F and hu MASP-3 p8 R) designed entirely within the serine protease domain (Fig.20) gave a product of 688 bp, which was subcloned into the TA cloning vector (pGEM-T easy, Promega CA) before being sequenced.

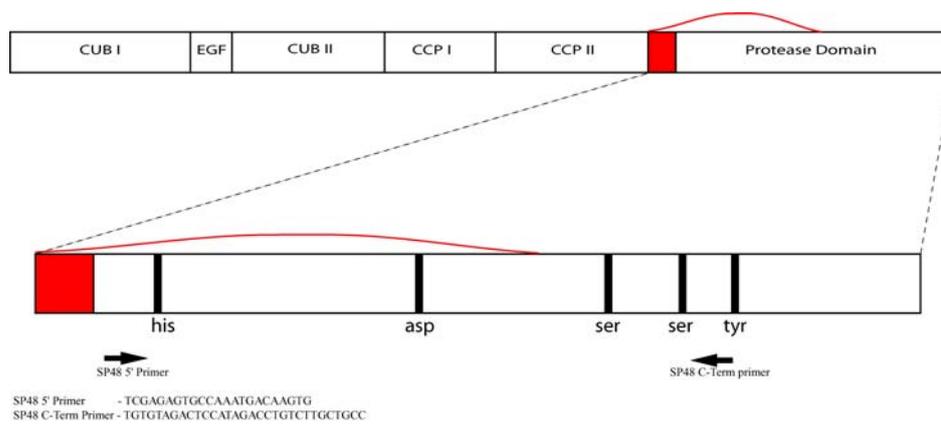


Fig.20 Human MASP-3 specific PCR primer binding positions with respect to MASP-3 protein domain structure.

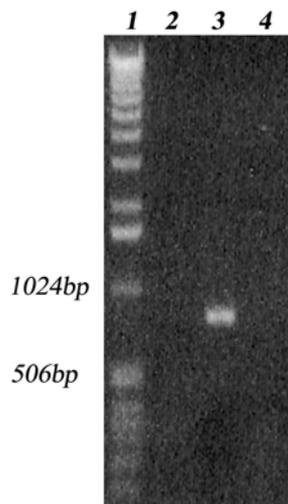


Fig.21 Gel electrophoresis showing product of human MASP-3 specific PCR utilising primer pairs. Lane 1 - 1kb DNA marker, lane 2 - Hu MASP-3 p5F + p8R, lane 3 - Hu MASP-3 p6F + p8F and lane 4 - negative control.

A specific human MASP-3 cDNA probe was generated of predicted size (Fig.21), and its sequence confirmed. The PCR product was entirely specific for the MASP-3 serine protease domain, spanning a region encompassing the active site residues.

### 3.2 Cloning of rat (*Rattus norvegicus*) MASP-3

A cDNA library constructed from rat liver undergoing an acute phase response was screened using a  $^{32}\text{P}$  dCTP radio labelled human MASP-3 probe, a PCR product specific for the human MASP-3 serine protease domain was generated from the template described by Dahl et al (Dahl, Thiel et al. 2001). The 688bp PCR product was subcloned into pGEM-T easy (promega, CA) and sequenced. The MASP-3 specific probe was excised from the TA cloning vector by EcoRI restriction digest, separated in and a specific band recovered from the agarose gel. 100ng of the MASP-3 specific DNA band was random primed  $^{32}\text{P}$  radio labelled and used as a probe to screen a cDNA library constructed from acute-phase rat liver (Sprague-Dawley rats, Stratagene (Cambridge, UK), Catalogue no. 936512) cloned in the phage vector  $\lambda$ ZAP. Approximately  $5 \times 10^6$  cfu's were screened as described, a minimum of three enrichment cycles yielded a number of isolated hybridizing phage plaques, the  $\lambda$ ZAP vector containing the cloned cDNA's were in vivo excised from phage genome allowing small scale plasmid preparations to be made. Individual clones were analysed by EcoRI/XhoI restriction digestion and southern blot (Fig.22) hybridization, clones of greatest length were further characterised by restriction endonuclease mapping.

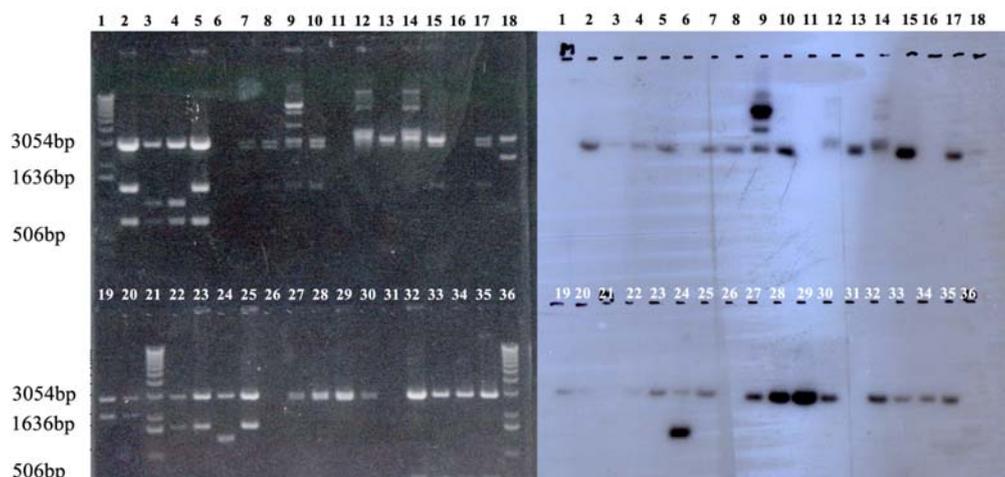


Fig.22 Restriction endonuclease (EcoRI/XhoI) digest of in vivo excised rat MASP-3 acute phase liver cDNA phage library clones (left panel). Southern blot hybridization (right panel) of restriction endonuclease (EcoRI/XhoI) digest of in vivo excised rat MASP-3 acute phase liver cDNA phage library clones (left panel) with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) outlined in Fig.15, Fig.16. (1 - 1kb DNA marker, lanes 2-5 phage clone 1.1/1A, lanes 7-10 phage clone 7-1A, lanes 12-15 phage clone 7-1B, lanes 16-20 phage clone 7-3B, lane 21 - 1kb DNA marker, lanes 22-25 phage clone 13-1/1A, lanes 27-30 phage clone 14-1/1A, lanes 32-35 phage clone 14-2/2A and lane 36 1kb marker)

Re-sampling and *in vivo* excision of an initially hybridising phage colony 1.1.1 isolated a further independent clone (Fig.23).



Fig.23 Restriction endonuclease digestion (EcoRI) of *in vivo* excised rat MASP-3 acute phase liver cDNA phage library clone 1.1.1 (left panel). Southern blot hybridization (right panel) of restriction endonuclease (EcoRI) digest of *in vivo* excised rat MASP-3 acute phase liver cDNA phage library clone 1.1.1 (left panel) with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) outlined in Fig.15, Fig.16. (lane 1 - 1kb DNA marker, lanes 2-13 1.1.1 plasmid preparations)

Clones 14.1, 1.1.1, 7.1 and 13.1 were deemed to be sufficiently different (Fig.23, Fig.24), based on the results of restriction endonuclease mapping and southern blotting to be sequenced.

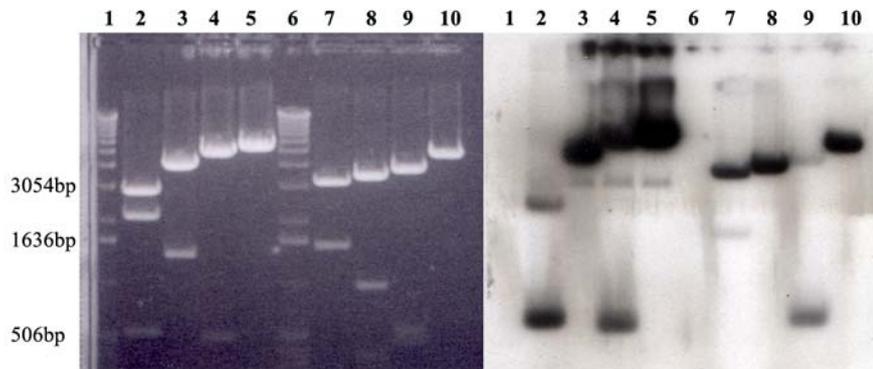


Fig.24 Restriction endonuclease characterisation of rat MASP-3 acute phase liver cDNA phage library clones (left panel) specifically hybridising with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) (right panel). Lane 1 – 1kb DNA marker, Rat MASP-3 cDNA clone 14.1 restriction endonuclease digestion lane 2 – AccI, lane 3 - Hind III, lane 4 – PstI, lane 5 – XhoI. Lane 6 – 1kb DNA marker. Rat MASP-3 cDNA clone 1.1.1 restriction endonuclease digestion lane 7 - AccI, lane 8 - Hind III, lane 9 - PstI, lane 10 - XhoI. Left panel - agarose gel electrophoresis, right panel - Southern blot hybridisation utilising human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) of right panel agarose gel electrophoresis.

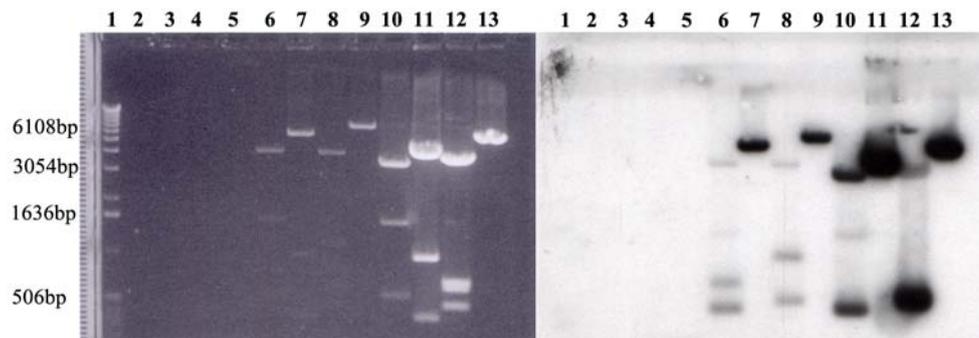


Fig.25 Restriction endonuclease analysis of rat MASP-3 acute phase liver cDNA phage library clones 7.1 and 13.1 specifically hybridising with human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F). Left panel - agarose gel electrophoresis, right panel - Southern blot hybridisation utilising human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F) of right panel agarose gel electrophoresis. Lane 1 – 1kb DNA marker. Rat MASP-3 cDNA clone 7.1 restriction endonuclease digestion lane 6 - *AccI*, lane 7 - *HinD III*, lane 8 - *PstI*, lane 9 - *XhoI*. Rat MASP-3 cDNA clone 13.1 restriction endonuclease digestion lane 10 - *AccI*, lane 11 - *HinD III*, lane 12 - *PstI*, lane 13 - *XhoI*.

Complete sequence of clones 13.1, 7.1 and partial sequence of clones 14.1, 1.1.1 were obtained. Sequence analysis indicated strong homology to published human MASP-3 transcript, the relative positions of each cDNA and their size.

Rat MASP-3 cDNA clone 7.1 (4224bp) beginning at amino acid position + 32 in the N-terminal region of CUBI encompassing the remaining coding region of MASP-3, 3' untranslated region and Poly-A tail (poly adenylation signal position 2550bp and 3928bp). An open reading frame extends from CUBI to the end of CCPII exon I, a region of retained intronic sequence (258bp of a 1389bp intron) terminates the open reading frame. The reading frame resumes at the next exon and continues to a TGA stop codon, a poly-A tail is preceded by 1.7kb of 3'untranslated region.

Rat MASP-3 cDNA clone 1.1.1 (2.7kb) and 13.1 (2.6kb) cDNA clones are entirely contained within 7.1, both sequences retain unspliced sequence from the intron between first and second exon of CCPII. Overlapping regions were found to be identical, the sequence information used to build a complete contig representing the rat MASP-3 mRNA.

Rat MASP-3 cDNA clone 14.1 (2.6kb) encompasses partial MASP-3 serine protease domain sequence, and retains intronic sequence from the region preceding the serine protease domain in contrast to the previously described clones.

The rat MASP-3 sequence obtained contains the entire MASP-3 specific region and all but the 3'UTR and first 93bp of the shared MASP-1/3 coding region.

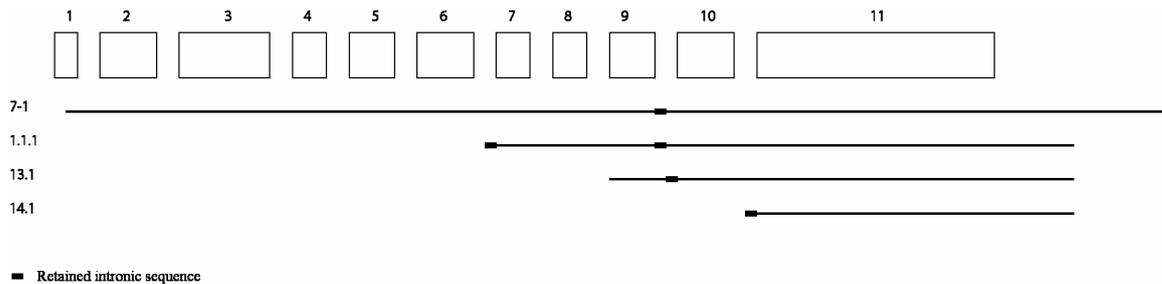


Fig.26 Schematic representation of rat MASP-3 cDNA clones characterised in Fig.24 and Fig.25, with respect to MASP-3 gene structure. Upper panel - MASP-3 exon structure (exons 1 – 11). Lower panel – represented length of rat MASP-3 cDNA clones 7.1, 1.1.1, 13.1 and 14.1. Regions of retained intronic sequence are highlighted in bold line.

The relative lengths and positions of rat MASP-3 cDNA clones derived from this work and discussed above are represented in fig.26, with respect to the MASP-3 exon structure. All rat MASP-3 cDNA clones obtained encompass a MASP-3 specific region, cDNA clones 7.1, 1.1.1 and 13.1 are contiguous with MASP-1/3 common sequence specific for exon 10. Rat MASP-3 cDNA clone 14.1 retains intronic sequence prior to the MASP-3 serine protease domain and as such cannot be directly linked to the common MASP-1/3 sequence.

Rat MASP-3 cDNA clones 7.1, 1.1.1 and 13.1 also exhibit retention of intronic DNA sequence between exons 9 and 10 in the common MASP-1/3 region, clone 1.1.1 also retains a intronic sequence at the start of the transcript corresponding to a region between exon 6 and 7.

Despite the incomplete nature of the obtained rat MASP-3 cDNA clones (7.1, 1.1.1, 14.1, 13.1), fig.26 clearly demonstrates a significant and reproducible proportion of the novel rat MASP-3 specific cDNA sequence was encompassed within the clones characterised in this work.

### 3.3 Rat MASP-3 5' RACE-PCR

All rat MASP-3 clones obtained from the acute phase liver library retained intronic sequence, to confirm if this was an artefact of the libraries construction or a specific splicing event a 5' RACE was performed spanning the region to conclusively link the MASP-3 specific exon to the MASP-1/3 shared region.

Total RNA prepared from rat liver was used as a template for first strand cDNA synthesis, utilising a gene specific primer (Rat MASP-3 GSPI) designed within the serine protease domain. The subsequent cDNA was extended at the 3' end with a dCTP tail. Specific amplification of rat MASP-3 sequence was achieved using a nested gene specific primer (Rat MASP-3 GSPII) also located within the serine protease domain and an abridged anchor primer (AAP), an oligonucleotide specific for the homopolymeric dCTP tail (Fig.27).

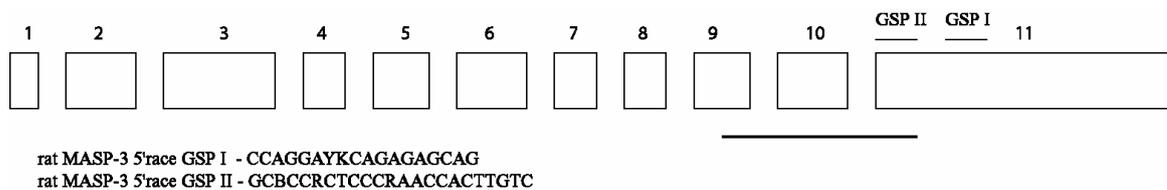


Fig.27 Schematic representation of rat MASP-3 5'RACE strategy and resulting cDNA clone. Upper panel – MASP-3 gene structure (exon 1-11), relative positions of gene specific primers GSP I and GSP II are indicated (primer sequences indicated in lower panel). Lower panel – rat MASP-3 5'RACE cDNA clone indicated as a bold line spanning exons 9-11 indicating a correctly spliced region and linking the serine protease domain to the common MASP-1/3 region.

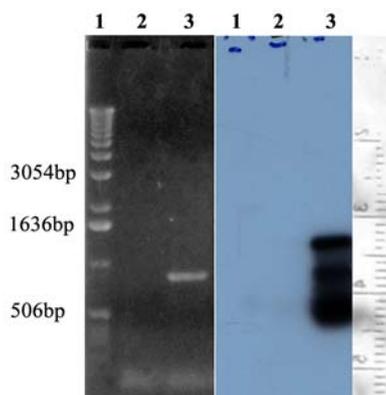


Fig.28 Agarose gel electrophoresis of rat MASP-3 5' RACE PCR (left panel), Southern blot hybridisation (right panel) of rat MASP-3 5' RACE PCR agarose gel (left panel) utilising human MASP-3 specific cDNA probe (Hu MASP-3 p6F + p8F). Lane 1 - 1kb DNA marker, lane 2 - negative control, lane 3 - rat MASP-3 5' RACE PCR.

The major visible product of the rat MASP-3 RACE was 800bp, although both 1.4kb and 500bp products could be detected by southern blot hybridisation (Fig.28). These potential products would not be long enough to contain the remaining MASP-3 specific region and entire MASP1/3 shared domains, but would be of a suitable length to resolve the retained intronic sequence present in the cDNA library screened rat MASP-3 clones. The major 800bp rat MASP-3 5' RACE product was cloned into pGEM-T easy and sequenced, the product spanned the region conclusively linking the rat MASP-3 specific region to the MASP-1/3 common area as a continuous reading frame (Fig.27).

A cDNA contig was assembled from the DNA sequence obtained, utilising the DNA analysis software SeqMan Pro (DNASTar).



BLAST (Altschul, Gish et al. 1990) DNA alignment of human MASP-3 (accession number AF284421) with the putative rat MASP-3 sequence obtained in this work exhibits an E-value of 0.0, evidence that the novel rat sequence is homologous to the previously characterised human MASP-3.

Rat MASP-3 cDNA predicted translation product exhibits conservation of a number of amino acid residues at functionally significant sites within the serine protease domain; within the remainder of the molecule residues important in orientation of the serine protease domain, dimer formation and associations with MBL and ficolin are also conserved.

Fundamental to the function of a serine protease is the catalytic triad, amino acid residues spatially arranged to drive the cleavage of a specific substrate. Histidine, aspartic acid and serine residues are conserved (highlighted cyan in fig.29); codon usage is variant at position 3 in all active site residues with respect to the human MASP-3 cDNA sequence. Human MASP-3 active site serine utilises an AGY codon (Y denoting T or C), this is also exhibited by Rat MASP-3 (fig.29). Rat MASP-1 is noted to retain an evolutionarily distant TCN (N denoting A, T, G or C) active site serine codon, requiring a minimum series of 2 mutations to transition from one codon to the other (Endo, Takahashi et al. 1998).

Serine protease residues at positions 214 and 225 also exhibit conservation and evolutionary significance of their codon usage. Residue 214 is a conserved serine residue (Ser214) (Krem, Prasad et al. 2002) residing in the primary binding pocket (S1) at close proximity to the active site, promoting efficient formation of an enzyme substrate complex. Rat MASP-3 conserves a serine residue at position 214; TCN (N denoting A, T, G or C) codon usage is exhibited in common with other MASP-like serine protease. Rat MASP-3 residue 225 is a conserved tyrosine amino acid (Krem and Di Cera 2001) identical to that seen in human MASP-3.

Residue 189 is a conserved aspartate responsible for coordination of Na<sup>+</sup> ion within the primary binding pocket (S1); this role is specific to sodium activated proteases of the complement and coagulation cascade. Rat MASP-3 conserves an aspartate residue at this position, in common with human MASP-3.

Residue at position 193 is responsible for interactions with the P1 residue of the substrate in conjunction with serine protease residue Ser195 via important hydrogen bonding, this crucial residue is a conserved glycine amino acid (fig.29). Rat MASP-3 conserves a glycine at position 193, in common with human MASP-3.

Rat MASP-3 in common with human MASP-3 retains five cysteine residues within the serine protease domain, the histidine loop disulphide bridge being absent. The disulphide bond configuration has evolutionary significance amongst vertebrate S1 peptidase, MASP-1 exhibits an alternative disulphide bond structure to that exhibited by MASP-3 (Kenesi, Katona et al. 2003).

The potential cleavage site for MASP-3 activation located within the linker region present in exon 11, arginine, isoleucine and isoleucine residues are conserved. There is variation at the codon third position for each codon between human and rat cDNA sequence, respective coding sequence for human MASP-3 RII cleavage site (AGG ATC ATT) and rat MASP-3 RII cleavage site (AGA ATT ATC).

Interaction between the CCPII domain and serine protease domain linker region at conserved residues stabilises the domains relative to each other, limiting flexibility and potentially modifying access to the serine protease active site (Dobo, Harmat et al. 2009). A proline-tyrosine-tyrosine (PYY) motif present within the CCPII domain is conserved between human and rat MASP-1/3 (fig.29), the reciprocal binding region within the MASP linker region consists of proline-valine-cysteine (PVC) with variation of the valine residue. There is conservation of the reciprocal PVC binding motif within the MASP linker region, exhibited in rat MASP-3 (highlighted pink fig.29).

Rat MASP-1/3 shared domains conserve residues important in homodimer formation via interactions within CUBI-EGF domains (Feinberg, Uitdehaag et al. 2003), highlighted in red and dark red fig.29.

Residues within the rat MASP-1/3 common domains important in the interaction with ficolin and MBL (highlighted in green and light green fig.29) are conserved with respect to human MASP-1/3.

Rat MASP-3 conserves the 11 exon structure encoding 6 protein domains, as exemplified by human MASP-3 (Dahl, Thiel et al. 2001). Analysis of available genomic sequence data, in conjunction with previously characterised rat MASP-1, confirms an identical MASP-1/3 loci structure. The intron and exon structure is consistent with the previously characterised human MASP1/3 locus. The codon phase of rat MASP-3 exons is conserved with that of human MASP-3, all exons of the MASP-1/3 locus utilise codon phase I (Stover, Lynch et al. 2003), splicing occurring between the first and second base of the spliced codon. A single poly-adenylation signal was identified within the 3' untranslated region of rat MASP-3 cDNA (accession number AJ487622), 1475bp downstream of the stop codon. No further poly-adenylation sites were predicted in the entire intronic region between rat MASP-3 serine protease domain and the first exon of rat MASP-1 serine protease domain.

CUBI-EGF-CUBII-CCPI-CCPII of the A-chain is encoded by 10 exons; MASP-3 serine protease domain is contained within a single exon (Fig.30). As seen in the human genomic sequence, rat MASP-1 serine protease is encompassed within 6 exons. Genomic sequence analysis utilising the 3' untranslated region of rat MASP-3 cDNA, indicate the loci resides on chromosome 11.

## Genomic organisation of rat MASP-1/3

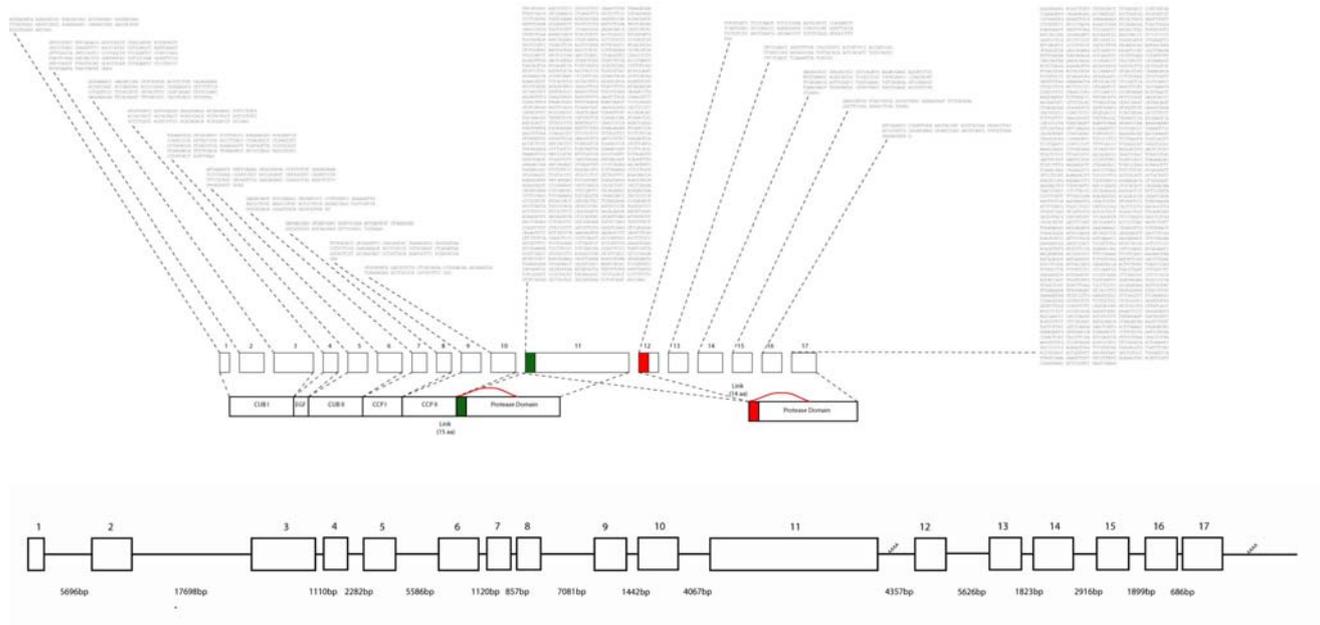


Fig.30 Rat MASP-1/3 protein domain structure (lower panel), linked to respective exons of the rat MASP-1/3 locus (central panel). cDNA sequence is presented in the upper panel associated with the respective exon.

Rat MASP-3 exhibits a high degree of identity with the previously characterised human MASP-3 sequence, both at the cDNA and translated primary amino acid sequence. This global level of similarity can be further dissected to reveal domains of the mosaic protein with important conserved function, individual protein domains determined on the basis of genomic exon structure were aligned in a pair wise manner utilising the ClustalW method (AlignX, Invitrogen). Analysis of the degree of sequence conservation between human and rat MASP-1/3 is presented below (Table.2).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	87%	95%	85%	88%	82%	85%	82%
Protein	90%	100%	88%	88%	77%	91%	81%

Table.2 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and rat MASP-1/3 cDNA (upper panel) and predicted translation products (lower panel).

As previously indicated there is a high degree of similarity between human and rat MASP-1/3, this is not uniform throughout the multiple domain structure (Table.2). CUBI and EGF domains exhibit a high degree of conservation at the amino acid level; CCP II appears to be less conserved. Comparing the alternatively spliced serine protease domain DNA sequence specific for MASP-1 and MASP-3 indicated a similar level of sequence identity, the primary amino acid sequence of MASP-3 exhibits a greater degree of conservation.

The primary translation product of rat MASP-3 cDNA was compared against a database of conserved protein families (Pfam) (Bateman, Birney et al. 2000), the predicted domain architecture was consistent to that exhibited by the previously characterised human MASP-3 (Fig.27). With reference to the consensus sequence for each predicted domain, rat MASP-3 exhibited probability values (E-value) of  $2^{-70}$ ,  $2$ ,  $4^{-48}$ ,  $1^{-8}$ ,  $9^{-8}$  and  $1^{-76}$  for domains CUBI-EGF-CUBII-CCPI-CCPII-Serine protease domain respectively.

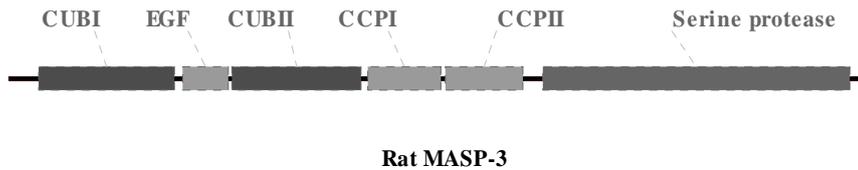


Fig.31 Rat MASP-3 protein domain structure predicted by comparison to the Pfam database, using Pfam analysis of rat MASP-3 primary amino acid sequence (Fig.29). Domains with maximal identity to Pfam consensus sequence are represented, indicating an identical architecture to that exhibited by human MASP-3 (Fig.3)

Analysis of the translational product of rat MASP-3 cDNA sequence (Fig.29) derived from this work indicates a polypeptide chain of 732aa including a signal peptide of 19aa. Potential N-glycosylation sites within MASP-3 are conserved between species, Rat MASP-3 retains 4 potential glycosylation sites within the A-chain and 3 within the serine protease domain as observed in the human primary amino acid sequence. The calculated molecular weight is 82,426Da without post-translational modification, the observed molecular weight of MASP-3 under reducing conditions was 110 kDa (zymogen) (Fig.32).

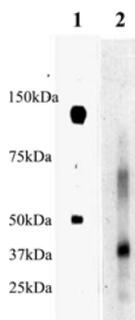


Fig.32 Western blot analysis of rat MASP-3 recombinant and plasma purified protein. Molecular weight markers (kDa) (left of panel), lane 1 - rat plasma lectin preparation (Mads Dahl), lane 2 - recombinant rat MASP-3 serine protease domain (Fig.66, Fig.68). Western blot analysed with anti-human MASP-3 antipeptide immune sera primary antibodies (Stover, Lynch et al. 2003).

Rat MASP1/3 locus is present on chromosome 11q23 spanning approximately 70kb, the region is bound by Somatostatin (Sst), Receptor transporter protein 4, a transcript similar to S60 Ribosomal protein L21 (LOC288005) and Receptor transporter protein 1 (RTP1) as in the human genomic region (Fig.33).

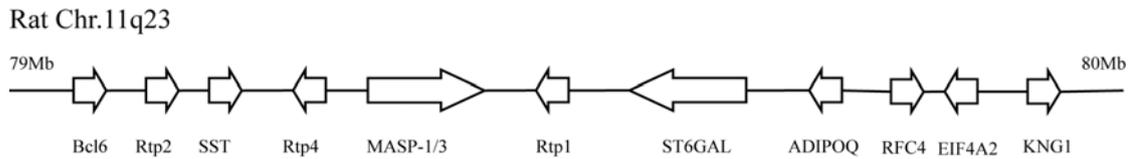


Fig.33 Rat chromosome 11q23, MASP-1/3 locus with respect to upstream and downstream gene loci. Relative size, position and orientation of each loci indicated by an arrow.

Rat MASP-3 shares the common intron exon structure as that found in human MASP-3, the MASP-1/3 locus as a whole exhibits identical structure to that seen in the human genomic sequence.

### 3.5 Cloning of mouse (*Mus musculus*) MASP-3

A cDNA library screening strategy was employed to obtain a mouse homologue of MASP-3, using a Lambda ZAPII mouse liver library (Stratagene). Numerous rounds of screening failed to produce any hybridising cDNA clones, utilising a human MASP-3 specific probe. An RT-PCR strategy was employed to obtain mouse MASP-3 sequence; human MASP-3 specific primer pairs were used under conditions of low stringency to amplify homologous regions of the mouse gene. MASP-3 specific PCR products were identified by Southern blot hybridisation utilising a rat MASP-3 cDNA  $^{32}\text{P}$  radio labelled probe (Fig.34).

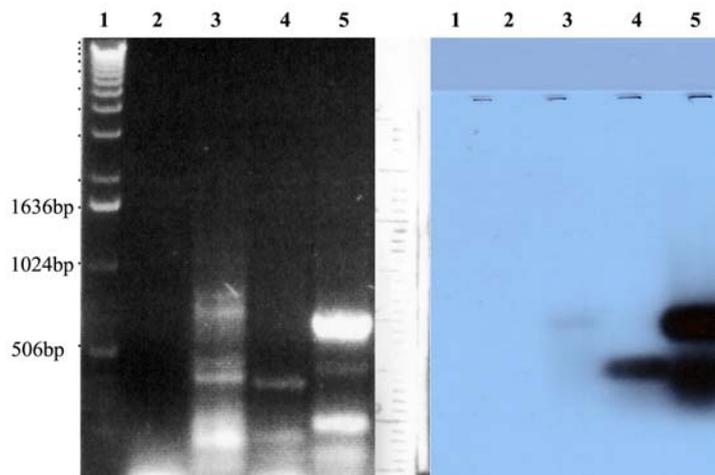


Fig.34 Agarose gel electrophoresis of mouse MASP-3 RT PCR utilising cDNA derived from mouse total liver RNA (left panel), Southern blot hybridisation (right panel) of mouse MASP-3 RT PCR agarose gel (left panel) utilising rat MASP-3 specific cDNA probe (rat MASP-3 PCR p6F + p8F).

Lane 1 - 1kb DNA marker, lane 2 - negative control, lane 3 - mouse MASP-3 PCR F1-BamHI + R1-EcoRI, lane 4 - mouse MASP-3 PCR TXP + R1-EcoRI, lane 5 - mouse MASP-3 PCR huMASP-3\_6 + huMASP-3\_8.

PCR products obtained in this way were hybridised with a  $^{32}\text{P}$  radioactively labelled rat MASP-3 specific cDNA probe, identifying specific amplification products. The most efficient primer pair, (human MASP-3\_6 and human MASP-3\_8 – Fig.34 lane 5) was chosen for further analysis. PCR products were subcloned into the pGEM T-easy vector (Promega, CA), individual clones were reselected by southern blotting using a rat MASP-3 specific probe. Positive candidates were selected for DNA sequencing. Sequencing of the putative mouse MASP-3 product (MoM36/8 - 690bp) indicated significant identity to the published human MASP-3 (accession number AF284421), a

BLAST alignment E-value of 0.0 confirming the novel sequence to be a mouse homologue of human MASP-3.

The novel mouse MASP-3 sequence resided entirely within the serine protease domain, encompassing the aspartic acid and serine residues of the active site. High quality cDNA sequence enabled predicted translation of a total of 219 amino acid residues.

Utilising the ClustalW method (Vector NTI 10.1.1, AlignX module) pairwise sequence alignment with the previously characterised human MASP-3 sequence identified an 81% sequence identity at the DNA level and 94% for the predicted translation product (Fig.37).

The newly obtained mouse MASP-3 specific sequence (MoM36/8) was used to screen a mouse liver poly-A cDNA library (Stratagene, CA) as previously described. One mouse specific clone (C57B16) was found to strongly hybridise with the mouse MASP-3 specific radioactively labelled cDNA probe. The potential mouse MASP-3 cDNA clone was recovered from the phage vector genome with specific primers flanking the multiple cloning site, after appropriate phage purification.

Putative mouse MASP-3 cDNA clone C57B16 spanned 2152bp (accession number AY135525S2). Sequencing data indicated overlapping identity at the 5' end to the previously characterised mouse MASP-3 (MoM36/8) RT-PCR product, determined by pairwise DNA sequence alignment (Vector NTI 10.1.1, AlignX module). A total of 71 amino acids completing the mouse MASP-3 serine protease coding region in the 3' direction from the original mouse MASP-3 RT-PCR product.

The resulting mouse MASP-3 specific contig (ContigExpress, Invitrogen) encompassed a total of 1193bp, including the stop codon (TGA) and 3' untranslated region of 1939bp.

### 3.6 Mouse MASP-3 5' RACE-PCR

Experiments to extend the sequence of mouse MASP-3 in a 5' direction were undertaken to link the MASP-3 specific serine protease domain to the shared MASP-1/3 region. Utilising the sequence data obtained from the mouse MASP-3 RT-PCR product, specific 5' primers were designed (MoGSP I, MoGSP II). A Rapid Amplification of cDNA Ends (RACE) methodology was employed to take advantage of the specific sequence information obtained in previous experiments. The 5' RACE PCR was optimised in an attempt to produce a clear product visible on agarose gel electrophoresis analysis, despite several rounds of optimisation the 5' RACE PCR product yield remained low (Fig.35). Detection of specific mouse MASP-3 5' RACE products was achieved by southern blot analysis, utilising a rat MASP-3 specific probe (rat MASP-3 PCR p6F + p8F) overlapping with the N-terminal region of the MASP-3 serine protease exon.

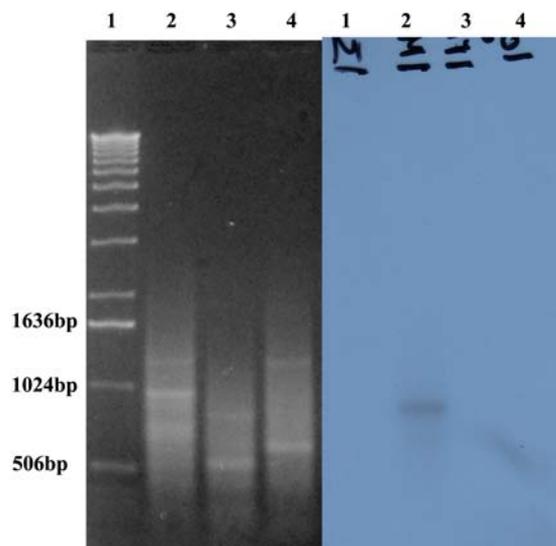


Fig.35 Agarose gel electrophoresis of mouse MASP-3 5' RACE PCR utilising cDNA derived from mouse total liver RNA (left panel), Southern blot hybridisation (right panel) of mouse MASP-3 5' RACE PCR agarose gel (left panel) utilising rat MASP-3 specific cDNA probe (rat MASP-3 PCR p6F + p8F). Lane 1 - 1kb DNA marker, lane 2 - mouse MASP-3 5'RACE PCR AAP + MoGSP II, lane 3 - mouse MASP-3 5'RACE PCR AAP + MoGSP I. (AAP – Abridged anchor primer), lane 4 - negative control.

Southern blot analysis of mouse MASP-3 5' RACE experiments clearly indicate a hybridising product that is of low yield and poorly visualised by agarose gel electrophoresis. Mouse MASP-3 5' RACE PCR products were isolated using pGEM-T

easy TA cloning vector, stabilising the PCR product and allowing further restriction endonuclease and sequence analysis.

Restriction endonuclease digestion (EcoRI) of potential mouse MASP-3 5' RACE pGEM-T easy constructs indicated excised DNA fragments of 850bp and 500bp (not visualised by agarose gel electrophoresis) in length, the largest clones were characterised by DNA sequencing.

Characterisation of the 5' RACE cDNA revealed an 850bp mouse specific sequence encompassing 144bp of the MASP-3 serine protease domain, continuous with the MASP1/3 CUB II, CCPI, and CCPII and domains. The 5'RACE cDNA clones obtained were not of sufficient length to complete the entire coding sequence of mouse MASP-3; a start codon was not identified. The sequence information was contiguous with the previously characterised mouse MASP-3, directly linking the mouse MASP-3 serine protease domain to the MASP-1/3 common domains.

Mouse MASP-3 5'RACE fragment contained 271aa, 48aa specific to the MASP-3 serine protease domain. The partial cDNA sequence extends in the 5' direction to the start of the CUBII domain, providing important information on the CCP domains and the link region between the MASP-1/3 region and the MASP-3 specific serine protease.

### 3.7 Mouse MASP-3

Mouse MASP-3 sequence data was derived from both cDNA phage library screening and from RT-PCR methodologies, utilising these techniques a 3493bp contig (accession number AY135525-135527) was constructed (ContigExpress, Invitrogen). Sequence analysis revealed a 1551bp reading frame terminating in a TGA stop codon, with a further 1939bp of 3' untranslated region. The mouse specific partial sequence contained the entire MASP-3 serine protease domain, and the shared MASP-1/3 CUBII, CCPI and CCPII domains. The overlapping cDNA sequences used to generate the mouse MASP-3 contig conclusively demonstrate the association of the MASP-3 serine protease domain and the shared MASP-1/3 domains in the mouse.

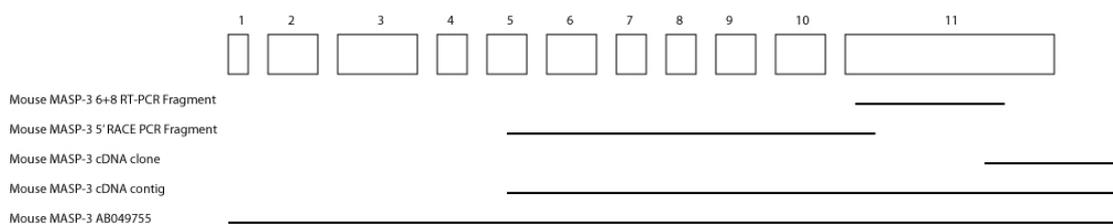


Fig.36 Schematic presentation of mouse MASP-3 cDNA clones generated in this work, characterised in Fig.34 and Fig.35, with respect to MASP-3 gene structure. Upper panel - MASP-3 exon structure (exons 1 – 11). Lower panel - represented length of mouse MASP-3 RT PCR products and cDNA clones, with respect to MASP-3 exon structure.

DNA sequence alignment of the mouse MASP-3 cDNA coding sequence with the characterised human MASP-3 indicated an 86% identity (Clustal W method), alignment of the derived translational product from both human and mouse MASP-3 gave an identity of 88% (Stover, Lynch et al. 2003) (Fig.37).

Subsequent to the identification of the partial mouse MASP-3 specific sequence presented in this work, a full length cDNA clone was obtained by the research group of Professor Fujita, (accession number AB049755) (Takahashi and Shiro 2000) and entered in to the GenBank sequence database. Alignment of this cDNA clone revealed an identity of 100%, for the overlapping region including coding and non-coding sequence. Two independent research groups have identified and characterised mouse MASP-3 specific transcripts. The sequence derived from this work extends the 3' untranslated region of mouse MASP-3 by a further 456bp.



DNA sequence similarity of the full length mouse MASP-3 cDNA (AB049755) in comparison to the human MASP-3 was 86%; derived translational products exhibited an identity of 90% (AlignX, Invitrogen).

Alignment of human MASP-3 with obtained mouse MASP-3 sequence, underlying with mouse MASP-3 AB049755 cDNA sequence (Fig.37).

In common with rat and human MASP-3 there is conservation of active site serine protease residues histidine, aspartic acid and serine (highlighted in cyan Fig.37). Mouse MASP-3 active site serine utilises an AGY codon (Y denoting T or C) as exhibited by the previously characterised rat and human MASP-3 sequence, mouse MASP-1 retains an evolutionarily primordial active site serine codon, TCN (N denoting A, T, G or C) (Endo, Takahashi et al. 1998).

Mouse MASP-3 residue 225 conforms to the previously characterised rat and human MASP-3 sequence, consisting of a tyrosine residue (Krem and Di Cera 2001) (fig.37). Serine 214 an evolutionary marker of the chymotrypsin-like serine protease family (SA family) is conserved by mouse MASP-3 with a TCN codon, as exhibited by the previously characterised human and rat sequence.

An aspartate residue is present at position 198 in mouse MASP-3 as in rat and human MASP-3 sequence; a conserved glycine amino acid is also present in mouse MASP-3 at position 193 (highlighted in grey fig.37).

Mouse MASP-3 conserves a complement of 5 cysteine residues within the serine protease domain, as exemplified by the MASP-3 homologues examined in the work the histidine loop is absent.

The point of cleavage during serine protease activation within the linker region at residues arginine, isoleucine and isoleucine (RII) are conserved in mouse MASP-3 (bold and underlined fig.37).

Characterised mouse MASP-1/3 common region conserves all residues contributing to the interaction between CUBI-EGF domains involved in dimer formation (highlighted green and light green, fig.37), amino acids involved in binding to MBL and ficolin also exhibit conservation within the mouse MASP-1/3 sequence (highlighted red and dark red, fig.37).

Mouse MASP-1/3 conserves residues involved in the interaction between the CCPII and serine protease domain, amino acids PYY within the CCPII domain and PVC within the serine protease domain (highlighted pink fig.37).

Mouse MASP-3 serine protease domain predicts 3 N-glycosylation sites; this is conserved amongst the currently characterised MASP-3 presented in this work. The glycosylation state of the serine protease domain differs between alternative splice variants MASP-1 and MASP-3.

Mouse MASP-3 conserves the 11 exon structure encoding 6 protein domains, as exhibited by previously characterised human MASP-3 and rat MASP-3 derived from this work (Fig.38). Analysis of the available genomic sequence data, in conjunction with the previously characterised mouse MASP-1 confirms an identical MASP-1/3 loci structure (Stover, Lynch et al. 2003). The codon phase of mouse MASP-3 exons is conserved with that of human MASP-3, all exons of the MASP-1/3 locus utilise codon phase I splicing occurring between the first and second base of the spliced codon. The common MASP-1/3 domain structure CUBI-EGF-CUBII-CCPI-CCPII is encoded by 10 exons, mouse MASP-3 serine protease domain is encoded within a single exon. Mouse MASP-1 serine protease domain is encoded by the splicing of 6 exons. The mouse MASP-1/3 locus is located on chromosome 16, as determined by alignment with genomic sequence database clones. Mouse MASP-3 shares the common intron exon structure as that found in human MASP-3, the MASP-1/3 locus as a whole exhibits identical structure to that seen in the human genomic sequence.

# Genomic organisation of mouse MASP-1/3

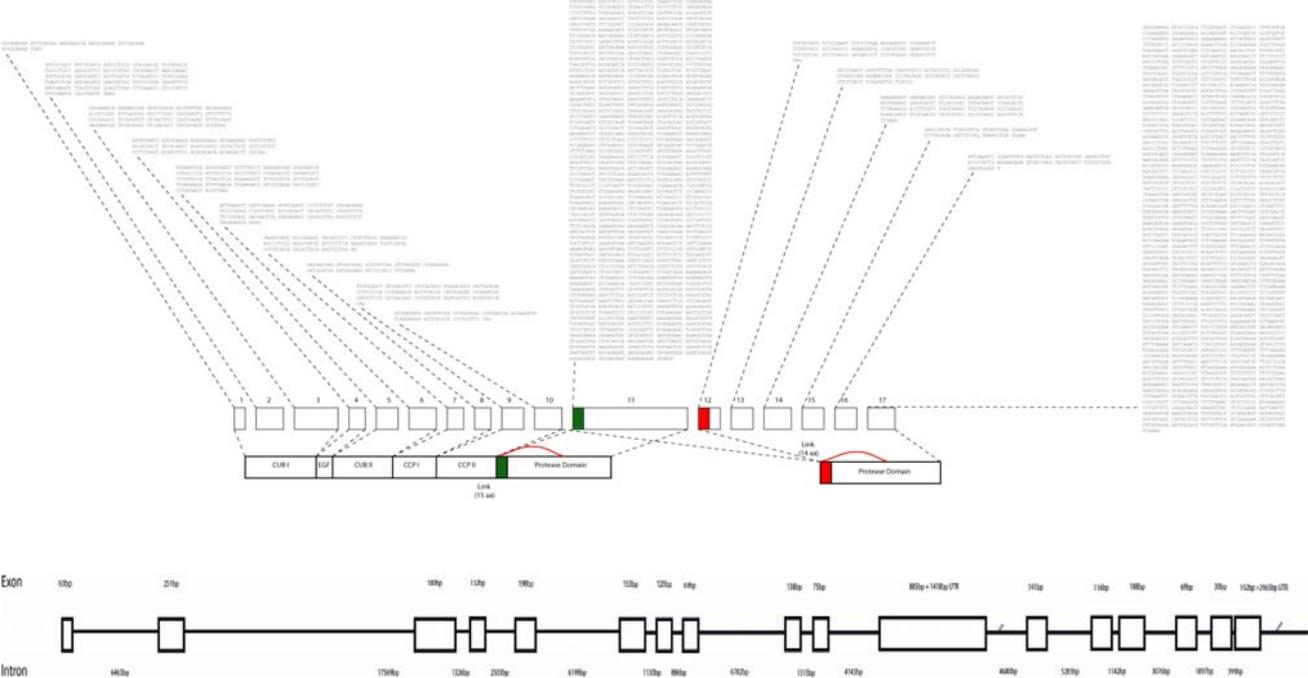


Fig.38 Mouse MASP-1/3 protein domain structure (lower panel), linked to respective exons of the mouse MASP-1/3 locus (central panel). cDNA sequence is presented in the upper panel associated with the respective exon.

Interrogation of available mouse genomic sequence databases (NCBI 2007) in conjunction with the mouse MASP-3 cDNA sequence obtained enabled characteristics of the mouse locus to be compared with the human MASP-1/3 counterpart. The mouse MASP-1/3 genomic region shares fundamental structure with that exhibited in human and rat species. A single polyadenylation site was predicted (PolyAH (Salamov and Solovyev 1997)) in the mouse MASP-3 3' untranslated region, 1463bp downstream of the stop codon.

Mouse MASP-3 exhibits a high degree of identity with the previously characterised human MASP-3 sequence, both at the cDNA and translated primary amino acid sequence. This global level of similarity can be further dissected to reveal domains of the mosaic protein with important conserved function, individual protein domains determined on the basis of genomic exon structure were aligned in a pair wise manner utilising the ClustalW method (AlignX, Invitrogen).

Analysis of the degree of sequence conservation between human and mouse MASP-1/3 is presented below (Table.3).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	88%	94%	85%	90%	82%	86%	82%
Protein	90%	100%	88%	89%	78%	91%	82%

Table.3 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and mouse MASP-1/3 cDNA (upper panel) and predicted translation products (lower panel).

Examining identity between human and mouse MASP-1 and MASP-3 serine protease domain, there is clearly greater conservation of MASP-3 specific amino acid sequence with respect to MASP-1. MASP-1/3 common domains exhibit a significant conservation; this is particularly marked within the CUBI and EGF domains.

Mouse MASP1/3 locus is present on chromosome 16B1 spanning approximately 71kb (Fig.39), the region is bound by Bcl6 (B-cell lymphoma 6 protein), Rtp2 (Receptor transporter protein 2), SST (Somatostatin) and Rtp4 (Receptor transporter protein 4). Upstream of the mouse MASP-1/3 loci are Rtp1 (Receptor transporter protein 1), ST6GAL (ST6 beta-galactosamide alpha-2, 6-sialyltransferase 1), BC106179 (predicted protein), LOC100043541 (hypothetical protein), EG625835 (predicted ribosomal protein), ADIPOQ (Adiponectin), RCF4 (Replication factor C subunit 4), EIF4A2 (eukaryotic translation initiation factor 4A, isoform 2), KNG1 (kininogen) and HRG (histidine-rich glycoprotein). Genes in linkage association in the mouse are conserved for rat and human species.



Fig.39 Mouse chromosome 16B1 indicating mouse MASP-1/3 locus with respect to upstream and downstream gene loci. Relative size, position and orientation of each loci indicated by arrow. Gene loci indicated in above text.

### 3.8 RT PCR cloning of rabbit (*Oryctolagus cuniculus*) MASP-3 cDNA

An RT-PCR cloning strategy was employed to identify a rabbit MASP-3 specific transcript; human MASP-3 specific primer pairs (hu MASP-3 p6 F and hu MASP-3 p8 R) were used under conditions of low stringency to amplify homologous regions of the rabbit cDNA. Total RNA prepared from rabbit liver was used as a template for first strand cDNA synthesis; reverse transcribed using an oligo dT primer. A predicted PCR product RabbitM3\_6/8 encompassing 687bp was generated (Fig.40, lane 3).

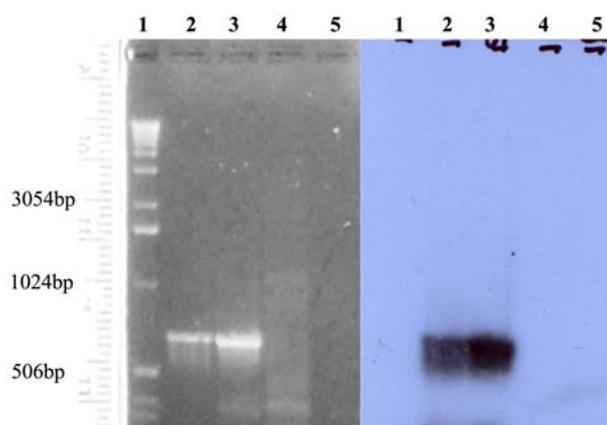


Fig.40 RT-PCR of rabbit cDNA utilising human MASP-3 specific primers. Agarose gel DNA electrophoresis of rabbit MASP-3 RT PCR utilising cDNA derived from rabbit total liver RNA (left panel), Southern blot hybridisation (right panel) of rabbit MASP-3 RT PCR agarose gel (left panel) utilising rat MASP-3 specific cDNA probe (rat MASP-3 PCR p6F + p8F). Lane 1 - 1kb DNA marker, lane 2 human MASP-3 RT PCR positive control (Hu MASP-3 p6F + p8F), lane 3 - rabbit MASP-3 RT-PCR (Hu MASP-3 p6F + p8F), lane 4 - rabbit cDNA negative control and lane 5 - PCR dH<sub>2</sub>O control.

PCR products obtained in this way were analysed by southern blot and hybridised with a rat MASP-3 specific cDNA probe, specific amplification products were identified. PCR products were subcloned into the pGEM T-Easy vector (promega, CA), stabilising the fragments for further sequence analysis. DNA sequence data of the putative rabbit MASP-3 specific PCR amplification product (RabbitM3\_6/8) indicated a region of novel sequence spanning 687bp, a BLAST (Altschul, Gish et al. 1990) alignment of human MASP-3 (accession number AF284421) with the potential rabbit MASP-3 confirmed the fragment to be a homologue of human MASP-3 with an E-value of 0.0. This is the first characterisation of MASP-3 sequence specific to the rabbit species, the

fragment encoding a predicted translation product of 229aa entirely within the serine protease domain.

### 3.9 Rabbit MASP-3 5' RACE-PCR

Extension of rabbit MASP-3 sequence was achieved in a 5' direction using a rapid amplification of cDNA ends (RACE) approach, enabling a conclusive link to be formed between the MASP-3 serine protease domain and the MASP-1/3 common sequence. Specific oligonucleotides were designed based on the previously obtained rabbit MASP-3 cDNA sequence. Total RNA was prepared from rabbit liver and used as a template for first strand cDNA synthesis, utilising a rabbit MASP-3 gene specific primer (Rb MASP-3 GSPI), a polymeric dCTP tail was added at the 3' end of the newly formed cDNA. Specific amplification was achieved utilising a downstream rabbit MASP-3 gene specific primer (Rb MASP-3 GSPII) and an abridged anchor primer (AAP) binding within the poly-dCTP tail.

A specific PCR product of approximately 700bp was obtained utilising this approach; these were subcloned into the TA cloning vector pGEM-T Easy for further analysis. Sequence data obtained from a number of clones was used to generate a contiguous cDNA sequence (Vector NTI, Contig express) with the previously characterised rabbit MASP-3 partial sequence clone rabbitM3\_6/8.

A rabbit MASP-3 specific contig was generated utilising sequence data obtained from the two experimental approaches; a cDNA contig of 1293bp was constructed.

Utilising the sequence information it was possible to conclusively link the rabbit MASP-3 serine protease domain to the rabbit MASP-1/3 common domains.

The PCR fragments generated did not achieve complete coverage of the MASP-1/3 common domains, but originated in the CUB II domain starting DHPEVP. Contained within the data obtained were all serine protease active site codons.

### 3.10 Rabbit MASP-3 3' RACE-PCR

Efforts were made to extend in a 3' direction, the rabbit MASP-3 sequence information obtained via the previous experimental approaches. A 3' RACE strategy was employed utilising total RNA derived from rabbit liver as a template, nested gene specific primers (GSP) primers were designed within the rabbit MASP-3 serine protease domain with reverse amplification via a standard oligo-dT primer. Despite several rounds of optimisation of PCR conditions and utilisation of alternate RNA preparations, it was not possible to obtain a suitable fragment for further analysis of rabbit specific sequence.

### 3.11 Rabbit MASP-3

Rabbit MASP-3 sequence data was derived from an RT-PCR methodology; overlapping rabbit MASP-3 fragments were aligned to generate a single contiguous sequence (accession number AJ457085).

Utilising the novel rabbit MASP-3 specific cDNA sequence to interrogate the current rabbit genomic sequence database (Broad\_Institute 2008) using the BLAST algorithm (Altschul, Gish et al. 1990), it was possible to identify a partially assembled genomic contig (AAGW01550346) containing rabbit MASP-3 exon sequence. Using this sequence information it was possible to predict the remaining MASP-3 serine protease domain sequence.

A translated identity database search (Tblastx) of the identified 18188bp rabbit whole genome shotgun (WGS) clone (AAGW01550346) using the partial rat MASP-3 (accession number AJ487622) sequence derived from this work, identified potential rabbit MASP-3 specific sequence information contiguous with that of the partial rabbit MASP-3 (accession number AJ457085) sequence presented in this work. Rabbit MASP-3 specific sequence information was extended by 168bp in the 3' direction.

The methodology was successful in identification of the rabbit MASP-3 specific cDNA sequence, and allowed predicted sequence information to be derived from public genomic DNA database. The complete predicted cDNA sequence could not be elucidated, due to the limitations of the genomic sequence database in both sequence coverage and stage of assembly.

Multiple sequences were aligned (ContigExpress, Invitrogen) to generate reliable sequence with a total length of 1638bp, including an open reading frame of 545aa.

A BLAST (Altschul, Gish et al. 1990) alignment analysis rabbit MASP-3 contig against human MASP-3 (accession number AF284421) confirmed the rabbit contig to be an ortholog of human MASP-3 with an E-value of 0.0.

Sequence analysis revealed the rabbit MASP-3 partial sequence encompassing the shared MASP-1/3 CUBII, CCPI and CCPII domains and a complete MASP-3 serine protease domain terminating in an ATG stop codon. The overlapping cDNA sequences used to generate the rabbit MASP-3 contig conclusively demonstrate the association of the MASP-3 serine protease domain and the MASP-1/3 common domains in the rabbit.

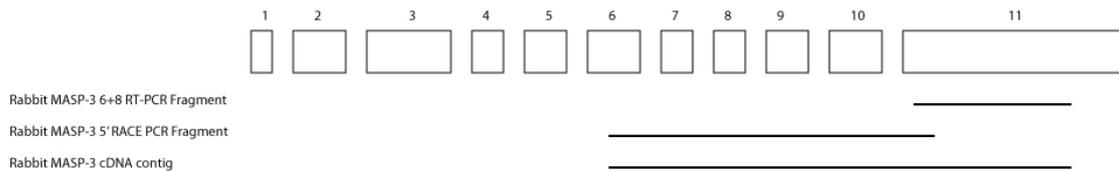


Fig.41 Schematic diagram of rabbit MASP-3 RT PCR products (Fig. 40) derived from this work (lower panel), with respect to MASP-3 exon structure (upper panel).

Figure.41 presents cDNA sequence obtained in this work, aligning them with respect to the MASP-3 exon structure. All cDNA sequence originates within the rabbit MASP-3 specific domain; RACE-PCR enabled extension of the sequence into the shared rabbit MASP-1/3 confirming an association between MASP-3 domain and the shared MASP-1/3 domains. Rabbit MASP-3 sequence does not encompass the entire MASP-3 sequence, as predicted from the previously characterised human, rat and mouse cDNA sequence.

DNA alignment of the rabbit MASP-3 cDNA sequence with the known human MASP-3 clone indicated an overall 89% identity (ClustalW method); alignment of the translated cDNA from both human and rabbit partial sequence gave an identity of 89%.



Rabbit MASP-3 serine protease domain conserves residues aspartate 198 and glycine 193 (highlighted in grey fig.42), consistent with previous MASP-3 sequences within this work.

Rabbit MASP-3 serine protease domain encompasses five cysteine residues, these are not correctly organised for the formation of a histidine loop structure.

Due to the partial rabbit MASP-3 sequence information available, particularly relating to the shared MASP-1/3 region, it is not possible to determine conservation of residues crucial to the interaction between homodimers, MBL and ficolin.

Comparison of the partial sequence information obtained within this work with the currently available genomic specific database indicates that the rabbit MASP-3 cDNA obtained encodes a total of 6 complete exons. The genomic structure of the rabbit MASP-3 partial cDNA sequence indicated an intron/exon utilisation consistent with the previously presented sequence in this study for the homologous regions.

The partial rabbit MASP-3 sequence obtained encodes a CUBII-CCPI-CCPII-SP (partial) domain architecture, encompassed within 6 complete exons. This is consistent with the previously analysed sequence information.

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	n/a	n/a	n/a	91%	69%	90%	n/a
Protein	n/a	n/a	n/a	92%	71%	94%	n/a

Table.4 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and rabbit MASP-1/3 cDNA (upper panel) and predicted translation products (lower panel). Sequence unavailable to complete alignment indicated by (n/a).

There is limited sequence information available to determine identity between human and rabbit MASP-1/MASP-3. Rabbit MASP-3 serine protease domain exhibits significant identity to human MASP-3 at the primary amino acid level (table.4).

Analysis of the rabbit MASP-3 predicted translation product against a database of known protein modules (Pfam) identified the conserved structural components of the newly characterised rabbit MASP-3 cDNA sequence. The partial sequence obtained clearly demonstrates the conserved structural motifs exemplified by the human MASP-3 predicted translation product, CUBII-CCPI-CCPII-Serine protease domain (E-values  $3^{-2}$ ,  $2^{-8}$ ,  $6^{-10}$ ,  $2^{-76}$ ). The domain pattern is also conserved amongst the other MASP-3 homologues presented in this work (Fig.43).



Fig.43 Rabbit MASP-3 protein domain structure predicted by comparison to the Pfam database, using Pfam analysis of rabbit MASP-3 primary amino acid sequence (Fig.42). Domains with maximal identity to Pfam consensus sequence are represented, indicating an identical architecture to that exhibited by human MASP-3 (Fig.6) for the available sequence obtained.

### 3.12 RT PCR cloning of guinea pig (*Cavia porcellus*) MASP-3 cDNA

An RT-PCR strategy was employed to obtain guinea pig specific MASP-3 cDNA sequence; human MASP-3 specific primer pairs were used under conditions of low stringency to amplify homologous regions of the guinea pig gene. Guinea pig total liver RNA was reverse transcribed using an oligo-dT primer, to provide a specific cDNA template.

PCR products obtained in this way were southern blot hybridised with a rat MASP-3 specific probe, identifying specific amplification products. PCR products were subcloned into the pGEM T-easy vector (promega, CA), individual clones were selected by southern blotting using a rat MASP-3 specific probe positive candidates selected for sequencing.

A guinea pig specific cDNA clone of 688bp was characterised, comparison of the partial sequence with those MASP-3 sequences obtained in prior experiments indicated homology entirely within the MASP-3 serine protease domain. This is the first confirmation of a MASP-3 specific transcript within the guinea pig species. Predicted translation product indicated a 229aa open reading frame, Pfam analysis (Sonnhammer, Eddy et al. 1997) conformed serine protease domain homology.

The partial guinea pig MASP-3 serine protease sequence obtained provided the basis for primer design to further extend the sequence information towards the shared MASP-1/3 common region and towards the termination codon and 3' end of the cDNA transcript.

### 3.13 Guinea pig MASP-3 5' RACE-PCR

Extension of guinea pig MASP-3 sequence was achieved in a 5' direction using a rapid amplification of cDNA ends (RACE) approach, enabling the acquisition of novel sequence data. Specific oligonucleotides were designed based on the previously obtained guinea pig cDNA sequence. Guinea pig liver total RNA was used as a template for first strand cDNA synthesis, utilising a guinea pig MASP-3 gene specific primer (gp MASP-3 GSPI), and a polymeric dCTP tail was added at the 3' end of the newly formed cDNA. Specific amplification was achieved utilising a downstream guinea pig MASP-3 gene specific primer (gp MASP-3 GSPII) and an abridged anchor primer (AAP) binding within the poly-dCTP tail (Fig.44).

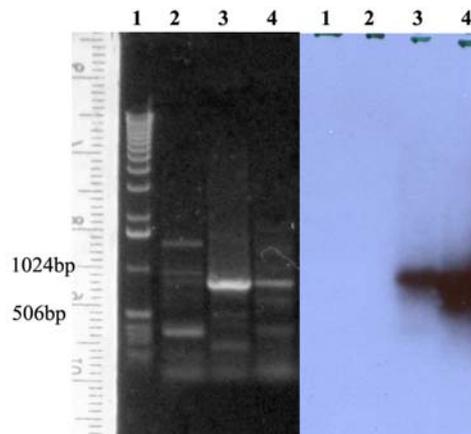


Fig.44 Guinea pig 5' RACE PCR utilising guinea pig MASP-3 gene specific primers (gp MASP-3 GSPI, gp MASP-3 GSPII). Agarose gel electrophoresis of guinea pig MASP-3 5' RACE PCR utilising cDNA derived from guinea pig total liver RNA (left panel), Southern blot hybridisation (right panel) of guinea pig MASP-3 5' RACE PCR agarose gel (left panel) utilising rat MASP-3 specific cDNA probe (rat MASP-3 PCR p6F + p8F). Lane 1 – 1kb DNA marker, lane 2 – negative control, lane 3 - Guinea pig MASP-3 5' RACE PCR, lane 4 – human MASP-3 5' RACE PCR positive control.

Restriction endonuclease digestion (EcoRI) of potential guinea pig MASP-3 5'RACE clones identified a range of fragments; to conclusively identify MASP-3 specific sequence a southern blot was performed. Southern blot hybridisation of guinea pig MASP-3 5' RACE clones with a rat MASP-3 specific clone showed a degree of cross hybridisation with vector DNA, a clear binding of a released fragment from clone 15 could be observed. Guinea pig MASP-3 5' RACE clone 15 was analysed by DNA sequencing, revealing a cDNA sequence of c.1kb overlapping with the previously characterised guinea pig MASP-3 sequence derived from this work. A contiguous

cDNA sequence specific to guinea pig MASP-3 was constructed, comprising 1.5kb and conclusively linking the MASP-3 specific serine protease domain to the shared MASP-1/3 domains.

### 3.14 Guinea pig MASP-3 3' RACE-PCR

Extension of the guinea pig MASP-3 sequence in a 3 prime direction was accomplished using mouse liver RNA template; first strand synthesis was initiated at the poly-A tail using the adaptor primer (AP) and the 3'RACE system for rapid amplification of cDNA ends (RACE) system (Gibco, CA). 3'RACE was performed utilising a touchdown PCR cycle, further optimisation of the PCR reaction failed to obtain a single clear band, identification of specific guinea pig MASP-3 3' RACE products was achieved by southern blotting and hybridisation. A human MASP-3 specific probe partially overlapping with the expected guinea pig MASP-3 3' RACE PCR products, hybridised on southern blotting with a single band of approximately 900bp. Guinea pig MASP-3 3' RACE PCR products were cloned into the TA cloning vector pGEM-T easy, analysed by DNA sequencing.

Guinea pig MASP-3 sequence obtained in this way formed a contig of 808bp, encompassing a region of the MASP-3 gene including 240bp of the serine protease domain and extending into the 3' untranslated region. Guinea pig 3' RACE cDNA sequence was used to generate a specific guinea pig MASP-3 contig.

### 3.15 Guinea pig MASP-3

Overlapping guinea pig MASP-3 sequences were aligned to generate a single cDNA contig (accession number AJ457086), encompassing the entirety of the available cDNA sequence. Multiple sequences were aligned to generate reliable sequence with a total length of 2181bp, including a reading frame of 523aa.

Sequence analysis revealed the guinea pig specific partial sequence contained a MASP-3 serine protease domain, and the shared MASP-1/3 CUBII, CCPI and CCPII domains.

Analysis of the available guinea pig genomic DNA sequence identified genomic scaffold sequence (7:52107348-52152008) containing MASP-3 specific sequence information, utilising a translated homology search it was possible to identify the remaining guinea pig specific exon sequence.

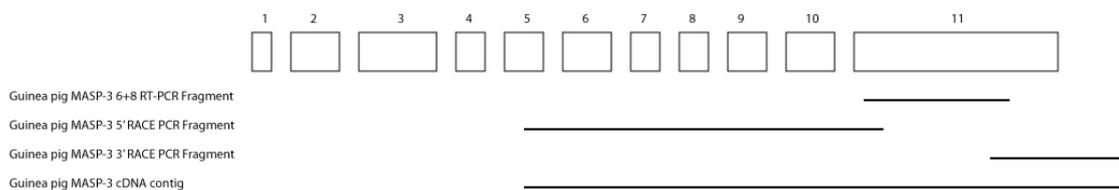


Fig.45 Schematic diagram of guinea pig MASP-3 specific cDNA clones derived from this work (lower panel), aligned with respect to MASP-3 exon structure (upper panel).

The overlapping cDNA sequences used to generate the guinea pig MASP-3 contig conclusively demonstrate the association of the MASP-3 serine protease domain and the MASP-1/3 domains in the guinea pig (Fig.45).



Guinea pig MASP-3 translated cDNA sequence retains a tyrosine residue at position 225 and serine residue 214 (Ser214) is encoded by a TCN (TCT) codon (highlighted grey fig.46), this combination is consistent with that observed in the previously characterised MASP-3 in this study and represents evolutionary conservation.

MASP-3 serine protease activation site is conserved within the guinea pig cDNA translation product; arginine-isoleucine-isoleucine (RII residues highlighted in bold and underlined fig.46) motif is evident as are the CCPII-Serine protease domains coordinating residues (highlighted pink fig.46) responsible for stabilisation of the interaction between domains.

Guinea pig MASP-3 encompasses five cysteine residues with the protease domain, a configuration unable to form a histidine loop structure. The loss of the stabilising effect of the histidine loop structure is an evolutionarily recent event, the structure remains in MASP-1 providing support to its antecedent nature.

MASP-1/3 common region of guinea pig translated cDNA product conserves all previously described residues vital for the formation of MASP dimers (highlighted in green and light green fig.46) and those amino acids required for interaction with both MBL and ficolin (highlighted red and dark red fig.46). This is consistent with the previously characterised MASP-3 sequence within this work.

The partial guinea pig MASP-3 sequence obtained encodes a CUBII-CCPI-CCPII-SP domain architecture, encompassed a total of 6 complete exons, with partial sequence from exon 5. Comparison of the partial sequence information obtained within this work with the currently available genomic specific database enabled the identification of the remaining exons, confirming a CUBI-EGF-CUBII-CCPI-CCPII-SP domain structure consistent with the previously analysed sequence information. Predicted cDNA sequence based on the available genomic DNA identity was able to identify the remaining exons, a total of 11 were identified. The genomic structure of the guinea pig MASP-3 partial cDNA sequence and that derived from genomic sequence databases, indicated an intron/exon utilisation consistent with the previously presented sequence in this study.

DNA alignment of the guinea pig MASP-3 cDNA sequence with the previously characterised human MASP-3 clone indicated an overall 87% identity (Clustal W method); alignment of the translated cDNA from both human and guinea pig gave an identity of 90% (Fig.46).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	86%	92%	88%	87%	84%	88%	82%
Protein	94%	100%	89%	89%	81%	90%	78%

Table.5 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and guinea pig MASP-1/3 cDNA (upper panel) and predicted translation products (lower panel).

Guinea pig MASP-1 and MASP-3 serine protease domains exhibit a high degree of identity to the respective human homolog; MASP-3 is conserved with respect to MASP-1 at the primary amino acid level. MASP-1/3 common domains CUBI and EGF also demonstrate a high degree of conservation in comparison to the remaining shared domains.

Analysis of the guinea pig MASP-3 predicted translation product utilising a database of known protein modules (Pfam) identified the conserved structural components of the newly characterised guinea pig MASP-3 cDNA sequence (Fig.47). The primary amino acid sequence obtained clearly demonstrates the conserved structural motifs exemplified by the human MASP-3 predicted translation product. Addition of the predicted cDNA sequence based on genomic sequence homology, confirms the 5' domain architecture. Identity to protein domain consensus sequences for the domains CUBI-EGF-CUBII-CCPI-CCPII-Serine protease domain are, (E-value)  $2^{-21}$ , 2,  $5^{-47}$ ,  $1^{-10}$ ,  $7^{-8}$  and  $3^{-74}$  respectively.

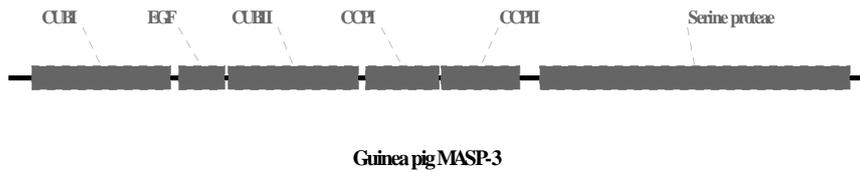


Fig.47 Guinea pig MASP-3 protein domain structure predicted by comparison to the Pfam database, using Pfam analysis of rabbit MASP-3 primary amino acid sequence (Fig.46). Domains with maximal identity to Pfam consensus sequence are represented, indicating an identical architecture to that exhibited by human MASP-3 (Fig.3).

This same structural architecture is also conserved amongst the other MASP-3 homologues presented in this work. Comparison of the guinea pig MASP-3 sequence derived from this work against an independent database of protein domain consensus sequences is further confirmation of the validity of the cDNA sequence obtained, in conjunction with the high degree of identity to known MASP-3 cDNA sequence and predicted translation products the architecture of the mosaic protein is also conserved (fig.47).

A comprehensive genomic map of the guinea pig MASP-1/3 region is not currently available to enable the surrounding genomic structure to be determined in this species.

### 3.16 Porcine (*Sus scrofa*) MASP-3

Analysis of the contemporary cDNA sequence databases utilising a translated homology search, for sequence resembling the rat MASP-3 specific sequence identified a number of porcine expressed sequence tags (EST's). EST clone BE030550 was obtained and restriction EcoRI/XbaI endonuclease digest from the parent vector pCMV-SPORT 6, two fragments of 500bp and 1.2kb were released (Fig.48).

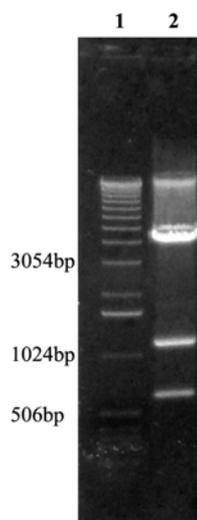


Fig.48 DNA agarose gel electrophoresis of porcine EST clone (BE030550) restriction endonuclease (EcoRI/XbaI) digestion. Lane 1 - 1kb DNA marker, lane 2 - porcine EST clone BE030550 EcoRI/XbaI restriction endonuclease digest.

A sequencing strategy was devised to obtain the entire EST cDNA clone (accession number BE030550) sequence. A contig of 1044bp was constructed from overlapping sequence using the ContigExpress (Invitrogen) software tool. The porcine contig contained an open reading frame of 564bp terminating in a TGA stop codon, encompassing 188aa specific to porcine MASP-3 serine protease (fig.48).

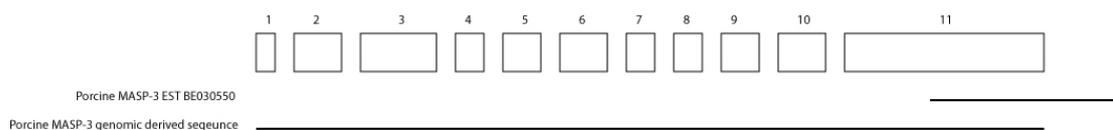


Fig.48 Schematic diagram of porcine MASP-3 cDNA sequence (lower panel) derived from this work. Porcine MASP-3 sequence derived from porcine genomic sequence database is indicated by the lower line. MASP-3 exon structure with respect to the derived cDNA sequence is represented in the upper panel.

Utilising the entire sequence information derived from the MASP-3 porcine cDNA clone, further analysis of the current sequence database identified a predicted cDNA sequence (XM\_001927110) derived from genomic sequence information LOC100152125. The predicted porcine MASP-3 cDNA sequence contained the entire MASP-3 sequence information (fig.48).

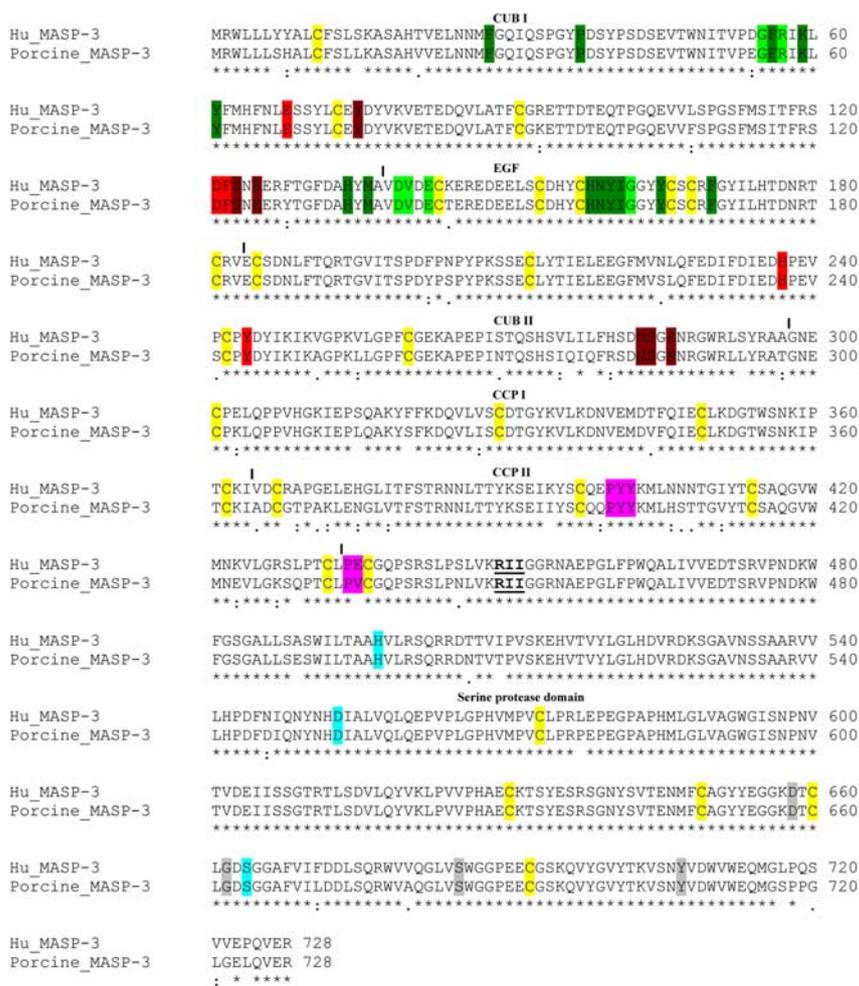


Fig.49 ClustalW 2.0.11 multiple sequence alignment of translated porcine MASP-3 (contig derived from this work) against human MASP-3 coding sequence (accession number AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCP II-SP interaction.

Porcine MASP-3 derived from genomic sequence encompassed an open reading frame of 2187bp, encoding 728aa. DNA alignment of porcine MASP-3 with the previously characterised human MASP-3 sequence identified an 89% identity (Clustal W method); alignment of the predicted cDNA translation products indicated a homology of 92%.

Analysis of the porcine MASP-3 predicted translation product (fig.49) confirms the conservation of the serine protease catalytic residues histidine, aspartic acid and serine. The active site serine residue at position 195 utilises an AGY (Y denoting T or C) codon, consistent with the previously characterised MASP-3 sequence in this work (Krem and Di Cera 2001). The use of the AGY codon at this position is evolutionarily recent in comparison to that exhibited by porcine MASP-1, which retains the evolutionarily distant TCN (N denoting A, T, G or C) active site codon.

Porcine MASP-3 predicted translation product in common with the previously characterised human MASP-3 sequence conserves residues within the serine protease but outside of the catalytic apparatus that influence interaction with the substrate, there is conservation of an aspartic acid residue at position 189 and a glycine residue at position 193 (highlighted grey in fig.49).

Residues at position 214 and 225 exhibit conservation with respect to human MASP-3 sequence (highlighted grey in fig.49), encoding a serine utilising a TCN codon and tyrosine residue respectively, maintenance of these residues confirms a common evolutionary origin.

Porcine MASP-3 predicted primary amino acid sequence exhibits an activation cleavage site encompassing residues arginine-isoleucine-isoleucine (highlighted RII, bold and underlined fig.49), amino acids responsible for interaction between the serine protease domains and the preceding CCPII domain (highlighted in pink fig.49) are conserved.

Within the serine protease domain and in common with the previously characterised MASP-3 sequence are 5 cysteine residues (highlighted yellow fig.49), indicating that the formation of a histidine loop structure is not possible as exhibited by MASP-1 serine protease.

All amino acid residues contributing to the formation of MASP dimer-dimer association are conserved in porcine MASP-3, mediated via interactions between opposing CUBI-EGF domains (highlighted green and dark green fig.49). Amino acids involved in the interactions with recognition molecules MBL and ficolin are also conserved (highlighted red and dark red fig.49), indicating the potential to form associations with these proteins is conserved within this species.

Porcine MASP-3 has the 11 exon structure of the previously characterised MASP-3 loci in this work, encoding 6 protein domains. Analysis of the available porcine genomic sequence information, utilising further BLAST homology searches to identify porcine MASP-1 specific sequence has revealed an identical MASP-1/3 loci structure to that exhibited by the previously characterised loci in this work. Porcine MASP-1/3 intronic and exonic structure is congruent with the previously characterised human MASP-1/3 loci (Dahl, Thiel et al. 2001). Porcine MASP-3 serine protease domain is encoded by a single exon; porcine MASP-1 serine protease domain is encoded within 6 exons. MASP-1/3 common region is encoded by 10 exons.

The percentage identity for individual MASP domains between human and porcine sequence, as determined by ClustalW alignment is presented below (table.6).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	91%	95%	88%	90%	84%	89%	87%
Protein	94%	98%	88%	92%	77%	95%	84%

Table.6 Pairwise identity as determined by ClustalW alignment of MASP-1/3 functional domains between human and porcine MASP-1/3 cDNA (upper panel) and predicted translation products (lower panel).

Analysis of the shared identity between human and porcine predicted primary amino acids sequence indicates significant conservation of porcine MASP-3 serine protease domain with respect to that of MASP-1 (table.6), this level of conservation is also exhibited by CUBI and EGF domains as previously demonstrated by MASP-3 sequence derived from this work.

Porcine MASP-1/3 locus is present on chromosome 13 spanning approximately of 55kb (Fig.50), the region is bound by Alpha-2-HS-glycoprotein precursor (Fetuin-A), Receptor transporter protein 1 (Rtp1) and Receptor transporter protein 4 (Rtp4).

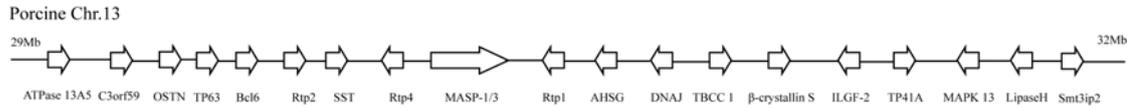


Fig.50 Porcine chromosome 13 indicating the porcine MASP-1/3 locus with respect to upstream and downstream gene loci. Relative size, position and orientation of each loci indicated by arrow. Gene loci indicated in above text.

The genomic region of the porcine MASP-3 locus exhibits a high degree of similarity with the previously characterised MASP-3 loci in this work; genes exhibit a high degree of linkage between species.

### 3.17 Bovine (*Bos taurus*) MASP-3

Analysis of the contemporary sequence databases utilising MASP-3 specific sequence information generated by this work, identified a number of bovine specific EST clone AW558440. Expressed sequence tag (EST) clones were obtained and a sequencing strategy devised to complete the sequence information. Restriction endonuclease digestion of the cDNA clone from the parent vector pCMV-SPORT 6, utilising EcoRI/XbaI released a fragment of approximately 2kb.

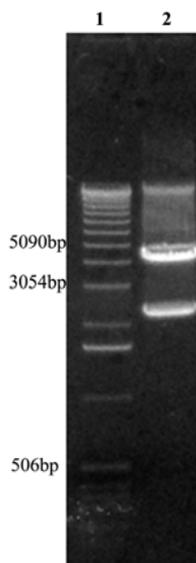


Fig.51. DNA agarose gel electrophoresis of bovine EST clone (AW558440) restriction endonuclease (EcoRI/XbaI) digestion. Lane 1 - 1kb DNA marker, lane 2 - bovine EST clone AW558440 EcoRI/XbaI restriction endonuclease digest.

A sequencing strategy was devised to obtain the entire EST cDNA clone sequence. Sequence derived from the EST AW558440 corresponded to the 3' end of the MASP-3 serine protease domain, including a termination codon. Limited sequence obtained from the cDNA clone using the vector specific sp6 primer, identified DNA sequence corresponding to the first 2 exons (fig.52). Analysis of the available genomic sequence utilising both homology for previously characterised MASP-3 sequence and sequence obtained from the EST clone AW558440, enable the identification of the entire Bovine MASP-3 cDNA sequence. Utilising the entire sequence information derived from the cDNA clones (fig.52), further analysis of the current sequence database including the expanding genomic sequence data allowed the location of the bovine MASP-1/3 locus to chromosome 1.

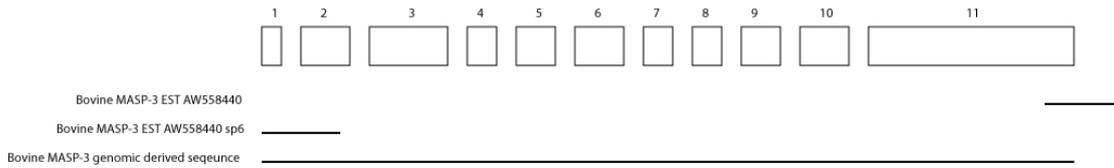


Fig.52 Schematic diagram of partial bovine MASP-3 sequence obtained in this work (lower panel), genomic sequence database derived bovine sequence is indicated by the lower line (lower panel). MASP-3 exon structure is represented in the upper panel, with respect to the derived cDNA sequence length.

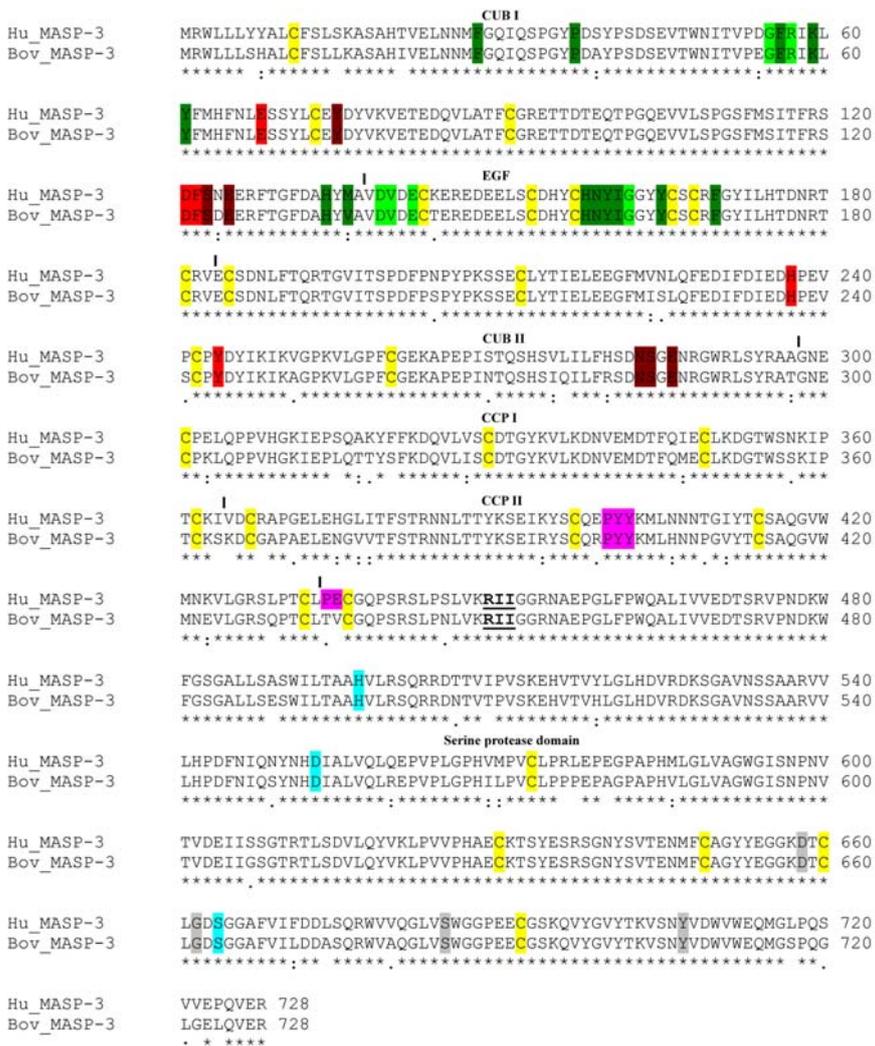


Fig.53 ClustalW 2.0.11 multiple sequence alignment of translated bovine MASP-3 (contig derived from this work) against human MASP-3 coding sequence (accession number AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCPII-SP interaction.

Bovine MASP-3 conserves all amino acid residues of the active site triad in comparison to human MASP-3 (highlighted cyan fig.53), the active site serine residue 195 utilises an AGY codon (Y denoting T or C) this is conserved within all MASP-3 sequence within this work.

All markers of chymotrypsin clad SA are conserved between human and bovine sequences, with respect to amino acid residue and codon usage (Krem and Di Cera 2001) at positions aspartic acid 189, glycine 193, tyrosine 225 and serine 214 which utilises a TCN (N denoting A, T, G or C) codon (highlighted grey fig.53).

Analysis of the complement of cysteine residues within the protease domain demonstrates the inability to form a histidine loop structure, due to the lack of a sixth residue. This conformation is shared by all characterised MASP-3 in this work, and places the serine protease as evolutionarily descendent of the MASP-1 protease that requires the histidine loop for stability within its structure.

Bovine MASP-3 contains all residues contributing to MASP dimer formation mediated by interactions between opposing CUBI-EGF domains (highlighted green and dark green fig.53), also conserved are amino acids contributing to the binding of MBL and ficolin (highlighted reed and dark red fig.53).

Amino acids residues responsible for the stabilisation of the association between CCPII and serine protease domains are preserved within the CCPII domain (highlighted pink fig.53) but absent within the bovine MASP-3 protease domain, with the potential for alteration in the flexibility between domains and modulation of substrate access to the catalytic site.

Alignment of human and bovine genomic derived sequence at the cDNA level indicates an 89% similarity. Alignment of the predicted translation products indicates an overall identity of 91% at the amino acid level.

Sequence homology for each protein domain with respect to the human MASP-1/3 cDNA and translation products are presented below (table.7).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	91%	94%	90%	92%	88%	87%	n/a
Protein	94%	98%	91%	88%	82%	92%	n/a

Table.7 Identity between human and bovine MASP-1/3 protein domains as determined by ClustalW pairwise sequence alignment. Percentage identity between predicted translation products (lower panel) and cDNA (upper panel). Sequence unavailable to complete alignment indicated by (n/a).

Bovine MASP-3 serine protease domain demonstrates a high degree of identity with the previously characterised human MASP-3 primary amino acids sequence. MASP-1/3 common domains CUBI and EGF also exhibit a high degree of sequence conservation between human and bovine species, highlighting the importance of these domains in both MASP dimer formation and interactions with lectin pathway recognition molecules (table.7).

The bovine MASP-1/3 locus is located on chromosome 1 (WGS AAFC02084419), it retains an identical MASP-1/3 structure to that exhibited by the preceding species studied in this work.

Bovine MASP-1/3 genomic region contains a conserved complement of flanking gene loci to those exhibited by the previously studied species (Fig.54).

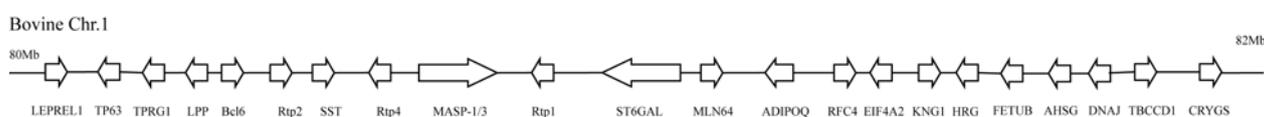


Fig.54 Bovine chromosome 1 indicating the porcine MASP-1/3 locus with respect to upstream and downstream gene loci. Relative size, position and orientation of each loci indicated by arrow. Gene loci indicated in above text.

### 3.18 *Xenopus* (*Xenopus tropicalis*) MASP-3

Analysis of the contemporary sequence database using a BLAST (Altschul, Gish et al. 1990) search, identified *Xenopus* EST sequence homologous to rat MASP-3 specific sequence derived from this work. An EST clone was identified using this method, *Xenopus tropicalis* clone (BG) BG4461412 was characterised.

*Xenopus tropicalis* EST clone BG encompassed 1027bp of cDNA sequence; recent homology database searches have identified a significant match to a hypothetical sequence (NM001097405). The *Xenopus tropicalis* EST clone BG4461412 sequence closely matches *Xenopus laevis* (mannose-binding lectin-associated serine protease) MASP-3a (NM\_001088873) (Strausberg, Feingold et al. 2002; Endo, Nonaka et al. 2003), and MASP-3b (NM\_001088871).

The genomic localisation of xenopus MASP-3 (ENSXETT00000042781) is not available due to the incomplete assembly of current sequence information; the locus is present on scaffold 81. A translated homology search (TBlastx) using sequence derived from this work was able to characterise the genomic structure of xenopus MASP-1/3 locus, in conjunction with the characterised xenopus MASP-1 (D83276) specific sequence (Endo, Takahashi et al. 1998; Endo, Nonaka et al. 2003). *Xenopus* MASP-1/3 locus exhibits an identical structure to that seen in mammalian species; MASP-1 and MASP-3 are alternative splice products originating from a single locus whereby a shared common MASP-1/3 heavy chain is spliced to a MASP-1 or MASP-3 specific serine protease domain. *Xenopus* MASP-1 serine protease domain is encoded by multiple exons, whilst MASP-3 serine protease domain is contained within a single exons structure.

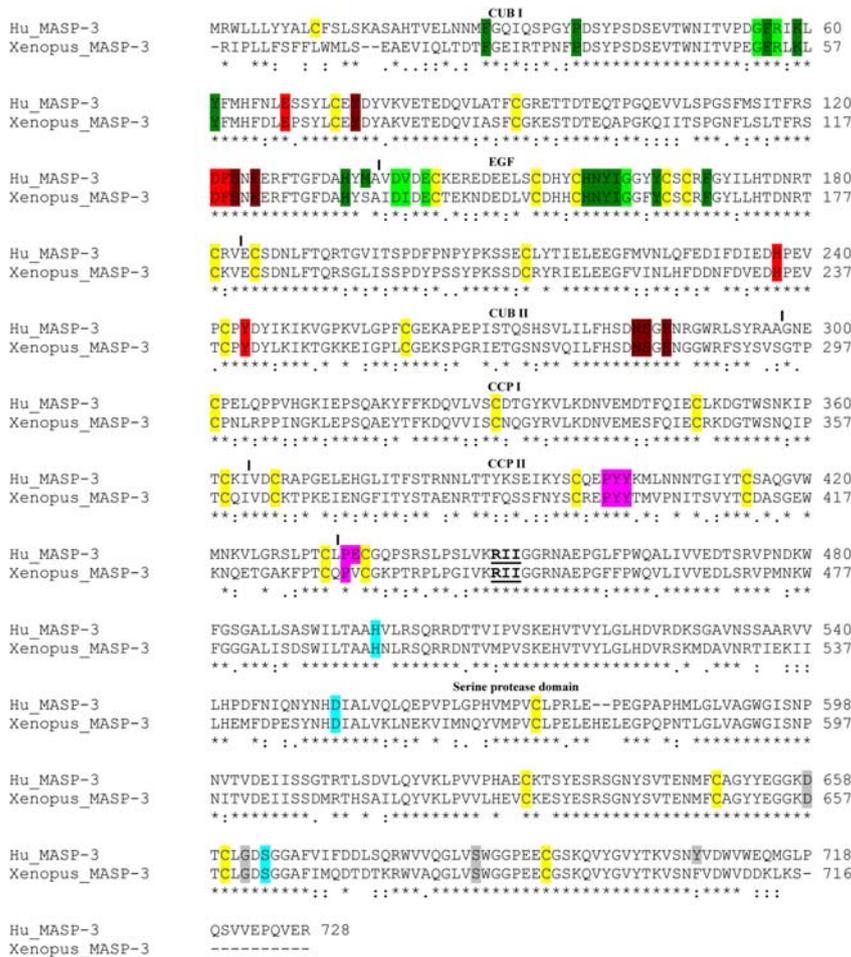


Fig.55 ClustalW 2.0.11 multiple sequence alignment of translated Xenopus MASP-3 (ENSXETT00000042781) against human MASP-3 coding sequence (accession number AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCPII-SP interaction.

Xenopus MASP-3 (ENSXETT00000042781) exhibits complete conservation of the serine protease active site residues histidine, aspartic acid and serine, with the serine 195 residue encoded by an AGY codon (highlighted cyan fig.55). Amino acid residues at positions 189 and 193 are conserved aspartic acid and glycine residues respectively, intimately involved in the interaction with a potential substrate. Serine 214 is encoded by a TCN codon as described in previously characterised MASP-3. A conserved tyrosine residue at position 225 in MASP-3 characterised in this work, is substituted for a phenylalanine residue in the genomic derived xenopus MASP-3 sequence.

Xenopus MASP-3 contains all vital amino acids contributing to MASP dimer formation (highlighted red and dark red fig.55), those interacting with MBL and ficolin to form a lectin pathway activation complex (highlighted green and dark green fig.55) and residues involved in interaction between CCP II the serine protease domain (highlighted pink fig.55).

In common with the MASP-3 currently investigated the serine protease domain encompasses 5 cysteine residues, indicating an inability to form a histidine loop structure as seen in evolutionarily antecedent serine protease (highlighted yellow fig.55).

	<b>CUB I</b>	<b>EGF</b>	<b>CUB II</b>	<b>CCP I</b>	<b>CCP II</b>	<b>MASP-3 SP</b>	<b>MASP-1 SP</b>
cDNA	52%	71%	70%	71%	62%	67%	61%
Protein	72%	77%	70%	71%	52%	75%	57%

Table.8 Identity between human and xenopus MASP-1/3 protein domains as determined by ClustalW pairwise sequence alignment. Percentage identity between predicted translation products (lower panel) and cDNA (upper panel). Sequence unavailable to complete alignment indicated by (n/a).

Xenopus MASP-3 identity to human MASP-3 indicates a relatively conserved coding sequence; in comparison to the previously discussed MASP-3 this is more uniform across the multiple protein domains. The MASP-3 serine protease domain retains the greatest identity with respect to human MASP-3 and is significantly more conserved than that of the comparable MASP-1 serine protease domain (see table.8).

### 3.19 cDNA cloning of zebra fish (*Danio rerio*) MASP-3

Several expressed sequence tag (EST) cDNA clones were identified from the current sequence databases utilising a translated homology search (TBlastx), utilising the rat MASP-3 cDNA sequence. Three EST's were identified as having homology in their predicted translational products in comparison to the rat MASP-3 primary amino acids sequence derived from this work. Matching cDNA clones were obtained from the IMAGE consortium (MRC), 3716657, 3732778 and 4759038. The clones were subsequently designated Zf AI, Zf AW and Zf BI respectively, these clones were fully sequenced and characterised.

EST cDNA clone Zf AI was sequenced, a total cDNA of 1877bp was characterised. The original EST sequence (IMAGE clone 3716657) of 490bp was fully contained within the cDNA sequence obtained. Homology searches of the current cDNA databases revealed at the DNA and predicted translation product level the Zf AI cDNA clone to have significant homology to *Cyprinus carpio* C1r/s B (AB042610). A larger *Danio rerio* cDNA clone has recently been identified in the cDNA database (BC081673), comprising 2249bp, containing the entire Zf AI cDNA clone. The locus is present on chromosome 16.

EST cDNA clone Zf AW comprised 916bp, the entire IMAGE EST clone 3732778 was contained within the sequence. BLASTn homology search of the current cDNA database indicated a homology to a 1648bp *Danio rerio* IMAGE clone (7250166), *Cyprinus carpio* C1r/s B (AB042610) is the closest currently characterised cDNA. Homology matches were obtained for the predicted translational product, the most significant sequence homology identified was the *Danio rerio* C1s subcomponent (AAI55075). The locus is present on chromosome 16.

EST cDNA clone pBK-CMV (4.5kb) Zf BI comprised a total of 2279bp, encompassing the entire IMAGE EST clone 4759038. BLASTn homology search of the current cDNA database indicated a complete homology to the predicted *Danio rerio* MASP-2

cDNA clone (XM001923846). Zf BI cDNA clone represents *Danio rerio* MASP-2 cDNA. The locus is present on chromosome 23.

Identification of a *Danio rerio* specific MASP-3 sequence was achieved by interrogation of the current zebra fish genomic sequence database, using a translated homology search (Tblastx) utilising rat MASP-3 specific sequence information derived from this work. Translated identity with human MASP-3 has an E-value of 0.0. *Danio rerio* predicted cDNA sequence XM\_001341900 with homology to MASP-3 was located on chromosome 15 (NCBI Map viewer LOC100002060). The locus consists of 10 exons with the terminal serine protease domain encoded by a single exon. The locus does not contain sequence specific for a MASP-1 serine protease domain, translated homology searches using the rat MASP-1 serine protease domain as a template identified only other MASP-like single exon serine protease transcripts within the *Danio rerio* genomic sequence. The potential *Danio rerio* MASP-3 sequence retains a CUBI-EGF-CUBII-CCPI-CCPII-SP domain structure (E-values  $2^{-17}$ , 0.3,  $3^{-39}$ ,  $2^{-9}$ , 0.05 and  $1^{-72}$  respectively), consistent with previously characterised sequence.

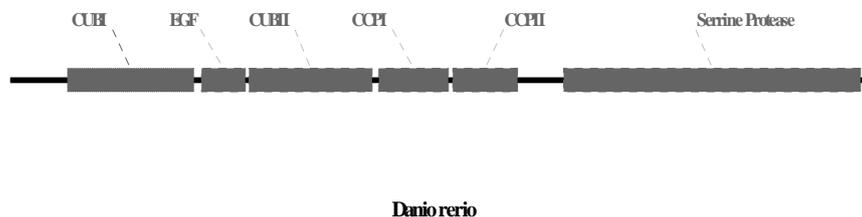


Fig.56 *Danio rerio* MASP-3 protein domain structure predicted by comparison to the Pfam database, using Pfam analysis of *Danio rerio* MASP-3 primary amino acid sequence (translation product of XM\_001341900). Domains with maximal identity to Pfam consensus sequence are represented, indicating an identical architecture to that exhibited by human MASP-3 (Fig.3).

DNA alignment of the *Danio rerio* MASP-3 cDNA sequence with the previously characterised human MASP-3 clone indicated an overall 62% identity (Clustal W method), alignment with human MASP-1 exhibits an identity of 57%. Alignment of translated human MASP-1 and MASP-3 cDNA, with that of *Danio rerio* XM\_001341900 produced a sequence identity of 47% and 56% respectively.



product conserves the evolutionary predictive amino acids at position 225 encoding a tyrosine residue and a serine at position 214 utilising a TCN codon (highlighted grey fig.57).

Examination of the serine protease domains reveals a complement of 5 cysteine residues, indicating an inability to form a histidine loop structure common to evolutionarily primitive serine protease.

Danio rerio MASP-1/3 common domains predicted translation product conserves all amino acids involved in the formation of MASP dimers an interaction (highlighted green and light green fig.57) mediated by binding within the CUBI-EGF domains. Similarly there is conservation of crucial amino acids residues required to associate with MBL and ficolin recognition molecules (highlighted red and dark red fig.57).

Residues conserved in previously characterised MASP-3 sequence implicated in a stabilising interaction between CCP1I domain and the serine protease domain are lacking in Danio MASP-3, additionally there are unique amino acid residues inserted within this region. Confirmation of this potential novel sequence is required with additional cDNA cloning experiments.

Analysis of the genes surrounding the zebra fish MASP-3 loci indicates significant differences to those previously described in this work, with the exception of the upstream SST (Somatostatin) locus (Fig.58).

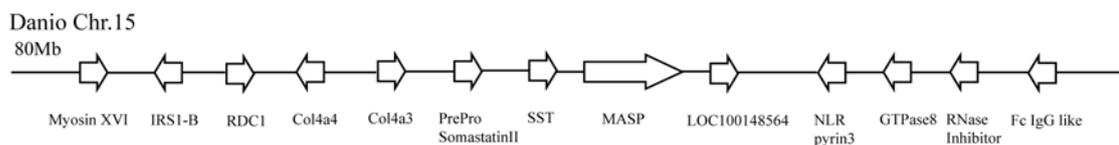


Fig.58 Danio rerio chromosome 15 indicating the Danio rerio MASP-1/3 locus with respect to upstream and downstream gene loci. Relative size, position and orientation of each loci indicated by arrow. Gene loci indicated in above text.

The genomic region of Danio rerio resides on chromosome 15, it exhibits considerable difference to those previously characterised in this work. Analysis of the MASP locus did not identify a multiple domain serine protease transcript, potentially indicating the loss of MASP-1 at this locus.

## 3.20 MASP sequences derived from genomic databases

The greatly increased availability and coverage of genomic sequence information has enabled the identification of potential cDNA sequence based on homology with previously characterised cDNA sequences. Utilising this approach it has been possible to identify potential MASP-3 cDNA and MASP-1/3 loci from a diverse range of species, augmenting the information gathered from traditionally isolated cDNA sequence. Predicted cDNA generated from genomic sequence is limited to the identification of coding sequence.

A systematic search of the current genomic sequence database utilising the sequence information previously generated within this work has enable the identification of 22 potential MASP-3 cDNA sequences and characterisation of the MASP-1/3 loci of a number of species, dependent on the coverage of the genomic sequence information. Findings for each species are discussed, with detailed sequence information presented in appendix. Sequence information derived from the genomic database will provide comparative data to enable phylogenetic relationships of the MASP-like serine protease to be elucidated.

Sequence derived from species closely related to those characterised in this work and those with important differences in MASP-1/3 loci structure are presented in the text, sequence derived from other species are included in appendices.

### 3.20.1 Chimpanzee (*Pan troglodytes*)

Based on sequence information derived from this work it was possible to analyse the current chimpanzee genomic sequence database ((WUGSC) 2006) to identify potential MASP-3 specific loci, using a translated homology search (Tblastx). A region on chromosome 3 was found to have a high degree of identity to rat MASP-3 in association with the previously predicted chimp MASP-1 (XM\_516941); predicted cDNA sequence for MASP-3 indicated open reading frames of 2160bp, encompassing 720 amino acid residues the sequence information is incomplete with an apparent insertion in the MASP-3 specific sequence due to alignment of genomic sequence clones. *Pan troglodytes* MASP-1 and MASP-3 specific sequences predicted from analysis of the current genomic sequence database reside on chromosome 3, spanning 70kb. The MASP-1/3 locus exhibits an identical genomic structure to those previously characterised in this work. Comparative analysis of human and chimpanzee MASP-1 and MASP-3 indicate a shared identity of 85% and 99% respectively, with both transcripts exhibiting a 99% identity at the predicted amino acid level (AlignX, Invitrogen) (Fig.59). Pfam analysis (Coghill, Finn et al. 2008) of the primary amino acid product indicates a MASP-like serine protease domain structure (CUBI-EGF-CUBII-CCPI-CCPII-Serine protease, E-value  $1^{-18}$ ,  $2$ ,  $7^{-53}$ ,  $7^{-10}$ ,  $3^{-8}$ ,  $5^{-71}$  respectively), chimpanzee MASP-3 retains features common to previously characterised MASP-3 transcripts.



determine if this amino acids and its respective codon usage is conserved or to confirm that the chimpanzee serine protease domain is encoded within a single exon. Cysteine residues within the serine protease domain are not aligned to construct a histidine loop, as determined from the available sequence information. There is conservation of serine protease activation site within the MASP-3 serine protease domain (highlighted bold and underlined fig.69), residues involved in the stabilisation of the interaction between CCPII and serine protease domain also exhibit conservation (highlighted pink fig.59).

Chimpanzee MASP-1/3 common region confirms complete conservation of amino acids residues contributing to MASP dimer formation through CUBI-EGF binding (highlighted green, dark green fig.59), and of residues involved in MBL and ficolin recognition molecule association (highlighted red and dark red fig.59).

### 3.20.2 Orang-utan (*Pongo pygmaeus*)

Utilising the rat MASP-3 cDNA sequence as a reference for a BLAST (Tblastx) translated DNA search identified a region of orang-utan chromosome 3 with significant identity. The region contained the entire orang-utan predicted MASP-3 sequence as identified by homology to rat MASP-3; similarly orang-utan MASP-1 specific sequence was also identified (ensemble gene model ENSPPYT00000016741). The MASP-1/3 locus spanned 76kb; the intron/exon structure of the locus was identical to the previously characterised loci.

The orang-utan predicted MASP-3 coding cDNA spanned 2187bp with an open reading frame of 728 amino acids. Comparison of orang-utan MASP-3 with human MASP-3 at the cDNA level indicates an identity of 98%; the predicted translation product exhibits an identity of 99% (AlignX, Invitrogen). Pfam analysis (Coggill, Finn et al. 2008) indicates the transcript encodes characteristic protein domains of the MASP-like serine protease, CUBI-EGF-CUBII-CCPI-CCPII-Serine protease (E-value  $8^{-20}$ ,  $2$ ,  $7^{-53}$ ,  $6^{-10}$ ,  $5^{-8}$ ,  $2^{-77}$  respectively). Examination of the genomic sequence indicates the MASP-3 serine protease domain is encoded within a single exon; characteristic evolutionary features are retained as discussed below.

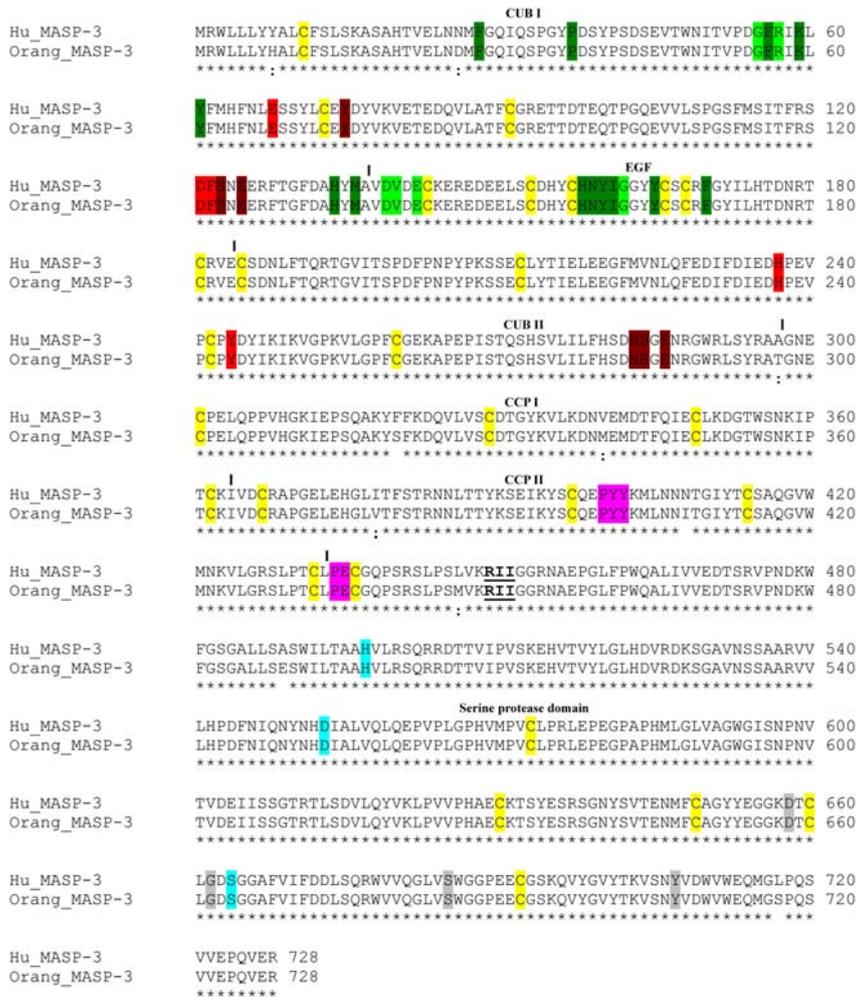


Fig.60 ClustalW 2.0.11 multiple sequence alignment of translated Pongo pygmaeus MASP-3 (ensemble gene model ENSPPYT00000016741) against human MASP-3 coding sequence (accession number AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Residues involved in MBL/ficolin interaction are highlighted in pink. Conserved residues involved in CCP II-SP interaction.

Orang-utan MASP-3 predicted translation product exhibits complete conservation of amino acids residues contributing to the catalytic triad, including codon usage at the evolutionarily discriminatory serine residue at position 195 (highlighted cyan fig.60). With respect to the characterised human MASP-3 sequence, orang-utan sequence retains important serine protease residues contributing to interactions within the S1 binding pocket, aspartic acid at position 189 and glycine 193 (highlighted grey fig.60). In common with previously discussed MASP-3 sequence Serine residue at position 214 utilises a TCN codon, position 225 is occupied by a tyrosine residue (highlighted grey

fig.60) and the protease domain contains five cysteine residues which is incompatible with the formation of a histidine loop structure (highlighted yellow fig.60). Stabilisation between CCPII and serine protease domains is enhanced by association between conserved residues (highlighted pink fig.60), as observed in previously identified MASP-3 sequence.

Orang-utan MASP-1/3 common region retains all amino acid residues contributing binding between CUBI-EGF domains during dimer formation (highlighted green and dark green fig.60), as are amino acids involved in MASP association with recognition molecules MBL and ficolin (highlighted red and dark red fig.60).

3.20.3 Rhesus monkey (*Macaca mulatta*)

Rhesus monkey MASP-1/3 genomic region was identified by translated identity (Tblastx) search utilising rat MASP-3 cDNA sequence derived from this work. The locus was identified on rhesus monkey chromosome 2, spanning 71kb. The intron/exon structure of the Rhesus macaca MASP-1/3 loci is identical to those previously described in this work. Rhesus macaca MASP-3 is incomplete due to the limitations of the genomic sequence, encompassing 2158bp, with an open reading frame of 719 amino acids missing the C-terminal region of the serine protease domain. Predicted translation products exhibit a 98% identity at the amino acid level in comparison with human MASP-3, identity at the cDNA level is 98%. Pfam analysis (Coggill, Finn et al. 2008) of the partial rhesus derived cDNA transcript predicted primary amino acid sequence indicates a domain structure conserved amongst the MASP-like serine protease, CUBI-EGF-CUBII-CCPI-CCPII-Serine protease (E-value  $8^{-20}$ ,  $2$ ,  $1^{-52}$ ,  $5^{-10}$ ,  $2^{-8}$ ,  $2^{-78}$ ).

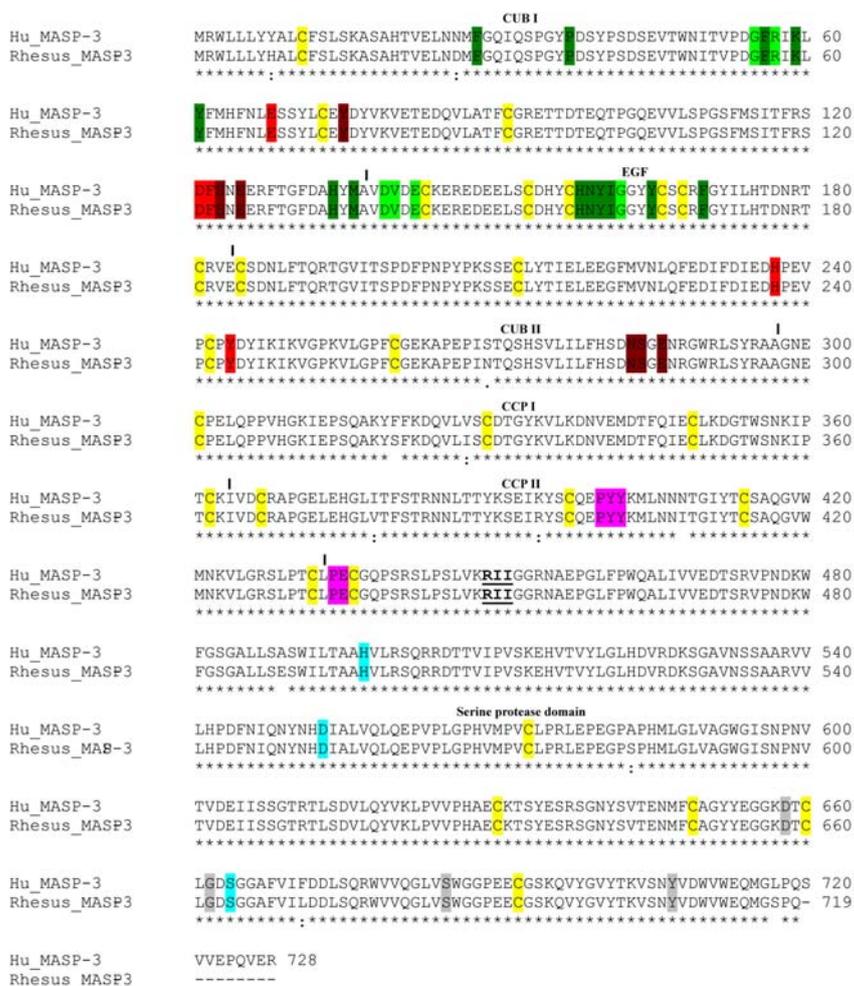


Fig.61 ClustalW 2.0.11 multiple sequence alignment of translated Rhesus macaca MASP-3 (cDNA sequence derived from this work) against human MASP-3 coding sequence (accession number

AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCPII-SP interaction.

Examination of rhesus MASP-3 sequence information indicates a domain structure consistent with the previously characterised MASP-3 characterised in this work, more detailed analysis confirms that residues important in the potential function of the protein are also retained. Catalytic residues histidine, aspartic acid and serine (highlighted cyan, fig.61) are conserved, the active site serine utilising an AGT codon. In common with previously described primate MASP-3 specific sequence information there is conservation of residues within the serine protease domains directly interacting and modulating the binding of potential substrates around the S1 binding pocket, highlighted grey in figure 61, residue 214 retains a TCN codon whilst residue 225 is occupied by a tyrosine residue indicating evolutionary concordance with previously described MASP. Rhesus MASP-3 exhibits previously conserved residues required at the activation cleavage site (highlighted in bold and underlined, fig.61) and amino acids associating to stabilise interaction between the CCPII and serine protease domain (highlighted pink fig.61).

MASP-1/3 shared domain of rhesus monkey presented complete conservation of amino acids residues involved in binding to lectin pathway recognition molecules MBL and ficolin (highlighted red and dark red fig.61), and in the formation of MASP dimers (highlighted green and dark green fig.61).

As a consequence of the incomplete nature of the current rhesus macaca genomic DNA assembly, it was not possible to confirm the terminal 9 residues of the MASP-3 serine protease domain and therefore if the domain is encoded within a single exon. Further cDNA cloning experiments and or correct assembly of the rhesus genomic sequence will address this anomaly.

#### 3.20.4 Zebra finch (*Taeniopygia guttata*)

Utilising the rat MASP-3 cDNA sequence as a reference for a BLAST (Tblastx) translated DNA search identified a region of zebra finch chromosome 9 with significant identity. The region contained an entire predicted MASP-3 specific sequence (accession number XP\_002191688), spanning 22kb. The exon structure of the MASP-3 gene is identical to those previously presented in this work, 9 exons encoding an A-chain whilst the serine protease domain is encompassed within a single exon. The zebra finch loci in common with the avian Gallus MASP-3 genomic loci (Lynch, Khan et al. 2005) does not contain a MASP-1 specific alternative splice partner. Analysis (Tblastx) of zebra finch genomic sequence utilising sequence derived from this work found identity with C1r (predicted gene model ENSTGUG00000013261), C1s (accession number XP\_002192681) and a truncated MASP-2 (accession number XP\_002186746) sequence. MASP-1 specific sequence could not be identified in the current zebra finch genomic sequence; the absence of MASP-1 and a truncated MASP-2 locus proposes MASP-3 as the only serine protease of the lectin pathway in this species. The predicted zebra finch MASP-3 (accession number XP\_002191688) spanned a total of 2270bp, with an open reading frame of 730aa. Predicted translation products exhibit a 77% identity at the amino acid level in comparison with human MASP-3, identity at the cDNA level is 76%. Protein domain homology analysis (Pfam) (Coggill, Finn et al. 2008) of the predicted primary amino acid sequence confirmed an identical domain structure to the previously characterised MASP-like serine protease, CUBI-EGF-CUBII-CCPI-CCPII-Serine protease (E-value  $7^{-18}$ ,  $1$ ,  $1^{-43}$ ,  $2^{-12}$ ,  $1^{-10}$ ,  $1^{-73}$ ).



described MASP-3 the formation of a histidine loop structure exhibited by MASP-1 is not possible. The MASP-3 activation site is conserved within the serine protease domain (highlighted bold and underlined fig.62), as are residues involved in the interaction with the CCPII domain which act to stabilise their relative positions (highlighted pink fig.62).

Zebra finch MASP-1/3 common domains exhibit conservation of amino acids residues contributing to the formation of MASP dimers through CUBI-EGF binding (highlighted green and dark green fig.62), interactions with MBL and ficolin are mediated via residues highlighted red and dark red (fig.62) which are also conserved.

### 3.20.5 Anole lizard (*Anolis carolinensis*)

The Anole lizard is currently the only reptile genomic sequencing project, examination of this important genomic sequence information allows new insights into the absence of MASP-1 in the avian species. Utilising sequence information derived from this work it was possible to identify a region of genomic sequence exhibiting identity to MASP-1 and MASP-3, this was present on Anole carolinensis scaffold\_656 encompassing contig 32922 and 32923. Ensembl gene prediction algorithm identified a potential MASP-1 sequence in this region (Ensembl gene ENSACAT00000006927), subsequent analysis of the genomic sequence revealed a MASP-3 specific serine protease exon. A translated identity search (TBlastx) predicted a MASP-3 single exon with a 282 amino acid match to rat MASP-3 (E-value of  $8.3^{-246}$ ); analysis of the genomic sequence indicated an open reading frame of 295 amino acids terminating in a TGA stop codon. Anole predicted MASP-1 specific sequence is encoded in 6 exons with identity to rat MASP-1 (TBlastx E-value  $4.0^{-146} - 8.2^{-160}$ ) (fig.63).

Anole MASP-1/3 genomic locus exhibits a configuration identical to that of mammalian species previously characterised in this work. A common MASP-1/3 region is downstream of an alternatively spliced MASP-3 specific serine protease exon; Anole MASP-1serine protease is upstream of the MASP-3 exon and is composed of multiple exons. Anole MASP-3 is encoded by a cDNA of 1956bp, with an open reading frame of 652aa. Anole MASP-3 exhibits an identity of 73% compared to human MASP-3 at the cDNA level; predicted translation products share a similarity of 75%.

Translated identity search (TBlastx) using sequence derived from this work identified other MASP-like serine protease present in the Anole genome, homologues of MASP-2, C1r and C1s were observed. With reference to rat MASP-3 serine protease domain the identity probability score, E-values were significantly lower for the predicted sequences  $5.6e^{-94}$ ,  $5.6e^{-72}$  and  $2.4e^{-46}$  respectively.



serine protease domains of Anole MASP-1 and MASP-3, indicates that anole MASP-1 preserves a primordial Ser195:TCC (TCN)/Ser214:TCT(TCN)/Tyr225 codon usage, whilst Anole MASP-3 Ser195:AGC (AGY)/Ser214:AGC(AGY)/Tyr225 corresponding to the most modern configuration. The primordial nature of MASP-1 with respect to MASP-3 is also predicted by the complement of five cysteine residues (highlighted yellow fig.63) within the MASP-3 protease domains inability to form a histidine loop structure, proposing an evolutionarily advanced structure. Anole MASP-3 retains highly conserved amino acids concerned with the interaction between CCP1I and the serine protease domain (highlighted pink fig.63).

Anole MASP-1/3 shared domains are incomplete with the absence of first exon coding for the CUB1 domain, therefore it is not possible to confirm the complete conservation of residues involve in MASP dimer formation (highlighted green and dark green fig.63) and interactions with lectin pathway recognition molecules (highlighted red and dark red fig.63). There is however conservation of these amino acid residues within the available sequence.

### 3.20.6 Tetraodon (*Tetraodon nigroviridis*)

Analysis of assembly TETRAODON 8.0, Mar 2007 of the *Tetraodon nigroviridis* genomic sequence database, identified two predicted gene sequences with a high degree of identity to rat MASP-3 (ENSTNIG00000011197 and ENSTNIG00000011199, E-value  $2.6^{-118}$  and  $2.0^{-120}$ ). Tetraodon sequences share a 99% identity at the DNA level, primary amino acids sequence identity of 94%. Predicted gene products are configured head to tail, spanning 14kb.

Serine protease domains are composed of multiple exons, ENSTNIG00000011197 preceded by CUBI-EGF-CUBII-CCPI-CCPII and ENSTNIG00000011199 by CUBI-EGF-CUBII-CCPI as determined by Pfam analysis of the predicted primary amino acid structure. Analysis of the serine protease domain for each of the derived MASP-like sequences indicates that both have a multiple exon structure Tetraodon-197 a dual exon whilst Tetraodon-199 consists of 3 exons.

Examination of tetraodon MASP-3 serine protease domain indicates that both derived sequences possess residues able to form a catalytic triad (highlighted cyan fig.64); there is also conservation of residues intimately involved in substrate binding at the S1 binding pocket at positions 189 and 193 an aspartic acid and glycine residue respectively (highlighted grey fig.64). Evolutionarily significant residues within the serine protease domain indicates that both tetraodon sequences possess a serine residue at position 214 (TCN codon) and a tyrosine at position 225 (highlighted grey fig.64), specifically Tetraodon-197 (Ser195:AGC (AGY)/Ser214:TCC (TCN)/Tyr225) and Tetraodon-199 (Ser195:AGC (AGY)/Ser214:TCC (TCN)/Tyr225). Tetraodon MASP-3 serine protease domain from both predicted transcripts encompasses five cysteine residues (highlighted yellow fig.64), therefore is unable to form a histidine loop structure as seen in more primitive serine protease including MASP-1.

Common MASP-1/3 region is incomplete for both predicted transcripts of the available sequence there is conservation of those residues involved in the interaction with lectin pathway recognition molecules MBL and ficolin (highlighted red and dark red fig.64). Examination of the derived sequence with respect to the amino acid residues

contributing to the inter MASP association mediated between CUBI and EGF domains, indicated that there are a number of residues that are not conserved with respect to the human MASP-1/3 sequence (highlighted green and dark green fig.64).

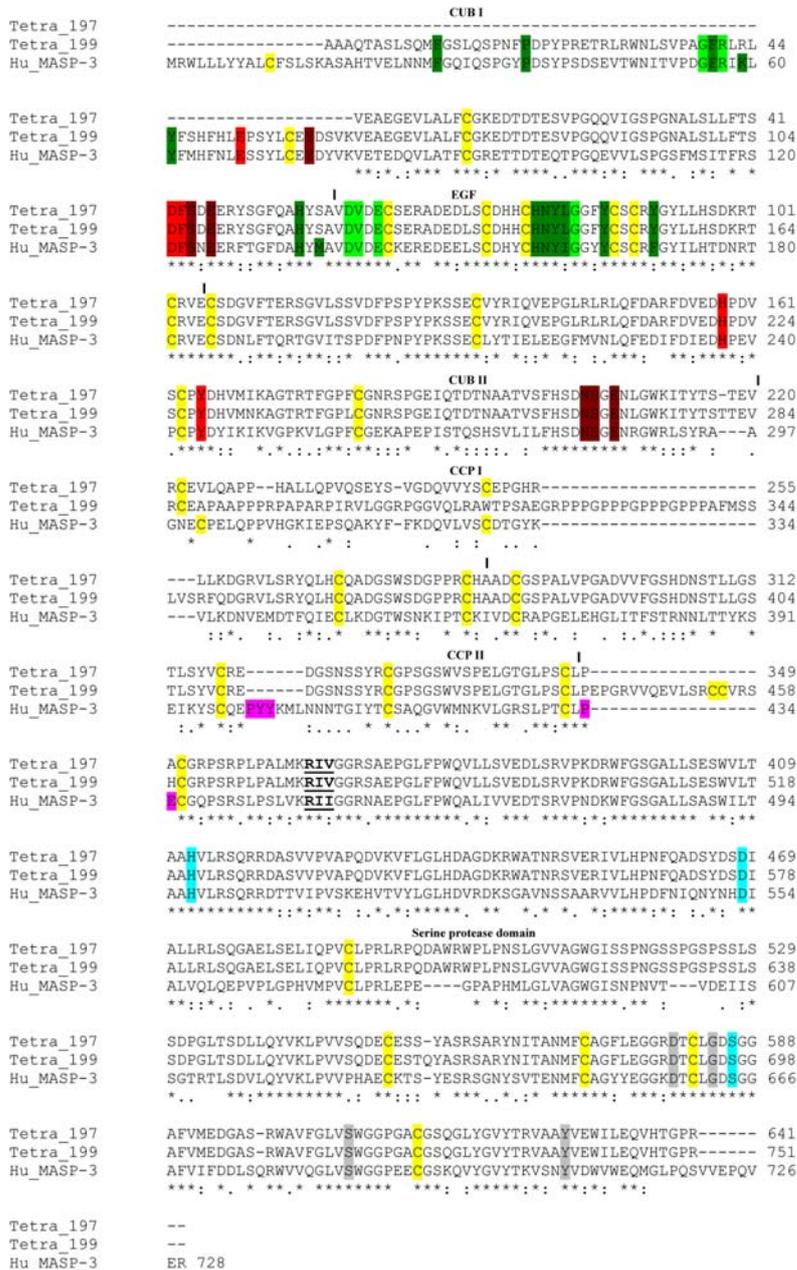


Fig.64 ClustalW 2.0.11 multiple sequence alignment of translated Tetraodon nigroviridis MASP-3 (Ensembl gene ENSTNIG00000011197 and ENSTNIG00000011199) against human MASP-3 coding sequence (accession number AF284421). Conserved amino acid residues are indicated by (\*), strongly conserved by (: ) and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCPII-SP interaction.

Translated identity searches for other MASP-like serine protease, utilising rat MASP-2, C1r and C1s sequences failed to identify complete homologues in the current Tetraodon genomics sequence database. Partial sequence information was observed for Tetraodon C1s (ENSTNIG00000002027), Pfam analysis revealed this to be limited to the A-chain. The presence of MASP-1 specific sequence remains in question due to the limited assembly of the tetraodon genomic sequence.

3.20.7 Stickleback (*Gasterosteus aculeatus*)

Analysis of the current (Jan 2009, 55.1j) *Gasterosteus* genomic sequence database (BROAD\_S1 2009) identified a number of predicted genes with a high degree of identity to MASP-3 specific sequence derived from this work. Translated identity searches using rat MASP-3 were able to locate several related predicted gene sequences, present on chromosome GpI and GpXX.

A predicted MASP-3 *Gasterosteus* gene is present on chromosome GpI (500,828-510,359), (ENSGACG00000004811) 2136bp with an open reading frame of 712 amino acids. Encompassing 13 exons, including 3 within the serine protease domain, Pfam (Bateman, Coin et al. 2004) analysis indicates a CUBI-EGF-CUBII-CCPI-CCPII-SP domain structure. The predicted sequence exhibits close homology to a number of Bony fish (Tetraodon, Carp, Rainbow trout and Danio) MASP-3 and to *Xenopus* MASP-3a. Identity to rat MASP-3 serine protease domain, (Tblastx) probability E-value of  $7.4^{-117}$ .

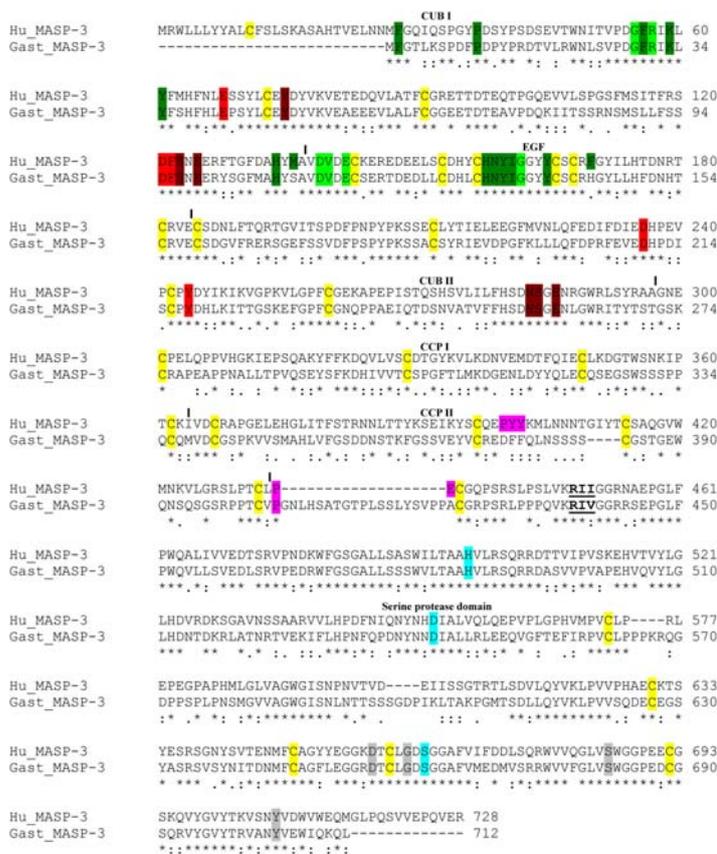


Fig.65 ClustalW 2.0.11 multiple sequence alignment of translated *Gasterosteus aculeatus* MASP-3 (ENSGACG00000004811) against human MASP-3 coding sequence (accession number AF284421).

Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Conserved residues with evolutionary significance highlighted grey. Conserved cysteine residues are highlighted yellow, underlined are cysteine residues of the methionine loop. Serine protease cleavage site RII, underlined and bold. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCP-II-SP interaction.

Examination of the predicted *Gasterosteus* MASP-3 serine protease domain indicates a catalytic triad encompassing histidine, aspartic acid and a serine residue encoded by an AGY codon. There is conservation of residues involved at the substrate binding site positions 189 and 193 encoded by aspartic acid and glycine residues respectively (highlighted grey fig.65). Characterisation of evolutionarily significant residues within the serine protease domain indicates the predicted sequence to possess a serine residue at position 214, utilising an AGY codon whilst a tyrosine residue occupies position 225, the most recent configuration in evolutionary terms (highlighted grey fig.65). Further evidence of the evolutionarily recent origin of the predicted *Gasterosteus* MASP-3 sequence is the configuration of five cysteine residues within the serine protease domain, indicating an inability to form a histidine loop structure. There is conservation of residues involved in the formation of MASP dimers via interaction between CUBI and EGF domains (highlighted green and dark green fig.65), as there is conservation of amino acids mediating interaction with MBL and ficolin (highlighted red and dark red fig.65)

No other MASP-like serine protease were identified in the genomic sequence surrounding the ENSGACG00000004811 locus, predicted Factor VII (ENSGACG00000015464, ENSGACG00000015472 – E-value  $8.6^{-80}$ ,  $3.0^{-74}$ ) and Factor X (ENSGACG00000015474 – E-value  $1.1^{-81}$ ) were identified on the same chromosome. Pfam analysis of Factor VII and Factor X predicted sequences indicates a homology to the trypsin serine protease consensus sequence (E-value  $3^{-78}$ ,  $9^{-76}$ ). There is an absence of MASP-1 specific sequence in this genomic region.

Two predicted genes are present on chromosome GpXX: 11,700,420 - 11,719,215, ENSGACT00000013414 and ENSGACT00000013445, with identity to bony fish C1s (Tetraodon CAF97835, E-value 0) and C1r (Tetraodon CAF97836 and Rainbow trout NP\_001117852, E-value 0). Predicted C1s and C1r both exhibit MASP-like serine

protease domain architecture, both loci are in the same orientation C1s is upstream of C1r. C1s serine protease domain is encoded in a single exon; C1r is encoded within 3 exons. C1r displays evolutionarily modern codon usage within the serine protease domain (Ser195:AGC (AGY)/Ser214:AGC (AGY)/Tyr225), C1s exhibits an intermediate form (Ser195:AGT (AGY)/Ser214:TCC (TCN)/Tyr225).

Analysis of the complement of MASP-like serine protease is complex within *Gasterosteus*; the predicted MASP-3 sequence certainly exhibits features unique to MASP-3 and the evolutionarily more recent lineage of MASP but appears to be contained within a multiple exon serine protease domain. The multiple domains nature appears to be exhibited by *Gasterosteus* C1r as discussed above; such findings require confirmation with further laboratory examination.

### 3.21 Truncated gene product of the MASP-1/3 locus

Analysis of the current genomic sequence databases has predicted novel splice products of the MASP-1/3 locus in rat, mouse, human and porcine species.

In addition to the alternative splice products MASP-1 and MASP-3, genomic alignment and gene prediction algorithms (MapViewer, NCBI) applied to the rat MASP-1/3 locus identified a truncated MASP-1/3 EST (accession number BC085685) differing at the C-terminal end. EST clone BC085685 includes a 3' untranslated region, poly-A tail and an open reading frame of 385aa with 17 novel amino acids at the C-terminal end. The predicted splice product does not include a serine protease domain, encompassing the first four protein domains of the MASP-1/3 common region and a novel C-terminal exon as described above (CUBI-EGF-CUBII-CCPI-KSEIDL EEEEELESEQVAE). This novel sequence information may represent a third alternative splice product of the MASP-1/3 locus; further work is needed to clarify this bioinformatic finding.

In common with the finding in rat, interrogation of the available genomic and EST sequence database (MapViewer, NCBI) identified a truncated gene product specific to the mouse, differing at the C-terminal end arising from the mouse MASP-1/3 locus (accession number AK043055). EST clone AK043055 includes a 3' untranslated region with an open reading frame of 385aa with 17 novel amino acids at the C-terminal end. The predicted splice product does not include a serine protease domain, encompassing the first four protein domains of the MASP-1/3 common region and a novel C-terminal exon as described above (CUBI-EGF-CUBII-CCPI-KSEIELEKELESEPVAE) (Fig.66).

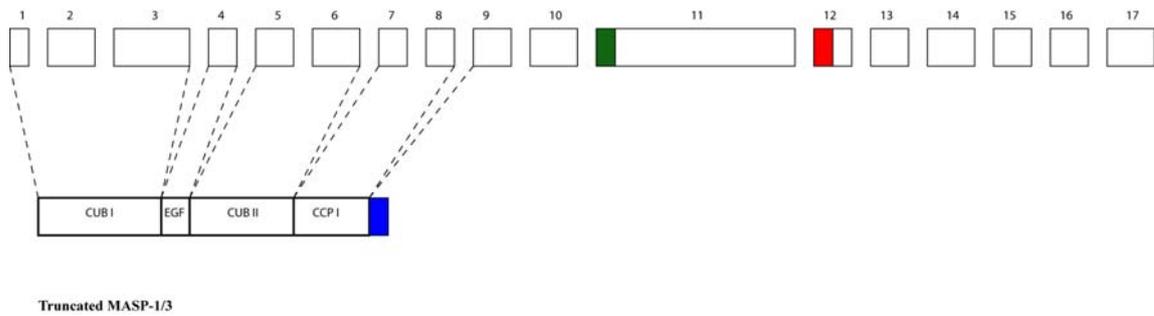


Fig.66 Schematic diagram of truncated MASP-1/3 cDNA product predicted domain structure (lower panel), respective MASP-1/3 exon structure (upper panel). Truncated MASP-1/3 specific amino acids indicated by a blue box, MASP-3 specific exon highlighted by a green box and MASP-1 specific exons indicated by a red box.

A further mouse EST clone (accession number AK032944) was identified as a further potential alternative splice product of the MASP-1/3 locus. The transcript includes a 3'untranslated region with a predicted poly-adenylation signal (PolyAH) 510bp upstream of the termination codon and an open reading frame of 284aa, encompassing 33 novel amino acids at the C-terminus. The predicted alternative splice product does not include a serine protease domain, encoding two N-terminal protein domains of the MASP-1/3 common region and the first exon of CUBII domain followed by 33aa of novel sequence (CUBI-EGF-CUBII/IKVSVQWLLLPRVHLGSIGKKMRILGLWLFIFS). The predicted transcript AK032944 was unique to mouse, interrogation of the current sequence databases failed to identify similar transcripts of sufficient identity in other species. With respect to the absence of the second exon of CUBII domain predicted in the truncated MASP product, there is loss of crucial amino acids residues involved in the interaction with recognition molecules MBL and ficolin (highlighted dark red, CUBII in fig.37).

Analysis of the human genomic and EST database using the NCBI Map Viewer tool identified three potential truncated alternative splice products of the human MASP-1/3 locus (BC039724, CR749615 and AK300663).

The predicted splice products do not include a serine protease domain; all four human transcripts encompassed a unique C-terminus, encoding a CUBI-EGF-CUBII-CCPI-KNEIDLESELKSEQVTE domain structure. Transcript AK300663 is 1528bp in length, with an open reading frame of 407aa and a 3' untranslated region of 174bp. Transcript

CR749615 includes 1778bp with an open reading frame of 462aa, terminating in a polyA tail 223bp downstream of the stop codon. Transcript BC039724 includes 2065bp with an open reading frame of 462 amino acids; the transcript includes 513bp of 3' untranslated region prior to polyadenylation. The identification of three independent transcripts each utilising alternative polyadenylation signals, suggests that the transcripts identified in the NCBI database are representative of expressed gene products.

Analysis of the Rhesus NCBI gene database utilising the Map Viewer tool identified a truncated product of the MASP-1/3 locus (accession number AB172986), the transcript includes 1880bp with an open reading frame of 382aa. In common with the previously described truncated transcripts the predicted primary amino acid product encodes a CUBI-EGF-CUBII-CCPI-ENEIDLESELKSEQVTE domain structure with 17 unique amino acids at the C-terminal end.

This novel sequence information may represent further alternative splice products of the MASP-1/3 locus; further work is needed to clarify this bioinformatic finding.

### 3.22 Evolutionary comparison of MASP-3 sequence

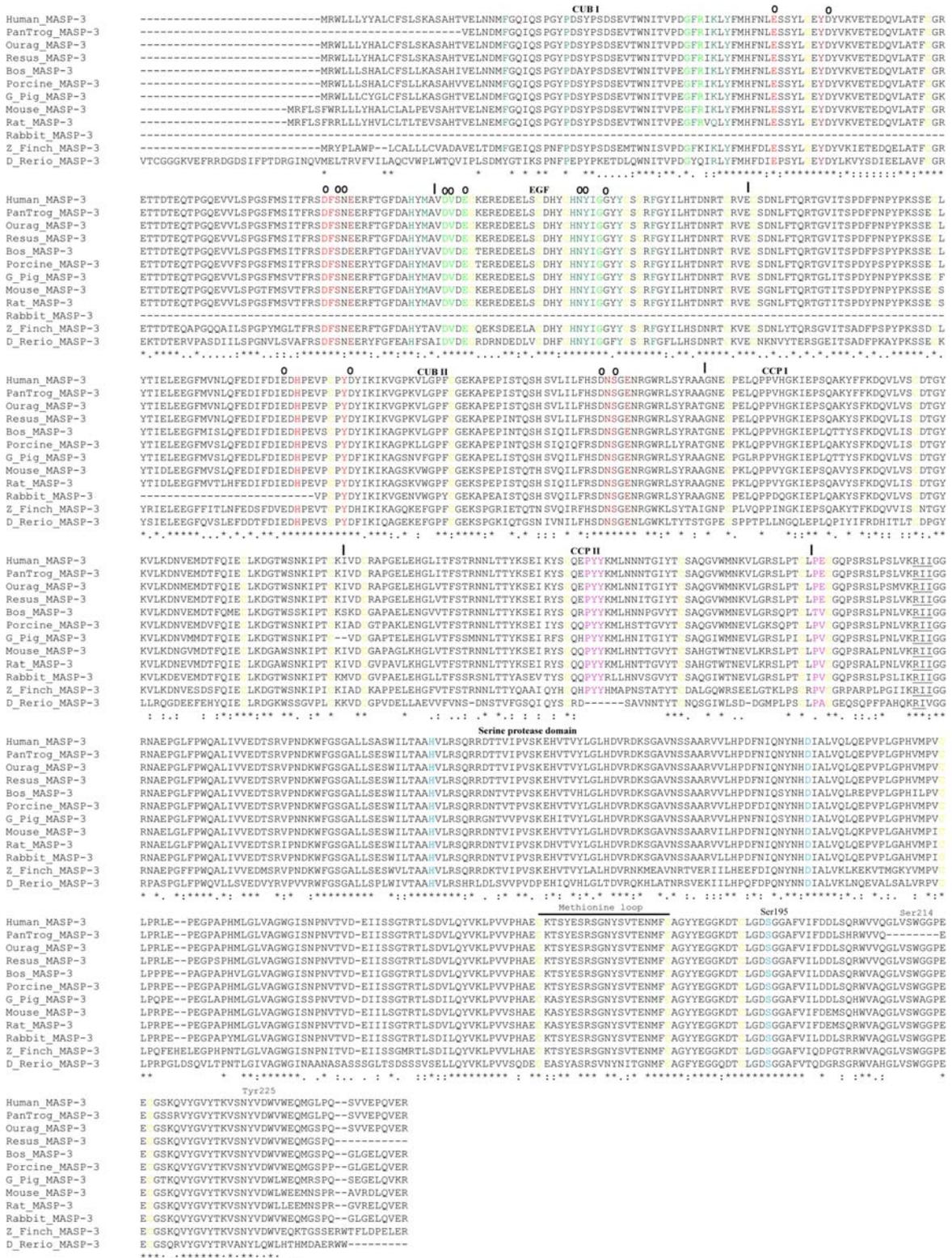


Fig.67 ClustalW 2.0.11 multiple sequence alignment of twelve translated MASP-3 (contig derived from this work), human MASP-3 coding sequence (accession number AF284421) as profile. MASP domain boundaries are indicated by vertical line and labelled accordingly. Conserved amino acid residues are indicated by (\*), strongly conserved by (:), and weakly conserved by (.). Serine protease active site

residues Histidine (H), Aspartic acids (D) and Serine (S), highlighted in blue. Ser241 and Tyr225 are indicated. Methionine loop structure is indicated with a horizontal line and labelled. Conserved cysteine residues are highlighted yellow, serine protease cleavage site RII, underlined and bold. Amino acids directly involved in  $\text{Ca}^{2+}$  binding are indicated by **0**. Residues involved in dimer-dimer formation of major and minor significance are highlighted in green and light green respectively. Residues involved in MBL/ficolin interaction are indicated in red and dark red, major and minor roles respectively. Highlighted in pink are conserved residues involved in CCP-II-SP interaction.

Multiple alignment of the predicted primary amino acid sequence derived from this work originating from a wide range of species, clearly demonstrates conservation of the domain architecture exemplified by human MASP-3 outlined in the figure above. With reference to the residues encompassing the serine protease catalytic triad (highlighted cyan fig.67) there is complete conservation of amino acid residues and as discussed in previous species specific sections a conservation of codon type with respect to the active site serine residue at position 195.

Amino acid at position 189 is invariably an aspartic acid residue (highlighted grey fig.67) in all species examined within this work, there is conservation of this amino acid due to its vital position at the base of the S1 binding pocket and role in modulating substrate interaction (Szabo, Bocskei et al. 1999).

There is conservation of a glycine residue at position 193 (highlighted grey fig.67) due to the interaction between substrate and protease domain, substitution of the glycine residue significantly impairs substrate binding and rate of catalysis (Bobofchak, Pineda et al. 2005).

A serine residue at position 214 is constant amongst all MASP-3 serine protease domains examined (highlighted grey fig.67), playing a crucial role in substrate binding (Krem, Prasad et al. 2002). Examination of the codon usage at this amino acid position has a further use as a marker of evolutionary phylogeny amongst serine protease (Krem and Cera 2002), of the MASP-3 specific sequences examined there was complete concordance with the use of a TCN codon at this position.

Residue at position 225 is conserved and exhibits evolutionary dichotomy amongst serine protease, MASP-3 serine protease exhibit a tyrosine residue at this position (Krem and Cera 2002) (highlighted grey fig.67). The tyrosine residue at this position

does not have direct interaction with the substrate, but functions to control water flow within the primary binding pocket of the serine protease (Enriqueta R. Guinto and Cera 1999).

Examination of cysteine residues within the MASP-3 serine protease domain (highlighted yellow fig.67) confirms a total of five residues in all species examined, the lack of a sixth cysteine residue as exhibited by MASP-1 and more primitive serine protease highlights MASP-3 inability to form a histidine loop structure required to stabilise the structure of the serine protease domain.

The activation cleavage site motif is conserved amongst MASP-3 serine protease examined (highlighted bold and underlined fig.67), an arginine, isoleucine, isoleucine predominates. MASP-3 sequence examined in *Danio rerio* showed the amino acid in the third position as an alternate valine residue, the impact on the ability of MASP-3 to be activated in this species remains to be determined and may indicate the ability to use an alternate cleavage motif.

Residues contributing to the interaction between CCPII and serine protease domain, (highlighted pink fig.67) exhibit conservation of sequence residing within the CCPII domain with the exception of that derived from *Danio rerio*. The corresponding amino acid residues of the protease domain exhibit more diversity of sequence, the proline residue exhibits most conservation with much variation at the adjacent amino acid. Retention of sequence at these specific positions highlights the importance of interactions between the serine protease domain and other associated domains of the mosaic protein.

The four C-terminal residues of all mammalian MASP-3 sequence examined where reliable sequence information is available, exhibit conservation of a QVER motif (see fig.67). The potential significance of this conservation remains to be determined; C-terminal residues are intimately linked with protein trafficking and post translational modification (Gatto and Berg 2003).

Examination of the common MASP-1/3 region in the alignment above (fig.67) highlights the complete conservation of amino acid residues involved in the

coordination of calcium ions (indicated by  $\circ$  fig.67) mediating interactions between MASP dimers (Teillet, Gaboriaud et al. 2008), this theme is continued with residues directly involved in binding between CUBI and EGF domains during dimer formation (highlighted green and dark green fig.67).

MASP-1/3 association with recognition molecules of the lectin pathway MBL and ficolin are mediated by a series of conserved amino acids residues (highlighted red and dark red fig.67), the alignment above confirms complete conservation of these residues proposing the importance of this interaction to the function of MASP-1 and MASP-3.

### 3.23 Phylogenetic analysis MASP-3 serine protease

Phylogenetic analysis of MASP-3 serine protease domain primary amino acid sequence for available and predicted cDNA transcripts was performed, using primordial Ascidiars MASPa and MASPB as an outgroup. The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 3.05534560 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Zharkikh and Li 1995). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 343 positions in the final dataset. Phylogenetic analyses were conducted using the MEGA4 (Tamura, Dudley et al. 2007) program.

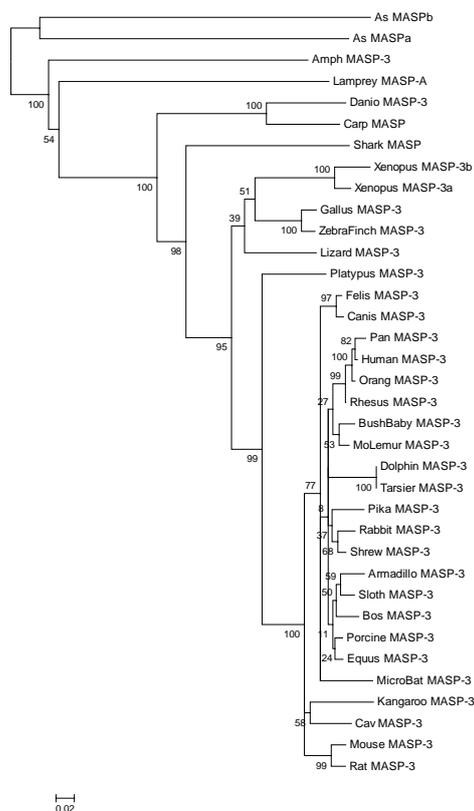


Fig.68 Evolutionary relationships of 35 MASP-3 serine protease domains presented in this work, as determined by the Neighbour-Joining method, bootstrap consensus tree was inferred from 500 replicates. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4.

The phylogenetic tree produced utilising MASP-3 serine protease domain sequence conforms to that predicted from the standard evolutionary model, with distinct placement of each species class. The topology of the MASP-3 constructed phylogenetic tree with high bootstrap values supports the reliability of both tree construction and the validity of the derived MASP-3 serine protease sequence. No member of the MASP-3 serine protease sequence presented here has experienced significantly greater evolutionary pressure that may have altered its sequence out of context of its predicted path of evolution, highlighting the conservation of sequence and potentially the conservation of function.

Vertebrate lineages are clearly separate from those of more primordial non-chordate species. MASP-3 phylogeny exhibits distinct grouping of Mammalian, Amphibian, Avian, Reptilian and Actinopterygian Class (Fig.68). The three classes of fish are also clearly separated, the jawless lamprey, jawed bony fish and jawed cartilaginous fish each present on a unique branch of the phylogenetic tree. There was specific clustering of all MASP-3 serine protease separate from the Ascidian MASPa and MAS Pb bootstrap percentage of 100%; these are at present the most primitive MASP-3 sequence available.

### 3.24 Phylogenetic analysis of full length MASP-like serine protease

Alignment of full length MASP-like serine protease primary amino acid sequence, derived from reference sequence and from predicted gene models based on genomic sequence data contained within the NCBI database was undertaken using the ClustalW method (Thompson, Gibson et al. 2002). A total of 87 amino acid sequences (accession numbers presented in appendices) were suitable for alignment, positions containing alignment gaps and missing data were eliminated in a pairwise manner (see fig.69).

Bootstrap confidence limit of 70% clearly defined four major sequence clusters corresponding to the out group of distant MASP-like protease encompassing Ascidian and Cnidaria derived sequence, the only non-chordate species currently characterised (Fig.68). The remaining three groups encompass the MASP-like serine protease loci identified in mammalian species, MASP-1, MASP-3, C1r, C1s and MASP-2 primary amino acids transcripts. Examination of the relative position and clustering of particular MASP-like serine protease sequences is revealing, MASP-1 is isolated on its own phylogenetic branch. MASP-3 and C1r although separated within a particular branch, are present on the same tree branch indicating a higher degree evolutionary identity to each other than to other MASP-like sequences. Similarly arranged are MASP-2 and C1s sequences, again highlighting their potential evolutionary relationship with respect to other MASP-like serine protease (fig.69).

The relative evolutionary relationships of MASP-like serine protease are of vital importance in delineating the events surrounding the emergence of a number of related proteins, thought to have emerged as a consequence of gene duplication.

Analysis of the phylogenetic associations of the MASP-like serine protease presented here proposes that MASP-1 is closely related to the primordial MASP, whilst a gene duplication event was required to generate the second member in each of the phylogenetic groups outlined above (MASP-3/C1r and MASP-2/C1s).

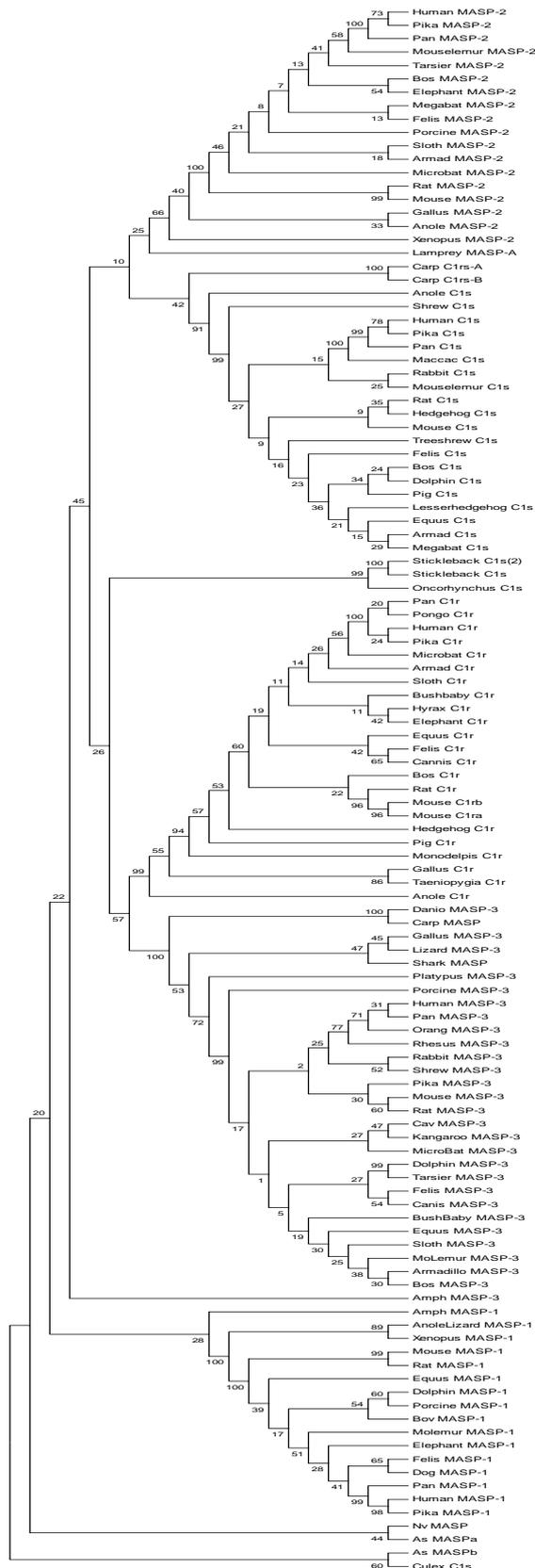


Fig.69 Cladogram of MASP-like serine protease primary amino acid sequence, derived from ClustalW alignment of 87 taxa. The evolutionary history was inferred using the Neighbor-Joining method, bootstrap consensus tree inferred from 500 replicates. All positions containing gaps and missing data were eliminated from the dataset, there were a total of 134 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

### 3.25 Phylogenetic analysis of MASP-like serine protease domain

Comparison was made between MASP-3 serine protease domain primary amino acid sequence and those of other MASP-like serine protease (MASP-1, MASP-2, C1r and C1s), both from cDNA sequence characterised in this work, reference sequence and predicted transcripts originating in the current NCBI sequence databases. The aim of this approach is to cluster sequence with those that are most similar, with the inference that they evolved from the most recent common ancestral sequence. Analysis of the serine protease domain in isolation aims to confirm the relationships proposed by the analysis of the entire MASP-like sequence, whilst excluding domains that are shared between alternative splice products that may experience differing evolutionary pressure. Elimination of the shared domain sequence excludes the problem of overestimation of identity when analysing members of an alternatively spliced gene locus, due to the dual evolutionary pressures experienced by the common sequence.

The primary amino acids structure from transcripts encompassing the entire protease domain for MASP-1, MASP-2, MASP-3, C1r and C1s were aligned using the ClustalW method (Thompson, Gibson et al. 2002). Alignments were analysed using the neighbour-joining method to determine associations between evolutionarily distant MASP-like serine protease sequences. Phylogenetic analysis of all available and predicted MASP-like serine protease domains clearly defines five board groups corresponding to the classical complement serine protease C1r and C1s, the lectin pathway MASP-2 and alternative splice products MASP-1 and MASP-3. Presentation of the data in this way enables an overview of relationships between groups to be determined.

An alternate pattern of evolutionary relationships emerges with respect to that determined by analysis of the entire MSAP-like sequence. MASP-1 and evolutionary primitive Ascidian and Nematostella MASP exhibit close identity. MASP-2 and C1s form separate groups on individual branches, whilst MASP-3 and C1r are present in close association sharing a common ancestral sequence.

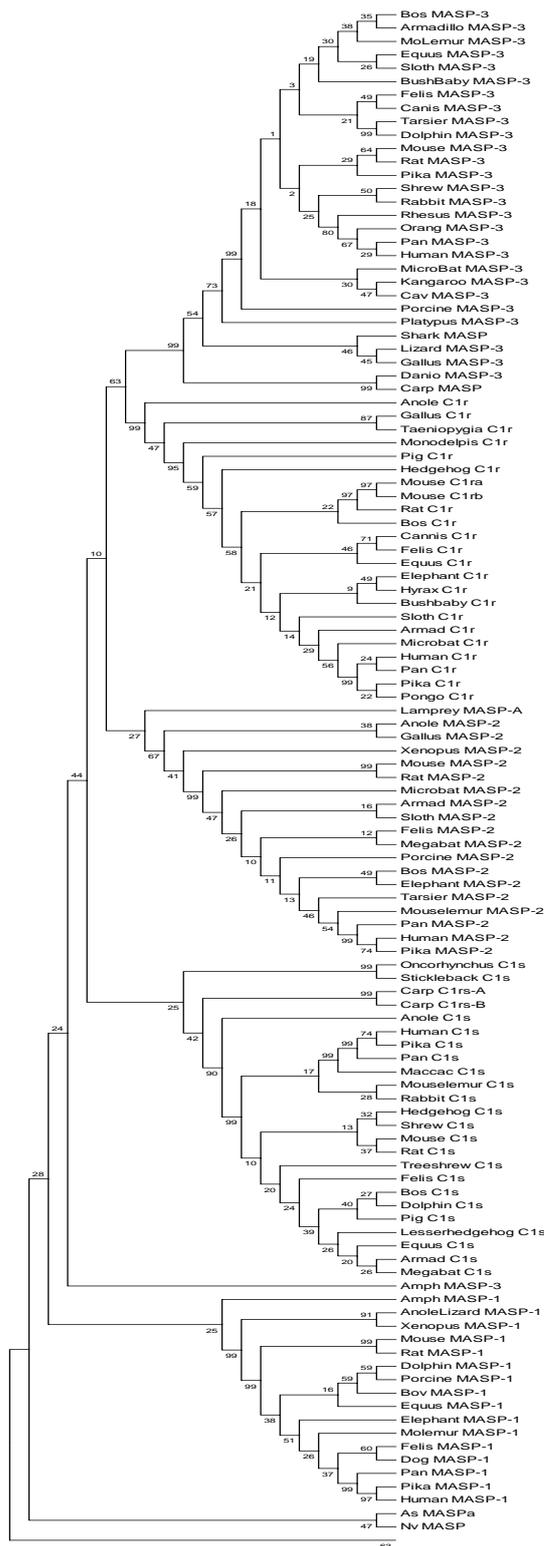


Fig.70 Phylogenetic tree analysis of MASP-like serine protease domain ClustalW alignment. Phylogenetic tree of MASP-like serine protease domain primary amino acid sequence, derived from ClustalW alignment of 87 taxa. The evolutionary history was inferred using the Neighbor-Joining method, bootstrap consensus tree inferred from 500 replicates. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset; there were 134 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

### 3.26 Northern blot analysis of rat and mouse MASP-3 mRNA transcripts

Northern blot analysis of total liver RNA in both rat and mouse species indicate a single transcript of 4kb hybridising with a human MASP-3 specific probe (Fig.71). Northern blot hybridisation supports the evidence from the previous work, identifying a single hybridising mRNA product of expected size for each species. No cross hybridisation is exhibited with MASP-1 mRNA species.

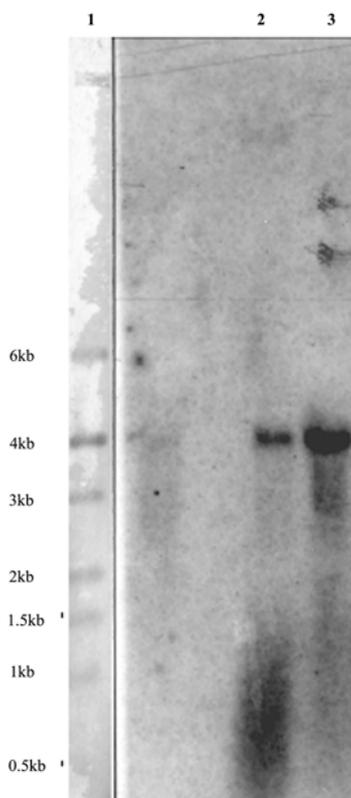


Fig.71 Northern blot analysis of 10 $\mu$ g total liver RNA from rat and mouse, hybridised with a radiolabeled human MASP-3 specific probe (Hu MASP-3 p6F + p8F, Fig.16). Lane 1 - 1kb RNA marker, lane 2 - rat total liver RNA, lane 3 - mouse total liver RNA.

Analysis of total RNA from rat and mouse liver successfully identified a single mRNA transcript specific for MASP-3, corresponding to the predicted size. Direct confirmation of the characteristics of an mRNA species is supportive of results obtained utilising a PCR cloning approach, enabling greater confidence in results. The single hybridising species correctly matches those obtain by both PCR and cDNA library cloning experiments.

### 3.27 Specific ratio of MASP-1: MASP-3 mRNA levels in rat liver

Sequence data obtained from rat MASP-3 and MASP-1 cDNA clones enabled the production of primer pairs spanning the common MASP-1/3 A-chain into specific MASP-1 and MASP-3 serine protease domains respectively. Oligonucleotide primers producing specific amplification products of approximately 300bp were designed and optimised for Quantitative PCR analysis.

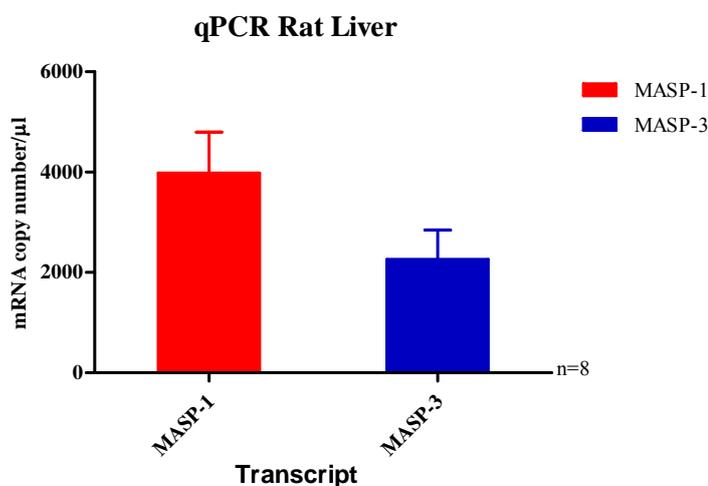


Fig.72 Quantitative RT PCR analysis of MASP-1 and MASP-3 mRNA expression in rat liver. Results are expressed as copies per microgram of RNA, and are the mean of 8 experiments. Error bars represent SD. MASP-1 represented by a red bar, and MASP-3 represented by a blue bar.

The relative expression of alternative splice variants of the MASP-1/3 locus were analysed in rat liver tissue. Rat liver total RNA was isolated from normal rat liver and reverse transcribed into cDNA by random oligonucleotide primed reverse transcription, as described. Rat liver cDNA was normalised with respect to the housekeeping gene GAPDH, a total of 8 rat liver RNA samples were analysed. Specific PCR products spanning the common MASP-1/3 region into the specific MASP-1 and MASP-3 sequence were generated, the quantity of PCR product analysed after each cycle via the Lightcycler (Roche) apparatus as described. Standard curves were generated using serial dilutions of cloned cDNA specific for each target sequence, standards were run in parallel. Rat liver exhibits an excess expression of MASP-1 mRNA with respect to its alternative splice product MASP-3. The relative expression of MASP-1 and MASP-3 alternative splice products in rat liver is consistent, with a ratio of 1.5:1 respectively.

### 3.28 Tissue distribution of human MASP-3 as determined by quantitative PCR

The relative expression of MASP-1/3 alternative splice products was analysed in human tissue total RNA by quantitative PCR, cDNA was generated by random primed reverse transcription and normalised with respect to the house keeping gene GAPDH. Specific PCR products spanning the common MASP-1/3 region into the specific MASP-1 and MASP-3 sequence were generated, the quantity of PCR product analysed after each cycle via the Lightcycler (Roche) apparatus as described. The abundance of specific PCR product was compared against serial dilutions of cloned MASP-1 and MASP-3 cDNA standard curve, run in parallel with test samples. Samples were analysed in triplicate.

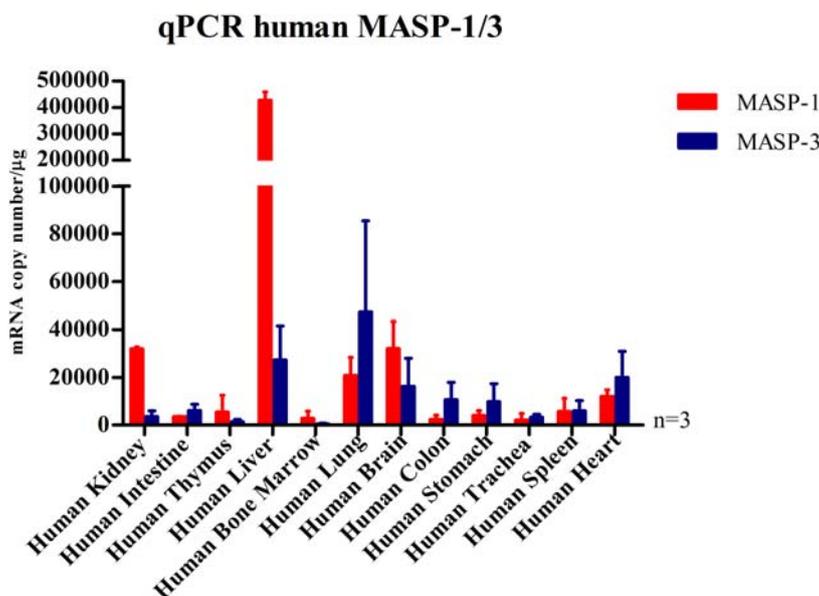


Fig.73 Quantitative RT PCR analysis of MASP-1 and MASP-3 mRNA expression in a panel of human tissue total RNA preparations. A total of 0.5µg total RNA was used per sample equally. Results are expressed as relative mRNA copy number, and are the mean of 3 experiments. MASP-1 represented by a red bar, MASP-3 represented by a blue bar and error bars represent SD.

Examining firstly MASP-1 mRNA, expression in human tissue is relatively constrained to the liver, with kidney, lung and brain tissue the next most abundant sites of mRNA production. MASP-3 expression is relatively less abundant, exhibiting a broader range of expression. Significant levels of expression can be detected in lung, liver, heart and brain with lesser expression in colon and stomach tissue. MASP-3 mRNA expression

in lung and heart tissue are exceptional in that they exceed that of the MASP-1 transcript. The significance of local serine protease expression is to be determined, with corroboration of the MASP-3 mRNA expression pattern in other species the first step in determining the function.

### 3.29 Recombinant expression of human and rat MASP-3 serine protease domain

The prokaryote expression system pRSET (Invitrogen, CA) was chosen to rapidly produce recombinant protein fragments of both human and rat MASP-3, to provide material for production of polyclonal antibodies and for potential structural and functional studies. The pRSET vector drives recombinant protein expression via the T7 promoter, downstream an N-terminal 6xHis residue peptide enables affinity purification of expressed proteins cloned in frame within the vector multiple cloning site (MCS). Utilising the sequence information derived from the work described in chapter 3 and that held in the public DNA databases, PCR primers were designed to amplify specific regions of the MASP-3 cDNA from both human and rat species. Included at the 5' end of the PCR primers were specific restriction endonuclease sequences enabling cloning of the MASP-3 serine protease domain and smaller sub fragments thereof, within the pRSET multiple cloning site (MCS) and in frame with the vectors N-terminal fusion peptide that includes a 6xHIS tag for affinity purification.

Primers including restriction sites were designed to generate expression cassettes encompassing the CCPII and serine protease domain of human and rat MASP-3 (hM3CCPII fwd *Bam*HI, rM3CCPII fwd *Bam*HI), the serine protease domain (hM3 fwd *Bam*HI, hM3 rev *Eco*RI, rM3 fwd *Bam*HI, rM3 rev *Eco*RI) and a C-terminal variable region (hM3 fwd TXP *Bam*HI, hM3 rev *Eco*RI, rM3 TXP fwd *Bam*HI, rM3 rev *Eco*RI).

Polymerase Chain Reaction template for the amplification of rat specific sequence was the cDNA clone 7.1, human total liver cDNA provided the template for human MASP-3 specific amplification. A series of MASP-3 specific regions were chosen for amplification (Fig.74, Fig.75).

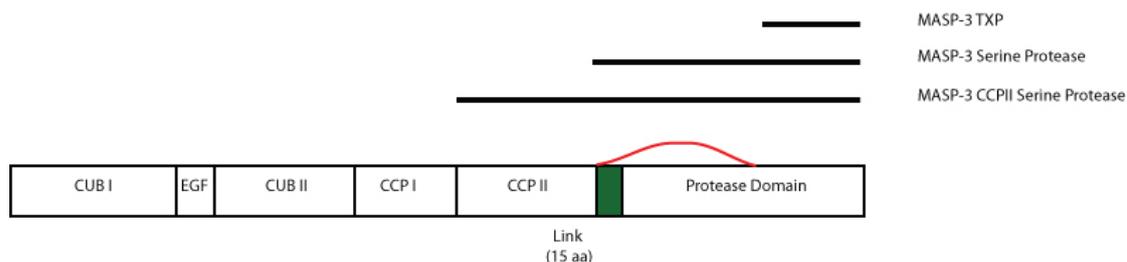


Fig.74 Schematic diagram of the cloning strategy for prokaryotic recombinant expression of rat and human MASP-3. Lower panel indicates MASP-3 protein domain structure, upper panel indicating the

relative length of each prokaryotic expression cassette (bold line, labelled accordingly) with respect to the MASP-3 domain structure.

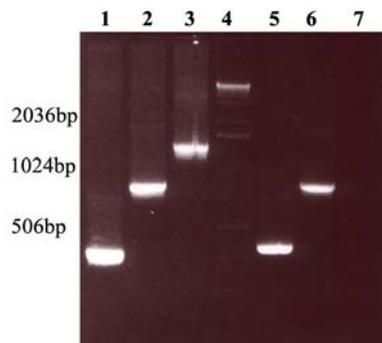


Fig.75 Agarose gel electrophoresis of human and rat MASP-3 expression cassette PCR (as presented in Fig.74) for restriction endonuclease directed cloning into pRSET prokaryote expression vector.

Lane 1 - human MASP-3 TXP, human MASP-3 serine protease domain, lane 3 - human MASP-3 CCPII-serine protease, lane 4 - 1kb DNA marker, lane 5 - rat MASP-3 TXP, lane 6 - rat MASP-3 serine protease domain, lane 7 - rat MASP-3 CCPII-serine protease domain.

Polymerase chain reactions utilising the above primer combinations, produced amplification products of the expected size as seen in Fig.74. TXP fragment 405bp, serine protease domain 887bp and CCPII serine protease domain 1.1kb. The above amplification strategy was successfully employed to generate expression cassettes for human and rat MASP-3 specific fragments encompassing the serine protease domain and the truncated C-terminal fragment (TXP), a human MASP-3 CCPII serine protease domain product was also generated. Despite PCR optimisation a Rat CCPII serine protease domain product could not be produced. Expression vector construction continued in parallel with human and rat MASP-3 constructs encompassing the serine protease domain and the TXP fragment.

Potential pRSET MASP-3 clones as determined by restriction endonuclease digestion were subject to DNA sequencing using a pRSET specific oligonucleotide. Enabling characterisation of the correct sequence and confirmation that the expression cassette was correctly cloned in frame with the pRSET N-terminal leader peptide. Appropriately constructed positive clones were transformed into the expression strain of *E.coli* BL21 and pilot expression experiments performed.

Small scale pilot expression experiments were performed to confirm the properties of the expressed MASP-3 protein fragments. Expression cassettes from both human and

rat, encompassing the serine protease domain and a C-terminal fragment (MASP-3 TXP), were examined. The experiment was performed as described, bacterial culture samples were taken prior to induction and at 2 hours post induction of recombinant protein expression by addition of IPTG. Bacterial pellets were prepared from each time point and the supernatant discarded, each sample was resuspended to an OD<sub>550</sub> of 10. Normalisation of each sample with respect to cell numbers allowed direct comparison of relative protein expression levels on coomassie gel staining.

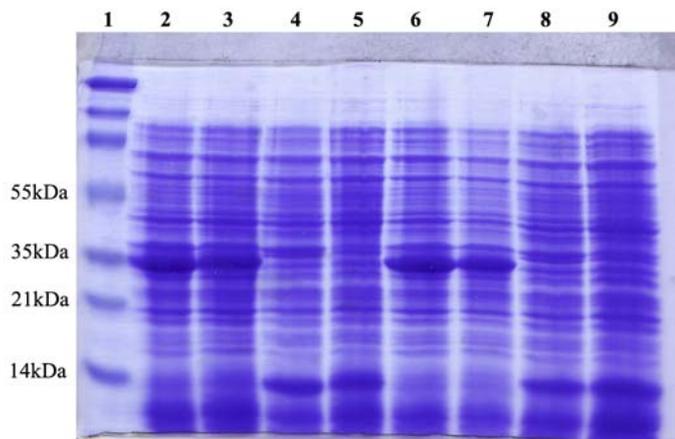


Fig.76 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human and rat MASP-3 serine protease domain prokaryotic cell lysate. Cell lysate of pRSET human and rat MASP-3 serine protease domain were harvested at T0 and T2 hours post induction with IPTG. Lane1 - protein standard, lane 2 - T0 human MASP-3 serine protease domain, lane 3 - T2 human MASP-3 serine protease domain, lane 4 - T0 human MASP-3 TXP, lane 5 - T2 human MASP-3 TXP, lane 6 - T0 rat MASP-3 serine protease domain, lane 7 - T2 rat MASP-3 serine protease domain, lane 8 - T0 rat MASP-3 TXP, lane 9 - T2 rat MASP-3 TXP.

Major protein bands could be identified for both MASP-3 serine protease and MASP-3 TXP expression constructs for human and rat species. These were of sizes 35kDa and 15kDa respectively, as predicted from the primary protein sequence including the N-terminal pRSET vector leader peptide (Fig.76).

Expression from both the rat and human MASP-3 constructs exhibited a high degree of promoter leakage, with almost constitutive expression of recombinant protein regardless of IPTG induction of the Lac operon. Strategies to reduce this leakage, including addition of glucose to the expression media, culture of expression clones at lower temperature and trial of different expression construct clones, failed to reduce this constitutive expression.

### 3.30 Expression time course of recombinant human MASP-3 serine protease domain and C-terminal region in pRSET vector

Larger scale expression experiments were undertaken using human MASP-3 constructs (MASP-3 serine protease and MASP-3 TXP). Successful production of recombinant protein was observed with expected molecular weights. Extended expression up to 4 hours from both constructs failed to significantly increase recombinant protein yield, and as observed with previous MASP-3 pRSET constructs there was considerable leakage from the *Lac* promoter prior to formal induction with IPTG.

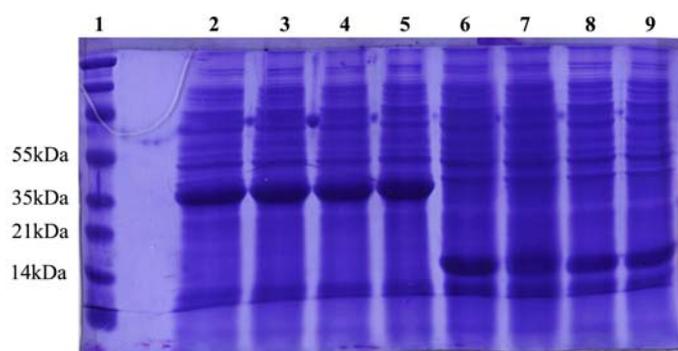


Fig.77 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human MASP-3 serine protease domain and pRSET human MASP-3 TXP prokaryotic cell lysate. Cell lysate of pRSET human serine protease domain and pRSET human MASP-3 TXP were harvested at T0, T1, T2, T3 and T4 hours post induction with IPTG. Lane 1 - protein standard, lane 2 - human MASP-3 serine protease domain T0, lane 3 - human MASP-3 serine protease domain T1, lane 4 - human MASP-3 serine protease domain T2, lane 5 - human MASP-3 serine protease domain T4, lane 6 - human MASP-3 TXP T0, lane 7 - human MASP-3 TXP T1, lane 8 - human MASP-3 TXP T2, lane 9 - human MASP-3 TXP T4.

Despite considerable constitutive expression, as evidenced in Fig.76 and Fig.77, from the expression constructs' promoter, substantial amounts of recombinant protein were produced of expected molecular weight using this method. These experiments show that during the time course of expression for both the MASP-3 serine protease and MASP-3 TXP fragment, there is no significant increase in the yield of recombinant protein with prolonged induction times. These experiments allowed optimisation of further recombinant protein production using these constructs.

### 3.31 Solubility screening of human MASP-3 serine protease expression products

The availability of recombinant MASP-3 serine protease domain in a native soluble form is vital for further studies of enzyme structure and function. Material from expression experiments was used to determine the solubility of recombinant MASP-3 fragments, shaping the subsequent purification strategy.

Recombinant expression products were sampled and individually disrupted by sonication in a series of buffer combinations, with the aim of identifying optimal conditions to release the recombinant protein in a soluble and native form. Visualisation of solubilised material using PAGE separation and coomassie blue protein staining from cell lysate, allowed the identification of conditions producing the greatest yield of solubilised native recombinant protein.

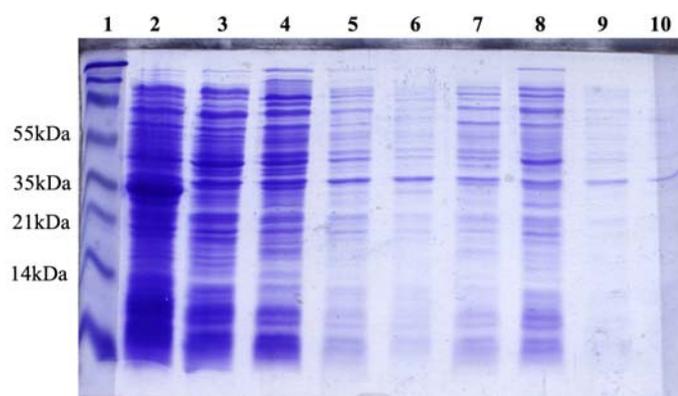


Fig.78 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET human MASP-3 serine protease domain cell lysate at T1 post induction. Prokaryotic cell pellets prepared in alternate buffers (each included 150mM NaCl) to maximise recombinant protein solubility. Lane 1 - protein standard, lane 2 - human MASP-3 serine protease domain T1, lane 3 - human MASP-3 serine protease domain T1 150mM Tris pH 6.8, lane 4 - human MASP-3 serine protease domain T1 M3 150mM Tris pH 7.0, lane 5 - human MASP-3 serine protease domain T1 M3 150mM Tris pH 7.5, lane 6 - human MASP-3 serine protease domain M3 150mM Tris pH 8.8, lane 7 - human MASP-3 serine protease domain T1 50mM Tris pH 6.8, lane 8 - human MASP-3 serine protease domain T1 M3 50mM Tris pH 7.0, lane 9 - human MASP-3 serine protease domain T1 50mM Tris pH 7.5, lane 10 - human MASP-3 serine protease domain T1 50mM Tris pH 8.8.

Expression cell pellets producing recombinant human MASP-3 serine protease were individually resuspended in a series of Tris buffers of varying pH, each were disrupted by ultrasonication to release intracellular proteins.

Fig.78 indicates the solubility of recombinant human MASP-3 in a series of Tris buffers, in comparison to a reference sample solubilised in 0.1% SDS (PAGE loading buffer), a series of sodium phosphate buffers using the same pH range gave identical results. The reference sample clearly exhibits an over expressed protein product of the expected size for the pRSET human MASP-3 recombinant protein. With reference to the major protein bands visualised in the series of Tris buffers of varying pH, there is no corresponding major protein visualised at the expected molecular weight. The buffer systems used to solubilise humans MASP-3 recombinant in this experiment were not able to produce significant quantities of soluble and native protein for further analysis.

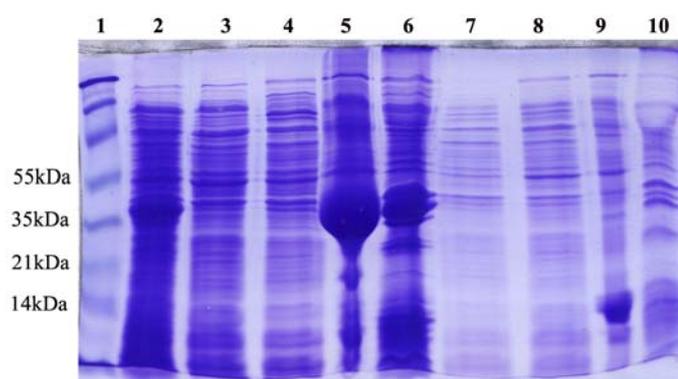


Fig.79 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of recombinant pRSET rat MASP-3 serine protease domain and rat MASP-3 TXP cell lysate at T1 post induction. Prokaryotic cell pellets prepared in alternate buffers (each included 150mM NaCl) to maximise recombinant protein solubility. Lane 1 - protein standard, lane 2 - rat MASP-3 serine protease domain T1, lane 3 - rat MASP-3 serine protease domain pH6.8, lane 4 - rat MASP-3 serine protease domain pH 7.5, lane 5 - rat MASP-3 serine protease domain 6M Urea, lane 6 - rat MASP-3 serine protease domain T1 pellet, lane 7 - rat MASP-3 TXP T1, lane 8 - rat MASP-3 TXP pH 7.5, lane 9 - rat MASP-3 TXP 6M Urea, lane 10 - rat MASP-3 TXP T1 pellet.

Experiments were undertaken to identify a suitable buffer to maintain recombinant rat MASP-2 serine protease domain and serine protease TXP fragment in a soluble form, enabling further functional and structural analysis.

Expression cell pellets were harvested and resuspended in appropriate buffer, recombinant proteins were release by ultrasonication and the soluble fraction made available for further analysis.

Figure 79 analyses the solubility of the rat recombinant proteins in a series of buffers, with reference to expression cell pellets solubilised in 0.1% SDS. Direct visualisation of the major expressed band in the reference sample in comparison to the buffer combinations clearly demonstrates that none are able to produce native soluble fractions

with significant amounts of recombinant material. In contrast for both rat MASP-3 serine protease domain and MASP-3 TXP fragment, a significant amount of protein could be maintained in the soluble fraction with 6M urea. The use of urea to solubilise recombinant proteins successfully increased the yield of proteins available for further purification, but was not suitable for structural or functional assays due to its denaturing properties.

Despite screening a series of buffer combinations none were identified that would produce enough soluble material for either structural or functional assays, further purification of recombinant MASP-3 was achieved under denaturing conditions. This method provided sufficient and suitable material for antibody production but was not appropriate for either structural or functional analysis of recombinant MASP-3 properties.

### 3.32 Purification of rat and human MASP-3 serine protease fragments under denaturing conditions

Large scale expression of recombinant MASP-3 fragments was undertaken as described, after suitable protein induction the cell pellet was harvested and resuspended in denaturing IMAC purification buffer prior to disruption via ultrasonication. The cell lysate was ultracentrifuged and the pellet discarded. The resulting soluble fraction was passed through a 22 $\mu$ M filter. The soluble filtrate was applied to an IMAC affinity column equilibrated in denaturing buffer; non-binding proteins were washed from the column with 15 column volumes of denaturing buffer. Specific elution of affinity bound protein was achieved by increasing concentrations of Imidazole + denaturing buffer being passed through the column. Eluted samples were collected and analysed by PAGE and coomassie staining. Recombinant fragments of MASP-3 were eluted at the lowest concentration used of 50mM Imidazole. The expression constructs generated allowed the production 2-4mg/l of recombinant protein per litre of culture medium (as determined by coomassie gel staining), this provided suitable and sufficient reagents for further use in a purified form.

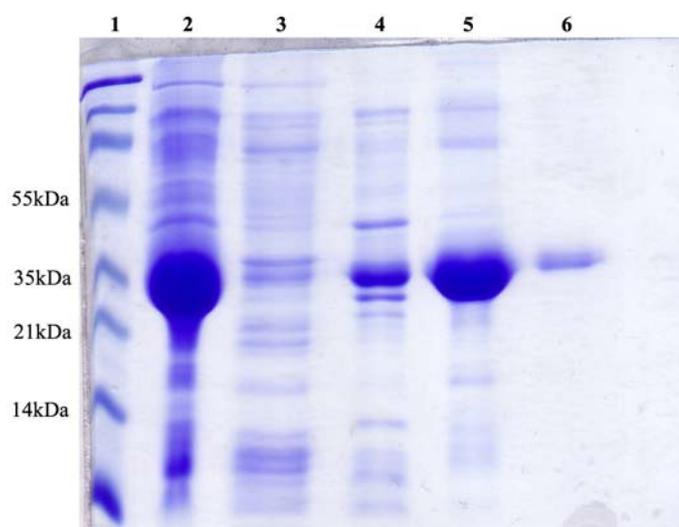


Fig.80 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET rat MASP-3 serine protease domain T1 post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea). Lane 1 - protein standard, lane 2 - rat MASP-3 serine protease domain T1, lane 3 - rat MASP-3 serine protease domain T1 non-bound, lane 4 - rat MASP-3 serine protease domain T1 50mM imidazole elution, lane 5 - rat MASP-3 serine protease domain T1 200mM imidazole elution, lane 6 - rat MASP-3 serine protease domain T1 500mM imidazole elution.

Rat MASP-3 serine protease pRSET expression construct was successful in the production of recombinant protein. Under denaturing conditions the recombinant protein was prepared in a soluble form for affinity purification, precolumn sample lane 1 figure 72. The 6xHis affinity tag fusion to the N-terminal Rat MASP-3 serine protease, exhibited specific binding to the IMAC affinity matrix. Specific elution of the Rat MASP-3 recombinant protein was achieved using imidazole at varying concentrations (lane 3, 4 and 5 figure 80), the greatest elution occurred at a concentration of 200mM imidazole.

This single step purification strategy under denaturing conditions was successful in the purification of a highly insoluble recombinant protein, producing large amounts of purified protein for further applications.

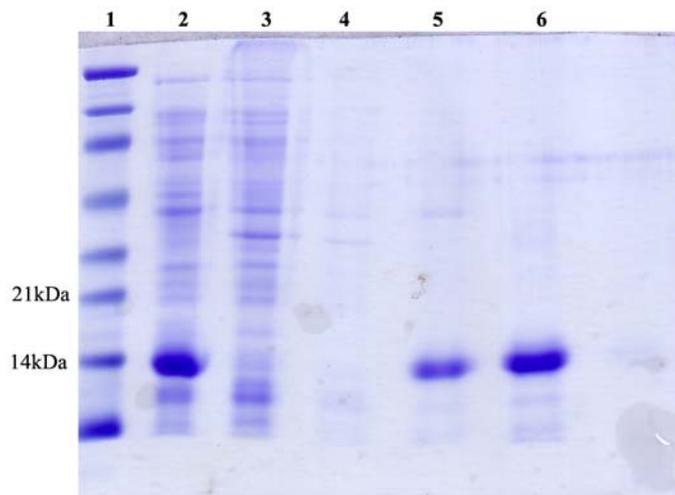


Fig.81 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET rat MASP-3 TXP post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea). Lane 1 - protein standard, lane 2 - rat MASP-3 TXP T1, lane 3 - rat MASP-3 TXP non-bound, lane 4 - rat MASP-3 TXP wash, lane 5 - rat MASP-3 TXP 50mM imidazole elution, lane 6 - rat MASP-3 TXP 250mM imidazole elution.

Rat MASP-3 serine protease domain TXP fragment was purified under denaturing conditions, utilising 6M urea. A soluble fraction of recombinant rat TXP protein was prepared as described, prior to application to the affinity matrix a major over expressed band could be visualised (Fig.81, lane 2) corresponding to the rat MASP-3 serine protease domain TXP fragment. Samples were analysed after application to the affinity matrix column, the non-binding fraction can be seen in lane 3 with the major visible protein band absent. Protein washed from the column can be seen in lane 4, representing non-specific binding to the column matrix. Specific elution from the

affinity column was achieved using increasing concentrations of imidazole. The major band is seen to specifically elute from the column in the presence of imidazole, the maximum elution occurring at a concentration of 200mM, this corresponds with the expected molecular weight predicted for rat MASP-3 serine protease TXP fragment. Fig.81 demonstrates the purification of rat MASP-3 TXP fragment under denaturing conditions, producing a highly purified protein for further analysis.

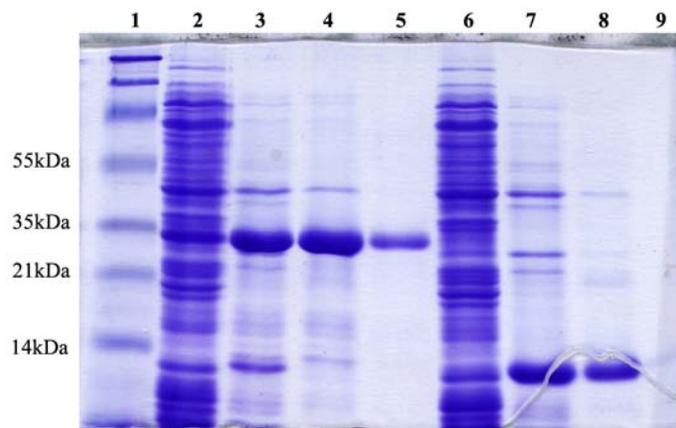


Fig.82 Coomassie blue stained polyacrylamide gel electrophoresis (PAGE) of IMAC purified recombinant pRSET human MASP-3 serine protease domain and human MASP-3 TXP post induction, under denaturing conditions (50mM Tris 250mM NaCl, 6M Urea).

Lane 1 - protein standard, lane 2 - human MASP-3 serine protease domain precolumn, lane 3 - human MASP-3 serine protease domain elution 50mM imidazole, lane 4 - human MASP-3 serine protease domain elution 250mM imidazole, lane 5 - human MASP-3 serine protease domain elution 500mM imidazole, lane 6 - human MASP-3 TXP precolumn, lane 7 - human MASP-3 TXP elution 50mM imidazole, lane 8 - human MASP-3 TXP elution 250mM imidazole, lane 9 - human MASP-3 TXP elution 500mM imidazole.

Human MASP-3 serine protease and human MASP-3 TXP fragment were successfully purified under denaturing conditions, utilising 6M urea. Soluble fractions were prepared from expression cell pellets, and applied to an IMAC affinity matrix. Specific protein binding was assessed visually via coomassie blue stained polyacrylamide gel electrophoresis (PAGE), in comparison to recombinant expression whole cell lysate solubilised in 0.1% SDS (Fig.82 lane 2 and lane 6, human MASP-3 SP, human MASP-3 TXP). Specific purification of recombinant protein was achieved; proteins of expected sizes were eluted from the affinity matrix (Fig.82 lane 3-5, lane 7-9). Proteins were prepared in a highly concentrated form, separated from many contaminating bacterial proteins.

Human MASP-3 serine protease domain and human MASP-3 serine protease domain TXP fragment were successfully purified under denaturing conditions, in a concentrated form for using in further applications.

Recombinant MASP-3 serine protease domain was tested using a Rat (Rat 65 sera, Mads Dahl) polyclonal antibody raised against the N-terminal region of purified MASP-3, exhibiting specific binding.

The above work enabled the successful generation of specific MASP-3 expression constructs for both rat and human protein products. Recombinant proteins were produced and a purification strategy developed to generate suitable material for subsequent use in the production of specific polyclonal antibodies. Due to the limitations of the prokaryotic expression system, it was not possible to produce protein material that was suitable for functional or structural studies.

# Discussion

## Chapter 4

This doctoral thesis presents the cloning, characterisation and recombinant expression of a novel complement serine protease, MASP-3 in a number of mammalian and non-mammalian species published in the work (Stover, Lynch et al. 2003). MASP-3 is an alternatively spliced transcript of the MASP-1/3 locus. Both MASP-like serine protease are constituents of multimolecular plasma complexes encompassing MBL or ficolins, MASP-2 and Map19. The degree of conservation amongst MASP-3 sequence derived from diverse species is examined as are sequence motifs important in protease activity, intra and intermolecular associations. The structure of MASP-1/3 locus is examined and compared between species. The relative expression of alternative splice transcripts MASP-1 and MASP-3 are compared as is their relative mRNA abundance in a wide range of organs. Recombinant expression and purification of MASP-3 aims to elucidate functional activity of the novel serine protease.

The identification of human MASP-3 as a novel component of the lectin pathway activation apparatus highlighted the importance of clarifying the nature of the lectin pathway components in other species particularly those model organisms required for the study of complement function. Many conflicting reports exist concerning the enzymatic function of lectin pathway activation complexes; the additional uncertainty of a previously unidentified component associating with these complexes required analysis and aimed to bring new light to previous experimental data.

The primary MASP-3 cDNA sequence of 7 mammalian species have been confirmed by laboratory experiment, a further 22 derived sequences have been determined utilising bioinformatic approaches based on the currently available genomic sequence databases. MASP-like serine protease sequences were sought in a similar manner if not currently available in reference sequence databases. Of the sequence information gathered, it was possible to confirm the existence and construct complete MASP-3 sequence information for 12 species, a further 10 partial sequences were available for analysis of the MASP-3 specific serine protease domain.

The diversity and scope of species in which MASP-3 specific transcripts were identified is appreciable when we consider the phylogenetic classification of organisms in which cDNA sequence was characterised, encompassing species of the class Mammalia, order Rodentia, rat, mouse (family muridae) and guinea pig (family caviidae). Order

Lagomorpha, Rabbit (family leporidae). Order Artiodactyla, Cow (family bovidae) and Pig (family suidae). Class Amphibia, order Anura, Xenopus (family pipidae). Class Actinopterygii, order cypriniformes, Danio rerio (family cyprinidae).

Similarly analysis of MASP-like serine protease cDNA sequence derived from genomic sequences further broadened the scope of organisms confirmed to possess MASP-3 specific loci, including species of the class Mammalia, order primate, chimpanzee, orang-utan (family hominidae), rhesus monkey (family cercopithecidae), marmoset (family callitrichidae). Order carnivore, dog (family canidae) and cat (family felidae). Order perissodactyla, horse (family equidae). Order proboscidea, elephant (family elephantidae). Order Didelphimorphia, opossum (family didelphidae). Order monotremata, platypus (family ornithorhynchidae). Class aves, order galliformes, chicken (family phasianidae) and order passeriformes, zebra finch (family estrildidae). Class reptilia, order squamata, Anolis carolinensis (family Polychrotidae). Taxonomic descriptions were obtained from the NCBI ([ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)).

This work clearly presents the identification, cloning and characterisation of the alternative splice product of the MASP-1/3 locus, MASP-3 in a series of mammalian and non-mammalian species, including the majority of laboratory model organisms. As discussed the prototypical MASP-3 cDNA was conclusively identified in humans by Dahl (Dahl, Thiel et al. 2001) as a component of the lectin pathway of complement activation residing in the MASP-1/3 loci as an alternative splice product, there is evidence in the literature that previous studies in other species may have identified the existence of a mRNA transcript closely related to MASP-1. The research groups of Kwakami and Ramadori investigating the complement serine protease MASP-1 (Ra reactive factor, P100) observed two species of mRNA transcripts of differing size on northern blot analysis of human (Takada, Takayama et al. 1993), mouse (Takayama, Takada et al. 1994) and rat (Knittel, Fellmer et al. 1997) liver RNA, at the time postulated to be alternatively polyadenylated MASP-1 mRNA are more convincingly due to alternatively spliced MASP-3 mRNA.

The knowledge that a homologous constituent of a complex multimolecular protein structure is present or absent within a particular organism under investigation is vital if the correct analysis of experimental data is to be achieved. The propensity of a

particular species to be susceptible to certain pathogens or disease processes is more appropriately assessed in the full knowledge of the components of an important arm of immune defence. In common with this the identification of species in which there is absence of MASP-3 or retention of MASP-3 in preference to other components of the lectin pathway cascade, taken in context with the stage of development of the adaptive immune response, provides vital clues to the role of MASP-3 within those specific species.

The results discussed in this section highlight the diversity of organisms in which it was possible to identify MASP-3 specific sequence, this systematic approach has enabled one of the primary aims of this thesis to be achieved by documenting the current knowledge regarding organisms proposed to possess MASP-3 specific sequence.

### **Genomic structure of MASP-1/3 loci**

All mammalian MASP-3 gene loci examined, where sufficient sequence information was available, were present as an alternatively spliced partner in a MASP-1/3 locus, conserving the prototypical organisation exhibited by human MASP-3 (Dahl, Thiel et al. 2001). Of the mammalian species where MASP-3 cDNA sequence was confirmed, it was possible to generate a complete MASP-1/3 locus model for 10 species. Analysis of the MASP-1/3 loci in mammals highlights not only the conservation of an association with MASP-1 specific sequence, but also of the surrounding loci which exhibit linkage in all mammalian genomes available for study.

A number of non-mammalian species in which MASP-3 specific sequence was identified did not share the prototypical human MASP-1/3 locus configuration, in avian species (gallus and taeniopygia) and bony fish (danio and gasterosteus) MASP-3 is present in absence of MASP-1. Extensive sequence analysis of the genomic region surrounding the identified MASP-3 locus in each species proposed to exhibit an alternative locus structure was unable to identify MASP-1 specific genomic sequence. The genomic structure of gallus MASP-3 has been characterised (Lynch, Khan et al. 2005), in conjunction with northern blot analysis, confirming the absence of a specific MASP-1 transcript. The alternative nature of the MASP-3 locus in the remaining

species does not preclude the organism expressing a MASP-1 specific mRNA from a locus at a distant site within the genome, although none were identified, the incomplete assembly of the genomic sequence allows for this possibility. Further experimental work as has been carried out in the case of *Gallus gallus* will be required to clarify the finding presented in this work.

The MASP-1 specific exons are proposed to be absent as a consequence of genomic rearrangement, with either direct loss during that event or subsequent loss during evolution as a result of the MASP-1 specific domains inability to form a transcript including the MASP-1/3 common domains. The complement of adjacent gene loci is disturbed in the avian genomes examined, the region of synteny between mammalian and avian genome ends downstream of the avian MASP-3 specific exon (Lynch, Khan et al. 2005). Findings indicate that the region may have been subject to a translocation event, separating MASP-3 and MASP-1 specific exons. MASP-1 specific exons may have been lost during the translocation event or at a later stage.

The configuration of avian MASP-3 loci is an example of the importance of correctly assessing the components of a particular system under investigation, and the light this may direct on to poorly understood roles of other proteins. This is especially relevant given the recent identification of the vital function MASP-1 has in the activity of the alternative pathway of complement activation, being responsible for the activation of factor D (Takahashi, Ishida et al. 2010) in the murine system. Therefore posing the question of which protein is responsible for Factor D cleavage activity in absence of MASP-1, and once identified does this protein retain Factor D activity in other species and also its usual functional activity. Further experimental data characterising complement activity in these species is required.

The characterisation of two divergent class of organisms that posses an alternate MASP-1/3 locus configuration and are proposed deficient of MASP-1, indicate that this may be a more general phenomenon within these groups of organism.

An important positive finding with respect to the anole lizard is the presence of the MASP-1 specific exon within the genome, confirming that the loss of the MASP-1 specific exon in avian species occurred after the divergence of these two species

common ancestor. This indicates that the events that formed the MASP-1/3 locus in bony fish are separate from those that formed the avian locus; caution must be applied to analysis of bony fish genomics given the genome wide duplications they have been subject to (Taylor, Van de Peer et al. 2001). To delineate the exact nature of the similarities and differences between avian and bony fish MASP-3 loci requires analysis of more complete genomic sequence data from a number of additional species of bony fish, this will provide conformation of the absence of MASP-1 specific sequence at an alternate site within the genome. The absence of MASP-1 within the bony fish would indicate that two class of organisms had independently removed the MASP-1 locus from their genome, this taken in context with the vital function of MASP-1 in the alternative pathway (Takahashi, Ishida et al. 2010). As C1r and C1s homologues are present in bony fish species, duplication of a MASP-3 like protein had already occurred in this class of organism (Boshra, Gelman et al. 2004). This indicate that there is loss of the MASP-1 specific sequence rather than a model whereby those sequence have yet to associate with the MASP-3 specific sequence, this finding does not invalidate the models proposed by Nonaka (Nonaka 2001).

### **Conservation of MASP-like serine protease domain architecture in diverse species**

MASP-3 was identified in all species examined within this work, expanding the number of species confirmed to possess this gene product from one to more than twenty. The structure of the MASP-3 specific transcript in all species where sufficient sequence information was available, with the exception of teleost fish, exhibited conservation with respect to the prototypical human MASP-3 (Dahl, Thiel et al. 2001).

In the majority of species examined MASP-3 contained a total of 11 exons encoding 6 distinct protein domains, the MASP-3 specific serine protease domain in common with the C1r and C1s MASP-like serine protease is encoded by a single exon. This is in contrast to the MASP-1 specific serine protease domain previously characterised (human, mouse and rat) in and genomic sequence information examined within this work (guinea pig, rabbit, cow, pig, equus and primate), which exhibit a conserved split exon structure consisting of 6 exons.

The split exon structure of MASP-1 serine protease domain is mirrored in the teleost fish, tetraodon transcript 197 utilising a dual exon protease domain whilst tetraodon transcript 199 and gasterosteus both exhibit 3 exons within the serine protease domain. The bioinformatic finding of multiple domain serine proteases with significant identity to MASP-3 in the teleost fish must be clarified by laboratory experimentation, to determine if they represent an intermediate step in the evolution of MASP-3 and MASP-like serine protease towards the modern single exon configuration or have arisen due to insertion of intronic sequence within an original single exon protease domain. The identification of MASP-like protease with intermediate gene structure would propose a model whereby there was a gradual loss of intronic sequence, with the eventual formation of a single exon serine protease transcript. A step wise mechanism would preclude the current retrotransposition model of the origin of MASP-like single exon serine protease (Nonaka and Miyazawa 2002).

As discussed previously the multiple whole genome duplications experienced by teleost fish introduces a great deal of plasticity to the evolution of its constituent genetic loci, there is significant relaxation of genetic pressure on a particular loci when multiple copies are present. Therefore the finding presented in this work of multiple domain MASP-3 like serine protease in the teleost fish may represent a unique finding in this class of organisms as a result of this greater genomic plasticity. The advantage gained by possessing a split exon protease domain remains unclear, but conservation of this feature amongst all teleost fish would propose an advantage. Ultimately the determination of the function of MASP-3 in these species maybe required to answer that question.

MASP-3 cDNA sequences derived from this work, enabled predicted translation products for each species to be generated and analysed. Comparisons of the primary amino acid sequences derived from this work with a protein family database (Pfam) (Finn, Tate et al. 2008), identified the conserved domain architecture exhibited by human MASP-3, and all members of the MASP-like serine protease. Serine protease of the classical and lectin pathway activation complex share common structural motifs at the protein level, CUBI-EGF-CUBII-CCPI-CCPII-Serine protease domain architecture is conserved amongst the MASP-like serine protease.

**Conservation of MASP-like serine protease primary structure in diverse species**

Analysis of the MASP-3 specific coding sequence derived from this work, encompassing a wide range of evolutionary diverse species, has enabled predictions to be made regarding the primary amino acid sequence of MASP-3 in these organisms. As presented in the pairwise sequence alignment tables (table.2 - table.8) and for all species examined in this work there is greater conservation of primary amino acid sequence than that of the proposed cDNA sequence. Conservation of sequence in this manner points to retention of function within the MASP-3 primary amino acids sequence. Further evidence of maintenance of function for the MASP-3 specific serine protease domain is proposed when comparing its conservation with that of the MASP-1 protease domain, in all species examined MASP-3 exhibits greater sequence identity with respect to MASP-1. This is a remarkable finding in the context of the deficient knowledge of MASP-3 activity; maintenance of primary amino acid sequence in this fashion would propose a significant function for this protein molecule.

Analysis of MASP-1/3 common protein domain primary amino acid sequence reveals a pattern of sequence conservation between diverse species. Comparing pairwise sequence identity between a distantly related species and the prototypical human primary amino acids sequence, it is clear that the CUBI and EGF domains exhibit greater conservation of sequence in comparison to the remaining shared MASP-1/3 domains. This finding reflects the vital function of these domains in the ability of MASP to associate to form dimers, and to form complexes with recognition components of the complement system. It is clearly demonstrated in this work that amino acid residues required for these interactions are present in all species examined and are localised within the CUBI and EGF domains. The finding that the CUBI and EGF domains exhibit a great degree of sequence conservation in primitive species indicates that the function attributed to modern MASP-like protease also applies to those primitive species, and indeed the last common ancestor of the divergent species. Therefore it would be proposed that MASP-like protease have retained a model of function whereby serine protease dimers are assembled to facilitate interaction with a complement recognition molecule (Phillips, Toth et al. 2009).

**Conservation of MASP-3 primary structure motifs**

The MASP-3 sequence studied within this work exhibit a number highly conserved amino acid residues encompassing the catalytic triad, associated with the active site substrate binding pocket and structural residues important in forming a functional serine protease. MASP-1/3 common region also conserves a number of crucial residues involved in MASP dimer formation and associations with the recognition molecules MBL and ficolin.

Analysis of the entire primary amino acid sequence of a serine protease produces an evolutionary relationship predominantly based on the surface structures and intermolecular associations afforded by residues exposed on the surface of a protein and those involved in modulating ligand association, rather than those involved in the fundamental catalytic mechanism of the serine protease (Krem and Di Cera 2001) whose sequence is invariant within a protein family. Residues involved in the highly conserved catalytic machinery provide useful markers with which to divide serine protease into discrete groups.

Determining the evolutionary relationships amongst serine protease can be facilitated by the analysis of these highly conserved amino acid residues intimately associated with the functional and structural mechanism of the catalytic machinery, such markers exhibit non-random residue selection and/or codon usage enabling progressive changes to be delineated. Correct selection of evolutionary markers enables the categorisation of serine protease according to evolutionarily distinct lineage, and to determine the relative age and order of emergence of those lineage (Krem and Di Cera 2001).

Members of the MASP-like serine protease can be divided into two distinct evolutionary groups based upon the codon usage of the active site serine residue, and the exon structure of the serine protease domain (Endo, Takahashi et al. 1998). An ancestral TCN codon is found in MASP-1 and ascidian MASP, and the serine protease domain is encoded by multiple exons. The descendent C1s, MASP-2 and MASP-3 utilise an AGY codon for the active site serine residue, and possess a serine protease domain encoded by a single exon. The teleost fish are divergent in that they appear to

possess a serine protease domain containing multiple exons, other features of AGY type MASP are present, and as discussed the origins of this structure are to be determined.

Ascidian (*Halocynthia roretzi*) occupies a position intermediate between invertebrate and vertebrate species, AsMASPa and AsMASPb are paralogous MASP gene species (Ji, Azumi et al. 1997) exhibiting structural features common to MASP-1. Ascidian MASP utilise a TCN codon at the active site serine residue, cysteine residues are able to form a histidine loop structure, and the nature of the protease domain exon structure is unknown. The recent characterisation of a single MASP-like serine protease in a sea anemone (*Nematostella vectensis*) NvMASP (Kimura, Sakaguchi et al. 2009) exhibiting features of a MASP-1 serine protease, a TCN active site codon, a split exon protease domain and a histidine loop structure. These ancient MASP-like serine protease represent the most distant ancestors of the MASP-like serine protease family currently characterised, single exon protease domain MASP-like transcripts were absent. The finding of a MASP-1 type protease in these ancient species in absence of other MASP provides further evidence of the primordial nature MASP-1 within the MASP-like serine protease family, indicating that a MASP-1 like protein was a key constituent of the ancestral complement system.

In determining that MASP-1 is in evolutionary terms older than MASP-3, it remains to be determined how the association as alternatively spliced products of a single gene locus arose.

As discussed the conservation of discrete amino acid residues within serine protease can provide valuable information regarding the origins of the related proteins MASP-1 and MASP-3, such evolutionary markers also enable us to position MASP-3 in the context of other serine protease evolution (Krem and Cera 2002). The model proposed by Krem would require that there were two distinct transitions from a common ancestor marker configuration required to achieve the divergent configurations exhibited by MASP-1 and MASP-3, proteins with an intermediate configuration have yet to be identified.

The transition between a TCN serine codon to an AGY type requires at least 2 base substitutions, a coding sequence with an important cellular function would have a significant selective pressure against the loss of function during this transition phase.

The transition between a primordial TCN codon to that of an AGY type, has been proposed to occur via codon capture event (Nonaka and Miyazawa 2002), potentially from a duplicate locus or pseudogene that is less constrained by pressure to maintain protein function and able to achieve nucleotide substitution at the catalytic site without detriment to the organism.

The primordial nature of MASP-1 with respect to MASP-3 is also indicated when examining cysteine residues of the serine protease domain. The ability to form specific disulphide bridge structures also has evolutionary significance. Disulphide bridges associated with a methionine residue as exhibited by the chymotrypsinogen A family, give rise to a methionine loop (Brown and Hartley 1966). MASP-1 is unique amongst the complement activating serine protease; unlike the previously characterised C1r and C1s the enzymatic domain cysteine residues are coordinated with a histidine residue to form a histidine loop structure. Human MASP-3 exhibits a cysteine pattern similar to that found in C1r and C1s, harbouring a methionine loop structure in relation to its active site residues. The number of cysteine residues identified within the rat MASP-3 serine protease domain cDNA sequence is identical to that seen within the human MASP-3 primary sequence. A total 5 located within the serine protease domain. The progressive loss of stabilising disulphide bonds within the catalytic domain of S1 serine protease is thought to be facilitated by the structural stability conferred by the associated light chain domains, in particular CCPII (Harmat, Gal et al. 2004).

Analysis of MASP-3 sequences derived from this work clearly indicate that although there is sequence variation in codon usage of marker residues amongst species, the amino acid residues remain invariant in all species examined and is in agreement with the previously characterised human MASP-3. The features exhibited by previously characterised human MASP-3 are conserved amongst all mammalian and non-mammalian MASP-3 sequences examined; there is highly conserved exon structure with the MASP-3 specific serine protease domain encoded within a single exon, the active site serine residue is invariably of the AGY type.

MASP-1 split-exon serine protease possess a distinct complement of marker residues, the non-random distribution of these markers divides MASP serine protease in to two distinct evolutionary groups.

Applying this classification approach to the MASP-like serine protease determine that MASP-1 serine protease is evolutionarily primordial to its alternative splice partner MASP-3, and to the other MASP single exon serine protease.

### **Phylogenetic Analysis**

The evolutionary history of serine protease have been a recurrent area of study due to their ubiquitous nature in a number of biological processes, and is particularly interesting in the case of the MASP-like protease multicomponent system employed by complement.

Phylogenetic analysis of the currently available MASP-like serine protease reference sequence and derived from genomic sequence databases, enable a comprehensive relationship analysis to be undertaken. The divergence of rat and mouse species is estimated to have occurred 33 million years ago and that of human and rodents 96 million years (Nei, Xu et al. 2001), the common ancestor of fish is significantly older the evolutionary rate is estimated to be lower compared to mammals.

Phylogenetic analysis of the MASP-3 primary amino acid structure, using the primordial MASP-like protease AsMASP as an outgroup, indicate that all transcripts identified remain in a common clad. The clustering of MASP-3 in this way represents a conservation of sequence between species, with the proposal that this conservation at the primary amino acid structure level is replicated at the functional level.

Phylogenetic analysis of the MASP-like serine protease primary amino acid structure ClustalW alignment clearly delineates a total of five major clades. These correspond to each of the MASP-like protease MASP-1, MASP-2, MASP-3, C1r and C1s, the primordial AsMASP and NvMASP are most closely associated with the MASP-1 clade.

Examination of the relationships between MASP-like serine protease aimed to illuminate potential evolutionary relationships between each of the serine protease, gene duplication events are proposed to have occurred in the development of more modern MASP-like protease. Results in this work indicate that MASP-3 is most closely related

to C1r, whilst MASP-2 has most identity to C1s. MASP-1 is separated on its own phylogenetic branch and as previously discussed is thought to be the primordial MASP-like serine protease. These findings propose that duplication of MASP-3 and MASP-2 were required to generate the classical complement components C1r and C1s respectively.

The phylogenetic approach facilitated the assignment of a number of MASP-like protease from distant species to the appropriate clade, based on the ClustalW alignment of the primary amino acids sequence.

### **Relative cDNA expression of the MASP-1/3 locus**

The relative abundance of alternative splice products, MASP-1 and MASP-3 were firstly examined in hepatic tissue. Across a range of concentrations the relative expression of the MASP-1 transcript was consistently 1.5 fold more abundant than that of the MASP-3 mRNA product as determined in total liver RNA (Stover, Lynch et al. 2003). Serum concentrations of both MASP-1 and MASP-3 have been determined, MASP-1 circulating at a mean protein concentration of 6µg/ml (range 1.5-13µg/ml) (Terai, Kobayashi et al. 1997) whilst MASP-3 has a mean concentration of 6.4µg/ml (range 2-12.9µg/ml) (Skjoedt, Palarasah et al. 2009). The difference in abundance of respective MASP-1/3 mRNA species is not reflected in the concentration of circulating protein, both proteins are present in almost identical concentration.

Both transcripts are expressed under the control of the same promoter apparatus, suggesting that differential expression of specific mRNA is mediated at the mRNA processing or post-transcriptional level.

Examination of the structures required for splice site selection across species have indicated a relatively more efficient branch point sequence in the intron preceding the MASP-3 specific exon (Stover, Lynch et al. 2003) composed of a longer polypyrimidine tract prior to the 3' splice site, this in conjunction with a negative regulator of splicing (GGXTpurine) (Helfman, Roscigno et al. 1990) associated with polypyrimidine tract in the intron prior to the first MASP-1 specific exon are at odds

with the relative abundance of MASP-1 transcript identified in this study. The mechanism of the differential expression of alternative splice transcripts MASP-1 and MASP-3 remains to be determined, investigation of the competition between the 3'splice sites of MASP-1 and MASP-3 for the 5'splice site of the second exon of CCPII, suppression of MASP-3 polyadenylation signal recognition and post-transcriptional stability of the MASP-3 mRNA propose questions to be answered.

The relative stability of each MASP-1/3 mRNA species may be a potential post-transcriptional mechanism mediating the observed abundance of MASP-1 mRNA with respect to MASP-3; MASP-1 transcripts with a greater half life are able to achieve a greater concentration due to a shift in the equilibrium between production and degradation of the specific mRNA.

The observed serum protein concentrations of MASP-1 and MASP-3 proteins are remarkably similar in the context of a 1.5x excess of MASP-1 mRNA transcript, implicating variation in translational mechanisms and/or posttranslational protein stability as factors influencing serum protein concentration.

It is important to delineate the contribution of each mechanism impacting on the relative abundance of mRNA and eventual serum protein concentration in the MASP-1/3 gene locus, in this way we can further understand the range of expression observed in a number of organs.

The relative expression of specific MASP-1/3 mRNA species in a panel of total RNA derived from a series of human tissues was examined, revealing the varied nature of mRNA expression between organs. Broadly MASP-3 exhibits a wider range of organ specific mRNA expression than that of MASP-1, the most abundant site of MASP-1 mRNA expression occurred in liver whilst MASP-3 was predominantly expressed in lung tissue. The other major organs of MASP-3 expression were heart and liver. In concordance with the finding in rat liver, MASP-1 is the more abundant transcript in human liver. As discussed above using the example of differential expression in rat total liver RNA, there are multiple mechanisms by which the production of MASP-1/3 mRNA transcripts can be adjusted.

The observation of differential expression of MASP-1/3 alternatively spliced mRNA transcripts in different tissues is intriguing, given the primary nature of both MASP-1 and MASP-3 proteins in that they circulate as serum components of the complement cascade. Therefore MASP-1 and MASP-3 protein expression need not be tissue specific to perform their primary function; merely they require access to the circulation. Restriction of mRNA expression to particular organs implies a local function of the protein product derived from the specific transcript.

The observation of MASP-3 expression in lung tissue is the most striking MASP-1/3 mRNA expression profile; there is significantly greater MASP-3 mRNA than that of MASP-1. Examination of relative MASP-1/3 expression in heart tissue exhibits similarities to that observed in lung tissue but with significantly less mRNA production. The reversal of the usual mRNA expression ratio in conjunction with large amounts of MASP-3 transcript expressed in lung tissue and to a lesser extent in heart tissue is significant and may propose a local function within the lung and heart for MASP-3, this finding is beneficial in directing future research effort. Determination of the specific function of MASP-3 within lung and cardiac tissue will provide valuable insight into the function of MASP-3 in circulation and of the lectin pathway as a whole.

Results within this section achieved two of the key aims of this thesis, namely determination of the differential expression of MASP-1/3 mRNA transcripts with respect to each other and to examine the range of expression in various organs. Results have also proposed a focused examination of lung and heart tissue as future areas of research that may illuminate the function of MASP-3.

### **Recombinant protein expression**

The identification and cloning of species specific MASP-3 cDNA transcripts provided the substrate for generation of recombinant protein expression constructs, with the aim of producing reagents for the further study of MASP-3 function.

Expression cassettes were successfully generated to produce recombinant fragments of the MASP-3 serine protease (human and rat), utilising an inducible prokaryotic

expression system. Optimum expression conditions were sought by examining recombinant protein production as a time course post induction of the Lac promoter. In the expression system used MASP-3 exhibited constitutive expression from the pRSET (Invitrogen, CA) Lac operon, this property was retained in the truncated MASP-3 TXP construct encoding the C-terminal 135aa. It was not possible to further optimise recombinant protein production, efforts to inhibit the Lac promoter by the addition of glucose to the growth media (Data not shown) failed to arrest constitutive protein expression. In all expression constructs, high levels of recombinant protein were observed of the expected molecular mass.

Efforts were undertaken to isolate recombinant MASP-3 protein fragments in a functional form, a series of buffer solutions failed to maintain a significant proportion of the expressed protein in a soluble state when released from the prokaryotic cell.

Eukaryotic extracellular proteins expressed in prokaryotic systems commonly form insoluble aggregates called inclusion bodies, a feature of the rate of protein production, lack of appropriate molecular chaperones and number of available cysteine residues with which to form intermolecular disulphide bonds (Fischer, Sumner et al. 1993).

Recombinant MASP-3 proteins were only held in solution under denaturing conditions, 6M urea maintained proteins in a form facilitating purification but excluded any further functional investigations.

Purification of expressed MASP-3 protein fragments was facilitated by an N-terminal 6xHIS tag cloned in frame with the expression vector promoter and MASP-3 expression cassette, functioning under denaturing conditions to form strong associations with Ni<sup>+</sup> charged Sepharose resin. Recombinant MASP-3 proteins successfully bound to the affinity resin, specific elution produced a purified protein for further application.

This approach enabled the rapid production and purification of large quantities of MASP-3 specific material.

Recombinant production of MASP-3 aimed to investigate some of the functional properties of the serine protease, as well as providing reagents to study MASP-3 protein expression.

Recombinant protein expression of MASP-3 was possible in a prokaryotic expression system, but generation of functional protein was not achieved. The ability to study the functional activity of MASP-3 requires the production of correctly folded protein products; this will require the use of a cell based expression system. The cell culture approach has a number of advantages, but in comparison to the rapid prokaryotic system the protein yields are often significantly reduced.



# Future work

## Chapter 5

Determination of the function of the highly conserved MASP-3 protein.

Confirmation of the absence of MASP-1 transcript in zebra finch and danio rerio and the impact this has on their innate immune function.

Characterisation of the differential expression mechanism of MASP-1/3 alternative splice transcripts.

Determination of the significance of local differential expression of MASP-1/3 gene products.

Characterisation of the truncated MASP-1/3 alternative splice product, predicted in human, mouse, rat and rhesus monkey genome.

# Appendices

## Chapter 6

## Oligonucleotides

## MASP-3 serine protease domain

hM3 fwd BamHI	GGATCCAGCCCTCCCCTCCCTG
hM3 rev EcoRI	GAATTCTCACCGTTCCACCTGGGGCT
rM3 fwd BamHI	GGATCCAGCCCTCCCCTCCCTG
rM3 rev EcoRI	GAATTCTCACCGTTCCACCTGGAGCT

## C-terminal variable region

hM3 fwd TXP BamHI	GGATCCAATGTGACAGTGGATGAGATC
hM3 rev EcoRI	GAATTCTCACCGTTCCACCTGGGGCT
rM3 TXP fwd BamHI	GGATCCAATGTGACAGTGGATGAGATC
rM3 rev EcoRI	GAATTCTCACCGTTCCACCTGGAGCT

hu MASP-3 p6 F	TCGAGAGTGCCAAATGACAAGTG
hu MASP-3 p8 R	TGTGTAGACTCCATAGACCTGCTTGCTGCC
rat MASP-3 5'race GSP I	CCAGGAYKCAGAGAGCAG
rat MASP-3 5'race GSP II	GCBCCRCTCCCRAACCACTTGTC
AI(B18) Fwd4	TCCCGTGTGACAGTCCAAAACCT
zfCUBII REV 2	GCACTGGATTTTCAGGATGAGTCTCCA
zfAI GSPI	GCGTGCACAAAGATTAACAAG
zfAI GSPII	GGCGTCATCCTCACAGTTAAAGCTTTCC
zfAW GSP I	GCAAATCACCTGAAACACA
zfAW GSP II	TCCCATCATGCCCTCTTTAACAGCTTCA
zfBI GSP I	TGGCAGAATAGATGACAGTGT
zfBI GSP II	TCGCCGTAATTTTTCTCTGTATCACCAC
huC1qB F1	CTTCGACCACGTGATCACCA
huC1qB R2	GTCGGTGGCCTGCAGGAAC
AI(B18) fwd5	AATGTAATCGGGAGGTTCAAAGTC
gp rb 5'race GSPI	GCAGCCCTGTACGAGAGC
gp rb 5'race GSPII	GGAACAGGATCTGGACACTGTGGCTCTG
hu SAAfwd	AGGCTTTTGATGGGGCTCGGGAC

hu SAArev	TCCTCCGCACCATGGCCAAAGAATC
gp A fwd	CTGGATGAATGAAGTGCTGGGGAGAA
rb A fwd	CGGGCTTCTCACCTTCTCCTCCAG
M1BCR	CCAAGACACMRYKCCSRCMAGRT
M1BCF	TGYTSCACCABBMACTBGATMC
GP3G 1	GGGAAACTACAGTGTACAGAG
GP3G 2	GCTGGCTACTACGAGGGTGGCAAGGAC
ZF BI FWD	CCGCTGAAGACATTGATGAGTGCCTAA
ZF BI FWD 2	TCCGAATCCCGTGGCTCCTCAA
ZF AI FWD 3	CCGAAGTAACTGGGTGTCTGAGAATGGT
ZF BI REV1	AACGCAAAGGGCCCTCCACTGTG
ZF CUBII FWD	TGGAGACTCATCCTGAAATCCAGTGC
SIL BG FWD2	TGCCTAGCAAGCAGCATGGAATCTTCTA
M3 CON FWD	AACTAYRACMAYGAYATMGCYMTSRTMMAG
M3 CON REV	AGGAACGGTGGGAGYAKCATMGGYSGYGT
RM3BAM	AGCCCTCCCGCGCCTGCCAAACC
RM3ECO	TCACCGTTCCACCTGGAGCTCGAGGGA
RM3RTR	ACCCAGTTCCAGTTCCCTCCA
RM3RTF	AAAGATAAGGCATCGCTGACACCAC
HM1SP R1	CATTCAGGTGTGACAGCATG
HM1SP R2	AGTGACTGGTGGAGGCAGTG
HM3SP R1	AGTGTCTCCACCACTATCAGG
HM3SP R2	CTTGTCATTTGGCACTCTCGA
POV M3R	GGGGGATCTGATGCCTTTCTTTTACA
BOVM3F	CGCAGGACGGGGAGGGAGCTAGA
BOVM3R	CCCGGGGAGGAGGACGATCAG
HSCRII F2	CAAGATGCTCAACAATAACACAGG
POVM3F	GGGCCCTGAAGAATGTGGCAGTAAG
SIL BG FWD	GGGAGCAAGCAGGTCTATGGCGTCTAT
MM3 SEQ	TCGAAATTCCCGCGCATGG
rmhGAPDH_F1	CACAGTCCATGCCATCACTG
gp 5'RACE G1	TATCCTGCAGTATGTCAAGTTA
gp 5'RACE G2	ATAGTGGTGGGGCCCTTGTCATCC
gp 3'RACE G1	TTCCGGCCTCCAATAAT
gp 3'RACE G2	AGAGACCGGGAGGGCTGACCAC
Gp 3' RACE GSPI	GGGAAACTACAGTGTACAGAG
Gp 3' RACE GSPII	GCTGGCTACTACGAGGGTGGCAAGGAC
Rb MASP-3 GSP I	GCGGAGCCTTTGTcATC
Rb MASP-3 GSP II	TGTGGCAAGCAAGCAGGTC



# Bibliography

## Chapter 7

- (WUGSC), W. U. G. S. C. (2006). Pan\_troglodytes-2.1, (WUGSC). **2006**.
- Aaij, C. and P. Borst (1972). "The gel electrophoresis of DNA." Biochim Biophys Acta **269**(2): 192-200.
- Akaiwa, M., Y. Yae, et al. (1999). "Hakata antigen, a new member of the ficolin/opsonin p35 family, is a novel human lectin secreted into bronchus/alveolus and bile." J Histochem Cytochem **47**(6): 777-86.
- Alder, M. N., I. B. Rogozin, et al. (2005). "Diversity and function of adaptive immune receptors in a jawless vertebrate." Science **310**(5756): 1970-3.
- Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.
- Annells, M. F., P. H. Hart, et al. (2004). "Interleukins-1, -4, -6, -10, tumor necrosis factor, transforming growth factor-beta, FAS, and mannose-binding protein C gene polymorphisms in Australian women: Risk of preterm birth." Am J Obstet Gynecol **191**(6): 2056-67.
- Appella, E., I. T. Weber, et al. (1988). "Structure and function of epidermal growth factor-like regions in proteins." FEBS Lett **231**(1): 1-4.
- Arora, M., E. Munoz, et al. (2001). "Identification of a site on mannan-binding lectin critical for enhancement of phagocytosis." J Biol Chem **276**(46): 43087-94.
- Arosa, F. A., O. de Jesus, et al. (1999). "Calreticulin is expressed on the cell surface of activated human peripheral blood T lymphocytes in association with major histocompatibility complex class I molecules." J Biol Chem **274**(24): 16917-22.
- Atkinson, A. P., M. Cedzynski, et al. (2004). "L-ficolin in children with recurrent respiratory infections." Clin Exp Immunol **138**(3): 517-20.
- Bajtaj, Z., M. Jozsi, et al. (2000). "Mannan-binding lectin and C1q bind to distinct structures and exert differential effects on macrophages." Eur J Immunol **30**(6): 1706-13.
- Bateman, A., E. Birney, et al. (2000). "The Pfam protein families database." Nucleic Acids Res **28**(1): 263-6.
- Bateman, A., L. Coin, et al. (2004). "The Pfam protein families database." Nucleic Acids Res **32**(Database issue): D138-41.
- Baumgarth, N., O. C. Herman, et al. (1999). "Innate and acquired humoral immunities to influenza virus are mediated by distinct arms of the immune system." Proc Natl Acad Sci U S A **96**(5): 2250-5.
- Bergmann, O. J., M. Christiansen, et al. (2003). "Low levels of mannose-binding lectin do not affect occurrence of severe infections or duration of fever in acute myeloid leukaemia during remission induction therapy." Eur J Haematol **70**(2): 91-7.
- Bersch, B., J. F. Hernandez, et al. (1998). "Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family." Biochemistry **37**(5): 1204-14.
- Bobak, D. A., T. A. Gaither, et al. (1987). "Modulation of FcR function by complement: subcomponent C1q enhances the phagocytosis of IgG-opsonized targets by human monocytes and culture-derived macrophages." J Immunol **138**(4): 1150-6.
- Bobofchak, K. M., A. O. Pineda, et al. (2005). "Energetic and structural consequences of perturbing Gly-193 in the oxyanion hole of serine proteases." J Biol Chem **280**(27): 25644-50.
- Boniotto, M., S. Crovella, et al. (2000). "Polymorphisms in the MBL2 promoter correlated with risk of HIV-1 vertical transmission and AIDS progression." Genes Immun **1**(5): 346-8.

- Bork, P. (1991). "Complement components C1r/C1s, bone morphogenic protein 1 and *Xenopus laevis* developmentally regulated protein UVS.2 share common repeats." *FEBS Lett* **282**(1): 9-12.
- Bork, P. and G. Beckmann (1993). "The CUB domain. A widespread module in developmentally regulated proteins." *J Mol Biol* **231**(2): 539-45.
- Boshra, H., A. E. Gelman, et al. (2004). "Structural and functional characterization of complement C4 and C1s-like molecules in teleost fish: insights into the evolution of classical and alternative pathways." *J Immunol* **173**(1): 349-59.
- Broad\_Institute (2008). Rabbit Genomic Database 55.1h, Ensembl.
- BROAD\_S1 (2009). "Stickleback genome database 55.1j."
- Brouwer, N., K. M. Dolman, et al. (2006). "Mannan-binding lectin (MBL)-mediated opsonization is enhanced by the alternative pathway amplification loop." *Mol Immunol* **43**(13): 2051-60.
- Brown, J. R. and B. S. Hartley (1966). "Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A." *Biochem J* **101**(1): 214-28.
- Budayova-Spano, M., W. Grabarse, et al. (2002). "Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism." *Structure (Camb)* **10**(11): 1509-19.
- Budayova-Spano, M., M. Lacroix, et al. (2002). "The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex." *Embo J* **21**(3): 231-9.
- Carter, R. H. and D. T. Fearon (1992). "CD19: lowering the threshold for antigen receptor stimulation of B lymphocytes." *Science* **256**(5053): 105-7.
- Cedzynski, M., J. Szemraj, et al. (2004). "Mannan-binding lectin insufficiency in children with recurrent infections of the respiratory system." *Clin Exp Immunol* **136**(2): 304-11.
- Chen, C. B. and R. Wallis (2001). "Stoichiometry of complexes between mannose-binding protein and its associated serine proteases. Defining functional units for complement activation." *J Biol Chem* **276**(28): 25894-902.
- Chomczynski, P. and N. Sacchi (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." *Anal Biochem* **162**(1): 156-9.
- Christiansen, O. B., H. S. Nielsen, et al. (2006). "Inflammation and miscarriage." *Semin Fetal Neonatal Med.*
- Chung, E. K., Y. Yang, et al. (2002). "Determining the one, two, three, or four long and short loci of human complement C4 in a major histocompatibility complex haplotype encoding C4A or C4B proteins." *Am J Hum Genet* **71**(4): 810-22.
- Clark, J. M. (1988). "Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases." *Nucleic Acids Res* **16**(20): 9677-86.
- Coggill, P., R. D. Finn, et al. (2008). "Identifying protein domains with the Pfam database." *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2 5.
- Collard, C. D., A. Vakeva, et al. (2000). "Complement activation after oxidative stress: role of the lectin complement pathway." *Am J Pathol* **156**(5): 1549-56.
- Cooper, N. R. and D. C. Morrison (1978). "Binding and activation of the first component of human complement by the lipid A region of lipopolysaccharides." *J Immunol* **120**(6): 1862-8.

- Cortese, C. L. and W. Jiang (2006). "Mannan-binding lectin-associated serine protease 3 cleaves synthetic peptides and insulin-like growth factor-binding protein 5." Arch Biochem Biophys **449**(1-2): 164-70.
- Cseh, S., L. Vera, et al. (2002). "Characterization of the interaction between L-ficolin/p35 and mannan-binding lectin-associated serine proteases-1 and -2." J Immunol **169**(10): 5735-43.
- Cutler, A. J., M. Botto, et al. (1998). "T cell-dependent immune response in C1q-deficient mice: defective interferon gamma production by antigen-specific T cells." J Exp Med **187**(11): 1789-97.
- Cyster, J. G., K. M. Ansel, et al. (2000). "Follicular stromal cells and lymphocyte homing to follicles." Immunol Rev **176**: 181-93.
- Dahl, M. R., S. Thiel, et al. (2001). "MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway." Immunity **15**(1): 127-35.
- Danilova, N. (2006). "The evolution of immune mechanisms." J Exp Zool B Mol Dev Evol **306**(6): 496-520.
- Davies, E. J., N. Snowden, et al. (1995). "Mannose-binding protein gene polymorphism in systemic lupus erythematosus." Arthritis Rheum **38**(1): 110-4.
- Dobo, J., V. Harmat, et al. (2009). "MASP-1, a promiscuous complement protease: structure of its catalytic region reveals the basis of its broad specificity." J Immunol **183**(2): 1207-14.
- Dodds, A. W. (2002). "Which came first, the lectin/classical pathway or the alternative pathway of complement?" Immunobiology **205**(4-5): 340-54.
- Downing, I., C. Koch, et al. (2003). "Immature dendritic cells possess a sugar-sensitive receptor for human mannan-binding lectin." Immunology **109**(3): 360-4.
- Downing, I., S. L. MacDonald, et al. (2005). "Detection of an autologous ligand for mannan-binding lectin on human B lymphocytes." Scand J Immunol **62**(6): 507-14.
- Du Pasquier, L. (2005). "Meeting the demand for innate and adaptive immunities during evolution." Scand J Immunol **62 Suppl 1**: 39-48.
- Edgar, P. F. (1995). "Hucolin, a new corticosteroid-binding protein from human plasma with structural similarities to ficolins, transforming growth factor-beta 1-binding proteins." FEBS Lett **375**(1-2): 159-61.
- Eggleton, P. and D. H. Llewellyn (1999). "Pathophysiological roles of calreticulin in autoimmune disease." Scand J Immunol **49**(5): 466-73.
- Eisen, D. P., M. M. Dean, et al. (2006). "Low mannose-binding lectin function is associated with sepsis in adult patients." FEMS Immunol Med Microbiol **48**(2): 274-82.
- Endo, Y., Y. Liu, et al. (2004). "Identification of the mouse H-ficolin gene as a pseudogene and orthology between mouse ficolins A/B and human L-/M-ficolins." Genomics **84**(4): 737-44.
- Endo, Y., M. Matsushita, et al. (2007). "Role of ficolin in innate immunity and its molecular basis." Immunobiology **212**(4-5): 371-9.
- Endo, Y., M. Nonaka, et al. (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**(9): 4701-7.
- Endo, Y., Y. Sato, et al. (1996). "Cloning and characterization of the human lectin P35 gene and its related gene." Genomics **36**(3): 515-21.

- Endo, Y., M. Takahashi, et al. (2002). "Functional characterization of human mannose-binding lectin-associated serine protease (MASP)-1/3 and MASP-2 promoters, and comparison with the C1s promoter." *Int Immunol* **14**(10): 1193-201.
- Endo, Y., M. Takahashi, et al. (1998). "Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates." *J Immunol* **161**(9): 4924-30.
- Enriqueeta R. Guinto, S. C., THIERRY ROSE, KLAUS FU"TTERER, GABRIEL WAKSMAN, and A. E. D. Cera (1999). "Unexpected crucial role of residue 225 in serine proteases." *Proc Natl Acad Sci U S A* **96**(March): 1852-1857.
- Ezekowitz, R. A., M. Kuhlman, et al. (1989). "A human serum mannose-binding protein inhibits in vitro infection by the human immunodeficiency virus." *J Exp Med* **169**(1): 185-96.
- Fang, Y., C. Xu, et al. (1998). "Expression of complement receptors 1 and 2 on follicular dendritic cells is necessary for the generation of a strong antigen-specific IgG response." *J Immunol* **160**(11): 5273-9.
- Feinberg, H., J. C. Uitdehaag, et al. (2003). "Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2." *Embo J* **22**(10): 2348-59.
- Finn, R. D., J. Tate, et al. (2008). "The Pfam protein families database." *Nucleic Acids Res* **36**(Database issue): D281-8.
- Fischer, B., I. Sumner, et al. (1993). "Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies." *Biotechnol Bioeng* **41**(1): 3-13.
- Fischer, M. B., M. Ma, et al. (1996). "Regulation of the B cell response to T-dependent antigens by classical pathway complement." *J Immunol* **157**(2): 549-56.
- Fisher, M. P. and C. W. Dingman (1971). "Role of molecular conformation in determining the electrophoretic properties of polynucleotides in agarose-acrylamide composite gels." *Biochemistry* **10**(10): 1895-9.
- Flajnik, M. F. (2002). "Comparative analyses of immunoglobulin genes: surprises and portents." *Nat Rev Immunol* **2**(9): 688-98.
- Flajnik, M. F. and M. Kasahara (2001). "Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system." *Immunity* **15**(3): 351-62.
- Gabay, Y., H. Perlmann, et al. (1979). "A rosette assay for the determination of C 1 q receptor-bearing cells." *Eur J Immunol* **9**(10): 797-801.
- Gaboriaud, C., V. Rossi, et al. (2000). "Crystal structure of the catalytic domain of human complement c1s: a serine protease with a handle." *Embo J* **19**(8): 1755-65.
- Gaboriaud, C., N. M. Thielens, et al. (2004). "Structure and activation of the C1 complex of complement: unraveling the puzzle." *Trends Immunol* **25**(7): 368-373.
- Gal, P., V. Harmat, et al. (2005). "A true autoactivating enzyme. Structural insight into mannose-binding lectin-associated serine protease-2 activations." *J Biol Chem* **280**(39): 33435-44.
- Garred, P., F. Larsen, et al. (2003). "Mannose-binding lectin deficiency--revisited." *Mol Immunol* **40**(2-4): 73-84.
- Garred, P., H. O. Madsen, et al. (1997). "Susceptibility to HIV infection and progression of AIDS in relation to variant alleles of mannose-binding lectin." *Lancet* **349**(9047): 236-40.
- Garred, P., H. O. Madsen, et al. (1999). "Mannose-binding lectin polymorphisms and susceptibility to infection in systemic lupus erythematosus." *Arthritis Rheum* **42**(10): 2145-52.

- Garred, P., H. O. Madsen, et al. (1995). "Increased frequency of homozygosity of abnormal mannan-binding-protein alleles in patients with suspected immunodeficiency." Lancet **346**(8980): 941-3.
- Garred, P., T. Pressler, et al. (2002). "Mannose-binding lectin (MBL) therapy in an MBL-deficient patient with severe cystic fibrosis lung disease." Pediatr Pulmonol **33**(3): 201-7.
- Garred, P., T. Pressler, et al. (1999). "Association of mannan-binding lectin gene heterogeneity with severity of lung disease and survival in cystic fibrosis." J Clin Invest **104**(4): 431-7.
- Garred, P., C. Richter, et al. (1997). "Mannan-binding lectin in the sub-Saharan HIV and tuberculosis epidemics." Scand J Immunol **46**(2): 204-8.
- Gatto, G. J., Jr. and J. M. Berg (2003). "Nonrandom tripeptide sequence distributions at protein carboxyl termini." Genome Res **13**(4): 617-23.
- Geyer, H., C. Holschbach, et al. (1988). "Carbohydrates of human immunodeficiency virus. Structures of oligosaccharides linked to the envelope glycoprotein 120." J Biol Chem **263**(24): 11760-7.
- Ghebrehiwet, B., B. L. Lim, et al. (1994). "Isolation, cDNA cloning, and overexpression of a 33-kD cell surface glycoprotein that binds to the globular "heads" of C1q." J Exp Med **179**(6): 1809-21.
- Ghiran, I., S. F. Barbashov, et al. (2000). "Complement receptor 1/CD35 is a receptor for mannan-binding lectin." J Exp Med **192**(12): 1797-808.
- Giclas, P. C., R. N. Pinckard, et al. (1979). "In vitro activation of complement by isolated human heart subcellular membranes." J Immunol **122**(1): 146-51.
- Gjerstorff, M., S. Hansen, et al. (2004). "The genes encoding bovine SP-A, SP-D, MBL-A, conglutinin, CL-43 and CL-46 form a distinct collectin locus on *Bos taurus* chromosome 28 (BTA28) at position q.1.8-1.9." Anim Genet **35**(4): 333-7.
- Gorski, J. P., T. E. Hugli, et al. (1979). "C4a: the third anaphylatoxin of the human complement system." Proc Natl Acad Sci U S A **76**(10): 5299-302.
- Gregory, L. A., N. M. Thielen, et al. (2004). "The X-ray structure of human MBL-associated protein 19 (MAP19) and its interaction site with mannan-binding lectin and L-ficolin." J Biol Chem.
- Guan, E., S. L. Robinson, et al. (1994). "Cell-surface protein identified on phagocytic cells modulates the C1q-mediated enhancement of phagocytosis." J Immunol **152**(8): 4005-16.
- Gulati, S., K. Sastry, et al. (2002). "Regulation of the mannan-binding lectin pathway of complement on *Neisseria gonorrhoeae* by C1-inhibitor and alpha 2-macroglobulin." J Immunol **168**(8): 4078-86.
- Guo, N., T. Mogue, et al. (1998). "The human ortholog of rhesus mannan-binding protein-A gene is an expressed pseudogene that localizes to chromosome 10." Mamm Genome **9**(3): 246-9.
- Guy Dodson, A. W. (1998). "Catalytic triads and their relatives." TIBS **23**: 347-352.
- Hajela, K., M. Kojima, et al. (2002). "The biological functions of MBL-associated serine proteases (MASPs)." Immunobiology **205**(4-5): 467-75.
- Hakansson, K. and K. B. Reid (2000). "Collectin structure: a review." Protein Sci **9**(9): 1607-17.
- Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J Mol Biol **166**(4): 557-80.
- Harmat, V., P. Gal, et al. (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): 1533-46.
- Hart, M. L., M. Saifuddin, et al. (2003). "Glycosylation inhibitors and neuraminidase enhance human immunodeficiency virus type 1 binding and neutralization by mannose-binding lectin." J Gen Virol **84**(Pt 2): 353-60.
- Harumiya, S., A. Omori, et al. (1995). "EBP-37, a new elastin-binding protein in human plasma: structural similarity to ficolins, transforming growth factor-beta 1-binding proteins." J Biochem (Tokyo) **117**(5): 1029-35.
- Haurum, J. S., S. Thiel, et al. (1993). "Complement activation upon binding of mannan-binding protein to HIV envelope glycoproteins." Aids **7**(10): 1307-13.
- Heise, C. T., J. R. Nicholls, et al. (2000). "Impaired secretion of rat mannose-binding protein resulting from mutations in the collagen-like domain." J Immunol **165**(3): 1403-9.
- Helfman, D. M., R. F. Roscigno, et al. (1990). "Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA." Genes Dev **4**(1): 98-110.
- Holmskov, U., S. Thiel, et al. (2003). "Collectins and ficolins: humoral lectins of the innate immune defense." Annu Rev Immunol **21**: 547-78.
- Hoppe, H. J. and K. B. Reid (1994). "Collectins--soluble proteins containing collagenous regions and lectin domains--and their roles in innate immunity." Protein Sci **3**(8): 1143-58.
- Hourcade, D. E. (2006). "The role of properdin in the assembly of the alternative pathway C3 convertases of complement." J Biol Chem **281**(4): 2128-32.
- Hsu, E., N. Pulham, et al. (2006). "The plasticity of immunoglobulin gene systems in evolution." Immunol Rev **210**: 8-26.
- Hu, G. (1993). "DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment." DNA Cell Biol **12**(8): 763-70.
- Hummelshoj, T., L. Munthe-Fog, et al. (2005). "Polymorphisms in the FCN2 gene determine serum variation and function of Ficolin-2." Hum Mol Genet **14**(12): 1651-8.
- Ichijo, H., U. Hellman, et al. (1993). "Molecular cloning and characterization of ficolin, a multimeric protein with fibrinogen- and collagen-like domains." J Biol Chem **268**(19): 14505-13.
- Ihara, I., Y. Harada, et al. (1982). "A new complement-dependent bactericidal factor found in nonimmune mouse sera: specific binding to polysaccharide of Ra chemotype Salmonella." J Immunol **128**(3): 1256-60.
- Ii, M., M. Wada, et al. (1988). "Isolation and characterization of lectins specific for mannose/fucose/N-acetylglucosamine from rat peritoneal macrophages." J Biochem (Tokyo) **104**(4): 587-90.
- Ikedo, K., T. Sannoh, et al. (1987). "Serum lectin with known structure activates complement through the classical pathway." J Biol Chem **262**(16): 7451-4.
- Jack, D. L. (2001). "Mannose-Binding Lectin Accelerates Complement Activation and Increases Serum Killing of Neisseria meningitidis Serogroup C." The Journal of Infectious Diseases **184**: 836-45.
- Ji, X., K. Azumi, et al. (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**(12): 6340-5.

- Jiang, H., T. F. Lint, et al. (1991). "Defined chemically cross-linked oligomers of human C-reactive protein: characterization and reactivity with the complement system." *Immunology* **74**(4): 725-31.
- Jung, D. and F. W. Alt (2004). "Unraveling V(D)J recombination; insights into gene regulation." *Cell* **116**(2): 299-311.
- Kelley, J., L. Walter, et al. (2005). "Comparative genomics of major histocompatibility complexes." *Immunogenetics* **56**(10): 683-95.
- Kelly, C. R., C. D. Dickinson, et al. (1997). "Ca<sup>2+</sup> binding to the first epidermal growth factor module of coagulation factor VIIa is important for cofactor interaction and proteolytic function." *J Biol Chem* **272**(28): 17467-72.
- Kenesi, E., G. Katona, et al. (2003). "Structural and evolutionary consequences of unpaired cysteines in trypsinogen." *Biochem Biophys Res Commun* **309**(4): 749-54.
- Kenjo, A., M. Takahashi, et al. (2001). "Cloning and characterization of novel ficolins from the solitary ascidian, *Halocynthia roretzi*." *J Biol Chem* **276**(23): 19959-65.
- Kilgore, K. S., E. Schmid, et al. (1997). "Sublytic concentrations of the membrane attack complex of complement induce endothelial interleukin-8 and monocyte chemoattractant protein-1 through nuclear factor-kappa B activation." *Am J Pathol* **150**(6): 2019-31.
- Kilpatrick, D. C., B. H. Bevan, et al. (1995). "Association between mannan binding protein deficiency and recurrent miscarriage." *Hum Reprod* **10**(9): 2501-5.
- Kilpatrick, D. C., T. Fujita, et al. (1999). "P35, an opsonic lectin of the ficolin family, in human blood from neonates, normal adults, and recurrent miscarriage patients." *Immunol Lett* **67**(2): 109-12.
- Kilpatrick, D. C., L. A. McLintock, et al. (2003). "No strong relationship between mannan binding lectin or plasma ficolins and chemotherapy-related infections." *Clin Exp Immunol* **134**(2): 279-84.
- Kimura, A., E. Sakaguchi, et al. (2009). "Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*." *Immunobiology* **214**(3): 165-78.
- Kirkitadze, M. D. and P. N. Barlow (2001). "Structure and flexibility of the multiple domain proteins that regulate complement activation." *Immunol Rev* **180**: 146-61.
- Knittel, T., P. Fellmer, et al. (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**(3): 221-30.
- Kojima, M. (2002). *University of Oxford*, University of Oxford.
- Krurup, A., S. Thiel, et al. (2004). "L-ficolin Is a Pattern Recognition Molecule Specific for Acetyl Groups." *J Biol Chem* **279**(46): 47513-9.
- Krem, M. M. and E. D. Cera (2002). "Evolution of enzyme cascades from embryonic development to blood coagulation." *Trends Biochem Sci* **27**(2): 67-74.
- Krem, M. M. and E. Di Cera (2001). "Molecular markers of serine protease evolution." *Embo J* **20**(12): 3036-45.
- Krem, M. M., S. Prasad, et al. (2002). "Ser(214) is crucial for substrate binding to serine proteases." *J Biol Chem* **277**(43): 40260-4.
- Kruse, C., A. Rosgaard, et al. (2002). "Low serum level of mannan-binding lectin is a determinant for pregnancy outcome in women with recurrent spontaneous abortion." *Am J Obstet Gynecol* **187**(5): 1313-20.
- Kuhlman, M., K. Joiner, et al. (1989). "The human mannose-binding protein functions as an opsonin." *J Exp Med* **169**(5): 1733-45.

- Lacroix, M., C. Dumestre-Perard, et al. (2009). "Residue Lys57 in the collagen-like region of human L-ficolin and its counterpart Lys47 in H-ficolin play a key role in the interaction with the mannan-binding lectin-associated serine proteases and the collectin receptor calreticulin." *J Immunol* **182**(1): 456-65.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-5.
- Lambiase, A., V. Raia, et al. (2006). "Microbiology of airway disease in a cohort of patients with cystic fibrosis." *BMC Infect Dis* **6**: 4.
- Lanzrein, A. S., K. A. Jobst, et al. (1998). "Mannan-binding lectin in human serum, cerebrospinal fluid and brain tissue and its role in Alzheimer's disease." *Neuroreport* **9**(7): 1491-5.
- Lau, Y. L., C. S. Lau, et al. (1996). "Mannose-binding protein in Chinese patients with systemic lupus erythematosus." *Arthritis Rheum* **39**(4): 706-8.
- Laursen, S. B., J. E. Hedemand, et al. (1995). "Collectin in a non-mammalian species: isolation and characterization of mannan-binding protein (MBP) from chicken serum." *Glycobiology* **5**(6): 553-61.
- Law, S. K. (1983). "Non-enzymic activation of the covalent binding reaction of the complement protein C3." *Biochem J* **211**(2): 381-9.
- Le, Y., S. H. Lee, et al. (1998). "Human L-ficolin: plasma levels, sugar specificity, and assignment of its lectin activity to the fibrinogen-like (FBG) domain." *FEBS Lett* **425**(2): 367-70.
- Liu, Y., Y. Endo, et al. (2005). "Ficolin A and ficolin B are expressed in distinct ontogenic patterns and cell types in the mouse." *Mol Immunol* **42**(11): 1265-73.
- Liu, Y., Y. Endo, et al. (2005). "Human M-ficolin is a secretory protein that activates the lectin complement pathway." *J Immunol* **175**(5): 3150-6.
- Loos, M., F. Clas, et al. (1986). "Interaction of purified lipoteichoic acid with the classical complement pathway." *Infect Immun* **53**(3): 595-9.
- Lu, J., P. N. Tay, et al. (1996). "Human ficolin: cDNA cloning, demonstration of peripheral blood leucocytes as the major site of synthesis and assignment of the gene to chromosome 9." *Biochem J* **313**(Pt 2): 473-8.
- Lynch, N., S. Roscher, et al. (2003). "L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of gram positive bacteria, and activates the lectin pathway of complement."
- Lynch, N. J., S. U. Khan, et al. (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**(8): 4998-5006.
- Lynch, N. J., S. Roscher, et al. (2004). "L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of Gram-positive bacteria, and activates the lectin pathway of complement." *J Immunol* **172**(2): 1198-202.
- Maas, J., A. M. de Roda Husman, et al. (1998). "Presence of the variant mannan-binding lectin alleles associated with slower progression to AIDS. Amsterdam Cohort Study." *Aids* **12**(17): 2275-80.
- Madsen, H. O., P. Garred, et al. (1995). "Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein." *J Immunol* **155**(6): 3013-20.
- Madsen, H. O., M. L. Satz, et al. (1998). "Different molecular events result in low protein levels of mannan-binding lectin in populations from southeast Africa and South America." *J Immunol* **161**(6): 3169-75.
- Madsen, H. O., V. Videm, et al. (1998). "Association of mannan-binding-lectin deficiency with severe atherosclerosis." *Lancet* **352**(9132): 959-60.

- Malik, S., M. Arias, et al. (2003). "Absence of association between mannan-binding lectin gene polymorphisms and HIV-1 infection in a Colombian population." *Immunogenetics* **55**(1): 49-52.
- Manderson A, M. C. P., Marina Botto, Mark J. Walport, Christopher R. Parish (2001). "Continual Low-Level Activation of the Classical Complement Pathway." *J Exp Med* **194**(6): 747-756.
- Matsumoto, A. K., J. Kopicky-Burd, et al. (1991). "Intersection of the complement and immune systems: a signal transduction complex of the B lymphocyte-containing complement receptor type 2 and CD19." *J Exp Med* **173**(1): 55-64.
- Matsushita, M., Y. Endo, et al. (2000). "Cutting edge: complement-activating complex of ficolin and mannan-binding lectin-associated serine protease." *J Immunol* **164**(5): 2281-4.
- Matsushita, M., Y. Endo, et al. (2001). "Activation of the lectin complement pathway by ficolins." *Int Immunopharmacol* **1**(3): 359-63.
- Matsushita M, E. Y., Taira S, Sato Y, Fujita T, Ichikawa N (1996). "A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin." *J Biol Chem* **271**: 2448-54.
- Matsushita, M. and T. Fujita (1992). "Activation of the classical complement pathway by mannan-binding protein in association with a novel C1s-like serine protease." *J Exp Med* **176**(6): 1497-502.
- Matsushita, M. and T. Fujita (2001). "Ficolins and the lectin complement pathway." *Immunol Rev* **180**: 78-85.
- Matsushita, M. and N. Hamasaki (2000). "[Structure and function of the Ficolin family]." *Tanpakushitsu Kakusan Koso* **45**(5): 664-70.
- Matsushita, M., M. Kuraya, et al. (2002). "Activation of the lectin complement pathway by H-ficolin (Hakata antigen)." *J Immunol* **168**(7): 3502-6.
- Matsushita, M., S. Thiel, et al. (2000). "Proteolytic activities of two types of mannan-binding lectin-associated serine protease." *J Immunol* **165**(5): 2637-42.
- McGreal, E. and P. Gasque (2002). "Structure-function studies of the receptors for complement C1q." *Biochem Soc Trans* **30**(Pt 6): 1010-4.
- Minchinton, R. M., M. M. Dean, et al. (2002). "Analysis of the relationship between mannan-binding lectin (MBL) genotype, MBL levels and function in an Australian blood donor population." *Scand J Immunol* **56**(6): 630-41.
- Mizuochi, T., T. J. Matthews, et al. (1990). "Diversity of oligosaccharide structures on the envelope glycoprotein gp 120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting N-acetylglucosamine residues." *J Biol Chem* **265**(15): 8519-24.
- Mogues, T., T. Ota, et al. (1996). "Characterization of two mannan-binding protein cDNAs from rhesus monkey (*Macaca mulatta*): structure and evolutionary implications." *Glycobiology* **6**(5): 543-50.
- Moller-Kristensen, M., J. C. Jensenius, et al. (2003). "Levels of mannan-binding lectin-associated serine protease-2 in healthy individuals." *J Immunol Methods* **282**(1-2): 159-67.
- Morgan, B. P. (1989). "Complement membrane attack on nucleated cells: resistance, recovery and non-lethal effects." *Biochem J* **264**(1): 1-14.
- Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." *Cold Spring Harb Symp Quant Biol* **51 Pt 1**: 263-73.
- Nadesalingam, J., A. W. Dodds, et al. (2005). "Mannan-binding lectin recognizes peptidoglycan via the N-acetyl glucosamine moiety, and inhibits ligand-induced

- proinflammatory effect and promotes chemokine production by macrophages." *J Immunol* **175**(3): 1785-94.
- Nagai, T., J. Mutsuro, et al. (2000). "A novel truncated isoform of the mannose-binding lectin-associated serine protease (MASP) from the common carp (*Cyprinus carpio*)." *Immunogenetics* **51**(3): 193-200.
- Naito, H., A. Ikeda, et al. (1999). "Characterization of human serum mannan-binding protein promoter." *J Biochem (Tokyo)* **126**(6): 1004-12.
- NCBI (2007). Mouse Genome, Build 37.  
[http://www.ncbi.nlm.nih.gov/mapview/map\\_search.cgi?taxid=10090](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=10090).
- Nei, M., P. Xu, et al. (2001). "Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms." *Proc Natl Acad Sci U S A* **98**(5): 2497-502.
- Neth, O., I. Hann, et al. (2001). "Deficiency of mannose-binding lectin and burden of infection in children with malignancy: a prospective study." *Lancet* **358**(9282): 614-8.
- Nonaka, M. (2001). "Evolution of initiating enzymes of the complement system." *Genomebiology* **3**(1): 1001.1 1001.5.
- Nonaka, M. and A. Kimura (2006). "Genomic view of the evolution of the complement system." *Immunogenetics* **58**(9): 701-13.
- Nonaka, M. and S. Miyazawa (2002). "Evolution of the initiating enzymes of the complement system." *Genome Biol* **3**(1): REVIEWS1001.
- Nonaka, M. and F. Yoshizaki (2004). "Evolution of the complement system." *Mol Immunol* **40**(12): 897-902.
- Oettinger, M. A., D. G. Schatz, et al. (1990). "RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination." *Science* **248**(4962): 1517-23.
- Ogden, C. A., A. deCathelineau, et al. (2001). "C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells." *J Exp Med* **194**(6): 781-95.
- Ohashi, T. and H. P. Erickson (1998). "Oligomeric structure and tissue distribution of ficolins from mouse, pig and human." *Arch Biochem Biophys* **360**(2): 223-32.
- Pancer, Z. (2000). "Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin." *Proc Natl Acad Sci U S A* **97**(24): 13156-61.
- Pangburn, M. K., R. D. Schreiber, et al. (1981). "Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3." *J Exp Med* **154**(3): 856-67.
- Papamichail, M., C. Gutierrez, et al. (1975). "Complement dependence of localisation of aggregated IgG in germinal centres." *Scand J Immunol* **4**(4): 343-47.
- Pastinen, T., K. Liitsola, et al. (1998). "Contribution of the CCR5 and MBL genes to susceptibility to HIV type 1 infection in the Finnish population." *AIDS Res Hum Retroviruses* **14**(8): 695-8.
- Paul, W. E. (2003). *Fundamental immunology*. Philadelphia, Pa. ; London, Lippincott Williams & Wilkins.
- Pepys, M. B. (1976). "Role of complement in the induction of immunological responses." *Transplant Rev* **32**: 93-120.
- Petersen, S. V., S. Thiel, et al. (2000). "Control of the classical and the MBL pathway of complement activation." *Mol Immunol* **37**(14): 803-11.

- Peterslund, N. A., C. Koch, et al. (2001). "Association between deficiency of mannose-binding lectin and severe infections after chemotherapy." Lancet **358**(9282): 637-8.
- Phillips, A. E., J. Toth, et al. (2009). "Analogous interactions in initiating complexes of the classical and lectin pathways of complement." J Immunol **182**(12): 7708-17.
- Porath, J., J. Carlsson, et al. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature **258**(5536): 598-9.
- Reid, K. B. and R. R. Porter (1976). "Subunit composition and structure of subcomponent C1q of the first component of human complement." Biochem J **155**(1): 19-23.
- Roos, A., P. Garred, et al. (2004). "Antibody-mediated activation of the classical pathway of complement may compensate for mannose-binding lectin deficiency." Eur J Immunol **34**(9): 2589-98.
- Rossi, V., S. Cseh, et al. (2001). "Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and-2." J Biol Chem **276**: 29.
- Rossi, V., F. Teillet, et al. (2005). "Functional Characterization of Complement Proteases C1s/Mannan-binding Lectin-associated Serine Protease-2 (MASP-2) Chimeras Reveals the Higher C4 Recognition Efficacy of the MASP-2 Complement Control Protein Modules." J Biol Chem **280**(51): 41811-8.
- Ruiz, S., A. H. Henschen-Edman, et al. (1999). "Digestion of C1q collagen-like domain with MMPs-1,-2,-3, and -9 further defines the sequence involved in the stimulation of neutrophil superoxide production." J Leukoc Biol **66**(3): 416-22.
- Rupert C. Wilmouth, K. E., Richard Neutze (2001). "X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate." nature structural biology **8**(689-694).
- Rus, H. G., F. Niculescu, et al. (1996). "Sublytic complement attack induces cell cycle in oligodendrocytes." J Immunol **156**(12): 4892-900.
- Saifuddin, M., M. L. Hart, et al. (2000). "Interaction of mannose-binding lectin with primary isolates of human immunodeficiency virus type 1." J Gen Virol **81 Pt 4**: 949-55.
- Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol Biol Evol **4**(4): 406-25.
- Salamov, A. A. and V. V. Solovyev (1997). "Recognition of 3'-processing sites of human mRNA precursors." Comput Appl Biosci **13**(1): 23-8.
- Sambrook, J., E. F. Fritsch, et al. (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.
- Sastry, K., G. A. Herman, et al. (1989). "The human mannose-binding protein gene. Exon structure reveals its evolutionary relationship to a human pulmonary surfactant gene and localization to chromosome 10." J Exp Med **170**(4): 1175-89.
- Sato, T., Y. Endo, et al. (1994). "Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein." Int Immunol **6**(4): 665-9.
- Schafranski, M. D., A. Stier, et al. (2004). "Significantly increased levels of mannose-binding lectin (MBL) in rheumatic heart disease: a beneficial role for MBL deficiency." Clin Exp Immunol **138**(3): 521-5.
- Schwaeble, W., M. R. Dahl, et al. (2002). "The mannan-binding lectin-associated serine proteases (MASPs) and MASP19: four components of the lectin pathway activation complex encoded by two genes." Immunobiology **205**(4-5): 455-66.

- Senior, R. M., W. F. Skogen, et al. (1986). "Effects of fibrinogen derivatives upon the inflammatory response. Studies with human fibrinopeptide B." J Clin Invest **77**(3): 1014-9.
- Seyfarth, J., P. Garred, et al. (2006). "Extra-hepatic transcription of the human mannose-binding lectin gene (mbl2) and the MBL-associated serine protease 1-3 genes." Mol Immunol **43**(7): 962-71.
- Sheriff, S., C. Y. Chang, et al. (1994). "Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil." Nat Struct Biol **1**(11): 789-94.
- Shum, B. P., L. Guethlein, et al. (2001). "Modes of salmonid MHC class I and II evolution differ from the primate paradigm." J Immunol **166**(5): 3297-308.
- Skjoedt, M. O., Y. Palarasah, et al. (2009). "MBL-associated serine protease-3 circulates in high serum concentrations predominantly in complex with Ficolin-3 and regulates Ficolin-3 mediated complement activation." Immunobiology.
- Skogen, W. F., R. M. Senior, et al. (1988). "Fibrinogen-derived peptide B beta 1-42 is a multidomained neutrophil chemoattractant." Blood **71**(5): 1475-9.
- Sonnhammer, E. L., S. R. Eddy, et al. (1997). "Pfam: a comprehensive database of protein domain families based on seed alignments." Proteins **28**(3): 405-20.
- Steinberger, P., A. Szekeres, et al. (2002). "Identification of human CD93 as the phagocytic C1q receptor (C1qRp) by expression cloning." J Leukoc Biol **71**(1): 133-40.
- Stengaard-Pedersen, K., S. Thiel, et al. (2003). "Inherited deficiency of mannan-binding lectin-associated serine protease 2." N Engl J Med **349**(6): 554-60.
- Stover, C. M., N. J. Lynch, et al. (2003). "Murine serine proteases MASP-1 and MASP-3, components of the lectin pathway activation complex of complement, are encoded by a single structural gene." Genes Immun **4**(5): 374-84.
- Stover, C. M., N. J. Lynch, et al. (2004). "Organization of the MASP2 locus and its expression profile in mouse and rat." Mamm Genome **15**(11): 887-900.
- Stover, C. M., W. J. Schwaeble, et al. (1999). "Assignment of the gene encoding mannan-binding lectin-associated serine protease 2 (MASP2) to human chromosome 1p36.3-->p36.2 by in situ hybridization and somatic cell hybrid analysis." Cytogenet Cell Genet **84**(3-4): 148-9.
- Stover, C. M., S. Thiel, et al. (1999). "The rat and mouse homologues of MASP-2 and MASP-3, components of the lectin activation pathway of complement." J Immunol **163**(12): 6848-59.
- Stover, C. M., S. Thiel, et al. (1999). "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**(6): 3481-90.
- Strausberg, R. L., E. A. Feingold, et al. (2002). "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." Proc Natl Acad Sci U S A **99**(26): 16899-903.
- Sugimoto, R., Y. Yae, et al. (1998). "Cloning and characterization of the Hakata antigen, a member of the ficolin/opsonin p35 lectin family." J Biol Chem **273**(33): 20721-7.
- Summerfield, J. A. and M. E. Taylor (1986). "Mannose-binding proteins in human serum: identification of mannose-specific immunoglobulins and a calcium-dependent lectin, of broader carbohydrate specificity, secreted by hepatocytes." Biochim Biophys Acta **883**(2): 197-206.

- Szabo, E., Z. Bocskei, et al. (1999). "The three-dimensional structure of Asp189Ser trypsin provides evidence for an inherent structural plasticity of the protease." Eur J Biochem **263**(1): 20-6.
- Takada, F., N. Seki, et al. (1995). "Localization of the genes for the 100-kDa complement-activating components of Ra-reactive factor (CRARF and Crarf) to human 3q27-q28 and mouse 16B2-B3." Genomics **25**(3): 757-9.
- Takada, F., Y. Takayama, et al. (1993). "A new member of the C1s family of complement proteins found in a bactericidal factor, Ra-reactive factor, in human serum." Biochem Biophys Res Commun **196**(2): 1003-9.
- Takahashi, M., Y. Endo, et al. (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**(5): 859-63.
- Takahashi, M., Y. Ishida, et al. (2010). "Essential role of mannose-binding lectin-associated serine protease-1 in activation of the complement factor D." J Exp Med **207**(1): 29-37.
- Takahashi, M., Y. Ishida, et al. (2010). "Essential role of mannose-binding lectin-associated serine protease-1 in activation of the complement factor D." J Exp Med **207**(1): 29-37, S1-3.
- Takahashi, M., D. Iwaki, et al. (2006). "Cloning and characterization of mannose-binding lectin from lamprey (Agnathans)." J Immunol **176**(8): 4861-8.
- Takahashi, M., Miura, S., Ishii, N., Sugamura, K., Shuichi, M., and S. Shiro, Endo, Y., Matsushita, M. and Fujita, T. (2000). "An essential role of MBL-associated serine protease-1 and -3 in activation of complement by lectin pathway." Unpublished.
- Takayama, Y., F. Takada, et al. (1994). "A 100-kDa protein in the C4-activating component of Ra-reactive factor is a new serine protease having module organization similar to C1r and C1s." J Immunol **152**(5): 2308-16.
- Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0." Mol Biol Evol **24**(8): 1596-9.
- Tauber, J. W., M. J. Polley, et al. (1976). "Nonspecific complement activation by streptococcal structures. II. Properdin-independent initiation of the alternate pathway." J Exp Med **143**(6): 1352-66.
- Taylor, J. S., Y. Van de Peer, et al. (2001). "Comparative genomics provides evidence for an ancient genome duplication event in fish." Philos Trans R Soc Lond B Biol Sci **356**(1414): 1661-79.
- Taylor, M. E., P. M. Brickell, et al. (1989). "Structure and evolutionary origin of the gene encoding a human serum mannose-binding protein." Biochem J **262**(3): 763-71.
- Tedesco, F., M. Pausa, et al. (1997). "The cytolytically inactive terminal complement complex activates endothelial cells to express adhesion molecules and tissue factor procoagulant activity." J Exp Med **185**(9): 1619-27.
- Teh, C., Y. Le, et al. (2000). "M-ficolin is expressed on monocytes and is a lectin binding to N-acetyl- D-glucosamine and mediates monocyte adhesion and phagocytosis of Escherichia coli." Immunology **101**(2): 225-32.
- Teillet, F., C. Gaboriaud, et al. (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): 25715-24.
- Teillet, F., M. Lacroix, et al. (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**(9): 5710-6.

- Teisberg, P., R. Jonassen, et al. (1988). "Restriction fragment length polymorphisms of the complement component C4 loci on chromosome 6: studies with emphasis on the determination of gene number." *Ann Hum Genet* **52**(Pt 2): 77-84.
- Tenner, A. J. (1989). "C1q interactions with cell surface receptors." *Behring Inst Mitt*(84): 220-9.
- Tenner, A. J. (1998). "C1q receptors: regulating specific functions of phagocytic cells." *Immunobiology* **199**(2): 250-64.
- Tenner, A. J., S. L. Robinson, et al. (1995). "Mannose binding protein (MBP) enhances mononuclear phagocyte function via a receptor that contains the 126,000 M(r) component of the C1q receptor." *Immunity* **3**(4): 485-93.
- Terai, I., K. Kobayashi, et al. (1997). "Human serum mannan-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**(2): 317-23.
- Thiel, S., S. V. Petersen, et al. (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**(2): 878-87.
- Thiel, S., T. Vorup-Jensen, et al. (1997). "A second serine protease associated with mannan-binding lectin that activates complement." *Nature* **386**(6624): 506-10.
- Thielens, N. M., K. Enrie, et al. (1999). "The N-terminal CUB-epidermal growth factor module pair of human complement protease C1r binds Ca<sup>2+</sup> with high affinity and mediates Ca<sup>2+</sup>-dependent interaction with C1s." *J Biol Chem* **274**(14): 9149-59.
- Thielens, N. M., P. Tacnet-Delorme, et al. (2002). "Interaction of C1q and mannan-binding lectin with viruses." *Immunobiology* **205**(4-5): 563-74.
- Thompson, J. D., T. J. Gibson, et al. (2002). "Multiple sequence alignment using ClustalW and ClustalX." *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2 3.
- Tsujimura, M., C. Ishida, et al. (2001). "Detection of serum thermolabile beta-2 macroglycoprotein (Hakata antigen) by enzyme-linked immunosorbent assay using polysaccharide produced by *Aerococcus viridans*." *Clin Diagn Lab Immunol* **8**(2): 454-9.
- Turner, M. W. (2003). "The role of mannan-binding lectin in health and disease." *Mol Immunol* **40**(7): 423-9.
- Turner, M. W. and R. M. Hamvas (2000). "Mannose-binding lectin: structure, function, genetics and disease associations." *Rev Immunogenet* **2**(3): 305-22.
- Uemura, K., H. Yamamoto, et al. (2004). "Superoxide production from human polymorphonuclear leukocytes by human mannan-binding protein (MBP)." *Glycoconj J* **21**(1-2): 79-84.
- van Emmerik, L. C., E. J. Kuijper, et al. (1994). "Binding of mannan-binding protein to various bacterial pathogens of meningitis." *Clin Exp Immunol* **97**(3): 411-6.
- Vincent, J. L., R. Moreno, et al. (1996). "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine." *Intensive Care Med* **22**(7): 707-10.
- Vorup-Jensen, T., J. C. Jensenius, et al. (1998). "MASP-2, the C3 convertase generating protease of the MBLectin complement activating pathway." *Immunobiology* **199**(2): 348-57.
- Vorup-Jensen, T., S. V. Petersen, et al. (2000). "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**(4): 2093-100.

- Wagner, S., N. J. Lynch, et al. (2003). "Differential expression of the murine mannose-binding lectins A and C in lymphoid and nonlymphoid organs and tissues." J Immunol **170**(3): 1462-5.
- Wallis, R. and R. B. Dodd (2000). "Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations." J Biol Chem **275**(40): 30962-9.
- Wallis, R., N. J. Lynch, et al. (2005). "Decoupling of carbohydrate binding and MASP-2 autoactivation in variant mannose-binding lectins associated with immunodeficiency." J Immunol **175**(10): 6846-51.
- White, R. A., L. L. Dowler, et al. (1994). "The murine mannose-binding protein genes (Mbl 1 and Mbl 2) localize to chromosomes 14 and 19." Mamm Genome **5**(12): 807-9.
- Wilkinson, B. J., Y. Kim, et al. (1981). "Factors affecting complement activation by Staphylococcus aureus cell walls, their components, and mutants altered in teichoic acid." Infect Immun **32**(1): 216-24.
- Yang, I. A., S. L. Seeney, et al. (2003). "Mannose-binding lectin gene polymorphism predicts hospital admissions for COPD infections." Genes Immun **4**(4): 269-74.
- Ytting, H., I. J. Christensen, et al. (2005). "Preoperative mannan-binding lectin pathway and prognosis in colorectal cancer." Cancer Immunol Immunother **54**(3): 265-72.
- Zharkikh, A. and W. H. Li (1995). "Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique." Mol Phylogenet Evol **4**(1): 44-63.
- Zundel, S., S. Cseh, et al. (2004). "Characterization of recombinant mannan-binding lectin-associated serine protease (MASP)-3 suggests an activation mechanism different from that of MASP-1 and MASP-2." J Immunol **172**(7): 4342-50.

# Publication

## Chapter 8

The published articles [p. 268-278] are not available in the electronic version of this thesis due to third party copyright restrictions.

The full version can be consulted at the University of Leicester Library.